

UE HMIN230

## Traitement Automatique des Langues

- Mercredi matin 8h 13h
- Christian Retoré
- Mathieu Lafourcade
- Violaine Prince
- Equipe TEXTE du LIRMM

# Des méthodes variées utiles au traitement automatique des langues ainsi que pour d'autres sujets

- Particularité: domaine très ancien de l'info (~1940) redevenu très à la mode (IA, machine learning)
- Théorie des langages, grammaires, automates
- Algo de graphes (réseaux lexicaux)
- Modèles de Markov cachés  
(= automates probabilistes, apprentissage par renforcement)
- (Un peu de) probabilités
- Unification (cf. logique, résolution)
- PROLOG (programmation logique),  
Grammaires de clauses définies
- Apprentissage (machine learning)
- Logique, lambda-calcul (plutôt en M2 avec la sémantique)
- **En lien avec: info fondamentale, bioinfo, compilation,...**

# Traitement Automatique des Langues

(≠ Recherche d'information, même si en rapport)

- **De quoi ça parle: connotations, champ sémantiques mots associés:**

RI

- L'avocat était marron. (fruit/juriste?)
- Petit ~ Grand ? Pourtant ces deux adjectifs vont avec les mêmes termes.
- Pierre a réussi son concours ~ Pierre n'a pas réussi son concours

- **Qui fait quoi, ce qui est affirmé, réfuté, supposé,...**

TAL

- Geach était il l'élève de Wittgenstein ?
- Wikipédia: Geach [...] En 1941, IL épousa la philosophe Elizabeth Anscombe, grâce à LAQUELLE IL entra en contact avec Ludwig Wittgenstein. Bien qu'IL N'ait JAMAIS suivi l'enseignement académique de CE DERNIER, cependant IL EN éprouva fortement l'influence.

# Des applications

- Traitement de l'information textuelle  
BigData: surtout des textes  
(Internet, mémoire d'entreprise)
- Interface homme-machine en langue naturelle
- Exemples:
  - Recherche d'information
  - Classification de textes
  - Correcteurs orthographiques
  - Interrogation en langage naturel,  
réponse en langage naturel (par ex. BD)
  - Traduction assistée par ordinateur

# Organisation du cours

- Surtout de la **syntaxe**, peu de sémantique  
pour cela il y a le cours de Master 2 sur la **sémantique**  
Intelligence artificielle et langage naturel (M2)
- Assez appliqué:
  - TP programmation,
  - Utilisation d'outils
- En particulier
  - Ecriture d'une grammaire d'unification en Prolog
  - Programmation d'algorithme d'analyse syntaxique
  - Graphes (réseaux lexicaux)
- Contrôle continu et examen