

DOCUMENT RESUME

ED 317 029

FL 017 783

AUTHOR Ladouceur, Jacques  
 TITLE Une Analyse automatique en syntaxe textuelle (An Automated Analysis of Textual Syntax). Publication K-5.  
 INSTITUTION Laval Univ., Quebec (Quebec). International Center for Research on Bilingualism.  
 REPORT NO ISBN-2-89219-186-6  
 PUB DATE 88  
 NOTE 84p.  
 AVAILABLE FROM International Center for Research on Bilingualism, Par Casault-Universite Laval, Quebec G1K 7P4, Canada.  
 PUB TYPE Reports - Research/Technical (143)  
 LANGUAGE French

EDRS PRICE MF01/PC04 Plus Postage.  
 DESCRIPTORS Artificial Intelligence; Computational Linguistics; \*Computer Software; \*Discourse Analysis; Foreign Countries; French; \*French Literature; Models; Programming; Statistical Analysis; \*Syntax; Tenses (Grammar); Verbs  
 IDENTIFIERS Flaubert (Gustave); \*Un Coeur Simple

ABSTRACT

This study reports the use of automated textual analysis on a French novel. An introductory section chronicles the history of artificial intelligence, focusing on its use with natural languages, and discusses its application to textual syntax. The first chapter examines computational linguistics in greater detail, looking at its relationship to several areas of language study (literature, artificial intelligence, and analysis of verb tense), the technological requirements, and specific aspects of the automated recognition of verb forms in a text. The second chapter outlines the automation of tense analysis using a model proposed by Harald Weinrich. Subsections discuss automated recognition of verb forms, statistical analyses, and the interactive role of the investigator with this model. The third chapter contains the results of an analysis performed on Gustave Flaubert's "Un Coeur Simple," accomplished in four stages: (1) analysis of the overall "landscape" of verbs in each chapter of the story; (2) the narrative tempo; (3) the speaker's attitude; and (4) transitions in tense within the text. The brief time taken for the computerized analysis is emphasized, and the approach used is seen as useful in bringing together literature and the study of natural languages through artificial intelligence. (MSE)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

ED317023

PUBLICATION  
K-5

(CIRB)

Centre international de recherche  
sur le bilinguisme

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

This document has been reproduced as  
received from the person or organization  
originating it.

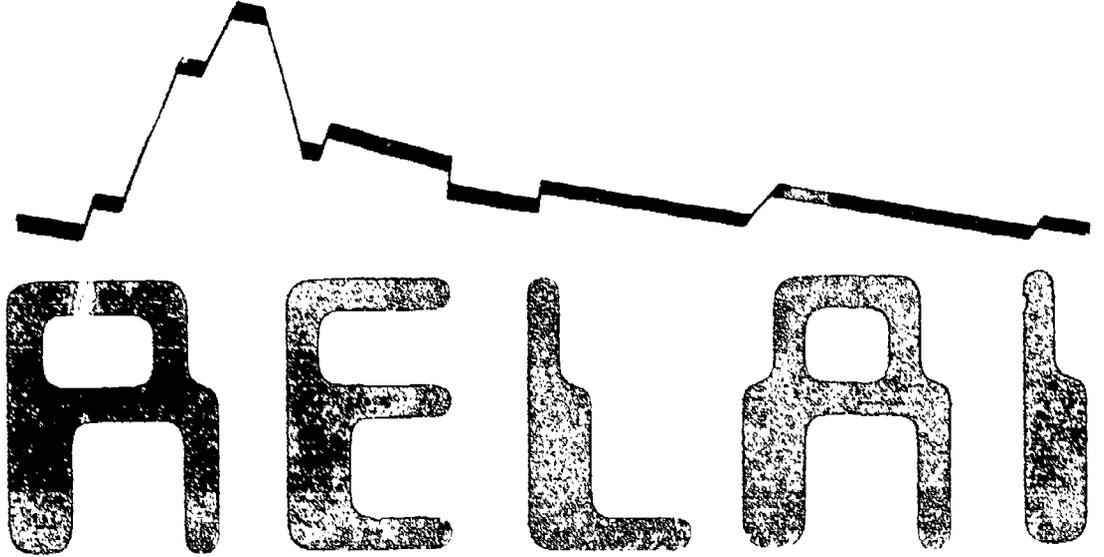
Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OERI position or policy.

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

L. Laforge

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."



Recherche  
en  
linguistique  
appliquée  
à  
l'informatique

UNE ANALYSE AUTOMATIQUE  
EN SYNTAXE TEXTUELLE

FL017783

Jacques Ladouceur

*Jacques LADOUCEUR*

**UNE ANALYSE AUTOMATIQUE  
EN SYNTAXE TEXTUELLE**

**Publication K-5**

**1988  
Centre international de recherche sur le bilinguisme  
International Center for Research on Bilingualism  
Québec**

*Le Centre international de recherche sur le bilinguisme est un organisme de recherche universitaire qui reçoit une contribution du Secrétariat d'État du Canada pour son programme de publication.*

*Ont contribué de façon spéciale à la publication de ce bulletin, l'Université Laval et la Compagnie IBM du Canada.*

*The International Center for Research on Bilingualism is a university research institution which receives a supporting grant from the Secretary of State of Canada for its publication programme.*

*Laval University and the IBM Company of Canada have also contributed, in a special manner, to the publication of this bulletin.*

© **1988 CENTRE INTERNATIONAL DE RECHERCHE SUR LE BILINGUISME**  
*Tous droits réservés. Imprimé au Canada*  
Dépôt légal (Québec) 1<sup>er</sup> trimestre 1988  
ISBN 2-89219-186-6

## **REMERCIEMENTS**

**Je tiens à exprimer ici mes remerciements les plus sincères:**

- **à Louis Francoeur et Michael Mepham, mes directeurs de thèse, pour leurs nombreux conseils et leur grande disponibilité.**
- **à mes parents, pour leur soutien financier qui m'a permis de faire l'achat d'un ordinateur.**
- **à Richard Bérubé, qui a toujours porté une oreille attentive à mes préoccupations.**
- **à Lise, qui a été une confidente infatigable, qui m'a apporté un support constant du début à la fin. Sans elle, je n'aurais pu mener à bien mes travaux.**

## TABLE DES MATIÈRES

<b>INTRODUCTION</b> . . . . .	1
1. <b>Intelligence artificielle</b> . . . . .	1
1.1 <b>Histoire</b> . . . . .	1
1.2 <b>La compréhension automatique des langues naturelles</b> . . . . .	2
2. <b>La syntaxe textuelle</b> . . . . .	7
2.1 <b>Monde raconté et monde commenté</b> . . . . .	8
2.2 <b>Attitude de locution</b> . . . . .	8
2.3 <b>Premier plan et arrière-plan</b> . . . . .	8
2.4 <b>L'analyse du comportement verbal et la compréhension automatique des langues naturelles</b> . . . . .	9
2.5 <b>L'automatisation de la méthode d'analyse de Weinrich</b> . . . . .	9
 <b>CHAPITRE 1:</b> <b>LINGUISTIQUE AUTOMATIQUE</b> . . . . .	 11
1. <b>Qu'est-ce que la linguistique automatique?</b> . . . . .	11
1.1 <b>Littérature et linguistique automatique</b> . . . . .	11
1.2 <b>Intelligence artificielle et linguistique automatique</b> . . . . .	12
1.3 <b>L'analyse des temps, intelligence artificielle et linguistique automatique</b> . . . . .	12
2. <b>L'ordinateur</b> . . . . .	13
2.1 <b>Manipulation de formes simples</b> . . . . .	13
2.2 <b>Reconnaissance des formes</b> . . . . .	13
2.3 <b>Les langages de programmation</b> . . . . .	14
3. <b>La reconnaissance automatique des formes verbales dans un texte</b> . . . . .	15
3.1 <b>L'homographie</b> . . . . .	15
3.2 <b>Techniques de reconnaissance automatique des verbes</b> . . . . .	15
3.3 <b>La méthode de Maegaard et Spang-Hanssen</b> . . . . .	15
 <b>CHAPITRE 2:</b> <b>L'AUTOMATISATION DE L'ANALYSE DES TEMPS D'APRÈS LE MODÈLE PROPOSÉ PAR HARALD WEINRICH</b> . . . . .	 27
1. <b>La reconnaissance automatique des formes verbales</b> . . . . .	27
1.1 <b>Emprunts à la méthode de Maegaard et Spang-Hanssen</b> . . . . .	27
1.2 <b>Nos dictionnaires</b> . . . . .	27
1.3 <b>Les tests de désambiguïsation</b> . . . . .	35

2.	L'analyse statistique . . . . .	35
2.1	Les coordonnées du texte . . . . .	36
2.2	Le graphique . . . . .	36
2.3	Table des valeurs . . . . .	37
2.4	Table des fréquences . . . . .	37
2.5	La matrice des transitions temporelles . . . . .	37
3.	L'Analyste . . . . .	38
3.1	Préparation d'un texte . . . . .	38
3.2	Affichage et correction du texte . . . . .	40
3.3	La reconnaissance des formes verbales potentielles . . . . .	40
3.4	La désambiguïsation des formes verbales potentielles . . . . .	42
3.5	La désambiguïsation interactive . . . . .	44
3.6	Le module d'analyse statistique . . . . .	44
<b>CHAPITRE 3:</b>		
<b>ANALYSE AUTOMATIQUE DE <i>UN COEUR SIMPLE</i> DE GUSTAVE FLAUBERT</b>		<b>47</b>
1.	Le relief . . . . .	47
1.1	Chapitre 1 . . . . .	49
1.2	Chapitre 2 . . . . .	51
1.3	Chapitre 3 . . . . .	55
1.4	Chapitre 4 . . . . .	59
1.5	Chapitre 5 . . . . .	62
1.6	Chapitre 6 . . . . .	64
2.	Le tempo narratif . . . . .	64
3.	L'attitude de locution . . . . .	66
4.	Les transitions temporelles . . . . .	67
4.1	La mise en relief . . . . .	68
4.2	L'attitude de locution . . . . .	71
5.	Commentaire . . . . .	74
CONCLUSION . . . . .		75
BIBLIOGRAPHIE . . . . .		79

# INTRODUCTION

## L'INTELLIGENCE ARTIFICIELLE ET LA SYNTAXE TEXTUELLE

### 1. Intelligence artificielle

#### 1.1 Histoire

L'intelligence artificielle, qui est "la discipline visant à comprendre la nature de l'intelligence en construisant des programmes d'ordinateur imitant l'intelligence humaine"<sup>1</sup>, est relativement jeune. Relativement, parce qu'on peut voir dès Socrate une préoccupation de l'Homme à produire des règles permettant d'anticiper ou de conditionner un comportement:

L'histoire de l'intelligence artificielle pourrait bien remonter à l'an 450 avant J.-C., le jour où Socrate, selon Platon, demande à son compatriote, l'Athénien Eutyphron, qui s'apprête à dénoncer pour meurtre son propre père: "Je veux savoir ce qui caractérise cette piété, qui fait de toute action un geste pieux...que je puisse faire appel à elle, et m'en servir comme étalon à l'aide duquel juger les actions et celles des autres hommes." Ce que Socrate demande ici à Eutyphron, c'est ce que nos modernes théoriciens de l'informatique appelleraient une "procédure opératoire", "un ensemble de règles qui nous dise avec précision, étape par étape, comment agir."<sup>2</sup>

Mais à proprement parler les premiers travaux en intelligence artificielle ont été faits vers la fin des années cinquante. C'est Anthony Oettinger qui fut l'un des premiers artisans. Après avoir conçu le premier dictionnaire automatique, il lançait ce qui allait être pendant une dizaine d'années la ruée vers la traduction automatique.

Les bailleurs de fonds de la traduction automatique ne s'intéressaient pas d'abord à leur propre langue, mais aux publications en langue étrangère pour des raisons de stratégie aussi bien politique que militaire. En France, ainsi qu'aux Etats-Unis d'Amérique, l'état, qui finance les recherches (Bien souvent d'ailleurs par le canal de la défense nationale. Le Centre d'Etudes pour la traduction Automatique a été financé pour 1/4 par le Ministère des Armées (D.R.M.E.)), s'intéressa en premier lieu à la traduction du russe en français.<sup>3</sup>

---

<sup>1</sup>Alain Bonnet, *L'intelligence artificielle*, Paris, InterEditions, 1984, p.17.

<sup>2</sup>Hubert L. Dreyfus, *Intelligence artificielle*, Flammarion, New York, 1984, p.3.

<sup>3</sup>Josse De Kock, *Linguistique automatique et langues romanes*, Editions Jean-Favard, Paris, 1977, p.16.

Bien que les résultats n'eurent rien d'éclatant, ils amorçaient une longue réflexion sur la résolution algorithmique des ambiguïtés. Les chercheurs réalisèrent qu'il n'était pas possible de traduire un texte par simple consultation dans un dictionnaire. *Porc* se traduira par *pig* s'il s'agit simplement de l'animal alors qu'il se traduira par *pork* s'il s'agit du mets. Une analyse lexicale ne peut à elle seule fournir suffisamment d'éléments d'information afin de préciser univoquement le sens de tous les mots d'une langue.

## 1.2 *La compréhension automatique des langues naturelles*

Aujourd'hui les applications en intelligence artificielle sont tellement variées qu'il est très difficile de toutes les répertorier. La conception de systèmes informatiques imitant l'intelligence humaine peut être la problématique d'un grand nombre de chercheurs provenant de disciplines diverses. Déjà, plusieurs travaux ont été faits en robotique, en bureautique, en médecine, en biologie, en linguistique, en astronomie etc.

"Individual research workers may have a background in physics, mathematics, linguistics, engineering, neurosciences, psychology, physiology, or some other discipline. Artificial intelligence is a field that crosses boundaries and embraces the approaches and insights of a host of different fields"<sup>4</sup>

Les travaux qui comportent le plus haut niveau de difficulté en intelligence artificielle sont sans doute ceux qui, pour deux raisons, portent sur la compréhension automatique des langues naturelles. D'abord parce que la compréhension du langage naturel est fondamentale dans un projet de simulation de l'intelligence humaine et ensuite parce que les débats sur les problèmes d'ambiguïtés sont loin d'être terminés.

### 1.2.1 *Problèmes reliés à la compréhension automatique des langues naturelles*

"Le lien unissant le signifiant au signifié est arbitraire, ou encore, puisque nous entendons par signe le total résultant de l'association d'un signifiant à un signifié, nous pouvons dire plus simplement: le signe linguistique est arbitraire."<sup>5</sup> Les mots ne portent en eux-mêmes aucun élément de sens. L'être humain lorsqu'il communique se réfère continuellement à un amas d'informations d'origine culturelle. A chaque mot correspond une "image", une définition ou un quelque chose qui git dans la mémoire de chaque individu. Mais le langage ne se limite pas qu'à cela. Un message pour être compris doit être formulé selon des règles de syntaxe précises. L'organisation des formes a une importance capitale. "Pierre tue Marie" est différent de "Marie tue Pierre" et pourtant nous avons dans les deux cas exactement les mêmes mots.

En s'attardant aux problèmes du langage ceux-ci semblent se multiplier. Que faire, par exemple, avec les ambiguïtés? Cette question en est une de taille. On pourrait même dire qu'elle pointe du doigt le point central de toute démarche dans les travaux sur la compréhension automatique du langage. *La* est un article dans "la porte" mais un pronom dans "il la porte". On ne peut pas passer

<sup>4</sup>F. David Peat, Artificial intelligence: How machines think, Baen Enterprises, New York, 1985, p.2.

<sup>5</sup>F. de Saussure, Cours de linguistique générale, Payot, Paris, 1983, p. 100.

outre ces réalités. Dans le premier cas l'article indique qu'il s'agit d'un substantif, d'une porte; pas de n'importe quelle porte mais une porte en particulier. Dans le second cas le pronom représente quelqu'un ou quelque chose qui n'est pas explicitement nommé, qui se situe dans un contexte plus large et qu'il conviendrait que l'ordinateur recherche. *Porte* n'est plus un substantif mais un verbe. Un verbe qui donne, entre autre, une indication sur la position temporelle de l'action. Bien sûr il y a des façons, des "trucs", afin de régler ces types d'ambiguïtés. La présence du pronom "il" nous montre que *porte* est un verbe donc que *la* est un pronom. Cependant, dans "Jean la porte" il n'y a plus de pronom. Il s'agirait que l'ordinateur puisse reconnaître en *Jean* un nom propre. La présence de la majuscule est une indication importante, mais que dire si *Jean* est le premier mot de la phrase et de plus s'il s'agissait plutôt de *Pierre* (ou de Pierre La porte)? Nous devinons bien tout le travail de réflexion qui s'amorce!

## 1.2.2 Description des travaux les plus marquants

### 1.2.2.1 La période pré-sémantique

Depuis le début des années soixante les travaux sur la compréhension automatique des langues naturelles sont nombreux et l'ensemble représente un apport considérable. De 1960 à 1970 il y a eu ce que Alain Bonnet appelle la période "pré-sémantique". Elle regroupe les approches qui ne cherchent pas "une "signification" de la phrase, c'est-à-dire une représentation symbolique en correspondance avec des objets du monde réel"<sup>6</sup>. Deux méthodes sont à retenir.

La première utilise des notions des grammaires transformationnelles: elle génère des phrases par transformation jusqu'à ce que la structure obtenue corresponde à celle de la phrase objet. Les résultats de cette méthode sont qualitativement intéressants mais le temps très long que nécessite l'analyse d'une phrase qui demanderait un nombre considérable de transformations la rend insatisfaisante.

La seconde méthode consiste à répertorier dans un dictionnaire une série de mots-clefs auxquels sont associés des parcelles de dialogue. Elle permet à un ordinateur de soutenir une conversation non-directive avec un interlocuteur. La procédure fonctionne de la façon suivante. Si une personne transmet à l'ordinateur la phrase "J'ai rencontré ma mère en fin de semaine" ce dernier cherchera chacun des mots dans son dictionnaire. Il trouvera *mère* auquel est associé la séquence "Comment va votre mère?" qu'il retransmettra en guise de réponse. L'ennui avec un tel programme c'est qu'à une phrase du type "La mère du prince Charles vient à Paris demain"<sup>7</sup> l'ordinateur répondrait "Parlez-moi de vos parents". On ne peut guère parler ici de compréhension. Il ne s'agit ni plus ni moins que d'un "jeu".

### 1.2.2.2 La période sémantico-syntaxique

À partir de 1970 a commencé la période "sémantico-syntaxique". Elle "reflète la fin des espoirs de représenter une grammaire de tout le langage. Les chercheurs se rabattent sur des syntaxes limitées, fortement reliées à des tests sémantiques"<sup>8</sup>. Les travaux les plus importants ont été faits

<sup>6</sup> Alain Bonnet, op. cit., p.38-39.

<sup>7</sup> Alain Bonnet, Op. cit., p.40.

<sup>8</sup> Ibid., p.42.

par Terry Winograd - avec SHRDLU, William Woods - avec la technique des ATNs (Augmented transition network, ou réseaux de transition augmentés) - et Roger Schank - avec la théorie de la dépendance conceptuelle.

SHRDLU est un programme qui simule le bras d'un robot qui reconnaît et déplace des formes variées. L'ordinateur est capable de "comprendre" les indications (données en langage naturel) d'un utilisateur et de transmettre des réponses. Voici un exemple d'un dialogue entre SHRDLU et un usager (Les phrases en majuscules sont celles de l'ordinateur):

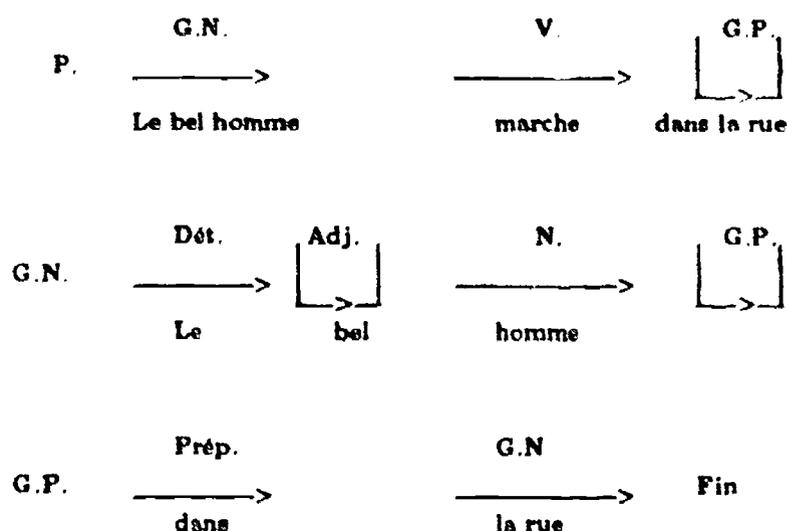
- Attrape un gros bloc rouge.
- O.K.
- Saisis la pyramide.
- JE NE COMPRENDS PAS DE QUELLE PYRAMIDE IL S'AGIT.
- Cherche un bloc plus grand que celui que tu tiens et place-le dans la boîte.
- JE PRÉSUME QUE PAR "LE" VOUS VOULEZ DIRE CELUI QUI EST PLUS GRAND QUE CELUI QUE JE TIENS.
- O.K.
- Que contient cette boîte?
- LA PYRAMIDE BLEUE ET LE BLOC BLEU.
- Sur quoi est posée la pyramide?
- SUR LA BOÎTE.
- La table peut-elle saisir des blocs?
- NON.
- Une pyramide peut-elle être posée sur un bloc?
- OUI.
- Une pyramide peut-elle être posée sur une pyramide?
- JE NE SAIS PAS.
- Empile deux pyramides.
- JE NE PEUX PAS (Le robot a essayé mais n'a pas réussi.)<sup>9</sup>

La réalisation de Winograd est très impressionnante mais elle consiste en ce que Hubert Dreyfus appelle "une manipulation des micromondes". Elle est inefficace ailleurs que dans le monde pour lequel Winograd l'a conçue. Toutefois elle montre une évolution très marquée dans la recherche sur la compréhension automatique.

Les ATNs, qui sont une amélioration des RTNs (Recursive Transition Network, ou réseaux de transition récursifs), "sont un moyen de spécifier une grammaire. Ils sont issus des automates d'état fini avec quelques adjonctions indispensables pouvant prendre en compte l'aspect récursif de certaines définitions"<sup>10</sup>. Un automate d'état fini permet de reconnaître la structure syntaxique d'une phrase objet. Il s'agit d'une procédure analytique de réduction des possibles. Par exemple si nous considérons la phrase "Le bel homme marche dans la rue" la procédure analyse d'abord *le*: il peut être un article ou un pronom. Les deux possibilités sont mises de côté et le premier élément est marqué comme n'étant pas encore précisé. La procédure analyse ensuite *bel* en tenant compte que l'élément précédent est un article ou un pronom. *Bel* ne peut pas être autre chose qu'un adjectif et un pronom ne peut pas le précéder. La procédure le marque comme étant définitivement un adjectif et retourne en arrière pour préciser que *le* est un article. A la fin de la procédure, on obtient un schéma syntaxique qui ressemble à ceci:

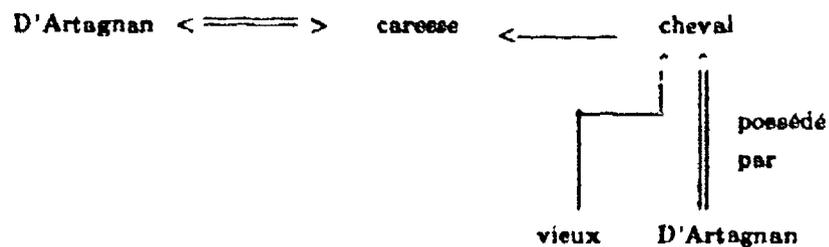
<sup>9</sup>Terry Winograd, "Understanding Natural Language", Cognitive Psychology n.3, Academic Press, New York, 1972, p.8-11.

<sup>10</sup>Alain Bonnet, op. cit., p. 52.



Les A.T.N., contrairement aux R.T.N., prévoient la possibilité de passer par dessus une étape. Par exemple si le programme avait prévu la présence d'un groupe nominal avant un verbe et qu'il n'y en a pas dans la phrase objet alors il peut passer directement à l'analyse du verbe. L'analyse n'est donc pas neutralisée par les imprévisibilités du langage.

La théorie de dépendance conceptuelle de Roger Schank permet d'établir des relations entre les différents éléments sémantiques dans une phrase. Par exemple la phrase "D'Artagnan caresse son vieux cheval":



- <====> : Double dépendance entre un nom et un verbe.
- <— : Dépendance à sens unique entre un nom et un verbe.
- <==== : Dépendance à sens unique entre deux noms.

Il y a double dépendance entre "D'Artagnan" et "caresse": l'actant et l'action ne prennent un sens que par leur union. Il y a une dépendance à sens unique entre "caresse" et "cheval": "cheval" pourrait être remplacé par n'importe quel autre nom équivalent (par exemple par "chien") sans que le sens global de la phrase ne soit modifié. Il y a dépendance à sens unique entre "cheval" et "vieux": on pourrait remplacer "vieux" par "jeune". Finalement il y a dépendance à sens unique entre deux noms, "cheval" et "D'Artagnan": ici encore le rapport est à sens unique parce que "cheval" peut être remplacé par un autre nom équivalent, mais de plus il y a un rapport d'appartenance qui est précisé par l'adjectif possessif "son". Cette démarche a pour but d'obtenir la même représentation conceptuelle pour deux phrases qui auraient la même signification mais des structures syntaxiques différentes.

### 1.2.2.3 L'analyse extra-phrastique

L'éclatement de la phrase en tant qu'unité d'analyse ne s'est fait véritablement pour la première fois qu'avec Propp et Rumelhart. Leurs travaux (qui portaient sur 100 contes merveilleux russes) consistaient à extraire des textes une structure générale. A partir d'une étude qui portait sur un très grand nombre de contes folkloriques Propp a introduit la notion de "morphologie du conte". Il démontra l'existence de règles précises desquelles dépendent les différentes structures de ces textes.

Personne n'a pensé à la possibilité de la notion et du terme de morphologie du conte. Dans le domaine du conte populaire, folklorique, l'étude des formes et l'établissement des lois qui régissent la structure est pourtant possible, avec autant de précision que la morphologie des formations organiques.<sup>11</sup>

Ces règles permettaient d'envisager la possibilité d'automatiser une analyse du conte. Cependant, les travaux de Propp et de Rumelhart ne pouvaient pas servir directement. Ils ne répondaient pas d'une façon satisfaisante aux exigences de l'informatique:

D'une part, elle (la méthode de Propp) est limitée à des structures très simples; d'autre part, il est plus facile de reconnaître ce qui satisfait une catégorie <groupe verbal> que la portion d'histoire satisfaisant <morale>; il n'y a pas eu (à ma connaissance) de programme fonctionnant sur ces principes.<sup>12</sup>

Toutefois, ces travaux présentent une nouvelle dimension dans la recherche sur la compréhension automatique du langage naturel. Ils démontrent la présence de règles structurales extra-phrastiques et la nécessité d'aborder un texte comme étant une entité plutôt qu'uniquement une suite de segments (de phrases).

Ce n'est qu'avec les travaux de Roger Schank et Robert Abelson que commencèrent les véritables pas (quoique très petits) vers la compréhension automatique du texte. Leur méthode consiste à répertorier un grand nombre de scripts (script *aller au restaurant*, script *accident* etc...) qui "décrivent des relations causales entre les différents épisodes et permettent de faire des inférences, de deviner "l'implicite" contenu dans la plupart des histoires, en quelque sorte de "remplir les blancs" comme fait tout être humain, avec d'ailleurs les mêmes risques d'erreurs"<sup>13</sup>. L'ordinateur peut alors faire ressortir d'un texte des éléments non-explicitement nommés. Il peut aussi faire un certain nombre de déductions à partir de règles qui accompagnent les scripts dans le répertoire. Voici un exemple du travail exécuté par SAM, un logiciel mis au point par Schank et ses collègues, qui utilise la méthode des scripts.

#### Histoire entrée

(Histoire tirée de SAM mais librement traduite et interprétée pour le contexte français).

Vendredi soir une automobile a quitté la Nationale 10 après avoir percuté un arbre. Le passager, David Farrain, demeurant en Normandie, a été tué sur le coup selon le

<sup>11</sup> Vladimir Propp, Morphologie du conte, Editions du Seuil, Paris, 1970, p.6.

<sup>12</sup> Alain Bonnet, op. cit., p.80.

<sup>13</sup> Ibid., p.80.

docteur Blanchard. Pierre Dutoit, le conducteur, demeurant 59, rue du Cherche-Midi à Paris a été conduit à l'hôpital Broussais par le SAMU. Après quelques jours, il a pu regagner son domicile. Aucune poursuite judiciaire n'a été engagée. L'inspecteur Dupuy a été chargé d'établir les circonstances de l'accident.

#### Résumé généré par le programme

Une automobile a percuté un arbre sur la Nationale 10 il y a quatre jours. David Parraïn, habitant en Normandie, le passager, est mort. Pierre Dutoit, le conducteur, habitant 59, rue du Cherche-Midi à Paris, a été légèrement blessé. La police n'a retenu aucune charge.

#### Explications

Le programme SAM utilise le script ACCIDENTS-AUTO ainsi que certains sous-scripts comme celui de l'AMBULANCE. C'est ce dernier qui lui permet notamment de déduire que Pierre Dutoit a été légèrement blessé, ce qui n'est pas dit dans l'histoire. Il établit aussi des relations causales notamment la relation entre "le véhicule a percuté l'arbre" et le fait que "Pierre Dutoit a été blessé" (inférence de bon sens).<sup>14</sup>

Cette méthode d'analyse, pour être efficace avec une très grande quantité de textes, demande un bon nombre de scripts. Cependant plus le nombre de scripts augmente plus il y a d'emboitements, de dépendances thématiques. Le réseau de relations entre les différents scripts devient tellement complexe à gérer qu'il en réduit considérablement l'efficacité de la méthode.

## 2. La syntaxe textuelle

Comme nous pouvons le constater les travaux sur la compréhension automatique du langage tendent vers l'analyse textuelle. Les spécialistes en intelligence artificielle sentent de plus en plus le besoin de faire éclater le cadre rigide de la phrase. Quoique l'analyse intra-phrastique restera toujours une étape essentielle dans la compréhension du langage, il faut considérer le texte comme une entité qui a ses propres règles et mécanismes. Harald Weinrich définit la notion de texte de la façon suivante:

Un texte est une succession signifiante de signes linguistiques entre deux ruptures manifestes de communication. Seront considérées comme "manifestes" les pauses de respiration ou celles qui traduisent la recherche des mots. Dans la communication écrite, ce seront, par exemple, les deux volets de la couverture d'un livre. Ce seront aussi ces coupes, délibérément introduites, et qui, dans un sens quasi-métalinguistique, ménagent des ruptures manifestes dans la communication. C'est pourquoi on peut avoir des textes très courts, en dehors même d'une situation "parlante" aidant aux délimitations. La limite inférieure est de deux monèmes. Il n'y a pas de limite supérieure<sup>15</sup>.

<sup>14</sup> Alain Bonnet, *op. cit.*, p.82.

<sup>15</sup> Harald Weinrich, *Le temps*, Editions du Seuil, Paris, 1973, p.13.

Issue de la linguistique structurale, "qui - sur la base du Cours de linguistique générale de Ferdinand de Saussure - étudie les signes de la langue d'après leur valeur positionnelle dans de larges ensembles structuraux"<sup>16</sup>, la syntaxe textuelle est l'étude des signes qui prennent un sens particulier par leur récurrence à travers un texte. Parmi ces signes, les formes verbales sont celles qui occupent quantitativement le premier rang. Il y a plus d'un verbe par ligne de texte en moyenne, en français.

### 2.1 *Monde raconté et monde commenté*

Harald Weinrich, dans son ouvrage *Le temps*, divise les formes temporelles en deux groupes: le monde commenté et le monde raconté. Le premier comprend le présent, le passé composé et le futur. Les textes où dominent ces temps sont dits commentatifs. Ils font savoir à l'interlocuteur que le texte mérite de sa part une attention vigilante. L'attitude de locution est plus tendue. La distance entre locuteur, auditeur et histoire est réduite au maximum. La barrière, érigée par le récit, qui fait en sorte que l'auditeur prenne une attitude d'écoute détachée parce qu'il sait bien qu'il s'agit de fiction, est brisée. Avec le présent de l'indicatif, le passé composé ou le futur simple le narrateur ne raconte plus, mais commente, donne son opinion, s'adresse directement à l'auditeur. Ce dernier ne peut plus jouer le rôle d'observateur invisible. En quelque sorte, il se fait interpeler et doit réagir, ce qu'il fait en adoptant une attitude d'écoute plus tendue. Le second groupe comprend le passé simple, l'imparfait, le plus-que-parfait et le conditionnel. Les textes où dominent ces temps sont dits narratifs. Ils requièrent de l'interlocuteur une écoute plus détachée, moins tendue. La distance entre locuteur, auditeur et histoire est plus grande: "L'une de nos attitudes en face du monde est de le raconter. Nous utilisons alors des signaux linguistiques à valeur narrative, en particulier les temps du récit. Ils ont pour fonction d'avertir l'auditeur que cet énoncé est "seulement" un récit et qu'il peut l'écouter avec un certain détachement"<sup>17</sup>

### 2.2 *Attitude de locution*

À l'intérieur de chacun des deux groupes de verbes on peut faire une nouvelle catégorisation. Dans le monde commenté "le présent est un temps, le plus fréquent des temps commentatifs; il caractérise donc une certaine attitude de locution"<sup>18</sup>. Habituellement, une histoire se raconte au passé simple et à l'imparfait alors qu'on utilise le présent pour la résumer. Le présent s'imbrique dans une situation commentative plus vaste. Il faut éviter de prendre les variations d'ordre temporelle pour des références à telles ou telles époques. L'emploi du Présent dans le résumé n'est que le signal d'une situation: ceci est un résumé, ceci est un commentaire etc... Le passé composé et le futur ont les mêmes fonctions que le présent. Ce qui les distingue c'est leur caractéristique à marquer l'antériorité ou la postériorité.

### 2.3 *Premier plan et arrière-plan*

À l'intérieur du monde raconté Weinrich distingue deux groupes de verbes. Le premier est constitué de l'imparfait et du plus-que-parfait. Ce sont les temps des circonstances secondaires, des descriptions, des réflexions et de tout ce que l'auteur désire repousser à l'arrière-plan. Le second est

<sup>16</sup>Ibid., p.12.

<sup>17</sup>Harald Weinrich, *op. cit.*, p. 33-34.

<sup>18</sup>Harald Weinrich, *op. cit.*, p.44.

constitué du passé simple, qui est le temps de l'action, du premier plan. En observant le jeu des transitions temporelles on peut constater des effets de mise en relief, une modulation du tempo narratif voire même un changement abrupt du débit narratif.

#### *2.4 L'analyse du comportement verbale et la compréhension automatique des langues naturelles*

Incorporer les règles issues de l'observation du comportement des formes verbales dans un algorithme de compréhension du langage c'est donner à l'ordinateur le moyen de reconnaître dans un texte les passages les plus importants (qui demandent par conséquent une analyse plus minutieuse) et les autres (qu'il pourrait jusqu'à un certain point mettre en veilleuse). C'est aussi lui donner des instruments qui lui permettent de générer des résumés, d'extraire les descriptions, les commentaires voire même de déterminer "l'atmosphère" dans laquelle se déroule l'action.

#### *2.5 L'automatisation de la méthode d'analyse de Weinrich*

Le traitement des formes verbales du point de vue de la syntaxe textuelle peut se faire à deux niveaux. Le premier est celui de l'intelligence artificielle. Il s'agirait de concevoir un logiciel ou un algorithme qui reconnaîtrait tous les verbes d'un texte donné et qui procéderait à un travail d'analyse afin de présenter un certain nombre de conclusions du type: "ce texte est très commentatif", "ce texte comporte un grand nombre de descriptions", etc. En fait le logiciel devrait être en mesure d'établir des correspondances avec le monde réel.

Le second, moins ambitieux, est celui de la linguistique automatique. Un logiciel ou un algorithme d'analyse des formes verbales du point de vue de la syntaxe textuelle n'aurait pas à tirer de conclusion. Il n'aurait par exemple qu'à reconnaître les verbes d'un texte et à transmettre des informations d'ordre statistique à partir desquelles un utilisateur pourrait tirer ses propres conclusions.

## CHAPITRE 1

### LINGUISTIQUE AUTOMATIQUE

#### 1. Qu'est-ce que la linguistique automatique?

L'expression "linguistique automatique" a donné lieu à d'interminables discussions. Certains préféreraient le terme "automatique linguistique", d'autres "mécano-linguistique". Nous n'avons pas l'intention de relancer la controverse: nous acceptons donc l'expression "linguistique automatique". L'important de toute évidence ne réside pas dans la combinaison de graphies qui nomment une discipline mais plutôt dans ce qui est décrit comme faisant partie d'une discipline. Fait partie de la linguistique automatique tout travail ou étude qui porte sur la langue, qui requiert des connaissances en linguistique, et qui nécessite l'utilisation d'un ordinateur ou la considération des règles qu'imposent l'utilisation d'un ordinateur. "Pour qu'il puisse être question de linguistique automatique il suffit et il est nécessaire à nos yeux qu'un ordinateur ait été impliqué dans le traitement de données linguistiques. La part de l'ordinateur dans la recherche et la façon dont il a été utilisé n'importent pas; le domaine linguistique dans lequel la recherche est ancrée et les intentions du chercheur ne nous semblent pas davantage des arguments d'exclusion ou d'inclusion."<sup>19</sup>

#### 1.1 Littérature et linguistique automatique

Un travail d'automatisation en littérature doit-il être situé en littérature ou en linguistique? Certes en littérature mais faute d'existence d'une sous-discipline en littérature où nous pourrions situer nos travaux nous nous voyons enclin à parler de linguistique automatique (notre but n'étant pas de créer une nouvelle discipline). Cela ne nous gêne nullement puisque la description des travaux dits de linguistique automatique fait par de Kock est suffisamment large pour nous plaire:

Depuis vingt ans la langue est traitée par l'ordinateur en vue de diverses exploitations pratiques telles que la traduction et la documentation, littéraires telles que l'authentification d'auteur et la confection de concordances pour des investigations thématiques ou stylistiques, ou voisines de la linguistique, comme l'établissement d'un vocabulaire fondamental, d'intérêt pédagogique, etc.<sup>20</sup>

Pour de Kock les travaux qui n'ont pas une finalité proprement linguistique supposent tout au moins une étude linguistique.

---

<sup>19</sup>Josse de Kock, *op. cit.*, p.14.

<sup>20</sup>*Ibid.*, p.17.

### 1.2 *Intelligence artificielle et linguistique automatique*

La linguistique automatique est à l'intelligence artificielle ce que la menuiserie est à l'architecture. En linguistique automatique les applications sont plus pratiques et ne requièrent pas de réflexion sur les mécanismes cognitifs de l'être humain. En général les linguistes n'ont pas à se préoccuper de l'intégration éventuelle de leurs travaux dans un projet plus vaste de reconstitution du comportement mental de l'être humain.

Certaines activités comme la traduction seront abordées à la fois par l'intelligence artificielle et la linguistique automatique et c'est en étudiant les deux approches que nous pouvons mieux comprendre les différences entre les deux disciplines. D'abord, un projet de traduction automatique entrepris par des chercheurs en intelligence artificielle comporte trois aspects essentiels. Le premier consiste en la compréhension d'un message dans une langue, le second en la reconstitution morpho-syntaxique d'un message dans une autre langue et le troisième en l'obtention d'un résultat semblable à celui qu'un être humain obtiendrait.

La linguistique automatique jouit d'une plus grande liberté. D'abord, elle n'exige pas une étape de compréhension automatique. Elle tente de faire le passage directement d'une langue à l'autre. Ensuite, elle n'exige pas un résultat de la qualité de celui qu'obtiendrait un traducteur humain. Elle comble les lacunes en incorporant une étape d'intervention humaine. En linguistique automatique il n'est pas vraiment question de traduction automatique mais plutôt de traduction assistée par ordinateur. Le type de résultat qu'on peut espérer à partir d'une méthode de travail comme celle-ci est représenté par le logiciel MACROCAT de traduction assistée par ordinateur qui a été développé par Weidner Communications.

Aux dires de TAO Canada, qui distribue les deux systèmes, ces derniers permettent aisément de traduire entre 3000 et 4000 mots par jour comparativement à la moyenne de 1200 à 1500 mots par jour qui est la norme à Montréal. Autrement dit, la productivité du traducteur est multipliée par un facteur de deux à trois ce qui est loin d'être négligeable.<sup>21</sup>

La procédure de traduction que propose le logiciel est constituée de six étapes. Dans la première il s'agit simplement de faire la saisie du texte à traduire à l'aide d'un logiciel de traitement de texte. Dans la seconde l'ordinateur, en consultant son dictionnaire, détecte tous les mots du texte qu'il ne connaît pas. Dans la troisième, l'utilisateur incorpore les mots de l'étape précédente dans le dictionnaire. Dans la quatrième il s'agit de la traduction proprement dite. Le résultat obtenu est dit une "traduction brute" et son style est très sec. Dans la cinquième étape l'utilisateur reprend le travail effectué par l'ordinateur et à l'aide du logiciel de traitement de texte il récrit le texte dans une forme plus stylisée et corrige les erreurs que le système a pu commettre. Finalement la dernière étape consiste à acheminer le résultat final sur un support approprié (papier ou fichier).

### 1.3 *L'analyse des temps, intelligence artificielle et linguistique*

Certains travaux se situent à la limite de l'intelligence artificielle et de la linguistique automatique. C'est le cas en ce qui nous concerne. Du point de vue de l'intelligence artificielle, nous l'avons vu, nos travaux peuvent être incorporés à un projet plus vaste de compréhension automatique des langues naturelles. Ils constitueraient en quelque sorte un instrument d'aiguillage de certains mécanismes de compréhension automatique. L'automatisation d'une analyse de récurrence des formes

<sup>21</sup> Gil Tocco, *Logiciels de Micro-Informatique*, "Informatique et bureautique", juillet-août 1984, p.16.

verbaux peut fournir à l'ordinateur des informations du type "ceci est un passage important", "ceci est un passage secondaire", "ceci est un commentaire", "ceci est une description" etc. Toutefois nous nous situons, quant à notre méthode et à nos objectifs, du côté de la linguistique automatique. Nous ne cherchons pas à mettre au point un algorithme produisant une automatisation complète de la méthode d'analyse ni à faire générer une représentation des données ayant un rapport avec des objets du monde réel.

## 2. L'ordinateur

Automatiser une analyse de la récurrence des formes verbales est une tâche qui comporte un grand nombre de difficultés. Cela demande d'abord des connaissances en syntaxe textuelle, en linguistique et bien entendu en informatique. Il faut être en mesure de manipuler un ordinateur, donc de connaître son fonctionnement externe, mais aussi son fonctionnement interne.

### 2.1 Manipulation de formes simples

D'abord, contrairement à ce que beaucoup de gens croient, l'ordinateur ne peut rien faire d'autre que manipuler des formes simples. Son "cerveau" n'est rien de plus que la multiplication d'un élément de base très simple: la cellule informatique. Celle-ci est constituée d'un groupe de bits (généralement huit ou seize, selon la conception des ordinateurs), qui peuvent prendre chacune deux aspects: positif ou négatif. En traduisant les chiffres, de 0 à 9, et les lettres alphabétiques, de a à z, en langage binaire, il est possible, en juxtaposant une série de cellules, d'inscrire un nombre complexe ou un mot, voire un texte, dans un ordinateur. Afin d'éviter à l'utilisateur d'écrire en langage binaire, ce qu'on appelle plus communément le langage machine, les concepteurs d'ordinateurs ont joué d'astuce. A un signal électrique provenant d'un endroit particulier (une des touches du clavier par exemple, correspond un code binaire particulier. Par exemple, à la lettre "A" correspond le code "1100 0001", qui se traduit, à l'intérieur d'une cellule (ou mot), par: bit négative, bit négative, bit positive, bit positive, bit positive, bit positive, bit positive, bit négative. L'affichage sur un écran, ou l'impression sur papier, procède sensiblement de la même manière. A un code binaire particulier correspond une organisation de points lumineux ou de points d'encre.

Nous avons évidemment simplifié à l'extrême la structure de l'ordinateur, mais toujours en respectant son principe de base. Ajoutons maintenant qu'un programme informatique n'est que la combinaison des opérations de base, elles aussi très simples. Prenons par exemple le cas d'une soustraction mathématique. Imaginons que nous voulons soustraire 5 de 7 (Les ordinateurs modernes ont des circuits qui font l'arithmétique binaire directement, sans décomposition. Nous avons choisi la description suivante dans le but de schématiser avec des termes simples la logique informatique). Il y aura d'abord une cellule qui prendra la valeur 5, nous allons l'appeler A, et une autre qui prendra la valeur 0, A'. Il y en aura une troisième qui prendra la valeur 7, B. L'opération consiste à faire prendre à la cellule B la valeur qui la précède, soit 6 dans ce cas-ci, et à la cellule A' la valeur qui la suit, soit 1. L'opération est répétée jusqu'à ce que la cellule A' prenne la même valeur que la cellule A. Le résultat donnera 2 comme valeur finale de B. L'addition procède sensiblement de la même façon. La multiplication et la division ne sont qu'une succession d'additions.

### 2.2 Reconnaissance des formes

Il est évident que l'ordinateur n'est pas apte à comprendre, mais plutôt à reconnaître. Dans le traitement informatique de la langue, l'opération de base consiste uniquement à comparer un groupe de cellules avec un autre et à en tirer une conclusion immédiate. Prenons l'exemple d'une méthode

de traduction automatique simple. Dans une première étape, il s'agit de structurer un ensemble de cellules sous la forme d'un dictionnaire. Celui-ci est constitué d'un nombre élevé d'enregistrements qui sont divisés chacun en deux secteurs (ou champs). Dans le premier, on place un mot. Dans le second, le mot équivalent dans une autre langue. Dans une deuxième étape, nous inscrivons les mots à traduire dans un fichier identique au premier, mais sans la traduction. L'opération consiste à prendre, tour à tour, les mots à traduire et à les comparer avec chacun des mots qui sont dans le dictionnaire. Lorsque le mot désiré a été trouvé, l'ordinateur transcrit la traduction dans le fichier des mots à traduire.

### 2.3 Les langages de programmation

Les langages de programmation mettent à notre disposition une série d'instructions complexes qui comportent chacune un sous-ensemble d'instructions simples. Par exemple, en BASIC, si nous voulons soustraire 5 de 7 il s'agit uniquement de transmettre à l'ordinateur l'instruction  $A = 7 - 5$ . L'interpréteur BASIC fera correspondre à notre instruction complexe la série d'instructions simples que nous avons décrite plus haut.

Les langages de programmation sont nombreux et possèdent chacun des caractéristiques précises. On peut les diviser en deux catégories générales. La première comprend les langages dit interprétés et la seconde les langages dit compilés. Avec un langage de programmation interprété les instructions sont lues, traduites et exécutées une à une. Si une instruction doit être exécutée 100 fois alors il y a 100 conversions en langage machine pour cette même instruction. Avec un langage compilé, le programme en entier est converti en langage machine avant son exécution. Il n'y a donc pas plus de traduction qu'il y a d'instructions. Le choix d'un programmeur quant au type de langage qu'il utilisera sera fait en considération du type de travail à faire. Un programme court qui comporte peu de répétitions sera plus aisément préparé avec un interpréteur, d'autant plus que la plupart des interpréteurs permettent à l'utilisateur de travailler en "mode direct" c'est-à-dire qu'il permettent à l'utilisateur de faire exécuter les instructions une à une d'un programme en voie de réalisation. L'utilisation d'un compilateur sera plus pertinente dans le cas d'un programme plus complexe qui comporte un grand nombre de répétitions. Puisque notre projet comporte un travail de programmation considérable, qui comprend un grand nombre de blocs d'instructions qui seront répétés plusieurs fois nous avons donc choisi un langage compilé.

Parmi les langages compilés il y a une grande variété. Certains ont été conçus surtout pour des travaux de programmation en administration, certains pour les sciences pures et d'autres pour le traitement des chaînes de caractères. Notre choix s'est arrêté sur le PASCAL, qui a été conçu pour le traitement des chaînes de caractères. Trois caractéristiques en particulier nous ont fait adopter ce langage. D'abord la possibilité de définir des ensembles à partir d'éléments non ordonnés. Par exemple, il nous a été possible de définir l'ensemble SÉPARATEURS qui est constitué des signes formels permettant de considérer les mots comme des unités dans un texte: l'espace, le point, la virgule, le point virgule etc. Nous avons également pu définir l'ensemble TEMPS qui est constitué des différents temps de verbe: présent, imparfait, futur, conditionnel etc. Ensuite la possibilité de rendre une procédure récurrente. Une procédure récurrente se caractérise par son aptitude à commander sa propre exécution. On peut ainsi par exemple avec un minimum d'instructions procéder à l'exploration d'une banque de données, structurée en forme arborescente, de grandeur variable. Finalement la possibilité de structurer les enregistrements d'un fichier avec une grande souplesse. Il est possible par exemple en PASCAL d'avoir une matrice en trois dimensions (ou plus, ou moins) dans un champ, un tableau dans un autre, un vecteur dans un autre et ainsi de suite.

Plusieurs versions du langage PASCAL sont disponibles sur le marché. Il y a entre autre le PASCAL d'ISO, le PASCAL de l'UCSD et le Turbo Pascal. Nous avons opté pour le Turbo Pascal, qui est distribué par la compagnie Borland. Ce dernier met à la disposition de l'utilisateur un éditeur qui comprend un grand nombre de facilités, un compilateur très rapide et un module d'aide à la mise au point de programmes qui est très efficace.

### 3. La reconnaissance automatique des formes verbales dans un texte

#### 3.1 *L'homographie*

Avant d'entreprendre un travail comme le nôtre il convient de faire une réflexion sur les ambiguïtés. Comme nous l'avons vu précédemment, elle est la plus grande source de difficultés dans le traitement des langues naturelles. Un ordinateur n'est apte qu'à reconnaître les formes. Dans le cas de deux homographes il n'est pas en mesure d'établir une distinction à moins de pouvoir jouir d'un supplément d'information. Dans certaines situations il sera possible de le lui fournir dans d'autres ce sera impossible. Dans "Pierre fit une marche" la présence de l'article **une** peut indiquer que **marche** est un substantif. Dans "Il marche dans la rue" la présence du pronom personnel **Il** peut servir à préciser que **marche** est un verbe. Par contre dans "Il dit "viens ici"" rien ne peut servir à préciser le verbe **dire** qui peut être à la troisième personne du présent de l'indicatif ou à la troisième personne du passé simple.

##### 3.1.1 Le travail interactif

La façon de régler les problèmes persistant reliés à l'homographie est tout simplement de prévoir dans un algorithme une intervention humaine. Cette solution est fréquemment utilisée en linguistique automatique. Afin de minimiser le travail de l'utilisateur il s'agit de reporter cette étape à la toute fin de la procédure. Pour cela cependant il faut prévoir un registre de tous les cas que l'ordinateur n'a pas été en mesure de traiter.

#### 3.2 *Techniques de reconnaissance automatique des verbes*

Reconnaître les verbes dans un texte n'est pas une chose facile pour un ordinateur. Nous pouvons répertorier toutes les terminaisons de verbes possibles et lui demander de détecter tous les mots dont la partie finale correspond à l'une de ces terminaisons. Cette méthode aurait une certaine efficacité en ce sens que toutes les formes verbales seraient détectées mais elle comporterait l'énorme inconvénient de prendre pour des verbes tous les mots qui ne sont pas des verbes et qui ont des terminaisons homographiques avec des verbes; c'est-à-dire qui finissent par exemple par -e, -es ou -ais. C'est le cas avec **Jamais**. Une autre méthode consisterait à répertorier tous les verbes conjugués. Celle-ci aurait un rendement satisfaisant mais nécessiterait un dictionnaire tellement volumineux qu'il faudrait un temps fou à l'ordinateur pour traiter un texte de volume considérable.

#### 3.3 *La méthode de Maegaard et Spang-Hanssen*

Bente Maegaard et Ebbe Spang-Hanssen dans leur ouvrage qui s'intitule *La segmentation automatique du français écrit* propose, dans un projet plus vaste d'analyse syntaxique automatique, une méthode de reconnaissance automatique des formes verbales. La charpente de la procédure consiste en deux dictionnaires (un pour les terminaisons des verbes et l'autre pour les radicaux) et d'une table de tous les types d'ambiguïtés possibles.

##### 3.3.1 Les dictionnaires

###### 3.3.1.1 Le dictionnaire des terminaisons

Le découpage des verbes en terminaisons et radicaux a été fait de sorte que l'espace requis par les dictionnaires est réduit au maximum. Cela signifie qu'il y a le moins de répétitions possible. Les

temps et les personnes de verbes sont divisés en six classes (Maegaard et Spang-Hanssen utilise l'expression "paradigme"):

(Les chiffres entre parenthèses correspondent aux personnes.)

- 1ère classe: le futur (1-6)  
le conditionnel (1-6)
- 2ème classe: le présent (4-5)  
le présent du subjonctif (4-5)  
l'imparfait (1-6)  
l'impératif (4-5)  
le participe présent
- 3ème classe: le présent (1,2,3,6)  
le présent du subjonctif (1,2,3,6)  
l'impératif (2)
- 4ème classe: le participe passé
- 5ème classe: le passé simple (1-6)  
l'imparfait du subjonctif (1-6)
- 6ème classe: l'infinitif

Sont regroupés ensemble les formes qui présentent des similitudes dans les règles de conjugaison. Par exemple, pour le présent 4 et 5 nous avons un type de terminaison: -ons, -ez. Pour le présent du subjonctif 4 et 5 nous avons: -ions, -iez. Ces terminaisons sont acceptées par les radicaux march- et finiss- (des verbes marcher et finir) qui sont catégorisés dans deux groupes différents par la grammaire traditionnelle. Pour ces deux mêmes verbes les désinences du présent de l'indicatif et du subjonctif 1,2,3 et 6 diffèrent. Nous avons -e, -es, -e, -ent, -e, -es, -e, -ent pour le radical march- (du verbe marcher) et les désinences -s, -s, -t, -ssent, -sse, -sses, -sse, -ssent pour le radical fini- (du verbe finir). Il convient donc de regrouper le présent de l'indicatif 4-5 et le présent du subjonctif 4-5 mais de les dissocier du présent de l'indicatif 1,2,3,6 et du présent du subjonctif 1,2,3,6.

À l'intérieur de chaque classe il y aura des variations, qui donnent lieu à des conjugaisons différentes, dans le découpage des verbes. Par exemple pour le verbe manger, dans la 2ème classe, nous avons les désinences -eons et -ez alors que pour le verbe marcher nous avons -ons et -ez. Dans la première classe il y a 3 conjugaisons, dans la seconde 9, dans la troisième 33, dans la quatrième 9, dans la cinquième 8 et dans la sixième 6.

<u>Classe 1.</u>		<u>Conjugaisons</u>		
Temps	Personne	1	2	3
Futur	1	-erai	-irai	-rai
	2	-eras	-iras	-ras
	3	-era	-ira	-ra
	4	-erons	-irons	-rons

	5	-eres	-ires	-rez
	6	-eront	-iront	-ront
	1	-erais	-irais	-rais
	2	-erais	-irais	-rais
Conditionnel	3	-erait	-irait	-rait
	4	-erions	-irions	-rions
	5	-eries	-iriez	-riez
	6	-eraient	-iraient	-raient

Classe 2.

Temps	Personne	1	2	3	4
Présent de l'indicatif	4	-ons	-issons	-isons	-vons
	5	-es	-ites	-ites	-ves
Présent du subjonctif	4	-ions	-issions	-issions	-chions
	5	-iez	-issiez	-issiez	-chiez
Imparfait de l'indicatif	1	-ais	-issais	-issais	-vais
	2	-ais	-issais	-issais	-vais
	3	-ait	-issait	-issait	-vait
	4	-ions	-issions	-issions	-vions
	5	-iez	-issiez	-issiez	-viez
	6	-aient	-issaient	-issaient	-vaient
Impératif	4	-ons	-issons	-isons	-chons
	5	-es	-ites	-ites	-ches
Participe présent		-ant	-issant	-issant	-chant
		5	6	7	8
Présent de l'indicatif	4	-oulons	-vons	-sommes	-eons
	5	-oulez	-vez	-êtes	-es
Présent du subjonctif	4	-oulions	-yons	-soyons	-ions
	5	-ouliez	-yez	-soyez	-iez
Imparfait de l'indicatif	1	-oulais	-vais	-étais	-eais
	2	-oulais	-vais	-étais	-eais
	3	-oulait	-vait	-était	-eait
	4	-oulions	-vions	-étions	-ions
	5	-oulies	-viez	-étiez	-iez
	6	-oulaient	-vaient	-étaient	-eaient
Impératif	4	-euillons	-yons	-soyons	-eons
	5	-euillez	-yez	-soyez	-es
Participe présent		-oulant	-yant	-étant	-çant

Temps	Personne	
		9
Présent de l'indicatif	4	-çons
	5	-ces
Présent du subjonctif	4	-cions
	5	-cies
Imparfait de l'indicatif	1	-çais
	2	-çais
	3	-çait
	4	-cions
	5	-cies
	6	-çaient
Impératif	4	-çons
	5	-ces
Participe présent		-çant

Classe 3.

Temps	Personne	1	2	3	4
Présent de l'indicatif	1	-e	-s	-ds	-us
	2	-es	-s	-ds	-us
	3	-e	-t	-d	-ut
	6	-ent	-eent	-dent	-ivent
Présent du subjonctif	1	-e	-see	-de	-ive
	2	-es	-ssee	-des	-ives
	3	-e	-see	-de	-ive
	6	-ent	-eent	-dent	-ivent
Impératif	2	-e	-s	-ds	-us
Présent de l'indicatif		5	6	7	8
	1	-s	-vais	-ieds	-s
	2	-s	-vas	-ieds	-s
	3	-t	-va	-ied	-t
	6	-ent	-vont	-eyent	-vent
Présent du subjonctif	1	-e	-aille	-eye	-ve
	2	-es	-ailles	-eyes	-ves
	3	-e	-aille	-eye	-ve
	6	-ent	-aillent	-eyent	-vent
Impératif	2	-s	-va	-ieds	-s

Temps	Personne	Personne			
		9	10	11	12
Présent de l'indicatif	1	-s	-s	-is	-ds
	2	-s	-s	-is	-ds
	3	-t	-t	-it	-d
	6	-illent	-eent	-issent	-eent
Présent du subjonctif	1	-ille	-ee	-isse	-ee
	2	-illes	-ees	-isses	-ees
	3	-ille	-ee	-isse	-ee
	6	-illent	-eent	-issent	-eent
Impératif	2	-s	-s	-is	-ds
		13	14	15	16
Présent de l'indicatif	1	-ns	-is	-s	-ais
	2	-ns	-is	-s	-ais
	3	-nt	-it	-t	-ait
	6	-gnent	-issent	-ment	-ont
Présent du subjonctif	1	-gne	-isse	-me	-asse
	2	-gnes	-isses	-mes	-asses
	3	-gne	-isse	-me	-asse
	6	-gnent	-issent	-ment	-assent
Impératif	2	-ns	-is	-s	-ais
		17	18	19	20
Présent de l'indicatif	1	-us	-is	-s	-ds
	2	-us	-is	-s	-ds
	3	-ut	-it	-t	-d
	6	-illent	-issent	-tent	-ent
Présent du subjonctif	1	-ille	-isse	-te	-le
	2	-illes	-isses	-tes	-les
	3	-ille	-isse	-te	-le
	6	-illent	-issent	-tent	-lent
Impératif	2	-us	-is	-s	-ds
		21	22	23	24
Présent de l'indicatif	1	-is	-eux	-ds	-is
	2	-is	-eux	-ds	-is
	3	-it	-eut	-d	-it
	6	-issent	-euvent	-nent	-vent

		1	-ise	-uisse	-ne	-che
<b>Présent du subjunctif</b>		2	-ises	-uisses	-nes	-ches
		3	-ise	-uisse	-ne	-che
		6	-isent	-uissent	-nent	-chent
	<b>Impératif</b>	2	-is	-eux	-da	-che
		25	26	27	28	
<b>Présent de l'indicatif</b>		1	-s	-ux	-ts	-ts
		2	-s	-ux	-ts	-ts
		3	-t	-ut	-t	-t
		6	-nent	-lent	-tent	-tent
<b>Présent du subjunctif</b>		1	-ne	-ille	-tte	-te
		2	-nes	-illes	-ttes	-tes
		3	-ne	-ille	-tte	-te
		6	-nent	-illent	-tent	-tent
<b>Impératif</b>	2	-s	-ux	-ts	-ts	
		29	30	31	32	
<b>Présent de l'indicatif</b>		1	-x	-uis	-ai	-suis
		2	-x	-eux	-as	-es
		3	-t	-eut	-a	-est
		6	-lent	-euvent	-ont	-sont
<b>Présent du subjunctif</b>		1	-ille	-uisse	-aie	-sois
		2	-illes	-uisses	-aies	-sois
		3	-ille	-uisse	-ait	-soit
		6	-illent	-uissent	-aient	-soient
<b>Impératif</b>	2	-ille	-eux	-aie	-sois	
		33				
<b>Présent de l'indicatif</b>		1	-cs			
		2	-cs			
		3	-c			
		6	-quent			
<b>Présent du subjunctif</b>		1	-que			
		2	-ques			
		3	-que			
		6	-quent			
<b>Impératif</b>	2	-cs				

Classe 4.

Dans ce paradigme, celui du participe passé, les chiffres de 1 à 4 n'indiquent pas la personne, mais le genre et le nombre, selon les conventions suivantes: 1 = masculin singulier, 2 = masculin pluriel, 3 = féminin singulier, 4 = féminin pluriel.

Temps	Personne	1	2	3	4	5	6	7
Participe passé	1	-é	-i	-u	-û	-s	-s	-t
	2	-és	-is	-us	-us	-s	-s	-ts
	3	-ée	-ie	-ue	-ue	-te	-se	-te
	4	-ées	-ies	-ues	-ues	-tes	-ses	-tes
		8	9					
Participe passé	1	-û	-I					
	2	-ûs	-Is					
	3	-ûe	-Ye					
	4	-ûes	-Yes					

Classe 5.

Temps	Personne	1	2	3	4
Passé simple	1	-ai	-is	-us	-ûs
	2	-as	-is	-us	-ûs
	3	-a	-it	-ut	-ût
	4	-âmes	-îmes	ûmes	-ûmes
	5	-âtes	-îtes	-ûtes	-ûtes
	6	-èrent	-irent	-urent	-ûrent
Imparfait du subjonctif	1	-asse	-isse	-usse	-ûsse
	2	-asses	-isses	-usses	-ûsses
	3	-ât	-ît	-ût	-ût
	4	-assions	-issions	-ussions	-ûssions
	5	-assiez	-issiez	-ussiez	-ûssiez
	6	-assent	-issent	-ussent	-ûssent
		5	6	7	8
Passé simple	1	-Is	-ins	-eai	-çai
	2	-Is	-îne	-eas	-ças
	3	-ît	-int	-ea	-ça
	4	-îmes	-inmes	-eâmes	-çâmes
	5	-îtes	-intes	-eâtes	-çâtes
	6	-îrent	-inrent	-èrent	-cèrent
		1	2		
		-Isse	-inse	-easse	-çasse
		-Isses	-innes	-easses	-çasses

Imparfait du subjonctif	3	-It	-Int	-eât	-çât
	4	-Issions	-inssions	-eassions	-çassions
	5	-Issies	-inssies	-eassies	-çassies
	6	-Issent	-inssent	-eassent	-çassent

### Classe 6.

Temps	1	2	3	4	5	6
Infinitif	-er	-ir	-re	-oir	-eoir	-Ir

Maegaard et Spang-Hanssen ont répertorié plus de 588 désinences. À chacune d'elles correspond un code de quatre caractères. Le premier indique la classe à laquelle appartient la désinence, le second la conjugaison à l'intérieur de la classe, le troisième le temps et le quatrième la personne. Par exemple il y aura une désinence *a* à laquelle correspondra le code 5-01-g-3 (cinquième classe, première conjugaison, passé simple, troisième personne du singulier) et une autre désinence *a* à laquelle correspondra le code 3-31-c-3 (troisième classe, 31ème conjugaison, présent de l'indicatif, troisième personne du singulier).

### 3.3.1.2 Le dictionnaire des radicaux

Le dictionnaire des radicaux de Maegaard et Spang-Hanssen a été fait à partir d'une banque de 4200 verbes et comprend plus de 5400 formes. A chaque radical correspond un code de 7 chiffres. Le premier indique le numéro de la conjugaison utilisée dans la première classe (ce sera 0 si le radical ne prend aucune désinence dans une classe donnée), le second le numéro de la conjugaison utilisée dans la seconde classe et ainsi de suite. Le septième chiffre représente un code d'ambiguïté.

Puisque un radical peut servir à reconstituer plus d'une forme alors il faut prévoir la possibilité d'ambiguïtés multiples. Par exemple, pour le radical *cour-* il y a deux ambiguïtés. Il peut s'agir d'un verbe à la deuxième personne du présent de l'indicatif (*cours*) ou d'un substantif qui est lui-même ambigu quant à son genre et à son nombre car il peut être au masculin singulier, au masculin pluriel ou au féminin pluriel. Il pourrait s'agir aussi d'un verbe à la troisième personne du singulier du présent de l'indicatif ou d'un adjectif. Plutôt que de faire correspondre à chaque radical une série de codes d'ambiguïtés Maegaard et Spang-Hanssen ont préparé un répertoire des combinaisons d'ambiguïtés. Ils se sont aperçus qu'une racine verbale, dans leur analyse, ne génère jamais plus de trois formes verbales homographes. Ils ont dressé une liste sous forme de table comprenant 43 combinaisons qui comportent chacune un maximum de 3 codes. Un code comprend trois éléments d'informations: le temps et la personne de la forme verbale et l'alternative. La table qu'ils ont préparée ressemble à ceci:

(Pour le premier caractère des triplets: *c* = présent de l'indicatif, *g* = passé simple, *i* = participe passé. Les chiffres, qui viennent en deuxième, correspondent à la personne verbale. Le dernier caractère des triplets représente l'alternative: *a* = substantif masculin singulier, *d* = substantif masculin pluriel, *j* = participe passé en *-s* ou un adjectif, *b* = substantif féminin singulier, *e* = substantif féminin pluriel.)

1-	c3a	c2d	
2-	g1j	i2j	
3-	c3b	c2e	c4d
etc...			

Dans le dictionnaire des radicaux, en guise de code d'ambiguïté, sera inscrit le chiffre qui renvoie à une série de triplets. Si pour un radical nous déterminons que les formes homographes qu'il est susceptible de générer sont représentées par les triplets c3a et c2d alors sera inscrit le numéro 1 comme code d'ambiguïté. Si par contre la racine ne peut pas générer d'homographe alors un chiffre bidon sera inscrit.

### 3.3.2 La procédure de reconnaissance des verbes

La procédure de reconnaissance des verbes est constituée de deux étapes principales: la reconnaissance des formes verbales potentielles et la désambiguïsation de celles-ci. En un premier temps il y a reconnaissance de toutes les formes susceptibles d'être des verbes (cela inclut évidemment des mots comme "marche" prit au sens de "faire une marche"). La désambiguïsation de ces formes est faite sous forme de filtrage. Les formes verbales potentielles sont soumises à sept tests: le test des particules verbales, le test des substantifs, le test des participes passés en -s et des adjectifs, le test des participes passés en -t, le test du participe-substantif "fait", le test de la préposition "entre" et le test des locutions adverbiales en -on.

#### 3.3.2.1 Reconnaissance des formes verbales potentielles

La première étape de la procédure de Maegaard et Spang-Hanssen consiste à faire reconnaître par l'ordinateur les formes verbales potentielles. En consultant d'abord le dictionnaire des désinences l'ordinateur détecte les mots qui ont une terminaison de verbe. Lorsqu'un tel mot est trouvé alors l'ordinateur extrait la partie restante et consulte le dictionnaire des radicaux. Si le résultat de la recherche est négatif alors c'est qu'il ne s'agit pas d'un verbe. Si la recherche est positive l'ordinateur compare les codes qui accompagnent désinence et racine afin de déterminer s'il s'agit bien d'une forme verbale potentielle. Dans un cas comme *dent* l'ordinateur trouvera la désinence -ent et le radical *d-* (pour le verbe *dire*). L'analyse des codes indiquera que la désinence -ent appartient à la classe 3 conjugaison 1 et à la classe 3 conjugaison 5 tandis que le radical *d-* n'accepte que les terminaisons de la classe 4 conjugaison 3 ou 2: il n'y a donc pas compatibilité.

#### 3.3.2.2 Le test des particules verbales

Lorsque les formes verbales potentielles ont été reconnues il s'agit de les désambiguïser. La deuxième étape de la procédure consiste à faire passer aux formes verbales potentielles un test dit des particules verbales. Il s'agit de regarder le contexte immédiat de la forme et selon la présence de telle ou telle particule l'ordinateur pourra préciser s'il s'agit d'un verbe ou non.

Voici une description du test des particules verbales:

- 1- L'ordinateur lit le mot qui précède une forme verbale potentielle.
- 2- L'ordinateur vérifie si le mot est en:
  - 2.1- Si oui l'ordinateur vérifie si le mot qui précède en est y:
    - 2.1.1- Si oui, alors la forme verbale potentielle est catégorisée comme étant un verbe et le test est terminé.
    - 2.1.2- Si non, alors le test continue.
- 3- L'ordinateur vérifie si le mot est y:
  - 3.1- Si oui, alors la forme verbale potentielle est catégorisée comme étant un verbe et le test est terminé.
  - 3.2- Si non, alors le test continue.

- 4- L'ordinateur vérifie si le mot est *lui*, ou *leur*:
  - 4.1- Si oui, alors la forme verbale potentielle est catégorisée comme étant un verbe et le test est terminé.
  - 4.2- Si non, alors le test continue.
- 5- L'ordinateur vérifie si le mot est *l'*, *le*, *la* ou *les*:
  - 5.1- Si oui, alors l'ordinateur lit le mot qui précède:
    - 5.1.1- L'ordinateur vérifie s'il s'agit de *m'*, *me*, *t'*, *te*, *s'*, *se*, *nous*, ou *vous*:
      - 5.1.1.1- Si oui, alors la forme verbale potentielle est catégorisée comme étant un verbe et le test est terminé.
      - 5.1.1.2- Si non, alors le test continue.
- 6- L'ordinateur vérifie si le mot est *m'*, *me*, *t'*, *te*, *s'*, *se*, *nous*, ou *vous*:
  - 6.1- Si oui, alors la forme verbale potentielle est catégorisée comme étant un verbe et le test est terminé.
  - 6.2- Si non, alors le test continue.
- 7- L'ordinateur vérifie si le mot est *m'* ou *ne*:
  - 7.1- Si oui, alors la forme verbale est catégorisée comme étant un verbe et le test est terminé.
  - 7.2- Si non, alors le test continue.
- 8- L'ordinateur vérifie si le mot est *J'*, *Je*, *tu*, *il*, *elle*, *nous*, *vous*, *ils*, *elles*, *ce*, *c'*, *cela*, *ça* ou *on*:
  - 8.1- Si oui, alors l'ordinateur lit le mot qui précède:
    - 8.1.1- L'ordinateur vérifie s'il s'agit d'une préposition:
      - 8.1.1.1- Si oui, alors la forme verbale est catégorisée comme n'étant pas un verbe et le test est terminé.
      - 8.1.1.2- Si non, alors la forme verbale est catégorisée comme étant un verbe et le test est terminé.
  - 8.2- Si non, alors le test continue.
- 9- L'ordinateur lit ce qui suit la forme verbale potentielle:
- 10- L'ordinateur vérifie si ce qui suit la forme verbale potentielle est un tiret:
  - 10.1- Si oui, alors l'ordinateur lit ce qui suit le tiret:
    - 10.1.1- L'ordinateur vérifie s'il s'agit de *Je*, *tu*, *il*, *elle*, *nous*, *vous*, *ils*, *elles*, *ce* ou *on*:
      - 10.1.1.1- Si oui, alors la forme verbale potentielle est catégorisée comme étant un verbe et le test est terminé.
      - 10.1.1.2- Si non, alors le test continue.
- 11- Si la forme verbale potentielle n'a pas encore été catégorisée alors elle demeure ambiguë.

L'importance de ce filtre est soulignée par les chiffres que nous avons trouvés dans nos sondages et qui nous ont surpris nous-mêmes: plus de la moitié des verbes finis (54%) comportent un pronom personnel sujet. Toutefois, comme il ne s'agit pas de tester seulement les formes verbales qui sont réellement des verbes finis, mais de tester tous les verbes finis potentiels, il est plus intéressant de trouver le nombre de ces derniers qui sont pourvus d'un pronom sujet. Des verbes finis potentiels, 40% ont un pronom sujet. Si l'on ajoute à ces 40% les formes verbales précédées d'autres particules qui prouvent qu'on a affaire à un verbe fini, on peut estimer que, dans la moitié des cas environ, le test des particules verbales suffit pour catégoriser de façon définitive les formes verbales homographes.<sup>22</sup>

<sup>22</sup>Maegaard, Bente et Spang-Hansen, Ebbe, La segmentation automatique du français écrit, France, éditions Jean-Favard, 1978, p.96.

### 3.3.2.3 Le test des substantifs

La troisième étape de la procédure de Maegaard et Spang-Hanssen est un test des substantifs. Si l'alternative d'une forme verbale potentielle (cette alternative est indiquée dans le code d'ambiguïté transmis par le dictionnaire des radicaux) est un substantif alors la forme subira ce test. Il s'agit ici que l'ordinateur aille chercher le mot qui précède la forme et vérifie s'il y a accord entre les deux. S'il y a accord alors la forme est catégorisée comme étant un substantif donc comme n'étant définitivement pas un verbe. Par exemple avec "Pierre fait une marche" **m**arche est un substantif car il y a accord entre **une** et **m**arche. Il faut dire que dans une situation comme "Il la porte" porte a été déjà classé comme étant un verbe avec le test des particules verbales et seuls les formes verbales non-précisées subissent les tests. "Sur 2377 homographes du type verbe-substantif, 122 ont été mal analysés; soit 5,1% d'échecs. La source d'erreurs la plus importante est l'insuffisance de la liste des déterminants."<sup>23</sup>

### 3.3.2.4 Le test des participes passés en -s et des adjectifs

La quatrième étape de la procédure consiste en un test des participes passés en -s et des adjectifs. Il ne s'agit pas vraiment d'un test. L'ordinateur, si la forme verbale potentielle dont l'alternative serait un participe passé en -s ou un adjectif, n'a pas été précisée définitivement comme étant un verbe avec le test des particules verbales, indique que la forme n'est pas un verbe. "Le test est basé sur le fait que les formes verbales potentielles qui sont des homographes de participes passés en -s ne peuvent être que de la première ou de la deuxième personne, qui exigent l'emploi d'un pronom personnel sujet."<sup>24</sup> Maegaard et Spang-Hanssen ont obtenu avec ce test un taux de succès de 98.6%.

### 3.3.2.5 Le test des participes passés en -t

La cinquième étape de la procédure est un test des participes passés en -t. L'ordinateur vérifie si le mot précédent est une forme des verbes avoir ou être ou s'il appartient à la liste suivante: bien, déjà, encore, jamais, mal, pas, plus, presque, rien, souvent, tout, toujours. Si oui alors c'est qu'il s'agit d'un participe, si non alors c'est un verbe fini. "Le pouvoir de discrimination de ce test est effectivement assez faible: 25 cas d'erreurs sur 168 cas relevés, soit 15% d'échecs."<sup>25</sup>

### 3.3.2.6 Le test du participe-substantif "fait"

La sixième étape est un test du participe-substantif fait. Puisque le participe fait est fréquemment employé comme substantif, la procédure, si la forme verbale n'a pas été identifiée par le test des participes en -t, lui fait passer le test des substantifs. S'il ne s'agit pas d'un substantif alors c'est un verbe. "Sur 115 cas relevés, on compte 6 cas d'erreurs, soit 5,2% d'échecs. 4 ou 5 des mauvaises analyses sont imputables à une erreur de programmation qui n'a été corrigée qu'en cours de route. On a donc des raisons sérieuses de croire qu'après cette correction le test est tout à fait satisfaisant."<sup>26</sup>

<sup>23</sup> *Ibid.*, p.46.

<sup>24</sup> Maegaard et Spang-Hanssen, *op. cit.*, p.48.

<sup>25</sup> *Ibid.*, p. 51.

<sup>26</sup> Maegaard et Spang-Hanssen, *op. cit.*, p.51.

### 3.3.2.7 Le test de la préposition "entre"

La septième étape est un test de la préposition *entre*. Si ce qui suit *entre* est une préposition, une virgule ou un point alors il est classé comme étant un verbe. "Toutes les occurrences de *entre* (71) ont été correctement analysées. Ce résultat s'explique en partie par les fréquences très inégales de la préposition et de la forme verbale: puisque celle-ci est bien moins fréquente, nos programmes exigent, soit par le test des particules verbales, ou par le test décrit ici, un indice sûr excluant la fonction prépositionnelle pour accepter *entre* comme une forme du verbe *entrer*."<sup>27</sup>

### 3.3.2.8 Le test des locutions adverbiales en -ons

La dernière étape est un test des locutions adverbiales en -ons. "Il s'agit ici des locutions à reculons et à tâtons. Si et seulement si le mot qui précède l'homographe n'est pas à, nous avons affaire à un verbe fini [...] Vu la grossièreté de nos tests, il peut sembler ridicule d'avoir créé un test spécial pour ces locutions dont nous n'avons enregistré qu'une seule occurrence (correctement analysée)."<sup>28</sup>

---

<sup>27</sup> *Ibid.*, p. 53.

<sup>28</sup> *Ibid.*, p. 53.

## CHAPITRE 2

### L'AUTOMATISATION DE L'ANALYSE DES TEMPS D'APRÈS LE MODÈLE PROPOSÉ PAR HARALD WEINRICH

L'automatisation de la méthode d'analyse de Harald Weinrich, en ce qui regarde les formes verbales, comporte trois étapes principales: La reconnaissance automatique des verbes, l'analyse statistique de ceux-ci et la présentation des résultats sous une forme telle qu'un littéraire non familier avec l'ordinateur et ses langages puisse directement les apprécier et les commenter. Pour la première étape nous avons repris l'essentiel de la méthode de Maegaard et Spang-Hanssen en y apportant un certain nombre de modifications en considérant le type de texte que nous voulions analyser, les temps de verbes dont nous avons besoin et du type de résultats que nous voulions obtenir. Pour la seconde nous avons conçu notre propre algorithme d'analyse statistique: les progiciels d'analyses statistiques disponibles sur le marché n'offrant pas les types d'analyses dont nous avons besoin. Pour la dernière étape nous nous sommes inspirés des graphiques que proposent les progiciels d'analyses statistiques et de la méthode que propose Weinrich pour la présentation des transitions temporelles.

#### 1. La reconnaissance automatique des formes verbales

##### 1.1 *Emprunts à la méthode de Maegaard et Spang-Hanssen*

De Maegaard et Spang-Hanssen nous n'avions qu'une description des divers tests qu'ils ont utilisés, de leur table des codes d'ambiguïtés et de leur dictionnaire des terminaisons. Nous avons fabriqué notre propre dictionnaire des radicaux, modifié quelque peu leur dictionnaire des terminaisons afin qu'il s'adapte parfaitement à notre dictionnaire des radicaux, modifié également leur table des codes d'ambiguïtés afin qu'elle corresponde à notre banque de verbes. Puisque nous n'avions qu'une description de la procédure nous avons fait tout le travail de programmation.

##### 1.2 *Nos dictionnaires*

###### 1.2.1 Dictionnaire des radicaux

La qualité d'un algorithme de reconnaissance des formes verbales dépend en grande partie du nombre de verbes que ses dictionnaires répertorient. Cela est d'autant plus vrai lorsque la procédure de reconnaissance est appliquée à des textes littéraires: le vocabulaire y étant généralement plus varié que dans des textes non-littéraires. Il n'était donc pas question de fabriquer un dictionnaire à partir d'une banque de verbes du français fondamental. Nous avons donc rédigé notre dictionnaire des radicaux à partir de la banque de 12,000 verbes proposée par Bescherelle dans son ouvrage *Le nouveau Bescherelle*. Compte tenu du nombre d'homographes, nous avons obtenu 8639 formes différentes qui se regroupent en 77 classes:

La codification fait référence à la banque de terminaisons: le premier chiffre correspond aux terminaisons de la conjugaison indiquée dans la première classe, le second aux terminaisons de la conjugaison indiquée dans la seconde classe, le troisième aux terminaisons de la conjugaison indiquée dans la troisième classe et le quatrième aux terminaisons de la conjugaison indiquée dans la quatrième classe.

1- au-	3 0 0 0	a-	0 6 0 0	xxx-	0 0 31 0
e-	0 0 0 3				
2- e-	1 0 0 0	f-	0 0 0 3	xxx-	0 7 32 0
3- aim-	1 1 1 1				
4- plac-	1 0 1 0	pla-	0 9 0 8		
5- mang-	1 8 1 7				
6- pès-	1 0 1 0	pes-	0 1 0 1		
7- céd-	1 1 0 1	céd-	0 0 1 0		
8- jett-	1 0 1 0	jet-	0 1 0 1		
9- modél-	1 0 1 0	model-	0 1 0 1		
10- cré-	1 1 1 1				
11- assiég-	1 8 0 7	assiég-	0 0 1 0		
12- appréci-	1 1 1 1				
13- pay-	1 1 1 1	pai-	1 0 1 0		
14- broi-	1 0 1 0	broy-	0 1 0 1		
15- enver-	3 0 0 0	envoy-	0 1 0 1	ex.voi-	0 0 1 0
16- fin-	2 0 0 2	finiss-	0 1 0 0	fini-	0 0 2 0
17- haf-	3 0 0 0	haïse-	0 1 0 0	ha-	0 0 18 5
18- all-	0 1 0 1	xxx-	2 0 6 0		
19- tiend-	3 0 0 0	ten-	0 1 0 0	tien-	0 0 25 0
t-	0 0 0 6				
20- acquier-	3 0 0 0	acquér-	0 1 0 0	acqui-	0 0 34 0
acqu-	0 0 0 2				
21- sent-	2 1 0 2	sen-	0 0 19 0		
22- vét-	2 1 0 2	vê-	0 0 28 0		
23- couvr-	2 1 1 2				

24- cueill-	1 1 1 2				
25- assaill-	2 1 1 2				
26- faill-	2 1 0 2	fa-	0 0 35 0	faud-	3 0 0 0
27- bouill-	2 1 0 2	bou-	0 0 9 0		
28- dorm-	2 1 0 2	dor-	0 0 15 0		
29- cour-	3 1 5 3				
30- mour-	3 1 0 3	meur-	0 0 5 0		
31- serv-	2 1 0 2	ser-	0 0 8 0		
32- fui-	3 0 5 0	fuy-	0 1 0 0	fu-	0 0 0 2
33- recev-	3 1 0 0	reçoi-	0 0 8 0	reç-	0 0 0 3
34- ver-	3 0 0 0	voy-	0 1 0 0	voi-	0 0 5 0
v-	0 0 0 2				
35- pourvoi-	3 0 5 0	pourvoy-	0 1 0 0	pourv-	0 0 0 3
36- sau-	3 0 0 0	sa-	0 4 24 0	s-	0 0 0 3
37- dev-	3 1 0 0	doi-	0 0 8 0	d-	0 0 0 3
38- pour-	3 0 0 0	pouv-	0 1 0 0	p-	0 0 22 3
p-	0 0 30 3				
39- mouv-	3 1 0 0	meu-	0 0 8 0	m-	0 0 0 3
40- pleuv-	3 1 0 0	pleu-	0 0 8 0	pl-	0 0 0 3
41- faud-	3 0 0 0	fa-	0 0 17 0	fall-	0 1 0 3
42- vaud-	3 0 0 0	val-	0 1 0 3	va-	0 0 26 0
43- voud-	3 0 0 0	voul-	0 1 0 3	veu-	0 0 29 0
44- assié-	3 0 0 0	assey-	0 1 0 0	ase-	0 0 7 2
assoï-	3 0 5 0	assoy-	0 1 0 0		
45- surseoi-	3 0 0 0	sursoy-	0 1 0 0	surseoi-	0 0 5 0
surs-	0 0 0 2				
46- choi-	3 0 5 0	ch-	0 0 0 3		
47- rend-	3 1 0 2	ren-	0 0 3 0		
48- prend-	3 0 0 0	pren-	0 1 23 0	pr-	0 0 0 2
49- batt-	3 1 0 2	ba-	0 0 27 0		

50- mett-	3 1 0 0	me-	0 0 27 0	m-	0 0 0 2
51- peind-	3 0 0 0	peign-	0 1 0 2	pei-	0 0 13 0
52- joind-	3 0 0 0	joign-	0 1 0 2	joi-	0 0 13 0
53- craind-	3 0 0 0	craign-	0 1 0 2	crai-	0 0 13 0
54- vainc-	3 0 0 0	vainqu-	0 1 0 2	vain-	0 0 33 0
55- trai-	3 0 5 0	tray-	0 1 0 0		
56- f-	1 0 0 2	fa-	0 3 16 0		
57- plai-	3 0 0 0	plais-	0 1 0 0	pla-	0 0 21 0
pl-	0 0 0 3				
58- connait-	3 0 0 0	connais-	0 1 0 0	conna-	0 0 36 0
conn-	0 0 0 3				
59- nait-	3 0 0 0	naiss-	0 1 0 0	na-	0 0 36 0
naqu-	0 0 0 2				
60- pait-	3 0 0 0	paiss-	0 1 0 0	pa-	0 0 36 0
61- crott-	3 0 0 0	croiss-	0 1 0 0	cro-	0 0 14 0
cr-	0 0 0 4				
62- croi-	3 0 5 0	croy-	0 1 0 0	er-	0 0 0 3
63- boi-	3 0 8 0	buv-	0 1 0 0	b-	0 0 0 3
64- clo-	3 0 0 0	clos-	0 1 0 0	cl-	0 0 37 0
65- conclu-	3 1 5 0	concl-	0 0 0 3		
66- absoud-	3 0 0 0	absolv-	0 1 0 0	absco-	0 0 4 0
67- coud-	3 0 0 0	cous-	0 1 0 2	cou-	0 0 12 0
68- moud-	3 0 0 0	moul-	0 1 0 3	mou-	0 0 20 0
69- suiv-	3 1 0 2	sui-	0 0 8 0		
70- viv-	3 1 0 0	vi-	0 0 8 0	véc-	0 0 0 3
71- li-	3 0 10 0	lis-	0 0 8 0	l-	0 0 0 3
72- di-	3 10 5 0	d-	0 0 0 3		
73- médis-	3 0 5 0	méd-	0 0 0 2	médis-	0 1 0 0
74- ri-	3 1 5 0	r-	0 0 0 2		
75- écri-	3 0 8 0	écriv-	0 1 0 2		

76- confi- 3 0 10 0    confis- 0 1 0 0    c if- 0 0 0 2  
 77- cui- 3 0 10 0    cuis- 0 1 0 2

Fabriquer un dictionnaire automatique demande un certain nombre de considérations. D'abord il faut mettre au point un programme qui nous permet de rédiger, corriger, trier, indexer et bien sûr d'afficher et imprimer le dictionnaire. Dans la mise au point de ce programme, les plus grandes tâches consistaient à réaliser un algorithme de tri et un algorithme d'indexation.

### 1.2.1.1 Le tri alphabétique et l'indexation

Pourquoi le tri et l'indexation? Les deux sont intimement liés. D'abord lorsque l'ordinateur recherche un mot dans le dictionnaire, si ce dernier n'est pas trié en ordre alphabétique, il doit, dans le cas par exemple où le mot recherché n'y apparaît pas, procéder à un balayage complet, mot par mot, du dictionnaire. Nous l'avons vu précédemment, l'ordinateur n'est apte qu'à reconnaître des formes. Il prend le premier mot du dictionnaire, le compare avec celui qu'il recherche: si c'est un mot différent alors il prend le deuxième et refait la vérification et ainsi de suite jusqu'à ce qu'il trouve le mot où qu'il atteigne la fin du dictionnaire. Si le dictionnaire est classé en ordre alphabétique alors on peut sauver du temps en donnant à l'ordinateur l'instruction d'arrêter la recherche s'il trouve le mot ou s'il le compare à un mot alphabétiquement plus grand. L'avantage du tri alphabétique apparaît plus nettement si on utilise la méthode de recherche dite dichotomique. Cette méthode consiste à dire à l'ordinateur d'aller d'abord lire le mot qui se situe exactement au milieu du dictionnaire et de le comparer au mot recherché. Si le mot est plus grand que celui qu'on cherche alors l'ordinateur va lire le mot qui se situe à la moitié de la première moitié du dictionnaire, s'il est plus petit alors l'ordinateur va à la moitié de la seconde moitié. L'ordinateur recommence cette opération jusqu'à ce qu'il trouve le mot désiré ou s'il cherche à aller à la moitié d'un écart de deux mots. Avec une telle méthode l'ordinateur ne fera que 14 comparaisons s'il recherche un mot absent d'un dictionnaire de près de 10,000 mots.

#### 1.2.1.1.1 L'indexation

L'indexation nous permet de réduire encore plus le nombre de comparaisons pour la recherche d'un mot. Il est vrai qu'à la vitesse à laquelle l'ordinateur travaille 14 comparaisons ne représentent qu'une fraction de seconde. Cependant s'il doit rechercher plus de 10,000, 20,000 ou 100,000 mots alors une simple multiplication nous montre vite que ces fractions de seconde deviennent facilement un temps fou! Il s'agit alors de diviser le dictionnaire en segments. Nous avons fabriqué un répertoire qui est constitué de toutes les combinaisons de deux caractères. A chaque combinaison nous avons placé les coordonnées de tous les mots qui commencent par ces deux caractères. Par exemple à ab correspond la position dans le dictionnaire du premier mot qui commence par ab (abaissier) et la position du dernier mot qui commence par ab (abuser). Lorsque l'ordinateur recherche un mot dans le dictionnaire il consulte d'abord notre répertoire, y lit les coordonnées correspondantes et procède à une recherche dichotomique à l'intérieur du nouveau segment. Cette méthode réduit le nombre de comparaisons maximum à 7 seulement.

#### 1.2.1.1.2 Le tri

Pour un être humain, trier est une opération relativement simple. Nous exécutons cette tâche sans grande réflexion, et elle ne cause de difficulté à peu près à personne. Cela est loin d'être le cas pour un ordinateur. Ce dernier peut facilement comparer deux mots et déterminer lequel est

alphabétiquement plus grand ou plus petit que l'autre. Les problèmes surviennent lorsque vient le moment de gérer une masse d'information pendant le tri et d'optimiser le nombre de comparaisons. En guise d'exemple, si nous voulons trier un fichier de dix mots on pourrait demander à l'ordinateur de faire un balayage complet du fichier et de trouver le mot le plus petit (alphabétiquement) et de le mettre à la première position. Ensuite, l'ordinateur refait la même opération mais en partant de la deuxième position et en plaçant le mot le plus petit à sa position de départ et ainsi de suite jusqu'à ce qu'il atteigne la fin du fichier. Cette opération exige 55 accès au fichier de la part de l'ordinateur c'est-à-dire  $N + (N - 1) + (N - 2) \dots (N - N)$  accès. Lorsqu'il s'agit d'un fichier de 10,000 mots le nombre d'accès (50,009,070) est trop important pour que le tri soit efficace. Il est donc nécessaire d'utiliser des algorithmes plus sophistiqués.

Les techniques de tri sont nombreuses et chacune comporte des avantages et des inconvénients. En gros nous pouvons les classer en deux groupes: celles qui exécutent le tri à partir d'un fichier transposé dans une table en mémoire vive et celles qui travaillent directement sur un disque. Les premières sont nettement plus rapides que les secondes mais ne peuvent pas trier un fichier plus grand que l'espace mémoire disponible dans l'ordinateur (cela va de soi!). Vu la grosseur de notre dictionnaire nous n'avons pu choisir une des premières cependant les secondes nécessitent quelques heures d'opération. Devant ce dilemme nous avons décidé de préparer notre propre algorithme de tri en combinant les avantages d'une technique (tri shell) faisant partie du premier groupe et les avantages d'une autre (Quicksort) faisant partie du second groupe.

Suivant le principe du dernier (Quicksort), l'ordinateur cherche aléatoirement un mot dans le dictionnaire et il s'en servira tout le long de la procédure pour faire ses comparaisons: appelons-le mot-source. Il interchange le mot-source avec le premier du fichier. Il compare ensuite le mot-source avec le dernier: si le dernier est plus gros l'ordinateur recule d'une position (c'est-à-dire qu'il prend l'avant dernier mot) et refait la comparaison. Cette opération est répétée jusqu'à ce que l'ordinateur trouve un mot plus petit. Lorsque c'est le cas il interchange les deux mots (le mot-source et le mot plus petit qu'il vient de trouver) et recommence la procédure mais en comparant le mot-source avec le deuxième en partant du début et en procédant à un nouvel échange uniquement lorsqu'il trouve un mot plus grand: lorsque cela est fait il recommence la procédure mais en continuant à partir de la dernière position qu'il occupait à la fin du fichier. A la fin de la procédure, le fichier est divisé en deux segments. Dans le premier il y a tous les mots plus petits que le mot-source et dans le second tous ceux qui sont plus grands. L'ordinateur place dans une pile les coordonnées du segment de droite (ou de la fin) puis vérifie si le segment de gauche (ou celui du début) est suffisamment petit pour être logé en mémoire vive. Si oui alors il le transpose dans une table en mémoire vive et procède à un tri shell, si non il reprend la procédure que nous avons décrite à partir de ce segment. Lorsque l'ordinateur a pu procéder à un tri shell il remet ensuite le segment trié à sa place et lit la dernière coordonnée placée dans la pile. Il vérifie si le nouveau segment est suffisamment petit pour être logé à son tour en mémoire vive et ainsi de suite jusqu'à ce que tout le fichier soit trié. Avec cette méthode l'ordinateur trie notre dictionnaire en dix minutes.

### 1.2.2 Dictionnaire des terminaisons

Au dictionnaire des terminaisons de Maegaard et Spang-Hanssen nous avons apporté un certain nombre de modifications. Les temps de verbes qui nous intéressent sont le futur, le conditionnel, le présent de l'indicatif, l'imparfait de l'indicatif, et le passé simple. Puisque le présent du subjonctif, l'imparfait du subjonctif et l'impératif ne semblaient pas ajouter à la difficulté de la mise au point de la procédure de reconnaissance des verbes, nous avons décidé de les conserver. Nous avons donc éliminé les terminaisons qui correspondent au participe passé, au participe présent et à l'infinitif. Plutôt que de diviser les terminaisons en six classes (paradigmes) nous les avons divisées en quatre et avons réduit la deuxième classe à quatre temps. L'élimination du participe présent a grandement joué en notre faveur puisque c'est le temps auquel correspondait le plus grand nombre d'erreurs dans la procédure de Maegaard et Spang-Hanssen.

Outre quelques erreurs de dactylographie, imputables sans doute à l'éditeur, que nous avons eu à corriger, nous avons dû ajouter un certain nombre de conjugaisons à l'intérieur des classes. A la troisième classe nous avons ajouté quatre conjugaisons:

Classe 3.

Temps	personne	34	35	36	37
Présent de l'indicatif	1	-ers	-ux	-is	-os
	2	-ere	-ux	-is	-os
	3	-ert	-ut	-it	-ôt
	6	-èrent	-illent	-issent	-oent
Présent du subjonctif	1	-ère	-ille	-isse	-ose
	2	-ères	-illes	-isses	-oses
	3	-ère	-ille	-isse	-ose
	6	-èrent	-illent	-issent	-oent
Impératif	2	-ers	---	-is	-os

La 34ème conjugaison va pour les verbes du type **acquérir**, la 35ème pour les verbes du type **faillir**, la 36ème pour les verbes du type **paitre** et la 37ème pour les verbes du type **clore**. Avec ces changements nous avons obtenu un dictionnaire de plus de 585 désinences.

Tout comme pour le dictionnaire des radicaux nous avons préparé un programme qui nous permet de faire la saisie du dictionnaire, de le corriger, de le trier, de l'indexer, de l'afficher et de le faire imprimer. Pour le tri et l'indexation nous avons repris les mêmes algorithmes que pour le dictionnaire des radicaux.

### 1.2.3 Table des codes d'ambiguïtés

Puisque notre banque de verbes est plus grande que celle de Maegaard et Spang-Hanssen nous avons dû ajuster la table des codes d'ambiguïtés en conséquence. Leur table contenait 42 codes, nous en avons ajouté huit pour caractériser les radicaux qui ont les types d'ambiguïtés suivants.

**Coch-** : présent de l'indicatif ou du subjonctif à la première ou à la troisième personne ou substantif au féminin singulier ou masculin singulier (coche), présent de l'indicatif ou du subjonctif à la deuxième personne ou substantif au féminin pluriel ou masculin pluriel (coches), présent de l'indicatif à la quatrième personne ou substantif masculin pluriel (cochons).

**Couv-** : présent de l'indicatif ou du subjonctif à la sixième personne ou substantif au masculin et au singulier (couvent).

**Crev-** : imparfait du subjonctif à la première ou à la deuxième personne ou substantif féminin singulier (crevasse), imparfait du subjonctif à la deuxième personne ou substantif féminin pluriel (crevasses).

**Pass-** : présent de l'indicatif ou du subjonctif à la première ou à la troisième personne ou substantif au masculin singulier ou féminin singulier (passe), présent de l'indicatif ou du subjonctif à

la deuxième personne ou substantif au masculin pluriel ou féminin pluriel (passes), imparfait de l'indicatif ou présent du subjonctif à la quatrième personne ou substantif au féminin pluriel (passions).

Pei- : présent du subjonctif à la première ou à la troisième personne ou substantif au masculin singulier (peigne), présent du subjonctif à la deuxième personne ou substantif au masculin pluriel (peignes), présent de l'indicatif à la troisième personne ou participe passé finissant avec un t (peint).

Viol- : présent de l'indicatif ou du subjonctif à la première ou à la troisième personne ou substantif au masculin singulier (viole), présent de l'indicatif ou du subjonctif à la deuxième personne ou substantif au masculin pluriel (violes), présent de l'indicatif à la quatrième personne ou substantif masculin pluriel (violons), présent de l'indicatif ou du subjonctif à la sixième personne ou un adjectif (violent).

Exempt- : présent de l'indicatif ou du subjonctif à la première ou à la troisième personne ou adjectif (exempte).

Parf- : présent de l'indicatif à la première ou à la troisième personne ou participe passé ou adjectif (parfait), présent de l'indicatif à la cinquième personne ou adjectif (parfaites).

#### Répertoire des déterminants

Lors du test des substantifs l'ordinateur doit vérifier si la forme verbale potentielle est précédée d'un déterminant. Si la vérification est positive alors il y a vérification de l'accord. S'il y a accord alors l'ordinateur peut préciser la forme comme n'étant pas une forme verbale. Pour procéder à la comparaison l'ordinateur doit être en mesure de connaître le genre et le nombre de l'alternative à la forme verbale et du déterminant. Pour la forme verbale toute l'information nécessaire est contenue dans le code d'ambiguïté, mais comment procéder avec les déterminants?

Nous avons préparé un répertoire de tous les déterminants à chacun desquels nous avons assigné un code représentant le genre et le nombre. Nous avons repris la même codification que pour l'alternative substantif des formes verbales potentielles:

a: masculin singulier	b: féminin singulier
c: masculin singulier, masculin pluriel	d: masculin pluriel
e: féminin pluriel	f: féminin singulier, féminin pluriel
g: masculin singulier, féminin singulier	h: masculin pluriel, féminin pluriel
i: masculin singulier, masculin pluriel, féminin pluriel	

Pour vérifier l'accord l'ordinateur n'a qu'à comparer les codes. Par exemple un substantif a s'accorde avec un déterminant a, c, g ou i. Un substantif b s'accorde avec un déterminant b, f ou g.

### 1.3 Les tests de désambiguïsation

Des tests de Maegaard et Spang-Hanssen nous n'avons retenu que le test des particules verbales, le test substantif, le test des participes passés en -s et des adjectifs et le test de la préposition "entre". Nous avons ignoré le test des participes passés en -t parce que son taux d'efficacité est assez faible: 15% d'échecs. Nous avons ignoré également le test du participe-substantif "fait" mais cette fois parce que le test des particules verbales et le test substantif semblent désambiguïser la grande majorité de ces formes. En ce qui concerne le test des locutions adverbiales en "-ons" Maegaard et Spang-Hanssen n'y voyait pas eux-mêmes son importance.

Une autre raison nous a poussé à rejeter si facilement un certain nombre de tests. Les résultats obtenus à la suite d'une série de tests comme ceux-ci ne sont jamais parfaits. Il y a toujours un certain pourcentage d'erreurs. Afin d'optimiser la qualité de nos résultats nous avons décidé d'ajouter à la procédure de reconnaissance automatique des formes verbales une étape de travail interactif où l'ordinateur présente à l'utilisateur, tour à tour accompagnés de leurs contextes, tous les verbes pour lesquels l'ambiguïté n'a pas été levée. Voici en exemple un message que l'ordinateur transmet à l'utilisateur dans le cas d'une ambiguïté persistante:

Ambiguïté	:48	Chapitre	:	1
Information	:c3	Ligne	:	8
		Ordre	:	4
paille. Contre le lambris, <u>peint</u> en blanc, s'alignaient				
Ce mot devrait-il être considéré comme un verbe? (O/N/A)				

Le travail interactif à la suite d'une analyse d'un texte de plus de 2,000 mots représente à peine une minute de travail.

## 2. L'analyse statistique

Une fois les verbes reconnus par l'ordinateur, il faut que ce dernier puisse les traiter afin de présenter des résultats qui nous permettent de décrire, commenter, voire interpréter la récurrence des formes verbales du point de vue de la syntaxe textuelle: plus précisément d'évaluer le tempo narratif, de discerner les mises en relief et de commenter les transitions temporelles. Ces résultats doivent être facilement compréhensibles par un littéraire non-familier avec les ordinateurs.

Les résultats sont présentés sous forme de dossiers. Chaque dossier comprenant les informations suivantes: les coordonnées du texte ou de la partie de texte étudié, un graphique représentant le nombre de verbes pour chaque temps par tranches de mots, une table des valeurs qui apporte la précision que le graphique ne peut pas rendre, une table des fréquences et une matrice des transitions temporelles (d'après le modèle proposé par Harald Weinrich).

## 2.1 Les coordonnées du texte

Lorsque l'ordinateur effectue une analyse statistique il garde en mémoire les coordonnées du texte ou de la partie du texte analysé. Ainsi pour chaque dossier l'utilisateur peut continuellement savoir à quelle partie exactement du texte correspondent les résultats qu'il obtient. La difficulté ici est de faire rédiger par l'ordinateur une phrase qui contient les coordonnées du texte. Pour cela nous avons préparé une phrase clef qui contient huit espaces vides: "La partie de texte commence au chapitre \_\_, phrase \_\_ : \_\_ \_\_\_\_\_ et se termine au chapitre \_\_, phrase \_\_ : \_\_ \_\_\_\_\_." Dans la première case l'ordinateur inscrit le chiffre qui correspond au chapitre, dans la seconde le chiffre qui correspond à la phrase, dans la troisième le chiffre qui correspond à l'ordre du mot dans sa phrase et dans la quatrième il inscrit *er mot* s'il s'agit du premier mot de la phrase ou *ème mot*. Pour les quatre autres cases c'est la même chose sauf que cette fois l'ordinateur y inscrit les coordonnées de la fin. Nous obtenons un résultat qui ressemble à ceci: "La partie de texte commence au chapitre 1, phrase 1: 1er mot et se termine au chapitre 2, phrase 31: 1er mot" ou "La partie de texte commence au chapitre 1, phrase 10: 6ème mot et se termine au chapitre 4, phrase 8: 4ème mot".

## 2.2 Le graphique

Fabriquer un graphique à partir de données statiques demande trois étapes. A la première on dessine la charpente du graphique: deux droites graduées dont l'une correspond à l'abscisse et l'autre à l'ordonnée. À la seconde on place les points aux coordonnées correspondantes. A la dernière on dessine les segments qui relient les points à l'intérieur du plan délimité par les droites.

Fabriquer un graphique à partir de données dynamiques pose un certain nombre de difficultés. La première est causée par la graduation des droites. Sur l'ordonnée nous inscrivons le nombre de verbes et sur l'abscisse le nombre de mots par tranches de texte. Avec un texte de 1024 mots nous pouvons diviser le texte par tranches de 64 mots: cela nous donne 16 tranches. Dans la première tranche nous pourrions avoir sept verbes à l'imparfait et deux au passé simple, dans la seconde un à l'imparfait et deux au passé simple et ainsi de suite. Avec une ordonnée graduée de 1 à 16 (tout comme l'abscisse) nous aurions un graphique avec des proportions raisonnables. Qu'arriverait-il si notre texte avait 10,240 mots? Nous pourrions avoir par tranches entre 20 et 100 verbes. Le graphique prendrait alors des proportions trop grande pour être imprimé sur une seule feuille et sa lecture serait très inconfortable.

Nous avons conçu et incorporé à l'intérieur de la procédure de la mise en forme des graphiques un module d'évaluation des données. Ce module, après analyse quantitative des données, transmet à la procédure un multiplicateur pour l'échelle de l'ordonnée et divise chaque élément numérique par le multiplicateur. L'ordonnée conserve donc toujours la même échelle mais celle-ci est accompagnée d'un multiplicateur. Si le graphique indique pour une tranche par exemple sept verbes à l'imparfait et que le multiplicateur indiqué est 2 alors cela signifie que nous avons en fait quatorze verbes à l'imparfait.

À l'abscisse nous faisons correspondre un chiffre qui représente le nombre de mots par tranches. Avec un texte de 10,240 mots nous aurons 640 mots par tranche, avec un texte de 1024 nous en aurons 64. Cette indication est nécessaire afin que la personne qui étudie le graphique puisse avoir une idée du nombre de mots analysés.

Une des particularités intéressantes du module de présentation des graphiques est la possibilité qu'il offre de focaliser à un endroit donné du graphique et de générer un nouvel ensemble de résultats ou nouveau dossier. Le focus est probablement l'élément qui nous a causé le plus de problèmes d'algorithmie et de programmation mais il est aussi l'élément le plus important de l'analyse statistique. Sans lui nous ne pourrions avoir d'un texte une vision trop générale du comportement verbale pour qu'elle puisse nous être de quelque utilité.

Les problèmes qui se sont présentés relèvent plus de notre manque d'expérience dans la mise au point d'algorithmes et de programmes (nous avons appris les techniques d'algorithmie et de programmation en échaffaudant notre projet, c'était là notre premier contact avec l'informatique). Incorporer un focus dans notre analyse consistait à rendre l'ensemble du programme de statistique récursif. Le focus est une option à l'intérieur du programme qui contient le programme en entier dont l'option de focalisation elle-même et ainsi de suite. Cette aptitude à la récursivité est une des caractéristiques du langage de programmation que nous avons choisi, mais encore faut-il savoir la manipuler.

Lorsqu'on demande à l'ordinateur de refaire toute l'analyse statistique sur un segment particulier il faut qu'il puisse reconnaître les nouvelles coordonnées textuelles. Pour cela nous demandons à l'ordinateur de faire correspondre à chacun des points du graphique leur emplacement exact dans le texte. De plus il faut prévoir une sauvegarde des résultats sur disque parce que les résultats de la nouvelle analyse remplaceront les résultats précédents. Nous aurions pu incorporer les nouvelles données à l'intérieur du dossier d'origine mais cela n'aurait pas été très commode. Si par exemple une analyse de texte nous amène à une dizaine de sous-niveaux alors dans une éventuelle reprise des travaux il faudrait refaire tout le cheminement antérieur afin d'étudier tel ou tel segment du texte. En divisant les données en dossiers indépendants nous n'avons qu'à reprendre le dossier qui nous intéresse.

### 2.3 Table des valeurs

Le graphique nous donne un aperçu schématisé du comportement verbal. Il est difficile parfois à partir uniquement du graphique d'obtenir des chiffres précis: nombre de verbes pour tel ou tel temps dans telle ou telle tranche. Pour pallier cet inconvénient nous faisons tenir un registre (ou une table) de toutes les données par tranche et pour tous les temps de verbe. Ces informations sont accessibles en pressant uniquement une touche au clavier.

### 2.4 Table des fréquences

Une autre option qu'offre la procédure de statistique est une table des fréquences. Pour le segment de texte analysé l'ordinateur compile le nombre total de verbes pour chaque temps. Il présente les résultats sous forme de table, laquelle comprend tous les temps de verbe accompagnés du nombre d'occurrences correspondantes.

### 2.5 La matrice des transitions temporelles

Outre le graphique, la matrice des transitions temporelles est la partie de l'analyse la plus importante. Nous avons repris le modèle que propose Harald Weinrich dans son ouvrage *Le temps*. Il s'agit ici de répertorier toutes les transitions d'un temps de verbe à un autre. Si le premier verbe du texte analysé est par exemple un passé simple il y aura une transition du type nul $\Leftrightarrow$ passé simple. Si le second verbe est un autre passé simple alors il y aura une transition du type passé simple $\Leftrightarrow$ passé simple. Si le troisième est un imparfait alors il y aura une transition du type passé simple $\Leftrightarrow$ imparfait et ainsi de suite. Voici un exemple de matrice des transitions temporelles (Elle a été faite à partir de l'analyse d'un segment d'*Un coeur simple* de Gustave Flaubert:

Transitions temporelles

	Nul	Futur	condi	pres	p.sub	imparf	i.sub	pas-simp	imper
Nul	0	0	0	0	0	0	0	1	0
Futur	0	0	0	0	0	0	0	0	0
Condi	0	0	0	0	0	0	0	0	0
Pres	0	0	0	0	0	0	0	0	0
P.sub	0	0	0	0	0	0	0	0	0
Imparf	0	0	0	0	0	47	0	19	0
I.sub	0	0	0	0	0	0	0	0	0
Pas-simp	1	0	0	0	0	19	0	17	0
Imper	0	0	0	0	0	0	0	0	0

Il faut lire la matrice en partant de gauche. On peut voir qu'il y a 47 transitions du type imparfait $\Leftrightarrow$ imparfait, 19 transitions imparfait $\Leftrightarrow$ passé simple, 19 transitions Passé simple $\Leftrightarrow$ imparfait et 17 transitions passé simple $\Leftrightarrow$ passé simple. Les premier et dernier verbes du segment analysé étaient au passé simple.

### 3. L'Analyste

L'aboutissement de nos travaux constitue un logiciel que nous avons baptisé *L'Analyste*. Il se divise en trois parties: un ensemble de modules secondaires permettant de faire la saisie et la gestion de nos différents dictionnaires et de préparer le texte à analyser, un module de reconnaissance des formes verbales et un module d'analyse statistique.

#### 3.1 Préparation d'un texte

Les textes, avant d'être soumis à L'ANALYSTE, doivent subir un traitement préalable qui comporte deux aspects importants. Le premier est la segmentation des textes en unités (mots). Le second est une structuration des données (des textes segmentés) sous une forme telle que L'ANALYSTE puisse les reconnaître. Chaque mot est placé dans ce qu'on pourrait appeler une fiche (enregistrement), qui comprend les informations suivantes: Le mot, la ponctuation, le numéro de ligne, l'ordre du mot dans la ligne, le numéro du chapitre, un espace (20 caractères) pour l'information relative aux divers temps et personnes de la forme verbale en voie de désambiguïsation, le code d'ambiguïté (initialisé à la valeur 99 qui représente l'absence d'ambiguïté) et finalement un code de contrôle (initialisé à "non" qui signifie qu'il ne s'agit pas d'un verbe). Voici un court texte avant et après la préparation:

Pendant un demi-siècle, les bourgeois de Pont-l'Évêque enviaient à Mme Aubain sa servante Félicité.

Pour cent francs par an, elle faisait la cuisine et le ménage, cousait, lavait, repassait, savait brider un cheval, engraisser les volailles, battre le beurre, et resta fidèle à sa maîtresse, -qui cependant n'était pas une personne agréable.<sup>29</sup>

<sup>29</sup>Gustave Flaubert, *Trois contes*, Paris, Garnier-Flammarion, 1965, p.27.

<u>mot</u>	<u>ligne</u>	<u>ordre</u>	<u>chap.</u>	<u>amb.</u>
1 pendant	1	1	1	99 non
2 un	1	2	1	99 non
3 demi	1	3	1	99 non
4 siècle	1	4	1	99 non
5 les	1	5	1	99 non
6 bourgeoises	1	6	1	99 non
7 de	1	7	1	99 non
8 pont	1	8	1	99 non
9 l	1	9	1	99 non
10 évêque	1	10	1	99 non
11 envièrent	1	11	1	99 non
12 à	1	12	1	99 non
13 mme	1	13	1	99 non
14 aubain	1	14	1	99 non
15 sa	1	15	1	99 non
16 servante	1	16	1	99 non
17 félicité	1	17	1	99 non
18 pour	2	1	1	99 non
19 cent	2	2	1	99 non
20 francs	2	3	1	99 non
21 par	2	4	1	99 non
22 an	2	5	1	99 non
23 elle	2	6	1	99 non
24 faisait	2	7	1	99 non
25 la	2	8	1	99 non
26 cuisine	2	9	1	99 non
27 et	2	10	1	99 non
28 le	2	11	1	99 non
29 ménage	2	12	1	99 non
30 cousait	2	13	1	99 non
31 lavait	2	14	1	99 non
32 repassait	2	15	1	99 non
33 savait	2	16	1	99 non
34 brider	2	17	1	99 non
35 un	2	18	1	99 non
36 cheval	2	19	1	99 non
37 engraisser	2	20	1	99 non
38 les	2	21	1	99 non
39 volailles	2	22	1	99 non
40 battre	2	23	1	99 non
41 le	2	24	1	99 non
42 beurre	2	25	1	99 non
43 et	2	26	1	99 non
44 resta	2	27	1	99 non
45 fidèle	2	28	1	99 non
46 à	2	29	1	99 non
47 sa	2	30	1	99 non
48 maîtresse	2	31	1	99 non
49 qui	2	32	1	99 non
50 cependant	2	33	1	99 non
51 n	2	34	1	99 non
52 était	2	35	1	99 non
53 pas	2	36	1	99 non

<u>mots</u>	<u>ligne</u>	<u>ordre</u>	<u>chap.</u>	<u>amb.</u>	
54 une		2	37	1	99 non
55 personne		2	38	1	99 non
56 agréable		2	39	1	99 non

### 3.2 Affichage et correction du texte

Lorsqu'il a subi le traitement préalable, le texte est prêt à être soumis à L'Analyste. Lorsqu'on met le logiciel en marche le premier menu qui apparaît (menu principal) est le suivant:

#### L'ANALYSTE

1-	RECONNAISSANCE DES FORMES VERBALES POTENTIELLES
2-	DÉSAMBIGUISATION DES FORMES VERBALES POTENTIELLES
3-	DÉSAMBIGUISATION INTERACTIVE
4-	CORRECTION DU CORPUS
5-	AFFICHAGE DU CORPUS
6-	FIN DE LA SESSION
CHOIX ?	

Si nous voulons faire afficher notre texte à l'écran alors nous choisissons l'option 5. Après vérification du texte, s'il y a des erreurs que nous voulons corriger, alors l'option 4 nous permet de le faire: L'ANALYSTE nous demande le numéro qui correspond au mot que nous voulons corriger puis affiche le mot. Nous apportons la correction puis le mot est remis à sa place.

### 3.3 La reconnaissance des formes verbales potentielles

La première étape dans le processus de détection des formes verbales est la reconnaissance des formes verbales potentielles. En choisissant l'option 1, L'ANALYSTE exécute ce travail. Au préalable toutefois, nous devons nous assurer que les dictionnaires (triés et indexés) soient sur la même disquette (ou disque ou répertoire) que L'ANALYSTE. Nous devons nous assurer aussi que la disquette contient les fichiers *vdisk.sys* et *config.sys* (avec les instructions *files=20*, *buffers=15*, *device=vdisk 100* (qui crée un disque virtuelle dont L'ANALYSTE a besoin).

Lorsque L'ANALYSTE a terminé la reconnaissance des formes verbales potentielles, le texte ressemble à ceci (les flèches ne font pas partie du fichier, mais sont générées par L'Analyste lors de l'affichage):

<u>mots</u>	<u>ligne</u>	<u>ordre</u>	<u>chap.</u>	<u>amb.</u>
1 pendant	1	1	1	99 non
2 un	1	2	1	99 non
3 demi	1	3	1	99 non
4 siècle	1	4	1	99 non

<u>mot</u>	<u>ligne</u>	<u>ordre</u>	<u>chap.</u>	<u>amb.</u>	
5	les	1	5	1	99 non
6	bourgeoises	1	6	1	99 non
7	de	1	7	1	99 non
8	pont	1	8	1	99 non
9	l	1	9	1	99 non
10	evêque	1	10	1	99 non
11	envièrent	1	11	1 g6	2 oui <=
12	à	1	12	1	99 non
13	mme	1	13	1	99 non
14	aubain	1	14	1	99 non
15	sa	1	15	1	99 non
16	servante	1	16	1	99 non
17	félicité	1	17	1	99 non
18	pour	2	1	1	99 non
19	cent	2	2	1	99 non
20	francs	2	3	1	99 non
21	par	2	4	1	99 non
22	un	2	5	1	99 non
23	elle	2	6	1	99 non
24	faisait	2	7	1 e3	41 oui <=
25	la	2	8	1	99 non
26	cuisine	2	9	1 c1c3d1d3h2	2 oui <=
27	et	2	10	1	99 non
28	le	2	11	1	99 non
29	ménage	2	12	1 c1c3d1d3h2	0 oui <=
30	cousait	2	13	1 e3	99 oui <=
31	lavait	2	14	1 e3	2 oui <=
32	repassait	2	15	1 e3	99 oui <=
33	savait	2	16	1 e3	23 oui <=
34	brider	2	17	1	99 non
35	un	2	18	1	99 non
36	cheval	2	19	1	99 non
37	engraisser	2	20	1	99 non
38	les	2	21	1	99 non
39	volailles	2	22	1	99 non
40	battre	2	23	1	99 non
41	le	2	24	1	99 non
42	beurre	2	25	1 c1c3d1d3h2	0 oui <=
43	et	2	26	1	99 non
44	resta	2	27	1 g3	0 oui <=
45	fidèle	2	28	1	99 non
46	à	2	29	1	99 non
47	sa	2	30	1	99 non
48	maitresse	2	31	1	99 non
49	qui	2	32	1	99 non
50	cependant	2	33	1	99 non
51	n	2	34	1	99 non
52	était	2	35	1 e3	99 oui <=
53	pas	2	36	1	99 non
54	une	2	37	1	99 non
55	personne	2	38	1	99 non
56	agréable	2	39	1	99 non

Le code 99 indique l'absence d'ambiguïté. On peut voir que *envièrent* a été reconnu comme étant une forme verbale potentielle à la sixième personne du passé simple avec le code d'ambiguïté 2. Ce code indique une ambiguïté à la deuxième ou troisième personne du présent de l'indicatif (*envie(s)*: l'alternative serait un substantif au féminin singulier ou au féminin pluriel). *A faisait* correspond le code 41 qui montre une ambiguïté à la troisième personne du présent de l'indicatif. *A cuisine* l'ordinateur indique qu'il pourrait s'agir d'un verbe à la première ou troisième personne du présent de l'indicatif ou du subjonctif ou de l'impératif à la deuxième personne. Le code d'ambiguïté montre qu'il pourrait s'agir également d'un substantif au féminin singulier. *A ménage* et *beurre* il pourrait s'agir d'un verbe à la première ou troisième personne du présent de l'indicatif ou du subjonctif. Si ces mots n'étaient pas des verbes il s'agirait alors de substantifs au masculin singulier. Pour *cousait*, *repassait* et *était* il n'y a pas d'ambiguïté: il s'agit de verbes à la troisième personne de l'imparfait. *A lavait* une ambiguïté est marquée à la deuxième ou troisième personne du présent de l'indicatif ou du subjonctif (*lave, laves*). Pour *savait* l'ambiguïté est à la quatrième personne du présent de l'indicatif (*savons*). Pour *resta* la forme ambiguë est à la deuxième ou troisième personne du présent de l'indicatif ou du subjonctif (*reste, restes*).

Sur onze formes verbales potentielles trois sont déjà reconnues comme étant des verbes finis (*cousait, repassait, était*). Pour cinq verbes l'ambiguïté ne s'applique pas à leur forme actuelle (*envièrent, faisait, lavait, savait* et *resta*). Trois formes seulement sont ambiguës et nous pouvons constater que dans les trois cas il ne s'agit pas de verbes mais de substantifs.

On pourrait se demander pourquoi les cinq verbes dont l'ambiguïté relève d'autres formes ne sont pas déjà reconnus comme étant des verbes finis. À cette étape L'ANALYSTE n'a fait qu'une analyse morphologique partielle. Il a reconnu les formes verbales potentielles en consultant les dictionnaires des terminaisons et des radicaux et les seules informations dont il disposait étaient celles qui correspondaient aux radicaux.

### 3.4 Désambiguïsation des formes verbales potentielles

Après cette étape nous demandons à L'ANALYSTE d'exécuter la deuxième phase de l'analyse: la désambiguïsation des formes verbales potentielles (l'option 2). À la fin de la procédure notre texte ressemble à ceci:

	<u>mots</u>	<u>ligne</u>	<u>ordre</u>	<u>chap.</u>	<u>amb.</u>
1	pendant	1	1	1	99 non
2	un	1	2	1	99 non
3	demi	1	3	1	99 non
4	siècle	1	4	1	99 non
5	les	1	5	1	99 non
6	bourgeoises	1	6	1	99 non
7	de	1	7	1	99 non
8	pont	1	8	1	99 non
9	l	1	9	1	99 non
10	evêque	1	10	1	99 non
11	envièrent	1	11	1 g6	99 oui <=
12	à	1	12	1	99 non
13	mme	1	13	1	99 non
14	subain	1	14	1	99 non
15	sa	1	15	1	99 non
16	servante	1	16	1	99 non

	<u>mot</u>	<u>ligne</u>	<u>ordre</u>	<u>chap.</u>	<u>amb.</u>	
17	félicité	.	1	17	1	99 non
18	pour		2	1	1	99 non
19	cent		2	2	1	99 non
20	francs		2	3	1	99 non
21	par		2	4	1	99 non
22	an	,	2	5	1	99 non
23	elle		2	6	1	99 non
24	faisait		2	7	1 e3	99 oui <=
25	la		2	8	1	99 non
26	cuisine		2	9	1	100 non
27	et		2	10	1	99 non
28	le		2	11	1	99 non
29	ménage	,	2	12	1	100 non
30	cousait	,	2	13	1 e3	99 oui <=
31	lavait	,	2	14	1 e3	99 oui <=
32	repassait	,	2	15	1 e3	99 oui <=
33	savait		2	16	1 e3	99 oui <=
34	brider		2	17	1	99 non
35	un		2	18	1	99 non
36	cheval	,	2	19	1	99 non
37	engraisser		2	20	1	99 non
38	les		2	21	1	99 non
39	volailles	,	2	22	1	99 non
40	battre		2	23	1	99 non
41	le		2	24	1	99 non
42	beurre	,	2	25	1	100 non
43	et		2	26	1	99 non
44	resta		2	27	1 g3	99 oui <=
45	fidèle		2	28	1	99 non
46	à		2	29	1	99 non
47	sa		2	30	1	99 non
48	maitresse	,	2	31	1	99 non
49	qui		2	32	1	99 non
50	cependant		2	33	1	99 non
51	n	,	2	34	1	99 non
52	était		2	35	1 e3	99 oui <=
53	pas		2	36	1	99 non
54	une		2	37	1	99 non
55	personne		2	38	1	99 non
56	agréable	,	2	39	1	99 non

Tous les mots sont bien catégorisés. Par convention, le code 100, en guise de code d'ambiguïté, indique qu'il y avait ambiguïté après l'étape de reconnaissance des formes verbales potentielles et que l'analyse syntaxique a reconnu le mot comme n'étant pas un verbe. C'est le cas pour *cuisine*, *ménage* et *beurre*. Cela permet de tenir un registre de ces mots.

La procédure semble parfaite mais cela n'est pas exact. Avec un texte de 1024 mots nous avons relevé six formes verbales pour lesquelles l'ambiguïté persistaient: peint (verbe à la troisième personne du présent de l'indicatif ou participe passé), close (verbe à la première ou troisième personne du présent du subjonctif ou participe passé), cuit (verbe à la troisième personne du présent de l'indicatif ou participe passé), livres (verbe à la deuxième personne du présent de l'indicatif ou du sub-

jonctif ou substantif), envie (verbe à la première ou troisième personne du présent de l'indicatif ou du subjonctif ou substantif) et dit (troisième personne du présent de l'indicatif ou troisième personne du passé simple).

Il faut indiquer, et cela dénote une certaine fragilité du travail de L'ANALYSTE, que pour le mot *livres* il s'agit ici d'une erreur. Le mot qui précédait notre forme verbale potentielle était *douz* (et non *douze*: une faute au moment de la saisie du texte ayant échappé à notre vigilance). S'il n'y avait pas eu de faute L'ANALYSTE aurait reconnu en *douze* un déterminant et aurait précisé la forme comme n'étant assurément pas une forme verbale.

### 3.5 Désambiguïsation interactive

Avec l'option 3 (désambiguïsation manuelle des formes verbales potentielles) l'ordinateur nous présente toutes les formes pour lesquelles une ambiguïté persiste. Apparaîtra par exemple à l'écran la suite de mots: "répondre et avait envie de s'enfuir aussitôt", avec le message suivant: "Ce mot doit-il être considéré comme étant un verbe? (O/N/A)". Si nous pressons, au clavier, la touche "O" cela indique à l'ordinateur qu'il doit prendre le mot pour un verbe. Si nous frappons la touche "N" cela signifie qu'il ne s'agit pas d'un verbe. L'option "A" indique à l'ordinateur que nous ne prenons aucune décision. Lorsque l'ambiguïté porte sur le temps ou la personne de la forme verbale alors l'ordinateur transmet le message suivant: "Quel est le nouveau code d'information?" Nous n'avons alors qu'à indiquer un nouveau code d'information. S'il pouvait s'agir soit de la troisième personne du présent de l'indicatif soit de la troisième personne du passé simple alors le code était c3g3. Si nous voulons préciser qu'il s'agit en fait de la troisième personne du passé simple alors nous inscrivons le code g3.

À la fin du traitement le taux de reconnaissance est très élevé. Certains mots ont été mal analysés comme *peut* dans "peut-être" et *soit* dans "soit! Allez-y". Il serait possible toutefois d'ajouter à la procédure une liste de mots qui doivent obligatoirement être présentés à l'opérateur lors de la désambiguïsation manuelle complémentaire.

### 3.6 Le module d'analyse statistique

Après l'étape de reconnaissance des verbes nous activons le module d'analyse statistique de *L'Analyste*. Un premier menu apparaît à l'écran avec les options suivantes: 1- Analyse, 2- Affichage, 3- Sauvegarde, 4- Récupération, 5- Fin. Si nous voulons consulter un dossier déjà existant alors l'option 4 nous permet de le récupérer. Si nous voulons conserver un nouveau dossier alors l'option 3 nous permet de le faire.

#### 3.6.1 L'analyse

L'option 1 nous permet de procéder à l'analyse statistique. L'ordinateur nous présente alors un nouveau menu avec les options suivantes: 1- Choisir une partie du texte, 2- Le texte en entier, 3- Retour au menu général. Si nous choisissons l'option 2, alors l'ordinateur procédera immédiatement à l'analyse de tout le texte. Si nous choisissons l'option 1 alors il nous demande à partir de quel chapitre et quelle phrase l'analyse doit commencer. Lorsque nous lui avons donné cette information il nous demande si nous voulons préciser une coordonnée pour la fin de l'analyse ou si nous voulons que celle-ci se termine à la fin du texte. Si nous voulons préciser la fin, alors il nous demande d'inscrire les numéros du chapitre et de la phrase. L'ordinateur procède alors à l'analyse de la partie de texte que nous avons spécifiée. À la fin de son analyse l'ordinateur nous indique qu'il a terminé et nous ramène au menu général.

### 3.6.2 L'affichage des résultats

En choisissant l'option 2 (Affichage) l'ordinateur nous demande pour quels temps de verbe nous voulons qu'il dessine le graphique des fréquences par tranche de texte. Lorsqu'il a reçu une réponse il dessine le graphique et nous présente le menu suivant: 1- Focus, 2- Information, 3- Transitions temporelles, 4- Impression, 5- Fin.

#### 3.6.2.1 La focalisation

L'option 1 nous permet une focalisation sur une partie du graphique. En la choisissant, l'ordinateur nous demande à partir de quelle tranche et jusqu'à quelle tranche nous voulons faire faire une nouvelle analyse. Lorsqu'il a cette information il régénère ses données en ne considérant que la nouvelle partie du texte et nous représente un nouveau graphique accompagné du même menu.

#### 3.6.2.2 Coordonnées du texte, valeurs, fréquences et légende

Si on choisit l'option 2 alors l'ordinateur nous présente un nouveau menu avec les options suivantes: 1- Légende, 2- Valeurs, 3- Position dans le texte, 4- Fin. Avec l'option 1 l'ordinateur nous présente la légende suivante:

a => futur  
 b => conditionnel  
 c => présent de l'indicatif  
 d => présent du subjonctif  
 e => imparfait de l'indicatif  
 f => imparfait du subjonctif  
 g => passé simple  
 h => impératif

Cette légende est un aide-mémoire pour les codes apparaissant dans le graphique. Si nous choisissons l'option 2 alors l'ordinateur nous demande pour quels temps de verbe nous voulons avoir les valeurs. Après lui avoir fourni l'information demandée il nous affiche les valeurs. Avec l'option 3 l'ordinateur nous indique à quelle proportion du texte correspond son analyse (cela est surtout important lorsque nous nous servons du focus). L'option 4 nous ramène au menu précédent.

Si nous choisissons l'option 3 alors l'ordinateur nous présente la matrice des transitions temporelles avec le menu suivant: 1- Fréquences, 2- Impression, 3- Fin. Si nous choisissons l'option 1 alors l'ordinateur nous présente la table des fréquences pour tous les temps de verbe. Si nous choisissons l'option 2 alors l'ordinateur imprime la matrice. L'option 3 nous ramène au graphique.

L'option 4 permet de faire imprimer le graphique. L'option 5 met fin à l'affichage et nous ramène au menu général.

## CHAPITRE 3

### ANALYSE AUTOMATIQUE DE *UN COEUR SIMPLE* DE GUSTAVE FLAUBERT

L'objectif que nous nous étions fixé dans le cadre de nos travaux était de réaliser un algorithme et un programme qui seraient en mesure de reconnaître les verbes dans un texte littéraire, de faire une analyse statistique de ceux-ci, conformément aux besoins d'une analyse de la récurrence des verbes d'après le modèle proposé par Harald Weinrich, et de présenter des graphiques ou tableaux qu'un littéraire, pas nécessairement familier avec le ou les langages des ordinateurs, puisse facilement comprendre, commenter et interpréter. Jusqu'à maintenant nous avons décrit notre algorithme: les dictionnaires, la procédure de reconnaissance des formes verbales potentielles, les divers tests de désambiguïsation (test des particules verbales, test du substantif, test des participes passés en -s etc.), la procédure de désambiguïsation complémentaire assistée par l'ordinateur et le module d'analyse statistique. Nous avons présenté également certaines sources de difficultés comme le tri alphabétique et l'indexation des dictionnaires. Finalement nous avons décrit le fonctionnement du logiciel issu de nos travaux et démontré la qualité de son analyse en ce qui a trait à la reconnaissance des formes verbales. Maintenant, en guise de démonstration de l'utilité de L'Analyste pour une analyse de la récurrence des verbes dans un texte littéraire nous procéderons à une brève analyse de *Un coeur simple* de Gustave Flaubert.

L'analyse comportera quatre étapes. D'abord, nous ferons une analyse du relief dans le conte. Nous ferons un balayage de chacun des chapitres afin d'extraire les passages qui contiennent une grande densité de verbes au passé simple. Le passé simple étant le temps de l'action, nous supposons être en mesure d'obtenir ainsi les moments les plus importants de l'histoire. Ensuite, nous traiterons du tempo narratif. Nous présenterons un texte qui se caractérise par un tempo narratif rapide et un autre au tempo narratif plus lent. Nous traiterons ensuite de l'attitude de locution. Nous présenterons un texte à l'intérieur duquel il y a deux attitudes de locution. Finalement, nous discuterons des transitions temporelles. Nous parlerons alors principalement de la mise en relief et de l'attitude de locution.

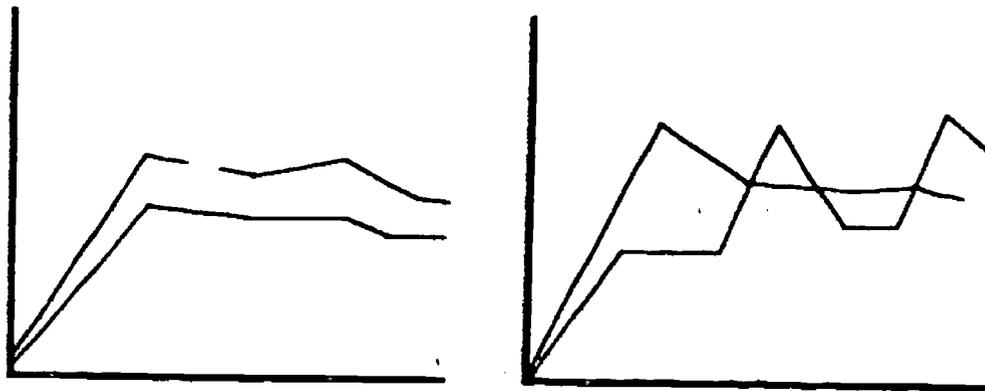
#### 1. Le relief

Dans un récit, il y a un certain nombre d'éléments que l'auteur désire mettre en arrière-plan parce qu'ils n'ont pas une valeur essentielle dans l'histoire et d'autres, plus importants, qu'il veut mettre en évidence. Pour ce faire, deux temps sont utilisés: l'imparfait et le passé simple. L'imparfait est le temps de l'arrière-plan et le passé simple est le temps du premier plan.

À partir des graphiques que trace L'Analyste, un verbe au passé simple entouré d'un grand nombre de verbes à l'imparfait passe généralement inaperçu lors d'une première observation. Cela ne signifie pas pour autant qu'il ne joue pas un rôle important dans son contexte. Pour faire ressortir

d'un graphique des verbes isolés, il s'agit de faire une focalisation ou de consulter la table des valeurs. C'est d'ailleurs ce que nous ferons lorsque nous aborderons la mise en relief du point de vue des transitions temporelles.

Ce qui ressort d'une première observation d'un graphique, pour autant que celui-ci couvre une partie de texte suffisamment vaste, par exemple un chapitre, ce sont les tendances générales du comportement verbal. Pour mieux comprendre ce que nous entendons par tendance générale, voici deux graphiques fictifs représentant la densité des verbes au passé simple et à l'imparfait dans un texte:



Dans le premier graphique, on observe des comportements constants. Nous dirons que la tendance générale des verbes à l'imparfait et au passé simple se maintient du début à la fin: les verbes à l'imparfait dominent ceux au passé simple.

Dans le second graphique, on constate que le comportement des formes verbales varie. À deux endroits, par exemple, il y a recrudescence de la densité des verbes au passé simple.

Du point de vue du relief, ce qui nous intéressera, c'est la courbe correspondant au passé simple. Quelle interprétation pouvons-nous faire de la variation de la densité des verbes au passé simple? D'abord, rappelons que le passé simple est le temps du premier plan et l'imparfait le temps de l'arrière-plan. Nous pouvons donc dire qu'un passage qui se caractérise par un grand nombre de verbes à l'imparfait et un petit nombre de verbes au passé simple contient plutôt des éléments secondaires à l'histoire. Inversement, un passage qui se caractérise par un grand nombre de verbes au passé simple contient plutôt des éléments importants de l'histoire.

Cela dit, il faut voir une histoire comme étant un enchaînement d'événements, d'actions, qui sont généralement constitués d'un ensemble de phrases et non pas d'une seule. Or, la probabilité qu'une partie d'une courbe, où il y a une accentuation de la densité des verbes au passé simple, corresponde à un événement majeur, ou tout au plus à un nombre restreint d'événements majeurs, est très forte.

L'hypothèse que nous émettons est la suivante: à l'ensemble des passages d'un texte qu'un graphique présente comme ayant une densité plus grande qu'à la normale de verbes au passé simple correspond les moments les plus importants de l'histoire.

Ce que nous nous proposons de faire est d'extraire de *Un coeur simple* les passages qui se caractérisent par une grande densité de verbes au passé simple. Nous devrions obtenir, en mettant côte à côte chacun des passages, les moments les plus importants du conte. Nous demanderons à L'Analyse de faire une analyse de chacun des chapitres du conte. Pour chaque chapitre, nous présenterons les résultats que le logiciel nous a transmis. Ces résultats seront un graphique, une table

des valeurs (qui rend la précision que le graphique ne peut donner) et une table des fréquences. Nous commenterons ensuite la table des fréquences et présenterons chacun des passages qui ressortent du graphique.

Lors de notre analyse nous utiliserons fréquemment deux expressions: segment et tranche. Nous entendons par tranche, une partie du texte dont les limites sont imposées par L'Analyste. Dans un graphique il y a toujours seize tranches de longueur égale. Nous entendons par segment une partie du texte qui est composée par une ou plusieurs tranches. Par exemple, si nous regardons un graphique et constatons que dans trois tranches consécutives il y a une prédominance des verbes au passé simple, nous parlerons alors d'un segment de texte, couvrant trois tranches, où il y a une prédominance des verbes au passé simple.

## 1.1 Chapitre 1

### 1- TABLE DES VALEURS:

	FUT.	CON.	PR.IND	PR.SUB	IMPA.IND.	IMPA.SUB.	PA.SIM.	IMPE.
tranche 1 =>	0	0	0	0	5	0	1	0
tranche 2 =>	0	0	0	0	2	0	1	0
tranche 3 =>	0	0	0	0	1	0	1	0
tranche 4 =>	0	0	0	0	1	0	1	0
tranche 5 =>	0	0	0	0	4	0	0	0
tranche 6 =>	0	0	0	0	2	0	0	0
tranche 7 =>	0	0	0	0	4	0	0	0
tranche 8 =>	0	0	0	0	1	0	0	0
tranche 9 =>	0	0	0	0	4	0	0	0
tranche 10 =>	0	0	0	0	2	0	0	0
tranche 11 =>	0	0	0	0	1	0	0	0
tranche 12 =>	0	0	0	0	3	0	0	0
tranche 13 =>	0	0	0	0	3	0	0	0
tranche 14 =>	0	0	0	0	3	0	0	0
tranche 15 =>	0	0	0	0	1	0	0	0
tranche 16 =>	0	0	0	0	3	0	1	0

### 2- TABLE DES FRÉQUENCES

futur	=>	0
conditionnel	=>	0
présent	=>	0
présent du subjonctif	=>	0
imparfait	=>	40
imparfait du subjonctif	=>	0
passé simple	=>	5
impératif	=>	0

#### 1.1.1 Fréquences

La table des fréquences indique qu'il y a quarante verbes à l'imparfait de l'indicatif et cinq au passé simple. Le monde commenté est totalement absent du premier chapitre. À l'intérieur du monde raconté, quarante verbes appartiennent à l'arrière-plan et cinq au premier plan.

La présence intense de l'arrière-plan au tout début du conte correspond à ce que Weinrich appelle l'ouverture du texte: "Au début d'une histoire, on ne peut se passer totalement d'exposition; aussi le récit a-t-il normalement une introduction, où le temps est le plus souvent de l'arrière-plan."<sup>30</sup> Le premier chapitre du conte est effectivement une introduction. Les personnages principaux y sont présentés. On y apprend que Mme Aubain n'est "pas une personne agréable", qu'elle a deux enfants et que son mari est mort très jeune. Félicité est une servante. Elle se contente d'un petit salaire, mange peu et travaille beaucoup.

### 1.1.2 Le graphique

Le premier chapitre ne comprend pas suffisamment de mots (585) pour que l'Analyste puisse tracer un graphique. Toutefois nous pouvons juxtaposer les valeurs correspondant aux seize tranches, qui auraient été utilisées pour former le graphique:

Tranche	imparfait	passé simple
1	5	1
2	2	1
3	1	1
4	1	1
5	4	0
6	2	0
7	4	0
8	1	0
9	4	0
10	2	0
11	1	0
12	3	0
13	3	0
14	3	0
15	1	0
16	3	1

En comparant les valeurs, on constate que l'imparfait domine quantitativement le passé simple tout le long du chapitre sauf pour une partie qui est représentée par les tranches trois et quatre. L'Analyste nous informe que ces tranches correspondent à la troisième et quatrième phrases du chapitre (soit les deux dernières phrases du troisième paragraphe):

Alors elle vendit ses immeubles, sauf la ferme de Touques et la ferme de Geffosses, dont les rentes montaient à 5000 francs tout au plus, et elle quitta sa maison de Saint-Melaine pour en habiter une autre moins dispendieuse, ayant appartenu à ses ancêtres et placée derrière les halles.<sup>31</sup>

Dans ce court passage, on peut voir comment Mme Aubain a redressé une situation financière peu reluisante après la mort de son mari. Les deux verbes au passé simple introduisent deux éléments importants: "elle vendit ses immeubles", "elle quitta sa maison". C'est vraisemblablement après ces

<sup>30</sup>Harald Weinrich, *op. cit.*, p.115

<sup>31</sup>Gustave Flaubert, *op. cit.*, p.27.

deux événements que l'histoire commence. La femme que nous connaissons tout le long du conte, est celle qui doit pourvoir seule aux besoins de sa famille. Une femme qui n'a pas le droit de succomber à ses émotions. La vente des immeubles et le départ de sa maison sont probablement les toutes premières décisions d'une femme seule.

## 1.2 Chapitre 2

1- GRAPHIQUE (voir page 52)

2- TABLE DES VALEURS:

	FUT.	CON.	PR.IND	PR.SUB	IMPA.IND.	IMPA.SUB.	PA.SIM.	IMPE.
tranche 1 =>	0	0	0	0	11	0	10	0
tranche 2 =>	0	0	1	0	12	0	16	0
tranche 3 =>	0	1	0	0	11	0	13	0
tranche 4 =>	0	2	0	0	7	0	14	0
tranche 5 =>	0	0	1	0	14	0	4	0
tranche 6 =>	0	0	1	0	15	0	0	0
tranche 7 =>	0	0	2	0	13	0	3	0
tranche 8 =>	0	0	1	0	18	0	3	1
tranche 9 =>	0	0	0	0	18	0	10	3
tranche 10 =>	0	0	0	0	7	1	14	0
tranche 11 =>	0	0	0	0	13	1	11	0
tranche 12 =>	0	0	0	0	10	1	1	0
tranche 13 =>	0	1	0	0	10	0	6	0
tranche 14 =>	0	0	0	0	20	0	0	0
tranche 15 =>	0	0	0	0	18	0	3	0
tranche 16 =>	0	1	0	0	12	0	11	0

4- TABLE DES FRÉQUENCES:

futur	=>	0
conditionnel	=>	5
présent	=>	7
présent du subjonctif	=>	0
imparfait	=>	209
imparfait du subjonctif	=>	3
passé simple	=>	119
impératif	=>	4

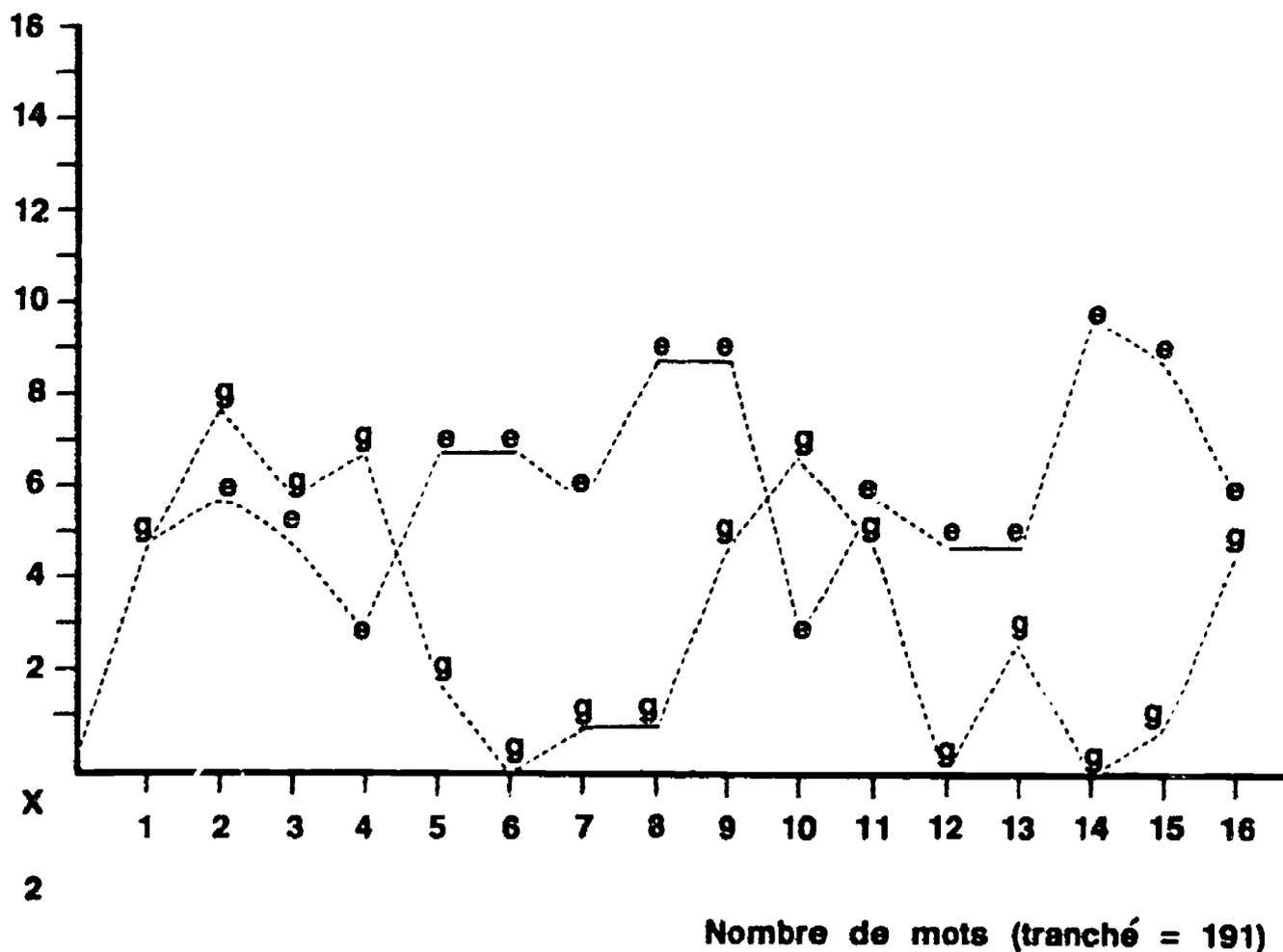
### 1.2.1 Fréquences

La table des fréquences nous indique qu'il y a sept verbes au présent de l'indicatif, 209 à l'imparfait, 119 au passé simple et finalement quatre à l'impératif. Nous avons onze verbes qui appartiennent au monde commenté en regard de 328 qui appartiennent au monde raconté. A l'intérieur du monde raconté, nous avons 119 verbes qui appartiennent au premier plan en regard de 209 qui appartiennent à l'arrière-plan.

### 1.2.2 Graphique

Le graphique nous montre quatre segments où on peut observer un nombre assez dense de verbes au passé simple. Le premier couvre les quatre premières tranches du graphique. Il y a quarante-et-un verbes à l'imparfait en regard de cinquante-trois au passé simple. Le second couvre les tranches neuf, dix et onze. Il y a trente-huit verbes à l'imparfait et trente-cinq au passé simple. Le troisième, correspond à la tranche treize. Il y a dix verbes à l'imparfait en regard de six au graphique passé simple. Finalement, le quatrième correspond à la dernière tranche du graphique. Il y a douze verbes à l'imparfait et onze au passé simple.

#### Nombre de verbes



e = imparfait  
g = passé simple

#### 1.2.2.1 Le premier segment:

Puisque ce segment correspond à une partie considérable du texte, soit le quart du chapitre, nous demandons à L'Analyse de faire une focalisation sur ce passage. Le segment ne contient pas suffisamment de mots pour que L'ANALYSTE puisse faire un graphique cependant nous pouvons consulter les tables des valeurs:

Tranches	Imparfait	Passé simple
1	3	4
2	3	3
3	3	2
4	2	1
5	1	5
6	3	4
7	3	3
8	5	4
9	5	2
10	2	4
11	2	3
12	2	3
13	2	2
14	2	3
15	0	7
16	3	3

Nous pouvons discerner cinq blocs. Le premier, de la tranche un à la tranche quatre, comprend à peu près autant de verbes à l'imparfait que de verbes au passé simple. Le deuxième, de la tranche cinq à la tranche sept, comprend plus de verbes au passé simple que de verbes à l'imparfait. Le troisième, les tranches huit et neuf, comprend plus de verbes à l'imparfait que de verbes au passé simple. Le quatrième, de la tranche dix à la tranche quinze, comprend plus de verbes au passé simple que de verbes à l'imparfait. Finalement le dernier bloc, la tranche seize, comprend autant de verbes à l'imparfait que de verbes au passé simple.

D'abord, après vérification, nous nous sommes aperçu que le segment de texte (l'ensemble des tranches) correspond à l'histoire de Félicité et de son amoureux. Le premier bloc de texte correspond à une introduction au récit à venir:

Elle avait eu, comme une autre, son histoire d'amour.

Son père, un maçon, s'était tué en tombant d'un échafaudage. Puis sa mère mourut, ses soeurs se dispersèrent, un fermier la recueillit, et l'employa toute petite à garder les vaches dans la campagne. Elle grelottait sous des haillons, buvait à plat ventre l'eau des mares, à propos de rien était battue, et finalement fut chassée pour un vol de trente sols, qu'elle n'avait pas commis. Elle entra dans une autre ferme, y devint fille de basse-cour, et, comme elle plaisait aux patrons, ses camarades la jalousaient.

Un soir du mois d'août (elle avait alors dix-huit ans), ils l'entraînèrent à l'assemblée de Colleville. Tout de suite elle fut étourdie, stupéfaite par le tapage des ménestriers, les lumières dans les arbres, la bigarrure des costumes, les dentelles, les croix d'or, cette masse de monde sautant à la fois, Elle se tenait à l'écart modestement, quand un jeune homme d'apparence cossue, et qui fumait sa pipe les deux coudes sur le timon d'un banneau, vint l'inviter à la danse.<sup>32</sup>

<sup>32</sup>Gustave Flaubert, *op. cit.*, p.30.

Il n'y a pas prédominance de l'imparfait sur le passé simple, mais le pourcentage d'imparfaits pour ce bloc est supérieur au pourcentage d'imparfaits pour le segment au complet.

Les deuxième, troisième et quatrième blocs correspondent au récit de l'aventure de Félicité avec son amoureux. En fait nous devrions parler d'un seul bloc où les temps du récit dominant particulièrement avec, à peu près au centre, une ponctuation, qui est constituée d'une séquence où ce sont les temps de l'arrière-plan qui dominant. L'Analyse, par son mode de fonctionnement, n'a pas été en mesure de le détecter. Toutefois notre attention est portée vers cette ponctuation: L'Analyse indiquant tout de même un passage avec une recrudescence de l'imparfait. La ponctuation dans ce récit est le passage suivant:

Elle baissa la tête. Alors il lui demanda si elle pensait au mariage. Elle reprit, en souriant, que c'était mal de se moquer. -"Mais non, je vous jure!" et du bras gauche il lui entourait la taille; elle marchait soutenue par son étreinte; ils se ralentirent. Le vent était mou, les étoiles brillaient, l'énorme charretée de foin oscillait devant eux; et les quatre chevaux, en trainant leurs pas, soulevaient de la poussière. Puis, sans commandement, ils tournèrent à droite. Il l'embrassa encore une fois. Elle disparut dans l'ombre.<sup>33</sup>

On constate dans ce passage un ralentissement marqué du tempo narratif, qui s'accorde bien avec la tendance romantique du passage. De plus nous sommes tenté de parler de relief étant donné le changement abrupt du comportement des formes verbales: cet événement revêt sûrement une grande importance dans la vie de Félicité.

Le dernier bloc constitue en quelque sorte une conclusion au récit de Félicité et de son amoureux: "(...) Cependant elle se trouvait heureuse. La douceur du milieu avait fondu sa tristesse."<sup>34</sup>

#### 1.2.2.2 Le deuxième segment:

Pour le deuxième segment, nous demandons à L'Analyse de faire une focalisation. Les résultats que nous obtenons sont constants tout le long du passage: les verbes au passé simple sont plus nombreux que ceux à l'imparfait dans une proportion de deux contre un.

Dans ce passage, Mme Aubain, Félicité et les enfants se font attaquer par un taureau:

C'était un taureau, que cachait le brouillard. Il avança vers les deux femmes. Mme Aubain allait courir. -"Non! non! moins vite!" Elles pressaient le pas cependant, et entendaient par-dérrière un souffle sonore qui se rapprochait. Ses sabots, comme des marteaux, battaient l'herbe de la prairie; voilà qu'il galopait maintenant! Félicité se retourna...<sup>35</sup>

<sup>33</sup> Gustave Flaubert, op. cit., p.32.

<sup>34</sup> Ibid., p.34.

<sup>35</sup> Gustave Flaubert, op. cit., p.37.

Le tempo narratif y est nettement plus rapide que dans les passages précédents. Tout se déroule très rapidement. L'auditeur est tenu en haleine du début à la fin de l'événement tragique, observant chacun des mouvements des personnages.

### 1.2.2.3 Le troisième segment

Le troisième segment marque l'arrivée de Mme Aubain, Félicité et les enfants à Trouville où Virginie allait reprendre des forces après l'aventure avec le taureau. Cette dernière avait été très troublée par ce fâcheux événement.

On fut encore une demi-heure avant d'atteindre Trouville. La petite caravane mit pied à terre pour passer les Ecores; c'était une falaise surplombant des bateaux; et trois minutes plus tard, au bout du quai, on entra dans la cour de l'Agneau d'or, chez la mère David.<sup>36</sup>

### 1.2.2.4 Le quatrième segment

Le deuxième chapitre se termine par une séparation. Pour assurer à son fils une bonne instruction, Mme Aubain se soumet à l'idée de l'envoyer au collège:

M. Bourais l'éclaira sur le choix d'un collège. Celui de Caen passait pour le meilleur. Paul y fut envoyé; et fit bravement ses adieux, satisfait d'aller vivre dans une maison où il aurait des camarades.

Mme Aubain se résigna à l'éloignement de son fils, parce qu'il était indispensable. Virginie y songea de moins en moins. Félicité regrettait son tapage. Mais une occupation vint la distraire; à partir de Noël, elle mena tous les jours la petite fille au catéchisme.<sup>37</sup>

## 1.3 Chapitre 3

1- GRAPHIQUE (voir page 55)

2- TABLE DES VALEURS:

	FUT.	CON.	PR.IND	PR.SUB	IMPA.IND.	IMPA.SUB.	PA.SIM.	IMPE.
tranche 1 =>	0	0	6	0	21	0	4	0
tranche 2 =>	0	0	0	0	17	0	15	0
tranche 3 =>	0	0	0	1	14	1	18	1

<sup>36</sup> *ibid.*, p. 41.

<sup>37</sup> Gustave Flaubert, *op. cit.*, p. 44.

	FUT.	CON.	PR.IND	PR.SUB	IMPA.IND.	IMPA.SUB.	PA.SIM.	IMPE.
tranche 4 =>	0	0	0	0	25	0	4	0
tranche 5 =>	0	2	1	0	12	0	17	0
tranche 6 =>	6	2	0	0	25	0	11	0
tranche 7 =>	0	0	2	0	21	1	8	1
tranche 8 =>	0	0	3	0	17	0	15	0
tranche 9 =>	0	0	1	0	14	0	18	0
tranche 10 =>	0	0	2	0	19	1	7	1
tranche 11 =>	0	1	1	0	15	0	20	0
tranche 12 =>	0	1	2	0	18	3	16	0
tranche 13 =>	0	1	1	0	28	0	5	0
tranche 14 =>	0	1	0	0	16	1	10	0
tranche 15 =>	0	0	0	0	18	0	17	0
tranche 16 =>	0	1	1	0	23	0	8	0

### 3- TABLE DES FRÉQUENCES:

futur	=>	0
conditionnel	=>	9
présent	=>	20
présent du subjonctif	=>	1
imparfait	=>	302
imparfait du subjonctif	=>	7
passé simple	=>	193
impératif	=>	3

#### 1.3.1 Fréquences

La table des fréquences nous indique qu'il y a vingt verbes au présent de l'indicatif, trois à l'impératif, 302 à l'imparfait de l'indicatif et 193 au passé simple. Nous avons donc vingt-trois verbes qui appartiennent au monde commenté en regard de 495 qui appartiennent au monde raconté. A l'intérieur du monde raconté nous avons 192 verbes appartenant au premier plan en regard de 302 appartenant à l'arrière-plan.

#### 1.3.2 Le graphique

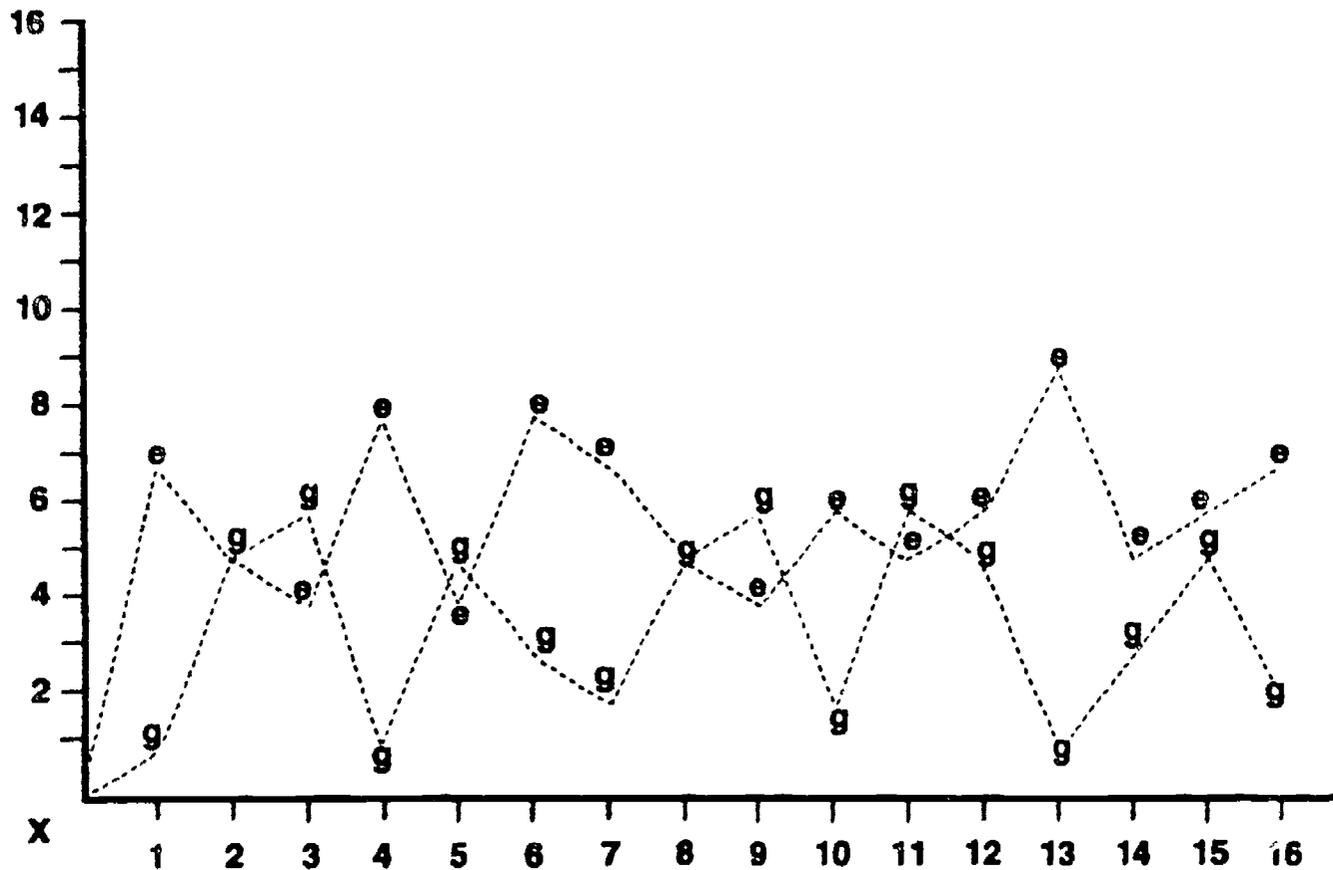
Dans le graphique nous pouvons observer cinq segments où le nombre de verbes au passé simple augmente beaucoup par rapport aux données générales du chapitre. Le premier couvre les tranches deux et trois du graphique. Il y a trente-trois verbes au passé simple en regard de vingt-sept à l'imparfait. Le second segment couvre la tranche cinq. Il y a plus de quinze verbes au passé simple en regard de douze à l'imparfait. Le troisième segment couvre les tranches huit et neuf. Il y a trente-trois verbes au passé simple en regard de vingt-sept à l'imparfait. Le quatrième segment couvre les tranches onze et douze. Il y a trente-trois verbes au passé simple et trente-trois à l'imparfait. Le cinquième segment couvre la tranche quinze. Il y a quinze verbes au passé simple et dix-huit à l'imparfait.

## 1.3.2.1 Le premier segment

Au premier segment, il se produit deux choses. Tout d'abord, Félicité assiste à la première communion de Virginie. Ici, cet événement revêt une importance toute spéciale. A travers Virginie, Félicité vit l'événement comme si c'était elle qui participait à la cérémonie "son éducation religieuse ayant été négligée dans sa jeunesse"<sup>38</sup>. Au moment culminant de la cérémonie elle se sentait vraiment à la place de Virginie: "Quand ce fut le tour de Virginie, Félicité se pencha pour la voir; et, avec l'imagination que donnent les vraies tendresses, il lui sembla qu'elle était elle-même cette enfant; sa figure devenait la sienne, sa robe l'habillait, son coeur lui battait dans la poitrine;"<sup>39</sup>.

Ensuite, vers la fin du segment, Virginie part pour le couvent. Sa mère, Mme Aubain, voulant pour sa fille une bonne instruction se résigna à l'envoyer chez les Ursulines de Honfleur. Le départ de la petite fut très douloureux tant pour Félicité que pour Mme Aubain.

## Nombre de verbes



Nombre de mots (tranché = 279)

e = imparfait

g = passé simple

<sup>38</sup>Gustave Flaubert, *op. cit.*, p. 46.

<sup>39</sup>*Ibid.*, p. 47.

### 1.3.2.2 Le deuxième segment

Au deuxième segment, Victor part pour un voyage en Amérique qui doit durer plus de deux ans. Félicité n'aime pas l'idée de voir partir son neveu pour un si long voyage: "La perspective d'une telle absence désola Félicité; et pour lui dire encore adieu, le mercredi soir, après le dîner de Madame, elle chaussa des galoches, et avala les quatre lieues qui séparent Pont-l'Evêque de Honfleur"<sup>40</sup>.

### 1.3.2.3 Le troisième segment

Au troisième segment, le nombre de verbes au passé simple augmente. Un événement qui bouleverse Félicité se produit. Liébard vient porter une lettre à Félicité: Victor est mort. "Il était mort, on n'en disait pas davantage"<sup>41</sup>. En perdant Victor, Félicité perd tout ce qu'elle a de famille. Après le décès de son neveu, Félicité n'entend plus parler de sa soeur.

### 1.3.2.4 Le quatrième segment

Au quatrième segment, Mme Aubain et Félicité apprennent le décès de Virginie: "La bonne soeur avec un air de componction dit qu' "elle venait de passer". En même temps, le glas de Saint-Léonard redoublait"<sup>42</sup>. Bouleversée, Félicité veille la morte pendant deux nuits: "Pendant deux nuits, Félicité ne quitta pas la morte. Elle répétait les mêmes prières, jetait de l'eau bénite sur les draps, revenait s'asseoir, et la contemplait"<sup>43</sup>.

### 1.3.2.5 Le cinquième segment

Au cinquième segment il y a plus de verbes à l'imparfait que de verbes au passé simple, toutefois le graphique montre une densité de verbes au passé simple nettement plus grande que la moyenne du chapitre. Dans ce segment, Félicité fait la connaissance du baron de Larsonnière et de sa femme. Cette rencontre aura pour Félicité une très grande importance. C'est d'eux que Félicité obtiendra son perroquet, qui l'accompagnera jusque dans sa mort:

Un sous-préfet nouveau, peu de jours après, fut nommé: le baron de Larsonnière, ex-consul en Amérique, et qui avait chez lui, outre sa femme, sa belle-soeur avec trois demoiselles, assez grandes déjà. On les apercevait sur leur gazon, habillées de blouses flottantes; elles possédaient un nègre et un perroquet.<sup>44</sup>

<sup>40</sup> *Ibid.*, p. 51.

<sup>41</sup> Gustave Flaubert, *op. cit.*, p. 55.

<sup>42</sup> *Ibid.*, p. 59.

<sup>43</sup> *Ibid.*, p. 59.

<sup>44</sup> Gustave Flaubert, *op. cit.*, p. 62.

## 1.4. Chapitre 4

## 1- GRAPHIQUE (voir page 60)

## 2- TABLE DES VALEURS:

	FUT.	CON.	PR.IND	PR.SUB	IMPA.IND.	IMPA.SUB.	PA.SIM.	IMPE.
tranche 1 =>	0	0	2	0	16	1	3	0
tranche 2 =>	0	0	0	0	19	1	2	0
tranche 3 =>	0	1	2	0	10	1	9	1
tranche 4 =>	0	1	1	0	10	0	8	0
tranche 5 =>	0	0	1	0	15	0	2	0
tranche 6 =>	0	0	0	0	15	0	6	1
tranche 7 =>	0	0	0	0	10	0	10	0
tranche 8 =>	0	0	0	0	8	0	14	0
tranche 9 =>	1	1	0	0	9	0	3	0
tranche 10 =>	0	0	1	0	14	0	3	0
tranche 11 =>	0	0	1	0	6	0	5	0
tranche 12 =>	0	0	0	0	7	0	14	0
tranche 13 =>	0	0	1	0	11	1	11	0
tranche 14 =>	0	0	0	0	15	1	3	0
tranche 15 =>	0	1	0	0	9	0	14	0
tranche 16 =>	0	0	2	0	15	1	10	3

## 3- TABLE DES FRÉQUENCES:

futur	=>	1
conditionnel	=>	4
présent	=>	11
présent du subjonctif	=>	0
imparfait du subjonctif	=>	6
imparfait	=>	189
passé simple	=>	117
impératif	=>	5

## 1.4.1 Fréquences

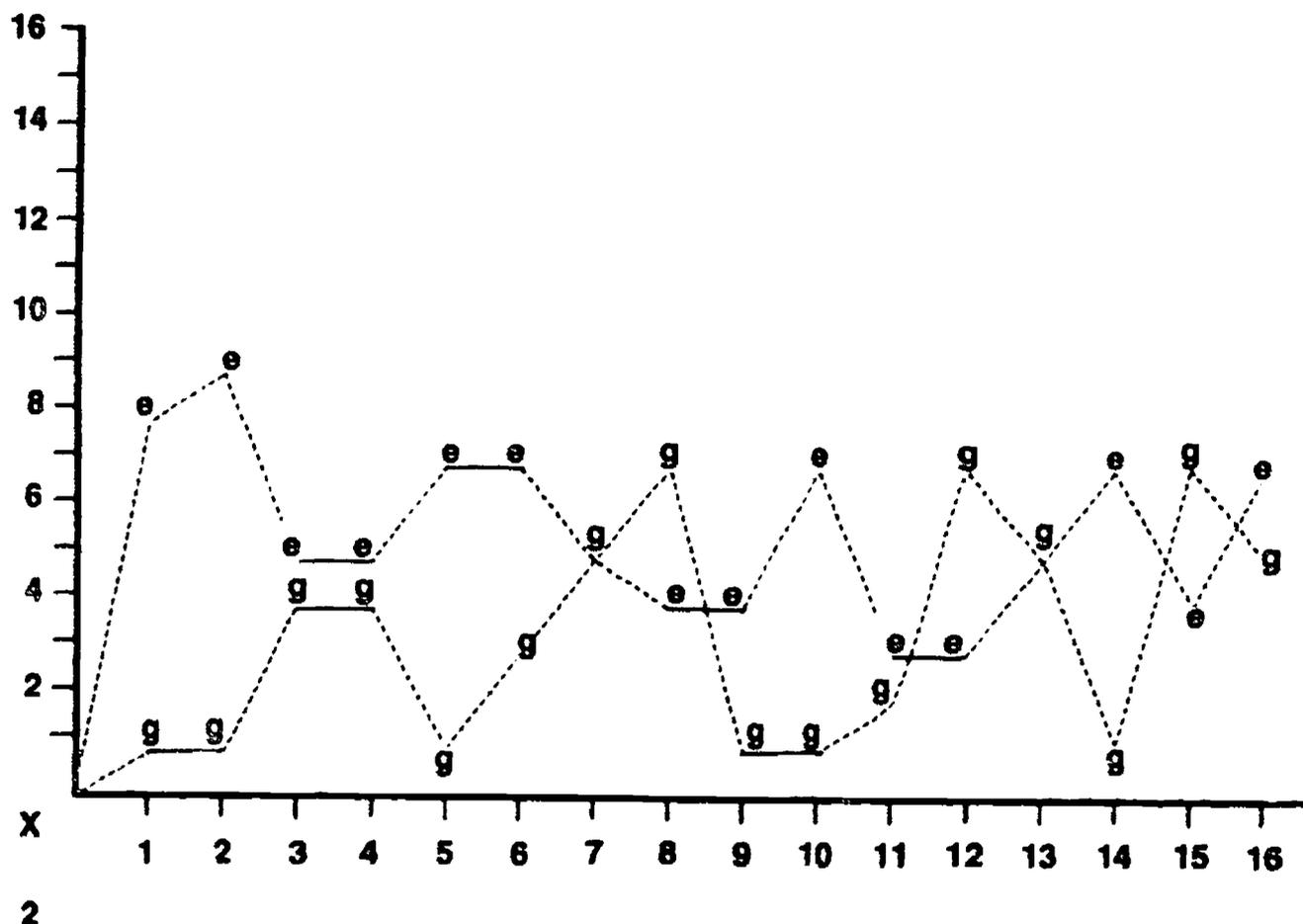
La table des fréquences nous indiquent qu'il y a un verbe au futur, onze au présent de l'indicatif, 189 à l'imparfait de l'indicatif, 117 au passé simple et finalement cinq à l'impératif. Nous avons donc dix-sept verbes qui appartiennent au monde commenté en regard de 306 qui appartiennent au monde raconté. A l'intérieur du monde raconté, nous avons 189 verbes qui appartiennent à l'arrière-plan en regard de 117 qui appartiennent au premier plan.

## 1.4.2 Le graphique

Dans le graphique nous pouvons observer quatre segments où le nombre de verbes au passé simple augmente beaucoup comparativement aux données générales du chapitre. Le premier couvre les tranches trois et quatre. Il y a seize verbes au passé simple et vingt à l'imparfait. Le second

segment couvre les tranches sept et huit. Il y a vingt-quatre verbes au passé simple et dix-huit à l'imparfait. Le troisième segment couvre les tranches douze et treize. Il y a vingt-quatre verbes au passé simple et seize à l'imparfait. Finalement, le quatrième segment couvre les tranches quinze et seize. Il y a vingt-six verbes au passé simple et vingt-quatre à l'imparfait.

### Nombre de verbes



Nombre de mots (tranché = 186)

e = Imparfait  
g = passé simple

#### 1.4.2.1 Le premier segment

Dans ce premier segment, Félicité perd et retrouve son perroquet. Après l'avoir posé sur l'herbe et s'être absentée quelques minutes, le perroquet disparût. Félicité le chercha partout, demandant à tout le monde si on l'avait vu. Soudain, après que Félicité mit fin à ses recherches, le perroquet vint se poser sur son épaule.

Cet événement marqua Félicité pour toujours:

Elle eut du mal à s'en remettre, ou plutôt ne s'en remit jamais.

Par suite d'un refroidissement, il lui vint une angine; peu de temps après, un mal d'oreilles. Trois ans plus tard, elle était sourde; et elle parlait très haut, même à l'église.<sup>45</sup>

<sup>45</sup>Gustave Flaubert, *op. cit.*, p.69.

#### 1.4.2.2 Le second segment

Au second segment, le perroquet de Félicité meurt. Celle-ci, sous les conseils de Mme Aubain, décide de le faire empailler. En allant porter l'oiseau à Honfleur d'où il sera acheminé par bateau au Havre, Félicité se fait asséner un coup de fouet par le conducteur d'une diligence:

Derrière elle, dans un nuage de poussière et emportée par la descente, une malle-poste au grand galop se précipitait comme une trombe. En voyant cette femme qui ne se dérangeait pas, le conducteur se dressa par-dessus la capote, et le postillon criait aussi, pendant que ses quatre chevaux, qu'il ne pouvait retenir, accéléraient leur train; les deux premiers la frôlaient; d'une secousse de ses guides, il les jeta dans le débord, mais furieux releva le bras, et à pleine volée, avec son grand fouet, lui cingla du ventre au chignon un tel coup qu'elle tomba sur le dos.<sup>46</sup>

#### 1.4.2.3 Le troisième segment

Au troisième segment trois événements majeurs se produisent. D'abord, Paul se marie: "un événement considérable surgit: le mariage de Paul"<sup>47</sup>. Ensuite, après le suicide de M. Bourais, qui avait faussé les comptes de Mme Aubain, celle-ci eut une douleur dans la poitrine et mourut:

Au mois de mars 1863, elle fut prise d'une douleur dans la poitrine; sa langue paraissait couverte de fumée, les saignées ne calmèrent pas l'oppression; et le neuvième soir elle expira, ayant juste soixante-douze ans.<sup>48</sup>

Finalement, après que Paul et sa femme vinrent prendre les meubles qu'ils voulaient et vendre les autres, la maison fut mise en vente.

Félicité est maintenant seule. Paul est marié et loin, Virginie, Mme Aubain et le perroquet sont morts. La maison, presque vide, est à vendre. Félicité n'a pas d'endroit où aller.

#### 1.4.2.4 Le quatrième segment

Au dernier segment, Félicité devient très malade. Un médecin lui apprit qu'elle avait une pneumonie. Peu de temps après, elle était sur le point de mourir: "Du mardi au samedi, veille de la Fête-Dieu, elle toussa plus fréquemment. Le soir son visage était grippé, ses lèvres se collaient à ses gencives, des vomissements parurent; et le lendemain, au petit jour, se sentant très bas, elle fit appeler un prêtre"<sup>49</sup>.

<sup>46</sup>*Ibid.*, p. 71-72.

<sup>47</sup>*Ibid.*, p. 76.

<sup>48</sup>Gustave Flaubert, *op. cit.*, p. 76.

<sup>49</sup>*Ibid.*, p. 79.

## 1.5 Chapitre 5

## 1- TABLE DES VALEURS:

	FUT.	CON.	PR.IND.	PR.SUB	IMPA.IND.	IMPA.SUB.	PA.SIM.	IMPE.
tranche 1 =>	0	0	0	0	6	0	1	0
tranche 2 =>	0	0	0	0	2	1	1	0
tranche 3 =>	0	0	0	0	1	0	0	0
tranche 4 =>	0	0	0	0	4	0	0	0
tranche 5 =>	0	0	0	0	2	0	1	0
tranche 6 =>	0	1	0	0	2	0	2	0
tranche 7 =>	0	0	1	0	1	0	5	0
tranche 8 =>	0	0	0	0	3	0	1	0
tranche 9 =>	0	0	0	0	3	0	2	0
tranche 10 =>	0	0	0	0	3	0	0	0
tranche 11 =>	0	0	0	0	2	0	0	0
tranche 12 =>	0	0	0	0	3	0	0	0
tranche 13 =>	0	0	0	0	2	0	1	0
tranche 14 =>	0	0	0	0	2	0	4	0
tranche 15 =>	0	0	0	0	1	0	4	0
tranche 16 =>	0	0	2	0	0	0	2	0

## 3- FRÉQUENCES

futur	=>	0
conditionnel	=>	1
présent	=>	3
présent du subjonctif	=>	0
imparfait	=>	37
imparfait du subjonctif	=>	1
passé simple	=>	24
impératif	=>	0

## 1.5.1 Fréquences

La table des fréquences nous indique qu'il y a trois verbes au présent de l'indicatif, trente-sept à l'imparfait et vingt-quatre au passé simple. Nous avons donc trois verbes qui appartiennent au monde commenté en regard de soixante-et-un qui appartiennent au monde raconté. À l'intérieur du monde raconté, nous avons trente-sept verbes qui appartiennent à l'arrière-plan en regard de vingt-quatre qui appartiennent au premier plan.

## 1.5.2 Le graphique

Le chapitre 5 ne comprend pas suffisamment de mots pour que l'Analyste puisse tracer un graphique. Toutefois nous pouvons juxtaposer les valeurs correspondant aux seize tranches, qui auraient été utilisées pour former le graphique:

Tranche	imparfait	passé simple
1	6	1
2	2	1
3	1	0
4	4	0
5	2	1
6	2	2
7	1	5
8	3	1
9	3	2
10	3	0
11	2	0
12	3	0
13	2	1
14	2	4
15	1	4
16	0	2

En observant les chiffres, nous pouvons discerner deux segments. Le premier couvre les tranches six et sept. Il y a sept verbes au passé simple en regard de trois à l'imparfait. Le second segment couvre les tranches quatorze, quinze et seize. Il y a dix verbes au passé simple en regard de trois à l'imparfait.

#### 1.5.2.1 Le premier segment

Au premier segment, Félicité, sur son lit de mort, s'inquiète une toute dernière fois pour son perroquet. Elle s'informe de l'état de l'oiseau: "Une fusillade ébranla les carreaux. C'étaient les postillons saluant l'ostensoir. Félicité roula ses prunelles, et elle dit, le moins bas qu'elle put: "Est-il bien?" tourmentée du perroquet"<sup>50</sup>.

#### 1.5.2.2 Le second segment

Au second segment, Félicité meurt. La densité plus élevée de verbes au passé simple met en relief cet événement qui est le seul dénouement possible du conte:

Elle avança les narines, en la humant avec une sensualité mystique; puis ferma les paupières. Ses lèvres souriaient. Les mouvements de son cœur se raientirent un à un, plus vagues chaque fois, plus doux, comme une fontaine s'épuise, comme un écho disparaît; et, quand elle exhala son dernier souffle, elle crut voir, dans les cieux entrouverts, un perroquet gigantesque, planant au-dessus de sa tête.<sup>51</sup>

<sup>50</sup>Gustave Flaubert, *op. cit.*, p. 81.

<sup>51</sup>*ibid.*, p. 83.

## 1.6 Conclusion

En portant une attention particulière aux passages qui se caractérisaient par une grande densité de verbes au passé simple nous avons pu extraire de *Un cœur simple* la plupart des événements autour desquels s'articule le conte. D'abord, pour le premier chapitre, qui est très court, il s'agissait d'une introduction. Les personnages principaux y sont décrits. Les lieux y sont dépeints. Ensuite, pour le reste du conte nous avons relevé les événements suivants:

- 1- L'aventure de Félicité et de son amoureux.
- 2- L'attaque du taureau qui est à l'origine de la santé précaire de Virginie.
- 3- La convalescence de Virginie à Trouville.
- 4- Le départ de Paul pour le collège.
- 5- La première communion de Virginie où Félicité fait la connaissance du Christ.
- 6- Le départ de Virginie pour le couvent.
- 7- Le départ de Victor pour l'Amérique.
- 8- La mort de Victor.
- 9- La mort de Virginie.
- 10- La rencontre entre Félicité et le baron de Larsonnière et sa femme. C'est d'eux que Félicité obtiendra son perroquet.
- 11- La disparition et la réapparition du perroquet de Félicité: on y constate l'importance qu'a ce perroquet pour Félicité.
- 12- La mort du perroquet.
- 13- Le mariage de Paul.
- 14- La mort de Mme Aubain.
- 15- La maladie de Félicité.
- 16- L'inquiétude de Félicité pour son perroquet qui la hante jusque sur son lit de mort.
- 17- La mort de Félicité.

Si nous regardons attentivement cette série d'événements nous remarquons qu'ils se rattachent principalement à trois thèmes. Quatre événements consistent en des départs: celui de Paul, de Virginie, de Victor et finalement du perroquet (qui s'enfuit). Trois événements se rattachent à la maladie: l'attaque du taureau qui est à l'origine de la faible santé de Virginie, la convalescence de Virginie à Trouville et finalement la maladie de Félicité. Quatre événements se rattachent à la mort: celle de Victor, de Virginie, du perroquet, de Mme Aubain et finalement celle de Félicité.

## 2. Le tempo narratif

Le tempo narratif est en quelque sorte le débit de l'histoire. Considérons deux exemples. On parle d'un tempo narratif rapide lorsque les événements d'un récit se succèdent à une vitesse relativement grande. Par opposition, on parle d'un tempo narratif lent lorsqu'on a l'impression que rien ne se passe. Avec un tempo narratif très rapide, l'auteur peut couvrir en quelques phrases, voire en quelques mots, une histoire s'échelonnant sur plusieurs années. Avec un tempo narratif très lent, l'auteur peut s'attarder pendant plusieurs pages, voire un roman complet, sur un événement ne couvrant guère plus que quelques heures.

Le jeu du tempo narratif se fait avec les verbes du monde raconté. Une grande densité de verbes à l'imparfait provoquant un tempo narratif plutôt lent alors qu'une grande densité de verbes au passé simple provoque un tempo narratif plus rapide. "Robert Champigny a appelé le Passé Simple "le

temps voltairien par excellence"<sup>52</sup>. L'aperçu est heureux. Voltaire, ses lecteurs le savent bien, conduit ses romans d'un style narratif sec et alerte où l'on ne flâne guère à l'arrière-plan. L'imparfait y est rare, souvent esquivé là-même où on l'attendait".

Dans *Un cœur simple* nous nous attendons donc à un tempo narratif rapide lorsque la courbe d'un de nos graphiques nous indique une présence intense de verbes au passé simple. Aussi, nous attendons-nous à un tempo narratif très lent lorsque la courbe d'un graphique nous indique une quasi-absence de verbes au passé simple. Nous comparerons deux passages du conte. Le premier est représenté par la sixième tranche du graphique schématisant le deuxième chapitre. Il ne contient aucun verbe au passé simple:

Chaque lundi matin, le brocanteur qui logeait sous l'allée étalait par terre ses ferrailles. Puis la ville se remplissait d'un bourdonnement de voix, où se mêlaient des hennissements de chevaux, des bêlements d'agneaux, des grognements de cochons, avec le bruit sec des carrioles dans la rue. Vers midi, au plus fort du marché, on voyait paraître sur le seuil un vieux paysan de haute taille, la casquette en arrière, le nez crochu, et qui était Robelin, le fermier de Geffosses. Peu de temps après, -c'était Liébard, le fermier de Toucques, petit, rouge, obèse, portant une veste grise et des housseaux armés d'éperons.

Tous deux offraient à leur propriétaire des poules ou des fromages. Félicité invariablement déjouait leurs astuces; et ils s'en allaient pleins de considération pour elle.

À des époques indéterminées, Mme Aubain recevait la visite du marquis de Gremanville, un de ses oncles, ruiné par la crapule et qui vivait à Falaise sur le dernier lo-in de ses terres. Il se présentait toujours à l'heure du déjeuner, avec un affreux caniche dont les pattes salissaient tous les meubles. Malgré ses efforts pour paraître gentilhomme jusqu'à soulever son chapeau chaque fois qu'il disait: "Fou mon père", l'habitude l'entraînant, il se versait à boire coup sur coup, et lâchait des gaillardises. Félicité le poussait dehors poliment: "Vous en avez assez assez, monsieur de Gremanville! A une autre fois!" Et elle refermait la porte.<sup>53</sup>

Le deuxième passage correspond à la dixième tranche du graphique schématisant le deuxième chapitre du conte. Il contient nettement plus de verbes au passé simple que de verbes à l'imparfait:

Mme Aubain descendit le fossé, poussa Virginie, Paul ensuite, tomba plusieurs fois en tâchant de gravir le talus, et à force de courage y parvint.

Le taureau avait acculé Félicité contre une claire-voie; sa bave lui rejaillissait à la figure, une seconde de plus il l'éventrait. Elle eut le temps de se couler entre deux barreaux, et la grosse bête, toute surprise, s'arrêta.

Cet événement, pendant bien des années, fut un sujet de conversation à Pont-l'Évêque. Félicité n'en tira aucun orgueil, ne se doutant même pas qu'elle eût rien fait d'héroïque.

Virginie l'occupait exclusivement; -car elle eut, à la suite de son effroi, une affection nerveuse, et M. Poupart, le docteur conseilla les bains de mer de Trouville.

<sup>52</sup>Harald Weinrich, *op. cit.*, p. 117.

<sup>53</sup>Gustave Flaubert, *op. cit.*, p. 34.

Dans ce temps-là, ils n'étaient pas fréquentés. Mme Aubain prit des renseignements, consulta Bourais, fit des préparatifs comme pour un long voyage.

Ses colis partirent la veille, dans la charette de Liébard. Le lendemain, il amena deux chevaux dont l'un avait une selle de femme, munie d'un dossier de velours; et sur la croupe du second un manteau roulé formait une manière de siège. Mme Aubain y monta, derrière lui. Félicité se chargea de Virginie, et Paul enforcha l'âne de M. Lechaptois, prêté sous la condition d'en avoir grand soin.<sup>54</sup>

Les deux textes ne parlent-ils pas d'eux-mêmes? Dans le premier texte, l'histoire est à un point mort. Le temps semble en suspens. "Chaque lundi matin, le brocanteur qui logeait sous l'allée étalait par terre ses ferrailles." Il s'agit d'un événement qu'on ne peut situer sur une échelle temporelle: cela se produit le lundi matin, tous les lundi matin. Cela semble toujours s'être produit et devoir se produire toujours. "A des époques indéterminées, Mme Aubain recevait la visite du marquis de Gremanville, un de ses oncles". On voit ici l'inutilité d'une précision temporelle. Tout ce dont il est question dans ce passage ne relève que de la description. Le narrateur décrit la routine quotidienne, rien de plus. À la fin de ce passage, nous ne sommes pas plus avancés dans le récit. Dans le texte, dix-neuf verbes sont à l'imparfait, aucun au passé simple.

Dans le deuxième passage, le tempo narratif n'est plus le même. Tout se déroule très rapidement. "Mme Aubain descendit le fossé, poussa Virginie, Paul ensuite, tomba plusieurs fois en tâchant de gravir le talus, et à force de courage y parvint". En une phrase, qui contient quatre verbes, le narrateur raconte un événement rempli de gestes de frayeur, de tentatives de fuite et d'échecs qui finalement a un dénouement heureux. Mme Aubain pousse Paul, celui-ci tombe. Mme Aubain le pousse à nouveau, Paul tombe à nouveau. Mme Aubain le pousse encore, Paul retombe... Toute cette série de tentatives est résumée en deux verbes: "Paul ensuite, tomba plusieurs fois...et à force de courage y parvint". Le tempo narratif très rapide a pour effet de mettre le lecteur en haleine.

Ensuite, on apprend que l'événement a été pendant plusieurs années un sujet de conversation. On apprend aussi que Virginie a été très affectée par l'incident, que Mme Aubain projette d'aller à Trouville, que les valises sont prêtes et que tout le monde part. Beaucoup de choses se sont produites en très peu de texte. Le tempo narratif y est très rapide. Ici le narrateur ne s'attarde pas. Il y a, dans le texte, plus de dix-neuf verbes au passé simple et seulement sept à l'imparfait.

### 3. L'attitude de locution

Dans une situation de communication, l'auditeur ne perçoit pas les messages toujours de la même façon. Parfois, l'attitude du locuteur est détendue, quoique le message puisse être revêtu d'une certaine gravité. A d'autres moments, l'écoute peut être d'une tension extrême. En français (et dans d'autres langues bien sûr), le jeu de l'attitude de locution se fait sur deux plans. Celui du monde commenté et celui du monde raconté. Le présent de l'indicatif, le passé composé et le futur simple tendent à créer un effet de tension chez l'auditeur, alors que les verbes du monde raconté créent un effet de détente. "Sont représentatifs des temps du monde commenté... Le locuteur y a une attitude tendue; ses propos s'en trouvent aiguisés, car ce dont il parle le touche de près, et il lui faut également toucher celui à qui il s'adresse"<sup>55</sup>. "Raconter est manifestement un comportement humain caractéristique - on songe à "ce récit fondamental dans lequel baigne notre vie entière", comme l'écrit Michel Butor -; l'une de nos attitudes en face du monde est de le raconter. Nous utilisons alors des

<sup>54</sup>Gustave Flaubert, *op. cit.*, p. 38-39.

<sup>55</sup>Harald Weinrich, *op. cit.*, p. 33.

signaux linguistiques à valeur narrative, en particulier les temps du récit. Ils ont pour fonction d'avertir l'auditeur que cet énoncé est "seulement" un récit et qu'il peut l'écouter avec un certain détachement<sup>56</sup>.

Nous présentons un texte qui contient une séquence de deux verbes à l'imparfait suivie par une séquence de trois verbes au présent de l'indicatif et qui se termine avec un verbe à l'imparfait:

Elle avait peine à imaginer sa personne; car il n'était pas seulement oiseau, mais encore un feu, et d'autres fois un souffle. C'est peut-être sa lumière qui voltige la nuit aux bords des marécages, son haleine qui pousse les nuées, sa voix qui rend les cloches harmonieuses; et elle demeurait dans une adoration, jouissant de la fraîcheur des murs et de la tranquillité de l'église.<sup>57</sup>

Dans la première phrase, nous suivons le récit d'une façon détendue. Ce qui y est dit ne nous touche pas. Félicité a de la difficulté à se représenter la personne du Christ. Soudain: "C'est peut-être sa lumière qui voltige la nuit aux bords des marécages, son haleine qui pousse les nuées, sa voix qui rend les cloches harmonieuses". Nous ne pouvons pas lire ce passage avec la même attitude de détachement. Ce passage se détache du récit. Le narrateur ne raconte pas, mais commente. Ce qui est dit n'a pas de valeur que pour l'histoire, mais également pour la vie réelle de l'interlocuteur. La frontière du récit est brisée. Il s'agit d'une image qui transperce le cadre du conte et qui vient se faire une place dans la tête du lecteur. C'est peut-être la lumière du Christ qui voltige la nuit aux bords des marécages.

Finalement, après ce passage, c'est le retour à l'imparfait: "et elle demeurait dans une adoration". On reprend une attitude détendue. Il s'agit de Félicité, personnage du conte, qui n'a rien à voir avec la réalité.

#### 4. Les transitions temporelles

Une transition temporelle consiste en le passage d'un verbe à un autre. Au début de la lecture d'un texte, on ne peut pas prévoir le type de transition auquel nous aurons affaire. Toutefois, plus nous avançons dans la lecture plus nous pouvons anticiper les transitions à venir. Si par exemple, nous rencontrons cinquante transitions consécutives du type imparfait / imparfait alors nous nous attendrons à l'apparition d'une transition imparfait / imparfait: cela relève d'un pur calcul des probabilités. Toutefois, si nous rencontrons une transition imparfait / passé simple alors que nous ne nous y attendons pas, cela aura valeur de signal. La théorie des transitions temporelles consiste tout simplement en ceci: si une transition temporelle dont la probabilité d'apparition est très faible survient tout de même, alors elle a une valeur de signal d'alerte.

Les transitions temporelles peuvent s'interpréter de trois points de vue: celui de la mise en relief (qui est le passage d'un verbe de l'arrière-plan à un verbe du premier plan ou vice versa), celui de l'attitude de locution (qui est le passage d'un verbe appartenant au monde raconté à un verbe appartenant au monde commenté ou vice versa) et finalement celui de la perspective de locution (qui se joue sur les plans de l'information rétrospective, de l'information anticipée et du degré zéro). Nous n'aborderons ici que les deux premiers. Rappelons que nous n'avons travaillé que sur les formes verbales simples. Pour être en mesure d'aborder la perspective de locution, il aurait fallu que le logiciel identifie les formes verbales composées.

<sup>56</sup> *Ibid.*, p 33-34.

<sup>57</sup> Gustave Flaubert, *op. cit.*, p. 46.

#### 4.1 La mise en relief

Nous présenterons deux textes dont les matrices des transitions temporelles indiquent des comportements différents. Nous choisissons nos textes en consultant les divers graphiques que nous a procurés L'Analyste jusqu'à maintenant. Le premier, correspond à la quatrième tranche du graphique schématisant le troisième chapitre du conte:

(...) et de cette façon comblait le vide des heures.

Le matin, par habitude, Félicité entra dans la chambre de Virginie, et regardait les murailles. Elle s'ennuyait de n'avoir plus à peigner ses cheveux, à lui lacer ses bottines, à la border dans son lit, - et de ne plus voir continuellement sa gentille figure, de ne plus la tenir par la main quand elle sortaient ensemble. Dans son désœuvrement, elle essaya de faire de la dentelle. Des doigts trop lourds cassaient les fils; elle n'entendait à rien, avait perdu le sommeil, suivant son mot, était "minée".

Pour "se dissiper", elle demanda la permission de recevoir son neveu Victor.

Il arrivait le dimanche après la messe, les joues roses, la poitrine nue, et sentant l'odeur de la campagne qu'il avait traversée. Tout de suite, elle dressait son couvert. Ils déjeunèrent l'un en face de l'autre; et, mangeant elle-même le moins possible pour épargner la dépense, elle le bourrait tellement de nourriture qu'il finissait par s'endormir. Au premier coup des vêpres, elle le réveillait, brossait son pantalon, nouait sa cravate, et se rendait à l'église, appuyée sur son bras dans un orgueil maternel.

Ses parents le chargeaient toujours d'en tirer quelque chose, soit un paquet de cassonade, du savon, de l'eau-de-vie, parfois de l'argent. Il apportait ses nippes à raccommoder; et elle acceptait cette besogne, heureuse d'une occasion qui le forçait à revenir.

Au mois d'août, son père l'emmena au cabotage.

C'était l'époque des vacances. L'arrivée des enfants la consola. Mais Paul devenait capricieux (...)<sup>58</sup>

Dans ce texte, il y a vingt-cinq verbes à l'imparfait en regard de quatre au passé simple. Les temps de verbe appartenant au monde commenté sont totalement absents. Voici la matrice des transitions temporelles que L'Analyste nous a préparée:

	nul	futur	condi	pres	p.sub	imparf	i.sub	pas-simp	imper
nul	0	0	0	0	0	1	0	0	0
futur	0	0	0	0	0	0	0	0	0
condi	0	0	0	0	0	0	0	0	0
pres	0	0	0	0	0	0	0	0	0
p.sub	0	0	0	0	0	0	0	0	0
imparf	1	0	0	0	0	20	0	4	0
i.sub	0	0	0	0	0	0	0	0	0
pas-simp	0	0	0	0	0	4	0	0	0
imper	0	0	0	0	0	0	0	0	0

<sup>58</sup> Gustave Flaubert, *op. cit.*, p. 49.

Nous pouvons constater qu'il y a vingt-deux transitions homogènes, soit du type imparfait / imparfait, en regard de huit transitions hétérogènes, qui sont du type imparfait / passé simple (il y en a quatre) ou du type passé simple / imparfait (il y en a quatre également). La transition nul / imparfait devrait être lue comme étant du type imparfait / imparfait car le dernier verbe avant ce passage (ce dont L'Analyste n'a pas tenu compte) est à l'imparfait. Également, la transition imparfait / nul devrait être lue comme étant du type imparfait / imparfait, pour la même raison évoquée plus haut. Il y a au total trente transitions temporelles. Donc, 73% des transitions sont homogènes contre seulement 27% qui sont hétérogènes. 13% des transitions sont du type imparfait / passé simple et 13% du type passé simple / imparfait.

À partir de ces résultats, nous pouvons dire que l'apparition d'une transition imparfait / imparfait est très prévisible. Si une telle transition survient, elle ne provoquera pas d'effet de relief. Toutefois, puisque les transitions imparfait / passé simple sont très rares, s'il y en a tout de même une qui survient, alors elle provoquera un effet de relief. En ce qui concerne les transitions passé simple / imparfait, il ne faut pas établir de relation directe entre leur faible pourcentage d'apparition et leur probabilité d'apparition. Ce qu'il faut voir ici, c'est qu'étant dans un contexte où il y a un très grand nombre de verbes à l'imparfait contre un très petit nombre au passé simple, il est très probable qu'immédiatement après une transition imparfait / passé simple suive une transition passé simple / imparfait. Pour ce texte-ci, 100% des transitions imparfait / passé simple sont suivies d'une transition passé simple / imparfait. Ce qui nous intéressera, en ce qui concerne l'effet de mise en relief, sera donc les transitions imparfait / passé simple.

Ce dont il est question dans le texte, c'est l'ennui qui hante Félicité après le départ de Virginie. Le narrateur y dépeint l'état d'âme de Félicité. Il décrit ensuite les visites de Victor, que Félicité fait venir pour se désennuyer. Il fait mention du départ de Victor, que son père emmène au cabotage. Finalement, les enfants (Paul et Virginie) arrivent pour les vacances. Ces quatre tableaux s'articulent autour des transitions imparfait / passé simple.

"Dans son désœuvrement, elle essaya de faire de la dentelle". On voit ici l'incapacité de Félicité à combattre son ennui. Cette tentative qui échoue montre son impuissance face à une situation douloureuse, mais aussi son désir de s'en sortir. Félicité ne peut se permettre un tel état: elle doit rendre des comptes à sa maîtresse. Elle n'a pas la liberté d'être faible à certains moments.

"Pour "se dissiper", elle demanda la permission de recevoir son neveu Victor". La seconde transition imparfait / passé simple, vient renforcer cette situation de non-liberté à laquelle Félicité doit faire face. Les visites de Victor ne jouent qu'un rôle secondaire dans l'histoire. Le narrateur en fait d'ailleurs mention en utilisant l'imparfait, temps de l'arrière-plan. Ce sur quoi l'attention du lecteur est attirée est la demande que Félicité fait à sa maîtresse: "elle demanda la permission". Félicité n'est pas libre de recevoir qui elle veut. Elle doit d'abord s'enquérir de l'autorisation de Mme Aubain.

"Au mois d'août, son père l'emmena au cabotage." La troisième transition imparfait / passé simple met un accent sur le départ de Victor, que son père emmène au cabotage. La présence de Victor aidait Félicité à surmonter son ennui, mais soudain celui-ci ne peut plus venir la voir. Le départ du neveu est douloureux pour elle.

"L'arrivée des enfants la consola". Avec cette transition le narrateur met l'accent sur l'arrivée des enfants. Cet événement met un terme à l'ennui de Félicité. Tout rentre enfin dans l'ordre.

Le deuxième texte correspond à la dernière tranche du graphique schématisant le cinquième chapitre, soit la toute fin du conte:

Les fabriciens, les chantres, les enfants se rangèrent sur les trois côtés de la cour. Le prêtre gravit lentement les marches, et posa sur la dentelle son grand soleil d'or qui rayonnait. Tous s'agenouillèrent. Il se fit un grand silence. Et les encensoirs, allant à pleine volée, glissaient sur leurs chaînettes.

Une vapeur d'asur monta dans la chambre de Félicité. Elle avança les narines, en la humant avec une sensualité mystique; puis ferma les paupières. Ses lèvres souriaient. Les mouvements de son coeur se ralentirent un à un, plus vagues chaque fois, plus doux, comme une fontaine s'épuise, comme un écho disparaît; et, quant elle exhala son dernier souffle, elle crut voir, dans les cieux entrouverts, un perroquet gigantesque, planant au-dessus de sa tête.<sup>59</sup>

Dans ce texte, il y a onze verbes au passé simple en regard de trois à l'imparfait. Voici la matrice des transitions temporelles:

	nul	futur	condi	pres	p.sub	imparf	i.sub	pas-simp	imper
nul	0	0	0	0	0	0	0	1	0
futur	0	0	0	0	0	0	0	0	1
condi	0	0	0	0	0	0	0	0	1
pres	0	0	0	1	0	0	0	1	1
p.sub	0	0	0	0	0	0	0	0	11
imparf	0	0	0	0	0	0	0	3	1
i.sub	0	0	0	0	0	0	0	0	0
pas-simp	1	0	0	1	0	0	0	6	0
imper	0	0	0	0	0	0	0	0	0

Nous pouvons constater qu'il y a, à l'intérieur du monde raconté, sept transitions homogènes, du type passé simple / passé simple (il y en a six) et du type imparfait / imparfait (il y en a une), et six transitions hétérogènes, du type passé simple / imparfait (il y en a trois) et du type imparfait / passé simple (il y en a trois également). La transition nul / imparfait doit être lue comme étant du type imparfait / imparfait car le dernier verbe avant ce passage (ce dont L'Analyste n'a pas tenu compte) était à l'imparfait. Il y a au total dix-sept transitions temporelles. Donc, 41% des transitions sont homogènes contre 35% qui sont hétérogènes. 18% des transitions sont du type passé simple / imparfait et 18% sont du type imparfait / passé simple.

À partir de ces résultats, nous pouvons dire que l'apparition la plus prévisible d'une transition temporelle est du type passé simple / passé simple. En ce qui concerne les transitions du type imparfait / passé simple, on ne peut pas faire de relation directe entre leur pourcentage d'apparition et leur probabilité d'apparition. Dans un contexte où il y a nettement plus de verbes au passé simple que de verbes à l'imparfait, il est fort probable qu'après l'apparition d'une transition passé simple / imparfait une transition imparfait / passé simple survienne. En fait, dans ce texte, 100% des transitions passé simple / imparfait sont suivies d'une transition imparfait / passé simple. Nous pouvons donc dire que le type de transition dont la probabilité d'apparition est la plus faible est du type passé simple / imparfait. Or, si une telle transition survient alors elle provoquera un effet de mise en relief.

<sup>59</sup>Gustave Flaubert, *op. cit.*, p. 82-83.

Ce dont il est question dans ce texte c'est la mort de Félicité. Elle reçoit l'extrême-onction puis meurt. Il serait maladroit ici de parler des effets de mise en relief sans replacer ce passage dans son contexte. Après la mort de Virginie, Victor et Mme Aubain, Félicité se retrouve seule. La maison de Mme Aubain, le seul endroit où elle peut demeurer, est à vendre. Pour Félicité, sa vie n'a plus de sens. Elle a passé toute sa vie à servir, maintenant elle n'a plus rien à faire. La mort sera pour elle une délivrance. Elle n'a plus rien à faire sur terre. Lorsqu'elle apprit la mort de sa maîtresse elle avait d'ailleurs de la difficulté à comprendre pourquoi le sort ne l'avait pas fait disparaître la première.

Dans le texte, presque tout se déroule au premier plan. Les verbes au passé simple sont très nombreux. L'événement joue un rôle très important dans le conte. A travers cette séquence de verbes au passé simple, trois transitions passé simple / imparfait viennent se glisser: "son grand soleil d'or qui rayonnait", "les encensoirs, allant à pleine volée, glissaient sur leurs chaînettes" et "ses lèvres souriaient". Ces transitions sont à l'origine d'images de splendeur, "rayonnait", de douceur, "glissaient", et de paisibilité, "souriaient". Si on extrait du texte ces trois images, on perd l'aspect de délivrance de la mort de Félicité. Ce sont essentiellement ces trois transitions qui donnent à la mort de Félicité toute sa dimension.

#### 4.2 *L'attitude de locution*

Les transitions temporelles peuvent s'interpréter du point de vue de l'attitude de locution. Il s'agit d'opposer les transitions homogènes, qu'elles soient à l'intérieur du monde commenté ou à l'intérieur du monde raconté, aux transitions hétérogènes, qui sont le passage du monde raconté au monde commenté ou l'inverse.

Nous présentons un texte qui a été tiré de la première tranche du graphique schématisant le chapitre trois. Nous avons choisi cette tranche de texte parce que le graphique nous indique la présence de verbes appartenant au monde commenté. Nous nous assurons donc la présence de transitions hétérogènes.

Quand elle avait fait à la porte une génuflexion, elle s'avancait sous la haute nef entre la double ligne des chaises, ouvrait le banc de Mme Aubain, s'asseyait, et promenait ses yeux autour d'elle.

Les garçons à droite, les filles à gauche, emplissaient les stalls du chœur; le curé se tenait debout près du lutrin; sur un vitrail de l'abside, le Saint-Esprit dominait la Vierge; un autre la montrait à genoux devant l'Enfant Jésus, et, derrière le tabernacle, un groupe en bois représentait Saint Michel terrassant le dragon.

Le prêtre fit d'abord un abrégé de l'Histoire Sainte. Elle croyait voir le paradis, le déluge, la tour de Babel, des villes tout en flammes, des peuples qui mouraient, des idoles renversées; et elle garda de cet éblouissement le respect du Très-Haut et la crainte de sa colère. Puis elle pleura en écoutant la Passion. Pourquoi l'avaient-ils crucifié, lui qui chérissait les enfants, nourrissait les foules, guérissait les aveugles, et avait voulu, par douceur, naître au milieu des pauvres, sur le fumier d'une étable? Les semences, les moissons, les pressoirs, toutes ces choses familières dont parle l'Évangile, se trouvaient dans sa vie; le passage de Dieu les avait sanctifiées; et elle aimait plus tendrement les agneaux par amour de l'Agneau, les colombes par l'esprit du Saint-Esprit.

Elle avait peine à imaginer sa personne; car il n'était pas seulement oiseau, mais encore un feu, et d'autres fois un souffle. C'est peut-être sa lumière qui voltige la nuit aux bords des marécages, son haleine qui pousse les nuées, sa voix qui rend les cloches harmonieuses.<sup>60</sup>

Dans ce passage, il y a six verbes appartenant au monde commenté en regard de vingt-cinq qui appartiennent au monde raconté. Voici la matrice des transitions temporelles qui correspond à ce texte:

	nul	futur	condi	pres	p.sub	imparf	i.sub	pas-sim	imper
nul	0	0	0	0	0	1	0	0	0
futur	0	0	0	0	0	0	0	0	0
condi	0	0	0	0	0	0	0	0	0
pres	1	0	0	3	0	2	0	0	0
p.sub	0	0	0	0	0	0	0	0	0
imparf	0	0	0	3	0	15	0	3	0
i.sub	0	0	0	0	0	0	0	0	0
pas-simp	0	0	0	0	0	3	0	1	0
imper	0	0	0	0	0	0	0	0	0

Dans cette matrice, il faut lire la transition nul / imparfait comme étant une transition du type passé simple / imparfait. Le dernier verbe, avant ce passage, étant au passé simple (L'Analyste n'en ayant pas tenu compte). La transition présent / nul doit être lue comme étant du type présent / imparfait, pour la même raison que celle évoquée plus haut.

Il y a, dans le passage, vingt-six transitions homogènes (soit 81% des transitions temporelles) contre seulement six transitions hétérogènes (soit 19% des transitions temporelles). A partir de ces chiffres, nous pouvons affirmer que lorsqu'une transition hétérogène survient, sa probabilité d'apparition étant très faible, elle met l'auditeur en alerte. Nous ne parlerons pas davantage de ce phénomène: nous en avons suffisamment discuté dans la partie traitant de la mise en relief. Nous allons plutôt nous pencher sur le rapport transitions homogènes / transitions hétérogènes et textualité.

Rappelons ces lignes de Weinrich:

Les transitions syntaxiques homogènes garantissent la consistance d'un texte, sa textualité. Cet aspect est déterminant pour la constitution du texte comme tel; c'est à partir de lui seulement qu'il devient possible de s'interroger sur le sens du texte. A un maximum de transitions homogènes correspond un maximum de textualité... Les transitions hétérogènes ne participent guère à la textualité, ou même pas du tout. Leur rôle n'en est pas moins indispensable, puisqu'elles modifient le niveau d'information de l'auditeur.<sup>61</sup>

Dans un récit, lorsqu'on passe du monde raconté au monde commenté, ou l'inverse, on change de niveau de locution. Le monde raconté contient l'histoire alors que le monde commenté contient des

<sup>60</sup>Gustave Flaubert, *op. cit.*, p. 45.

<sup>61</sup>Harald Weinrich, *op. cit.*, p. 204.

commentaires ou des réflexions. D'une façon générale, on peut dire que le monde raconté, du point de vue de l'histoire, peut se passer totalement du monde commenté. L'inverse est également vrai. Reprenons par exemple une partie de notre texte, mais en enlevant les passages qui relèvent du monde commenté:

Pourquoi l'avaient-ils crucifié, lui qui chérissait les enfants, nourrissait les foules, guérissait les aveugles, et avait voulu, par douceur, naître au milieu des pauvres, sur le fumier d'une étable? Les semailles, les moissons, les pressoirs, toutes ces choses familières (...) se trouvaient dans sa vie; le passage de Dieu les avait sanctifiées; et elle aima plus tendrement les agneaux par amour de l'Agneau, les colombes à cause du Saint-Esprit.

Elle avait peine à imaginer sa personne; car il n'était pas seulement oiseau, mais encore un feu, et d'autres fois un souffle (...).

Rien ne manque à l'histoire. Les passages que nous avons enlevés relèvent d'un niveau méta-historique, qui lui-même peut se passer du monde raconté. Voici, avec un minimum d'arrangement syntaxique, les passages du monde commenté: L'Évangile parle des semailles, des moissons, des pressoirs, de toutes ces choses familières (...) "C'est peut-être sa lumière qui voltige la nuit aux bords des marécages, son haleine qui pousse les nuées, sa voix qui rend les cloches harmonieuses".

Une transition hétérogène du type monde raconté / monde commenté nous fait passer de l'histoire à un niveau méta-historique. Une transition hétérogène du type monde commenté / monde raconté nous fait passer d'un niveau méta-historique à l'histoire. En ce sens, les transitions hétérogènes consistent en une brisure du flot narratif ou commentatif. Avec l'apparition d'une transition monde raconté / monde commenté, il y a interruption du récit. Avec l'apparition d'une transition monde commenté / monde raconté, il y a interruption du commentaire (ou de la réflexion).

Regardons maintenant la matrice des transitions temporelles qui porte sur l'ensemble du conte:

	nul	futur	condi	pres	p.sub	imparf	i.sub	pas-sim	imper
nul	0	0	0	0	0	0	0	1	0
futur	0	0	0	0	0	1	0	0	0
condi	0	1	3	0	0	5	1	9	0
pres	0	0	1	8	0	20	1	10	1
p.sub	0	0	0	0	0	0	0	1	0
imparf	0	0	9	19	0	540	8	196	5
i.sub	0	0	0	3	0	8	1	5	0
pas-simp	1	0	6	10	1	201	5	230	4
imper	0	0	0	1	0	2	1	6	2

Nous constatons qu'il y a 1198 transitions homogènes en regard de trente-sept transitions hétérogènes, soit 90% de transitions homogènes contre 3% de transitions hétérogènes. D'une part, ces résultats confirment l'hypothèse de Weinrich sur la rareté des transitions hétérogènes dans un texte: "Quelle que soit la catégorie syntaxique examinée, les transitions, dans les textes, sont en général plus souvent homogènes qu'hétérogènes"<sup>62</sup>.

<sup>62</sup>Harald Weinrich, *op. cit.*, p.204.

D'autre part, nos chiffres indiquent un niveau de textualité très grand dans *Un coeur simple*. La textualité étant proportionnelle à la rareté des transitions hétérogènes.

## 5. Commentaire

L'utilisation d'un ordinateur pour un traitement quelconque se justifie toujours par la qualité du travail qu'il est en mesure de faire et par le temps qu'il fait économiser. Il serait presque absurde d'utiliser L'Analyste dans le cadre d'une analyse en syntaxe textuelle s'il devait prendre plus de temps qu'un être humain à exécuter ses tâches. Il s'avère donc opportun de donner quelques indications sur le temps qu'a duré le travail de notre logiciel en ce qui regarde l'analyse de *Un coeur simple*:

- 1- La segmentation du texte: quelques minutes.
- 2- La reconnaissance automatique des formes verbales potentielles: trois heures et demi.
- 3- La désambiguïsation automatique des formes verbales: vingt minutes.
- 4- La désambiguïsation interactive des formes verbales ambiguës: environ trente minutes.
- 5- L'analyse statistique de tous les chapitres: une vingtaine de minutes.

Au total, le travail de l'ordinateur a été fait en moins de cinq heures. Cela est très appréciable, mais il faut souligner que l'ordinateur que nous avons utilisé possédait un microprocesseur 8088 de Intel et n'était équipé que de lecteurs de disquettes. De toute évidence, si nous avions disposé d'un PC AT, qui possède un microprocesseur 80286 de Intel, et d'un disque rigide, le travail aurait été fait beaucoup plus rapidement.

## CONCLUSION

L'automatisation de la méthode d'analyse de Weinrich, en ce qui regarde les formes verbales, a été pour nous beaucoup plus qu'une simple réalisation. Ce fut une longue aventure qui nous a fait connaître et aimer des disciplines auxquelles nous ne portions qu'une attention farouche: la linguistique automatique, l'informatique et l'intelligence artificielle. Ce fut une aventure aussi qui nous a amené à réfléchir aux problèmes reliés au traitement automatique des langues naturelles: les ambiguïtés, les analyses syntaxiques, morphologiques, la fabrication de dictionnaires automatisés, la codification et la représentation graphique des données etc.

Afin d'en venir à réaliser L'Analyste, logiciel d'analyse de la récurrence des formes verbales d'après le modèle proposé par Weinrich, nous avons dû faire une longue démarche. D'abord, nous nous sommes familiarisés avec la syntaxe textuelle, la linguistique automatique, l'informatique et l'intelligence artificielle. Nous devons acquérir, dans les trois premières disciplines, une bonne compétence. Sans cela, nous n'aurions pu mener à bien notre projet. En ce qui concerne l'intelligence artificielle, nous voulions être en mesure de situer nos travaux par rapport à elle.

La première étape de nos travaux fut une longue réflexion sur les problèmes reliés à l'homographie. Un certain nombre d'homographes peuvent être désambiguïsés à l'aide d'analyses syntaxiques relativement simples, mais d'autres persistent. Nous avons donc eu à choisir entre la complexification des analyses syntaxiques et la mise au point d'une étape d'analyse interactive permettant à un usager de régler lui-même, en cours de traitement, les ambiguïtés persistantes. Nous avons choisi la deuxième solution, sachant que de toute manière les analyses syntaxiques automatiques, aussi sophistiquées soient-elles, ne sont pas en mesure de régler d'une façon systématique toutes les ambiguïtés.

Nous avons ensuite rédigé et automatisé nos dictionnaires. Nous avons commencé avec les dictionnaires des radicaux et des terminaisons. Le dictionnaire des radicaux de Maeggard et Spang-Hanssen, que nous n'avions pas de toute façon, ne répertoriait que 4200 verbes. Etant donné que notre analyse devait porter sur des textes littéraires, il nous a semblé opportun de répertorier le plus grand nombre de verbes possible. Nous avons donc utilisé la banque de 12000 verbes de Bescherelle. En éliminant les homographes, nous avons obtenu plus de 8639 radicaux. Pour le dictionnaire des terminaisons, nous avons utilisé celui de Maeggard et Spang-Hanssen auquel nous avons dû apporter quelques modifications pour qu'il s'adapte à notre répertoire des radicaux. Nous avons ensuite préparé un dictionnaire des déterminants. Finalement, nous avons identifié toutes les ambiguïtés pouvant être associées à chacun des radicaux et répertorié toutes les combinaisons possibles d'ambiguïtés pouvant être rattachées à une même forme.

La mise au point des dictionnaires nous a amené à réfléchir sur certains aspects des méthodes d'accès à l'information. Cela nous a amené à préparer des algorithmes de tri alpha-numérique et d'indexation.

Une fois notre matériel en place, nous avons pu commencer à travailler sur la mise au point de L'Analyste. Le logiciel est constitué de quatre modules: le premier consiste en la reconnaissance automatique des formes verbales potentielles, qui est essentiellement en une analyse morphologique, le second consiste en la désambiguïsation automatique des formes ambiguës, qui est une analyse syntaxique, le troisième consiste en la désambiguïsation interactive et, finalement, le dernier consiste en une analyse statistique.

La reconnaissance automatique des formes verbales potentielles est un travail qui se fait en trois étapes: découpage d'un mot en radical et terminaison, consultation des dictionnaires afin de déterminer si le radical et la terminaison existent et finalement, lorsque le radical et la terminaison figurent dans les dictionnaires, vérification de la compatibilité des deux formes.

La désambiguïsation automatique des formes verbales est un ensemble de quatre analyses syntaxiques, ou quatre tests. Le premier est un test des particules verbales. Toutes les formes ambiguës doivent le subir. L'application des trois autres tests, substantif, participe passé en -s et préposition "entre" est déterminée par un code d'ambiguïté qui accompagne le radical.

La désambiguïsation interactive nous permet de régler tous les cas d'ambiguïtés qui persistent. Bien que le taux de succès est déjà suffisamment intéressant avant que l'on entreprenne cette étape, la désambiguïsation interactive permet au logiciel d'obtenir un pourcentage d'efficacité de plus de 99.9%. Cette étape a un mode de fonctionnement très simple. Lorsque le logiciel n'a pas suffisamment d'éléments d'information pour déterminer d'une façon non équivoque si un mot est un verbe ou non, alors il le présente à l'utilisateur. Le mot en question est accompagné de son contexte et de l'ensemble de l'information que le logiciel a pu compiler. Il s'agit pour l'utilisateur, d'indiquer au logiciel s'il s'agit d'un verbe ou non. Très souvent, lorsqu'il s'agit d'un verbe, le logiciel est en mesure de déterminer la personne et le temps. S'il ne peut pas le faire, il présente à nouveau le mot à l'utilisateur et lui demande de préciser la personne et le temps.

L'analyse statistique consiste à compiler les formes verbales et à en donner une représentation permettant de les interpréter du point de vue de la syntaxe textuelle. L'Analyste génère d'abord un graphique qui nous permet d'observer le comportement des formes verbales tout le long d'un texte. Il génère ensuite une table des valeurs qui donne la précision que le graphique ne peut pas rendre afin que nous puissions travailler à partir de chiffres précis. L'Analyste produit également une table des fréquences qui nous indique le nombre de verbes appartenant à chaque temps pour le texte ou la partie de texte analysée. Finalement, il présente une matrice des transitions temporelles. Si, à partir des résultats que présentent le logiciel, plus particulièrement à partir des graphiques, un sous-ensemble du texte attire notre attention, nous pouvons, à partir de l'option de focalisation, faire faire une nouvelle analyse sur le sous-ensemble. La focalisation n'a théoriquement pas de limite. Il s'agit que le sous-ensemble comporte au moins deux mots. Nous pourrions, par exemple, analyser un texte de 1000 pages, focaliser un sous-ensemble de 800 pages à partir duquel nous focaliserions un nouveau sous-ensemble de 700 pages et ainsi de suite. La grandeur du sous-ensemble est déterminée par l'utilisateur.

Une fois le logiciel réalisé, il s'agissait maintenant de l'appliquer à un texte et de démontrer son utilité pour une analyse des formes verbales en syntaxe textuelle. Pour cela nous avons fait l'analyse de *Un coeur simple* de Gustave Flaubert.

L'analyse de *Un coeur simple* a été faite en quatre étapes: le relief, le tempo narratif, l'attitude de locution et les transitions temporelles. L'étude du relief, qui consiste en l'extraction du texte des passages importants (du premier plan), a été faite chapitre par chapitre. Les résultats que nous avons obtenu sont surprenants. En mettant côte à côte les passages que L'Analyste nous a permis de détecter, nous avons pu reconstituer les grandes lignes de l'histoire. De plus, de ces passages s'est dégagé le thème général du conte: la séparation.

À partir des graphiques tracés par L'Analyste, nous avons été en mesure de détecter des passages qui se distinguaient par des tempos narratifs différents. Puisque le logiciel nous donnaient toujours un rapport complet sur le nombre de verbes, classés par temps, il était très facile de faire le lien entre le nombre de verbes au passé simple et le tempo narratif.

En ce qui concerne l'attitude de locution, L'Analyste a fait montre d'une certaine fragilité. Lorsqu'il s'agit d'étudier des formes verbales à caractère hautement récurrent, le logiciel s'avérait un instrument très utile. Toutefois, lorsque certaines formes étaient rares et éparpillés tout le long du texte, comme cela était le cas pour les verbes appartenant au monde commenté, les résultats de L'Analyste étaient nettement moins utilitaires. Cela ne signifie pas cependant qu'on ne puisse pas faire une analyse assistée par ordinateur de formes peu récurrentes. Toutes les formes verbales ont été identifiées, il ne s'agirait que d'apporter quelques modifications à L'Analyste pour qu'il puisse donner une représentation plus utile de ces formes. Cela est un travail que nous n'avons pas envisagé dans le cadre de nos travaux, ceux-ci prenant déjà suffisamment d'ampleur. Nous y reviendrons fort probablement dans des travaux ultérieurs.

Finalement, à partir des matrices des transitions temporelles que le logiciel nous transmettaient, nous avons pu étudier les transitions temporelles de deux points de vues: de la mise en relief et de l'attitude de locution. Nous n'avons pas abordé les transitions temporelles du point de vue de la perspective de locution, et cela pour deux raisons. D'abord, dès le départ nous savions que la reconnaissance des formes verbales composées représentaient une grande difficulté. Comme c'était notre première expérience dans le traitement automatique des langues naturelles, nous nous sommes enclins à ne pas vouloir résoudre tous les problèmes d'une même haleine. Il valait mieux, à notre sens, entreprendre moins de choses afin d'obtenir un résultat de meilleure qualité. De plus, en syntaxe textuelle, outre pour les études de perspective, le passé composé, le futur simple et le présent de l'indicatif ont la même valeur. Il en va de même pour l'imparfait et le plus-que-parfait ainsi que pour le passé simple et le passé antérieur, le futur simple et le futur antérieur et finalement pour le conditionnel et le conditionnel passé. Le logiciel, par son mode de fonctionnement, associe les temps composés aux temps simples correspondant. Par exemple, les verbes *chantait* et *avais chanté* sont tous deux catégorisés imparfait de l'indicatif. De cette façon, du point de vue du monde commenté ou raconté, ou encore du point de vue du premier plan ou de l'arrière-plan, cela n'a aucune incidence. Les deux formes font partie de la même catégorie. *Chantais* et *avais chanté* sont deux verbes appartenant à l'arrière-plan. Sans avoir à aborder les temps composés, nous étions en mesure de couvrir une bonne partie de l'analyse des temps en syntaxe textuelle.

Automatiser la méthode d'analyse de Weinrich, en ce qui regarde les formes verbales, a été une expérience des plus excitantes. Elle nous a fait découvrir la puissance et la beauté de l'ordinateur. Elle nous a fait voir les innombrables avenues de l'automatisation dans le traitement des langues naturelles. D'une façon immédiate, beaucoup de travail reste à faire dans l'automatisation de la méthode d'analyse de Weinrich. La reconnaissance automatique des temps composés permettrait d'aborder l'analyse de la perspective de locution. Une refonte de notre module d'analyse statistique permettrait d'approfondir grandement l'analyse de l'attitude de locution.

Au delà de la méthode d'analyse de Weinrich, tout un monde ne demande qu'à être découvert. Il y a une contribution certaine que la littérature peut apporter aux travaux qui portent sur la compréhension automatique des langues naturelles. La plupart des travaux, jusqu'à maintenant, ont porté sur une analyse intra-phrastique. Un texte pourtant est loin de n'être qu'une succession de mots dont la portée sémantique est restreinte aux limites arbitraires de la majuscule et du point. Plusieurs chercheurs se sont aperçus de l'inaptitude à résoudre tous les problèmes reliés à la compréhension automatique des langues naturelles en se restreignant à des analyses intra-phrastiques. Quelques-uns ont fait éclater le cadre rigide de la phrase. Peu à peu, au sein de la discipline qu'on appelle intelligence artificielle, se fait une place pour les littéraires.

Incorporer une analyse automatique de la récurrence des formes verbales à un projet plus vaste de compréhension automatique des langues naturelles, par exemple, serait une entreprise des plus

intéressantes. À tout le moins, elle permettrait d'extraire, avec une relative facilité, les passages les plus importants d'un texte. À toute fin pratique, ce pourrait être, pour un interface de communication, un système d'aiguillage permettant de sélectionner les éléments d'informations à emmagasiner en mémoire pour usage ultérieur. Après tout, est-ce qu'un être humain, si on nous permet ce parallèle, retient tout ce qu'il lit ou entend? Probablement pas. De toute façon, si l'être humain mémorisait tout, du moins le niveau de mémorisation n'est pas le même pour toute l'information qui lui parvient. Après lecture d'un conte ou d'un roman, lorsqu'on raconte l'histoire à quelqu'un, même avec un sens profond du détail, on ne dit pas tout. Curieusement, on s'en tient au *monde raconté*.

## BIBLIOGRAPHIE

- BONNET, Alain, *L'intelligence artificielle*, Paris, InterÉditions, 1984.
- DE KOCK, Josse, *Linguistique automatique et langues romanes*, Éditions Jean-Favard, Paris, 1977.
- DE SAUSSURE, Ferdinand, *Cours de linguistique générale*, Payot, Paris, 1983.
- DREYFUS, Hubert L., *Intelligence artificielle*, Flammarion, New York, 1984.
- FLAUBERT, Gustave, *Trois contes*, Paris, Garnier-Flammarion, 1965.
- MAEGAARD, Bente et SPANG-HANSEN, Ebbe, *La segmentation automatique du français écrit*, France, éditions Jean-Favard, 1978.
- PEAT, F. David, *Artificial Intelligence: How Machines Think*, Baen Enterprises, New York, 1985.
- PROPP, Vladimir, *Morphologie du conte*, Éditions du Seuil, Paris, 1970.
- TOCCO, Gil, *Logiciels de Micro-Informatique*, "Informatique et bureautique", juillet-août 1984.
- WEINRICH, Harald, *Le temps*, Éditions du Seuil, Paris, 1973.
- WINOGRAD, Terry, "Understanding Natural Language", *Cognitive Psychology* n.3, Academic Press, New York, 1972.