

MA1 Traitement Automatique du Langage (TALN): Approches Linguistiques et Approches Statistiques

Paola Merlo

MA1 - 2020-2021

- L'analyse du langage naturel tente de donner à un ordinateur la faculté de comprendre des langues naturelles comme l'anglais, le français ou le japonais.
- Par « comprendre », nous ne voulons pas faire croire que l'ordinateur acquiert un mode de pensée, des sensations et des connaissances humaines. Nous voulons seulement dire que l'ordinateur peut reconnaître et utiliser des informations exprimées à l'aide d'une langue naturelle.

Applications du TAL

- L'anglais comme langage de commande – c'est-à-dire l'usage d'une langue naturelle en lieu et place d'un langage artificiel comme c'est encore le cas dans les langages de commande des ordinateurs.
- Siri: le dialogue humain-machine.
- La traduction assistée par ordinateur de documents scientifiques et techniques ou bien d'informations commerciales d'une langue naturelle vers une autre.
- L'analyse automatique des rapports médicaux, médecine personnalisé.
- IBM Watson: l'extraction d'information et les systèmes question-réponse.

Google Duplex



<https://www.youtube.com/watch?v=znNe4pMCsD4>

L'informatique demain



Réalité augmentée

Applications du TAL: dialogue humain-machine

- <https://sites.google.com/site/olemon/Home/videos>

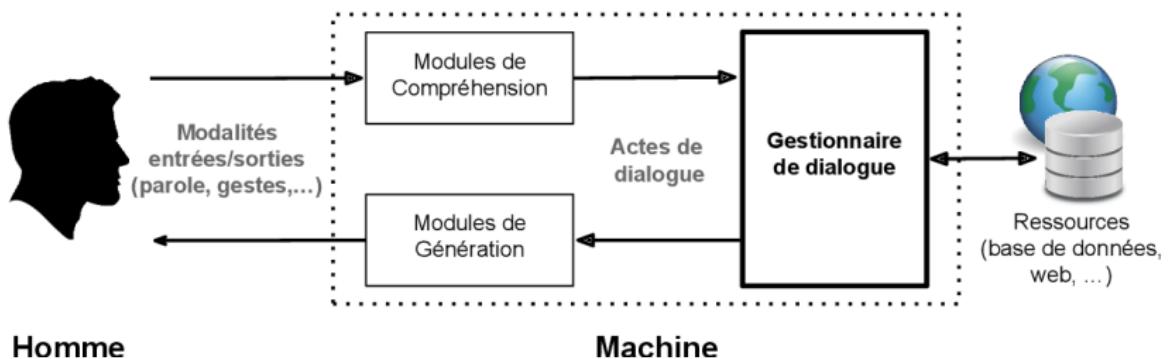
Applications du TAL: IBM Watson

- https://www.youtube.com/watch?v=WFR3lOm_xhE

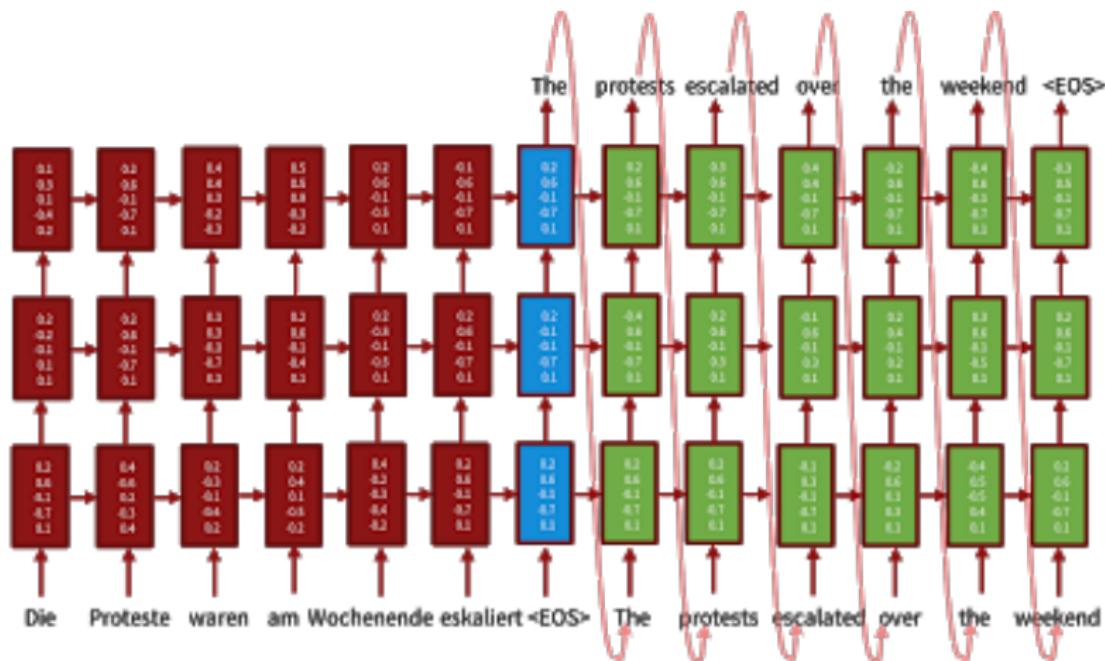
Applications du TAL: Google Translate

- <https://translate.google.com>

Architecture d'un système dialogue



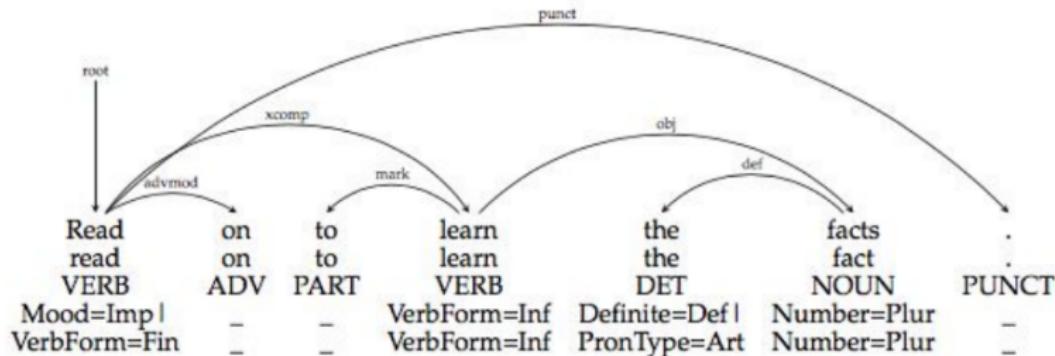
Architecture d'un système de traduction automatique



Word level - parts of speech tagging

ID	String	Lemma	UPOS	Feat	Dep	Rel
1	Read	read	VERB	Mood=Imp VerbForm=Fin	0	root
2	on	on	ADV	_	1	advmod
3	to	to	PART	_	4	mark
4	learn	learn	VERB	VerbForm=Inf	1	xcomp
5	the	the	DET	Definite=Def PronType=Art	6	det
6	facts	fact	NOUN	Number=Plur	4	obj
7	.	.	PUNCT	_	1	punct

Sentence level - parsing



- Course 1 Admin, intro to NLP
- Course 2 Introduction to basic notions about language
- Course 3 Language Models
- Course 4 Tagging problems and dynamic graphical models (HMMs).
- Course 5 Annotated treebanks, Statistical Parsing
- Course 6 Introduction to Distributional Semantics.
- Course 7 Word embeddings: Word2vec.
- Course 8 The Transformer, BERT, GPT-3
- Course 9 Semantic roles
- Course 10 Summary and revision, exam prep

Enseignantes Paola Merlo, Haozhou Wang

Heures de réception sur rendez-vous

Horaire Cours Jeudi 10-12 salle 316 Battelle

TP Jeudi 12-14 salle 319 Battelle

Evaluation La note finale sera basée sur les TPs (30%) et l'examen finale (70%).

- Admission** Sont admis au cours les étudiantEs de MA en Informatique pour les Sciences Humaines, et les étudiantEs de la maitrise en Sciences.
- Langue** Le cours a lieu en français mais une bonne compréhension passive des l'anglais pour les lectures est indispensable. Les trasparents seront aussi en anglais.

- Le plagiat – la copie du travail autrui sans citation des sources – est interdit et passible de sanctions.
Le plagiat aux TPs entraîne la note 0 pour tous les TPs.
- Ceci s'applique tant aux textes écrits que aux programmes
- Le plagiat à l'examen entraîne la note 0 et l'annulation de la session d'examen.

Support du Cours et Bibliographie

Support Documents distribués sur Moodle et sur les web.

Vos notes. Ce cours n'a pas de polycopie, vous devez prendre des notes des discussions au cours si vous voulez vous en souvenir.

Références

- B. Habert, A. Nazarenko, et A. Salem, *Les linguistiques de corpus*, Armand Colin
- T. Mc Enery et A. Wilson, *Corpus Linguistics*, Edinburgh Press
- C. Manning et H. Schuetze, *Foundations of Statistical Natural Language Processing*, MIT Press
- D. Jurafsky et J. Martin, *Speech and Language Processing*, Prentice Hall
- Jacques Moescheler, Antoine Auchlin, *Introduction à la linguistique Contemporaine*, Colin, 2005
- Victoria Fromkin editor *Linguistics: an Introduction to Linguistic Theory*, Blackwell, Part 2, chapter 3.
- Emily M. Bender *Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax* Morgan-Claypool, San Rafael, CA, USA. Sous réserve à la bibliothèque