

Traitement automatique des langues

Brigitte Grau (LIMSI & ENSIIE)

Anne-Laure Ligozat (LIMSI & ENSIIE)

Vincent Guigue (LIP6 & UPMC)

Introduction

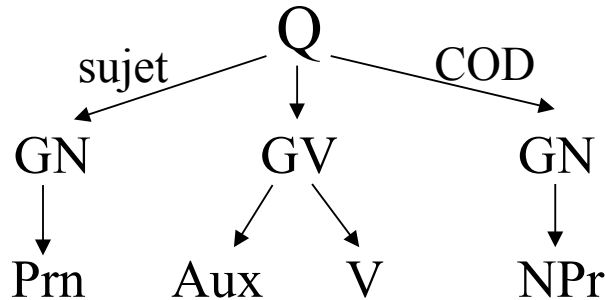
- **Qu'est-ce que le traitement automatique de la langue ?**
- **Historique**
- **Applications**

Problème de TAL

Répondre à des questions

Qui a poignardé Henri IV ?

NOM de PERSONNE



Méthodes

Patron d'extraction :

Personne <poignarder> Henri IV

Entité nommée

Critères syntaxiques :

Personne <poignarder> Henri IV

The text shows the same extraction pattern as above, but with brackets underneath. A bracket labeled 'SUJET' spans the word 'Personne'. Another bracket labeled 'COD' spans the words '<poignarder> Henri IV'.

... C'est à Paris, rue de la Ferronnerie, qu'Henri IV a rendu son dernier soupir...
D'autre part, on peut rappeler que les rues encombrées et étroites du Paris d'avant
Hausmann étaient très favorables aux guets-apens. La facilité avec laquelle
Ravaillac a pu poignarder Henri IV peut paraître incroyable ! ...

Problème de TAL

Répondre à des questions

Méthodes

Qui a tué Henri IV ?



Critères sémantiques

Tuer synonyme Poignarder

Où a été tué Henri IV ?



LIEU dans le contexte



LIEU (VILLE, PAYS, etc)

ou



Critères sémantiques

Tuer => Mort

Rendre son dernier soupir => Mort

... C'est à Paris, rue de la Ferronnerie, qu'Henri IV a rendu son dernier soupir...

D'autre part, on peut rappeler que les rues encombrées et étroites du Paris d'avant Haussmann étaient très favorables aux guets-apens. La facilité avec laquelle Ravallac a pu poignarder Henri IV peut paraître incroyable ! ...

Problème de TAL

- **Question-réponse : Qu'est-ce qu'une bonne réponse ?**
 - Qui est Picasso ? Un peintre, un sculpteur, ...
 - Où est né Picasso ? Le pays, la ville, l'adresse ?
 - Cf. dialogue extrait de POLITICS (Carbonell, 80)

- **Et plus généralement,**
 - Que signifie comprendre ?
 - Approche « cognitive »
 - Approche par la « tâche »

Trouver un SENS (1)

■ Niveau des mots :

- Ambiguités de formes :
 - nourrissons
- Formes variables -> même concept
 - clé et clef
- Formes différentes -> même concept : synonymes
 - soulier et chaussure
- Une forme -> concepts différents : polysémie
 - permis -> autorisation/certificat
- Une forme -> concepts différents : homonymes
 - crêpe -> gâteau/tissu

Trouver un SENS (2)

■ Niveau des phrases :

- Ambiguïté de forme : je regarde le mur de la cuisine
- Formes variables -> même sens : allotaxie + paraphrases
 - ➔ le vent agite le drapeau et le drapeau s'agite au vent
- Formes identiques -> sens différents : homotaxie
 - ➔ prendre son train et prendre son parapluie
- Repérage (déictique)
 - ➔ son livre -> écrit/possède/utilise
 - ➔ en 65, mon directeur était un économiste

Trouver un SENS (3)

■ Niveau des phrases :

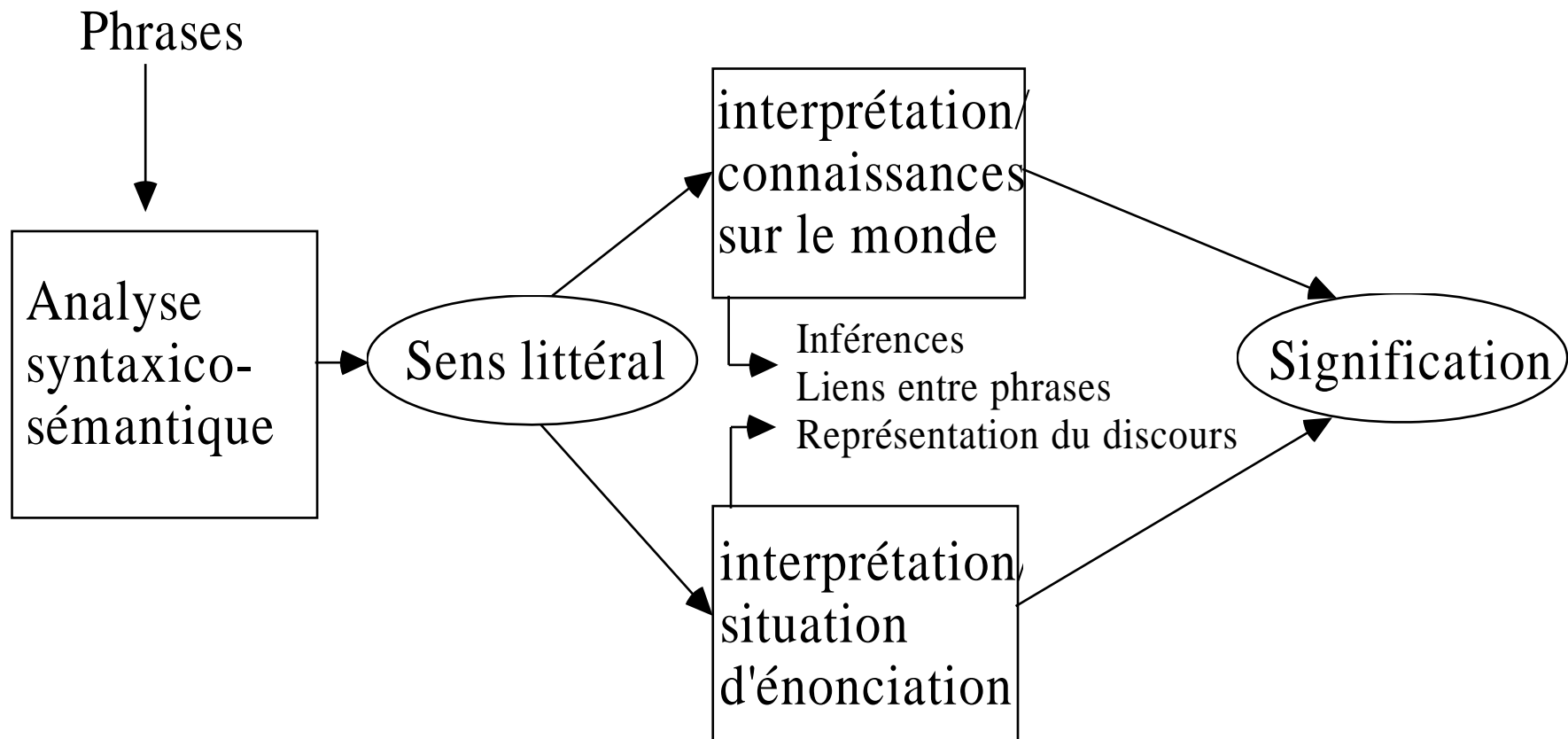
- Tropes (métaphores, métonymie)
 - soigner un moteur, boire un verre

■ Niveau du discours

- Anaphores et référence (pronoms et groupes nominaux définis)
 - Ma grand-mère se repose près de la cheminée.
 - Elle chauffe bien.
 - Elle se réchauffe
- Ellipses
 - Où vas-tu ? À Evry.

Compréhension

■ Transformation en différentes représentations



Historique

- **1946 : Discussion sur l'utilisation de calculateurs comme dictionnaires (Weaver-Roth)**
- **1949 : Memorandum Weaver**
- **1952 : Première conférence sur la traduction automatique (MIT, Bar Hillel)**
- **1957 : Théorie des langages formels (Chomsky)**
- **1960 : Bar Hillel : Traduction automatique impossible**

Réalisations : années 60 à 80

- **1960-1965 : Premières réalisations concrètes**
 - Baseball, Student, Sir, Eliza
 - Fondées sur mots clefs et patrons syntaxiques
- **Approche sémantique**
 - 1970 : SCHRDLU (Winograd)
 - 1973 : MARGIE (Schank), Traduction Anglais-Français (Wilks)
- **Approche pragmatique**
 - Frames (Minsky) (1974)
 - Compréhension d'histoires par application de Scripts et Plans : SAM-PAM, Schank, Wilensky (1977), BORIS, Dyer (1983)
 - Dialogues dirigés par la tâche (1977): Grosz

Actuellement ...

- **Approches robustes**
 - Vrais textes, large couverture
 - Approches statistiques vs approches analytiques par règles
 - Très décrites auparavant
 - Maintenant : de nombreux textes disponibles, pouvant être annotés
→ matériel pour l'apprentissage automatique
- **Evaluation**
- **Disponibilité d'outils et ressources :**
 - Analyseur syntaxique, bases de connaissances
 - Dépend de la langue
- **Problèmes ouverts**
 - Compréhension / raisonnement / inférence
 - Acquisition de connaissances sémantiques et pragmatiques

Applications

- **Enjeux économiques : Accès à l'information**
 - Manuels d'utilisation: Interrogation, Traduction
 - 10 000 pages pour un ordinateur, 100 000 pages pour un central téléphonique, un wagon pour un mirage
 - WEB: veille technologique ou stratégique
 - Documents dans les entreprises
- **Multilinguisme : Traduction automatique**
- **Génération de documents, aide à la rédaction**
- **Recherche d'information : approche sémantique**
- **Résumé automatique**
- **Extraction d'information, population de bases de connaissances**
- **Interfaces en langue (ex. avatars) → oral**

Organisation du cours

- **Problèmes et solutions**
 - phénomènes linguistiques
 - principaux modèles d'apprentissage
- **Manipulation d'outils et réalisation d'une application**
- **Cours:**
 - Morphologie et syntaxe, application à la simplification
 - Entités nommées et extraction de relations, application à la construction de bases de connaissances et question-réponse
 - Modèles HMM et CRF
 - Sémantique distributionnelle et LSA
 - Classification de textes, application à l'analyse de sentiments
 - Analyse thématique de texte
 - Apprentissage de représentations par modèles de langue
 - Evaluation

Organisation du cours

- **TD/TP : application des modèles d'apprentissages**
- **TDM fondés sur un travail personnel encadré des étudiants**
 - Effectuer une **synthèse bibliographique** sur l'un des thèmes proposés
 - Développer une **application** associée
 - Présentation
- **Les thèmes à développer sont les suivants :**
 - morphologie
 - terminologie : extraction de termes / collocations / entités nommées
 - syntaxe : analyseurs
 - sémantique : représentation/désambiguïsation/ressources
 - variations de sens (termes / phrases) : paraphrase, implication textuelle
 - anaphore et coréférence
 - analyse thématique de textes
 - résumé automatique
 - analyse du discours

Thématiques

Analyse lexicale

Termes & Collocations

Entités nommées

Il y a quelques années, je me trouvais dans un grand magasin de Harlem, entouré de quelques centaines de personnes. J'étais en train de dédicacer des exemplaires de mon livre "Stride toward Freedom", qui relate le boycottage des autobus de Montgomery en 1955-56. Soudain, tandis que j'apposais ma signature sur une page,

Analyse syntaxique



Séance
de dédicace

Attentat

je sentis quelque chose de pointu s'enfoncer brutalement dans ma poitrine. Je venais d'être poignardé à l'aide d'un coupe-papier, par une femme qui devait être reconnue folle par la suite. On me transporta d'urgence à l'Hôpital de Harlem

Sémantique

lit, nom / lire, verbe

où je restai de longues heures sur un lit tandis qu'on faisait mille préparatifs pour extraire l'arme de mon corps

Discours

Résumé automatique

Analyse thématique