

CHAPITRE I : INTRODUCTION

1.1 QU'EST CE QUE LA TALN ?

L'objectif du traitement automatique du langage naturel (TALN) est la conception de logiciels, capables de *traiter de façon automatique des données linguistiques*, c'est-à-dire des données exprimées dans une *langue* (dite "naturelle"). Ces données linguistiques peuvent être des *textes écrits*, ou bien des *dialogues écrits* ou *oraux*, ou encore des *unités linguistiques* de taille inférieure à ce que l'on appelle habituellement des textes (par exemple : des phrases, des énoncés, des groupes de mots ou simplement des mots isolés).

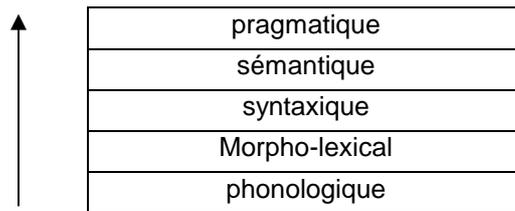
1.2 DOMAINES D'APPLICATION DU TALN ?

Parmi les diverses applications possibles, on peut citer principalement :

- La traduction automatique.
- Le résumé automatique des textes.
- La réalisation d'interface en langue naturelle.
- L'indexation automatique de documents.
- L'enseignement assisté par ordinateur.
- Serveurs vocaux
- Correction lexicale et syntaxique
- Outils d'aide à la rédaction
- Dictée vocale
- ... etc.

1.3 LES NIVEAUX DE TRAITEMENT :

Pour traiter le langage naturel, on a besoin d'informations coordonnées et pertinentes sur la langue à des niveaux divers. Le plus souvent on a recours à cinq niveaux de connaissances sur une langue: phonologique, morpho-lexical, syntaxique, sémantique et pragmatique. Ces niveaux se superposent, chacun apportant des problèmes spécifiques à résoudre relatif à un niveau donné. Cela nous donne la hiérarchie suivante :



Le *niveau phonologique* : La machine doit reconnaître les signaux acoustiques (domaine de la phonétique) et les identifier en tant que mots.

Exemple : « Vincent mit l'âne dans un pré ou vingt cent mille ânes dans un pré ».

Plus précisément, il s'agit de reconnaître dans le flot sonore les unités acoustiques élémentaires (phonèmes). La difficulté est que la forme acoustique d'un phonème varie selon plusieurs facteurs : le sexe, l'âge, la région, la fatigue, la peur, l'intensité de la parole, etc...

- Le *niveau morpho-lexical* : il concerne l'étude de la formation des mots et de leur variation de formes.
Exemple : « Aujourd'hui, parce qu'il a acheté un micro, Paul a payé la T.V.A ».
- Le *niveau syntaxique* : il s'intéresse à l'agencement des mots et à leurs relations structurelles.
Exemple : « Le boucher sale la tranche ».
- Le *niveau sémantique* : se consacre au sens des énoncés.
Exemple : « Elle a mangé du poisson avec des amis ».
- Le *niveau pragmatique* : prend en compte le contexte d'énonciation.
Exemple : « La mousse aux fraises est sur la table de l'avocat ».

1.4 L'ANALYSE VS LA GENERATION :

En fonction des objectifs attendus d'un système de TALN, on peut le classer dans l'un des domaines suivants : *analyse* et *génération*.

Le processus d'*analyse* consiste à démarrer de la structure de surface (locution ou texte écrit) pour arriver à la structure profonde équivalente.

Le processus de *génération* est le processus inverse : il consiste à passer de la structure profonde à la structure de surface.

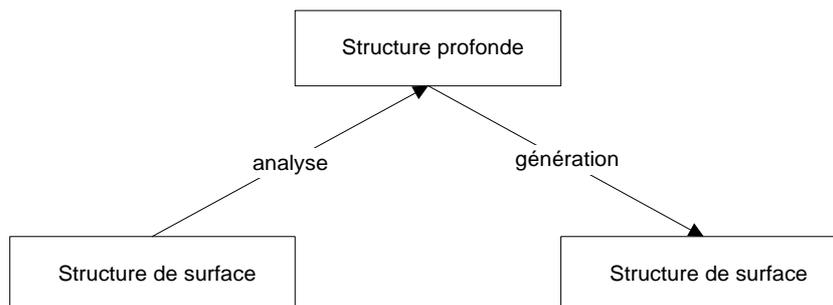


Figure 1.2 Analyse-Génération

Remarque : il est possible qu'une application de TALN appartienne en même temps aux deux domaines : analyse et génération; c'est le cas par exemple d'une application de traduction automatique.

1.5 LE PROBLEME CENTRAL DU TALN : L'AMBIGUÏTÉ.

Parmi les problèmes les plus importants du traitement des langues naturelles, on trouve en premier rang le problème de l'ambiguïté (on dit qu'il y a ambiguïté lorsqu'il y a plus d'une interprétation pour une structure linguistique donnée). En effet, l'être humain parlant d'une langue donnée, passe à côté de la plupart des ambiguïtés, puisqu'il est très doué en la manière de les résoudre : il utilise ses connaissances qu'il a accumulées à propos du monde et du contexte. Mais, malheureusement, les machines ne sont pas encore aussi compétentes. Les applications du TALN sont souvent confrontées aux ambiguïtés à tous les niveaux d'analyse.

Exemple : **Le secrétaire vole des livres**

Cette phrase comporte cinq mots ambigus, comme le montre le tableau suivant :

N°	Mot	Sens
01	Le	Article ou pronom
02	Secrétaire	Homme, meuble, oiseau
03	Voler	Planer, dérober
04	Des	Article contracté ou article partitif
05	Livre	Bouquin, monnaie, poids

Elle correspond donc à 108 possibilités. Le programme d'analyse devra lever ces ambiguïtés de diverses façons complémentaires, en considérant des règles morphologiques, syntaxiques, sémantiques, ... etc.