

II.1. Introduction

Traduction automatique classique qui comprend des règles faites par des linguistes dans des grammaires reposent sur l'utilisation de nombreuses règles linguistiques et de millions d'entrées de dictionnaires pour chaque paire de langues

Dans ce chapitre nous présentons traduction automatique à base de règles et les défieront approche de traduction et on fin les étapes de traduction

II.2. La traduction à base de règle (Rule-Based Machine Translation (RbMT))

Les logiciels de traduction automatique à base de règles reposent sur l'utilisation de nombreuses règles linguistiques et de millions d'entrées de dictionnaires pour chaque paire de langues.

Le logiciel parcourt le texte à traduire et crée une représentation intermédiaire à partir de laquelle la traduction est générée. Ce processus nécessite l'utilisation de dictionnaires volumineux, de données syntaxiques, morphologiques et sémantiques, et de nombreuses règles linguistiques. Le logiciel utilise ces règles pour transférer la structure grammaticale du texte source dans le texte cible (traduction).

Les traductions sont construites à partir de dictionnaires gigantesques et de règles linguistiques sophistiquées. Les utilisateurs peuvent améliorer la qualité de traduction en intégrant leur propre terminologie ; celle-ci sera prise en compte au cours du processus de traduction. L'utilisation de ces dictionnaires utilisateurs prime sur les paramètres de base du logiciel.

Dans la plupart des cas, ce processus de personnalisation se fait en deux étapes : un investissement initial qui améliore sensiblement la qualité de traduction, puis des investissements continus pour améliorer la qualité de manière incrémentale. Les logiciels de traduction automatique à base de règles

permettent d'atteindre un niveau de qualité très élevé, mais le processus de personnalisation peut se révéler long et onéreux.[sys]

II.3. Les avantages de RBMT [sys]

- ♦ Traductions prévisibles et homogènes
- ♦ Bonne qualité des traductions généralistes
- ♦ Connaissance des règles grammaticales
- ♦ Performance et robustesse
- ♦ Cohérence entre les versions

II.4. Les différentes approches

La Traduction Automatique désigne le fait de traduire un texte d'une langue source vers une langue cible sans aucune intervention humaine. Depuis ses premiers pas, deux grandes approches coexistent : une première approche, dite experte, fondée sur les connaissances d'experts humains et une seconde approche, dite empirique, qui extrait les connaissances à partir de quantités importantes de données textuelles.

Le processus naturel de traduction peut être vu comme la succession de trois étapes. La première étape est l'analyse du texte source où l'être humain cherche à définir comment le structurer ou le segmenter pour le traduire. Une fois le texte source structuré, la deuxième étape est le transfert vers la langue cible. Ce dernier consiste à appliquer des règles permettant de passer de la langue source à la langue cible. Enfin, une fois le transfert effectué, la troisième étape est la génération d'une traduction en langue cible du texte source initial. Le triangle de Vauquois schématisé dans la figure 2.1 représente ces trois étapes d'analyse, de transfert et de génération. [figure 1.1]

Chacune des approches envisagées, quelle soit experte ou empirique, traite ces étapes de façon particulière avec des niveaux d'analyse, de transfert et de génération différents. Nous décrivons dans les sections suivantes les principales

méthodes de TA issues de chacune des deux approches.

II.4.1. L'approche experte

L'approche experte repose sur les connaissances professionnelles des êtres humains. Ces connaissances sont utilisées durant chaque étape du processus de traduction. Lors de l'analyse d'abord, la phrase source est segmentée en unités liées les unes aux autres. Selon le niveau d'analyse, ces unités peuvent être morpho-syntaxiques, grammaticales ou encore sémantiques. La phase de transfert se charge ensuite de convertir les unités sources en unités cibles en utilisant notamment des règles et des dictionnaires bilingues. Le module de génération, enfin, transforme les unités cibles en texte cible. Les trois principales méthodes dérivées de cette approche sont la Traduction Automatique (TA) directe, la TA à base de règles de transfert et la TA fondée sur l'interlangue. Ces méthodes sont détaillées dans les sections suivantes:

II.4.1.1. TA directe

L'approche directe de la TA correspond à la base du triangle de Vauquois. La traduction s'opère ici mot à mot et ne nécessite qu'une analyse partielle de la phrase source. Le processus de traduction repose essentiellement sur l'utilisation d'un grand dictionnaire bilingue qui va spécifier pour chaque mot source un ensemble de règles permettant de le traduire. Les mots de la phrase source, une fois traduits, sont ensuite réordonnés à l'aide de simples règles afin de générer une traduction. Le procédé est illustré par la figure 2.2

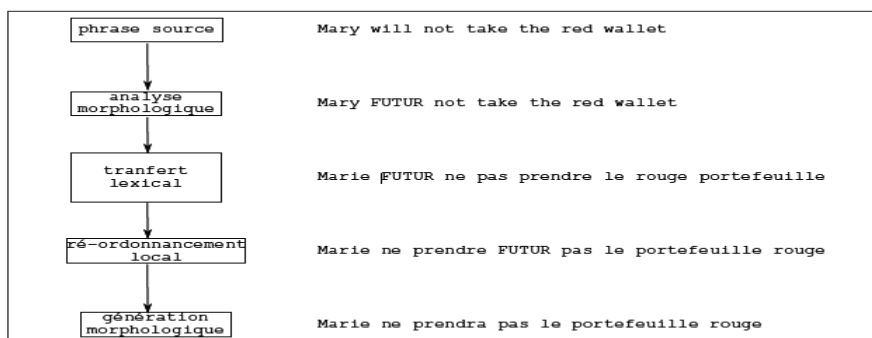


Figure 2.2TA directe [Jur 06]

L'approche directe de la TA est simple à mettre en place en ce sens qu'elle ne nécessite pas de connaissances approfondies ni de la langue source, ni de la langue cible. Elle ne fait appel qu'à des règles primaires. L'analyse seulement partielle de la phrase source présente l'avantage de permettre au système de toujours fournir une traduction de la phrase source même si cette dernière n'est pas grammaticalement correcte. En effet, les connaissances utilisées ne sont pas suffisamment importantes pour détecter un phénomène incorrect grammaticalement. Si une structure grammaticale n'est pas reconnue lors de l'analyse de la phrase source, elle est écrite directement en mots cibles sans réordonnement.

Cependant, le manque d'analyse de la phrase source ainsi que l'absence de lien entre les représentations source et cible du texte à traduire diminuent fortement la fiabilité des résultats obtenus. De plus, la TA directe ne possédant pas de grammaire explicite pour la langue cible, contrairement aux approches dites indirectes (règles de transfert, interlangue), la bonne construction de la phrase, du point de vue de la langue cible n'est pas garantie.

Dans les années 80, des systèmes de traduction tels que SYSTRAN [Wheeler 84] à ces débuts et PAHO [Vasconcellos 85] utilisaient l'approche directe. Aujourd'hui, du fait de sa simplicité et de part la complexité des langues, l'approche directe n'est plus guère utilisée pour la TA grande échelle. Toutefois, elle peut s'avérer efficace et appropriée dans le cadre d'applications spécifiques pour lesquelles le vocabulaire est restreint et standardisé et où une substitution de mots peut suffire à produire une traduction correcte. Nous pouvons notamment citer le système METEO [Chandioux 96] qui traduit des bulletins météorologiques avec un vocabulaire restreint d'un millier de mots mais aussi le grand fabricant d'appareils médicaux Medtronic [Uwe 06] qui utilise un système de TA directe pour traduire dans plusieurs langues sa base de données de produits.

II.4.1.2. TA à base de règles de transfert

Contrairement à l'approche directe, l'approche à base de règles de transfert requiert un niveau d'analyse de la phrase source plus poussé. En effet, comme l'illustre la figure 2.3, la phrase source est analysée à l'aide d'un analyseur syntaxique et d'une grammaire. Cette phase d'analyse donne lieu à une représentation arborescente qui est ensuite convertie dans la langue cible. Cette étape de transfert, lors du passage de la langue source vers la langue cible fait intervenir des règles bilingues supposées être applicables dans les deux sens source-cible et cible-source.

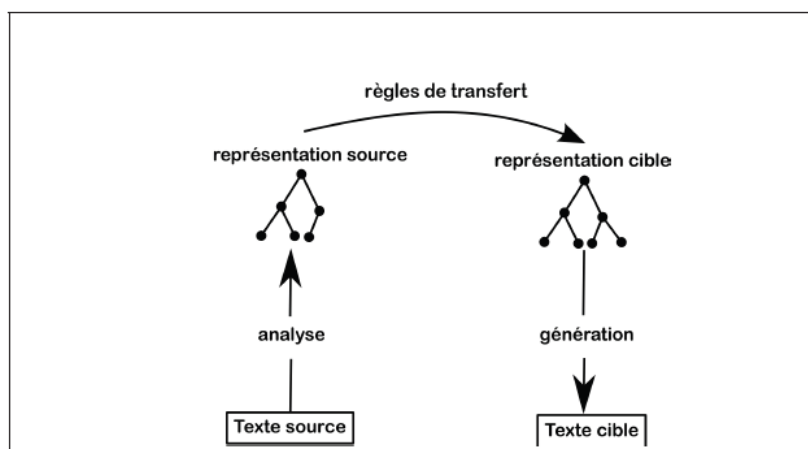


Figure 2.3 TA à base de règles de transfert [Jur 06]

Il s'agit en fait de règles de correspondance entre les nœuds de l'arbre source et les nœuds de l'arbre cible. Elles peuvent être syntaxiques, lexicales ou encore sémantiques. Elles assurent la transformation entre la représentation de la phrase source et sa représentation dans la langue cible. Enfin, une grammaire de la langue cible va générer la phrase cible à partir de cette représentation. Un cas simple de traduction à base de règles de transfert est illustré dans la figure 2.4

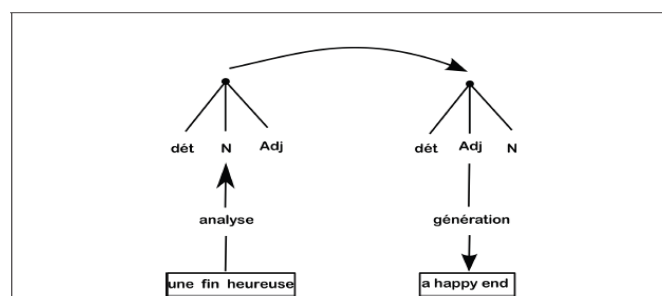


Figure 2.4 Exemple simple de TA à base de règles de transfert du Français vers

l'Anglais de la phrase "une fin heureuse"

Dans cet exemple, l'analyse faite de la phrase source est simplement syntaxique. Cependant, suivant les langues engages dans la traduction, l'analyse peut être beaucoup plus profonde. Pour une traduction de l'Anglais vers le Chinois par exemple, il est nécessaire de connaître le rôle sémantique des groupes prépositionnels afin de savoir s'il se place avant ou après le verbe en chinois. En effet, si le groupe prépositionnel exprime une entité qui bénéficie de l'action induite par le verbe, celui-ci sera placé avant le verbe. A l'inverse, s'il exprime une entité passivement impliquée dans les événements dénotés par le verbe, il sera plutôt placé après le verbe.

La TA à base de règles de transfert implique une connaissance avancée de la langue source et de la langue cible ainsi que des liens existant entre ces deux langues. L'inconvénient majeur de cette méthode réside donc en la difficulté de formaliser toute la grammaire réelle d'une langue, ainsi que les grammaires de transfert. Une phrase source trop complexe grammaticalement, par exemple, peut conduire à l'échec du processus de traduction si la grammaire de la langue source n'a pas été suffisamment décrite.

Les systèmes de TA à base de règles sont relativement simples à mettre en place et facilement adaptables pour de nouvelles paires de langues. En effet, les langues sont des modules bien séparés et indépendants. Permettre à un système de traduire vers ou depuis une nouvelle langue, revient donc simplement à ajouter une grammaire de cette nouvelle langue et des grammaires de transfert.

De part le manque de formalisme dans les grammaires, cette approche, au départ, a surtout montré de bons résultats dans des domaines restreints comme par exemple pour la réservation d'hôtels. Mais la disponibilité des grammaires et de lexiques de grande couverture, en fait de nos jours un des paradigmes majeurs en Traduction Automatique. De nombreux systèmes de traduction automatique basée sur cette approche sont disponibles aujourd'hui. C'est notamment le cas de SYSTRAN [Senellart 01], de Babelfish ou encore de PROMT pour n'en citer

que quelques-uns.

II.4.1.3. TA fondée sur une interlangue

Comme le montre la figure 2.1, l'analyse de la phrase source se fait à des niveaux plus ou moins profonds. Plus l'analyse de la phrase est poussée, plus la représentation en devient abstraite. Le but de la TA fondée sur l'interlangue est de trouver une analyse telle que le transfert des représentations de la langue source vers la langue cible n'aurait plus lieu d'être. A un tel niveau de représentation, la sortie du composant d'analyse du texte source serait directement l'entrée du composant de génération du texte cible sans passer par l'application de règles de transfert. Cette méthode tend vers une représentation du sens de la phrase indépendante de la langue utilisée d'où le terme d'approche interlangue. Et la différence par rapport à la TA à base de règles de transfert tient plus du degré de représentation que d'une réelle distinction.

La principale difficulté de cette méthode est de trouver un vocabulaire de tous les concepts possibles pour l'interlangue. En effet, pour cela il faut trouver les concepts primitifs de la représentation du sens des phrases quelles que soient les langues considérées. Les chercheurs dans ce domaine sont confrontés au problème d'identification de ces concepts primitifs. Il faut définir tous les concepts possibles selon toutes les langues et ceci peut conduire à la génération d'informations inutiles dans certains cas. Il se peut notamment qu'il y ait des différences entre certains concepts de l'interlangue qui n'ont pas lieu d'être. Par exemple, le concept frère correspond en Japonais à deux concepts selon que le frère soit une personne plus jeune ou plus vieille. Cette différence doit être spécifiée au sein de l'interlangue même si elle est inutile pour traduire du Français vers l'Anglais.

L'avantage principal que présente la TA fondée sur l'interlangue, une fois celle-ci bien définie, est la facilité à mettre en place un système de traduction pour de nouvelles paires de langues. Contrairement à la méthode à base de règles de transfert qui nécessite une grammaire pour la langue cible, une pour la source

et deux grammaires de transfert, l'ajout d'une nouvelle paire de langues dans le cas d'un système interlangue ne requiert qu'une grammaire supplémentaire pour chacune des deux langues.

Aujourd'hui, le seul système opérationnel basé sur l'interlangue est celui résultant du projet KANT (Knowledge-based Accurate Natural-language Translation) [T. Mitamura 91]. Un autre grand projet, aujourd'hui terminé, suivant cette approche était le projet DLT (Distributed LanguageTranslation) [Witkam 88] qui utilisait notamment l'Esperanto comme interlangue au sein de son système comprenant douze langues européennes.

II.4.2. L'approche empirique

L'approche empirique est née des suites de l'abondance croissante de données textuelles. De nombreux chercheurs ont vu, au travers de cette recrudescence, une porte ouverte à l'élaboration d'un nouveau type d'architectures dans lequel la connaissance nécessaire à la traduction n'émane plus d'experts linguistiques mais provient directement de données réelles. Deux grandes familles se distinguent alors : l'approche par analogie, également appelée approche à base d'exemples, et l'approche statistique. Toutes deux tirent profit des données brutes et contrairement aux méthodes de l'approche experte, ne nécessitent aucune connaissance a priori pour mettre en place un système de traduction.

II.5. À quoi faut-il faire attention avant de traduire

La traduction est une tâche très complexe. Pour le fonctionnement optimal d'un programme de traduction, il faut prendre en considération les points

suivants :

- ❖ Le texte à traduire doit être grammaticalement correct.
- ❖ Les expressions argotiques, les mots mal orthographiés, les signes de ponctuation manquants,
- ❖ les phrases complexes ou trop longues peuvent conduire à une mauvaise

traduction.

II.6. Les étapes d'une traduction [.net]

La figure suivante explique de nouveau le processus de traduction :

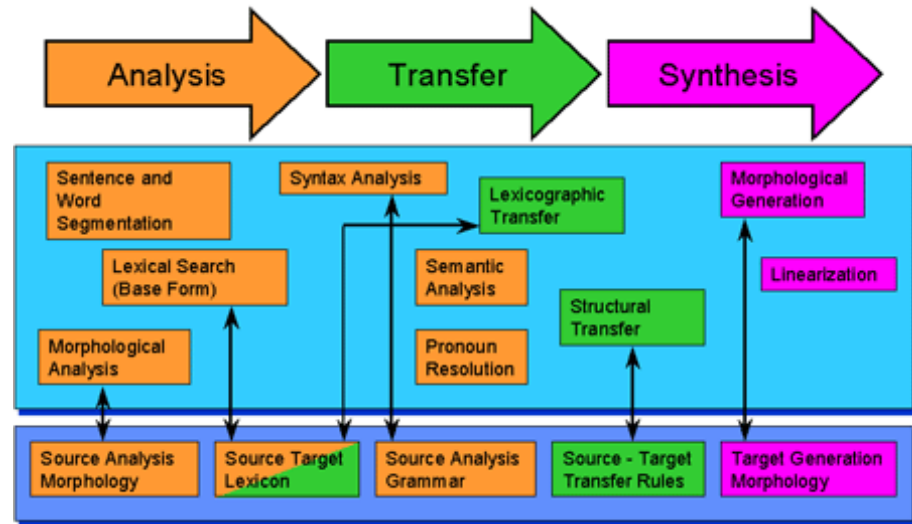


Figure 2.5 fonctionne la traduction automatique

- Lors de la traduction, le logiciel procède comme suit:
- **1^{ère} étape** : l'analyse du texte source

Le logiciel examine le texte à traduire, dit texte source, comme le ferait un traducteur humain. Il décortique la structure de chaque phrase et étudie les fonctions des groupes de mots, disséquant sujet, verbe et compléments. Tous ces éléments sont envoyés les uns à la suite des autres dans le « moteur de traduction ».

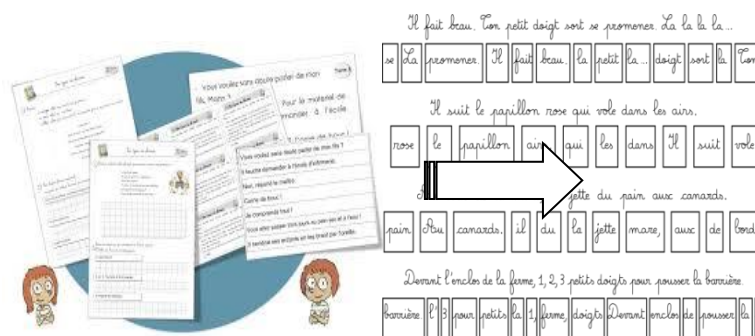


Figure 2.6 l'analyse du texte source

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 de ces mots, 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.

- **2 eme Etape : l'application des règles**

Le logiciel applique alors à tous les éléments analysés, les « règles » syntaxiques, sémantiques et morphologiques, propres à la langue source, qu'il possède dans sa base de données. Une succession d'algorithmes résout les problèmes les uns après les autres.

Pour cela, le logiciel utilise tous les indices présents dans la phrase, y compris la ponctuation et les majuscules. Ainsi, s'il rencontre le mot Paris avec une majuscule au milieu d'une phrase, il sait qu'il s'agit de la ville et non du nom commun. Si le mot est en début de phrase, il analyse alors les mots qui suivent. Dans « Paris est une belle ville », comme Paris est suivi d'un verbe au singulier et n'est pas précédé d'un article, il en déduit (peut-être un peu vite) qu'il se trouve devant un nom propre.

- **3 eme Etape : l'intervention du dictionnaire**

Pendant qu'il mouline pour appliquer les règles au texte, le logiciel pioche dans son dictionnaire afin de passer en revue tous les cas possibles pour un mot.

La richesse et la structure du dictionnaire se révèlent très importantes. L'utilisation de dictionnaires thématiques ou spécialisés enrichit une traduction de façon très nette en jouant le rôle de filtre supplémentaire. De quoi éviter certains contre-sens. Ainsi, le mot anglais *board* désigne à la fois une planche, un tableau, le fait d'être à bord d'un véhicule ou un conseil d'administration. Si le

logiciel comporte un dictionnaire spécialisé en finance et que l'utilisateur choisit de s'en servir, le logiciel interprète le mot, sans hésitation, dans le sens conseil d'administration.

Si le logiciel ne comporte pas de dictionnaire spécialisé, il a des chances de se tromper. D'ailleurs, c'est cette présence de dictionnaires spécialisés pour affiner la traduction qui fait toute la différence entre un logiciel et un service de traduction en ligne. Parfois, le logiciel propose à l'utilisateur de choisir lui-même entre différents homophones. De plus, certains logiciels évolués sont capables d'apprendre. On dit qu'ils possèdent une « mémoire de traduction », qui leur permet d'enregistrer des textes qu'ils ont déjà traduits et de les identifier lors d'une autre traduction. Il est aussi parfois proposé à l'utilisateur d'enrichir lui-même le dictionnaire.

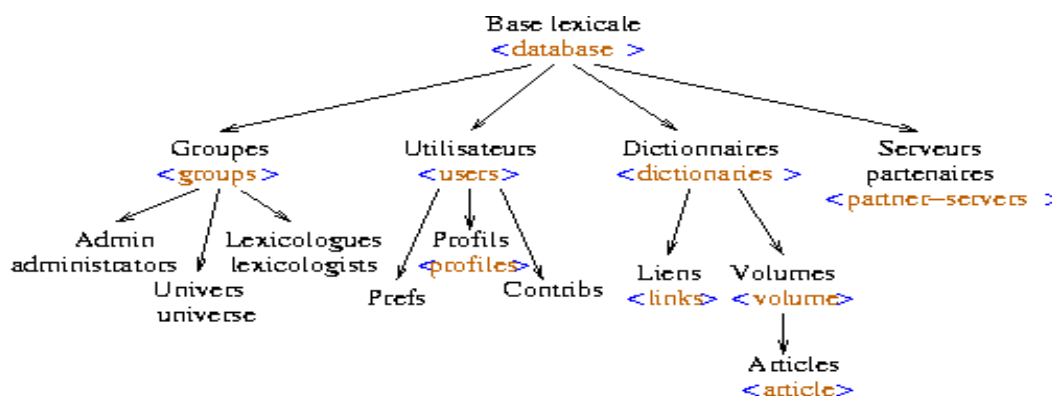


Figure 2.7 : Le transfert lexical

- **Etape 4 : le transfert vers le texte cible**

Une fois que le logiciel a décortiqué et analysé chaque portion de phrase ou chaque mot, qu'il les a traduits dans ce qu'il estime être le bon sens, il lui reste à effectuer ce qu'on appelle le « transfert » pour reconstituer le texte cible, traduction complète du texte source.

De nouvelles difficultés apparaissent dans la construction de cette

synthèse. Car le logiciel doit maîtriser aussi bien les subtilités de la langue source que celles de la langue cible. Afin de traduire, par exemple, l'expression française « il pleut des cordes » par son équivalent anglais « it's raining cats and dogs » (les Anglais font, eux, pleuvoir des chats et des chiens) et non pas « it's raining ropes » .

C'est généralement là que pèchent le plus les logiciels de traduction automatique. La qualité de la traduction dépend donc de la simplicité syntaxique du texte source et de l'absence de fautes d'orthographe. Le logiciel ne comprend pas le sens du texte, il applique les règles phrase par phrase, sans vision globale. En revanche, il ne traduit pas mot à mot, mais il fait l'analyse de la phrase en entier.

Exemple anglais → français

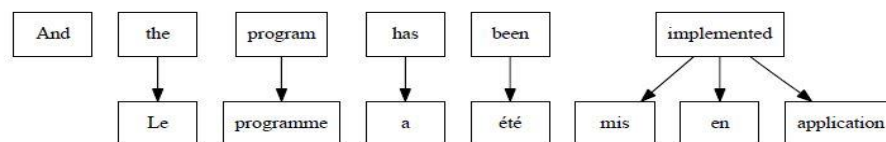


Figure 2.8 : le transfert structurel

II.7. La vérification automatique d'orthographe :

La correction automatique d'orthographe a été conçue pour éliminer automatiquement les fautes les plus évidentes et les plus gênantes pour permettre une traduction plus précise. Lorsque plusieurs corrections sont possibles, il vous indique que c'est à vous de choisir.

La mise en place d'un système de correction dépend beaucoup du type de corpus que l'on veut corriger. C'est pourquoi nous ferons état de la grande

diversité de ces derniers dans le paragraphe qui suit avant de dresser un panorama des travaux réalisés dans le domaine

II.8. Conclusion

Dans ce chapitre, nous avons essayé de construire une idée sur le domaine de traduction automatique, ou bien la traduction assistée par ordinateur, ainsi que nous avons présenté les différents approches utilisés dans ce domaine, en suite et pour des raison appropriés pour notre travail , nous avons choisis La traduction à base de règle (Rule-Based Machine Translation (RbMT) qui nous avons la plus détaillé pour l'utiliser dans notre travail .