

Thèse

Le document numérique :  
la complexité des formes  
et  
les formes de la complexité

*Systèmes et interfaces utilisant des descripteurs sémiotiques  
de présentation de l'information*

présentée devant  
**L'Institut National des Sciences Appliquées de Lyon**

Pour obtenir  
le grade de docteur

Ecole doctorale :  
**Informatique et Mathématiques**

Spécialité :  
Documents numériques, Images et Systèmes d'Information Communicants

par  
Joël GARDES

Jury

---

Rapporteur	R. INGOLD	Professeur, Université de Fribourg
Rapporteur	B. BACHIMONT	Professeur, UTC
Directeur de Thèse	H. EMPTOZ	Professeur, INSA Lyon
Examineur	C. JUTTEN	Professeur, Université Joseph Fourier, Grenoble
Examineur	G. NAGY	Professeur, ESCE Troy University
Examineur	C. GARCIA	Ingénieur R&D, Orange Labs
Examineur	D. CHENE	Ingénieur R&D, Orange Labs

---

Laboratoire :

Laboratoire d'InfoRmatique en Image et Système d'information (LIRIS)





*A ma Mère, partie de ce monde mais toujours présente dans mon cœur.*

*A mon Père, toujours prêt à dire une "formule magique" pour aider son petit.*

*A mes Amis de si longue date qu'ils font partie de la famille.*

## *Remerciements*

Préparer une thèse de Doctorat quand on a passé le cap du demi-siècle n'est pas un évènement banal. Il n'y a pas que le sujet de travail captivant et en totale adéquation avec mes convictions. Il y a un "flashback" sur une longue période de ma vie ravivant dans ma mémoire, toute la succession d'évènements qui m'ont permis d'arriver à ce stade important.

Nombreuses sont les personnes que je me dois de remercier de m'avoir fait vivre cette expérience forte, aussi, je ne pourrai pas toutes les citer.

En tout premier lieu, il y a trois institutions que je remercie :

Mon employeur, Orange Labs, qui, en organisant en 2006, une campagne de candidatures à des thèses pour ses ingénieurs m'a permis d'effectuer ce travail de valorisation de mon expérience professionnelle. M. Gérard Eude a été, en particulier, la véritable cheville ouvrière qui a réussi à me convaincre de me lancer dans l'aventure.

L'Ecole Doctorale "Informatique et mathématiques" de Lyon, et en particulier, son Directeur, M. le Professeur Alain Mille, qui, malgré mon cursus assez particulier, a autorisé cette thèse.

L'INSA de Lyon et le LIRIS, qui m'ont accueilli durant ces presque trois années en m'offrant des conditions plus que favorables.

Sans ces trois institutions, cette thèse n'aurait jamais pu voir le jour.

A ces remerciements viennent s'ajouter ceux que j'adresse chaleureusement aux membres de mon jury :

MM. Les Professeurs Bruno Bachimont, Rolf Ingold, Christian Jutten, George Nagy, MM. Christophe Garcia, Denis Chêne ont vraiment pris le temps d'analyser en profondeur mon travail lié à un sujet que je sais complexe et quelque peu à contre-courant pour certains aspects.

Malgré une rédaction qui, j'en ai conscience, peut donner l'impression d'avoir défriché un terrain qui reste à cartographier, leur jugement et leurs conseils ont toujours été très constructifs et représentent autant d'idées sur la façon d'explorer ce terrain de recherche que je pense être très fertile.

M. Le Professeur Hubert Emptoz, mon Directeur de thèse dont la bienveillance, la patience et la disponibilité ont été sans faille. Nos discussions scientifiques furent intenses et captivantes ; chez lui, l'humain n'est pas seulement au centre d'une boucle de traitements informatiques. Elles m'ont guidé dans ce travail tout en me laissant une très grande liberté dans les orientations choisies ; aucun dogmatisme n'y a trouvé sa place. Tout ceci fait, qu'au fil du temps, j'ai vu naître et croître une complicité et une amitié qui m'honorent.

Je ne saurais oublier mes collègues que désormais j'appelle mes amis : Yann Leydier, Jean Duong, Christophe Maldivi, Eric Petit, Sébastien Roux, et tous ceux que je n'ai pas cités ici mais qui se retrouveront. Ils ont su toujours être présents; quitte à supporter presque quotidiennement un sacré personnage, avec ses idées décalées, ses états d'âmes, pour ne pas dire ses caprices.

*INSA Direction de la Recherche - Ecoles Doctorales – Quadriennal 2007-2010*

<b>SIGLE</b>	<b>ECOLE DOCTORALE</b>	<b>NOM ET COORDONNEES DU RESPONSABLE</b>
<b>CHIMIE</b>	<b>CHIMIE DE LYON</b> <a href="http://sakura.cpe.fr/ED206">http://sakura.cpe.fr/ED206</a>  M. Jean Marc LANCELIN  Insa : R. GOURDON	M. Jean Marc LANCELIN Université Claude Bernard Lyon 1 Bât CPE 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.43 13 95 Fax : <a href="mailto:lancelin@hikari.cpe.fr">lancelin@hikari.cpe.fr</a>
<b>E.E.A.</b>	<b>ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE</b> <a href="http://www.insa-lyon.fr/eea">http://www.insa-lyon.fr/eea</a> M. Alain NICOLAS Insa : C. PLOSSU <a href="mailto:ede2a@insa-lyon.fr">ede2a@insa-lyon.fr</a> Secrétariat : M. LABOUNE AM. 64.43 – Fax : 64.54	M. Alain NICOLAS Ecole Centrale de Lyon Bâtiment H9 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60 97 Fax : 04 78 43 37 17 <a href="mailto:eea@ec-lyon.fr">eea@ec-lyon.fr</a> Secrétariat : M.C. HAVGOUDOUKIAN
<b>E2M2</b>	<b>EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION</b> <a href="http://biomserv.univ-lyon1.fr/E2M2">http://biomserv.univ-lyon1.fr/E2M2</a>  M. Jean-Pierre FLANDROIS Insa : H. CHARLES	M. Jean-Pierre FLANDROIS CNRS UMR 5558 Université Claude Bernard Lyon 1 Bât G. Mendel 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cédex Tél : 04.26 23 59 50 Fax 04 26 23 59 49 06 07 53 89 13 <a href="mailto:e2m2@biomserv.univ-lyon1.fr">e2m2@biomserv.univ-lyon1.fr</a>
<b>EDISS</b>	<b>INTERDISCIPLINAIRE SCIENCES- SANTÉ</b>  Sec : Safia Boudjema M. Didier REVEL Insa : M. LAGARDE	M. Didier REVEL Hôpital Cardiologique de Lyon Bâtiment Central 28 Avenue Doyen Lépine 69500 BRON Tél : 04.72.68 49 09 Fax :04 72 35 49 16 <a href="mailto:Didier.revel@creatis.uni-lyon1.fr">Didier.revel@creatis.uni-lyon1.fr</a>
<b>INFOMATHS</b>	<b>INFORMATIQUE ET MATHÉMATIQUES</b> <a href="http://infomaths.univ-lyon1.fr">http://infomaths.univ-lyon1.fr</a> M. Alain MILLE  Secrétariat : C. DAYEYAN	M. Alain MILLE Université Claude Bernard Lyon 1 LIRIS - INFOMATHS Bâtiment Nautibus 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72. 44 82 94 Fax 04 72 43 13 10 <a href="mailto:infomaths@bat710.univ-lyon1.fr">infomaths@bat710.univ-lyon1.fr</a> - <a href="mailto:alain.mille@liris.cnrs.fr">alain.mille@liris.cnrs.fr</a>
<b>Matériaux</b>	<b>MATERIAUX DE LYON</b>  M. Jean Marc PELLETIER  Secrétariat : C. BERNAVON 83.85	M. Jean Marc PELLETIER INSA de Lyon MATEIS Bâtiment Blaise Pascal 7 avenue Jean Capelle 69621 VILLEURBANNE Cédex Tél : 04.72.43 83 18 Fax 04 72 43 85 28 <a href="mailto:Jean-marc.Pelletier@insa-lyon.fr">Jean-marc.Pelletier@insa-lyon.fr</a>
<b>MEGA</b>	<b>MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE</b>  M. Jean Louis GUYADER  Secrétariat : M. LABOUNE PM : 71.70 –Fax : 87.12	M. Jean Louis GUYADER INSA de Lyon Laboratoire de Vibrations et Acoustique Bâtiment Antoine de Saint Exupéry 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél :04.72.18.71.70 Fax : 04 72 43 72 37 <a href="mailto:mega@lva.insa-lyon.fr">mega@lva.insa-lyon.fr</a>
<b>ScSo</b>	<b>ScSo*</b>  M. OBADIA Lionel  Insa : J.Y. TOUSSAINT	M. OBADIA Lionel Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.69.72.76 Fax : 04.37.28.04.48 <a href="mailto:Lionel.Obadia@univ-lyon2.fr">Lionel.Obadia@univ-lyon2.fr</a>

\*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie



# Sommaire

<b>AVANT PROPOS</b>	<b>13</b>
<b>Idées directrices et organisation du document</b>	<b>21</b>
<b>CHAPITRE 1</b>	
<b>LES MUTATIONS DE L'ECOSYSTEME DE L'INFORMATION NUMERIQUE</b>	<b>25</b>
<b>I. D'une information diffusée à une information pervasive</b>	<b>26</b>
I.1. - Trois ères des systèmes d'information	26
I.2. - De l'interconnexion physique à l'interconnexion logique des systèmes d'information	27
I.3. - D'une médiation matérielle à une médiation logicielle dans les interfaces	28
<b>II. L'émergence de la complexité dans la relation homme machine</b>	<b>29</b>
II.1. - Ambiguïté et décision	30
II.2. - Des formes de la complexité, de la complexité des formes	31
<b>III. Physique et information, deux évolutions parallèles</b>	<b>31</b>
III.1. - Un modèle réductionniste de l'information aux limites	31
III.2. - Trois expériences de pensées	34
III.3. - Point sur les théories de l'information et de la communication	40
<b>IV. Au croisement du multimédia et de la multimodalité, la présentation de l'information</b>	<b>45</b>
IV.1. - Document numérique et interaction	45
IV.2. - Vers un modèle adaptatif de l'interaction	50
IV.3. - Des maquettes révélant des potentialités techniques	52
<b>CHAPITRE 2</b>	
<b>DE LA NUMERISATION... A LA PRESENTATION DE L'INFORMATION</b>	<b>55</b>
<b>I. Numériser des documents patrimoniaux et personnels</b>	<b>56</b>
I.1. - Les limites de la numérisation industrielle	57
<b>II. Un témoignage de 10 ans sur la numérisation</b>	<b>58</b>
<b>III. Le glissement des domaines de la médiation</b>	<b>60</b>
III.1. - Les ruptures induites par l'information numérique banalisée	60
III.2. - Le renversement des privilèges des "consomm'acteurs" de documents	62
III.3. - Conséquences sur la présentation de l'information	63
III.4. - Un darwinisme documentaire	64
<b>IV. La numérisation industrielle de masse</b>	<b>65</b>
IV.1. - La numérisation industrielle de bibliothèques	67
IV.2. - Des hypothèses d'évolutions pour la numérisation	68
<b>V. La numérisation de plans de réseaux</b>	<b>69</b>
<b>VI. Le traitement de documents hétérogènes</b>	<b>72</b>
VI.1. - Le projet DocMining	72
VI.2. - Les concepts du projet	73
VI.3. - Le profil d'une application basée sur DocMining	74
<b>VII. Les trois niveaux de présentation de l'information dans le document numérique</b>	<b>75</b>
VII.1. - Document numérique et accessibilité	75
VII.2. - Le cas de la numérisation	76
<b>VIII. L'arbitraire de l'interprétation et le déterminisme technique</b>	<b>76</b>
<b>IX. Interaction et prise de décision</b>	<b>77</b>
<b>X. Sémiotique et information</b>	<b>78</b>



## CHAPITRE 3

### APPROCHE SYSTEMIQUE DE LA PRESENTATION DE L'INFORMATION \_\_\_\_\_ 81

<b>I. L'interface homme machine</b>	<b>82</b>
<b>II. Segmentation et prise de décision</b>	<b>83</b>
<b>III. Présentation de l'information et perception</b>	<b>85</b>
III.1. - L'interface homme machine comme canal de communication	86
III.2. - Sémiotique et présentation de l'information	88
<b>IV. Résoudre les trois problèmes de Weaver reste d'actualité</b>	<b>89</b>
IV.1. - Le problème technique de la communication	89
IV.2. - Le problème sémantique de la communication	89
IV.3. - Le problème d'effectivité de la communication	90
IV.4. - Reformulation sémiotique des trois problèmes de Weaver	90
<b>V. Information et cybernétique</b>	<b>91</b>
V.1. - L'interface homme machine et la dynamique de l'interaction	93
V.2. - Le confinement de la sémantique en dehors des interfaces	96
<b>VI. Systémique de la présentation de l'information</b>	<b>101</b>
VI.1. - La présentation de l'information et la multimodalité	101
VI.2. - L'indexation multimédia par alignement sur un document statique	103
VI.3. - L'indexation multimédia par alignement transmodal	103
VI.4. - Un système où la place de l'homme est située	104

## CHAPITRE 4

### LA COMPLEXITE D'UN CALCUL COMME INSTRUMENT DE MESURE \_\_\_\_\_ 109

<b>I. Un point sur les systèmes complexes :</b>	<b>110</b>
I.1. - La complexité computationnelle	110
I.2. - L'utilité de la complexité	113
I.3. - Les dimensions de la complexité	115
I.4. - Les propriétés fondamentales d'un système complexe	119
<b>II. Les interfaces et la complexité</b>	<b>120</b>
II.1. - La complexité de calcul comme mesure de distance	121
II.2. - Discussion sur la classification et l'apprentissage	122

## CHAPITRE 5

### UNE DISTANCE UNIVERSELLE ENTRE INFORMATIONS \_\_\_\_\_ 125

<b>I. Une reformulation du problème "reconnaître une forme"</b>	<b>126</b>
I.1. - Distance entre informations	127
I.2. - L'analogie avec la bioinformatique	128
I.3. - Un exemple de recherche d'informations par le contenu : le séquençage du code génétique	128
I.4. - La complexité de Kolmogorov comme instrument de mesure de distance entre informations	129
I.5. - La mesure de la complexité dans le traitement de l'information	130
I.6. - Les éléments structurants utilisés pour évaluer cette distance d'information	130
<b>II. La distance de compression</b>	<b>131</b>
II.1. - L'indécidabilité calculatoire de la complexité de Kolmogorov et son approximation	132
II.2. - Quelques propriétés de la distance normalisée de compression	132
<b>III. Applications de la distance normalisée de compression</b>	<b>133</b>
<b>IV. Le potentiel d'applications utilisant un calcul de distance universelle</b>	<b>138</b>
IV.1. - Classement multimédia par le contenu	138
IV.2. - Propriétés de la complexité dans les compresseurs	138
IV.3. - Mise en équations formelles des articulations de modalités	139
IV.4. - Perspectives de recherche à plus long terme	139
<b>V. Conclusions sur les applications de la complexité</b>	<b>139</b>

## CHAPITRE 6

<b>MISE EN ŒUVRE DE LA PRESENTATION DE L'INFORMATION</b>	<b>141</b>
<b>I. Révéler les potentialités techniques de notre approche</b>	<b>142</b>
<b>II. Conception et réalisation d'un système d'OCR intuitif</b>	<b>143</b>
II.1. - Contraintes initiales pour notre OCR intuitif	143
II.2. - Modularité de l'OCR intuitif	144
II.3. - Description fonctionnelle de l'OCR intuitif	147
II.4. - Discussions sur l'OCR intuitif	150
II.5. - Application de l'OCR intuitif à la Bible glosée de Gutenberg	152
<b>III. Projet AdHoc : conception et réalisation d'un système de construction incrémentale de réseaux sémantiques</b>	<b>159</b>
<b>IV. Conception d'un système de lecture nomade de contenus multimédias</b>	<b>161</b>
<b>V. Discussion sur ces 3 cas d'applications</b>	<b>163</b>
V.1. - Un peu d'étymologie des mots forme et document	163
V.2. - Les formes et les documents	164
V.3. - Les signes d'une forme	164
V.4. - L'alignement de modalités	166
V.5. - La présentation de l'information	167
<b>CONCLUSIONS ET PROSPECTIVES</b>	<b>171</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b>	<b>178</b>
<b>ANNEXES</b>	<b>181</b>
<b>ANNEXE 1</b>	
<b>ILLUSTRATIONS DES LIMITES DE LA NUMERISATION DE MASSE</b>	<b>183</b>
<b>ANNEXE 2</b>	
<b>DESCRIPTION FONCTIONNELLE DU PROJET DOCMINING</b>	<b>191</b>





## **Avant propos**

*"Il devrait être bien clair aujourd'hui qu'on n'expliquera pas l'univers dans tous ses détails par une seule formule ou par une seule théorie. Et pourtant le cerveau humain a un tel besoin d'unité et de cohérence que toute théorie de quelque importance risque d'être utilisée de manière abusive et de dériver vers le mythe."*

**François Jacob** - "Le jeu des possibles"

## 167 téraoctets de données visibles... X pétaoctets de données invisibles ?

Le document numérique est le principal d'accès à l'information en réseaux. Selon le préambule de la norme ISO 190005-1, en 2005, le web de surface (visible par les moteurs de recherche) représente 167 téraoctets. Hormis le format HTML, le format PDF d'Adobe représente à lui seul 9,7 téraoctets. Les autres formats sont principalement Microsoft Office, XML, Open Office et la "galaxie" des fichiers multimédia. Il ne faut pas oublier que MPEG-4 contient des spécifications relatives au texte et au graphisme.

Cet ordre de grandeur situe par lui-même l'importance stratégique du document numérique. Il met aussi en évidence une forte hétérogénéité des formats. Le chiffre annoncé par l'ISO est la partie visible d'un iceberg, il s'agit des documents numériques susceptibles d'être indexés par un moteur de recherche. La partie immergée n'est pas quantifiable par définition mais il est pertinent d'estimer qu'elle est très importante, car elle correspond clairement à des documents faiblement structurés ou sous la forme d'images non indexées.

La partie visible de cet iceberg documentaire pose le problème de l'hétérogénéité de présentation de l'information du fait de la diversité des formats et des pratiques de composition des documents.

La partie immergée de cet iceberg pose le problème de la mise en visibilité de ces documents et reprend l'ensemble de la problématique posée par la numérisation de documents.

Dès 1996, date à laquelle France Télécom a mis en place une activité d'étude sur le document numérique, il est apparu une frontière existante entre au moins trois grands domaines d'études :

### - l'acquisition et le traitement d'images

Ce domaine couvre traditionnellement le traitement du signal et pose les questions techniques de la numérisation. Elles sous-tendent en particulier, les contraintes mathématiques et physiques de qualité d'une image numérique produite à partir d'un document "analogique" inscrit sur support papier ou microfilmé. Même si l'image numérique d'un document est un signal discret bidimensionnel, ces contraintes doivent prendre en considération deux éléments communs à la numérisation de tout signal analogique : la règle de Shannon et le critère de Nyquist.

La règle de Shannon nous renvoie vers la résolution de l'image produite par un scanner ; cette règle, rappelons-le, permet de choisir la fréquence d'échantillonnage d'un signal analogique en vue de sa numérisation et cette fréquence d'échantillonnage doit correspondre au double de la fréquence maximale du signal analogique que l'on veut numériser.

Transposée à l'image, cette fréquence d'échantillonnage correspond à la résolution de l'image qui devra être en rapport avec la taille de la plus petite part de forme que l'on désire reconnaître. A titre d'exemple, on pourra dire qu'un document comportant des traits importants pour l'interprétation, de 1/10 de millimètres, soit 100  $\mu\text{m}$ , d'épaisseur, devra être numérisé de manière à ce que ces traits soient "couverts" par au moins deux pixels. Ceci nous donnera des pixels de 50  $\mu\text{m}$  de côté.

Si nous rapportons cette taille de pixel à l'unité de mesure traditionnelle de la résolution des scanners en pixels par pouces ("dots per inch" ou en abrégé "dpi"), cela nous donne une résolution de 500 dpi.

Le critère de Nyquist est un peu plus délicat à exprimer dans le cas d'une image numérique de document, mais il n'en est pas moins important. Ce critère nous donne la fréquence de coupure à appliquer au signal analogique pour obtenir un signal numérique conforme.

Cette fréquence de coupure est en relation avec la vitesse de phase de chacune des composantes du signal analogique et doit correspondre à la moitié de la fréquence d'échantillonnage appliquée à la numérisation. Si aucun filtrage n'est appliqué au signal analogique, les composantes d'ordres supérieures produisent des effets de déphasage ou de "gigue de phase" qui aggravent le bruit de quantification en incluant des décalages de phase des composantes du signal.

Dans le cas de l'image, nous avons deux causes de bruits, l'une intrinsèque à la numérisation et l'autre extrinsèque, car due au choix du format de stockage de l'image numérisée, mais qui produisent le même effet : un bruit de quantification qui se traduit dans l'image, par des artéfacts sur les formes et induit de l'ambiguïté dans la segmentation des formes à reconnaître.

Les causes intrinsèques ne sont pas contrôlables, car elles sont données par le positionnement du capteur par rapport à une forme à numériser : un caractère présenté au capteur n'est pas correctement calé sur la grille du capteur. Cet effet ne peut être minimisé que dans le choix d'une résolution suffisante qui préserve l'information de contour des formes.

Les causes extrinsèques, quant à elles, dépendent principalement du choix du format de stockage de l'image, notamment de la méthode de compression utilisée ; en particulier, les méthodes de compression avec pertes, sont basées sur des critères psychovisuels et agissent sur les fréquences spatiales de l'image, la profondeur de pixels. La restitution d'une image codée selon ce mode de compression (par exemple JPEG) est satisfaisant pour l'aspect visuel, étant donné que l'information soustraite à l'image correspond à des détails que notre vision ne sait pas percevoir.

En revanche, ce mode de compression induit un bruit de quantification lors de la restitution de l'image qui se traduit par des contours de formes dégradés rendant plus délicate leur reconnaissance. Ce bruit est d'autant plus important que les formes sont proches de la taille minimale des pixels issus de la numérisation.

Dans un souci de préservation de la conformité de l'image, il apparaît nécessaire de transposer la règle de Shannon et le critère de Nyquist à l'image de documents en choisissant des critères de qualité de numérisation suffisants pour obtenir une image numérique d'un document conforme au document original.

Dans cette thèse nous ne traiterons pas du choix des paramètres techniques de la numérisation, mais plutôt, de la détermination des conditions de conformité de présentation de l'information dans un document numérisé pour un usage donné, conditions à partir desquels sont déduits les paramètres techniques de la numérisation.

## - l'ingénierie des données et des connaissances en numérisation,

Traditionnellement, on considère une chaîne de numérisation comme l'association d'outils et de méthodes d'acquisition d'une image, de traitement de celle-ci pour en dégager des formes et de reconnaissance de ces formes. Cette association est organisée classiquement de manière **séquentielle** et le produit de cette chaîne de numérisation est formé de deux parties :

- des données susceptibles d'être utilisées dans une application informatique, ce qui veut dire, conformes à un modèle de données, à une ontologie, qui est celle de l'application.
- des représentations de ces données qui permettent à l'utilisateur de les manipuler dans une interface, donnant à ces données leur tangibilité.

Par exemple, la numérisation d'un document textuel a pour objectif de construire un texte à partir des formes inscrites dans un document. Cette construction se fait soit par des procédures de reconnaissance de caractères ou de transcription manuelle du texte original. Mais ce texte ne suffit pas à obtenir un document numérique conforme à l'original ; les composantes graphiques et les informations de mise en page sont également des attributs du document original. Leurs transcriptions numériques permettent de rendre tangibles les structures physiques et logiques du document dans une interface homme machine.

Cette remise en forme de la présentation du document implique de traiter également, les formes non textuelles (dessins, images, graphismes et traces issus de la dégradation du document), avec deux objectifs :

- préserver la conformité du texte reconnu par rapport au texte inscrit sur le document en évitant notamment que les traces non textuelles présentes sur le document, ne soient pas des artéfacts ambigus mal reconnu par l'OCR,
- construire une représentation calculable de la structure du document en vue de sa manipulation par l'utilisateur. Nous verrons dans le cas de l'interaction homme machine, que cette représentation de la structure d'un document est importante pour sa manipulation, car c'est elle qui conditionne la tangibilité du contenu.

Globalement, le produit d'une numérisation, vu sous l'angle des données et des connaissances, revient à disposer d'une présentation de l'information contenue dans un document qui répond à deux propriétés fondamentales :

- produire des formes qui font sens pour l'utilisateur,
- associer ces formes à des données dont le codage doit faire sens pour une application, c'est-à-dire, être conforme à une ontologie.

Les problèmes soulevés par ces propriétés sont nombreux : ils touchent à l'ensemble des domaines du traitement des données et des connaissances :

- dans le codage élémentaire de l'information : la valeur des caractères reconnus par un OCR doit être conforme à la table de codage utilisée dans l'application cible et ne pas induire d'ambiguïté ou de "bruit" sémantique.
- dans la segmentation des formes reconnues : le texte doit être manipulable à partir des caractères, mais aussi, selon la nature de l'opération voulue par l'utilisateur, à partir des mots, phrases, paragraphes en fonction de leurs présentations ou mise en forme électronique.



- dans l'indexation des formes du document : le texte inscrit dans un document a été produit dans une langue donnée, avec un type d'écriture ou de typographie qui correspond lui-même à de l'information pouvant faire sens pour l'application.

L'hypothèse théorique d'une chaîne de numérisation idéale permettant d'extraire toute l'information contenue dans un document et de la transposer dans n'importe quel formalisme informatique, conduit à un problème incalculable. Ce problème correspond à l'extrême complexité de description de toutes les formes possibles de documents, allant du support jusqu'à la sémantique, en passant par sa mise en forme.

Si nous posons ce problème en termes de conformité de l'information, on s'aperçoit que chacune des formes présentes dans un document n'est signifiante que pour une famille usages. Classiquement, on tente de simplifier le problème de la numérisation de documents, en le ramenant à celui d'une reconnaissance de texte.

Transposé en termes d'usages cela signifie qu'on limite la lecture d'un document au déchiffrement des caractères. Or la structure d'un document est porteuse de l'information permettant au lecteur de naviguer dans le contenu.

Si cette simplification est applicable à des documents dont la structure est élémentaire, elle trouve rapidement ses limites dans le cas de documents à la structure complexe comme les pages de journaux, les documents richement illustrés, les documents anciens (contenant des lettrines, des enluminures, des annotations). Reconnaître ces formes de structure pose le problème de leur symbolique : à un caractère correspond une valeur unique transposable dans un langage compris par la machine, mais qu'en est-il d'une illustration? des formes correspondant à la structure logique d'un article de journal?

Le sens d'une forme reconnue est obtenu par une combinaison complexe d'un ensemble important de critères, chacun d'entre eux étant obtenus par des programmes logiciels variés (acquisition, segmentation et classification). La reconnaissance d'un seul critère de présentation d'une forme ne suffit pas à en déduire le sens, les relations entre critères sont tout aussi saillantes. Ceci évoque une reconnaissance de formes à placer à l'aune des systèmes complexes. En particulier, pour résoudre l'extrême difficulté d'établir a priori un schéma de relations entre critères saillants d'une forme et sa valeur.

L'élément structurant qui nous permet de déduire les propriétés et les paramètres d'un système de numérisation est la détermination des critères de conformité de l'information.

Cette détermination portera sur l'identification de l'information que l'on veut extraire d'un document. A cette intention on mettra en place un processus de reconnaissance de cette information, garantissant cette conformité en des termes d'utilisabilité dans une application. L'utilisateur de cette application pourra être tantôt expert, tantôt candide par rapport à l'ontologie de l'application.

Cette approche par la complexité, nous conduit à revisiter le concept de reconnaissance de formes en le situant dans un contexte de conformité de présentation de l'information. Dans un tel contexte, l'utilisateur final représente l'élément pilote de la reconnaissance de formes au travers de l'observation de son activité.

Ainsi, d'un processus d'initialisation d'un système d'information, la numérisation et la reconnaissance de formes tendent elles-mêmes à devenir des composants "intelligents" d'interface.

### **- l'interaction homme machine.**

A l'issue d'une numérisation d'un document, l'utilisateur en manipule le contenu à partir de formes calculées par reconnaissance de celles-ci. Ces actions de manipulation correspondent à des intentions produites à partir de la perception du contenu restitué à l'interface.

Ces intentions sont instrumentées par des composants logiciels d'interfaces qui sont présentées à l'utilisateur sous la forme d'actionneurs (boutons, widgets, mais aussi les formes présentes dans le document), et sont aussi des effecteurs qui exécutent des tâches sur des données présentes dans la mémoire de l'application.

Ceci met en évidence le fait que toute interaction homme machine effective doit se faire sur un objet incluant une relation entre des données calculables d'une application et des formes de présentation de l'information saillantes pour l'utilisateur.

Ainsi, la conception même des interfaces ne saurait se limiter à une organisation ergonomique des actionneurs de celles-ci ; elle devra prendre en compte ce que l'on appelle en design des "facteurs de formes" qui porteront à la fois sur la présentation des outils de l'interaction, mais aussi, sur la présentation des objets de l'interaction porteurs de l'information que désire traiter notre utilisateur.

Le problème de présentation de l'information devient ici celui de la construction d'une structure d'objets soumise à deux formes d'interprétations :

- une manipulation par un utilisateur humain, au travers de formes ayant une affordance pour ce dernier,
- un calcul par une machine, au travers de données ayant une cohérence pour l'application concernée.

Or, un processus de reconnaissance de formes a précisément pour objectif de construire cette relation entre formes et données ; cette relation s'exprime en termes de conformité de l'information.

Par conséquent, quel que soit le domaine scientifique abordé, touchant au concept d'information, la conformité de l'information représente une véritable clé de voûte du problème de la reconnaissance de formes.

Cette notion de conformité de l'information est directement héritée des théories de l'information, notamment celle de Shannon et Weaver [SHA48] . Elle se justifie par la nature primaire du document, à savoir, qu'il s'agit d'un support, voire d'un véritable canal de communication.

La communication par un document se fait au moyen des formes inscrites dans celui-ci au même titre que la communication dans un canal se fait au moyen des signaux codés venant perturber ce canal.

Aujourd'hui, nous assistons à une profonde mutation de la notion même de document. De purement textuel à l'origine, il est devenu totalement multimédia, avec une véritable explosion combinatoire des modes de présentation de

l'information. Ceci induit dans nos réseaux de télécommunication l'émergence d'une forme de complexité qu'il convient d'analyser comme objet d'étude par lui-même.

L'écosystème des réseaux a profondément et brutalement évolué, tout comme la définition des moyens d'accès : d'une logique "client serveur" dans laquelle le schéma de communication est celui de la diffusion à partir d'un nombre de serveurs limités vers des terminaux dits légers, nous sommes passés à une logique d'interconnexion totale de "terminaux" ayant des capacités de traitement faisant que la notion de "serveur" se dilue.

L'utilisateur n'est plus un lecteur passif, il est devenu un élément actif, producteur d'informations.

La conséquence en est une démultiplication des moyens de production d'informations et donc, une démultiplication des modes de présentation de l'information, que les approches réductionnistes de la standardisation ne résolvent désormais que partiellement.

La thèse développée ici, tente de dresser un panorama de cette évolution de la présentation de l'information vers la complexité. Elle tente également de proposer, au travers de la question de la numérisation, une voie alternative du traitement de l'information, voie qui se rapproche des théories modernes de l'information, c'est-à-dire celles qui sont apparues après Shannon.

Dans cette approche :

- La machine redevient un instrument de calcul et perd l'"aura" donnée par la notion "d'intelligence artificielle".
- La sémantique est confinée à un domaine qu'elle n'aurait jamais dû quitter : celui de donner à un utilisateur la capacité d'interpréter des formes. Il devient difficile, voire impossible, d'évaluer la pertinence d'un résultat de numérisation sans faire intervenir directement l'utilisateur dans la boucle de traitement.
- Il n'y a pas de connaissance exprimée explicitement dans la machine. La connaissance n'existe qu'au travers de traces ou de signes que sont des données informatiques.
- Ces traces ou données informatiques sont toutes le résultat d'un calcul plus ou moins complexe. Dans le cas le plus extrême, on peut imaginer que l'inscription ou l'effacement d'un bit sera le résultat de l'exécution d'un programme correspondant à un problème à la complexité factorielle. Or l'inscription d'une donnée fait perdre la trace de la complexité qui est, en fait, la présentation de la connaissance, au même titre que le bit de notre exemple est la présentation d'une information. Dès lors, comment imaginer, par exemple, une évaluation d'un processus de reconnaissance de formes, c'est-à-dire l'évaluation d'une cohérence entre données et connaissances, à partir des seules données produites par ce processus? La nécessité d'apprendre à quantifier la complexité transparaît ici.
- Cette complexité se situera toujours dans un domaine d'acceptabilité que seuls les ergonomes seront capables de nous fournir. En effet, cette acceptabilité sera toujours déduite des usages et de l'interaction homme machine. Paradoxalement, nous verrons que cette approche conduit à une grande simplification algorithmique des applications de numérisation ; elle permet de substituer une complexité computationnelle difficilement acceptable et utilisable pour l'utilisateur par des

actions élémentaires de ce dernier : si la production d'un résultat de reconnaissance de caractères fait appel à une complexité de calcul inapplicable au contexte d'utilisation de la machine, ce résultat est donné en substitution par une action simple de l'utilisateur.

- Enfin, cette thèse, en abordant le document comme support de la présentation de toute information, le situe à la croisée de plusieurs domaines traditionnels de recherche en lui conférant un statut particulier, à savoir, celui d'assurer l'interface entre ces différents domaines. Comme l'illustre le schéma ci-dessous, le document dans sa forme numérique, est au carrefour de toutes les disciplines de l'informatique. Ceci peut expliquer les raisons pour lesquelles, chacune de ces disciplines tente d'aborder le sujet du document numérique, mais les descriptions qu'ils nous en donnent ne sont pas satisfaisantes sur un plan holistique. En effet, chacune des descriptions faites du document dans ces disciplines fait appel à des modèles qui seront toujours réductionnistes pour les autres disciplines.

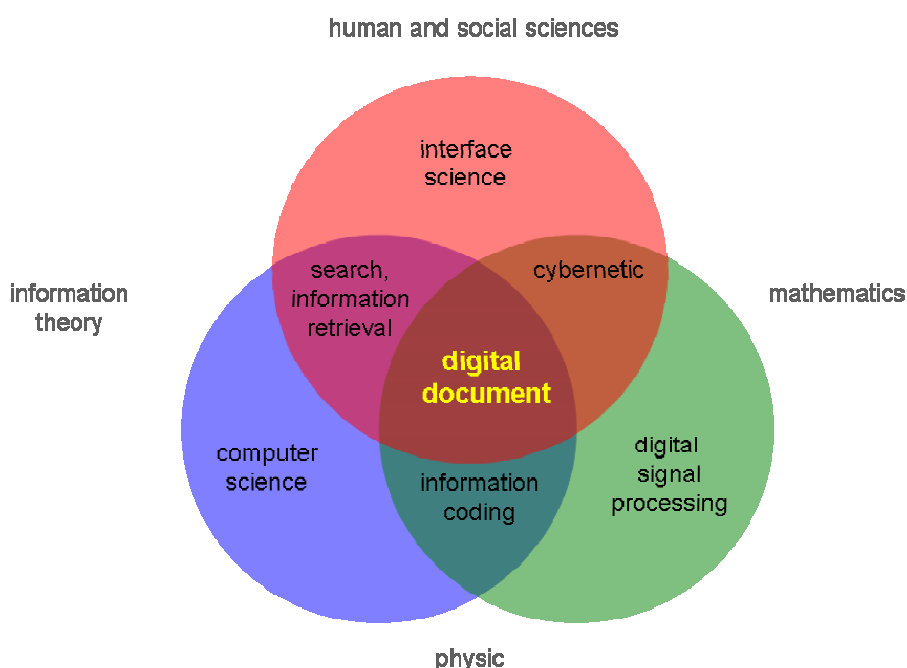


Fig. 1 : Le document numérique comme élément de convergence du traitement de l'information et comme objet d'étude spécifiquement pluridisciplinaire

Ceci milite clairement pour considérer le document numérique comme un objet complexe, dont le contenu l'est également et dont l'étude se doit d'être pluridisciplinaire. Dans cette approche, nous voyons apparaître une problématique qui ne sont pas sans évoquer celle de la physique contemporaine (quantique et statistique en particulier) ; cette thèse abordera cette problématique au travers de la complexité.

Cette approche du document numérique comme support de la présentation de l'information évoque la question posée par Charles Bennett :

**"L'information est elle physique ou la physique est-elle information ?"**

## *Idées directrices et organisation du document*

Comme je viens de l'expliquer dans cet avant-propos, dès que France Télécom m'a confié l'étude de la numérisation des documents techniques, en 1996, j'ai constaté l'existence d'une frontière entre trois grands domaines d'études :

- l'acquisition et le traitement d'images,
- L'ingénierie des données et des connaissances,
- L'interaction homme-machine,

J'en ai rapidement déduit que, dans le cas de la numérisation des documents, cette frontière générerait de nombreux verrous. J'appréhendais le document numérique comme étant un objet complexe, associant un signal, des données et des usages, décrivant et formant son contenu, dont l'étude se doit d'être pluridisciplinaire.

Ces frontières dans le traitement du document allaient aussi à l'encontre de l'évolution de l'informatique et des nouveaux potentiels qui s'y rattachaient : on évoquait déjà l'émergence des réseaux numériques tels qu'ils se présentent aujourd'hui, avec pour conséquence une informatique devenant pervasive.

Ce constat est d'autant plus vrai si l'on aborde le traitement de l'information en partant des besoins et contraintes de l'utilisateur final et non de ceux d'une source émettant de l'information.

La décennie 1996-2006 n'a fait que renforcer ma conviction qu'il fallait aborder le traitement du document de manière plus pragmatique :

- en réunissant les domaines d'études, incluant l'ergonomie pour les usages,
- en se mettant en phase avec l'évolution des matériels toujours plus personnalisables et portables, ouvrant la porte à l'information pervasive.

La culture d'entreprise de France Télécom n'était pas étrangère à ce sentiment.

Quand Orange Labs m'a accordé la possibilité de faire la présente thèse, j'ai pensé à développer et valider ces idées. Ces idées, qui sont à contre-courant de l'automatisation forcenée, étaient partagées par l'équipe document du LIRIS et, en particulier, par mon Directeur de Thèse qui, au-delà de la recherche académique, est souvent sur le terrain de la numérisation industrielle des bibliothèques et des fonds d'archives.

Mon objectif dans cette thèse, est de montrer qu'on peut, au niveau des logiciels de traitement des documents, créer des produits qui correspondent aux matériels d'aujourd'hui, à savoir des produits :

- simples et légers,
- d'un accès simple et intuitif utilisant des moyens de saisie et d'interaction devenus banalisés (le stylet, par exemple)
- modulables
- ayant la capacité de faire évoluer ses données a priori par un apprentissage permanent.

Ce choix sous-entend que l'interface homme machine prend une place importante, peut être même plus importante que les moteurs de traitements des images et de

reconnaissance, accordant un rôle central à l'utilisateur qui devient le pilote de la machine.

De façon concomitante à l'élaboration de quelques réalisations logicielles applicatives, j'ai souhaité donner une base théorique, voire philosophique à mon travail.

Cela a conduit à la rédaction des six chapitres suivants :

Dans le chapitre 1, j'ai rappelé l'évolution de l'informatique et de l'information numérique (pour reprendre la différenciation anglo-saxonne de "computer science" et "information theory"), des années 50 à nos jours. Evolution au cours de laquelle on a pu assister à une banalisation et une diversification de la technique qui devient accessible à tout un chacun et progressivement, partout où il se trouve.

Le chapitre 2 est un retour sur mon expérience personnelle des projets de numérisation. Ce retour a fortement accentué ma perception du contraste existant entre l'état d'évolution des matériels et réseaux prenant en charge les besoins de l'utilisateur, et notre manière de concevoir des chaînes de traitement de l'information qui correspondent encore trop au mythe de "l'ordinateur pensant à notre place".

Dans le chapitre 3, je propose des solutions de ruptures aux problèmes classiques de la numérisation et de la présentation de l'information, en abordant la question sous l'angle de la systémique. Ceci me permet de démontrer que la validation d'une reconnaissance de formes doit à la fois inclure des données utilisables par l'application finale et une présentation de l'information accessible et perceptible par l'utilisateur.

Dans cette optique, il s'agissait de définir les critères de conformité de l'information par rapport aux usages en fonction du support, du codage et des connaissances.

Ce chapitre m'a permis de mettre en avant un critère important intervenant dans l'évaluation de la conformité de l'information, mais peu abordé jusqu'ici dans le domaine du document numérique : la complexité de calcul.

Cette complexité fait l'objet du chapitre 4. De manière très pragmatique, il s'est agi de mettre en œuvre un équivalent technique du principe d'économie propre à toute application ergonomique (un "rasoir d'Occam à deux lames") : plus la présentation d'une information est évocatrice pour l'utilisateur, moins il doit dépenser d'énergie pour en interpréter le sens. Corollairement, plus la description d'une information en objets informatiques est simple et non ambiguë, moins la machine doit faire de calcul pour donner une valeur fiable à cette information.

Le chapitre 5 est une application de la complexité comme instrument de mesure, qui m'a été inspirée par des travaux de recherche sur le génome en bioinformatique, selon une approche très similaire à celle que j'ai développée. Le problème traité est initialement celui du classement des espèces à partir de l'analyse de leur séquence d'ADN.

La solution proposée se base sur une application originale de cette complexité de calcul pour évaluer le degré de similarité entre deux séquences. En l'appliquant à différents contenu multimédias (gestes, visages, documents anciens), j'en ai déduit qu'un système, incluant la complexité comme donnée d'un problème, conduisait à des solutions logicielles s'avérant, à performances équivalentes, voire supérieures,

nettement plus légères sur le plan algorithmique, que les approches classiques tentant de restreindre cette complexité en restreignant le domaine d'application.

Il ne me restait plus qu'à mettre en œuvre le principe de "l'humain dans la boucle" dans le chapitre 6, tel qu'il transparaît tout au long de la thèse. Cela m'a conduit à concevoir une application de reconnaissance supervisée de mots reprenant l'approche systémique décrite dans le chapitre 3. Le résultat est un "OCR léger" par les algorithmes utilisés, mais efficace sur des images de qualités disparates, par la méthode d'évaluation, de correction et d'apprentissage de la reconnaissance, totalement contrôlée par l'utilisateur via une interface stilet et des gestes graphiques. L'application de cet OCR à une Bible imprimée de Gutenberg, n'a nécessité que l'ajout d'un lexique de formes latines ce qui démontre la modularité du système.

Cette modularité pourra être renforcée par l'expérimentation du système sur différentes écritures modernes ou anciennes, comme par exemple, le Syriaque (L'écriture de l'Araméen, la "langue du Christ").

L'objectif est d'obtenir un modèle d'interaction adaptatif et évolutif, nous rendant capable de traiter une information pervasive et hétérogène dans sa présentation.

Ainsi, des contributions à la valorisation de patrimoines historiques permettent de trouver des solutions pour les applications logicielles de notre quotidien ou de notre futur proche d'"homo sapiens telecommunicans".





# Chapitre 1

## Les mutations de l'écosystème de l'information numérique

*"On ne connaît un objet qu'en agissant sur lui et en le transformant."*

**Jean Piaget** - "Psychologie et Epistémologie"

L'informatique est dans un état permanent de mutation. En même temps que la bande passante des réseaux augmente, les puissances de calcul des clients et des serveurs se rééquilibrent et l'accès à des ressources informatiques se banalise. Quelles peuvent être les conséquences de ces mutations sur la conception et le développement de nos applications et interfaces ?

## ***I. D'une information diffusée à une information pervasive***

La mutation brutale de notre rapport à l'information numérique met en avant le besoin d'un schéma conceptuel d'interface capable de rendre compte de la diversité numérique désormais observée sur le web. Celle-ci se caractérise par le fait qu'une même information peut être codée d'un très grand nombre de manières en termes de média utilisés, de modalités de présentation et de descriptions sémantiques. L'une des causes de cette diversité est le nombre élevé de sources d'informations.

D'une logique de diffusion de l'information à partir d'un nombre réduit de serveurs spécialisés, nous sommes passés à une logique d'interconnexion complète de nœuds d'un réseau au maillage stochastique, ceux-ci étant tour à tour clients ou serveurs d'informations.

L'utilisateur, muni d'un terminal qui est désormais un appareillage multifonction, multimodal et multimédia se trouve immergé dans un environnement d'informations pervasives dont la présentation, établie par un nombre élevé de sources, revêt un caractère fortement hétérogène.

Cela induit la nécessité d'établir une forme d'intermédiation entre l'utilisateur destinataire de l'information et les différentes sources. Cette intermédiation doit porter conjointement sur la présentation de l'information communiquée par les sources et sur celle produite par les actions de l'utilisateur, de manière à ce que chacune des sources puissent interpréter ces actions et les traduire en exécution de programmes logiciels.

### **I.1. - Trois ères des systèmes d'information**

Le rapport scientifique et prospectif "Being human", rédigé par un groupe de chercheur de Microsoft, met en avant un profond changement dans nos moyens d'accès à l'information par rapport à l'évolution de la définition même des ordinateurs. Cette évolution des machines suit exactement l'évolution de la puissance de calcul liée à la miniaturisation des composants.

Elle correspond à la loi de Moore et se traduit par les trois principales périodes de l'histoire de l'informatique.

#### ***1.1.1. - L'ère du mainframe (des origines aux années 80)***

Dans cet environnement, nous avons peu de machines et des utilisateurs qui accèdent à l'information via un terminal léger (la plupart du temps, synchrone). Cet environnement est la conséquence de la technique : le concept d'ordinateur correspondait à une machine volumineuse ; la mise en œuvre d'un programme informatique nécessitait un haut niveau d'expertise, tout comme l'administration d'un système d'informations. Inversement, l'utilisateur disposait d'une technique peu sophistiquée (application en mode texte ou, plus tard, videotext) ; le rôle du terminal se limitait à l'affichage d'une vue de l'information, codée au niveau du serveur, avec de faibles capacités d'interaction, à savoir, les commandes prévues par le serveur. Globalement, l'utilisateur était un lecteur passif de l'information contenue dans un mainframe. En outre, les réseaux étaient du type synchrone avec une connexion physique permanente à un serveur.

### ***1.1.2. - L'ère de l'ordinateur personnel (des années 80 à 2000)***

La technique devient diffusable sur une machine de taille réduite pour un coût abordable. Dans ce contexte, le paradigme de l'accès à l'information devient celui d'un utilisateur lié à sa machine personnelle. Les capacités de connexion via un réseau de télécommunication sont encore limitées, car basées sur la même logique que celle du réseau téléphonique, c'est-à-dire, l'établissement d'une connexion temporaire de point à point qui s'établit dans un réseau hiérarchisé. Si internet existe, son accès reste limité pour le grand public pour des raisons de débit (réseau terminal encore analogique et essentiellement filaire), et de principe de facturation basé sur celui des communications téléphoniques, qui correspondait à l'utilisation de ressources techniques peu disponibles et coûteuses à mettre en œuvre. Cela signifiait un coût basé sur la durée de la connexion, la distance, puis, progressivement, le volume d'informations échangées. L'ordinateur personnel connecté contenait essentiellement des données et programmes personnels et la connexion vers des serveurs reprenait le schéma des architectures clients serveurs.

### ***1.1.3. - L'ère de l'intelligence ambiante (de 2000 à 2008)***

La technique a atteint un degré de sophistication telle qu'elle devient portable. Cela signifie un nouveau paradigme de l'accès à l'information qui se traduit par un utilisateur muni d'un terminal portable, ayant de fortes capacités de calcul et de connexion à un réseau. Ce dernier n'est plus obligatoirement filaire ; la contrainte d'immobilité disparaît donc, au même titre que la notion physique de machines interconnectées. : L'utilisateur "voit" une seule machine, dont il ignore le degré de complexité, ce dernier étant dilué, voire complètement masqué par les outils (logiciels et interfaces embarqués dans son terminal). Mais, d'un point de vue définition d'un système d'information, cela correspond à un utilisateur ayant à sa disposition une infinité de machines, toutes équivalentes en termes de puissance et d'accès physique transparent à l'information.

Ces trois périodes de l'informatique se résument comme suit :

- 1948-1980 : l'ère du "one machine, many users",
- 1980-2000 : l'ère du "one machine, one user",
- 2000-2028 : l'ère du "many machines, one user"

Du fait des capacités d'interconnexions physiques généralisées et permanentes, il conviendrait mieux de parler de l'ère du "many machines, many users", mais la définition proposée par le groupe de chercheurs de Microsoft met bien en avant le passage d'une organisation des systèmes d'information centrée machine à une organisation centrée utilisateur immergé dans un environnement d'informations ambiantes et pervasives.

*Cette analyse de la situation actuelle des systèmes d'information pose le cadre de la présente thèse.*

## **1.2. - De l'interconnexion physique à l'interconnexion logique des systèmes d'information**

Si les réseaux de télécommunication ont atteint des capacités de bande passante, et les machines, des capacités de calcul faisant que l'accès physique à l'information ne

pose plus aucun problème, qu'en est-il de l'interconnexion logique entre applications ?

Cette question ne se poserait pas si nous étions dans une logique de normalisation forte du codage des contenus avec un mode unique de présentation de l'information, pour un profil unique d'utilisateur qui conduirait à :

- une modalité unique de présentation de l'information : le texte
- une restitution unique de l'information : le visuel
- une ontologie centralisée des connaissances : wordnet, par exemple.

Mais la réalité informatique est toute autre : la standardisation des contenus que l'on observe est, en fait, une compétition sélective dans laquelle il n'y a pas de véritables normes pérennes ; nous sommes face à une cohabitation conjoncturelle de standards de faits qui conduisent à un "écosystème numérique" évoluant en permanence.

### **I.3. - D'une médiation matérielle à une médiation logicielle dans les interfaces**

De plus, la capacité des machines a permis le développement d'applications et d'interfaces graphiques dans lesquelles l'interaction homme machine fait de moins en moins appel à une technicité propre à l'informaticien : la souris et sa logique de pointage/sélection est depuis longtemps un outil banal. Cependant, cet appareillage se voit progressivement concurrencé ; nous assistons actuellement à l'émergence forte des interfaces tactiles dans lesquelles l'utilisateur désigne directement à l'écran une forme porteuse de l'information qui l'intéresse. Il n'y a plus la médiation technique et matérielle de la souris, mais une reconnaissance directe du geste pour une action naturelle.

L'évolution sur laquelle nous pouvons parier est celle de la tangibilité des contenus : actuellement, l'objet que désigne l'utilisateur est en fait, une représentation d'un fichier (par exemple, une photo). Le seul cas d'usage courant où cette désignation porte sur les objets d'un fichier, donc sur de vrais contenus, est celui du texte structuré ou, dans une moindre mesure, des objets graphiques ; dans ces deux cas, il s'agit de formes agencées selon une structure correspondant à des objets informatiques bien identifiés lors de la conception du fichier contenant ces formes.

Or cet agencement de formes structurées présentes dans n'importe quel document numérique restitué par une interface est précisément ce que nous appellerons **la présentation de l'information** faite à l'utilisateur.

Inversement, le moyen traditionnel de saisie d'information mis à disposition de l'utilisateur reste l'archaïque clavier, héritage de la télégraphie et du télex ; il assure une médiation technique qui permet à l'utilisateur de coder directement et simplement, après un apprentissage, des données textuelles et d'exécuter des commandes élémentaires.

Une évolution, concomitante à celle des interfaces tactiles, est aujourd'hui largement entamée : la saisie d'informations au travers de gestes de l'utilisateur sans cette médiation matérielle du clavier. Il ne s'agit plus seulement de saisir un texte par reconnaissance d'un tracé manuscrit, mais aussi, de considérer toute action de l'utilisateur comme porteuse d'informations correspondant à ses intentions. Intentions qui correspondent à des actions dont le but est de naviguer dans un

contenu, lancer une commande de traitement (par exemple, un copier coller d'un extrait de phrase pour exécuter une recherche d'information, etc.).

De plus, les terminaux sont tous équipés aujourd'hui de capteurs multimédias (audio, image fixe, vidéo) qui sont autant de dispositifs pouvant potentiellement saisir de l'information.

Ainsi, nous voyons se dessiner un terminal épuré pour l'utilisateur, qui pourrait bien être, à terme, dépourvu de toute commande mécanique assurant une médiation technique et matérielle pour la saisie et le traitement d'informations (Fig. 2), médiation matérielle qui se verrait substituée par une médiation logicielle, en considérant que toutes les actions de l'utilisateur sont également des présentations de l'information.



Fig. 2 : vue d'artiste montrant un terminal épuré de ses commandes mécaniques : toute l'interaction homme machine est faite par des commandes exécutées par reconnaissance de gestes sur un contenu capturé à la volée (source : Google)

## ***II. L'émergence de la complexité dans la relation homme machine***

Un écueil technique difficile freine cette tendance : l'ordinateur reste une machine qui effectue des calculs dans une logique de programmes qui n'a pas changé depuis la thèse de Church-Turing [TUR36][BER02] ; la communication d'informations se fait toujours au travers de signaux codés dont la logique est la même depuis la théorie mathématique de la communication de Shannon et Weaver [SHA48]. Or, dans un contexte multimédia et multimodal, une séquence binaire peut représenter une infinité d'informations : un symbole, un pixel, un phonème, un graphème, un segment de geste, une force de retour d'effort, etc.

Il apparaît là une source latente d'ambigüités pour l'interprétation de ces informations, ambigüités qui ne peuvent être levée que par la connaissance de l'information de structure du document numérique porteur d'information et communiqué entre une source et un destinataire. Cette connaissance de la structure du document devra être négociée entre source et destinataire, du fait des propriétés et contextes fortement hétérogènes de production et d'utilisation des contenus.

## II.1. - Ambiguïté et décision

Cette substitution logicielle de la médiation matérielle des actions de l'utilisateur s'accompagne d'une complexification de la présentation de l'information. Par exemple, analysons le geste d'un utilisateur traçant un cercle à l'écran, avec l'intention de sélectionner un fragment d'un contenu. Ce geste doit être détecté et mesuré par le système : l'utilisateur a bien tracé quelque chose qui est une forme, ce qui pose le problème de la différenciation entre un signal intentionnel et un bruit.

La mesure de cette forme ressemble à un cercle : la description produite par le calcul de cette forme doit permettre de reconnaître un cercle dans l'ensemble des formes connues du système avec le moins d'ambiguïté possible, ce qui pose le problème de la saillance de la forme.

La finalité de cette forme correspond à la sélection d'un fragment d'un contenu : seul le contexte de l'interaction, donné par l'application et le contenu restitué à l'interface, permet de déduire que l'intention d'action consistait à sélectionner un fragment de contenu. Dans une interface générique de saisie et de manipulation, le cercle tracé par l'utilisateur peut tout aussi bien signifier

- la sélection d'un fragment de contenu,
- la saisie d'un chiffre zéro ou de la lettre O,
- une commande de "zooming" ou de "dézooming" d'une image,
- une commande d'ajustement du volume d'un signal audio,
- etc.

Cette signification du geste correspond à un problème de sémantique et l'on voit ici apparaître une relation non bijective entre la valeur d'une forme (ici un cercle) et son sens (ici la signification de ce cercle pour une application donnée). Nous avons ici une ambiguïté sémantique à résoudre.

Mais nous avons vu aussi que la reconnaissance d'un cercle à partir d'un tracé de l'utilisateur posait le problème de la représentativité, ou de la saillance de la description de ce tracé. Ainsi, le mode de description d'un tracé peut induire une ambiguïté sémiotique qu'il faut également résoudre.

Enfin, la séparation du signal et du bruit pose un problème d'ambiguïté physique, qui lui, sera principalement le fait des propriétés du capteur de geste.

Si nous nous plaçons dans le contexte de la numérisation de documents ou dans celui de la vision par ordinateur, nous retrouvons exactement ces trois problèmes d'ambiguïté physique, sémiotique et sémantique qui viennent pénaliser les performances d'un système de reconnaissance.

Mais en cherchant à analyser les causes de ces mauvaises performances, nous observons une subtile intrication entre les causes propres à chacune de ces ambiguïtés. La résolution de ces problèmes d'ambiguïtés nécessite d'aborder la reconnaissance de formes selon l'approche des systèmes complexes ; une action correctrice permettant d'optimiser un des composants d'un processus de reconnaissance de formes a des effets difficilement prévisibles sur les autres constituants de ce processus. Ce comportement connu des systèmes de reconnaissance de formes est l'apanage des systèmes dynamiques complexes dans

lesquels la connaissance intime de chacun des éléments constituant ce système ne permet en aucun cas de déduire le comportement global de l'ensemble.

## **II.2. - Des formes de la complexité, de la complexité des formes**

Il apparaît intéressant d'explorer des hypothèses données par la théorie des systèmes pour étudier et comprendre un système complexe tel que celui d'une reconnaissance de formes. L'un des principaux aspects de l'étude d'un système complexe est la recherche et l'identification d'éléments structurants de ce dernier, indépendamment de ces composants. Ces éléments structurants permettant d'appréhender le comportement du système sont des points de vue à inclure dans l'interface homme machine de ce système.

Nous avons choisi cette approche en nous calquant sur l'évolution de la théorie de l'information et de la communication postérieure à Shannon [SHA48]. Le fait de considérer que toute interaction homme machine se base sur la présentation de l'information rend légitime cette approche, comme nous tenterons de le démontrer.

Les hypothèses qui seront développées se basent sur la nécessité d'inclure une approche sémiotique dans le processus même d'interaction avec le contenu ; une présentation de l'information est toujours le résultat d'un calcul opéré par la machine.

La conformité de ces formes par rapport aux intentions d'usages est la condition sine qua non pour obtenir des contenus "tangibles" sur lesquels un utilisateur peut effectuer des actions ou des opérations de traitements allant d'un simple copier/coller ou d'une recherche d'information, jusqu'à une reformulation complète de la présentation du contenu. L'objectif sera de transposer la présentation de l'information dans une modalité compatible avec les aptitudes d'un utilisateur muni de ses instruments matériels et logiciels.

## **III. Physique et information, deux évolutions parallèles**

### **III.1. - Un modèle réductionniste de l'information aux limites**

En préliminaire au développement de ces principes, il est nécessaire de mettre en évidence l'évolution profonde, voire la mutation de notre rapport à l'information telle que nous l'observons aujourd'hui sur le web.

En effet, la question de la présentation de l'information comme conditionnant son interprétation nous renvoie sur celle de la dépendance entre le calcul de la valeur d'une information (donc du résultat d'une interprétation) et les moyens mis en œuvre pour observer et mesurer une présentation de l'information.

Cette forme d'intrication subtile entre l'utilisateur, ses moyens techniques et les contenus sur le web, remet en cause un certain déterminisme du codage des documents : un contenu ne sera interprétable que si les contextes (source, destinataire, réseau) le permettent. En ce sens, nous nous écartons du déterminisme classique des bases de données.

Les effets de ce non déterminisme induisent de l'ambiguïté du sens d'une information que l'on observe déjà dans le contexte du web (valeurs d'une information imprécise, contradictoire, voire absente selon le contexte).

Les causes sont à voir du côté de la démultiplication :

- des sources d'informations ayant chacune leurs propriétés,
- des moyens de production de contenus,
- des choix de mise en forme de la présentation de l'information dans des documents,
- des ontologies relatives à différents domaines de connaissances, de savoirs qui ne se recouvrent pas exactement.

Ainsi, la valeur de l'information contenue dans un document dépend étroitement des modèles mis en œuvre lors de la production d'un contenu.

Vu du destinataire de l'information, cela se traduit par la remise en question :

- de la confiance que peut accorder ce dernier à la légitimité du contenu en termes de sens ou de validité de la source,
- de sa capacité à déchiffrer ce contenu par les formes qui lui sont présentées,
- de sa capacité de simplement décoder le format du contenu.

Si les questions liées au codage physique de l'information, sont aujourd'hui facilement résolues au moyen d'outils de conversion de formats (wrappers), celles relatives à l'interprétation posent un problème majeur :

**Le destinataire d'une information ne peut s'en approprier le sens que s'il dispose des savoirs, savoir-faire et connaissances relatives au sens de l'information qu'il reçoit. Il doit aussi être capable de s'approprier la syntaxe utilisée pour interpréter le contenu.**

Revenons à la définition de ce qu'est une syntaxe, non pas au sens informatique, mais au sens linguistique.

La syntaxe est la description de la manière dont des **morphèmes libres** se combinent pour former des **syntagmes** menant à des **propositions** formant elles mêmes des énoncés.

En linguistique, un **morphème** libre est un mot, mais, si l'on se penche sur l'étymologie de ce nom, un morphème est une forme. Dans un texte, cette forme représente un concept, une idée, un témoignage, elle revêt l'aspect symbolique d'un mot formé de caractères. Nous pouvons parfaitement étendre cette définition de morphème à tout objet mis en forme d'un contenu numérique, qu'il s'agisse d'un fragment d'image représentant un visage, d'un graphique représentant une séquence de nombres, d'une ligne de séparation des colonnes d'un journal, etc.

Un groupe de formes, avant de correspondre à un sens, a une signification que l'on peut qualifier d'opérationnelle. Il s'agit de groupes nominaux et ou groupes verbaux, selon que ces groupes de formes représentent un objet ou une action sur un objet.

Cette définition générale d'une syntaxe met en évidence plusieurs points :

- Les morphèmes libres sont les formes les plus élémentaires d'une langue et la capacité de les lire conditionne toute l'interprétation que l'on peut en faire. Transposés à notre problématique de présentation de l'information, ces morphèmes libres doivent à la fois être perceptibles par un humain qui en comprend la signification et reliés à des données calculables par un ordinateur. Ce sont des Unités Techniques de Manipulation (Bachimont) dont l'existence est subrogée à cette double propriété d'affordance et de calculabilité.



- Les syntagmes sont le résultat d'une combinaison de morphèmes. Généralisée à la présentation de l'information, une combinaison de morphèmes correspond à la structure physique d'un document qui comprend un ensemble de relations topologiques que l'on peut décrire sous la forme d'un graphe.

- En appliquant à ces syntagmes des propriétés d'objets ou d'actions et en construisant des relations entre ces syntagmes, on obtient une structure logique d'une proposition (par exemple, sujet, verbe, complément), qui ne donne pas encore le sens de cette dernière mais son articulation en vue de son interprétation. Dans le domaine du document, on parlera de forme regroupée sur le plan topologique et indexée sur le plan logique.

- Enfin, l'énoncé sera le document lui-même, avec tous les attributs de formes, de relations topologiques et logiques qui permettent au destinataire d'en interpréter le sens.

Ainsi, dans tout le processus de construction d'une présentation de l'information, nous devons être capables de donner une certaine matérialité perceptive pour le destinataire à tous les éléments composant ce document.

La question qui se pose ici est celle du langage de description de toutes ces formes ou morphèmes, donc, de la construction d'une syntaxe capable de former un "discours graphique" pour les objets vectoriels d'un document, d'un "discours haptique" pour les phénomènes mesurés dans un geste et un retour d'effort, etc.

Or cette matérialité va dépendre des capacités du destinataire à la percevoir, à l'analyser et à l'interpréter, ce qui signifie qu'il doit connaître et partager les mêmes conventions de mise en forme d'un document que la source.

Mais, nous l'avons vu, le destinataire est de plus en plus soumis à une démultiplication du nombre de sources d'information ayant chacune des moyens et pratiques de mise en œuvre d'une syntaxe destinée à former un message de communication (ou "syntaxe véhiculaire" selon Bachimont), à commencer par :

- les formes et leur inscription sur un média physique qui permettront de qualifier le type de document (image, graphisme, texte, mais aussi, un son, une force, etc.) ; le destinataire doit disposer des mêmes moyens techniques que la source pour pouvoir identifier ces formes ou avoir recours à un "interprète" qui reformule cette présentation physique dans un format compatible avec les instruments du destinataire.

- ensuite, les structures physiques, logiques et le document lui-même sont représentés par un ensemble de formes. Ainsi la présentation d'une information revient à construire la relation entre des formes et des groupes de formes, avec une fonction précise dans un document. Les définitions de ces fonctions correspondent à ce que Bachimont appelle des Unité Sémiotiques d'Interprétation dont la matérialité dépend des connaissances du destinataire ; ces dernières doivent être conformes à celles de la source pour que la communication d'un document conduise à une conformité de l'information. Si cette conformité n'est pas établie, il s'agira également d'avoir recours à un interprète pour la reformulation sémiotique du contenu.

- enfin, à partir des connaissances liées à ces fonctions notre destinataire va pouvoir interpréter le sens. Ces connaissances et savoirs doivent être partagés ou négociés entre la source et le destinataire d'une information.

Ainsi, quel que soit le niveau de description ou de présentation d'une information, le prérequis indispensable est le partage négocié et consensuel de la syntaxe véhiculaire utilisée dans un document entre la source et le destinataire. Ceci donne à l'intermédiation sa définition la plus précise et met en évidence sa complexité. C'est cette complexité que nous allons tenter d'illustrer au travers de trois expériences de pensées.

### III.2. - Trois expériences de pensées



Fig. 3 : Cette illustration de Granville de la fable "le chat, la belette et le petit lapin", de Jean de la Fontaine, rassemble les trois "héros" des expériences de pensées que nous abordons :

- le "chat de Schrödinger" montrant l'intrication entre la valeur d'une information, sa présentation et les instruments logiciels utilisés pour observer l'information,
- le "lapin cavernicole" qui est un raccourci du mythe de la caverne de Platon montrant toute l'ambiguïté induite par le contexte d'observation de toute information,
- la "belette dans un labyrinthe" qui permet d'expliquer les trois modes d'observation du comportement d'un programme logiciel dans une interface.

Les possibilités de construction d'une présentation de l'information représentent une véritable combinatoire entre les niveaux physique, sémiotique et sémantique que nous venons d'évoquer.

Cette combinatoire fait perdre au document, donc à la présentation de l'information son statut et ses propriétés déterministes. Un document est produit par une source en fonction de règles et pratiques fixées a priori. Elles ne sont pas obligatoirement partagées par un destinataire, avec pour conséquence une confrontation entre la technique de mise en forme d'un document et l'arbitraire de son interprétation.

Ceci apporte une dimension stochastique à l'interprétation des contenus qui n'est pas sans rappeler l'expérience de pensée du chat de Schrödinger (Fig. 4) en mécanique quantique.

### III.2.1. - La reproduction des chats de Schrödinger

Ce modèle de pensée ("Gedankenexperimente" en philosophie), illustré en (Fig. 4) décrit l'indécidabilité de la valeur d'état d'un objet (en l'occurrence, la valeur d'une fonction d'onde), tant que l'on n'a pas observé cet état.

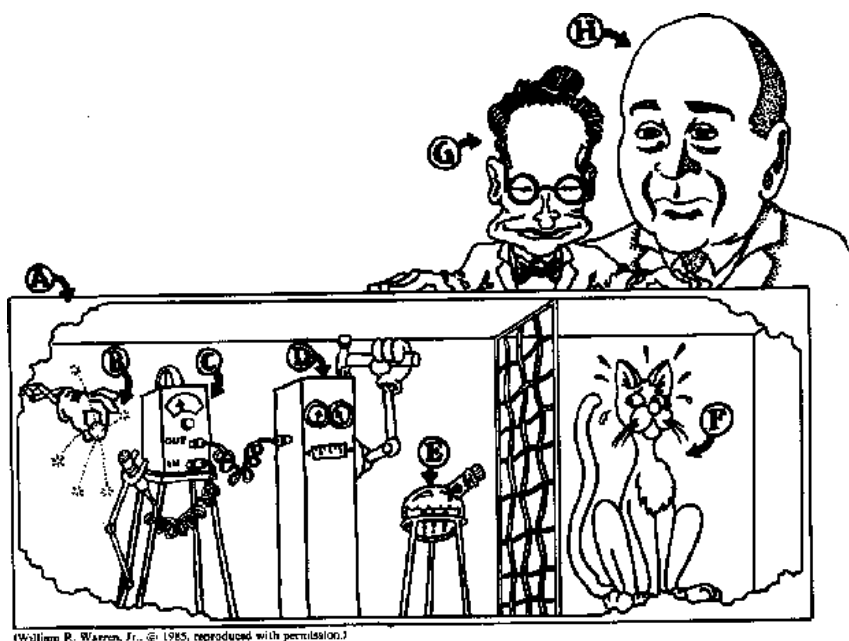


Fig 4 : L'expérience du chat de Schrödinger.

Il s'agit d'une expérience théorique consistant en une boîte étanche (A) dans laquelle on place une source radioactive (B) au comportement statistique (équiprobabilité d'émettre ou ne pas émettre un rayonnement). Ce rayonnement est mesuré par un compteur Geiger (C) qui déclenche, en fonction de la valeur de rayonnement un mécanisme (D) brisant un "flacon d'acide prussique" (E). Dans ce dispositif, on place un chat (F), on referme la boîte et l'on attend... L'observateur "Erwin" (G) constate que l'état du chat (mort ou vivant) est indécidable car il suit la probabilité d'émission d'un rayonnement. Seule l'ouverture du dispositif permet de statuer sur l'état du chat. Avant cette observation on ne peut considérer que le chat aurait deux états superposés, à savoir à la fois mort ET vivant. Mais on peut également partir du principe que l'observateur (G) a également un comportement en "chat de Schrödinger", ce qui signifie que l'on a besoin d'un observateur "Léo" (H) pour déterminer l'état de l'observateur (G), etc...

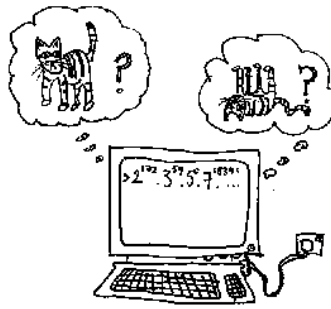


Fig 5 : Une vision "informatique" du chat de Schrödinger.

La conséquence de ce modèle d'expérience de pensée est double (Fig. 5) :

- a) la valeur d'une information n'est obtenue qu'au moment de son observation, ce qui, dans notre cas, correspond à un sens de l'information qui émerge de l'interaction entre un humain et un contenu,
- b) cette valeur d'une information dépend des instruments d'observation, des techniques qui sont mises en œuvre pour identifier, analyser et interpréter les formes présentées à une interface.

En physique quantique, ce problème de l'observation et de la mesure est dû au fait qu'il s'agit d'observer sur un plan macroscopique (ici, la perception de la présentation de l'information par un observateur), des phénomènes microscopiques intriqués et au comportement aléatoire.

La transposition de ceci en information numérique va considérer l'ensemble des sources d'informations du web comme l'analogie des phénomènes aléatoires d'un système quantique et la recherche d'information comme l'analogie des méthodes d'observation et de mesure de ce système quantique à un niveau macroscopique.

Dès lors, nous n'avons accès à chaque source d'informations qu'au travers d'un moteur de recherche recevant des requêtes de l'utilisateur/observateur ; requêtes formées à partir et au moyen de l'environnement technique de traitement de l'information par ce dernier.

Il apparaît qu'en dehors de ce contexte d'observation et de mesure de la valeur d'une information, on ne peut plus établir a priori la valeur de celle-ci.

L'une des meilleures illustrations de la maîtrise des méthodes d'observation de la valeur d'une information réside dans une exploitation frauduleuse du "ranking" du moteur de recherche Google, en augmentant massivement et artificiellement les références hypertextuelles vers un site donné. C'est ce phénomène, dénommé "Google bombing" qui conduisait à des résultats de requêtes surprenants (l'exemple le plus connu ayant été pendant un temps la requête "The worst man in the world" qui retournait comme premier résultat le site de la Maison Blanche).

Outre son aspect anecdotique, cet exemple montre bien le contexte actuel de l'information sur le web, fortement intriqué avec les méthodes d'observation de cette dernière.

### *III.2.2. - Le lapin et la caverne de Platon*

Les formes, quant à elles, doivent être suffisantes pour que leur interprétation ne soit pas ambiguë. Cela signifie que l'on doit avoir suffisamment d'instruments d'observation pour donner à la forme sa matérialité perceptive.



*Fig. 6 : L'ombre projetée représentant un Lapin. Les signes correspondant à cette ombre sont ambigus et ne permettent pas de reconnaître l'objet.*

Une autre expérience de pensée nous permet d'appréhender cette matérialité perceptive. Cette expérience évoque très schématiquement le mythe de la caverne de Platon (Fig. 7a), et plus prosaïquement, l'interprétation d'une ombre projetée d'un lapin dansant (Fig. 6). Cette l'interprétation est ambiguë puisque Cette ombre portée est une forme évoquant plutôt une main.

Cette expérience transposée à l'information numérique montre le rôle fondamental de la description d'une information donc, de la syntaxe véhiculaire utilisée par la source. Décrire une information c'est en construire une présentation qui fait sens pour son interprétation. Cela revient à "projeter" l'information au moyen de techniques (de numérisation, de saisie, etc.). Le résultat de cette "projection" doit être suffisamment discriminant pour que l'interprétation de l'objet projeté soit univoque.

En numérisation de documents, on pourra citer comme exemple, le choix de la valeur d'un paramètre aussi élémentaire que la résolution de l'image produite. Une résolution trop faible fait perdre de l'information, à l'instar de l'ombre portée d'un objet éclairé en fait perdre sa morphologie fine. C'est précisément cette information perdue qui est source d'ambiguïté pour l'interprétation (ici la segmentation/classification de formes).

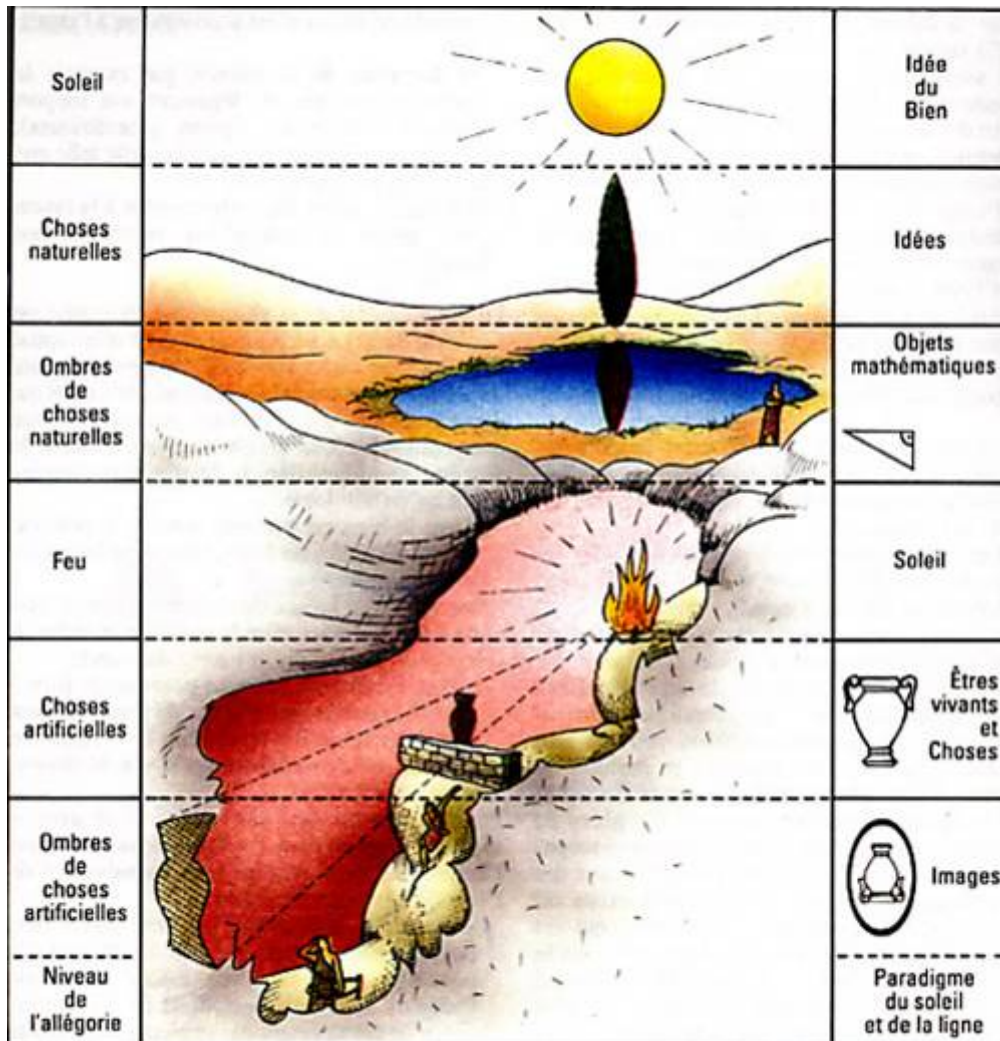


Fig. 7a : Représentation du mythe de la caverne de Platon : dans notre problématique, la présentation de l'information est métaphorisée par les ombres de choses naturelles ou artificielles. Ces ombres sont des images ou des objets mathématiques représentant "quelque chose".



Fig. 7b : La version "moderne" du mythe de la caverne de Platon ?

### III.2.3. - Un programme traitant des données complexes : une belette dans un labyrinthe

Enfin, il y a le processus d'analyse et d'interprétation lui-même qui, dans une interaction homme machine, repose sur l'exécution de programmes logiciels, eux-mêmes générateurs de formes devant être analysées et interprétées.

La dynamique de ce processus va conditionner la charge cognitive à mettre en œuvre par le destinataire dans l'ensemble de l'interaction, selon la lisibilité des formes, le séquençement de celles-ci lors de l'exécution d'un traitement et l'attention que devra apporter ce dernier dans la supervision/régulation du dialogue homme machine.

Ici, une troisième expérience de pensées permet d'appréhender ce problème de supervision/régulation du dialogue. Imaginons une belette dans un labyrinthe. Le labyrinthe représente l'ensemble des données d'un traitement et la belette représente le programme exécutant ce traitement. On peut imaginer la supervision selon trois principes dans lesquels nous allons retrouver plusieurs paradigmes de l'ergonomie des interfaces : les seuls points de vue offerts à l'utilisateur sont l'entrée et la sortie du labyrinthe. Ce dernier est considéré comme une "boîte noire". L'utilisateur fait entrer notre belette dans le labyrinthe et attend que cette dernière en ressorte. La question qui se pose ici est : que faire quand la belette ne sort pas du labyrinthe et comment en analyser les causes ? La solution est d'accéder à une vision complète de ce labyrinthe et de rechercher à quel endroit notre mustélide s'est arrêté. Cette recherche nécessitera aussi d'avoir une connaissance complète du comportement de l'animal.

On peut imaginer un labyrinthe comme étant une "boîte transparente" dans laquelle le manipulateur peut suivre en continu la progression de la belette. Cette observation continue, après de nombreuses expériences, aboutit à une connaissance exhaustive de ce labyrinthe, mais requiert une charge cognitive importante pour l'observateur.

Une approche intermédiaire de notre expérience de pensée est celle d'un labyrinthe dans lequel existent des points d'observation avec un dispositif qui permet de signaler au manipulateur que la belette a franchi cet endroit. La supervision de la progression de notre animal devient discrète. Elle ne nécessite pas une attention soutenue de la part du manipulateur et, de surcroît, limite la recherche des causes d'arrêt de l'animal à une partie seule du labyrinthe.

En termes d'interaction homme machine, le concept de "boîte noire" revient à concevoir l'interface en se reposant entièrement sur la technicité du programme logiciel en cours d'exécution et en appliquant des contraintes d'exécution fortes limitant le risque d'exceptions. Ce concept de boîte noire atteint vite ses limites quand le traitement demandé par l'utilisateur fait appel à un contexte complexe à définir (labyrinthe avec de nombreuses solutions de parcours). Même s'il permet de masquer la technicité ou la complexité d'exécution d'un programme, il ne tolère pas d'exceptions dues à des situations de contexte non prévues a priori.

Le concept de boîte transparente présente toute la complexité des données à l'utilisateur à l'interface et la charge d'identifier une configuration d'exception repose entièrement sur l'expertise du manipulateur. Ce concept de boîte transparente est typiquement celui des interfaces pour experts d'un problème à traiter. L'observation en continu du système permet de repérer les situations d'exception en cours de traitement au prix d'une connaissance exhaustive des données et du comportement du programme. Toute la complexité de situation est présentée à l'interface ce qui limite le profil d'usage à des utilisateurs experts.

Le concept intermédiaire auquel conduit notre expérience de pensée met en avant un point important : la nécessité d'établir une stratégie de supervision globale du

système programme et données (ou belette et labyrinthe) qui permet d'informer rapidement l'utilisateur d'un déroulement du traitement conforme à ses attentes.

Dans cette approche, il n'est pas nécessaire d'avoir une connaissance exhaustive de tous les comportements possibles du programme ni de toutes les configurations des données. Les points de visites du labyrinthe sont des points de contrôles à l'interface d'un système ayant un nombre élevé d'états possibles. Ce concept autorise un apprentissage par la détermination des points de contrôles fait à partir de l'analyse des traces d'activités du couple homme machine. Il semble bien approprié à la définition d'interfaces de systèmes d'information complexes.

### **III.3. - Point sur les théories de l'information et de la communication**

L'idée d'un principe de causalité entre le sens interprété d'une information (l'effet) et la présentation perçue par un utilisateur (la cause), laisse supposer que la méthode d'observation elle-même peut influencer la valeur d'état d'un objet.

Dans le contexte de l'information présente dans le "système web", la démultiplication des sources conduit à des valeurs probabilistes de l'information dont la présentation se comporte comme une variable aléatoire. On peut émettre comme hypothèse que nos outils d'observation de l'information ont un effet direct sur la valeur d'une information.

La valeur de l'information a un comportement de chat de Schrödinger si l'on considère l'information comme constituée d'un triplet comprenant :

- un objet informatique sous la forme de données, intervenant dans un calcul représentant un traitement résultant d'une action de l'utilisateur,
- un programme produisant des formes de présentation de cet objet qui sont autant de signes conditionnant l'interprétation par un humain dans son contexte,
- les formes correspondant à la tangibilité de l'information pour l'utilisateur, en relation avec leur décidabilité pour la machine.

Ainsi, le sens ne sera pas décodé par l'humain s'il n'y a pas un programme logiciel (l'instrument d'observation et de mesure) capable de construire une présentation de l'information affordante pour cet humain en fonction de ses intentions, buts, capacités et connaissances.

Dans ce cas, le principe de conformité de la communication d'information est transgressé, et le mécanisme de transgression en est ici, la technique d'observation et de mesure qui conduit à une indécidabilité de calcul de la valeur d'une information.

Cette hypothèse justifie que l'on se penche sur l'évolution des modèles théoriques de l'information afin de voir si l'on ne peut pas y trouver de solution acceptables permettant de lever cette indécidabilité. Cette évolution est marquée par trois grands jalons : mathématique, analogie thermodynamique et assimilation à la physique.

*La théorie mathématique de l'information et de la communication* (Shannon) qui a rendu l'information calculable (en termes de codage et d'entropie), donc descriptible pour la coder sous la forme de signaux perturbant le canal de communication et pour décoder ces signaux du côté du destinataire afin que ce dernier reconstruise



l'information. Cette théorie ramène le problème de la communication à un problème de décision.

*La théorie thermodynamique de l'information* (Brillouin) qui, appliquée au web, nous permet de définir un état entropique qui peut évoluer dans le temps : l'analogie avec la thermodynamique provient du fait que l'information ajoutée à un système agit comme un "démon de Maxwell", c'est-à-dire, qu'elle en fait baisser l'entropie. Ainsi l'information est dénommée par Brillouin "néguentropie". En physique, une baisse d'entropie viole le second principe de la thermodynamique, mais Maxwell a imaginé une expérience de pensée mettant en évidence cette situation paradoxale. Cette condition a été dénommée par la suite "le paradoxe du Démon de Maxwell".



Fig. 8 : Illustration du "paradoxe du Démon de Maxwell" qui permet d'expliquer la transgression apparente de la deuxième loi de la thermodynamique (l'entropie d'un système thermodynamique ne peut qu'augmenter).

On se représente un système thermodynamique cloisonné en deux compartiments à l'interface desquels se trouve un "observateur" qui décide de laisser passer une particule d'un compartiment à l'autre en fonction de l'état mesuré de celle-ci. La conséquence de cette décision prise par l'observateur est l'observation "apparente" d'une baisse de l'entropie dans le premier compartiment.

La vitesse des particules d'un gaz (traduisant l'entropie thermodynamique du système) d'un des deux compartiments est mesurée par un observateur (le "Démon") qui, en fonction de sa valeur, agit sur une vanne laissant passer la particule observée. Cette valeur de vitesse est une information observée (une présentation de l'information) à un instant donné et son interprétation conduit à une action (l'ouverture de la vanne par le "Démon"). Cette information induit une action qui fait baisser l'entropie du compartiment où la température est la plus élevée. Ce paradoxe représente un bon modèle pour comprendre la relation entre :

- la présentation de l'information : les moyens d'observer celle-ci utilisé par le "Démon",
- la prise de décision : la "vanne" du système représente les instruments d'une interface permettant d'instrumenter la prise de décision en une action,
- la conséquence de l'action : l'information mesurée qui est ajoutée au système dans sa globalité et qui a été appelée par Brillouin un "facteur de néguentropie".

La mesure faite par l'observateur est précisément de l'information conduisant à une prise de décision et c'est la détermination de cette valeur qui conditionne une action

dont le résultat fait baisser l'entropie du premier compartiment, d'où l'appellation de néguentropie pour qualifier l'information.

Par rapport à notre problématique, l'utilisateur avec ses moyens techniques d'observation de l'information joue le rôle du "Démon de Maxwell" (Fig. 8) et c'est en fonction de sa capacité prendre une décision d'action à partir d'un état observé et mesuré de l'information qu'il pourra organiser un système d'informations.

Il faut noter que cette théorie de l'information voyait la thermodynamique comme une analogie et ne considérait pas encore l'information comme un phénomène physique à part entière.

*Les théories modernes de l'information* ont franchi le cap d'assimiler l'information à un phénomène physique en ajoutant la dimension d'énergie. Produire une information ou une présentation de l'information nécessite un calcul auquel est associé un bilan énergétique. Landauer a montré que l'effacement définitif d'un bit, c'est-à-dire, une suppression d'information (comme par exemple, un traitement par une porte logique "ET"), conduisait à une dépense d'énergie mesurable et dissipée en chaleur. Bennett a mis en évidence des situations de calculs irréversibles ou réversibles (qui, en mécanique quantique, correspondent à la notion de "flèche du temps"). Pour expliquer ceci, il faut considérer un système au travers de ses variables d'états. Un calcul revient à porter ce système d'un état initial à un état final.

Un calcul irréversible ne permet pas de retrouver un état initial unique à partir d'un état final du fait d'une perte d'information (d'une augmentation de l'entropie du système). Ceci laisse entendre qu'un calcul irréversible conduit, sur le plan statistique, à des états superposés du système, c'est-à-dire, qu'à un instant donné, une information a une équiprobabilité de valeurs possibles (chat de Schrödinger).

Un calcul réversible, quant à lui, permet à tout moment de revenir à l'état initial et donc ne dépend plus de la flèche du temps, parce qu'il n'y a pas de perte d'information lors du traitement. Cette hypothèse a conduit à la définition de nouvelles architectures de portes logiques (par exemple les portes de Tofoli). Elle est aussi à la base de la cryptographie quantique.

Pour mieux comprendre ces états réversibles ou irréversibles, Charles Bennett utilise l'image d'un cadenas pourvu d'une clé ou d'un cadenas dépourvu d'une clé pour enchaîner un objet quelconque.



*Fig. 9 : Illustration de la réversibilité et de l'irréversibilité d'un calcul selon Bennett. Le cadenas de gauche permet de revenir à l'état initial du système (objet non enchaîné), alors que le cadenas de droite (en forme de cœur) ne le permet pas. Ici la clé du cadenas correspond à l'information perdue dans le cas d'un processus de calcul irréversible.*

Cette théorie quantique de l'information met en avant le principe de causalité qui est précisément celui que nous avons décrit en parlant des relations entre un espace de formes mesurables et un espace de décision correspondant à l'alphabet et au vocabulaire d'une application.

Le seul élément déterministe est précisément l'alphabet et le vocabulaire de l'application du destinataire, les formes perçues par ce destinataire, ont aujourd'hui un caractère aléatoire certain. Dans un tel environnement, la question du maintien de la conformité de l'information devient délicate car elle fait intervenir la notion de contexte d'un triplet formé par l'utilisateur lui-même, son environnement incluant sa situation, ses matériels et logiciels permettant de construire des présentations de l'information, et l'information elle-même

Ces trois grands jalons ne se remettent pas en cause l'un l'autre, mais correspondent plutôt à une solution de continuité, une théorie correspondant à la généralisation de la précédente, à l'instar de ce que l'on a observé en physique où la mécanique newtonienne n'a été remise en cause par la mécanique relativiste que par les limites des modèles (masse, énergie et vitesse).

Dans nos applications informatiques et nos interfaces, nous avons besoin de revenir sur les constituants d'un canal de communication :

- la saisie d'informations par des dispositifs autres que le clavier est l'exemple le plus représentatif d'une évolution du problème de décision posé par l'analyse et l'interprétation d'une information. En effet, le signal produit par ces dispositifs (interfaces tactiles, vocales, etc.) n'est pas univoque, contrairement à celui produit par un clavier.
- la présentation d'une information multimédia et multimodale nécessite la connaissance du contexte de production du contenu (une des variables d'état du

système). Cela concerne la description syntaxique de toutes les formes inscrites dans un document : les formes libres, les relations entre formes, les fonctions des formes et leurs valeurs symboliques.

Ces descriptions syntaxiques concourent, mais ne la garantissent pas, à la conformité de l'information communiquée au travers d'un canal de communication. Cette conformité n'est établie que si les modes de présentation de l'information sont soit partagés, soit négociés, par intermédiation entre une source et un destinataire.

Ainsi, nous nous trouvons confrontés désormais à une complexité des systèmes d'information, quelle que soit l'étendue de ceux-ci (information domestique, un intranet d'entreprise et internet), qui ne porte plus seulement sur des questions de métadonnées, mais qui s'est étendue à des questions de construction de la présentation d'une information.

La construction d'une présentation de l'information peut être soit mentale par sémiose, soit computationnelle par un calcul plus ou moins complexe.

La question de convergence des usages de l'information numérique qui induit celle de son accessibilité, correspond à la faculté de communiquer une information conforme, quels que soient le matériel utilisé, le média support et les moyens d'interaction. Ceci revient à établir des modes de présentation de l'information qui font sens pour le destinataire.

Elle ne peut plus se limiter à des questions d'implémentation technique ou de standardisation des contenus, donc de maîtrise des syntaxes, parce que le destinataire choisit ses propres outils d'observation (terminaux, logiciels et interfaces). Le calcul, et notamment sa complexité permettant d'aboutir à une présentation de l'information, est une donnée tout aussi importante que le choix d'une syntaxe véhiculaire.

Les critères orientant les choix techniques faits par le destinataire ne sont plus seulement établis en fonction des préconisations des sources d'information, mais dépendent de ses aptitudes, moyens, cultures et connaissances. Or ce sont précisément ces paramètres, non quantifiables par la technique qui vont conditionner cette complexité de calcul.

Nous verrons que cette complexité est parfaitement quantifiable et donc, permet d'établir des seuils de complexité acceptable ou non pour la production d'une présentation de l'information qui fait sens pour le destinataire.

Limiter la problématique de la présentation de l'information au choix d'une syntaxe véhiculaire revient à oublier la connaissance ayant conduit au programme logiciel permettant de construire cette présentation de l'information. Nous pouvons affirmer que si les données informatiques et les formes d'une interface sont la présentation de l'information, le calcul et sa complexité sont la présentation de la connaissance.

Nous illustrerons ceci au travers de cette thèse et, en particulier dans les premiers cas d'application que nous avons maquetés à titre d'illustration.

Dans les modes de construction d'une présentation de l'information, nous trouvons, bien entendu, la numérisation de contenus analogiques. C'est précisément l'étude et la conception de plateformes de numérisation qui nous ont conduits à émettre ces hypothèses.

## *IV. Au croisement du multimédia et de la multimodalité, la présentation de l'information*

### **IV.1. - Document numérique et interaction**

Toute action d'une interaction homme machine porte sur une forme que l'utilisateur pointe avant d'effectuer un traitement. Cette forme peut soit représenter un objet précis du contenu ou le document lui-même, soit correspondre à un fragment arbitraire d'un contenu.

Dans un document textuel, l'information de segmentation permet en soi, de pointer n'importe quelle partie du texte en sélectionnant la suite de caractères voulue par l'utilisateur. On s'aperçoit que cette opération peut devenir difficile, voire impossible dans des cas de figures de plus en plus fréquents ; tout document dont la composition ne correspond pas avec les intentions d'usage du destinataire.

Les formes ne se limitent plus seulement aux caractères. Tout document composé a aujourd'hui une structure qui conditionne directement son accessibilité. Cette structure de présentation est un ensemble de formes codées dans un descripteur tout comme les caractères. Dans les autres objets structurants d'un texte, on trouve une répartition topologique en blocs de textes qui vont représenter, après déchiffrement par l'utilisateur, des paragraphes, des titres, ainsi que tout élément de la structure logique d'un document.

Or un document ne se limite pas aujourd'hui à un texte. C'est une mise en forme selon des conventions données par une charte de présentation, traduite par un style applicable à chacun des objets qui le composent.

Un document est devenu aujourd'hui un objet largement composite, dans lequel nous trouvons des objets graphiques, donc des représentations vectorielles, mais aussi des images, donc des tableaux de pixels, pour lesquels, l'information de segmentation peut devenir très complexe à détecter ou à calculer.

Ainsi, on s'aperçoit que le document numérique a acquis une dimension multimédia qui en fait un objet complexe à manipuler, étant donné que chacune de ses composantes a une mise en forme qui lui est propre, dictée par les propriétés de la technique ayant permis sa production. Aujourd'hui la technique tend encore à prendre davantage en compte les contraintes de la source d'information que celles du destinataire.

Cette pratique conduit à présenter l'information d'une façon fortement hétérogène, ce qui en diminue sa visibilité par le destinataire, que ce dernier soit humain (un utilisateur d'une interface) ou artificiel (un programme logiciel d'indexation).

Si l'on se place du côté du destinataire, nous observons plusieurs filtres qui masquent cette hétérogénéité en masquant de l'information.

Tout d'abord, seuls les documents indexables par un logiciel sont visibles au travers d'un moteur de recherche.

Ensuite, les codages de documents font l'objet d'une normalisation forte qui permet de tenter d'uniformiser les syntaxes de codage et donc, de donner une visibilité étendue aux contenus.

### ***IV.1.1. - Coder de l'information : matérialiser des concepts***

Par la suite nous nous pencherons sur la partie "visible et matérielle" de l'information, à savoir le codage de celle-ci en données et nous dégagerons les principales propriétés des codes en matière d'accessibilité et de tangibilité des contenus.

Cette complexité renforcée par la large diffusion des outils de production de contenus fait que l'on doit prendre en compte, non seulement la diversité des formats et des jeux de métadonnées, mais également, une diversité sémiotique de présentation de l'information ; il faut rajouter la diversité qui se manifeste aussi par la visualisation d'un texte selon une grande variété de polices de caractères, d'attributs de mise en forme.

Ceci nous conduira à tenter de définir les critères de tangibilité des contenus, et à mettre en relief le lien étroit existant entre tangibilité et documentation du contenu.

Ce terme de documentation ("documentarisation" chez Roger T. Pédaque [PED07] pour mieux refléter l'idée d'action dynamique par rapport à un état statique) correspond à l'ajout de descripteurs relatifs à toutes les formes présentes dans un contenu ; la forme étant ici entendue dans une acception générale de "patron" ("pattern") ou de "gabarit", résultant d'une opération de segmentation, et non plus en tant qu'objet matériel repéré a priori. L'information de documentation, quant à elle, est l'instrumentalisation de la relation effective entre une forme et sa signification dans une application.

En effet, n'importe quelle opération de segmentation conduit à la production de formes, mais ces dernières n'auront de validité pour une application que si leurs descripteurs font sens par rapport à l'interaction : un texte ne devrait contenir que des éléments textuels, une route sur une carte ne devrait contenir que les éléments topologiques prévus dans le modèle de l'objet "route" de l'application, etc.

On met en avant ici une difficulté inhérente à tout système de numérisation, à savoir, la pertinence des résultats obtenus dans n'importe quelle chaîne de traitement. On retrouve ces difficultés au niveau de la saisie d'information faite par l'utilisateur, notamment lors de la saisie d'un geste (pour écrire ou commander l'application).

Ces difficultés sont directement liées à la capacité de discriminer le signal du bruit. Ceci peut être illustré par les deux exemples suivants :

- Le résultat d'une segmentation dans une image conduit à fragmenter celle-ci et un fragment peut être saillant pour l'utilisateur (la forme d'un caractère par exemple) ou la machine (le descripteur de la forme peut être classifié). Ce fragment ne peut être saillant s'il ne fait référence à aucune symbologie "connue" de l'utilisateur ou de la machine.
- Un geste effectué par un utilisateur ne fait sens, dans une interface, que si le programme de l'application arrive à détecter, segmenter et interpréter ce geste par rapport à des traces de références apprises. Ces traces de référence sont des signes ou messages émis par l'utilisateur et leur sémantique dépend étroitement de l'application. La symbolisation d'un geste revient donc à rapprocher un geste reconnu à une action ou une donnée dans l'application : un geste dont la géométrie est un cercle peut être polysémique et correspondre, par exemple, à l'intention de tracer la lettre "O", d'effectuer un zoom sur une image, etc.

#### ***IV.1.2. - Le document numérique structure l'interaction***

Le maître mot est ici la symbologie qui nous renvoie à une convention de code partagée entre la source et le destinataire. Ce partage de convention peut être connu de manière a priori (l'écriture en est l'archétype) ou négocié (de manière consciente ou inconsciente) lors de l'acte de communication. Ce partage du code porte sur :

- ***Le niveau physique*** ou la capacité de stimuler le récepteur. L'humain "voit" une image, "entend" un son, "ressent" une vibration, alors que la machine "perçoit" un flux d'octets contenant toute l'information codée selon différents modes ; à chacun des modes correspondent des formes inscriptibles sur un support selon des spécifications de format (pour l'image, ce sera, par exemple, un codage entrelacé de cette dernière). En l'absence de ces informations relatives à la forme physique de l'information, le récepteur est incapable de percevoir le contenu.
- ***Le niveau sémiotique*** ou la capacité pour le récepteur à discriminer les signaux porteurs d'information. A titre d'exemple, l'information saillante permettant de repérer les caractères d'un texte dans une image peut être spectrale (caractères noirs sur fond blanc, par exemple) ou spatiale quand elle décrit la géométrie de ces formes (taille, proportions, etc.) ou la distance entre ces formes permettant d'identifier les mots, les paragraphes, etc. Cette capacité de l'humain de discriminer ces formes a une transposition dans la machine, sa capacité de segmenter le contenu à partir de mesures faites sur le flux d'octets. Le résultat de cette segmentation est restitué sous la forme de signes tangibles pour l'utilisateur à la condition que ces formes soient reliées à des données décidables pour la machine.
- ***Le niveau sémantique*** ou la capacité pour le récepteur de déchiffrer le sens de ces signes. Pour l'humain, il s'agira de comprendre le contenu en reconstruisant l'information émise par la source et pour la machine, d'affecter la valeur symbolique de l'information convenue (telle forme est le caractère "A" qui a telle valeur en ASCII et qui est représenté en Garamond 12). En termes d'interaction homme machine, la reconstruction du sens global du contenu se fait par un ensemble d'itérations entre le contenu et l'utilisateur en faisant appel au code sémiotique (information spectrale, morphologique et spatiale), pour retrouver, par exemple, les mots d'un texte et le sens global du texte. Du côté de la machine, le sens (si tant est que l'on peut parler de sens ici), est donné par la capacité d'un programme de décoder ce flux de données. On peut aisément constater que cette capacité sera directement inféodée à la nature symbolique de l'information : la symbolique d'un texte codé en ASCII est "calculable" directement ; celle d'un texte dans un document image ne l'est pas et doit faire l'objet de traitements pour reconstruire le code sémiotique de présentation de l'information.

Nous verrons, en explorant les différentes formes de codage, les conséquences de ces dernières en termes d'interaction et tenterons de déduire un schéma conceptuel d'interface capable de prendre en considération une symbologie explicite d'un contenu (celle qui est directement codée dans le document), par rapport à une symbologie latente (celle qui devra faire l'objet d'un calcul dans le document).

L'ensemble des codes (physiques, sémiotiques et sémantiques) relatifs à la capacité de reconstruire le sens d'un contenu à partir de signaux définit une modalité de présentation de l'information. Si nous partons du destinataire d'une information, nous devons prendre en considération l'ensemble des codes et instruments qu'il

utilise. Cet ensemble constitue ce que l'on appellera une **modalité préférentielle de présentation de l'information**.

#### ***IV.1.3. - Codage exprimé et codage explicite de l'information***

Par rapport à la capacité donnée à tous les nœuds d'un réseau de coder de l'information dans un formalisme de plus en plus libre, nous observons une contradiction de plus en plus flagrante qui se répercute directement sur l'accessibilité de l'information : pour la plupart des contenus textuels, chaque mot est potentiellement un lien hypertexte vers de l'information et donc, un document textuel est naturellement une interface de navigation pour peu que les codes physiques, sémiotiques et sémantiques sont explicites pour le destinataire, alors que pour la plupart des contenus multimédia, la navigation s'arrête à l'affichage du document, du fait que son codage de présentation est latent en ce sens qu'il résulte de l'expression d'un code ou de la mise en œuvre du système de codage.

Enfin, si ces codes physiques, sémiotiques et sémantiques ne sont pas explicites pour le destinataire, on se retrouve exactement dans la même situation que pour des documents multimédias. C'est probablement cette mésestimation de l'adéquation entre la présentation de l'information et les conditions d'accessibilité pour le destinataire qui sont l'une des causes de l'échec patent du web sémantique.

#### ***IV.1.4. - Convergence numérique et co-alignement modal***

La convergence de l'information numérique, vue du destinataire de l'information, nécessite de pouvoir aligner les différentes modalités de présentation de l'information par rapport à ses modalités préférentielles ; ceci nous conduira à définir une couche spécifique et encore embryonnaire dans les interfaces, que nous appellerons "couche d'intermédiation" ou "couche de redocumentation" ; elle aura pour objectif, d'une part, de valider la présentation d'un contenu par rapport à son accessibilité par le destinataire et, d'autre part, de procéder à des reformulations négociées entre le destinataire et la machine lesquelles permettront d'arriver aux conditions de validation d'accessibilité du contenu.

Ces reformulations correspondent ici à des co-alignements modaux, consistant, si besoin est, à convertir le format physique, à ressementer le contenu pour obtenir un codage sémiotique conforme à l'accessibilité et à restructurer (ou même redéfinir) les métadonnées par rapport aux connaissances de l'utilisateur.

Ici se pose clairement la question de la validation de ces reformulations ou transcodage de l'information ; nous proposerons une méthode de validation indirecte de cette reformulation, se basant non pas sur la performance algorithmique pure, mais sur la définition de critères de tangibilité d'un segment d'information. Cette approche est similaire à celles rencontrée dans les systèmes complexes (comme ceux à tolérance de pannes), elle propose une validation plus globale du processus de reformulation au lieu de s'intéresser à la performance algorithmique pure de chacun des composants.

Un système de reconnaissance de caractères (OCR) que nous avons implémenté, qui est capable de retrouver des mots sans avoir à en reconnaître tous les caractères et dont le fonctionnement est assisté par l'utilisateur, en représente un bon exemple d'étude.



#### ***IV.1.5. - Numérisation, présentation de l'information et accessibilité***

A ce stade du développement, il apparaîtra nécessaire de revenir sur la chaîne globale de traitement d'une information numérique en partant de la numérisation même et de réaffirmer les contraintes d'ordre physique qu'impose l'acte de numériser un contenu, que ce dernier soit un flux audiovisuel ou un document statique.

Cette chaîne globale conduit systématiquement à la production de documents numériques eux-mêmes porteurs de contenus. Ces documents, les porteurs des codes physiques, sémiotiques et sémantiques, sont les interfaces naturelles d'accès à l'information et toute requête sur le web retourne toujours un ensemble de documents numériques. Ce sont toutes les informations et méta-informations associées aux contenus qui en donnent leur accessibilité (visibilité et tangibilité).

Ainsi, la documentation d'un contenu devient l'élément structurant de la complexité de présentation de l'information et donc, sa perception par un utilisateur correspondrait au démon de Maxwell décrit dans notre survol des théories de l'information.

Le fait que la documentation d'un contenu agisse comme élément structurant ne veut pas dire qu'elle donne une accessibilité généralisée à l'information. En effet, le document numérique peut être vu aujourd'hui comme un canal de communication porteur de signaux multiplexés, supportant chacun d'une modalité de présentation de l'information.

Le document représente donc le l'interface naturelle d'accès à l'information. C'est lui qui porte les informations de visibilité et de tangibilité des contenus. La forme classique du document en fait un objet conçu à partir des spécifications et contraintes propres à la source d'information. Elle impose son accessibilité par le destinataire.

Si nous ajoutons une dimension dynamique à ce codage de documentation en y incluant les propriétés du destinataire, le document devient un "intermédiaireur" rendant possible l'interopérabilité de l'information (en considérant la source et l'émetteur comme des sous-systèmes d'informations interconnectés ouverts).

Aujourd'hui, cette intermédiation existe déjà dans les sites web à base de pages dynamiques (sous PHP, ou bien, au travers de mécanismes de transcodage en image à la volée de données cartographiques, par exemple)

Cette fonction d'intermédiation existe également déjà, à l'état embryonnaire, au sein des navigateurs de certains terminaux mobiles. Il s'agit des navigateurs SAFARI (présent dans le iPhone d'Apple) et DEEPFISH (de chez Microsoft). Ces navigateurs proposent un mode d'affichage d'un document passant par une image codée en JPEG2000 [TAU02] de manière à bénéficier des performances excellentes en termes de multi résolution, de tuilage de l'image et de la gestion de son transport ; ceci conduit à une navigation visuelle grandement facilitée (zooming, panning).

Cette couche de visualisation peut être vue comme une couche d'intermédiation de la présentation de l'information portée par un document, en ce sens que, en plus des attributs visuels, certaines informations de navigation sont transposées, les liens hypertextes en particulier. Cependant, ceux-ci doivent être explicitement codés dans le document source.

Toutefois, en termes d'accessibilité, cette fonction d'intermédiation est fortement limitative car elle impose au destinataire sa modalité préférentielle. En l'occurrence, seule la modalité visuelle est présente pour l'interaction, selon un codage sémiotique classique n'autorisant qu'un pointage d'un segment d'information proprement formaté a priori.

Par cet exemple, on voit à nouveau l'importance de la présentation de l'information dans la définition d'une interface acceptable pour une famille usage et dont les contenus sont accessibles et tangibles.

Corollairement, c'est la conception même de l'interface qui va initier l'interaction homme machine : l'utilisateur ne peut saisir de l'information qu'à partir du moment où il perçoit que la machine, en face de lui, est en état de recevoir quelque chose. L'efficacité d'une interaction homme machine est fortement tributaire de l'information présentée à l'interface.

Cette matérialité des conditions de l'interaction stimule l'utilisateur et correspond au codage sémiotique de la présentation de l'information sous la forme de données décrivant à la fois l'information recherchée et celle relatant l'état du système : il n'y a pas d'interaction possible sans stimulation préalable de l'utilisateur. L'IHM ne traite pas l'autopoïèse ou l'énaction mais ne fait que capter des comportements d'un utilisateur stimulé par des présentations de l'information.

## **IV.2. - Vers un modèle adaptatif de l'interaction**

L'observation des propriétés du web en matière de codage hétérogène de l'information nous confronte à une difficulté dans la définition de services liés à la convergence de l'information numérique.

Nous avons vu que cette convergence est de plus en plus difficilement obtenue par la normalisation des formats d'échanges, du fait de la cohabitation de standards dans une logique de concurrence quasi darwinienne : la prépondérance d'une norme par rapport à une autre est donnée par les propriétés de l'écosystème que représente l'ensemble des systèmes d'informations interconnectés du web.

Il n'y a donc pas seulement une concurrence frontale et directe entre standards, mais plutôt un contexte d'environnement donné par les propriétés à un instant donné, des réseaux, des moyens informatiques et des usages ; elles agissent sur la sélection des standards, au même titre que l'environnement naturel agit sur la sélection des espèces.

Si ces conditions de convergence de l'information numérique ne se situent pas préférentiellement dans les actions de normalisation, il faut trouver l'élément structurant venant en substitution de cet état de fait. Ceci nous conduit à émettre cinq hypothèses

- **L'interface comme moyen d'intermédiation entre systèmes d'informations** : la première hypothèse que l'on peut émettre est que, si nous sommes confrontés à des sources codantes d'informations hétérogènes, il devient indispensable de concevoir des interfaces de présentation de l'information capables d'assurer la communication entre systèmes. Le rôle de ces interfaces de présentation de l'information est d'assurer un transcodage restaurant la conformité de communication de l'information entre sources et destinataires.

- **Prise en compte du glissement vers le destinataire des domaines de la médiation** : l'hétérogénéité de codage de l'information n'est observable que du côté du destinataire, il apparaît donc que le pilotage des fonctions de transcodage de la présentation de l'information doit être le fait de ce dernier.

Or ce destinataire est également un système d'information (humain ou artificiel) qui est doté de propriétés perceptives physiques, sémiotiques et sémantiques individuelles. Par conséquent, ces fonctions de transcodage doivent être relativisées par rapport aux propriétés du destinataire et non plus par rapport à celles des sources d'informations.

Ceci va dans le même sens que le glissement des domaines de la médiation que nous observons dans la dynamique de la communication numérique : on ne télécharge plus un document fabriqué par une source, mais on télécharge les codes permettant de reconstruire un document dans un environnement technique.

- **Identification des modalités préférentielles d'interaction du destinataire.**

Les propriétés perceptives du destinataire induisent la notion de modalités préférentielles de présentation de l'information (pour ce destinataire). Ces modalités préférentielles sont elles mêmes contextuelles, en ce sens qu'elles sont obérées par le contexte d'usage (capacités de perception, formalisation des intentions et instruments disponibles pour l'interaction).

Ainsi, la définition a priori des propriétés physiques, sémiotiques et sémantiques d'une modalité préférentielle est difficile à mettre en œuvre ; par contre, elle peut être déduite de l'analyse d'activité du destinataire et de l'apprentissage des signes saillants résultant de cette activité. Une action de correction d'une valeur erronée produite par une reconnaissance de caractères entre bien dans le cadre de cette analyse de l'activité et, par extension, l'évaluation de toutes les méthodes de transcodage de l'information.

- **Reformulation sémiotique du contenu et des actions du destinataire.**

Le destinataire émet des informations dans un mode de présentation donné qui doit pouvoir être également perçu par l'interface de transcodage.

Nous déduisons que la présentation de l'information émise par une source est conforme aux propriétés du destinataire si nous obtenons les conditions de co-alignement modal entre ce destinataire et cette source. Au-delà de la conversion de format, ceci se traduira, premier exemple, par une re-segmentation d'un contenu de manière à le rendre tangible, ce qui signifie également, donner une évidence sémiotique à un fragment de contenu pour le destinataire.

Un deuxième exemple peut être donné par la nécessité de reconnaître toutes les formes d'un document pour en faire une présentation vocale qui ne se limite pas seulement à une synthèse vocale linéaire du texte. La prosodie devra s'appuyer sur la sémiotique de structure d'un document pour permettre à l'utilisateur de segmenter le discours de la machine, et faciliter l'interprétation de ce discours.

- **Intermédiation négociée entre la source et le destinataire**

De manière à intégrer toute la chaîne de communication dans un service; il est nécessaire d'inclure une négociation entre une source et le destinataire portant à la fois sur la présentation et les droits d'usages de l'information.

Au même titre que l'acquisition d'un texte imprimé dans un livre donne à ce dernier une légitimité par l'imprimatur faite par l'éditeur, la légitimité d'un contenu numérique ne saurait être donnée que par une marque sémiotique perçue lors de la reconstruction de l'information par le destinataire.

Cette marque sémiotique est une forme présente dans un contenu, au même titre qu'un texte ou une image et les métadonnées liées à cette forme pourront contenir les droits d'utilisation et de transcodage.

Si l'on reprend globalement ces hypothèses, on s'aperçoit que l'on se place dans un contexte identique à celui qui a conduit à la normalisation de l'interconnexion de systèmes ouverts dans les réseaux.

Cette normalisation a porté non pas sur la recherche d'un mode unique de codage en signaux, mais plutôt sur la définition de protocoles de conversion de l'information dont la structure normative est donnée par les couches hiérarchisée du modèle ISO de l'OSI [GAR04].

Une telle approche revient à déterminer une interface entre deux extrémités d'un canal de communication. Il s'agit d'une interface de transcodage de la présentation de l'information. Une telle approche est applicable si l'on considère que l'adaptation de la présentation de l'information revient à articuler entre eux deux canaux de communication symétriques :

- celui partant d'une source vers le destinataire (que l'on pourra appeler "canal de restitution"),
- celui partant du destinataire vers une source (que l'on pourra appeler "canal de requête").

### **IV.3. - Des maquettes révélant des potentialités techniques**

Nous présenterons au chapitre 6, à titre d'illustration, trois cas d'applications dont la conception a pris en compte le modèle théorique de traitement sémiotique de l'information développé dans le cadre de cette thèse.

#### ***IV.3.1. - Un OCR intuitif***

L'approche sémiotique permet de construire un OCR robuste, avec un apprentissage incrémental compensant par un apprentissage a priori minimal. Le co-alignement modal réside dans l'hybridation d'une reconnaissance de caractères et d'une reconnaissance de tracés en ligne se substituant à la saisie au clavier.

Ceci permet, par exemple, de revenir sur la segmentation dans l'image de manière extrêmement simple pour l'utilisateur. On rapprochera cet OCR des travaux de Jie Zou et George Nagy [ZOU06]. Il s'agira aussi de mettre en avant le fait que ce co-alignement induit un effet Gestalt réellement observable. En effet, les performances globales de cet OCR (robustesse par rapport à la qualité de l'image et la variabilité des caractères) sont sans rapport avec les performances individuelles de chaque algorithme le constituant.

Enfin, en matière d'interface, on se place dans un paradigme proche de la machine pilotée par un utilisateur qui permet de ne pas interrompre le dialogue homme machine en cas d'erreurs de communication : une erreur de communication fait partie intégrante du dialogue et donc, sa correction doit faire partie intégrante de ce

dernier par des actions de reformulation. La principale innovation réside ici dans le fait que la reformulation se fait au niveau sémiotique donné par la segmentation de l'image et non pas au niveau de l'étiquette sémantique de la forme (saisie de la valeur du caractère reconnu).

### ***IV.3.2. - La construction incrémentale de terminologies à partir de corpus documentaires***

Ici, nous sommes dans le cas d'une application qui travaille dans un mode d'ingénierie inverse de gestion de documents numériques. L'hypothèse de départ est que nous n'avons aucune connaissance a priori sur la terminologie utilisée dans la rédaction de document et l'application a pour objectifs de :

- construire cette terminologie de manière incrémentale,
- déduire des indices de qualification d'un document par analyse de l'activité des utilisateurs.

Si nous utilisons des méthodes issues du traitement automatique de la langue naturelle (lemmatisation, construction d'abrévés, calculs de cooccurrences), la phase d'affectation d'un terme dans une terminologie se fait au fil de l'eau et peut être remise en question à tout moment. Ici ce sont les résultats de traitement qui sont pris comme indices sémiotiques de présentation de l'information. L'utilisateur apporte d'autres indices sémiotiques qui, associés à ceux produits par la machine, conduisent au classement sémantique des termes et des documents.

### ***IV.3.3. - La lecture sur écran***

Il s'agit d'une étude dans laquelle notre modèle sémiotique de traitement de l'information met en évidence les lacunes des systèmes de lecture sur écran et les causes du relatif échec des e\_books. Nous démontrerons que, même dans un processus, a priori non interactif en matière de saisie de données, comme la lecture, nous trouvons :

- un couplage hapto-visuel venant guider la lecture correspondant à une appropriation spatio-temporelle du contenu par l'utilisateur radicalement différente de celle d'un contenu audio-vidéo.
- un traitement spécifique du balisage visuel d'un texte en fonction des modes de lecture (lecture séquentielle, "verticale", etc.)

A partir de ces propriétés, nous pouvons en déduire des architectures d'interfaces que nous commençons à trouver dans des logiciels de navigation avancés (SAFARI sur iPhone et DEEPFISH chez Microsoft). Le paradoxe de ces navigateurs est le passage systématique à l'image pour permettre une exploration visuelle intuitive. Ce passage à l'image rend les contenus intangibles, seul le contenant (l'image) le reste. La question devient ici de définir comment restaurer cette tangibilité des contenus.



## Chapitre 2

### De la numérisation...

#### ...à la présentation de l'information

*"La pierre angulaire de la méthode scientifique est le postulat de l'objectivité de la Nature. C'est-à-dire le refus systématique de considérer comme pouvant conduire à une connaissance "vraie" toute interprétation des phénomènes donnée en termes de causes finales, c'est-à-dire de "projet".*

**Jacques Monod** - "Le hasard et la nécessité"

La production de contenus, et donc, la numérisation de documents préexistants, nécessite la prise en compte des usages. Dans le chapitre précédent nous avons vu quelles étaient les changements profonds induits par la généralisation de l'information numérique en réseau. Ce chapitre montre, au travers de retours d'expériences, que plusieurs questions relatives à la qualité et à l'évaluation de contenus numérisés restaient non résolues. Ces questions, toutes reliées aux usages et interface de manipulation de contenus numérisés, sont à la base des mes travaux de recherche en tirant les leçons de différents projets industriels de numérisation.

Ce chapitre présente l'ensemble des projets antérieurs à cette thèse auxquels j'ai contribué ou dont j'ai été l'instigateur. Ces projets m'ont permis de construire progressivement, une vision de la problématique de présentation de l'information au travers des thèmes liés à la numérisation du document et à l'interaction homme-machine.

Cette vision peut paraître à contre-courant de ce que nous propose un état de l'art classique, mais elle constitue une forme de "tabula rasa" du sujet, en ce sens que ces projets sont partis d'une problématique de terrain où l'objectif était de trouver des solutions techniques viables.

Au fil des projets, cette recherche a montré que la conception d'un système de numérisation de documents en fonction des seuls critères propres à l'ingénierie industrielle classique ne permettait pas de rendre compte des usages de manière réellement satisfaisante.

**Pourquoi associer aussi étroitement numérisation, présentation de l'information et interaction homme machine ?**

*Parce que l'objectif de toute plateforme de numérisation est précisément de construire une présentation de l'information par la reconnaissance de formes issues d'une image numérisée.*

## ***I. Numériser des documents patrimoniaux et personnels***

L'information existe dans un document papier, elle est perceptible et interprétable par un lecteur humain. Dans sa forme d'image numérique, le document reste perceptible et interprétable par un humain, pour peu que ce dernier dispose de l'appareillage adéquat de visualisation. Mais cette perception ne rend pas compte de la faculté de manipuler ces objets en termes d'interaction homme machine. Cette dernière doit être associée à des données utilisables par un programme logiciel pour que les actions de l'utilisateur puissent être opérationnelles.

Un document restitué par une interface doit donc être vu sous son angle sémiotique. Les objets du document au sens informatiques du terme, sont lus par un programme logiciel dont l'objectif est de calculer une présentation de l'information. Cette présentation de l'information doit être perceptible et interprétable par un humain.

Si l'interprétation faite par l'humaine se limite à la lecture du contenu, l'objet "image" est suffisant. Mais si cette interprétation conduit à une interaction, le document doit contenir des objets faisant références aux formes inscrites dans le document et à leur tangibilité.

Ainsi, du codage technique du document dépendent les capacités d'interactions avec ce dernier. Ceci nous évoque ce que Bachimont [BAC04][BAC96] appelle "l'association du déterminisme technique a priori et l'arbitraire de l'interprétation a posteriori" en définissant le document composé d'unités techniques de manipulation et un mode d'interprétation conduisant à la notion d'unités sémiotiques de manipulation. C'est ce concept que nous développerons au chapitre suivant sur lequel nous nous baserons pour définir une approche systémique de l'interface homme machine.



## **I.1. - Les limites de la numérisation industrielle**

L'approche classique de l'ingénierie industrielle consiste à définir un processus de traitement pour produire un résultat prévisible et répétitif, à partir d'un matériau de départ.

Ici, le matériau de départ est un ensemble d'images issues de la numérisation de documents, le processus mis en œuvre est celui de la reconnaissance de formes.

Un processus de reconnaissance de formes conçu selon une approche industrielle a des propriétés techniques déduites d'un modèle de traitements. Ce modèle est formé d'un ensemble de règles applicables à un standard de composition du document. La robustesse de ce processus dépend d'une marge de tolérance entre un document et le document standard.

Ce processus de reconnaissance de forme doit avoir comme principale propriété l'exactitude, qui conduit à minimiser les fausses reconnaissances et produire du rejet en signalant les formes et documents non susceptibles d'être traités.

Ceci correspond à la définition de la qualité dans tout processus industriel de fabrication d'un produit.

Afin de minimiser les erreurs dues à de fausses reconnaissances, on tente d'améliorer le fonctionnement de chacune des étapes du processus. Ceci se traduit par un processus capable de traiter un maximum d'exceptions par rapport au document standard, avec pour conséquences des fonctions plus complexes sur le plan des algorithmes.

Inversement, on recherche à appliquer des contraintes à chacune des étapes du processus pour limiter les risques de fausse reconnaissance. Or ces contraintes sont données par la sélection des documents en fonction de leur proximité de présentation par rapport au document modèle.

### ***1.1.1. - Les défaillances de la numérisation industrielle***

Une telle approche pose des problèmes a priori difficilement solubles dans le cas de la numérisation industrielle classique des documents<sup>1</sup>, à plus d'un titre :

- le matériau initial n'est pas standard en raison de la variation de présentation dans les documents.
- les étapes d'un processus de reconnaissance de formes, ne sont pas indépendantes, mais fortement intriquées : une modification d'une valeur de paramètre agit sur le comportement global du processus et non sur le seul composant du processus impliquant ce paramètre. Cela nous donne l'intuition de cette intrication.

Ces dépendances sont subtiles car elles ne sont pas facilement traduisibles en fonctions mathématiques. Elles ne permettent pas de catégoriser de manière exacte une famille de paramètres pour la rapporter à une heuristique de traitement. Elles évoquent une intrication entre composants d'un système complexe.

---

<sup>1</sup> Voir en ANNEXE 1 un exemple de résultat de numérisation industrielle sur un ouvrage du XIX<sup>ème</sup> siècle

- le résultat d'une reconnaissance de formes n'est pas univoque. Il dépend du système d'informations auquel il est destiné. La seule vision du système d'information qu'a le processus de reconnaissance de formes est fournie par l'espace de décision correspondant aux valeurs possibles du classifieur. Ces données forment le contexte d'usage, elles ne peuvent pas être considérées isolément du reste de la chaîne de traitement.

Il faut voir ces causes de défaillance dans l'essence même de la conception d'une plateforme industrielle de numérisation, à savoir :

- la détermination de propriétés standards à partir du matériau traité
- l'étude du comportement isolé de chacun des composants d'un processus de reconnaissance de formes.

Les modèles issus de ce travail de conception sont réductionnistes. Ils ne rendent pas compte de l'effet Gestalt induit par l'intrication des composants dans un système qui, dans le cas de la reconnaissance de formes, a toutes les propriétés d'un système complexe que nous abordons au chapitre 3.

## ***II. Un témoignage de 10 ans sur la numérisation***

Pour corroborer cette hypothèse de la complexité comme objet d'étude dans le problème de la présentation de l'information, nous allons exposer les principales contributions et travaux antérieurs à la thèse. Ces travaux et projets reflètent tous un état de l'art qui était celui du moment où ils ont été réalisés entre 1997 et aujourd'hui.

Le premier projet concernait la rétroconversion de plans techniques décrivant l'infrastructure de réseaux de télécommunication. Il avait été abordé dans une logique d'ingénieur, en recherchant la meilleure combinaison possible de traitements assemblés séquentiellement.

Le corpus de plans, très hétérogène, avait mis en évidence la difficulté de l'analyse et de l'interprétation d'un plan technique par une machine, selon une approche purement ascendante, tel que l'on conçoit classiquement un processus industriel de numérisation.

Ceci m'a conduit par la suite, à aborder le sujet selon deux approches :

- Partir de l'utilisateur et de l'application finale pour tenter de mieux identifier les points critiques de la numérisation
- Explorer la problématique de la numérisation sous un angle systémique, c'est-à-dire, ne pas analyser indépendamment chaque composant d'un processus de reconnaissance de forme, mais observer ce processus global dans son contexte d'exécution.

Ces deux approches m'ont conduit à définir la plateforme "DocMining" qui fut un projet labellisé par le Réseau National des Technologies Logicielles (RNTL). Le résultat a été une plateforme logicielle répondant à plusieurs questions dures du traitement du document numérique, notamment, la prise en compte de l'hétérogénéité de présentation de l'information.

Corollairement, l'une des questions dures de l'analyse documentaire, reste celle de la place de l'utilisateur dans la boucle de traitement. Cette place ne saurait se limiter à un "simple problème technique" car elle fait intervenir des questions complexes

d'ergonomie et de définition des interfaces en réponse à ce que Bachimont [BAC04][BAC96] a appelé

*"L'association de l'arbitraire de l'interprétation et du déterminisme technique".*

C'est une reformulation du classique paradoxe de Sayre [SAY73] sur la reconnaissance de formes généralisée à tous contenus, à savoir

*"Segmenter pour reconnaître, reconnaître pour segmenter".*

La finalité même des interfaces homme machine est de communiquer et de présenter de l'information. La sémantique, les connaissances, ne sont jamais exprimées en tant que telles dans une interface. Elles émergent au travers de la boucle d'interaction qui induit une sémiose pour l'utilisateur et qui s'exprime au travers de la mesure de l'activité de l'utilisateur par la machine.

Ainsi, nous remplaçons la notion d'interface homme machine dans le contexte de la communication, ce qui nous permet de faire un point par rapport aux théories de l'information.

Ici, l'information est redéfinie selon une vision pourtant très classique. Il s'agit de formes ou signes, résultant d'un calcul de la machine à destination d'un utilisateur ; ces formes représentent soit un contenu soit des traces de l'activité de l'utilisateur.

Dans cette approche par la communication, on voit apparaître trois aspects importants :

- Le seul point commun de partage et de médiation de l'information entre un humain et une machine sont les formes présentées à l'interface qui constituent la présentation de l'information.
- Ces formes sont TOUJOURS le résultat d'un calcul, que ce soit celui fait à partir des données de l'application ou de la mesure de l'activité de l'utilisateur.
- Le concept de forme développé ici nous renvoie vers la notion de triade sémiotique, en ce sens qu'une forme n'est signifiante d'une information que si l'on peut l'interpréter, soit par le raisonnement fait par l'utilisateur, soit par l'exécution d'un programme logiciel.

En ergonomie, on parle d'affordance facilitant l'interprétation par un humain. Dans le cas de la machine, nous avons le pendant technique de l'affordance qui est la complexité de calcul permettant de construire la relation entre une forme et les objets qu'elle représente.

Ces hypothèses que nous développons dans le chapitre 3 sur la complexité, sont le résultat des réflexions qui ont guidé mes contributions au Réseau Thématique Pluridisciplinaire du CNRS sur le document numérique (RTP-DOC).

Le chapitre développé ici est une synthèse de l'ensemble des problèmes posés par le concept, apparemment banal de document numérique. Ces problèmes dépassent le cadre de l'ingénierie logicielle liée à la conception de plateformes technologiques ; les périmètres d'usages des documents et, par extension, de n'importe quelle présentation de l'information, évoluent profondément. Ceci nous oblige désormais à aborder le document numérique sous un angle résolument pluridisciplinaire en n'hésitant pas à revenir sur des concepts théoriques.

En premier lieu, nous aborderons ce témoignage sous l'angle de la théorie du document, il apportera un meilleur éclairage sur le couplage entre technique et usages du document.

### ***III. Le glissement des domaines de la médiation***

L'une des principales conclusions des travaux du réseau de recherche "RTP-DOC" du CNRS sur le document numérique [PED06][PED07][GAR04], a trait à la mutation brutale dans nos rapports, pratiques et usages d'une information devenue essentiellement numérique.

Pour mettre en évidence ces mutations ou ruptures, il est nécessaire de confronter les problèmes de la technique à ceux de l'ergonomie et des usages en posant ces deux questions :

- **Quel était le "statut technique" du document analogique et sur quelles propriétés sont venus s'adosser des réglementations et usages ?**
- **Quel est le "statut technique" actuel du document numérique et quelles sont les réglementations et usages qui ne peuvent plus s'adosser sur une propriété technique ?**

A travers ces deux questions, on voit apparaître une évolution de la forme même du document.

Un document analogique est le résultat de la mise en œuvre de tout un arsenal de processus complexes, difficilement diffusables en termes de savoir-faire, de métiers et de matériels concourant à sa fabrication. Dans une logique de production, c'est un produit entièrement manufacturé.

C'est ce produit que vient acquérir un utilisateur, et cet utilisateur n'a aucun accès à la technique ayant conduit à la technique de fabrication de ce produit.

Un document numérique n'a pas le même statut technique. C'est un objet immatériel, alors que le document analogique est inscrit de manière indélébile, sur un support matériel.

Il est reconstruit lors de son chargement dans une application. Les résultats de cette reconstruction sont restitués sur une interface qui vient se substituer au support physique du document analogique.

En ce sens, le document numérique n'est plus un produit «manufacturé» ; l'utilisateur fait l'acquisition d'un "kit de fabrication" du document, le fichier électronique correspondant, et reconstruit ce document au moyen de ses propres outils.

Nous voyons clairement apparaître ici ce glissement des domaines de la médiation. D'un document analogique dont le contenu, initialement intangible, en fixait les usages et permettait d'appliquer une réglementation, nous sommes passés à un document numérique dont le contenu devient adressable par une technique accessible à l'utilisateur.

#### **III.1. - Les ruptures induites par l'information numérique banalisée**

A ce glissement des domaines de la médiation, vient s'ajouter la banalisation des moyens de fabrication de documents.

La numérisation de contenus analogiques (documents, enregistrements audio, photos, films), propose aujourd'hui d'un arsenal technique qui tend, de plus en plus, à se banaliser.

N'importe quel microordinateur courant peut héberger une suite logicielle permettant de numériser nos "archives familiales" avec un niveau de qualité proche de celui des stations équipant des ateliers de numérisation industrielle.

Les principales différences qui existent aujourd'hui entre un atelier professionnel et une suite logicielle domestique résident dans

- la capacité de numériser un volume important de contenus dans un intervalle de temps réduit
- la possibilité de mettre en œuvre des dispositifs de numérisation spécifiques à un support donné, comme par exemple, la numérisation de livres anciens pour les bibliothèques numériques.
- la conception d'interfaces différentes, selon que l'application est destinée à un utilisateur expert ou "naïf". L'utilisateur expert saura exploiter une présentation de l'application et de l'information qu'elle traite, plus proche des données et des programmes, que ne saurait le faire un utilisateur "naïf". Inversement, une interface ayant un niveau de métaphore éloigné de la technique mais proche des usages, ne permettra pas de gérer facilement les exceptions de l'application [SAI01].

Les technologies induisent également des mutations au niveau des terminaux, avec l'apparition de nouvelles pratiques et de nouveaux besoins.

L'appareil photo numérique a supplanté l'appareil à film argentique, rendant la production d'images immédiate. Il se trouve décliné dans une version communicante avec les terminaux mobiles, avec des performances, rendant plausible à court terme son utilisation comme moyen de saisie d'informations contenues dans une image captée sur le vif.

Jamais la photographie n'a autant mérité le qualificatif d'instantané : avec l'argentique, l'instantanéité résidait dans la prise de vue, avec le numérique, on inclut la dimension d'instantanéité de communication.

Ainsi, une plaque de rue, un code barre, une couverture de livre dans un magasin, sont des objets porteurs d'informations susceptibles d'être utilisées dans la production de requêtes vers un moteur de recherche.

Nous commençons à trouver des logiciels d'OCR dans les mobiles et des projets existent, notamment à Orange Labs, mais aussi chez la plupart des opérateurs et fabricants de terminaux mobiles, dont l'objectif est de produire des "assistant vocaux" à destination des malvoyants, en de vocalisant un texte reconnu dans un document numérisé à la volée.

Nous avons là une importante source de production de contenus numériques dont l'exploitation dans des services et l'indexation passe par la reconnaissance des contenus.

Cependant, cette banalisation de la numérisation induit des changements de contextes radicaux par rapport à une chaîne de numérisation industrielle.

Si des logiciels de reconnaissance de formes sont utilisés en appui de cette nouvelle forme de numérisation, la robustesse des résultats devient critique et les critères d'évaluation de cette robustesse, difficiles à établir en raison de l'instabilité des contextes de numérisation :

- images de qualité disparate,
- diversité des objectifs d'usages des contenus, et donc, de l'indexation de ceux-ci,
- diversité technique des interfaces, applications et terminaux mis en œuvre.

Le fait d'accepter la diversité de présentation de l'information dans des contenus aussi hétérogènes que les contenus personnels ou capturés sur le vif pose de nouvelles contraintes par rapport aux plateformes industrielles classiques de numérisation. La résolution des problèmes posés par ce contexte offre des voies de recherche alternatives réellement prometteuses dans la définition d'un processus de numérisation et l'indispensable validation des résultats pilotée par l'usage des contenus ; validation dans laquelle reste à définir la place de l'utilisateur.

### **III.2. - Le renversement des privilèges des "consomm'acteurs" de documents**

Le néologisme de "consomm'acteur" de document introduit ici la conséquence de la mutation des domaines de la médiation. Elle se traduit par le fait que l'utilisateur ne peut plus être considéré comme un lecteur passif de contenus définis et préformatés par un serveur d'informations. L'utilisateur est à la fois consommateur et producteur d'informations ; il devient un acteur du cycle de vie des documents.

Le document numérique n'est pas un produit manufacturé. Il est reconstruit par l'utilisateur sur son terminal à l'aide de ses outils informatiques. Ceci fait que la légitimité du contenu est remise en question de manière permanente ; ce que l'on peut appeler le "privilège d'imprimatur"<sup>2</sup> de l'émetteur du document s'est considérablement dilué dans la chaîne de communication.

#### **III.2.1. - L'imprimatur**

Dans le droit Canon de la religion catholique, l'imprimatur signifie qu'un document peut être publié. Cet imprimatur est un ensemble de marques et de signes distinctifs, apposés au document sous la forme de cachets accompagnés de signature. Ce sont donc des marques distinctives d'un document qui en tracent sa légitimité pour une organisation.

Ces cachets peuvent être au nombre de trois, selon l'autorité religieuse qui a donné son aval :

- - "Imprimi potest" (il peut être imprimé)
- - "Nihil obstat" (rien ne s'y oppose)
- - "Imprimatur" (qu'il soit imprimé)

---

<sup>2</sup> Il y a une forte analogie entre les règles de l'édition documentaire générale et celle des écrits religieux par le droit Canon

Si cette validation porte sur la nature du contenu, l'imprimatur lui, est valide au travers de marques reconnaissables sur le document. C'est la présentation de cette information qui permet de donner au document sa valeur légale par rapport à une communauté d'intérêts.

Si nous étendons cette notion de marque à toutes les formes présentes dans un document, on arrive à la notion de charte de présentation qui marque l'identité de l'éditeur et certifie le document.

Ainsi, un billet de banque porte lui aussi des marques distinctives lui conférant sa valeur fiduciaire qui, si l'on se réfère à l'étymologie de cet adjectif, correspond à une valeur de confiance. Celle-ci sera d'autant plus forte que ces marques sont à la fois connues et repérables par le lecteur, tout en étant difficilement reproductibles ou falsifiables.

Toutes ces marques sont une forme d'imprimatur qui donne sa valeur de confiance au document et à son contenu en permettant d'identifier la source et de préserver la conformité de l'information portée par un document.

### ***III.2.2. - Un récepteur actif de documents***

Le problème de l'intangibilité de l'imprimatur est posé sur le document numérique dont les codes de fabrication sont ouverts.

L'exemple de l'imprimatur est intéressant, en ce sens que si l'utilisateur dispose des moyens techniques nécessaires permettant d'intervenir dans la présentation du document, il peut parfaitement en modifier sa valeur de confiance, tout comme l'information contenue dans ce dernier.

Ainsi, dans un contexte de communication par des documents numériques et, par extension, de n'importe quel contenu numérique, le lecteur destinataire de l'information ne peut plus être vu comme un élément passif dans une chaîne de communication. Il dispose d'un ensemble de matériels (terminaux, ordinateurs, etc.), de logiciels et de connaissances qu'il met en œuvre pour reconstruire le document lors de sa réception. Cet ensemble de moyens définit une "infosphères" [GAR04] .

En revenant sur le modèle de la communication de Shannon, une infosphère représente les équipements d'extrémités d'un canal : les signaux arrivent à l'interface de cette infosphère, ils sont transformés en messages perceptibles au travers de l'interface utilisateur d'un logiciel, puis l'utilisateur, par sémiose (c'est-à-dire, l'interaction homme machine), reconstruit le sens à partir des messages.

### **III.3. - Conséquences sur la présentation de l'information**

Le rôle d'un canal de communication, tel que le voyaient Shannon et Weaver [SHA48], était d'assurer la mise en œuvre de moyens pour garantir la conformité de communication de signaux codés, représentant des messages. Cette conformité s'exprime via trois problèmes qui correspondent aux propriétés des signaux, des messages et de leurs relations par rapport à la sémantique du contenu.

Il ne peut y avoir de "conformité sémantique" que si les infosphères se "comprennent" ; la source et le destinataire doivent partager les mêmes conventions de signaux (pour assurer le transport des signaux porteurs de messages) et les mêmes conventions de messages (pour assurer une interprétation conforme de l'information).

Dans le cas de la communication d'un texte, cette conformité est simple à établir : les signaux forment des séquences binaires de longueur fixe dont le séquençement correspond à une forme unique, cette forme à pour valeur unique, un symbole.

Ainsi toute séquence binaire dans une communication en mode texte fait l'objet d'une restitution d'une forme non ambiguë pour l'analyse de cette séquence. Puis, cette forme est comparée à celles présentes dans un espace de décision qui permet d'en déduire la valeur symbolique.

L'espace de décision est organisé en fonction de la symbolique construite a priori et cet ordonnancement est mesuré par l'entropie de l'information dans cet espace.

Globalement, nous avons un ensemble de bijections entre un signal perçu et une forme analysée, puis entre cette forme et une forme de référence étiquetée avec une valeur correspondant à son interprétation dans une application.

Ces bijections, sont les conditions de conformité de la communication.

- **Qu'en est-il quand on envisage d'autres modes de communication, basés sur le graphique, l'image, l'audio ?**
- **Comment généraliser cette question de la conformité de la communication, quand la source et le destinataire ne partagent pas les mêmes conventions ?**

Ramener la question de la conformité de la communication à un problème de décision peut conduire à des situations incalculables et c'est cette incalculabilité qui est la cause principale de l'ambiguïté car elle correspond à des exceptions au principe de relations bijectives entre signaux, messages et valeurs.

Cependant, c'est bien cette question de conformité de la communication d'informations qui est le fondement de toute interface homme machine, mais aussi, de la validation de tout processus de reconnaissance de formes.

Sous cet éclairage, il devient intéressant d'explorer deux pistes peu représentées actuellement dans la recherche de méthodes d'évaluation de traitements de l'information :

- établir une **mesure de la complexité de calcul pour résoudre un problème** comme moyen d'évaluation d'un processus de reconnaissance de formes.
- définir des **modalités de présentation de l'information venant en substitution dans un contexte** conduisant à une situation d'incalculabilité ou de trop grande complexité de calcul par rapport à un contexte d'usage.

#### **III.4. - Un darwinisme documentaire**

Le principal changement induit par le numérique l'irruption de l'hétérogénéité dans la définition des données et de leur présentation, et par la redistribution des domaines de la médiation. Les causes de ces mutations sont dues à l'évolution de nos machines, de nos réseaux et des techniques logicielles : la notion de la relation mainframe-terminal a évolué vers la notion d'architecture client serveur avec ou sans médiation. Et à présent, les ressources informatiques accessibles à tout utilisateur sont devenues telles que clients et serveurs peuvent se confondre.

Dans cet écosystème, la notion même de normalisation par les formats de documents n'a progressivement plus sa place car elle repose sur des critères



techniques identiques à ceux que sont l'imprimatur du droit canon : une norme n'existe que par sa facilité de s'adosser à une technique peu accessible.

Or ce paradigme devient difficile à appliquer au document numérique. Aussi, dans un tel environnement volatil, **comment retrouver les critères de conformité de communication d'une information ?**

La réponse que nous proposons se trouve dans les termes d'intermédiation, de négociation entre source et destinataire sur la nature des messages (ou formes, ou traces) échangées dans une communication. L'objet de cette intermédiation est ici le transcodage (la reformulation) d'une présentation de l'information.

Cette notion d'intermédiation n'est pas sans évoquer ce que nous avons pu observer dans les réseaux de télécommunication, quand est apparue la nécessité de l'interconnexion physique de réseaux utilisant des protocoles différents de codages en signaux.

L'interaction homme machine est considéré ici sous l'angle d'une interconnexion logique, dans laquelle un des éléments agissant ou sous-système, est l'humain dont le comportement est éloigné du déterminisme a priori d'un modèle conceptuel de données.

#### ***IV. La numérisation industrielle de masse***

La numérisation en masse est traitée comme tout processus industriel, dans une logique de reproductibilité des processus, la rationalité ainsi mise en œuvre est, nous le verrons, paradoxalement très sélective, car deux contraintes antinomiques sont à prendre en compte :

- des contenus au caractère fortement hétérogène en termes de support, de présentation de l'information et de thématique traitée dans le document,
- des algorithmes de traitements nécessitant un paramétrage dépendant de la nature des contenus à analyser et à reconnaître.

Une chaîne de numérisation est un processus séquentiel dont l'efficacité de chaque étape dépend des performances obtenues dans les étapes précédentes.

Cette dépendance est difficilement quantifiable et la mise au point d'une plateforme de numérisation relève d'une véritable approche empirique par la méthode des essais et des erreurs, dans le cas d'une recherche de l'automatisation des processus : l'opérateur de numérisation choisit un échantillon de contenus, le met à l'épreuve dans une configuration de la plateforme de numérisation et, à partir des résultats, ajuste les paramètres de réglage des outils mis en œuvre dans cette plateforme.

Face à cette complexité, de nombreux façonniers de la numérisation préfèrent la solution d'une retranscription manuelle des contenus, reportant l'efficacité de la chaîne de traitement sur l'organisation d'un workflow logiciel, en instituant des processus de saisie/validation en double aveugle.

Cependant, quelle que soit la stratégie de plateforme retenue, le point d'entrée reste toujours le même, à savoir, des contenus dématérialisés obtenue par une numérisation en mode image.

Dans l'hypothèse d'un traitement automatisé, la présentation de l'information portée par ces contenus correspond à des formes dont la morphologie et la topographie se traduisent par des descripteurs géométriques et topographiques.

Ces descripteurs de formes sont calculés par un algorithme de segmentation dont le paramétrage conditionne les performances de la classification qui a pour objectif de reconnaître les formes du document pour leur affecter une valeur. Cette opération de classification va, elle-même, dépendre étroitement de l'entraînement préalable de l'outil de classification, donc d'un échantillon de documents de tests.

Dans une retranscription manuelle, ce sont les connaissances et savoirs de l'opérateur de saisie, son aptitude à déchiffrer le contenu, qui vont conditionner totalement la qualité de cette transcription.

Dans un cas, comme dans l'autre, la prise en compte d'une diversité de la présentation de l'information reste un écueil.

Cet écueil se traduit par une complexification technologique susceptible de rapidement devenir inacceptable sur le plan économique en termes d'efficacité d'un processus industriel incluant une automatisation des traitements.

La contrainte de répétitivité imposée par cette approche conduit à la spécification de processus devant fonctionner dans une marge de manœuvre préétablie.

Cette marge est d'autant plus étroite que les processus sont nombreux et interdépendants ; les contraintes d'utilisation d'un processus vont inféoder les autres processus de la chaîne de traitement.

L'ensemble conduit à une chaîne de traitement adaptée à un groupe restreint de contenus, conduisant à une intolérance aux exceptions d'autant plus forte que la spécialisation de la chaîne est grande. Cet effet "boule de neige" fait que la complexité des outils industriels de reconnaissance de formes s'est accrue au fil du temps.

Aujourd'hui, un OCR industriel est, en fait, une combinaison de plusieurs OCR dont les composants sont mis en compétition par un superviseur. Malgré cette complexité technique, on observe que l'amélioration des performances a atteint un niveau asymptotique ; le gain de quelques dixièmes de points de performance se traduit par une dépense "d'énergie logicielle" disproportionnée

Donc, ce n'est pas dans la sophistication des traitements que se trouve la solution pour contrer cet effet "boule de neige". C'est ce qui motive la sélection des contenus.

N'importe quel OCR du commerce a un fonctionnement correct sur des documents modernes, numérisés dans de bonnes conditions et dont la mise en forme fait appel à des modèles de polices de caractères et de structures de documents courants et relativement bien uniformisés. Ainsi, la conception d'une chaîne de numérisation comporte toujours, dans son cahier des charges, la description d'un document standard sur lequel repose toute l'organisation de la chaîne de traitement. Définir ce document standard induit une sélection des contenus par des critères techniques.

Cette sélection technique est acceptable, voire utile et nécessaire, pour un corpus de documents techniques, mais elle pose un problème grave pour les documents du patrimoine culturel. En effet, elle devient prééminente par rapport à la valeur culturelle ou patrimoniale du document à numériser.

Ainsi, pour résumer, une plateforme de numérisation massive est adaptée à la numérisation de documents composés dans une logique industrielle, ce qui sous tend une logique de standardisation portant à la fois sur les processus et sur les objets traités par ces processus. Tout document s'écartant de cet étalon qu'est le document standard défini lors de la conception de la chaîne de numérisation ne sera pas traité.

#### IV.1. - La numérisation industrielle de bibliothèques

L'opération de numérisation faite par la BNF (Bibliothèque Nationale de France) pour la bibliothèque numérique "Gallica" est intéressante à analyser et elle permet de comprendre les difficultés de mise en œuvre d'une opération de numérisation de masse dans une logique industrielle.

Ces difficultés sont, nous le verrons, principalement dues à l'hétérogénéité des documents sujets à la numérisation. Les données chiffrées ci-dessous montrent bien la problématique posée par la numérisation d'une bibliothèque, par rapport à la définition d'un standard de numérisation.

Le premier projet Gallica<sup>3</sup> reposait sur 80 000 volumes numérisés, dont 80% de monographies et 20% de revues

La répartition de ces volumes à travers le temps est présentée dans le tableau ci-après.

Antiquité	3%
Moyen Age	5%
XVI au XVIII siècle	34%
XIX siècle	40%
XX siècle	18%

La numérisation a été faite selon trois modalités :

Numérisation directe sans destruction du livre	55%
Numérisation directe avec destruction du livre	10%
Numérisation à partir de micro-supports	35%

Le taux de documents disposant d'une retranscription du texte est très faible : 1,5%

Cela correspond à environ 1250 ouvrages dont 1100 qui ont été numérisés par le CNRS avec une retranscription incomplète et 140 œuvres de Chateaubriand et Balzac retranscrits par un éditeur, Acamedia, qui n'existe plus aujourd'hui.

---

<sup>3</sup> Les données chiffrées présentées ici, sont extraits du site Gallica ([gallica.bnf.fr](http://gallica.bnf.fr)) en date du 01/10/2008

Cette très faible proportion de documents retranscrits en mode texte avait été relevée par la "commission Moati" pour la Bibliothèque Numérique Européenne [BNU06] comme posant un problème réel de valorisation des contenus de bibliothèques.

Ceci a conduit à l'établissement de recommandations applicables au projet Gallica 2. Ces recommandations portent sur :

- la définition d'une chaîne de numérisation privilégiant la production de documents numériques avec transcription du texte, soit par OCR, soit par transcription manuelle,
- la reprise du corpus numérisé du premier Gallica pour sa transcription en mode texte.

Cependant, ces recommandations portent sur une reprise au maximum de 80% des fonds de Gallica, du fait de la difficulté que représente une opération de retranscription, quelle que soit la procédure retenue.

Ainsi, force est de constater qu'une numérisation massive de documents patrimoniaux, organisée dans une logique purement industrielle sur des critères de standardisation technique, induit toujours une perte de valeur de l'information. Perte d'autant plus grande que le document à numériser s'écarte des standards établis pour la chaîne de numérisation, comme nous l'avons déjà signalé.

Pour la BNF, les ouvrages antérieurs au XIXème siècle, sont considérés comme étant «de transcription difficile par une procédure d'OCR».

Le problème vient du fait que les spécifications de normalisation des documents sont aujourd'hui celles de l'édition électronique ; les contenus sont produits à partir d'une composition électronique du document qui est accessible via des contrats de partenariats avec l'éditeur.

Dès lors, l'intérêt de la numérisation diminue fortement si l'on définit des contrats d'accès au document électronique source, en lieu et place de la numérisation de sa version imprimée.

## **IV.2. - Des hypothèses d'évolutions pour la numérisation**

Succomber à la tentation de l'ingénierie industrielle pour la numérisation, revient à négliger les questions d'affordance des contenus, donnée par leur présentation et donc, à négliger le contexte du destinataire dans notre canal de communication qui représente l'interaction homme machine.

Même en disposant du document source, les usages sont susceptibles de fortement conditionner une reformulation de la présentation du contenu :

- le récepteur du destinataire ne connaît ni ne maîtrise, l'intégralité du format physique du document,
- la mise en forme du document peut est inadaptée à la modalité préférentielle de l'utilisateur,
- les métadonnées peuvent se révéler hors du champ des connaissances de ce dernier lors de la recherche d'information.

Dans tous les cas, une intermédiation entre source et destinataire est nécessaire. Elle doit être matérialisée par des protocoles de transcodage de la présentation de l'information entre l'émetteur et le récepteur. Ces protocoles seront négociés et appris. Cet ensemble d'hypothèses semble représenter la seule solution du respect de la contrainte de conformité de l'information dans un contexte de documents hétérogènes.

Dans une plateforme industrielle de numérisation, les spécifications de numérisation sont précisément des spécifications de codage de l'information : l'émetteur produit une image qu'un récepteur doit analyser et interpréter.

Le canal de communication que représente une chaîne de numérisation est un système séquentiel fermé. Or du fait de la diversité des usages, il peut apparaître intéressant de considérer une chaîne de numérisation comme un système ouvert, assurant l'interconnexion de deux canaux de communication.

Du fait de la diversité des contenus à traiter, de leurs relations directes avec le paramétrage des processus composant la chaîne de traitement et de l'interdépendance de ces processus, il peut apparaître opportun d'adopter une stratégie nettement plus orientée cybernétique à chaque sous-système d'une chaîne numérisation.

Dans cette stratégie, les informations de feedback et de feedforward sont les traces d'activités de chacun des processus de la chaîne de traitement ; ces traces peuvent faire l'objet d'une analyse et d'un apprentissage incrémental, soit par le calcul, soit par intervention de l'utilisateur.

Ces hypothèses mettent en avant l'intérêt de considérer la numérisation industrielle selon une approche systémique et non plus selon la seule approche technologique d'assemblage de composants logiciels. L'idée sous-tendue ici est, à complexité technologique égale, d'étudier l'organisation fonctionnelle d'une plateforme de numérisation en adoptant un modèle cybernétique.

De premières tentatives de mise en œuvre avaient été proposées dans les années 90 au travers des travaux sur les cycles perceptifs de reconstruction d'objets [OGI94] et d'études de plateformes de rétroconversion (Den Hartog, Vaxivière...) et d'utilisation de systèmes multi agents pour la rétroconversion des plans de réseau de télécommunications.

## ***V. La numérisation de plans de réseaux***

L'étude de la numérisation de plans de réseaux de télécommunications a correspondu à un projet dont j'avais la charge et qui a servi de cadre aux thèses de Vincent Grenier [GRE00a][GRE00b] et de Sébastien Adam [ADA00].

Cette plateforme avait pour objectif d'automatiser les processus de reconnaissance d'éléments graphiques et symboliques constitutifs d'un réseau et décrivant l'itinéraire de ce réseau, c'est-à-dire, son emprise géographique sur le terrain.

Très rapidement, nous nous sommes rendu compte qu'une approche par le seul critère de l'ingénierie logicielle se révélait insuffisant.

En effet, concevoir un processus informatique séquentiel revient à définir des cas d'usages dans lesquels, le type de données (ici, le type de document), définit un

périmètre standard de faisabilité technique qui donne au processus son périmètre d'utilisation.

Pour les documents traités, cela nous a conduits à tenter de décrire un "document standard" adapté à la chaîne de traitement.

Or, en appliquant cette stratégie de conception, deux difficultés sont apparues :

- La définition de ce "document standard" en faisait un objet technique très complexe :
  - image issue de la numérisation de plans au format A0,
  - symbologie riche, incluant des objets à la géométrie et aux relations topologiques très variées (éléments connectés ou non, multi-orientation des symboles et chaînes de caractères, superposition de "couches thématiques" sur un seul plan...).
- Existence de nombreuses exceptions par rapport au "document standard", malgré l'existence d'un cahier des charges de spécifications de la documentation technique :
  - qualité du support variable provenant de l'âge des différents plans et de leurs manipulations,
  - nature des fonds de plans cartographiques disponibles selon la région concernée par la documentation réseau,
  - règles de dessins variant d'un établissement à l'autre selon les normes d'ingénierie des réseaux ou selon les "habitudes métier" du dessinateur".

Ces propriétés communes à tous les plans de réseaux ont rendu l'étude du processus de reconnaissance très complexe, en ce sens qu'il a fallu prendre en compte les exceptions aux documents de tests, c'est-à-dire, les plans conformes aux spécifications exactes de l'opérateur France Télécom (Fig. 11). Ce décalage réel entre la mise en forme d'un "document étalon" issu de ces recommandations et la vérité terrain était un fort verrou empêchant toute idée de généralisation d'une plateforme classique de numérisation des plans.

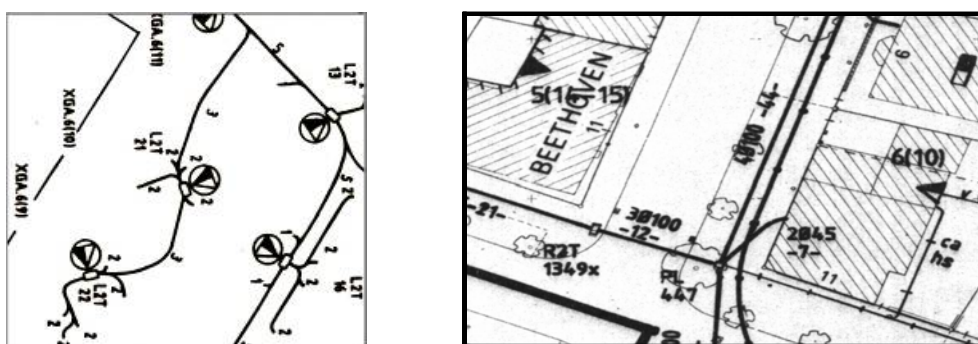


Fig. 11 : extraits de plans d'itinéraires de réseaux de télécommunication (à gauche, Toulouse, à droite Strasbourg).

Cela a conduit au développement d'un système multi-spécialiste, dans lequel nous avons défini des classes d'agents correspondant à chacune des étapes de la reconnaissance de formes.

Cette plateforme reposait sur cinq concepts :

- une modélisation en couche spécifiant la présentation d'un objet métier en termes de données graphiques et images,
- un modèle de traitements reprenant cette structure en couche, assurant ainsi la conversion de données images (segmentation) en données graphiques (formes) et leur relation avec des données métier,
- dans chacune des couches, spécification d'agents logiciels ayant chacun leur domaine de compétence,
- mémoire partagée de type "blackboard", permettant à chaque agent de prendre comme contexte de sa propre compétence un résultat de traitement d'autres agents,
- initiation d'un traitement par désignation du type d'objet à reconnaître (propagation et reformulation d'une intention de reconnaissance dans chacune des couches), puis exécution du processus par un mécanisme ascendant,

Globalement, cette plateforme était capable de traiter la topologie complexe d'un plan conforme aux règles de dessin établis par l'opérateur France Télécom (Fig.12). Mais le passage d'un type de plan à un autre nécessitait de revoir totalement la cartographie des agents et de leurs compétences. Ceci a mis en évidence que "l'intelligence" du système ne se limitait pas aux données décrivant les documents traités, mais également à la définition même des composants, de leur paramétrage et du séquençement de ces composants.

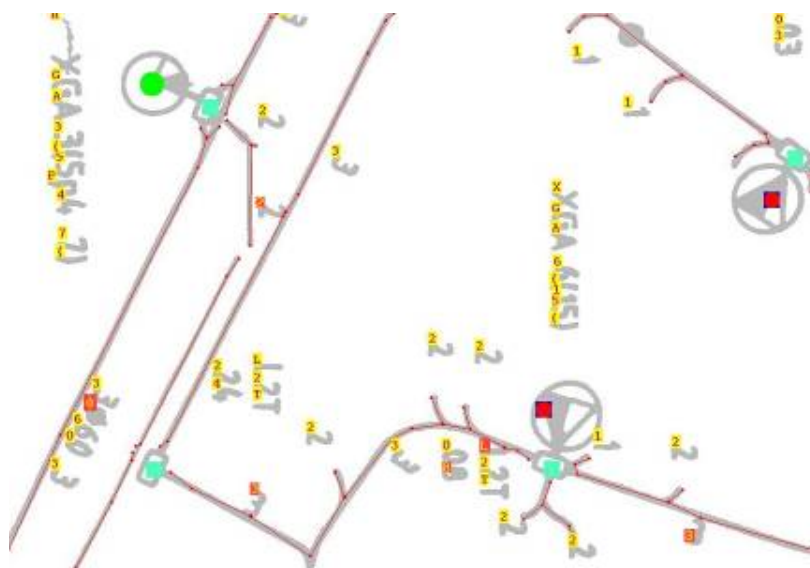


Fig. 12 : Reconnaissance de symboles connectés, vectorisation d'éléments linéaires et reconnaissance de caractères assistée par les données métier pour la nomenclature

Hormis cette incapacité de transposer cette plateforme d'un type de documents à un autre, nous avons été confrontés aux problèmes suivants :

- la gestion de la mémoire du système par blackboard qui a conduit à un modèle de données complexe à gérer et difficilement maintenable, car trop décorrélée des formes de l'image,

- la nécessité de construire un langage de requêtes commun à tous les agents quel que soit leur niveau de compétence,
- la difficulté d'évaluer un résultat de traitement ; dans le cas des plans de réseaux, cette difficulté a nécessité d'avoir recours à des utilisateurs experts de la documentation des réseaux,
- une plateforme très complexe en termes de développement et de maintenance du code (en relation avec les technologies de génie logiciel des années 90 : C++, bus de communication CORBA, base de données Oracle 8, langage de requêtes entre agents KQML

Cependant, ce premier projet de numérisation de masse de documents techniques a permis de démontrer l'intérêt de coupler les approches descendantes et ascendantes de la reconnaissance de formes, la phase descendante correspondant à une propagation d'intention, la phase ascendante à la mise en œuvre du traitement. L'approche développée constituait un système piloté par des connaissances actionnables, c'est-à-dire dans lequel la présentation des connaissances était portée par des outils logiciels et non les données.

## ***VI. Le traitement de documents hétérogènes***

A la suite de cette conclusion sur la numérisation des plans, la problématique de l'hétérogénéité des documents et de la présentation de leurs contenus est devenue ma thématique de recherche principale. Elle a conduit à la constitution d'un projet collaboratif labellisé par le Réseau National des Technologies Logicielles (RNTL) de 2001 à 2003 : le projet DocMining

### **VI.1. - Le projet DocMining**

Les objectifs du projet devaient apporter des réponses aux problèmes posés par la numérisation des plans de réseau tout en ouvrant la plateforme à tout document faiblement ou non structuré. De plus, en raison de la diversité des traitements à mettre en œuvre, il s'est agi de produire un "framework" capable de

- s'interfacer avec différentes bibliothèques logicielles correspondant aux données d'entrées/sorties,
- aux composants logiciels mis en œuvre pour les traitements et les interfaces,
- au séquençement des traitements sous la forme de scénarios

Le résultat du projet a été une plateforme de démonstration d'un système d'acquisition des contenus inscrits dans des documents faiblement structurés.

Ce système se place à l'interface d'une base d'informations documentaires et des entrepôts de données susceptibles d'être interrogés par des médiateurs de services (moteurs de recherche, par exemple).

Du fait de sa position particulière dans un urbanisme complexe de systèmes d'informations documentaires, le projet DocMining doit assurer des fonctions d'intermédiation dans les échanges de documents entre différents services internet ou intranet.

Ce projet, défini et piloté à France Télécom R&D avait pour partenaires :



- Le LORIA (Karl Tombre et al.) : ce laboratoire devait dresser la cartographie des outils et méthodes de reconnaissance de formes et mettre à notre disposition sa plateforme QGAR de traitement de documents techniques.
- L'équipe DIVA du Département d'Informatique de l'Université de Fribourg (Rolf Ingold et al.) : ce laboratoire avait pour tâche de fournir les composants logiciels d'interfaces et nous a mis à disposition sa bibliothèque XML de construction de présentation de documents.
- Le laboratoire PSI de l'université de Rouen (Yves Lecourtier et al.) et le laboratoire L3i de l'université de La Rochelle (Jean Marc Ogier et al.). Ces deux laboratoires devaient définir les solutions d'architecture et de modèles de données correspondant aux spécifications de la plateforme.

## VI.2. - Les concepts du projet

Cette plateforme technologique repose sur 3 concepts :

- **Architecture reposant sur une approche document-centrée.**

Les traitements de documents communiquent à travers un point de vue unique du document lui-même; cette approche permet d'éviter les problèmes de dispersion des caractéristiques calculées, habituellement constatés dans les chaînes classiques de traitement de documents et correspondant à la difficulté de gérer un blackboard commun dans notre plateforme précédente.

Cette architecture était la réponse à la question suivante :

**Pourquoi ne pas travailler dans une logique "métadonnée", à savoir, construire une relation incluant la topologie, entre une forme produite par la segmentation et ses différentes valeurs?**

- **Plateforme logicielle basée sur une architecture modulaire orientée plug-in.**

Les développeurs peuvent facilement ajouter un nouveau traitement rendant, de ce fait, la plateforme facilement extensible. Les outils de visualisation et de manipulation de documents sont eux-mêmes des plug-ins, de sorte que les possibilités d'interaction s'adaptent à la structure de document.

Ici la question traitée a été celle de la multiplicité des traitements. Au-delà des problèmes techniques soulevés par l'intégration de bibliothèques logicielles hétérogènes, elle sous-tendait la définition de l'interface utilisateur. Nous nous sommes basé sur les solutions JAVA/beans pour l'intégration de bibliothèques logicielles. Les outils de l'interface, quant à eux, sont invoqués de manière contextuelle pour adapter à la volée la segmentation d'un document à partir des intentions de l'utilisateur, traduites en actions. Chaque composant de la plateforme est "objectisé" (fonctions de marshalling/unmarshalling) via un "contrat de lecture" décrivant le mode de dialogue entre un composant et la plateforme.

- **Scénario encapsulés.**

La plateforme manipule des fonctions de type scénarios encapsulables dans d'autres scénarios. L'exécution d'un scénario peut enregistrer les actions de l'utilisateur, qui deviennent ainsi une partie intégrante du scénario qui

pourra être reproduit. Le scénario peut alors être transformé en un nouveau traitement par rapport à un scénario de plus haut niveau.

Ici la question a été celle de la **co-construction d'un scénario de traitement par les actions de l'utilisateur** en vue de reproduire ce traitement avec d'autres données et de **moduler le paramétrage des étapes d'un traitement**. Ceci a permis d'utiliser DocMining dans quelques applications d'acquisition et de constitution de bases de vérité terrain (Clavier [CLA04] )

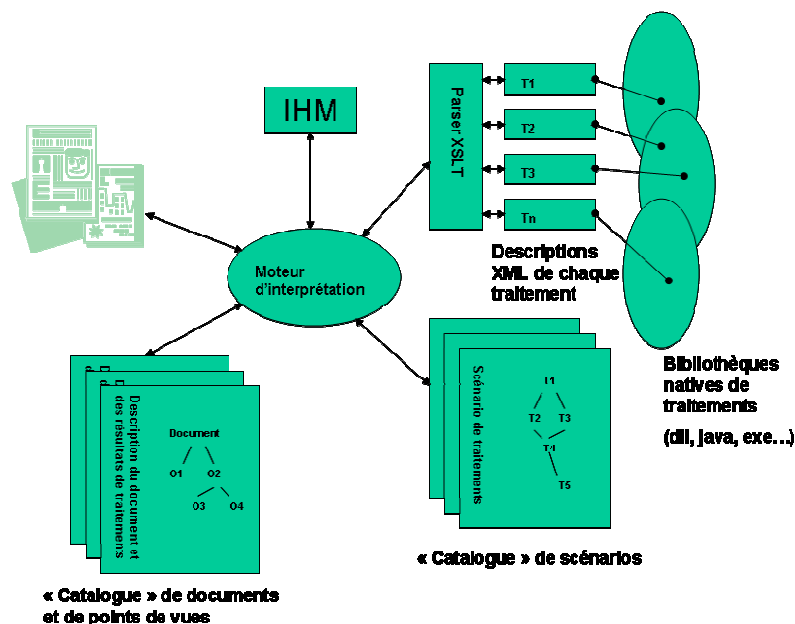


Fig. 13 : Schéma global de l'architecture de la plateforme DocMining (pour plus de détails, se reporter à l'ANNEXE 2)

### VI.3. - Le profil d'une application basée sur DocMining

Globalement une application basée sur DocMining a une structure en 3 couches :

- une couche de visualisation qui est l'image du document,
- une couche de segmentation (les signes saillants, donc manipulables, de présentation d'un document, adaptable selon les besoins de l'utilisateur/récepteur), qui correspondent aux formes canoniques du document
- une couche de relation/classification, contenant les opérations de fusion et d'éclatement des formes canoniques présentes dans la couche de segmentation et son contenu en fonction de l'usage. Cette couche contient, par ailleurs, les opérations de traitement susceptibles d'être exécutées par l'utilisateur.

La notion de "syntaxe véhiculaire" d'XML prend ici tout son sens.

La présentation d'un document calculée par DocMining et la composition d'un scénario d'application, sont en effet, des documents XML sur lesquels nous appliquons des masques de vues (XSL, XSLt, DTD).

DocMining est une proposition de solution technique de plateforme réutilisable quel que soit le type de documents. Cependant, plusieurs problèmes n'ont pas été résolus par la conception de cette plateforme, notamment, le problème de la validation et de l'évaluation des résultats de reconnaissance de formes.

Ces problèmes sont apparus comme étant ceux de la relation à construire entre une forme issue d'une segmentation et sa valeur donnée par un classifieur. Pour résumer ce problème de décision, il s'agissait à présent d'étudier de plus près la relation effective entre une forme restituée à une interface, son interprétabilité par un utilisateur humain, et les objets la décrivant par rapport au modèle de calcul utilisé par un système d'informations.

## *VII. Les trois niveaux de présentation de l'information dans le document numérique*

Une syntaxe ne permet que de fixer le langage de description d'un document multimédia en fixant une norme de codage sur trois niveaux :

- un niveau physique : l'inscription d'un contenu sur un support physique,
- un niveau sémiotique : la constitution de délimiteurs de formes présentant une information,
- un niveau sémantique : les métadonnées.

Mais ce langage, quel qu'il soit, conduit toujours à la production d'objets pour une famille d'usages délimités et donc, contraint l'utilisateur destinataire, ainsi que la technique mise en œuvre dans son récepteur, à devoir adhérer aux contraintes de la source pour garantir la conformité de l'information.

Le besoin identifié ici est celui d'une fonction d'intermédiation, dont une proposition d'implémentation a été faite dans DocMining. Cette intermédiation consiste à adapter, en la reformulant, la présentation de toute l'information d'un document en fonction des contextes d'usages.

Ces contextes correspondent à des contraintes données par :

- l'usage du contenu dans une application utilisée par le destinataire,
- la composition du document à partir des propriétés de la source,
- le transport des signaux liés au réseau de communication.

Ces trois contextes peuvent être antinomiques :

- la structure du document peut ne pas être conforme aux propriétés de l'application finale,
- certaines utilisations peuvent violer les droits d'usages octroyés par la source et traduits dans la mise en forme du document,
- la bande passante du réseau peut ne pas permettre le transport de toute la présentation de l'information contenue dans un document.

### **VII.1. - Document numérique et accessibilité**

Le handicap est le contexte le plus illustratif de l'accessibilité à un document. Le changement de modalité, nécessaire pour rendre un document accessible à une

forme de handicap, peut remettre totalement en question la segmentation initiale d'un document. Les difficultés induites par ces changements de modalités proviennent de la non superposition des périmètres sémantiques et d'usages entre deux modes de présentation de l'information (p.ex. l'acoustique et le visuel), (Sharon Oviatt).

A ces périmètres sémantiques sont associés des informations de structure du document ou de segmentation qui sont indispensables pour l'interprétation et la manipulation d'un contenu. Or cette information de segmentation peut être totalement absente dans une modalité donnée ou très difficile à transcrire.

L'exemple le plus représentatif peut être donné par la vocalisation d'un texte à destination des non-voyants. La mise en forme d'un texte en paragraphes, titres, etc. est faite selon un modèle topographique sur un document plan, donc, un espace bidimensionnel.

Cette mise en forme n'est pas transposable dans le cas d'une vocalisation du même texte. On doit passer d'une structure bidimensionnelle de la mise en forme à une structure purement séquentielle, où seules les intonations placées à la manière de jalons séquentiels, peuvent permettre à l'auditeur de reconstruire mentalement la structure d'un discours.

## **VII.2. - Le cas de la numérisation**

La numérisation des documents est également un exemple représentatif de cette difficulté soulevée par le codage de la présentation de l'information. En effet, les procédures de numérisation conduisent à la production d'images de documents. Or une image est une mesure, une quantification d'éléments de surface d'un document sur support papier et l'unité élémentaire d'information à la sortie d'un scanner est le pixel contenant la valeur de cette mesure.

Ainsi, l'information de segmentation d'un texte, bien que ce dernier soit parfaitement lisible par un humain, est totalement absente du document image. Cette information est obligatoire, d'une part, pour permettre une reconnaissance de formes dans le document (caractères, dessins, toutes formes de graphismes) et, d'autre part, permettre à un utilisateur humain de manipuler les contenus pour, par exemple, corriger et valider les résultats d'un OCR (Reconnaissance Optique de Caractères).

Les traitements de segmentation d'images produisent une proposition de structure physique du document à partir de critères morphologiques et topographiques. Elle conditionne fortement la validation d'une procédure d'OCR, en dehors de toutes considérations sémantiques sur le sens du texte présent dans l'image.

## ***VIII. L'arbitraire de l'interprétation et le déterminisme technique***

Dans cette approche, nous voyons que la segmentation conduit à produire une structure de l'image sous la forme d'objets manipulables, tant par l'utilisateur que par un programme logiciel. Ces objets peuvent donc être vus comme des unités techniques de manipulation du contenu (Bachimont). Par rapport à la reconnaissance de caractères, cette segmentation est une information technique construite a priori.

La validation, quant à elle, résultera toujours d'une interprétation qui sera faite a posteriori de la segmentation.

Cette validation est toujours le résultat d'une interaction homme machine, même à l'usage de post traitements linguistiques qui ont pour objectif de fixer la valeur d'une forme en fonction de son sens : L'humain perçoit une présentation de l'information calculée par la machine ; la machine perçoit des données résultant d'une mesure ou d'un calcul représentant l'activité de l'utilisateur.

Le résultat d'une reconnaissance de formes validée est une information ajoutée correspondant à la documentation de cette forme par rapport à son usage. Il s'agit donc d'une métadonnée qui doit non seulement définir la valeur sémantique de cette forme, mais également son affordance par rapport à un usage.

Par exemple, une propriété d'usage d'un caractère peut être appelée une unité sémiotique d'interaction. Son affordance sera potentialisée par sa vraisemblance de forme par rapport à un usage.

Ainsi, on devrait considérer tout résultat de transcription (par OCR ou manuelle) comme un travail éditorial de production de métadonnées sur le sens du texte, mais aussi, sa tangibilité (ou affordance) par rapport à une interface utilisateur.

A partir de ces deux exemples, celui de la conversion de modalité et celui de la numérisation de documents, nous voyons apparaître l'importance de la présentation de toute information dans la capacité de reconstruire son sens, que ce soit par un programme logiciel ou un utilisateur humain.

En termes d'interaction homme machine, cela se traduit par la capacité de mettre en relation un espace de formes issues d'une segmentation proposant une structure physique du contenu avec un espace de valeurs résultant de l'ontologie de l'application. Cette mise en relation permet à l'utilisateur de superviser son dialogue avec la machine.

## ***IX. Interaction et prise de décision***

L'image est le support le plus complet de la présentation de l'information en termes de perception, d'analyse et d'interprétation humaine. En revanche, sa perception par la machine pour l'indexation, la segmentation en unités techniques de manipulation, nécessite un calcul.

Ceci met en avant un fort déséquilibre sémiotique entre la représentation mentale du contenu de l'humain et la représentation technique du contenu pour la machine. La résolution de ce problème nous incite à redéfinir le concept de reconnaissance de forme le plaçant dans le concept phénoménologique d'une triade sémiotique (objet, signes, interprétant)

La perception, l'analyse et l'interprétation du contenu d'une image par un humain est une sémiologie qui est facilitée par l'affordance des formes inscrites dans cette image. Le concept d'affordance (Gibson, Norman) traduit la notion de présentation d'une information dont la mise en forme facilite l'interprétation du contenu. On peut résumer ce concept en disant qu'une forme doit évoquer un sens par elle-même et correspondre à la signification d'un objet présenté, mais aussi, à sa tangibilité dans une interface.

La production de formes affordantes passe par un calcul sur des données. Dans le cas de l'image, il s'agira d'un processus de reconnaissance de formes.

Il y a là, une symétrie intéressante à exploiter dans la conception d'une interface et dans l'évaluation d'un résultat de reconnaissance de formes : à l'affordance humaine, le signe par rapport à son interprétant correspond une affordance computationnelle, le signe calculé à partir d'un objet ou l'objet calculé à partir de signes. Cette notion d'affordance computationnelle que nous introduisons ici, correspond à la complexité d'un calcul permettant, à partir d'un ensemble de données, de construire une présentation de l'information. Cet ensemble organisé de données est un objet au sens informatique du terme. Il s'agit d'une présentation de l'information adaptée à un programme logiciel. Le programme, quant à lui, produit des formes restituées à l'interface, et ces formes sont une présentation de l'information adaptée à l'utilisateur.

Ceci nous conduit à reconsidérer la reconnaissance de formes. Il ne s'agit plus d'un processus purement ascendant de traitement d'un signal, mais d'une opération, revenant à transcoder la présentation d'une information. D'une présentation computationnelle, il s'agit de passer à une présentation perceptible par l'humain. Ce transcodage se fait en effectuant une intermédiation homme machine au cours de laquelle, il s'agira de capter les traces d'activités comme support d'un apprentissage.

On peut résumer cette approche par une rupture du paradoxe de Sayre rappelé précédemment :

**"segmenter pour reconnaître, reconnaître pour segmenter",**

En y insérant la notion d'affordance le paradoxe disparaît :

**"segmenter pour manipuler, manipuler pour interpréter, interpréter pour reconnaître".**

## ***X. Sémiotique et information***

Cette symétrie d'affordance humaine et computationnelle a de fortes accointances avec la sémiotique de C.S. Peirce. Dans cette approche phénoménologique, l'élément central de la reconstruction d'un sens est le signe : tout objet est représenté par des signes et le sens dépend de l'interprétation de ces signes. La composition d'un document, quel qu'il soit, revient à transcrire de l'information sous la forme de signes.

Dans une optique de communication, ces signes sont le strict équivalent des messages définis par Shannon. Ces signes/messages sont amenés à être interprétés par un programme logiciel ; dans ce cas, les messages seront traduits en données susceptibles d'être lues par un programme, ou par un humain, et donc, il s'agira donc d'une mise en forme perceptible et tangible de ces données.

Ainsi, nous développerons trois principes :

- **L'équilibre sémiotique nécessaire entre la présentation du contenu faite à l'utilisateur et sa segmentation codée dans la machine** (relation entre forme et valeur de la forme). Il faut remarquer ici que cet équilibre sémiotique n'est autre que la déclinaison contemporaine des "trois problèmes de la communication" développés par Warren Weaver en avant-propos de la théorie mathématique de la communication de Shannon en 1948.

- **La connaissance de la, ou des modalités de présentation de l'information préférentielles pour l'utilisateur** et les possibilités de les co-aligner avec celles mises en œuvre lors du codage par les différentes sources d'informations. Nous verrons au chapitre 4 que cette connaissance nécessite la construction d'une distance entre informations qui fait abstraction du format et de la finalité du contenu.
- **La mise en place d'outils et de méthodes issues de la reconnaissance de formes et de l'apprentissage actif** pour obtenir les conditions de co-alignement des modes de présentation de l'information.





## Chapitre 3

### **Approche systémique de la présentation de l'information**

*"La société dans laquelle l'homme vit et se pose comme être humain n'est rien d'autre qu'un vaste et complexe système de systèmes de signes."*

**Umberto Eco** - "Le Signe"

Les chapitres 1 et 2 ont eu pour objectif de dresser un constat montrant un décalage entre les capacités offertes par les matériels et réseaux actuels et celles mises en œuvre dans nos habitudes de conception d'applications. Ce chapitre qui suit est un positionnement du problème de la présentation de l'information liée aux usages.

Ce positionnement s'adosse aux théories de l'information en ce centrant le test de décision qui revient à donner une valeur à une forme arrivant sur le récepteur d'un destinataire. Ne pouvant prévoir tous les cas d'exceptions à la configuration des données prévue lors de la conception d'une application, le rôle de l'interface comme outils de contrôle des traitements de l'information devient important tout en situant mieux l'utilisateur et ses usages dans la boucle de traitements.

## *I. L'interface homme machine*

Toute interface homme-machine a pour finalité d'échanger de l'information provenant soit d'un utilisateur humain, soit d'un programme logiciel. L'acceptabilité d'une application informatique passe toujours par l'évaluation de l'information échangée entre un utilisateur et un programme logiciel. Cette mesure nécessite de pouvoir définir et quantifier l'information en termes de maintien de la conformité de l'information dans la communication et de facilités de dialogue.

Si l'on souhaite procéder à une évaluation complète de la conformité de l'information communiquée dans une interface homme machine, on se heurte à une difficulté qui peut tenir lieu de paradoxe : toute l'information communiquée par une machine est codée sous la forme de données correspondant à un espace de décision connu a priori et construit à partir

- d'un modèle des données,
- d'organisation des traitements,
- des syntaxes mises en œuvre pour former des objets calculables par une application informatique.

Ces objets représentent, comme l'explique Bachimont, des "Unités Techniques de Manipulation".

A l'autre extrémité du canal de communication que représente une interface homme-machine, se trouve un utilisateur humain agissant en mettant en œuvre ses connaissances, cultures et croyances, associées à ses capacités de manipulation d'objets représentant une information afin d'en reconstruire le sens.

Le paradoxe se situe dans le fait que l'utilisateur n'a que rarement la connaissance des règles et méthodes ayant conduit à la production des "Unités Techniques de Manipulation" et donc, ne partage pas toujours les référentiels techniques et de connaissances intervenu dans la conception de l'application logicielle et de son interface, ainsi que dans l'ingénierie de ses contenus.

Les saillances d'une forme, caractérisant son affordance, dépendent étroitement de ce que l'on appellera ici pour simplifier, un modèle cognitif d'interprétation de l'utilisateur, et ce dernier n'est pas unique. Le mécanisme d'interprétation est une succession d'analyse et de reconnaissance de formes porteuses de sens. Elles sont reconnues par l'utilisateur dans son contexte cognitif, ce qui nous évoque un mécanisme constructiviste de sémiologie au sens de Peirce.

Les formes ou signes donnent à l'information sa matière tangible et perceptible pour l'utilisateur. Elles correspondent à la présentation de l'information. L'évaluation des critères de tangibilité d'une présentation de l'information revient à définir la capacité pour un utilisateur de manipuler celle-ci dans son propre contexte. Aussi, ces affordances, qui sont également des formes ou signes, peuvent être assimilées aux "Unités Sémiotiques d'Interprétation" décrites également par Bachimont<sup>4</sup>.

---

<sup>4</sup> Dans le modèle de la triade sémiotique de Peirce, tout est signe. Ceci permet de construire un treillis sémiotique dans lequel nous distinguons les signes en tant que tels en relation, les signes en tant que représentation d'objets et les signes en tant que représentation des interprétants. L'Unité

Ici nous faisons l'hypothèse que la construction d'un sens passe obligatoirement par la mise en contexte (ou "manipulation" dans un sens très large du terme) d'une forme quelconque. Cette mise en contexte revient à extraire (lecture d'un contenu) ou calculer (segmentation ou re-segmentation d'un contenu) ce qui rend cette forme techniquement adressable par un logiciel et dont la manipulation fait sens pour l'utilisateur.

La difficulté réside ici dans l'imprédictibilité des états possibles de couplages entre Unités Techniques de Manipulation et Unités Sémiotiques d'Interprétation : l'interprétation est un mécanisme qui est propre à l'utilisateur destinataire et indépendant de la source d'information.

La conformité d'interprétation de l'information réside dans la compatibilité existante entre la définition des espaces de décision propres à la source et au destinataire. Ces espaces définissent les conventions de codage de l'information pour la communication et comportent des valeurs affectées à des messages/formes/signes et traduits en signaux codés et ordonnés en respectant des critères de conformité avec une faible entropie.

## ***II. Segmentation et prise de décision***

Dans cette approche, cette notion d'espace de décision des états possibles de l'information doit pouvoir être capable de prendre en compte les exceptions, c'est-à-dire, des états imprévus dans la conception d'une application.

Ceci est d'autant plus vrai dans un système de communication multimédia et multimodal dans lequel une forme est équivoque en l'absence de contextes. De plus, selon le mode de présentation utilisé, ces espaces de décision ne se combinent pas de manière exacte (Oviatt).

Cette non complétion, sur le plan syntaxique, tout comme sur le plan sémantique, entre modalités de présentation de l'information induisent de l'ambiguïté, de l'inexactitude, voire des contradictions. Elles se traduisent par une relation inexacte ou non bijective entre les formes et les valeurs de l'information dans chacune des modalités traitées dans une interface.

Si l'on prend l'exemple de la numérisation, et en particulier, celui de la reconnaissance des caractères (OCR), nous nous plaçons déjà dans ce contexte imprécis : reconnaître les caractères d'un texte contenu dans l'image d'un document, revient à effectuer une conversion de modalités en passant d'une modalité visuelle, celle de l'image numérique, à une modalité textuelle, de nature vectorielle.

Dans l'image, l'Unité Technique de Manipulation est le pixel et dans le texte, l'Unité Technique de Manipulation (que nous abrègerons désormais par l'acronyme d'UTM comme chez Bachimont) est le caractère. L'OCR agit comme un mécanisme interprétant qui établit des relations entre les UTM élémentaires de l'image pour en faire des formes représentatives de caractères. Ces formes sont ensuite soumises à des outils de classification dont le rôle est de donner la valeur des caractères.

---

Sémiotique d'Interprétation, décrite par Bachimont comme un concept virtuel, a ici une matérialisation comme étant les saillances d'une forme caractérisant ses affordances.

La manifestation des exceptions évoquée précédemment est ici, l'apparition d'une "forme exotique" à l'espace de décision fixé a priori dans une base de classification. La difficulté est que l'exotisme de la forme d'un caractère a deux causes principales :

- **La "forme exotique" est issue du document** : il peut s'agir d'une forme correspondant à une police de caractères inconnue du système. Il s'agit d'une exception que l'on qualifiera de globale, car elle est reproductible. Elle nécessite un nouvel entraînement de l'OCR, voire la mise en œuvre d'une nouvelle méthode de classification en raison de descripteurs non représentatifs des propriétés morphologiques de la forme et, par voie de conséquence des méthodes de calcul de distances de classification.

Cette gestion des exceptions pénalise fortement le fonctionnement d'un OCR ; Elle nécessite la redéfinition des outils, méthodes et données mis en œuvre dans cet OCR. Cette révision induit l'apprentissage d'un nouvel ensemble d'Unités Techniques de Manipulations et de leurs moyens de calcul.

- **La "forme exotique" est un artéfact produit par la chaîne de numérisation elle-même** : les traitements de l'image d'un document numérisé préalables à la classification, génèrent des exceptions singulières (par opposition aux exceptions globales). Elles ne peuvent que très difficilement être traitées par la seule approche technique :
  - la singularité de l'exception d'une forme candidate à la reconnaissance est elle évaluable, donc quantifiable et reproductible ?
  - Fait-elle sens à son interprétation ?
  - Doit-on traiter cette singularité comme un artéfact parfaitement aléatoire ?

La réponse à ces questions requiert l'intervention de l'utilisateur dans le processus de validation de l'OCR, mais également, dans tout processus de reconnaissance de formes. L'hypothèse faite ici est que ces exceptions de la reconnaissance de formes ne sont pas calculables a priori car elles ne dépendent pas seulement des Unités Techniques de Manipulation fixées a priori par la source. Elles dépendent également de la capacité de l'utilisateur de les percevoir et d'en interpréter le sens dans son propre contexte.

Vus de la machine, il s'agit de facteurs arbitraires car ayant une dimension stochastique, par opposition au déterminisme de la source. Ce n'est qu'à partir de l'observation des traces de l'activité du destinataire (ce qu'il fait à la réception d'un message ou d'une forme), que l'on pourra valider la conformité soit d'une reconnaissance de formes, soit, plus généralement, d'une communication d'informations. L'interprétation n'est pas conditionnée par les propriétés techniques de la source mais par celles du destinataire.

Cette dimension stochastique des actions comportementales de l'utilisateur ne peut pas être réduite et rapportée à un modèle déterministe. Vu de la source, le contexte du destinataire fait que les états de son espace de décision (relation entre une forme et sa valeur) sont statistiques et probabilistes, tout comme les axes définissant cet espace de décision.

### *III. Présentation de l'information et perception*

L'information restituée par la machine est devenue largement multimédia et multimodale en intégrant progressivement les modalités sensorielles de la vision, de l'audition et du toucher, alors que la saisie d'information par l'humain est essentiellement haptique et gestuelle par l'usage du clavier et la souris.

Ce décalage est en relation avec l'organisation fonctionnelle de notre propre système nerveux qui est multimodal sur le plan des chaînes sensitives au travers de nos 5 sens, et monomodal sur le plan de nos chaînes motrices puisque nous ne disposons que d'une seule catégorie d'effecteur : notre l'appareil locomoteur. Certes, la variété d'actions que peut générer notre appareil locomoteur est riche, mais l'humain est essentiellement haptique quand il agit d'émettre de l'information<sup>5</sup>.

Le dispositif traditionnel de capture d'informations gestuelles, le clavier, tend à disparaître avec l'apparition de dalles tactiles sur l'écran, comme par exemple, les tablet-PC, les umpc et, bien entendu, les mobiles de dernière génération. Bien que le clavier tende à persister dans une forme virtuelle à l'écran, son usage est peu ergonomique du fait de l'absence de feedback mécanique vers l'utilisateur et des contraintes liées à la taille disponible pour l'affichage.

Le feedback mécanique de l'appui d'une touche peut être simulé par des dispositifs générant des vibrations perceptibles par la pulpe du doigt, mais la contrainte de taille disponible pour l'affichage reste problématique. Aussi, la solution d'un clavier virtuel ne représente pas une solution réellement viable pour effectuer une saisie massive d'informations, comme par exemple, rédiger un texte, former une requête complexe, prendre des notes. Ainsi, il convient de revenir sur la définition même d'une interface homme machine haptique.

La forme la plus simple de ce codage est celle des signaux générés à partir du clavier. Il n'y a aucune ambiguïté de présentation de l'information : à une action sur une touche correspond un signal unique dont la valeur est un code unique. Nous obtenons une bijection entre code et signal transmis et les seules contraintes à appliquer pour préserver la conformité de l'information sont :

- Pour le récepteur : partager le même espace de décision que l'émetteur, espace qui est représenté par une table de codage de symboles,
- Pour le réseau : garantir la conformité du signal en le discriminant du bruit.

Le clavier assure la segmentation de l'information en message du fait de la spécialisation des commandes. En l'absence de la médiation technique du clavier, les signaux émis par un utilisateur vers une interface doivent systématiquement faire l'objet d'un traitement logiciel qui revient à en analyser un flux pour en extraire des formes. Ces formes doivent être représentatives par rapport à un codage de l'information fixé a priori dans la machine.

---

<sup>5</sup> Puisque les commandes nerveuses de la phonation empruntent les mêmes voies que celles de la motricité volontaire, la parole peut être considérée comme une spécialisation de nos commandes haptiques.

### III.1. - L'interface homme machine comme canal de communication

La construction de modèles théoriques de l'information et de la communication a résulté d'un besoin éminemment pragmatique de déploiement économique de ces réseaux. En effet, tout réseau communique de l'information et il s'est rapidement avéré indispensable d'apprendre à quantifier et de qualifier cette "matière première" qu'est l'information.

Cette efficacité d'un réseau de télécommunications repose donc sur l'analyse de la quantité et de la vitesse des informations échangées dans chaque partie d'un réseau. La mesure de l'efficacité d'une interface passe par cette même logique si l'on admet le principe pragmatique qu'une interface acceptée par un utilisateur est le siège d'un échange rapide et facile d'un volume important d'informations entre l'homme et la machine.

Dans ce contexte, la mesure de l'efficacité de l'interface passe par la mesure de l'affordance des formes qu'elle a restituées ou capturées.

Quelques exemples simples viennent étayer cette approche pragmatique de l'évaluation des interfaces :

- **l'envoi de messages par SMS** : les contraintes apportées par le dispositif de saisie d'un téléphone mobile, associées à la taille réduite de l'écran, destinent ce genre de service à l'échange de messages courts, donc, de peu d'informations. En outre, l'accès indirect au codage des caractères fait que la vitesse de composition d'un message est lente, ce qui se traduit par un faible débit d'informations au niveau de l'interface.
- **les logiciels de reconnaissance de formes** : l'évaluation de l'efficacité d'un logiciel de reconnaissance de formes reste une problématique importante. Cette évaluation se heurte toujours à l'évaluation du contexte d'usage de la forme par son destinataire. Qu'est ce qu'un forme non ou mal reconnue ? Est-ce une véritable erreur logicielle ou la manifestation d'un comportement du programme dû à un contexte d'utilisation imprécis, voire absent ? Nous verrons par la suite, qu'il s'agit de critères d'évaluation d'un système de reconnaissance de formes. Ces critères sont exactement les mêmes que ceux de la définition d'un système de communication d'informations : une mise en forme saillante associée à un système de codage univoque de l'information.

Ainsi, que ce soit pour une saisie directe d'informations codées en caractères ou n'importe quelle action couplée à une reconnaissance de formes, nous sommes dans le même contexte de la communication d'une information dont la présentation est à quantifier et à qualifier par le récepteur.

Cette rationalisation de l'ingénierie des réseaux par application d'un principe d'économie, se base sur la vitesse de transmission de signaux ainsi qu'une adéquation complète entre les propriétés physiques du support de communication, sa capacité à être perturbé par un signal physique. Cette adéquation conduit à caractériser un signal par :

- sa mise en forme
- la détermination du code que transporte ce signal

Ce sont ces deux grandeurs, définies par Hartley dans les premières approches théoriques de l'information, qui déterminent le caractère de décidabilité d'un système de communication, à savoir :

- la capacité de différencier le signal (partie utile de la communication) d'un bruit (existant même en l'absence de toutes communications),
- la capacité de mettre en œuvre un code univoque dans lequel exempt d'ambiguïtés entre les différents états possibles d'un symbole et sa représentation sous la forme d'un signal.

Dans le contexte de l'interaction homme machine, nous retrouvons exactement les mêmes grandeurs fondamentales de caractérisation d'une information ainsi que la méthode de présenter celle-ci à l'utilisateur. En effet, la mise en forme d'un message dans une interface se fait à partir de signaux correspondant aux données. Elle permet à l'utilisateur de lire ce dernier pour le déchiffrer en se basant sur ses connaissances relatives au code utilisé pour composer ce message. La présentation de l'information est inscrite par un support adapté à une modalité d'interaction. Chez l'homme ces modalités correspondent à ses chaînes sensorimotrices à savoir, pour la capture d'informations :

- la vue,
- l'audition
- le toucher

NB : les autres sens ne font pas partie des moyens usuels de communication

Et pour la restitution d'informations :

- l'haptique
- le vocal

Ainsi pour chacune de ces modalités de communication, nous retrouvons les grandeurs de mise en forme d'un signal et la référence à un système de codage symbolique.

La communication vocale produit des phonèmes (formes de signaux), regroupés en allophones, qui correspondent schématiquement à des syllabes et font référence à une forme de table de codage de la parole. Entre deux humains, la communication peut être établie s'ils partagent un même ensemble de conventions de signes ou bien, s'ils sont capables de traduire une convention en une autre :

- le canal de communication est acoustique, à savoir qu'il est perturbé par des ondes sonores, ce qui définit le média physique utilisé pour la communication.
- les signaux transportés sont segmentés en phonèmes dont la mise en forme dépend d'un code symbolique. Le phonème définit le codage sémiotique de l'information, en ce sens que ce code est adapté aux propriétés de perception d'un signal par le destinataire.
- la symbolique ayant conduit à la définition du code sémiotique de la communication orale fait référence à des informations de syntaxe et donc, de sémantique.

Une communication vocale est, la plupart du temps, immédiate et temporelle, elle se fait en temps réel et manière séquentielle ; le destinataire doit "attendre" une interruption pour répondre à la source.

La lecture d'un livre, d'un document électronique ou la navigation sur le web, sont également des actes de communication mais les situations, contextes et propriétés sont différents.

- En premier lieu, nous observons une articulation de deux modalités : l'haptique dans le sens de l'émission et le visuel dans le sens de la réception.
- Ensuite, la communication est médiée, en ce sens que les formes sont inscrites ou enregistrées sur un support qui n'est autre qu'un document.
- Enfin, contrairement au vocal, la dynamique de lecture d'un document n'est pas séquentielle, mais entièrement pilotée par le lecteur destinataire de l'information.

Nous retrouvons là encore les notions de support physique porteur de l'information, de mises en forme de saillances attachées à une symbologie.

### **III.2. - Sémiotique et présentation de l'information**

Il est parfaitement possible d'étendre la triade sémiotique de Peirce pour le codage et de décodage de l'information dans n'importe quelle modalité de communication.

Nous aurons toujours des signaux codés pour perturber un milieu physique ; ces signaux sont porteurs de message dont l'interprétabilité (la décidabilité de l'information) dépend de leur mise en forme et de leur codage symbolique.

Ce codage sémiotique de l'information permet, en outre, de rendre compte d'une difficulté rémanente de la validation en reconnaissance de formes. En effet, une forme reconnue doit répondre à deux critères a priori antinomiques :

- l'information doit être renseignée, c'est-à-dire décrite dans un langage et un formalisme compatible avec les données calculables par une machine,
- l'information doit être présentée dans une forme interprétable par un utilisateur.

Le formalisme de la présentation de l'information est une question de grandeurs mesurables. Il s'exprime au moyen d'une syntaxe, celle du langage de description utilisé pour présenter une information. Elle est opérationnelle au travers de la capacité d'effectuer des calculs sur une description syntaxique d'une forme. Cette description syntaxique prend en considération les modèles de données et d'organisation des traitements d'une application logicielle.

Ces connaissances sur le langage et la programmation sont connues a priori. Par conséquent, la validation d'une forme reconnue sur le plan de sa documentation pour la machine, est déterministe. La finalité est de produire des objets techniquement manipulables par un programme logiciel. Cela signifie que le programme logiciel est conçu en prédéterminant un ensemble d'actions sur des objets porteurs de sens.

La construction d'une présentation de l'information fait intervenir, quant à elle, l'arbitraire de l'interprétation par un utilisateur humain. Ici, la validation nécessite de prendre en compte des critères qualitatifs dont la grandeur n'est pas quantifiable



par la seule approche technique, déterministe ou heuristique. L'arbitraire de l'interprétation sous-tend l'imprévisibilité de la valeur de la présentation d'une information.

#### ***IV. Résoudre les trois problèmes de Weaver reste d'actualité***

Dans l'introduction de la "théorie mathématique de la communication" de Shannon, Warren Weaver expose la question de la communication comme la résolution de trois problèmes dans lesquels nous retrouvons le paradoxe de l'arbitraire de l'interprétation confrontée au déterminisme du codage de l'information.

##### **IV.1. - Le problème technique de la communication**

Le "**problème technique**" est celui de la combinaison de signaux nécessaires pour communiquer une certaine quantité d'information. Ce "problème technique" est celui qui conduit à la définition d'une syntaxe de codage de l'information. L'objectif est de produire un système de codage non ambigu dans lequel, chaque élément syntaxique est une forme (ou un message) unique et indivisible parmi un ensemble de formes possibles regroupées dans un espace de décision. Cette notion de "problème technique" nous renvoie vers la production d'Unités Techniques de Manipulation si nous l'appliquons à l'interprétation de l'information dans une interface homme machine.

Ce "problème technique" qui est traité par Shannon établit des critères de décidabilité partagés entre la source et le destinataire d'une information. Ici, la conformité de l'information revient à préserver la décidabilité de la description d'une forme par rapport à un éventail de formes possibles.

La préservation de la capacité de discriminer une forme pour le récepteur revient d'une part, à connaître l'espace de décision utilisé par la source d'information pour composer son message et, d'autre part, à être capable de discriminer les formes représentatives (les segmenter) par rapport au bruit.

Ce bruit peut être induit par le canal lui-même, et correspond à un bruit physique. Mais il peut aussi être induit par les éléments constitutifs du système intervenant dans le codage/décodage de l'information. On parlera alors de bruit "sémantique" dans la mise forme des messages et de bruit "sémiotique" dans le codage de ces messages en signaux.

##### **IV.2. - Le problème sémantique de la communication**

Le "**problème sémantique**" est celui de la valeur symbolique susceptible d'être affecté à une forme. Il s'agira de la valeur d'un caractère, mais aussi, d'une métadonnée descriptive d'un contenu et donc, d'une manière plus générale, d'un ensemble de valeurs possibles d'interprétations. Ce problème sémantique n'est pas traité par Shannon qui limite la conformité de l'information à sa dimension technique de préservation des messages et de discrimination du signal par rapport au bruit.

*"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem" (Shannon).*

Cette approche place la sémantique d'interprétation en dehors de la question de la communication. En considérant l'interface homme machine comme support ou canal de communication, il est intéressant de voir quelles peuvent être les conséquences de ce confinement sémantique.

Une interface homme machine agit bien comme un système de communication. Soit la machine émet des formes qui font sens à l'utilisateur, soit l'utilisateur matérialise ses intentions par des actions qui sont également des formes analysées et interprétées par la machine.

Seules ces formes (ou messages) sont effectivement codées en signaux et non leur sémantique.

De même, les données d'un programme logiciel représentent une information et le programme logiciel, lui même n'est que la transcription d'une sémantique d'action et de traitement de l'information.

Cette sémantique ne se trouve pas exprimée de manière explicite dans les données, mais de manière latente par les modèles de données et de traitements ayant conduit à la production d'objets informatiques selon une certaine logique structurelle.

C'est donc l'ensemble de toute la structure logicielle et matérielle d'une application et de son interface qui représente une sémantique. Nous verrons par la suite que la grandeur quantifiable permettant d'exprimer cette sémantique correspond à la mesure de la complexité de ce système. Elle ne correspond pas à la seule validité des traces ou messages résultant de la mise en œuvre de ce système.

### **IV.3. - Le problème d'effectivité de la communication**

Weaver a eu l'intuition du "problème d'effectivité", mais l'a quelque peu éludé, en le plaçant dans le contexte abstrait de la psychologie, nous pouvons, toutefois, ramener ce problème à une dimension très technique, celle du lien entre la présentation et la valeur d'un l'information.

Pour ce faire, nous devons considérer la présentation de l'information comme étant la trace d'une organisation et d'une activité inscrite sur un support physique. S'il s'agit d'un document, ce sera, par exemple, la charte graphique de ce dernier, s'il s'agit d'un comportement de l'utilisateur, ce sera la mesure de cette dernière au travers d'une action.

Dans ce contexte, le "problème d'effectivité" devient alors celui de la relation existante entre une forme (une Unité Technique de Manipulation) et une valeur (une Unité Sémiotique d'Interprétation), cette valeur. A cette valeur est associée l'énergie dépensée pour l'interprétation d'une forme dans une application.

### **IV.4. - Reformulation sémiotique des trois problèmes de Weaver**

Les trois problèmes de Weaver évoquent également la triade sémiotique de Peirce en ce sens que

- le problème sémantique traite des objets à mettre en forme,
- le problème technique traite des messages porteurs d'informations,
- le problème d'effectivité traite de la relation entre formes/messages et valeurs symboliques.

La communication est désormais devenue multisource, multidestinataire, multimédia et multimodale. Elle place le système de communication dans le contexte de la complexité en raison de la diversité. Cette diversité nécessite de comprendre non seulement les méthodes de production de signaux, mais également, celles de production des messages pour identifier les éléments structurants permettant d'appréhender cette complexité.

Ces éléments structurants sont les suivants :

- Une communication homme machine (mais aussi, machine machine) est portée par un système composé de deux sous-systèmes matérialisés par deux canaux de communication, l'un correspondant à une modalité de communication de la source, l'autre correspondant à une modalité de communication du destinataire.
- L'interface entre ces deux sous-systèmes a pour objectif de transcoder la présentation de l'information en des messages perceptibles par chacun de ces deux sous-systèmes.
- L'information est codée mise sous la forme de signaux représentatifs d'un symbole, ce qui conduit à définir l'information comme une combinaison d'une description physique (le support), sémiotique (les formes) et sémantique (la symbolique d'interprétation).
- Les seuls éléments réellement communiqués dans le système sont les formes. La détermination de l'interprétabilité d'une forme-message correspond à des facteurs de décidabilité du codage sémiotique de l'information. Par rapport aux trois problèmes de Weaver, c'est ici que nous trouvons la formulation technique du problème d'effectivité.

## ***V. Information et cybernétique***

Si l'on observe les théories de l'information sous l'aspect système, on constate une évolution importante qui n'est pas encore bien assimilée dans la conception de nos services, applications et interfaces de manipulation de contenus numériques.

Cette évolution concerne la prise en compte de l'information sur l'état du système avant l'interaction (information de feedforward) et de contrôle d'une action sur une interface, autrement dit, l'information de feedback. Il s'agit donc de la dimension cybernétique d'une interface homme machine. Cet aspect est important pour la présentation de l'information, car elle conditionne la planification d'une action et son déroulement. Du fait qu'une action est une information apportée au système et qu'elle porte sur les formes de l'interface, il convient de revenir sur les systèmes à boucles de contrôle.

On peut schématiquement classer ces systèmes en systèmes à boucles ouvertes ou à boucles fermées. Selon la conformation du système, l'information de contrôle va concerner la planification d'une action ou son exécution.

Un système à boucle ouverte est un système dans lequel un processus de traitement de l'information s'exécute de manière séquentielle. L'information de feedback signale la fin de l'exécution d'une tâche. On parle de feedback à longue période. A une action correspond une réaction, mais la réponse du système ne permet pas d'assurer le suivi fin d'une action.

Dans un système à boucle fermée, l'information de feedback module continuellement l'exécution d'une action. On parle de feedback à courte période. C'est le complémentaire du feedback d'une action de planification, nécessaire au contrôle et à la modulation de l'ensemble du processus mis en œuvre dans tout système.

La systématisation de l'interaction homme machine sous l'angle cybernétique est à la fois possible en raison de l'évolution des techniques matérielles et logicielles pour le développement d'interface et nécessaire pour mettre en adéquation l'interaction homme machine avec les propriétés de nos propres effecteurs biologiques.

Notre système physiologique est un ensemble de sous-systèmes à boucle fermée qui permet de préserver sa stabilité (température : homéothermie, équilibre chimique : homéostasie).

Notre système nerveux moteur, quant à lui a une organisation fonctionnelle mixte adapté à la nature du geste effectuée.

Les gestes balistiques, comme par exemple, un retrait réflexe, ou une saccade oculaire sont pilotés en boucle ouverte : l'information de feedback est un "acquiescement" de fin de geste

Les gestes dits "en rampe" sont des gestes associés à un feedback continu. L'information de contrôle a, schématiquement, deux origines :

- origine interne : l'information de tonus musculaire (feedback proprioceptif),
- origine externe : l'ensemble des informations résultant de notre perception multimodale de l'environnement via nos sens (feedback extéroceptif).

Cette organisation fonctionnelle de notre système nerveux est illustrée dans la fig.14.

L'information de contrôle (feedback et feedforward) donne au système sa plasticité d'adaptation ; elle permet de moduler ou de réguler son comportement, notamment dans un environnement non déterministe. En partant d'un ensemble de comportements prévus a priori à l'état initial du système, capturer l'information de contrôle autorise une mesure de l'activité effective de ce système, en particulier, lors de la survenue d'un évènement imprévu.

L'analyse de cette information de contrôle aboutit à des formes issues du contexte de l'interaction ; leur interprétation et leur inscription dans la mémoire du système constituent un apprentissage incrémental et actif.

De tels systèmes, dotés de fonctions de traitement de l'information de contrôle, peuvent être qualifiés de stochastiques (c'est-à-dire, dont le comportement est dynamique et lié à un **contexte**).

Dans le cas particulier de la présentation de l'information numérique, le travail de détermination a priori des contextes sont corrélé à de fortes limitations relatives aux capacités d'accès et de traitement de l'information. En outre, le codage hétérogène de l'information induit une complexité ayant des conséquences importantes pour la définition des modes d'accès à l'information.

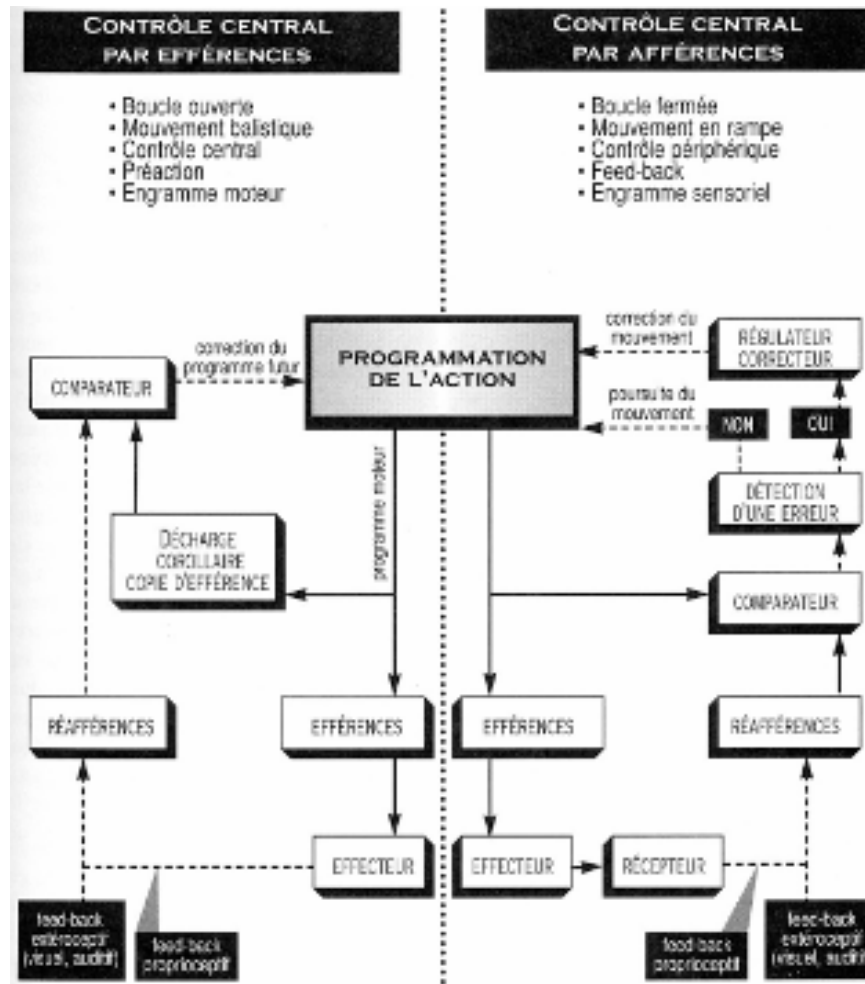


Fig. 14 : Organigramme fonctionnel du système nerveux humain, mettant en évidence des deux modes de contrôle de nos chaînes sensorimotrices

Par conséquent, l'apprentissage actif à partir de la mesure de l'activité d'un système cybernétique d'interface via l'information de contrôle, représente une solution complémentaire, voire substitutive à un système déterministe, basé sur un modèle de connaissances construites a priori.

### V.1. - L'interface homme machine et la dynamique de l'interaction

Le rôle de l'arsenal logiciel mis en œuvre par le destinataire est prédominant pour la détermination du sens et de la valeur d'une information. L'utilisateur ne perçoit les objets que grâce à ses outils d'observation et, inversement, les outils d'observation induisent des présentations de l'information qui en influencent fortement son interprétation.

Ainsi, ce n'est pas seulement la composition d'un document qui conditionne le sens de l'information, mais les conditions de perception et d'analyse de ce document par le destinataire utilisateur ; ces conditions de perception et d'analyse sont données par la puissance de calcul disponible sur son terminal.

Ceci évoque à nouveau, une intrication, au sens de la mécanique quantique, existant entre un observateur, le phénomène qu'il observe et les moyens d'observation accessibles à cet utilisateur. Cette intrication met en évidence que l'arsenal logiciel du destinataire contribue fortement à la détermination de la valeur et du sens d'une

information. Le contexte d'une diversité combinatoire (multisource, multimodale et multimédia) des modes de codage de l'information ne fait qu'accentuer cette intrication. Il y a un réel effet analogue au "chat de Schrödinger" car :

- l'émergence du sens n'intervient que lors de l'interaction (l'observation) des messages issus du système,
- l'affordance d'un message est conditionnée par sa présentation telle que prévue par la source,
- l'outil ou composant d'interface permettant l'observation des messages du système conditionne leur interprétation.

Ainsi, la dynamique des processus de capture ou de restitution d'une présentation de l'information semble tout aussi importante que les descripteurs documentant cette présentation. Or ces processus décrivent la quantification et le codage en formes ou messages. Ils peuvent être considérés comme une projection d'un phénomène observé dans un espace de représentation, correspondant à un espace de phases d'un système dynamique.

Les axes de mesures de cet espace de phases correspondent quant à eux, les descripteurs propres à chaque modalité concernée : Les systèmes de l'émetteur et du récepteur peuvent être modélisés par ces espaces de phases

En réception, le destinataire reçoit des signaux qui sont autant de porteurs de messages utilisés par lui, pour reconstruire l'information. La reconstruction des messages à partir des signaux et la reconstruction d'une information conforme au phénomène décrit par la source sont les critères de réussite de la communication entre source et destinataire.

On voit de suite que la nature des axes de l'espace de phases introduit ici, va conditionner fortement la conformité de l'information reçue.

L'interaction homme machine peut être réduite à un système déterministe si, et seulement si, les espaces de phases de la source et du destinataire sont identiques. C'est-à-dire, que les référentiels de présentation de l'information sont communs.

Dans le cas d'une communication où la connaissance de ces valeurs n'est pas certaine, nous passons à un système non déterministe dans lequel on pourra mettre en œuvre différentes approches statistiques, probabilistes ou stochastiques sur l'espace de phase et sur la fonction de calcul de la forme perçue par le destinataire.

Dans le développement qui suit, nous allons considérer que cette fonction de calcul est portée par une machine de Turing universelle, l'objectif étant ici de mettre en évidence des questions de complexité et de calculabilité d'une forme par rapport au contexte du destinataire de l'information.

Ce modèle conceptuel permet simplement de dire qu'il existe une fonction de calcul sur les signaux de la communication, sans préjuger a priori de sa nature. Ce calcul est un interprétant au sens sémiotique.

L'existence de cette forme correspond à sa calculabilité et sa tangibilité à l'interface dépend de la complexité de ce calcul. Cette complexité, quant à elle se déduit de l'observation des états initiaux, finaux et de transition d'un système dont les degrés de libertés correspondent aux différents modes de présentation de l'information.

Ce modèle inclut la complexité de calcul comme valeur quantitative de l'évaluation d'un traitement d'information. Cette complexité ne fait pas partie aujourd'hui des paramètres utilisés pour l'évaluation, par exemple, d'un résultat de reconnaissance de formes. Cependant, il s'agit d'un paramètre important car il conditionne l'acceptabilité d'un traitement, qui se situera dans une marge de valeurs donnée par des critères d'ergonomie et d'utilisabilité dans une interface. Ce paramètre de complexité permettra, en outre, de mettre en évidence des situations indécidables, c'est-à-dire, non traduisibles en un calcul, par application du Théorème de Rice (dit "problème de l'arrêt")

L'inclusion du calcul et la mesure de sa complexité nous conduit à un modèle dynamique du traitement de l'information, car il fait intervenir conjointement des complexités en espace et en temps, notions qui seront développées au chapitre 5. Le modèle proposé ici (schématisé en Fig. 15) est une hypothèse de travail pour de futures études.

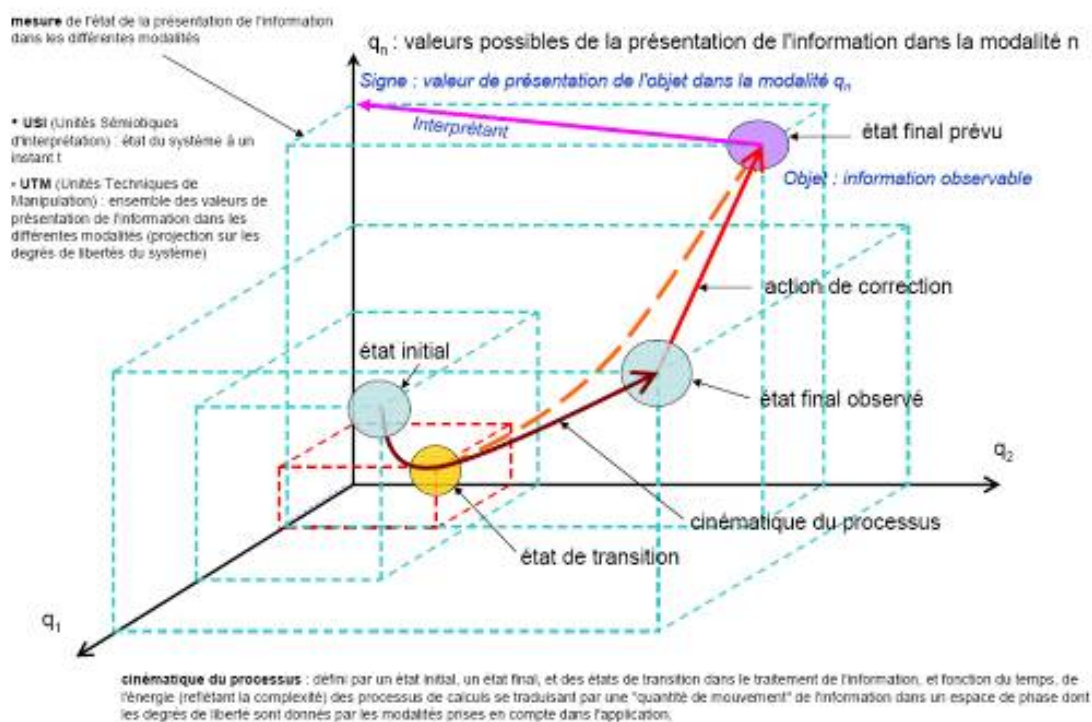


Fig. 15 : Une proposition de modèle de système dynamique de traitement de l'information dans lequel, un objet  $n$  a de sens et de tangibilité qu'au travers de ses projections sur les degrés de liberté du système. Ces degrés de liberté correspondent aux modes de présentation de l'information et sont ordonnés selon les règles de codage de l'information.

Ainsi pour la question de la validation d'une reconnaissance de formes qui, nous le rappelons, revient à définir la validité d'une solution à un problème de décision, nous nous rendons compte que la mesure de cette validation fait bien intervenir les deux critères d'arrêt et de complexité de cette fonction :

- L'arrêt consistera en la mesure des états finaux du système (s'ils existent) et de la mesure de conformité ou de confiance par rapport à l'espace de décision du récepteur.
- La complexité sera celle de l'ensemble des fonctions logicielles mises en œuvre pour aboutir à un ensemble d'états finaux, à partir des états initiaux d'un processus. Cette complexité est la représentation technique de critères ergonomiques utiles à la conception d'une interface ; elle permet de quantifier l'acceptabilité d'une application logicielle portant le processus de calcul, en termes de temps d'exécution, de charge de la machine et de fiabilité des résultats par rapport au problème de l'arrêt.

Si l'on choisit de développer une interface à ergodicité faible, un résultat fiable mais nécessitant une trop forte complexité de calcul par rapport à un usage et une situation donnée, correspondra à un état inacceptable du système.

On recherchera alors, à partir de cette classification par la complexité, des états de transition plus ou moins coûteux en calcul (comme par exemple la substitution d'une action de la machine par une action de l'utilisateur) aboutissant au même état final.

Si par contre, on choisit de développer un système à ergodicité forte, nous lèverons la contrainte de l'économie logicielle pour privilégier la robustesse des états finaux.

## **V.2. - Le confinement de la sémantique en dehors des interfaces**

En traitement numérique de l'information, un message correspond à une quantification et/ou une description d'une information, quelle qu'elle soit.

Cette quantification peut se faire dans diverses modalités de présentation de l'information. Le résultat en est une forme ou message résultant de la projection d'un objet (au sens cognitif du terme) dans une modalité donnée, dont les états possibles peuvent être rapportés à un axe d'un espace de phases ; cet espace de phases correspond à l'ensemble des modalités accessibles à la source ou au destinataire.

Les valeurs possibles de chaque axe correspondent aux conventions de codage de l'information établies à priori lors de la conception d'une application. Dans le strict cadre de la communication, l'établissement de cette convention de codage des messages se base sur la nature des combinaisons de signaux à transmettre via le canal de communication.

### ***V.2.1. - Discussion par rapport à la sémantique***

Nous avons vu que la conformité d'une communication apportée par les fonctions de codage Messages/Signaux n'aborde pas la question de la "conformité sémantique" (Shannon).

Les trois problèmes de Weaver sur la communication confirment ce confinement de la sémantique en dehors de la communication via une interface homme machine.

#### **Le problème physique :**

- C'est celui qui, à partir des contraintes de conformité physique de communication de l'information à travers des signaux, conduit à définir à priori une table de codage des messages :



- Dans le cas d'un texte brut, il s'agira des tables de caractère. En ASCII ce sont les différentes pages de codes et en UNICODE les valeurs abstraites des caractères. On rappellera ici qu'en ASCII, le mode de présentation d'une forme est imposé et unique, à chaque caractère, on associe un symbole ; en UNICODE, ce mode de présentation est défini par la relation entre la valeur abstraite d'un caractère et le symbole choisi.
- Dans les autres cas, on parlera de primitives morphologiques (morphèmes) et des propriétés affectées par la mesure de l'objet lui conférant sa présentation.

#### **Le problème sémantique :**

- L'organisation des états possibles de la valeur d'une forme dans une modalité est la trace de la sémantique de l'application considérée. Le sens n'est pas décrit en tant que tel, mais guide la structuration de cet axe de représentation. Si la forme représente de l'information, la structure, l'organisation des valeurs possibles pour cette forme, représente les connaissances et la sémantique.

#### **Le problème d'effectivité :**

- Dans le schéma de présentation que nous proposons, l'objet "information" n'existe que si l'on a la capacité de le projeter dans un espace de représentation, et, inversement, qu'on soit capable d'approximer cet objet après interprétation des messages constitués lors de l'émission par la source.

La réception d'un message directement identifiable par le récepteur par rapport à son espace de décision, ne remet pas en cause l'entropie du récepteur.

Par contre, la réception d'un message non identifiable, est un facteur d'augmentation de l'entropie ; son apprentissage par le récepteur est un facteur de négentropie. Le mécanisme d'apprentissage de tout système agit donc comme un démon de Maxwell, en ce sens, qu'il agit comme un observateur capable de décider d'une nouvelle action par rapport au contexte de réception de ce message. Cette nouvelle action consiste à apporter de l'information (négentropie, selon Brillouin) au système.

En reconnaissance de formes, une erreur de reconnaissance accroît l'entropie du système de reconnaissance, l'expression de sa correction et son inscription dans l'espace des possibles la fait décroître.

Les erreurs de reconnaissance sont principalement dues à trois causes :

- formelles, liées à la quantification (acquisition, segmentation)
- structurelles, liées à l'organisation des états possibles (classification)
- efficaces, liées à l'instrumentation d'observation et de mesure (la projection sur un axe de représentation).

Faire baisser l'entropie d'un système revient à lui apporter de l'énergie sous la forme d'un calcul exécuté par un programme logiciel, lié ou non à des actions utilisateur.

Ce programme logiciel a pour objectif d'une part, d'identifier des combinaisons de signaux révélateurs de nouveaux messages et, d'autre part, d'insérer tout nouveau

message dans la structure de l'espace de décision de présentation de l'information selon la modalité donnée. La quantité d'énergie correspondant à ce calcul est en relation directe avec la complexité de ce dernier.

Selon le principe d'économie, on déterminera des seuils établissant une plage de complexité acceptable (critères de temps, de bande passante, par exemple) qui conditionneront le choix de la méthode de calcul la plus économique par rapport au contexte d'utilisation d'une application.

Ainsi, en partant de l'hypothèse de la sémantique comme élément structurant de l'organisation de ce système, on verra transparaître des indices structurants traduisibles en autant de paramètres de fonctionnement d'une application. Ces paramètres ne dépendent plus de la nature technique des données, mais des usages permis d'une application traduits en fonctions portées par des algorithmes.

La question sera ici celle de l'instrumentation de ces indices structurants.

Le calcul permettant de prendre une décision sur la valeur d'un message perçu dans un système de communication est porteur de la sémantique, au même titre qu'un signal est porteur d'un message.

Nous pouvons dire que la sémantique n'a pas d'existence propre au système, mais qu'elle en conditionne son organisation. En ce sens, l'organisation en données et programmes représente les traces de la sémantique.

Aussi, la question qui se pose ici est de savoir comment compléter le critère de conformité d'une communication d'une information en incluant la sémantique.

Pour parler de conformité sémantique, on doit faire intervenir l'utilisateur qui est en situation de dialogue avec une machine. Cela signifie un canal de communication différent de celui utilisé par la machine pour l'application et son interface, et dans lequel nous devons identifier :

- Des signaux capables de stimuler une chaîne sensorimotrice du sous-système "homme"
- Des signaux capables d'induire une action de la part du sous-système "machine"

### ***V.2.2. - Vers des axiomes de la présentation de l'information***

Afin de tenter de systématiser et d'axiomatiser la présentation de l'information dans le contexte de la complexité, il nous faut extraire des références à partir de l'évolution des théories de l'information. On peut déterminer trois grands jalons historiques permettant de dégager progressivement un cadre systémique au concept de la présentation de l'information. Ces grands jalons sont :

- **L'approche mathématique de la théorie de l'information (Shannon-Weaver, Wiener)**

La théorie mathématique de la communication (Shannon, Weaver) met en place la notion d'information comme grandeur calculable et catégorisée dans un espace de représentation/décision.

L'ordonnement de cet espace et la complexité de prise de décision pour la détermination de la valeur de l'information sont représentés par l'entropie de l'information. La communication de l'information repose sur un canal perturbé par

des signaux résultant d'un codage adapté de l'information aux propriétés du canal, effectué par une source et d'un décodage effectué par un récepteur. Dans le canal, seuls des signaux codés représentant l'information sont véhiculés et présentés au destinataire.

Un tel schéma s'applique de plein droit à une interface homme machine dans laquelle nous avons une source et un destinataire qui, selon la phase d'interaction, peut être soit la machine (en fait, l'application hébergée par la machine), soit un utilisateur.

La modélisation d'une interface homme machine par le modèle du canal de communication, montre qu'une interface met en œuvre au minimum deux canaux, du fait de la nature physique des signaux véhiculés lors d'une interaction. Nous avons des signaux représentant des données qui sont des séquences binaires et des signaux représentant des formes perceptibles par l'utilisateur. L'association de ces canaux constituera le problème à résoudre.

De plus, une interface n'est pas seulement un dispositif de saisie, elle est aussi le siège d'une boucle d'interaction entre l'homme et la machine ; l'efficacité d'un dialogue homme machine dépend des points de contrôle (feedbacks et feedforwards) qui agissent comme des régulateurs de ce dialogue. Ceci nous rapproche de la cybernétique de Wiener, du fait de ces informations venant agir sur l'exécution des programmes logiciels intervenant dans l'interaction homme machine.

- **L'approche thermodynamique de la théorie de l'information (Brillouin, Szilard,...)**

Cette théorie a commencé à faire le rapprochement entre l'information et la physique. Ici la thermodynamique est plutôt considérée comme un modèle analogue à celui de l'information. L'information est vue ici comme une variable d'état dans un système physique et l'énergie de ce système est l'instrument de mesure permettant de déduire les valeurs de ces variables d'état. Elle permet de prévoir ou d'évaluer les propriétés du système ou de ces composants en reprenant des lois fondamentales de la thermodynamique, tout en mettant en avant des paradoxes, comme par exemple, celui de "l'exorcisation du Démon de Maxwell".

L'information apporte une énergie au système qui tend à faire baisser l'entropie, ce qui fait qu'elle porte le nom de néguentropie chez Brillouin.

- **Les théories modernes de l'information (von Neumann, Landauer, Bennett, Feynman,...)**

Les théories de l'information antérieures reposaient sur une analogie avec la physique newtonienne et le réductionnisme de Descartes ; tout traitement de l'information pouvait se ramener à l'exécution d'un programme séquentiel sur une machine et un système d'information pouvait être ramené à un ensemble de machines ayant chacune des fonctions et buts déterminés à partir desquels il était possible de déduire des lois régissant le comportement de ce système.

Cette approche atteint ses limites si l'on observe le web. Nous y rencontrons des propriétés surprenantes de valeurs simultanées d'informations (des informations contradictoires).

En outre, la présentation de l'information est devenue multimédia et multimodale ; une même information s'instancie sur le web dans différentes modalités d'usages et

différents supports et média. La recherche exhaustive d'informations sur le web montre la nécessité de mettre en œuvre une large panoplie d'outils capables de décoder ces différents modes de présentation.

Ce problème de la valeur sémantique de l'information, associée à celui de sa présentation évoquent deux aspects imprévisibles de la mécanique newtonienne.

L'information a un comportement conforme à l'expérience de pensée du chat de Schrödinger. Sa valeur n'est plus déterministe, du fait que sur le web, nous avons une démultiplication des sources susceptibles de produire chacune des valeurs de l'information qui leur sont propres. Ainsi, tant que l'on n'a pas observé cette information, sa valeur ne peut être déterminée.

En outre, nous avons une authentique intrication entre l'information et les instruments d'observation. On peut très facilement illustrer cette propriété, également non prédictible par l'analogie newtonienne avec l'exemple des sites Wikipedia indiqués en résultats de requêtes Google. En effet, selon le paramétrage de la langue dans le navigateur web, la liste de résultats nous fera pointer vers différentes instances de Wikipedia dans lesquelles, il n'y a aucune uniformisation des contenus d'une langue à l'autre.

Un autre exemple est donné par le problème de la validation d'une reconnaissance de forme ; les causes de reconnaissance erronée dépendent de l'ensemble des étapes du processus de reconnaissance et ces processus ont des relations de dépendances, notamment en termes de valeurs de paramètres, qui ne sont pas triviales à établir.

Ces situations montrent les limites des théories newtoniennes de l'information. Actuellement, la modélisation théorique de l'information privilégie une approche par les systèmes complexes en lieu et place des approches réductionnistes.

En outre, notamment avec Von Neumann, Landauer et Bennett (voir chapitre 4), l'information est désormais vue comme étant un phénomène physique mesurable dans lequel, le calcul et sa complexité représentent un instrument de mesure à part entière. En particulier, nous verrons que la réversibilité du calcul proposée par Bennett représente un outil de mesure ayant de grandes potentialités pour le traitement d'une information sans connaissances a priori.

## VI. Systémique de la présentation de l'information

### VI.1. - La présentation de l'information et la multimodalité

Rechercher de l'information par le contenu pose le problème de la multimodalité de présentation de l'information ; ce problème n'est pas résolu aujourd'hui. Il est contourné par un important travail éditorial de construction d'un jeu de métadonnées qui serviront à l'indexation. Ce travail d'indexation manuelle nécessite des applications pour des utilisateurs experts, en particulier dans l'évaluation de la qualité des métadonnées produites. Ce travail éditorial revient à affecter des valeurs symboliques à des formes. Il porte plus sur le contenant lui-même que sur le contenu. Il s'agit d'une description globale des éléments présents dans le contenu mais, la plupart du temps, il n'y a pas de relation entre une métadonnée et la forme que décrit cette métadonnée ; ceci est particulièrement vrai pour les contenus de type images numériques.

Il est intéressant de discuter de cette question à l'aune de la norme MPEG-7. En effet, dans la déclinaison des normes MPEG (Motion Experts Picture Group, norme fortement structurée de codage des contenus multimédias), MPEG-7 représente la couche de spécification et de standardisation qui traite précisément du problème de l'effectivité de Weaver, à savoir, la relation entre une forme et son sens en contexte. On peut présenter la "pile de normes" MPEG de manière très synthétique comme dans le tableau ci-après.

Niveau	Résumé des spécifications principales
MPEG-1	<ul style="list-style-type: none"><li>- Norme "racine" de stockage/diffusion (partie 1), codage compressif vidéo (partie 2) et audio (partie 3) plus tests de conformités (partie 4) et logiciels de référence (partie 5).</li><li>- Principalement dédiée au support CD-vidéo (débit 1,5 Mbits/s)</li></ul>
MPEG-2	<ul style="list-style-type: none"><li>- Extension de MPEG-1 pour les parties 1 à 5 (notamment la télédiffusion numérique hertzienne et satellite et le support DVD)</li><li>- Gestion d'interfaces temps réel pour le téléchargement par exemple, (parties 6, 9 et 10) et codage avancé audio AAC (Dolby surround, multicanal)</li></ul>
MPEG-4	<ul style="list-style-type: none"><li>- Extension et enrichissement de l'offre de codeurs décodeurs (prise en compte de la haute définition et transport sur IP) par rapport à MPEG-2</li><li>- Multiplexage de scènes vectorielles et multimédia : texte, graphismes, toutes représentations symboliques de l'information.</li><li>- Métadonnées de protection intellectuelles</li></ul>
MPEG-7	<ul style="list-style-type: none"><li>- Norme pour l'indexation des flux et objets MPEG-4 (descripteurs visuels et audio, décomposition en séquences spatio-temporelles, informations contextuelles et contextuelles)</li><li>- Langage définissant des descripteurs (syntaxe des différents objets), des schémas (structure et relation entre objets et aussi entre schémas), dérivé d'XML</li></ul>
MPEG-21	Norme définissant un système interopérable pour l'utilisation transparente de représentations audiovisuelles numériques.

Cette description simplifiée de la norme MPEG, dont la philosophie se retrouve également dans la définition de la norme JPEG<sup>6</sup> (Joint Photographic Experts Group), notamment dans sa déclinaison JPEG-2000, révèle trois principaux sujets qui correspondent aux trois problèmes de Weaver :

- Le problème technique, qui est celui du codage à proprement parlé des contenus. Celui-ci est d'avantage vu sous l'angle du stockage et de la diffusion (compression, gestion des flux) que sous l'angle de l'interaction. Il s'agit des normes MPEG 1 à 4.
- Le problème sémantique, qui est celui de la définition de la valeur affectée aux objets est quant à lui, décrit dans la norme MPEG-7. Cette norme fournit l'ensemble des syntaxes permettant l'organisation des flux multimédias par rapport à une application donnée, sur un terminal donné.

Il reste le problème d'effectivité qui n'est pas abordé dans MPEG, mais qui correspond pleinement à la relation entre un objet d'un flux MPEG-4 et une donnée d'indexation en MPEG-7.

La présentation d'une information dans un flux MPEG-4 peut se faire dans n'importe quel mode entre l'image, le graphisme ou le texte, ce qui donne à cette norme une dimension multimodale dans laquelle nous retrouvons la question du recouvrement sémantique imprécis propre à la multimodalité, soulevé par Sharon Oviatt.

Par ailleurs nous constatons une très faible utilisation de MPEG-7, il n'existe encore aucune application grand public utilisant ce niveau de normalisation pour l'indexation.

Deux hypothèses permettent d'expliquer les causes de cette relative désaffection de MPEG-7 :

- Certaines parties de MPEG-4 ne sont pas encore stabilisées, notamment celles concernant les spécifications de codage du texte et des graphismes et, actuellement, MPEG-4 reste perçue comme la "norme de codage audiovisuelle de la télévision haute définition", malgré les extensions portant sur le type de contenus.
- L'indexation de tout objet, faite en respectant les spécifications MPEG-7, nécessite de pouvoir identifier cet objet dans n'importe quel flux (audio, vidéo, texte, graphismes). Dans le strict cas de l'audiovisuel, nous avons des mécanismes de synchronisation des flux qui assurent la relation temporelle entre l'audio et la vidéo.

---

<sup>6</sup> JPEG est historiquement la norme de codage de l'image numérique fixe, mais on constate une convergence dans la définition de cette norme par la prise en compte dans JPEG2000 du format Motion-JPEG dans la partie 3 des spécifications. Il faut noter que ce travail de normalisation décrit, non seulement le codage compressif de contenus, mais également les spécifications de synchronisation de contenus multimédias, de stockage, de diffusion et de gestion des droits. Les questions de traitement de l'information apparaissent si l'on considère une norme sous un angle systémique, à savoir, en observant les corrélations et dépendance entre chacune de ses parties.

## **VI.2. - L'indexation multimédia par alignement sur un document statique**

Lalanne, Ingold et al. proposent de partir d'un document statique (par exemple, les transparents projetés lors d'une réunion), comme base d'alignement des contenus audio et vidéo en recherchant des occurrences de mots saillants issus de la reconnaissance vocale, ou de la reconnaissance de caractères. Cette approche est intéressante, car elle permet de travailler dans une logique de "wordspotting" qui permet de ne pas à faire de reconnaissance exhaustive des textes présents dans les flux audio et vidéo. Cette technique revient à former un motif (forme ou "pattern") de mots et à en rechercher les occurrences dans chacun des modes de présentation de l'information dans un contenu multimédia. L'un de ces contenus est un document statique, principalement associé au contenu multimédia, mais pouvant être également multiplexé en MPEG-4. Ce document offre par définition une information de segmentation qui est celle du texte du document et de sa structure. Cette présentation de l'information est un ensemble d'Unités Techniques de Manipulation qui font sens pour l'indexation en raison même du codage des documents : le codage d'un document textuel est computationnel car symbolique.

L'alignement de contenus multimédias sur un document statique est principalement temporel, car il permet de savoir à quel moment telle expression ou tel mot a été évoqué. A cet alignement temporel, peut être associée l'identification ou la reconnaissance du locuteur et toute une série d'informations de contexte qui fournissent ensemble des indices sémiotiques dont la mise en relation constitue l'indexation du contenu.

Ici le problème d'effectivité est résolu en prenant comme règle a priori le fait que le document statique contient une présentation une information symbolique, compatible avec l'indexation : le texte et sa structure.

## **VI.3. - L'indexation multimédia par alignement transmodal**

La détection d'une forme recherchée dans un flux multimédia pour son indexation et l'accessibilité à un document statique contenant le texte servant à l'alignement des contenus ne résout que partiellement le problème d'effectivité posé par une indexation croisée de tout contenu. Le document statique peut être absent ou non accessible à l'enregistrement d'une scène multimédia.

Si l'on veut étendre cette stratégie d'alignements de contenus dans n'importe quelle modalité, il devient nécessaire de passer par une recherche de contenus via leur présentation sémiotique.

Ceci nécessite de trouver une méthode de calcul d'une distance entre informations et non plus seulement entre descripteurs d'informations segmentés a priori. Cela signifie, trouver un espace de représentation dans lequel le calcul de distance reste applicable à n'importe quel mode de présentation de l'information.

En effet, une véritable recherche d'informations par le contenu nécessite de pouvoir désigner n'importe quel fragment d'un document multimédia ou bien, de construire un mode de présentation de l'information, puis de rechercher les occurrences de ce fragment dans un corpus.

Cependant, le codage des contenus reste un écueil important qui interdit actuellement l'exhaustivité d'un résultat de recherche. La raison en est la difficulté de transposer contextuellement une information d'un mode de représentation à

l'autre<sup>7</sup>. Nous avons vu que le passage par les métadonnées résout partiellement cet écueil en posant sous la forme d'heuristiques, la question de la préexistence des formes symboliques structurant un alphabet et un vocabulaire. Mais cet écueil persiste dans les cas suivants :

- au niveau physique de codage du contenu : un document peut se présenter sous la forme d'un tableau de valeurs (les pixels), d'un texte, ou de vecteurs (contenus graphiques). Une même information peut être présentée dans chacun de ces modes portés par un même contenu.
- au niveau sémiotique, qui est la présentation de l'information : dans le mode image, se pose toujours le problème de la segmentation du contenu pour en extraire des formes. Dans une image brute, la seule forme techniquement manipulable est le pixel. De plus, une requête est susceptible de rassembler plusieurs formes pour en faire un motif ou "pattern", au quel cas, une information de voisinage topologique fait également partie de la requête.

Ceci montre les difficultés posées par la généralisation de l'indexation par alignement de modalités en partant du problème de la construction de requêtes textuelles pour l'indexation par alignement sur un document statique : un texte dans un document à une structure physique et logique. Ce problème se retrouve également dans la segmentation et la construction de topologies faisant sens entre des primitives géométriques telles que lignes, polygones, surfaces formant des objets graphiques, donc des formes vectorielles.

#### **VI.4. - Un système où la place de l'homme est située**

En revenant aux trois problèmes de Weaver, nous trouvons une généralisation du concept de forme que l'on peut définir ainsi :

*Toute forme est une présentation de l'information décrite par des données adressables par la technique pour le calcul et l'interaction.*

L'interaction elle-même est représentée par une forme issue d'un calcul et résultante d'une action de l'utilisateur. Cette action est consécutive à une décision prise par ce dernier à partir de l'information qu'il a reconstruit par sémiologie. Cette sémiologie est l'intégration d'informations faite à partir des signes ou messages interprétés émis par l'interface du terminal, par le terminal lui-même et par l'environnement de l'utilisateur.

La décision de déclencher une action repose sur l'affordance (Gibson, Norman) de la présentation de toute l'information que perçoit l'utilisateur, par rapport à ses connaissances ou croyances. Parmi ces présentations de l'information qui forment le contenu de l'interface, nous trouvons des présentations d'objets correspondant à de

---

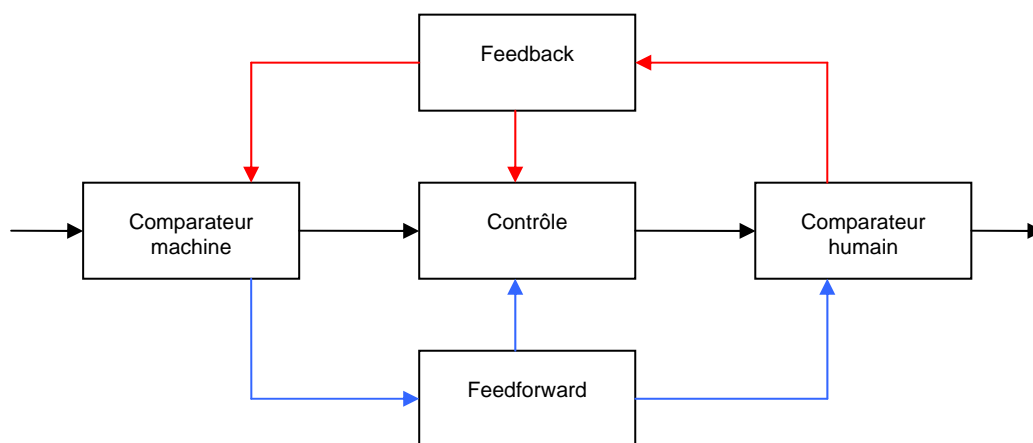
<sup>7</sup> Cette difficulté explique pourquoi la numérisation de documents entre totalement dans le périmètre des interfaces et de la présentation de l'information : au-delà de sa dématérialisation, numériser un document revient à construire des objets informatiques décrivant son contenu et manipulables via une interface. Il s'agit donc de construire une tangibilité numérique du contenu.



l'information (un texte, une image, un son, etc.), ou à des commandes (les actionneurs de l'interface). A chacune correspond un ensemble de saillances qui évoquent (ou n'évoquent pas) quelque chose à l'utilisateur dans son interprétation.

Dans un système homme machine, les "objets" de l'utilisateur ne sont pas directement adressables par la technique. Par conséquent, l'information provenant de l'utilisateur ne peut être obtenue que par une mesure de son activité. Ainsi, le seul moyen de vérifier une affordance et donc, de valider une présentation de l'information calculée à partir de données, passe par l'observation et la mesure de ce que fait l'utilisateur : il s'agira donc d'une mesure d'un feedback humain, qui peut être vu comme le symétrique du feedback de la machine. Cette mesure du feedback revient à analyser le comportement au travers des traces d'activité de l'utilisateur.

On peut représenter le schéma générique de toute interface homme machine, incluant cette symétrie de feedbacks comme suit :



- L'interface représente l'unité de contrôle du dialogue homme machine. Elle présente toute l'information intervenant dans ce dialogue sous la forme de messages qui doivent à la fois être capables de stimuler la sémiose chez l'utilisateur (par les affordances) et correspondre à des données calculables par la machine.
- Les comparateurs interprètent les informations de feedback provenant de l'utilisateur pour "réguler" la production de messages à l'interface, ainsi que les informations de présentation de feedforward indiquant l'état du système.

La détermination des messages porteurs des informations de feedback et de feedforward au niveau de l'interface est à la base de l'ergonomie de cette dernière. En effet, c'est là que nous rencontrons les deux grands paradigmes de l'interface :

- Une interface agissant en "boîte noire" ne présentera de l'information qu'à l'issue du traitement. Par conséquent, l'information de feedforward se résumera à la présentation de l'état initial du système, notamment au fait qu'il est prêt à recevoir de l'information de la part de l'utilisateur,

- Une interface agissant en "boîte transparente", qui présente en continu l'évolution de son état.

Selon le paradigme d'interface, cette information de feedforward conditionne les actions de l'utilisateur, donc, les feedbacks qu'il fait à destination de la machine :

- Dans une logique de "boîte noire", ce seront des actions de sollicitation à destination de la machine, avec un système se comportant en boucle ouverte, ne traitant pas de feedbacks à courte période. Une telle logique est donc celle de la seule planification des actions.
- dans une logique de "boîte transparente", ce seront des opérations d'analyse et d'interprétation des messages de feedforward, avec un système se comportant en boucle fermée, permettant sa supervision via les feedbacks à courte période. Une telle logique associe donc la planification des actions et leur régulation par le contexte d'activité.

L'efficacité d'une interaction homme machine revient à trouver le meilleur équilibre entre ces deux logiques et dans la présentation des informations de feedforward et de feedback intervenant dans la régulation du dialogue.

A travers la question d'une mesure des informations de feedforward et de feedback, nous avons affaire à l'évaluation de l'affordance du système, et à celle des actions de l'utilisateur.

En positionnant cette boucle formée des feedbacks de l'utilisateur et des feedbacks de la machine dans le contexte de la communication selon Shannon, on obtient un modèle de l'interaction applicable à tout traitement de l'information (Fig. 16).

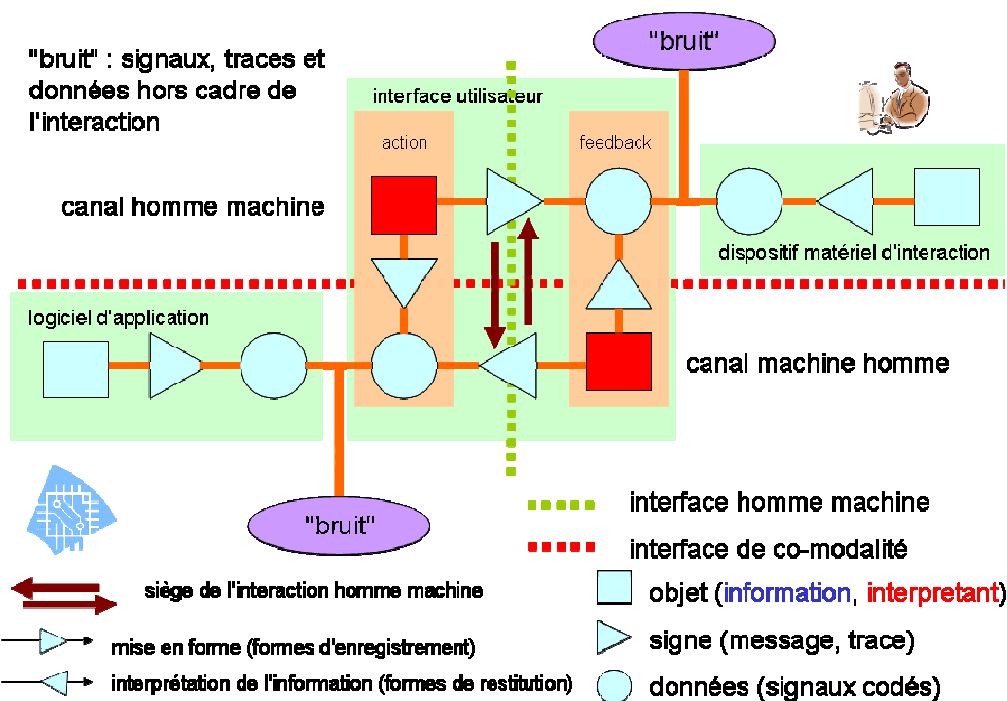


Fig. 16 : Proposition d'un schéma de modèle générique de l'interaction homme machine.

Ce modèle satisfait aux principes suivants :

- l'interface homme machine s'articule entre deux canaux de communication, homme-machine et machine-homme
- les seuls éléments partagés par l'homme et la machine sont les formes à l'interface homme machine, si bien que cette dernière constitue un point de contrôle et de supervision de l'interaction par l'homme.
- la boucle d'imbrication incluant actions et feedbacks permet de rendre compte de la double nature des feedbacks (en boucle ouverte et en boucle fermée), tout comme de l'information de feedforward montrant l'état du système pendant le dialogue.
- la "mémoire" du système est donnée par les interprétants issus des deux canaux de communication.

Ce schéma constitue le système sur lequel va se baser notre raisonnement sur la complexité comme moyen de mesure d'une distance entre informations. Ce système associe deux canaux de communication, matérialisant la différence de signaux perceptibles par l'homme ou la machine.

- Pour l'homme : ce sont les signaux susceptibles de stimuler une de nos voies sensori-motrices. Si l'on considère une interaction multimodale, il conviendra d'instancier ce canal en autant de modalités concernées
- Pour la machine : ce sont les signaux binaires représentant des données.

Chacun des canaux a la structure du canal de communication de Shannon et Weaver, notamment en termes d'extrémités.

#### *Dans le sens homme machine :*

Une action d'un utilisateur, un geste portant sur une information ou une commande, est capturé par le dispositif matériel invoqué sur le terminal. Cette trace est ensuite mesurée et quantifiée (numérisée), sous la forme d'un message codé en données qui interviennent dans des calculs effectués par des algorithmes (le traitement et la reformulation de la présentation de l'action)

#### *Dans le sens machine homme :*

L'application effectue un calcul sur des données, en entrée d'un programme ; ces données représentent les objets techniques intervenant dans un traitement. Elles peuvent être également le résultat de l'analyse et de l'interprétation des actions de l'utilisateur, si l'interface reformule la présentation des actions de l'utilisateur (par exemple, traitement en "boîte transparente" d'un tracé).

Le résultat de ce calcul se présente également sous la forme de données calculable pour la machine qui produit une présentation de l'information affordante à l'interface, dans une modalité perceptible par l'utilisateur. Ici, nous pouvons commencer à traduire le concept ergonomique d'affordance comme le calcul d'une présentation de l'information permettant à l'utilisateur de percevoir la forme, d'en analyser la structure pour en déduire un sens. Ces trois niveaux de l'affordance que nous présentons ici, forment des propriétés de tangibilité de l'information.

Nous y retrouvons les trois niveaux du document numérique définis chez Roger T. Pédauque, à savoir

- Un niveau physique (perception du contenu)
- Un niveau sémiotique (lecture du contenu dans le sens de l'analyse)

- Un niveau sémantique (compréhension du sens du contenu)

Il est également frappant de constater qu'à nouveau, cette approche évoque les trois problèmes de la communication posés par Weaver.

*La régulation de l'interaction homme machine :*

Le seul point commun de partage d'informations entre un homme et une machine se situe au niveau des messages. Ceux-ci doivent être abordables pour l'humain, c'est-à-dire tangibles au sens que nous venons de définir. Ils doivent correspondre à des données calculables pour l'application, c'est-à-dire, interprétables par un programme logiciel.

Ainsi, ce modèle propose une relation entre le concept d'affordance propre aux ergonomes et le concept de calculabilité qui est propre à la technique. Un contenu faiblement abordable pour l'utilisateur correspond à des messages dont l'analyse (la structure) et l'interprétation (la classification) sont complexes. Symétriquement, on peut dire qu'un contenu sera faiblement abordable sur le plan technique si sa reconnaissance (analyse de la structure et classification en objets informatiques) nécessite un calcul complexe.

Cette symétrisation de l'affordance et de complexité de calcul pour produire un message abordable pour l'utilisateur est un apport réellement important du modèle proposé. Elle évoque, en outre, la complexité descriptive et interprétative de Löfgren.

Sur le plan de la dynamique de l'interaction, nous mettons en évidence deux types de feedbacks. En effet, Les données sont décodées au niveau de l'interface de l'application qui génère des objets informatiques correspondant aux "messages" susceptibles d'être traités par l'application. A ce stade, l'interface restitue

- un feedback relatif à l'action de l'utilisateur (interaction à courte période, "en rampe continue" pour un tracé, par exemple),
- un feedback relatif à l'acquiescement du message de l'utilisateur (interaction à longue période discrète).

On notera que cette dualité de messages feedbacks stimule la sensibilité de l'utilisateur en corrélation directe avec les fonctionnalités de notre système nerveux en termes de boucles de régulations sensorimotrices (Outrequin, Boutiller, Rouvière, etc.). En effet, l'anatomie fonctionnelle du système nerveux (central et périphérique) montre deux types de contrôles portés par des structures anatomiques spécialisées :

- un contrôle en boucle ouverte, correspondant à des mouvements balistiques. Il est potentialisé par les feedbacks à longue période. Ce type de contrôle correspond plus à une planification de combinaisons de gestes qu'à un contrôle précis du geste (par exemple, l'utilisateur désigne un objet)
- un contrôle en boucle fermée, correspondant à des mouvements en rampe. Il est potentialisé par les feedbacks à courte période. Ce type de contrôle correspond plus à un contrôle d'un geste précis qu'à une planification d'une combinaison de gestes (par exemple, l'utilisateur effectue un tracé)

En ce sens, ce modèle, incluant les deux types de feedbacks, potentialise la perception de l'utilisateur de manière optimale.

## Chapitre 4

### La complexité d'un calcul comme instrument de mesure

*"C'est un immense problème que de savoir si l'homme pourra, indéfiniment, s'adapter à ce qu'il s'ajoute."*

**Jean Rostand** - "Carnet d'un biologiste"

Dans un système incluant le contrôle d'un traitement par l'utilisateur, il se pose la question légitime de l'efficacité et de la simplicité de ce contrôle. Le point de départ est la présentation de l'information qui doit être suffisamment évidente pour l'utilisateur afin que ce dernier puisse décider des actions qu'il peut accomplir. Cette évidence de perception a son symétrique vu de ma machine : une action de l'utilisateur est d'autant moins ambiguë qu'elle est simple à quantifier, analyser et quantifier par les programmes de l'application.

Ce chapitre montre que cette évidence d'interprétation, tant par l'homme que par la machine, peut se quantifier au travers de la mesure de la complexité d'un calcul. Ainsi, dans l'évaluation d'une forme, que cette dernière soit produite par un outil d'édition ou résulter d'une reconnaissance de formes, se doit d'inclure la mesure de la complexité du calcul ayant conduit à lui donner une valeur. Et si plusieurs solutions permettent d'affecter une valeur à une forme, le système mémoriserait la solution la plus simple.

## ***I. Un point sur les systèmes complexes :***

Un système complexe est un système formé de plusieurs entités en situation d'interaction dans lequel un observateur ne peut pas prévoir le comportement ou l'évolution en appliquant de simples règles de calcul à partir de chacun des composants de ce système.

D'une manière générale, un système complexe s'oppose à un système réductionniste ; la connaissance parfaite des composants élémentaires d'un système complexe ne permet pas de prévoir son comportement autrement que par l'expérience ou la simulation.

Or, expérimenter ou simuler le fonctionnement d'un système complexe nécessite d'en mesurer des états observables, c'est-à-dire, susceptibles d'être décrits dans un langage formel, puis quantifiés et rapprochés de valeurs de références connues (étalonnage d'une mesure, classification de formes ou d'évènements).

L'impossibilité de prévoir le comportement d'un système complexe provient de trois causes :

- Une trop grande complexité computationnelle par rapport aux seuils de complexité admissibles par l'observateur. Ces seuils dépendent des contraintes initiales appliquées au système qui, dans le cas de la conception d'applications informatiques sont les prérequis matériels (telle application doit fonctionner sur tel terminal) ainsi qu'aux prérequis ergonomiques de l'application (le temps de calcul doit rester dans les normes acceptables d'attente de l'utilisateur).
- Le comportement aléatoire d'un composant du système qui donne à ce dernier un caractère probabiliste par opposition à un système déterministe.
- Une sensibilité du système aux conditions initiales comme les systèmes chaotiques.

Dans cette étude, nous aborderons principalement les deux premières causes qui correspondent à notre problématique de traitement sémiotique de l'information. Elles permettent de considérer que la complexité d'une fonction traduite en algorithme par rapport aux données traitées, peut servir de point de mesure.

### **I.1. - La complexité computationnelle**

La compréhension de la complexité est couverte par un champ de l'informatique théorique qui décrit formellement la difficulté intrinsèque d'un problème définissant un algorithme. Il s'agit d'un formalisme mathématique dans lequel interviennent trois concepts qui sont, le problème algorithmique, la réponse algorithmique, la complexité du problème algorithmique. Ce formalisme permet de quantifier la relation entre une taille de donnée et un temps d'exécution en relation avec le nombre d'états de transition entre deux instructions élémentaires d'un algorithme.

Cette quantification est donc une mesure de la complexité faite à partir d'évènements ou de faits observables sans avoir à tenir compte de la nature même de l'algorithme, de sa finalité. On peut assimiler cette mesure à l'analyse du

comportement dynamique d'un algorithme, donc de la fonction logicielle portant cet algorithme, dans un contexte donné.

En fonction du résultat de cette mesure de la complexité, une classification de cet algorithme sera faite à partir de ces propriétés dites d'espace (la taille de la donnée en entrée), et de temps (la dynamique de l'exécution) en établissant une fonction de dépendance entre ces deux propriétés. Pour schématiser, un algorithme pourra avoir une fonction de temps linéaire pour traiter une certaine taille de donnée constante, ou bien, cette fonction de temps pourra avoir être de nature logarithmique, exponentielle, factorielle, etc.

Pour aboutir à cette mesure, il est nécessaire de définir des modèles décrivant le problème à résoudre par un algorithme, le calcul fait par cet algorithme, ainsi qu'une machine abstraite exécutant les instructions portées par cet algorithme.

### ***1.1.1. - Le problème algorithmique***

Un problème algorithmique est la transposition en programme d'un problème mathématique et comprend toujours des hypothèses, des données ou instances et une question.

Résoudre un problème algorithmique conduit à catégoriser deux sortes de problèmes :

- Le problème de décision : l'algorithme doit conduire à une réponse booléenne (oui/non)
- Le problème de recherche d'une solution ou problème d'existence : l'algorithme doit trouver un élément tel qu'un ensemble de conditions se trouvent réunies.

On notera qu'un problème d'existence peut être transformé en un problème de décision. Il s'agira d'une reformulation du problème.

Par exemple, considérons le problème du voyageur de commerce qui consiste à visiter un ensemble de villes en définissant le trajet le plus court. Cela revient à construire un graphe dont chaque arête est étiquetée avec une valeur de distance, à calculer un cycle ne passant qu'une seule fois par tous les nœuds et en couvrant une distance minimale. Autrement dit, calculer le cycle hamiltonien de poids minimal dans un graphe. Posé comme cela, le problème du voyageur de commerce est bien un problème d'existence ou de recherche de solutions.

Si nous le reformulons comme suit : Existe-t-il un cycle passant une seule fois par chaque nœud tel que tout autre cycle passant par tous les nœuds on un coût supérieur? Autrement dit, après avoir calculé l'ensemble des cycles hamiltoniens avec leur poids respectif, existe-t-il un cycle de poids minimum?

Nous posons un problème de décision dont la réponse est oui ou non et qui est strictement équivalent au problème de départ. En effet, cette reformulation se base sur un argument constructif élémentaire : si un algorithme est capable d'avoir pris une décision sur un problème, c'est que la donnée en solution du problème existe et a été traitée par l'algorithme.

Le passage d'un problème d'existence à un problème de décision se fait donc en posant ou en calculant des prédicats, traduisant des hypothèses conduisant à une décision d'existence d'une solution.

Ainsi, prenons l'expression quelconque

$$\forall x \in R, \exists y = f(x) / y = K\theta$$

Déterminer l'ensemble des solutions qui satisfont l'expression est un problème de décision sur la valeur de  $y$ . Cette capacité de décision et sa complexité seront à la base des méthodes de validation de l'algorithme opérationnalisant cette expression.

Rechercher la fonction  $f$  conduisant cet ensemble de solution est un problème d'existence de la fonction  $f$ . Cette capacité de pouvoir calculer la fonction  $f$ , et l'ensemble des prédicats traduisant son contexte, sera à la base d'un mécanisme d'apprentissage.

### ***1.1.2. - La réponse algorithmique***

Il s'agit de la réponse au problème. Selon sa nature, le problème sera qualifié de décidable si cette réponse sera donnée par un algorithme de décision. Il sera qualifié de calculable si cette réponse est un algorithme de recherche de solution.

A titre d'exemple, prenons le cas d'une méthode de classification par calcul de distance, quelle qu'elle soit. La classification est un algorithme venant en solution d'un problème de décision, en ce sens que l'on décide de donner une valeur à une forme en fonction par rapport à des formes de référence. Le problème posé ici sera le calcul de distance entre formes mesurées et formes de référence.

Valider une méthode de mesure de la distance reviendra donc à connaître les familles de solutions conduisant à un problème de classification décidable et correspondra à établir le contexte d'utilisation du calcul de distance mis en œuvre dans cette classification.

Symétriquement, la remise en question de ce contexte d'utilisation, notamment lors de la mise en œuvre de méthodes d'apprentissage, conduira à la résolution de problèmes de calculabilité de la méthode : à tel contexte correspond une méthode calculable ou non, c'est-à-dire, qu'il existe une fonction permettant de trouver une famille de solutions par rapport aux prédicats. Dans notre exemple de la classification, il s'agira de déterminer la fonction de calcul de distance dans un environnement de prédicats et sa validation reviendra à connaître son domaine d'application à un niveau de complexité donné.

On notera que la validation d'un apprentissage sur une méthode de classification est un problème de décision combiné à un problème d'existence.

### ***1.1.3. - La complexité d'un problème algorithmique***

L'objectif de l'analyse de la complexité des problèmes est de pouvoir évaluer les ressources à mettre à disposition pour la mise en œuvre de l'algorithme en termes de temps et d'espace mémoire alloué. Ces critères sont en relation directe avec le contexte d'utilisation d'une application informatique, en ce sens que l'on déterminera des seuils d'acceptabilité pour l'allocation en ressources. Ces seuils seront déduits des usages acceptables sur le plan ergonomique et du contexte matériel (type de terminal, bande passante calculatoire, de mémoire, etc.). On voit donc que l'utilisation de la théorie de la complexité à des retombées directes sur la définition de toute application informatique dès sa conception.



## **I.2. - L'utilité de la complexité**

A t'on besoin d'un tel formalisme mathématique donné par la théorie de la complexité pour toute application informatique ? La réponse est dans la question en ce sens que ce formalisme n'a vraiment d'intérêt que pour les problèmes conduisant à des systèmes informatiques complexes, faisant cohabiter plusieurs fonctions logicielles dans un ensemble de données lui-même complexe. Nous verrons par la suite, que le problème de l'interaction multimodale, la numérisation de corpus, la recherche d'information dans les corpus numériques sont typiquement des problèmes pour lesquels l'analyse de la complexité représente un outil nécessaire du fait d'un système et d'une dynamique difficile à rationaliser selon un modèle réductionniste. En effet, un tel modèle conduirait à devoir rapidement gérer des exceptions, comme nous avons pu l'observer régulièrement dans les systèmes experts à base d'inférences à partir de règles.

### ***1.2.1. - Relation entre la complexité et la définition d'un système***

Ces critères de complexité en espace et en temps définissent des contraintes applicables au système lors de sa conception. L'analyse des cas d'usages consiste à identifier et à poser les problèmes. L'étude algorithmique consiste à définir s'il s'agit d'un problème de décision ou d'existence. Enfin la mesure de la complexité permet de mettre la solution technique du problème dans l'environnement technique adapté aux ressources nécessaires. Nous sommes dans le contexte de la conception universelle d'applications informatiques dans lesquels nous avons une étroite dépendance entre usages, technologies mis en œuvre et environnement. Par exemple, l'attente d'une réponse d'un algorithme de décision peut rapidement devenir inacceptable pour l'utilisateur. Cette attente sera prédictible si l'on sait décrire et positionner la complexité de cet algorithme de décision.

Dans un système complexe, c'est un ensemble d'algorithmes qui sont mis en situation d'interactions avec chacun des propriétés de décidabilité et de calculabilité. On établit une hiérarchie des difficultés entre chaque problème algorithmique de manière à pouvoir organiser et structurer la perception que l'on veut avoir du système.

Cela conduit à la définition de familles et de classes de complexité que nous évoquerons plus loin.

Dans chaque classe on trouvera des ramifications dépendant de la nature déterministe ou non du calcul à effectuer.

Un calcul déterministe est un calcul dont l'état suivant peut être déduit par une mesure de l'état courant, alors qu'un calcul non déterministe n'est pas prédictible, c'est-à-dire, qu'à partir d'un état courant correspondent plusieurs états suivants sans dépendances stochastiques.

### ***1.2.2. - Modèles de calcul et machines abstraites***

Classiquement, on tend à considérer un calcul comme une succession d'étapes élémentaires. A chacune de ces étapes correspond un ensemble d'actions élémentaires possibles parmi lesquelles une sera choisie en fonction de l'état de la machine. L'entité effectrice est une machine abstraite qui sera soit déterministe, si à

un état donné de la mémoire correspond une action unique, soit non déterministe si, pour un état donné de la mémoire, plusieurs actions sont possibles.

Ce modèle de calcul se situe en droite ligne de la machine de Turing et de la thèse de Church-Turing.

Une machine de Turing est, nous le rappelons, une machine abstraite capable de lire, d'écrire et de se déplacer sur un ruban qui correspond à la mémoire du système.

Elle est définie par le sextuplet  $(Q, \Sigma, \Gamma, B, q_0, \delta, F)$  où

- $Q$  est un ensemble fini d'états
- $\Sigma$  est un alphabet fini
- $\Gamma$  est l'alphabet de travail avec  $(\Sigma \subseteq \Gamma)$
- $B \in \Gamma - \Sigma$  est un symbole spécial (dit "blanc")
- $q_0 \in Q$  est l'état initial
- $\delta: Q \times \Gamma \rightarrow Q \times \Gamma \times \{\leftarrow, \uparrow, \rightarrow\}$  est la fonction de transition (les flèches représentent les trois déplacements possibles de la tête de lecture : droite vers gauche, immobile, gauche vers droite)
- $F \subset Q$  est l'ensemble des états terminaux (ou états acceptants ou finaux)

Exemple de fonctionnement :

$$\delta(q_1, x) = (q_2, y, \leftarrow)$$

Cela signifie que si la machine est dans un état  $q_1$  et qu'elle lit le symbole  $x$ , elle écrit le symbole  $y$  à la place de  $x$ , passe à l'état  $q_2$  et déplace sa tête de lecture à gauche.

On remarquera que d'autres machines abstraites ont pu être définies, comme par exemple, la machine RAM qui est la plus proche de nos microprocesseurs actuels. La machine de Turing à un seul dispositif de lecteur écriture et un seul registre de mémoire. Elle se définit comme un automate parcourant un ruban selon une valeur qu'elle a lue sur un "ruban" (sa mémoire) et y inscrit une nouvelle valeur dépendant du déplacement qu'elle a effectué, alors que la machine RAM a deux registres, donc deux rubans, un pour la lecture et un pour l'écriture. On a pu montrer que la machine RAM pouvait intégralement être transposable en machine de Turing.

Schématiquement, à chaque machines de Turing incluses dans un système, est associée une fonction définissable et récursive. Cela signifie que les fonctions traduites d'expressions mathématiques, peuvent être formalisées en méthodes effectives de calcul devant satisfaire aux conditions suivantes :

- Etre un ensemble d'instructions simples et précises, utilisant un nombre restreint de symboles
- Produire un résultat en un nombre fini d'étapes ou d'itérations
- Produire des traces susceptibles d'être suivies par un humain
- L'exécution séquentielle de cette méthode effective de calcul ne requiert de l'humain pas d'autre intelligence que celle liée à la compréhension des instructions élémentaires de la méthode.

On voit que les propriétés des méthodes effectives de calcul ont un caractère intuitif, en ce sens qu'il n'y a pas de définition formelle de ce que sont, par exemple, une "instruction simple et précise", "un nombre restreint de variables", "l'intelligence liée à la compréhension des instructions". C'est ce qui fait que l'on parle, non pas d'un "théorème de Church-Turing", mais de la "thèse de Church-Turing".

Cependant l'effectivité des méthodes de calcul se traduit par des algorithmes qui eux, ont une représentation formelle, si bien que l'on a une concordance entre ces méthodes effectives de calcul et les algorithmes.

Un tel modèle de calcul et de machines abstraites est légitime dans notre étude de la complexité du fait que les langages de programmation utilisés aujourd'hui (C, C++ et même JAVA) sont des langages qualifiés de "Turing-complet" car ils permettent de représenter n'importe quelle fonction calculable sous la forme d'un programme respectant les propriétés de la thèse de Church-Turing.

### **1.3. - Les dimensions de la complexité**

L'objectif de l'analyse de la complexité d'un système est, nous le rappelons, de pouvoir évaluer les ressources nécessaires à l'exécution d'un programme incluant un algorithme. Si bien que l'analyse de la complexité d'un système va conduire à trouver un compromis entre un temps de calcul qui devra être acceptable et une occupation de la mémoire compatible avec l'environnement technique. Ainsi on parlera de complexité en taille et en espace pour caractériser un système.

#### ***1.3.1. - Complexité en taille***

Tout programme utilise des données et, ramené à la notion d'instruction élémentaire, chaque problème a une donnée en entrée à traiter ayant une taille. Il peut s'agir du nombre de données à traiter, de la taille d'un fichier, du codage d'un entier, du degré d'un polynôme, du nombre d'arêtes d'un graphe, etc.

Cette taille est un entier naturel et communément dénommé  $n$ .

Ici, la mesure de la complexité revient à évaluer la place mémoire nécessaire pour l'exécution de l'algorithme en dehors de l'espace occupé par les instructions du programme et les données.

Cette évaluation de la variation de la taille mémoire par rapport à la taille du problème définit une fonction que l'on notera  $s(n)$ .

#### ***1.3.2. - Complexité en temps***

Il s'agira d'un temps d'exécution qui est une fonction du nombre d'opérations fondamentales (d'instructions) composant un algorithme. Ce temps est un ordre de grandeur, une mesure faite à partir des états de transition d'une machine de Turing entre deux opérations élémentaires.

Ce temps; noté  $t(n)$  est donc un indicateur de complexité en temps d'un problème.

#### ***1.3.3. - Expressions de la complexité en temps et en taille***

Considérons un problème de taille  $n$  et posons la fonction :

$$t(n) = 4n^2 - 2n + 2$$

Cette fonction indique le temps de résolution du problème nécessaire en fonction de la taille de la donnée.

Cette expression nous permet de déduire que  $t(n)$  a le même sens de variation de  $4n^2$ . Autrement dit, si  $n$  varie, la fonction  $t(n)$  varie de l'ordre de  $4n^2$ .

D'autre part, on élimine les constantes multiplicatrices qui dépendent des machines utilisées et non de la complexité intrinsèque de l'algorithme (deux ordinateurs de même architecture, mais de puissances différentes vont traiter le même problème dans des temps proportionnels).

Pour exprimer cet ordre de grandeur de la variation de  $t(n)$  on utilise la notation de Landau "O" qui traduit l'évaluation asymptotique d'une fonction et qui s'écrit dans notre exemple

$$t(n) = O(n^2)$$

Posons à présent un algorithme permettant de calculer la puissance  $n$ -ième d'un entier avec  $n$  positif. La fonction en C de cet algorithme peut s'écrire

```
int puissance (int a, int n)
{
    int i, x ;
    x=1 ;
    for (i = 0; i < n; i=i+1) // boucle de n iterations
        x = x * a ; // opération fondamentale de
multiplication
    return x ;
}
```

On voit ici que les performances de la machine, correspondant à des constantes de la fonction  $t(n)$  n'interviennent pas directement dans l'ordre de grandeur de la complexité.

La variable  $n$ , correspondant au nombre d'itérations faite par la fonction, est la seule variable susceptible de faire varier le temps d'exécution de la fonction "puissance" et ce temps est principalement conditionné par le calcul de la  $n$ -ième puissance. Plus  $n$  est élevé, plus le temps de calcul est important.  $n$  représente ici la taille du problème et on observe une croissance linéaire du temps qui suit celle de  $n$ .

En utilisant la notation de Landau, la complexité de cet algorithme s'écrit

$$t(n) = O(n)$$

Nous pouvons tenir le même raisonnement quant à la variation de la fonction de complexité en espace  $s(n)$ .

Le tableau ci-après présente différents type de complexité

Notation	Type de complexité
$O(1)$	Complexité constante
$O(\log(n))$	Complexité logarithmique
$O(n)$	Complexité linéaire
$O(n.\log(n))$	Complexité quasi-linéaire
$O(n^2)$	Complexité quadratique
$O(n^3)$	Complexité cubique
$O(n^p)$	Complexité polynômiale ( $p =$ fonction polynôme)
$O(n^{\log(n)})$	Complexité quasi-polynômiale
$O(2^n)$	Complexité exponentielle
$O(n!)$	Complexité factorielle

#### ***1.3.4. - Les familles de classes de complexité***

Selon l'évaluation asymptotique de la fonction  $t(n)$  ou de la fonction  $s(n)$  et la nature déterministe ou non déterministe de la machine, on obtient 4 familles de complexités.

$TIME(t(n))$  (ou  $DTIME(t(n))$ ) : classe des problèmes de décision qui peuvent être résolus dans un temps de l'ordre de grandeur  $t(n)$  sur une machine déterministe.

$NTIME(t(n))$  : classe des problèmes de décision qui peuvent être résolus dans un temps de l'ordre de grandeur  $t(n)$  sur une machine non déterministe.

$SPACE(s(n))$  : classe des problèmes de décision qui nécessitent un espace de l'ordre de grandeur  $s(n)$  sur une machine déterministe.

$NSPACE(s(n))$  : classe des problèmes de décision qui nécessitent un espace de l'ordre de grandeur  $(s(n))$  sur une machine non déterministe.

Du fait que chaque problème d'existence peut être transformé en un problème de décision, ces quatre classes s'appliquent également.

#### ***1.3.5. - Les classes de la complexité***

Une classe de complexité est un ensemble de problèmes pouvant être résolus en utilisant une certaine quantité de certaines ressources de calcul.

Deux exemples :

- classe P : problèmes de décision pouvant être résolus par une machine déterministe dans un temps polynomial

$$P = \bigcup_{k \geq 0} TIME(n^k)$$

remarque : un temps polynômial peut s'écrire  $t(n) = O(n^k)$  avec  $k \in \mathbb{N}$

- classe NP : problème de décision pouvant être résolu par une machine non déterministe dans un temps polynomial

$$NP = \bigcup_{k \geq 0} NTIME(n^k)$$

- classe EXPTIME : problème de décision pouvant être résolu par une machine déterministe dans un temps exponentiel

$$EXPTIME = \bigcup_{k \geq 0} TIME(2^{n^k})$$

Intuitivement, la classe P correspond à celle des problèmes pouvant être résolus efficacement dans le pire des cas. Par exemple, la recherche de la connexité d'un graphe par un algorithme qui en analyse les sommets deux à deux de manière exhaustive est un problème de complexité quadratique.

La classe NP est typiquement où l'on trouvera la validation d'un calcul par marquage des résultats au moyen d'un certificat (par exemple, "TRUE"). Intuitivement, la validation d'une méthode de classification en reconnaissance de formes aura les propriétés d'un problème dans NP. On lui associe sa classe duale Co-NP correspondant aux problèmes de réfutation d'un résultat.

Selon la nature de la fonction de variation en temps et en espace de la taille du problème traité par une machine abstraite, on obtient un système de classification des algorithmes sur lequel nous pouvons nous baser pour déterminer la stratégie la plus efficace de traitement de l'information. A état initial et état final connu, cette stratégie cherchera à :

- trouver un compromis entre les complexités en espace et en temps,
- minimiser la complexité, et, dans notre cas, utiliser cette complexité minimale comme unité de mesure représentative du calcul d'une présentation de l'information.

La recherche de cette optimisation d'un système complexe utilisera, en outre, une algèbre décrivant les propriétés des classes de complexité entre elle pour transposer un problème et réduire sa complexité.

Ce système de classification est lui-même complexe, notamment, Kuperberg et Aaronson parlent d'un zoo de la complexité en ayant recensé environ 400 classes.

Cette caractérisation de la complexité est un domaine de recherche à part entière de l'informatique théorique et ce domaine est ouvert, en ce sens que toutes les classes prévues de la complexité n'ont pas encore trouvé de réponse sur le plan de leurs propriétés mathématiques, et encore moins sur le plan de l'implémentation. Ces classes forment ce que l'on appelle les questions ouvertes de la complexité.

#### I.4. - Les propriétés fondamentales d'un système complexe

**Propriété 1, intrication des éléments** : un système complexe est composé de plusieurs parties qui sont interconnectées par intrication (J. Moses). Sa définition passe par le nombre et la nature de ces interconnexions et la mesure d'intrication est donnée par la quantité d'informations présentes dans le système.

**Propriété 2, instabilité de la relation de causes à effets** : Un système peut présenter une complexité dynamique quant les relations entre causes à effets sont subtiles (non triviale) et variant en fonction du temps (P. Senge). Des causes évidentes peuvent produire des effets non évidents (non prédictibilité d'un résultat de traitement, comme par exemple, la reconnaissance d'une forme)

**Propriété 3, imprédictibilité des émergences de comportements** : Un système complexe en composé de plusieurs sous-systèmes dont les relations ne sont pas triviales (déterministes) (J. Sussmann). Même si le comportement de chacun des sous-systèmes est prédictible, l'émergence d'un comportement du système entier ne l'est pas.

**Propriété 4, effet Gestalt** : Une fonction exécutée dans un système complexe résultant des relations et des connexions d'un ensemble d'éléments ne peut être reproduite par la somme des résultats de l'exécution de chaque élément pris isolément (Rechtin et Maier)

**Propriété 5, profondeur logique** : Tout état final (un effet) d'un système complexe est la conséquence de transitions à partir d'un état initial (une cause). Le modèle de communication de Shannon prévoit qu'à un effet ne correspond qu'une cause. Le modèle généralisé de l'information numérique transgresse cette loi en nous montrant que l'on ne peut pas déduire une seule cause à partir d'un effet. La profondeur logique (Bennett) consistera à choisir la cause à partir du chemin le plus économique conduisant à un effet donné. On notera que cette définition se rapproche du principe du rasoir d'Occam mais qu'elle évoque fortement aussi, la complexité de Kolmogorov.

La présentation de l'information pour l'interaction fait apparaître des propriétés de la complexité. La numérisation, opération qui consiste à calculer une présentation de l'information contenue dans une image, utilise un "ensemble subtil" de paramètres pris dans la multiplicité des composantes d'une image (composantes globales et composantes formelles locales). Le traitement pour l'extraction et la reconnaissance de formes, nécessite un ensemble de processus liés et dont la dépendance évoque une intrication (acquisition, segmentation/détection, mesure/classification).

C'est précisément dans cette intrication que l'on observe des phénomènes de propagation d'erreurs rendant difficile la validation des processus de reconnaissance de formes et, d'autre part, l'ensemble des propriétés évoquées ici, évoquent fortement les causes de l'ambiguïté d'un résultat de reconnaissance. Ainsi, en abordant la question de la reconnaissance de formes sous l'angle des systèmes complexe, nous pouvons en déduire que la validation d'un processus de reconnaissance doit faire intervenir un ensemble de paramètres et de propriétés non déductibles en prenant isolément chacune des étapes de la reconnaissance. Il conviendra d'adopter une approche systémique de la validation.

L'idée développée ici est de considérer l'image, les algorithmes de traitement et la machine exécutant les traitements comme un système complexe et dynamique, dont l'analyse des propriétés portera donc sur l'identification d'éléments structurants (donc des propriétés globales du système) et l'établissement d'instruments d'observation sur ces éléments structurants. Les résultats de ces mesures d'observations permettant de déduire ici, une distance entre informations.

La complexité sous-tend des idées paradoxalement simples :

- **La taille** : que ce soit celle du problème posé, ou celle des données traitées, la taille donne une indication de la difficulté de dialoguer avec le système.
- **L'ignorance** : la complexité est la cause de l'ignorance (Ferreira). La connaissance d'un système complexe ne nécessite pas l'exhaustivité de la connaissance de tous ces éléments et propriétés.
- **La recherche de descripteurs de longueur minimale** : c'est la complexité computationnelle de Kolmogorov. Elle consiste à trouver, dans un langage de programmation donné, le plus petit programme permettant de construire une séquence binaire donnée.
- **La variété** : une diversité d'éléments d'un système complexe est nécessaire pour arriver à un seuil de complexité.
- **L'ordre et le désordre** : un système complexe est à mi-chemin entre un système ordonné et un système chaotique (totalement désordonné)

## *II. Les interfaces et la complexité*

Dans la définition de toute interface, une présentation d'une information est le résultat d'un calcul. Ce dernier peut avoir une complexité en temps ou en espace qui le situe en dehors de cadre de l'interaction. Schématiquement, une présentation de l'information peut correspondre à un calcul trop complexe ou à la résolution d'un problème non calculable.

Ainsi, la mesure de la complexité d'un calcul devient un paramètre de la mesure du dialogue homme machine :

- Un résultat de traitement inattendu par l'utilisateur correspond à une présentation de l'information insuffisante du fait de la mise en œuvre d'un calcul dont la complexité est inadaptée à la situation ou au contexte.
- La mesure de complexité et sa traduction en mesure de distance, permet d'envisager la détermination des seuils d'acceptabilité au-delà desquels, la machine passera le contrôle à l'utilisateur pour qu'il lui fournisse l'information manquante ayant induit cette complexité de calcul.
- L'apprentissage d'une situation pourra s'appuyer ici sur une recherche de solution de calcul la moins complexe pour construire la relation entre données et présentation de l'information.

Si nous appliquons ces principes à la reconnaissance de formes, on définit par là une stratégie de validation d'un résultat de reconnaissance générique : il n'y a pas d'erreur de reconnaissance de formes au sens "machine" du terme, mais plutôt une mesure de l'état du système que l'humain ne peut pas valider (par exemple, un caractère non ou mal reconnu par un OCR).



Cet état du système est consécutif à un calcul effectué à partir d'un état initial, l'image du caractère, et dont la complexité a conduit à un état final, la valeur du caractère produit par l'OCR. Or, cet état final, présenté à l'utilisateur, ne lui est pas affordant (caractère erroné pour l'interprétation que l'utilisateur en fait). Cela correspond à un manque d'information dans le système que l'on peut assimiler à une complexité de calcul.

Ce manque d'information peut être apporté par l'utilisateur par ses actions à différents moments du traitement (désignation, segmentation, classification).

A titre d'exemple, nous citerons deux démonstrateurs que nous avons développés dans le cadre de ces travaux et que nous décrirons dans le chapitre 6 à savoir, le développement d'un système de reconnaissance intuitive de caractères, ainsi qu'un système de construction de bases terminologique pour l'évaluation sémantique de documents.

Dans l'état de l'art, on peut citer un projet portant sur une application de recherche d'information par le contenu assisté par l'utilisateur dans la désignation de l'objet à reconnaître dans une scène. Il s'agit du projet CAVIAR (Lopresti, Nagy et al.).

L'application proposée comporte une interface permettant de détourner la forme requête dans une image au moyen d'outils infographiques assistés par l'utilisateur, en lieu et place de méthodes de détection. L'objet détourné correspond à un état intermédiaire du système et la complexité de calcul qui aurait permis d'obtenir le même état intermédiaire a été substitué par la mesure d'actions de l'utilisateur pour détourner l'objet, c'est-à-dire, segmenter la partie de l'image correspondant à une zone d'intérêt et la désigner comme objet de requête.

## **II.1. - La complexité de calcul comme mesure de distance**

Si, dans CAVIAR, l'action d'un utilisateur permet de lever l'ambiguïté de la forme désignée à l'interface, il reste la question de la représentativité de cette forme, par rapport à celles connues par le système.

Par exemple, la reconnaissance d'une fleur ou d'un visage détournés dans une image, permet de lever l'ambiguïté du problème technique donné par la segmentation en une forme à évaluer dans un classifieur, mais le problème d'effectivité persiste.

Ce problème correspond à l'établissement d'une métrique utilisée pour affecter une valeur à cette forme et la représentativité de la base de référence construite pour ce classifieur, ainsi que la stratégie même de classification (par exemple, le nombre de plus proches voisins).

La littérature permet de dénombrer un vaste ensemble de distances (de la distance de Manhattan à la divergence de Kullback Leibler) dont la définition et le calcul sont déduits du mode de présentation de l'information utilisé par l'application, c'est-à-dire, à partir de descripteurs représentant une géométrie, une densité, un voisinage, voire des mesures de probabilités, grandeurs toutes dépendantes de la segmentation.

Réciproquement, la définition de bases de classification se fait en ramenant le problème de la classification à un problème de décision, c'est-à-dire : un problème dont la solution est une réponse "VRAI ou FAUX", "OUI ou NON". Dans le cas d'une reconnaissance de formes, le problème se pose en ces termes : **quelles sont les**

**formes connues du système les plus proches en termes de distance par rapport à la forme mesurée et présentée au classifieur?**

Ramené à un problème de décision, la question se reformule ainsi : **existe t'il une forme proche en distance de la forme mesurée et présentée ?**

L'ambiguïté apparaît ici dans la réponse à cette question, selon que l'on ait une réponse unique (pas d'ambiguïté de classification) ou plusieurs réponses (ambiguïté de mesure et de quantification de la forme) ou bien, pas de réponse (forme non représentative de l'espace de décision correspondant aux valeurs de formes connues de l'application).

Une réponse unique correspond à une résolution du problème d'effectivité par des relations bijectives entre une forme et sa valeur. Le cas le plus représentatif est celui de la traduction en caractères d'une séquence binaire correspondant à un code ASCII ou Unicode.

Toutes les autres solutions non bijectives sont porteuses d'ambiguïté se traduisant par une entropie élevée du codage de l'information.

## **II.2. - Discussion sur la classification et l'apprentissage**

Cela nous conduit classiquement à redéfinir la distribution des classes de l'espace de décision, et la seule solution pour obtenir une base de classification cohérente, est de tendre vers une entropie minimale de l'information dans cette base. Cela se traduit par

- un nombre de classe correspondant à l'alphabet ou au vocabulaire de l'application. Pour le texte, il s'agit de la symbologie utilisée par l'application, mais pour les modes images ou graphique, cet alphabet peut être nettement plus complexe et faire appel à des propriétés sémiotiques ("la graphique" de Bertin, "information display" selon Tuffte) difficilement énumérables et quantifiables (exemple de la carte de Minard Fig. 17).

- une distribution des membres d'une classe idéalement répondant à une loi normale, avec un recouvrement minimisé entre classes. Ces propriétés se traduisent par la détermination de frontières dans un espace de décision qui reste un important problème de recherche.

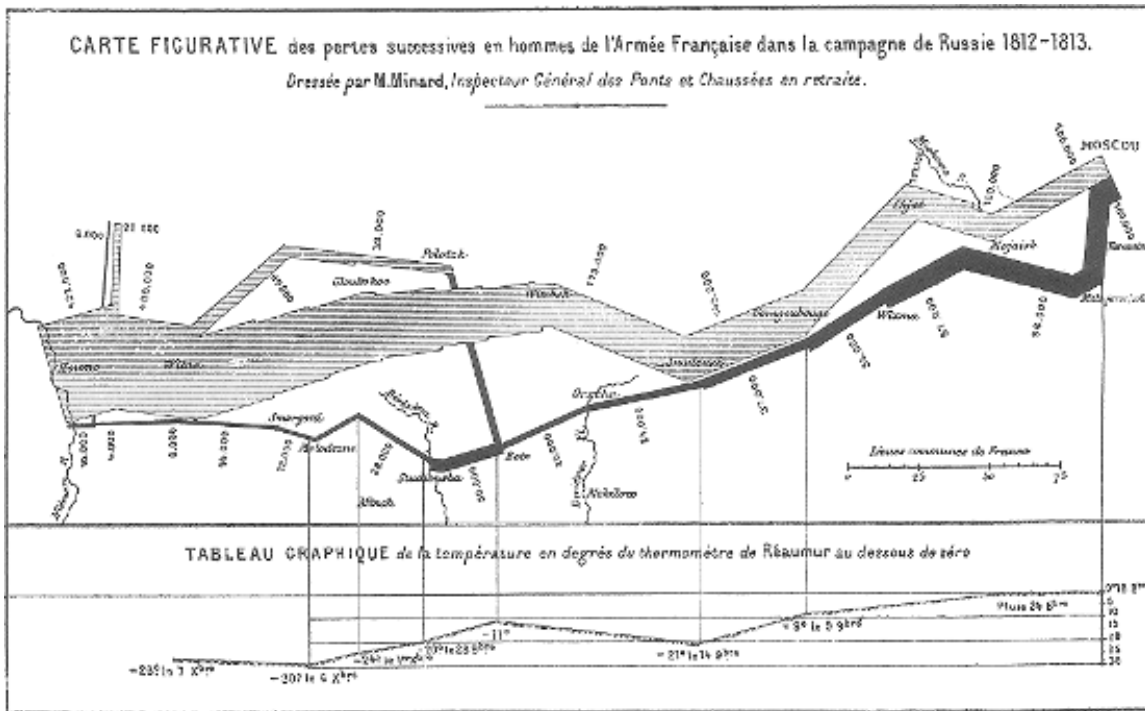


Fig. 17 : Carte de Minard : exemple de présentation d'une information dont la présentation ne se limite par à la définition d'un alphabet symbolique

La spécialisation du calcul de distance à partir des valeurs des descripteurs de formes fait qu'un tel système devient lui-même spécialisé et n'est pas transposable à une autre famille de problèmes de décision. Soit la mesure faite sur les formes ne permet pas d'appliquer la distance prévue dans le problème initial, soit la loi de distribution des classes dans l'espace de décision, représentant "l'alphabet de l'application" n'est pas applicable à cette famille de problème. En outre, on constatera que cette spécialisation de la méthode de calcul de distance ne permet pas facilement de prendre en compte les actions de l'utilisateur pour valider une reconnaissance de formes. Par conséquent, l'évolution d'une base de classification par apprentissage reste problématique.

En effet, hormis la saisie directe de la valeur d'une forme, où l'on force la valeur de décision, l'absence de mesure de distance entre informations présentées dans des modes différents ne permet pas de déduire un apprentissage sans avoir recours à des heuristiques de traitement d'exceptions.

L'absence de distance entre informations signifie l'impossibilité de définir une métrique à partir de seuls descripteurs hétérogènes de formes et donc, l'impossibilité d'effectuer une indexation multimodale ou une recherche par le contenu, à partir des seuls critères de formes.

Afin d'éviter l'écueil de la description des formes générant une ambiguïté sémantique d'un mode de présentation à l'autre, il est intéressant de revenir sur l'approche sémiotique de l'information en formulant le problème suivant :

**La présentation de toute information résulte d'une mesure de cette dernière. Cette mesure produit des formes, perceptibles par un utilisateur et résultant d'un calcul par un programme logiciel.**

En appliquant le modèle sémiotique à ce problème, nous obtenons :

- l'objet est le phénomène mesuré,
- les signes sont soit les formes présentées à l'utilisateur, soit les données calculables par la machine (**symétrie de présentation**),
- l'interprétant correspond, pour l'humain, au mécanisme de sémiologie que met en œuvre ce dernier pour donner sens à la forme qu'il perçoit et, pour la machine, au programme logiciel lui-même qui effectue un calcul pouvant être celui produisant la présentation de l'information pour l'utilisateur (**symétrie entre affordance et complexité de calcul**).

Les approches classiques de mesure de distance portent sur les signes et les données elles-mêmes. Nous avons vu que cela ne résolvait pas le problème d'effectivité dans un contexte multimodal, comme celui de la reconnaissance de formes.

Il reste l'interprétant qui est, vu de la machine correspond à un calcul et vu de l'utilisateur à la constitution d'une affordance. C'est cet interprétant qui permet d'obtenir la symétrie entre **affordance et complexité de calcul** dans notre modèle d'interaction

L'objectif est ici de voir sous quelles conditions, les mesures faites sur l'interprétant machine, donc sur un programme logiciel, permettent aboutir à une distance.

Dans cette hypothèse, nous considérons ce programme logiciel comme le moyen de passer d'un état initial d'un objet à un état final. La distance correspond, quant à elle, à la complexité du programme logiciel permettant de passer de l'état initial à l'état final.

Dans le cas d'une reconnaissance d'un caractère, il s'agirait de mesurer la complexité de calcul permettant de passer d'une forme de requête représentant un caractère, à une forme représentative d'une classe dans un espace de décision, la classe correspondant à la valeur du caractère.

Mais cette hypothèse de prendre la complexité de calcul comme valeur de distance, peut également s'appliquer à la mesure d'une distance entre allographes, entre phonèmes, entre objets graphiques, ce qui sous-entend une notion de distance universelle entre information dont nous allons tenter de démontrer la validité au sein d'une même modalité<sup>8</sup>.

---

<sup>8</sup> On insistera bien sur le fait que cette notion de distance universelle ne s'applique qu'à des informations présentées dans une même modalité. Mesurer la distance entre deux informations codées dans des modalités différentes n'est pas trivial car il n'y a pas de contenus communs entre deux présentations différentes d'une même information. Nous proposons, pour se faire, de partir de la propriété d'inégalité triangulaire de la distance de compression et de l'appliquer dans un espace de phase hamiltonien représentant une interface multimodale, dont chaque degré de liberté correspond à un mode de présentation de l'information.

## Chapitre 5

### Une distance universelle entre informations

*"Ce que nous observons, ce n'est pas la Nature en soi, mais la Nature exposée à notre méthode d'investigation."*

**Werner Heisenberg** - "La Nature dans la physique contemporaine"

La complexité, abordée dans le chapitre précédent, n'est pas qu'un instrument théorique. Le chapitre qui suit est une mise en œuvre de la complexité comme moyen de calculer le degré de similarité entre deux contenus, à partir de leur mode de présentation le plus élémentaire : le codage binaire.

Ainsi, une méthode de calcul de distance a été mise en œuvre, dont le résultat est une mesure de la longueur normalisée des séquences de bits communes entre deux contenus.

Si l'aspect théorique conduisant à l'équation de distance proposée est assez complexe à appréhender, sa mise en œuvre (ce que l'on appelle la "réponse algorithmique" dans la théorie de la complexité) est paradoxalement simple à implémenter et applicable à plusieurs types de contenus.

## I. Une reformulation du problème "reconnaître une forme"

Reconnaître une forme signifie résoudre le problème suivant :

"Le fragment extrait d'un contenu ressemble t'il à une forme déjà inscrite dans le système ?".

La ressemblance est caractérisée par des indices qui sont évalués à partir de saillances d'une forme comparée à un ensemble de formes connues du système. On désignera ces indices sous le vocable de distance, même si cela ne concorde pas exactement avec l'axiomatique de cette notion en mathématiques.

La Fig. 18 montre deux exemples représentatifs de la résolution d'un problème de la reconnaissance de formes classique.

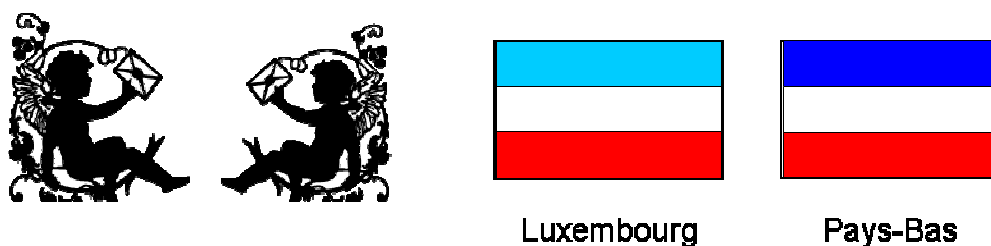


Fig. 18 : Quels données et programmes pour établir les critères de ressemblance entre ces formes ?

Nous voyons du premier coup d'œil que les formes se ressemblent mais, pour transposer cette ressemblance à un calcul, les descripteurs du couple de figures de gauche devront inclure la symétrie spéculaire des formes, et, ceux du couple de figures de droite, une table de couleur incluant suffisamment de nuances pour éviter toute ambiguïté.

Ces exemples montrent que le calcul d'un vecteur de caractéristiques décrivant la structure d'une forme n'est pas trivial. Cette opération dépend des saillances d'une forme qui en révèle sa spécificité. De plus, ces saillances se situent à différents niveaux de détails de la description d'une forme ce qui, en image numérique, se traduit par des résolutions différentes.

A ce premier problème vient s'ajouter la méthode de calcul des indices de ressemblance (ou dissemblance) qui est toujours choisie en fonction de la nature des descripteurs ainsi, le processus de reconnaissance de formes applicable au couple de figures de gauche ne l'est pas au couple de figures de droite sans être profondément redéfini.

Si, pour l'homme, les formes présentées Fig. 18 sont des présentations d'informations très différentes, elles sont inscrites dans un formalisme commun pour la machine : des séquences binaires, tout comme n'importe quel contenu numérique.

Ainsi, toute séquence binaire est le point de vue "machine" de la présentation de toute information.

La stratégie que nous avons élaborée part des travaux de différents auteurs comme M. Li, P. Vitanyi, R. Cilibrasi sur le séquençement du génome et J. P. Delahaye sur le clustering de documents textuels. Elle s'applique à la reconnaissance de formes et à la classification non supervisée de corpus numériques.

L'approche repose sur des calculs effectués directement sur les séquences binaires qui représentent des contenus et sont traitées comme étant des variables aléatoires.

L'hypothèse fondatrice se base sur la complexité de Kolmogorov qui s'exprime par la longueur du plus petit programme permettant de calculer une séquence, en l'étendant au concept d'information mutuelle, c'est-à-dire :

**Soient deux séquences binaires A et B, mesurer la longueur du plus petit programme permettant de produire une séquence binaire A connaissant la séquence binaire B est une mesure de la distance entre A et B.**

Du fait de la non-calculabilité de la complexité de Kolmogorov, M. Li a proposé une approximation satisfaisante de celle-ci par la mesure de la longueur de la séquence produite par la compression sans perte de deux séquences binaires. Cette approximation est légitimée par le concept de calcul réversible de Ch. Bennett qui s'applique aux compresseurs sans perte.

J.P. Delahaye propose une compréhension intuitive de cette méthode en disant que la compression sans perte de deux séquences concaténées va regrouper leurs motifs similaires dans les mêmes classes calculées par un algorithme de compression.

Ceci nous permet de comprendre que l'approximation de la complexité de Kolmogorov faite par Li, réside dans la complexité du calcul effectué par un compresseur dépendant de la loi de distribution interne des séquences binaires soumises à un calcul de distance entre informations contenues dans ces séquences.

## **I.1. - Distance entre informations**

Classiquement, une mesure de distance, ne porte pas sur l'information elle-même, mais sur un mode de présentation structuré dans une syntaxe de codage de cette information. Cette syntaxe de codage est une quantification de la forme et résulte d'une segmentation de cette dernière dans un contenu.

Si nous souhaitons étendre la notion de distance à une mesure entre informations, il apparaît nécessaire de pouvoir :

- faire abstraction du format des données pour que la même mesure de distance soit applicable dans un mode de présentation donné quel que soit ce mode,
- de rechercher la plus petite unité commune de codage existante entre tous les modes de présentation de l'information.

Déterminer la plus petite unité commune de présentation d'une information numérique est facile : Il s'agit du bit ; une information est toujours représentée par une séquence de bits.

Ainsi, une distance universelle entre informations ne peut se déterminer qu'à partir des séquences binaires qui définissent à la fois des données pour la machine et des formes de présentation de l'information pour l'utilisateur.

La question qui se pose ici est :

**Peut-on calculer une distance permettant de calculer une distance entre deux informations, basée les propriétés de séquences binaires décrivant toutes formes de contenus ?**

## I.2. - L'analogie avec la bioinformatique

Pour répondre à cette question, une analogie est à faire avec la biologie, en particulier la génétique. En effet, en biologie, le séquençement de l'ADN (acide désoxyribonucléique) pose des problèmes analogue, voire similaires, à celui de la distance entre informations :

- les séquences binaires sont l'analogie de séquences de nucléotides qui forment le génotype de tout individu. "L'alphabet de quatre lettres" que représentent les bases nucléiques que sont l'Adénine (A), la Guanine (G), la Cytosine (C) et la Thyminine (T) est universel et les séquences de ces suites de nucléotides est propre à chaque espèce animale. En ce sens, le nucléotide est le plus petit élément commun de description d'une espèce et ce plus petit élément commun est tétravalent. Le bit, quant à lui, est bivalent par définition (0 ou 1), et une séquence binaire est une variable aléatoire permettant de caractériser n'importe quelle information en numérique, au même titre qu'une séquence d'ADN est une variable aléatoire permettant de caractériser n'importe quelle espèce animale.
- les présentations de l'information qui résulte d'un calcul, sont l'analogie du phénotype qui est le résultat de l'expression des gènes sous la forme des protéines formant tout individu. Le calcul est ici l'équivalent des mécanismes de synthèse des protéines dont s'est inspiré Bennett pour élaborer le concept de machine balistique.

## I.3. - Un exemple de recherche d'informations par le contenu : le séquençement du code génétique

Quelques propriétés du code génétique nous révèlent un problème complexe à résoudre sur la phylogénie qui est "l'indexation des individus et des espèces" à partir de l'analyse de leur séquence d'ADN (Fig. 19).

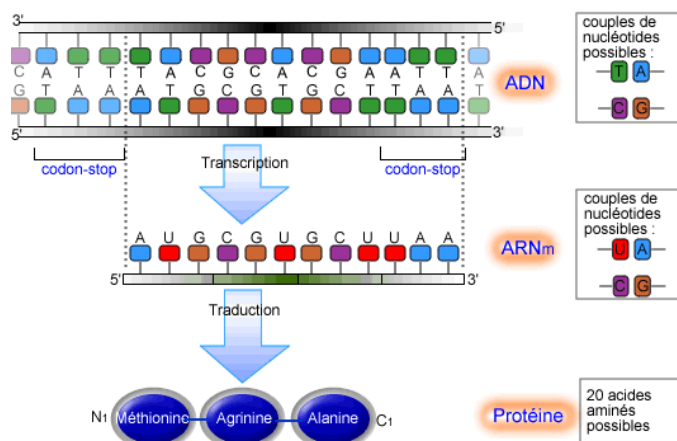


Fig. 19 : Schéma simplifié de la synthèse des protéines à partir de l'ADN (acide désoxyribonucléique)

- L'ADN est le support de l'inscription du code génétique.



- Le code génétique est la signature propre à chaque individu dans chaque espèce.
- L'ADN est codé selon un alphabet de 4 lettres avec des "symboles" de 3 lettres consécutives
- Un "symbole" ou allèle, code un acide aminé et une séquence de symboles code une protéine constitutive de l'organisme.
- problème 1 : le génome humain comporte plus de 3 milliard de paires de nucléotides (soit un "texte" d'1 milliard de symboles correspondant à notre "signature naturelle")
- problème 2 : les séquences sont de longueur variables et toutes ne sont pas codantes (20 acides aminés, 64 "unités codante"), mais la "signature" de l'individu est décrite par l'ensemble de la séquence.

En transcrivant le code génétique d'un individu dans une chaîne alphanumérique formée de quatre lettres représentant chacune un nucléotide, on obtient un contenu, et c'est précisément ce contenu qui sert à notre "indexation génétique".

Une approche par la segmentation et la classification n'est pas applicable en termes de complexité : matching complexe dû à une structure aléatoire des séquences d'ADN du génome entre espèces, équiprobabilité de triplets des nucléotides "symboles" conduisant à code génétique à entropie élevée. Reconnaître une séquence de nucléotide dans ces conditions revient à poser un problème de complexité tel que celui du calcul d'une décimale élevée du nombre  $\pi$ .

L'approche retenue [CIL05a] pour l'élaboration d'une distance entre séquences d'ADN s'est basée sur la théorie des systèmes complexes et a conduit à la définition d'une distance universelle, c'est-à-dire, une distance qui se calcule à partir de la plus petite unité codante d'information, le nucléotide.

La méthode est largement opérationnelle en bioinformatique. Son application aux contenus multimédias représente une voie intéressante [ADR07][CEB05], pour l'indexation sémiotique et la recherche d'informations par le contenu.

#### **I.4. - La complexité de Kolmogorov comme instrument de mesure de distance entre informations**

La détermination d'une mesure de distance proposée par Li se base sur la complexité de Kolmogorov qui s'énonce comme suit [CIL05b][VIT07] :

Soit un ensemble de programmes  $p$ , permettant de construire une séquence binaire  $x$ . La **complexité de Kolmogorov**, notée ici  $K$ , de cette séquence correspond à la longueur en bits du plus petit programme permettant de construire cette séquence. Elle s'exprime :

$$K_U(x) = \min_p \{ |p| : U(p) = x \}$$

$U$  est ici une machine de Turing universelle exécutant le programme  $p$  et dont le résultat est la séquence  $x$ .

Pour évaluer la distance entre une information portée par une séquence  $x$  et celle portée par une séquence  $y$ , cette complexité est relativisée par la **complexité conditionnelle de Kolmogorov** qui est l'analogue de l'entropie d'informations liées ou entropie jointe :

$$H(X, Y) = H(X) + H(Y|X)$$

C'est-à-dire, que la combinaison des deux séquences  $X$  et  $Y$ , considérés comme deux variables aléatoires discrètes, est la somme de la distribution aléatoire de  $X$  plus celle de  $Y$  connaissant  $X$ .

Exprimé par la complexité de Kolmogorov on obtient (Li)

$$H(X, Y) = H(X) + H(Y | X) \Leftrightarrow K(X, Y) = K(X) + K(Y | X) + O(\log(K(X, Y)))$$

On notera que l'on n'a pas une équivalence complète entre l'entropie jointe et la complexité conditionnelle de Kolmogorov en raison d'un facteur  $\log$  résiduel d'ordre de grandeur de  $K(X | Y)$ .

- L'**information mutuelle** se déduit de la complexité conditionnelle de Kolmogorov en posant  $I_K(X, Y) = K(X) + K(Y|X)$ . Cette information mutuelle, correspond à une divergence de Kullback Leibler [VIT07, et a les propriétés d'une distance quand

$$I_K(X, Y) \approx I_K(Y, X)$$

### 1.5. - La mesure de la complexité dans le traitement de l'information

Cette mesure revient à estimer les ressources en temps et en espace mémoire utilisées par un programme logiciel pour exécuter un algorithme. Elle repose ici sur la complexité de Kolmogorov qui correspond à la longueur en bit du plus petit programme permettant de construire une séquence. Cette complexité peut s'appliquer de manière conditionnelle à une séquence  $X$  connaissant une séquence  $Y$ . Dans ce cas, elle correspond à l'entropie jointe des deux séquences (concept d'information mutuelle). Le problème est que, si cette mesure de la complexité a toutes les propriétés mathématiques d'une distance, son implémentation en algorithme est impossible, car correspondant à une situation d'incalculabilité démontrée au travers du théorème de Rice. Toutefois, nous allons voir que l'on peut effectuer une approximation satisfaisante de cette complexité en définissant des propriétés structurante au système (l'application exécutée par un ordinateur).

### 1.6. - Les éléments structurants utilisés pour évaluer cette distance d'information

- le **coût thermodynamique de l'effacement d'un bit** (von Neumann, Landauer) : ce coût correspond à une dépense d'énergie  $E$  dissipée en chaleur et transformée en entropie dans le système. Landauer a démontré que cet accroissement de l'entropie du système correspondait à l'accroissement de l'entropie de Shannon consécutive à l'effacement d'un bit. Ceci a pour conséquence de permettre de considérer le traitement de l'information comme étant un problème de physique avec une relation entre la variation de l'entropie de Shannon et la variation de l'entropie thermodynamique engendrée par un calcul.

Ceci est à rapprocher de la théorie thermodynamique de l'information développée par Léon Brillouin et permet d'appliquer le modèle physique de la thermodynamique au traitement de l'information.

$$E = kT(\log 2) \quad [K = 1,3806 \cdot 10^{-23} \text{ J} \cdot \text{K}^{-1} \text{ constante de Boltzmann, } T : \text{température}]$$

- le **coût énergétique  $E$**  d'un programme  $p$  pour calculer une séquence  $y$  à partir d'une séquence  $x$  (Bennett). On y considère que la séquence  $x$  correspond à un état

initial du système et la séquence  $y$ , à son état final. Ce coût énergétique, exprimé en bits, correspond au calcul effectué par une machine de Turing universelle et réversible  $U$  pour partir de la séquence  $x$  et arriver à la séquence  $y$ .

$$E_U(x, y) = \min\{p \mid U(x, p) = y, U(y, p) = x\}$$

- la **réversibilité d'un calcul**, qui correspond à la possibilité de "revenir en arrière" à tout moment de l'exécution d'un programme. Il s'agira donc d'un programme sans dissipation d'énergie, sans perte d'information et donc, sans augmentation de l'entropie de Shannon. Les compresseurs sans pertes de données répondent à cette propriété, ceux avec perte, non. Cette réversibilité est la condition pour maintenir les propriétés de distance de la mesure que nous décrivons. Ceci se démontre en indiquant qu'un état final donné est consécutif à un état initial unique dans un processus réversible. A contrario, un calcul non réversible correspond à plusieurs états initiaux susceptibles de conduire à un état final unique.

## II. La distance de compression

L'évaluation est faite en calculant la fonction de coût minimal en bits engendrée par un programme  $p$  permettant de calculer une séquence  $X$  connaissant  $Y$  et une séquence  $Y$  connaissant  $X$ .

Cilibrasi [CIL05b] démontre que cette fonction de coût est une distance absolue entre les informations portées par les séquences  $X$  et  $Y$ . En effet, elle vérifie les propriétés mathématiques de symétrie, d'inégalité triangulaire, d'identité, de densité et d'énumérabilité propre à toute distance ; de plus cette distance est normalisée par la maximisation de la fonction de coût.

$$E_U(x, y) = \min\{p \mid U(x, p) = y, U(y, p) = x\} \Leftrightarrow E(x, y) = \max(C(x|y), C(y|x))$$

Le programme  $p$  est donc l'instrument de mesure permettant de calculer cette distance. Il s'agira d'un programme de compression sans pertes de données. La mesure portera, finalement, sur un calcul arithmétique sur les longueurs en bits des séquences à mesurer après compression.

Rappelons brièvement les propriétés de  $E$  :

$$E(x, y) = E(y, x); E(x, x) = 0; E(x, y) \geq 0 \quad (\text{symétrie, identité})$$

$$E(x, y) = \max\{C(x|y), C(y|x)\}$$

$$\leq \max\{C(x|z) + C(z|y), C(y|z) + C(z|x)\}$$

$$\leq \max\{C(x|z), C(z|x)\} + \max\{C(z|y), C(y|z)\}$$

$$= E(x, z) + E(z, y) \quad (\text{inégalité triangulaire})$$

$$|\{y : E(x, y) < d\}| < 2d \Rightarrow \{y : E(x, y) \leq d\} \quad (\text{densité et énumérabilité})$$

On obtient une distance normalisée en posant

$$d(x, y) = \frac{E(x, y)}{\max(C(x), C(y))} \Leftrightarrow d(x, y) = \frac{\max\{C(x|y), C(y|x)\}}{\max\{C(x), C(y)\}}$$

## II.1. - L'indécidabilité calculatoire de la complexité de Kolmogorov et son approximation

Si l'on fait référence au théorème de Rice (le problème de l'arrêt), on constate que le problème de la complexité de Kolmogorov est indécidable. Cela signifie que l'on est incapable de savoir à partir de quel moment on obtient le programme le plus court.

Cependant, si l'on se place dans les conditions de calcul réversible, donc, dans un système à entropie constante (Vitanyi, Cilibrasi), on obtient une approximation valide de la complexité de Kolmogorov par la mesure de la longueur des séquences de bits sortant d'un compresseur sans pertes.

Ainsi cette approximation s'écrit

$$d(x, y) = \frac{Comp(xy) - \min\{Comp(x), Comp(y)\}}{\max\{Comp(x), Comp(y)\}}$$

*Comp* est la fonction de compression sans perte qui retourne une valeur de longueur exprimée en bits. Elle se substitue directement à la fonction de coût déduite de la complexité de Kolmogorov. *x* et *y* sont les séquences porteuses d'informations dont on veut mesurer la distance et *xy* est la concaténation des deux séquences.

## II.2. - Quelques propriétés de la distance normalisée de compression

- Cilibrasi [CIL05b] a démontré que cette distance est une mesure de la divergence de Kullback-Leibler totale de la moyenne des distributions en *X* et *Y*.

- [HES07] En théorie, l'application de ce calcul de distance est indépendante du format, de la mise en forme et de la modalité de présentation de l'information. Il s'agit d'une distance universelle. Toutefois, si l'on arrive effectivement à calculer une distance entre n'importe quelles séquences binaires, encore faut-il que cette distance fasse sens en termes d'usages et d'applications. Par conséquent, le choix du corpus par rapport au type de mesure que l'on veut effectuer, est important.

- l'implémentation du calcul de cette distance nécessite peu de ressources informatiques. Il s'agit d'une opération arithmétique effectuée sur des données compressées au moyen d'outils classiques (zlib, bzip2, ppmd, lzma). La charge de travail est principalement portée par la compression.

Appliquée aux séquences d'ADN, l'algorithme de calcul d'une distance normalisée de compression conduit à des applications permettant de reconstruire, par exemple, l'arbre phylogénétique entre espèces de la classe des mammifères (Fig. 20 : essai effectué les séquences d'ADN mitochondrial de dix espèces différentes). Les essais que nous avons effectués ont porté jusqu'à une cinquantaine de séquences d'ADN mitochondrial.

```

gatcacaggtctatcacccatattaaccactcacgggagctctccatgcatttggtatcttctgctctgggggatgcacgcatagcattgcga
gacgctggagccgggagcacccctatgtcgcagtatctgtcttttgattcctgctcatcctattatctatcgcacctacgttcaatattacaggc
gaacatacttactaaagtgtgtaattaattaatgcttggtaggacataataaacaattgaaatgtctgcacagccactttccacacagacat
catacaaaaaatttccaccaaaacccccctccccgcttctggccacagcacttaaacacatctctgccaacccccaaaaaacaagaaccct
aacaccagcctaaccagatttcaaattttatcttttggcggatgcacttttaacagtcaccccccaactaacacattattttccctcccac
tcccatactactaatctcatcaatacaacccccgcccctctaccagcacacacacaccgctgctaaccocataccccgaaccaaccaaac
ccaaagacacccccacagtttatgtagcttacctcctcaagcaatacactgaaaatggttagacgggctccatcaccccataaaacaata
ggtttggtcctagcctttctattagctcttagtaag...

```

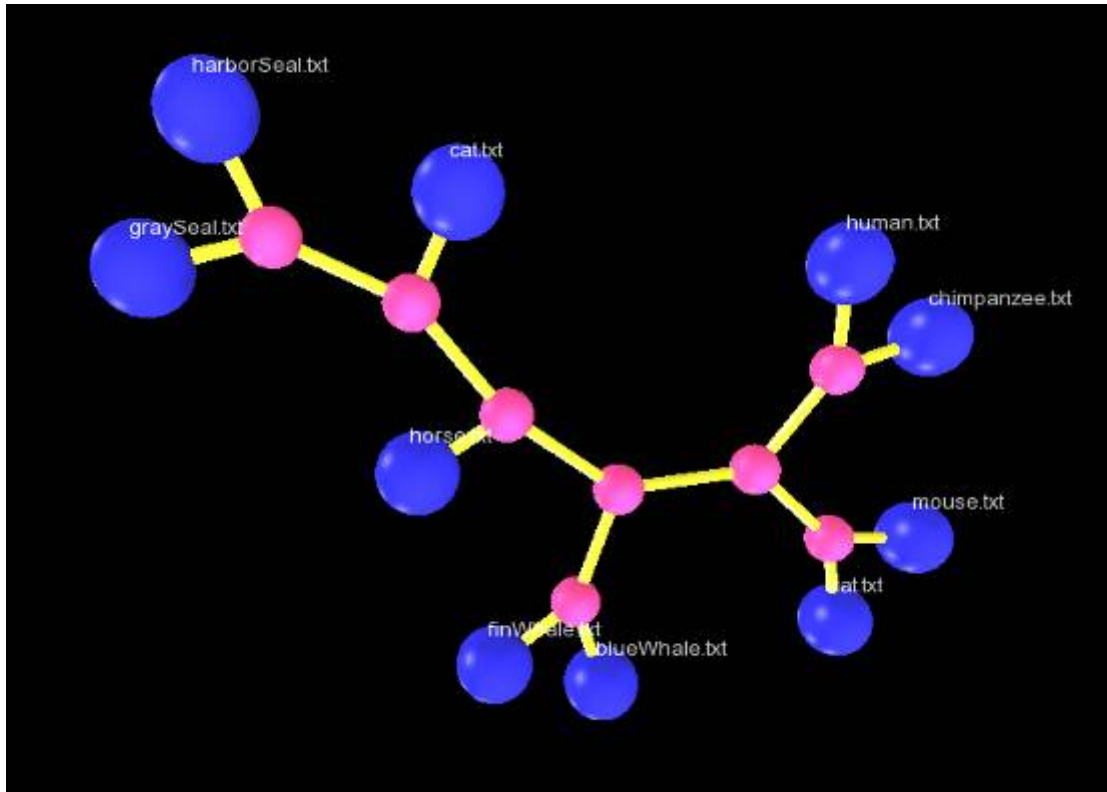


Fig. 20 : application de la distance de compression normalisée sur l'ADN mitochondrial de mammifères. La distance mesurée traduit la proximité phylogénique entre deux espèces et correspond à des similarités dans le séquençement de l'ADN.

En haut : extrait d'une séquence d'ADN mitochondrial humain.

En bas : classification non supervisée utilisant la distance de compression normalisée, portant ici sur 10 séquences d'ADN mitochondrial correspondant à des mammifères.

Le résultat est un classement phylogénique des espèces. Les feuilles de cet arbre sont étiquetées avec le nom de l'espèce et les nœuds identifient les classes. Par exemple, l'ADN humain est proche de celui du chimpanzé. On notera que cet arbre est obtenu après 15 secondes de traitement sur un PC ordinaire pour des séquences d'une longueur de l'ordre de 20000 nucléotides.

### III. Applications de la distance normalisée de compression

Toute opération de clustering non supervisé est théoriquement envisageable par application de ce calcul de distance. Outre la reprise des tests effectués par Vitanyi et Cilibrasi, nous avons procédé à différents essais sur des contenus hétérogènes.

Parmi les exemples que nous allons présenter, celui de la reconnaissance de visages montre l'intérêt de l'approche : la reconnaissance de visage est encore aujourd'hui un problème mal résolu qui suscite l'intérêt de nombreuses équipes de recherche. La base BioID regroupe un corpus d'images prises dans des conditions standards qui pose un problème difficile : la variabilité en termes de pose et d'éclairage.

D'une manière générale, les techniques actuelles, spécialisées et plus complexes à mettre en œuvre, cherchent à répondre au premier plus proche voisin. Cette complexité de mise en œuvre sert notamment à lever les verrous liés aux variations d'éclairage, de pose, en proposant des modèles non linéaire de la texture des visages. De plus, elles assurent une détection automatique en amont de la phase de reconnaissance [LEF07] . D'autres approches historiques telles que Turk et Pentland [TUR91] se basent sur une analyse en composante principale ("eigenfaces"), représentant un espace de caractéristiques de grande dimension.

L'approche que nous avons expérimentée est de type globale et ne nécessite un espace de caractéristiques complexe : il est formé par les classes résultant de la compression sans pertes de données binaires. Il n'y a pas de segmentation structurelle guidée par les buts (on recherche d'abord le visage que l'on veut reconnaître). Son principal inconvénient est qu'elle ne produit pas l'objet "visage". Mais elle est suffisante pour opérer un classement non supervisé via un algorithme très simple et applicable à différents types de données.

### *Visages*

La base BioID utilisée pour la reconnaissance de visages (Fig. 21) : le clustering à partir de la distance de compression montre des classes bien représentative des visages contenus dans cette base de photos. Le temps d'exécution, incluant le calcul de la matrice carrée de distances et la construction du graphe de distance sur une base de l'ordre de 1500 visages, est de l'ordre de 10 minutes sur un DELL précision 490, en utilisant un compresseur à prédiction par reconnaissance partielle (algorithme PPM). D'autre part, l'utilisation de cette distance pour reconnaître un visage, conduit à des résultats tout à fait satisfaisant, en regard des méthodes classiques de détection/reconnaissance mais au prix d'un calcul beaucoup plus simple. Le résultat obtenu est une reconnaissance systématiquement exacte sur les 5 plus proches voisins.

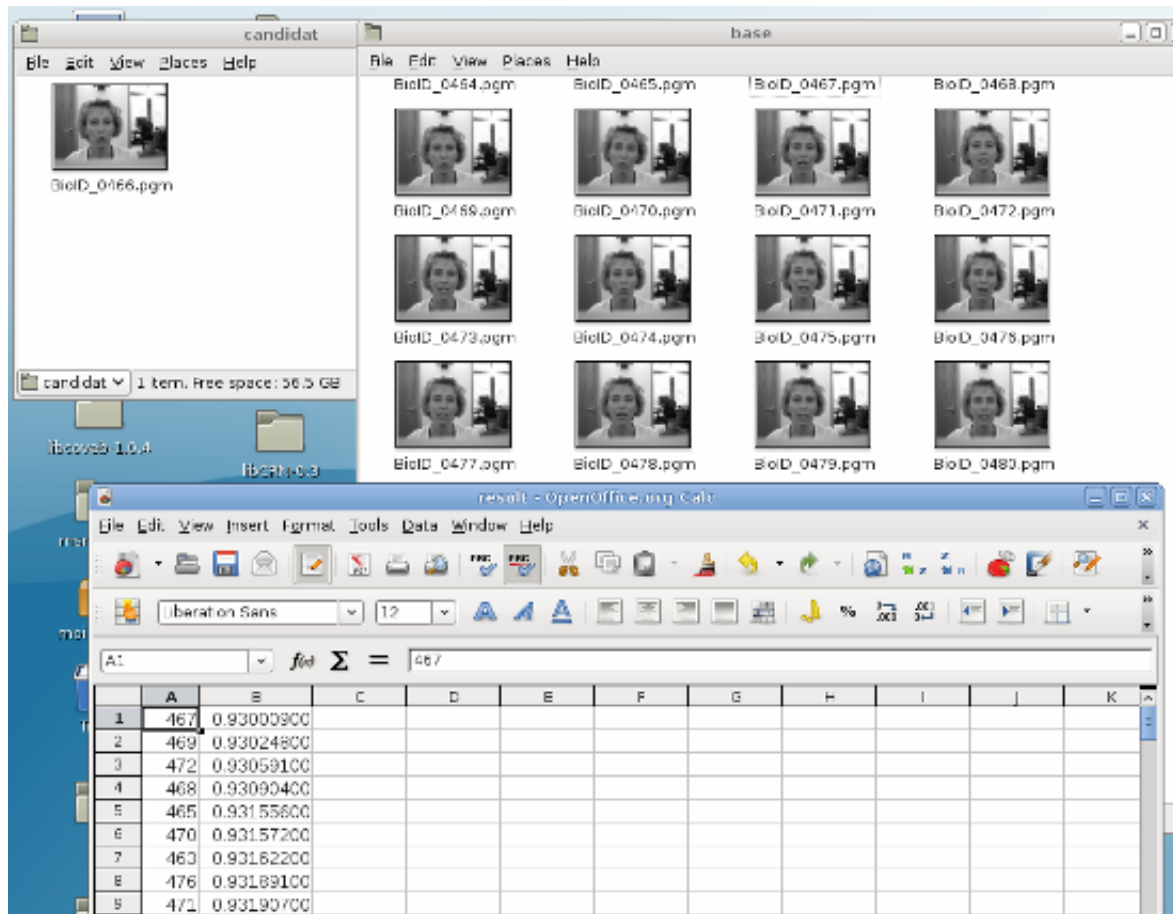


Fig. 20 : résultat d'une reconnaissance de visage dans la base BioID par application de la distance de compression : une image de visage (fenêtre "candidat") a été extraite de la base (fenêtre "base"). Chacune des distances de l'image candidate par rapport aux images de la base, a été importée dans un tableau dans lequel un tri par distance croissante a été effectué.

### Gestes 2D

Une base de 256 tracés graphiques utilisée dans un prototype d'interface tactile développée à Orange Labs (Fig. 22): le clustering de ces échantillons de tracé est de l'ordre de quelques minutes. Le graphe obtenu correspond à une classification cohérente des tracés relevés en coordonnées (x,y,t). Il révèle, en outre, les erreurs d'étiquetage et les doublons dans cette base : deux tracés identiques avec deux étiquettes différentes affichent une distance de zéro.

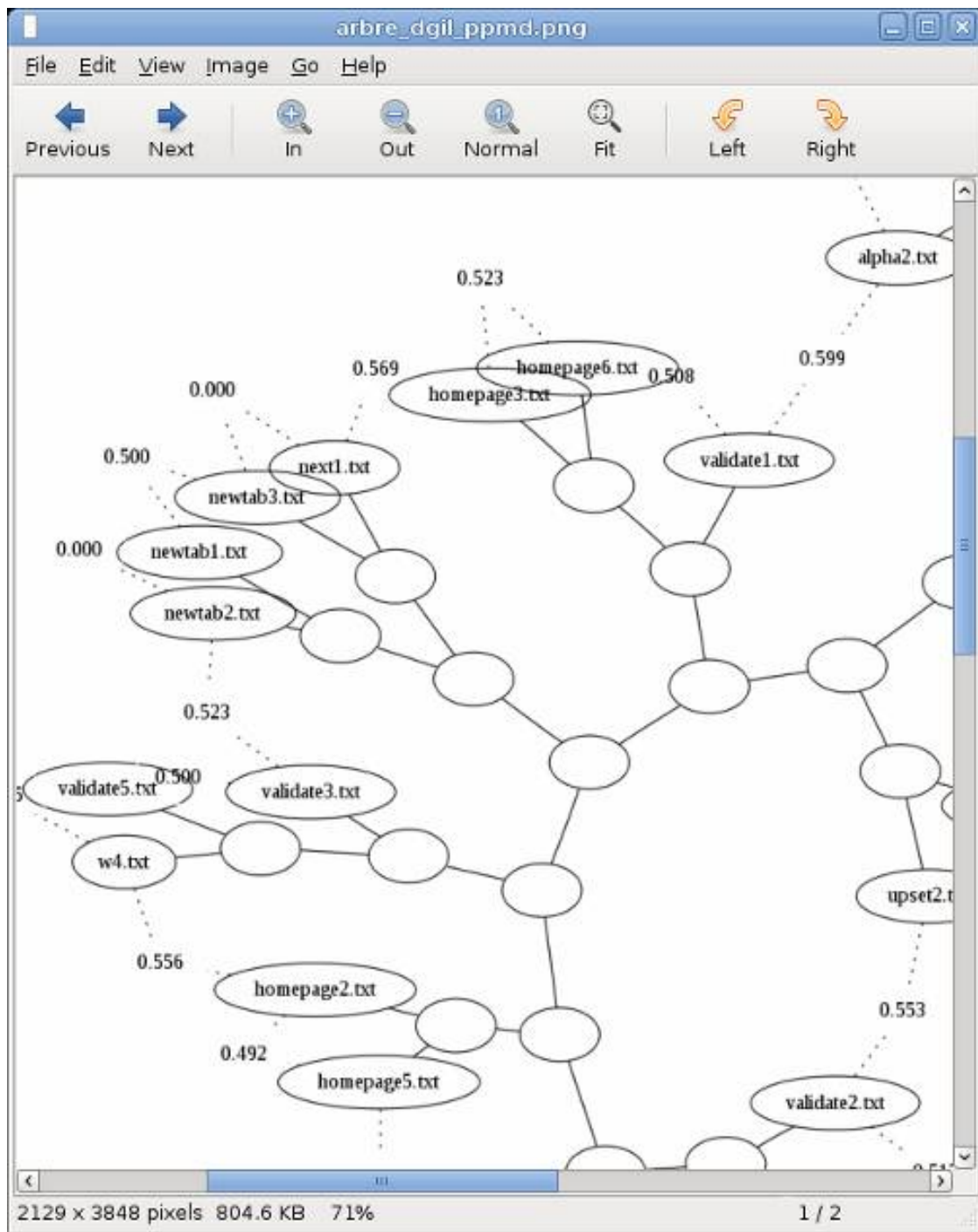


Fig. 22 : clustering d'un ensemble de tracés graphiques (relevé de coordonnées  $t=f(x,y,t)$ ) par analyse de la matrice de confusion obtenue à l'aide de la distance de compression : les tracés ont été étiquetés (nom des fichiers).

### Journaux anciens

Un répertoire de pages de journaux anciens numérisés (Fig. 23a et 23b). Cette expérience portait sur des documents ayant une mise en page hétérogène incluant un texte multicolonné et des illustrations dont la segmentation est particulièrement difficile à faire par les méthodes classiques.





Fig. 23a : échantillon de pages de journaux anciens traités

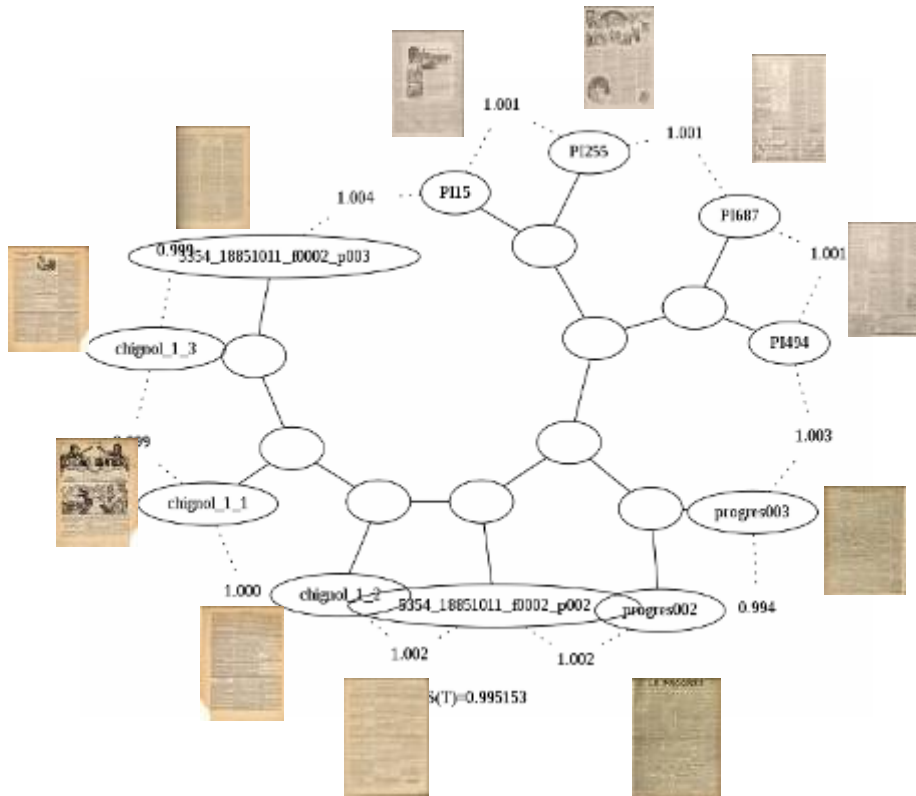


Fig. 23b : Arbre de distance obtenu par la distance de compression sur des pages de journaux numérisé.

## IV. Le potentiel d'applications utilisant un calcul de distance universelle

### IV.1. - Classement multimédia par le contenu

Le classement de photos numériques reste problématique du fait de la cohabitation de deux familles de paramètres : ceux, globaux, liés à l'histogramme de l'image et ceux, locaux, liés aux formes présentes dans l'image. Des problématiques similaires se rencontrent dans le classement de contenus audio.

L'hypothèse faite ici est que la compression de la séquence de bits conduit à effectuer une présentation de l'information mixant ces paramètres et calculant des caractéristiques calculables déduites de l'information jointe des deux contenus à mesurer. Quelques essais ont été faits sur des photos faite sur des scènes différentes et le classement proposé semble cohérent. Le fait que cette méthode de calcul de distance, basée sur la complexité, est utilisée avec succès pour construire des arbres phylogéniques à partir du séquençement de l'ADN, rend plausible cette hypothèse de cohérence dans le clustering de contenus multimédias.

### IV.2. - Propriétés de la complexité dans les compresseurs

On dénote une variation des valeurs de distances dépendant du compresseur utilisé. Il s'agira de déterminer la relation entre l'algorithme de compression mis en œuvre et la nature du contenu pour déterminer le cluster le plus pertinent (en termes de distribution inter-classes et intra-classe) dans une famille d'applications donnée.

Dans un premier temps, en déterminant le choix du compresseur et, dans un deuxième temps, la recherche de traitements conduisant à une "amplification sémiotique" du contenu, se traduisant par une meilleure discrimination des distances entre informations.

### **IV.3. - Mise en équations formelles des articulations de modalités**

La distance normalisée de compression présente deux propriétés avantageuses pour envisager cette mise en équation. Elle se calcule à partir de l'élément commun à toute information numérique, quelle qu'en soit sa modalité de présentation et possède toutes les propriétés d'une distance, notamment celle de l'inégalité triangulaire qui se rapproche de l'inégalité de Cauchy-Schwartz dans les espaces vectoriels.

D'autre part, du fait de la démonstration de la relation entre entropie de Shannon et entropie thermodynamique, on entre dans un modèle mécanique et thermodynamique du traitement de l'information, ce qui nous permet d'émettre l'hypothèse qu'un système d'informations se comporte comme un espace hamiltonien ayant autant de degrés de liberté que de modalités de présentation de l'information.

La mise en équation d'une articulation ou d'un alignement de modalités de présentation de l'information reviendrait à décrire les propriétés de ce système hamiltonien. En particulier, établir la variable aléatoire permettant de passer d'une présentation de l'information d'une modalité à l'autre.

### **IV.4. - Perspectives de recherche à plus long terme**

La distance normalisée de compression, hormis son universalité qui permet d'envisager des applications à courte échéance, est la première mise en application de la théorie de la complexité de Kolmogorov bien que ce soit au prix d'une approximation d'un problème incalculable.

Les lemmes et hypothèses ayant conduit à démontrer cette applicabilité, renforcent le lien des théories de l'information avec la physique.

Selon le niveau de complexité des problèmes à formaliser en algorithmes, les questions de décidabilité et de calculabilité peuvent rapidement nous confronter à des situations de calculs non linéaires et des états stochastiques et superposés de l'information. Cela correspond à des contextes dans lesquels les théories de Shannon Weaver ne peuvent plus s'appliquer, ou aboutissent à des modèles fortement réductionnistes : à un effet, peut correspondre plusieurs causes qui sont autant d'état superposés de l'information. Toutefois, des éléments de solution dans l'application des modèles de la physique statistique et quantique.

On notera que la théorie du "calcul réversible" de Charles Bennett est une des bases de la théorie quantique de l'information, ce dernier ayant appliqué cette propriété de réversibilité pour la cryptographie quantique.

## ***V. Conclusions sur les applications de la complexité***

L'étude de la complexité conduit à une simplification algorithmique radicale.

Par ces travaux, nous n'avons pas introduit de changement de paradigme, mais nous sommes plutôt revenus à des concepts de traitement de l'information par le

calcul. La nécessité d'appliquer la théorie de la complexité est devenu indispensable pour rendre compte de la complexité de nos ordinateurs actuels.

Il s'agit d'un travail de recherche appliquée qui s'inscrit en continuité avec l'évolution des théories de l'information. Elle n'induit pas de changement de paradigme aussi violent qu'à pu être celui de l'intelligence artificielle et du web sémantique des années 1990.

L'approche proposée dans ce travail de recherche induit un repositionnement du sujet de la présentation de l'information pour l'interaction et l'indexation, par rapport aux évolutions des théories de l'information contemporaines. Les premiers résultats de cette approche par les systèmes complexes, sont, a priori, surprenant, car ils conduisent à des solutions d'implémentation très simples, voire triviale. Ici, nous obtenons une divergence de Kullback-Leibler à partir d'une formule arithmétique simple.

Pour illustrer le potentiel de l'approche de la reconnaissance de formes par la complexité, nous avons développé un démonstrateur de reconnaissance intuitive de caractères que nous présentons en détail au chapitre 6. Cette maquette inclut des méthodes simples de segmentation (décomposition en composantes connexes), de mesure (projection de la composante connexe en x et y) et de classification (distance euclidienne sur le premier plus proche voisin). Elle est associée à une méthode de reconnaissance graphique basée sur une classification markovienne des graphèmes, pour la correction des résultats de reconnaissance de caractères. Ce couplage d'éléments logiciels simples, comporte une stratégie élémentaire d'apprentissage.

Le système obtenu est un logiciel d'OCR (Reconnaissance de Caractères Optiques) performant en termes d'utilisabilité dans une interface homme machine. Les performances d'un tel système sont observables mais difficilement quantifiables dans une logique classique de l'évaluation, car ce système potentialise les interactions entre chacun de ses composants et dispose d'une profondeur logique dépendante de sa mémoire des actions passées ; il s'agit de deux propriétés d'un système complexe.

## Chapitre 6

### Mise en œuvre de la présentation de l'information

*"Si nous examinons les lois générales de la perception, nous voyons qu'une fois devenues habituelles, les actions deviennent aussi automatiques."*

**Vaclav Havel** -"Anatomie du gag"

Le chapitre précédent montre que l'étude d'un problème (ici, celui de la distance entre informations) traitée sous l'angle des systèmes complexes, conduisait à une solution simple sur le plan algorithmique. L'OCR présenté dans le chapitre qui suit résulte de la même approche, en ce sens que c'est l'architecture assemblant des composants logiciels qui donne au système son "intelligence" et non la complexité propre à chacun de ses composants.

L'"absence de complexité" dans les composants est compensée par l'interface utilisateur qui devient un véritable outil de contrôle de l'exécution d'un programme : tout programme produit des données restituées à un utilisateur selon un mode de présentation de l'information, ce dernier module en validant ou invalidant au fil de l'eau les données produites par un calcul.

Un test de décision sur la qualité de la segmentation par rapport aux données de référence du classifieur, est effectué pour déclencher ou non un apprentissage a posteriori à partir de l'action de l'utilisateur alors qu'il est dans le contexte de calcul produisant ce résultat.

L'application de cet OCR original sur des images dégradées de documents, tout comme sur des documents anciens, montre la robustesse de l'approche.

## *I. Révéler les potentialités techniques de notre approche*

Dans ce chapitre nous présentons des applications issues des développements théoriques de cette thèse :

- la maquette d'un OCR intuitif utilisable pour des images de basse qualité de documents hétérogènes,
- un système de construction incrémentale de bases terminologiques à partir de documents,
- la conception d'un système de lecture nomade incluant un auxiliaire de navigation contextuelle.

La méthodologie appliquée s'appuie sur des contraintes d'usage, à partir desquelles nous avons déduit les briques technologiques et leurs associations. Ces contraintes reposent sur la prise en compte du contexte des terminaux mobiles et de l'accessibilité à l'information.

Ceci nous a conduit à la définition d'un système qui a toutes les propriétés de la complexité, puisque l'on est toujours dans le contexte d'une information numérique multimédia et multimodale, associé à une méconnaissance a priori des actions de l'utilisateur.

Les éléments structurants permettent de définir un tel système. Ils doivent exprimer la présentation d'un contexte des activités de l'utilisateur et de la machine, en particulier, les traces d'exécutions articulées à celles de l'activité de l'utilisateur. Ces présentations de l'information sont portées par les messages de feedbacks et de feedforward (homme ou machine) dont la dynamique d'interaction joue le rôle de régulateur de l'exécution d'un programme par des informations apportées par l'utilisateur.

Le type et la nature de ces informations de contrôle représentent le lien que l'on peut établir entre l'ergonomie d'une interface et la technique mise en œuvre dans une application. Leurs catégorisations et leurs présentations conduisent ensuite à la mise en place d'une boucle d'interaction pour laquelle, l'apparition d'une exception (un évènement inattendu) correspond à la nécessité de remettre en cause les feedbacks traités par l'interface et non la remise en cause de la boucle d'interaction elle-même. Cette remise en cause sera :

- pour l'ergonome : dictée par l'affordance d'un message par rapport au dialogue homme machine,
- pour l'informaticien : dictée par le coût en complexité de calcul permettant de produire ou d'analyser un message de l'interface.

Cette méthodologie recommande, en tout premier lieu, d'identifier les boucles d'interactions permises d'un système en fonction de ces propriétés structurantes ; ces boucles d'interactions représentent des "points de visite" ou plutôt, des points de mesures du comportement du système au travers de la présentation de l'information.

Si la conception d'un système de lecture nomade en est, pour le moment à un stade purement conceptuel, l'OCR intuitif, quant à lui, a été développé au stade de démonstrateur. C'est la raison pour laquelle nous décrivons cette application dans ces détails.

Le système de construction incrémentale de base terminologique, quant à lui, est évoqué ici pour montrer que la méthodologie mise en œuvre est applicable à plusieurs formes de projets informatiques : la connaissance sur les contenus et leur structure est portée par un calcul fait sur des données pour présenter l'information ; elle n'est plus déterminée a priori par des heuristiques.

## ***II. Conception et réalisation d'un système d'OCR intuitif***

Les terminaux mobiles actuels portent des caméras qui permettent de prendre des clichés de documents ; certains d'entre eux proposent même une reconnaissance de caractères en intégrant des versions réduites d'OCR.

Un terminal mobile impose plusieurs contraintes à l'utilisateur, en particulier, la taille réduite de l'écran, la définition d'un clavier réduit et un dispositif de pointage tendant à devenir tactile avec ou sans retour haptique.

Les interfaces des versions bureautiques des OCR ne sont pas adaptées à ces contraintes, leur conception repose sur l'environnement graphique d'un ordinateur personnel.

Nous proposons la conception d'un système d'OCR en prenant en compte ces contraintes. Pour se faire, nous avons adopté une démarche résolument systémique en nous fixant des contraintes initiales dont, en particulier, une grande simplicité des algorithmes. Nous voulions démontrer que l'efficacité d'usage d'un OCR repose d'avantage sur la conception de son interface que sur le renforcement des contraintes et paramètres de ses constituants.

L'application obtenue est un OCR dont la prise en main est facilitée sur le plan qualitatif.

Sur le plan quantitatif, cette maquette démontre le besoin de revoir les méthodes d'évaluation d'un résultat de reconnaissance de caractères : l'assemblage d'outils et d'algorithmes simples auquel nous avons procédé a conduit à des performances que ne laissaient pas prévoir les propriétés de chacun des composants pris isolément.

### **II.1. - Contraintes initiales pour notre OCR intuitif**

L'implémentation d'un module d'OCR dans un terminal portatif pose les contraintes initiales suivantes :

- l'acquisition de l'image se fait au moyen du capteur d'image présent sur le terminal (Fig. 24).
- la prise de vue du document se fait dans les conditions ambiantes de l'utilisateur au moment du cliché.
- on n'applique aucun critère de sélection a priori sur la nature du document que l'on veut numériser : les images documents (textes et bandes dessinées) proviennent indifféremment de captures sur le vif de requêtes internet, de livres, revues ou notes techniques.
- la finalité de l'application est la reconnaissance d'un fragment du texte par sélection d'une zone d'intérêt dans l'image pour en faire une requête vers un moteur de recherche.

- le terminal, support de ce démonstrateur, est doté d'une interface stylet et dépourvu de clavier mécanique.
- le traitement de la validation des résultats de l'OCR ne doit pas être bloquant et ne doit changer le schéma de l'interface présenté à l'utilisateur que dans le cas d'un problème insoluble. En outre, l'application doit préserver la fluidité du dialogue homme machine et éviter, dans la mesure du possible, l'affichage d'un clavier virtuel.
- le module de reconnaissance de caractère se doit d'être évolutif par la mise en œuvre d'une stratégie d'apprentissage simple pour l'utilisateur, après correction d'un caractère erroné.
- la réalisation technique de ce démonstrateur doit être la moins coûteuse en ressources et en volume de données de référence.

Ces contraintes conduisent à produire une **application légère** de reconnaissance de caractères, **portable** sur différents terminaux et dont l'ergonomie est **intuitive** pour l'utilisateur. Cette portabilité sur différents type de terminaux est assurée par le recours à des bibliothèques standards pour la construction de l'interface. Nous avons développé deux solutions (GTK+ et JAVA) autour d'une bibliothèque de manipulation d'images et de graphismes en C++ standard.



Fig. 24 : Présentation de l'OCR intuitif implémenté sur un umc (ultra mobile PC) sous Linux Debian.

## II.2. - Modularité de l'OCR intuitif

L'OCR intuitif que nous avons développé n'est pas un logiciel "monobloc". Il s'agit d'un assemblage logiciel de composants et de données, effectué avec un fort souci de modularité de manière à permettre d'échanger facilement l'un des modules selon le type de texte à reconnaître.



Ces composants logiciels, qui sont de fait les éléments structurants du système, sont interchangeables sans redéveloppement de l'architecture du système.

#### *L'outil de segmentation*

Dans notre maquette, la segmentation repose classiquement sur une binarisation de l'image suivi d'une analyse en composantes connexes avec regroupement des caractères en mots par mesure d'écartement.

#### *L'outil de mesure des formes*

Nous avons volontairement choisi une métrique rudimentaire, à savoir les histogrammes de projection en x et en y des composantes connexes. S'il est habituellement admis que cette méthode de mesure induit de l'ambiguïté de présentation des formes, nous avons pu observer que ce type rudimentaire de descripteur apportait au système sa tolérance au changement de polices de caractères.

#### *L'outil de classification*

Là encore, nous avons volontairement choisi un outil rudimentaire qui se contente de rechercher le premier plus proche voisin par une distance euclidienne. Par conséquent, le système ne gère pas d'hypothèses multiples, le problème de la validation des hypothèses étant intégralement transposé dans l'interface homme machine.

#### *L'outil de reconnaissance de mots*

Cet outil minimise le recours à la fonction de reconnaissance de caractère qui représente l'élément le moins fiable de la chaîne de traitement et repose sur l'emploi d'un lexique. L'idée consiste à opérer une reconnaissance de caractères suffisante pour obtenir un résultat. La recherche d'un mot dans le lexique se fait de manière itérative en prenant comme données le nombre de caractères (correspondant ici, au nombre de composantes connexes formant le "pattern" mot), ainsi que la valeur du premier et du dernier caractère. L'incrémentation de la recherche de mot se fait en ajoutant progressivement d'autres valeurs de caractères jusqu'à ce que la recherche retourne une valeur de mot unique. La composition de la requête de recherche dans le lexique, se fait indifféremment à la suite d'une reconnaissance de caractères ou d'une correction de reconnaissance par un tracé graphique.

#### *L'outil de saisie et de reconnaissance graphique*

Il s'agit d'un module intégré développé à Orange Labs (E. Petit, S. Vidal), à partir des travaux de T. Artières et P. Gallinari. Schématiquement, le tracé de l'utilisateur est segmenté en primitives graphiques classifiées par rapport à un ensemble limité de "caricatures" de graphèmes. La description d'une forme est donnée par le chaînage de ces caricatures et la classification se fait selon une approche markovienne incrémentale. On pourra noter que c'est également dans cette simplification des descripteurs de formes que réside la souplesse de ce système de reconnaissance de symboles graphiques ; cette méthode est omniscriteur.

#### *L'outil de correction et d'apprentissage*

L'OCR intuitif est supervisé de manière dynamique et reprend l'expérience de pensée de la belette dans le labyrinthe. L'utilisateur est simplement alerté par la présentation du résultat de reconnaissance de mot (mot incomplet ou erroné).

L'action correctrice consiste à "redessiner" l'un des caractères du mot pour en soumettre le tracé à une reconnaissance graphique. L'effet sur le système est triple :

- La surface englobant le tracé est une information permettant de corriger à la volée la segmentation due à des troncatures ou des caractères collés. Ceci permet d'ajuster le nombre de caractères d'un mot dans la requête vers le lexique.

- La valeur reconnue par reconnaissance graphique est affectée à la forme de caractère en cours de traitement. L'utilisateur a droit à trois tentatives au-delà desquelles, le système considère que la reconnaissance est insoluble et affiche une zone de saisie au clavier.

- Si la relation d'effectivité à une solution unique (à une valeur de caractère correspond une forme unique), la forme traitée est injectée dans la classe étiquetée par la valeur reconnue, de la base de classification. C'est cette formalisation du problème d'effectivité qui constitue le déclencheur de l'apprentissage.

*Les données de l'OCR intuitif*

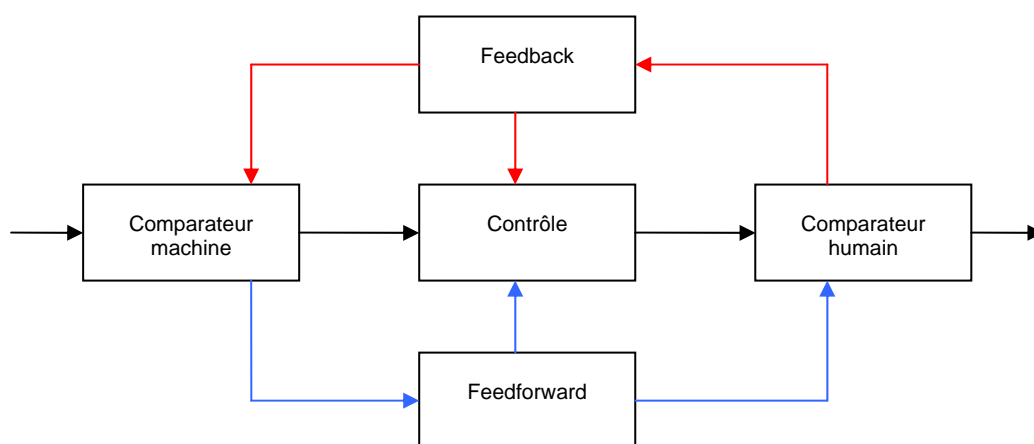
Ces données sont externes au système :

- Une base de caractères images étiquetés : cette base a été élaborée avec Yann Leydier du LIRIS à l'INSA de Lyon. Sa structure est simple : à l'installation de la maquette, les classes correspondant aux caractères d'un alphabet ne contiennent qu'une seule forme représentative. Ces classes sont enrichies de manière incrémentale lors de la correction faite par l'utilisateur d'une reconnaissance erronée d'un caractère par la reconnaissance d'un tracé graphique équivalent de ce caractère.

- Une base de tracés de symboles graphiques étiquetés : cette base fait partie du module de reconnaissance de tracés développé à Orange Labs. Elle se compose d'un ensemble de fichiers comportant pour chaque symbole, un ensemble de primitives graphiques. A l'heure actuelle, cette base est invariable ; il n'y a pas d'apprentissage des tracés graphiques intégrés à l'OCR intuitif. Il s'agit d'une évolution que nous envisageons de faire mais qui nécessite de se pencher sur les mécanismes déclencheurs de l'apprentissage.

- Un lexique de formes linguistiques : ce lexique des formes linguistiques françaises se présente sous l'aspect d'une liste d'environ 200000 entrées au format ASCII. Ce lexique est celui utilisé par Microsoft Word pour son module de correction orthographique. Nous disposons plusieurs lexiques sous cette forme, selon la spécialisation du vocabulaire.

Le fonctionnement de cet OCR est celui d'une machine assistée par l'utilisateur, conformément au schéma présenté au chapitre 3 que nous reproduisons ici :



L'IHM visualise la présentation des informations de feedforward traduisant l'exécution du programme d'OCR et celle des informations de feedback traduisant les actions de corrections éventuellement faites par l'utilisateur.

Le système a été conçu selon une approche globale ; on traite les problèmes de la reconnaissance de caractères à partir des contraintes d'usages et non à partir de la complexification des algorithmes. En ce sens, la mémoire de l'OCR intuitif est enrichie selon le principe de profondeur logique décrit dans les propriétés fondamentales des systèmes complexes : on garde la trace de l'action la plus économique pour l'apprentissage en prenant en compte la résolution du problème d'effectivité.

Associer quand nécessaire, une reconnaissance de caractères à une reconnaissance graphique correspond à une stratégie de co-alignement de modalités ; cela permet, précisément, d'établir la profondeur logique du système que nous rappelons être la solution la moins complexe (graphique ou image) de reconnaissance d'un caractère. Les modalités en présence sont l'image et le geste symbolique.

L'évolution prévue pour l'OCR intuitif réside dans l'approfondissement des propriétés de cette articulation de modalités. Actuellement, l'apprentissage ne concerne que la reconnaissance d'images de caractères. Il s'agira de mettre en place une stratégie pour déclencher l'apprentissage d'un nouveau tracé graphique ou d'une nouvelle forme linguistique.

### II.3. - Description fonctionnelle de l'OCR intuitif

Notre approche conduit à un système assisté par l'utilisateur (son organigramme simplifié est présenté en Fig. 25). Les points d'actions de ce dernier sont

- les des interruptions du processus de reconnaissance correspondant à des demandes d'informations faites à l'utilisateur,
- les temporisations permettant à l'utilisateur de "corriger le cap" en cours de traitement, par ajout d'informations.

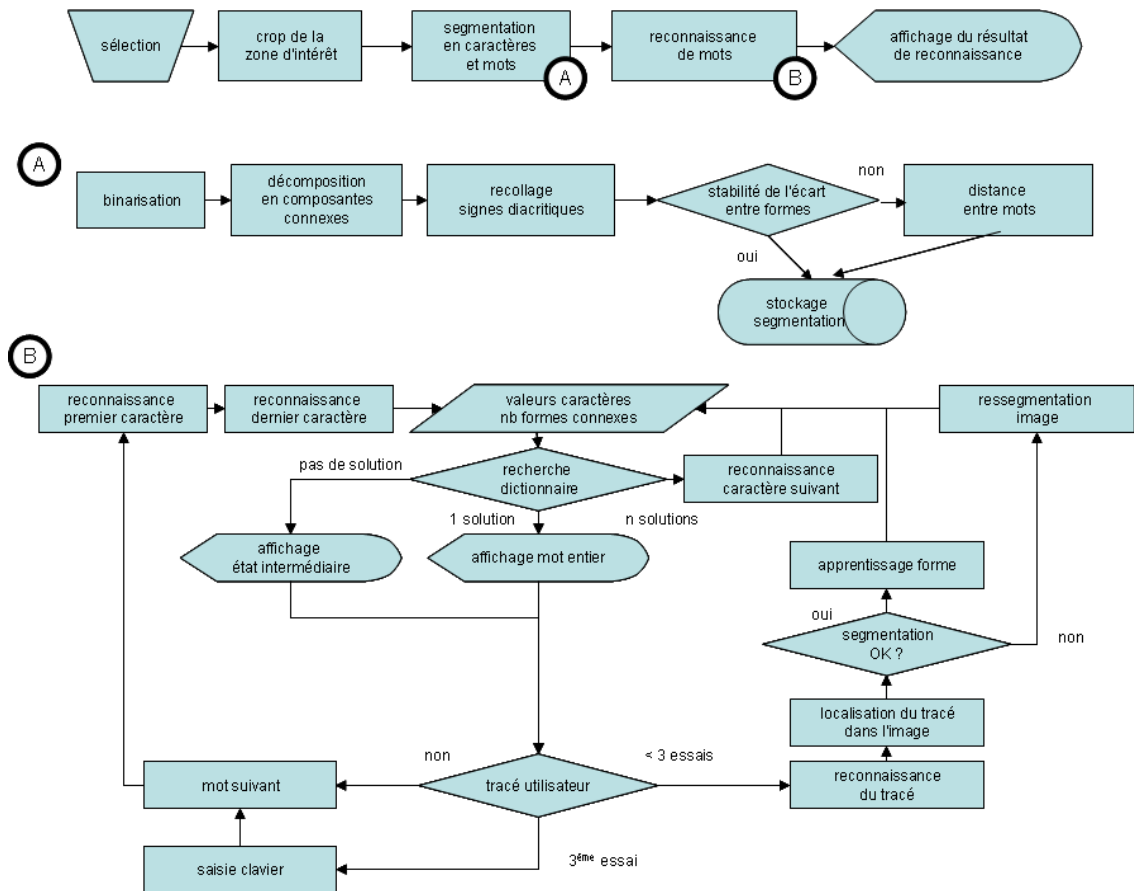


Fig. 25 : Schéma fonctionnel de notre OCR intuitif. Les renvois A et B correspondent à une description plus détaillées.

L'interface de l'application est présentée en Fig. 25.



Fig. 26 : interface du démonstrateur de l'ocr intuitif.

La Fig. 27 montre le synoptique du fonctionnement de l'OCR intuitif en situation de correction d'une erreur d'OCR par le tracé de l'utilisation : le processus de reconnaissance du mot "ornithorynque" s'interrompt en raison d'une confusion sur la reconnaissance du troisième caractère ; il n'y a pas de solution correspondante dans le lexique des formes linguistiques.

La correction effectuée par l'utilisateur consiste à redessiner le caractère "en erreur" afin de proposer ce tracé à une reconnaissance graphique.

Le résultat de cette reconnaissance graphique vient se substituer au résultat de l'OCR et une nouvelle recherche d'occurrence de mot est faite dans le lexique.

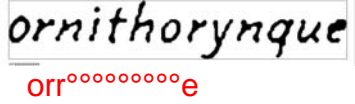
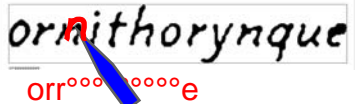

	Affichage de la zone d'intérêt binarisée et du résultat intermédiaire de reconnaissance du mot.
	L'utilisateur retrace la forme de la lettre "n" au stylet
	Le tracé de la lettre "n" a été reconnu par et une nouvelle recherche est faite dans le dictionnaire de l'ocr conduisant à une solution unique

Fig. 27 : Synoptique de déroulement d'une procédure de correction dans l'ocr intuitif

La Fig. 28 montre l'interruption du processus de reconnaissance de mot dû à une erreur de segmentation en composantes connexes. Les 5<sup>ème</sup> et 6<sup>ème</sup> caractères du mot récurrence sont collés, ce qui conduit à une hypothèse de mot de 9 caractères. L'action correctrice consiste à redessiner l'un des caractères collés. Le tracé est soumis à une reconnaissance graphique, tout comme dans la Fig. 26, la surface englobant ce tracé permet de ressegmenter le mot et une nouvelle recherche d'occurrence est faite dans le lexique.

Remarques :

- La même logique de traitement s'applique pour les caractères tronqués. La surface englobant le tracé permet de regrouper les composantes connexes d'un caractère tronqué.
- Dans tous les cas une temporisation de la reconnaissance de caractère est faite pour laisser le temps à l'utilisateur de corriger toute erreur de reconnaissance.
- Si le caractère en cours de correction n'est toujours par reconnu après trois tentatives de tracé graphique, un secours est donné par une zone de saisie couplée à un clavier virtuel est affichée pour ne pas bloquer l'application.

	Extrait du document lu. Le mot sélectionné par l'utilisateur est "récurrence"
---	---

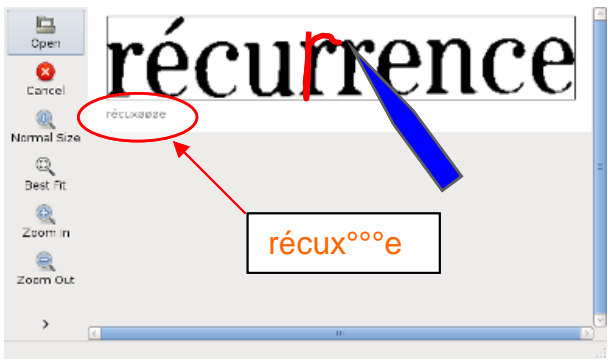
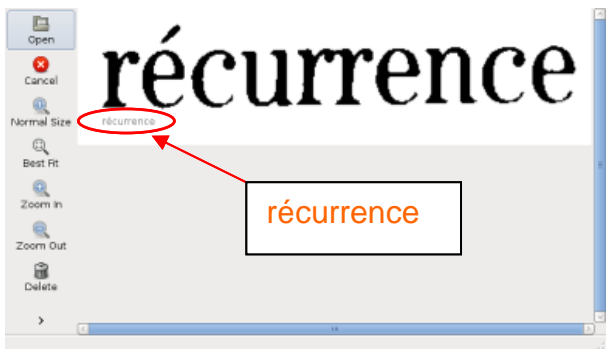
	<p>Après découpage de la zone d'intérêt et binarisation, il apparaît que deux caractères sont collés ("rr"). La tentative de reconnaissance conduit à une proposition insoluble que voit l'utilisateur pendant le traitement. Ce dernier retrace un des caractères</p>
	<p>La surface englobante du tracé de l'utilisateur a été utilisée pour couper la composante connexe en défaut, ce qui a permis de recalculer le nombre de lettre du mot. Le tracé a été reconnu comme étant un "r", ce qui a suffi pour que l'information ajoutée aboutisse à une solution unique de recherche dans le dictionnaire</p>

Fig. 28 : Exemple de correction de la segmentation par le tracé de l'utilisateur

#### II.4. - Discussions sur l'OCR intuitif

Notre approche a mis en évidence deux propriétés intéressantes pour de futurs OCR embarqués dans une interface :

##### 1) l'apport d'un OCR supervisé par un couplage de modalités

Nous avons considéré l'image comme une modalité de présentation de l'information et le tracé graphique comme une modalité associée de présentation de la même information. L'articulation de ces deux modalités permet à la fois, de moduler en continu le comportement de l'OCR et d'induire un mécanisme d'apprentissage incrémental de ce dernier. Du point de vue de l'utilisateur, les paramètres techniques de l'OCR sont entièrement masqués, ce dernier ne devant qu'observer et corriger les résultats de manière intuitive. La présentation de l'interface ne change pas de schéma d'organisation : une interface styler utilisée à la fois pour désigner la zone d'intérêt, ajuster le comportement de l'OCR et effectuer la requête vers le moteur de recherche.

##### 2) l'apport de la complexité dans l'OCR intuitif

Bien que les algorithmes mis en œuvre soient simples, le couplage de modalité nous conduit à un OCR performant dont les propriétés globales ne peuvent être déduites à partir des propriétés élémentaires de ces composants : l'algorithme de reconnaissance d'images de caractères est volontairement rudimentaire et celui de reconnaissance graphique dispose de peu de données de classification et le dictionnaire se réduit à une liste de mots.

Ce n'est qu'en étudiant la systémique de notre maquette que l'on peut expliquer ses performances.

L'affichage d'un résultat de reconnaissance de mot est l'information de feedforward du système, indiquant à l'utilisateur l'avancement du processus.

Le tracé de l'utilisateur est l'indice indiquant à la machine que le processus a une anomalie de fonctionnement. Ce tracé est une information de feedback dans le sens homme machine, relative à la validation de l'OCR.

- l'utilisateur désigne une zone d'intérêt posant problème en redessinant le caractère devant être reconnu.
- La mesure de la surface englobant ce tracé donne un indice saillant de la qualité de la segmentation.
- en tentant de reconnaître le tracé, on obtient un indice saillant sur la valeur de la forme désignée par l'utilisateur.

Du point de vue de l'interaction, nous sommes dans le contexte d'une machine assistée par l'homme au travers de la mesure de feedbacks.

Le tracé reconnu résout un problème d'effectivité de manière transparente pour l'utilisateur, en construisant une relation entre une forme redéfinie par le tracé de l'utilisateur et sa valeur dans l'espace de décision que représente l'alphabet ; la correction classique d'une erreur d'OCR revient à forcer la valeur d'une forme par ressaisie au clavier, sans remettre en question la segmentation.

La combinaison de deux modalités, visuelle (l'image) et gestuelle (graphique) nous permet d'identifier 4 états possibles du système :

- L'OCR a affecté une valeur à la forme et l'utilisateur n'effectue aucune action sur ce caractère reconnu :

Ceci correspond à une bonne reconnaissance du caractère. L'utilisateur n'a apporté aucune information complémentaire pour ajuster le comportement de l'OCR.

- la zone d'intérêt circonscrit une seule composante connexe dans l'image :

Cela correspond à une forme qui a été correctement segmentée, qui n'a pas été reconnue, ou mal reconnue par l'OCR. C'est cet état qui déclenche un apprentissage sur la forme après correction.

- La zone d'intérêt est englobée dans une composante connexe :

Cela correspond à une sous-segmentation de l'image dans laquelle les formes ne sont pas signifiantes par rapport à l'espace de décision. Actuellement, il n'y a pas de déclenchement d'un apprentissage. La seule opération effectuée est une coupure de la composante connexe qui ajuste le nombre de caractères pour la recherche du mot.

- La zone d'intérêt englobe plusieurs composantes connexes :

Cela correspond à une sur-segmentation de l'image dans laquelle les formes ne sont également pas signifiantes pour l'espace de décision. Actuellement, il n'y a pas de déclenchement d'un apprentissage. La seule opération effectuée ici est une fusion des composants connexes pour ajuster le nombre de caractères pour la recherche du mot.

## **II.5. - Application de l'OCR intuitif à la Bible glosée de Gutenberg**

L'exemplaire numérisé que nous avons soumis à l'OCR intuitif, nous a été fourni par Marjorie Burghart (EHESS pôle de Lyon). Il s'agit d'une Bible glosée imprimée à Strasbourg en 1458 par Gutenberg dont nous pouvons voir un exemple en Fig. 29.

### ***II.5.1. - Evaluation de la qualité du support et de sa numérisation***

Le support est remarquablement bien conservé mais nous y retrouvons les déformations inhérentes à la numérisation de livres anciens, notamment au niveau de l'axe de reliure.

Le papier présente également une certaine transparence posant le problème de transvision (on distingue le texte imprimé au verso d'une page).

La typographie est relativement régulière pour les techniques d'impression de l'époque.

La définition des fontes est totalement archaïque, de type gothique, avec les caractères en lettrines sur deux, voire trois couleurs. La taille des caractères est variable d'un fragment de texte à l'autre, selon qu'il s'agisse du texte biblique, de la glose ou d'annotations.

La langue utilisée est le latin ecclésiastique de la fin du Moyen-âge avec une utilisation d'abréviations dans l'édition proposée. La glose correspond à des annotations du texte biblique.

La numérisation est de bonne qualité en termes de résolution et de profondeur de pixels (600 dpi, image codée sur 24 bits). L'image n'est pas compressée.



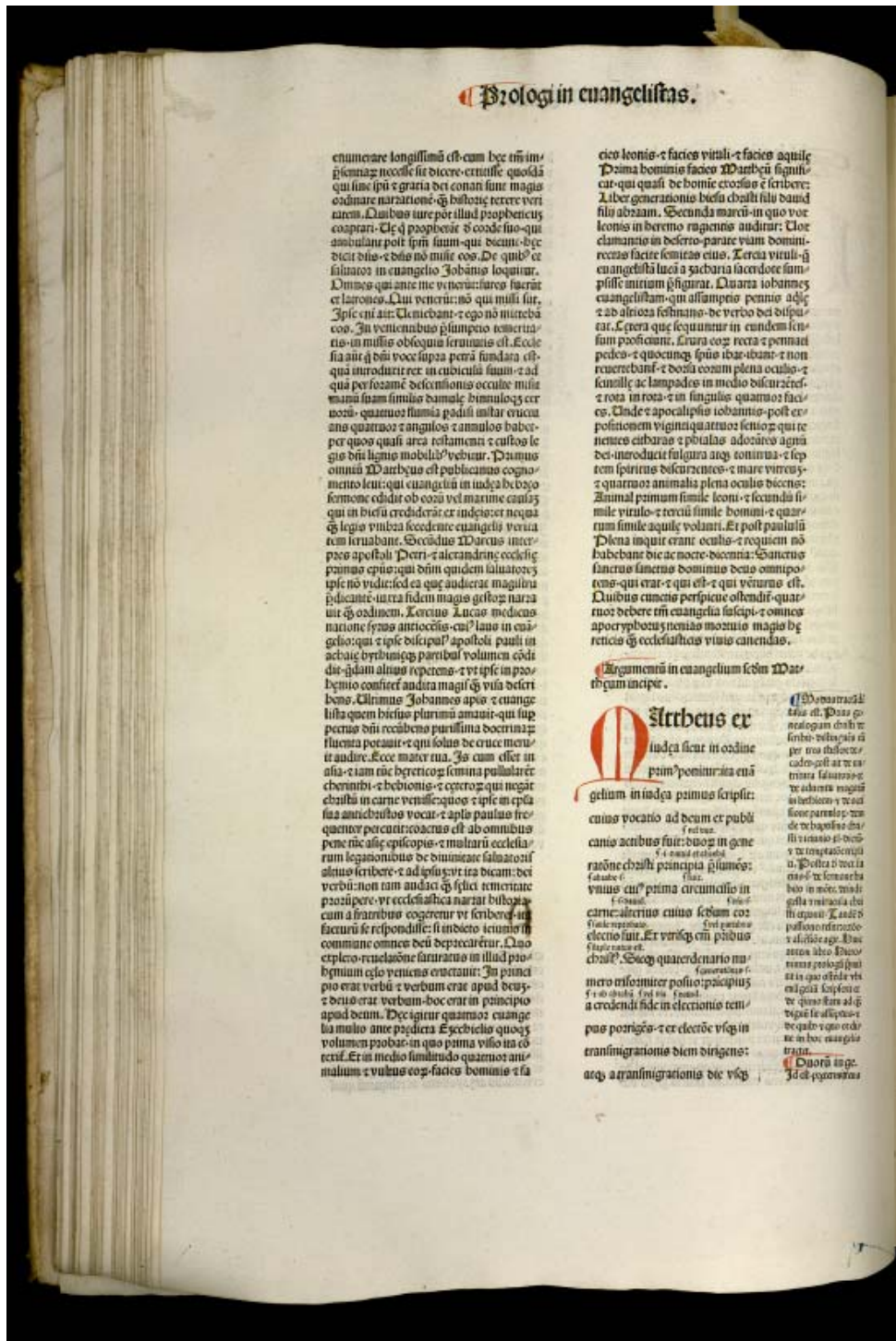


Fig. 29 : Exemple de page numérisée de la Bible glosée imprimée à Strasbourg en 1458 par Gutenberg.

### II.5.2. - Reconnaissance de mots par l'OCR intuitif

Nous avons remplacé le fichier de lexique des formes linguistiques françaises, par celui d'un lexique d'environ 400000 formes linguistiques latines médiévales, fourni par Marjorie Burghard, pour la prise en compte de cette langue par l'OCR. La différence de taille par rapport au lexique français s'explique par les déclinaisons de la grammaire latine ; cette taille est moins importante qu'on aurait pu le prévoir en raison du vocabulaire.

Il n'y a pas eu d'entraînement préalable ni de remplacement de la base de caractères de l'OCR ni de celle de la reconnaissance graphique.

Les manipulations auxquelles nous avons procédé ont consisté à sélectionner un mot dans le texte pour le soumettre à l'OCR intuitif selon la même démarche que celle décrite au paragraphe III.3 du présent chapitre.

L'objectif a été de nous focaliser principalement sur la méthode de correction et non sur le taux de reconnaissance.

#### a) reconnaissance du nom "Mattheus" avec apprentissage

L'exemple présenté ici correspond à une correction d'un résultat d'OCR suivi d'un apprentissage. Il a pour objectif de montrer l'effet du couplage entre la reconnaissance d'un caractère dans une image et la reconnaissance d'un tracé graphique

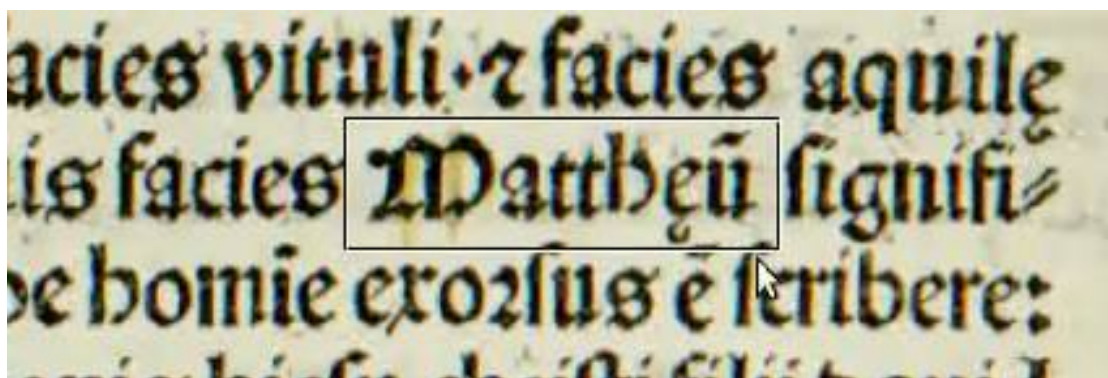


Fig. 30a : sélection du mot "Mattheus"

Après sélection du mot "Mattheu" par l'utilisateur, le fragment d'image est découpé, binarisé et une détermination des composantes connexes est faite. La reconnaissance du mot révèle que le premier caractère n'a pas été reconnu, ce qui suspend le processus. La Fig. 30b montre un extrait de copie d'écran de l'interface.



Fig 30b : proposition de reconnaissance par l'OCR intuitif. ("°°°°°u" en bas à gauche de l'image)

L'utilisateur trace une forme correspondant au caractère à reconnaître au moyen du stylet de l'interface. Ce tracé est reconnu comme un "m" minuscule ; à l'heure actuelle, les bases de références ne différencient pas les majuscules et les minuscules (Fig. 30c).

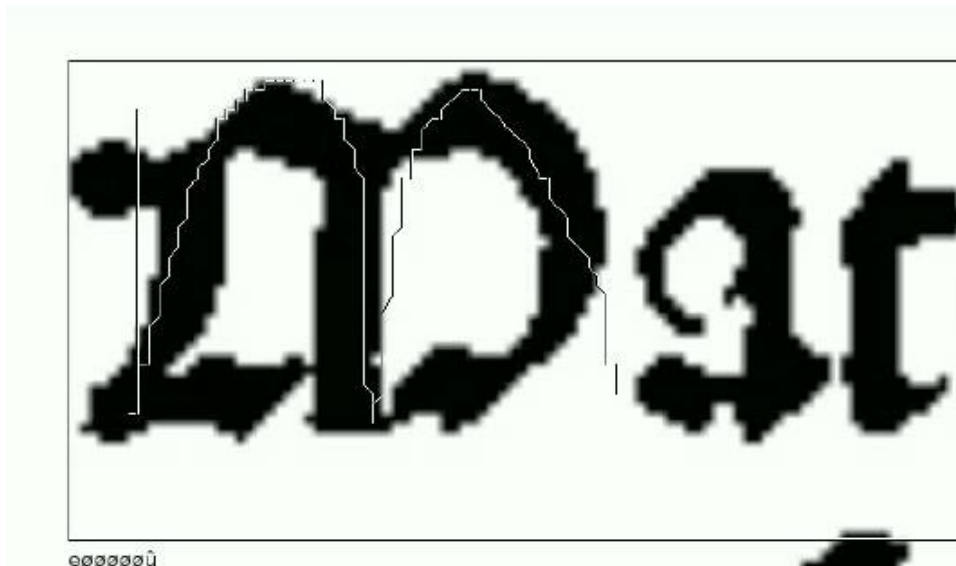


Fig. 30c : tracé graphique fait par l'utilisateur pour corriger le résultat de reconnaissance

Cette valeur est affectée à la première composante connexe de l'image et conduit à une proposition suffisante pour la recherche dans le lexique dont le résultat retourné est le mot "mattheù" (Fig. 30c).



Fig. 30d : reprise de la reconnaissance par l'OCR après correction.

L'état de l'OCR à l'issue de la correction conduit à une résolution correction du problème d'effectivité de la relation entre la composante connexe traitée et la valeur ajoutée par l'utilisateur via une reconnaissance graphique. Ceci induit un apprentissage de la base d'image de caractères en insérant la composante connexe traitée dans la classe "m" de cette base.

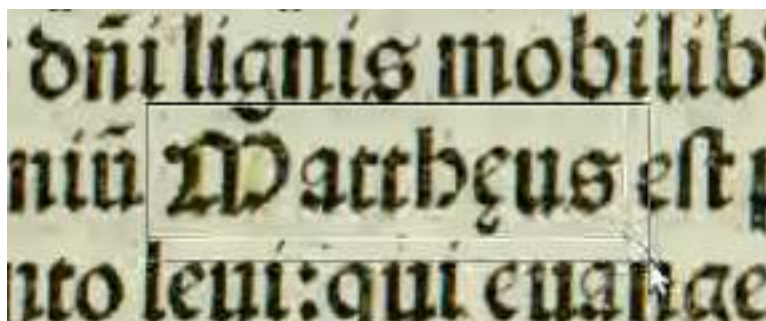


Fig. 30e : sélection et reconnaissance d'une autre occurrence du mot "Mattheus" dans la page

La Fig. 30e montre la reconnaissance immédiate du mot "Mattheus" après apprentissage de la forme gothique du caractère m. D'autres essais de reconnaissance du caractère "m" ont été refaits sur d'autres documents et n'ont pas montré de régression.

#### b) correction du mot "david" sans apprentissage

L'exemple montré ici est une correction d'un résultat d'OCR par un tracé graphique sans apprentissage consécutif sur la forme non reconnue (Fig. 31a, 31b, 31c et 31d).

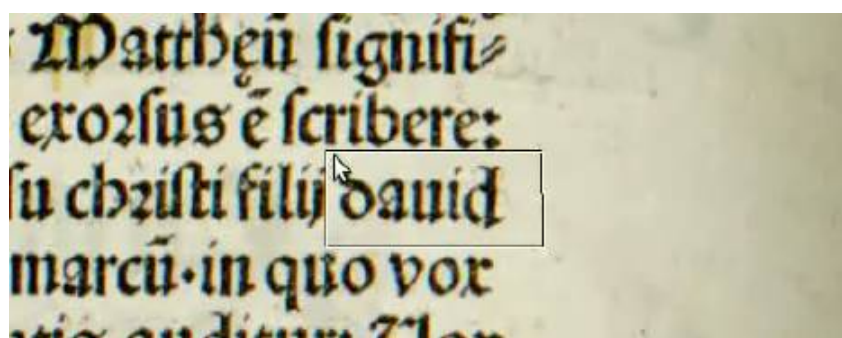


Fig. 31a : sélection du mot "david"

L'utilisateur trace la zone d'intérêt contenant le mot "david". Cette zone est traitée comme indiqué dans les exemples précédents (Fig. 31a).



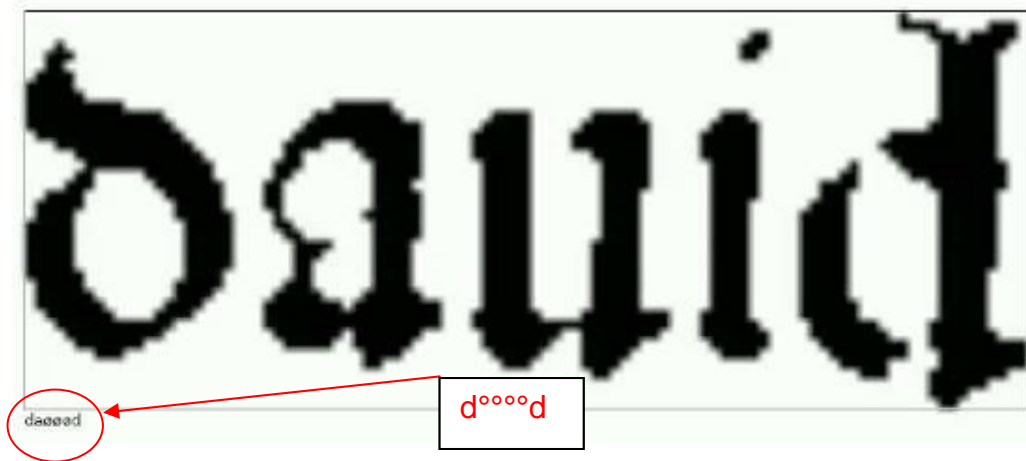


Fig 31b : proposition de reconnaissance par l'ocr intuitif.

Le résultat de la binarisation et du calcul des composantes connexes tronque le dernier caractère du mot à reconnaître. La conséquence est une proposition de reconnaissance n'ayant pas de solution dans le lexique provoquant la suspension du processus (Fig. 31b).



Fig. 31c : correction par tracé graphique fait par l'utilisateur

La surface englobant le tracé de l'utilisateur indique au système l'erreur de segmentation et conduit à une fusion des deux composantes connexes du caractère en cours de correction. Ceci permet d'ajuster le nombre de caractères dans la proposition de reconnaissance du mot (Fig. 31c).

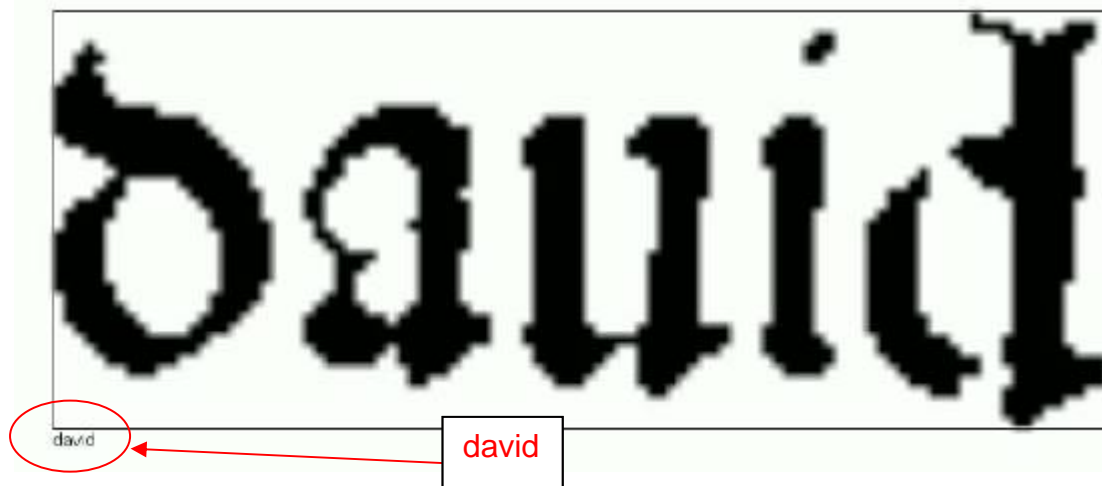


Fig. 31d : reprise de l'OCR intuitif après correction.

La valeur correcte de caractère retournée par la reconnaissance du tracé est affectée la composante connexe re-segmentée. La reconnaissance de mot reprend et retourne une valeur de mot correcte (Fig. 31d). Toutefois, l'opération de fusion des composantes connexes provoquée par le tracé de l'utilisateur ne satisfait pas à la résolution du problème d'effectivité traité par l'OCR pour cette forme. En effet, le système d'OCR ne connaît qu'une relation de bijection entre une forme et une valeur uniques.

### ***II.5.3. - Un projet d'OCR "tout terrain"***

Les deux exemples que nous venons de présenter montrent les potentialités techniques de l'OCR intuitif dans son utilisation comme outils d'aide à la transcription de livres à la typographie ancienne. Une tolérance à la variabilité des formes a pu être observée en utilisant un jeu réduit de document.

Il est tout à fait certain qu'une évaluation plus approfondie est à conduire. Cette évaluation est prévue pour la suite des travaux inclus dans une proposition de projet de retranscription de documents médiévaux. Dans cette future application, la sélection d'une zone d'intérêt sera revue en incluant la possibilité de choisir des paragraphes. D'autre part, nous allons étudier une approche incluant également une solution alternative à l'OCR, développée au LIRIS et appelée "Transcription Assistée par Ordinateur". Elle consistera à répercuter le résultat de la reconnaissance d'un caractère à l'ensemble des formes ressemblant à ce caractère dans le document ; l'évaluation se fera par une relecture du document par l'utilisateur en affichant mot à mot les formes et les mots reconnus. La correction se fera selon le principe de couplage de modalités. L'apprentissage actuel sera renforcé par les facultés de pouvoir effectuer un apprentissage sur les tracés et d'enrichir le lexique.

L'objectif de cette future application sera de fournir une plateforme de reconnaissance de documents anciens ayant de fortes capacités d'adaptation au contexte ; cette adaptation se fera dans une logique modules de données externes, insérés au système dans une logique de "plug-in", sans remise en question du moteur de reconnaissance lui-même.

### *III. Projet AdHoc : conception et réalisation d'un système de construction incrémentale de réseaux sémantiques*

Ce projet avait pour objectif la construction d'une plateforme d'évaluation de tous documents intervenant dans la relation clientèle de l'opérateur Orange. L'évaluation portait sur la qualité rédactionnelle des documents et se basait sur un ensemble de recommandations faites par les ergonomes, à partir des retours clients. Il s'agissait donc d'un projet pluridisciplinaire à vocation opérationnelle, couplant technologie et ergonomie.

La difficulté, qui se traduit en complexité des données, provient du fait que les documents entrant dans le périmètre du projet couvrent plusieurs domaines terminologiques allant du commercial au technique, en passant par le juridique.

Les documents entrant dans le périmètre d'application d'AdHoc, ont des cycles de vie instables qui dépendent des offres faites à la clientèle. Ceci se traduit par l'évolution permanente des formes de présentation et du vocabulaire.

Les résultats de ce projet ont été transférés vers des applications opérationnelles.

Le contexte du projet est le suivant :

- Inexistence de bases terminologiques définissant le vocabulaire de la relation client de France Télécom. Cette inexistence provient du fait de l'évolution rapide et constante du vocabulaire attaché aux produits et services de l'opérateur, mais aussi d'une prise de conscience tardive des questions liées à la mémoire d'entreprise.
- Existence de documents à destination des clients, rédigés par des spécialistes d'un métier de la relation client (principalement, commerciaux, ergonomes et maintenance).

L'approche retenue pour la conception de cette plateforme reprend certains aspects théoriques qui ont été mise en œuvre dans l'OCR intuitif. En particulier, le développement d'un processus assisté par l'utilisateur :

La machine soumet à l'utilisateur une synthèse calculée du document comprenant un abrégé et une liste d'expressions clés. Cette synthèse est une information de feedforward qui traduit les saillance du document en cours de traitement ; ce sont des formes syntaxiques nominales.

L'utilisateur affecte des attributs à ces formes selon leur valeur terminologique (contexte d'usage, signification, relations de synonymie) dans la mémoire du système.

L'état initial est ici, le texte d'un document et l'état final un "atome terminologique", c'est-à-dire, un mot ou une expression ramenée à un lemme de définition avec ses relations de dépendances et de synonymies. Ces atomes terminologiques sont des formes linguistiques obtenues par un calcul.

L'état de transition est le résultat d'une mesure de cooccurrences de termes conduisant à l'affichage d'un abrégé du texte et d'une liste de mots clés.

L'action régulatrice de l'utilisateur est ici la supervision des attributs donnés par le système à chaque mot ou expression. Ces attributs sont le classement terminologique a posteriori du mot ou de l'expression détectée dans le document.

Comme pour l'OCR intuitif, les informations de feedforward de la machine sont présentées à l'utilisateur qui, par ses actions de feedback, joue le rôle de régulateur.

La détection des formes linguistiques est faite au moyen d'outils simples issus du TALN (Traitement Automatique de la Langue Naturelle). Ces outils consistent en un abrégéur/extracteur de mots clés, basé sur la mesure de cooccurrence du second ordre entre termes. Le vocabulaire propre à chaque métier est inscrit dans les documents ; ces indices correspondent à des traces d'activités relatives à un métier ou une spécialité particulière.

Leur intégration dans une base terminologique revient à qualifier un terme lors de sa première présentation à l'utilisateur d'AdHoc, cette qualification fait appel soit à un rapprochement sémantique de bases terminologiques associées (bases techniques en particulier), soit à la construction incrémentale un nouveau référentiel terminologique. Dans ce cas, la détection d'un terme inconnu aura servi de révélateur d'une évolution du vocabulaire de la relation client.

L'un des autres objectifs d'AdHoc est la qualification de relations entre termes (syntagmes, mots composés, etc.). La présentation de ces termes est, en effet, variable d'une charte de recommandations d'usage d'un domaine terminologique à l'autre ; elle dépend également des habitudes rédactionnelles de l'auteur d'un document.

Ceci rend difficile l'utilisation d'un système à base de règles comme par exemple, celui de Word dans Microsoft Office qui détecte et tente de corriger à la volée un terme "mal composé".

La principale cause de ces difficultés est donnée par les contraintes initiales du projet qui rend impossible une approche déterministe d'un recensement exhaustif a priori de toutes les règles d'une charte évoluant en permanence, ainsi que leur transposition technique dans la plateforme.

L'approche constructiviste et incrémentale d'AdHoc permet d'une part, de détecter les traces ou saillance d'une évolution du vocabulaire ou de la mise en forme d'expressions et, d'autre part de qualifier ces traces en les inscrivant dans une base terminologique évoluant en permanence.

L'évolution de cette application consistera en l'établissement d'une méthode non supervisée d'identification d'un document et d'évaluation de sa qualité rédactionnelle.

Ce grading utilise un vecteur hétérogène de saillances de la qualité d'un document à partir de la combinaison de propriétés sémiotiques de mise en forme d'un terme d'un document et de propriétés sémantiques d'utilisation de syntaxes et de vocabulaires appropriés à la thématique portée par le document analysé.

L'ensemble forme un ensemble de données très hétérogène qui rend difficile la mise en œuvre d'une approche descriptive et structurelle. Cependant, ces propriétés d'usage se traduisent par des traces ou formes codées sur le document. L'outil qui devra permettre cette classification de document est la distance de compression, décrite au chapitre 5.

Elle nous permettra de classer les documents selon leur type (technique, commercial, juridique, etc.) dans un mode non supervisé et, dans un second temps, chaque



document sera évalué par sa distance par rapport au référentiel terminologique construit dans AdHoc.

#### *IV. Conception d'un système de lecture nomade de contenus multimédias*

Ce travail s'inscrit dans un projet de nouvelle interface pour terminaux légers incluant des fonctions d'aides à la navigation dans des contenus hétérogènes. Cet auxiliaire de navigation est le support d'une intermédiation entre le document et l'utilisateur. Son objectif est double :

- faire abstraction du format des documents en adaptant la présentation des contenus à un schéma d'interface unique,
- adapter l'organisation de l'interface du terminal à différentes formes de handicaps ou à des situations rendant l'utilisation de l'interface difficile.

La conception d'un système de lecture nomade apporte de nouvelles contraintes dans la définition de l'interface homme machine. Le contexte de la mobilité pose, outre la question de la spécificité des systèmes d'exploitation et des langages de programmation, celle de leur utilisation dans des appareils dépourvus de clavier et de systèmes de pointage identiques à ceux, désormais classiques d'un ordinateur individuel.

On peut établir des contraintes initiales pour de tels développements de services comme suit :

- la mobilité : elle implique des terminaux portatifs dotés d'écrans de taille réduite et dépourvus de clavier ou de souris. Ces dispositifs de pointage et de saisie d'informations sont remplacés par une interface tactile. Dans de telles situations, le codage des contenus, généralement conçu pour des microordinateurs, peut s'avérer totalement inadapté.
- l'hétérogénéité des contenus dus à la généralisation de l'information numérique, à la banalisation des moyens de production et à l'amollissement des actions de normalisation conduisant à une cohabitation compétitive de standards. Cette hétérogénéité est elle-même favorisée par la nécessité de prendre en compte un large panel de profils d'utilisateurs, dont en particulier les handicapés et les séniors.

Ces contextes sont donc ceux liés à la convergence de l'information numérique appliquée à la définition de nouvelles interfaces "hommeS machineS" pour lesquelles l'accès multimodal à une information codée de manière multimédia en est la propriété principale.

Dans cette approche, la présentation de l'information est le principal critère d'affordance des contenus, en ce sens qu'elle en conditionne l'accessibilité et la tangibilité par un utilisateur, quel qu'il soit. Du fait de la multiplicité des combinaisons possibles de présentations de contenus et d'usages, il devient impossible de rester dans le paradigme classique du document numérique basé sur une Unité Documentaire Globale contenant l'ensemble des modes de présentations et métadonnées permettant tous les modes d'interaction, sans imposer des contraintes devenant inacceptables pour certaines catégories d'utilisateurs.

Ces contraintes sont induites par le lien existant entre la technicité des contenus fixée à priori (leur codage physique, leur segmentation et métadonnées fixées) et leur adéquation à un usage connu a posteriori (sur quelle catégorie d'objets d'un document peut porter une action d'un utilisateur dans une modalité dictée par des contextes d'usages). C'est ce que Bruno Bachimont appelle "l'arbitraire de l'interprétation" en proposant de revoir le paradigme du document, définissant celui-ci en Unités Techniques de Manipulation (en fait l'information de segmentation) et en Unité Sémiotiques d'Interprétation (en fait, les spécifications d'interaction, traduisant les affordances attendues dans l'interface).

Les contributions faites avec les équipes travaillant au développement de démonstrateurs de services pour les terminaux mobiles ont consisté, dans un premier temps, à regarder du côté de la mise en contexte des gestes de commandes ou de saisies de données.

Le modèle ayant servi de base est sémiotique, en ce sens que nous avons un geste qui sera reconnu pour sa forme à partir non seulement de la mesure de la géométrie du tracé, mais aussi de la mesure de la vitesse, des accélérations, des points d'inflexions, etc. Cette forme est ensuite interprétée par un superviseur, qui, en fonction de la nature du contenu affiché, va exécuter la commande appropriée. Ainsi, un même geste, comme par exemple, le tracé d'un cercle, va conduire à exécuter un zoom sur une image numérique, augmenter ou diminuer le volume d'un fichier audio en cours de lecture.

L'implémentation de ce modèle est faite dans un formalisme RDF (Ressource Description Framework d'XML) qui est vu ici comme l'inscription technique d'une triade sémiotique.

La syntaxe RDF permet d'inscrire des relations ternaires (sujet-objet-prédicat) dans des fichiers respectant la syntaxe XML, la structure logique d'une expression RDF est :

*"Subject IS A Object WHERE predicat"*

Cette syntaxe a pour objectif initial de modifier la présentation d'une URI en fonction du terminal. Dans notre approche, le sujet d'une expression RDF est un descripteur de forme dont la valeur d'interprétation est conditionnée par un prédicat représentant le contexte du système.

Dans l'implémentation qui a été faite dans notre équipe à Orange, le sujet est représenté par la forme d'un tracé fait par l'utilisateur. Ce tracé est reconnu par le système, et l'objet est une commande à exécuter. Pour le moment, le prédicat se limite, au type MIME du document.

A titre d'exemple, nous citerons la reconnaissance d'un cercle tracé par l'utilisateur qui est reconnu en tant que commande de volume si le document chargé est un document sonore, de fonction de zoom si le document chargé est une image ou un cercle graphique si le document chargé est infographique.

Cette approche est d'une implémentation relativement simple, due en particulier, à la disponibilité des parsers RDF.

## V. Discussion sur ces 3 cas d'applications

Au travers de ces trois cas d'applications, on voit transparaître un retour aux fondamentaux dans les définitions des concepts de forme et d'information.

### V.1. - Un peu d'étymologie des mots forme et document

La définition d'une forme que nous avons utilisée dans nos trois exemples d'applications, est tout à fait conforme à son sens étymologique.

Le mot forme vient du latin *forma*, qui signifie précisément tout ce qui est produit par la mise en forme d'un objet réel ou abstrait. Accessoirement, *forma*, en latin, signifiait également, la beauté.

Une forme est aussi une manifestation perceptible et tangible d'une idée résultant de la fabrication, autrement dit, la mise en œuvre d'une technique de mise en forme, avec ses règles et ses outils ; que cette idée porte sur quelque chose d'abstrait, comme un témoignage, un concept, ou que cette idée porte sur quelque chose de concret, c'est-à-dire, tout objet (un livre, une statue, un tableau, etc.).

Le processus qui conduit à une forme est un processus d'in-formation<sup>9</sup>.

Dans ce concept de forme, le document est lui-même une forme. En effet, il est porteur de traces d'informations résultant de la mise en œuvre d'une technique de fabrication.

Mais un document est aussi un support de communication si l'on se réfère à son étymologie. En effet, le mot document provient du verbe "*doceo, docere*" qui signifie à la fois "enseigner" et "apprendre" et qui a donné le mot *documentum*, signifiant une inscription (sur un support), dans le but de communiquer de l'information à des fins d'apprentissage ou d'enseignement.

Ainsi, tout au long de l'histoire des langues, les concepts de forme, document et communication sont étroitement liés.

Qu'en est-il dans notre vision contemporaine et numérique ?

**Les moyens de mise en forme ou "d'in-formation" se sont à la fois démultipliés et banalisés, mais la finalité du document reste la même : communiquer de l'information de manière conforme. Cette conformité ne s'obtient que si le mode de présentation de l'information est acceptable pour le destinataire. Dans sa forme numérique, cela signifie la nécessité d'une intermédiation entre la source et le destinataire pour adapter la mise en forme du document.**

---

<sup>9</sup> Là encore, nous restons proches de l'étymologie du mot forme :

- Pour les objets concrets : la mise en forme d'un objet s'appelait un "enformage" dont dérive précisément le mot information. Par exemple, la mise en forme d'un chapeau par moulage s'appelle, chez les chapeliers, un "enformage" ; par ailleurs, tout gastronome devrait se souvenir que le mot "fromage" est une déformation du mot "formage".
- Pour les objets abstraits : l'information, dans son sens premier, avait une forte connotation juridique, car elle signifiait au Moyen-âge ce qu'aujourd'hui nous appelons l'instruction d'une affaire, c'est à dire, la constitution d'un dossier d'accusation à l'aide de témoignages constituant des présomptions de preuves tangibles.

## **V.2. - Les formes et les documents**

Quel que soit le support de restitution, une forme est inscrite dans un document qui est lui-même une forme. Cette inscription résulte toujours de la mise en œuvre d'une technique de fabrication, que cette dernière soit le fait d'une mesure, d'une quantification et d'un codage (les trois étapes d'une numérisation), ou d'un travail de composition.

Ainsi, toute forme présente de l'information, à partir du moment où elle peut être interprétée par un lecteur, ici, le destinataire de l'information.

La forme correspond à ce que Bachimont appelle une "unité technique de manipulation" qui est le résultat d'un processus technique mis en œuvre par la source, mais aussi, qui doit être adressable par une technique (celle du destinataire).

Cette forme est interprétée par le destinataire pour en construire son sens. Or, ce mécanisme d'interprétation ne dépend plus des propriétés de la source, mais de celle du destinataire.

Dans le cas d'un document analogique, le lecteur/destinataire doit obligatoirement adhérer aux propriétés de la source pour s'approprier le sens du contenu, c'est-à-dire, interpréter les formes inscrites dans le document.

Dans le cas d'un document numérique, nous avons un glissement de la médiation entre la source et de destinataire (Pédauque), qui se traduit par le fait que le destinataire reconstruit les formes du document à partir du codage de l'information faite par la source : un fichier HTML ou Word ne devient un document que lorsqu'il est décodé par un logiciel.

Aussi, c'est bien l'adéquation entre le récepteur du destinataire et le code décrivant le document reçu qui permet de reconstruire ce dernier.

Pour la numérisation, le calcul d'une présentation de l'information se place dans le même contexte. Ici, le destinataire est le système d'information qui utilisera les données produites par la chaîne de numérisation.

Ainsi, que ce soit pour l'adaptation et la restitution d'un document numérique dans une interface ou pour la validation d'un processus de reconnaissance de formes, la validité d'un descripteur de forme et sa valeur seront le fait de la validité du moyen utilisé pour produire cette forme.

## **V.3. - Les signes d'une forme**

Une logique ternaire apparaît systématiquement dans nos développements. Elle définit la présentation de l'information, que ce soit dans le cas de la numérisation ou celui de l'interaction homme machine.

Cette logique ternaire correspond en tous points à une instance de la triade sémiotique de Peirce, avec les éléments indissociables entre eux, d'objet, de signes et d'interprétant.

La description des trois problèmes de la communication de Weaver peut facilement s'apparier avec la triade sémiotique :

- le signe est la trace du problème technique,
- l'objet est la trace du problème sémantique,

- l'interprétant est la trace du problème d'effectivité.

Vu sous cet aspect, la reconnaissance de forme revient à trouver des familles de solutions à chacun des problèmes de construction d'une présentation de l'information :

- Le problème technique traite de la segmentation en formes et de leur inscription dans un formalisme syntaxique évitant l'ambiguïté de présentation.

- Le problème sémantique traite de la constitution d'une base de classification, avec les questions d'acquisition et de maintenance d'une vérité terrain représentative des données capables d'être traitées par une application informatique.

- Le problème d'effectivité traite de la relation entre une forme et sa valeur. Ainsi ce problème est propre à l'interprétation. La question que l'on peut se poser est sur le formalisme syntaxique décrivant cet interprétant ou cette relation entre une forme et sa valeur.

Nous pouvons établir un parallèle avec un modèle d'interprétation humain :

- L'équivalent du problème technique est, chez l'humain, la perception et la capacité de discriminer les formes par rapport au bruit, mais aussi, par rapport aux formes entre elles (ambiguïté de présentation).

- L'équivalent du problème sémantique correspond aux savoir et aux connaissances de l'humain intervenant dans sa capacité de nommer les formes.

-- L'équivalent du problème d'effectivité est la sémiose, à savoir, le processus de reconstruction mentale d'un objet à partir des signes perçus.

A partir de là, nous pouvons déduire une forme de symétrie entre l'interprétation d'une forme par le calcul dans la machine ou par une sémiose faite par l'utilisateur. On prendra pour exemple un terminal à l'interface tactile dans lequel vient s'afficher une image d'un document numérisé, un document électronique dont la feuille de style est inconnue, ou des graphismes sans relations topologiques.

Dans tous ces exemples, l'interaction/régulation du dialogue homme machine, pose problème en l'absence de présentation d'informations de feedback et de feedforward, notamment, le changement d'état d'un fragment sélectionné dans un document, n'est pas possible à faire parce que l'objet informatique et sa présentation n'existent pas pour la machine.

Le problème sémantique abordé ici est un peu particulier car, chez l'humain en situation de communication (avec un autre humain ou une machine), la sémantique est représentée par le comportement.

Ce comportement se mesure au travers de l'activité, en particulier par les traces mesurables de cette activité. Ces traces ont une valeur symbolique, au même titre que l'ensemble des valeurs possibles dans une base de classification.

Par conséquent, dans une interface homme machine, la communication se fait toujours au travers des formes calculées par la machine et ces formes correspondent soit à de l'information issue d'un système d'information, soit de la mesure et de l'interprétation, donc un calcul à partir de données, des traces d'activités de l'humain.

Cette approche nous donne une symétrie d'interface. Nous avons vu que dans ce modèle d'interaction (chapitre 3), cette symétrie conduit à une notion d'interface agissant comme un régulateur du dialogue homme machine, dans lequel le contrôle offert à l'utilisateur se fait au travers de formes restituées par l'interface, la régulation se faisant au travers des informations de feedback et de feedforward sur l'état du système d'information et l'activité de l'humain.

On prendra pour exemple un terminal à l'interface tactile dans lequel vient s'afficher une image d'un document numérisé, un document électronique dont la feuille de style est inconnue, ou des graphismes sans relations topologiques.

Dans tous ces exemples, l'interaction/régulation du dialogue homme machine, pose problème en l'absence de présentation d'informations de feedback et de feedforward, notamment, le changement d'état d'un fragment sélectionné dans un document, n'est pas possible à restituer si l'objet informatique et sa présentation n'existent pas pour la machine.

#### **V.4. - L'alignement de modalités**

Ici le mécanisme que nous avons mis en œuvre se base sur la reconnaissance d'une valeur d'une forme présentée selon deux modalités jointes. Par exemple, dans l'OCR intuitif :

- la forme d'un caractère obtenue par segmentation d'une image
- la forme d'un caractère obtenue par un tracé de l'utilisateur

C'est l'alignement de ces modalités jointes qui sert de mesure de l'évaluation d'une reconnaissance de caractères en validant ou invalidant les solutions proposées aux trois problèmes de Weaver.

- au problème technique : la saisie d'un tracé remet en question l'information de segmentation et déclenche une mesure d'isomorphisme entre une composante connexe et un tracé de l'utilisateur,
- au problème d'effectivité : la saisie d'un tracé invalide la reconnaissance d'un caractère et l'utilisateur valide la valeur obtenue par la reconnaissance de son tracé, valeur venant se substituer à celle obtenue par l'OCR.
- au problème sémantique : en fonction des propriétés des résultats obtenus dans la résolution du problème technique et du problème d'effectivité, le système vient enrichir la base de classification de l'OCR.

Ainsi la seule mesure de l'évaluation de l'OCR intuitif réside dans la mesure de l'activité de l'utilisateur en cours de dialogue avec la machine.

Afin de valider et de généraliser cette approche de co-articulation de modalités de présentation de l'information, un protocole d'expérimentation, avec sa métrologie associée reste à établir en généralisant le tracé manuscrit comme une trace de l'activité de l'utilisateur, afin d'affiner la stratégie d'apprentissage pour la reconnaissance de n'importe quelle forme visuelle.

L'un des avantages que l'on peut escompter de cette généralisation sera la possibilité de faire abstraction de la valeur symbolique d'une forme : l'apprentissage consistera à demander à un utilisateur de schématiser ce qu'il voit à l'écran. Ce schéma de la

forme servira de valeur de référence à la reconnaissance de cette dernière dans l'image.

La détermination de la valeur de cette forme se fera, non pas à partir de l'image, mais à partir du schéma fait par l'utilisateur. Ce schéma revient à réduire fortement l'espace de représentation d'objets, ce qui devrait contribuer à abaisser l'ambiguïté de formes ressemblantes mais de sens différents.

Ainsi, on devrait obtenir à terme une indexation des contenus à deux niveaux :

- une indexation sémantique à partir des métadonnées affectées aux schémas de formes produits par l'activité de l'utilisateur
- une indexation sémiotique à partir du regroupement de formes ressemblantes dans des classes étiquetées par les schémas de présentation.

### V.5. - La présentation de l'information

La présentation de l'information correspond à l'inscription d'un ensemble de formes et de leurs descriptions, dans un document dans le but de représenter un objet ou ensemble d'objets abstraits ou concrets.

Cette définition est conforme au modèle d'Ackoff (Fig. 32) et applicable au canal de communication de Shannon Weaver (Fig. 33), ainsi qu'à la théorie mathématique de la communication sous-tendue dans ce canal.

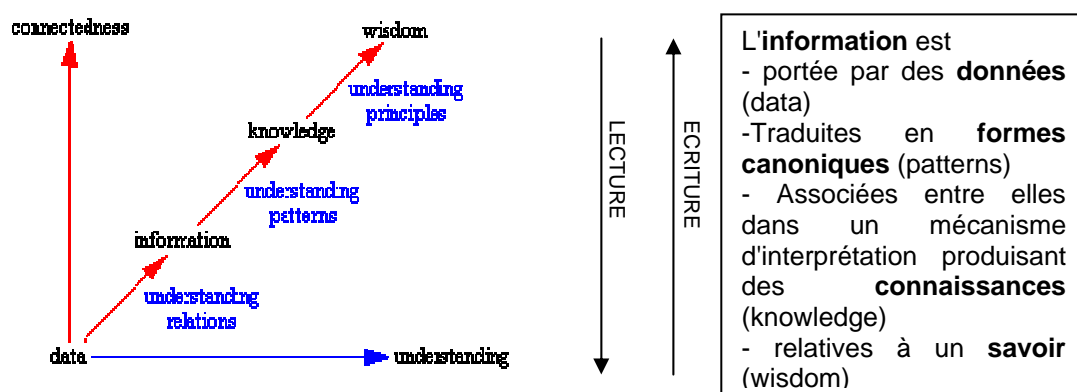
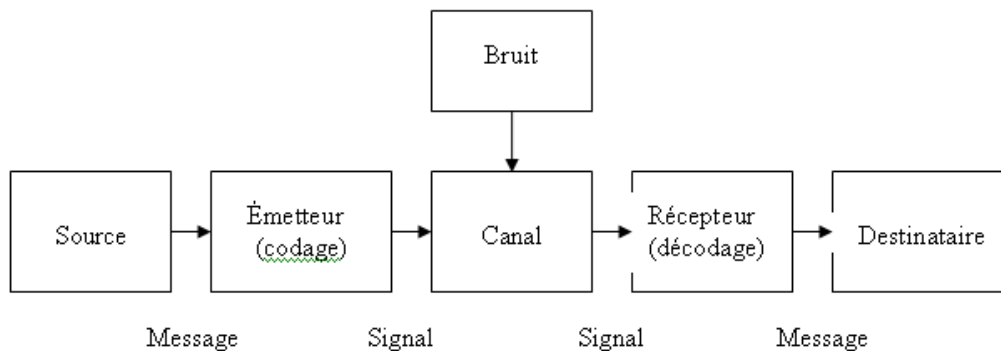


Fig. 32 : Le modèle d'interprétation d'Ackoff



- 1) La source d'information énonce un message ...
- 2) ... que l'émetteur va encoder et transformer en signal,
- 3) lequel va être acheminé par le canal,
- 4) puis décodé par le récepteur, qui reconstitue un message à partir du signal
- 5) et le transmet enfin au destinataire.

Fig. 33 : Le canal de communication de Shannon et Weaver

On peut critiquer l'approche technique et réductrice de ces définitions par rapport à ce qu'est une information dans le domaine des sciences humaines (Jakobson, Riley, etc.), mais ces définitions s'avèrent nettement plus appropriées à une interaction homme machine du fait que tous les systèmes de codage informatiques se basent encore sur le modèle de Shannon.

En outre, l'avantage de cette revisite des fondamentaux de l'information permet de construire une relation directe avec le modèle sémiotique de Peirce, basé sur une triade (objet-signe-interprétant) définissant le phénomène information. Or les seuls éléments intervenants dans une communication homme machine, voire machine machine sont, précisément les signes codés en signaux venant perturber un canal de communication.

Ces signes sont à la fois les messages du canal de Shannon et la présentation de l'information dans l'approche inspirée d'Ackoff et celle développée dans cette thèse. Ainsi, le traitement sémiotique de l'information dans une interface, revient à étudier les mécanismes de sémiologie tant par l'utilisateur dans sa perception et la compréhension des messages émis par la machine, que par la machine elle-même, dans sa reconnaissance des gestes et actions de l'utilisateur. Ici la notion d'interprétant est le programme logiciel avec ses spécifications de données d'entrée/sortie.

Si les approches proposées font références à des définitions très classiques du traitement de l'information, leur utilisation en est, par contre, innovante :

- Dans cette optique, la reconnaissance de formes revient à effectuer un transcodage de l'information d'une modalité de présentation vers une autre.

Cette notion de transcodage de la présentation de l'information ne se cantonne pas au traitement d'un signal en limitant les connaissances sur la source d'information à des connaissances a priori ou des heuristiques : une reconnaissance de forme devient un acte négocié, une intermédiation entre un émetteur avec sa technique de production de contenus et un récepteur avec ses capacités d'interprétation.



- En limitant le concept d'information au périmètre strict des interfaces comme étant les traces d'activités associées d'un utilisateur et d'une machine (en fait, un calcul), on tend vers une approche constructiviste de l'interaction.

L'objectif est d'évoluer vers des interfaces adaptatives dont l'apprentissage est conditionné par une éthologie de l'action, c'est-à-dire l'observation de comportements, celui de la machine et celui de l'humain conduisant à mesurer leur activité.

- Ceci nous rapproche des systèmes dynamiques dans lesquels un processus est mesuré par son état initial, son état final et des états de transitions représentatifs.

Ce sont ces seuls états qui sont présentés à l'utilisateur dans une interface. Dans le cas notre l'OCR intuitif, l'état initial est la zone d'intérêt et l'état final est la valeur du mot à reconnaître. L'état de transition représentatif du processus sur lequel agit l'utilisateur, est la segmentation en composante connexes.

- Enfin, ce recadrage du traitement de l'information par une vision nettement plus orientée système que technologique, conduit, à performances d'utilisation et finalités égales, à une grande simplification algorithmique par rapport aux approches classiques.

Les outils mis en œuvre sont classiques, voire simplistes, mais c'est leur périmètre d'utilisation qui est mieux cerné du fait qu'ils n'agissent plus en "boite noire". L'interface joue le rôle d'un régulateur que l'utilisateur pilote au travers des présentations d'informations de feedforward et de feedback.

L'étude et la compréhension des causes de cette simplification algorithmique nécessitent les contributions de plusieurs disciplines notamment, l'ergonomie, la systémique et les sciences de la complexité. Les applications que nous avons proposées illustrent un cadre théorique pluridisciplinaire de recherche sur les corpus numériques en tenant compte de leurs propriétés actuelles et réelles : la masse, l'hétérogénéité et le polymorphisme d'usages.



## **Conclusions et perspectives**

*"L'évolution procède comme un bricoleur qui pendant des millions et des millions d'années, remanierait lentement son œuvre, la retouchant sans cesse, coupant ici, allongeant là, saisissant toutes les occasions d'ajuster, de transformer, de créer."*

**François Jacob** - "Le jeu des possibles"

Au travers des différents projets de R&D que nous avons construit ou piloté depuis 1996, est apparu un problème commun à toutes les solutions techniques de numérisation que nous avons éprouvées : l'évaluation technique d'une méthode de reconnaissance de formes trop souvent dissociée de l'utilisabilité de son résultat dans une interface.

Une reconnaissance de formes produit des données qui permettent à un programme logiciel de faire des traitements. L'un d'entre eux, consiste en le calcul d'une présentation de l'information qui doit permettre à l'utilisateur de retrouver le sens de cette information. Nous nous trouvons donc confronté à l'arbitraire de l'interprétation d'une forme associée à son évaluation technique au travers de données.

De plus, l'inscription des données reconnues fait perdre toute notion de la complexité des calculs qui a conduit à les produire.

**Calculer une présentation de l'information et la restituer via une interface représente l'interprétant du sens du contenu et non ce sens lui-même.**

Nous avons abordé le problème de la présentation de l'information sous l'angle de la communication homme machine, en reprenant la théorie mathématique de la communication de Shannon et Weaver. Plus particulièrement nous avons repris les trois problèmes de la communication posés par Weaver pour en proposer une formulation applicable à la présentation de l'information :

- **Le problème technique**, qui est ici celui du codage de la description des formes dans un document numérique.
- **Le problème sémantique**, qui est ici celui des valeurs d'usages possibles affectées à des formes selon l'ontologie, le vocabulaire et la symbologie construits à priori dans un système d'information,
- **Le problème d'effectivité**, qui assure les conditions de relation entre la résolution du problème technique et la résolution du problème sémantique, à savoir, la relation entre une forme et sa valeur obtenue par classification ou interprétation.

La solution de ces trois problèmes détermine un point de vue de l'information conforme aux moyens, attentes et objectifs du destinataire.

Les propriétés du destinataire et de son récepteur, peuvent être antinomiques à celles d'une source émettant de l'information. Le cas des bibliothèques numériques est presque caricatural de cet état de fait : les contraintes, objectifs et besoins d'un moine copiste du XIVème siècle n'étaient certainement pas les mêmes que celle d'un lecteur contemporain.

Sans aller jusqu'à cette situation, nous nous rendons compte quotidiennement que ces questions d'accès et de présentation de l'information nécessitent de généraliser, à défaut d'unifier, des concepts connus de nos activités de recherche en les abordant dans une vision résolument pluridisciplinaire :

- Le concept d'affordance des ergonomes doit être complété par celui d'affordance computationnelle que nous avons introduit dans cette thèse.
- La mesure de la complexité d'un calcul est un paramètre important intervenant dans l'évaluation d'un processus de reconnaissance de formes.

- Enfin, le document numérique et par extension, les corpus numériques a acquis une dimension multimodale de présentation de l'information que nous devons prendre en considération. C'est précisément cette dimension multimodale qui confère à un corpus numérique et aux outils permettant d'en extraire de l'information, les propriétés d'un système complexe.

### *Symétrie entre affordances humaine et computationnelle*

En définissant cette symétrie d'affordances, on obtient une interface incluant un mécanisme constructiviste de présentation de l'information, en ce sens qu'il résout au cas par cas un problème d'effectivité dont la solution peut être mise en mémoire du système en fonction de contextes.

La description d'une forme donne à celle-ci ses propriétés d'affordance. L'affordance correspond à la présentation d'une information qui "suggère" à l'utilisateur l'interaction possible avec cette forme.

Une forme est l'équivalent des signes représentant un objet dans une triade sémiotique. Elle est restituée par une interface au travers d'un calcul fait à partir de données informatiques. Nous avons là une première symétrie entre un objet au sens sémiotique du terme et un objet au sens informatique du terme.

L'affordance d'une forme correspond à ses propriétés descriptives ou perceptives qui permettent à l'utilisateur d'en interpréter le sens (que ce soit pour "déchiffrer" un symbole ou pour lancer l'exécution d'une commande). Cette affordance est d'autant plus évidente que l'interprétation de la forme par sémiose, est facile pour l'utilisateur.

Cette sémiose peut être vue comme le symétrique d'une complexité de calcul côté machine : une même forme est, soit calculée par un programme logiciel et interprétée par un humain, soit produite par une action d'un utilisateur puis analysée et interprétée par la machine.

Ici, la complexité de calcul pour la machine devient le symétrique de la sémiose pour l'utilisateur. Cette symétrie permet de concevoir des interfaces dans lesquelles, une complexité de calcul peut être substituée par une action de l'utilisateur. Ainsi, un même résultat ou état final du système est obtenu par au moins deux méthodes : un calcul ou une action de l'utilisateur. Le choix de la méthode la moins complexe correspond bien à la notion de profondeur logique du système décrite par Bennett. Cette profondeur est à rapprocher du principe d'économie ou de faible ergodicité des interfaces selon la terminologie des ergonomes

Les exemples d'applications, l'OCR intuitif et AdHoc, que nous avons développés dans le cadre de cette thèse montrent cette symétrie d'affordance.

Dans les deux cas, l'action d'un utilisateur vient en substitution d'une hypothèse de calcul conduisant à un même résultat. Dans AdHoc, il s'agit de donner les propriétés terminologiques d'une expression calculée par un abrégéur/extracteur de mots clés et dans l'OCR intuitif, il s'agit d'une méthode de correction/validation agissant à la fois sur la segmentation et la classification.

### *La complexité comme indicateur de mesure*

Dans le chapitre 4, nous avons abordé les systèmes complexes dans le but d'en extraire certaines propriétés utiles à l'élaboration d'un modèle générique de l'interaction homme machine.

Nous avons montré, dans le chapitre 5, que la complexité pouvait nous fournir un moyen de calcul de distance entre informations, applicable à n'importe quel type de contenus (images, textes, graphismes).

**L'évaluation de la reconnaissance de formes incluant la mesure de la complexité d'un calcul effectué sur une forme est un axe de recherche qu'il paraît important d'explorer.**

Les propositions de méthodes proposées actuellement reposent toutes sur la validité d'un résultat de reconnaissance de formes, c'est-à-dire, sur la validité de la valeur produite par un calcul. Le calcul lui-même n'entre pas en ligne de compte comme paramètre de cette évaluation.

Or, un problème peut avoir un énoncé et une solution simple mais nécessiter un calcul très complexe pour sa résolution. Calculer la nième décimale d'un nombre retournera un entier (énoncé et solution simples), mais selon la donnée à traiter, le calcul sera plus ou moins complexe.

Il apparaît clairement une relation entre la complexité de la donnée en entrée et la complexité d'un calcul. En reconnaissance de formes, cette complexité de la donnée initiale correspond à la complexité de la présentation du document ou du fragment de document à reconnaître.

**Savoir établir des seuils de complexité pour tout calcul d'une présentation d'information dans une interface est la traduction technologique de l'ergodicité des interfaces.**

Selon la nature du document et la modalité de présentation de l'information, ces opérations de reconnaissance ont une certaine complexité en temps et en espace, dont la mesure devrait intervenir dans l'évaluation d'une méthode de reconnaissance de formes en termes de complexité acceptable par rapport à l'usage d'un résultat de reconnaissance de formes.

Cette notion de complexité acceptable prend une importance particulière pour les méthodes de reconnaissance de formes à la volée, c'est-à-dire, incluses dans une IHM : cette mesure devrait, à terme, permettre de quantifier un seuil au-delà duquel, on pourra substituer un calcul par une action de l'utilisateur.

Cette action est une information manquante, apportée au système selon une modalité préférentielle de l'utilisateur.

Dans l'application d'OCR intuitif que nous avons proposée au chapitre 6, la décision d'enrichir la base de classification après reconnaissance du tracé fait par l'utilisateur pour corriger l'OCR, repose sur une heuristique relative à l'unicité de la composante connexe mal reconnue par l'OCR.

Une évolution possible consistera à évaluer la complexité de calcul manquante pour le regroupement ou la coupure de composantes connexes comme seuil de décision d'un apprentissage.

On notera, enfin, que dans notre approche de la reconnaissance de formes à la volée, la validation d'un résultat se fait selon des critères d'acceptabilité par l'utilisateur de la valeur retournée par le processus de reconnaissance.

*Humaniser l'interaction homme machine*

Définir la place de l'utilisateur dans tout processus de traitement de l'information faisant intervenir des documents numériques est un vœu humaniste et coutumier que l'on trouve fréquemment formulé dans toutes les disciplines associées au document numérique.

Exaucer ce vœu nécessite de reprendre certaines variables du document numérique. En particulier celles qui ont été d'emblée écartées ces dernières décennies dans la conception de système de numérisation et de traitement de documents numériques. L'objectif était de concevoir des systèmes d'informations selon une approche essentiellement technologique et réductionniste en demandant à l'utilisateur de s'adapter aux contraintes techniques du système.

Désormais on devra prendre en compte : la complexité d'un système et la multimodalité d'interaction avec ce dernier.

Dans l'écosystème actuel de l'information numérique, l'homme n'est plus un lecteur passif accédant à un nombre limité de sources d'informations. Il devient un élément actif immergé dans une forme d'intelligence ambiante. C'est un nœud actif d'un réseau de communication planétaire. Dans ce réseau, circule de l'information dont la présentation atteint un degré de complexité tel que la standardisation des contenus devient inopérante.

Que signifie humaniser l'interaction homme machine dans un tel contexte ?

Il s'agit de mettre à disposition des moyens techniques de lecture et d'écriture permettant à tout utilisateur destinataire d'informations, de communiquer avec un système complexe, selon ses modalités préférentielles et non plus selon les contraintes techniques dictées par la machine source d'informations.

Ces nouveaux moyens de lecture et d'écriture (incluant l'indexation et la recherche d'information), résolument adaptés à n'importe quel utilisateur, sont les outils permettant d'appréhender n'importe quel corpus numérique, quels que soient sa taille et ses contenus. Le moyen de cette appropriation des contenus sera toujours un document numérique et c'est la présentation de l'information portée par ce document qui conditionne l'interaction.

Ces mutations ont été l'une des motivations principales de cette thèse pour tenter de traduire techniquement quelle était cette nouvelle place de l'homme dans ce contexte.

Les réponses que nous proposons ne sont que partielles car le nombre de questions posées dans cette étude est immense, tout comme la tâche qui attend les chercheurs dans les prochaines années.

Cependant, les enjeux liés à cette tâche sont particulièrement importants, car ils concernent la pérennité de la conservation et de la communication de l'information.



Fig. 34 : "le dernier banquet des rats" de Gustave Doré.

*Vu sous un angle allégorique, les rats représentent l'hétérogénéité des moyens d'accès à l'information (le sac de grains,) non fédérés par la maîtrise de la complexité de la présentation de l'information. Le chat, quant à lui, représente cette complexité et se serait échappé de sa boîte du laboratoire d'Erwin Schrödinger*

Au stade actuel de la réflexion, nous pouvons déduire deux axes de travail important à explorer.

#### 1) Document numérique et sciences de la complexité

Le premier axe de travail que nous proposons est tiré de notre expérience de la numérisation au travers de plusieurs projets passés. Il s'agit de prendre en compte la complexité comme une donnée du problème, et non plus comme un obstacle à contourner en simplifiant ce problème par un modèle réductionniste.

La complexité est un verrou quand on cherche à résoudre l'impossibilité de transposer une chaîne de numérisation classique à différents corpus de documents en redéfinissant isolément ses composants internes, car cette méthode ne rend pas compte de l'intrication entre composants.

Elle devient un atout si l'on adopte une démarche systémique en observant et mesurant le comportement dynamique de ses composants afin d'exploiter au mieux les propriétés d'émergence propres aux systèmes complexes. Ces propriétés d'émergence ne sont pas quantifiables a priori, par conséquent, un système d'information ayant les propriétés d'un système complexe, évoluera en permanence ; cette évolution sera guidée par l'information contenue dans la mémoire de ce système et les interactions faites par les utilisateurs sur la présentation de cette information. C'est en soit, une évolution du paradigme du web 2.0 car elle inclut les concepts de forme, d'affordance et d'outils d'accès à l'information.



Cette proposition suit de près l'évolution des théories de l'information, qui, en devenant filles de la physique contemporaine, nous donne les outils permettant d'appréhender et de quantifier la complexité.

## 2) la cybernétique de l'interaction avec un document numérique

Le second axe de travail que nous proposons est un parallèle à établir entre la capacité de produire une information à la présentation multimodale et notre capacité de la percevoir pour l'analyser et l'interpréter.

La présentation de l'information est le stimulus qui conditionne toute l'interaction homme machine. Si dans le cas d'un texte, la construction d'une présentation de l'information reste dans le cadre de la théorie mathématique de la communication de Shannon, le passage à une information multimédia et multimodale nécessite que l'on revisite plusieurs concepts fondamentaux de la communication.

L'approche que nous proposons consiste à prendre en compte un modèle sémiotique de présentation de l'information qui semble être le seul capable à établir ce parallèle. La triade sémiotique permet d'établir des relations entre objet et information, signe et message ; elle induit la notion d'interprétant, correspondant aux mécanismes de reconstruction d'un sens.

Cet interprétant est, pour l'humain, la sémiologie conduisant à la reconstruction mentale d'un objet à partir des signes le représentant. Pour la machine, cet interprétant est précisément un processus de reconnaissance de formes, si l'on accepte de généraliser le concept de formes à tout fragment d'un document présentant de l'information. Dans les deux cas, l'interprétant se quantifie par sa complexité. C'est ce qui nous a conduits à associer à la notion d'affordance humaine, la notion d'affordance computationnelle.

Pour l'humain, la présentation de l'information par ses formes et leur structure conditionne la lisibilité du document et donc, la facilité d'en interpréter le contenu.

Pour la machine, la segmentation et la classification seront d'autant plus complexes à mettre en œuvre que la saillance "technique" des formes est faible. Dans les deux cas, la complexité d'un calcul est le symétrique de la complexité d'une interprétation.

La solution que nous proposons est encore de l'ordre de la systémique, mais cette fois, vue sous l'angle des systèmes dynamiques. Il s'agira de mieux appréhender et quantifier les propriétés dynamiques d'un calcul pour la machine et d'une sémiologie pour l'homme.

**Cette recherche conduit à l'établissement d'un dialogue homme machine multimodal et redonne au document sa première raison d'être : un objet de communication porteur de traces persistantes qui sont le seul moyen d'accès à l'information contenue dans un corpus numérique.**

**Si les formes présentent l'information, la complexité de calcul de ces formes pourrait bien représenter la capacité de les interpréter, autrement dit, représenter la connaissance.**

## Références bibliographiques

- [ADA00] **S. Adam, R. Mullot, J.M. Ogier, C. Cariou, J. Gardes**, "Interprétation de documents du réseau téléphonique : approche multi-spécialistes." Congrès Francophone RFIA'2000, Vol 3, pp 357-364, 2000
- [ADR07] **P. Adriaans** : Learning as data compression, in "Computation and Logic in the real world", LNCS Vol 4497/2007, pp.11-24
- [BAC04] **B. Bachimont**, " Arts et Sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle" Mémoire d'Habilitation à Diriger des Recherches de l'Université de Technologie de Compiègne, 2004
- [BAC96] **B. Bachimont**, " Herméneutique matérielle et Artéfacture : des machines qui pensent aux machines qui donnent à penser ; Critique du formalisme en intelligence artificielle." Thèse d'épistémologie de l'Ecole Polytechnique, 1996
- [BER02] **D. Berthier**, "Le savoir et l'ordinateur", p.148-171, l'Harmattan Ed. ISBN 2747533506, 9782747533508, 2002
- [BNU06] **Synthèse et plan d'actions BNUE**, Secrétariat Général du Comité de pilotage- janvier 2006
- [CEB05] **Cebri'an, M., Alfonseca, M., Ortega, A.**: Common pitfalls using normalized compression distance: what to watch out for in a compressor. *Communications in Information and Systems* 5(4), 367-384 (2005)
- [CIL05a] **Rudi Cilibrasi, Leo van Iersel, Steven Kelk and John Tromp**, On the complexity of several haplotyping problems, *Algorithms in Bioinformatics: 5th International Workshop, WABI 2005, Mallorca, Spain, October 3-6, 2005*, LNCS 3692 / 2005, pp. 128-139
- [CIL05b] **R. Cilibrasi, P.M.B. Vitanyi**, Clustering by compression, *IEEE Trans. Information Theory*, 51:4(2005)
- [CLA04] **E. Clavier, G. Masini, M. Delalandre, M. Rigamonti, K. Tombre, J. Gardes**, "DocMining a Cooperative Platform for Heterogeneous Document Interpretation According to User-Defined Scenarios" in *Lecture Notes in Computer Science*, vol 3088; pp 13-24 (2004)
- [DUF07] **Duffner S., Garcia C.**, "Face Recognition using Non-Linear Image Reconstruction", *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2007)*, London, UK, Septembre 2007.
- [GAR04] **J. Gardes, D. Dionisi, E. Trupin, J. Labiche**, "Le document dans un réseau de communautés d'intérêts : un avatar ?"; *La semaine du document numérique*, La Rochelle 2004, Europa ISBN 2-909285-29-4; pp 69-92, 2004
- [GRE00a] **V. Grenier, R. Mullot, J.M. Ogier, S. Adam, J. Gardes, Y. Lecourtier**, "Une architecture distribuée pour l'interprétation des documents techniques" Congrès Francophone RFIA'2000, Vol 1, pp 427-436, 2000

- [GRE00b] **V. Grenier, R. Mullet, J.M. Ogier, S. Adam, J. Gardes, Y. Lecourtier**, "Distribution d'opérateurs pour l'analyse de documents techniques" Congrès International Francophone, CIFED 2000; Lyon France, pp 151-160, 2000
- [HES07] **B. Hescott, D. Koulomzin**, On Clustering Images Using Compression, Technical Report, CS Department, Boston University, 2007
- [LEF07] **Lefebvre G., Garcia C.**, "Facial Biometry by Stimulating Salient Singularity Masks", *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2007)*, London, UK, September 2007.
- [MIC08] **R. Harper, T. Rodden, Y. Rogers, A. Sellen** "Being Human: Human-Computer Interaction in the year 2020" – Microsoft research, ISBN: 978-0-9554761-1-2, 2008
- [OGI00] **J.M. Ogier**, "De l'Image au Document Technique : Problème d'interprétation", Habilitation à diriger des recherches de l'Université de Rouen
- [OGI94] **J.M. Ogier**, "Contributions à l'analyse automatique de documents cartographiques : Interprétation de données cadastrales" Thèse de doctorat de l'Université de ROUEN – 1994
- [PED06] **Roger T. Pédaque**<sup>10</sup>, *Le document à la lumière du numérique: Forme, texte, medium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité*, C & F Ed, ISBN 2915825041, 9782915825046, 2006
- [PED07] **Roger T. Pédaque**, "La redocumentarisation du monde", Cépaduès éd., ISBN 2854287282, 9782854287288, 2007
- [SAI01] **Y.Saïdali, E. Trupin, J. Gardes, J. Labiche** "Dialogue expert-machine naïf : expérimentation sur une plate-forme de traitement d'images" Conférence Ingénierie des connaissances 2001, IC2001, Grenoble France, pp 35-35, 2001
- [SAY73] **K. M. Sayre**, "Machine recognition of handwritten words" in *Pattern Recognition J.*, vol. 5, no. 3, pp. 213-228, Sept. 1973
- [SHA48] **C. E. Shannon**, "A mathematical theory of communication", *The Bell System Technical Journal*, Vol. 27, pp. 379-423, 623-656, July, October, 1948.
- [TAU02] **D. Taubman, M. Marcellin**, "JPEG2000 Image Compression Fundamentals, Standards and Practice", Series : The International Series in Engineering and Computer Science, 2002, 800 p., ISBN 079237519X
- [TUR36] **A.M. Turing**, "On computable Numbers, with an application to the EntscheidungsProblem", *Proceedings of the London Mathematical Society*, 2:42, 230-265, 1936

---

<sup>10</sup> Roger T. Pédaque est un nom collectif d'un groupe de chercheurs ayant animé le réseau thématique pluridisciplinaire du CNRS sur le document numérique. Ce nom collectif regroupe : *Responsables de la synthèse* : Dominique Boullier, Jean Charlet, Dominique Cotte, Yves Jeanneret, Joël Gardes, Niels W. Lund, Jean-Michel Salaün, Monique Slodzian et Jean-Yves Vion-Dury.

Contributeurs :

Evelyne Broudoux, Marie-Anne Chabin, Ghislaine Chartron, Carol Chovsky, Olivier Ertzscheid, Raja Fenniche, Gabriel Gallezot, Brigitte Guyot, Julien Laflaquière, Thierry Lafouge, Hervé Le Crosnier, Olivier Le Deuff, Yannick Prié, Eric Thivant, Christian Vandendorpe, Manuel Zacklad.

- [TUR91] **M. Turk and A. Pentland**, "Face recognition using eigenfaces". *IEEE Conference on computer Vision and Pattern Recognition* pp. 586-591, 1991.
- [VIT07] **P.M.B. Vitanyi**, Analysis of Sorting Algorithms by Kolmogorov Complexity (A Survey). Pp.209--232 in: In: *Entropy, Search, Complexity*, Bolyai Society Mathematical Studies, 16, I. Csiszar, G.O.H. Katona, G. Tardos, Eds., Springer-Verlag, 2007.
- [ZOU06] **J. Zou, G. Nagy**, "Visible models for interactive pattern recognition", *Pattern Recognition Letters*, Vol.28, Issue 16, p. 2335-2342, Elsevier Ed. ISBN 0167-8655, 2006

## **ANNEXES**



## **ANNEXE 1 :**

### **Illustrations des limites de la numérisation de masse**

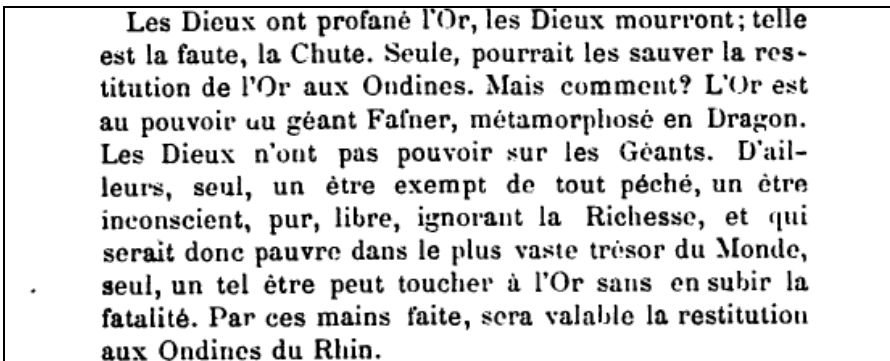
*Cette annexe a pour objectif d'illustrer les limites d'une numérisation de masse et leurs conséquences sur la présentation de l'information pour l'interaction et l'indexation des contenus.*

### *Extrait 1(BNF) : paragraphe de texte*

L'extrait de document numérisé que nous présentons ici, est tiré d'une édition du XIXème siècle de l'étude et la traduction française des livrets des 4 opéras formant la tétralogie de "l'anneau des Nibelungen" de Richard Wagner. L'objectif de cette analyse est de mettre en évidence

- les limites d'une numérisation industrielle centrée sur une forte standardisation de la présentation des documents,
- le "poids sémantique" que représente une erreur de reconnaissance de caractère, selon qu'elle porte sur un mot de liaison du texte ou un mot clé.

### **Image originale**

A rectangular box containing a scan of a printed page. The text is in French and discusses the gods' quest for the Golden Horn. The text is: "Les Dieux ont profané l'Or, les Dieux mourront; telle est la faute, la Chute. Seule, pourrait les sauver la restitution de l'Or aux Ondines. Mais comment? L'Or est au pouvoir au géant Fafner, métamorphosé en Dragon. Les Dieux n'ont pas pouvoir sur les Géants. D'ailleurs, seul, un être exempt de tout péché, un être inconscient, pur, libre, ignorant la Richesse, et qui serait donc pauvre dans le plus vaste trésor du Monde, seul, un tel être peut toucher à l'Or sans en subir la fatalité. Par ces mains faite, sera valable la restitution aux Ondines du Rhin." The text is centered and uses a serif font.

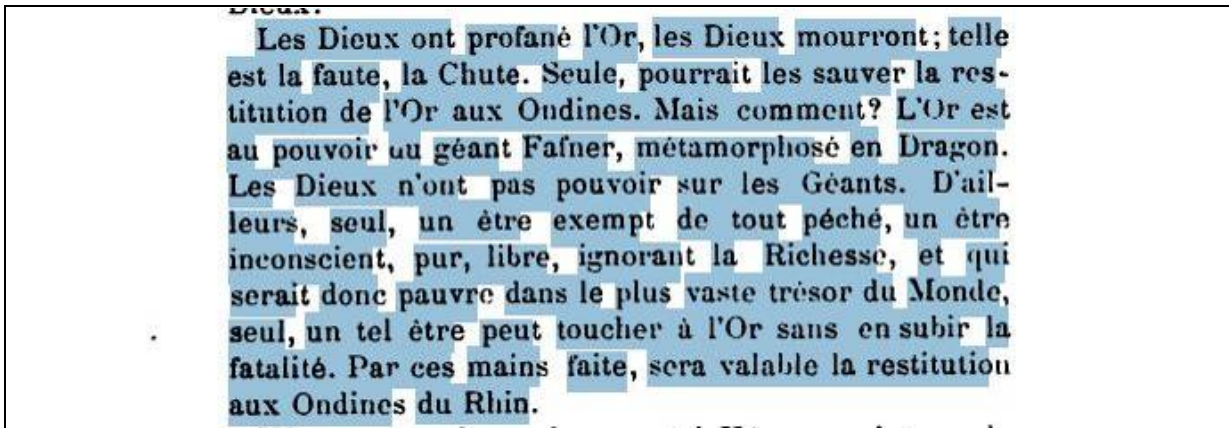
Commentaires :

Le document source est au format PDF et affiché via Acrobat Reader. Il résulte d'une numérisation en noir et blanc à la résolution de 300 dpi d'un support microfilmé (l'indication de la source est présente dans les métadonnées des fichiers XML transcrits en PDF).

Le texte a été composé en utilisant des polices de caractères et une mise en page modernes (1894). Il a fait l'objet d'une reconnaissance de caractère industrielle.

L'objectif de la manipulation a consisté à évaluer la qualité de la relation entre les formes de l'image et les données issues de l'OCR. Cette manipulation est aisée à reproduire, étant donné qu'il s'agit d'utiliser les deux instances possibles de la fonction copier-coller présente dans Acrobat Reader : extraire un fragment de l'image et extraire un fragment des données de texte.

### **Segmentation du texte (Présentation de l'information)**

A rectangular box containing the same text as the previous image, but with a blue highlight effect applied to the text. The text is: "Les Dieux ont profané l'Or, les Dieux mourront; telle est la faute, la Chute. Seule, pourrait les sauver la restitution de l'Or aux Ondines. Mais comment? L'Or est au pouvoir au géant Fafner, métamorphosé en Dragon. Les Dieux n'ont pas pouvoir sur les Géants. D'ailleurs, seul, un être exempt de tout péché, un être inconscient, pur, libre, ignorant la Richesse, et qui serait donc pauvre dans le plus vaste trésor du Monde, seul, un tel être peut toucher à l'Or sans en subir la fatalité. Par ces mains faite, sera valable la restitution aux Ondines du Rhin." The text is centered and uses a serif font.



Commentaire :

L'image affichée ci-dessus montre le résultat de la désignation d'un fragment de texte. Les surfaces englobant les données du texte sélectionnées apparaissent en vidéo inverse. Nous pouvons constater qu'il n'y a pas une bonne superposition entre les formes de l'image perçues par l'utilisateur et celles produites par segmentation en mot du texte après OCR.

### Le texte reconnu

"Les Dieux ont profané l'Or, les Dieux mourront; telle est la faute, la Chute. Seule, pourrait les sauver la restitution de **FOr** aux Ondines. Mais comment? L'Or est au pouvoir **uu** géant Fafner, **métamorph)osé** en Dragon. Les Dieux n'ont pas pouvoir sur les Géants. D'ailleurs, seul, un être exempt de tout péché, un être inconscient, pur, libre, Ignorant la Richesse, et qui serait donc pauvre dans le plus vaste trésor du Monde, seul, un tel être peut toucher à l'Or sans en subir la fatalité. Par ces mains faite, sera valable la restitution aux **Ondincs** du Rhin."

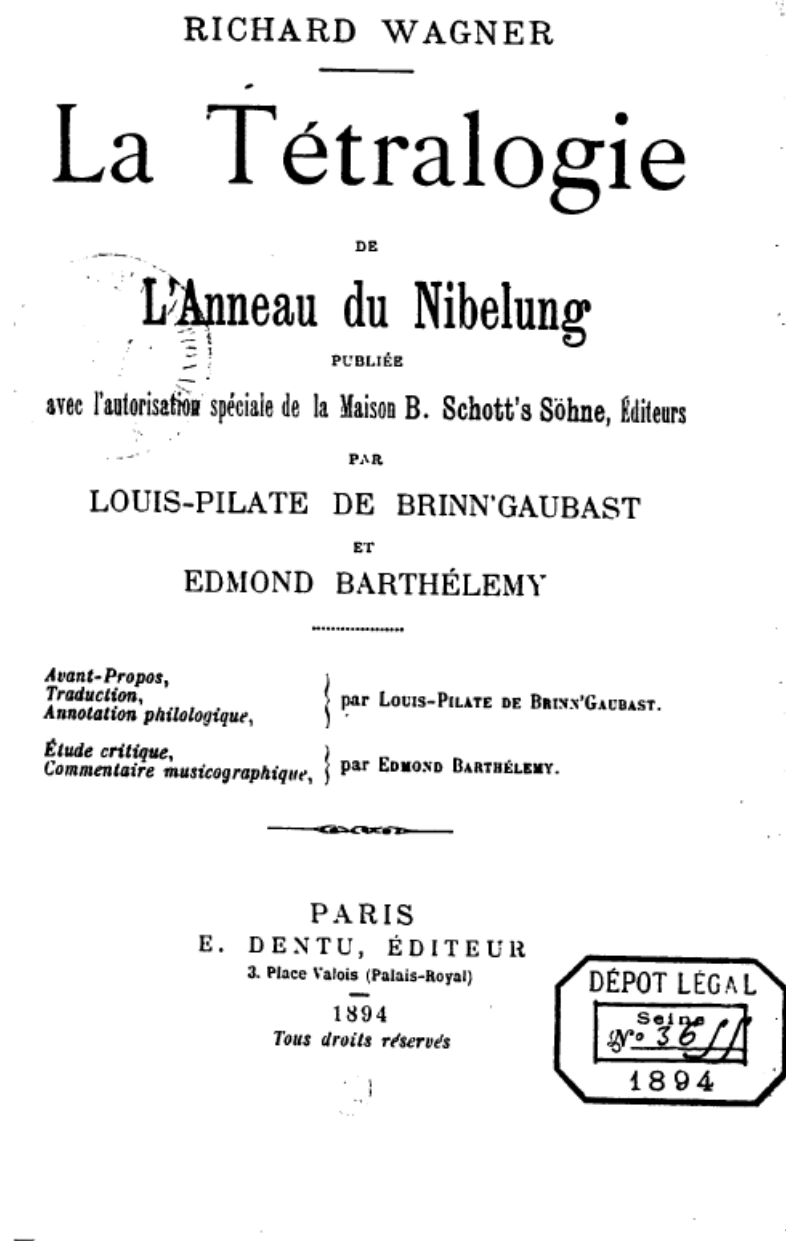
Commentaires :

Le texte sélectionné a été copié puis collé dans un fichier Word par les fonctions classiques de ce traitement de texte. Il comporte 559 caractères, 97 mots dont 4 en erreur correspondant à un taux d'erreur de 4,1% qui est représentatif du taux de reconnaissance que nous obtenons en effectuant ce calcul sur l'ensemble du corps de texte (nous verrons que pour les autres attributs de mise en forme du document, ce taux s'aggrave considérablement).

Ce taux d'erreurs est distribué indifféremment sur des mots de liaisons ou des mots clés du texte. En termes techniques le taux de reconnaissance obtenu est satisfaisant par rapport au cahier des charges de la plateforme mais il induit des anomalies qui pénalisent l'usage du document. En effet, ce seul paragraphe contient des mots clés importants du document (Or, métamorphose, Ondines) qui échappent à la recherche d'information.

Extrait 2 (BNF) : page de couverture

Image originale



■

Commentaires :

Le document présenté ici est la première page de garde du document qui en donne le titre, les auteurs et des informations sur l'éditeur. En outre, la page est surchargée des traces de timbre humide utilisées à la BNF pour le référencement de l'ouvrage. La mise en page est celle d'une couverture (centrage des paragraphes, utilisation de fontes de tailles, proportions et graisses différentes). On notera également la présence d'entrefilets.

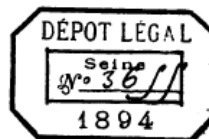
Les conditions de numérisation et les manipulations effectuées sont exactement les mêmes que dans le premier fragment.

Segmentation du texte (présentation de l'information)

RICHARD WAGNER  
DE  
La Tétralogie  
PUBLIÉE  
avec l'autorisation spéciale de la Maison B. Schott's Söhne, Éditeurs  
PAR  
LOUIS-PILATE DE BRINN'GAUBAST  
ET  
EDMOND BARTHÉLEMY

Avant-Propos,  
Traduction,  
Annotation philologique, } par LOUIS-PILATE DE BRINN'GAUBAST.  
Étude critique,  
Commentaire musicographique, } par EDMOND BARTHÉLEMY.

PARIS  
E. DENTU, ÉDITEUR  
3, Place Valois (Palais-Royal)  
1894  
Tous droits réservés



Commentaires :

La segmentation du texte reconnu n'est pas alignée avec le texte dans l'image. Si l'on veut sélectionner dans Acrobat Reader le mot "Tétralogie", il faut chercher le segment en dessous du mot à l'image dans notre exemple.

Ce document devient très difficilement utilisable dans une interface combinant lecture et extraction du contenu.

Texte reconnu par l'OCR

RICHARD WAGNER

La

Tétralogie

DE

L~ineau PdUuBLNIÉiEbelung

avec!'MiorisatsMpéBciadleelaNaisoBn. Schott'sSbhne,Éditeurs

PAR

LOUIS-PILATE DE BRINN'GAUBAST

Er

EDMOND BARTHÉLÉMY

Avant-Propos,

ATrnadnuOctiäotni,on philologique, par L.S-P.LATE DE B-G.ST.

ÉCtOudMeMcr~it!i<qautefC,MM~t'CO~rapA~MjpCa,Pr~ ~0x0 IBiARTH~ÉLLBÉrMiYY..

E. DENTU, EDITEUR

3. Ptaee Vatois(Patais-Royat)

PARIS

1894

7'OM~~rot!~ r/s~r~~

)

Commentaires :

On observe un véritable décrochage des performances de l'OCR, rendant inopérant tout post traitement linguistique. Ce décrochage est probablement dû à une erreur de segmentation causée par

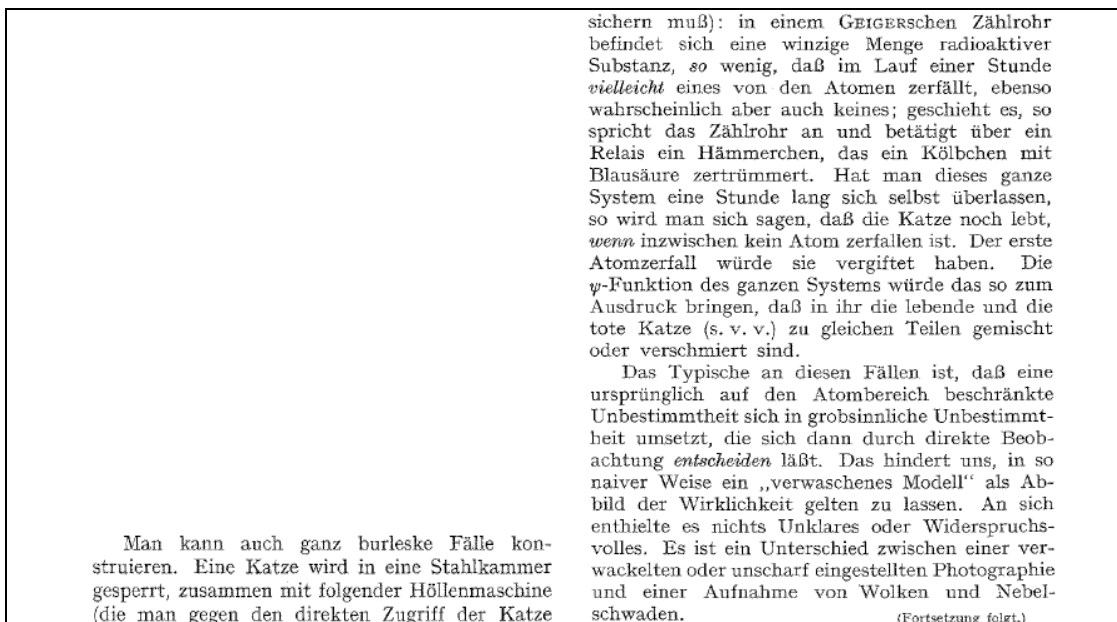
- une mise en page structurée,
- la trace d'un tampon de référencement

Il faut noter que les informations présentes sur cette page de l'ouvrage sont précisément celles qui servent à référencer le document.

### Extrait 3 (Internet) : article scientifique édité en 1938

Les problèmes vus précédemment ne trouvent pas qu'en explorant les ouvrages numérisés par la BNF. Ils se retrouvent dans la recherche de documents sur le web. L'extrait présenté ci-dessous est issu de la version numérisée en pdf de l'article écrit par Erwin Schrödinger dans la revue "Natur und Wissenschaft" en 1938 dans lequel il explique sa célèbre expérience de pensée impliquant un chat dans une boîte. Malgré une qualité visuelle assez bonne de l'image numérisée à 150 dpi, on s'aperçoit que la retranscription du texte par OCR rend très difficile sa compréhension. Le taux de bonne reconnaissance de caractères est estimé ici à 95%. Ce taux de reconnaissance reste correct sur le plan technique mais nuit fortement à la lisibilité du texte (tant par un humain que par la machine).

#### L'image de l'extrait



#### Le texte obtenu par OCR

Man kann auch ganz burleske Fälle konstruieren. Eine Katze wird in eine Stahlkammer gesperrt, zusammen mit folgender Höllenmaschine (die man gegen den direkten Zugriff der Katze sichern muß): in einem GEIGERSchen Zählrohr befindet sich eine winzige Menge radioaktiver Substanz, so wenig, daß im Lauf einer Stunde *vielleicht* eines von den Atomen zerfällt, ebenso wahrscheinlich aber auch keines; geschieht es, so spricht das Zählrohr an und betätigt über ein Relais ein Hämmerchen, das ein Kölbchen mit Blausäure zertrümmert. Hat man dieses ganze System eine Stunde lang sich selbst überlassen, so wird man sich sagen, daß die Katze noch lebt, *wenn* inzwischen kein Atom zerfallen ist. Der erste Atomzerfall würde sie vergiftet haben. Die  $\psi$ -Funktion des ganzen Systems würde das so zum Ausdruck bringen, daß in ihr die lebende und die tote Katze (s. v. v.) zu gleichen Teilen gemischt oder verschmiert sind. Das Typische an diesen Fällen ist, daß eine ursprünglich auf den Atombereich beschränkte Unbestimmtheit sich in grobsinnliche Unbestimmtheit umsetzt, die sich dann durch direkte Beobachtung *entscheiden* läßt. Das hindert uns, in so naiver Weise ein „verwaschenes Modell“ als Abbild der Wirklichkeit gelten zu lassen. An sich enthielte es nichts Unklares oder Widerspruchsvolles. Es ist ein Unterschied zwischen einer verwackelten oder unscharf eingestellten Photographie und einer Aufnahme von Wolken und Nebelschwaden. (Fortsetzung folgt.)



## **ANNEXE 2 :**

### **Description fonctionnelle du projet DocMining**

*Cette annexe contient des extraits de la documentation d'implémentation technique de la plateforme DocMining citée dans la thèse. Ces extraits illustrent la philosophie retenue dans la conception de la plateforme.*

## *Structure des documents dans DocMining*

La plate-forme DocMining est bâtie autour de trois entités élémentaires : le document, le traitement et le scénario. Le document est le plus passif de ces trois composants dans le sens où il n'a aucune influence directe sur les deux autres. Néanmoins, son rôle est central puisqu'il est le seul canal de communication possible entre traitements, le scénario ayant en charge le contrôle de la communication. Aussi, plus le document aura une structure flexible est modulable, plus les interactions entre traitements seront possibles et plus les scénarios pourront être complexes.

La recherche d'une structure de document flexible est modulable abouti logiquement au choix du format XML (aussi bien en raison de sa nature que des outils permettant de le manipuler). Le but de ce chapitre est de décrire certaines structures de document pouvant être manipulés par les traitements. En effet, même si la plate-forme actuelle accepte en entrée toute structure XML valide, certains traitement font appel à des structures XML particulières spécialement définies pour des tâches de reconnaissance de document. A l'heure actuelle, deux structures sont principalement utilisées dans la plate-forme:

- une structure de document basée sur un schéma XML
- une structure XML résultant de la sérialisation des propriétés des objets de la JAI.

Dans une première partie, la nature du contenu d'un document est abordé. Par la suite, le schéma XML est présenté ainsi que l'API Java associée. Finalement, a troisième partie présente comment s'appuyer sur les propriétés de sérialisation des objets de la JAI pour produire une sortie XML.

### *Nature des documents traités*

Avant d'étudier les différentes structure de document développées pour la plate-forme, il est important d'explicitier quelle peut être la nature de l'information contenue dans un document DocMining. En effet, les documents manipulables par la plate-forme DocMining peuvent avoir deux natures :

- Le document correspond à une représentation structurée du document traité. La structure du document XML est le reflet d'un aspect de la structure physique du document traité. Les éléments contenus dans le document correspondent exactement aux objets que la tâche de reconnaissance cherche à extraire et identifier.
- Le document n'est qu'une coquille vide, les informations qu'il contient ne sont que des éléments de contrôle indiquant que les différentes étapes de la tâche de reconnaissance ont eu lieu. Au mieux, les éléments contiennent un lien sur des objets réellement en rapport avec la tâche de reconnaissance. La hiérarchie du document obtenu n'a aucun rapport avec la structure des documents réels.

L'extrait 1 illustre le premier aspect de ce que peut être un document. La structure correspond à une tâche de segmentation en blocs et lignes de texte. L'extrait 2 quant à lui représente le résultat d'une tâche s'apparentant à un traitement batch d'images.



```

<object_doc source=" c:\v3\test\liberte\liberte.pdf" type="PdfDoc">
<object_position h="1290" w="848" x="0" y="0"/>
</object_doc>
<object_doc type="TextLine">
<object_position h="11" w="81" x="711" y="53"/>
<object_data>
<ascii_data>LUNDI 19 JUIN 2000 </ascii_data>
</object_data>
</object_doc>
<object_doc source="c:\v3\test\liberte\liberte.pdf" type="TextBlock">
<object_position h="96" w="540" x="14" y="82"/>
<object_doc object_id="995199304" type="TextLine">
<object_position h="53" w="540" x="14" y="82"/>
<object_data>
<ascii_data>Le musical made in Switzerland </ascii_data>
</object_data>
</object_doc>
<object_doc object_id="995199358" type="TextLine">
<object_position h="53" w="420" x="14" y="125"/>
<object_data>
<ascii_data>est né à Bonnefontaine! </ascii_data>
</object_data>
</object_doc>
</object_doc>

```

**Extrait 1: document DocMining correspondant à une approche structurée**

```

<object_doc source="" type="ImageBase">
<object_doc source="C:\corpus\poulet\aille\0000.pgm" type="GrayImage">
  <object_doc source="C:\corpus\poulet\aille\0000.pbm" type="BinaryImage">
</object_doc>
<object_doc source="C:\corpus\poulet\aille\0001.pgm" type="GrayImage">
  <object_doc source="C:\corpus\poulet\aille\0001.pbm" type="BinaryImage">
</object_doc>
<object_doc source="C:\corpus\poulet\aille\0002.pgm" type="GrayImage">
  <object_doc source="C:\corpus\poulet\aille\0002.pbm" type="BinaryImage">
</object_doc>
<object_doc source="C:\corpus\poulet\aille\0003.pgm" type="GrayImage">
  <object_doc source="C:\corpus\poulet\aille\0003.pbm" type="BinaryImage">
</object_doc>

```

**Extrait 2: document DocMining correspondant à un scénario batch**

Un des intérêts de DocMining est que la plate-forme permet de manipuler aussi bien un document unique qu'un lot de document. Ce type d'approche peut s'avérer très utile pour un utilisateur souhaitant modéliser une tâche de reconnaissance au moyen d'un scénario. En effet, lors de la mise au point d'un traitement, un concepteur alterne entre phase de test unitaire et phase de test batch.

### ***Structure de document définie par un schéma XML***

Les premières recherches sur la plate-forme étaient orientées vers l'étude d'une structure pivot et la définition d'objets du domaine pouvant être partagés. Ces travaux ont abouti sur la définition d'un schéma XML décrivant une structure de document assez générale. De nombreux traitements actuels sont basés sur ce schéma. La figure 2 présente sous forme de diagramme la structure du schéma et les objets qu'il contient.

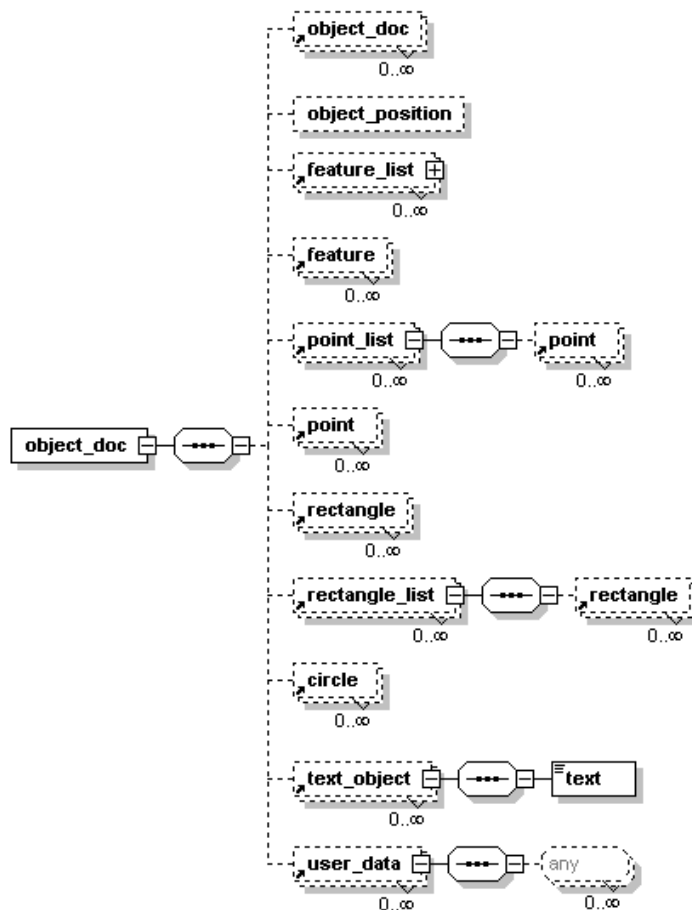


Figure 1: schéma XML définissant la structure de document

L'extrait 3 donne le schéma correspondant à la figure 2. La particularité de la structure ainsi définie est que les objets sont différenciables non pas en fonction de leur tag (`<object_doc>`) mais en fonction de leurs attributs (`source, name, type`).

```

<xs:element name="object_doc">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="object_doc" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element name="object_position" type="object_position" minOccurs="0"/>
      <xs:element ref="feature_list" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="feature" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="point_list" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="point" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="rectangle" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="rectangle_list" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="circle" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="text_object" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="user_data" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="source" type="xs:string"/>
    <xs:attribute name="name" type="xs:string"/>
    <xs:attribute name="type" type="xs:string"/>
    <xs:attribute name="object_id" type="xs:string"/>
  </xs:complexType>
</xs:element>

```

### Extrait 3 : Schéma XML d'une structure physique de document

La description d'un objet peut être complétée par un certain nombre de descripteurs prédéfinis correspondant à des propriétés de forme basiques. L'extrait 4 illustre la structure des objets <feature>, <feature\_list> et <text\_object>. Comme la liste des descripteurs est loin d'être exhaustive, l'élément <user\_data> permet d'ajouter des descripteurs additionnels tout en préservant la cohérence du schéma.

```

<xs:complexType name="intrinsic_data">
  <xs:attribute name="name" type="xs:string"/>
  <xs:attribute name="value" type="xs:string"/>
</xs:complexType>
<xs:element name="text_object">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="text" type="xs:string"/>
    </xs:sequence>
    <xs:attribute name="name" type="xs:string"/>
    <xs:attribute name="value" type="xs:string"/>
  </xs:complexType>
</xs:element>
<xs:element name="feature" type="intrinsic_data"/>
<xs:complexType name="label">
  <xs:complexContent>
    <xs:extension base="intrinsic_data">
      <xs:attribute name="confidence" type="xs:float"/>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
<xs:element name="feature_list">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="label" type="label" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="feature" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="name" type="xs:string"/>
    <xs:attribute name="value" type="xs:string"/>
  </xs:complexType>
</xs:element>

```

#### Extrait 4: Exemples de descripteurs de formes

Associée au schéma XML on trouve une API Java permettant d'interagir avec les objets de la structure. L'API permet de charger, créer et sauvegarder directement en XML des instances d'éléments de la structure. Une telle API est beaucoup plus intuitive qu'une API XML classique telle que DOM dans le sens où l'on manipule directement les éléments et non des nœuds XML. De nombreux outils permettent de générer à partir du schéma XML une API ayant de telles fonctionnalités. L'intérêt d'utiliser ces outils est qu'ils permettent de régénérer automatiquement la librairie dès que le schéma XML subit une modification.

Initialement, la génération de code était faite au moyen du package JAXB de Sun. Néanmoins, le non respect des wildcards de la spécification XML schéma ainsi que l'aspect non open source du produit et ont rendu son utilisation rédhibitoire. Pour autant, l'idée de générer automatiquement une API de manipulation en parallèle du schéma n'a pas été abandonnée. L'outil XMLBeans de BEA remplace avantageusement JAXB puisqu'il est open source et supporte les wildcards.

La figure 3 donne un aperçu des interfaces générées automatiquement par l'outil XMLBeans, il apparaît clairement que le code résultant correspond directement à la structure du schéma XML.

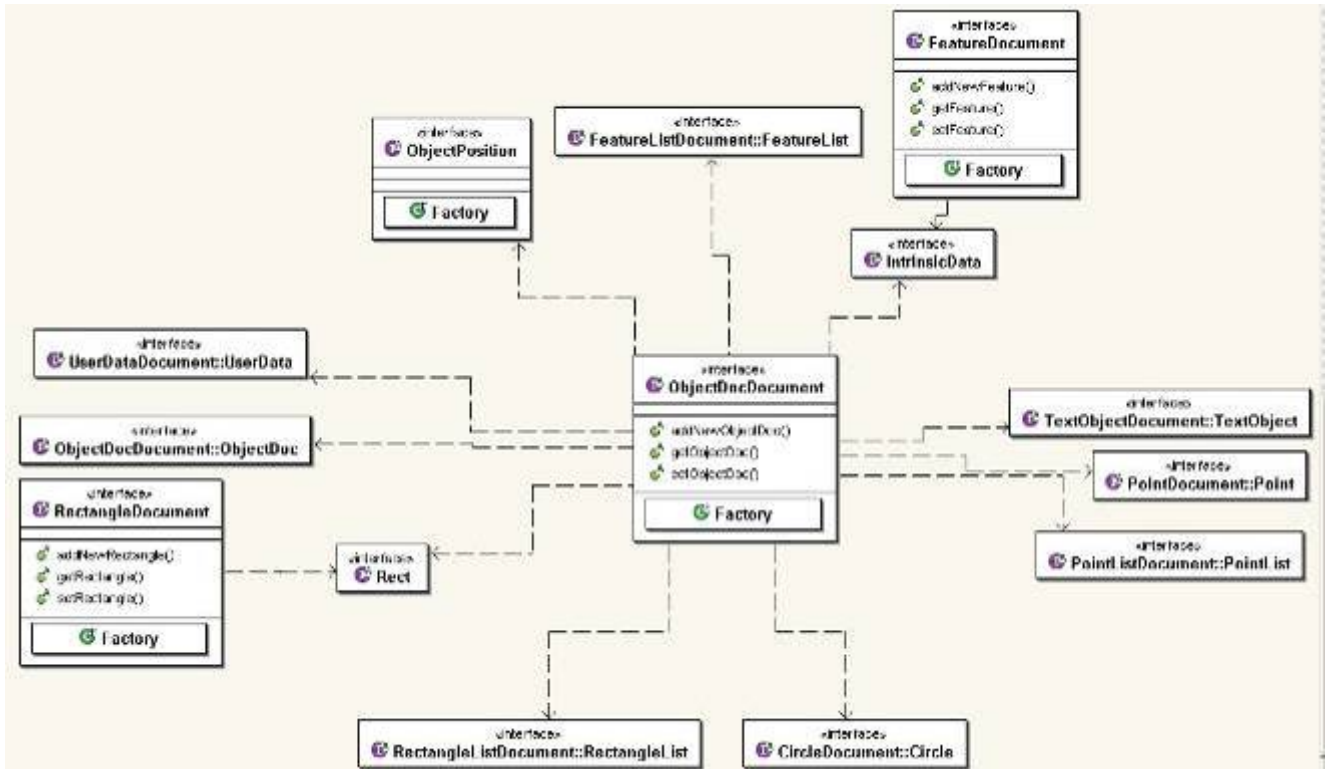


Figure 2 : API Java correspondant au schéma de document

L'extrait suivant fournit un exemple d'utilisation du code généré et donne un aperçu des possibilités offertes par la librairie.

```

//creates a new document
ObjectDocDocument doc=ObjectDocDocument.Factory.newInstance();
ObjectDocDocument.ObjectDoc objDoc=doc.addNewObjectDoc();
//sets the object_doc name
objDoc.setName("Sample");
//sets the type
objDoc.setType("SampleType");
//sets the source
objDoc.setSource("image.src");

for(int i=0;i<10;i++) {
    //add a descriptor
    IntrinsicData feat=objDoc.addNewFeature();
    feat.setName("Descriptor "+Integer.toString(i));
    feat.setValue(Integer.toString(i));
}

//saves the Document
doc.save(new File("c:\\test.xml"),new XmlOptions().setSavePrettyPrint());

```

#### Extrait 5 : Utilisation de l'API de manipulation de documents XML

Le document produit est le suivant :

```

<?xml version="1.0" encoding="UTF-8"?>
<object_doc source="image.src" type="SampleType" name="Sample">
  <feature value="0" name="Descriptor 0"/>
  <feature value="1" name="Descriptor 1"/>
  <feature value="2" name="Descriptor 2"/>
  <feature value="3" name="Descriptor 3"/>
  <feature value="4" name="Descriptor 4"/>
  <feature value="5" name="Descriptor 5"/>
  <feature value="6" name="Descriptor 6"/>
  <feature value="7" name="Descriptor 7"/>
  <feature value="8" name="Descriptor 8"/>
  <feature value="9" name="Descriptor 9"/>
</object_doc>

```

#### Extrait 6 : Document produit par l'API

### *Création de document XML par sérialisation des objets JAI*

Une des particularités des objets de la JAI est de pouvoir définir des propriétés propres à l'objet « à la volée » qui sont alors stockées dans l'objet et auxquelles on peut accéder ultérieurement. On peut donc s'appuyer sur cette caractéristique pour définir une structure de document adaptée à des tâches de traitements d'image. Par la suite, il suffit de définir une classe capable d'inspecter les propriétés de l'objet et de produire une sortie XML. L'exemple suivant fourni un exemple de sérialisation d'un objet JAI.

```
String sourceFile="c:\\java\\docmining\\jai\\rotule.pgm";
    PlanarImage source=JAI.create("fileload",sourceFile);
    //adds a property
    source.setProperty("file_name",sourceFile);
    //generates an XML representation based on the object properties
IMarshallableObjectDoc
jai=JAIMarshaller.getJAIMarshaller().createMarshallable(source);
//outputs the properties
jai.marshal(System.out);
```

#### Extrait 7 : S rialisation dun objet de la JAI

La g n ration de la structure XML se fait en scrutant tout ou partie des propri t s de l'objet source ; la classe JAIMarshaller est responsable de cette t che. L'extrait suivant correspond   la repr sentation XML produite en sortie.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<PlanarImage
data_type="byte"
file_name="c:\\java\\docmining\\jai\\rotule.pgm"
image_height="512"
image_min_x_coord="0"
image_min_y_coord="0"
image_width="512"
nbands="1"
pixel_size="8" />
```

#### Extrait 8 : Document produit par s rialisation de la JAI

### Structure d'un sc nario

Pour pouvoir  tre appliqu  un sc nario doit respecter une certaine syntaxe qui est sp cifi e par un sch ma XML. La Figure 3 illustre sous forme de graphe la structure d'un sc nario. Un sc nario est compos  de deux parties principales :

- les  l ments de configuration qui permettent d'initialiser les composants destin s    tre utilis s dans la suite du sc nario.
- les  tapes qui correspondent   la r alisation de la t che de reconnaissance de document..

#### Structure des  l ments de configuration

Les  l ments de configuration sont d finis par les tags suivants :

- **<interpreter>** : ce tag permet de d clarer un composant capable d'interpr ter une instruction (cf **<instruction\_step\_type>**)
- **<trigger\_listener>** : ce tag permet de d clarer un composant de type g chette
- **<support\_listener>** : ce tag permet de d clarer un composant de type adaptateur de support.
- **<init\_scenario>** : ce tag permet de d clarer une instruction d'initialisation.
- **<user\_defined>** : ce tag permet d'ins rer des informations qui pourront  tre r utilis e par des instructions ou par tout autre composant propri taire.

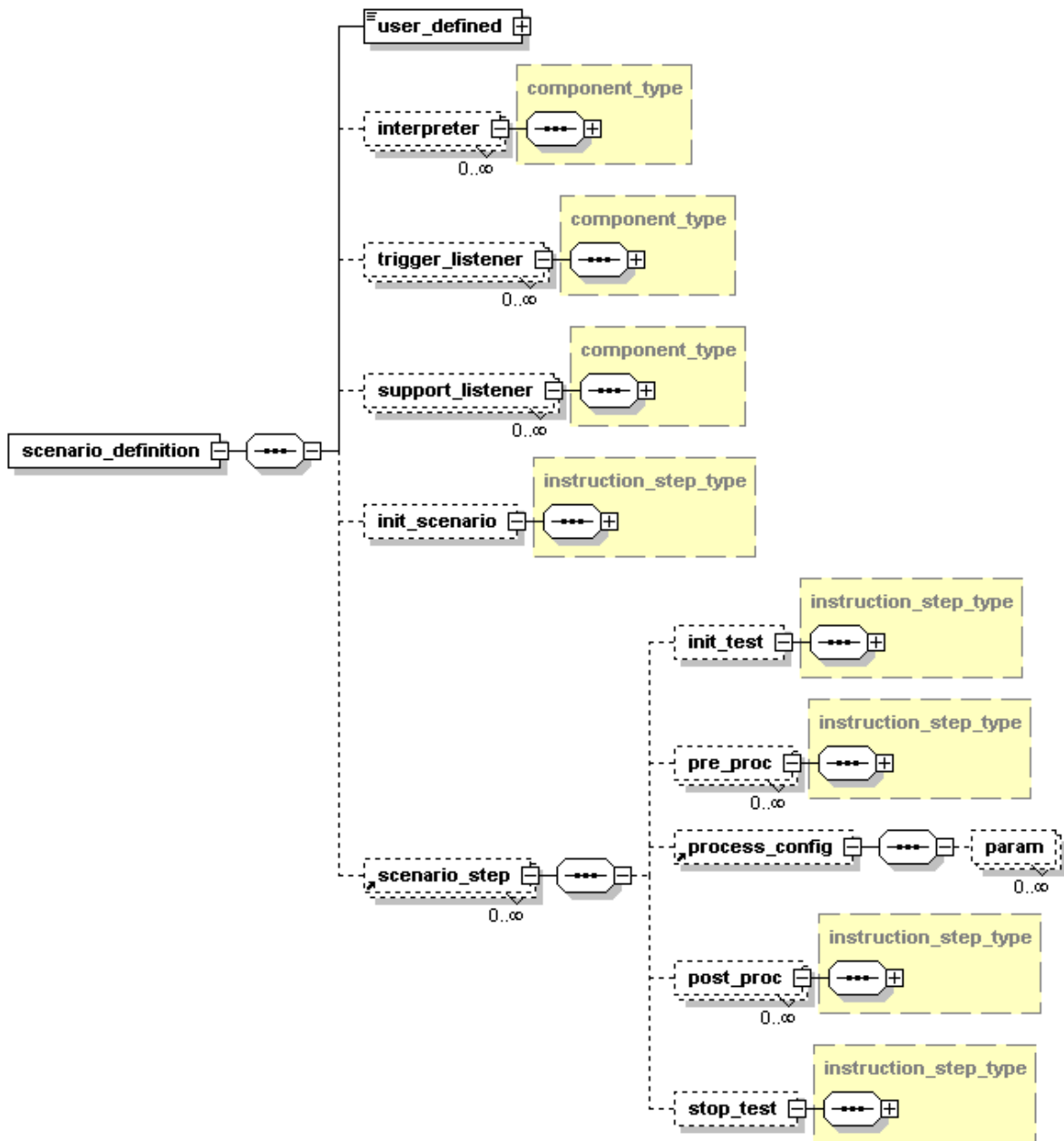


Figure 3 : Structure d'un scénario

Le type <component\_type> correspond à la représentation des éléments <interpreter>, <trigger\_listener> et <support\_listener> et est détaillé dans l'Extrait 1. La principale caractéristique d'un composant est la classe qui doit être instanciée (attribut class\_name). L'objet ainsi créé peut ensuite être initialisé par des paramètres (éléments <cpt\_param>). Le référencement, le chargement et l'appel des composants sont exposés en détail dans la troisième partie de ce chapitre.



```

<xs:complexType name="component_type">
  <xs:sequence>
    <xs:element name="cpt_param" type="simple_param_type" minOccurs="0"
      maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="name" type="xs:string"/>
  <xs:attribute name="class_name" type="xs:string"/>
</xs:complexType>

<xs:complexType name="simple_param_type">
  <xs:attribute name="key" type="xs:string"/>
  <xs:attribute name="value" type="xs:string"/>
</xs:complexType>

```

Extrait 9: schéma d'un composant

La figure 4 illustre la structure d'une instruction telle que contenue dans le tag `<init_scenario>`.

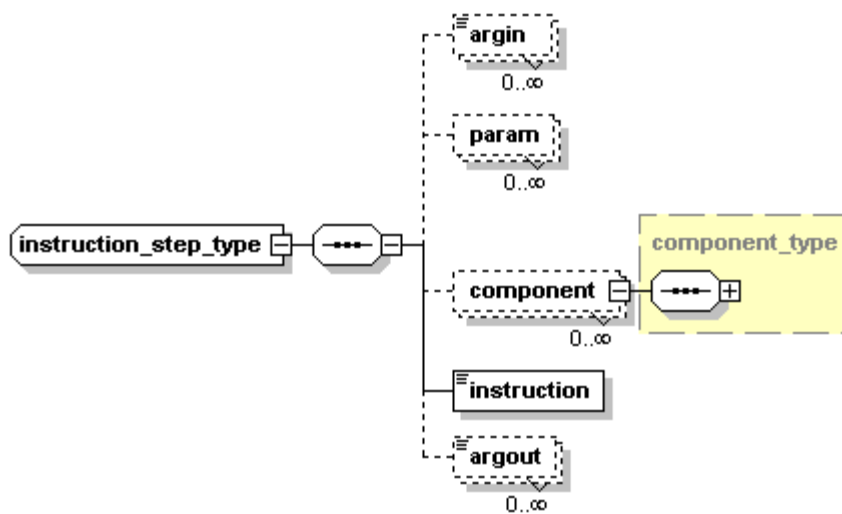


Figure 4: Structure d'une instruction

Une instruction est décomposée en trois entités : les arguments d'entrées, l'instruction à exécuter et les arguments de sorties, le détail de sa représentation XML est donnée par l'Extrait 10.

Les arguments d'entrée d'une instruction peuvent être :

- des variables référencées dans le scénario (tag `<argin>`).
- des composants (tag `<component>`).
- des paramètres de traitement (tag `<param>`).

Une instruction peut également produire des variables référencées dans le scénario et réutilisable par une autre instruction (tag `<argout>`). Seule l'instruction `<init_scenario>` permet de déclarer des variables globales, les instructions contenues dans les étapes ne produisent que des variables locales.

L'instruction elle-même est définie par le tag `<instruction>` et contient le code à interpréter.

```

<xs:complexType name="instruction_step_type">
  <xs:sequence>
    <xs:element name="argin" type="arg_desc" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="param" type="param_type" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="component" type="component_type" minOccurs="0"
      maxOccurs="unbounded"/>
    <xs:element name="instruction" type="xs:string"/>
    <xs:element name="argout" type="arg_desc" minOccurs="0" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="interpreter" type="xs:string"/>
</xs:complexType>

<xs:complexType name="arg_desc" mixed="true">
  <xs:attribute name="name" type="xs:string"/>
</xs:complexType>

```

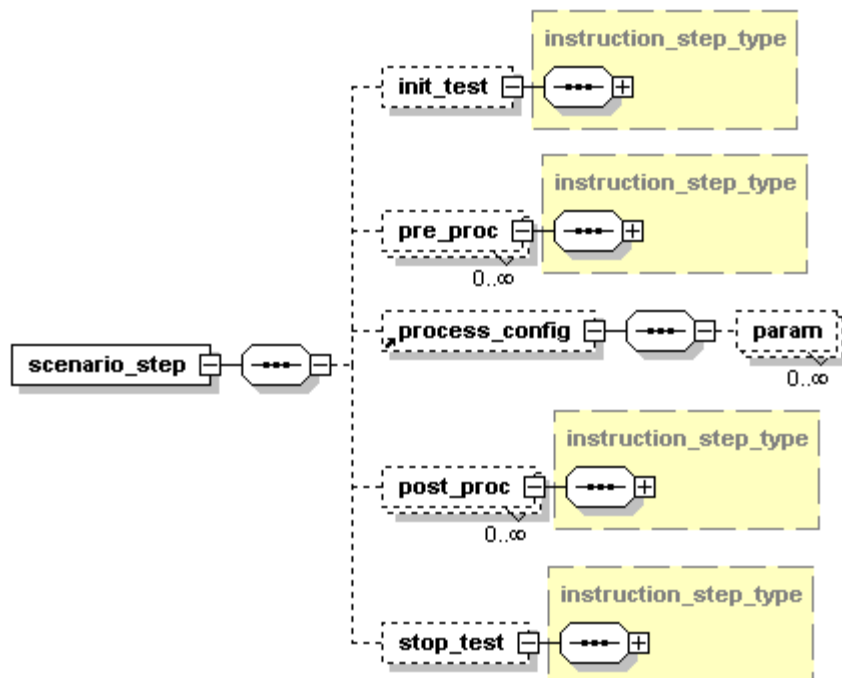
**Extrait 10 : schéma d'une instruction**

Un aperçu des possibilités offertes par les instructions d'initialisation est donné dans la troisième partie de ce chapitre.

### Structure des étapes d'un scénario

Outre des éléments d'initialisation, un scénario est surtout constitué d'étapes. En effet, les composants et variables déclarés lors de la phase d'initialisation ont pour objectif principal d'être réutilisés dans une ou plusieurs étapes.

La Figure 5: Structure d'une étape de scénario donne un aperçu de la structure d'une étape qui est représentée dans un scénario par le tag <scenario\_step>.



**Figure 5: Structure d'une étape de scénario**

Une étape est composée des éléments suivants :

- des instructions pré et post traitements (tag <pre\_proc> et <post\_proc>).

- un traitement dont les paramètres sont contenus dans l'élément `<process_config>`, la classe de référence du traitement est contenue dans l'attribut `process_classname`.
- Des instructions permettant de répéter conditionnellement l'exécution d'un traitement ainsi que les pré et post-traitement (tag `<init_loop>` et `<test_loop>`).

Les paramètres du traitement lui-même sont contenus dans l'élément `<process_config>` composé de `<param>` détaillés dans l'extrait suivant

```
<xs:complexType name="param_type">
  <xs:attribute name="type" default="ParamIn">
    <xs:simpleType>
      <xs:restriction base="xs:NMTOKEN">
        <xs:enumeration value="Input"/>
        <xs:enumeration value="Output"/>
        <xs:enumeration value="ParamIn"/>
        <xs:enumeration value="ParamOut"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name="name" type="xs:string" use="required"/>
  <xs:attribute name="support" type="xs:string"/>
  <xs:attribute name="param_value" type="xs:string"/>
  <xs:attribute name="info" type="xs:string"/>
  <xs:attribute name="read_only" default="no">
    <xs:simpleType>
      <xs:restriction base="xs:NMTOKEN">
        <xs:enumeration value="yes"/>
        <xs:enumeration value="no"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
</xs:complexType>
```

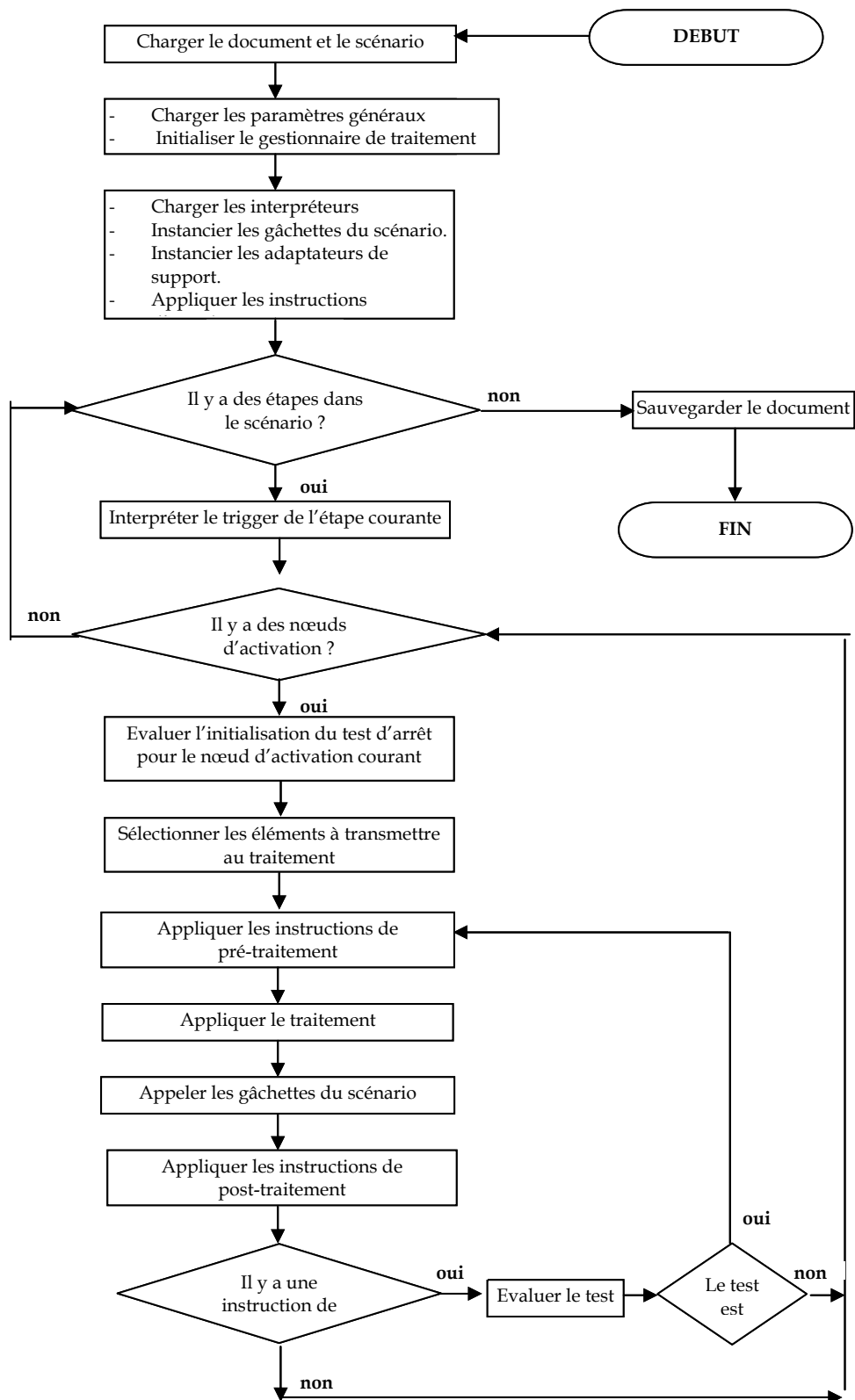
#### Extrait 11 : schéma d'un paramètre de traitement

Un paramètre est composé de différents attributs :

- `name` : le nom du paramètre, on transmet au traitement le couple nom du paramètre / valeur interprétée du paramètre.
- `info` : un attribut non obligatoire décrivant le paramètre.
- `param_value` : la valeur du paramètre, en fonction de l'attribut `support`, cette chaîne de caractère est interprétée différemment avant d'être transmise au traitement.
- `support` : Cet attribut indique comment doit être interprétée la valeur du paramètre ainsi un support « Double » indique que la chaîne de caractère correspond à un double. Un support « ObjectDoc » indique de la chaîne de caractère correspond à une expression XPath. Les adaptateurs de support sont les composants en charge de l'interprétation de la valeur d'un paramètre en fonction de son support. Une liste d'adaptateurs de base est fournie en annexe du document.
- `type` : à l'heure actuelle deux types sont supportés « ParamIn » et « Input ». Le premier indique que la valeur du paramètre est transmise en tant que paramètre au traitement, le second que la valeur est transmise en tant qu'objet à traiter.

La structure de scénario fournit un formalisme pour déclarer et agencer les éléments constitutifs d'un scénario. Seul le moteur ImTrAC est en mesure d'interpréter la structure d'un scénario, d'exécuter les traitements et de mettre à jour le document.

## Organigramme d'un scénario DocMining





## FOLIO ADMINISTRATIF

### THESE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

NOM : GARDES

DATE de SOUTENANCE : 24/07/2009

(avec précision du nom de jeune fille, le cas échéant)

Prénoms : Joël

TITRE : Le Document Numérique : la Complexité des Formes, les Formes de la Complexité

NATURE : Doctorat

Numéro d'ordre : 2009-ISAL-0061

Ecole doctorale : Informatique et Mathématiques

Spécialité : Document numérique, Images et systèmes d'Informations Communicants

Cote B.I.U. - Lyon : T 50/210/19 / et bis

CLASSE :

RESUME :

Les réseaux de télécommunications actuels offrent une interconnexion généralisée de terminaux dont les capacités de calculs et de mémoire évoluent sans cesse. De nouveaux comportements émergent face à une information devenue pervasive. L'utilisateur n'est plus passif, ni unique, c'est un acteur local du traitement de l'information. Une interface doit prendre en considération la variété des modes d'accès et d'usages qui conditionnent la présentation de tout contenu.

La présentation de l'information se situe au carrefour de trois grands domaines indissociables : l'acquisition et le traitement du signal, l'ingénierie des données et des connaissances, l'interaction homme machine. Ce thème pluridisciplinaire, très proche du document numérique nous permet de comprendre et prévoir nos manières d'appréhender l'information numérique. Nous retrouvons des sujets tels que la qualité de la numérisation et son évaluation.

Cette thèse présente un socle théorique identifiant les problèmes posés par la communication numérique dans le contexte évoqué. Les éléments de solutions sont présentés par des réalisations techniques innovantes sur la reconnaissance de mots et la distance entre informations. Ces applications, issues des théories de la complexité et de la cybernétique, ont pour caractéristiques remarquables leur simplicité algorithmique face à la complexité des problèmes traités, et leur facilité d'appropriation par un utilisateur totalement inclus dans la boucle de traitement.

MOTS-CLES :

Théorie de l'information, interaction homme machine, document numérique, numérisation, ergonomie, systèmes complexes, cybernétique

Laboratoire (s) de recherche :

LIRIS – INSA Lyon

Directeur de thèse: M.. Hubert Emptoz

Président de jury : M. Christian Jutten

Composition du jury :

MM. Rolf Ingold, Bruno Bachimont, Christian Jutten, George Nagy, Christophe Garcia, Denis Chêne, Hubert Emptoz