

Défis scientifiques du Traitement Informatique de la Langue

G rard Huet

INRIA

17 Septembre 2002, ENS Cachan

Petit historique

- Traduction Automatique
- Le rapport ALPAC (1966)
- Chomsky-Schützenberger
- Théorie des langages formels
- Compilation des langages de programmation
- PROLOG
- Transducteurs rationnels
- Méthodes statistiques

Applications

- Reconnaissance de la parole
- Interface homme-machine
- Localisation/internationalisation
- Traitement de textes
- Recherche d'informations
- Philologie
- Assistance à la traduction
- Génération de textes multilingues

Niveaux de représentation du langage

- Phonologie - du signal au flux de phonèmes
- Segmentation - des phonèmes aux mots
- Morphologie - flux de mots étiquetés
- Syntaxe - structure arborescente
- Sémantique propositionnelle
- Sémantique intentionnelle et modale
- Structure informationnelle du discours
- Structure rhétorique
- Pragmatique - ce que parler veut dire

Théories linguistiques

- Saussure - dualité du signe
- Tesnière - dépendances
- Chomsky - grammaire générative
- Colmerauer - programmation logique
- Lambek - grammaires catégorielles
- Joshi - TAG
- Chomsky - \bar{X} , G&B, Minimalisme
- Psychologie, Sciences cognitives

Informatique théorique

- Hiérarchie des langages formels
- Automates et transducteurs d'état fini
- Analyse syntaxique
- Algorithmique, complexité
- Sémantique formelle
- λ -calcul
- Modèles catégoriques

Théorie de la démonstration

- Mathématisation de la logique
- Boole Frege
- Hilbert Brouwer
- Cantor Russell
- Heyting
- Gödel Church Kleene
- Prawitz Herbrand
- Tarski Scott
- Curry Howard de Bruijn
- Martin-Löf Girard

Entre la Logique et l'Algèbre

- Logique Linéaire comme Fondements Logiques
- Ni Contraction, Ni Affaiblissement
- Séquents comme Sacs de Littéraux
- Lambek Syntactic Calculus \ /
- LL Non Commutative
- LL Non Associative
- Pré-groupes

Entre la Syntaxe et la Sémantique

- Lambek Syntactic Calculus : colle syntaxique
- \approx LL non-commutative
- Montague Semantics comme sémantique propositionnelle
- \approx Logique d'ordre supérieur + modalités
- définissable en LL avec modalités
- Réseaux de démonstration
- Géométrie de l'interaction
- Machine de contraintes

Problèmes ouverts

- Hors contexte vs Gram. Catég.
- Complexité de la reconnaissance
- Langages d'arbres d'analyse
- Langages faiblement contextuels
- Gram. Univ. vs lexicalisée
- Apprentissage
- Recodage des TAGs
- Codage des Gram. minimalistes
- Synthèse à partir de langage pivot

Plus loin dans la Sémantique

- Pronoms
- Ellipses, Focus, Accent
- Anaphores
- Déictiques
- Quantificateurs généralisés
- DRT
- Situations
- Logique dynamique, épistémologique
- Sémantique de Kripke
- Logique hybride

Focus sur les anaphores

- Résolution d'équations
- Unification de 1er ordre
- Equations entre mots
- $Xa=aX$
- $aXa=YbY$
- Equations polynomiales
- Théorie des types de Church
- Unification d'ordre supérieur

Résolution d'anaphores par unification

- John likes Mary and Peter does too.
- $like(j, m) \wedge R(p)$
- $R(j) = like(j, m)$
- $R \leftarrow \lambda x. like(x, m)$
- $*R \leftarrow \lambda x. like(j, m)$

Exemples gradués

- John likes his wife and Peter too.
- John says that he likes his mother and Peter does too.
- John spent his paycheck but Peter saved it.
- Mary wants to go to Spain and Fred wants to go to Peru but because of limited resources, only one of them will.

Pas trop loin dans la Sémantique

- Connaissances linguistiques
- Connaissances sur le monde
- Bon sens
- Cyc
- Wordnet
- Nomenclatures, taxonomies
- Thesaurus, ontologies
- Logiques de description

Entre la Parole et la Linguistique

- The Watergate tapes
- Langue comme jeu de linguistes
- Publications typiques en CL

Le défi applicatif (1)

- Pauvreté des logiciels libres
- Rareté des ressources linguistiques
- Manque de protocoles d'évaluation
- Prééminence de l'anglais
- Recherche orientée phantasmes
- Pâtée technocratique
- La valorisation mal comprise
- Des start-ups à FT
- Un tissu industriel fragile

Le défi applicatif (2)

- Implémenter les algorithmes
- Les tester sur des données en vraie grandeur
- Construire des bibliothèques modulaires
- Coopérer avec les meilleures équipes
- Importer les ressources linguistiques
- En dernier ressort, les construire
- Ni Prolog, ni C++
- Ne pas courir derrière les modes
- Mais utiliser la technologie récente
- Travailler sur la durée

Le défi applicatif (3)

- Souvenirs d'un implémenteur
- Mentaur/Centaure vs Emacs et Word
- ML, Caml, Objective Caml, Pidgin ML
- Zen
- <http://paillac.inria.fr/> / huet/ZEN
- Gerard.Huet@inria.fr

Conseils à un jeune chercheur

- Choisir un domaine pas trop mou
- Acquérir les bases conceptuelles
- Implémenter des maquettes documentées
- Ecoles d'été, tutoriels
- Changer de sujet de recherche tous les 5 ans
- Change de discipline au moins une fois dans sa vie
- Explorer en largeur
- Seek applications for the Joy of it
- Do what only you can do
- Ne soyez pas un mouton de Panurge