

## Quelques idées sur la linguistique informatique

Auteurs : *Olga Tcaci, Olga Severin, Daniela Istrati*

La linguistique est la discipline qui se concentre sur l'étude du langage. Plusieurs branches composent la linguistique et ensemble, elles tentent de cerner les phénomènes par lesquels les êtres humains peuvent communiquer entre eux.

Les découvertes linguistiques tendent à montrer que la langue se comporte comme un système. Cependant, c'est un système difficile à cerner, quelquefois flou (jugements grammaticaux et d'acceptabilité de Chomsky). Il faut admettre que les êtres humains ont besoin d'un système sous-jacent pour permettre un encodage et un décodage uniforme chez des locuteurs d'une même langue.

On peut dire qu'avec le développement rapide des technologies informatiques, on a senti le besoin de s'appuyer sur les techniques linguistiques pour faciliter la communication homme-machine. A la fois, la linguistique a pu profiter de la puissance des ordinateurs pour acquérir une nouvelle dimension, et ouvrir la voie à de nouveaux domaines de recherche.

En français, « linguistique informatique » et « informatique linguistique » sont des termes que l'on rencontre dans la littérature pour désigner les champs de recherche qui ont comme objet le traitement automatique de données langagières.

La linguistique informatique fait partie intégrante des techniques informatiques, et intervient également dans des sous-domaines de l'intelligence artificielle. Parmi les applications concrètes on peut parler de :

- **La conception de logiciels en langages informatiques pour l'analyse d'une langue naturelle** : Noam Chomsky a établi la similitude entre les langues naturelles et les langages informatiques. Ainsi, un langage informatique est constitué d'un dictionnaire et d'une grammaire. L'interprétation d'un texte passe par son *analyse lexicale*, puis *syntactique*, ensuite *sémantique*. C'est le travail qu'effectuent les interprètes et les compilateurs.
- **La traduction automatique** : ce problème longtemps sous-estimé s'est en fait avéré l'un des plus délicats à effectuer pour un ordinateur. Aux niveaux lexicaux et syntaxiques, à peu près maîtrisés, s'ajoutent une analyse *sémantique*, puis *pragmatique*, qui tentent de déterminer le sens particulier d'un mot, dans le contexte où il apparaît, le contexte lui-même pouvant s'étendre à l'ensemble du texte traduit.
- **L'analyse de requêtes en langage naturel** : cette idée fut considérée comme une solution satisfaisante au problème de communication entre l'homme et la machine. Quoi de mieux en effet pour un non-spécialiste que de pouvoir adresser ses commandes et ses questions à l'ordinateur, dans son propre langage? L'arrivée des interfaces graphiques, de la souris, et de la métaphore du bureau ont quelque peu éclipsé cette technique. Elle réapparaît sous d'autres formes, notamment comme composante nécessaire de l'analyse et de la reconnaissance vocale ou encore de la reconnaissance d'écriture, popularisée par certains *assistants personnels*.

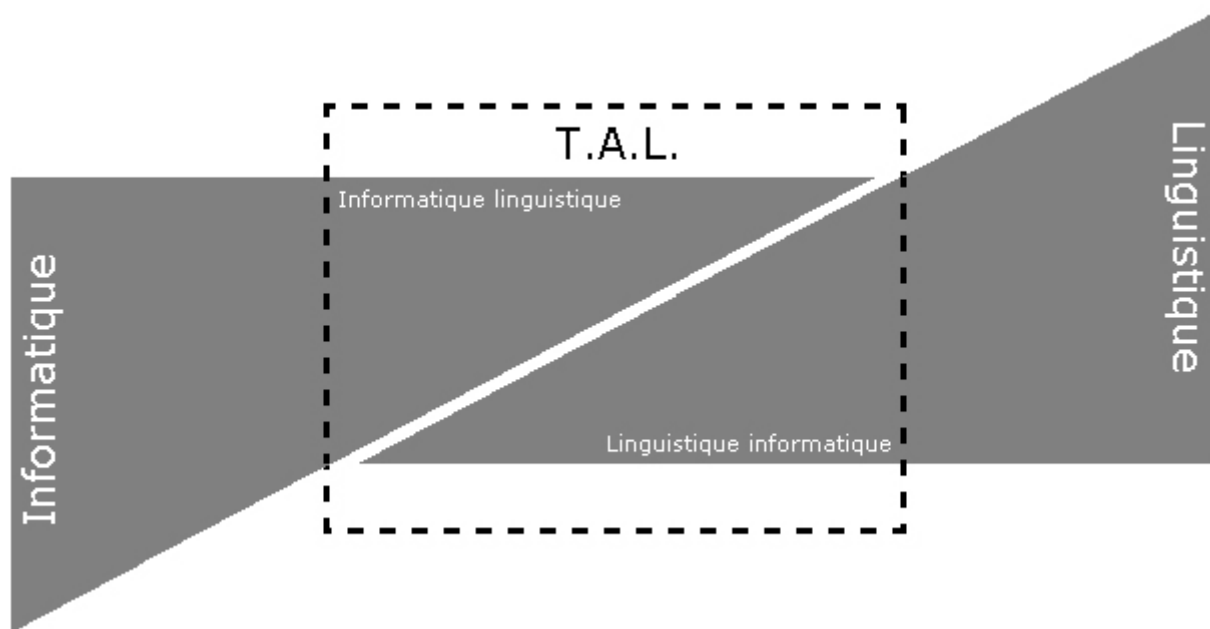
Une analyse formaliste pour un langage simplifié, c'est en faveur de quoi plaident les dernières directives européennes. D'autres approches d'analyse textuelle, en particulier développées dans le cadre de la sémiologie, refusent cette générativité et vont même jusqu'à affirmer que le texte n'a qu'une surface.

La linguistique et l'informatique sont, à prime abord, des disciplines qui n'ont rien à voir l'une avec l'autre. Pourtant, elles unissent leurs efforts pour doter une machine, l'ordinateur, de réelles

capacités de communication. Le terme *traitement automatique du langage* (T.A.L.) désigne toutes les activités de conception d'outils informatiques de manipulation automatique de données textuelles.

De nos jours, l'informatique inclut des aspects matériels, logiciels en plus de l'aspect théorique toujours présent.

En conséquence, on pourrait représenter les recherches sur un continuum allant de la linguistique à l'informatique, en supposant toujours un mélange des deux, comme le montre l'illustration ci-dessous.



**Fig. L'informatique et la linguistique s'associent pour le traitement automatique de la langue (T.A.L.)**

En conclusion, on peut dire qu'à l'extrémité de la portion du T.A.L., on retrouve les buts suivants: pour la linguistique informatique il faut explorer le langage à l'aide d'outils informatiques et pour l'informatique linguistique, manipuler l'information linguistique et automatiser des tâches linguistiques.

En informatique, la règle de base est la suivante : ce qui est difficile pour l'humain est facile pour la machine et ce qui est facile pour l'humain est difficile pour la machine.

**Bibliographie :**

1. Fuchs et al. 1993, Wehrli 1997
2. Le Français dans le monde, février 2006