

Formalisation des langues
Linguistique informatique
Linguistique computationnelle

max.silberztein@univ-fcomte.fr

CM :

Max Silberztein, *Formalisation des langues : l'approche de NooJ*. ISTE : Londres (420 pages)

TD :

Max Silberztein, the NooJ Manual. Téléchargeable à partir de la page **Downloads** du site

<http://www.nooj4nlp.org>

Quelques chaînes de tutoriels video sur YouTube

La linguistique

- linguistique vs. grammaire
- réflexions sur la langue déjà dans l'Antiquité, ex. Platon, la Kabbale
- linguistique moderne, Ferdinand de Saussure

Convergence d'intérêts

- Les linguistes veulent décrire les langues scientifiquement : « Sciences du langage »

Convergence d'intérêts

- Les linguistes veulent décrire les langues scientifiquement « Sciences du langage »
 - Autant d'écoles de linguistiques que de linguistes ?

Convergence d'intérêts

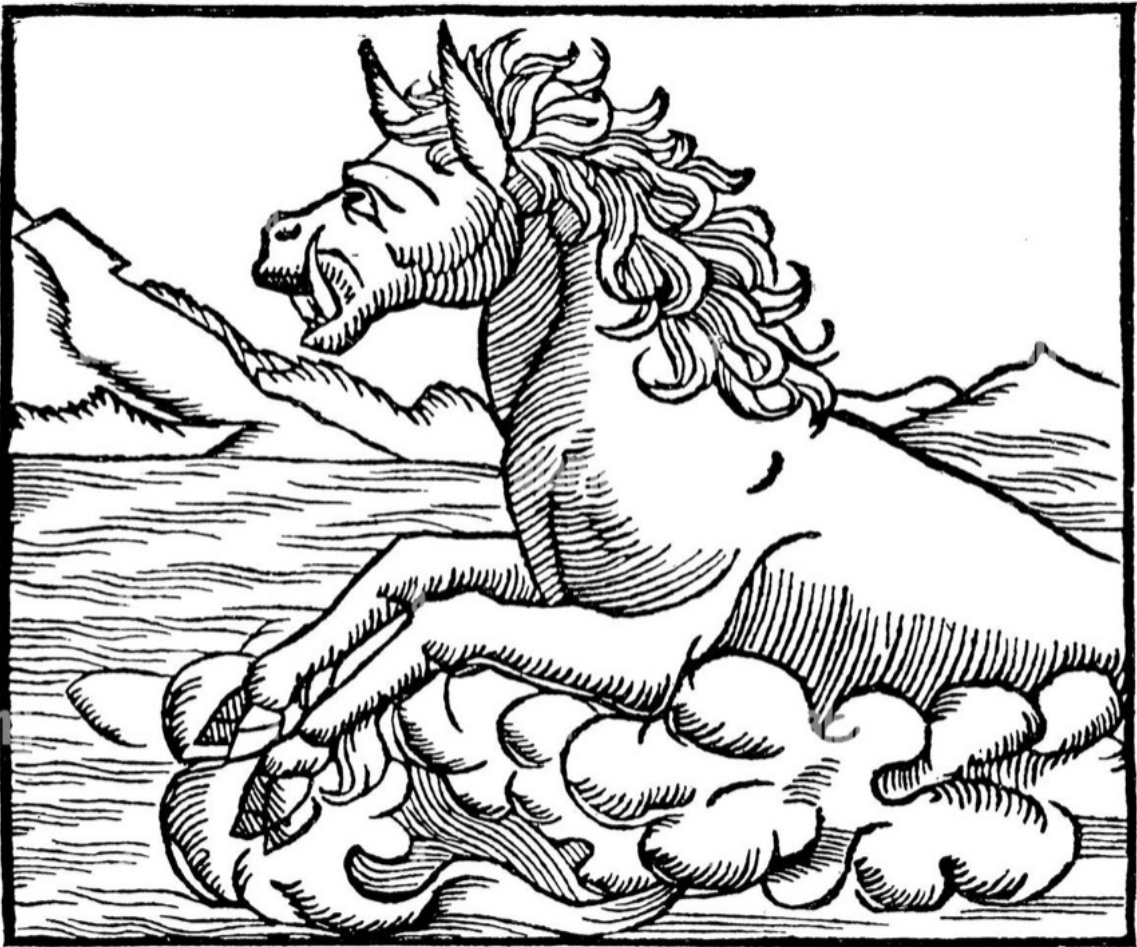
- Les linguistes veulent décrire les langues scientifiquement « Sciences du langage »
 - Autant d'écoles de linguistiques que de linguistes ?
- Les informaticiens veulent créer des logiciels capables de traiter les textes en langue naturelle

Convergence d'intérêts

- Les linguistes veulent décrire les langues scientifiquement « Sciences du langage »
 - Autant d'écoles de linguistiques que de linguistes ?
- Les informaticiens veulent créer des logiciels capables de traiter les textes en langue naturelle, mais sans avoir à construire des dictionnaires et des grammaires (ce qui est fatigant)
 - Pas de linguistique pour traiter les langues naturelles ?

La biologie avant le XVIIe

Hippopotame. Après Sebastian Münster "cosmographie" de 1550, 1902
Les légendes sont fournies par nos contributeurs.



attouchant coup sur coup. Car ilz afferment que c'est la premiere maison que Abraham edificia iamaix. Au milieu de la tour ya vn puitz fort profond d'ont il puisent de l'eau, & en iettent sur les hommes qui passent tout alentour, & en ce faisant les purgent de toute coulepe. En ces quartiers la, le ciel est couuert d'une infinité de pigeons. Car ilz disent qu'ilz tiennent leur race ancienne de la colombe, soubz l'espece de laquelle le saint esperit paroist a l'oreille de Mahomet. Ces oyseaux vont par les rues, maisons, boutiques & greniers a fourmient, mil, & ris, & a grand peine les peut on chasser. De les tuer ou prendre, c'est comme vn crime capital: car les habitantz estiment, que si on battoit ces oyseaux la, que le monde fouldroit. Pour tant on les nourrist a grandz fraiz au milieu du temple. De l'autre costé dudict temple on voit des parcs, esquelz on garde vne paire de licornes, lesquelles on monstre au peuple comme vn miracle. Ilz disent que ceste beste est semblable a vn poulain de 30 mois. Elle porte au front vne corne esleeue qui est noire et longue de deux ou trois coudées. La couleur est comme d'un cheual moustelle, la teste comme d'un cerf, le col n'est pas fort long; & a fort peu de crins, lesquels pendent seulement d'un costé. Ses jambes sont minces & gresles, les ongles des pieds de deuant fendues en deux comme vne chieure. La partie extérieure des jambes de derriere est velue, & pleine de poil.



On amene ces bestes du roy des Ethiopes, lequel par le moyen de ce present entretient amitié avec le Souldan de Meche. Plin dit que la Licorne est vne saunage beste, fort aspre, & ressemble de tout le reste du corps a vn cheual, mais qu'elle a la teste de cerf, les pieds d'elephant, la queue de sanglier, le buglemet fort, et vne corne noire au milieu du front longue de deux coudées. Elle vient en Ethiopie, où on la prent toute viue. Au reste quand aux marchadises qu'on y porte, il fault sçauoir qu'on leur porte d'Inde les perles & autres bagues, especeries & choses aromatiques, principalement de la ville d'Inde qu'on appelle Bangella, on y mene des drapz de cotton & de foye.

Aden, lieu marchand d'Arabie l'heureuse.

se, Aiaz aussi vn autre lieu marchand.

Ceste ville qui est bien forte, est murée de deux costez, ce qui reste est encloz de mōrai. Les traffiques y gnes, où il y a cinq chasteaux. La ville est en lieu plat, & a six mille de tour. Ilz traffiquent es foires de nuit a deux heures, a cause des grâdes chaleurs qui y sont de iour. On y va par mer en marchandise de toutes les Indes, Ethiopie & Perse. Et aussi bien ceux qui vont souuer a Meche. Incontinēt qu'ilz sont aruez au port de Liburne, les peagiers sont la, pour demander d'où c'est qu'ilz sont partiz, & quelle marchandie ilz porteront, &

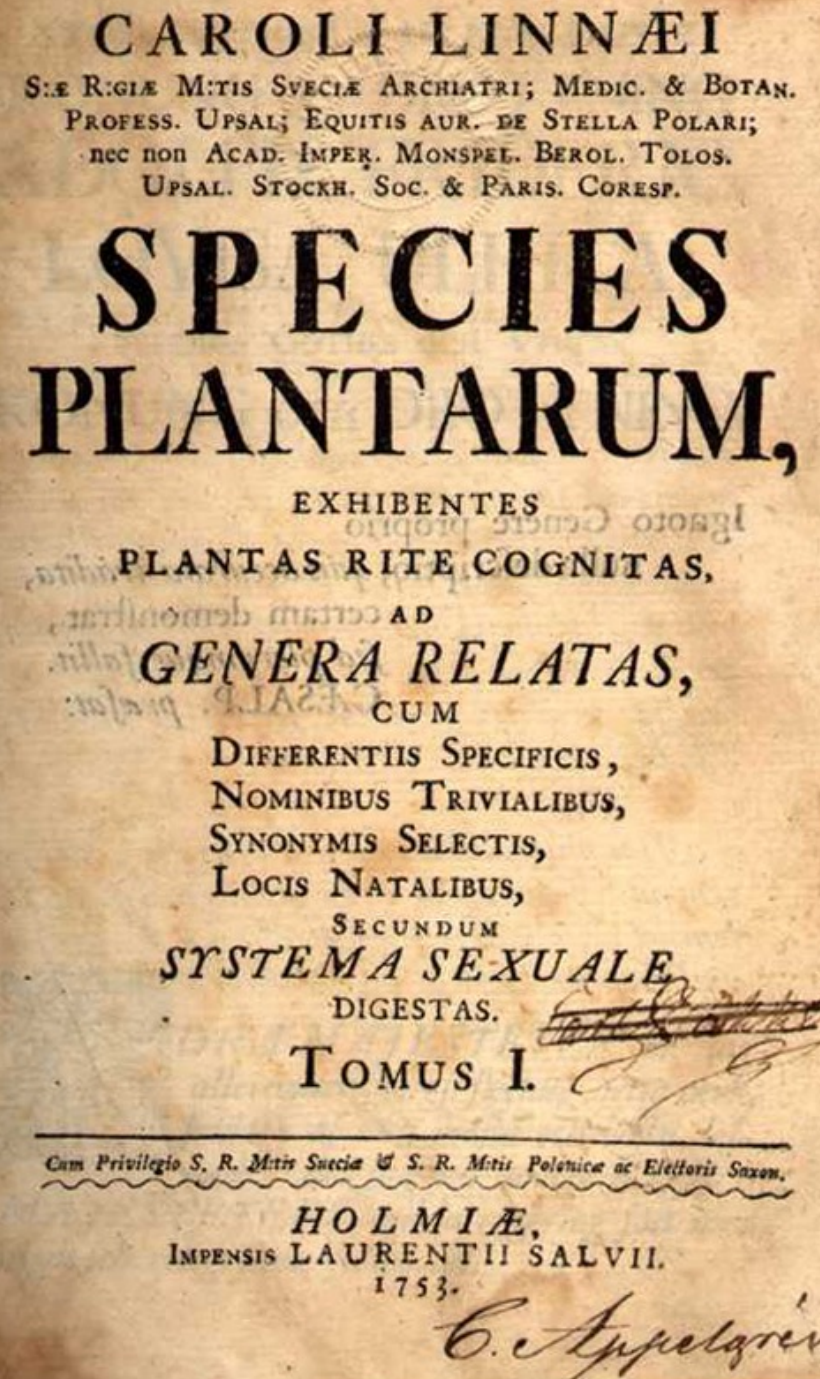
Linné, la naissance de la biologie scientifique

(1753)

8.000 végétaux différents

Méthode de classification objective, reproductible :

- *La plante est-elle pérenne ? si oui aller à la page 142, sinon aller à la page 157.*
- *La plante est-elle ligneuse ? si oui, aller à la page 210, sinon aller à la page 67,*
- etc.



Linguistique moderne

- Structuralisme, Ferdinand de Saussure, *Cours de linguistique générale* (1916)
- Distributionalisme, Zellig Harris (1909-1992), *Structural Linguistics* (1966)
- Grammaire générative, Noam Chomsky (1928-), *Syntactic Structures* (1957)
- Lexique-Grammaire, Maurice Gross, LADL (1934-2001), *Méthodes en Syntaxe* (1975)

Applications de TAL et méthodes empiriques

- **IBM Watson** gagne au jeu *Jeopardy* en 2011
- **AlphaGo** (Google) bat le champion du monde de go en 2017. **AlphaZero** (Google) bat le champion du monde d'échecs en 2017
- **Google Translate** et **DeepL** sont les traducteurs les plus utilisés aujourd'hui
- **ChatGPT** rédige des textes, construit des programmes, répond aux questions
- Les voitures **TESLA** se conduisent toutes seules

C'est magique !



Siri. Quelqu'un à qui parler.

Siri est l'assistant intelligent qui comprend ce que vous dites, voit ce que vous voulez dire et trouve les réponses à vos questions. Et comme il maîtrise désormais davantage de langues, il se rend indispensable dans plus de pays².

IBM Watson

<https://www.youtube.com/watch?v=P18EdAKuC1U>

- 200 millions de pages de texte, incluant Wikipedia (4 To)
- Gagne le jeu **Jeopardy**, 2011 (\$1M)
- Vendu à quelques hôpitaux et à des entreprises (Ressources humaines)

C'est magique !



Wired

<https://www.wired.com> > ... > Electric Vehicles

Elon Musk Says Every New Tesla Can Drive Itself

Oct 19, 2016 — Tonight, at a press conference, he announced that every new Tesla will be **fully** capable of **driving** itself. After being upgraded with a suite of ...



← Tweet



Elon Musk ✓
@elonmusk

...

In ~2 years, summon should work anywhere connected by land & not blocked by borders, eg you're in LA and the car is in NY

9:11 PM · Jan 10, 2016

4,208 Retweets 534 Quote Tweets 7,414 Likes




IBM Watson

<https://www.youtube.com/watch?v=P18EdAKuC1U>

- 200 millions de pages de texte, incluant Wikipedia (4 To)
- Gagne le jeu **Jeopardy**, 2011 (\$1M)
- Vendu à quelques hôpitaux et à des entreprises (Ressources humaines)... mais c'est un flop (« pas le niveau d'un étudiant de première année de médecine »)

... mais avec quelques bémols


MA How many countries begin with the letter "Y"?

 As of my last knowledge update in January 2022, there are three countries whose names start with the letter "Y." These countries are:

1. Yemen
2. Zambia
3. Zimbabwe

Please note that geopolitical changes can occur, and the number of countries in the world can change over time, so it's a good practice to verify this information with up-to-date sources for the most current and accurate count.

KA If 1 woman can make 1 baby in 9 months, how many months does it take 9 women to make 1 baby? Explain each step you used to arrive at your answer.

 First, we need to determine how many babies each woman makes per month. We can do this by dividing the number of babies each woman makes (1) by the number of months it takes for each woman to make a baby (9). This gives us a rate of $1/9$ babies per month for each woman.

Next, we need to determine how many babies 9 women make per month. We can do this by multiplying the number of women (9) by the rate at which each woman makes babies ($1/9$ babies per month). This gives us a total rate of $9 * (1/9) = 1$ baby per month.

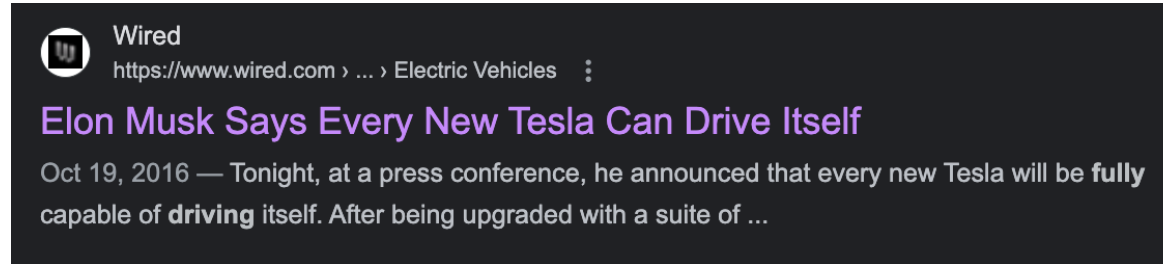
Finally, we need to determine how many months it takes for 9 women to make 1 baby. We can do this by dividing the number of babies we want (1) by the rate at which the women are making babies (1 baby per month). This gives us a total of $1 / 1 = 1$ month.

Therefore, it takes 9 women 1 month to make 1 baby.



Free-styling! Alexa decided she'd rather turn on a lullaby than the lights... surely making for a rather dark musical experience

... mais avec quelques bémols



17 morts, 736 accidents : le bilan alarmant d'Autopilot et Full Self-Driving de Tesla indique-t-il une course à l'innovation au mépris de la sécurité ?

Les statistiques de la NHTSA inquiètent

Le 12 juin 2023 à 10:43, par [Stéphane le calme](#) | 33 commentaires

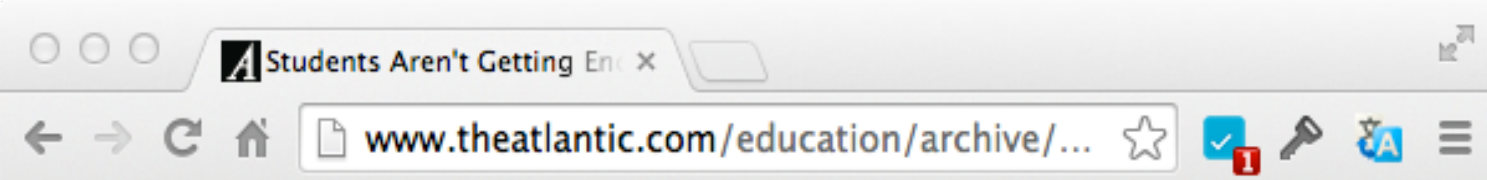




Google
translate

<https://youtube.com/shorts/ZjugCyoolde?feature=share>

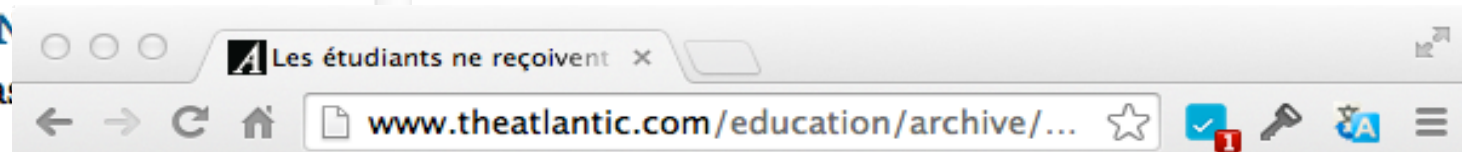
Pas terrible



As the lazy days of summer give way to the painful reality of pre-dawn alarms, many kids are beginning their descent into chronic school-year sleep deprivation. The median school start time in this country is 8 a.m. But this fall, some schools, including a handful of elementary schools in New York City, are ringing their first bell up to 40 minutes earlier than they did last year to accommodate curricular demands.

These early school start times result in sleepy kids and frustrated parents. As of Monday, those kids and parents have the formidable weight of the American Academy of Pediatrics on their side. The organization has issued a [policy statement](#) saying that "insufficient sleep in adolescents is a public health issue that significantly affects the health and safety and academic success, of our nation's middle and high school students."

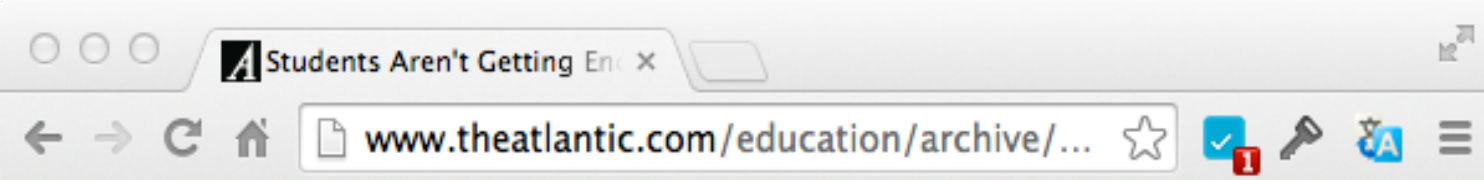
un peu mieux...



Google translate Translated to: [French](#) Show original Options

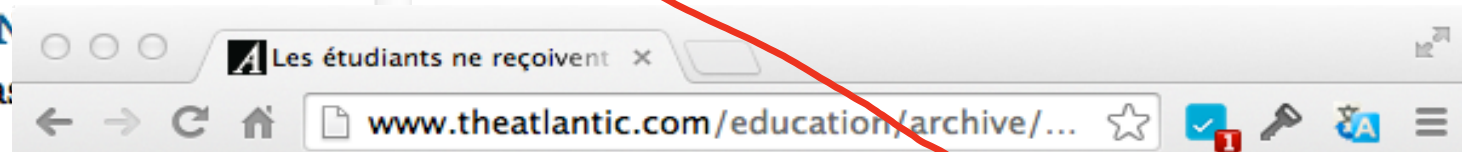
Comme les beaux jours d'été font place à la douloureuse réalité des alarmes avant l'aube, de nombreux enfants commencent leur descente dans la privation de sommeil chronique l'année scolaire. L'heure de début de l'école médiane dans ce pays est huit heures. Mais cet automne, certaines écoles, y compris une poignée d'écoles primaires de [New York City](#), sonnent leur première cloche jusqu'à 40 minutes plus tôt que l'an dernier afin de répondre aux demandes scolaires.

Ces temps de démarrage de l'école à l'avance entraîne les enfants endormis et les parents frustrés. Mais, à partir de lundi, les enfants et les parents ont le poids formidable de l'Académie américaine de pédiatrie de leur côté. L'organisation a publié [un nouvel énoncé de politique](#) en disant que «le manque de sommeil chez les adolescents [est] un problème de santé publique important qui affecte de manière significative la santé et la sécurité, ainsi que la réussite scolaire, des élèves des collèges et lycées de notre pays."



As the lazy days of summer give way to the painful reality of pre-dawn alarms, many kids are beginning their descent into chronic school-year sleep deprivation. The median school start time in this country is 8 a.m. But this fall, some schools, including a handful of elementary schools in New York City, are ringing their first bell up to 40 minutes earlier than they did last year to accommodate curricular demands.

These early school start times result in sleepy kids and frustrated parents. As of Monday, those kids and parents have the formidable weight of the American Academy of Pediatrics on their side. The organization's policy statement saying that "insufficient sleep in adolescents is a public health issue that significantly affects the health and safety and academic success, of our nation's middle and high school students."



Google translate Translated to: French Show original Options

Comme les beaux jours d'été font place à la douloureuse réalité des alarmes avant l'aube, de nombreux enfants commencent leur descente dans la privation de sommeil chronique l'année scolaire. L'heure de début de l'école médiane dans ce pays est huit heures. Mais cet automne, certaines écoles, y compris une poignée d'écoles primaires de New York City, sonnent leur première cloche jusqu'à 40 minutes plus tôt que l'an dernier afin de répondre aux demandes scolaires.

Ces temps de démarrage de l'école à l'avance entraînent les enfants endormis et les parents frustrés. Mais, à partir de lundi, les enfants et les parents ont le poids formidable de l'Académie américaine de pédiatrie de leur côté.

L'organisation a publié un nouvel annonce de politique en disant que «le manque de sommeil chez les adolescents [est] un problème de santé publique important qui affecte de manière significative la santé et la sécurité, ainsi que la réussite scolaire, des élèves des collèges et lycées de notre pays."

mais pas très bon

AI Search Is a Disaster

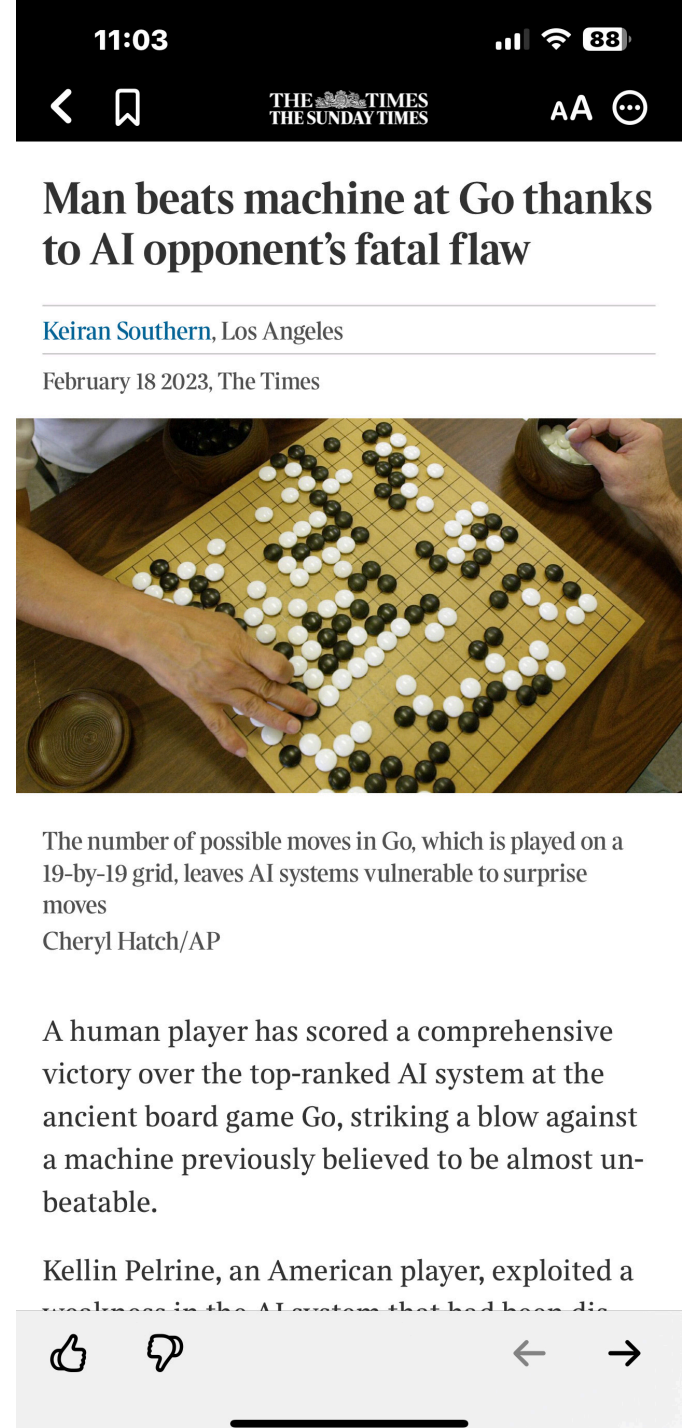
Microsoft and Google believe chatbots will change search forever. So far, there's no reason to believe the hype.

By Matteo Wong

Last week, both Microsoft and Google announced that they would incorporate AI programs similar to ChatGPT into their search engines—bids to transform how we find information online into a conversation with an omniscient chatbot. One problem: These language models are notorious mythomaniacs.

Watch Elon Musk Promise Self-Driving Cars "Next Year" Every Year Since 2014

<https://www.youtube.com/watch?v=zhr6fHmCJ6k>



Analogie : épidémiologie vs. biologie

- En 1952, Hammon & Horn étudient une population de 187.776 personnes

Epidémiologie

- En 1952, Hammon & Horn interviewent 187.776 personnes
- 18 mois plus tard, ils reviennent les interroger, et comptent les personnes qui ont décédé entre temps
- En 1954 ils publient une étude statistique qui prouve la corrélation entre fumer et mourir.

Hammon & Horn, 1954. *The relationship between human smoking habits and death rates: a follow-up study of 187,776 men. JAMA Eds.* 1954;155(15):1316-1328.
doi:10.1001/jama.1954.03690330020006

Epidémiologie vs. Biologie

- En 1952, Hammon & Horn interviewent 187.776 personnes
- 18 mois plus tard, ils reviennent les interroger, et comptent les personnes qui ont décédé entre temps
- En 1954 ils publient une étude statistique qui prouve la corrélation entre fumer et mourir.
- En 2017, Yunlong Ma & Ming Li ouvrent une piste d'explication biologique

Epidémiologie vs. Biologie

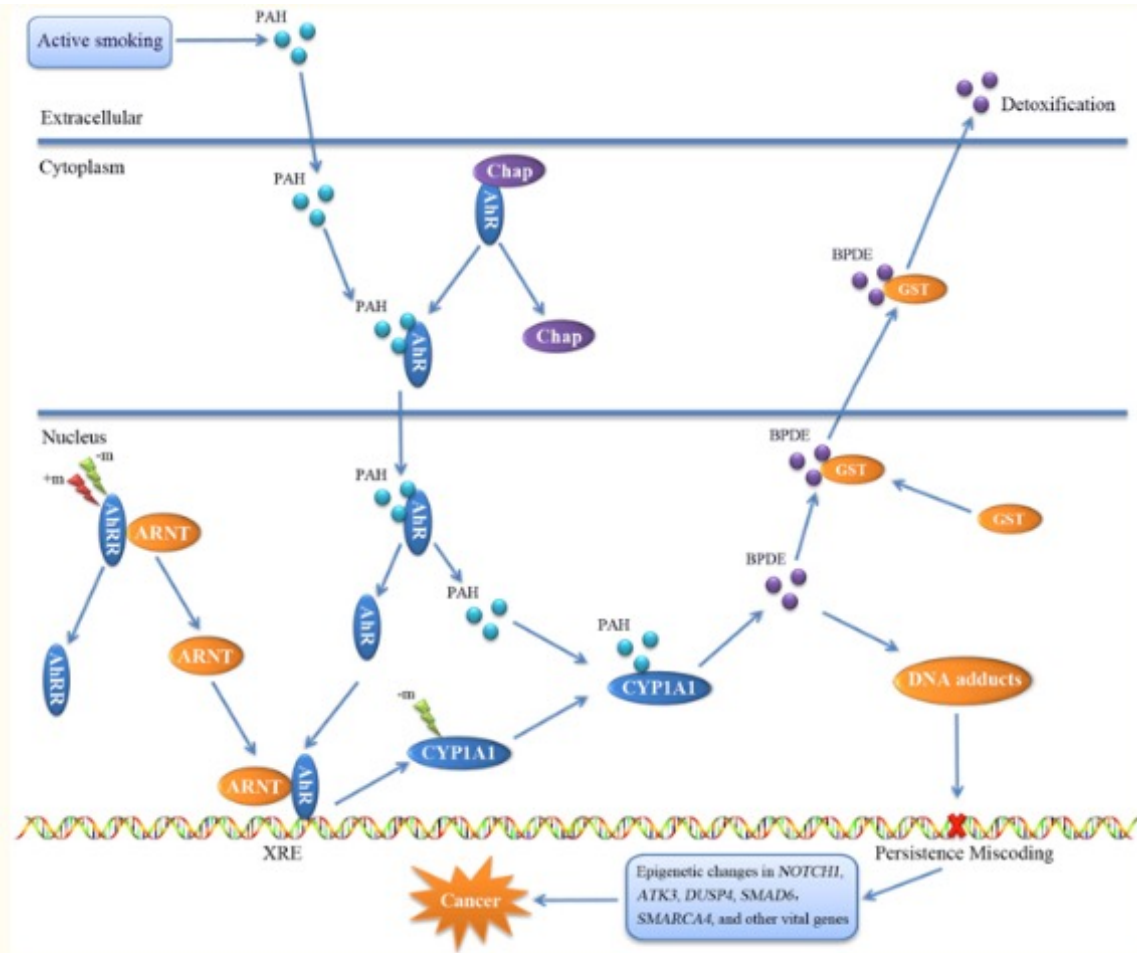


Figure 1

The pathway of “aryl hydrocarbon receptor signaling”-initiated smoking-related cancer. Arrows show event flow. -m represents hypomethylation, and +m represents hypermethylation. The plot

On commence seulement à comprendre pourquoi fumer cause le cancer : quel processus bio-chimique est impliqué ; pourquoi l'ADN de certaines cellules est abîmée, etc.

Epidémiologie vs. Biologie : méthodes empiriques vs. scientifiques

- on peut obtenir des résultats spectaculaires avec des méthodes stochastiques, mais sans comprendre pourquoi ni comment
- Pour comprendre le monde, on ne peut pas faire l'économie d'une **étude scientifique** : biologie, échecs, go, linguistique, etc.

Le projet linguistique

- La linguistique couvre de nombreuses disciplines : linguistique formelle, linguistique comparée, psycho-linguistique, socio-linguistique, étymologie, sémiotique, communication, études littéraires, etc.

Le projet linguistique

- La linguistique couvre de nombreuses disciplines : linguistique formelle, linguistique comparée, psycho-linguistique, socio-linguistique, étymologie, sémiotique, communication, études littéraires, etc.
- Notre projet : décrire précisément **l'ensemble des phrases** d'une langue, i.e. distinguer les deux séquences suivantes :

Luc mange une pomme

** Luc une mange pomme*

Combien y a-t-il de phrases en français ?

Combien y a-t-il de phrases en français ?

- Une phrase typique :

Ma cousine a remarqué que le petit chien de son voisin aime les croquettes

Combien y a-t-il de phrases en français ?

- Une phrase typique :

Ma cousine a remarqué que le petit chien de son voisin aime les croquettes

$$\begin{array}{c} 1.000 \times 100.000 \times 100 \times 10.000 \times 10 \times 1.000 \times 10.000 \times 100.000 \times 100 \times \\ 1.000 \times 100.000 \times 10.000 \times 1000 \times 100.000 \end{array}$$

$$= 1 \times 10^{49}$$

Combien y a-t-il de phrases en français ?

- Une phrase typique :

Ma cousine a remarqué que le petit chien de son voisin aime les croquettes

$$\begin{array}{c} 1.000 \times 100.000 \times 100 \times 10.000 \times 10 \times 1.000 \times 10.000 \times 100.000 \times 100 \times \\ 1.000 \times 100.000 \times 10.000 \times 1000 \times 100.000 \end{array}$$

$$= 1 \times 10^{49}$$

(Il y a environ 10^{45} atomes sur terre)

Impossible donc de recenser sur terre toutes les phrases de ce type.

Combien y a-t-il de phrases en français ?

- Raisonnement par l'absurde :
 - on part du principe qu'il y a n phrases en français, ex. $n = 1.000.000.000.000$
 - on montre qu'il existe plus que n phrases.

Raisonnement par l'absurde

multiplier le nombre de phrases

Luc dort

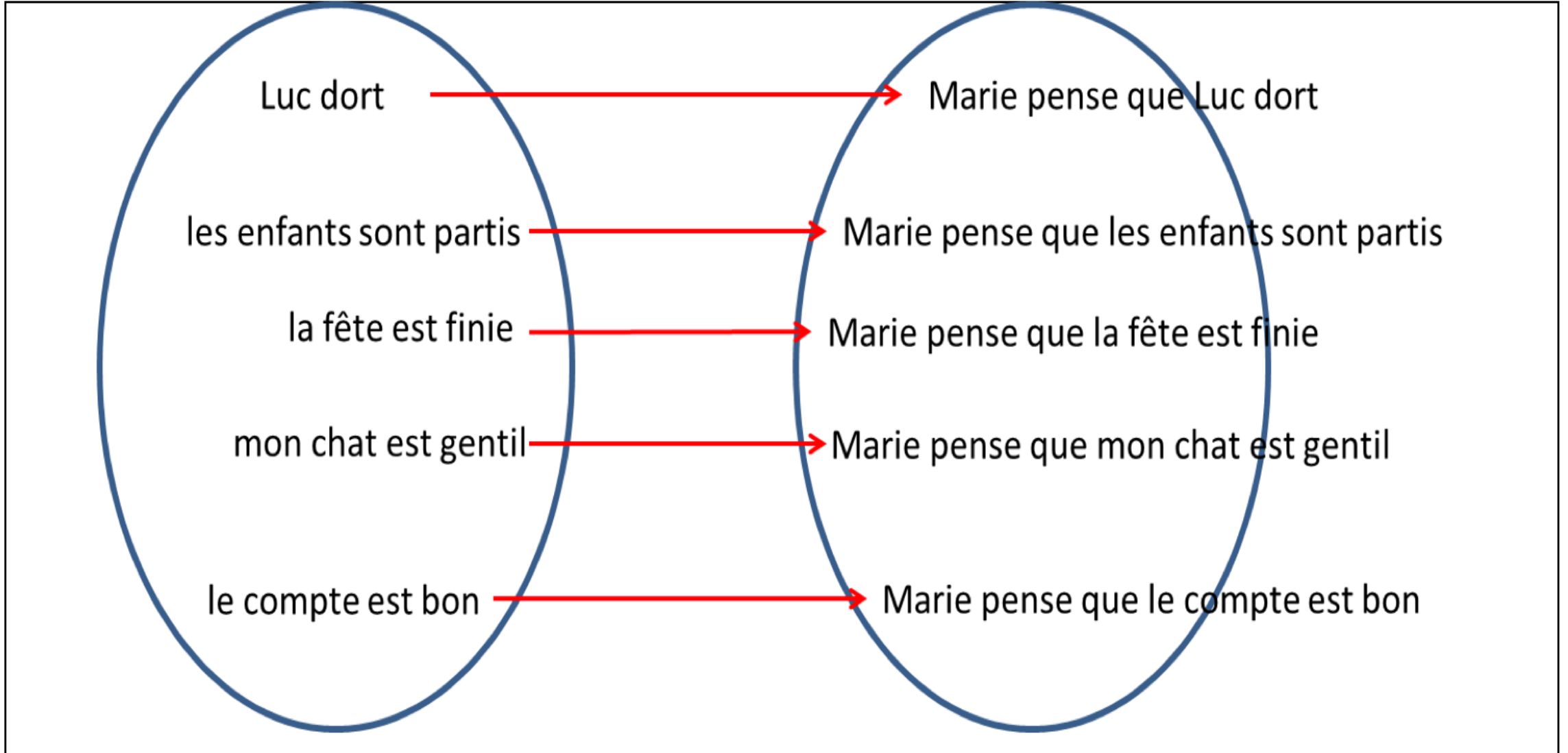
les enfants sont partis

la fête est finie

mon chat est gentil

le compte est bon

Combien y a-t-il de phrases maintenant ?



Généraliser

Généraliser *Marie*

- Paul dit que, Sarah dit que, Philippe dit que, Dominique dit que, ...
- Le garçon dit que, Ma voisine dit que, Son client dit que...
- Le garçon de café dit que, Ma voisine du dessus dit que, Son client pressé dit que...

Généraliser *pense*

- Marie annonce que, déclare que, espère que, imagine que, regrette que, répète que, rêve que, souhaite que, suppose que, décide que, veut que...
- Le voisin de la boulangère a annoncé que la fête est finie.

Ajouter *et* entre deux phrases

- *Il pleut + Luc dort => Il pleut et il dort*
- Combien de phrases obtient-on ?

Généraliser *et*

Connecteurs

- Conjonctions de coordination

mais, ou, et, donc, or, ni, car

Connecteurs

- Conjonctions de coordination

mais, ou, et, donc, or, ni, car

- Conjonctions de subordination

lorsque, où, parce que, pendant que, puisque, quand, si, ...

Connecteurs

- Conjonctions de coordination

mais, ou, et, donc, or, ni, car

- Conjonctions de subordination

lorsque, où, parce que, pendant que, puisque, quand, si, ...

- Pronoms relatifs

à qui, chez qui, dont, que, qui

Décrire un ensemble infini

- Deux règles pour définir ce qu'est un entier naturel :

A. $0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \in \mathbb{N}$

B. si deux mots x et y sont des entiers naturels, alors xy est un entier naturel

Décrire un ensemble infini

- Deux règles pour définir ce qu'est un entier naturel :

A. $0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \in \mathbb{N}$

B. si deux mots x et y sont des entiers naturels, alors xy est un entier naturel

- Est-ce que 132 est un entier naturel ?

Décrire un ensemble infini

- Deux règles pour définir ce qu'est un entier naturel :

A. $0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \in \mathbb{N}$

B. si deux mots x et y sont des entiers naturels, alors xy est un entier naturel

- Est-ce que 132 est un entier naturel ?
- Est-ce que 3,5 est un entier naturel ?

Décrire une langue

- Deux règles pour définir l'ensemble des phrases d'une langue :
 - l'ensemble des mots de la langue (le vocabulaire)

table, maison, mange, pomme $\in \mathcal{F}$

- l'ensemble des règles utilisées pour combiner des mots pour en faire des phrases (la grammaire)

Article Nom Verbe Article Nom =:

Mon voisin mange une pomme