

Traitement Automatique du Langage Naturel

Doter les ordinateurs d'une capacité de compréhension et de production d'énoncés en langue naturelle écrite/parlée/signée

Patrick Paroubek

LISN (Laboratoire Interdisciplinaire des Sciences du Numérique)
CNRS - U. Paris-Saclay, Campus universitaire bât. 507, Rue du Belvédère, Orsay
patrick.paroubek@lisn.upsaclay.fr *https://perso.limsi.fr/pap*

27/10/2023 - Confs. de rentrée du dépt. info. de l'ENS Paris-Saclay

Plan

1. Langage naturel, langues et technologies
2. Abrégé de l'histoire du TAL
3. TAL et apprentissage

1 Langage naturel, langues et technologies

Evolution technologique

- ▶ le site WEB *Ethnologue*¹ rapporte l'existence de 7168 langues pour 7,6 milliards de personnes dans le monde.
- ▶ l'Inde a 22 langues officielles, mais 270 langues maternelles².
- ▶ l'Europe compte 24 langues officielles (et deux langues d'états : le luxembourgeois et le turc, parlé à Chypre).
- ▶ en France il y a 140 variétés de langues régionales³.



Philippe Boula de Mareüil, Albert Rilliard, Frédéric Vernier, A Speaking Atlas of the Regional Languages of France, LREC 2018,

<https://aclanthology.org/L18-1652>

1. <https://www.ethnologue.com>
2. source : https://fr.wikipedia.org/wiki/Langues_en_Inde
3. <https://atlas.limsi.tr>

1/6 Langage naturel, langues et technologies

Evolution technologique

- ▶ Premiers systèmes d'écriture vers 3500 av. JC (cunéiforme en Mésopotamie, hiéroglyphes en Égypte).
- ▶ Vers 1452 impression de la bible de Guterberg à Mayence.
- ▶ construction à la Moore School of Electrical Engineering de l'U. Pennsylvania à partir de mai 1944 du premier ordinateur, l'ENIAC ⁴
- ▶ 1971, première spécification de FTP (File Transfer Protocole),
- ▶ 1993, premier navigateur WEB grand-public (Mosaic NCSA)
- ▶ 2018 (juin), l'ITU compte 4,2 milliards d'internaute soit 55% de la population mondial ⁵



<https://france3-regions.francetvinfo.fr/paris-ile-de-france/paris/>

[paris-games-week-plus-grand-salon-jeux-video-francais-debute-ce-mercredi-1742067.html](https://france3-regions.francetvinfo.fr/paris-ile-de-france/paris-games-week-plus-grand-salon-jeux-video-francais-debute-ce-mercredi-1742067.html)

4. <http://www.seas.upenn.edu/about-seas/eniac/>

5. source : https://fr.wikipedia.org/wiki/Internet_dans_le_monde

1/6 Langage naturel, langues et technologies

Premier ordinateur

ENIAC - 13/février/1946 - *Moore School of Elec. Engineering UPenn*



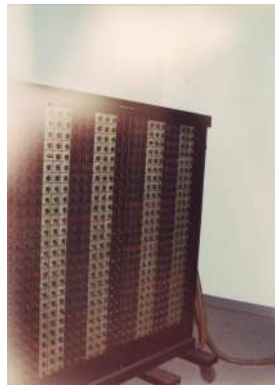
1



2



3



4

sources des images : 1,2,4 PParoubek ; 3 wikipedia

1/6 Langage naturel, langues et technologies

Genèse orale (et/ou proprioceptive)

- ▶ Les enfants disposent de capacités cognitives d'apprentissage à large spectre leur permettant d'apprendre n'importe quel langage.
 - ▶ 2 mois après la naissance, discrimination des phonèmes,
 - ▶ entre 5 et 12 mois, patrons d'accentuation et règles phonotactiques,
 - ▶ dès 12-14 mois interaction entre règles phonotactiques et processus lexico-sémantiques,
 - ▶ à 14 mois sémantique des mots dans des contextes imagés,
 - ▶ vers les 30 mois, sémantique des mots dans des contextes phrastiques,
 - ▶ dès 32 mois, réponses électrophysiologiques aux violations syntaxiques comparables à celles des adultes.

Les similitudes des réactions cérébrales observées chez les enfants et chez les adultes confortent l'hypothèse de la continuité dans le développement du langage.⁶

6. Angela D. Friederici, Marqueurs neurophysiologiques de l'acquisition précoce du langage : des syllabes aux énoncés, 2005, trad. fr.

<https://books.openedition.org/editions-cnrs/5957>

Développement du langage chez les enfants

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence : A roadmap for reverse-engineering the infant language learner. *Cognition*, 173,

43-59, doi :10.1016/j.cognition.2017.11.008., arXiv :1607.08723v4

E. DUPOUX

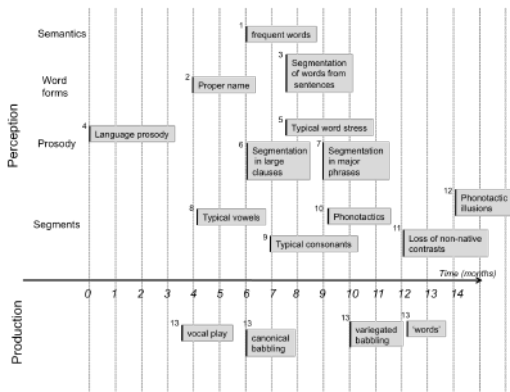


Figure 1. Sample studies illustrating the *time line* of infant's language development. The left edge of each box is aligned to the earliest age at which the result has been documented. ¹ Tincoff & Jusczyk, (1999); Bergelson & Swingley, (2012); ² Mandel et al. (1995); ³ Jusczyk & Aslin (1995) ⁴ Mehler et al. (1988) ⁵ Jusczyk et al. (1999) ⁶ Hirsh-Pasek et al. (1987) ⁷ Jusczyk et al (1992) ⁸ Kuhl et al. (1992) ⁹ Eilers et al. (1979) ¹⁰ Jusczyk et al. (1993) ¹¹ Werker & Tees (1984) ¹² Mazuka et al. (2011) ¹³ Stark (1980).

1/6 Langage naturel, langues et technologies

Genèse prénatale

- ▶ Dans les premiers mois de la vie, le système linguistique des enfants s'est adapté au système phonétique de leur langue maternelle.
- ▶ Des résultats de recherche récents⁷ sur la perception de la parole, il semble que les fœtus apprennent plus que ce que l'on pensait pendant la période de sa vie pré-natale.
- ▶ Il existe un ensemble grandissant d'observations qui laissent à penser que les nouveau-nés sont déjà familiarisés avec la prosodie des langues qu'ils entendent pendant leur vie pré-natale.
- ▶ L'apprentissage d'une langue, commencerait donc avant la naissance par un pré-apprentissage prosodique.

7. Judith. Gervains, The role of prenatal experience in language development, Volume 21, June 2018, pp 62-67, Current Opinion in Behavioral Sciences, <https://doi.org/10.1016/j.cobeha.2018.02.004>

1/6 Langage naturel, langues et technologies

Images et mots dans le cerveau

Des travaux récents en imagerie neuronale semblent indiquer que les représentations visuelles et linguistiques sont alignées dans le cerveau.

- ▶ Jordana Cepelewicz, *New Map of Meaning in the Brain Changes Ideas About Memory*, Quanta Magazine, février 2022,
<https://www.quantamagazine.org/new-map-of-meaning-in-the-brain-changes-ideas-about-memory-20220208/>
- ▶ Popham, S.F., Huth, A.G., Bilenko, N.Y. et al. Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat Neurosci* 24, 1628–1636 (2021).
<https://doi.org/10.1038/s41593-021-00921-6>,
<https://www.nature.com/articles/s41593-021-00921-6>
- ▶ Michael F. Bonner & Russell A. Epstein, Object representations in the human brain reflect the co-occurrence statistics of vision and language *NATURE COMMUNICATIONS* (2021) 12 :4081,
<https://doi.org/10.1038/s41467-021-24368-2>

Quelques caractéristiques du langage

Le langage est :

- ▶ **appris** ,
- ▶ d'abord **oral et dialogique** ,
- ▶ **linéaire** pour la langue parlée ou écrite,
et **multilinéaire** pour la langue des signes,
- ▶ **représenté** par des séquences de signes, symboles,...
- ▶ cotextuel et **contextuel**,
- ▶ **réflexif**,
- ▶ en **évolution** permanente et virtuellement **infini**,
- ▶ et **partiellement** formalisable (cf Noam Chomsky).

cœur 4 ou 5 caractères ?

une pomme de terre 2 ou 4 mots ? 1 ou 2 groupes syntaxiques ?

Le pilote ferme la porte en disant : "Ahh qu'elle est lourde !" .

1 ou 2 phrases ? **ferme** verbe ou adjectif ?

1/6 Langage naturel, langues et technologies

Histoire du TAL

- ▶ 1948 premières recherches sur la faisabilité de Traduction Automatique (TA) en Grande-Bretagne et aux Etats-Unis,
- ▶ 1954 premiers travaux en Union Soviétique,
- ▶ 1959 en France, création par le CNRS du CETAP (Paris, dir. Aimé Sestier) et du CETAG (Grenoble, dir. Bernard Vauquois)⁸ et naissance de l'ATALA⁹
- ▶ pour une histoire du TAL en France voir l'ouvrage¹⁰ de Jacqueline Léon, *Histoire de l'automatisation des sciences du langage*, Lyon, ENS Éditions, 2015,

8. <https://journals.openedition.org/histoire-cnrs/3461>

9. Association pour le Traitement Automatique des Langues
<https://www.atala.org>

10. <https://journals.openedition.org/rhsh/691>

1/6 Langage naturel, langues et technologies

Jeux de langage

- ▶ D'après le philosophe Ludwig Wittgenstein, l'apprentissage du langage se fait par le biais jeux interactifs avec les parents, les « **Language Games** », "*Philosophical Investigations*"(1953)
- ▶ The De 1999 à 2001, Luc Steels a dirigé l'expérience des « têtes parlantes »¹¹ (*Talking Heads*), une **population ouverte d'agent ancrés situés** et d'humains a créé un nouveau vocabulaire partagé en jouant à des **jeux de langage** concernant des scènes réelles qu'ils observaient.



11. Luc Steels, The Talking Heads experiment : Origins of words and meanings, 2015, <https://langsci-press.org/catalog/book/49>

Exemples de réalisations

Nous pouvons construire...

Des systèmes de traduction parole vers parole quasi-temps-réel¹²

	<i>Bùshí zǒngtǒng zài Mòsīkē yǔ Pǎijīng huìwǔ</i> 布什 总统 在 莫斯科 与 普京 会晤 Bush president in Moscow with/and Putin meet
(a) simultaneous: our wait-2	<i>...wait 2 words...</i> pres. bush met with putin in moscow
(b) non-simultaneous baseline <i>wait whole sentence</i> pres. bush met with putin in
(c) simultaneous: test-time wait-2	<i>...wait 2 words...</i> pres. bush in moscow and pol- ite meeting
(d) simultaneous: non-predictive	布什 总统 在 莫斯科 与 普京 会晤 <i>...wait 2 words...</i> pres. bush <i>wait 5 words</i> met with putin in moscow

12. <https://www.aclweb.org/anthology/P19-1289.pdf>

Exemples de réalisations

Nous pouvons construire...

Des systèmes de traduction parole vers parole quasi-temps-réel¹³

The screenshot shows a web browser window displaying the title page of a research paper. The title is "STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework". The authors listed are Mingbo Ma, Liang Huang, Hao Xiong, Renjie Zheng, Kaibo Liu, Baogang Zheng, Chuanqiang Zhang, Zhongjun He, Haifeng Liu, Xing Li, Beixi Wu, and Haifeng Wang. The paper is associated with Baidu Research, Baidu Inc., and Oregon State University. The abstract discusses simultaneous translation and proposes a wait-to-see framework. A diagram titled "Diagram" shows a grid representing source and target sentences with arrows indicating the flow of information and anticipation. The caption below the diagram explains that the model emits target words based on source-side prefixes and a wait-to-see strategy.

STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework*

Mingbo Ma^{1,2} Liang Huang^{1,2} Hao Xiong¹ Renjie Zheng¹ Kaibo Liu^{1,2} Baogang Zheng¹ Chuanqiang Zhang² Zhongjun He¹ Haifeng Liu¹ Xing Li¹ Beixi Wu¹ Haifeng Wang¹

¹Baidu Research, Sunnyvale, CA, USA
²Baidu, Inc., Beijing, China
³Oregon State University, Corvallis, OR, USA
{mingbo.ma, liang.huang, hao.xiong@baidu.com, renjie.zheng}@osu.edu

Abstract

Simultaneous translation, which translates sentences before they are finished, is useful in many scenarios but is notoriously difficult due to word-order differences. While the conventional seq-to-seq framework is only suitable for full-sentence translation, we propose a wait-to-see framework for simultaneous translation that implicitly learns to anticipate in a single translation model. Within this framework, we present a very simple yet surprisingly effective "wait-to-see" policy trained to generate the target sentence concurrently with the source sentence, but always k words behind. Experiments show our strategy achieves low latency and reasonable quality (compared to full-sentence translation) on 4 directions: zho-en and de-zho.

1 Introduction

Diagram

Figure 1: Our wait-to-see model emits target word y_t given source-side prefix $x_1 \dots x_{t+k-1}$, often before using the corresponding source word (here $x_t=2$, suggesting y_t ="red" before zho ="house"). Without anticipation, a 5-event wait is needed (dashed arrows). See also Fig. 2.

13. <https://www.aclweb.org/anthology/P19-1289.pdf>

Exemples de réalisations

Nous pouvons construire...

Des avatars dialoguants (simulateur de patient en consultation),¹⁴

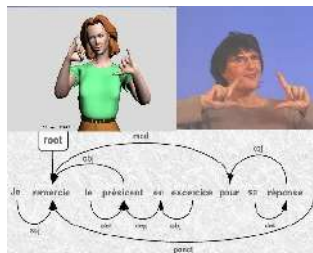


14. <https://pvdial.limsi.fr>

Exemples de réalisations

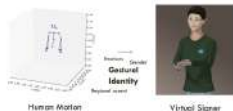
Les langues des signes

Formalisation¹⁵, analyse, reconnaissance, transcription, génération, anonymisation¹⁶, etc.



Signing Avatar Motion: Combining Naturality and Anonymity

Filix Dignel filix.dignel@univ-lille.fr LIML/CMIS, Paris-Saclay University, Paris-Saclay University Orsay, France	Antoine Brault antoine.brault@univ-lille.fr LIML/CMIS, Paris-Saclay University Orsay, France
Élise Prigent elise.prigent@univ-lille.fr LIML/CMIS, Paris-Saclay University, Paris-Saclay University Orsay, France	Bastien Berret bastien.berret@univ-lille.fr LIML/CMIS, Paris-Saclay University, Paris-Saclay University, ICF Orsay, France



15. <https://www.sign-lang.uni-hamburg.de/lrec/pub/22014.pdf>

16. <https://hal.science/hal-02400929/document>

Mais toujours pas d'ordinateur conscient

Le projet CYC (“encyclopédie”) de Doug Lenat (Cycorp Inc., Austin-Texas, 1984), qui espérait rendre “conscient” un ordinateur par la simple accumulation de règles d’inférences du sens commun dans un système expert.

La base de connaissances est passée de 1 million en 1994 à 24.5 million de règles pour sa dernière version en 2017, elle représente un effort de 1.000 personne-an (source : <https://en.wikipedia.org/wiki/Cyc>)

Le TAL et l'apprentissage...

From language to knowledge (and back)

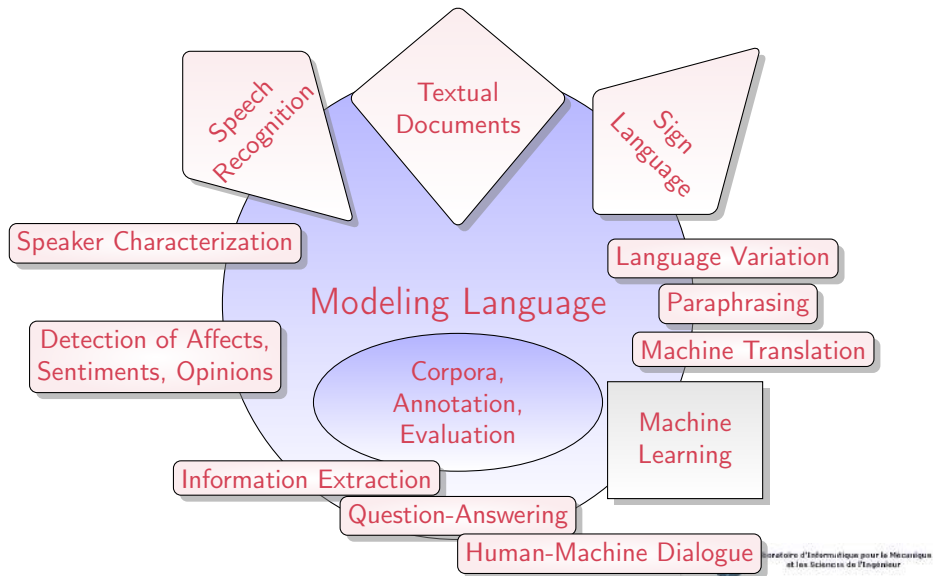
Context: wealth of information and knowledge in language form



Fundamental Issues

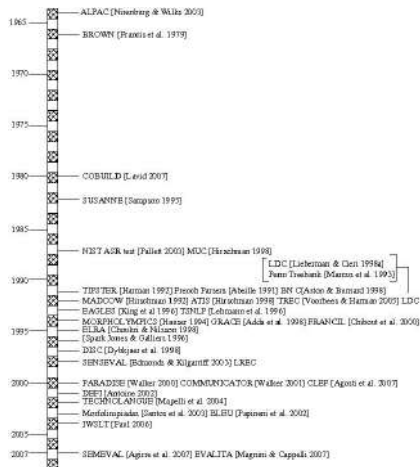
- Semantic representation vs linguistic variation
- Language modalities
- Understanding — inference — synthesis

Exemples de tâches en TAL



L'arrivé des corpus et de l'évaluation en TAL

https://perso.limsi.fr/pap/hdr_memoir_pap.pdf



Evolution du TAL

- ▶ approches symboliques
- ▶ approches statistiques
- ▶ approches neuronales profondes

Joseph-Jean Mariani, Gil Francopoulo, Patrick Paroubek, The NLP4NLP Corpus (I) : 50 Years of Publication, Collaboration and Citation in Speech and Language Processing (2019) <http://dx.doi.org/10.3389/frma.2018.00036>

Joseph-Jean Mariani, Gil Francopoulo, Patrick Paroubek, Frédéric Vernier, The NLP4NLP Corpus (II) : 50 Years of Research in Speech and Language Processing (2019) <http://dx.doi.org/10.3389/frma.2018.00037>

Une grammaire est un ensemble de règles qui définissent un langage. On peut définir une grammaire comme un système de réécriture qui part d'un symbole unique, source de tous les messages appartenant au langage que l'on peut générer en appliquant les règles de réécritures. On distingue les symboles intermédiaires (non-terminaux), S, D et U dans l'exemple ci-dessous, et les symboles terminaux 0 et 1 dans l'exemple. Par ex. la grammaire suivante génère une suite arbitraire de 1 se terminant toujours par un 0.

$$S \rightarrow 1D$$
$$D \rightarrow 1U$$
$$D \rightarrow 1S$$
$$U \rightarrow 0$$

Les langages peuvent être classés en fonction des caractéristiques des règles qui les définissent selon une hiérarchie dite de de Chomsky (aussi appelée hiérarchie de Chomsky-Schützenberger), de la classe des plus complexe (dite de type 0), qui regroupe les langues naturelles, jusqu'à la classe la plus simple, celle des langages réguliers (dits de type 3), qui sont définis par les expressions régulières (et reconnus par les automates à états finis déterministes).

type 0	pas de restriction sur les règles	langues naturelles
type 1	la partie droite des règles doit contenir au moins autant de symboles que la partie gauche ¹⁷	langages sensibles au contexte
type 2	toutes les règles doivent avoir un seul symbole non-terminal en partie gauche et il n'y a pas de restriction sur les parties droites.	langages hors-contextes
type 3	deux types de règles sont admises, tous deux avec un unique symbole non-terminal en partie gauche de règle : les règles dont la partie droite est uniquement composée de symboles terminaux et les règles qui contiennent un seul symbol non terminal soit en première soit en dernière position de partie droite sans contrainte sur le nombre de symboles terminaux présents de la partie droite	langages réguliers

17. Il est possible d'avoir une règle ne générant au symbole à condition que la partie gauche soit seulement le symbole initial et que ce dernier n'apparaisse jamais dans une partie droite de règle.

Annotations syntaxiques

les principales représentations

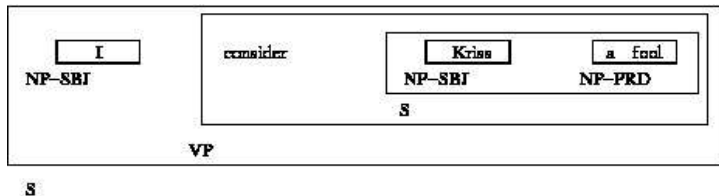
Le but de l'analyse syntaxique automatique est de fournir une analyse complète ou partielle de la structure d'un énoncé en termes de :

- ▶ “**chunks**” (lit. morceaux), des séquences de mots ayant une signification/fonction syntaxique commune,
- ▶ **constituants**, des séquences de mots qui représentent une unité fonctionnelle au sein d'une structure hiérarchique,
- ▶ **dépendances**, des relations liant un mot donné (appelé “**la tête**”) et un de ses dépendants,
- ▶ ou tout autre représentation propre à la théorie syntaxique utilisée...

Un exemple d'analyse en constituants extrait du corpus arboré (corpus avec des annotations représentant des arbres syntaxiques) PennTreebank pour l'anglais, pour l'énoncé :

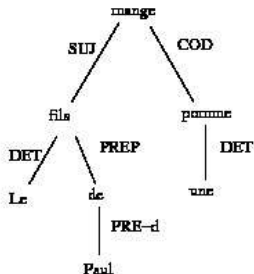
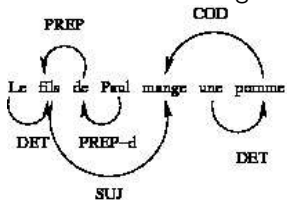
" *I consider Kriss a fool* " .

- ▶ S : clause déclarative simple (pas introduite par une conjonction de subordination ni par un pronom interrogatif et qui n'a pas d'inversion sujet-verbe)
- ▶ NP-SBJ : constituant nominal sujet de surface (sujet syntaxique),
- ▶ VP : constituant verbal
- ▶ NP-PRD : constituant nominal prédicatif

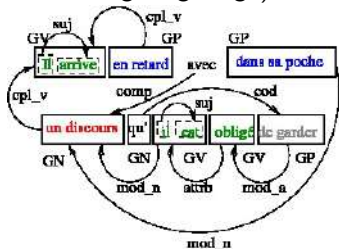


Un exemple d'analyse en dépendances avec les annotations de l'analyseur SYNTAX (D. Bourrigault) :

“Le fils de Paul mange une pomme”.



Le formalisme de la campagne d'évaluation PASSAGE des analyseurs syntaxiques du français (2006-2009) associe dans sa représentation un niveau "chunks" (groupes syntaxiques) et des relations grammaticales (triplets ou quadruplet de la forme relation-arg1-arg2 ou relation-arg1-arg2-arg3).



Groupes Syntaxiques

- GV = groupe verbal
- GP = groupe prépositionnel
- GN = groupe nominal

Relations syntaxiques

- suj = sujet → verbe
- cpl_v = complément → verbe
- comp = complémentateur → complémenté
- mod_n = modifieur → nom
- atrb = attribut → verbe
- mod_a = modifieur → adjectif
- cod = cod → verbe

En analyse automatique, les représentation syntaxiques combinent des éléments atomiques du formalisme syntaxique au moyen d'une ensemble d'opérateurs ou règles sur les structures syntaxiques pour construire l'arbre syntaxique qui correspond à un énoncé (approche de la grammaire générative). Par exemple le formalisme des grammaires d'arbres adjoints (Tree Adjoining Grammars) de A.K. Joshi, combine des arbres au moyens de 2 opérations :

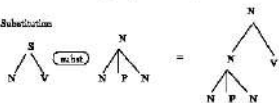
- ▶ la substitution (remplacement d'une feuille par un arbre dont la racine porte la même étiquette que la feuille qu'il remplace)
- ▶ l'adjonction remplacement d'un noeud interne de l'arbre par un arbre dont la racine porte la même étiquette que la feuille qu'il remplace et dont l'une des feuille porte la même étiquette que la racine de l'arbre.

Analyse syntaxique

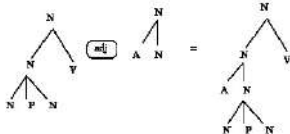
TAG

Tree Adjoining Grammar Operations

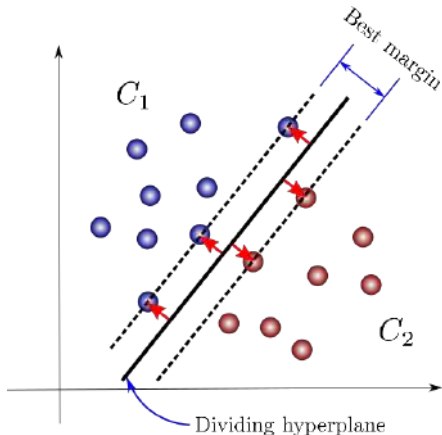
Substitution



Adjoining



Features based Machine Learning systems (1990-2014)



Machine Learning modules - Limitations

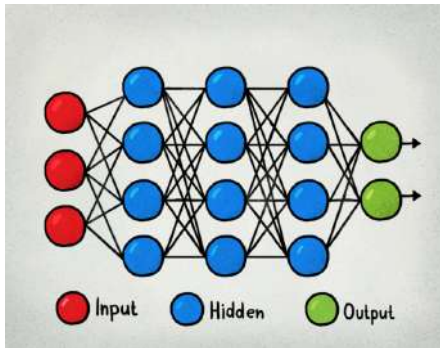
- System performance depends mainly on input features.
- Several NLP tools are used for extracting those features.
- Domain expertise is required for feature extraction.
- NLP tools used for pre-processing may contribute to error propagation.

Apprentissage neuronal

- ▶ 1949 Donald Hebb : règle de modification des coefficients synaptiques
- ▶ 1957 Frank Rosenblatt : le modèle du perceptron
- ▶ 1969 Marvin Lee Minsky & Seymour Papert : perceptron sans couche cachée = classification linéaire)
- ▶ 1986 Rumelhart & McClelland : perceptron multicouches, retropropagation
- ▶ 1990 J. L. Elman : réseaux convolutionnels et récurrents
- ▶ 2012 début du raz-de-marée de l'apprentissage profond

Regardons maintenant de plus près l'apport au TAL des derniers progrès des modèles neuronaux profonds avec BERT (encodeur-décodeur) et les plongements lexicaux (*word embeddings*) de grande dimension (*thèse de Sanjay Kamath, Question-réponse utilisant des données et modèles hybrides, 06/02/2020, <http://theses.fr/s163545>*).

Neural Networks based systems (2014-2018)



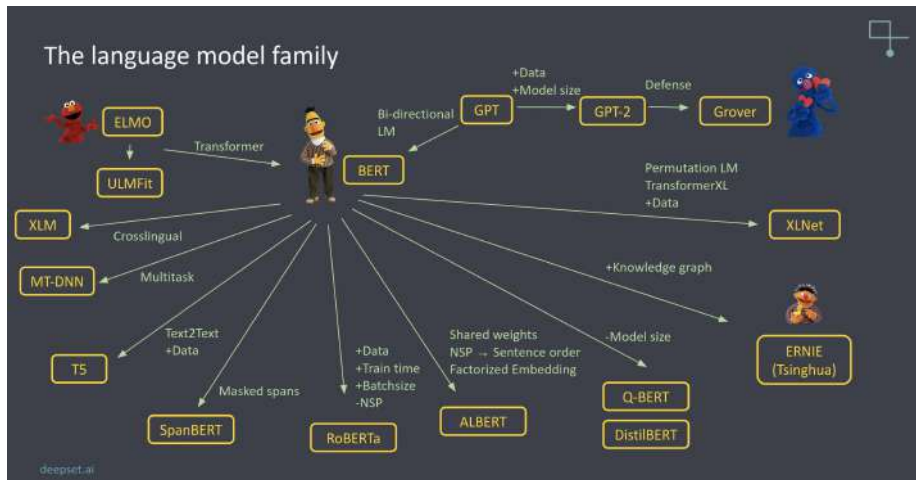
Deep Learning modules - Limitations

- Choosing right architecture, hyper parameters, optimizer etc. is very important.
- Feature Engineering → → Architecture Engineering

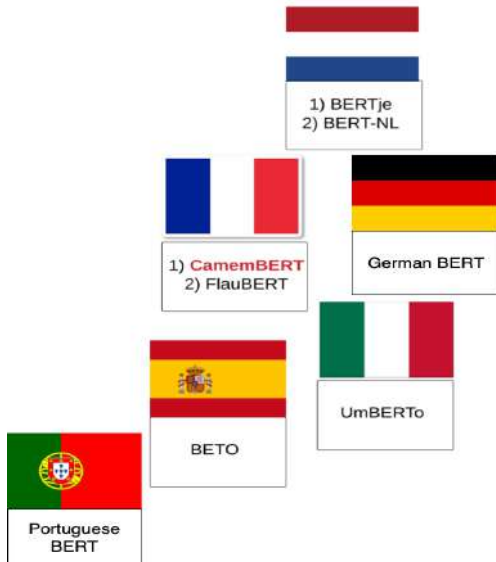
Large Scale Language Models (2018 - Present)



Large Scale Language Models (2018 - Present)



Large Scale Language Models (2018 - Present)



- ▶ Les systèmes de dialogue (chatbots) se divisent en deux grande famille : les systèmes orientés-tâche pour des applications comme la réservation de restaurant, de voyage, la vente en ligne, etc. d'une part et d'autre part les systèmes conversationnels (systèmes de chit-chat) comme Siri, Alexa, Ok Google et plus récemment ChatGPT.
- ▶ Historiquement les systèmes de dialogue orientés tâche ont une achitecture divisé en 3 parties :
 1. transformation du signal d'entrée (audio, text, vidéo) d'un énoncé de l'utilisateur en représentation interne (transcription),
 2. gestion du dialogue (stratégie du dialogue, cohérence et complétude de la tâche) (interprétation et calcul de la réponse),
 3. génération du message de retour en réponse à l'énoncé de l'utilisateur (conversion de la réponse en signal de sortie).

mais depuis les progrès récents des modèles de langage les approches dites de bout-en-bout (end-to-end) qui apprennent à prédire, directement à partir du signal d'entrée, la réponse à l'utilisateur, sont une alternative à l'architecture classique.

voici dans ce qui suit, quelques transparents présentant l'architecture type d'un système de dialogue orienté tâche et les données qu'il utilise (corpus annotés), issus de la soutenance de la thèse de Léon-Pau Schaub¹⁸, *Dimensions mémorielles de l'interaction écrite humain-machine : une approche cognitive par les modèles mnémoniques pour la détection et la correction des incohérences du système dans les dialogues orienté-tâche* (2022).

18. <https://www.theses.fr/2022UPASG023>

Dialogue chez Akio



le colis semble avoir été **perdu** lors du transport. Je vais procéder **immédiatement au remboursement**



je vous remercie, j'avais payé avec **paypal**, vous me remboursez par paypal ?



je viens de procéder au remboursement **via paypal**



merci beaucoup, mais je reste **très décue** car votre site ne m'a pas avertie du problème. Je pense que vous venez de perdre une cliente.

Dialogue dans DSTC2 [Henderson et al., 2013]



I need a **moderately** priced restaurant



What part of town do you have in mind?



Um doesnt matter I want **panasian**



What part of town do you have in mind?



I need panasian

Dialogue dans MultiWOZ [Budzianowski et al., 2018]



Yes, I need a place to stay that is
expensive

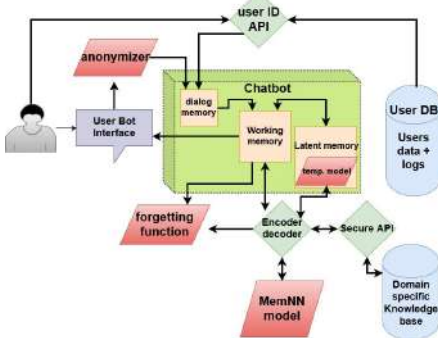
I have 5 different hotels that meet
your needs.

Do all of them include **free**
parking?

Yes, all of these hotels include
parking.

Ethique et aspects légaux [schaub et al.,2020]

GCCP : (GDPR Compliant Chatbot Protocol)



- Nouvelles réglementations (RGPD)
- Anonymisation
- Séparation des informations et des données personnelles
- Fonction d'oubli
- Pas d'accès direct à la base de données

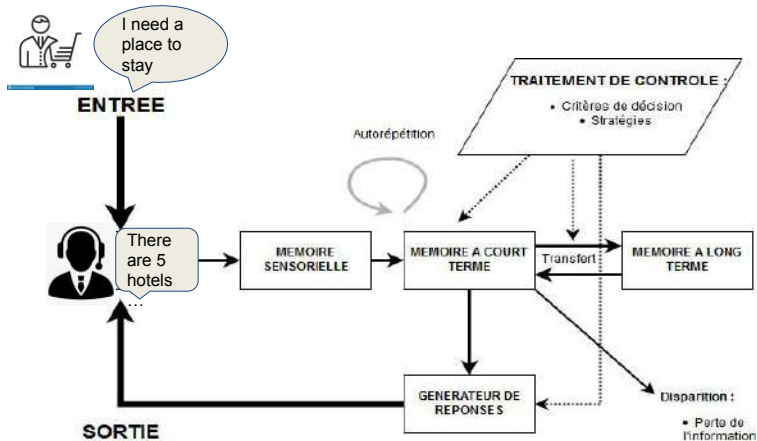
Dialogue et Tâche

<p>I need a place to stay that is expensive</p> <p>I have 5 different hotels that meet your needs.</p>	<p>Modèle de langue</p> <p>Production langagière</p> <p>Progression dans le dialogue</p>
<ul style="list-style-type: none">• You are looking for a place to stay.• The hotel should be in the expensive price	<p>Contexte opérationnel</p> <p>Etat de croyance</p> <p>Progression dans la tâche</p>

Modèle de mémoire [Atkinson et Shiffrin, 1968]

Nom de mémoire	Durée	Fonction
Perceptive/sensorielle	< 1 seconde	Enregistre les percepts
De Travail	entre 30-60 secondes	Transforme les informations pour réaliser une tâche
Episodique (déclarative)	illimitée	Retient les événements autobiographiques
Sémantique (déclarative)	illimitée	Connaissance du monde
Procédurale (non-déclarative)	illimitée	Apprentissage des actions

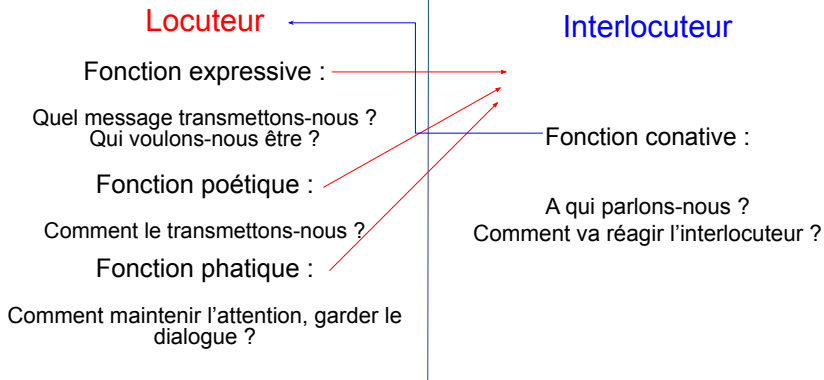
La mémoire dans le dialogue [Schaub et Vaudapiviz, 2019]



La Co-correction dans le dialogue [Chernyshova, 2018]

	initiée	réparée
<i>“On se voit demain...”</i>		
auto (soi)	“euh..”	“Non, ce soir, pardon !”
hetero (l'autre)	“Tu es sûr ?”	“Tu veux dire ce soir, non ?”

Liens avec les fonctions langagières de Jakobson [Jakobson, 1960]



Erreur ou incohérence ?

● Erreur

- Situationnelle/ contextuelle
- Comportement humain
- Affecte le succès de la tâche



ce que je cherche c'est
Quatre en Rose



excusez moi, j'avais mal
compris.

● Incohérence

- Langagière/ sémantique
- Comportement mécanique
- Affecte la continuité dialogue



What I want is thai food !



I'm sorry there is no
fusion food.

Une typologie des incohérences [Schaub et al., 2021]

Compréhension

Gestion

Intention

dans quelle zone est-il situé ?

Erana est un bon restaurant.

Négation

pas de cuisine française

*Un bon resto français,
c'est ça ?*

Insultes

*Putain, je m'en fous
d'un resto fusion.*

Fin du dialogue

au revoir

Petit Paris sert une cuisine française.

Répétition

*Ok, mais quelle zone
est Erana?*

*Erana est un bon
restaurant.*

Entités

Je veux manger thai

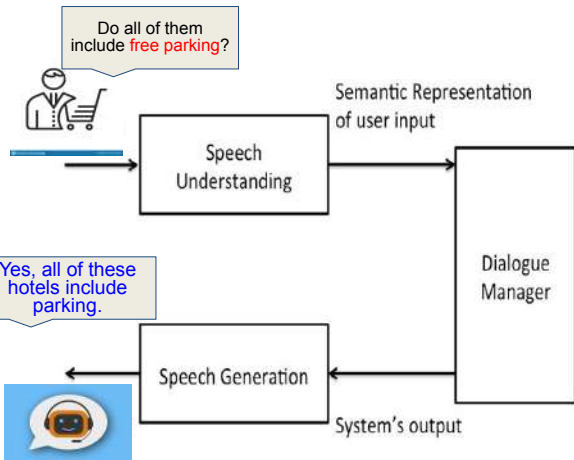
*On a pas de resto
français ici*

Incomplétude

*téléphone et adresse svp
Le téléphone est *****

Mauvaise action (API)

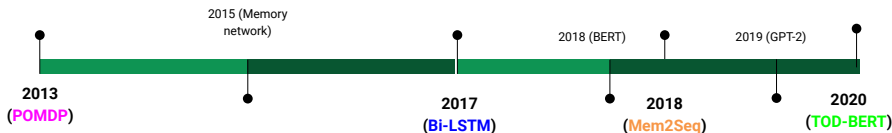
Système de dialogue orienté-tâche [Young et al., 2013]



1. Compréhension du langage
 - a. classification des actes
 - b. affectation des valeurs
2. Gestion du dialogue
 - a. pistage d'état
 - b. mise à jour de la politique du dialogue
3. Génération de texte
 - a. réponse extraite
 - b. ou générée mot à mot

Quelques systèmes de dialogue [Schaub et Vaudapiviz, 2019b]

- **POMDP** [Young et al., 2013]
(Young et al. "Pomdp-based statistical spoken dialog systems: A review." 2013)
- **Bi-LSTM** [Wen et al., 2017]
(Wen et al. "A Network-based End-to-End Trainable Task-oriented Dialogue System" 2017)
- **Transformers (TOD-BERT)**
[Wu et al., 2020]
(Wu et al., "TOD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogue" 2020)
- **Memory Network (Mem2Seq)**
[Madotto et al., 2018]
(Madotto et al., "Mem2Seq: Effectively Incorporating Knowledge Bases into End-to-End Task-Oriented Dialog Systems" 2018)



Corpus de travail

DSTC2 [Henderson et al., 2013] :

- 3000 dialogues
- domaine du restaurant
- homme-machine
- annoté en :
 - actes de dialogue
 - scénario
 - entités
 - succès
 - contraintes et pré-requis
 - contient des incohérences

MultiWOZ [Budzianowski et al., 2018] :

- 10 000 dialogues
- 7 domaines différents
- Magicien d'Oz
- annoté en :
 - actes de dialogue
 - scénario
 - entités
 - contraintes et pré-requis
 - ne contient PAS d'incohérence

Comment détecter et éviter les incohérences ?

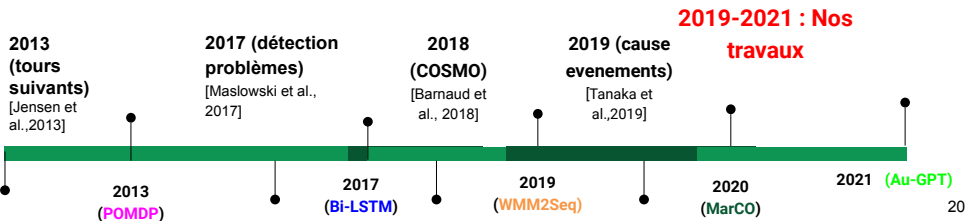
- . Humain se met à la place de son interlocuteur (COSMO)
- Est capable de se corriger en temps réel
- Peut prédire le temps restant pour la fin du dialogue
- La machine a besoin d'un modèle de l'autre
- La machine a besoin d'un détecteur de ses propres incohérences
- La machine a besoin d'un prédicteur

Positionnement de nos travaux

- **MarCO** [Wang et.al 2020]

- **Transformers (AuGPT)** [KulhaneK et.al 2021]

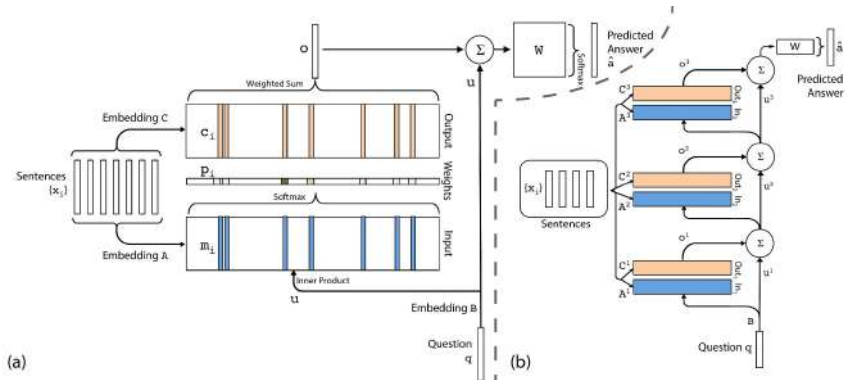
- **Memory Network (WMM2Seq)** [Chen et.al 2019]



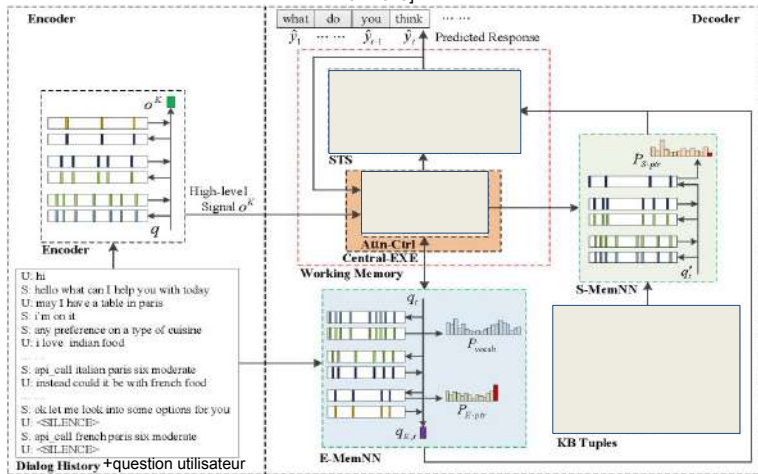
Les Memory Networks (MemNN) [Sukhbaatar et al.,2016]

- Réseaux récurrents
 - Avec sauts de réflexion
 - De bout-en-bout
 - Avec un système de pointeur pour les mots hors vocabulaire
- Utilisables dans plusieurs tâches de TAL
 - Question-réponse
 - Classification
- Adaptés au dialogue...
 - Gestion de l'historique de dialogue

Exemple de MemNN



Le WMM2Seq : système de bout-en-bout [chen et al. 2019]



Parmi les derniers développements¹⁹ très récents des *Large Language Models*, ChatGPT, LLama-2, etc il faut citer le *prompting*²⁰, qui consiste à utiliser un LLM pour effectuer une tâche dont l'entrée et la sortie sont en langue naturelle, sous-forme d'une question contextualisée :

- ▶ soit directement à partir d'un modèle pré-entraîné, sans effectuer d'apprentissage complémentaire spécifique (dans ce cas on parle de *zero-shot prompting*),
- ▶ soit en lui donnant avec la question un (ou quelques) exemple(s) de réponse(s) avec explications (dans ce cas on parle de *few-shots prompting*).

On parle de *prompt engineering* pour la conception des prompts, qui peut être manuelle ou automatique, voire obtenue par l'intermédiaire d'un modèle de langue (large).

19. <https://arxiv.org/pdf/2307.09288.pdf>

20.

Large Language Models (LLMs)

Exemple de prompting (source A. Stricker)

Voici dans les transparents suivants un exemple de prompting issu des travaux de thèse (en cours) d'A. Stricker (LISN), dans cette expérience, l'objectif est d'utiliser un LLM pour donner la capacité à un chatbot-orienté tâche d'être capable d'engager avec l'utilisateur des sous-dialogues de type conversationnel (sur des sujets sans rapport avec la tâche) tout en restant performant sur sa tâche, en la réalisant dans un temps raisonnable. L'interaction sera ainsi plus conviviale, et pourra donc être qualifiée de *plus humaine*.

Large Language Models (LLMs)

Prompting (source A. Stricker)

Can you help me find an expensive Thai restaurant?



Sala Thong is an expensive Thai restaurant on the west side. Would you like to book there?



That would be great. I need a table on Sunday at 15:45 for 2 people, please.
My wife and I are planning to fly to Thailand for our honeymoon. We wanted to go to a Thai restaurant to prepare our taste buds for the food served there !



Congratulations on your recent marriage! I hope you both have a wonderful time in Thailand.
I have made your reservation. Is there anything else I can assist you with today?



User Chitchat interference

System Chitchat Reaction

Large Language Models (LLMs)

Example de prompting (source A. Stricker)

The following are excerpts from conversations with an AI system. From the exchange, the user's utterances have been summarized to describe the user's situation. Importantly, information given by the user has not been changed, only condensed into a single passage, recounting the situation.

Exchange:

User: Recently, I have been given a long break by my boss to go on a holiday.

System: That's great news! There's a plethora of holiday destination out there, have you made up your mind?

User: I heard that Bishops Stortford is an ideal place for one to relax and enjoy.

System: Do you need help getting there?

Summarized Situation:

<situation>I was recently given a long break by my boss to go on holiday. I heard that Bishops Stortford is an ideal place for one to relax and enjoy. <endsituation>

x 3

- Instruction
- Few-Shot Examples

Large Language Models (LLMs)

Prompting (source A. Stricker)

Exemple d'architecture pipeline pour générer un mélange d'énoncés orientés-tâche et d'énoncés conversationnels à l'aide d'un LLM

