

## PRESENTATION

Le figement est l'un des problèmes majeurs auxquels sont confrontés les linguistes et les informaticiens travaillant à la mise en place de systèmes de traitement automatique des langues naturelles du fait que ces dernières comportent une part non négligeable de structures figées. Dans la première partie, différents aspects de ce problème seront traités.

Une première série d'articles fait état des nécessaires descriptions linguistiques, relatives au figement, dans une langue donnée. Michel Mathieu-Colas rend compte de l'absence d'homogénéisation des majuscules des noms propres composés du français et, ce faisant, pose la question de leur détection automatique du fait de l'extrême variété des formes proposées par les lexicographes. La prise en compte des faits de graphie exposés est indispensable si l'on souhaite développer des systèmes performants. D'une manière plus générale, tous les faits linguistiques ayant trait au figement doivent être pris en compte dans la perspective du traitement automatique des langues naturelles. C'est le cas, par exemple, des phénomènes syntaxiques et sémantiques relevant de la détermination de l'espagnol, du français et du grec étudiés par Xavier Blanco, d'une part, et Zoé Gavriilidou, d'autre part.

Une seconde série d'articles expose les possibilités d'applications qui résultent des descriptions linguistiques préalables. Pierre-André Buvet évoque la possibilité de recourir à des informations linguistiques standardisées, en l'occurrence les classes d'objets, pour améliorer l'extraction automatique des termes complexes. Les formes répétées dans des textes permettent à Pierre-Yves Foucou de préciser les conditions de détection automatique des formes figées. Patrick Drouin propose de coupler les analyses linguistiques avec des analyses statistiques afin de développer des automates de reconnaissance et d'acquisition en terminologie.

Les travaux préalables d'analyses linguistiques comme les applications qui découlent supposent la mise en œuvre de bases de données considérable relatives aux structures figées. Peter Greenfield évoque la structuration de l'une de ces bases de données : un dictionnaire spécialisé de collocations multilingues. Il fait état des possibilités d'exploitation informatique des invariants inhérents à ce type de dictionnaire.

La deuxième partie regroupe les textes de communications présentées au colloque FRACTAL97 qui n'ont pu être publiés dans le numéro spécial de BULAG de l'année 1997. Sur le thème « informatique et linguistique : théories et outils pour le traitement automatique des langues naturelle », de nombreux participants ont fait état de la diversité des approches possible en linguistique computationnelle.