

République Algérienne Démocratique et Populaire
Ministère de L'enseignement Supérieur et de la Recherche scientifique



Université d'Oran
Faculté des lettres, des langues et des arts
Ecole doctorale de Français
Pole ouest-Oran

Intitulé :

TRAITEMENT AUTOMATIQUE DU LANGAGE, OU INGENIERIE LINGUISTIQUE

ETUDES ET ILLUSTRATIONS A TRAVERS QUATRE OUTILS ET
APPLICATIONS DU GÉNIE LINGUISTIQUE

Mémoire de Magister

Présenté par :

Titt Karima

Sous la direction de :

M^{me} SARI Fawzia

Membres du jury :

M^{me} BOUTALEB Jamila MC UNIVERSITE D'ORAN PRESIDENTE

M^{me} SARI FAWZIA PROFESSEUR UNIVERSITE D'ORAN ENCADREUR

M. FARI Gamal-El-Hak MC UNIVERSITE D'ORAN EXAMINATEUR

M^{me} HAMIDOU Nabila MC UNIVERSITE D'ORAN EXAMINATRICE

Année 2010 /2011

SOMMAIRE

Remerciements

Avertissement * 1

Sommaire.....4

Introduction.8

CHAPITRE I : Aperçu historique, définition, famille d'outils et niveaux d'analyse.....12

1.1-Aperçu historique12

1.2- Appellations..... 14

1.3- Définitions.....14

1.4 -Liste thématique des outils du traitement automatique du langage13

1.5- Famille des outils linguistiques du traitement automatique de l'information..... 14

1.6-Grammaires et formalismes syntaxiques16

1.7-Niveaux d'analyse17

CHAPITRE II : Correcteur automatique / logiciel ou outil de correction orthographique, grammaticale et stylistique /correcticiels19

2.1-Correcteur automatique et nouveau mode d'écriture..... 19

2.2- Brève typologie historique des correcteurs automatique.....21

2.3-Les outils existants22

2.3.1 -Les correcticiels intégrés dans les traitements de texte22

2.3.2-Les correcticiels indépendants.....23

*Prière de lire l'avertissement.

2.4 Application : le logiciel Word	24
2 .4.1-Description du logiciel.....	24
2 4- Expérimentation : correction automatique.....	25
2.5- Fonctionnement du correcticiel	26
2 .5-Analyse du processus de correction	26
2-6 Typologie des erreurs détectées par les correcteurs :.....	28
2.7 Etapes du processus de correction	
2.8 Un champ d applications particulier des correcteurs : leur utilisation pour l'apprentissage/enseignement du français.....	29

CHAPITRE III : Ingénierie linguistique et interdisciplinarité. Le cas de l'analyse automatique statistique et documentaire des texte.....36

3.1- Interdisciplinarité et interactions :

3.1.1- Le pourquoi de l'interdisciplinarité : (qu' est ce qui fait que l' ingénierie linguistique est interdisciplinaire)

3 .1.2 – Le comment de l'interdisciplinarité (par quelles procédure le traitement automatique est- il interdisciplinaire ?)

3.2-L'analyse automatique des textes

3.2-Interactions, points en commun et importances des interactions.

3 .3 Analyse (linguistique) statistique

3.3.1-Le model du logiciel d'analyse automatique statistique Hyperbase.....

3. 3.2-Les trois logiciels d'analyse automatique statistique et documentaire de la gamme
.d' « Ascetic »

CHAPITRE IV : 4.1- Traduction automatique

Traduction automatique et interdisciplinarité	53
4.2.-Question épistémologique	55
4. 3: logiciel de traduction	56
4 . 3 .2 Composants d'un logiciel de traduction.....	57
4.3.2.1- Les règles linguistiques	58
4.3.2.2- Dictionnaires : (spécifiques).....	58
4.3.2.3- L'interface	59
4.4 Comment fonctionne la traduction automatique ?	60
4.5-Cas pratique : Le traducteur « Reverso »	63
4.6-Limites du traducteur : (a quel point est-il performant ?).....	66

CHAPITRE V : Apprentissage et enseignement assisté par ordinateur (Logiciel d'apprentissage du Français)

5.1-Présentation, exploration et expérimentation du logiciel	71
5.2 - L'informatique, le numérique, le traitement automatique au service de l'acquisition, apprentissage, diffusion et promotion des langues 74:	
5 .2 .1 Les prémices à partir de 1976	74
5 .2.2-L'après guerre froide, les années 80, une nouvelle interaction	75
5.2.3- Les années 90 , Flexibilité et adaptabilité	77
5.2.4- A partir de l'an 2000, l'autonomie	78
5.2.5-L'accessibilité aux ressources	79
5.2.6-Une deuxième lecture du progrès en question	81

5.2.7-Le pour et le contre (véhicule de communication ou de marginalisation ?).....	82
5.2.8-Complémentarité entre l'Homme et la machine	82
6-Conclusion et (perspectives)	85
7-Bibliographie	88
8-Annexes	92
8. 1- Apprentissage des langues	
8.2- Dictionnaire électronique	
8.3- Recherche en langage naturel	
8.4 Gestion électronique de documents et de l'information	
8.5- Reconnaissance et synthèse vocales	
8.6-- Traduction automatique	
8.7- Indications sur l'état de l'art technologique (Associations, agences et relais) .	
Tables des illustrations :	
Figure 1_: Représentation des niveaux d'analyse du langage	17
Figure 2 : Logiciel de traitement de texte simplifié «aide TTX PAPA NOEL ».....	34
Figure 3_ : Une page du dictionnaire (ou index).....	48
Figure 4 : exemple de la fonction « Contexte »	49
Figure 5: exemple d'une représentation en forme d'histogramme	50
Figure 6 :Schéma récapitulatif du processus de traduction.....	62

Figure 7: Les principaux constituants de l'interface du logiciel de traduction automatique Reverso	60
Figure 8 :Environnement informatique d'apprentissage BLCC : exemple d'une plate forme d'apprentissage	79
Tableau 1 : Exemple de correction automatique	25

Introduction

« ...Au commencement était le vers »

Depuis le début de l'histoire de la communication entre les Hommes; le signe et le mot furent au centre de l'univers et notamment celui (l'univers) des savoirs et des connaissances. Avec la mise en place des premiers alphabets ; la fonction ainsi que le pouvoir du mot sont devenus incontestables.

Vinrent ensuite les sciences et disciplines en succession l'une après l'autre ; jusqu'au XIX^{ème} siècle ; qui a vu naître la linguistique en tant que science à part entière. Depuis Saussure cette science n'a cessé d'évoluer, de se diviser en plusieurs spécialités, et dernièrement d'interagir avec d'autres disciplines, afin de donner naissance à de nouveaux champs et domaines d'étude.

L'une des disciplines scientifiques et techniques avec laquelle les sciences du langage ont entretenu d'étroits rapports et continuent est l'informatique. Depuis le début des années quarante, l'idée de fusionner les deux domaines s'est manifestée d'abord aux Etats-Unis, conceptualisée, ensuite opérationnalisée et concrétisée dans d'autres territoires au niveau des plus grands laboratoires, instituts et centres de recherche et notamment en France.

La nécessité des réciproques interrogations et interactions entre les deux disciplines en question; et les quelques points en commun qui existent entre eux ont donné naissance à un nouveau domaine connu sous le nom de linguistique-informatique, traitement automatique du langage ou encore ingénierie linguistique. (cf. présentations des différentes appellations du domaine dans le § 2 du chapitre I) .Toutes ces interactions se sont avérées incontournables pour plusieurs raisons, nous citons entre autres ; d'une part ; le souci de répondre aux préoccupations grandissantes des sociétés en plein expansion, d'autres part comme résultats logique de

l'accélération de la communication de l'information suite à l'essor et au progrès des nouvelles technologies.

Le fait que cette nouvelle discipline se base à 60% sur les sciences du langage ,le développement rapide de cette discipline caractérisé par l'apparition d'un certain nombre de nouveaux outils et de nouvelles applications et pratiques intégrés dans la vie au quotidien , par exemple le traitement de texte, l'utilisation de différents outils et logiciels de traitement automatique du langage ;comme la correction et la traduction automatique, les analyseurs automatiques de texte, ou encore les logiciels d'apprentissage de langue (cf. voir liste des outils et applications dans le § 4 et 5 du chapitre I), l'effervescence des manifestations et activités de recherche , ainsi que la profusion de différents spécialistes investies dans le développement de cette discipline furent toutes des raisons nous ayant incité à essayer de traiter d'une certaine manière ce thème récent et sensible .

Technologies du langage ,traitement automatique du langage, Ingénierie linguistique, ou encore linguistique informatique sont toutes des appellations pour un domaine émergeant, d'actualité et d'avenir- comme l'a si bien noté Jocelyne Pierre dans son rapport à la « Délégation générale à la langue française et aux langues de France » en parlant du numérique et de toutes ces nouvelles technologies- que nous tenterons d'approcher à travers un certain nombre de questions et d'interrogations auxquelles nous essayerons de répondre à savoir ; qu'est ce que l'ingénierie linguistique ou le traitement du langage au juste ? Quelles valeurs ajoutées, nouveaux défis ,enjeux et perspectives ouvre ce domaine ? A quel point les outils existants sont-ils performants ? Quelle est la dimension interdisciplinaire de ce nouveau champ ? Et à quel lendemain on s'attend ?

Afin de parvenir à cet objectif, nous procéderons comme suit :

Une entrée en matière, à travers le premier chapitre dans lequel nous présenterons un bref aperçu historique croisé de la linguistique et de l'informatique, les différentes appellations du domaine , trois de ses définitions pour tenter de cerner notre champ d'étude, une liste thématique des applications, et une présentation des plus importants outils du traitement automatique de l'information organisés en famille d'outils , ainsi que les grammaires et les différents niveaux du traitement du langage effectuées automatiquement .

En deuxième lieu, nous présentons le premier outil d'étude, un correcteur automatique intégré au logiciel de traitement de texte « Word ». Nous commençons par attirer l'attention sur le nouveau mode d'écriture et rapport à l'erreur caractérisé par le recours grandissant aux logiciels de traitement de texte (qui font le travail à la place du rédacteur ...) présentons ensuite un bref historique et typologie historique des correcticiels¹. Après quoi nous donnons une idée sur le paysage actuel en matière de correction automatiques. Nous passerons après cela à un cas pratique ; qui est donc le logiciel « Word », nous le décrirons, expérimenterons en lui soumettant un petit passage à corriger, analyserons les résultats de la correction en nous approchant de son fonctionnement et de son processus de correction. Nous citerons la typologie des erreurs détectées par les logiciels de correction automatique, ainsi que les étapes du processus de correction automatique et un champ d'application particulier de ces outils ; leur utilisation pour l'apprentissage /enseignement du français, à travers quelques exemples et points de vue types, et conclurons .

Dans le troisième chapitre, nous évoquerons un point sensible et pragmatique des approches de notre travail ; il s'agit de s'interroger sur un aspect profondément pluridisciplinaire et interdisciplinaire de ce nouveau domaine ; dans la perspective de savoir qu'est ce qui octroi à l'ingénierie linguistique ce statut de discipline interdisciplinaire, pourquoi est-elle interdisciplinaire et comment? Et quelles sont les résultats de cette interactivité scientifique (interdisciplinarité) ? Une fois avoir répondu à ces interrogations, nous sélectionnons l'un des plus importants résultats des interactions en question, un des domaines émergents du traitement automatique du langage qui est l'analyse automatique des textes, qui comme son nom l'indique est l'aboutissement de la fusion entre informatique et analyse humaine des texte , nous tenterons de mettre en exergue les points en commun entre ces deux domaines et l'importances des interactions en question , ce qui nous mènera par conséquent vers le domaine de l'analyse statistique et documentaire automatique de textes et de corpus.

¹ Les correcteurs automatiques sont également appelés logiciel ou outil de correction automatique (traitement de texte) ou encore correcticiels ; terme obtenu de la contraction des mots correcteur et logiciel .

Pour illustrer concrètement et pratiquement ces interactions, nous présentons le model du logiciel d'analyse automatique statistique et documentaire des textes « Hyperbase »,et brièvement la gamme des logiciels d'analyse automatique de l'intégrateur de technologie « Ascétic.».

Nous effectuerons un passage par la traduction automatique dans le quatrième chapitre, nous évoquerons la question de l'interdisciplinarité de cette application, nous ouvrons une parenthèse épistémologique liant la traduction automatique à l'ingénierie linguistique ,nous la définissons, présentons les composants d'un logiciel de traduction et comment fonctionne-elle(la traduction automatique) ?Nous passons enfin à un cas pratique qui est le traducteur automatique « Reverso »,nous présentons son interface² avant et après le processus de traduction automatique ,l'expérimentons en lui soumettant quelques phrases à traduire , et évaluons les résultats ;Nous finissons par une brève analyse de l'un des principaux problèmes détectés que rencontre la traduction automatique , soixante ans après sa mise en place, le problème de l'ambiguïté ; qui se pose sur différents niveaux du langage .

Le dernier chapitre sera consacré à un outil automatique pédagogique, il s'agit d'un logiciel d'apprentissage de la langue française « Logitron » .Nous expliquerons d'abord que les logiciels d'apprentissage des langues font également partie intégrante des outils du traitement automatique du langage. Et qu'il contribuent à l'instauration de ce qui est connu comme étant l'apprentissage et enseignement des langues assisté par ordinateur qui peut se faire sans être en ligne , comme ils peuvent également constitué les interfaces, plates formes ³ d'apprentissage et portails de formation en ligne .

Nous justifierons notre choix du logiciel, le testons pour vérifier son fonctionnement et ses performance, à quel niveau il peut intervenir pour aider

² (cf. voir définition d'interface dans les chapitres IV et V).

³ Une plateforme de formation est un logiciel installé sur un serveur informatique. C'est autrement dit un site web auquel l'utilisateur peut se connecter et accéder notamment aux différentes formations prévues par le formateur ou le concepteur du logiciel en question .

l'apprenant, (quels objectifs, ressources, cours, et exercices il contient ?) et quelles valeur ajoutée apporte-il ?

On s'interrogera sur les nouvelles voies, enjeux et perspectives pour le traitement, l'apprentissage, l'enseignement, la promotion et la diffusion des langues en appui avec des moyens informatiques et numériques ?

Pour y répondre, nous élargirons notre étude à travers la présentation et la confrontation d'un certain nombre de références, ensuite statistiques et expériences qui ont été effectuées ailleurs dans des pays ayant officiellement entrepris d'intégrer ou qui sont en cours d'intégration de ces nouvelles technologies et techniques informatiques dans leur différents systèmes éducatifs⁴.

Nous présenterons une conclusion à ce travail en ouvrant à la fin la voie aux perspectives que ce domaine promet d'offrir à l'humanité.

⁴ Dans ce contexte, il s'agit des systèmes pédagogique et didactique destinés à l'apprentissage des langues .

CHAPITRE I :

Aperçu historique, définitions, famille d'outils

Grammaires et niveaux d'analyse.

Dans ce premier chapitre introductif, nous présentons un bref aperçu historique de la linguistique ainsi que de l'informatique, ensuite les différentes appellations du domaine - précisons que notre thème d'étude ; l'ingénierie linguistique ou le traitement automatique du langage (TAL) est connue sous plusieurs désignations- il y a donc lieu de les citer ; avant de passer à trois de ses définitions, suivi de la liste thématique des principales applications et d'un certain nombre d'outils du traitement automatique de l'information¹, les plus répandus ; organisés en familles d'outils, ainsi que les grammaires et niveaux d'analyse du langage dans le domaine en question .

1.1 Aperçu historique :

Les dates suivantes représentent et soulignent des faits importants surtout de l'histoire de la linguistique ensuite de l'informatique :

Avant de citer une quelconque date, l'honneur est à cette majestueuse mise en place de l'écriture, sans quoi aucune science n'aurait pu trouver sa voie ni à l'existence, encore moins au développement. Une mise en place qui a été amplifiée, puis préservée par l'imprimerie.

Avant le XVIII^{ème} : les précurseurs de la linguistique sont les auteurs de la grammaire descriptive.

1660 : publication de l'ouvrage *Grammaire générale et raisonnée*, d'Arnaud et Lancelot. Sa perspective était de décrire les règles du langage en termes de principes rationnels universels communs à toutes les langues.

Entre XVIII^{ème} et XIX^{ème} : libre champ à la linguistique historique et comparative. On compare les langues entre elles et on tente de mettre en place les lois d'évolution générales.

1916 : publication posthume du célèbre *Cours de linguistique générale* du linguiste suisse Ferdinand de Saussure (1857- 1913) qui est considéré comme étant le père fondateur de la linguistique en tant que science à part entière .

¹ Le traitement automatique de l'information est l'une des dénominations globalisantes de l'ingénierie linguistique.

.Années 30-40 : La constitution du « cercle de Prague » avec notamment Roman Jakobson (1896-1954) et Nicolas Troubetzkoy qui plaident pour une « linguistique structurale ».

Allan Turing (1912-1954) : l' avènement de ce mathématicien anglais et de sa proposition « machine de Turing » en 1936 sont considérées comme la date de naissance de l'informatique.

1952 : première conférence sur la traduction automatique, organisée au MIT (*Massachusetts Institute of Technology*) par Yehoshua Bar-Hillel (1915-1975) .

André Martinet (1908 -1999) : linguiste français ,apporte sa brique à l'édifice en se faisant connaître à travers la caractérisation des langues par la propriété de la double articulation.

Noam Chomsky (né en 1928) : linguiste et activiste politique américain d'une grande productivité , professeur de linguistique au MIT depuis 1961 .Il a argumenté en faveur de la primauté de la syntaxe sur tous les niveaux d'analyse du langage. En 1957, il publie *Structures syntaxiques* ; ouvrage fondateur en la matière. Dans les années soixante ; l'approche se raffine dans la théorie des « grammaires génératives et transformationnelles » , qui deviendra dans les années soixante-dix la théorie standard de la syntaxe .

Chomsky est aussi connu et notamment des informaticiens via sa hiérarchie issue de la grammaire générative, qui une fois mathématisée puis informatisée via la programmation informatique constitue le fondement de la théorie des langages formels ; branche des mathématiques, logique et informatique dans laquelle sont étudiés les langages artificiels ainsi que leurs propriétés, comme les langages de programmation informatique .

Les années soixante-dix , quatre-vingt sont marquées par l'effervescence de la sémantique formelle qui peut être défini comme étant la sémantique des langages de programmation , pour représenter des connaissances et formaliser des raisonnements : les nouvelles logiques (flou, modale) , les « frames » , les réseaux sémantiques , les graphes conceptuels ; naissent à cette époque . La pragmatique est aussi prise en compte dans ces modélisations. Les systèmes experts apparaissent également à cette époque.

Depuis les années quatre-vingt-dix ; l'augmentation considérable de la capacité de stockage et de calcul des ordinateurs , ainsi que le développement et l'expansion d'internet , permettent l' émergence d'une linguistique de corpus caractérisée par le traitement de données linguistiques en très grandes quantités et fondée par conséquent- pour faciliter le traitement- sur l'exploitation des textes au format numérique ;en faisant usage

des méthodes, calculs, programmes et techniques statistiques et informatiques et donc concrètement le traitement automatique du langage .

1.2 Appellations (Désignations) ² :

Avant de passer à la définition de la discipline en question ; nous devons effectuer un passage par les différentes appellations de celles-ci ; on retiendra :

« Ingénierie linguistique » , « traitement automatique du langage » , « traitement automatique des langues naturelles » , « linguistique informatique » , « informatique linguistique » , « technologies de la langue » , « génie linguistique » , parfois même « industrie de la langue » , « linguistique computationnelle » ou encore « nouvelles technologies d'analyse de l'information » .

1.3 Définition :

Il n'existe selon bon nombre d'auteurs et de chercheurs, que peu de définitions unanimes du domaine de l'ingénierie linguistique .

Nous essayons cependant d'en retenir deux et tentons d'en présenter une troisième à partir de recueil de différentes approches de concepteurs , ainsi que de différentes lectures que nous avons effectuées :

Soit cette définition , donnée par Robert Tabory lors de la première « *International Conference of Computational Linguistics* » ³ en 1961 :

« Une nouvelle science à l'intersection de la linguistique proprement dite, de l'appareil formel offert par les sciences logico-mathématique et certaines contraintes disons de -règles de jeu- suggérées par les machines destinées au traitement automatique de l'information »

(Robert Tabory ,1961)

² Durant ce travail, nous utiliserons à la fois ces différentes appellations pour faire référence au domaine en question.

³ La première conférence internationale de linguistique computationnelle. Du mot « computer » en Anglais .

FRANCIL (réseau francophone de l'ingénierie de la langue) qui est un programme mis en place en juin 1994 par l'Agence universitaire de la francophonie (AUPELF-UREF) pour renforcer ses activités dans le domaine du génie linguistique justement le défini comme suit :

« Le traitement automatique des langues est un secteur ⁴ en plein développement. Il comprend les recherches et développements en matière d'analyse et de génération de texte⁵, de reconnaissance, de compréhension et de synthèse de la parole⁶. Il inclut les applications relatives à la gestion de documents, la communication entre l'humain et la machine, à l'aide à la rédaction, à la traduction assistée par ordinateur. Il comporte des enjeux de type industriel⁷ et économique, de type scientifique et technologique mais présente aussi une dimension culturelle très spécifique. Il est toujours préférable de maîtriser la langue qui sert à exprimer les résultats des recherches, surtout si cette recherche porte sur la langue elle-même. La langue française et la francophonie sont donc très directement concernées par ces enjeux. » (FRANCIL, 1994)

Et nous pouvons ajouter :

L'ingénierie linguistique est l'aboutissement de la rencontre de la linguistique, de l'informatique et de toute forme d'intelligence artificielle. C'est une manière originale, scientifique et technique d'outiller, d'équiper et de préserver la pérennité d'une langue - en partant du principe qu' «...aujourd'hui une langue, qui n'a pas de dictionnaire est une langue morte, demain une langue qui n'aura pas de grammaire numérique sera une langue morte ». (Joceyline Pierre, 2007) - et ce à travers l'application d'un certain nombre de techniques et programmes informatiques à tous les aspects du langage humain.

⁴ Certains spécialistes et concepteurs du domaine le considère parfois d'un angle industriel et commercial.

⁵ Production automatique des textes.

⁶ La reconnaissance, la compréhension et la synthèse de la parole (cf. voir § chapitre suivant)

⁷ D'où l'appellation « Industries de la langue »

Nous obtenons, après application de ces techniques un certain nombre d'applications d'outils dont, nous présentons brièvement la liste thématique suivante (cf. nous reprendrons les exemples de chaque groupe thématique dans l'annexe.).

1.4 Liste thématique des outils du traitement automatique du langage :

Les applications du traitement automatique du langage pourront ⁸ être classés dans les groupes thématiques suivants, ces ensembles sont sans cesse appelés à évoluer en fonction des nouvelles avancées du domaine.

1.4.1- Les groupes thématiques d'applications sont :

1-L'apprentissage des langues.

2-Les dictionnaires électroniques.

3-La gestion électronique de documents et de l'information (gérer les flux de documents de documents qui pénètrent, sortent ou circulent à l'intérieur d'un espace donné. Ces techniques ont pour fonction de capturer des documents, afin de gérer, indexer, stocker, rechercher, consulter, traiter et transmettre des fichiers numériques de toutes origines.)

4-La reconnaissance, compréhension et synthèse vocale : conversion de la voix en fichier numérique et traitement automatique ou semi-automatique de la langue parlée

5-La traduction automatique.

Et nous présentons dans ce qui suit , afin de se placer dans le contexte des nouvelles avancées de la spécialité en question et donner une idée sur le paysage actuel du domaine du traitement automatique du langage et de l'information , organisés en famille d'outils les plus répandus outils ainsi que leurs fonctions , et nous présenterons pour chaque famille d'outils les éléments dont elle est constituée :

1.5 Familles des outils linguistiques du traitement automatique de l'information:

1.5.1-La famille des analyseurs :

Comme leur nom l'indique ; cette famille est constitué de l'ensemble des outils automatiques destinés aux différentes fonctions d'analyse (cf. exemple chapitre III); elle contient :

⁸ Le progrès technologique en perpétuel mouvement

1.5.2-Les analyseurs de documents :

L'ensemble des outils automatiques rendant possible la compréhension et l'analyse globale d'un document.

1.5.3-Les analyseurs morphologiques :

A travers un certain nombre d'opérations analytiques, comme la précision de la catégorie grammaticale, de la racine et l'association aux affixes des informations sémantiques et flexionnelles, se sont l'ensemble des programmes visant la reconnaissance et l'analyse des mots d'un texte d'après leur forme ou leur graphie.

1.5.4- Les analyseurs pragmatiques :

Ensemble de programmes informatiques ayant pour objectif de déterminer le sens d'un texte selon les contextes pragmatiques qui peuvent le caractériser.

1.5.5-Les Analyseurs syntaxiques

Ensemble de programmes informatiques visant à analyser la structure grammaticale d'un texte.

2.5. La famille des correcteurs :

L'ensemble des logiciels remplissant les différentes fonctions de correction (cf. exemple chapitre II), elle se divise en trois parties :

2.5.1-Les correcteurs grammaticaux:

L'ensemble de logiciels informatiques permettant de corriger la structure grammaticale d'un texte. .

2.5.2-Les correcteurs orthographiques / lexicaux :

L'ensemble des logiciels permettant de détecter les erreurs orthographiques d'un texte et d'en proposer la correction.

2.5.3-Les correcteurs stylistiques :

L'ensemble des programmes qui corrigent le style d'un texte. Ils éliminent les répétitions, constatent les mêmes ensembles de mots souvent présents dans le texte, signalent les phrases trop longues, etc.

3.5- La famille des étiqueteurs :

Ensemble des programmes et dispositifs ayant pour fonction d'associer des informations (étiquettes) distinctives à des mots , phrases , et passages d'un texte ou document . (cf. exemple l'étiqueteur « Hybridtagger ») pour le français. Elle est constituée de trois catégories :

3.5.1-Les étiqueteurs morphosyntaxiques :

Outils permettant d'ajouter ou d'associer des informations morphologiques et syntaxiques aux mots d'un corpus, d'un texte ou d'un document écrit ou orale.

3.5.2-Les étiqueteurs prosodiques :

L'ensemble d'outils permettant d'ajouter ou d'associer des informations prosodiques aux mots d'un corpus, d'un texte ou d'un document. Désormais aux corpus oraux.

3.5.3-Les étiqueteurs sémantiques

Outil permettant d'ajouter ou d'associer des informations sémantiques aux mots d'un corpus, d'un texte ou d'un document écrit ou orale..

4.5-La famille des outils de reconnaissance de données à partir de corpus :

Nous trouvons dans cette famille (cf. exemple chapitre V) :

4.5.1- Les outils de reconnaissance automatique de la parole :

Elle contient des outils qui concernent l'identification et l'interprétation du message vocal par une machine.

4.5.2- Les outils de reconnaissance de caractères imprimés :

Procédé qui transforme les textes imprimés en code textuel, la rendant donc lisible par une machine.

Après avoir présenté ce bref aperçu des familles et groupes d'outils qu'offre le domaine du traitement automatique des langues naturelles , la logique implique d'opérer un passage par le fonctionnement de cette discipline .

1.6 Grammaires et formalismes syntaxiques :

Les premiers travaux de recherches, ayant porter sur l'ingénierie linguistique et initiés notamment par l'approche et les travaux de Chomsky à travers les notion de

compétence (phrases grammaticales/ phrases agrammaticales), de performance , sa hiérarchie issue de la grammaire générative , ses grammaires non contextuelles (traduction de « *Context-free grammar* » qui bien entendu ont largement inspirer les informaticiens et notamment durant les années soixante-dix visaient à créer un modèle qui serait capable de théoriser, formaliser puis automatiser le langage et sa compréhension dans leur globalité .

Pour le traitement automatique des langues, une grammaire serait l'ensemble des informations en forme de règles, de paramètres et de modèles programmables.

Voici donc quelques grammaires ayant contribué à mettre en place les assises du traitement automatique du langage.

1.6.1-La grammaire générative :

Vers la fin des années cinquante, les travaux et écrits de Noam Chomsky, en particulier son ouvrage *Structures syntaxique* (1957) ont profondément marqué les débuts du traitement automatique des langues. Il proposait de décrire la langue de façon formelle, en partant du même principe des langages formelles informatiques ; et en commençant par ce qui constitue d'après lui l'élément clef de la langue : la syntaxe. L'objectif étant de « générer » toutes les phrases grammaticales de la langue. C'est à dire toutes les phrases grammaticalement correctes, d'où l'appellation de la grammaire générative.

1.6.2-Les grammaires transformationnelles :

Pour Chomsky *Structures syntaxique* (1957) et *Aspects* (1965) , il était question à l'origine de poser un niveau de structuration dit « sous-jacent » où seraient spécifiées des relations et des fonctions syntaxiques non immédiatement appréhensibles à partir des structures de « surface » et nécessaire pour une interprétation sémantique ultérieure .

Pour Harris *Methods in Structural linguistics* , (1971) , la méthode des transformations vise à décrire une langue comme un sous-ensemble de phrases qui constituent le noyau de la langue à partir duquel des opérations qui sont des transformations, permettent d'établir toutes les autres phrases de la langue .

1.6.3-Grammaire lexicale fonctionnelle :

Dans « *Lexical Functional Grammar* » Lancée par J.Bresnan et R.Kaplan à partir de 1982, les informations et les constructions syntaxiques sont pour la majeure partie des éléments ; directement codée dans le lexique, où elles sont décrites en termes de fonctions (sujet, objet direct, action) et non pas en termes de catégories .

1.6.4-La théorie du Gouvernement et du liage :

Le modèle de Gouvernement et de liage « *Gouvernement and Binding* » est un model également élaboré par Chomsky à partir de 1981.

Dans ce model, le lexique joue un rôle pivot ; autour duquel gravitent tous les autres éléments .Ce qui le distingue du reste des modèles, c'est cette conception de la grammaire en tant que système de principes abstraits plutôt que systèmes de règles de réécriture énumérant les constructions de la langue.

Ces grammaires ont contribué donc au développement de grammaires et sémantiques formelles, qui constituent le fondement des langages de programmation informatiques. Ces langages ont permis par la suite le développement de programmes et de logiciels d'analyse et de traitement des langues .Le principe générale des logiciels et analyseurs automatiques en question décompose l'analyse du texte en quatre étapes successives, comme est le cas dans la méthode traditionnelle , les étapes d'analyse sont donc les suivantes :

1.7-Niveaux (étapes d'analyse) du texte dans le traitement automatique du langage :

Les étapes d'analyse automatique du texte sont :

L'analyse morpho-lexicale : qui se préoccupe de la structure des mots.

L'analyse syntaxique : consacrée à l'analyse de la structure des phrases.

L'analyse sémantique, qui s'intéresse au sens des phrases considérées individuellement.

L'analyse pragmatique : qui s'approche et prends en charge l'analyse de la phrase placée dans son contexte.

1.7.1-Analyse morpho-lexicale :

L'analyse morpho-lexicale prend en charge l'identification des mots du texte (simples, composés, noms propres, abréviations) et leurs traits (genre et nombre). L'analyse morpho-lexicale se décompose en trois étapes :

- 1- La segmentation, dont le but est de découper le texte en phrases puis en mots distincts.
- 2 La lemmatisation, qui permet de déterminer les règles et les formes canoniques qui régissent les mots séparés dans l'étape précédente.
- 3- L'étiquetage, dont l'objectif est d'identifier l'adéquate catégorie morphosyntaxique (verbe, nom...) des mots selon le contexte.

Cette dernière étape est considérablement importante car elle conditionne le processus de l'interprétation du texte. Il se peut qu'elle soit parfois équivoque, car, il y a parfois confusion quant à l'attribution d'une catégorie à un mot. Prenons l'exemple de l'expression « les laissez-passer » ; doit-on l'étiqueter comme un verbe ou comme un nom ?

1.7.2- Analyse syntaxique

Après le stade de l'identification des unités minimales, on s'intéresse à la structure syntaxique des phrases. On s'intéresse à l'analyse des phrases et leur cohérence par rapport à une grammaire qui spécifie les structures correctes et autorisées du langage.

L'un des objectifs de l'analyse syntaxique est d'abord de trouver si la phrase en entrée au système automatique par la saisie est correctement formée ; donc vérifier l'ordre des mots dans la phrase selon les règles de la grammaire .Il sera ensuite question d'associer à chaque mot de la phrase sa catégorie grammaticale parmi celles qui lui ont été assignées par l'analyseur lexicale.

Cette étape est la plus importante dans le processus d'analyse automatique étant donnée que c'est celle qui détermine les structures syntaxiques qui vont permettre ensuite de calculer les divers interprétations sémantiques et pragmatiques .

1.7.3- Analyse sémantique

Après la forme et la structure, on analyse le sens des phrases du texte. Cette étape s'appelle l'analyse sémantique.

Celle-ci traite les phrases de manière isolées pour tenter d'identifier leur sens hors contexte.

Pour définir le sens d'une phrase, une première étape va se préoccuper du sens de chacun des mots constituant la phrase. Ensuite, à l'aide des informations fournies par l'analyse syntaxique, le sens complet de la phrase pourra être déduit grâce à l'application des relations existantes entre les mots selon la grammaire.

1.7.4- Analyse pragmatique

Au stade pragmatique, on intègre le sens littéral de chaque phrase dans son contexte.

En vue d'une compréhension globale d'un texte, le lecteur doit être au courant d'un certain nombre d'éléments qui ne sont pas explicitement mentionnés dans le texte : des connaissances et des données propres au contexte, à la culture générale, au sujet abordé, etc.

C'est l'implicite du langage, les étapes précédente aboutissent parfois à des ambiguïtés qu'il est possible d'explicitier en utilisant l'analyse pragmatique.

Pour pouvoir caractériser ces niveaux ; leurs données et leurs traitements , nous Essayons de proposer le schéma suivant :

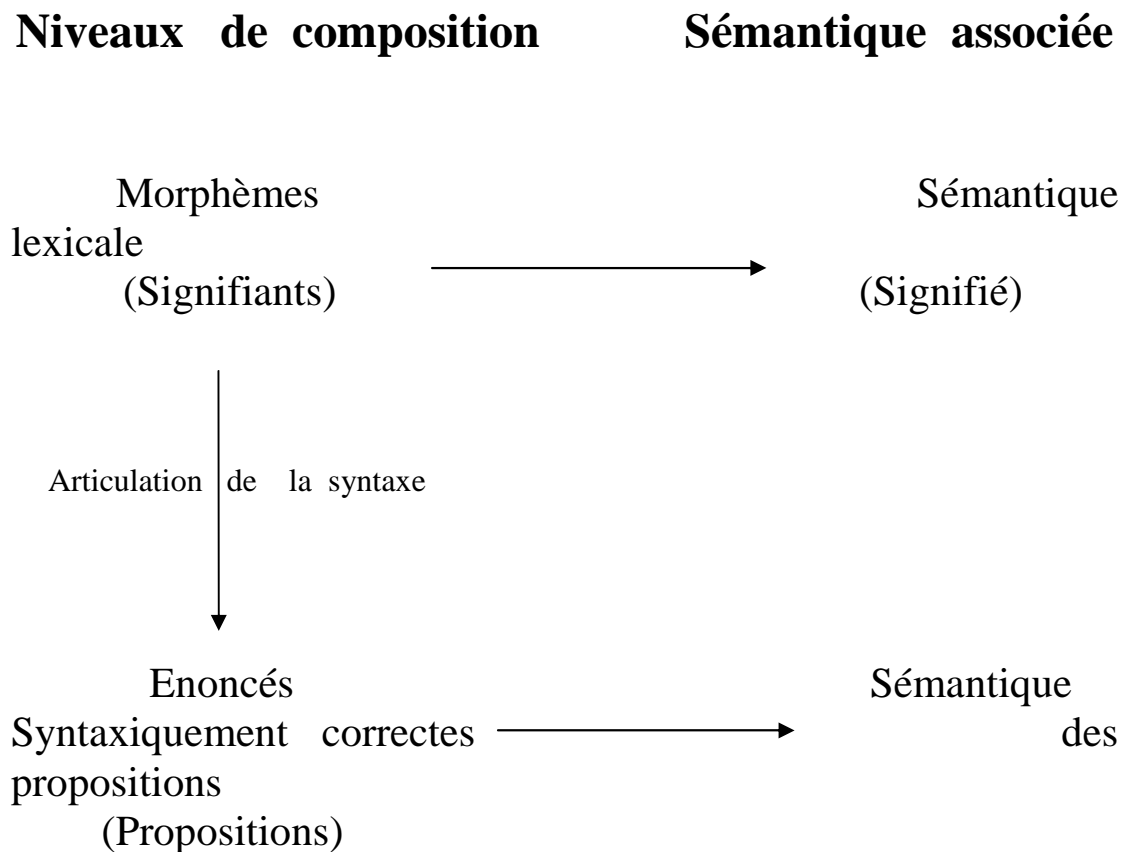


Figure 1 : Représentation des niveaux d'analyse du langage

La double articulation du langage se retrouve dans les deux niveaux de traitements combinatoires successifs qui occupent l'axe syntagmatique. Afin de réaliser et de programmer différents logiciels du traitement automatique du langage, la

première fusion qui s'opère entre linguistique et informatique est celle du recours aux langages de programmation informatiques. Une sémantique des langages de programmation informatique qu'on appelle sémantique formelle, devra être jointe ainsi que la syntaxe informatique à tout les niveaux pour pouvoir réaliser et programmer les logiciels du traitement automatiques du langage. Les langages de programmation les plus utilisés en ingénierie de la langues sont le (PROLOG, PERL, JAVA). Les langages de programmation peuvent également être doublement articulé. La première articulation est celle des règles à suivre impérativement pour écrire un programme syntaxiquement correcte. La deuxième articulation caractérise la construction des unités lexicales de ce programme (soit mots clés du langage, soit identifiants de variables ou noms de fonctions) à partir des unités distinctives élémentaires que sont les caractères alphanumériques autorisés

A travers cet entrée en matière, et en ayant présenté les différentes appellations du domaine, ainsi que la liste thématique des applications et familles des outils linguistiques du traitement automatique de l'information, nous avons montré que le traitement automatique du langage est une discipline interdisciplinaire et que celui-ci peut être exploité par différents domaines et qu'il reprend les mêmes étapes d'analyse humaine du langage, mais d'une manière informatisée. L'un des outils du traitement de texte et d'analyse automatique des erreurs, un correcteur automatique sera présenté et étudié dans le second chapitre.

CHAPITRE II

Correcteur automatique / logiciel ou outil de correction automatique / correcticiel⁹

Dans ce second chapitre ; nous tentons d'approcher notre premier outils d'étude, un des plus répandus et des plus utilisées des outils du TAL¹⁰ un correcteur automatique intégré dans un logiciel de traitement de texte ; le Word .Cette partie du logiciel a été programmée pour prendre en charge entre autres les corrections orthographiques, grammaticales et stylistiques (cf. Voir la définition de l'outil en question dans la liste des familles d'outils linguistiques du traitement automatique de l'information,§5 chapitre I).

Nous commençons par une brève introduction au chapitre présentant le nouveau rapport à l'erreur qu'engendre l'emploi grandissant des correcticiels , ensuite présentons un aperçu et une typologie historique des correcteurs, donnons une idée sur le paysage actuel en matières de correction automatique de texte avant de passer à une concise description du logiciel en question .Nous l'expérimentons en lui soumettant une phrase à corriger, présentons l'analyse des résultats de différents angles ,nous verrons comment fonctionne le correcticiel, et comment procède-t-il pour corriger ,nous présenterons les typologies des erreurs détectées par les correcteurs , et nous nous attaquerons à un cas particulier de l'emploi de ces logiciels, un cas présentant des situations qui devrait remédier au mauvais et à la dépendance à l'emploi des correcticiels ,celui de leur utilisation dans l'enseignement ,pour la réorientation de leur emploi après quoi nous concluons .

2.1 Correcteur automatique et nouveau mode d'écriture :

Les énormes progrès et développements technologiques et informatiques qui ont provoqué une accélération de la communication de l'information ont entraîné le développement de

⁹ Contraction des termes correcteurs et logiciel .Ce sont les trois désignations faisant référence à cet outil. Nous aurons la possibilité de les employer toutes les trois durant ce travail.

¹⁰ Traitement automatique du langage

nouvelles pratiques au quotidien dues au développement de nouveaux outils spécialisés (logiciels de traitement de textes, correcteurs orthographiques et grammaticaux, dictionnaires en ligne, outils d'aide à la traduction, les outils d'aide à la recherche d'information, la synthèse et la reconnaissance de la parole ...) et donc des modifications importantes du mode de vie et d'écriture, où il apparaît que la notion d'erreur se multiplie et trouve une dimension privilégiée, puisque l'une des fonctions les plus utilisées est celle des outils de correction et d'aide à la correction également appelées correcteurs automatiques ou correcticiels (ceux-ci peuvent être des correcteurs orthographiques, grammaticaux ou encore stylistiques à savoir Word, Word Perfect, Quark Xpress...)

Ces outils qui permettent de plus en plus aux utilisateurs d'avoir un travail prêt corrigé et facile sans grands efforts font en sorte que les utilisateurs prennent vraisemblablement goût à cet avantage qui les rend encore moins sensibles à la nécessité de développer leur capacités à rédiger des textes sans erreurs. Nous assistons à une nouvelle forme de « dépendance » : la dépendance aux correcticiels.

Conçus à la base pour des rédacteurs professionnels (bureautiques, secrétariats et administrations), ils ont plus rapidement été utilisés par un plus vaste public et notamment celui des apprenants pour des tâches variées allant de la correction rapide de texte à l'analyse fine de documents complexes et à l'utilisation comme ressources pédagogiques en classe scolaire dans certains pays.

Les étudiants principalement – rares sont les étudiants qui optent pour la préparation et surtout la rédaction de leur travaux d'une manière manuscrite (ne serait-ce que dans un premier temps, sachant qu'aujourd'hui le traitement de texte à tendance à devenir obligatoire pour la présentation des devoirs...) ou autonomes, c'est-à-dire personnelle -et professionnels sont passés de l'écriture manuelle à la frappe à la machine et à la sortie avec traitement de texte. Ils écrivent vite sans véritable soin – du moins dans un premier temps – aussi bien de l'orthographe que de la stylistique en ayant à l'esprit que chaque rectification sera prise en charge et facile à faire grâce à ces outils " magiques" conçus pour, c'est une nouvelle, exagérée et pathologique forme de dépendance à la correction automatique. Le temps de relecture est par conséquent de plus en plus réduit. Cela fait que ce nouveau mode

d'écriture entretient avec l'erreur un nouveau rapport alarmant: le risque d'erreur s'accroît ; et l'objectif premier des logiciels en question qui est d'aider l'utilisateur dans son intérêt et dans la perspective de réduction d'erreurs à l'air de se transformer au contraire ,puisque ces logiciels ne sont pas toujours performant et il n'incitent pour ainsi dire pas –sauf exception- l'apprenant à se perfectionner puisqu'il font le travail à sa place. C'est l'une des motivations majeurs nous ayant incité à commencer nos chapitre par cet outil .

2.2- Brève typologie historique des correcteurs automatique :

Avant de poursuivre ; nous présentons une brève typologie historique des correcteurs afin d'avoir une idée sur leur évolution:

Essayer de trouver des moyens ou des auxiliaires extérieurs aux différents rédacteurs pour les aider à corriger une partie ou complètement leurs erreurs relève d'une préoccupation ancienne qui date d'avant l'informatique à proprement dite :

2.1-La première génération, les mots isolés :

Les premiers programmes destinés à corriger des erreurs orthographiques ont été développés à partir de 1957 dans des mots isolées , nous citons le cas des mots figurants sur des listes , par exemple les noms mentionnés sur les listes de passagers des compagnies aérienne (Galntz,1957). Ils reposent sur le principe de la base de données qui est une liste de mots ordonnée a la quelle est comparé une liste non ordonnée alphabétiquement des mots du texte à corriger .La comparaison se fait entre des chaines de caractères : si le mot du texte n'est pas dans la liste de référence ,le correcteur signale une erreur .Ce type de correcteurs est aujourd'hui le fondement de la plupart des correcteurs intégrés aux logiciels de traitement de texte .

2.2- La deuxième génération, le mot et son contexte immédiat

Les correcteurs dit de deuxième génération représentent un progrès considérable sur les précédents ; dans la mesure où il font intervenir des grammaires locales-basées sur des dictionnaires de mots contenant des informations supplémentaires permettant d'étiqueter la catégorie grammaticale de chaque mot et de définir le genre

et le nombre des noms, pronoms, des adjectifs et des déterminants ainsi que la ou les personnes pour les verbes.- qui analysent le

contexte immédiat du mot et permettent de déterminer par exemple les règles d'accord propres à la langue pour laquelle ils ont été élaborés .Ce type de correcteurs est utilisé dans les caractères grammaticaux intégrés dans les traitements de texte .

2.3-La troisième génération, la phrase complète :

Enfin les correcteurs dits de troisième génération fonctionnent avec une analyse grammaticale complète de phrases complexes. Cette génération a recours à des analyseurs morphosyntaxiques puissants. Le dictionnaire de mots est une base de données de référence comprenant des informations syntaxiques servant à valider la construction des mots. Ils sont maintenant intégrables dans plusieurs logiciels de traitement de texte et d'aide à la rédaction (messagerie) , comme le Word ou le Wordperfect .On s'attend bientôt à une quatrième génération .

2.3 - Les outils existants :

Jusqu'à présent- 2010, les correcticiels peuvent être classés en deux catégories :

2.3.1- Les correcteurs automatiques intégrés dans les traitements de texte :

Comme leur nom l'indique, ils sont intégrés dans d'autres logiciels de traitement de texte. L'efficacité des correcticiels intégrés dans les traitements de texte est notamment limitée par le fait qu'ils ont été initialement développés pour la langue anglaise, dont les structures grammaticales sont différentes de celles de la langue française. Ils déçoivent parfois par leurs insuffisances syntaxiques et sémantiques. Entre autres, les traitements de texte suivants disposent d'un correcticiel intégré (la liste est incomplète : le but n'étant pas ici de recenser tous les traitements de textes...)

Dans l'environnement PC ¹¹ : (de Microsoft)

Word, (de Microsoft),

Word Pro (de Lotus)

Wordperfect (de Corel).

Appleworks (anciennement ClarisWorks) (de Claris)

Dans l'environnement Mac ¹²

ClarisWorks (de Claris)

Word, (de Microsoft),

Wordperfect (de Corel).

2.3.2- Les correcticiels indépendants

Indépendant, c'est-à-dire et à l'encontre des correcticiels intégrés, leur atout est qu'ils sont élaborés par des spécialistes de la langue française et que leurs capacités syntaxiques et sémantiques sont dès lors plus performantes, sans pour autant être complètement performants.

Pour information, il semble selon des statistiques que les produits canadiens sont, à l'heure actuelle, les plus performants. Voici des exemples :

Dans l'environnement PC :

Correcteur 101 (de Machina Sapiens)

Correcteur 101 didactique (de Machina Sapiens)

Cordial (de Synapse)

Antidote (de Druide informatique)

Dans l'environnement Mac

Hugo Plus (de Logidisque)

Prolexis (de Editions Diagonal).

¹¹ L'emploi d'environnement dans un contexte informatique fait référence entre autre et notamment dans le contexte des exemples en question, au système d'exploitation utilisé, dans ce cas Windows de Microsoft corporation

¹² Mac est le système d'exploitation de mac Os de l'entreprise « Apple » computer.

2.4 Application : le logiciel « Word »

2.4.1-Description du logiciel

Dans ce qui suit une description du logiciel a pour objectif de montrer quelles corrections il prend en charge :

Le logiciel Word – sous sa version 2007- est un programme de traitement de texte qui se présente sous forme d'interface graphique¹³ récapitulant un certain nombre de fonctionnalités. La partie qui nous intéresse est celle correspondante à l'onglet " Révision" .En y cliquant surgissent les aides et corrections suivantes :

Grammaire et orthographe, recherche dans le dictionnaire des synonymes et traduction.

La correction grammaticale et orthographique comprend :

- L'accord des noms, adjectifs et déterminant.
- L'accord du sujet et du verbe .
- Forme de l'adjectif attribut et du participe passé.
- Temps et mode des verbes.
- Négation,
- Article définis contractés.
- Harmonie phonétique (ce/cet)
- Comparatif/superlatif
- Emploi des pronoms
- Emploi des auxiliaires .-

¹³ (cf. voir définition chapitre 5)

La correction stylistique comprend

Les expressions archaïques et littéraires .

La longueur de la phrase.

La ponctuation.

2.4.2- Expérimentation : correction automatique

Nous présentons l'exemple suivant : « Les aprenan réviz les lesson », de l'orthographe correcte: « Les apprenants révisent les leçons ».

En saisissant cette phrase, l'analyseur automatique qui est dans ce cas le logiciel Word ; identifie les erreurs en les soulignant en rouge, et propose des alternatives de correction : nous classons ses propositions de correction pour chaque erreur dans un tableau et passerons ensuite à l'analyse :

Erreur	Proposition de correction
aprenan	Apprenant , apprenante , apprenants , arénant
réviz	Reviez , rémiz , revis , revif
lesson	Lessonie, besson, laissons , tessons

Tableau 1 : Exemple de correction automatique

2.5- Fonctionnement du correcticiel :

Pour la première erreur « aprenan », le logiciel de traitement de texte a proposé trois rectifications :

Apprenant ; apprenante, apprenants , arénant

Pour la seconde erreur « réviz » il propose : Reviez , rémiz , revis , revif

Pour la troisième « lesson » il propose : Lessonie , besson, laissons , tessons

« Il est important de noter que l'élément de base que traite un correcticiel est toujours une chaîne de caractères, laquelle est par défaut délimitée par un caractère blanc (un espacement), ou par d'autres caractères conventionnels (traditionnellement un signe de ponctuation). » (Fernand. Berten 1999) . Ici la chaîne de caractère est l'équivalent de mot. L'emploi de chaîne de caractère est emprunté à la terminologie des langages de programmation.

Dans « Word » lorsqu'une chaîne de caractères ne correspond à aucune des formes du lexique interne, une erreur est soulignée en rouge pour les erreurs grammaticales et d'orthographe, en vert pour les erreurs stylistiques. Autrement dit, le correcteur automatique ne traite pas les mots comme des unités lexicales, mais comme des formes ou unités alphanumériques (chaîne de caractères).

Pour procéder et afin d'aider l'utilisateur, le logiciel propose une liste de graphies alternatives susceptibles de remplacer la forme erronée comme dans l'exemple que nous avons cité (aprenan : apprenant, apprenante, apprenants) ,à l'utilisateur de sélectionner la forme correcte ; nous parlons dans ce cas d'une correction assistée par ordinateur . Mais dans nombre de cas, aucune proposition ne peut être faite, c'est par exemple le cas des néologismes ou encore des noms propres ou simplement de toute autre mot qui n'est pas implémenté¹⁴ dans la base de données¹⁵ du logiciel. La deuxième erreur n'a également pas été prise en charge, car elle n'a à son tour pas été programmée , (réviz : réviez ,rémiz , révez , revis revif) alors que le terme voulu et correcte est « révisent », donc il n'y a pas d'identification du mot

¹⁴ Programmé .

¹⁵ Une base de données (son abréviation est BD, en anglais DB, *database*) est une entité dans laquelle il est possible de stocker des données de façon structurée et avec le moins de redondance possible. Ces données doivent pouvoir être utilisées par des programmes, par des utilisateurs différents

mal orthographié ni au niveau lexicale ,(le mot réviser), ni au niveau syntaxe (marque ou terminaison du pluriel)et aucune forme ne peut être à l'heure qu'il est -les logiciels étant en perpétuelle évolution ;ceci explique les différentes versions réalisées d'un programme donnée , dont le principal critère de différenciation est les nouveaux éléments implémentés qui se classe donc par rapport au temps , on obtient par exemple pour le Word les version 98,2003, 2007 et 2010 . Plus la version ou ce qu'on appelle également l'édition est récente plus le nombre de fonctionnalités ainsi que leur performances augmentent – proposée par les données avec lesquelles le logiciel a été programmé.

Cette limite traduit l'incapacité du correcteur automatique d'atteindre pour ainsi, dire la pensée profonde de l'apprenant , pour se contenter de traiter l'aspect extérieur de l'erreur, ce qui est à l'encontre de l'intérêt de l'apprenant du moment où l'objectif de l'opération pédagogique n'est pas de faire en sorte que l'apprenant produise un texte sans erreurs de n'importe quelle façon qui soit , mais de lui apprendre les mécanismes et règles qui lui permettent d'être autonome tout en développant chez lui la compétence , de produire des textes sans erreurs, avec toutes les performances qui accompagne celle-ci.

Le constat est identique dans le troisième exemple (Lesson : besson ,letson, lésons, lessonie , laissons) , le correcteur automatique n'a pas proposé le mot adéquat qui est (leçon) ; d'un point de vue lexicale, on ne peut donc pas s'attaquer à l'analyse des niveaux orthographe ou syntaxe (leçons, « s »la marque du pluriel) ; du moment où il n'a pas reconnu de quel mot il s'agit. Le processus de vérification et de suggestion de corrections d'un logiciel de traitement de texte indépendant ou intégré, dépend complètement des données linguistiques qu'ils lui ont été implémentées et des règles de calculs avec lesquelles il a été programmé. Les performances des outils de traitement automatique des langues sont étroitement donc liées à leurs bases de données. S'il y a un problème à un niveau précis (mots inconnu, néologisme , nom propre ...), s'il donnent une information autre que celle voulu ,insuffisante ou parfois erronées cela peut être expliqué par rapport à deux facteurs principaux : les données à traiter sont inconnues ou ignorées , c'est-à-dire qu'elles n'ont pas été prises en considération par le programme , ou les règles de calcul – en d'autres termes de programmation- sont incorrectes ou incomplètes. Pour traiter ce problème Henri Madec voit que la dimension sémantique doit être prise en considération par les

opérations de programmation. : « Le problème est de créer des correcteurs “sensibles au texte” c’est à dire qui ne considèrent plus le texte comme une suite de caractères mais comme des mots et des phrases produisant des effets sémantiques et pragmatiques » (Henri Madec 2004). Ceci est le point de vue scientifique.

Cependant ces insuffisances ont parfois leur explications et motivations commerciales, car le facteur économique a eu son mot à dire, dans la mesure où l’on programme beaucoup plus de correcteurs automatiques en pensant à leur commercialisation et aux chiffres d’affaires qu’ils réaliseront : « Depuis environ quinze ans, la correction de texte assistée par ordinateur est passée du laboratoire au terrain commercial . » (Frédéric DOLL et Claude COULOMBE .2004) . La correction automatique devient de plus en plus un domaine négligé par les recherches en Traitement Automatique des Langues, elle a quitté les laboratoires de recherche universitaire pour devenir un véritable champ industriel et commercial, d’où l’appellation industrie de langues .

Donc, quelles sont les erreurs qu’un correcticiel peut à l’heure qu’il est ; corriger ?

2.6- Erreurs repérées (prise en charge) par les correcteurs automatiques :

(cf. voir § 2 de ce chapitre)

Dans un logiciel de traitement de texte ; est considéré comme « erreur » uniquement ce qui n’est pas reconnu par le système et donc :

Les formes graphiques ne correspondant pas à une forme mémorisée ou stockée dans le lexique du correcticiel même si elle est juste ,donc les mots justes reconnues comme erronées , mots étrangers, néologismes...etc.

Les mots mal orthographiés reconnus comme erronés .Les erreurs d’orthographe relève soit du niveau lexical, soit du niveau flexionnel. Les correcteurs orthographiques ne font ; en l’état actuel ; aucune hypothèse catégorielle et encore moins sémantique, sur un mot non stocké dans leur lexique.

Le niveau lexical concerne toutes les erreurs portant sur l’écriture du mot hors contexte et la vérification des règles d’accord et de conjugaison intervient au niveau de l’analyse syntaxique et rend nécessaire la mise en œuvre du processus de correction grammaticale.

2. 7-Etapes du processus de correction :

Le fonctionnement des correcteurs automatiques suit un processus qui se décompose en six étapes :

- 1-La vérification lexicale.
- 2-La proposition et le choix du mot à retenir
- 3-La vérification grammaticale.
- 4-La proposition et le choix de l'alternative.
- 5-La vérification stylistique
- 6-La proposition et le choix de l'alternative

Ces étapes du processus de correction automatiques correspondent aux trois types de correcticiels qui existent à présent, à savoir :

- 1-Les correcteurs lexicaux : Traitement lexicale
- 2-Les correcteurs syntaxiques : Traitement syntaxique
- 3-Les correcteurs stylistiques: Traitement stylistique

2.8- Un champ et une solution d'application particulière des correcteurs automatique : leur utilisation dans l'enseignement des langues :

Le fait d'avoir attiré l'attention sur ce champ d'application particulier de l'emploi des correcticiels, vient d'une part, en accompagnement et comme résultat logique du développement technologique auquel nous assistons.

D'autre part en réponse et dans une tentative de remédier à l'état alarmant du recours croissant à la Toile¹⁶ et en l'occurrence aux logiciels de traitement de texte et notamment ceux intégrant des correcteurs automatiques pour les plus simples travaux de rédaction(fait pouvant être constaté à tous les niveaux scolaires)- tentant vers une dépendance anti-pédagogique à ces outils . Nous estimons dans la poursuite des

¹⁶ Internet .

estimations faites par les spécialistes, que pédagogie et didactique doivent conjuguer les efforts afin de concilier entre technologie et développement d'une part et bonne évolution de l'apprentissage et notamment des langues d'autre part.

Linguistique, traitement automatique du langage et apprentissage des langues entretiennent désormais d'étroits rapports. Les sciences du langage et la didactique avaient bien avant partagé d'intimes liens ; notamment pour ce qui relève des théories psycholinguistiques, neurolinguistiques ainsi que neuropsycholinguistiques et leurs rôles dans l'« acquisition », l'apprentissage, et par la suite l'enseignement des langues.

De nos jours, et comme par héritage des précédentes générations de disciplines ; l'ingénierie linguistique ; avec ses différents domaines ; est à son tour sollicitée et mise au service de l'apprentissage des langues. Le perpétuel et réciproque mouvement interrogatoire entre ces deux domaines est présent. Le traitement automatique des langues interroge la didactique au niveau de certaines , théories, méthodes et règles afin de produire des logiciels conformes et qui sauraient répondre aux préoccupations et besoins des apprenants .Et la didactique fusionne avec l'ingénierie linguistique et l'informatique par rapport à certaines techniques , et notamment pour ce qui est de la mise en place de logiciels d'apprentissage .Il devient de plus en plus difficile de nos jours de contester l'élargissement du spectre de l'utilisation des différentes applications et outils d'ingénierie linguistique par le public des apprenants et étudiants et ce pour différentes tâches , nous citons à titre d'exemple, les dictionnaires électroniques , l'apprentissage via les didacticiels et plates formes ,¹⁷les logiciels de traitement de texte(correcticiel d'orthographe, de grammaire , les conjugueurs) ...etc.

Cette utilisation devient de plus en plus importante, c'est pour cela que l'apprentissage et enseignement des langues doit par anticipation mettre en place des plans et des projets qui serait bénéficié de ces outils avant qu'ils deviennent de vraies ennemies et entraves à la bonne évolution du processus pédagogique.

Comme l'a noté Bibeau dans un bon nombre de ses travaux et notamment dans « Il apprennent à lire à l'aide de l'ordinateur » (1998) dans le cadre de son

¹⁷ (cf. voir chapitre IV) .

observation dans le domaine des technologies de l'information et de la communication éducative depuis 1983, il constate que c'est l'évolution même de la didactique qui a entraîné une profonde mutation du pédagogique et devrait encore beaucoup l'enrichir : l'enseignement met désormais au cœur de ses préoccupations la compétence et le paradigme constructiviste, dont l'axe directeur est apprendre à apprendre. L'élève doit apprendre à corriger, comme il doit apprendre à rédiger. -Or la réalité est tout autre, car une majorité d'élèves devient quasi-dépendante à la Toile et à l'emploi des correcticiels, et n'arrivent ni à se corriger, encore moins à rédiger.

C'est pour cela et dans la perspective d'un apprentissage permettant à l'apprenant de rédiger convenablement et de manière autonome un texte, qu'il conviendrait de placer le recours aux logiciels d'aide à la rédaction¹⁸ – nous réduisons notre approche analytique dans ce chapitre selon l'outil étudié aux logiciels de traitement de texte- c'est-à-dire aux correcteurs-.

Cependant la réalité sur le terrain est différente, c'est pourquoi concrètement et pédagogiquement parlant, la question se pose alors de savoir si ce type d'aide à la rédaction est réellement au service ou au contraire aux dépens de l'apprentissage de la compétence orthographique et grammaticale des élèves.

Il serait donc idéal mais surtout pertinent, par exemple en Algérie, où le nombre des utilisateurs de ces outils ne cessent également de croître comme partout ailleurs de s'inspirer des expériences ou du moins des hypothèses de recherche, et théories- prouvées- des prédécesseurs afin de mettre en place un plan et des méthodes intelligemment conçues dans la perspective de mettre ces outils au service de l'apprentissage des langues et des apprenants et leurs apprendre à en tirer profit dans le bon sens.

Dans un certain nombre de pays, où l'utilisation des correcticiels a été intégrée ou est en cours d'intégration des méthodes d'apprentissage des langues; linguistes, didacticiens et Talistes se mettent d'accord dans la perspective de poser la question de l'utilisation des correcteurs en milieu scolaire essentiellement en fonction de deux paramètres principaux :

¹⁸ Les logiciels contenant des outils permettant d'aider l'utilisateur dans ses différentes tâches de rédaction. Il s'agit entre autres des logiciels de traitement de texte.

1-Utiliser le correcticiel comme une aide à la correction

Doit-on et peut-on permettre aux apprenants, de s'aider de ces outils pour apprendre à produire un texte sans erreurs (encore faudrait-il être capable de les empêcher ou de les contrôler en permanence) ? A ce moment on sera amenés à s'interroger sur le comment de cette démarche et donc quelle procédure adopter ? On se préoccupera de savoir si les apprenants ne risquent-ils pas d'apprendre à trop compter sur ces outils, d'en devenir dépendants et de perdre par conséquent leurs jugements et capacités critiques ?

L'utilisation de ces outils ne risque-elle pas même d'induire des erreurs dans les cas où ils détectent des erreurs là où il n'y en a pas ? Et donc finir par fausser l'évolution convenable du processus d'apprentissage ? Cela nous mène vers :

2-Utiliser le correcticiel comme une aide à la réflexion, comme un outil, un soutien dans le cadre de l'apprentissage d'une langue.

Nous élargissons les exemples que nous présentons ci-dessous vers d'autres correcteurs automatiques, pour illustrer ce rôle. Avec l'intervention de l'enseignant, et en exploitant l'aspect interactif, et relativement motivant (via les supports informatiques), on pourra :

1-Les utiliser pour développer les compétences lexicales :

Dans une séance de cours, et avec l'intervention de l'enseignant, le correcteur automatique pourra alors être utilisé pour faciliter une séance de travail sur les mots, le logiciel « antidote » qui est l'un des plus grands logiciels de correction grammaticale et d'aide à la rédaction en français, propose par exemple un million de termes classés par sens sur 54000 entrées. Il englobe des synonymes, des homographes, des définitions, la conjugaison... Il restera à préciser de quelle façon l'utiliser (un plan pédagogique d'utilisation) et en fonction de quels objectifs pédagogiques pour en tirer le meilleur profit.(Certains paramètres rentrent en jeu , la tranche d'âge des utilisateurs, le contexte dans lequel et pour lequel il est exploité ,les pré-requis et connaissances préalables...), la recherche de synonyme pour remplacer, par exemple un mot utilisé plusieurs fois , ou trop peu expressif , cela contribuera à l'enrichissement du vocabulaire, en gardant toujours l'aspect interactif .

2-En ayant recours au « feed-back » en orthographe :

Les théories de l'apprentissage reconnaissent le rôle important de la rétroaction dans la formation .C'est l'aspect diagnostique de l'évaluation qui est important dans l'acquisition de l'orthographe: un apprenant va progresser s'il se rend compte qu'il fait toujours les mêmes erreurs (exemple: bontée, santé, qualité,), mais à la condition qu'on lui permette de s'en rendre compte et de lui donner les moyens de reconstruire correctement la règle erronée qu'il applique. C'est à ce moment que l'enseignant interviendra, également dans le cas des correcticiels que se soit « Word », «Correcteur 101 didactique » ou autre pour expliquer à l'apprenant que si le logiciel de correction automatique lui souligne par exemple les mêmes erreurs, c'est non seulement parce qu'il est programmé en fonction, mais surtout car l'apprenant répète lui aussi la même erreur et n'a pas pu connaître les raisons qui l'ont menées à faire cette erreur , on devrait aussi lui indiquer comment il doit la corriger .

3- Les employer pour faire de la grammaire

Le développement et la réflexions sur la grammaire pourront être conduits de plusieurs points de vue en analysant les suggestions faites par les différents logiciels (toujours en exploitant l'aspect interactif et avec l'intervention de l'enseignant) . Dans le cas de Word, les propositions de corrections grammaticale (grand / gri)z , forme correct (grand / grise) le féminin. (idéau), forme correcte(idéaux) le pluriel des noms ou encore. (ils discut, révis / discutent, révisent) la terminaison du pluriel ,en proposant plusieurs alternatives , il faudra choisir la bonne forme ,et il y aura naturellement lieu d'expliquer à l'apprenant pourquoi il faut écrire ces mots comme ça,(marque du féminin, du pluriel , les catégories de mots auxquelles il faut ajouté x...etc.) ou d'en faire le rappel si les cours ont été déjà faits. Lui demander par exemple de les réécrire. Pour qu'il y ai ancrage. Ceci constituera un moyen interactif et relativement motivant de développer chez l'apprenant certaines compétences et une première connaissance des techniques critiques.

Si les interfaces des logiciels citées plus haut seraient compliqués pour les plus jeunes , il y aura lieu de recourir à des logiciels le plus possible adaptées à cette tranche d âge. Tout en adoptant des méthodes pédagogique qui pourraient accompagner ce recours. Nous citerons à titre d'exemple les logiciels pour les jeunes apprenants : « WordPad » « « je lis puis j'écris » « aide traitement de texte PAPA

NOEL » etc. Ils ont justement une interface graphique¹⁹ et un contenu adapté aux plus jeunes (cf. retour aux logiciel en question) .Nous présentons l’interface du logiciel « aide traitement de texte PAPA NOEL » pour les plus jeunes :



Figure 2 : logiciel de traitement de texte simplifié «aide TTX PAPA NOEL »

D’un point de vue général, si ces d’outils ouvrent de nouvelles perspectives et procurent aux élèves une certaine autonomie, ou une « aide » , ils ne devraient pas être une fin en soi , les élèves sont invités ,plus encor il doivent comprendre, et agir en considérant ces outils comme une aide pour se perfectionner. Cette tache relève du ressort d’un spécialiste, notamment enseignant, ou tout autre connaisseur qui permettra à l’apprenant de comprendre le fonctionnement de l’outil sur les plans notamment linguistiques. Cette démarche d’accompagnement, interactive entre l’enseignant et l’apprenant est nécessaire pour qu’il y ait véritablement apprentissage (compréhension, intégration, transfert- sur d’autres cas) et l'autonomie d'un élève ou de toute autre apprenant en matière d'orthographe est obtenue, seulement s'il a la possibilité d'une part de comprendre son fonctionnement, mais surtout d'évaluer correctement la validité des propositions que fait un logiciel. Dans ce sens Berten,

¹⁹ (cf. voir chapitre IV)

affirme que « Si la machine fait de l'orthographe à sa place, l'élève n'acquerra jamais la capacité de détecter ses propres erreurs » (Berten .1999)

En conclusion à ce chapitre, nous constatons qu'il y a un changement dans le mode d'écriture caractérisé par le recours aux correcticiels .Cependant ces outils ne présentent pas toujours des performances, n'empêche que l'évolution en matière d'intelligence artificielle est telle qu'il faut s'attendre à d'énormes progrès en matière de correcteurs automatiques, dans les toutes prochaines années.

L'apport des correcticiels actuels dans la pédagogie de la langue française est minime, et leur utilisation doit être bien réfléchi, sous peine de conduire à une aggravation des erreurs, surtout dans le cas des élèves qui connaissent de réelles difficultés en orthographe. Pour cette raison, il est essentiel que l'élève ou tout autre apprenant aient pu en mesurer quelques limites, et en comprendre le mécanisme du fonctionnement. Il ne demeure pas moins que ces derniers pourraient s'avérer très utiles dans la mesure ou comme nous l'avons cité il sera nécessaire avant tout recours à ces logiciels d'adopter une méthodologie d'emploi. Cette dernière pourrait être mise en place par les enseignants ou toute autre spécialiste.

CHAPITRE III

Ingénierie linguistique et interdisciplinarité

Le cas de l'analyse automatique des textes.

Ce chapitre est à double vocation :

Dans un premier temps, et avant de poursuivre notre étude de deux autres applications du traitement automatique du langage ; (quatrième et cinquième chapitre); nous ferons en sorte d'ouvrir une parenthèse sur un point sensible et pragmatique des approches de notre travail ; il s'agit de s'interroger sur un aspect profondément pluridisciplinaire et interdisciplinaire²⁰ de cette nouvelle discipline ; dans la perspective de savoir qu'est ce qui octroi à l'ingénierie linguistique ce statut de discipline interdisciplinaire - pourquoi est-elle interdisciplinaire et par rapport à quoi-? Comment est-elle interdisciplinaire ? Et quelle sont les résultats de cette interactivité scientifique (interdisciplinarité) ?

Il est important dans ce sens ; et pour montrer quelques dimensions importantes de ce caractère scientifique interdisciplinaire ;une fois avoir présenter la réponse aux questions ci-dessus ; de souligner un des plus conséquents résultats de cette interactivité qui existe entre linguistique et informatique. C'est ce qui nous mènera ,en second lieu -la conséquence ou le résultat étant le fait d'aboutir logiquement à de nouvelles spécialités et applications- à en sélectionner une pouvant être mise au service de différents domaines à savoir la linguistique et la littérature , il s'agit de l'analyse automatique des texte qui est donc le résultat ou l'aboutissement de la fusion et du dialogue entre l'analyse des textes et les méthodes informatique. Nous en ferons la présentation ; et tenterons de montrer en partie comment s'effectue toutes ces interactions, avant d'illustrer par deux cas pratiques, et concluons.

²⁰ Le terme interdisciplinaire peut être définie comme étant un adjectif relatif aux rapports entre plusieurs disciplines, il est le second palier d'une échelle scientifique de signification classant les termes : pluridisciplinaire, interdisciplinaire et transdisciplinaire .

3.1- Interdisciplinarité et interactions :

3.1.1- Le pourquoi de l'interdisciplinarité : (qu' est ce qui fait que l' ingénierie linguistique est interdisciplinaire ?) :

L'un des constats pouvant être détecté par tout observateur et qui est considérablement intéressant de mentionner aussi historique qu'il puisse être est que l'aube des savoirs et des connaissances – toute disciplines confondues - s'est effectué avec la mise en place des premières alphabets et par la suite « écritures » en Egypte et en Mésopotamie . Qui dit alphabet, dit mot, et qui dit mot, dit langue, et qui dit langue dit ensuite communication et surtout écriture, et qui dit production écrite dit savoir et connaissances.

Autrement exprimé, le mot, les différentes constructions linguistiques de la plus petite à la plus grande, et la langue sont au cœur des savoirs. Ils sont le pivot et l'essence même de la science et des connaissances. Sans mot ni langage ; il n'y a point de savoir. C'est en quelque manière ce qui rappelle la dialectique des « langues de spécialités », qui permet de voir que la linguistique est finalement sollicitée et au retour peut être mise au service de toutes les disciplines. Puisque le langage et les langues sont le véhicule privilégié de l'information, toutes natures confondues :

« Dans la masse d'informations à laquelle l'homme est confronté, la forme la plus courante est bien entendu le langage, qu'il soit écrit ou parlé. Le langage possède une faculté d'adaptation unique, permettant de tout représenter, jusqu'au langage lui-même. » (Jerome Cardot , 2002)

Leur étude devient donc par analogie implicitement ou explicitement, un anneau commun à tous les savoirs et à toutes les disciplines ; d'où cette interdisciplinarité.

L'ingénierie linguistique ou la linguistique informatique puisqu'elle est considérée comme étant l'informatisation de certaines spécialités des sciences du langage, prend le relais de cet état de fait. Les différents outils et domaines du traitement automatique du langage- puisqu'ils reposent sur l'informatique qui à l'instar des mathématiques, sont des disciplines auxquelles toutes les autres ont

recours- sont à leur tour interrogés par la quasi-totalité des disciplines et le nombre d'utilisateurs ayant recours à ces derniers est de plus en plus croissants.

En fait, si l'on considère la question d'un angle simple, il suffit de remarquer que toutes les applications du traitement automatique du langage, sont le résultat de la rencontre et du dialogue entre les spécialistes des différentes spécialités linguistiques et branches informatiques pour saisir la dimension interdisciplinaire du domaine en question.

C'est donc d'abord par un accord tacite, que ces interactions prennent forme. (scientifiques, administratifs, et littéraires) tous participent à un processus interdisciplinaire d'abord implicite en interagissant et en utilisant tous ces outils. Par exemple dans le cas des logiciels de traitement de texte - rencontre(mariage comme le considère certains spécialistes) de l'orthographe, la grammaire, la stylistique et de l'informatique-, ces outils sont utilisés à l'heure qu'il est - et notamment le Word- par des utilisateurs appartenant à différentes spécialités -premier niveau d'interdisciplinarité- d'une part. D'autres part ces mêmes utilisateurs contribuent indirectement, et parfois directement à l'amélioration de ces outils et notamment à travers différentes campagnes de statistiques et d'évaluation organisée par des institutions et des agences spécialisées en génie linguistique comme l'agence « Elda ».

C'est un cas à l'image de la double articulation de « Martinet », de double interdisciplinarité. C'est-à-dire qu'il y a : d'abord d'un point de vue conception de l'outil un premier niveau d'interdisciplinarité, le principe d'interaction et de collaboration entre spécialistes de différents domaines, ensuite un second niveau d'interdisciplinarité, l'utilisation de ces techniques par différents domaines.

Ce travail de collaboration entre différentes disciplines et groupes de recherche nous mène droit vers la réponse à la seconde question, celle de savoir comment le traitement automatique des langues est-il interdisciplinaire, par quelles procédures ?

3.1.2 – Le comment de l’interdisciplinarité (par quelles procédures le traitement automatique des langues est-il interdisciplinaire) :

D’abord par le fait que l’objet ainsi que la structure scientifique du domaine en question est le résultat de la fusion de deux domaines principaux qui sont les sciences du langage avec ses différentes spécialités et l’informatique avec également quelques une de ses branches et techniques .

Ensuite à travers l’ouverture vers d’autres domaines plus vaste à l’image de la statistique, la traduction , les sciences cognitives, d’une manière générale , les science de l’éducation , la recherche d’information, l’analyse des genre et des texteetc.

Et ce à travers ; en troisième lieu des efforts conjugués de spécialistes travaillant en collaboration ainsi que des opérations réciproques d’interrogation et des « aller- retours » effectués sans-cesse entre ces différentes disciplines .

Chacune des spécialités concernées apporte sa brique à cet édifice qu’est l’ingénierie linguistique en différents termes :

Il peut s’agir pour les science du langage de :

-Corpus oraux ou écrits,

-Différentes théories d’analyse .

-Règles grammaticales , orthographiques et de conjugaison

Toutes données , ressources et principes linguistiques par exemple phonétiques , phonologiques , morphologiques , ou encore sémantique morphosyntaxiques , pragmatique, discursives...

L’informatique intervient avec :

Les techniques et langages de programmation

Les algorithmes.

Les différents types de calcul.

Les bases de données.

L'intelligence artificielle et ses techniques.

Les mathématiques à leur tour interviennent avec :

Les différentes catégories de logique.

Les statistiques et leurs différentes techniques. .

Les méthodes probabilistes...

Les sciences de l'éducation avec leur :

Différentes méthodes et théories (didactiques et pédagogiques).

L'analyse des textes et la catégorisation des genres via

Les différentes approches et théories d'analyse.

3.1.3-Résultat (quelles sont les résultats de cette interdisciplinarité):

Après avoir essayé de montrer ce qui fait du traitement automatique des langues une discipline interdisciplinaire , et par quelles procédures il l'est , nous aboutissons logiquement en troisième lieu au résultat de cette interdisciplinarité : ce sont l'ensemble des nouvelles spécialités, outils et applications qui ont vu jour (cf.la liste thématique d'applications et les familles d'outils , chapitre I) dans un premier niveau , et dans un second , c'est leur utilisation et mise à disposition d'autres domaines . Comme nous l'avons cité plus haut, nous allons essayer de nous approcher de l'un des résultats en question, le premier cas d'interdisciplinarité que nous présentons étant celui de :

3.2-L'analyse automatique des textes et de corpus :

L'analyse automatique des textes est un domaine du traitement automatique du langage . Il est une application à caractère pluridisciplinaire - en fait , parce qu'il est l'aboutissement de la fusion de l'informatique et de l'analyse des textes ,qui implique d'ailleurs un étroit travail de collaboration entre analystes et informaticiens ,une collaboration justement que les données de la société dans laquelle nous vivons nécessite ,et tend à devenir obligatoire , pas uniquement d'un point de vue technique et scientifique, mais également d'un point de vue pragmatique ,dans ce sens nous retiendrons le point de vue d'Anne Condamine dans son article « Discours , corpus , et traitement automatique » qui précise que :

« Le traitement de gros corpus pour l'acquisition de données plus riches implique quant à lui la mise en œuvre de méthodes (semi-) automatiques ou informatiques qui ont l'avantage d'être reproductibles. Ainsi union entre linguistique de corpus et outils informatiques est-elle scellée , contribuant au mouvement de fond vers un empirisme qui sous des formes plus ou moins radicale , peut être vu comme définitoire » (Anne Condamine, 2005 , 182)

Condamine, linguiste, enseignante et Directrice de Recherche , au sein de l'unité de recherche mixte de l'université de Toulouse, UMR 5263 ; « Cognition Langues Langage Ergonomie » ; équipe de Recherche en Syntaxe et Sémantique du centre nationale de recherche scientifique de France(le CNRS) et développant des projets de recherches dans le cadre de l'axe « [Structures sémantiques : des catégories lexicales au discours](#) » évoque l' « implication » entre traitement linguistique - qui devient informatisé- de corpus (le traitement à ce niveau sous-entend toutes les opérations linguistiques d'analyse , tout paliers et niveaux compris) et les multiples méthodes et techniques de traitement et d'analyse informatiques. .On parle d'informatique linguistique . Dans ses différents travaux de recherche ; Leech, quant à

lui évoque précisément un nouveau domaine qu'il nomme « Computer Corpus Linguistic » qui serait traduit en linguistique computationnelle de corpus: .Une implication qui devient de plus en plus importante, voire obligatoire, car elle est l'essence même de cette discipline qu'est le traitement automatique du langage (cf. .voir le caractère interdisciplinaire §1de ce chapitre) d'une manière générale , et de la catégorisation et de l'analyse automatique des textes précisément .

L'une des raisons pour lesquelles la mise en œuvre de ces méthodes (semi-) ou complètement automatiques et informatiques ;est aussi intéressante et avantageuse aussi bien pour les littéraires et linguistes que pour les développeurs des méthodes informatiques concernées (en l'occurrence les algorithmes et logiciels d'analyse informatique comme le logiciel Hyperbase qu'on va présenter plus loin) ,pour le traitement d'un nombre de plus en plus croissant de corpus, de données et de contenus , est le fait que ces méthodes et techniques sont concrètement très pratiques et opérationnelles -c'est l'une des raisons d'ailleurs par analogie qui a fait que le nombre de disciplines et d'utilisateurs ayant recours à l'informatique est de plus en plus considérable- l'aspect et le caractère opérationnel des différentes spécialités et techniques informatiques et notamment de programmation des multiples logiciels ; toutes spécialités confondues a permis en fait au genre humain d'obtenir des résultats , de gagner en temps et en moyens, plus encore ; d'atteindre un certain niveau d'aisance et de précision que jamais les humains auraient pu atteindre sans.

L'une des manifestations de cette opérationnalité ; est la capacité de ces méthodes et techniques à être facilement et rapidement reproduites et reproductibles ,on entend par reproductibles ici le fait de pouvoir appliquer à répétition et surtout dans des espaces de temps réduits ; les mêmes techniques et analyses a de très grand nombre de corpus (toutes natures, catégories et supports compris) ce qui logiquement permet de gagner en temps et en moyens lorsque le nombre de documents ou de textes à analyser est considérables . Quant à la qualité des résultats obtenus, elle s'améliore en fait et mesure que les méthodes utilisées se développent.

Ce mouvement de performance étant en perpétuelle évolution et c'est principalement par rapport au temps qu'on peut le caractériser. Autrement dit, plus le temps passe, plus les techniques et les approches deviennent de plus en plus performantes, c'est ce qui explique par ailleurs les différentes générations et versions d'un même logiciel, système, programme ou outil informatique, des versions éditées

chaque années ou chaque deux jusqu'à quatre ans, par exemple, « Word », ou encore « Windows » qui présente à chaque reprise des nouvelles versions.

Dans la poursuite de cette implication, et en la reprenant ; avec des dimensions plus universelles et plus globalisantes, on pourra se rendre compte qu'il n'est pas uniquement question d'aspect pluridisciplinaire, ou encore, interdisciplinaire caractérisant les domaines concernés, mais d'un lien plus important qu'est celui de l'union. Une union entre, linguistique et informatique dans un premier temps, et une fois de plus entre sciences du langage, littérature et informatique. La fusion et l'union de ces disciplines donne naissance à une toute nouvelle informatique, appelé désormais « Informatique des Sciences Humaines », une discipline qui avant toute autre considération, -et c'est là une réalité tangible à partir des nouvelles pratiques qui naissent au quotidien- vient comme suite logique et inévitable aux progrès technologiques, aux profondes mutations que connaissent les sociétés et toutes leurs structures, et aux besoins qui lui sont parallèles.

L'union (ou le mariage, comme le considère certains spécialistes) que nous évoquons se justifie notamment par tous les efforts de travail collectif et interactif et collaboratif qui unie un nombre de plus en plus important de spécialistes et d'équipes de recherches qui tendent à devenir de plus en plus mixtes ; comme celles dont est constituées par exemple les équipes de recherche du Centre Nationale de Recherche Scientifique (CNRS) et l'Institut National de Recherche en Informatique et en Automatique de France (INRIA), ou encore des différents laboratoires spécialisés en la matière à l'échelle universelle.

Elle s'explique également – et c'est d'ailleurs d'un point de vue épistémologique ce qui a fait que le rapprochement et les liens s'effectuent entre informatique et linguistique d'une manière aussi étroite - par les points en commun qui existent entre ces deux domaines. D'abord les Sciences du langage et l'informatique utilisent le langage comme moyen de communication (naturelles pour les langues, les langages de programmation pour l'informatique). Ce même langage ; qui est comme nous l'avons cité plus haut le détonateur de l'existence des disciplines de savoirs.

Ensuite, elles se basent sur le traitement de l'information, d'où le Traitement automatique des langues- et entrent en jeu quand celle-ci a déjà été mise en place. Elles ont entre autre comme objectif d'étudier et d'analyser comment différentes

entités communiquent entre elles en ayant recours à plusieurs niveaux d'étude du langage (soit naturel ou informatique de programmation) structurés., d'où l'appellation nouvelles technologies d'analyse de l'information.

Un troisième élément important est que les deux disciplines en question se rapprochent également par leur méthodologie d'étude et de recherche qui pourrait se caractériser par ces mots d'ordre qui représentent sommairement le principe du traitement automatique du langage et qui sont : délimiter/chercher , analyser/traiter , classer . Les linguistes cernent l'information, une fois délimitée, ils effectuent un certain nombre d'opérations analytiques (qui diffèrent selon la spécialité et le niveau ou l'angle d'analyse envisagé), avant de les classer par exemple dans des ouvrages et ressources . La machine également permet de rechercher , mais aussi reçoit des informations en formes d'instructions , données ou des éléments programmables , les analyse et les traite - en les transformant en un langage qu'elle comprend ,c'est-à-dire des séquences numériques ou des langages de programmation- ensuite les classe de différentes manières à savoir en base de données afin qu'elles puissent être consultables à chaque fois qu'un utilisateur en a besoin.

Toutes ces opérations foisonnantes de rapprochement et d'interactions ont des objectifs et des résultats importants, voire capitales, elles permettent par exemple de mieux comprendre les mécanismes du langage , de l'écriture ; de faciliter leurs compréhensions, acquisition et exploitation .Sans oublier le fait qu'elles permettent de gagner énormément de temps.

Cette analyse automatique ou informatique de textes et de corpus²¹, a recours aux méthodes statistiques pour effectuer ces études linguistiques. C'est ce qu'on appelle l'analyse, les méthodes ou encore la linguistique statistique. Les analyses peuvent s'effectuer sur un ou plusieurs contenus ou textes, appartenant à différents domaines composant ainsi un corpus textuel.

3.3- Linguistique statistique :

Connue sous plusieurs noms à savoir, la linguistique statistique ,la statistique lexicale ,ou encore la linguistique quantitative ,Charaudeau & Maingueneau la définissent comme suit : « La statistique lexicale qui se veut exhaustive,

²¹ Corpus oraux également

systématique et /ou automatisée a pour but le comptage des unités lexicales d'un texte ou d'un corpus » (Charaudeau & Maingueneau)

Elle a ; en d'autres termes ; comme but général d'identifier les structures mathématico-statistiques propres au langage, considérées en tant que phénomène objectif. Les modèles et résultats computationaux ,quantitatifs , numérisés ou encore informatisés finissent par déboucher sur une linguistique nommée « la linguistique de corpus » qui fonde ses enquêtes et ses traitements sur des recueils de données linguistiques de différentes typologies au format électronique ou numérique .Et c'est le trait pertinent et ce qui caractérise selon les chercheurs à savoir Geoffrey Williams (université de Bretagne sud) cette nouvelle linguistique.

Cette linguistique fondé sur l'étude, le traitement et l'exploitation des textes au format numérique et informatique conçoit et définit le corpus linguistique justement comme étant « une collection de textes informatisés » .

Il faut donc créer un corpus informatisé, avant de procéder concrètement à l'analyse textuelle, statistique, lexicale ou documentaire automatique voulue .La création des corpus informatisés est considérée comme étant l'une des caractéristiques et l'un des plus importants résultats de la communication Homme-Machine(CHM). Elle est la phase intermédiaire entre le corpus à son état original et automatisé.

Nous pouvons repérer trois phases principales durant cette création :

1- Le choix des textes, leur typologies ainsi que différentes caractéristiques : généraux, spécialisées, de référence, archives et leur formats, ainsi que les critères d'analyse organisés éventuellement en grille d'analyse..

2-Une fois le choix précisé et la sélection effectuée, vient la phase d'acquisition et de transformation de données qui nécessitent de rendre le texte en MRF (Machine readable form), c'est-à-dire le format que la machine peut déchiffrer. A travers un code précis (cf, le paragraphe suivants) . Cette phase pourra s'effectuer à travers la frappe (la saisi), le scanning, la dictée, dans le cadre des techniques de traitement, synthèse et reconnaissance vocale.

3-La troisième étape est celle de la codification des données : lors du passage de la forme écrite a la forme déchiffrable par l'ordinateur ,le texte perd quelques caractéristiques , raison pour laquelle la nécessité pour l'ordinateur d' avoir recours à l'utilisation de codes précis comme le ASCII(American Standard Code for

Information Interchange) l'équivalent en français du « Code américain normalisé pour l'échange d'information » ,pour le traitement informatique de ces textes, se fait ressentir et s' avère nécessaire.

Ce code constitué de caractères alphanumérique est considéré comme étant la norme de codage de caractères en informatique la plus connue ,et la plus largement et relativement- aux changements et caractéristiques des langues- compatibles .Inventée en 1961 par Bob Berner , Elle est à la base de nombreuses autres normes nationales ou internationales qui l'étendent telles que « ISO/CEI 8859 » , « GB 18030 » ou encore « Unicode ».

Une fois le corpus créé et informatisé, l'utilisateur pourra à ce moment procédé à l'analyse proprement dite. A travers cette brève présentation du fonctionnement du logiciel Hyperbase, ainsi que les logiciels de la gamme d'Ascetic, nous allons montrer comment les données citées précédemment se conjuguent pratiquement afin d'aboutir à ces deux logiciels d'analyse automatiques statistiques et documentaires. :

3. 3.1-Le premier model : le logiciel d'analyse automatique Hyperbase :

Description, illustration et présentation du fonctionnement :

Le premier model que nous présentons est celui du logiciel d'analyse automatique statistique et lexicale Hyperbase.

Hyperbase est un logiciel pour le traitement documentaire et l'analyse statistique des corpus textuel. Conçu et programmé en Français , il s'applique cependant indifféremment à toute langue qui utilise l'alphabet latin .

Une fois installé, le logiciel crée dans le disque dur une base de donnée composée de textes de différentes typologies et différentes périodes, et en permet divers analyses .On peut par exemple traité, analyser et comparer différents textes ,et différentes bases de documents et l'interrogation s'enrichit et s'avantage des diverses fonctions et notamment hypertextuelles²² du logiciel .Il suffit par exemple de cliquer sur un mot pour connaître sa répartition dans le texte de travail ou d'analyse et être conduit dans les passages où le mot est employé. Ces opérations

²² L'hypertexte désigne un mode d'organisation des documents textuels informatisés caractérisé par l'existence de liens dynamiques entre ses différentes sections menant vers d'autres sections. Dans le Web, des mots ou des expressions soulignés ou encore des images indiquent les liens sur lesquels on clique à l'aide de la souris .

peuvent se faire à partir de l'index (cf. voir exemple figure 4) .L'hypertextualité du logiciel se manifeste également dans la possibilité de se connecter à la Toile pour comparer les résultats ou approfondir les analyses .Le menu du sommaire est constitué horizontalement des fonctions documentaires, et verticalement des fonctions statistiques.

L'analyseur peut avoir recours à une base textuelle déjà créée dans le logiciel , comprenant des textes littéraires français comme « Zadig » ou « Emile » et étudier et expérimenter de cette manière les différentes fonctionnalités du logiciel ou alors créer sa propre base, c'est-à-dire son propre corpus d'analyse. Il existe aussi une option Vers pour la poésie. On peut également comparer le corpus traité à l'un des corpus littéraire de « Frantext » qui est une base de donnée lexicale et textuelle qui comprend 117 million de mots, et s'étend sur cinq siècles .On a la possibilité d'utiliser le corpus en entier , ou bien de se limiter à une tranche ou une période temporelle précise. Dans le cas d'une langue étrangère, un dictionnaire de référence et d'informations relatives au corpus propre à la langue choisie peut être installé en plus .Si un nouveau texte est introduit dans la base de donnée, le logiciel procède à la reconnaissance et à l'analyse des formes du texte, à l'indexation proprement dite , à la création d'un dictionnaire de fréquences alphabétique et hiérarchique permettant la visualisation des résultats en tableau par exemple. Par ailleurs il calcule les spécificités externes et internes propres au texte , et en mesure la richesse lexicale ou l'utilisation du vocabulaire .Une fois que la base est créée , l'utilisateur peut exploiter les différentes fonctions d'Hyperbase.

Les sous- fréquences dans les textes sont réparties en binômes, dont le premier élément indique le numéro d'ordre du texte, le second la fréquence du mot dans ce texte.	16 modestement , 1 1 3 1 4 1 6 1 7 1 9 2 10 3 11 1 16 1 19 2 21 1 22 1	N° 2 Paysan 2	N° 3 Zadig 1
	11 modestes , 5 1 6 3 7 1 11 1 12 1 17 1 19 1 21 1	N° 4 Candide 6	N° 5 Héloïse 13
	35 modestie , 4 1 5 5 6 14 8 19 6 10 2 12 1 13 2 1	N° 6 Emile 43	N° 7 Atala 9
	1 modesties , 8 1	N° 8 Rancé 16	N° 9 Chouans 6
	2 modicité , 9 1 10 1	N° 10 Pons 4	N° 11 Indiana 6
	2 modifia , 17 1 21 1	N° 12 Mare 5	N° 13 Bovary 4
	1 modifiaient , 22 1	N° 14 Bouvard 6	N° 15 Une Vie 3
	1 modifiant , 20 1	N° 18 Bête 4	N° 20 Stortz 3
	4 modification , 19 2 20 1 21 1	N° 21 Swann 2	N° 22 Temps 4
	7 modifications , 5 2 10 1 14 2 15 1 17 1	TOUS LES TEXTES	
	6 modifie , 20 1 21 1 22 4		
	10 modifié , 15 1 19 4 21 1 22 4		
	1 modifiée , 22 1		
	2 modifiées , 9 1 21 1		
	1 modifiant , 22 1		
12 modifier , 11 1 16 1 19 1 20 2 21 5 22 2			
2 modifierent , 12 1 16 1			
2 modifiés , 19 1 22 1			
2 modique , 5 1 10 1			
2 modiste , 13 1 14 1			
2 modulait , 13 1 15 1			
2 modulation , 5 1 21 1			
7 modulations , 5 1 10 2 13 3 22 1			
1 modulés , 22 1			
1 modus , 6 1			
2 Moedler , 19 2			
1 Moedler , 19 1			
1 Moelder , 19 1			
1 moele , 5 1			
5 moelle , 14 2 19 1 21 1 22 1			
3 moelles , 15 2 18 1			
3 moelleuse , 9 1 14 1 21 1			
1 moelleuses , 9 1			
5 moelleux , 11 2 13 1 16 1 21 1			
1 moellon , 14 1			
2 moellons , 21 2			
137 moeurs , 2 2 3 1 4 6 5 13 6 43 7 9 8 16 9 6 10 4 11 6 12 5 13 4 14 6 15 3 18 4 20 3 21 2 22 4			

moeurs **fréquence totale: 137**

CLIQUEZ SUR UN TEXTE (ou sur TOUS) pour y repérer les contextes du mot "moeurs"

Cliquer AILLEURS dans cette fenêtre pour l'effacer

Figure 4 : Une page du dictionnaire (ou index)

Exploitation et représentation graphique des résultats :

Pour ce qui est des mots, le chercheur a la possibilité, de voir la fréquence répartie dans le texte ou dans le corpus entier. A partir de ce mot, il peut voir la représentation graphique de sa fréquence en la comparant à d'autres mots et il pourra consulter un index hiérarchique ordonné par fréquence. Les fonction hypertextuelles permettent de voir sur le champ le contexte d'un mot.

Hyperbase propose des outils permettant d'assumer une exploitation pratique, méthodique et pragmatique de la documentation. Par exemple l'outil « Contexte », en sollicitant cette option, par rapport à un mot ou une expression désirées, l'utilisateur se voit offrir la possibilité de le voir placé à l'intérieur de différents textes où ils sont employés. S'il clique sur une des passages affichées, il a la possibilité de voir la page entière où le mot apparaît. Le logiciel effectue des allés-retours permettant une vérification ultérieure des résultats et une interrogation plus

ample. Une autre fonction documentaire est représentée par « Concordance » qui extrait également des contextes plus étroits qui tiennent par exemple en une ligne.

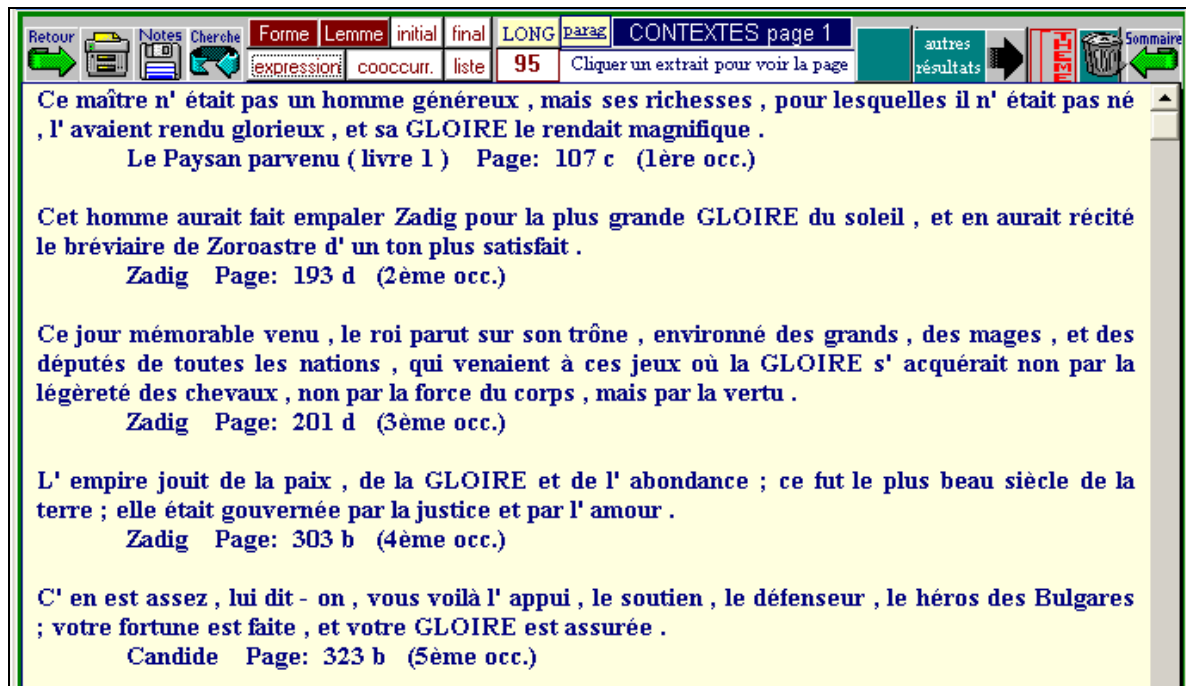


Figure 5 ; exemple du résultat de la fonction « Contexte »

Le recours aux illustrations

L'une des caractéristiques du logiciel Hyperbase est le recours à la représentation visuelle et graphique pour illustrer les résultats statistiques. Il y a par exemple la possibilité de voir sous forme d'histogramme les écarts entre les fréquences d'un mots observé dans le texte , et la fréquence théorique qu'on aurait du s'attendre à obtenir en considérant la proportion du texte par rapport à l'ensemble .

Les listes de résultats possèdent à leurs tour un certain nombre de fonctionnalités et de représentations graphiques , en colonnes ou en lignes , et on peut les soumettre à deux types d'analyse ,factorielle et arboré.

D'autres outils permettent de voir l'évolution du langage, la richesse du vocabulaire. c'est le programme factorielle qui calcule les spécificités internes de la langue.

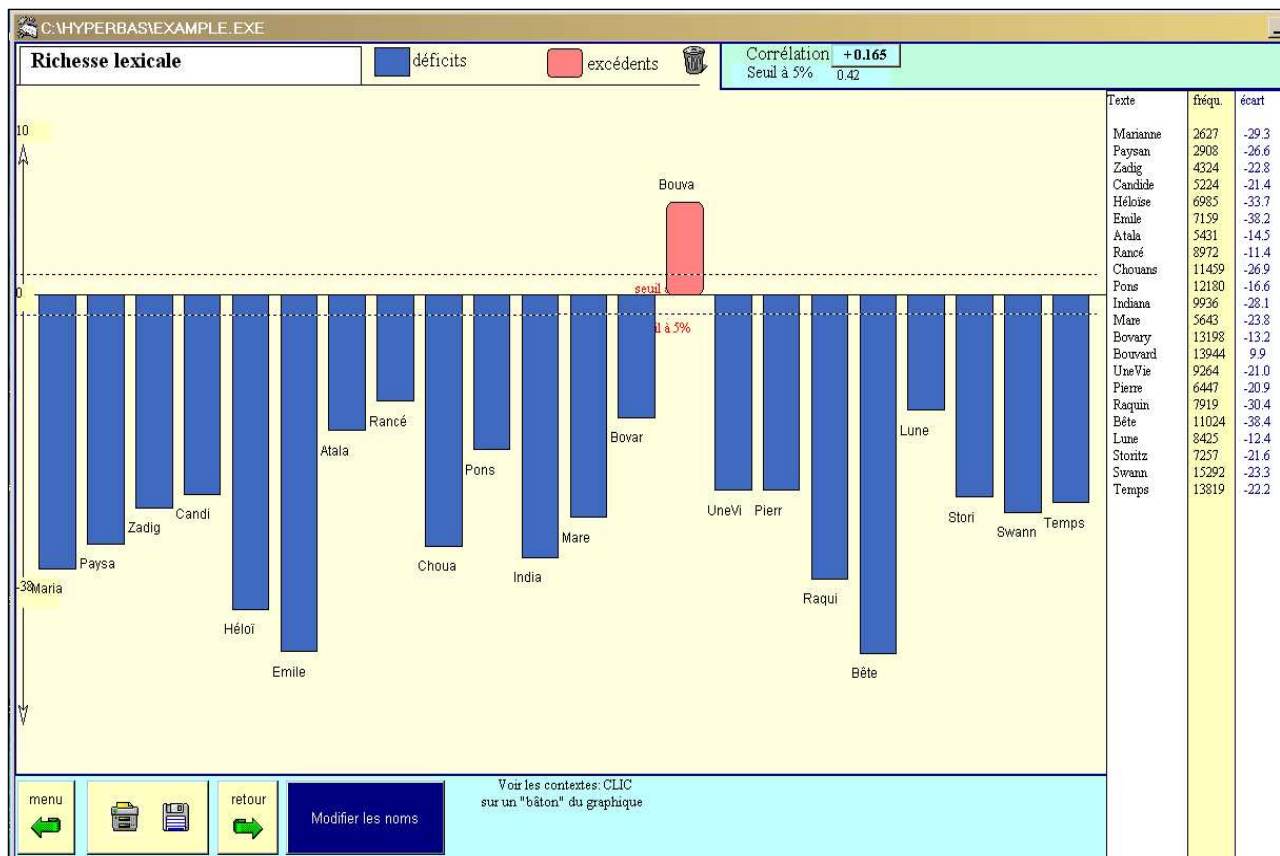


Figure 6 : exemple d'une représentation en forme d'histogramme

3.3.2-Les trois logiciels d'analyse automatique statistique et documentaire de la gamme d' « Ascetic » :

Nous citons brièvement un deuxième modèle à titre plus ample d'exemple :

L'ensemble des logiciels d'analyse automatique de documents de la gamme « Ascetic ». Pouvant être exploité et mis au service de la documentation, la gestion des connaissances, l'intelligence économique, la veille technologique, la revue de presse, les analyses stratégiques, scientifiques, éducatives, linguistiques...la gamme « Ascetic » est à la fois constituée :

D'un le logiciel d'analyse de textes Tropes :

Conçu pour la classification sémantique, l'analyse qualitative, l'extraction de mots clés et la constitution de thésaurus, le logiciel Tropes est un logiciel pour le traitement documentaire, les études de marché, l'analyse d'entretiens, les études littéraires, ...etc.

D'un moteur de recherche [Zoom](#) -

Avec son moteur d'indexation en langage naturel, ses classifications sémantiques intégrées, Zoom est un moteur de recherche conçu pour la gestion électronique de documents, l'Intelligence Economique, l'analyse de presse, la veille technologique

D'un système d'indexation- ; classification Sémantique pour serveurs

Index est un système de classification et d'indexation conçu pour la gestion de différentes applications et systèmes comme la documentation en langage naturel, la Gestion des connaissances...etc. Pour les grandes entreprises ou les grands projets notamment scientifiques.

Apportant une approche sémantique et pragmatique- entre autre en ayant recours à l'instar d'Hyperbase a des comparaisons entre les résultats d'une analyse donnée effectuée et la base de donnée contenue dans le logiciel qui est constituée des différents textes appartenant à différentes périodes ou genres- , Tropes permet d'obtenir des analyses reproductibles (deux personnes étudiant un même texte et observant la même technique d'analyse obtiendront des résultats semblables).

Parmi les outils d'analyse sémantique contenus dans Tropes, on peut citer :

- La classification automatique des mots du texte.
- La détection du contexte global du document.
- La catégorisation des mots-outils (analyse qualitative) .
- La localisation de séries chronologiques (analyse du discours) .

Le calcul des relations entre thèmes (lexicométrie) .

Des dictionnaires sémantiques personnalisables, appelés Scénarios, conçus pour fabriquer rapidement des outils d'interprétation ou des thesaurus automatiques ;
L'extraction terminologique (couplée à l'analyse sémantique) .

Le logiciel d'analyse de textes Tropes est interconnecté avec le moteur de recherche sémantique Zoom, pour constituer Tropes Zoom, une suite complète de recherche et de traitement automatique et documentaire de l'information.

Nous avons essayé à travers ce chapitre de mettre en exergue le caractère interdisciplinaire de l'ingénierie linguistique, ce caractère donnant lieu à de nouvelles pratiques et disciplines, qui montrent à chaque fois un peu plus à quel point il est intéressant et important de créer un dialogue entre les différentes disciplines. Nous avons rendu compte des points en commun qui lie la linguistique à l'informatique, et que se sont justement les passerelles entre ces deux disciplines, donnant lieu à des interactions aboutissant à de nouvelles applications et nouvelles formes d'analyse.

Parmi ces nouvelles formes d'analyse automatique figure l'analyse automatique statistique textuelles et par conséquent la linguistique de corpus. Les logiciels Hyperbase et la gamme des logicielles d'« Ascetic » en sont deux exemples par excellence.

Nous poursuivrons en nous attaquant à autre cas d'interdisciplinarité dans le chapitre suivant, celui de la traduction automatique.

CHAPITRE IV

Traduction automatique

Un travail relevant du traitement automatique du langage, doit nécessairement, pour des considérations historiques, épistémologiques et techniques que nous tenterons d'expliquer, opérer un passage par la Traduction Automatique (TA). Ceci dit, la Traduction Automatique est également un des domaines interdisciplinaires du traitement automatique du langage puisqu'elle est le résultat de l'informatisation du processus de traduction. Dans ce chapitre ; nous évoquons donc brièvement la question de l'interdisciplinarité de la Traduction Automatique, ouvrons une parenthèse épistémologique sur l'origine du traitement automatique du langage, dont la traduction automatique en est la première application, présentons le logiciel de traduction automatique et ses composants, ainsi que son principe de fonctionnement. Et finissons par présenté un des problèmes que rencontre la Traduction Automatique, celui de l'ambiguïté et illustrons à travers le traducteur automatique « Reverso » , ensuite concluons.

Les ouvrages traitants de la Traduction Automatique et entre autre de son caractère interdisciplinaire étant relativement rares, sinon écrits en anglais, nous avons eu recours beaucoup plus aux différents travaux et rapports de recherches des laboratoires spécialisées ainsi qu'aux articles authentiques .

4.1- Traduction automatique et interdisciplinarité :

La première des constatations à faire au sujet de la Traduction Automatique est le fait qu'elle soit également une discipline interdisciplinaire (cf. rappelons que tout le traitement automatique du langage est un domaine interdisciplinaire), simplement par le fait qu'elle consiste comme nous l'avons déjà dit en l'automatisation à travers la programmation informatique ²³de cette pratique qu'est la traduction. Ainsi Roland Raoul Kouassi, précise que :

²³ Par opposition aux autres formes de programmation.

« La traduction automatique pose le problème des universaux du langage et celui des paramètres de fonctionnement inhérents à chaque langue naturelle...la traduction automatique est basée fondamentalement sur des approches qui s'organisent autour du fait de la langue en tant que phénomène, c'est-à-dire en tant que fait réalisé ou produit. On établit des correspondances entre des mots et expressions. On écrit des algorithmes de traitement informatiques pour programmer ces correspondances ». (Roland Raoul Kouassi .2005)

Il appartient au champ de la traduction et désormais de la traduction automatique de traiter tout ce qui est en rapport avec les universaux du langage ainsi que les mécanismes et principes (les règles) de fonctionnement de chaque langue .Ce traitement qui consiste à établir des correspondances et parallélismes entre les mots, les phrases et les paragraphes appartenant à deux ou plusieurs langues ; et donc traduire, est programmé à l'aide de langages ,d' algorithmes et de techniques de traitement informatique destinées pour. C'est une fois de plus à ce niveau , c'est-à-dire le travail de collaboration , ce mouvement interactif et ce dialogue entre traducteurs, linguistes ou ce qu'on appel dorénavant « Talistes »²⁴ ainsi qu'informaticiens , que se traduit le caractère interdisciplinaire de la Traduction automatique.

Nous essayerons de montrer plus bas , après le passage sur la question épistémologique ; à travers les composantes, ainsi que quelques uns des principes et paramètres de fonctionnement des traducteurs automatiques, comment se conjugue et se concrétise ce caractère interdisciplinaire nous expliciterons de la sorte l'apport de la linguistique ainsi que celui de l'informatique .

²⁴ Spécialistes du traitement automatique du langage.

4.2.-Questions épistémologiques :

D'un point de vue chronologique , la traduction automatique a précédé l'ingénierie linguistique à proprement dite car l'élaboration du traitement automatique du langage a vu naitre le jour avec les premières expérimentations effectuées en traduction automatique ainsi que les premiers prototypes de calculateurs utilisés en cryptographie durant la seconde guerre mondiale « L'histoire du Traitement Automatique du langage commence par les expérimentation en traduction automatique qui débutèrent à la fin des années 1940 , aux Etats-Unis et en grande Bretagne » (Jacqueline Léon 2001.) . La première démonstration a eu lieu à New York en 1954, ce fut une double et historique traduction d'un passage de l'évangile (de l'anglais au russe puis à l'anglais) , les deux fameux passage sont :

« L'esprit est ardent mais la chair est faible »

« L'alcool est fort mais la viande ne vaut rien »

L'informatique a donc ouvert de nouvelles perspectives pour différents domaines et notamment l'étude des langues et très exactement la Traduction Automatique durant cette période qui marquent le début du traitement automatique du langage du moment où :

" Les premières recherches dans ce domaine été entamée dans l'euphorie des premiers grands succès de l'informatique : les réussites remportées dans le traitement des données numériques avaient alors amené les informaticiens à s'intéresser à d'autres domaines que les domaines purement mathématiques, en particulier celui du langage (...)l'idée qui animait les chercheurs de l'époque était que pour travailler sur le langage , il s'agissait de procéder à une simple substitution de données : les données numériques seraient remplacées par les données alphanumériques que constituaient les mots ". (Frédérique LAB .1988)

Frédérique LAB précise également que :

"La Traduction Automatique des débuts était essentiellement une traduction de mots , de fait la part dévolue aux dictionnaires était énorme , celle consacrée à la syntaxe au minimum. Les problèmes relatifs au sens d'une phrase , d'un texte , n'étaient pas abordés en tant que domaine de recherche spécifique, pas plus que l'activité de traduction n'était envisagée comme une activité ayant sa problématique propre. (*ibid*)

C'est après plusieurs tentatives et travaux de recherches et d'échecs de systèmes , comme il est habituellement le cas dans le développement des disciplines -,de la première génération pour que les ingénieurs impliqués dans ces travaux accordent aux recherches linguistiques(syntaxe et sémantique) la part qui leur revenait, afin de justement permettre le développement de logiciels présentant des résultats comme le font ceux disponibles à présent, comme par exemple les logiciels de Traduction Automatique "Systran" ou " Reverso" .

Jusqu' en 1960, pour des raisons politiques , militaires , et stratégiques, et notamment durant la guerre froide ²⁵ ; le développement de la traduction automatique a atteint un niveau mondial , et c'est pendant cette décennie qu'apparaissent la plus part des idées et des méthodes encore en cours de nos jours dans la plus part des théories et applications du traitement automatique des langues naturelles .

Et pour une approche plus pragmatique :

4.3 Logiciel de traduction

4.3.1-Définition :

Un logiciel de traduction peut être définie comme étant une application ou un

²⁵ Les outils informatiques et notamment la traduction automatique étaient considérés durant la guerre froide, comme un moyens stratégique de défense au même au même titre que le reste des politiques et des armes.

outil du traitement automatique du langage qui permet d'obtenir de façon automatique (informatique) une traduction de textes (lettres, rapports, articles, sites web...) d'une langue (source) vers une autre langue (cible).

Le souci majeur des spécialistes, des groupes de recherches et développeurs ; est de faire en sorte que le texte traduit doit être autant que possible correcte dans la langue cible et restituer l'information contenue dans le texte d'origine .

Jusqu'à présent (le mouvement de développement en perpétuelle évolution) les logiciels de traduction peuvent fonctionner en environnement PC ,Internet , Intranet²⁶ soit sous la forme d'une application autonome (logiciels de traduction par exemple) , soit sous la forme de fonctions de traduction directement intégrées dans une application (dans un Intranet, dans les moteur de recherche ;comme Google, dans les barres d'outils de certains logiciels qu'on peut ajouter à l'ensemble des barres d'outils qu'on installe dans un ordinateur, ou encore dans les logiciels de traitement de texte Word ou Excel...).

Il existe par ailleurs d'autres outils d'aide à la traduction qui sont complémentaires à un logiciel de traduction ce sont les mémoires de traduction comme (Trados) et dictionnaires électroniques bilingues comme (Collins) ou multilingues comme(EuroDico).

4.3.2- Composants d'un logiciel de traduction

En règle générale ; un logiciel de traduction est composé :

1-De règles linguistiques.

2-De dictionnaires conçus spécifiquement pour le logiciel de traduction, sachant que pour chaque application, des dictionnaires spécifiques y sont implémentés.

3-D'une interface : permettant selon les produits (selon les traducteurs) de conserver le format, réviser la traduction, adapter et personnaliser les dictionnaires...

²⁶Un intranet est un réseau locale et privé proper par exemple à une (entreprise) ou à un espace donné qui utilise les technologies de l'Internet., mais ne s'ouvre pas aux connexions publiques contrairement à la Toile.

4.3.2.1- Les règles linguistiques :

Le moteur de traduction est le cœur technologique du logiciel de traduction. C'est là où se réalise concrètement la traduction automatique. On y applique les règles linguistiques de transformation propre à chaque langue. Les étapes de traduction automatique sont les suivantes :

1-L' analyse du texte source (cf. c'est le principe d'analyse cité dans le § 7 du chapitre I)

2-L' application de règles linguistiques de transformation et de désambiguïsation

On entend par transformation à ce stade ; la transformation linguistique qui s'opère lors du passage d'une langue à une autre, chaque langue ayant ses propriétés et qui mène entre autre vers la désambiguïsation qui est le retour sur les constructions linguistiques équivoques dans le souci de les expliciter. L'ambiguïté est considérée comme étant l'un des problèmes principaux de la traduction automatique (cf. nous y reviendrons dans le §6 du présent chapitre).

Les règles linguistiques appliquées peuvent être : l'ordre des mots, les informations morphologiques, syntaxiques et grammaticales, la sémantique, l'analyse pragmatique des composants de la phrase. Plus ces règles sont fines ; plus la traduction obtenue est précise.

3-En troisième lieu : la génération de la traduction : la production de la traduction à travers des algorithmes de traitement informatique précis.

Dans la poursuite de l'interdisciplinarité que nous avons évoquée plus haut, l'analyse du texte et l'application des règles linguistiques constituent la partie linguistique du processus de traduction automatique, donc le travail des traducteurs et linguistes. La génération (la programmation) de la traduction, est ce qui relève de l'informatique.

4.3.2.2- Dictionnaires : (spécifiques)

Pour ce qui est des dictionnaires intégrés aux logiciels de traduction, Ils ne doivent pas être comme dans les premières générations de traducteurs automatiques, de simple dictionnaire de mots faisant uniquement office d'une liste mots ou d'expressions avec leurs traductions, Chaque construction linguistique doit être définie avec des informations linguistiques précises (syntaxe, sémantique

pragmatique...) dans la langue source puis dans la langue cible. Ces informations sont ensuite gérées par le moteur de traduction. Plus les dictionnaires sont riches, plus la traduction obtenue est précise.

4.3.2.3- L'interface

L'interface²⁷ des logiciels de traduction a un rôle important, puisque c'est elle qui permet la réalisation et l'exploitation de la traduction. En fait une fois le texte dans la langue cible est obtenu, l'utilisateur souhaitera le relire, le vérifier et le personnaliser pour l'adapter à son style, à son activité, à ses besoins, bref à son niveau... Ce besoin de révision et d'adaptation dépend de l'utilisation finale de la traduction (compréhension, diffusion, publication externe...) et s'effectue via l'interface.

L'interface d'un logiciel de traduction se compose :

- 1- De différents modules de traduction.
- 2- D'outils de traduction et de révision qui facilitent le processus de traduction et de révision du texte.

1-Modules de traduction :

Les modules de traduction peuvent être définis comme étant l'ensemble des éléments, des options et des fonctionnalités qui constituent d'un point de vue général le logiciel de traduction. Nous pouvons citer la :

Présentation du texte source et de sa traduction sous forme de deux fenêtres l'une en dessus de l'autre ou l'une à côté de l'autre, (comme nous allons montrer dans l'exemple plus bas)

Conservation de la mise en page , option que ne fournissent pas l'ensemble des traducteurs.

²⁷ (Cf. voir définition d'interface dans le chapitre V)

Intégration dans des logiciels Microsoft Word, Excel (95, 98, 2000), dans les barres d'outils d'Internet Explorer (4, 5), ou des moteurs de recherche.

Traduction de documents en intégral ou encore de sites web en partie ou complètement.

2 Outils de traduction et de révision :

Comme leur nom l'indique, les outils de traduction et de révision sont l'ensemble des fonctionnalités permettant une relative amélioration du résultat de la traduction automatique, il s'agit de :

Alignement des paragraphes source et cible,

Liste et marquage des mots inconnus.

Création et marquage des mots réservés (mots ou parties de textes que l'on ne souhaite pas traduire) .

Proposition d'alternatives de traduction quand le contexte est ambiguë.

Enregistrement du texte en mode bilingue.

Possibilité de traduction paragraphe par paragraphe ou texte entier.

4.4-Fonctionnement de la traduction automatique :

« La traduction implique deux messages équivalents dans deux codes différents. L'équivalence dans la différence est le problème cardinal du langage et le principal objet de la linguistique »

(Roman JACOBSON, T.1, 1963, p 80)

A l'encontre des théories de la traduction humaine et/ou classique ; issues de la linguistique comparative, et qui ne semblent pas rencontrer autant de problèmes que la traduction automatique, -bien qu'il y en ai- celle-ci se heurte en fait à des difficultés aussi bien d'ordre lexicale et syntaxique que sémantique et pragmatique.

C'est pour cela qu'à fin de fonctionner convenablement et obtenir les meilleurs résultats possibles; durant la phase de programmation , en plus des règles fondamentales de grammaire de la langue source et de la langue cible

ainsi que des exceptions à ces règles, tout vocabulaire doit être accompagné, - comme nous l'avons précisé dans la nature des dictionnaires des traducteurs automatiques (cf. §

5.3.2) -d'informations des plus diverses. En font partie :

-La catégorie grammaticale et les formes dérivées (substantif, verbe...) ;

-La signification dans le contexte ou la désambiguïsation (par ex. «Driver» signifie « chauffeur » dans un contexte général ou « pilote » dans un contexte informatique.

-Pour les substantifs, on détermine s'il s'agit d'un être animé ; son genre et son nombre, d'un objet inanimé, d'un lieu, ou encore d'un concept, c'est-à-dire les caractéristiques grammaticales. Pour les verbes, il faut déterminer par exemple s'ils sont transitifs ou intransitifs ou quel type de complément d'objet suit. Toutes ces procédures se passent durant la programmation.

4.4.1-Le principe de fonctionnement de la traduction automatique :

Lors de la traduction, le moteur de traduction procède comme suit : il découpe chaque phrase en mots pour les analyser et les recherche dans le dictionnaire. Le programme détermine alors les propriétés grammaticales, les traductions possibles de chacun des mots (en fonction du lexique contenu dans les dictionnaires), ainsi que les connaissances supplémentaires. Ensuite, il effectue une analyse syntaxique de la phrase. Il découpe la phrase dans les différents éléments qui la constituent. Il détermine ensuite la fonction de chaque mot dans la phrase comme sujet, verbe, complément, et leur affecte une valeur sémantique.

Ensuite, la traduction proprement dite a lieu en deux étapes :

Première étape : le transfert lexical qui peut se résumer au fait qu' à chaque mot est affectée sa traduction appropriée dans le contexte de la phrase (prise en considération du contexte). Outre les propriétés grammaticales, le programme tient compte des informations supplémentaires enregistrées dans le dictionnaire pour chaque mot. Plus il existe d'entrées pour chaque mot, plus le programme prend des décisions fondées et la traduction est meilleure.

En second lieu le transfert structurel qui assure le bon agencement des mots dans la phrase et entreprend d'autres adaptations structurelles. Par adaptations structurelles, on entend la mise en correspondance des schémas structuraux des langues, et catégories de transformations comme par exemple la transposition d'une proposition interrogative française en son équivalent Anglais ou inversement.

Le programme crée ensuite les formes de constructions linguistiques correctes et confère à la traduction sa forme définitive.

La figure suivante explique de nouveau le processus de traduction :

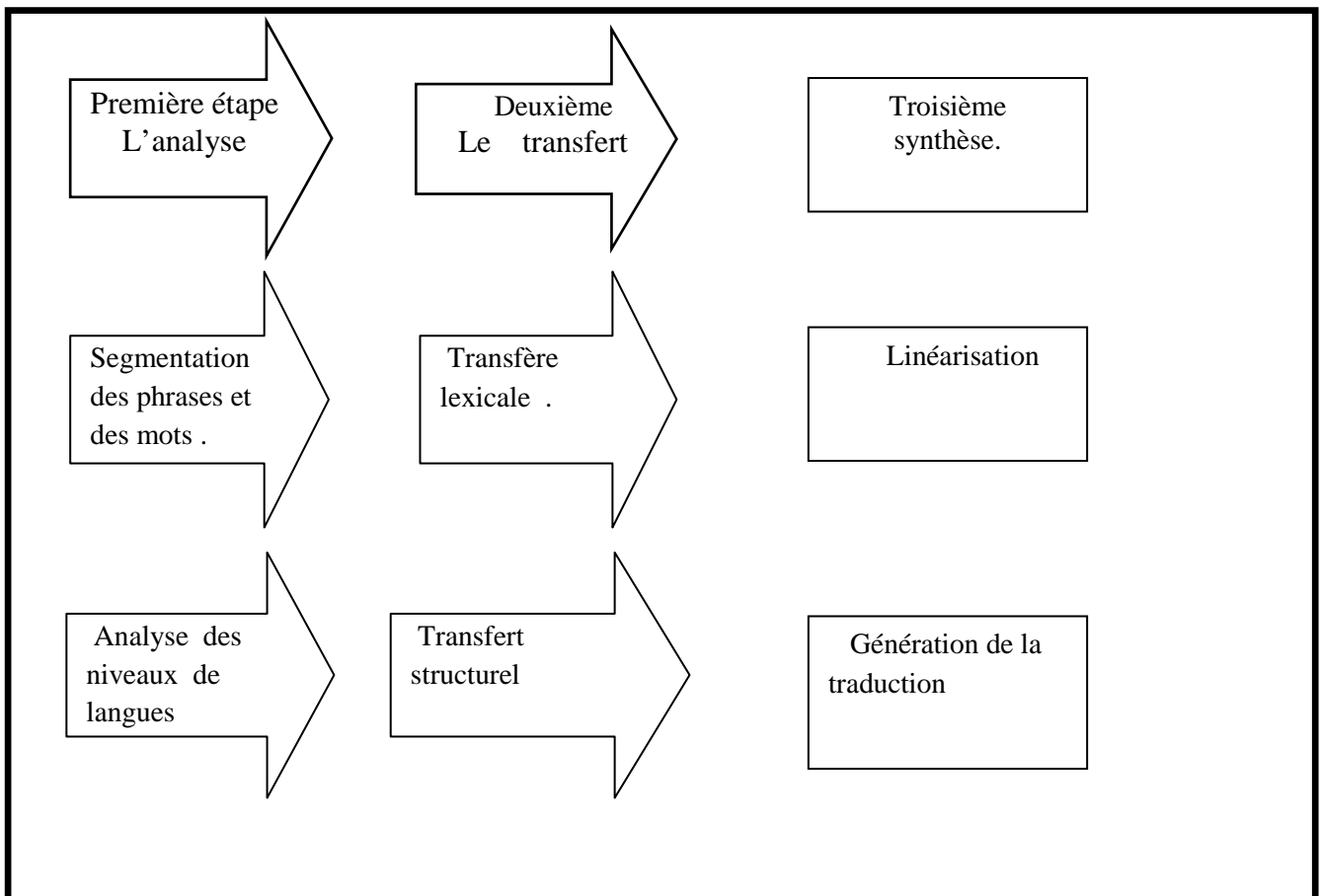


Figure 6 : Schéma récapitulatif du processus de traduction automatique.

4.5 Cas pratique : Le traducteur « Reverso » :

Parmi le nombre croissant des traducteurs automatiques qui ne cessent d'évoluer, nous avons opté pour le traducteur automatique « Reverso » du groupe « Sofitissimo ».

« Sofitissimo » est un éditeur et intégrateur de technologies linguistiques et de solutions multilingues destinées aux entreprises, aux administrations et au grand public.





Le logiciel en question est disponible sur la Toile à partir de plusieurs sites :

Nous l'avons atteint à partir du lien :

<http://www.reverso.com/> , dont l'interface est la suivante :

Nous présentons l'interface du logiciel constitué de plusieurs petits espaces afin de savoir ce qu'un logiciel de traduction peut à l'heure qu'il est faire comme activités (en l'occurrence Réverso). , nous présentons nos commentaires des espaces en forme d'intitulés soulignés en gras :

1- L'interface du logiciel avant la traduction:

<p><u>Le premier espace , celui de la traduction .</u></p>	 <p><u>Un espace conjugaison</u></p> <p>Tapez un verbe à l'infinitif ou une forme conjuguée: Langue : Conjuguer</p> <p></p>
<p>Tapez ou collez le texte à traduire:</p> <p>Direction de traduction :</p> <p>Traduire :</p> <p></p>	 <p><u>Un espace</u></p> <p>Mot ou expression : <u>réservé</u></p>




 <p><u>grammaire</u></p> <p>« Tout ce que avez voulu savoir à propos de la grammaire française: règles, astuces et quizz pour tester vos connaissances ! »</p>	<p><u>pour</u></p> <p>Rechercher : <u>les définitions .</u></p> <p>Dictionnaire :</p> <p>+ Webmaster Tools</p>
 <p><u>le traducteur sur portable</u></p> <p>Utilisez la traduction, le dictionnaire Reverso directement depuis votre téléphone portable sur: http://mobile.reverso.net</p>	<p><u>Là ou l'on retrouve</u> + Webmaster Tools</p> <p><u>C'est l'espace réservé au concepteur du site</u></p>


Figure 7 : Les principaux constituants de l'interface du logiciel de traduction automatique Reverso

(Avant la traduction)

2- Voici en quoi consiste (les éléments ajoutés) l'interface après la traduction :

Texte d'origine 

Le haut parleur est l'icône de la lecture orale

Traduction de Reverso en Anglais 

Le second est pour la lecture

du texte traduit

L'utilisateur aura la possibilité d'évaluer les résultats :

<p>Merci de votre avis. Vous pouvez aussi suggérer une meilleure traduction ou apporter un commentaire.</p>	<p>Selon vous, cette traduction est : Compréhensible, passable, mauvaise.</p> <hr/>
<p><u>Une possibilité d'évaluation du résultat</u></p>	<p>Vous pouvez modifier votre note en cliquant sur l'étoile correspondante .</p>

Il aura également la possibilité de l'améliorer ou la modifier

<p>Votre texte d'origine:</p>	<p>Proposez une meilleure traduction (optionnel)</p>
<p>Afficher la traduction de Reverso</p>	<p><u>La possibilité de l'améliorer</u></p>

Ainsi que de la commenter :

<p><u>Possibilité de la commenter</u></p>	<p>Ajouter un commentaire (optionnel):</p>
--	--

Les éléments ajoutés à l'interface après la traduction .

Le choix du traducteur « Reverso » se justifie par rapport au fait qu'il figure parmi l'un des meilleurs traducteurs automatiques à présent. Nous l'avons donc choisi de telle sorte à montrer que ce meilleur traducteur automatique, aussi performant qu'il puisse l'être, présente encore des failles et lacunes. Les langues traitées par celui-ci sont le : français, anglais, allemand, espagnol, russe.

L'interface de Reverso :

Les modules et les outils disponibles dans les produits « Reverso » dépendent de la version du logiciel et de l'environnement choisi.

4.6. Limites du traducteur :

La traduction automatique rencontre certes à l'heure qu'il est pas mal de problèmes et on y détecte beaucoup de failles. L'une des plus importantes lacunes, est le problème de la non-univocité des emplois et expressions linguistique et donc l'ambiguïté, qui résulte de l'abondance des nuances, des équivalents et des différentes occurrences et alternatives qu'offre les différents niveaux et emplois de la langue. En parlant de l'ambiguïté, Xavier Tannier voit que :

« Le problème qui se pose est que plusieurs catégories peuvent très souvent convenir à une même forme. En Français, les formes ambiguës sont estimées à environ 25 % du lexique, voire plus pour les mots : Les plus courants » (TANNIER 2006).

Il existe selon les chercheurs plusieurs catégories et types d'ambiguïtés, celles-ci peuvent être classées justement en fonction de différents paramètres et niveaux d'analyse, autrement dit la catégorisation reste relative aux approches envisagées. On peut notamment trouver ces quatre types d'ambiguïté :

4.6.1- Ambiguïtés lexicales :

L'un des cas les plus pertinents est le cas de la polysémie : le vocabulaire anglais recense par exemple 31 sens différents pour le verbe « to make », 28 pour le verbe « to go » et 23 pour le nom « head ». Il est actuellement impossible pour un traducteur automatique, à savoir « Reverso » en traduisant, de repérer sur les 23

sens par exemple du mot « head » ; ou les 28 du verbe « to make » quel est celui que l'utilisateur veut employé .

Pour Yves Gentillome :

" ...Le monde et les cultures évoluent .Tous les jours apparaissent des concepts nouveaux .Grace aux glissements sémantiques, on se tire d'affaire, mais , ce faisant ; on laisse la porte ouverte à l'ambigüité L'ambigüité devient , dans cette perspective , comme un cadeau du ciel , qui permet à la langue de s'adapter à des situations nouvelles , et qui permet de parler de ce qui paraissait indicible avec le lexique disponible. , (Yves gentillome. 1996 .p 115) .

4.6. 2- Ambigüité sémantique :

Prenons l'exemple : Une femme a dit que ce sénateur était anormal.

Un francophone sait qu'anormal peut avoir plusieurs sens figurés, par contre, même si Reverso parvient à fournir cette traduction « A woman said that this senator was abnormal. » qui est sémantiquement et syntaxiquement correcte, il n'en demeure pas moins que le rédacteur peut entendre plusieurs autres sens d' « anormal » dans ce contexte. Ce qui revient à dire dans ce cas que le traducteur automatique ne parvient pas à traduire le sens connoté ou figuré , du moment ou il ne contient pas de données qui lui permettent d'atteindre cette objectif .

4.6.3-Ambigüités syntaxiques :

Il y a ambigüité syntaxique chaque fois que l'analyseur détecte plus d'une représentation grammaticalement correcte pour la même phrase, comme dans l'exemple

" Time flies ".

L'homme ne trouverait pas de difficulté ; pour interpréter et traduire cet énoncé par le temps s'envole ou passe trop vite, mais la machine (d'autres traducteurs automatiques) l'interpréterons et traduirons de deux façons :

Soit par : Le temps s'envole

Ou alors : chronométrer des mouches. « fly » étant mouche en anglais.

La traduction de Reverso à donné :

Le temps passe vite .

Un second exemple d'ambiguïté syntaxique :

Je suis un grand homme .

ci , " suis" peut être auxiliaire présent de être ou présent de suivre .Le traducteur automatique ne peut pas prendre en charge les deux cas de figure.

La traduction de Reverso à donné :

I am a great (tall) man. Alors que suivre en Anglais est traduit par “ Follow”, et on devrait obtenir également ;” I follow a great man.”

Notre objectif n'étant pas de s'attarder sur ce problème d'ambiguïté dans ce modeste travail ,mais d'ouvrir une parenthèse (les promotions et génération futures pourront développer davantage ce thème) sur l'un des obstacles que rencontre la Traduction Automatique à l'heure qu'il est , nous , concluons ce chapitre en soulignant que la Traduction Automatique est historiquement et concrètement la première application du traitement automatique du langage , qu'elle est dans la poursuite et à l'instar de toutes les applications du Traitement Automatique du Langage, le résultat de collaboration entre différents spécialistes, mais qu'après soixante ans d'existence, elle présente encore des lacunes et rencontre des problèmes à savoir l'ambiguïté qui se pose presque à tout les niveaux du langage. Elle pose quelques enjeux qu'elle partage avec le domaine de l'apprentissage des langues assisté par ordinateur, certains aspects de ce dernier ainsi que quelques uns de ses enjeux seront traités dans le dernier chapitre.

CHAPITRE V

Apprentissage et enseignement assisté par ordinateur

Logiciel d'apprentissage du Français

Avant d'entamer ce chapitre, et pour relever toutes sorte de confusion, nous tenons à rappeler que les logiciels d'apprentissage de la langue, appelés également « didacticiels »²⁸ font également partie des outils et applications du génie linguistique (cf. la liste thématique des applications § 4, chapitre I).

Cela peut s'expliquer par rapport à plusieurs paramètres entre autres ; la programmation informatique : tout logiciel est un programme informatique, un logiciel d'apprentissage d'une langue est un programme qui à été réalisé à partir de différentes données et règles linguistiques (lexicales, grammaticales, sémantiques,...etc.) destinés à des fins pédagogiques. Les fonctionnalités de la synthèse vocale ainsi que la reconnaissance de la parole étants intégrés à un nombre de plus en plus important de didacticiels, et d'interfaces et plates formes d'apprentissage en lignes²⁹ sont tous les deux des outils d'ingénierie linguistique permettant l'exercice et l'apprentissage orale de la langue , pour ce qui est de la prononciation les utilisateurs s'entraînent à répéter des séquences ou encore à interagir oralement avec l' interface par exemple en répondant à des questions , ou se présentant....etc.

Ces logiciels contribuent à l'instauration de ce qui est connu comme étant l'apprentissage et enseignement des langues assisté par ordinateur qui peut se faire sans être en ligne- uniquement avec un ordinateur qui n'est pas lié à la Toile³⁰-, ils

²⁸ Contraction des termes didactique et logiciel.

²⁹ En **informatique**, une interface est le dispositif permettant le **dialogue homme-machine**, c'est en d'autres termes l'ensemble des moyens qui constituent les logiciels inclus dans les systèmes d'exploitations .

³⁰ On interent .

peuvent également constituer les interfaces et plates formes ³¹ d'apprentissage et de formation en ligne.

Dans ce chapitre , nous présentons un logiciel d'apprentissage du français destiné aux premières tranches d'âge « Logitron », le testons afin de cibler et voir son fonctionnement et ses performances , quel avantages , quel plus , ou valeur ajoutée apporte- il ? Ensuite , et en essayant d'élargir un peu le champ d'étude et notamment en présentant des statistiques et expériences qui ont été effectuées ailleurs dans des pays ayant officiellement entrepris d'intégrer ces nouvelles technologies et techniques dans leur différents systèmes éducatifs et voir un peu quelles nouvelles voies , perspectives et enjeux pour le traitement (apprentissage, enseignement diffusion, production, et promotion) des langues en appui avec des moyens informatiques ..

Le choix de ce logiciel n'a pas été effectué en prenant en compte un très grand nombre de paramètres. Il n'est à ce stade qu'à titre indicatif et non récapitulatif. L'objectif étant d'en faire une introduction permettant d'attirer l'attention sur une réalité en pleine expansion, laissant ainsi le champ ouvert aux futurs collègues qui souhaiteraient développer davantage la question. Nous avons tout de même fait en sorte qu'il soit destiné aux plus jeunes –et parfois aux débutants d'un point de vue général- et qu'il soit gratuit et téléchargeable, autrement dit mis à la disposition des apprenants. Il en existe pas mal d'autres plus développés qui contiennent en plus des fonctionnalités que présente celui-ci, la possibilité de s'entraîner et d'apprendre oralement la langue grâce aux techniques de synthèse et de reconnaissance vocale (cf. la liste thématique détaillée en annexe). Cependant leur disponibilité est relative, ils sont parfois pas permis au téléchargement, pour des raisons principalement commerciales, et donc pas accessibles à la totalité du public d'apprenants, dans la mesure où ils sont parfois chers.

³¹ Une plateforme d'apprentissage ou de formation est un logiciel installé sur un serveur informatique. C'est autrement dit un site web auquel l'utilisateur peut se connecter et accéder notamment aux différents cours et formations prévues par le formateur ou le concepteur du logiciel en question.

5.1-Présentation et exploration du logiciel :

Le logiciel en question se présente sous forme d'une interface constitué de cinq parties.

Sur les cinq parties qui constituent le logiciel, nous en présentons trois :

La première : « Des mots plein la tête » :

Contenu

Avant de commencer à l'utiliser, l'utilisateur ou l'apprenant est amené à procéder à un certain réglage d'options auxquelles il aura recours, comme les critères pour qu'un mot soit acquis, la durée de l'affichage des mots ... etc.

La partie « Des mots plein la tête », à l'image des cinq autres parties se présente sous forme d'une interface graphique constituée de plusieurs activités adapté au jeune âge des apprenants. Il contient quatre mille mots (la grande quantité de ressources et de données , (cf. Cerquiglini ,§5 du présent chapitre) répartis en 31 niveaux. L'apprentissage de la graphie du vocabulaire s'effectue à l'intérieur de dix activités différentes.

Quant à l'écriture des mots, l'élève doit soit utiliser le clavier ou cliquer sur des lettres présentées à l'écran à l'aide de la souris. C'est la première forme d'interactivité entre l'apprenant et l'ordinateur, toujours présente. Chaque mot réussi lors de sa première présentation est considéré comme maîtrisé et ne sera plus demandé. Quand un terme est mal orthographié, le logiciel donne un message qui présente la graphie exacte. Ce mot manqué devra être réussi deux ou trois fois de suite dans une autre activité pour être versé dans la liste des termes acquis, (l'aspect plus ou moins ludique) .Plusieurs paramètres permettent d'adapter le logiciel aux besoins particuliers des apprenants ,c'est l'adaptabilité au niveau de l'apprenant.

L'apprenant a la possibilité de visualiser le cheminement de son évolution (progression méthodique). Ses résultats peuvent être conservés, c'est-à-dire sauvegardé et imprimés .Un gestionnaire permet de limiter l'accès aux menus, de

gérer les inscriptions – l'apprenant doit s'inscrire et s'identifier avant chaque utilisation- et de modifier les paramètres³² du logiciel.

Objectifs :

Ce module vise à mémoriser la graphie des mots les plus fréquemment utilisés. Il développe aussi la compréhension de ces termes en les présentant dans un contexte significatif , à travers un certain nombre d'activités a caractère interactive

Public cible :

Les jeunes de 6 à 13 ans

Environnement informatique requis :

Ce logiciel fonctionne sous Windows 95, 98, XP, Vista et Windows7 (32 et 64 bits).

La seconde partie : « Lecto »

Contenu :

Egalement, à la première utilisation, l'utilisateur est amené à effectuer un certain nombre de réglage d'options en rapport avec le contenu du logiciel.

Pour ce qui est du module Lecto-mots, il présente 495 leçons regroupées en neuf groupes d'activités. Les trois premières activités développent principalement les habiletés préalables à la lecture : la discrimination visuelle, la concentration, la mémoire visuelle, et la mémoire des séquences (en ayant justement recours à des jeux de séquences) .Les cinq autres groupes d'activités améliorent l'habileté du lexique en développant la lecture globale des mots, puis des groupes de mots .

Le module Lecto-livres interroge l'élève sur ses lectures (l'aspect interactif toujours présent), et lui donne un certificat confirmant le nombre de livres qu'il a lus (motivation). L'enseignant peut gérer facilement tous les questionnaires créés pour le logiciel.

L'apprenant peut visualiser l'évaluation de toutes les leçons (apprendre à s'auto-corriger et à s'évaluer). Ce bilan indique si les exercices terminés répondent au

³² Dans ce cas les éléments et options de réglage.

critère de réussite fixé par l'enseignant. Les résultats de chaque utilisateur peuvent être conservés et imprimés.

Objectifs :

Ce logiciel contient deux modules. Le premier permet d'augmenter la vitesse de lecture et la compréhension, tandis que le second vise à encourager les élèves à lire des livres de la bibliothèque et à long terme toutes lectures

Public cible :

Les élèves du primaire et tout jeune ou adulte désirant améliorer sa vitesse de lecture.

Environnement informatique requis :

Ce logiciel fonctionne sous Windows 95, 98, XP, Vista et Windows7 (32 et 64 bits).

La troisième : « J'écris »

Contenu :

Même procédure pour le réglage d'options .En même temps module du didacticiel, et éditeur de texte. Cet éditeur de texte indique immédiatement après l'écriture de chaque mot s'il est bien orthographié et s'il respecte les règles grammaticales déjà apprises (efficacité de la méthode). Une erreur d'orthographe d'usage sera indiquée par un trait jaune et une faute grammaticale sera soulignée en rouge. En cliquant sur un mot souligné, une aide appropriée est fournie à l'utilisateur afin qu'il puisse se corriger. (Autocorrection et autonomie, cf. Gerbault. 2002)

Objectifs :

Ce logiciel permet à l'utilisateur d'améliorer sa maîtrise de l'orthographe d'usage et grammaticale à partir de ses productions écrites. Pour réaliser cet objectif, il souligne les fautes et peut illustrer les mots écrits correctement.

Public cible :

Les jeunes de 6 à 16 ans

Environnement informatique requis :

Ce logiciel fonctionne sous Windows 95, 98, XP, Vista et Windows7 (32

5.2 - L'informatique, le numérique, le traitement automatique au service de l'acquisition, apprentissage, diffusion et production des langues :

Nouvelles voies et perspectives :

Les nouvelles technologies appliquées aux différents domaines, l'informatique, la programmation notamment, les logiciels (didacticiels) qui en résultent, les interfaces, la toile, plates formes en ligne, les réseaux, les divers moyens et instruments informatiques ouvrent de nouvelles perspectives, défis et voies pour les différents traitements : apprentissage, acquisition, production, diffusion des langues. Il semble selon bon nombre de chercheurs et de

spécialistes (Cohen, Kern, Diki-Kidiri, Bibeau, Jocelyn Pierre et bien d'autres) que l'avenir, et la « survie » même et le développement des langues se jouent considérablement dans leur capacité à investir des supports et espaces technologiques et notamment informatiques, il y a par conséquent de sérieuses motivations et de réels enjeux à s'interroger sur les actions en question. L'informatisation de l'apprentissage et des différents traitements des langues atteint un tel degré d'importance qu'il est judicieux, voire obligatoire de la considérer et la traiter sous différents angles et lectures parallèles

5.2.1 Les prémices « en France »³³ : à partir de 1976 :

Nous commencerons par le président de la république Valéry Giscard d'Estaing qui en 1976 mandate S. Nora, inspecteur général des finances, afin de rédiger un rapport devant « faire progresser la réflexion sur les moyens de conduire l'informatisation de la société ». L'informatisation à ce niveau, comprend tous les aspects de la vie et notamment les aspects scientifiques et pédagogiques.

³³Ce développement s'est manifesté par exemple aux Etats-Unis bien avant (cf. Chapitre I et IV)

Le fait que se soit le président en personne qui ait pris cette initiative déjà depuis 1976 permet de se faire une idée sur l'impact et le caractère sensible de ces différentes opérations et procédures d'informatisation et automatisations. Il savait par une vision anticipée, qui se concrétise de plus en plus de nos jours que la survie des savoirs, entre autres la langue allait dépendre de leur degré d'investissement des différentes technologies. Déjà à l'époque, la compétition était déclarée et féroce entre les grands pôles universaux et nations durant la guerre froide ³⁴ pour le développement de l'outil informatique.

Nous rappelons juste que les moyens informatiques dont il s'agit peuvent être différents logiciels d'ingénierie linguistique: des didacticiels, interfaces des dictionnaires, des bases de données terminologiques électroniques, des interfaces et plates formes numériques, par la synthèse vocale et la reconnaissance de la parole; des aides à la compréhension orale et à la prononciation., des outils de production et de manipulation de texte (ce qu'on appelle plus communément les logiciels d'aide à la rédaction ou traitement de texte),des ressources authentiques sur la toile, à des cours d'autoformation en ligne, la communication en ligne à distance entre apprenants avec des natifs de la langues, des francophones ou des même des publics non francophone.

5 .2.2-L'après guerre froide, les années 80 , une nouvelle interaction :

En parlant du rôle de l'ordinateur dans l'apprentissage, Rachel Cohen ; psychanalyste, enseignante, et auteur de nombreux articles interdisciplinaires et notamment concernant l'apprentissage des langues par le recours aux didacticiels; dans *Les jeunes enfants ,la découverte de l'écrit et l'ordinateur* (1987) et *Quand l'ordinateur parle.*(1992) ; fait état de travaux de recherches et applications autour de l'ordinateur, du didacticiel et la synthèse vocale au service de l'apprentissage de la langue écrite et orale pour des publics notamment non francophones -c'est d'ailleurs à ce sujet que les deux ouvrages sont davantage intéressants, le fait de s'attaquer à des publics d'apprenants non francophone -des les années 1980.

³⁴ Les outils informatiques et notamment ceux consacrés au traitement automatique du langage étaient considérés durant la guerre froide, comme un moyen stratégique au même titre que le reste des moyens et politiques.

Nous constaterons comment l'ordinateur à travers les différents moyens (didacticiels,) et possibilités qu'il offre , par exemple l'interactivité ,l'enfant ou l'apprenant utilise le clavier et donc interagit d'une manière active dans un premier temps car il est tout le temps en train de manipuler soit le clavier, soit la souris, ensuite à travers les système de question-réponse³⁵ ; il pose une question et reçoit au retour une réponse , ou inversement, les exercices du logiciel contiennent des questions, auquel l'apprenant répond ,ces activités créent action, échange, et par conséquent motivation .

Les fonctionnalités, exercices , interfaces , et illustrations qu'on retrouve dans les logiciels ;dans l'exemple que nous avons présenté, les parties qui constituent le logiciel logitron, ont des interfaces interactives ,qui contiennent des illustrations et des exercices adaptées au jeune âge des apprenants, des interfaces qui ont été mise en place justement par des équipes interdisciplinaires de spécialistes(pédagogues, didacticiens, informaticien, talistes, cognitiens...), plus encore la possibilité de régler les paramètres et options du logiciel avant son utilisation,-ces paramètres qui sont généralement en relation avec le niveau de difficulté s'apprentissage souhaité- comme par exemple les critères pour qu'un mot soit acquis (cf. retour au logiciel en question) .

La synthèse vocale dans le cas des logiciels qui l'intègre (par exemple la gamme du logiciel, j'apprend à lire et à écrire , cf. retour aux sites du logiciel en question) qui est l'une de ces possibilités les plus remarquables , joue un rôle très important en ouvrant d'une part la voie à de nouvelles formes de pratiques au quotidien par l'ouverture de la voie à l'exercice et à la pratique orale de la langue dans différentes situations .Ce qui constitue généralement et une fois de plus une solution motivante pour les apprenants et notamment dans le cas où il ont accès à des logiciels leur permettant d'abord d'entendre les prononciations correctes par exemple de mots ou de toutes autres constructions linguistiques , ensuite de s'entendre soi même en la prononçant , donc pratiquer l'exercice orale de langue. Ce qui permet d'établir une comparaison entre les deux formes vocales, et par la suite apprendre à s'auto-corriger.

³⁵ Question-réponse , une des formes du dialogue Homme-machine .

5.2.3- Les années 90 , Flexibilité et adaptabilité :

Dans son article « La communication médiatisée par ordinateur en langue : recherches et application récentes aux USA », Kern , professeur associé en français à l'université de Californie, et directeur du Berkeley Language Center .Ses recherches portent sur la lecture, l'écriture, et l'acquisition des langues étrangères ainsi que sur l'apprentissage médiatisées par les nouvelles technologies , ne situe qu'au début des années 1990 aux Etats-Unis les premiers projets et les premières recherches autour de la communication assistée par ordinateur au service des cours et de l'apprentissage des langues .

Nous entendons par cours tous les niveaux et spécialités linguistiques qui rentrent en jeu pour l'apprentissage de la langue. Ces premières études relevaient que la communication assistée par ordinateur - nous rappelons que la communication assistée par ordinateur représente toute forme d'échange écrite ou orale entre un utilisateur et un ordinateur, c'est une forme de dialogue H-M ³⁶-dans l'exemple de « Logitron » , elle consiste en toute forme d'interactivité entre l'utilisateur et le logiciel (cf. retour au logiciel et aux trois parties présentés plus haut § 4.1) ; donnait plus d'occasions aux apprenants de s'exprimer, plus de temps pour concevoir les idées- puisqu' il n y a relativement pas de limites de temps ; il n'y pas de durées limites de séances comme est le cas dans les séances en classes . Nous pouvons constaté par exemple que dans le cas des logiciels d'apprentissage (Logitron ou tout autre didacticiel) , l'apprenant est libre de les utiliser à l'heure que bon lui semble , avec les fréquences qui l'arrange et les moments et durées d'utilisation qu'il souhaite ,donc une flexibilité qui était auparavant inaccessible , de plus grandes possibilités de précision de l'expression et de réalisation de bons résultats, puisque il y a un champ plus grand , d'exercice, et de pratique des activités langagières .

Il considère par ailleurs qu'elle favorise l'esprit de participation et d' « encouragement » puisque toutes les occasions et séances d'expression sont proportionnelles non seulement à la qualité et la nature du contexte d'utilisation, c'est-à-dire les fonctionnalités des logiciels,-l'utilisateur est amené à différentes reprises à participer activement à une activité donnée- , leur paramétrage (paramétrage de logitron) et

³⁶ Homme-machine

parfois les instruments et matériel utilisés, mais également à la fréquence d'utilisation et de recours aux moyens de communication et d'apprentissage assistée par ordinateur. Il est clair que plus le recours est fréquent, plus le temps d'exercice est long, et plus les possibilités de réaliser de meilleurs résultats sont importantes, et donc une certaine forme de réconfort et de satisfaction, ce qui mène et comme résultat logique, vers une diminution de l'anxiété liée à l'emploi des langues étrangères, d'où l'encouragement, la motivation et la progression.

5.2.4- A partir de l'an 2000, l'autonomie :

L'autonomie de l'apprenant ; est ce que met en exergue entre autre Jeanine Gerbault (2002), à partir de son travail d'enquête auprès d'institutions culturelles francophones dans le monde (519 répondants); note que les logiciels d'apprentissage et de formation (logitron, en l'occurrence) ou d'autres, les interfaces numériques, ainsi que les plates formes d'apprentissage en ligne (cf. retour aux différents didacticiels cités dans la liste thématique d'applications de l'annexe) la communication à distance, ou encore les espaces numériques de travail et d'apprentissage comme par exemple « Auralog », ou « BLCC », ou ce qui est connu sous le nom du « e-learning » ou « l'open and distance learning » en anglais, « encouragent l'autonomie des apprenants », puisque comme nous l'avons brièvement cité, l'apprenant à, entre autres, la possibilité de gérer son emploi de temps, et aussi plus de possibilités de choisir quelques modalités de son propre apprentissage, à savoir les contenus et activités, plus encore, les adapté à son niveau, comme dans la première utilisation de « Logitron »

Nous constaterons à travers ces statistiques que l'offre de formation s'adapte mieux à la demande avec le déploiement des moyens informatiques.-touts moyens confondus- Cette demande se manifeste travers la motivation des utilisateurs quant à ces nouveaux moyens et dispositifs informatiques, ainsi que le besoin enrichissant - pour ce qui est des portails ou des espaces numériques de travail en ligne qui comme nous l'avons précisé plus haut sont constitués entres autres de didacticiels- d'accès à la culture et de contacts avec d'autres apprenants qui utilisent les mêmes portails ou plates formes par exemple.

Sur l'ensemble des personnes interrogées, deux-tiers déclarent ce désir « d'autonomie », l'un des enjeux serait donc d'adapter ces techniques, possibilités disponibles et de les mettre au service de l'apprentissage et des apprenants, afin d'en tirer le meilleur profit.



Figure 8 : Environnement informatique d'apprentissage BLCC :
exemple d'une plate forme d'apprentissage
Différentes fonctionnalités, ressources et paramètre

5.2.5-L'accessibilité aux ressources :

Le déléguée générale à la langue française et aux langues de France ; linguiste et Professeur Cerquiglini (2004), souligne quant à lui l'accessibilité aux ressources en énorme quantité. C'est l'une des caractéristiques de ces nouveaux moyens de traitement automatique du langage, et des outils informatique d'un point de vue

générale. Il évoque le double avantage des outils et moyens informatiques et notamment ceux mis au service de l'apprentissage et des traitements de la langue (cf., « Logitron », ou toute autre site pouvant relever du domaine et être scientifiquement consacré à l'apprentissage).

D'abord leur efficacité concrète en terme d'accessibilité aux ressources : Cette concrète efficacité peut d'une part s'expliquer par rapport au nombre de ressources que l'on peut trouver ou nombre de leçons, il y a profusion de documentation électronique sur une seule et même question ou information. Dans le cas des sites que nous avons classés en annexe ou des plates formes d'apprentissage en ligne comme par exemple le centre de formation virtuel de la « FOAD » formation ouverte à distance (il y en a plusieurs autres) ou « Je vous ai compris » de « Neuro concept », ou le cas de « Logitron », le nombre des données , informations, leçons , cours et ressources disponibles et pouvant être consultées est concrètement considérable . (cf. Retour aux sites, plate forme et portails en question) .Les ressources, données et informations sont classées d'une manière hypertextuelle. dans l'exemple de Logitron , (chaque partie contient un grand nombre de leçons et d'exercices organisé hypertextuellement,) .

D'autre part, par la relative aisance et facilité d'accès à ces ressources (du moment où ils sont classés, répertoriés et présentés en très grand nombre, et d'une manière méthodique et hypertextuelle ; dans de très petits espaces sur les interfaces des sites en question , il suffit d'y cliquer pour pouvoir les consulter) ce qui permet de gagner en temps, et en efforts, un temps et des efforts qui pourront et devront être exploités et investis pour le développement d'autres performances. Cet aspect méthodique peut également apprendre la discipline en entraînant à la recherche méthodique et ciblée .

Le second avantage, est la possibilité de mise en commun des cours, à travers les logiciels (didacticiels), et plates forme en ligne, ainsi que les différents portails des espaces numériques de travail, comme par exemple la plate forme de l'EDAF .Une mise en commun qui pourra être employé pour des objectifs et raisons pédagogiques ou parfois empiriques précises, et qui était bien entendu impossible de réaliser il y a peine deux décennies.

Un troisième avantage plus symbolique, se poserait en terme de survie des langues nationales en jouant le « jeu de la modernité » par leur disponibilité et diffusion sur des supports informatiques en ligne ou hors ligne, et leur activité sur les réseaux, et les espaces numériques ce qui rejoint un peu l'idée des enjeux stratégiques et décisifs que nous allons évoquer dans le paragraphe suivant .

5.2.6- Une deuxième lecture du progrès en question :

Un aspect encore une fois plutôt pragmatique , voire stratégique , avec des connotations politiques de la problématique du développement des moyens informatiques de traitement et apprentissage et la promotion des langues , pour qui considère la question d'un œil critique ; le fait que ce développement soit justement indissociable du rayonnement et de l'influence des langues vivantes .Les travaux de recensement des locuteurs de chaque langue sur internet , quels que soient les problèmes méthodologiques ou techniques liés à ce type d'études et de statistiques , convergent tous vers le même constat : sur les six mille langues environ du monde , seule une douzaine d'entre elles sont vraiment présentes sur la Toile - avec 98 % des pages web selon les statistiques effectuées par (Paolillo,Unesco 2005)

Ce constat ne passe pas inaperçu.Le caractère polyglotte d'internet est donc très relatif, marquant des écarts flagrants. L'état de fait en question pourra être interprété en prenant en compte certains paramètres et critères précis et intelligents, et en étant placé dans l'un de ses véritables contextes, à savoir celui de la mondialisation,(certains spécialistes préfèrent employé globalisation) , car la présence d'une langue sur le réseau renvoie davantage à sa position dans la hiérarchie des langues des mondes, en incitant ainsi de plus en plus de personnes à l'apprendre .La question du politique est donc au cœur du développement des langues « par et pour » les réseaux informatiques .

Ce sont en partie les mêmes motivations qui ont incité les grandes nations, et notamment durant la guerre froide, et jusqu'à présent à déclarer dans un premier temps une compétition d'abord pour le développement et le progrès technologiques dans tous les domaines et davantage l'informatique. Et par la suite la compétition pour

le développement de ces outils d'informatique linguistique ou (traitement automatique du langage) .

5.2.7-Le pour et le contre :

D'un point de vue plus général , qui a surtout peser le pour et le contre de ces nouvelles techniques et technologies, le docteur et chargée de recherche au CNRS Diki-Kidiri Marcel (2007) constate que tous ces moyes informatiques et numériques (didiacticiels , logitron) et plates forme en ligne (cf. Annexe) espace numérique de travail, ou encore les réseaux sociaux ou de spécialistes ont un rôle capital et décisif dans « les transformations linguistiques » .

Il manifeste par contre en plus deux de leurs potentialités contradictoires : ils peuvent être un véhicule important de communication entre et à l'intérieur des communautés linguistiques, puisqu'elles facilitent et accélèrent la communication par exemple, et notamment sur la toile et les réseaux, par exemple les portails éducatifs en ligne ou les réseaux sociaux de spécialistes à savoir les divers « Listes de diffusion » ou « Relais » à titre d'exemple le « RISC » (relais d'information pour les sciences cognitives) qui est , la « Linguiste liste » « lien et facilitent les échanges de connaissances entres autres linguistique et culturels, ,pour ceux qui en ont accès, c'est ce qu'on peut constaté de plus en plus dans les milieu des jeunes apprenants, un peu partout dans d le monde. Comme ils pourraient être un facteur aggravant de la marginalisation des langues, dans le cas des communautés linguistiques qui n'ont pas encore accès à ces technologies ou encore des apprenants qui ne savent pas manipuler ces outils informatiques .C'est le cas de certaines communautés en Afrique, Asie, Amérique latine et même Europe, mais beaucoup plus l'Europe de l'est, ainsi que la Russie.

5.2.8-Cependant, le rôle de l'enseignant ne s'estompe pas :

Bien entendu, les capacités de l'outil informatique , si toute fois sont performantes et innovantes, c'est à l'origine l'Homme qui les mis en place le premier, dans la poursuite de ce constat et dans son article « L'apprentissage/enseignement des

langues à l'ère du numérique : tendances récentes et défis », Piet Desmet, traite la question de l'apprentissage et enseignement des langues assisté par les nouvelles technologies de plusieurs angles, ce que nous voulons très exactement souligner, c'est le fait que l'épanouissement et la montée en percé de toutes ces technologies et nouvelles voies d'apprentissage des langues ne remplacera aucunement le rôle de l'enseignant, dans ce sens :

« Il serait bien évidemment aberrant de penser un seul instant que l'ordinateur puisse remplacer l'enseignant, idée qui refait surface régulièrement. Bien au contraire, l'utilisation croissante d'ordinateurs dans l'enseignement, ne fera qu'augmenter la demande en matière de personnel enseignant. Plus que jamais, l'enseignement des langues a besoin de professeurs enthousiastes pour encadrer et assurer le suivi des tâches faisant appel aux TIC. »
(Piet Desmet,)

S'il fallait une fois de plus expliquer pourquoi, on se contentera rien que du fait que l'usage justement des Technologies en question ; nécessite au préalable des acquis fondamentales dans le domaine informatique, ces acquis sont la plus part du temps transmis par des enseignants (humains) spécialistes.

En conclusion au dernier chapitre, nous constatons une certaine convergence entre les différents spécialistes sur ce que pourrait apporter l'informatique pour les multiples traitements des langues (apprentissage, diffusion et promotion) à travers ce domaine. qu'est le traitement automatique des langues) .

Pour ce qui est de l'apprentissage, avec l'intervention de l'outil informatique, il est devenu davantage interactif, souple, et plus adaptable aux états et niveaux des apprenants. Il s'est également enrichi à travers l'accessibilité et la profusion du nombre de ressources pédagogiques- toutes catégories confondues -.ainsi que de la possibilité de mise en commun des différents contenus pédagogique.

La diffusion, l'expansion et la promotion des langues ont à leur tour largement bénéficié du numérique, de l'informatique, de la Toile et de toutes les technologies qui contribuent à l'instauration de cette discipline.

Ce progrès peut tout de même avoir des points négatifs à savoir la marginalisation, mais il pose un défi et un enjeu de taille ; en rapport avec la dominance dans la hiérarchie des langues à l'échelle internationale.

Enfin, ce progrès technologique n' s'épanouit pas sans l'intervention et l'accompagnement de l'enseignant, ce qui revient à dire qu'il y ou qu'il devrait y avoir complémentarité entre l'Homme et l'outil technologique.

Conclusion

Force est de constater, au terme de modeste travail que sciences du langage et informatique entretiennent d'étroits rapports caractérisés par une permanente interaction, à caractère interdisciplinaire et qui donne naissance à une troisième discipline, l'ingénierie linguistique.

Des liens justifiés d'une part par un certain nombre de points communs existants entre ces deux domaines, à commencer par les travaux de Chomsky et notamment sa hiérarchie issue de la grammaire générative qui a permis de mettre en place ; une fois informatisée, les assises de la théorie des langages formels ; branche des mathématiques, logique et informatique dans laquelle sont étudiés les langages artificiels ainsi que leurs propriétés, comme les langages de programmation informatique. Une nouvelle aire est née annonçant les débuts du traitement automatique du langage.

A travers l'entrée en matière, et en ayant présenté les différentes appellations du domaine, ainsi que la liste thématique des applications et famille des outils linguistiques du traitement automatique de l'information, nous avons montré que le traitement automatique est une discipline interdisciplinaire faisant appel à plusieurs autres disciplines et celui-ci peut être exploité par différents domaines et qu'il reprend les mêmes étapes d'analyse humaine du langage, mais d'une manière informatisée.

Notre sélection des applications et outils d'ingénierie linguistique étudiés est venue dans la poursuite de l'interactivité qui a fait naître cette nouvelle discipline.

En ayant commencé par l'un des outils de traitement automatique les plus utilisés et notamment par le public d'apprenants, nous avons attiré l'attention sur une situation alarmante, présentant le nouveau rapport à l'erreur qu'engendre l'emploi grandissant des correcticiels qui ne sont pas toujours performant et de voir comment il est possible d'éviter ce genre de dépendance négative à ces logiciels de telle sorte à bien garder à l'esprit qu'il ne sont qu'un moyen constituant une aide à la rédaction et non une fin en soi.

L'apport des correcticiels actuels dans la pédagogie de la langue française est minime, et leur utilisation doit être bien réfléchi, sous peine de conduire à une aggravation des erreurs, surtout dans le cas des élèves qui connaissent de réelles difficultés en orthographe ou alors en syntaxe, d'où cette comparaison et emploi du terme « dépendance ». Pour cette raison, il est essentiel que les élève ou toute autre apprenant aient pu en mesurer quelques limites, et en comprendre le mécanisme du fonctionnement.

Il ne demeure pas moins que ces derniers pourraient s'avérer très utiles dans la mesure où comme nous l'avons cité il sera nécessaire avant tout recours à ces logiciels d'adopter une stratégie d'emploi et des plans ou/et méthodes pédagogique précises. Cette dernière pourrait être mise en place par les enseignants ou toute autre spécialiste.

Le fait d'avoir mis en exergue le caractère interdisciplinaire de l'ingénierie linguistique en ayant expliqué pourquoi et comment est-elle pluridisciplinaire et en montrant quels étaient les résultats de cette interactivité scientifique interdisciplinaire, nous a permis de montrer quel point il est intéressant et important de créer un dialogue entre les différentes disciplines.

Dans le cas du traitement automatique du langage ; ce dialogue et cette interactivité se justifie par des points en communs existants entre l'informatique et la linguistique, auxquelles se joint l'analyse des textes pour aboutir à l'analyse automatique des textes et de corpus via les techniques de statistique et statistique lexicale .

Les logiciels d'analyse automatique lexicale et documentaire Hyperbase et Tropes sont des exemple par excellence .

A l'heure qu'il est, les applications et outils du traitement automatique du langage, ne sont pas toujours , ni tout à fait performant , nous l'avons constaté à travers les correcteurs automatiques , mais également les logiciels de traductions automatique . A son tour domaine et application interdisciplinaire , à l'instar de toutes les autres et la plus anciennes des applications du TAL, la , rencontre tout comme la traduction humaine , des problèmes et lacunes. L'un des principaux problèmes est celui de l'ambigüité qui se pose sur les différents niveaux du langage.

Pour ce qui est de l'apprentissage, avec l'intervention de l'outil informatique, il est devenu davantage interactif, souple, et plus adaptable aux états et niveaux des apprenants. Il s'est également enrichi à travers l'accessibilité et la profusion du nombre de ressources pédagogiques- toutes catégories confondues-.ainsi que de la possibilité de mise en commun des différents contenus pédagogique.

La diffusion, l'expansion et la promotion des langues ont à leur tour largement bénéficié du numérique, de l'informatique, de la Toile et de toutes les technologies qui contribue à l'instauration de cette discipline.

Ce progrès peut tout de même avoir des points négatifs à savoir la marginalisation, mais il pose un défi et un enjeux de taille ; en rapport avec la dominance dans la hiérarchie des langues à l'échelle internationale .

Enfin, ce progrès technologique n' s'épanouit pas sans l'intervention et l'accompagnement de l'enseignant, ce qui revient à dire qu'il y ou qu'il devrait y avoir complémentarité entre l'Homme et l'outil technologique.

Le débat concernant les nouvelles technologies et notamment celles du traitement automatique du langage peut être élargit de plusieurs points de vue .Ce qu'il y a lieu de noter , est que le domaine en question à l'instar de tous les savoirs , en en perpétuel progression .

Cette progression est porteuse de nombreux défis certes, mais également de perspectives pouvant toutes êtres « en principe » au service du bien-être de l'humanité d'un point de vue général. Ce domaine à offert à l'homme des possibilités qui lui était auparavant inaccessibles, et a parmi une ouverture vers d'autres domaines plus vaste ;créant ainsi une interdisciplinarité enrichissante.

REFERENCES BIBLIOGRAPHIQUES

OUVRAGE :

BIBER Douglas. « *Variation across speech and writing* »³⁷ Cambridge University Press, 1988.

CHOMSKY Noam . « *Syntactic Structures* » .La Hay : Mouton ,1957 .

CHOMSKY Noam « *Aspect of the theory of Syntax* » Cambridge University. Press, 1965 .

CHOMSKY Noam « *Lectures on Government and Binding* » Dordrecht ,Foris, 1981

PIERRE Jocelyn . *La langue au cœur du numérique : les enjeux culturels des technologies de la langue* . [2007] [en ligne], Rapport présenté par et à la Délégation générale à la langue française et aux langues de France, l'url paraissait incorrect sous certaines versions du word. consulté le [04/10/2009] .

COHEN Rachel . *Les jeunes la découverte de l'écrit et l'ordinateur*. PUF, 1987.

COHEN Rachel . *Quand l'ordinateur parle* . Paris PUF ,1992.

CONDAMINE Anne *Sémantique et Corpus*, Paris Hermès Science , 2005

SABAH Gérard . *L'intelligence artificielle et le langage* .Paris ; Hermès, 1984.

HARRIS Z. « *Methods in structural Linguistics* ».University of Chicago Press ,1971.

³⁷ Les titres d'ouvrages entre guillemets sont ceux des ouvrages en Angalis .

JACOBSON Roman « Essais de linguistique générale : Les fondations du langage » Paris, Editions de Minuit, 1963.

ARTICLES SCIENTIFIQUES :

GLANTZ H.T. « On the recognition of information with a digital computer » in *Journal des ACM* . 1957. Vol 4 ,n °2 . P. 178-188 .

DOLL Frédéric et COULOMBE Claude « L'avenir des correcteurs grammaticaux : un point de vue industriel » in *Bulag* . 2004 . N ° 29 .

HENRI « une approche cognitive de la correction automatique des fautes de syntaxe » in *Bulag*. 2004. N ° 29 .P. 85-105.

HARRIS Z« *Coocurrence and Transformations on Linguistic Structure* » in *Langage* ,1957 . N 33 , P 283-340 .

GERBAULT. « Technologies de l'information et de la communication et diffusion du français : usages, représentations, politiques. » In *Alsic*, vol. 5 n° 2, 2002. p. 183-207 [en ligne] http://alsic.u-strasbg.fr/Num09/gerbault/alsic_n09-pra1.htm consulté le [10/05/2010]

KERN Richard. « La communication médiatisée par ordinateur en langue : recherches et application récentes aux USA » In *Le français dans le monde. Recherches et applications*, 2006. n° 40, P . 17-29 .

LAB Frédérique « La Traduction Automatique » in *Le Bulletin de l'EPI* ,1988 n° 52 P165-170

LEON Jacqueline «Le traitement automatique des langues » in *Histoire Épistémologie Langage* 2001. P. 3-6

BERTEN Fernand . « Correcteurs orthographiques et enseignement du français » [en ligne], URL <http://users.skynet.be/ameurant/francinfo/correcteur/correcteur.html> consulté le [15/5/2010]

GENTILHOMME Yves « L ambiguïté est-elle une tare ? » in *Bulag* 1996 n ° 21 .

DESMET Piet « L'apprentissage/enseignement des langues à l'ère du numérique : tendances récentes et défis », article accepté dans la revue française de linguistique appliquée

CARDOT Jerome « Introduction au traitement automatique de la langue naturelle » [en ligne] URL <http://geai.univ-brest.fr/~cardot/CMntic-ha/node52.html> consulté le [20/01/2010]

CERQUIGLINI Bernard « Les outils informatique » In *Les études françaises. Séminaire international*, ADPF 2004.

TABORY Robert. « *International Conférence of Computational Linguistics* »* 1961 , USA .

KOUASSI Roland Raoul « *La Traduction automatique* » [en ligne], URL http://www.ltml.ci/files/articles4/article_traduction_automatique.pdf consulté le du [12/04/2010]

BIBEAU Robert « Ils apprennent à lire à l'aide de l'ordinateur . L'intégration des TIC en classe de français » [en ligne] URL source : http://netia59.ac-lille.fr/ref/pedagogie/Robert_Bibeau/ecrit.html. Consulté le [25/04/2010]

ORGANISME ET CONSTITUTIONS :

PAOLILLO John & al. *Mesurer la diversité linguistique sur Internet*. UNESCO, 2005.

DIKI-KIDIRI Marcel. *Comment assurer la présence d'une langue dans le cyberspace ?*
UNESCO, 2007.

RAPPORTS DE RECHERCHE

TANNIER .X « Traitement automatique du langage naturel pour l'extraction et la recherche d'information ».Rapport de recherche , Saint-Etienne , France ,2006 .

GHENIMI Abdelwafi « Traduction Automatique » Mémoire de magistère, ILE , Oran ; 2000

TABET AOUL Zoulikha « Approche de l'écriture Quignardienne par l'analyse du discours et la démarche lexico-statistique » ILE , ORAN 2005.

DICTIONNAIRES :

Patrick Charaudeau et Dominique Maingueneau .Paris .*Dictionnaire d'Analyse du Discours* Seuil, 2002

Bibliographie supplémentaire (complémentaire) :

Nous avons eu recours durant ce travail a des sources bibliographiques qui nous pas directement fait office de références , c'est-à-dire dont nous avons pas puisé pour la rédaction du mémoire, mais qui nous généralement enrichi et donc indirectement contribué à l'élaboration de ce travail.

CHAPITRE II

Correcteur automatique / logiciel ou outil de correction automatique / correcticiel¹

Dans ce second chapitre ; nous tentons d'approcher notre premier outils d'étude, un des plus répandus et des plus utilisées des outils du TAL² un correcteur automatique intégré dans un logiciel de traitement de texte ; le Word .Cette partie du logiciel a été programmée pour prendre en charge entre autres les corrections orthographiques, grammaticales et stylistiques (cf. Voir la définition de l'outil en question dans la liste des familles d'outils linguistiques du traitement automatique de l'information,§5 chapitre I).

Nous commençons par une brève introduction au chapitre présentant le nouveau rapport à l'erreur qu'engendre l'emploi grandissant des correcticiels , ensuite présentons un aperçu et une typologie historique des correcteurs, donnons une idée sur le paysage actuel en matières de correction automatique de texte avant de passer à une concise description du logiciel en question .Nous l'expérimentons en lui soumettant une phrase à corriger, présentons l'analyse des résultats de différents angles ,nous verrons comment fonctionne le correcticiel, et comment procède-t-il pour corriger ,nous présenterons les typologies des erreurs détectées par les correcteurs , et nous nous attaquerons à un cas particulier de l'emploi de ces logiciels, un cas présentant des situations qui devrait remédier au mauvais et à la dépendance à l'emploi des correcticiels ,celui de leur utilisation dans l'enseignement ,pour la réorientation de leur emploi après quoi nous conclurons .

2.1 Correcteur automatique et nouveau mode d'écriture :

Les énormes progrès et développements technologiques et informatiques qui ont provoqué une accélération de la communication de l'information ont entraîné le développement de

¹ Contraction des termes correcteurs et logiciel .Ce sont les trois désignations faisant référence à cet outil. Nous aurons la possibilité de les employer toutes les trois durant ce travail.

² Traitement automatique du langage

nouvelles pratiques au quotidien dues au développement de nouveaux outils spécialisés (logiciels de traitement de textes, correcteurs orthographiques et grammaticaux, dictionnaires en ligne, outils d'aide à la traduction, les outils d'aide à la recherche d'information, la synthèse et la reconnaissance de la parole ...) et donc des modifications importantes du mode de vie et d'écriture, où il apparaît que la notion d'erreur se multiplie et trouve une dimension privilégiée, puisque l'une des fonctions les plus utilisées est celle des outils de correction et d'aide à la correction également appelées correcteurs automatiques ou correcticiels (ceux-ci peuvent être des correcteurs orthographiques, grammaticaux ou encore stylistiques à savoir Word, Word Perfect, Quark Xpress...)

Ces outils qui permettent de plus en plus aux utilisateurs d'avoir un travail prêt corrigé et facile sans grands efforts font en sorte que les utilisateurs prennent vraisemblablement goût à cet avantage qui les rend encore moins sensibles à la nécessité de développer leur capacités à rédiger des textes sans erreurs. Nous assistons à une nouvelle forme de « dépendance » : la dépendance aux correcticiels.

Conçus à la base pour des rédacteurs professionnels (bureautiques, secrétariats et administrations), ils ont plus rapidement été utilisés par un plus vaste public et notamment celui des apprenants pour des tâches variées allant de la correction rapide de texte à l'analyse fine de documents complexes et à l'utilisation comme ressources pédagogiques en classe scolaire dans certains pays.

Les étudiants principalement – rares sont les étudiants qui optent pour la préparation et surtout la rédaction de leur travaux d'une manière manuscrite (ne serait-ce que dans un premier temps, sachant qu'aujourd'hui le traitement de texte a tendance à devenir obligatoire pour la présentation des devoirs...) ou autonomes, c'est-à-dire personnelle – et professionnels sont passés de l'écriture manuelle à la frappe à la machine et à la sortie avec traitement de texte. Ils écrivent vite sans véritable soin – du moins dans un premier temps – aussi bien de l'orthographe que de la stylistique en ayant à l'esprit que chaque rectification sera prise en charge et facile à faire grâce à ces outils "magiques" conçus pour, c'est une nouvelle, exagérée et pathologique forme de dépendance à la correction automatique. Le temps de relecture est par conséquent de plus en plus réduit. Cela fait que ce nouveau mode

d'écriture entretient avec l'erreur un nouveau rapport alarmant: le risque d'erreur s'accroît ; et l'objectif premier des logiciels en question qui est d'aider l'utilisateur dans son intérêt et dans la perspective de réduction d'erreurs à l'air de se transformer au contraire ,puisque ces logiciels ne sont pas toujours performant et il n'incitent pour ainsi dire pas –sauf exception- l'apprenant à se perfectionner puisqu'il font le travail à sa place. C'est l'une des motivations majeurs nous ayant incité à commencer nos chapitre par cet outil .

2.2- Brève typologie historique des correcteurs automatique :

Avant de poursuivre ; nous présentons une brève typologie historique des correcteurs afin d'avoir une idée sur leur évolution:

Essayer de trouver des moyens ou des auxiliaires extérieurs aux différents rédacteurs pour les aider à corriger une partie ou complètement leurs erreurs relève d'une préoccupation ancienne qui date d'avant l'informatique à proprement dite :

2.1-La première génération, les mots isolés :

Les premiers programmes destinés à corriger des erreurs orthographiques ont été développés à partir de 1957 dans des mots isolées , nous citons le cas des mots figurants sur des listes , par exemple les noms mentionnés sur les listes de passagers des compagnies aérienne (Galntz,1957). Ils reposent sur le principe de la base de données qui est une liste de mots ordonnée a la quelle est comparé une liste non ordonnée alphabétiquement des mots du texte à corriger .La comparaison se fait entre des chaines de caractères : si le mot du texte n'est pas dans la liste de référence ,le correcteur signale une erreur .Ce type de correcteurs est aujourd'hui le fondement de la plupart des correcteurs intégrés aux logiciels de traitement de texte .

2.2- La deuxième génération, le mot et son contexte immédiat

Les correcteurs dit de deuxième génération représentent un progrès considérable sur les précédents ; dans la mesure où il font intervenir des grammaires locales- basées sur des dictionnaires de mots contenant des informations supplémentaires permettant d'étiqueter la catégorie grammaticale de chaque mot et de définir le genre et le nombre des noms, pronoms, des adjectifs et des déterminants ainsi que la ou les personnes pour les verbes.- qui analysent le

contexte immédiat du mot et permettent de déterminer par exemple les règles d'accord propres à la langue pour laquelle ils ont été élaborés .Ce type de correcteurs est utilisé dans les caractères grammaticaux intégrés dans les traitements de texte .

2.3-La troisième génération, la phrase complète :

Enfin les correcteurs dits de troisième génération fonctionnent avec une analyse grammaticale complète de phrases complexes. Cette génération a recours à des analyseurs morphosyntaxiques puissants. Le dictionnaire de mots est une base de données de référence comprenant des informations syntaxiques servant à valider la construction des mots. Ils sont maintenant intégrables dans plusieurs logiciels de traitement de texte et d'aide à la rédaction (messagerie), comme le Word ou le Wordperfect .On s'attend bientôt à une quatrième génération .

2.3 - Les outils existants :

Jusqu'à présent- 2010, les correcteurs peuvent être classés en deux catégories :

2.3.1- Les correcteurs automatiques intégrés dans les traitements de texte :

Comme leur nom l'indique, ils sont intégrés dans d'autres logiciels de traitement de texte. L'efficacité des correcteurs intégrés dans les traitements de texte est notamment limitée par le fait qu'ils ont été initialement développés pour la langue anglaise, dont les structures grammaticales sont différentes de celles de la langue française. Ils déçoivent parfois par leurs insuffisances syntaxiques et sémantiques.

Entre autres, les traitements de texte suivants disposent d'un correcteur intégré (la liste est incomplète : le but n'étant pas ici de recenser tous les traitements de textes...)

Dans l'environnement PC ³ : (de Microsoft)

Word, (de Microsoft),

Word Pro (de Lotus)

Wordperfect (de Corel).

Appleworks (anciennement ClarisWorks) (de Claris)

³ L'emploi d'environnement dans un contexte informatique fait référence entre autre et notamment dans le contexte des exemples en question, au système d'exploitation utilisé, dans ce cas Windows de Microsoft corporation

Dans l'environnement Mac⁴

ClarisWorks (de Claris)

Word, (de Microsoft),

Wordperfect (de Corel).

2.3.2- Les correcticiels indépendants

Indépendant, c'est-à-dire et à l'encontre des correcticiels intégrés, leur atout est qu'ils sont élaborés par des spécialistes de la langue française et que leurs capacités syntaxiques et sémantiques sont dès lors plus performantes, sans pour autant être complètement performants.

Pour information, il semble selon des statistiques que les produits canadiens sont, à l'heure actuelle, les plus performants. Voici des exemples :

Dans l'environnement PC :

Correcteur 101 (de Machina Sapiens)

Correcteur 101 didactique (de Machina Sapiens)

Cordial (de Synapse)

Antidote (de Druide informatique)

Dans l'environnement Mac

Hugo Plus (de Logidisque)

Prolexis (de Editions Diagonal).

⁴ Mac est le système d'exploitation de mac Os de l'entreprise « Apple » computer.

2.4 Application : le logiciel « Word »

2.4.1-Description du logiciel

Dans ce qui suit une description du logiciel a pour objectif de montrer quelles corrections il prend en charge :

Le logiciel Word – sous sa version 2007- est un programme de traitement de texte qui se présente sous forme d'interface graphique⁵ récapitulant un certain nombre de fonctionnalités. La partie qui nous intéresse est celle correspondante à l'onglet " Révision" .En y cliquant surgissent les aides et corrections suivantes :

Grammaire et orthographe, recherche dans le dictionnaire des synonymes et traduction.

La correction grammaticale et orthographique comprend :

- L'accord des noms, adjectifs et déterminant.
- L'accord du sujet et du verbe .
- Forme de l'adjectif attribut et du participe passé.
- Temps et mode des verbes.
- Négation,
- Article définis contractés.
- Harmonie phonétique (ce/cet)
- Comparatif/superlatif
- Emploi des pronoms
- Emploi des auxiliaires .-

⁵ (cf. voir définition chapitre 5)

La correction stylistique comprend

Les expressions archaïques et littéraires .

La longueur de la phrase.

La ponctuation.

2.4.2- Expérimentation : correction automatique

Nous présentons l'exemple suivant : « Les aprenan réviz les lesson », de l'orthographe correcte: « Les apprenants révisent les leçons ».

En saisissant cette phrase, l'analyseur automatique qui est dans ce cas le logiciel Word ; identifie les erreurs en les soulignant en rouge, et propose des alternatives de correction : nous classons ses propositions de correction pour chaque erreur dans un tableau et passerons ensuite à l'analyse :

Erreur	Proposition de correction
aprenan	Apprenant , apprenante , apprenants , arénant
réviz	Reviez , rémiz , revis , revif
lesson	Lessonie, besson, laissons , tessons

Tableau 1 : Exemple de correction automatique

CHAPITRE III

Ingénierie linguistique et interdisciplinarité

Le cas de l'analyse automatique des textes.

Ce chapitre est à double vocation :

Dans un premier temps, et avant de poursuivre notre étude de deux autres applications du traitement automatique du langage ; (quatrième et cinquième chapitre); nous ferons en sorte d'ouvrir une parenthèse sur un point sensible et pragmatique des approches de notre travail ; il s'agit de s'interroger sur un aspect profondément pluridisciplinaire et interdisciplinaire¹ de cette nouvelle discipline ; dans la perspective de savoir qu'est ce qui octroi à l'ingénierie linguistique ce statut de discipline interdisciplinaire - pourquoi est-elle interdisciplinaire et par rapport à quoi-? Comment est-elle interdisciplinaire ? Et quelle sont les résultats de cette interactivité scientifique (interdisciplinarité) ?

Il est important dans ce sens ; et pour montrer quelques dimensions importantes de ce caractère scientifique interdisciplinaire ;une fois avoir présenter la réponse aux questions ci-dessus ; de souligner un des plus conséquents résultats de cette interactivité qui existe entre linguistique et informatique. C'est ce qui nous mènera ,en second lieu -la conséquence ou le résultat étant le fait d'aboutir logiquement à de nouvelles spécialités et applications- à en sélectionner une pouvant être mise au service de différents domaines à savoir la linguistique et la littérature , il s'agit de l'analyse automatique des texte qui est donc le résultat ou l'aboutissement de la fusion et du dialogue entre l'analyse des textes et les méthodes informatique.

¹ Le terme interdisciplinaire peut être définie comme étant un adjectif relatif aux rapports entre plusieurs disciplines, il est le second palier d'une échelle scientifique de signification classant les termes : pluridisciplinaire, interdisciplinaire et transdisciplinaire .

Nous en ferons la présentation ; et tenterons de montrer en partie comment s'effectue toutes ces interactions, avant d'illustrer par deux cas pratiques, et concluons.

3.1- Interdisciplinarité et interactions :

3.1.1- Le pourquoi de l'interdisciplinarité : (qu' est ce qui fait que l' ingénierie linguistique est interdisciplinaire ?) :

L'un des constats pouvant être détecté par tout observateur et qui est considérablement intéressant de mentionner aussi historique qu'il puisse être est que l'aube des savoirs et des connaissances – toute disciplines confondues - s'est effectué avec la mise en place des premières alphabets et par la suite « écritures » en Egypte et en Mésopotamie . Qui dit alphabet, dit mot, et qui dit mot, dit langue, et qui dit langue dit ensuite communication et surtout écriture, et qui dit production écrite dit savoir et connaissances.

Autrement exprimé, le mot, les différentes constructions linguistiques de la plus petite à la plus grande, et la langue sont au cœur des savoirs. Ils sont le pivot et l'essence même de la science et des connaissances. Sans mot ni langage ; il n'y a point de savoir. C'est en quelque manière ce qui rappelle la dialectique des « langues de spécialités », qui permet de voir que la linguistique est finalement sollicitée et au retour peut être mise au service de toutes les disciplines. Puisque le langage et les langues sont le véhicule privilégié de l'information, toutes natures confondues :

« Dans la masse d'informations à laquelle l'homme est confronté, la forme la plus courante est bien entendu le langage, qu'il soit écrit ou parlé. Le langage possède une faculté d'adaptation unique, permettant de tout représenter, jusqu'au langage lui-même. » (Jerome Cardot , 2002)

Leur étude devient donc par analogie implicitement ou explicitement, un anneau commun à tous les savoirs et à toutes les disciplines ; d'où cette interdisciplinarité.

L'ingénierie linguistique ou la linguistique informatique puisqu'elle est considérée comme étant l'informatisation de certaines spécialités des sciences du langage, prend le relais de cet état de fait. Les différents outils et domaines du traitement automatique du langage- puisqu'ils reposent sur l'informatique qui à l'instar des mathématiques, sont des disciplines auxquelles toutes les autres ont recours- sont à leur tour interrogés par la quasi-totalité des disciplines et le nombre d'utilisateurs ayant recours à ces derniers est de plus en plus croissants.

En fait, si l'on considère la question d'un angle simple, il suffit de remarquer que toutes les applications du traitement automatique du langage, sont le résultat de la rencontre et du dialogue entre les spécialistes des différentes spécialités linguistiques et branches informatiques pour saisir la dimension interdisciplinaire du domaine en question.

C'est donc d'abord par un accord tacite, que ces interactions prennent forme . (scientifiques, administratifs, et littéraires) tous participent à un processus interdisciplinaire d'abord implicite en interagissant et en utilisant tous ces outils. Par exemple dans le cas des logiciels de traitement de texte - rencontre(mariage comme le considère certains spécialistes) de l'orthographe, la grammaire , la stylistique et de l'informatique-, ces outils sont utilisés à l'heure qu'il est - et notamment le Word- par des utilisateurs appartenant à différentes spécialités –premier niveau d'interdisciplinarité- d'une part .D'autres part ces mêmes utilisateurs contribuent indirectement, et parfois directement à l'amélioration de ces outils et notamment à travers différentes campagnes de statistiques et d'évaluation organisée par des institutions et des agences spécialisées en génie linguistique comme l'agence « Elda » .

C'est un cas à l'image de la double articulation de « Martinet », de double interdisciplinarité. C'est-à-dire qu'il y a : d'abord d'un point de vue conception de l'outil un premier niveau d'interdisciplinarité, le principe d'interaction et de collaboration entre spécialistes de différents domaines, ensuite un second niveau d'interdisciplinarité, l'utilisation de ces techniques par différents domaines.

Ce travail de collaboration entre différentes disciplines et groupes de recherche nous mène droit vers la réponse à la seconde question , celle de savoir comment le traitement automatique des langues est-il interdisciplinaire, par quelles procédures ?

3.1.2 – Le comment de l’interdisciplinarité (par quelles procédure le traitement automatique des langues est-il interdisciplinaire) :

D’abord par le fait que l’objet ainsi que la structure scientifique du domaine en question est le résultat de la fusion de deux domaines principaux qui sont les sciences du langage avec ses différentes spécialités et l’informatique avec également quelques une de ses branches et techniques .

Ensuite à travers l’ouverture vers d’autres domaines plus vaste à l’image de la statistique, la traduction , les sciences cognitives, d’une manière générale , les science de l’éducation , la recherche d’information, l’analyse des genre et des texteetc.

Et ce à travers ; en troisième lieu des efforts conjugués de spécialistes travaillant en collaboration ainsi que des opérations réciproques d’interrogation et des « aller- retours » effectués sans-cesse entre ces différentes disciplines .

Chacune des spécialités concernées apporte sa brique à cet édifice qu’est l’ingénierie linguistique en différents termes :

Il peut s’agir pour les science du langage de :

-Corpus oraux ou écrits,

-Différentes théories d’analyse .

-Règles grammaticales , orthographiques et de conjugaison

Toutes données , ressources et principes linguistiques par exemple phonétiques , phonologiques , morphologiques , ou encore sémantique morphosyntaxiques , pragmatique, discursives...

L'informatique intervient avec :

Les techniques et langages de programmation

Les algorithmes.

Les différents types de calcul.

Les bases de données.

L'intelligence artificielle et ses techniques.

Les mathématiques à leur tour interviennent avec :

Les différentes catégories de logique.

Les statistiques et leurs différentes techniques. .

Les méthodes probabilistes...

Les sciences de l'éducation avec leur :

Différentes méthodes et théories (didactiques et pédagogiques).

L'analyse des textes et la catégorisation des genres via

Les différentes approches et théories d'analyse.

3.1.3-Résultat (quelles sont les résultats de cette interdisciplinarité):

Après avoir essayé de montrer ce qui fait du traitement automatique des langues une discipline interdisciplinaire, et par quelles procédures il l'est, nous aboutissons logiquement en troisième lieu au résultat de cette interdisciplinarité : ce sont l'ensemble des nouvelles spécialités, outils et applications qui ont vu jour (cf. la liste thématique d'applications et les familles d'outils, chapitre I) dans un premier niveau, et dans un second, c'est leur utilisation et mise à disposition d'autres domaines. Comme nous l'avons cité plus haut, nous allons essayer de nous approcher de l'un des résultats en question, le premier cas d'interdisciplinarité que nous présentons étant celui de :

3.2-L'analyse automatique des textes et de corpus :

L'analyse automatique des textes est un domaine du traitement automatique du langage. Il est une application à caractère pluridisciplinaire - en fait, parce qu'il est l'aboutissement de la fusion de l'informatique et de l'analyse des textes, qui implique d'ailleurs un étroit travail de collaboration entre analystes et informaticiens, une collaboration justement que les données de la société dans laquelle nous vivons nécessitent, et tend à devenir obligatoire, pas uniquement d'un point de vue technique et scientifique, mais également d'un point de vue pragmatique, dans ce sens nous retiendrons le point de vue d'Anne Condamine dans son article « Discours, corpus, et traitement automatique » qui précise que :

« Le traitement de gros corpus pour l'acquisition de données plus riches implique quant à lui la mise en œuvre de méthodes (semi-) automatiques ou informatiques qui ont l'avantage d'être reproductibles.

Ainsi union entre linguistique de corpus et outils informatiques est-elle scellée , contribuant au mouvement de fond vers un empirisme qui sous des formes plus ou moins radicale , peut être vu comme définitoire » (Anne Condamine, 2005 , 182)

Condamine, linguiste, enseignante et Directrice de Recherche , au sein de l'unité de recherche mixte de l'université de Toulouse, UMR 5263 ; « Cognition Langues Langage Ergonomie » ; équipe de Recherche en Syntaxe et Sémantique du centre nationale de recherche scientifique de France (le CNRS) et développant des projets de recherches dans le cadre de l'axe « Structures sémantiques : des catégories lexicales au discours » évoque l' « implication » entre traitement linguistique - qui devient informatisé- de corpus (le traitement à ce niveau sous-entend toutes les opérations linguistiques d'analyse , tout paliers et niveaux compris) et les multiples méthodes et techniques de traitement et d'analyse informatiques. .On parle d'informatique linguistique . Dans ses différents travaux de recherche ; Leech, quant à lui évoque précisément un nouveau domaine qu'il nomme « Computer Corpus Linguistic » qui serait traduit en linguistique computationnelle de corpus: .Une implication qui devient de plus en plus importante, voire obligatoire, car elle est l'essence même de cette discipline qu'est le traitement automatique du langage (cf. .voir le caractère interdisciplinaire §1 de ce chapitre) d'une manière générale , et de la catégorisation et de l'analyse automatique des textes précisément .

L'une des raisons pour lesquelles la mise en œuvre de ces méthodes (semi-) ou complètement automatiques et informatiques ;est aussi intéressante et avantageuse aussi bien pour les littéraires et linguistes que pour les développeurs des méthodes informatiques concernées (en l'occurrence les algorithmes et logiciels d'analyse informatique comme le logiciel Hyperbase qu'on va présenter plus loin) ,pour le traitement d'un nombre de plus en plus croissant de corpus, de données et de contenus , est le fait que ces méthodes et techniques sont concrètement très pratiques et opérationnelles -c'est l'une des raisons d'ailleurs par analogie qui a fait que le nombre de disciplines et d'utilisateurs ayant recours à l'informatique est de plus en plus considérable- l'aspect

et le caractère opérationnel des différentes spécialités et techniques informatiques et notamment de programmation des multiples logiciels ; toutes spécialités confondues a permis en fait au genre humain d'obtenir des résultats , de gagner en temps et en moyens, plus encore ; d'atteindre un certain niveau d'aisance et de précision que jamais les humains auraient pu atteindre sans.

L'une des manifestations de cette opérationnalité ; est la capacité de ces méthodes et techniques à être facilement et rapidement reproduites et reproductibles ,on entend par reproductibles ici le fait de pouvoir appliquer à répétition et surtout dans des espaces de temps réduits ; les mêmes techniques et analyses a de très grand nombre de corpus (toutes natures, catégories et supports compris) ce qui logiquement permet de gagner en temps et en moyens lorsque le nombre de documents ou de textes à analyser est considérables . Quant à la qualité des résultats obtenus, elle s'améliore en fur et mesure que les méthodes utilisées se développent.

Ce mouvement de performance étant en perpétuelle évolution et c'est principalement par rapport au temps qu'on peut le caractériser. Autrement dit, plus le temps passe, plus les techniques et les approches deviennent de plus en plus performantes, c'est ce qui explique par ailleurs les différentes générations et versions d'un même logiciel, système, programme ou outil informatique, des versions éditées chaque années ou chaque deux jusqu'à quatre ans, par exemple, « Word », ou encore « Windows » qui présente à chaque reprise des nouvelles versions.

Dans la poursuite de cette implication , et en la reprenant ; avec des dimensions plus universelles et plus globalisantes , on pourra se rendre compte qu' il n'est pas uniquement question d'aspect pluridisciplinaire, ou encore, interdisciplinaire caractérisant les domaines concernés, mais d'un lien plus important qu'est celui de l'union .Une union entre, linguistique et informatique dans un premier temps, et une fois de plus entre sciences du langage, littérature et informatique . La fusion et l'union de ces disciplines donne naissance à une toute nouvelle informatique, appelé désormais « Informatique des Sciences Humaines » , une discipline qui avant toute autre considération , -et c'est là une réalité tangible à partir des nouvelles pratiques qui naissent au quotidien- vient comme suite logique et inévitable aux progrès technologiques, aux profondes mutations que connaissent les sociétés et toutes leurs structures, et aux besoins qui lui sont parallèles.

L'union (ou le mariage, comme le considère certains spécialistes) que nous évoquons se justifie notamment par tous les efforts de travail collectif et interactif et collaboratif qui unie un nombre de plus en plus important de spécialistes et d'équipes de recherches qui tendent à devenir de plus en plus mixtes ; comme celles dont est constituées par exemple les équipes de recherche du Centre Nationale de Recherche Scientifique (CNRS) et l'Institut National de Recherche en Informatique et en Automatique de France (INRIA) , ou encore des différents laboratoires spécialisés en la matière à l'échelle universelle .

Elle s'explique également – et c'est d'ailleurs d'un point de vue épistémologique ce qui a fait que le rapprochement et les liens s'effectuent entre informatique et linguistique d'une manière aussi étroite - par les points en commun qui existent entre ces deux domaines. D'abord les Sciences du langage et l'informatique utilisent le langage comme moyen de communication (naturelles pour les langues, ,les langages de programmation pour l'informatique). Ce même langage ; qui est comme nous l'avons cité plus haut le détonateur de l'existence des disciplines de savoirs.

Ensuite, elles se basent sur le traitement de l'information, d'où le Traitement automatique des langues- et entrent en jeu quand celle-ci a déjà été mise en place. Elles ont entre autre comme objectif d'étudier et d'analyser comment différentes entités communiquent entre elles en ayant recours à plusieurs niveau d'étude du langage(soit naturel ou informatique de programmation) structurés., d'où l'appellation nouvelles technologies d'analyse de l'information.

Un troisième élément important est que les deux disciplines en question se rapprochent également par leur méthodologie d'étude et de recherche qui pourrait se caractérisée par ces mots d'ordre qui représentent sommairement le principe du traitement automatique du langage et qui sont : délimiter/chercher , analyser/traiter , classer . Les linguistes cernent l'information, une fois délimitée, ils effectuent un certain nombre d'opérations analytiques (qui diffèrent selon la spécialité et le niveau ou l'angle d'analyse envisagé), avant de les classer par exemple dans des ouvrages et ressources . La machine également permet de rechercher , mais aussi reçoit des informations en formes d'instructions , données ou des éléments programmables , les analyse et les traite - en les transformant en un langage qu'elle comprend ,c'est-à-dire des séquences numériques ou des langage de programmation- ensuite les classent de différentes manières à savoir en base de donnée afin qu'elles puissent être consultables à chaque fois qu'un utilisateur en a besoin.

Toutes ces opérations foisonnantes de rapprochement et d'interactions ont des objectifs et des résultats importants, voire capitales, elles permettent par exemple de mieux comprendre les mécanismes du langage, de l'écriture ; de faciliter leur compréhension, acquisition et exploitation. Sans oublier le fait qu'elles permettent de gagner énormément de temps.

Cette analyse automatique ou informatique de textes et de corpus², a recours aux méthodes statistiques pour effectuer ces études linguistiques. C'est ce qu'on appelle l'analyse, les méthodes ou encore la linguistique statistique. Les analyses peuvent s'effectuer sur un ou plusieurs contenus ou textes, appartenant à différents domaines composant ainsi un corpus textuel.

3.3- Linguistique statistique :

Connue sous plusieurs noms à savoir, la linguistique statistique, la statistique lexicale, ou encore la linguistique quantitative, Charaudeau & Maingueneau la définissent comme suit : « La statistique lexicale qui se veut exhaustive, systématique et /ou automatisée a pour but le comptage des unités lexicale d'un texte ou d'un corpus » (Charaudeau & Maingueneau)

Elle a ; en d'autres termes ; comme but général d'identifier les structures mathématico-statistiques propres au langage, considérées en tant que phénomène objectif. Les modèles et résultats computationaux, quantitatifs, numérisés ou encore informatisés finissent par déboucher sur une linguistique nommée « la linguistique de corpus » qui fonde ses enquêtes et ses traitements sur des recueils de données linguistiques de différentes typologies au format électronique ou numérique. Et c'est le trait pertinent et ce qui caractérise selon les chercheurs à savoir Geoffrey Williams (université de Bretagne sud) cette nouvelle linguistique.

Cette linguistique fondée sur l'étude, le traitement et l'exploitation des textes au format numérique et informatique conçoit et définit le corpus linguistique justement comme étant « une collection de textes informatisés ».

Il faut donc créer un corpus informatisé, avant de procéder concrètement à l'analyse textuelle, statistique, lexicale ou documentaire automatique voulue. La création des corpus informatisés est

² Corpus oraux également

considérée comme étant l'une des caractéristiques et l'un des plus importants résultats de la communication Homme-Machine(CHM). Elle est la phase intermédiaire entre le corpus à son état original et automatisé.

Nous pouvons repérer trois phases principales durant cette création :

1- Le choix des textes, leur typologies ainsi que différentes caractéristiques : généraux, spécialisées, de référence, archives et leur formats, ainsi que les critères d'analyse organisés éventuellement en grille d'analyse..

2-Une fois le choix précisée et la sélection effectué, vient la phase d'acquisition et de transformation de données qui nécessitent de rendre le texte en MRF (Machine readable form), c'est-à-dire le format que la machine peut déchiffrer. A travers un code précis (cf, le paragraphe suivants) . Cette phase pourra s'effectuer à travers la frappe (la saisi), le scanning, la dicté, dans le cadre des techniques de traitement, synthèse et reconnaissance vocale.

3-La troisième étape est celle de la codification des données : lors du passage de la forme écrite a la forme déchiffrable par l'ordinateur ,le texte perd quelques caractéristiques , raison pour laquelle la nécessité pour l'ordinateur d' avoir recours à l'utilisation de codes précis comme le ASCII(American Standard Code for Information Interchange) l'équivalent en français du « Code américain normalisé pour l'échange d'information » ,pour le traitement informatique de ces textes, se fait ressentir et s' avère nécessaire.

Ce code constitué de caractères alphanumérique est considéré comme étant la norme de codage de caractères en informatique la plus connue ,et la plus largement et relativement- aux changements et caractéristiques des langues- compatibles .Inventée en 1961 par Bob Berner , Elle est à la base de nombreuses autres normes nationales ou internationales qui l'étendent telles que « ISO/CEI 8859 » , « GB 18030 » ou encore « Unicode ».

Une fois le corpus créé et informatisé, l'utilisateur pourra à ce moment procéder à l'analyse proprement dite. A travers cette brève présentation du fonctionnement du logiciel Hyperbase, ainsi que les logiciels de la gamme d'Ascetic, nous allons montrer comment les données citées précédemment se conjuguent pratiquement afin d'aboutir à ces deux logiciels d'analyse automatiques statistiques et documentaires. :

3. 3.1-Le premier model : le logiciel d'analyse automatique Hyperbase :

Description, illustration et présentation du fonctionnement :

Le premier model que nous présentons est celui du logiciel d'analyse automatique statistique et lexicale Hyperbase.

Hyperbase est un logiciel pour le traitement documentaire et l'analyse statistique des corpus textuel. Conçu et programmé en Français , il s'applique cependant indifféremment à toute langue qui utilise l'alphabet latin .

Une fois installé, le logiciel crée dans le disque dur une base de donnée composée de textes de différentes typologies et différentes périodes, et en permet divers analyses .On peut par exemple traité, analyser et comparer différents textes ,et différentes bases de documents et l'interrogation s'enrichit et s'avantage des diverses fonctions et notamment hypertextuelles ³ du logiciel .Il suffit par exemple de cliquer sur un mot pour connaître sa répartition dans le texte de travail ou d'analyse et être conduit dans les passages où le mot est employé. Ces opérations peuvent se faire à partir de l'index (cf. voir exemple figure 4) .L'hypertextualité du logiciel se manifeste également dans la possibilité de se connecter à la Toile pour comparer les résultats ou approfondir les analyses .Le menu du sommaire est constitué horizontalement des fonctions documentaires, et verticalement des fonctions statistiques.

L'analyseur peut avoir recours à une base textuelle déjà créée dans le logiciel , comprenant des textes littéraires français comme « Zadig » ou « Emile » et étudier et expérimenter de cette

³ L'hypertexte désigne un mode d'organisation des documents textuels informatisés caractérisé par l'existence de liens dynamiques entre ses différentes sections menant vers d'autres sections. Dans le Web, des mots ou des expressions soulignés ou encore des images indiquent les liens sur lesquels on clique à l'aide de la souris .

manière les différentes fonctionnalités du logiciel ou alors créer sa propre base, c'est-à-dire son propre corpus d'analyse. Il existe aussi une option Vers pour la poésie. On peut également comparer le corpus traité à l'un des corpus littéraire de « Frantext » qui est une base de donnée lexicale et textuelle qui comprend 117 million de mots, et s'étend sur cinq siècles. On a la possibilité d'utiliser le corpus en entier, ou bien de se limiter à une tranche ou une période temporelle précise. Dans le cas d'une langue étrangère, un dictionnaire de référence et d'informations relatives au corpus propre à la langue choisie peut être installé en plus. Si un nouveau texte est introduit dans la base de données, le logiciel procède à la reconnaissance et à l'analyse des formes du texte, à l'indexation proprement dite, à la création d'un dictionnaire de fréquences alphabétique et hiérarchique permettant la visualisation des résultats en tableau par exemple. Par ailleurs il calcule les spécificités externes et internes propres au texte, et en mesure la richesse lexicale ou l'utilisation du vocabulaire. Une fois que la base est créée, l'utilisateur peut exploiter les différentes fonctions d'Hyperbase.

Les sous- fréquences dans les textes sont réparties en binômes, dont le premier élément indique le numéro d'ordre du texte, le second la fréquence du mot dans ce texte.	16 modestement, 1 1 3 1 4 1 6 1 7 1 9 2 10 3 11 1 16 1 19 2 21 1 22 1 11 modestes, 5 1 6 3 7 1 11 1 12 1 17 1 19 1 21 1 35 modestie, 4 1 5 5 6 14 8 1 9 6 10 2 12 1 13 2 1 1 modesties, 8 1 2 modicité, 9 1 10 1 2 modiffa, 17 1 21 1 1 modiffaient, 22 1 1 modifiant, 20 1 4 modification, 19 2 20 1 21 1 7 modifications, 5 2 10 1 14 2 15 1 17 1 6 modifie, 20 1 21 1 22 4 10 modifié, 15 1 19 4 21 1 22 4 1 modifiée, 22 1 2 modifiées, 9 1 21 1 1 modifiant, 22 1 12 modifier, 11 1 16 1 19 1 20 2 21 5 22 2 2 modifièrent, 12 1 16 1 2 modifiés, 19 1 22 1 2 modique, 5 1 10 1 2 modiste, 13 1 14 1 2 modulait, 13 1 15 1 2 modulation, 5 1 21 1 7 modulations, 5 1 10 2 13 3 22 1 1 modules, 22 1 1 modus, 6 1 2 Moedeler, 19 2 1 Moedler, 19 1 1 Moelder, 19 1 1 moele, 5 1 5 moelle, 14 2 19 1 21 1 22 1 3 moelles, 15 2 18 1 3 moelleuse, 9 1 14 1 21 1 1 moelleuses, 9 1 5 moelleux, 11 2 13 1 16 1 21 1 1 moellon, 14 1 2 moellons, 21 2 137 moeurs, 2 2 3 1 4 6 5 13 6 43 7 9 8 16 9 6 10 4 11 6 12 5 13 4 14 6 15 3 18 4 20 3 21 2 22 4	N° 2 Paysan 2 N° 4 Candide 6 N° 6 Emile 43 N° 8 Rancé 16 N° 10 Pons 4 N° 12 Mare 5 N° 14 Bouvard 6 N° 18 Bête 4 N° 21 Swann 2 TOUS LES TEXTES	N° 3 Zadig 1 N° 5 Héloïse 13 N° 7 Atala 9 N° 9 Chouans 6 N° 11 Indiana 6 N° 13 Bovary 4 N° 15 Une Vie 3 N° 20 Stortiz 3 N° 22 Temps 4
Chercher un mot	<p>moeurs fréquence totale: 137</p> <p><i>CLIQUEZ SUR UN TEXTE (ou sur TOUS) pour y repérer les contextes du mot "moeurs"</i></p> <p><i>Cliquer AILLEURS dans cette fenêtre pour l'effacer</i></p>		

Figure 4 : Une page du dictionnaire (ou index)

Exploitation et représentation graphique des résultats :

Pour ce qui est des mots, le chercheur a la possibilité, de voir la fréquence répartie dans le texte ou dans le corpus entier. A partir de ce mot, il peut voir la représentation graphique de sa fréquence en la comparant à d'autres mots et il pourra consulter un index hiérarchique ordonné par fréquence. Les fonction hypertextuelles permettent de voir sur le champ le contexte d'un mot.

Hyperbase propose des outils permettant d'assumer une exploitation pratique, méthodique et pragmatique de la documentation. Par exemple l'outil « Contexte », en sollicitant cette option, par rapport à un mot ou une expression désirées, l'utilisateur se voit offrir la possibilité de le voir placé à l'intérieur de différents textes où ils sont employés. S'il clique sur une des passages affichées, il a la possibilité de voir la page entière où le mot apparaît. Le logiciel effectue des allés-retours permettant une vérification ultérieure des résultats et une interrogation plus ample. Une autre fonction documentaire est représentée par « Concordance » qui extrait également des contextes plus étroits qui tiennent par exemple en une ligne.

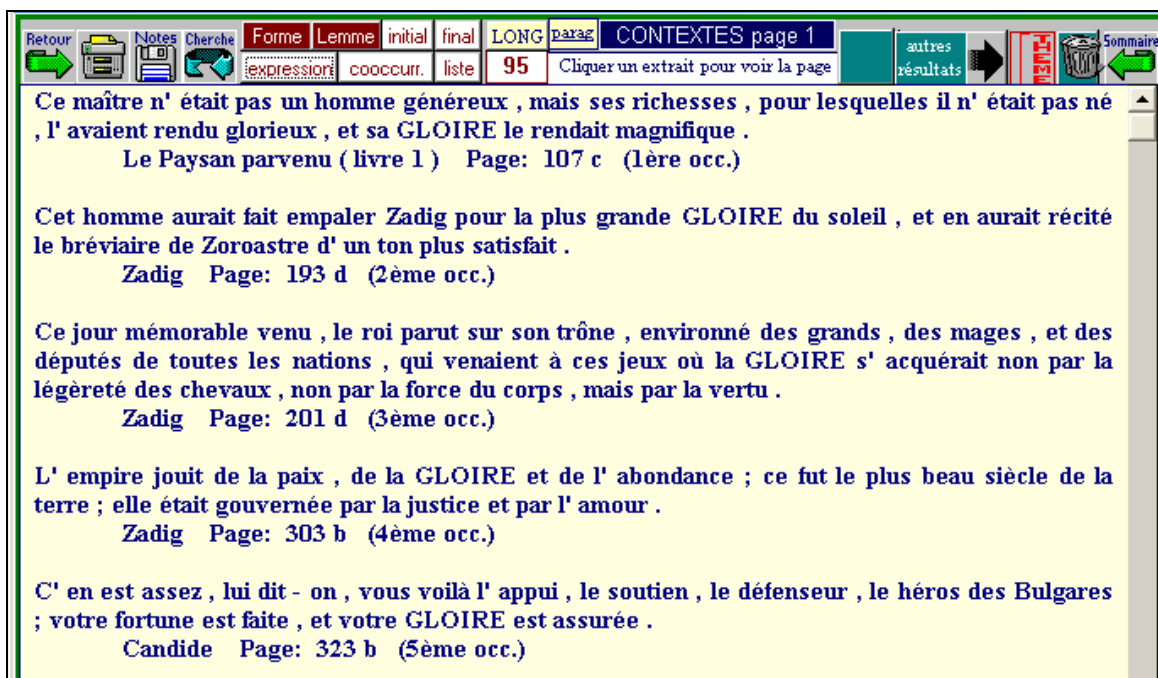


Figure 5 ; exemple du résultat de la fonction « Contexte »

Le recours aux illustrations

L'une des caractéristiques du logiciel Hyperbase est le recours à la représentation visuelle et graphique pour illustrer les résultats statistiques. Il y a par exemple la possibilité de voir sous forme d'histogramme les écarts entre les fréquences d'un mots observé dans le texte , et la fréquence théorique qu'on aurait du s'attendre à obtenir en considérant la proportion du texte par rapport à l'ensemble .

Les listes de résultats possèdent à leurs tour un certain nombre de fonctionnalités et de représentations graphiques , en colonnes ou en lignes , et on peut les soumettre à deux types d'analyse ,factorielle et arboré.

D'autres outils permettent de voir l'évolution du langage, la richesse du vocabulaire. c'est le programme factorielle qui calcule les spécificités internes de la langue.

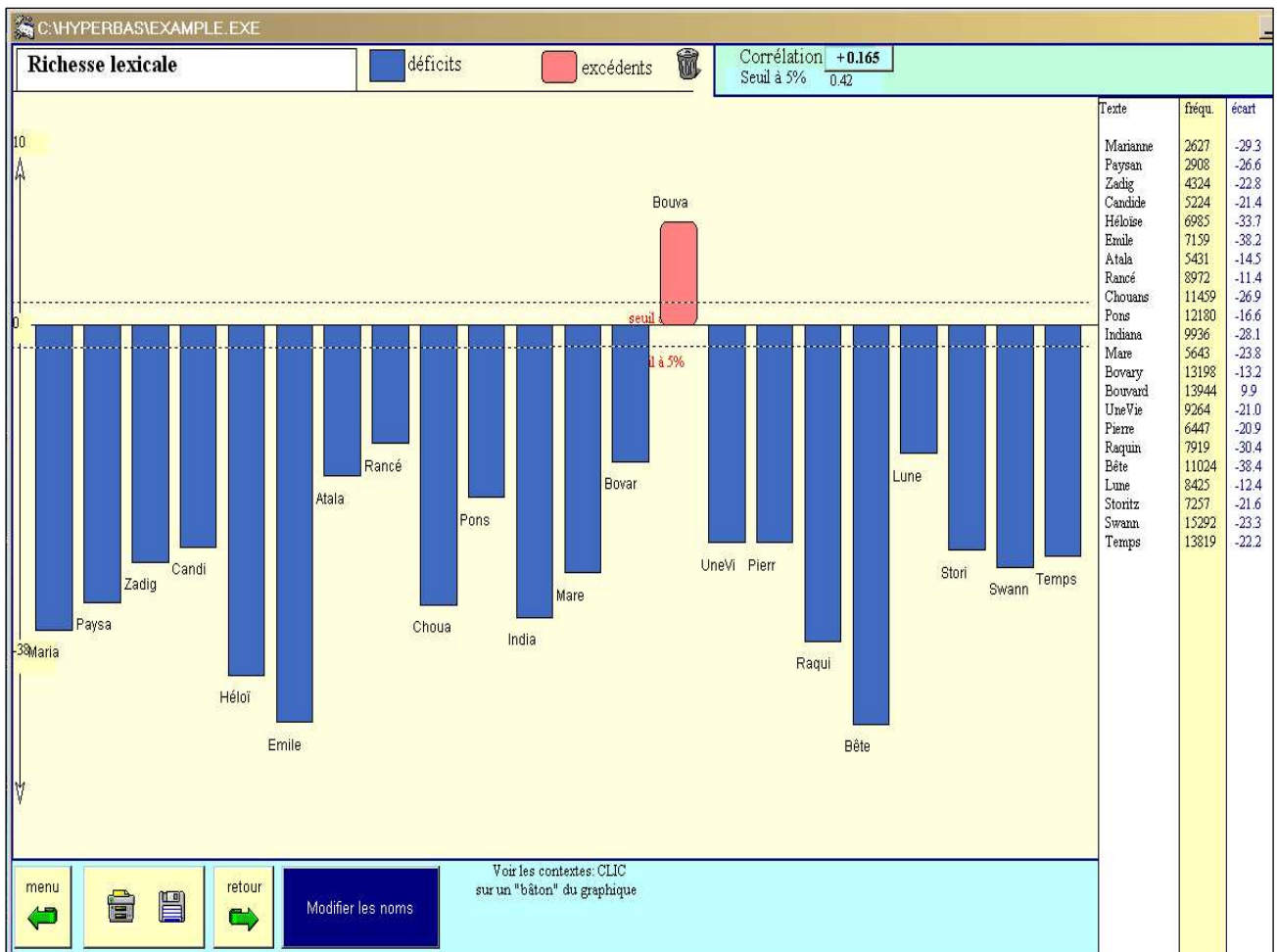


Figure 6 : exemple d'une représentation en forme d'histogramme

3.3.2-Les trois logiciels d'analyse automatique statistique et documentaire de la gamme d' « Ascetic » :

Nous citons brièvement un deuxième modèle à titre plus ample d'exemple :

L'ensemble des logiciels d'analyse automatique de documents de la gamme « Ascetic ».

Pouvant être exploité et mis au service de la documentation, la gestion des connaissances,

l'intelligence économique, la veille technologique, la revue de presse, les analyses stratégiques, scientifiques, éducatives, linguistiques...la gamme « Acsétic » est à la fois constituée :

D'un logiciel d'analyse de textes Tropes :

Conçu pour la classification sémantique, l'analyse qualitative, l'extraction de mots clés et la constitution de thésaurus, le logiciel Tropes est un logiciel pour le traitement documentaire, les études de marché, l'analyse d'entretiens, les études littéraires, ...etc.

D'un moteur de recherche Zoom -

Avec son moteur d'indexation en langage naturel, ses classifications sémantiques intégrées, Zoom est un moteur de recherche conçu pour la gestion électronique de documents, l'Intelligence Economique, l'analyse de presse, la veille technologique

D'un système d'indexation- ; classification Sémantique pour serveurs

Index est un système de classification et d'indexation conçu pour la gestion de différentes applications et systèmes comme la documentation en langage naturel, la Gestion des connaissances...etc. Pour les grandes entreprises ou les grands projets notamment scientifiques.

Apportant une approche sémantique et pragmatique- entre autre en ayant recours à l'instar d'Hyperbase a des comparaisons entre les résultats d'une analyse donnée effectuée et la base de donnée contenue dans le logiciel qui est constituée des différents textes appartenant à différentes périodes ou genres- , Tropes permet d'obtenir des analyses reproductibles (deux personnes étudiant un même texte et observant la même technique d'analyse obtiendront des résultats semblables).

Parmi les outils d'analyse sémantique contenus dans Tropes, on peut citer :

- La classification automatique des mots du texte.
- La détection du contexte global du document.
- La catégorisation des mots-outils (analyse qualitative) .
- La localisation de séries chronologiques (analyse du discours) .

Le calcul des relations entre thèmes (lexicométrie) .

Des dictionnaires sémantiques personnalisables, appelés Scénarios, conçus pour fabriquer rapidement des outils d'interprétation ou des thesaurus automatiques ;

L'extraction terminologique (couplée à l'analyse sémantique) .

Le logiciel d'analyse de textes Tropes est interconnecté avec le moteur de recherche sémantique Zoom, pour constituer Tropes Zoom, une suite complète de recherche et de traitement automatique et documentaire de l'information.

Nous avons essayé à travers ce chapitre de mettre en exergue le caractère interdisciplinaire de l'ingénierie linguistique , ce caractère donnant lieu à de nouvelles pratiques et disciplines ,qui montrent à chaque fois un peu plus à quel point il est intéressant et important de créer un dialogue entre les différentes disciplines .Nous avons rendu compte des points en commun qui lie la linguistique à l'informatique , et que se sont justement les passerelles entre ces deux disciplines, donnant lieu à des interactions aboutissant à de nouvelles applications et nouvelles formes d'analyse .

Parmi ces nouvelles formes d'analyse automatique figure l'analyse automatique statistique textuelles et par conséquent la linguistique de corpus .Les logiciels Hyperbase et la gamme des logicielles d' « Ascetic » en sont deux exemples par excellence.

Nous poursuivrons en nous attaquant à autre cas d'interdisciplinarité dans le chapitre suivant, celui de la traduction automatique.

CHAPITRE IV

Traduction automatique

Un travail relevant du traitement automatique du langage, doit nécessairement, pour des considérations historiques, épistémologiques et techniques que nous tenterons d'expliquer, opérer un passage par la Traduction Automatique (TA). Ceci dit, la Traduction Automatique est également un des domaines interdisciplinaires du traitement automatique du langage puisqu'elle est le résultat de l'informatisation du processus de traduction. Dans ce chapitre ; nous évoquons donc brièvement la question de l'interdisciplinarité de la Traduction Automatique, ouvrons une parenthèse épistémologique sur l'origine du traitement automatique du langage, dont la traduction automatique en est la première application, présentons le logiciel de traduction automatique et ses composants, ainsi que son principe de fonctionnement. Et finissons par présenté un des problèmes que rencontre la Traduction Automatique, celui de l'ambiguïté et illustrons à travers le traducteur automatique « Reverso » , ensuite concluons.

Les ouvrages traitants de la Traduction Automatique et entre autre de son caractère interdisciplinaire étant relativement rares, sinon écrits en anglais, nous avons eu recours beaucoup plus aux différents travaux et rapports de recherches des laboratoires spécialisés ainsi qu'aux articles authentiques .

4.1- Traduction automatique et interdisciplinarité :

La première des constatations à faire au sujet de la Traduction Automatique est le fait qu'elle soit également une discipline interdisciplinaire (cf. rappelons que tout le traitement automatique du langage est un domaine interdisciplinaire), simplement par le fait qu'elle consiste comme nous l'avons déjà dit en l'automatisation à travers la

programmation informatique ¹de cette pratique qu'est la traduction. Ainsi Roland Raoul Kouassi, précise que :

« La traduction automatique pose le problème des universaux du langage et celui des paramètres de fonctionnement inhérents à chaque langue naturelle...la traduction automatique est basée fondamentalement sur des approches qui s'organisent autour du fait de la langue en tant que phénomène, c'est-à-dire en tant que fait réalisé ou produit. On établit des correspondances entre des mots et expressions. On écrit des algorithmes de traitement informatiques pour programmer ces correspondances ». (Roland Raoul Kouassi .2005)

Il appartient au champ de la traduction et désormais de la traduction automatique de traiter tout ce qui est en rapport avec les universaux du langage ainsi que les mécanismes et principes (les règles) de fonctionnement de chaque langue .Ce traitement qui consiste à établir des correspondances et parallélismes entre les mots, les phrases et les paragraphes appartenant à deux ou plusieurs langues ; et donc traduire, est programmé à l'aide de langages ,d' algorithmes et de techniques de traitement informatique destinées pour. C'est une fois de plus à ce niveau , c'est-à-dire le travail de collaboration , ce mouvement interactif et ce dialogue entre traducteurs, linguistes ou ce qu'on appel dorénavant « Talistes » ² ainsi qu'informaticiens , que se traduit le caractère interdisciplinaire de la Traduction automatique.

Nous essayerons de montrer plus bas , après le passage sur la question épistémologique ; à travers les composantes, ainsi que quelques uns des principes et paramètres de fonctionnement des traducteurs automatiques, comment se conjugue et se

¹ Par opposition aux autres formes de programmation.

² Spécialistes du traitement automatique du langage.

concrétise ce caractère interdisciplinaire nous expliciterons de la sorte l'apport de la linguistique ainsi que celui de l'informatique .

4.2.-Questions épistémologiques :

D'un point de vue chronologique , la traduction automatique a précédé l'ingénierie linguistique à proprement dite car l'élaboration du traitement automatique du langage a vu naître le jour avec les premières expérimentations effectuées en traduction automatique ainsi que les premiers prototypes de calculateurs utilisés en cryptographie durant la seconde guerre mondiale « L'histoire du Traitement Automatique du langage commence par les expérimentation en traduction automatique qui débutèrent à la fin des années 1940 , aux Etats-Unis et en grande Bretagne » (Jacqueline Léon 2001.) . La première démonstration a eu lieu à New York en 1954, ce fut une double et historique traduction d'un passage de l'évangile (de l'anglais au russe puis à l'anglais) , les deux fameux passage sont :

« L'esprit est ardent mais la chair est faible »

« L'alcool est fort mais la viande ne vaut rien »

L'informatique a donc ouvert de nouvelles perspectives pour différents domaines et notamment l'étude des langues et très exactement la Traduction Automatique durant cette période qui marquent le début du traitement automatique du langage du moment où :

" Les premières recherches dans ce domaine été entamée dans

l'euphorie des premiers grands succès de l'informatique : les réussites

remportées dans le traitement des données numériques avaient

alors amené les informaticiens à s'intéresser à d'autres domaines que

les domaines purement mathématiques, en particulier celui du langage (...)

l'idée qui animait les chercheurs de l'époque était que pour travailler

sur le langage , il s'agissait de procéder à une simple substitution de

données : les données numériques seraient remplacées par les données alphanumériques que constituaient les mots ". (Frédérique LAB .1988)

Frédérique LAB précise également que :

"La Traduction Automatique des débuts était essentiellement une traduction de mots , de fait la part dévolue aux dictionnaires était énorme , celle consacrée à la syntaxe au minimum. Les problèmes relatifs au sens d'une phrase , d'un texte , n'étaient pas abordés en tant que domaine de recherche spécifique, pas plus que l'activité de traduction n'était envisagée comme une activité ayant sa problématique propre. (*ibid*)

C'est après plusieurs tentatives et travaux de recherches et d'échecs de systèmes , comme il est habituellement le cas dans le développement des disciplines -,de la première génération pour que les ingénieurs impliqués dans ces travaux accordent aux recherches linguistiques(syntaxe et sémantique) la part qui leur revenait, afin de justement permettre le développement de logiciels présentant des résultats comme le font ceux disponibles à présent, comme par exemple les logiciels de Traduction Automatique "Systran" ou " Reverso" .

Jusqu' en 1960, pour des raisons politiques , militaires , et stratégiques, et notamment durant la guerre froide ³ ; le développement de la traduction automatique a atteint un niveau mondial , et c'est pendant cette décennie qu'apparaissent la plus part des idées et des méthodes encore en cours de nos jours dans la plus part des théories et applications du traitement automatique des langues naturelles .

³ Les outils informatiques et notamment la traduction automatique étaient considérés durant la guerre froide, comme un moyens stratégique de défense au même au même titre que le reste des politiques et des armes.

Et pour une approche plus pragmatique :

4.3 Logiciel de traduction

4.3.1-Définition :

Un logiciel de traduction peut être définie comme étant une application ou un outil du traitement automatique du langage qui permet d'obtenir de façon automatique (informatique) une traduction de textes (lettres, rapports, articles, sites web...) d'une langue (source) vers une autre langue (cible).

Le souci majeur des spécialistes, des groupes de recherches et développeurs ; est de faire en sorte que le texte traduit doit être autant que possible correcte dans la langue cible et restituer l'information contenue dans le texte d'origine .

Jusqu'à présent (le mouvement de développement en perpétuelle évolution) les logiciels de traduction peuvent fonctionner en environnement PC ,Internet , Intranet⁴ soit sous la forme d'une application autonome (logiciels de traduction par exemple) , soit sous la forme de fonctions de traduction directement intégrées dans une application (dans un Intranet, dans les moteur de recherche ;comme Google, dans les barres d'outils de certains logiciels qu'on peut ajouter à l'ensemble des barres d'outils qu'on installe dans un ordinateur, ou encore dans les logiciels de traitement de texte Word ou Excel...).

Il existe par ailleurs d'autres outils d'aide à la traduction qui sont complémentaires à un logiciel de traduction ce sont les mémoires de traduction comme (Trados) et dictionnaires électroniques bilingues comme (Collins) ou multilingues comme(EuroDico).

4.3.2- Composants d'un logiciel de traduction

En règle générale ; un logiciel de traduction est composé :

1-De règles linguistiques.

⁴Un intranet est un réseau locale et privé proper par exemple à une (entreprise) ou à un espace donné qui utilise les technologies de l'Internet., mais ne s'ouvre pas aux connexions publiques contrairement à la Toile.

2-De dictionnaires conçus spécifiquement pour le logiciel de traduction, sachant que pour chaque application, des dictionnaires spécifiques y sont implémentés.

3-D'une interface : permettant selon les produits (selon les traducteurs) de conserver le format, réviser la traduction, adapter et personnaliser les dictionnaires...

4.3.2.1- Les règles linguistiques :

Le moteur de traduction est le cœur technologique du logiciel de traduction. C'est là où se réalise concrètement la traduction automatique. On y applique les règles linguistiques de transformation propre à chaque langue. Les étapes de traduction automatique sont les suivantes :

1-L' analyse du texte source (cf. c'est le principe d'analyse cité dans le § 7 du chapitre I)

2-L' application de règles linguistiques de transformation et de désambiguïsation

On entend par transformation à ce stade ; la transformation linguistique qui s'opère lors du passage d'une langue à une autre, chaque langue ayant ses propriétés et qui mène entre autre vers la désambiguïsation qui est le retour sur les constructions linguistiques équivoques dans le souci de les expliciter. L'ambiguïté est considérée comme étant l'un des problèmes principaux de la traduction automatique (cf. nous y reviendrons dans le §6 du présent chapitre).

Les règles linguistiques appliquées peuvent être : l'ordre des mots, les informations morphologiques, syntaxiques et grammaticales, la sémantique, l'analyse pragmatique des composants de la phrase. Plus ces règles sont fines ; plus la traduction obtenue est précise.

3-En troisième lieu : la génération de la traduction : la production de la traduction à travers des algorithmes de traitement informatique précis.

Dans la poursuite de l'interdisciplinarité que nous avons évoquée plus haut, l'analyse du texte et l'application des règles linguistiques constituent la partie linguistique du processus de traduction automatique, donc le travail des traducteurs et linguistes. La génération (la programmation) de la traduction, est ce qui relève de l'informatique.

4.3.2.2- Dictionnaires : (spécifiques)

Pour ce qui est des dictionnaires intégrés aux logiciels de traduction, Ils ne doivent pas être comme dans les premières générations de traducteurs automatiques, de simple dictionnaire de mots faisant uniquement office d'une liste mots ou d'expressions avec leurs traductions, Chaque construction linguistique doit être définie avec des informations linguistiques précises (syntaxe, sémantique pragmatique...) dans la langue source puis dans la langue cible. Ces informations sont ensuite gérées par le moteur de traduction. Plus les dictionnaires sont riches, plus la traduction obtenue est précise.

4.3.2.3- L'interface

L'interface ⁵des logiciels de traduction a un rôle important, puisque c'est elle qui permet la réalisation et l'exploitation de la traduction. En fait une fois le texte dans la langue cible est obtenu, l'utilisateur souhaitera le relire, le vérifier et le personnaliser pour l'adapter à son style, à son activité, à ses besoins, bref à son niveau... Ce besoin de révision et d'adaptation dépend de l'utilisation finale de la traduction (compréhension, diffusion, publication externe...) et s'effectue via l'interface.

L'interface d'un logiciel de traduction se compose :

- 1- De différents modules de traduction.
- 2- D'outils de traduction et de révision qui facilitent le processus de traduction et de révision du texte .

1-Modules de traduction :

Les modules de traduction peuvent être définis comme étant l'ensemble des éléments, des options et des fonctionnalités qui constituent d'un point de vue général le logiciel de traduction. Nous pouvons citer la :

Présentation du texte source et de sa traduction sous forme de deux fenêtres l'une en dessus de l'autre ou l'une à côté de l'autre, (comme nous allons montrer dans l'exemple plus bas)

Conservation de la mise en page , option que ne fournissent pas l'ensemble des traducteurs.

Intégration dans des logiciels Microsoft Word, Excel (95, 98, 2000), dans les barres d'outils d'Internet Explorer (4, 5),,ou des moteurs de recherche.

⁵ (Cf. voir définition d'interface dans le chapitre V)

Traduction de documents en intégral ou encore de sites web en partie ou complètement.

2 Outils de traduction et de révision :

Comme leur nom l'indique, les outils de traduction et de révision sont l'ensemble des fonctionnalités permettant une relative amélioration du résultat de la traduction automatique, il s'agit de :

Alignement des paragraphes source et cible,

Liste et marquage des mots inconnus.

Création et marquage des mots réservés (mots ou parties de textes que l'on ne souhaite pas traduire) .

Proposition d'alternatives de traduction quand le contexte est ambiguë.

Enregistrement du texte en mode bilingue.

Possibilité de traduction paragraphe par paragraphe ou texte entier.

4.4-Fonctionnement de la traduction automatique :

« La traduction implique deux messages équivalents dans deux codes différents. L'équivalence dans la différence est le problème cardinal du langage et le principal objet de la linguistique » (Roman JACOBSON, T.1, 1963, p 80)

A l'encontre des théories de la traduction humaine et/ou classique ; issues de la linguistique comparative, et qui ne semblent pas rencontrer autant de problèmes que la traduction automatique, -bien qu'il y en ait- celle-ci se heurte en fait à des difficultés aussi bien d'ordre lexicale et syntaxique que sémantique et pragmatique.

C'est pour cela qu'à fin de fonctionner convenablement et obtenir les meilleurs résultats possibles; durant la phase de programmation , en plus des règles fondamentales de grammaire de la langue source et de la langue cible ainsi que des exceptions à ces règles, tout vocabulaire doit être accompagné,- comme nous l'avons précisé dans la nature

des dictionnaires des traducteurs automatiques (cf. § 5.3.2) -d'informations des plus diverses. En font partie :

- La catégorie grammaticale et les formes dérivées (substantif, verbe...) ;
- La signification dans le contexte ou la désambiguïsation (par ex. «Driver» signifie « chauffeur » dans un contexte général ou « pilote » dans un contexte informatique.
- Pour les substantifs, on détermine s'il s'agit d'un être animé ; son genre et son nombre, d'un objet inanimé, d'un lieu, ou encore d'un concept, c'est-à-dire les caractéristiques grammaticales. Pour les verbes, il faut déterminer par exemple s'ils sont transitifs ou intransitifs ou quel type de complément d'objet suit. Toutes ces procédures se passent durant la programmation.

4.4.1-Le principe de fonctionnement de la traduction automatique :

Lors de la traduction, le moteur de traduction procède comme suit : il découpe chaque phrase en mots pour les analyser et les recherche dans le dictionnaire. Le programme détermine alors les propriétés grammaticales, les traductions possibles de chacun des mots (en fonction du lexique contenu dans les dictionnaires), ainsi que les connaissances supplémentaires. Ensuite, il effectue une analyse syntaxique de la phrase. Il découpe la phrase dans les différents éléments qui la constituent. Il détermine ensuite la fonction de chaque mot dans la phrase comme sujet, verbe, complément, et leur affecte une valeur sémantique.

Ensuite, la traduction proprement dite a lieu en deux étapes :

Première étape : le transfert lexical qui peut se résumer au fait qu' à chaque mot est affectée sa traduction appropriée dans le contexte de la phrase (prise en considération du contexte). Outre les propriétés grammaticales, le programme tient compte des informations supplémentaires enregistrées dans le dictionnaire pour chaque mot. Plus il existe d'entrées pour chaque mot, plus le programme prend des décisions fondées et la traduction est meilleure.

En second lieu le transfert structurel qui assure le bon agencement des mots dans la phrase et entreprend d'autres adaptations structurelles. Par adaptations structurelles, on entend la mise en correspondance des schémas structuraux des langues, et catégories de

transformations comme par exemple la transposition d'une proposition interrogative française en son équivalent Anglais ou inversement.

Le programme crée ensuite les formes de constructions linguistiques correctes et confère à la traduction sa forme définitive.

La figure suivante explique de nouveau le processus de traduction :

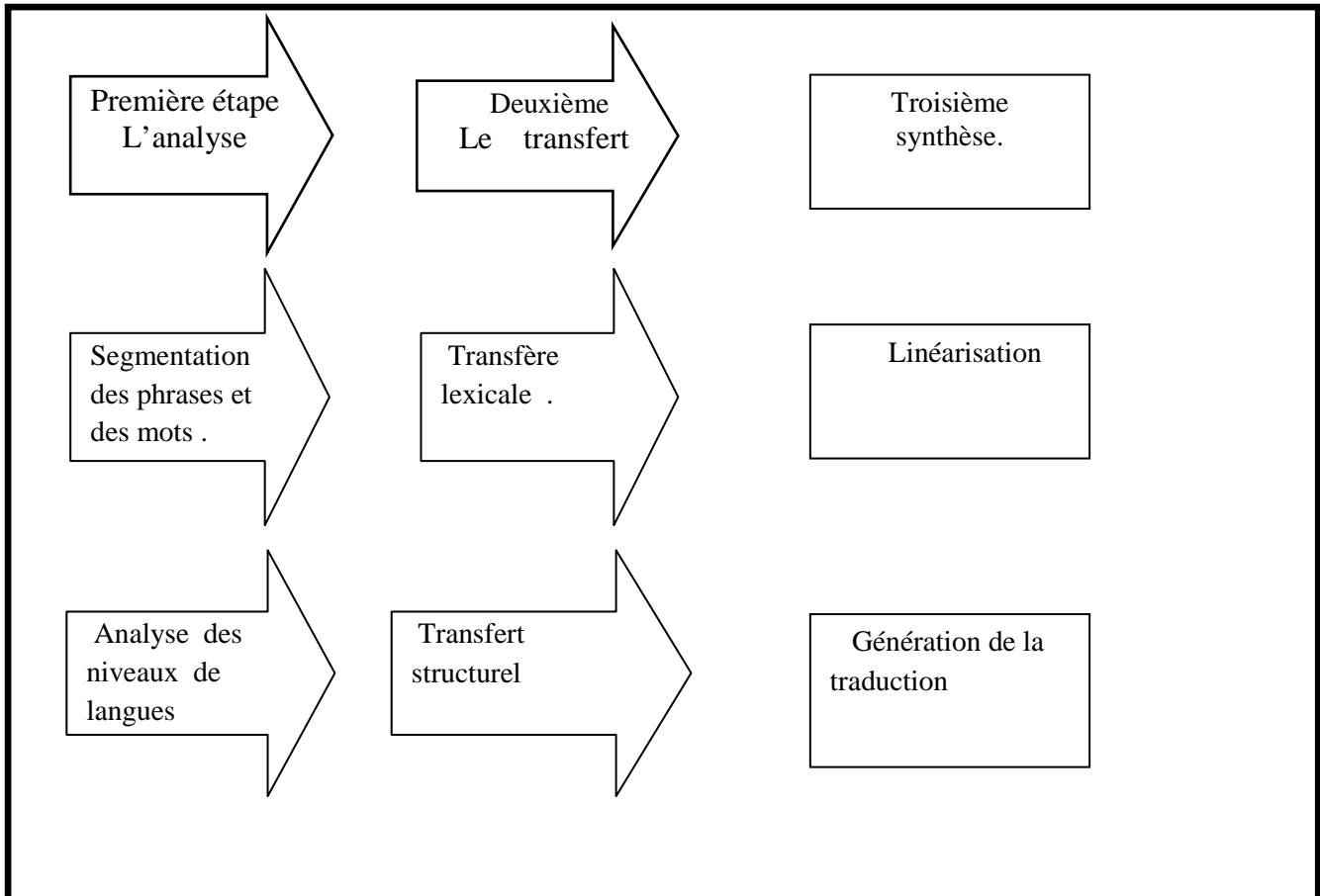


Figure 6 : Schéma récapitulatif du processus de traduction.

4.5 Cas pratique : Le traducteur « Reverso » :

Parmi le nombre croissant des traducteurs automatiques qui ne cessent d'évoluer, nous avons opté pour le traducteur automatique « Reverso » du groupe « Sofitissimo ».

« Sofitissimo » est un éditeur et intégrateur de technologies linguistiques et de solutions multilingues destinées aux entreprises, aux administrations et au grand public.





Le logiciel en question est disponible sur la Toile à partir de plusieurs sites :

Nous l'avons atteint à partir du lien :

<http://www.reverso.com/> , dont l'interface est la suivante :

Nous présentons l'interface du logiciel constitué de plusieurs petits espaces afin de savoir ce qu'un logiciel de traduction peut à l'heure qu'il est faire comme activités (en l'occurrence Réverso). , nous présentons nos commentaires des espaces en forme d'intitulés soulignés en gras :

1- L'interface du logiciel avant la traduction :

<p><u>Le premier espace , celui de la traduction .</u></p> <p>Tapez ou collez le texte à traduire:</p> <p>Direction de traduction :</p> <p>Traduire :</p> <p> Webmaster Tools</p>	<p></p> <p><u>Un espace conjugaison</u></p> <p>Tapez un verbe à l'infinitif ou une forme conjuguée:</p> <p>Langue : Conjuguer</p> <p></p>
	<p></p> <p>Mot ou expression :</p> <p>Rechercher :</p> <p>Dictionnaire :</p> <p><u>Un espace réservé pour les définitions .</u></p>

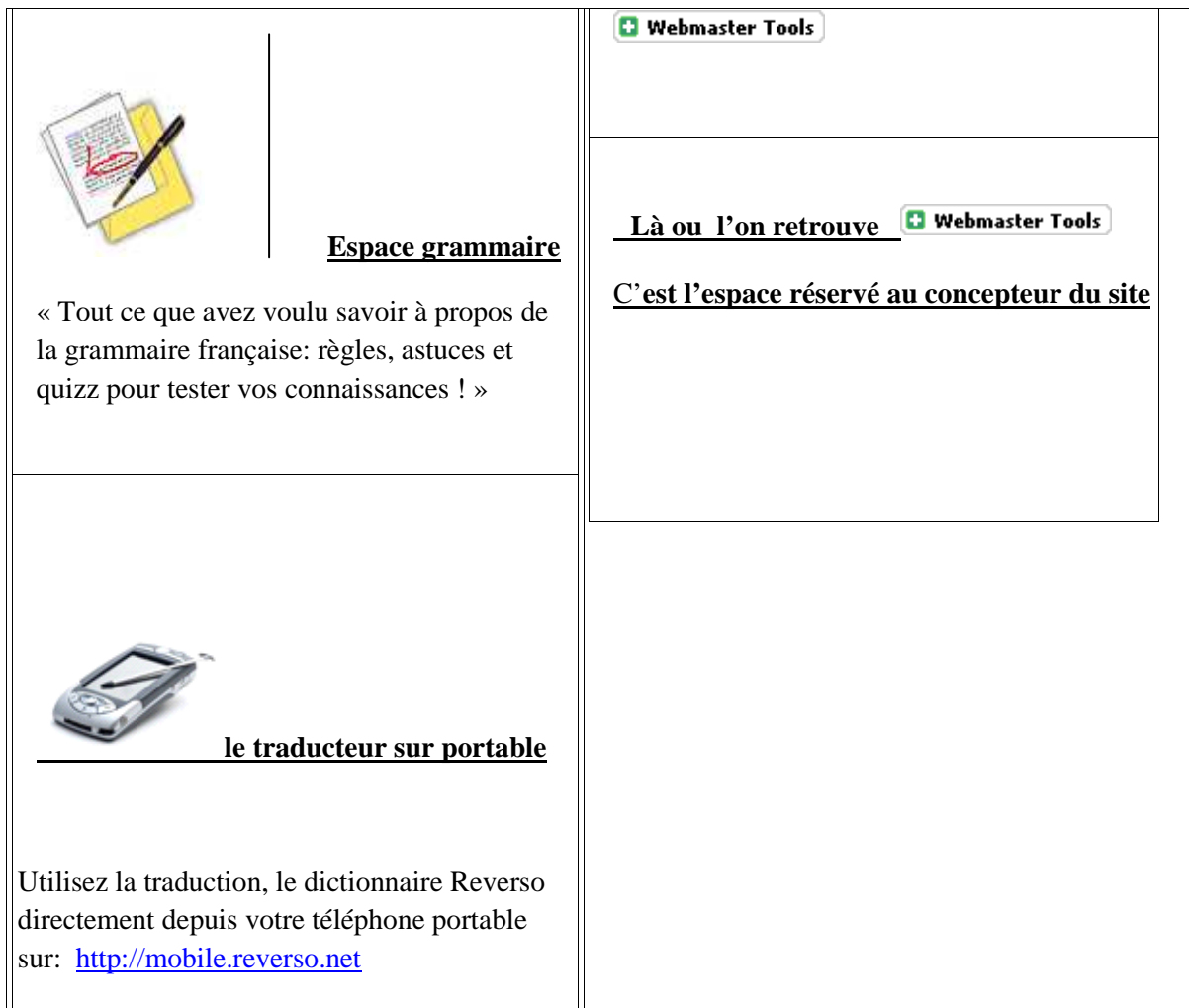



Figure 7 :Les principaux constituants de l'interface du logiciel de traduction automatique Reverso
(Avant la traduction)

2- Voici en quoi consiste (les éléments ajoutés) l'interface après la traduction :

Texte d'origine 

Le haut parleur est l'icône de la lecture orale

Traduction de Reverso en Anglais 

Le second est pour la lecture

du texte traduit

L'utilisateur aura la possibilité d'évaluer les résultats :

Merci de votre avis. Vous pouvez aussi suggérer une meilleure traduction ou apporter un commentaire.	Selon vous, cette traduction est : Compréhensible, passable, mauvaise. _____
<u>Une possibilité d'évaluation du résultat</u>	Vous pouvez modifier votre note en cliquant sur l'étoile correspondante .

Il aura également la possibilité de l'améliorer ou la modifier

Votre texte d'origine:	Proposez une meilleure traduction (optionnel)
Afficher la traduction de Reverso	<u>La possibilité de l'améliorer</u>

Ainsi que de la commenter :

<u>Possibilité de la commenter</u>	Ajouter un commentaire (optionnel):
---	-------------------------------------

Les éléments ajoutés à l'interface après la traduction .

Le choix du traducteur « Reverso » se justifie par rapport au fait qu'il figure parmi l'un des meilleurs traducteurs automatiques à présent. Nous l'avons donc choisi de telle sorte à montrer que ce meilleur traducteur automatique, aussi performant qu'il puisse l'être, présente encore des failles et lacunes. Les langues traitées par celui-ci sont le : français, anglais, allemand, espagnol, russe.

L'interface de Reverso :

Les modules et les outils disponibles dans les produits « Reverso » dépendent de la version du logiciel et de l'environnement choisi.

4.6. Limites du traducteur :

La traduction automatique rencontre certes à l'heure qu'il est pas mal de problèmes et on y détecte beaucoup de failles. L'une des plus importantes lacunes, est le problème de la non-univocité des emplois et expressions linguistique et donc l'ambiguïté, qui résulte de l'abondance des nuances, des équivalents et des différentes occurrences et alternatives qu'offre les différents niveaux et emplois de la langue. En parlant de l'ambiguïté, Xavier Tannier voit que :

« Le problème qui se pose est que plusieurs catégories peuvent très souvent convenir à une même forme. En Français, les formes ambiguës sont estimées à environ 25 % du lexique, voire plus pour les mots : Les plus courants » (TANNIER 2006).

Il existe selon les chercheurs plusieurs catégories et types d'ambiguïtés, celles-ci peuvent être classées justement en fonction de différents paramètres et niveaux d'analyse, autrement dit la catégorisation reste relative aux approches envisagées. On peut notamment trouver ces quatre types d'ambiguïté :

4.6.1- Ambiguïtés lexicales :

L'un des cas les plus pertinents est le cas de la polysémie : le vocabulaire anglais recense par exemple 31 sens différents pour le verbe « to make », 28 pour le verbe « to go » et 23 pour le nom « head ». Il est actuellement impossible pour un traducteur automatique, à savoir « Reverso » en traduisant, de repérer sur les 23 sens par exemple du mot « head » ; ou les 28 du verbe « to make » quel est celui que l'utilisateur veut employé.

Pour Yves Gentillome :

" ...Le monde et les cultures évoluent. Tous les jours apparaissent des concepts nouveaux. Grâce aux glissements sémantiques, on se tire d'affaire, mais, ce faisant ; on laisse la porte ouverte à l'ambiguïté

L'ambiguïté devient , dans cette perspective , comme un cadeau du ciel ,
qui permet à la langue de s'adapter à des situations nouvelles , et qui
permet de parler de ce qui paraissait indicible avec le lexique disponible.
, (Yves gentillome. 1996 .p 115) .

4.6. 2- Ambiguïté sémantique :

Prenons l'exemple : Une femme a dit que ce sénateur était anormal.

Un francophone sait qu'anormal peut avoir plusieurs sens figurés, par contre, même si Reverso parvient à fournir cette traduction « A woman said that this senator was abnormal. » qui est sémantiquement et syntaxiquement correcte, il n'en demeure pas moins que le rédacteur peut entendre plusieurs autres sens d' « anormal » dans ce contexte. Ce qui revient à dire dans ce cas que le traducteur automatique ne parvient pas à traduire le sens connoté ou figuré , du moment ou il ne contient pas de données qui lui permettent d'atteindre cette objectif .

4.6.3-Ambiguïtés syntaxiques :

Il y a ambiguïté syntaxique chaque fois que l'analyseur détecte plus d'une représentation grammaticalement correcte pour la même phrase, comme dans l'exemple " Time flies ".

L'homme ne trouverait pas de difficulté ; pour interpréter et traduire cet énoncé par le temps s'envole ou passe trop vite, mais la machine (d'autres traducteurs automatiques) l'interpréterons et traduirons de deux façons :

Soit par : Le temps s'envole

Ou alors : chronométrer des mouches. « fly » étant mouche en anglais.

La traduction de Reverso à donné :

Le temps passe vite .

Un second exemple d'ambiguïté syntaxique :

Je suis un grand homme .

Ici , " suis" peut être auxiliaire présent de être ou présent de suivre .Le traducteur automatique ne peut pas prendre en charge les deux cas de figure.

La traduction de Reverso à donné :

I am a great (tall) man. Alors que suivre en Anglais est traduit par “ Follow”, et on devrait obtenir également ;” I follow a great man.”

Notre objectif n'étant pas de s'attarder sur ce problème d'ambiguïté dans ce modeste travail ,mais d'ouvrir une parenthèse (les promotions et génération futures pourront développer davantage ce thème) sur l'un des obstacles que rencontre la Traduction Automatique à l'heure qu'il est , nous , concluons ce chapitre en soulignant que la Traduction Automatique est historiquement et concrètement la première application du traitement automatique du langage , qu'elle est dans la poursuite et à l'instar de toutes les applications du Traitement Automatique du Langage, le résultat de collaboration entre différents spécialistes, mais qu'après soixante ans d'existence, elle présente encore des lacunes et rencontre des problèmes à savoir l'ambiguïté qui se pose presque à tout les niveaux du langage. Elle pose quelques enjeux qu'elle partage avec le domaine de l'apprentissage des langues assisté par ordinateur, certains aspects de ce dernier ainsi que quelques uns de ses enjeux seront traités dans le dernier chapitre.

CHAPITRE V

Apprentissage et enseignement assisté par ordinateur

Logiciel d'apprentissage du Français

Avant d'entamer ce chapitre, et pour relever toutes sorte de confusion, nous tenons à rappeler que les logiciels d'apprentissage de la langue, appelés également « didacticiels »¹ font également partie des outils et applications du génie linguistique (cf. la liste thématique des applications § 4, chapitre I).

Cela peut s'expliquer par rapport à plusieurs paramètres entre autres ; la programmation informatique : tout logiciel est un programme informatique, un logiciel d'apprentissage d'une langue est un programme qui à été réalisé à partir de différentes données et règles linguistiques (lexicales, grammaticales, sémantiques,...etc.) destinés à des fins pédagogiques. Les fonctionnalités de la synthèse vocale ainsi que la reconnaissance de la parole étants intégrés à un nombre de plus en plus important de didacticiels, et d'interfaces et plates formes d'apprentissage en lignes² sont tous les deux des outils d'ingénierie linguistique permettant l'exercice et l'apprentissage orale de la langue , pour ce qui est de la prononciation les utilisateurs s'entraînent à répéter des séquences ou encore à interagir oralement avec l' interface par exemple en répondant à des questions , ou se présentant....etc.

Ces logiciels contribuent à l'instauration de ce qui est connu comme étant l'apprentissage et enseignement des langues assisté par ordinateur qui peut se faire sans être en ligne-

¹ Contraction des termes didactique et logiciel.

² En informatique, une interface est le dispositif permettant le dialogue homme-machine, c'est en d'autres termes l'ensemble des moyens qui constituent les logiciels inclus dans les systèmes d'exploitations .

uniquement avec un ordinateur qui n'est pas lié à la Toile³-, ils peuvent également constitué les interfaces et plates formes ⁴ d'apprentissage et de formation en ligne.

Dans ce chapitre , nous présentons un logiciel d'apprentissage du français destiné aux premières tranches d'âge « Logitron », le testons afin de ciblé et voir son fonctionnement et ses performances , quel avantages , quel plus , ou valeur ajoutée apporte- il ?Ensuite , et en essayant d'élargir un peu le champ d'étude et notamment en présentant des statistiques et expériences qui ont été effectuées ailleurs dans des pays ayant officiellement entrepris d'intégrer ces nouvelles technologies et techniques dans leur différents systèmes éducatifs et voir un peu quelles nouvelles voies , perspectives et enjeux pour le traitement (apprentissage, enseignement diffusion, production, et promotion) des langues en appui avec des moyens informatiques ..

Le choix de ce logiciel n'a pas été effectué en prenant en compte un très grands nombre de paramètres. Il n'est à ce stade qu'à titre indicatif et non récapitulatif. L'objectif étant d'en faire une introduction permettant d'attirer l'attention sur une réalité en plein expansion, laissant ainsi le champ ouvert aux futurs collègues qui souhaiteraient développer davantage la question. Nous avons tout de même fais en sorte qu'il soit destiné aux plus jeunes –et parfois aux débutants d'un point vu général- et qu'il soit gratuit et téléchargeable, autrement dit mis à la disposition des apprenants. Il en existe pas mal d'autres plus développés qui contiennent en plus des fonctionnalités que présente celui-ci, la possibilité de s'entraîner et d'apprendre oralement la langue grâce aux techniques de synthèse et de reconnaissance vocale (cf.la liste thématique détaillée en annexe).Cependant leur disponibilité est relative, ils sont parfois pas permis au téléchargement, pour des raisons principalement commerciales, et donc pas accessibles à la totalité du public d'apprenants, dans la mesure ou ils sont parfois chers.

5.1-Présentation et exploration du logiciel :

Le logiciel en question se présente sous forme d'une interface constitué de cinq parties.

³ On interent .

⁴ Une plateforme d'apprentissage ou de formation est un logiciel installé sur un serveur informatique. C'est autrement dit un site web auquel l'utilisateur peut se connecter et accéder notamment aux différents cours et formations prévues par le formateur ou le concepteur du logiciel en question.

Sur les cinq parties qui constituent le logiciel, nous en présentons trois :

La première : « Des mots plein la tête » :

Contenu :

Avant de commencer à l'utiliser, l'utilisateur ou l'apprenant est amené à procéder à un certain réglage d'options auxquelles il aura recours, comme les critères pour qu'un mot soit acquis, la durée de l'affichage des mots ... etc.

La partie « Des mots plein la tête », à l'image des cinq autres parties se présente sous forme d'une interface graphique constituée de plusieurs activités adapté au jeune âge des apprenants. Il contient quatre mille mots (la grande quantité de ressources et de données , (cf. Cerquiglini ,§5 du présent chapitre) répartis en 31 niveaux. L'apprentissage de la graphie du vocabulaire s'effectue à l'intérieur de dix activités différentes.

Quant à l'écriture des mots, l'élève doit soit utiliser le clavier ou cliquer sur des lettres présentées à l'écran à l'aide de la souris. C'est la première forme d'interactivité entre l'apprenant et l'ordinateur, toujours présente. Chaque mot réussi lors de sa première présentation est considéré comme maîtrisé et ne sera plus demandé. Quand un terme est mal orthographié, le logiciel donne un message qui présente la graphie exacte. Ce mot manqué devra être réussi deux ou trois fois de suite dans une autre activité pour être versé dans la liste des termes acquis, (l'aspect plus ou moins ludique) .Plusieurs paramètres permettent d'adapter le logiciel aux besoins particuliers des apprenants ,c'est l'adaptabilité au niveau de l'apprenant.

L'apprenant a la possibilité de visualiser le cheminement de son évolution (progression méthodique). Ses résultats peuvent être conservés, c'est-à-dire sauvegardé et imprimés .Un gestionnaire permet de limiter l'accès aux menus, de gérer les inscriptions – l'apprenant doit s'inscrire et s'identifier avant chaque utilisation- et de modifier les paramètres⁵ du logiciel.

Objectifs :

Ce module vise à mémoriser la graphie des mots les plus fréquemment utilisés. Il développe aussi la compréhension de ces termes en les présentant dans un contexte significatif , à travers un certain nombre d'activités a caractère interactive

⁵ Dans ce cas les éléments et options de réglage.

Public cible :

Les jeunes de 6 à 13 ans

Environnement informatique requis :

Ce logiciel fonctionne sous Windows 95, 98, XP, Vista et Windows7 (32 et 64 bits).

La seconde partie : « Lecto »

Contenu :

Egalement, à la première utilisation, l'utilisateur est amené à effectuer un certain nombre de réglage d'options en rapport avec le contenu du logiciel.

Pour ce qui est du module Lecto-mots, il présente 495 leçons regroupées en neuf groupes d'activités. Les trois premières activités développent principalement les habiletés préalables à la lecture : la discrimination visuelle, la concentration, la mémoire visuelle, et la mémoire des séquences (en ayant justement recours à des jeux de séquences) .Les cinq autres groupes d'activités améliorent l'habileté du lexique en développant la lecture globale des mots, puis des groupes de mots .

Le module Lecto-livres interroge l'élève sur ses lectures (l'aspect interactif toujours présent), et lui donne un certificat confirmant le nombre de livres qu'il a lus (motivation). L'enseignant peut gérer facilement tous les questionnaires créés pour le logiciel.

L'apprenant peut visualiser l'évaluation de toutes les leçons (apprendre à s'auto-corriger et à s'évaluer). Ce bilan indique si les exercices terminés répondent au critère de réussite fixé par l'enseignant. Les résultats de chaque utilisateur peuvent être conservés et imprimés.

Objectifs :

Ce logiciel contient deux modules. Le premier permet d'augmenter la vitesse de lecture et la compréhension, tandis que le second vise à encourager les élèves à lire des livres de la bibliothèque et à long terme toutes lectures

Public cible :

Les élèves du primaire et tout jeune ou adulte désirant améliorer sa vitesse de lecture.

Environnement informatique requis :

Ce logiciel fonctionne sous Windows 95, 98, XP, Vista et Windows7 (32 et 64 bits).

La troisième : « J'écris »

Contenu :

Même procédure pour le réglage d'options .En même temps module du didacticiel, et éditeur de texte. Cet éditeur de texte indique immédiatement après l'écriture de chaque mot s'il est bien orthographié et s'il respecte les règles grammaticales déjà apprises (efficacité de la méthode). Une erreur d'orthographe d'usage sera indiquée par un trait jaune et une faute grammaticale sera soulignée en rouge. En cliquant sur un mot souligné, une aide appropriée est fournie à l'utilisateur afin qu'il puisse se corriger. (Autocorrection et autonomie, cf. Gerbault. 2002)

Objectifs :

Ce logiciel permet à l'utilisateur d'améliorer sa maîtrise de l'orthographe d'usage et grammaticale à partir de ses productions écrites. Pour réaliser cet objectif, il souligne les fautes et peut illustrer les mots écrits correctement.

Public cible :

Les jeunes de 6 à 16 ans

Environnement informatique requis :

Ce logiciel fonctionne sous Windows 95, 98, XP, Vista et Windows7 (32

5.2 - L'informatique, le numérique, le traitement automatique au service de l'acquisition, apprentissage, diffusion et production des langues :

Nouvelles voies et perspectives :

Les nouvelles technologies appliquées aux différents domaines, l'informatique, la programmation notamment, les logiciels (didacticiels) qui en résultent, les interfaces, la toile, plates formes en ligne, les réseaux, les divers moyens et instruments informatiques ouvrent de nouvelles perspectives, défis et voies pour les différents traitements : apprentissage, acquisition, production, diffusion des langues. Il semble selon bon nombre de chercheurs et de

spécialistes (Cohen, Kern, Diki-Kidiri, Bibeau ; Jocelyn Pierre et bien d'autres) que l'avenir, et la « survie » même et le développement des langues se jouent considérablement dans leur capacité à investir des supports et espaces technologiques et notamment informatiques, il y a par conséquent de sérieuses motivations et de réels enjeux à s'interroger sur les actions en question. L'informatisation de l'apprentissage et des différents traitements des langues atteint un tel degré d'importance qu'il est judicieux, voire obligatoire de la considérer et la traiter sous différents angles et lectures parallèles

5.2.1 Les prémices « en France »⁶ : à partir de 1976 :

Nous commencerons par le président de la République Valéry Giscard d'Estaing qui en 1976 mandate S. Nora, inspecteur général des finances, afin de rédiger un rapport devant « faire progresser la réflexion sur les moyens de conduire l'informatisation de la société ». L'informatisation à ce niveau, comprend tous les aspects de la vie et notamment les aspects scientifiques et pédagogiques.

Le fait que se soit le président en personne qui ait pris cette initiative déjà depuis 1976 permet de se faire une idée sur l'impact et le caractère sensible de ces différentes opérations et procédures d'informatisation et d'automatisation. Il savait par une vision anticipée, qui se concrétise de plus en plus de nos jours que la survie des savoirs, entre autres la langue allait dépendre de leur degrés d'investissement des différentes technologies. Déjà à l'époque, la

⁶Ce développement s'est manifesté par exemple aux États-Unis bien avant (cf. Chapitre I et IV)

compétition était déclarée et féroce entre les grands pôles universaux et nations durant la guerre froide ⁷ pour le développement de l'outil informatique.

Nous rappelons juste que les moyens informatiques dont il s'agit peuvent être différents logiciels d'ingénierie linguistique: des didacticiels , interfaces des dictionnaires , des bases de données terminologiques électroniques , des interfaces et plates formes numériques , par la synthèse vocale et la reconnaissance de la parole ; des aides à la compréhension orale et à la prononciation. , des outils de production et de manipulation de texte (ce qu'on appelle plus communément les logiciels d'aide à la rédaction ou traitement de texte),des ressources authentiques sur la toile, à des cours d'autoformation en ligne , la communication en ligne à distance entre apprenants avec des natifs de la langues , des francophones ou des même des publics non-francophone.

5 .2.2-L'après guerre froide, les années 80 , une nouvelle interaction :

En parlant du rôle de l'ordinateur dans l'apprentissage, Rachel Cohen ; psychanalyste , enseignante , et auteur de nombreux articles interdisciplinaires et notamment concernant l'apprentissage des langues par le recours aux didacticiels; dans *Les jeunes enfants ,la découverte de l'écrit et l'ordinateur* (1987) et *Quand l'ordinateur parle.*(1992) ; fait état de travaux de recherches et applications autour de l'ordinateur , du didacticiel et la synthèse vocale au service de l'apprentissage de la langue écrite et orale pour des publics notamment non francophones -c'est d'ailleurs à ce sujet que les deux ouvrages sont davantage intéressants , le fait de s'attaquer à des publics d'apprenants non francophone -des les années 1980 .

Nous constaterons comment l'ordinateur à travers les différents moyens (didacticiels,) et possibilités qu'il offre , par exemple l'interactivité ,l'enfant ou l'apprenant utilise le clavier et donc interagit d'une manière active dans un premier temps car il est tout le temps en train de manipuler soit le clavier, soit la souris, ensuite à travers les système de question-réponse-⁸ ; il

⁷ Les outils informatiques et notamment ceux consacrées au traitement automatique du langage étaient considérés durant la guerre froide, comme une moyens stratégique au même au même titre que le reste des moyens et politiques .

⁸ Question-réponse , une des formes du dialogue Homme-machine .

pose une question et reçoit au retour une réponse , ou inversement, les exercices du logiciel contiennent des questions, auquel l'apprenant répond ,ces activités créent action, échange, et par conséquent motivation .

Les fonctionnalités, exercices , interfaces , et illustrations qu'on retrouve dans les logiciels ;dans l'exemple que nous avons présenté, les parties qui constituent le logiciel logitron, ont des interfaces interactives ,qui contiennent des illustrations et des exercices adaptées au jeune âge des apprenants, des interfaces qui ont été mise en place justement par des équipes interdisciplinaires de spécialistes(pédagogues, didacticiens, informaticien, talistes, cogniticiens...), plus encore la possibilité de régler les paramètres et options du logiciel avant son utilisation,-ces paramètres qui sont généralement en relation avec le niveau de difficulté s'apprentissage souhaité- comme par exemple les critères pour qu'un mot soit acquis (cf. retour au logiciel en question) .

La synthèse vocale dans le cas des logiciels qui l'intègre (par exemple la gamme du logiciel, j'apprend à lire et à écrire , cf. retour aux sites du logiciel en question) qui est l'une de ces possibilités les plus remarquables , joue un rôle très important en ouvrant d'une part la voie à de nouvelles formes de pratiques au quotidien par l'ouverture de la voie à l'exercice et à la pratique orale de la langue dans différentes situations .Ce qui constitue généralement et une fois de plus une solution motivante pour les apprenants et notamment dans le cas où il ont accès à des logiciels leur permettant d'abord d'entendre les prononciations correctes par exemple de mots ou de toutes autres constructions linguistiques , ensuite de s'entendre soi même en la prononçant , donc pratiquer l'exercice orale de langue. Ce qui permet d'établir une comparaison entre les deux formes vocales, et par la suite apprendre à s'auto-corriger.

5.2.3- Les années 90 , Flexibilité et adaptabilité :

Dans son article « La communication médiatisée par ordinateur en langue : recherches et application récentes aux USA », Kern , professeur associé en français à l'université de Californie, et directeur du Berkeley Language Center .Ses recherches portent sur la lecture, l'écriture, et l'acquisition des langues étrangères ainsi que sur l'apprentissage médiatisées par les nouvelles technologies , ne situe qu'au début des années 1990 aux Etats-Unis les premiers projets et les

premières recherches autour de la communication assistée par ordinateur au service des cours et de l'apprentissage des langues .

Nous entendons par cours tous les niveaux et spécialités linguistiques qui rentrent en jeu pour l'apprentissage de la langue. Ces premières études relevaient que la communication assistée par ordinateur - nous rappelons que la communication assistée par ordinateur représente toute forme d'échange écrite ou orale entre un utilisateur et un ordinateur, c'est une forme de dialogue H-M⁹-dans l'exemple de « Logitron » , elle consiste en toute forme d'interactivité entre l'utilisateur et le logiciel (cf. retour au logiciel et aux trois parties présentés plus haut § 4.1) ; donnait plus d'occasions aux apprenants de s'exprimer, plus de temps pour concevoir les idées- puisqu' il n'y a relativement pas de limites de temps ; il n'y pas de durées limites de séances comme est le cas dans les séances en classes . Nous pouvons constaté par exemple que dans le cas des logiciels d'apprentissage (Logitron ou tout autre didacticiel) , l'apprenant est libre de les utiliser à l'heur que bon lui semble , avec les fréquences qui l'arrange et les moments et durées d'utilisation qu'il souhaite ,donc une flexibilité qui était auparavant inaccessible , de plus grandes possibilités de précision de l'expression et de réalisation de bons résultats, puisque il y a un champ plus grand , d'exercice, et de pratique des activités langagières .

Il considère par ailleurs qu'elle favorise l'esprit de participation et d' « encouragement » puisque toutes les occasions et séances d'expression sont proportionnelles non seulement à la qualité et la nature du contexte d'utilisation, c'est-à-dire les fonctionnalités des logiciels,- l'utilisateur est amené à différentes reprises à participer activement à une activité donnée- , leur paramétrage (paramétrage de logitron) et parfois les instruments et matériel utilisés, mais également à la fréquence d'utilisation et de recours aux moyens de communication et d'apprentissage assistée par ordinateur. Il est clair que plus le recours est fréquent , plus le temps d'exercice est long , et plus les possibilités de réaliser de meilleurs résultats sont importantes, et donc une certaine forme de réconfort et de satisfaction, ce qui mène et comme résultat logique, vers une diminution de l'anxiété liée à l'emploi des langues étrangères, d'où l'encouragement ,la motivation et la progression.

⁹ Homme-machine

5.2.4- A partir de l'an 2000, l'autonomie :

L'autonomie de l'apprenant ; est ce que met en exergue entre autre Jeanine Gerbault (2002) , à partir de son travail d'enquête auprès d'institutions culturels francophones dans le monde (519 répondants) ; note que les logiciels d'apprentissage et de formation (logitron, en l'occurrence) ou d'autres , les interfaces numériques ,ainsi que les plates formes d'apprentissage en ligne (cf. retour aux différents didacticiels cités dans la liste thématique d'applications de l'annexe)la communication à distance , ou encore les espaces numériques de travail et d'apprentissage comme par exemple « Auralog », ou « BLCC », ou ce qui est connu sous le nom du « e-learning » ou « l'open and distance learning » en anglais, « encouragent l'autonomie des apprenants » , puisque comme nous l'avons brièvement cité , l'apprenant à, entre autres ,la possibilité de gérer son emploi de temps , et aussi plus de possibilités de choisir quelques modalités de son propre apprentissage , à savoir les contenus et activités , plus encore , les adapté à son niveau ,comme dans la première utilisation de « Logitron »

Nous constaterons a travers ces statistiques que l'offre de formation s'adapte mieux à la demande avec le déploiement des moyens informatiques.-touts moyens confondus- Cette demande se manifeste travers la motivation des utilisateurs quant à ces nouveaux moyens et dispositifs informatiques, ainsi que le besoin enrichissant -pour ce qui est des portails ou des espaces numériques de travail en ligne qui comme nous l'avons précisé plus haut sont constitués entres autres de didaciciels- d'accès à la culture et de contacts avec d'autres apprenants qui utilisent les mêmes portails ou plates formes par exemple.

Sur l'ensemble des personnes interrogées, deux-tiers déclarent ce désir « d'autonomie » , l'un des enjeux serait donc d'adapter ces techniques, possibilités disponibles et de les mettre au service de l'apprentissage et des apprenants, afin d'en tirer le meilleur profit.

Liens sociaux


<div data-bbox="95 436 726 929"> <p>Contenu du cours</p> <p>Méthode pilotée</p> <p>Suivez le <u>trajet préétabli</u>. Découvrez le cours par un trajet préétabli par les tuteurs.</p>  <p>Méthode personnelle</p> <p>Table des matières Fureter l' <u>aperçu total</u> du cours.</p> <p>Consulter les documents Fureter les <u>documents</u> spécifiques du cours.</p> <p>Faire les exercices Compléter les <u>exercices</u> sans suivre le trajet préétabli.</p> <p>Votre boîte à devoirs</p> </div>	<div data-bbox="742 436 1117 929"> <p>La communauté</p> <p>Vos tuteurs pour ce cours</p> <p>Les participants du cours</p> <p>Calendrier contact</p> <p>Courrier électronique</p> <p>Forum du cours</p> <p>Bavardoir du cours</p> </div>	<div data-bbox="1141 436 1476 929"> <p><i>Communication</i></p> <p>Progrès / résultats</p> <p>Centre de ressources</p> </div>
---	---	---

Figure 8 : Environnement informatique d'apprentissage BLCC : exemple d'une plate forme d'apprentissage

Différentes fonctionnalités, ressources et paramètre

5.2.5-L'accessibilité aux ressources :

Le déléguée générale à la langue française et aux langues de France ; linguiste et Professeur Cerquiglini (2004), souligne quant à lui l'accessibilité aux ressources en énorme quantité. C'est l'une des caractéristiques de ces nouveaux moyens de traitement automatique du langage, et des outils informatique d'un point de vue générale. Il évoque le double avantage des outils et moyens informatiques et notamment ceux mis au service de l'apprentissage et des traitements de la langue (cf., « Logitron », ou toute autre site pouvant relevé du domaine et être scientifiquement consacré à l'apprentissage).

D'abord leur efficacité concrète en terme d'accessibilité aux ressources : Cette concrète efficacité peut d'une part s'expliquer par rapport au nombre de ressources que l'on peut trouver ou nombre de leçons, il y a profusion de documentation électronique sur une seule et même question ou information. Dans le cas des sites que nous avons classés en annexe ou des plateformes d'apprentissage en ligne comme par exemple le centre de formation virtuelle de la « FOAD » formation ouverte à distance (il y en a plusieurs autres) ou « Je vous ai compris » de « Neuro concept », ou le cas de « Logitron », le nombre des données , informations, leçons , cours et ressources disponibles et pouvant être consultées est concrètement considérable . (cf. Retour aux sites, plateforme et portails en question) . Les ressources, données et informations sont classées d'une manière hypertextuelle. dans l'exemple de Logitron , (chaque partie contient un grand nombre de leçons et d'exercices organisés hypertextuellement,) .

D'autre part, par la relative aisance et facilité d'accès à ces ressources (du moment où ils sont classés, répertoriés et présentés en très grand nombre, et d'une manière méthodique et hypertextuelle ; dans de très petits espaces sur les interfaces des sites en question , il suffit d'y cliquer pour pouvoir les consulter) ce qui permet de gagner en temps, et en efforts, un temps et des efforts qui pourront et devront être exploités et investis pour le développement d'autres performances. Cet aspect méthodique peut également apprendre la discipline en entraînant à la recherche méthodique et ciblée .

Le second avantage, est la possibilité de mise en commun des cours, à travers les logiciels (didacticiels), et plateformes en ligne, ainsi que les différents portails des espaces numériques de travail, comme par exemple la plateforme de l'EDAF . Une mise en commun qui pourra être employée pour des objectifs et raisons pédagogiques ou parfois empiriques précises, et qui était bien entendu impossible de réaliser il y a peine deux décennies.

Un troisième avantage plus symbolique, se poserait en terme de survie des langues nationales en jouant le « jeu de la modernité » par leur disponibilité et diffusion sur des supports informatiques en ligne ou hors ligne, et leur activité sur les réseaux, et les espaces numériques ce qui rejoint un peu l'idée des enjeux stratégiques et décisifs que nous allons évoquer dans le paragraphe suivant .

5.2.6-Une deuxième lecture du progrès en question :

Un aspect encore une fois plutôt pragmatique , voire stratégique , avec des connotations politiques de la problématique du développement des moyens informatiques de traitement et apprentissage et la promotion des langues , pour qui considère la question d'un œil critique ; le fait que ce développement soit justement indissociable du rayonnement et de l'influence des langues vivantes .Les travaux de recensement des locuteurs de chaque langue sur internet , quels que soient les problèmes méthodologiques ou techniques liés à ce type d'études et de statistiques , convergent tous vers le même constat : sur les six mille langues environ du monde , seule une douzaine d'entre elles sont vraiment présentes sur la Toile - avec 98 % des pages web selon les statistiques effectuées par (Paolillo,Unesco 2005)

Ce constat ne passe pas inaperçu.Le caractère polyglotte d'internet est donc très relatif, marquant des écarts flagrants. L'état de fait en question pourra être interprété en prenant en compte certains paramètres et critères précis et intelligents, et en étant placé dans l'un de ses véritables contextes, à savoir celui de la mondialisation,(certains spécialistes préfère employé globalisation) , car la présence d'une langue sur le réseau renvoie davantage à sa position dans la hiérarchie des langues des mondes, en incitant ainsi de plus en plus de personnes à l'apprendre .La question du politique est donc au cœur du développement des langues « par et pour » les réseaux informatiques .

Ce sont en partie les même motivations qui ont incité les grandes nations, et notamment durant la guerre froide, et jusqu'à présent à déclarer dans un premier temps une compétition d'abord pour le développement et le progrès technologiques dans tous les domaines et davantage l'informatique. Et par la suite la compétition pour le développement de ces outils d'informatique linguistique ou (traitement automatique du langage) .

5.2.7-Le pour et le contre :

D'un point de vue plus général , qui a surtout peser le pour et le contre de ces nouvelles techniques et technologies, le docteur et chargée de recherche au CNRS Diki-Kidiri Marcel (2007) constate que tous ces moyes informatiques et numériques (didiacticiels , logitron) et

plateforme en ligne (cf. Annexe) espace numérique de travail, ou encore les réseaux sociaux ou de spécialistes ont un rôle capital et décisif dans « les transformations linguistiques » .

Il manifeste par contre en plus deux de leurs potentialités contradictoires : ils peuvent être un véhicule important de communication entre et à l'intérieur des communautés linguistiques, puisqu'elles facilitent et accélèrent la communication par exemple, et notamment sur la toile et les réseaux, par exemple les portails éducatifs en ligne ou les réseaux sociaux de spécialistes à savoir les divers « Listes de diffusion » ou « Relais » à titre d'exemple le « RISC » (relais d'information pour les sciences cognitives) qui est , la « Linguiste liste » « lien et facilitent les échanges de connaissances entre autres linguistique et culturels, ,pour ceux qui en ont accès, c'est ce qu'on peut constaté de plus en plus dans les milieu des jeunes apprenants, un peu partout dans le monde. Comme ils pourraient être un facteur aggravant de la marginalisation des langues, dans le cas des communautés linguistiques qui n'ont pas encore accès à ces technologies ou encore des apprenants qui ne savent pas manipuler ces outils informatiques .C'est le cas de certaines communautés en Afrique, Asie, Amérique latine et même Europe, mais beaucoup plus l'Europe de l'est, ainsi que la Russie.

5.2.8-Cependant, le rôle de l'enseignant ne s'estompe pas :

Bien entendu, les capacités de l'outil informatique , si toute fois sont performantes et innovantes, c'est à l'origine l'Homme qui les mis en place le premier, dans la poursuite de ce constat et dans son article « L'apprentissage/enseignement des langues à l'ère du numérique : tendances récentes et défis », Piet Desmet, traite la question de l' apprentissage et enseignement des langues assisté par les nouvelles technologies de plusieurs angles, ce que nous voulons très exactement souligner, c'est le fait que l'épanouissement et la montée en percé de toutes ces technologies et nouvelles voies d'apprentissage des langues ne remplacera aucunement le rôle de l'enseignant, dans ce sens :

« Il serait bien évidemment aberrant de penser un seul instant que l'ordinateur puisse remplacer l'enseignant, idée qui refait surface régulièrement. Bien au contraire, l'utilisation croissante d'ordinateurs

dans l'enseignement, ne fera qu'augmenter la demande en matière de personnel enseignant. Plus que jamais, l'enseignement des langues a besoin de professeurs enthousiastes pour encadrer et assurer le suivi des tâches faisant appel aux TIC. » (Piet Desmet,)

S'il fallait une fois de plus expliquer pourquoi, on se contentera rien que du fait que l'usage justement des Technologies en question ; nécessite au préalable des acquis fondamentales dans le domaine informatique, ces acquis sont la plus part du temps transmis par des enseignants(humains) spécialistes.

En conclusion au dernier chapitre, nous constatons une certaine convergence entre les différents spécialistes sur ce que pourrait apporter l'informatique pour les multiples traitements des langues (apprentissage, diffusion et promotion) à travers ce domaine. qu'est le traitement automatique des langues) .

Pour ce qui est de l'apprentissage, avec l'intervention de l'outil informatique, il est devenu davantage interactif, souple, et plus adaptable aux états et niveaux des apprenants. Il s'est également enrichi à travers l'accessibilité et la profusion du nombre de ressources pédagogiques- toutes catégories confondues-.ainsi que de la possibilité de mise en commun des différents contenus pédagogique.

La diffusion, l'expansion et la promotion des langues ont à leur tour largement bénéficier du numérique, de l'informatique, de la Toile et de toutes les technologies qui contribue à l'instauration de cette discipline.

Ce progrès peut tout de même avoir des points négatifs à savoir la marginalisation, mais il pose un défi et un enjeux de taille ; en rapport avec la dominance dans la hiérarchie des langues à l'échelle internationale.

Enfin, ce progrès technologique n' s'épanouit pas sans l'intervention et l'accompagnement de l'enseignant, ce qui revient à dire qu'il y ou qu'il devrait y avoir complémentarité entre l'Homme et l'outil technologique.

Conclusion

Force est de constater, au terme de modeste travail que sciences du langage et informatique entretiennent d'étroits rapports caractérisés par une permanente interaction, à caractère interdisciplinaire et qui donne naissance à une troisième discipline, l'ingénierie linguistique.

Des liens justifiés d'une part par un certain nombre de points communs existants entre ces deux domaines, à commencer par les travaux de Chomsky et notamment sa hiérarchie issue de la grammaire générative qui a permis de mettre en place ; une fois informatisée, les assises de la théorie des langages formels ; branche des mathématiques, logique et informatique dans laquelle sont étudiés les langages artificiels ainsi que leurs propriétés, comme les langages de programmation informatique. Une nouvelle aire est née annonçant les débuts du traitement automatique du langage.

A travers l'entrée en matière, et en ayant présenté les différentes appellations du domaine, ainsi que la liste thématique des applications et famille des outils linguistiques du traitement automatique de l'information, nous avons montré que le traitement automatique est une discipline interdisciplinaire faisant appel à plusieurs autres disciplines et celui-ci peut être exploité par différents domaines et qu'il reprend les mêmes étapes d'analyse humaine du langage, mais d'une manière informatisée.

Notre sélection des applications et outils d'ingénierie linguistique étudiés est venue dans la poursuite de l'interactivité qui a fait naître cette nouvelle discipline.

En ayant commencé par l'un des outils de traitement automatique les plus utilisés et notamment par le public d'apprenants, nous avons attiré l'attention sur une situation alarmante, présentant le nouveau rapport à l'erreur qu'engendre l'emploi grandissant des correcticiels qui ne sont pas toujours performant et de voir comment il est possible d'éviter ce genre de dépendance négative à ces logiciels de telle sorte à bien garder à l'esprit qu'il ne sont qu'un moyen constituant une aide à la rédaction et non une fin en soi.

L'apport des correcticiels actuels dans la pédagogie de la langue française est minime, et leur utilisation doit être bien réfléchie, sous peine de conduire à une aggravation des erreurs, surtout dans

le cas des élèves qui connaissent de réelles difficultés en orthographe ou alors en syntaxe, d'où cette comparaison et emploi du terme « dépendance ». Pour cette raison, il est essentiel que les élève ou toute autre apprenant aient pu en mesurer quelques limites, et en comprendre le mécanisme du fonctionnement.

Il ne demeure pas moins que ces derniers pourraient s'avérer très utiles dans la mesure ou comme nous l'avons cité il sera nécessaire avant tout recours à ces logiciels d'adopter une stratégie d'emploi et des plans ou/et méthodes pédagogique précises. Cette dernière pourrait être mise en place par les enseignants ou toute autre spécialiste.

Le fait d'avoir mis en exergue le caractère interdisciplinaire de l'ingénierie linguistique en ayant expliqué pourquoi et comment est-elle pluridisciplinaire et en montrant quels étaient les résultats de cette interactivité scientifique interdisciplinaire, nous a permis de montrera quel point il est intéressant et important de créer un dialogue entre les différentes disciplines.

Dans le cas du traitement automatique du langage ;ce dialogue et cette interactivité se justifie par des points en communs existants entre l'informatique et la linguistique, auxquelles se joint l'analyse des textes pour aboutir à l'analyse automatique des textes et de corpus via les techniques de statistique et statistique lexicale .

Les logiciels d'analyse automatique lexicale et documentaire Hyperbase et Tropes sont des exemple par excellence .

A l'heure qu'il est, les applications et outils du traitement automatique du langage, ne sont pas toujours , ni tout à fait performant , nous l'avons constaté à travers les correcteurs automatiques , mais également les logiciels de traductions automatique .A son tour domaine et application interdisciplinaire , à l'instar de toutes les autres et la plus anciennes des applications du TAL, la , rencontre tout comme la traduction humaine , des problèmes et lacunes. L'un des principaux problèmes est celui de l'ambigüité qui se pose sur les différents niveaux du langage.

Pour ce qui est de l'apprentissage, avec l'intervention de l'outil informatique, il est devenu davantage interactif, souple, et plus adaptable aux états et niveaux des apprenants. Il s'est également enrichi à travers l'accessibilité et la profusion du nombre de ressources pédagogiques- toutes

catégories confondues-.ainsi que de la possibilité de mise en commun des différents contenus pédagogique.

La diffusion, l'expansion et la promotion des langues ont à leur tour largement bénéficié du numérique, de l'informatique, de la Toile et de toutes les technologies qui contribuent à l'instauration de cette discipline.

Ce progrès peut tout de même avoir des points négatifs à savoir la marginalisation, mais il pose un défi et un enjeu de taille ; en rapport avec la dominance dans la hiérarchie des langues à l'échelle internationale .

Enfin, ce progrès technologique n' s'épanouit pas sans l'intervention et l'accompagnement de l'enseignant, ce qui revient à dire qu'il y ou qu'il devrait y avoir complémentarité entre l'Homme et l'outil technologique.

Le débat concernant les nouvelles technologies et notamment celles du traitement automatique du langage peut être élargi de plusieurs points de vue .Ce qu'il y a lieu de noter , est que le domaine en question à l'instar de tous les savoirs , en en perpétuel progression .

Cette progression est porteuse de nombreux défis certes, mais également de perspectives pouvant toutes être « en principe » au service du bien-être de l'humanité d'un point de vue général. Ce domaine à offert à l'homme des possibilités qui lui était auparavant inaccessibles, et a parmi une ouverture vers d'autres domaines plus vaste ;créant ainsi une interdisciplinarité enrichissante.

REFERENCES BIBLIOGRAPHIQUES

OUVRAGE :

BIBER Douglas. « *Variation across speech and writing* »¹ Cambridge University Press, 1988.

CHOMSKY Noam . « *Syntactic Structures* » .La Hay : Mouton ,1957 .

CHOMSKY Noam « *Aspect of the theory of Syntax* » Cambridge University. Press, 1965 .

CHOMSKY Noam « *Lectures on Government and Binding* » Dordrecht ,Foris, 1981

PIERRE Jocelyn . *La langue au cœur du numérique : les enjeux culturels des technologies de la langue* . [2007] [en ligne], Rapport présenté par et à la Délégation générale à la langue française et aux langues de France, URL : www.dglf.culture.gouv.fr > consulté le [du 04/10/2009] .

COHEN Rachel . *Les jeunes la découverte de l'écrit et l'ordinateur*. PUF, 1987.

COHEN Rachel . *Quand l'ordinateur parle* . Paris PUF ,1992.

CONDAMINE Anne *Sémantique et Corpus*, Paris Hermès Science , 2005

SABAH Gérard . *L'intelligence artificielle et le langage* .Paris ; Hermès, 1984.

¹ Les titres d'ouvrages entre guillemets sont ceux des ouvrages en Angalis .

HARRIS Z. « *Methods in structural Linguistics* ».University of Chicago Press ,1971.

JACOBSON Roman « Essais de linguistique générale : Les fondations du langage » Paris, Editions de Minuit, 1963.

ARTICLES SCIENTIFIQUES :

GLANTZ H.T. « On the recognition of information with a digital computer » in *Journal des ACM* . 1957. Vol 4 ,n °2 . P. 178-188 .

DOLL Frédéric et COULOMBE Claude « L'avenir des correcteurs grammaticaux : un point de vue industriel » in *Bulag* . 2004 . N ° 29 .

HENRI « une approche cognitive de la correction automatique des fautes de syntaxe » in *Bulag*. 2004. N ° 29 .P. 85-105.

HARRIS Z« *Coocurrence and Transformations on Linguistic Structure* » in *Langage* ,1957 . N 33 , P 283-340 .

GERBAULT. « Technologies de l'information et de la communication et diffusion du français : usages, représentations, politiques. » In *Alsic*, vol. 5 n° 2, 2002. p. 183-207 [en ligne] http://alsic.unstrasbg.fr/Num09/gerbault/alsic_n09-pra1.htm consulté le [10/05/2010]

KERN Richard. « La communication médiatisée par ordinateur en langue : recherches et application récentes aux USA » In *Le français dans le monde. Recherches et applications*, 2006. n° 40, P . 17-29 .

LAB Frédérique « La Traduction Automatique » in *Le Bulletin de l'EPI*, 1988 n° 52 P165-170

LEON Jacqueline « Le traitement automatique des langues » in *Histoire Épistémologie Langage* 2001. P. 3-6

BERTEN Fernand . « Correcteurs orthographiques et enseignement du français » [en ligne],
URL <http://users.skynet.be/ameurant/francinfo/correcteur/correcteur.html>
consulté le [15/5/2010]

GENTILHOMME Yves « L ambiguïté est-elle une tare ? » in *Bulag* 1996 n ° 21 .

DESMET Piet « L'apprentissage/enseignement des langues à l'ère du numérique : tendances récentes et défis », article accepté

CARDOT Jerome « Introduction au traitement automatique de la langue naturelle » [en ligne] URL
<http://geai.univ-brest.fr/~cardot/CMntic-ha/node52.html> consulté le [20/01/2010]

CERQUIGLINI Bernard « Les outils informatique » In *Les études françaises. Séminaire international*, ADPF 2004.

TABORY Robert. « *International Conference of Computational Linguistics* »* 1961 , USA .

KOUASSI Roland Raoul « *La Traduction automatique* » [en ligne], URL
http://www.ltml.ci/files/articles4/article_traduction_automatique.pdf consulté le du [12/04/2010]

BIBEAU Robert « Ils apprennent à lire à l'aide de l'ordinateur . L'intégration des TIC en classe de français » [en ligne] URL source : http://netia59.ac-lille.fr/ref/pedagogie/Robert_Bibeau/ecrit.html. Consulté le [25/04/2010]

ORGANISME ET CONSTITUTIONS :

PAOLILLO John & al. *Mesurer la diversité linguistique sur Internet*. UNESCO, 2005.

DIKI-KIDIRI Marcel. *Comment assurer la présence d'une langue dans le cyberspace ?* UNESCO, 2007.

RAPPORTS DE RECHERCHE

TANNIER .X « Traitement automatique du langage naturel pour l'extraction et la recherche d'information ». Rapport de recherche , Saint-Etienne , France ,2006 .

GHENIMI Abdelwafi « Traduction Automatique » Mémoire de magistère, ILE , Oran ; 2000

TABET AOUL Zoulikha « Approche de l'écriture Quignardienne par l'analyse du discours et la démarche lexico-statistique » ILE , ORAN 2005.

DICTIONNAIRES :

Patrick Charaudeau et Dominique Maingueneau .Paris .*Dictionnaire d'Analyse du Discours* Seuil, 2002

Bibliographie supplémentaire (complémentaire) :

Nous avons eu recours durant ce travail a des sources bibliographiques qui nous pas directement fait office de références , c'est-à-dire dont nous avons pas puisé pour la rédaction du mémoire, mais qui nous généralement enrichi et donc indirectement contribué à l'élaboration de ce travail.

Annexes

8.1 Annexe 1 :

Apprentissage des langues :

Ace I Learner : [De Wilde CBT](#)

Assistant : [NEURO CONCEPT](#)

Click & speak : Infogrammes

Createst - Evalutest : [NEURO CONCEPT](#)

French Lesson : [EURASAM](#)

Je vous ai compris : [NEURO CONCEPT](#)

L.E.M. (Laboratoire d'Enseignement Multimédia) : [NEURO CONCEPT](#)

Nego Sphère : [NEURO CONCEPT](#)

Newspeak : [ASCOLA](#)

Phone-in : [NEURO CONCEPT](#)

Pour tout dire : [NEURO CONCEPT](#)

Speaker auteur : [NEURO CONCEPT](#)

Trefle : [ASCOLA](#)

Talk to me : [AURALOG](#)

Tell me more : [AURALOG](#)

- Dictionnaire électronique

Cordial 7 : [SYNAPSE](#)

Dicologique : [MÉMODATA](#)

Dictionnaire intégral : [MÉMODATA](#)

Duplex : [BILINGUA INGÉNIERIE LINGUISTIQUE](#)

Eurodico : [SOFTISSIMO](#)

Hugo plus : [SOFTISSIMO](#)

Ideoptima : [MÉMODATA](#)

Lexidiom : [MÉMODATA](#)

Myriade : [DIAGONAL](#)

Semiographe : MÉMODATA

XTS (modules Xerox TermFinder et Xerox TermOrganizer) : [XEROX-MKMS](#)

Recherche en langage naturel :

Darwin : [SINEQUA](#)

Dig Out For You : [ARISEM](#)

Illico : [PROLOGIA](#) .

Intuition : [SINEQUA](#)

Kombe : [PROLOGIA](#)

LIRIX XRCE : [Xerox Research Centre Europe](#)

Noematics : [NOEMATICS](#)

Recherche SAP : [NOEMATICS](#)

Search'97 : [VERITY France](#)

Search'97 Agent server : [VERITY France](#)

Search'97 Information Server : [VERITY France](#)

Search'97 Personal : [VERITY France](#)

Spirit : [T-G.I.D.](#)

Terminology Manager : [LEXIQUEST](#)

Tool Kit For You : [ARISEM](#)

Tropes : [ACETIC](#)

Retour à la page [Outils](#)

- Gestion électronique de documents et de l'information :

AskOnce : [XEROX MKMS](#)

Cin Doc : [CINCOM](#)

Cincom Smalltak : [CINCOM](#)

Cin Web : [CINCOM](#)

Darwin : [SINEQUA](#)

Intranet Suite For You : [ARISEM](#)

Information Miner For You : [ARISEM](#)

Intuition : [SINEQUA](#)

Knowledge Network : [FULCRUM](#)

Lexiguide : [LEXIQUEST](#)

Lexirespond : [LEXIQUEST](#)

Leximine : [LEXIQUEST](#)

Lex : [LCI](#)

Lexpro : [LCI](#)

Noematics : [NOEMATICS](#)

Recherche SAP : [NOEMATICS](#)

Route For You : [ARISEM](#)

Search Server : [FULCRUM](#)

SERware : [SER France](#)

Taurus : [D.C.I.](#)

Taurusweb : [D.C.I.](#)

Théo-FINDER : [THÉOROM PRODUCTIONS](#)

Tropes : [ACETIC](#)

Zoom : [ACETIC](#)

- Reconnaissance et synthèse vocales :

Dial & play mail : [ELAN Informatique](#)

Dragon Naturally Speaking : [MYSOFT](#)

Dragon Dictate Express : [MYSOFT](#)

Holmes : [HOLISTIQUE COMMUNICATIONS](#)

Model Creator : [HOLISTIQUE COMMUNICATIONS](#)

Proverbe speech platform : [ELAN Informatique](#)

Quick Translator : [SOFTISSIMO](#)

Reverso Voice : [SOFTISSIMO](#)

Reverso Pro : [SOFTISSIMO](#)

Speech engine : [ELAN Informatique](#)

Speech cube : [ELAN Informatique](#)

Tell me more : [AURALOG](#)

Tetravox : [VECSYS](#)

Théo-TALK : [THÉOROM PRODUCTIONS](#)

Voice Type dictée : [IBM France](#)

Via Voice : [IBM France](#)

Voice Type : [SOFTISSIMO](#)

Retour à la page [Outils](#)

- Traduction automatique :

Duplex : [BILINGUA INGÉNIERIE LINGUISTIQUE](#)

L&H Power Translator Pro : [LERNOUT & HAUSPIE](#)

Power Translator : [SOFTISSIMO](#)

Quick Translator : [SOFTISSIMO](#)

Reverso Voice : [SOFTISSIMO](#)

Reverso Pro : [SOFTISSIMO](#)

Systran : [SYSTRAN](#)

TeXnology Showroom : [Xerox Research Centre Europe](#)

Webview 2000 : [SOFTISSIMO](#)

XTS (module Xerox Web@ssistant) : [XEROX-MKMS](#)