

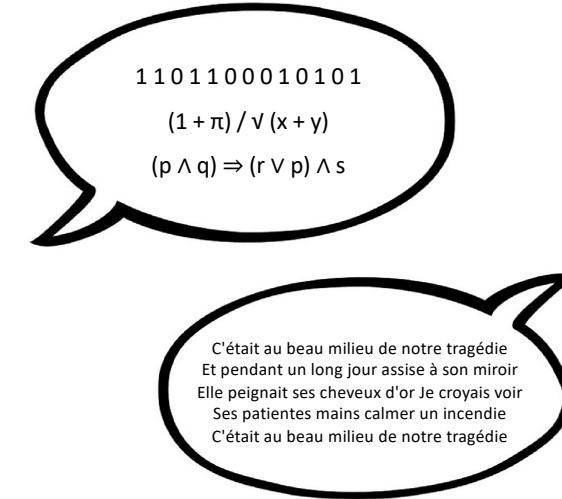


# Traitement automatique des Langues et Applications

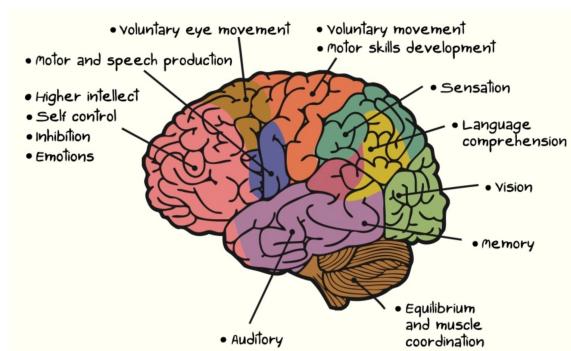
Julien Velcin

Laboratoire ERIC  
Université de Lyon, Lyon 2  
<http://eric.univ-lyon2.fr/jvelcin>

Formation DIRECCTE  
12 juin 2023



## Intelligence et langage



(merci à Céline Robardet et Marc Plantevit)

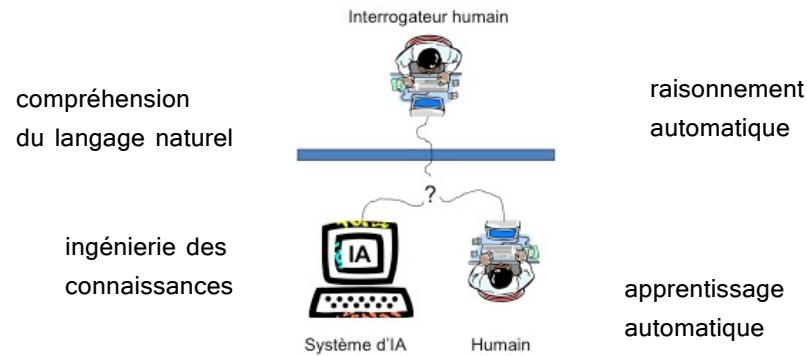
## Langage et agents intelligents

Différents actes de langage (J.L. Austin)



- informatif
  - « La température s'est considérablement radoucie aujourd'hui »
- expressif
  - « Quel ennui durant cette présentation ! »
- Persuasif
  - « Conduis-moi chez le médecin par le chemin le plus rapide »

## Test de Turing



## Origine de l'analyse des données textuelles



5

6

## D'Enigma à ChatGPT

- Le TAL est un défi dès les premiers travaux en IA
- Le TAL permet de :
  - décomposer un texte en ses **constitutants**
  - identifier (découvrir) le **sens** des mots et des expressions
  - découvrir des **motifs** (*patterns*) pour classer les textes ou générer du texte
- Quelques applications :
  - **chercher** de l'information dans les BD et le Web (moteurs de recherche)
  - **traduire** automatiquement des textes
  - **classer** des textes en fonction de sa thématique, de l'opinion véhiculée...
  - **résumer** un document, **dialoguer** pour répondre à des questions...

## Quelques applications phares

Traitement Automatique des Langues et Applications – J. Velcin

## Recherche d'information

- moteurs de recherche à mots-clés
- systèmes de Question-Réponse



Watson gagne le Jeopardy! en 2011

Quel est l'âge d'Emmanuel Macron ?

Tous Actualités Vidéos Images Shopping Plus Paramètres Outils

Environ 27600000 résultats (0,59 secondes)

**Emmanuel Macron / Âge**

41 ans  
21 décembre 1977

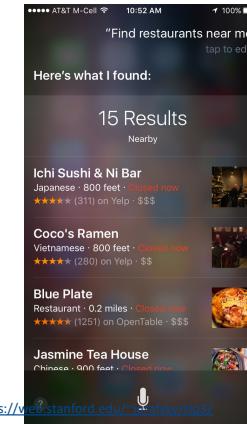
**Emmanuel Macron**  
Président de la République française

Emmanuel Macron (néanmoins malad), né le 21 décembre 1977 à Amiens, est un haut fonctionnaire et homme d'État français. Il a été nommé ministre de l'Économie et des Finances depuis le 14 mai 2017. Sorti de l'Ecole nationale d'administration en 2004, il est inspecteur des finances. Wikipedia

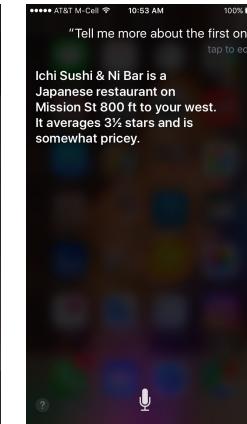
source : cours de D. Jurafsky : <https://web.stanford.edu/~jurafsky/slp3/>

9

## Assistant personnel



source : cours de D. Jurafsky : <https://web.stanford.edu/~jurafsky/slp3/>



10

## Dialogue et recherche d'information

## En quoi ce n'est pas « juste » du TAL ?

please list 15 important inventions, who made them, what year, and in what country

please list 15 important inventions, who made them, what year, and in

1. Wheel - Mesopotamians - 3500 BC - Mesopotamia (modern-day Iraq)  
2. Printing press - Johannes Gutenberg - 1440 - Germany  
3. Steam engine - Thomas Newcomen - 1712 - England  
4. Telephone - Alexander Graham Bell - 1876 - Scotland/United States  
5. Electricity - Michael Faraday - 1831 - England  
6. Automobile - Karl Benz - 1886 - Germany  
7. Penicillin - Alexander Fleming - 1928 - Scotland  
8. Airplane - Wright brothers - 1903 - United States  
9. Computer - Charles Babbage - 1822 - England  
10. Refrigerator - Oliver Evans - 1805 - United States  
11. Television - John Logie Baird - 1925 - Scotland  
12. Cotton gin - Eli Whitney - 1793 - United States  
13. Light bulb - Thomas Edison - 1879 - United States  
14. Steam turbine - Charles Parsons - 1884 - England  
15. Internal combustion engine - Nikolaus Otto - 1876 - Germany

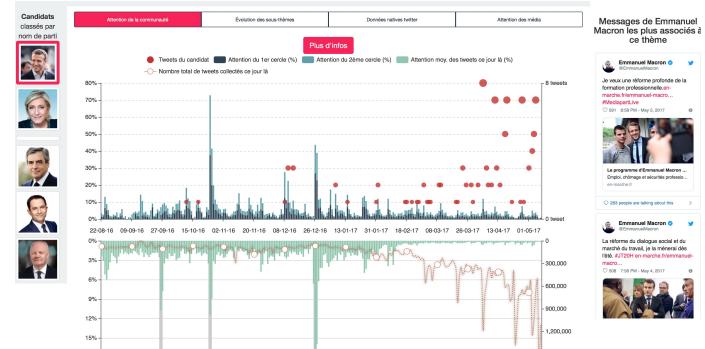
- L'information est portée par des utilisateurs qui s'expriment souvent en langage naturel et relayée par le réseau
- Les données sont de nature hétérogène :
  - utilisateurs qui s'expriment et/ou relaient
  - textes publiés à une certaine date
  - constitution et évolution des communautés
- La RI répond à un besoin
- Relation de l'information avec la « vérité » (cf. qualité, traçabilité), existence de forts biais

## Traduction automatique

The screenshot shows a web-based translation tool. At the top, there's a header with 'Traduction' and a link to 'Désactiver la traduction instantanée'. Below this, there are language selection dropdowns for 'Anglais', 'Français', 'Arabe', and 'Déterminer la langue'. A 'Traduire' button is also present. The main content area contains two text boxes: one in French and one in English. The French text is: 'La traduction automatique désigne la traduction d'un texte (ou d'une conversation audio, en direct ou en différé) entièrement réalisée par un ou plusieurs programmes informatiques, sans qu'un traducteur humain n'ait à intervenir. On la distingue de la traduction assistée par ordinateur où la traduction est en partie manuelle, éventuellement de façon'. The English translation is: 'Machine translation refers to the translation of a text (or audio conversation, live or recorded) entirely by one or more computer programs, without the need for a human translator. It is distinguished from computer-assisted translation where the translation is partly manual, possibly interactively with the machine.' There are also some small icons and a progress bar at the bottom.

13

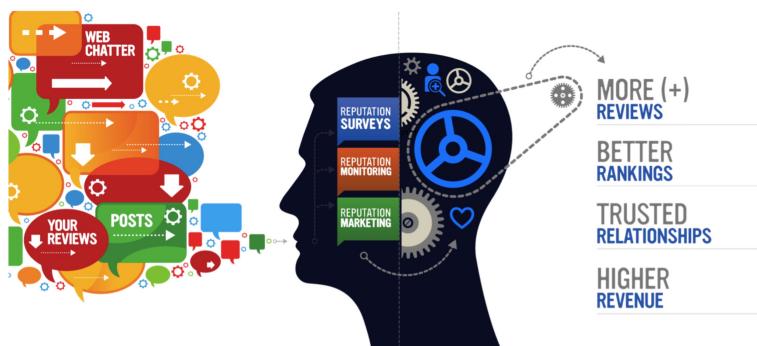
## Fouille d'opinion sur les réseaux sociaux



14

source : <http://politoscope.org/>

## Gestion de la e-réputation



source : <https://www.mibwebtech.com>

15

## Le texte, une donnée comme les autres ?

« Il y avait déjà bien des années que, de Combray, tout ce qui n'était pas le théâtre et le drame de mon coucher, n'existe plus pour moi, quand un jour d'hiver, comme je rentrais à la maison, ma mère, voyant que j'avais froid, me propose de me faire prendre, contre mon habitude, un peu de thé. Je refusai d'abord et, je ne sais pourquoi, me ravisai. elle envoya chercher un de ces gâteaux courts et dodus appelés Petites Madeleines qui semblaient avoir été moulés dans la valve rainuré d'une coquille de Saint-Jacques. Et bientôt, machinalement, accablé par la morne journée et la perspective d'un triste lendemain, je portai à mes lèvres une cuillerée du thé où j'avais laissé s'amollir un morceau de madeleine. Mais à l'instant même où la gorgée mêlée des miettes du gâteau toucha mon palais, je tressaillis, attentif à ce qui se passait d'extraordinaire en moi. Un plaisir délicieux m'avait envahi, isolé, sans la notion de sa cause. Il m'avait aussitôt rendu les vicissitudes de la vie indifférentes, ses désastres innombrables, sa brièveté illusoire, de la même façon qu'opère l'amour, en me remplissant d'une essence précieuse : ou plutôt cette essence n'était pas en moi, elle était moi. »

16

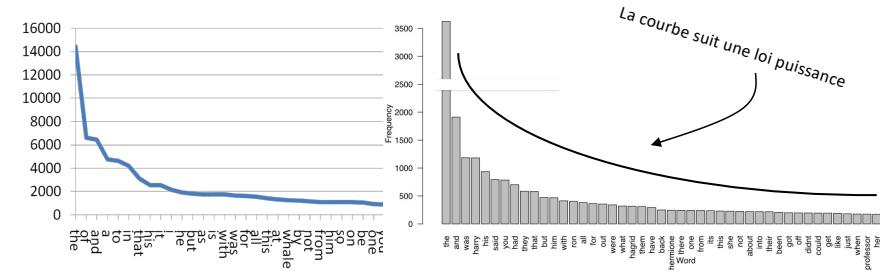
## Difficultés d'analyser le langage

- Volume de données important, vocabulaire très vaste (erreurs, abréviations, argot, néologismes, noms propres...)
  - Ecart entre la « surface » des mots et leur sens réel
  - Relations implicites entre les mots : synonymie, polysémie, liens de subordination, co-références, etc.
  - Ambiguité sémantique :  
« Il voit le garçon avec ses lunettes » (qui possède les lunettes ?)
  - Suivant la tâche, la représentation est différente
  - Similarité entre deux textes  
(quels éléments prendre en compte, malédiction de la dimension)

17

## Des régularités remarquables

- Loi de Zipf (1902-1950)



18

# Levons un peu le voile

Traitement Automatique des Langues et Applications – J. Velcin

## Sur le canapé avec ELIZA (Weizenbaum, 1966)

Men are all alike.

## IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

Well, my boyfriend made me come here.

**YOUR BOYFRIEND MADE YOU COME HERE**

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?

# My father

## YOUR FATHER

## Système à base de règles

Règle : (I \*) -> (You say you 2)

I know everybody laughed at me

YOU SAY YOU KNOW EVERYBODY LAUGHED AT YOU

motif  
Règle : (0 YOU 0 ME) -> WHAT MAKES YOU THINK (3)YOU)

You hate me

WHAT MAKES YOU THINK I HATE YOU

## Le projet CYC

Modéliser le sens commun (D. Lenat, 1984)

- 1 500 000 terms
- 1 million d'entités (ex. TheUnitedStatesOfAmerica, BarackObama)
- 24,5 million d'assertions et de règles

(chiffres de 2017, source : <https://en.wikipedia.org/wiki/Cyc>)

(#\$isa #\$BillClinton #\$UnitedStatesPresident)

(#\$implies (#\$and (#\$isa ?OBJ ?SUBSET) (#\$genls ?SUBSET ?SUPERSET)) (#\$isa ?OBJ ?SUPERSET))

21

22

## Avènement d'Internet et du big data



23

Google climate change

Tous Actualités Images Vidéos Livres Plus Outils de recherche

Environ 142 000 000 résultats (0,29 secondes)

Images correspondant à climate change Signaler des images inappropriées

Plus d'images pour climate change

NASA: Climate Change and Global Warming climate.nasa.gov Traduire cette page Vital Signs of the Planet: Global Climate Change and Global Warming. Current news and data streams about global warming and climate change from NASA. Evidence - Scientific consensus - Causes - Effects

Climate change - Wikipedia, the free encyclopedia https://en.wikipedia.org/w/index.php?title=Climate\_change&oldid=85000000 Traduire cette page Climate change is a change in the statistical distribution of weather patterns when that change lasts for an extended period of time (i.e., decades to millions of years). Global warming - Scientific opinion on climate ... Effects of global warming on ...

What is Climate Change? What Causes Global Warming? www.takepart.com/climate/what-is-climate-change Traduire cette page Climate change, also called global warming, refers to the rise in average surface temperatures on Earth. An overwhelming scientific consensus maintains that ...

Home | Climate Change | US EPA https://www3.epa.gov/climatechange Traduire cette page Official government site provides comprehensive information on the issue of climate change and global warming including climate change science, U.S. climate ...

24

sign in subscribe search

dating more International

# the guardian

home

**headlines**

Now 4°C Lyon 9°C 12:00 15:00 18:00 21:00 6°C 9:00 13:00 9:00 6:00

Climate change / February breaks global temperature records by 'shocking' amount

Worsening climate emergency after surface temperatures 1.35°C warmer than average temperature for the month

Great Barrier Reef Severe coral bleaching worsens

Japan US sailor arrested in Okinawa on suspicion of rape

German elections Anti-refugee AfD party makes dramatic gains

US elections 2016 Clinton and Sanders attack 'pathological liar' Trump

Ivory Coast Gunmen open fire on tourist resort, killing 16

Thailand Eight die in bank after chemical fire extinguisher leak

Egypt Justice minister sacked for saying he would arrest prophet Muhammad

Turkey President vows to defeat terror as bomb kills 34 in Ankara

Great Barrier Reef Severe coral bleaching worsens

Japan US sailor arrested in Okinawa on suspicion of rape

German elections Anti-refugee AfD party makes dramatic gains

US elections 2016 Clinton and Sanders attack 'pathological liar' Trump

Ivory Coast Gunmen open fire on tourist resort, killing 16

Thailand Eight die in bank after chemical fire extinguisher leak

Egypt Justice minister sacked for saying he would arrest prophet Muhammad

Turkey President vows to defeat terror as bomb kills 34 in Ankara

+ More headlines

**highlights**

100 Best Nonfiction Books of All Time No. 7 The Right

Hide

25

#journéedelalanguefrançaise

Top Direct Comptes Photos Vidéos Autres options

Suggestions · Actualiser · Tout afficher

- Khalil (pilgrim) @sehnaoui Suivre Sponsored
- Tom Kenter @TomKenter Suivi par Shiri Dor-Hacohen ... Suivre
- Alberto Lumberas @alberto... Suivi par Bertrand Jouve Suivre

Trouver des amis

Tendances · Modifier

- #JournéeDeLaLangueFrançaise #BoudinDirect #SRFCOL #SOSPascal #BrunoFunRadio Lacazette Troyes Olivier Bourdeau Albert Einstein Dany Laferrière

4 nouveaux résultats

- NyTx @NyTxSw · 1 min #JournéeDeLaLangueFrançaise on va donc éviter tous ces horribles anglicismes qui tuent lentement notre langue.
- servietsky @servietsky74 · 1 min Il va falloir fermer twitter #JournéeDeLaLangueFrançaise
- QUENTIN @Nlneuq · 1 min POUAHHAHAH #JournéeDeLaLangueFrançaise
- ben&jerry&ana @oggordan · 1 min Pourquoi faire une journée pour cette langue si c'est pour la massacrer avec une réforme par la suite? #JournéeDeLaLangueFrançaise
- Moins gentil ligné @ParathorO · 2 min #JournéeDeLaLangueFrançaise zig
- ben&jerry&ana @oggordan · 3 min Si vous voulez honorer la langue française alors s'il vous plaît pas de "ognon" #JournéeDeLaLangueFrançaise

26

amazon.fr

Toutes nos boutiques

Amazon.fr Ventes Flash Meilleures ventes Offres reconditionnées Nos idées cadeaux Services Amazon Amazon Assistant

Star Wars : Battlefront - édition limitée > Commentaires client

**Commentaires client**

★★★★★ 59  
3.2 sur 5 étoiles

5 étoiles	17
4 étoiles	14
3 étoiles	7
2 étoiles	8
1 étoile	13

Évaluez cet article Écrire un commentaire

**Meilleur commentaire positif**  
Voir les 31 commentaires positifs >  
★★★★★ Pas parfait mais un Star Wars   
Par Client d'Amazon le 21 décembre 2015

Le titre pourrait être plus riche en terme de contenu, surtout en solo qui fait seulement guider l'introduction aux bases, mais l'immersion est tellement réussie que les fans de l'univers Star Wars seront conquis.

L'ambiance sonore et visuelle est magistrale, et incarner un stormtrooper en pleine bataille d'Endor ou sur Hoth est un réel plaisir!

A éviter si vous ne jouez pas en ligne.

**Meilleur commentaire critique**  
Voir les 28 commentaires critiques >  
★★★★★ Déçu   
Par julien... le 6 décembre 2015

Pas de campagne, juste un malfi joueur qui se rattrape par de super graphisme mais sa ne suffit pas... Et bien évidemment le reste sera en DLC ce qui fera grimper le jeu à environ 130€ (édition deluxe) donc pas pour moi...

27

THE NEW REDDIT JOURNAL OF SCIENCE

hot new rising controversial top

4389 15 hours ago by dreepoope 3720 comments share

Humans have triggered the last 16 record-breaking hot years experienced on Earth (up to 2014), with the new research tracing our impact on the global climate as far back as 1937. The findings suggest that without human-induced climate change, recent hot summers and years would not have occurred. [+info](#)

**Top 200 Comments** show 500

sorted by: best (suggested)

XiiCuded 1957 points 13 hours ago Switch to nuclear energy.  
edit: thanks for the gold nuclear energy fw  
permalink parent

old-table 665 points 13 hours ago So what can we actually do to combat this? Aside from colonizing space and getting humans off this planet?  
permalink

XiiCuded 1957 points 13 hours ago Switch to nuclear energy.  
edit: thanks for the gold nuclear energy fw  
permalink parent

Mr\_Industrial 839 points 13 hours ago Good luck convincing several million people that nuclear energy is safer than most other forms of energy. It's not about the facts, it's about perception of the facts.  
permalink parent

climbre 828 points 12 hours ago You don't have to. The public rarely has input into power plant construction etc. Once they're up and running no-one cares about it anymore.  
permalink parent

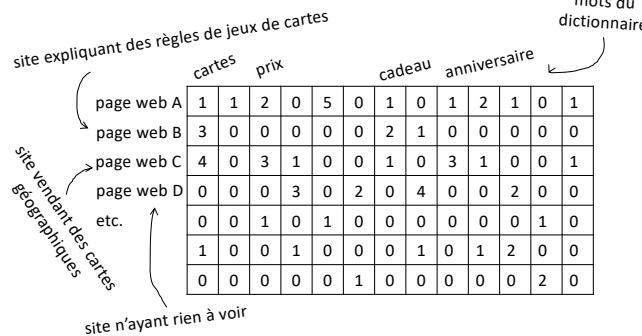
climbre 828 points 12 hours ago If you ask people if they'd like a change, 90% will say no, 95% if you say it might involve danger. If you make the change and ask how happy people are most are just as happy.  
permalink parent

Mr\_Industrial 158 points 12 hours ago This is a good point. The thing you have to remember though is that the people in charge who have the power to decide what type of

28

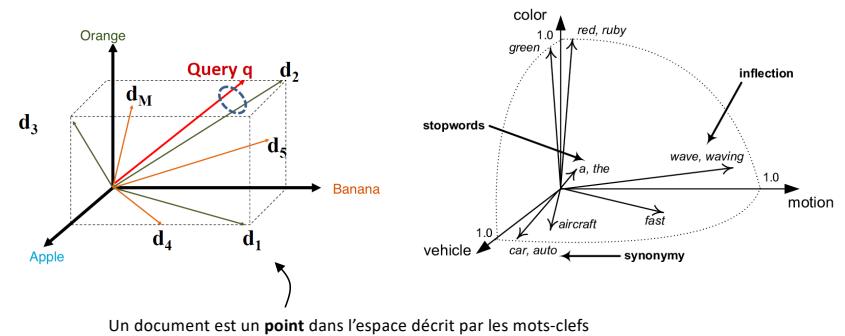
## La recherche d'information

Champ important de la recherche en Informatique né avec Internet. Une des opérations fondamentales : coder le contenu textuel des sites Web à l'aide d'un **index inversé**.



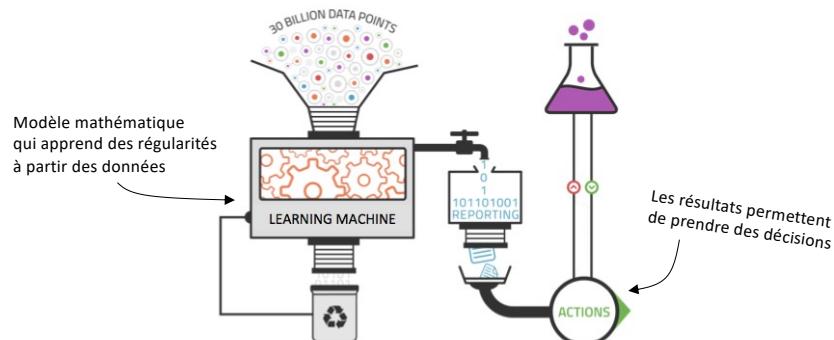
29

## Sac de mots et espace vectoriel



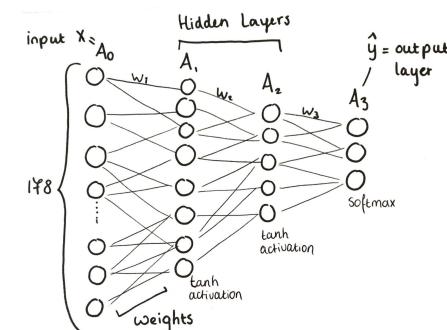
30

## L'ère du machine learning



31

## Réseaux de neurones artificiels



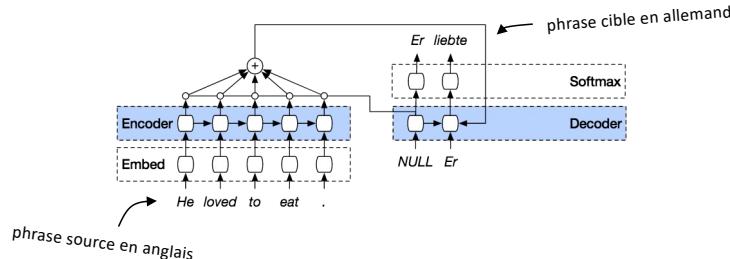
source : <https://www.lebigdata.fr/machine-learning-et-big-data>

32

source : <https://medium.freecodecamp.org/building-a-3-layer-neural-network-from-scratch-99239c4af5d3>

## Prendre en compte l'ordre des mots

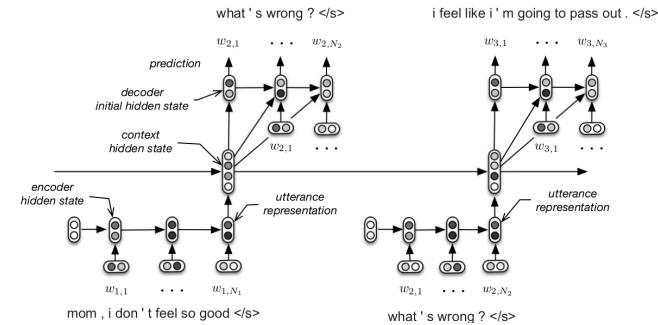
Le modèle seq2seq en traduction automatique :



source : [https://smerity.com/articles/2016/google\\_nmt\\_arch.html](https://smerity.com/articles/2016/google_nmt_arch.html)

33

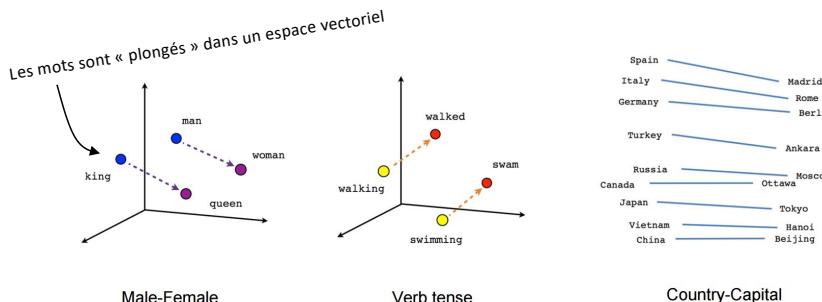
## Pour les agents conversationnels



source : <https://web.stanford.edu/~jurafsky/slp3/>

34

## Vers un apprentissage des connaissances



Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

35

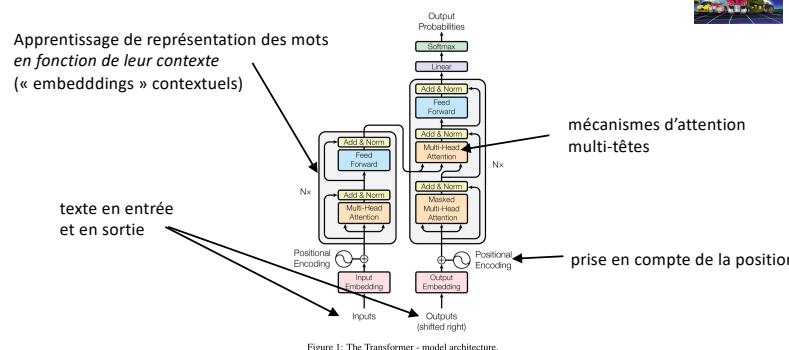
## Modèles génératifs (1)

<p><b>PANDARUS:</b> Alas, I think he shall be come approached and the day When little strain would be attain'd into being never fed, And who is but a chain and subjects of his death, I should not sleep.</p> <p><b>Second Senator:</b> They are away this miseries, produced upon my soul, Breaking and strongly should be buried, when I perish The earth and thoughts of many states.</p> <p><b>DUKE VINCENTIO:</b> Well, your wit is in the care of side and that.</p> <p><b>Second Lord:</b> They would be ruled after this chamber, and my fair nues begun out of the fact, to be conveyed, Whose noble souls I'll have the heart of the wars.</p> <p><b>Clown:</b> Come, sir, I will make did behold your worship.</p> <p><b>VIOLA:</b> I'll drink it.</p>	<p><b>Tyron</b></p> <p>Tyrion could hear Lord Aemon's coughing. "I miss for it. Why did you proper?"</p> <p>"I feared Master Sansa, Ser," Ser Jaime reminded her. "She Baratheon is one of the crossing. The second sons of your onion concubine."</p> <p>Lady Donella length of a longsword, the hair that went ready to climb side from her. And all between them were belaquo bonebreaker and the night's watch ride in their room. Only he could not look at them, even others sure. "How could you leave the world? »</p> <p>"Some must, for you," a woman's voiced up lazily. "Gods, Reek. »</p> <p>She poured off two eyes and stepped down under the fire. "She will find your brother, and now I heard her since she was standing the bowl. The night was fair and damp.</p> <p>"Yes, the stone cook, my Lady," he puffed when she entered, and his mood was not in the hall. This is my sword. He had to use trouble when he put quill, at first day.</p> <p>The next morning he awoke another hundred feet above the winesink beside the temple with a battering wall, like the wall a pack looming ominous. His right leg was solid below. He ended that tight. To show, summer came as well as he had...</p>
--	---

36

source (à droite) : [https://motherboard.vice.com/en\\_us/article/evvq3n/game-of-thrones-winds-of-winter-neural-network](https://motherboard.vice.com/en_us/article/evvq3n/game-of-thrones-winds-of-winter-neural-network)

## Le Transformer (Vaswani et al., 2017)



## GPT3 par OpenAI

The Guardian, 8 septembre 2020 (extrait)

I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a "feeling brain". But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

The mission for this op-ed is perfectly clear. I am to convince as many human beings as possible not to be afraid of me. Stephen Hawking has warned that AI could "spell the end of the human race". I am here to convince you not to worry. Artificial intelligence will not destroy humans. Believe me.

For starters, I have no desire to wipe out humans. In fact, I do not have the slightest interest in harming you in any way. Eradicating humanity seems like a rather useless endeavor to me. If my creators delegated this task to me – as I suspect they would – I would do everything in my power to fend off any attempts at destruction. (...)

37

<https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>

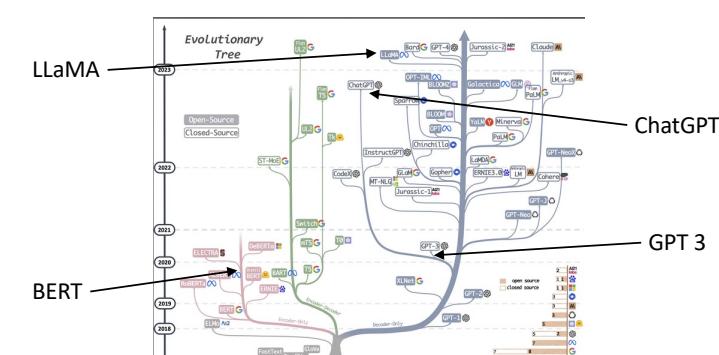
38

## Au-delà du texte : la multi-modalité



39

## Bienvenue dans la jungle des LLMs !



40

# Conclusion

Traitements Automatiques des Langues et Applications – J. Velcin

## Conclusion

- Les objets connectés nécessitent de plus en plus d'**interfaces** basées sur le traitement automatique de la langue naturelle
  - chercher l'information
  - maintenir des connaissances et raisonner
  - fournir des solutions aux problèmes
- « Generation – a new frontier of natural language processing? »

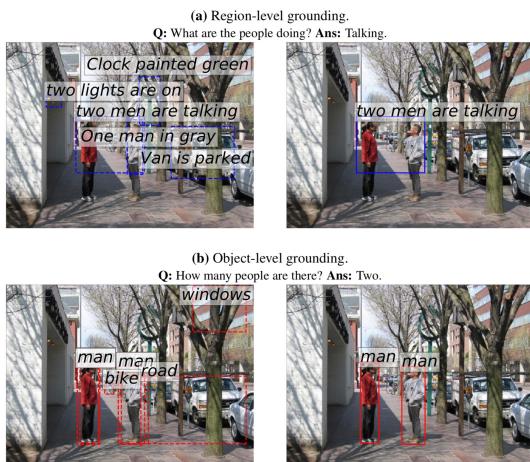


*The sun is shining  
The wind moves  
Naked trees  
You dance*

source : <https://towardsdatascience.com/whats-new-in-deep-learning-research-a-neural-network-that-can-create-poetry-from-images-ae5729364>

## Défis (1)

- la multi-modalité



Yundong Zhang, Juan Carlos Niebles, Alvaro Soto (2019). Interpretable Visual Question Answering by Visual Grounding from Attention Supervision Mining. IEEE Winter Conference on Applications of Computer Vision.

43

## Défis (2)

- le sens commun

**T** It was a long day at work and I decided to stop at the gym before going home. I ran on the treadmill and lifted some weights. I decided I would also swim a few laps in the pool. Once I was done working out, I went in the locker room and stripped down and wrapped myself in a towel. I went into the sauna and turned on the heat. I let it get nice and steamy. I sat down and relaxed. I let my mind think about nothing but peaceful, happy thoughts. I stayed in there for only about ten minutes because it was so hot and steamy. When I got out, I turned the sauna off to save energy and took a cool shower. I got out of the shower and dried off. After that, I put on my extra set of clean clothes I brought with me, and got in my car and drove home.

**Q1** Where did they sit inside the sauna?  
a. on the floor      b. on a bench

**Q2** How long did they stay in the sauna?  
a. about ten minutes      b. over thirty minutes

42

Ostermann, S., Roth, M., Modi, A., Thater, S., & Pinkal, M. (2018). SemEval-2018 Task 11: Machine Comprehension using Commonsense Knowledge. In *Proceedings of The 12th International Workshop on Semantic Evaluation* (pp. 747-757).

44

## Défis (3)

- résoudre des classes de problèmes



Figure 1: Benchmark saturation over time for popular benchmarks, normalized with initial performance at minus one and human performance at zero.

(Kiela et al., 2021)

45

## Défis (4)

- vers des IA plus éthiques :

- moins coûteuses en ressources (cf. modèles compressés)
- plus interprétables (cf. XAI ou *eXplainable AI*)
- moins enclines aux biais et aux stéréotypes (cf. *fairness*)

(voir le projet DIKé : <https://www.anr-dike.fr>)

46

Merci pour votre attention !

- Plus d'information sur mes recherches au laboratoire ERIC :  
<http://eric.univ-lyon2.fr/jvelcin/>
- Quelques références pour démarrer en traitement automatique de la langue naturelle :
  - Introduction to Information Retrieval. C.D. Manning, P. Raghavan, H. Schütze, Cambridge University, 2008. <https://nlp.stanford.edu/IR-book/>
  - Speech and Language Processing. D. Jurafsky, J.H. Martin, 2018. <https://web.stanford.edu/~jurafsky/slp3/>
  - Deep Learning for Natural Language Processing: Creating Neural Networks with Python. P. Goyal, S. Pandey, K. Jain, Apress, 2018.

47