Résumé

Le mythe des industries de la langue.

J.C Milner (Milner 1989) caractérise une science par *"la mathématisation de l'empirique"* et par *"la constitution d'une technique, telle que la technique se définisse comme l'application pratique de la science (d'où le terme de science appliquée) et que la science se définisse comme la théorie de la pratique (d'où le terme de science fondamentale)"*.

Une industrie suppose préalablement une science achevée (provisoirement). Il ne saurait en être ainsi pour les langues naturelles : les descriptions sont toujours partielles, partiales et destinées à le rester. Le Traitement Automatique du Langage Naturel (TALN) doit placer au centre de ses préoccupations la nécessité d'outils, de dispositifs permettant de mettre au point les connaissances à utiliser, de tenir compte de leur caractère avant tout provisoire, daté et évolutif. L'objectif de la thèse est de donner l'intuition de ce type de dispositif, (en donnant à voir) par des morceaux de réalisation, en s'appuyant sur deux approches informatiques traitant différemment de la classification : la PàO (Programmation à Objets) et la PàP (Programmation à Prototypes), et en utilisant le paradigme de la réflexivité. Le point d'ancrage langagier est tout d'abord le traitement d'une série d'expressions figées puis le traitement de savoirs lexicaux extraits à partir de corpus. On traitera de l'organisation des connaissances (ici lexicales au sens d'entrées lexicales dans toutes leurs dimensions : morphologie, sémantique, pragmatique), et de l'utilisation des connaissances (règles de combinaison et de structuration).

Notre travail s'inscrit plus particulièrement dans une approche informatique manipulant des prototypes. Cette approche de représentation ne préprogramme pas de manière figée ni toutes les opérations pour un traitement automatique de la construction du sens ni les processus de représentation des unités lexicales manipulées. Il faut faire en sorte que les programmes construits puissent évoluer aussi bien dans les opérations que ces derniers permettent que dans les représentations du domaine construites. La PàP (Programmation à Prototypes) conduit à penser différemment pour construire une représentation informatique d'un certain domaine de connaissances. Il ne s'agit pas de construire une représentation informatique d'un domaine donné à partir d'une somme de connaissances figées et connues par avance. Ce modèle de représentation permet de construire progressivement les entités informatiques suivant les connaissances dont on dispose sur le domaine visé : si de nouvelles connaissances sur le domaine sont mises à jour, on peut affiner le processus de représentation déjà amorcé en tenant compte de ces nouvelles informations sans avoir à reconstruire entièrement de nouvelles structures. Le processus de représentation lié à la PàP ne nécessite donc pas une définition exacte des concepts à représenter. La PàP permet d'introduire une dimension temporelle dans le processus de la représentation : ce dernier peut évoluer ou faire évoluer les résultats construits de manière continue.