



Linguistique computationnelle : entre sciences cognitives et traitement automatique des langues

Maxime Amblard, Greg Kobele, Christian Retoré

► To cite this version:

Maxime Amblard, Greg Kobele, Christian Retoré. Linguistique computationnelle : entre sciences cognitives et traitement automatique des langues. ARCo, Mar 2006, Bordeaux, France. 2006. <hal-01079258>

HAL Id: hal-01079258

<https://hal.archives-ouvertes.fr/hal-01079258>

Submitted on 31 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Linguistique computationnelle : entre sciences cognitives et traitement automatique des langues

Maxime AMBLARD, Greg KOBELE et Christian RETORE

ARCo 2006

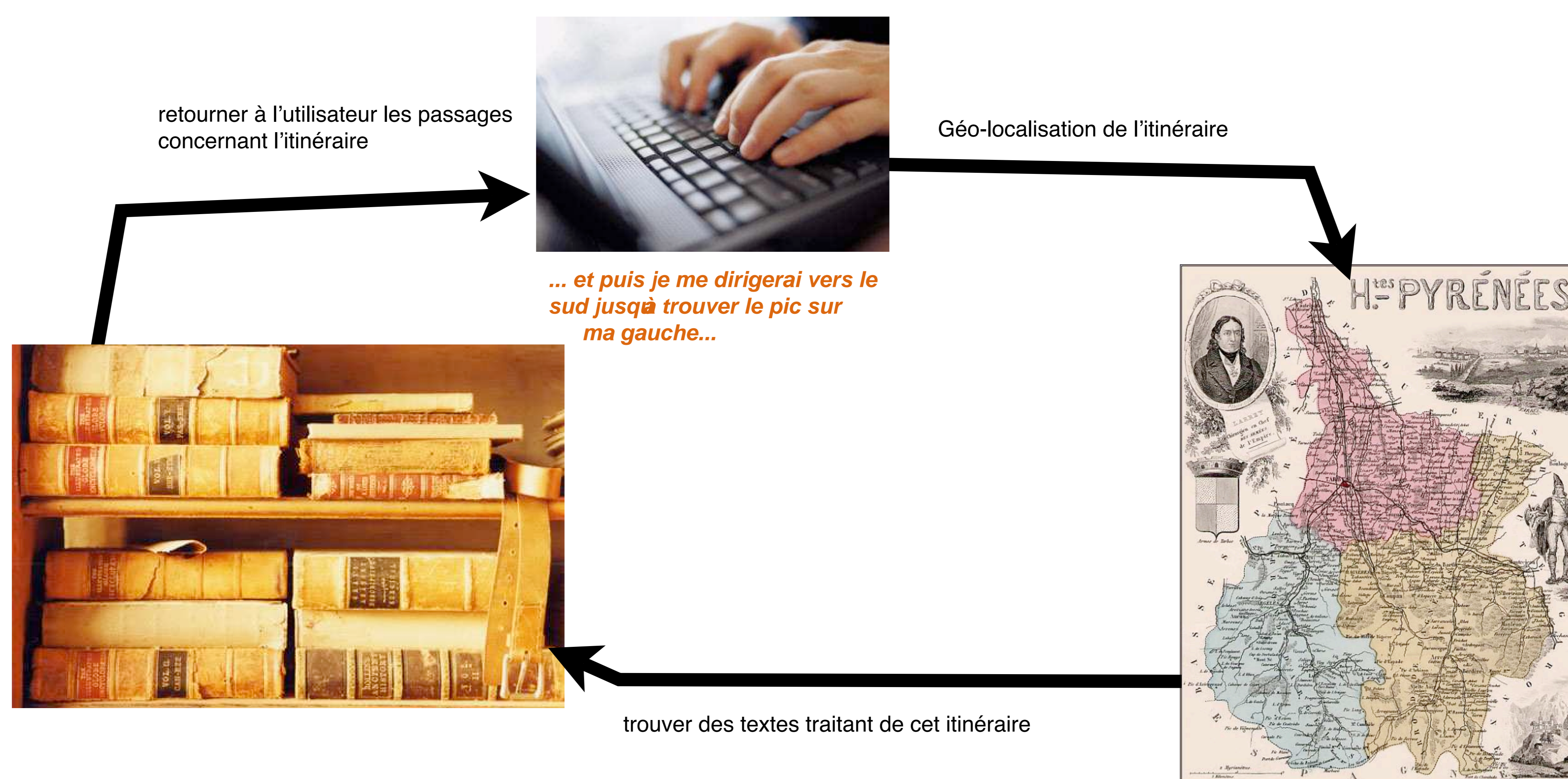


SIGNES - LaBRI, Université de Bordeaux 1, INRIA Futurs, ENSEIRB

	Sciences cognitives (SC)	Traitement automatique des langues (TAL)
Buts :	Description du processus langagier humain <ul style="list-style-type: none"> • mécanismes de production (du sens aux sons) • mécanismes d'analyse et de compréhension (des sons au sens) 	Réalisation d'outils <ul style="list-style-type: none"> • similaires au comportement humain (IHM) • utiles mais non humains (recherche sur Internet)
Modèles	<ul style="list-style-type: none"> • grammaires formelles, machines virtuelles (automates) • logique de description et de raisonnement • sémantique formelle et philosophie du langage 	<ul style="list-style-type: none"> • grammaires régulières ou algébriques • statistiques • système d'information, représentation des connaissances
Conception du langage :	<ul style="list-style-type: none"> • un phénomène biologique/psychologique individuel • un ensemble fini de mécanismes décrivant l'infini potentiel des dire (compétence) • limité par les capacités de la mémoire à court terme (performance) 	<ul style="list-style-type: none"> • un phénomène collectif • un ensemble fini de productions langagières (copus)
Succès:	description adéquate (prédiction) et explicative: expression, compréhension, acquisition	fonctionnalité: couverture, efficacité, pertinence pour d'autres applications
Limites	problème de passage à l'échelle	résultats souvent inexploitable par d'autres applications
Intérêt :	Leurs particularités peuvent être appliquées pour deux points : <ol style="list-style-type: none"> 1. la syntaxe des langues possède une structure plus riche que celle reconstruite par les grammaires hors-contextes, donc l'utilisation de représentations linguistiques permet de représenter directement des analyses qui ne pouvaient être atteinte que statistiquement. 2. les représentations cognitives peuvent permettre de décider localement d'une meilleure solution. 	Les applications du traitement automatique des langues ont deux aspects : <ol style="list-style-type: none"> 1. un noyau génératif (une grammaire hors-contexte) 2. une suite d'outils spécifiques qui de manière stochastique cherche une représentation optimale du lexique pour un petit fragment de l'ensemble des productions.

Exemple d'architecture mixte TAL/SC

extraction automatique des paragraphes géographiquement pertinents dans un grand corpus régional



Conclusion

constat:

- objectifs différents

- manque de communication TAL / SC

idée pour dépasser les limites actuelles:

utilisation raisonnée de modèles cognitifs linguistiques pour

- améliorer les applications

- tester les performances algorithmiques des modèles.

(cf. l'amélioration de la compression d'image par incorporation de structures cognitives)

References

- [ManningSchutze 99] Manning C. D. et Schütze H. (1999), *Foundations of Statistical Natural Language Processing*.
- [Moot 06] Moot R. (2006), Automated extraction of type-logical supertags from the spoken Dutch corpus. In *The complexity of Lexical Descriptions and its Relevance to NLP: A Supertagging Approach*.
- [Pollock 97] Pollock J.-Y. (1997), *Langage et Cognition: la programme minimaliste de la grammaire générative*.
- [Retore 05] Retore C. (2005), Syntaxe et traitement automatique des langues. In F. Lambert et H. Nolke, editors, *La syntaxe au coeur de la grammaire: mélanges offerts à Claude Muller*.
- [Schmalz 05] Schmalz M. S. (2005), On the convergence of image compression and object recognition. In *Proceedings of the 43rd Annual Association for Computing Machinery Southeast Conference*.