

TAL DEVOIR N°3

Introduction

Créer une machine qui parle est un rêve depuis toujours pour les chercheurs et cela depuis les débuts de l'informatique, pourtant ce fantasme est loin d'être gagné, en effet malgré de nombreux travaux de recherche le langage reste extrêmement complexe à programmer sur un ordinateur.

Cela commence dans les années 50 avec la mise en place de la traduction automatique après de nombreuses tentatives, vient dans les années 70-80 les « systèmes experts » où l'on voit se développer l'intelligence artificielle, le TALN garde encore aujourd'hui de nombreux concepts datant de cette époque. Nait ensuite dans les années 90 l'ingénierie linguistique, celle-ci se met au service de la « fouille de textes » (ce qui aide à classer, traiter ou extraire les contenu des textes.)

Dans un premier temps nous étudierons les liens entre la linguistique et l'informatique ainsi que sur les niveaux d'analyse, dans un second temps nous passerons en revue les différents niveaux. Enfin, nous nous focaliserons sur les principales tâches de la « fouille de textes »

Traitement automatique du langage naturel

1) Introduction

Parler reste le propre de l'homme, aucun ordinateur n'est encore capable de tenir une conversation courante. Les langues humaines nommées aujourd'hui langues naturelles s'opposent aux langues artificielles. Les objets des sciences du langage est l'étude des langues naturelles et donc de la « faculté de langue » ; les linguistes tentent de dégager les règles communes qui régissent les langues, ils se contentent d'observer et non de contrôler les langues naturelles. Pour cela l'informatique joue un rôle de plus en plus important de part le TALN (traitement automatique du langage naturel). L'informatique naît quant à elle d'une tradition mathématique et logique, en l'introduisant dans le domaine de la linguistique cela revient à se poser toujours la même question : *Quels sont les données et traitements pertinents de ce domaine, comment les coder ?* Les linguistiques caractérisent les données de maîtrise d'une langue, les informaticiens eux contribuent à les modéliser.

2) Histoires croisées de la linguistique et de l'informatique

Points de repère importants concernant la linguistique :

- Publication de la grammaire générale et raisonnée de Lancelot et Arnaud en 1660
- Règne de la linguistique comparative et historique (on compare les langues et on en déduit des règles générales cela aboutit à un rapprochement entre plusieurs langues)
- 1916 : Publication du cours de linguistique générale, Saussure
- Linguistique structurale : années 30-40
- 1956 : naissance du terme « intelligence artificielle »
- La double articulation par Martinet
- 1966 : naissance d'Elizia
- La grammaire universelle et la théorie de l'innéisme par Chomsky
- **Programme SHRDLU par T. Winograd en 1972**

- La sémantique formelle (70-80)
- La linguistique de corpus (1990)

Points de repère important Concernant l'informatique :

- XIXe s : Période de progrès en mathématiques et naissance de la logique booléenne + automatisation du travail
- La machine de Turing(1936) → naissance de l'informatique
- Rapport de plan de construction d'un ordinateur (1945) par V.Neumaan
- 1956 : naissance du terme « intelligence artificielle »
- 1966 : naissance d'Elizia
- Programme SHRDLU par T.Winograd en 1972
- La sémantique formelle (70-80)
- La linguistique de corpus (1990)

3) Les niveaux d'analyse du langage

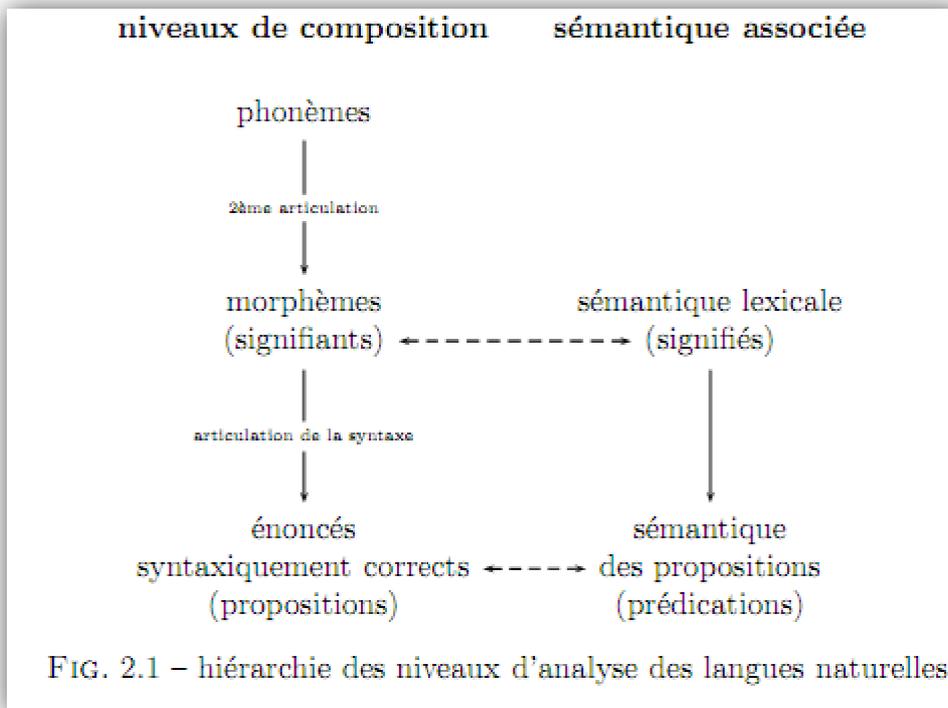


FIG. 2.1 – hiérarchie des niveaux d'analyse des langues naturelles

Ce schéma montre :

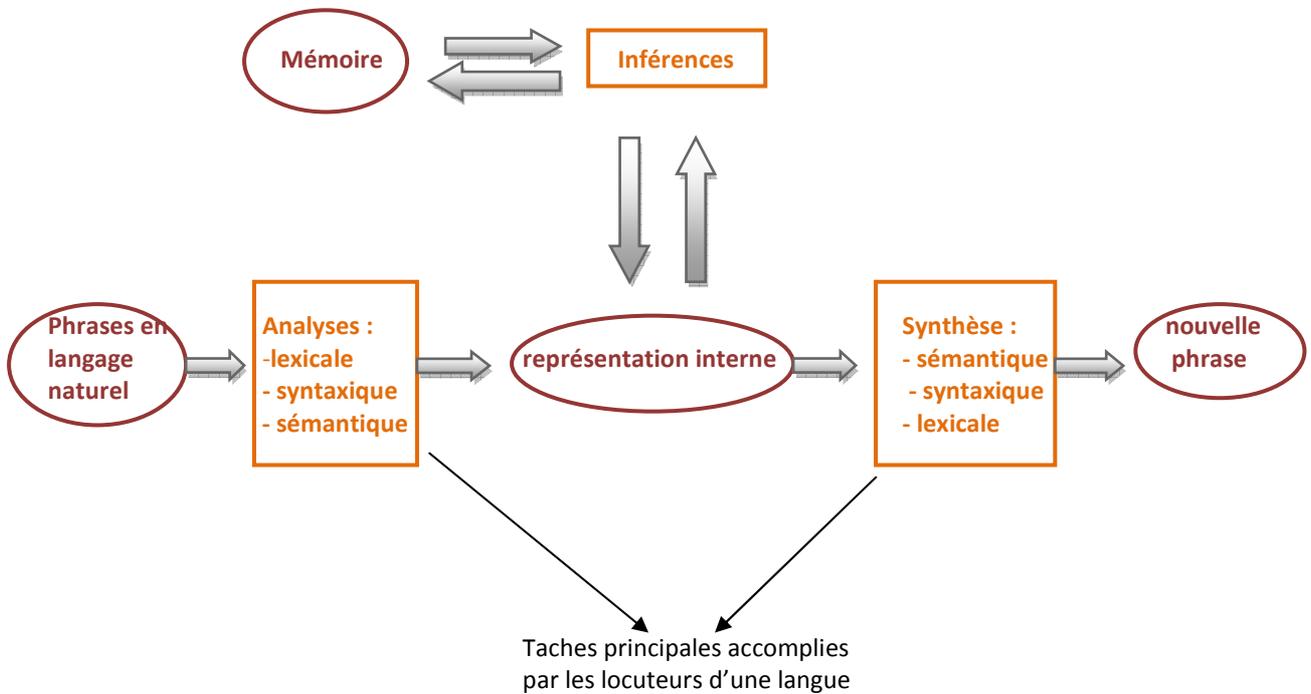
- L'axe syntagmatique (à la verticale)
- L'axe paradigmatisque (à l'horizontal)
- La double articulation du langage.
- La distinction entre sémantique lexicale et sémantique propositionnelle.

Caractérise les langues naturelles

On peut pourtant dire que les langages de programmation informatique dispose de la double articulation. La première repose sur les règles à suivre pour écrire un programme correct, la seconde sur la construction d'unités lexicales du programme. On peut associer une sémantique aux programmes informatiques.

4) La chaîne du traitement standard

Schéma représentant le fonctionnement de l'esprit humain :



 Traitement

 Données

Comprendre un énoncé nécessite une analyse tandis que produire un énoncé demande une synthèse.