

Informatique Linguistique et Programmation Logique

ILPL

■ Problématique et résultats

Les activités de l'équipe de recherche en Informatique Linguistique (ILPL) se concentrent autour du traitement automatique de la langue française écrite. De manière très générale, notre approche associe langue et raisonnement, aux niveaux formels et applicatifs.

Nos travaux sont organisés selon deux axes : (1) un travail sur les notions de sens et de polysémie, concrétisés par le développement de ressources linguistiques liées aux formes prédicatives (verbes et prépositions), et (2) les modèles et applications en langue et raisonnement pour l'accès intelligent aux bases de données, aux textes et au Web (systèmes de question-réponses coopératives, résumé automatique, grammaires génératives pour le texte et extraction de connaissances).

Notre activité en sémantique lexicale est une analyse approfondie de la syntaxe et de la sémantique lexicale des formes prédicatives et leur modélisation dans une dynamique générative du langage.

Ces travaux permettent une extraction de connaissances plus avancée des textes, prenant en compte les actions (prédicats) et leurs participants (arguments).

La dimension générative modélise l'émergence d'usages non-standards, reflets de la créativité et de la dynamique du langage, propres à chaque locuteur. Les axes principaux de notre recherche sur les ressources linguistiques s'articulent autour de :

LA MODÉLISATION DE CONNAISSANCES LINGUISTIQUES

- En syntaxe et en sémantique lexicale (représentations en structure lexicale conceptuelle, traitement de la compositionnalité, sous-détermination).
- L'analyse de la variation du sens, les modèles formels pour la polysémie et les métaphores, le lexique génératif et ses évolutions.
- Le développement de ressources linguistiques pour le traitement du langage naturel : structures syntaxique et conceptuelle des verbes (projet VOLEM) et des prépositions (projet PrepNet) [4714, 5956].
- La programmation logique (contraintes, modélisation, stratégies d'analyse et de génération pour le traitement du langage naturel).

PERSONNEL

Directeur de recherche
Patrick Saint-Dizier

Maîtres de conférence
Farah Benamara
Mouna Kamel

Doctorants
Farida Aouladomar (01/04→)
Olfa Jenhani El Jed (09/02→)
Véronique Moriceau (09/03→)

RÉFÉRENCES

[4692]

Farah Benamara,
Patrick Saint-Dizier.

*Advanced Relaxation
for Cooperative
Question Answering.*

Dans : *New Directions in Question
Answering*. Mark Maybury (Eds.),
MIT Press, 234-254, août 2004.

[4698]

Farah Benamara,
Patrick Saint-Dizier.

*Lexicalisation Strategies
in Cooperative Question-
Answering Systems.*

Dans : *COLING 04*, Genève,
22 au 26 août 2004, pp. 345-352.

[4714]

Patrick Saint-Dizier,

*Syntax and semantics
of prepositions*, Kluwer Academic,
Dordrecht, *speech and language
processing series*.

[5946]

Aouladomar, F.,

*Towards Answering Procedural
Questions*, Workshop KRAQ05,
IJCAI05, Edinburgh,
p. 1-14, July 2005.

[5956]

Saint-Dizier, P.,

*PrepNet: a Framework
for Describing Prepositions:
preliminary investigation results*,
IWCS05, Tilburg.LNCS,
Springer, p. 128-139.

LANGUE ET RAISONNEMENTS POUR LES APPLICATIONS INTELLIGENTES SUR LE WEB

- Extraction de connaissances dans les pages Web, organisée autour des formes prédicatives. Production dynamique de représentations (frames + logique) à partir de l'ontologie du domaine aux noeuds de laquelle sont associées des grammaires locales.
- Dans le cadre d'un accès intelligent au WEB, développement de réponses coopératives : analyse de fausses présuppositions, de termes flous, questions par l'exemple, etc. Dans les questions : développement d'un savoir-faire coopératif incluant des techniques de réponses relaxées, intentionnelles, conditionnelles et narratives (projet WebCoop) [4692].
- Développement d'autres types de question-réponses, toujours dans un cadre coopératif : les questions en Comment ? appliquées aux textes procéduraux et l'intégration de données lorsque l'on obtient à partir du Web des réponses avec contradictions. Travaux en grammaire générative pour les textes procéduraux en vue d'améliorer l'extraction de réponses [5946].
- Analyse plus formelle des notions d'explicitation des connaissances, d'explication et d'argumentation dans les réponses coopératives : modèles d'interprétation (par exemple des implicatures conversationnelles), pragmatique de la coopérativité, connecteurs et templates associés.
- Résumé automatique de pages Web produit à partir d'une requête sous forme de mots-clés (projet WebSum). L'objectif est de produire un résumé en langue naturelle organisé selon l'ontologie du domaine au lieu des liens produits par les moteurs classiques. Nous étudions des modes de résumés allant d'une simple organisation ontologique des liens à du langage naturel.
- Génération de langue naturelle associée aux projets ci-dessus, avec un accent particulier sur la lexicalisation, l'agrégation et l'introduction d'hyper-liens comme mode d'expression du discours [4698].

Les ressources développées en syntaxe et sémantique lexicale sont utilisées de façon intensive pour les projets en langue et raisonnement. Ce que nous développons se veut à la fois approfondi sur le plan de la langue et opérationnel dans des applications.

RÉSULTATS IMPORTANTS

Ces années récentes nous ont permis de développer un ensemble de principes et de modèles concernant les réponses coopératives, qui ont abouti au projet WebCoop [4692]. Ce projet a stimulé des recherches complémentaires sur d'autres types de questions et la fusion d'information. Nous avons pu ainsi constituer un pôle de compétence autour des question-réponses avancés, qui s'est fait toute sa place au niveau international.

Dans une autre direction, nous avons ouvert la voie, via le projet PrepNet [4714], à un travail approfondi sur la syntaxe et la sémantique des prépositions dans une perspective pluri-lingue. Ce travail est particulièrement utile en TALN : extraction de connaissances, traduction automatique, génération de langue. Il est totalement original, et nous avons été le moteur de workshops dans ce domaine. Une ouverture sur les langues de l'Asie est en cours.

■ Prospective

Nos perspectives pour les deux années à venir incluent la continuité des activités ci-dessus. Un effort particulier visera :

- Les question-réponses coopératives dans des univers de connaissances incomplètes ou de fiabilité limitée
- La syntaxe et la sémantique des textes procéduraux et les questions en comment ? et en pourquoi ? associées
- La fusion de données venant du Web et la génération de réponses synthétiques en langue naturelle
- Les techniques de génération à base de templates sous-spécifiés, modèles pour le choix lexical, l'agrégation et la micro-planification, la génération d'explications
- La syntaxe et la sémantique des prépositions (PrepNet) dans un cadre multilingue (français, espagnol, anglais, langues de l'Asie)

■ Thèses et habilitations

- Farah Benamara. WebCoop : un système de question-réponses coopératif sur le Web, Thèse UPS, 11/2004

■ Collaborations, contrats et transfert

- TCAN-CNRS : génération de réponses coopératives
- GENI : ARC Inria : génération et inférence
- VOLEM : projet international en sémantique des verbes
- GDR Sémantique linguistique : rédaction de fiches encyclopédiques en sémantique linguistique
- Projet UPS-BQR : l'argumentation et la gestion des connaissances dans les systèmes question-réponses
- Collaborations internationales : Portugal (CLUL), Espagne (UdL, UdB), Inde (IIIT-Hyderabad, IIT-Kharagpur), Thaïlande (Kasertart), USA (Maryland)

Diffusion de nos ressources linguistiques (verbes et prépositions) sur notre page Internet : www.irit.fr/recherches/ILPL

■ Animation, gestion et vulgarisation de la recherche

- Membres de nombreux comités de programme internationaux (conférences et revues),
- Organisation de plusieurs conférences et workshops : ACL'01, 1er workshop sur les prépositions, IJCAI-Kraq05 (question réponses et IA),
- Activité éditoriale : ouvrage édité chez Kluwer sur la syntaxe et la sémantique des prépositions, actes de la conférence KRAQ'05.