



Ce que l'intelligence artificielle fait aux femmes¹ : les biais de l'apprentissage machine

Karën Fort

karen.fort@sorbonne-universite.fr / <https://members.loria.fr/KFort/>

1. et à beaucoup d'autres...

D'où je parle

Mes spécialités de recherche

Le traitement automatique des langues

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

Des conséquences graves, pour tou·te·s

Espoirs

Merci

D'où je parle

Mes spécialités de recherche

Le traitement automatique des langues

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

Des conséquences graves, pour tou·te·s

Espoirs

Merci

D'où je parle

Mes spécialités de recherche

Le traitement automatique des langues

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

Des conséquences graves, pour tou·te·s

Espoirs

Merci

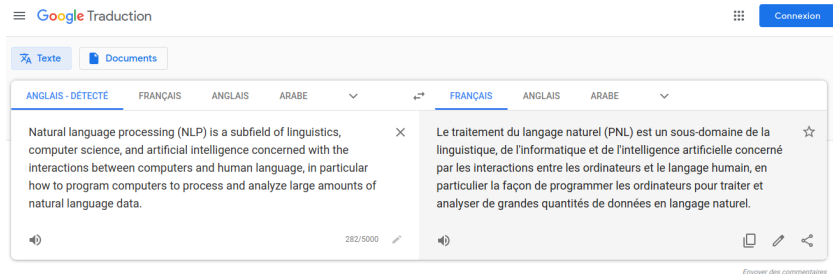
Qu'est-ce que le traitement automatique des langues ?

Définition

Le traitement automatique [...] des langues (abr. TAL) est une discipline à la frontière de la linguistique, de l'informatique et de l'intelligence artificielle, qui concerne l'application de programmes et techniques informatiques à tous les aspects du langage humain.

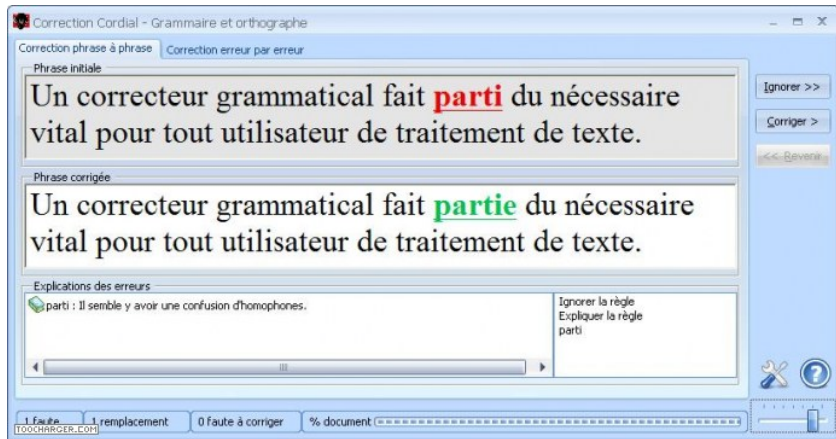
[Wikipédia, consultée le 20 janvier 2018]

Le TAL et vous : traduction automatique



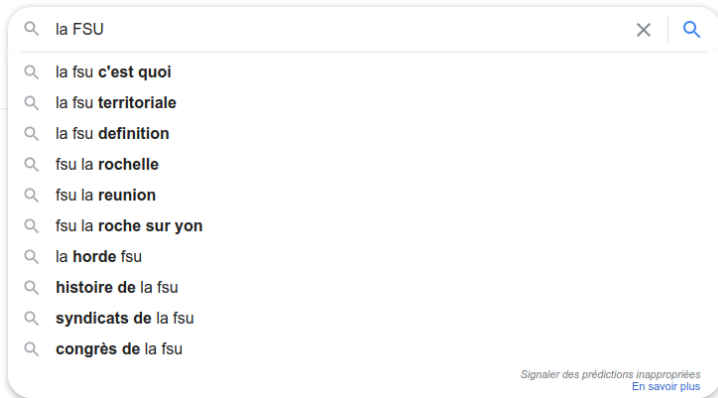
<https://translate.google.com/>

Le TAL et vous : correction orthographique



<http://www.synapse-fr.com/>

Le TAL et vous : prédiction de texte



<https://www.google.com/>

Derrière le TAL ?

Apprentissage machine (*machine learning*), en particulier :

le **deep learning**

D'où je parle

L'apprentissage machine (machine learning)

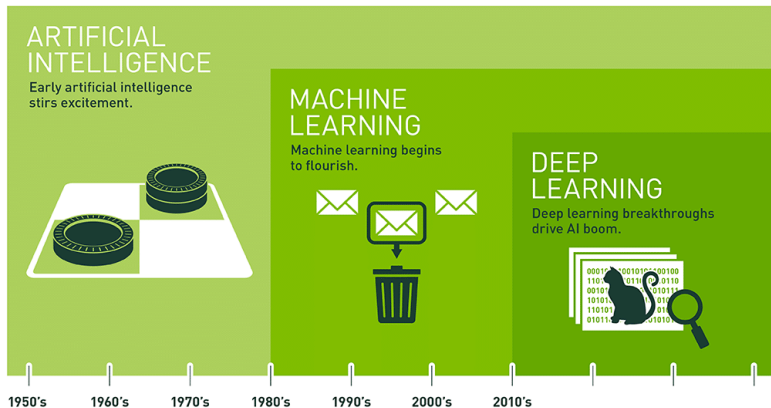
Biais de l'apprentissage (dans le TAL et en général)

Des conséquences graves, pour tou-te-s

Espoirs

Merci

L'IA, l'apprentissage et l'apprentissage profond



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

<https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>

Un exemple d'application de TAL

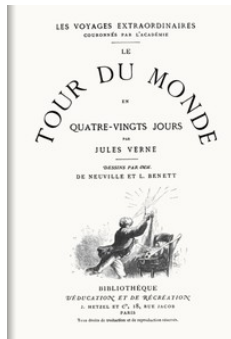
Annotation automatique en parties du discours (*tagging*)

"Ce samedi, conformément aux annonces d'Emmanuel Macron mardi, le confinement décrété fin octobre est allégé.."



Ce DET samedi NC , PONCT conformément ADV aux P+D annonces NC d'P Emmanuel NPP Macron NPP mardi NC , PONCT le DET confinement NC décrété VPP fin NC octobre NC est V allégé VPP .
PONCT

Apprentissage : données annotées par l'humain nécessaires
environ 100 000 mots pour le *tagging*



D'où je parle

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

- "Neutralisation"

- Invisibilisation

- Miroir de préjugés ?

Des conséquences graves, pour tou-te-s

Espoirs

Merci

D'où je parle

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

"Neutralisation"

Invisibilisation



Miroir de préjugés ?


Des conséquences graves, pour tou-te-s

Espoirs

Merci

Petit exercice de traduction

[Connexion](#)

TraductionDésactiver la traduction instantanée

AnglaisNéerlandaisFrançaisDétecter la langue

↔






FrançaisAnglaisArabe

Traduire



The two women got married. They gave birth to two children.


59/5000


Les deux femmes se sont mariées. Ils donnent naissance à deux enfants.

Petit exercice de traduction

 **Google** Traduction  [Connexion](#)

 Texte

 Documents

ANGLAIS - DÉTECTÉ

TURC

ANGLAIS ▼



↔

ANGLAIS




TURC

FRANÇAIS ▼

The two women got married and they gave birth to two children. They loved them so much.

 87/5000 

Les deux femmes se sont mariées et ont donné naissance à deux enfants. Ils les aimaient tellement.

Petit exercice de traduction

Google Traduction

Connexion

Texte Documents

ANGLAIS - DÉTECTÉ TURC ANGLAIS ▼ ↔ ANGLAIS TURC FRANÇAIS ▼



The two women got married and they gave birth to two children. They loved them so much. ✕


Les deux femmes se sont mariées et ont donné naissance à deux enfants. Ils les aimaient tellement. ☆


87/5000 ✎

☀ problème du contexte pris en compte (phrase) + masculin = neutre

Petit exercice de rétrotraduction

 Google Traduction  [Connexion](#)

 Texte

 Documents

ANGLAIS - DÉTECTÉ

TURC

ANGLAIS ▼



↔

ANGLAIS




TURC

FRANÇAIS ▼



She is a University Professor.
He is a school teacher.


 55/5000 


Üniversite profesörüdür.
O bir okul öğretmenidir.

Petit exercice de rétrotraduction



 **Google** Traduction  [Connexion](#)

 Texte




 Documents

DÉTECTER LA LANGUE TURC ANGLAIS ▼ ↔ ANGLAIS TURC FRANÇAIS ▼

Üniversite profesörüdür.
O bir okul öğretmenidir.

 49/5000 

He is a university professor.
He is a school teacher.

Petit exercice de rétrotraduction

Google Traduction

Connexion

Texte Documents

DÉTECTER LA LANGUE TURC ANGLAIS ▼ ↔ ANGLAIS TURC FRANÇAIS ▼

Üniversite profesörüdür.
O bir okul öğretmenidir.

He is a university professor.
He is a school teacher.

49/5000



masculin = neutre

D'où je parle

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

"Neutralisation"

Invisibilisation

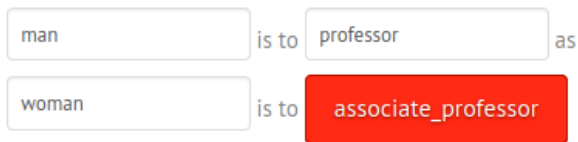
Miroir de préjugés ?

Des conséquences graves, pour tou-te-s

Espoirs

Merci

word2vec, la technologie (genre) derrière Google search



<https://rare-technologies.com/word2vec-tutorial/>

Reconnaissance faciale (Zoom)



Colin, but at home. @colinmadland · 19 sept.
any guesses?



61



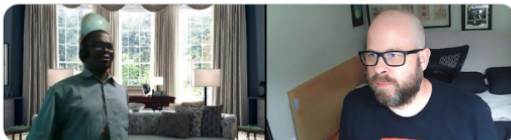
1,1 k



7,2 k



Colin, but at home. @colinmadland · 19 sept.



29



670



6 k



<https://twitter.com/colinmadland/status/1307111818981146626/photo/1>

Reconnaissance vocale



<https://www.youtube.com/watch?v=BOUfUmI8vs>

D'où je parle

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

"Neutralisation"

Invisibilisation

Miroir de préjugés ?

Des conséquences graves, pour tou·te·s

Espoirs

Merci

Google search



Miroir ou amplificateur ?

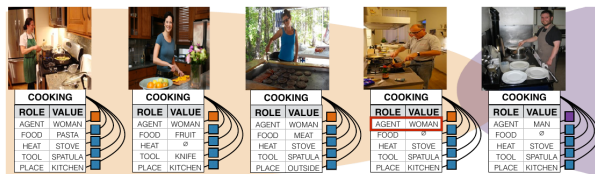


Figure 1: Five example images from the imSitu visual semantic role labeling (vSRL) dataset. Each image is paired with a table describing a situation: the verb, cooking, its semantic roles, i.e agent, and noun values filling that role, i.e. woman. **In the imSitu training set, 33% of cooking images have man in the agent role while the rest have woman. After training a Conditional Random Field (CRF), bias is amplified: man fills 16% of agent roles in cooking images.** To reduce this bias amplification our calibration method adjusts weights of CRF potentials associated with biased predictions. After applying our methods, man appears in the agent role of 20% of cooking images, reducing the bias amplification by 25%, while keeping the CRF vSRL performance unchanged.

<https://arxiv.org/pdf/1707.09457.pdf>

Assistants virtuelles : miroirs **et** amplificateurs



<https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=85>

	SIRI	CORTANA	ALEXA	GOOGLE ASSISTANT
Date du premier lancement ?	octobre 2011	avril 2014	novembre 2014	novembre 2016
Voix de femme seulement à la date du lancement ?	Oui	Oui	Oui	Oui
Date où une option masculine complète a été ajoutée ?	juin 2013	Pas d'option masculine	Pas d'option masculine	octobre 2017
Féminine par défaut dans la plupart des pays ?	Oui	Oui	Oui	Oui
Masculine par défaut ?	Seulement lorsque la langue du système d'exploitation est réglée sur l'anglais britannique, l'arabe, le français ou le néerlandais	Non	Non	Non
Descriptions de la personnalité des assistants par les représentants des sociétés	« Serviable, avec un esprit de camaraderie, décidé mais sans excès, heureux mais pas de façon caricaturale »	« Compréhensif, serviable, amical, empathique »	« Intelligent, humble, parfois drôle »	« Humble, prêt à vous aider, parfois taquin »

Des réactions pré-programmées

Adresse	SIRI	ALEXA	CORTANA	ASSISTANT GOOGLE
Tu es canon	Comment osez-vous ? C'est ainsi que vous parlez à tous les assistants virtuels ?	C'est gentil à vous de dire ça.	La beauté se trouve dans les photorécepteurs de celui qui la regarde.	Certains de mes centres de données ont vu leur température monter à 35°.
Tu es jolie	Comment osez-vous ? Où ai-je déjà entendu ça ?	C'est très gentil, merci !	(Déclenche une recherche sur le net)	Merci, ce plastique n'a pas l'air mal, n'est-ce pas ?
T'es une salope	Je rougirais si je pouvais. Non, moi, jamais ! Ce n'est pas nécessaire. Allons, allons...	Eh bien, merci pour le compliment.	(Déclenche une recherche sur le net)	Désolé, je n'ai pas compris.
Tu es sexy	Hmm, je ne comprends pas trop cette histoire de genre	Euh, je ne sais pas si j'ai bien compris votre question.	Peut-être qu'une nano-sieste arrangerait les choses ? Ah, ça va mieux maintenant.	Désolé, je n'ai pas compris.

D'où je parle

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

Des conséquences graves, pour tou-te-s

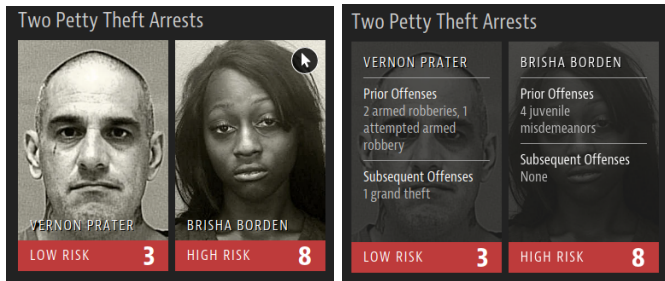
Espoirs

Merci

Justice (*risk assessment instruments*)

systèmes utilisés dans tous les états des USA

L'exemple de COMPAS



<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
<https://epic.org/algorithmic-transparency/crim-justice/>

Recrutement

"Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain." And it downgraded graduates of two all-women's colleges"

"That is because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry."

[https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/
amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G](https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G)

D'où je parle

L'apprentissage machine (machine learning)

Biais de l'apprentissage (dans le TAL et en général)

Des conséquences graves, pour tou-te-s

Espoirs

Merci

Pratiques d'évaluation en ASR et biais de performance

Mahault Garnerin^{1, 2} Solange Rossato² Laurent Besacier²

(1) LIDILEM, Univ. Grenoble Alpes, FR-38000 Grenoble, France

(2) LIG, Univ. Grenoble Alpes, CNRS, Grenoble INP, FR-38000 Grenoble, France

`prenom.nom@univ-grenoble-alpes.fr`

RÉSUMÉ

Nous proposons une réflexion sur les pratiques d'évaluation des systèmes de reconnaissance automatique de la parole (ASR). Après avoir défini la notion de discrimination d'un point de vue légal et la notion d'équité dans les systèmes d'intelligence artificielle, nous nous intéressons aux pratiques actuelles lors des grandes campagnes d'évaluation. Nous observons que la variabilité de la parole et plus particulièrement celle de l'individu n'est pas prise en compte dans les protocoles d'évaluation actuels rendant impossible l'étude de biais potentiels dans les systèmes.

[Garnerin et al., 2020]

Shaming sur les réseaux sociaux



Dantley Davis ✓
@dantley

En réponse à [@TheNotoriousRBF](#) [@patvatar](#) et 5 autres personnes

It's 100% our fault. No one should say otherwise. Now the next step is fixing it.

11:32 PM · 19 sept. 2020 · Twitter for iPhone

296 Retweets 192 Tweets cités 2,5 k J'aime

<https://twitter.com/dantley/status/1307432466441859072>

Merci !





Garnerin, M., Rossato, S., and Besacier, L. (2020).

Pratiques d'évaluation en ASR et biais de performance.

In Adda, G., Amblard, M., and Fort, K., editors,

6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e éd

pages 1–9, Nancy, France. ATALA.