

Lundi 28 novembre 2022



# Plateformes numériques, IA et TALN

Comment pourrait-on faire comprendre  
à une IA ce qu'on lui dit  
ou ce qu'elle lit ?



Mathieu Lafourcade  
Université de Montpellier, LIRMM, équipe Texte  
[mathieu.lafourcade@lirmm.fr](mailto:mathieu.lafourcade@lirmm.fr)





# des plateformes numériques

---

motivation : construire des connaissances (depuis des formes textuelles)

médiation par la machine (IA + TALN)

## Co-construction entre humains et machines

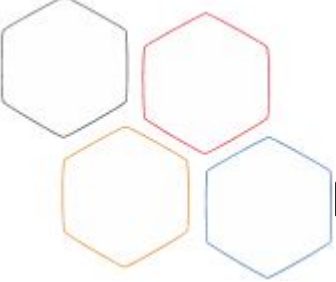
**AREN** : débats numériques (scolaires / société civile)

**OACS** : détermination de clauses abusives (droit de la consommation)

**NîmesProstate** : détermination des causes du cancer de la prostate / diagnostics

**JeuxDeMots** : Associations sémantiques / vocabulaire de sens commun et de spécialité

*enjeu crucial* dans les humanités numériques      - considérer l'IA comme une partenaire  
de la construction de connaissances



# des plateformes numériques

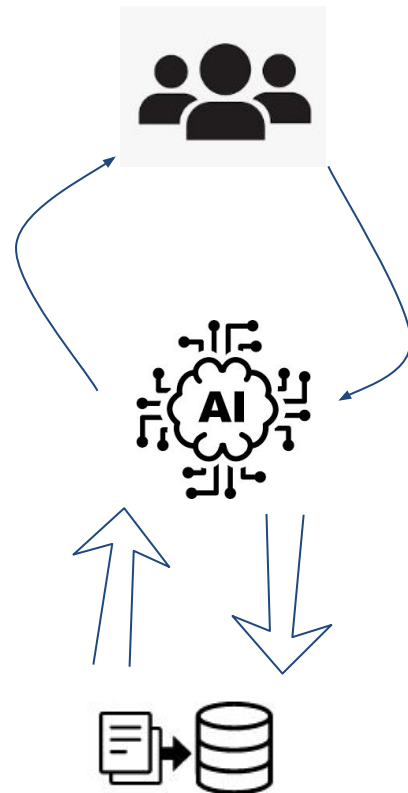
## Principes de la co-construction entre humains et machines

humains et machines apprennent en-même temps  
mais pas la même chose

la machine peut simuler de la bêtise pour *intriguer* l'humain  
parfois même pas besoin de simuler :-)

l'humain peut *challenge* la machine  
elle doit expliquer ce qu'elle comprend / prédit

une ressource se crée incrémentalement  
une trace / une mémoire / une base de connaissances



# La langage humain est plein d'ambiguïtés

Ambiguïté des mots

*je mange un avocat*

*je mange avec un avocat*

Ambiguïté des structures de phrases

*il regarde la fille avec un télescope*

*il regarde la fille avec un cheval*

Ambiguïté d'intention

*Avez-vous l'heure ?*

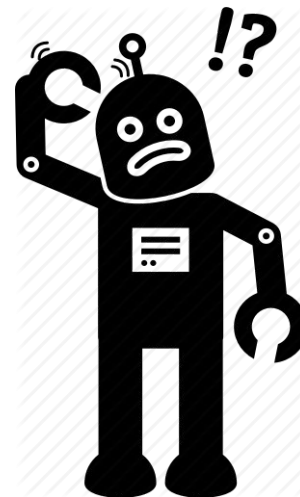
*Peux-tu me passer le sel ?*

Référence

*La chienne de la voisine est en chaleur. Elle braille sans arrêt.*

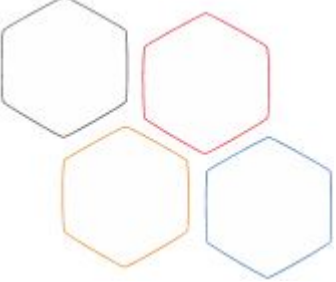
*Pourtant elle lui donne la pilule.*

*(qui est 'elle' dans chaque cas ?)*



problème





# problème

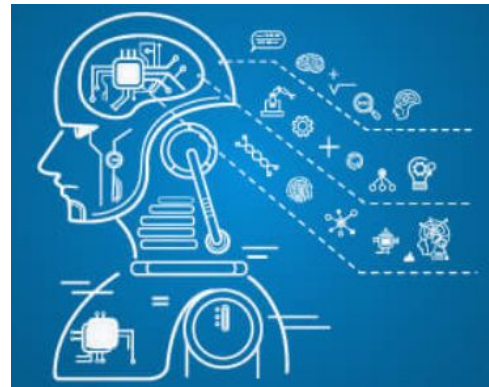
## Comment lever ces ambiguïtés ?

De quoi une IA a besoin ? va savoir...

Décrire la langue à la machine ? comment ?

La lui faire apprendre ? comment ?

C'est si difficile que ça ? oui?



Google Translate

Text

Documents

DETECT LANGUAGE

RUSSIAN

ENGLISH

FRENCH



RUSSIAN

FRENCH

ENGLISH

je mange un avocat accompagné de crevettes



I eat a lawyer with shrimp

De quoi une IA a besoin ?



de connaissances du monde

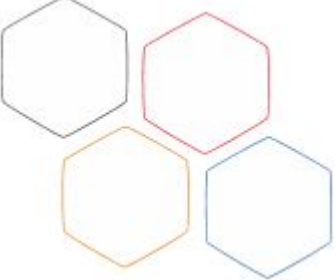


gérer le sens

passer des formes aux significations

mot 'chat'  $\Rightarrow$  animal, félin, poils, griffes, ...





# Comment acquérir de la connaissance du monde ?

---

## Manuellement ?

trop long, trop cher, trop compliqué...

## En analysant des textes ?

possible oui, avec une approche incrémentale

apprendre des choses simples, les retenir

utiliser ces connaissances pour apprendre des choses plus compliquées

Problème de la connaissance implicite nécessaire pour comprendre

*Elle insiste pour qu'il noue ses lacets avant d'aller jouer dehors*

Qui est "elle" ?

Qui est "il" ?

Pourquoi ?

problème



# Comment acquérir de la connaissance du monde ?

## Avec des contributeurs volontaires ?

Projet JeuxDeMots



Faire jouer des gens

Un **terme** va vous être présenté ainsi qu'une **consigne**. Vous devez entrer autant de **propositions** que possible conformément à la consigne.

D'**autres joueurs** vont être confrontés au même terme. Vous gagnerez des crédits et peut-être le mot, lorsque les termes que vous avez donnés **correspondent** à ceux donnés par les autres joueurs. Plus un terme est précis, plus vous gagnerez de points, mais l'autre joueur y aura-t-il pensé ?

<http://www.jeuxdemots.org>

solutions ?





# Comment acquérir de la connaissance du monde ?

DONNER DES IDEES ASSOCIEES AU TERME QUI SUIT :

kaput  
Niveau : 209  
Crédits : 199400  
Honneur : 194682  
2782  
1000



Niveau : 51

lutin



Temps  
49 s

30s

OK



Dernier terme proposé : petit

Raffinements possibles :

1. petit (taille)
2. petit (jeune)
3. petit (minuscule)
4. petit (modique)

Ce terme a plusieurs sens ou il en manque ? Demandez de l'aide à vos amis

petit >>

agaçant

taquin

pétulant

espiègle

enfant >

gobelin

gnome

farfadet

tourmenter

esprit follet

Bretagne

être fantastique

lutiner

petit >

15/28





# Comment acquérir de la connaissance du monde ?

## lutin

Niveau : 51



Réponses données par kaput : petit • agaçant • taquin • pétulant • espiègle • enfant • goblin • gnome • farfadet • tourmenter • esprit follet • Bretagne • être fantastique • lutiner • petit

Réponses données par egaillat : grelot • oreille • habit • Schtroumpfs • malicieux • nez • créature légendaire • taquin • oreilles • elfe • fée • forêt • nuton • Puck • Père Noël • espiègle • mythologie • légende • vert • esprit follet • bonnet • créature • conte • être imaginaire • gnome • chapeau • petit • troll • korrigan • farfadet

petit • taquin • espiègle • gnome • farfadet • esprit follet

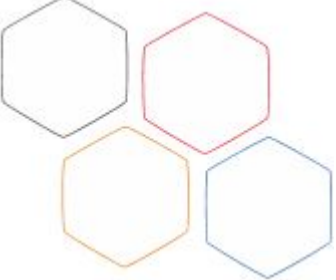


**Vous gagnez 300 crédits et 8 point(s) d'honneur**



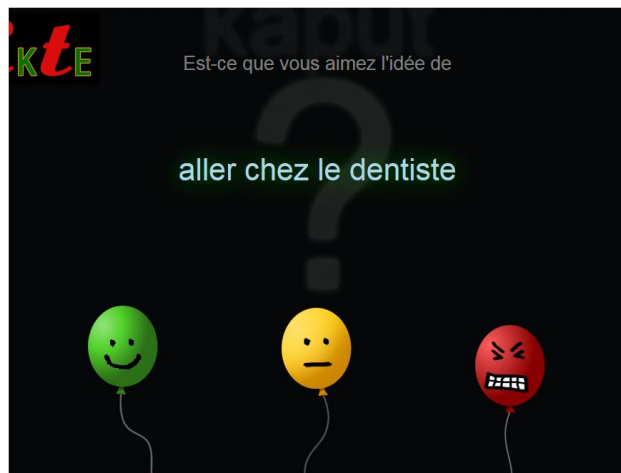
Soyez le premier de vos amis à indiquer que vous aimez ça.

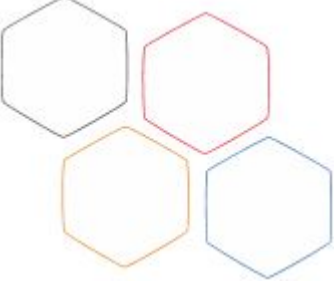




# Galaxie de jeux et contre-jeux

- Relations sémantiques spécifiques
- Questions fermées
- Polarité





# Galaxie de jeux et contre-jeux

---

## A propos de JeuxDeMots

<https://www.jeuxdemots.org/jdm-about.php>

- Totaki • TierXical • Polilt • Likelt • Emot
- Colorlt • Asklt • Selemot • Top10 • YaKaDiRou
- Semintox • Quicky Goose •
- AskYou • Dicasso • Diko • Hélix • Rezo •
- Une liste de listes •

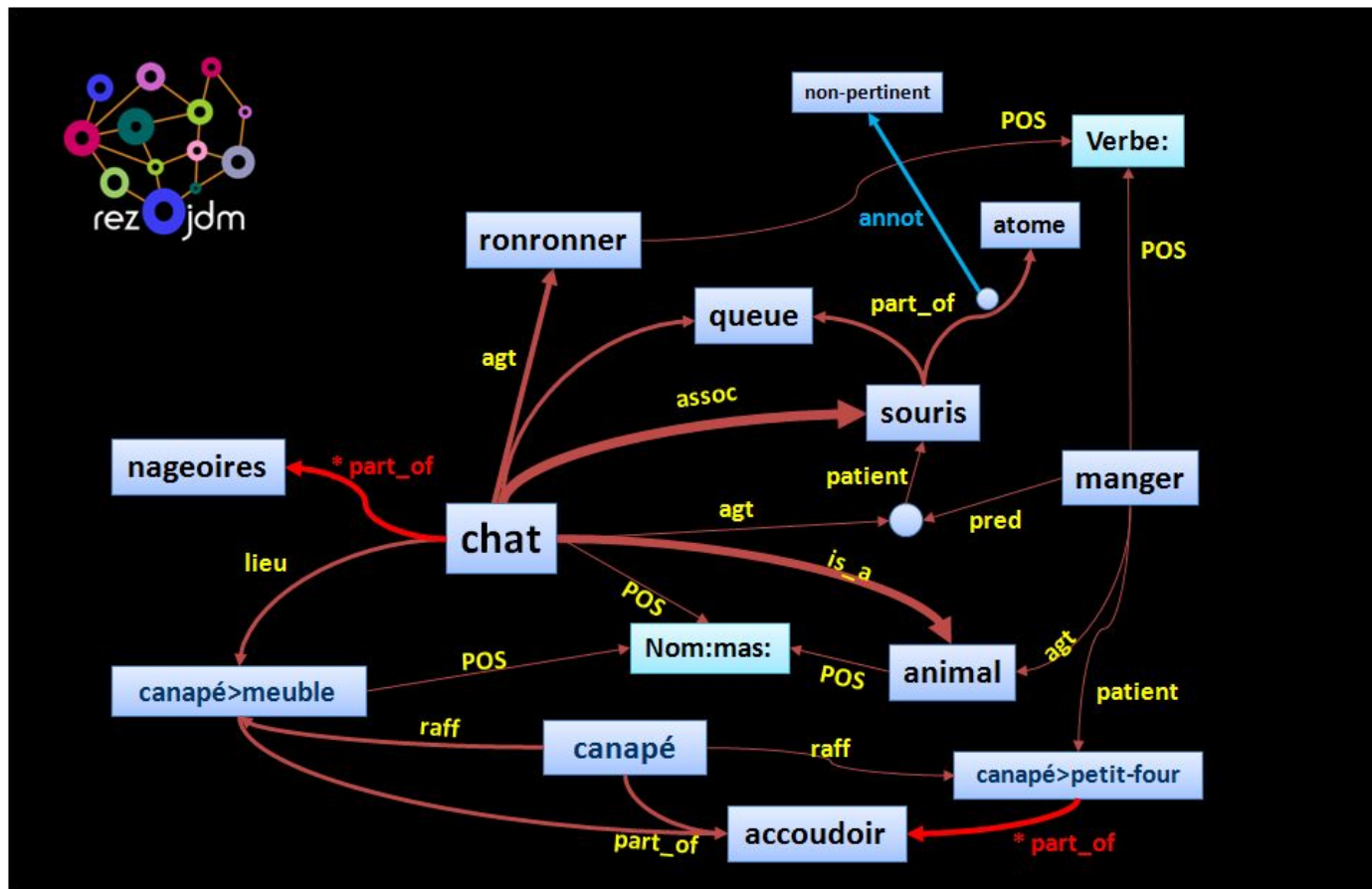


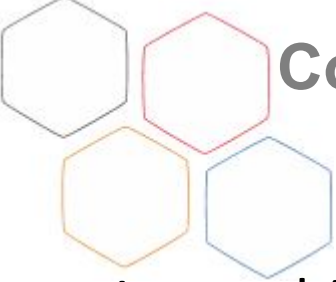
**cliquez et jouez !**



Millions de relations

> 120 types  
de relation





# Comment acquérir de la connaissance du monde ?

---

La machine apprend à partir des informations collectées...

...et pose des questions (naïves) pour consolider ses connaissances

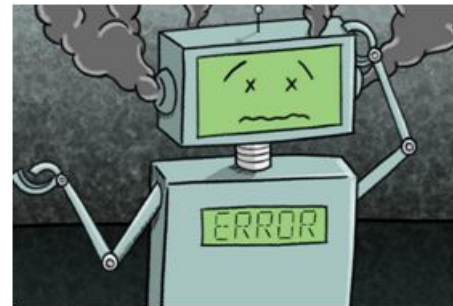
*Est-ce qu'une **frégate (oiseau)** peut avoir des **voiles** ?*

*Est-ce qu'un **générique de fin** est la même chose qu'un **médicament palliatif** ?*

Faire des inférences (déductives et abductives)

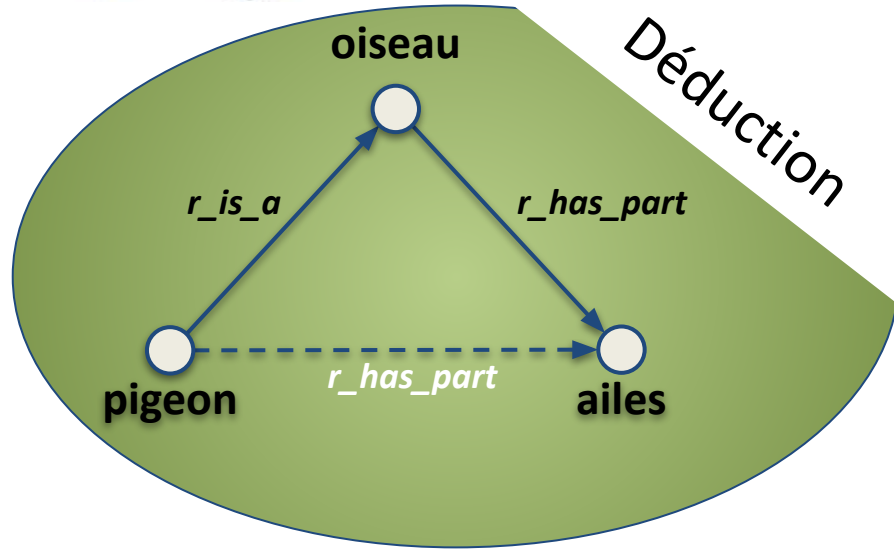
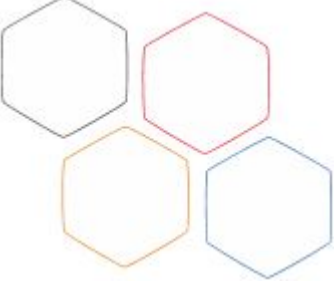
*Un pigeon est un oiseau, donc il a des ailes !*

*Une voiture a des ailes. Une voiture est-elle un oiseau ?*



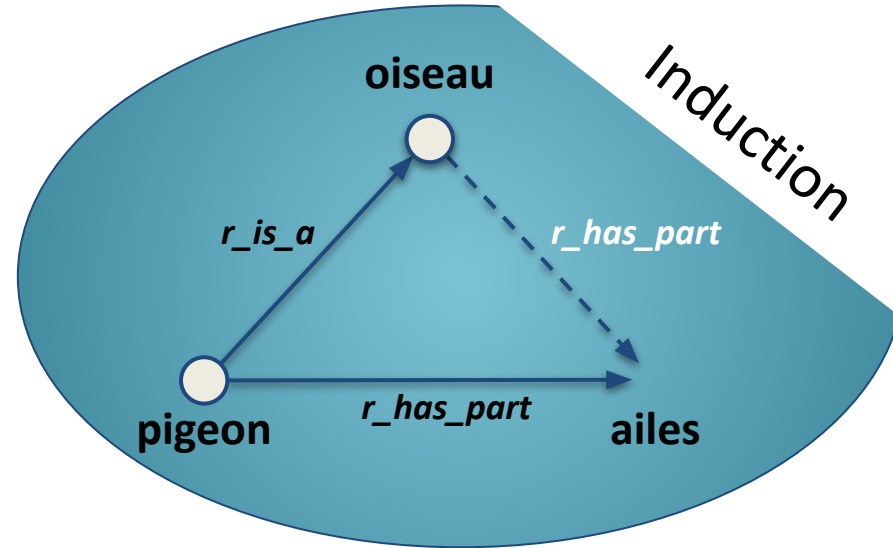


# Des inférences ?



raisonnement du générique  
vers le spécifique

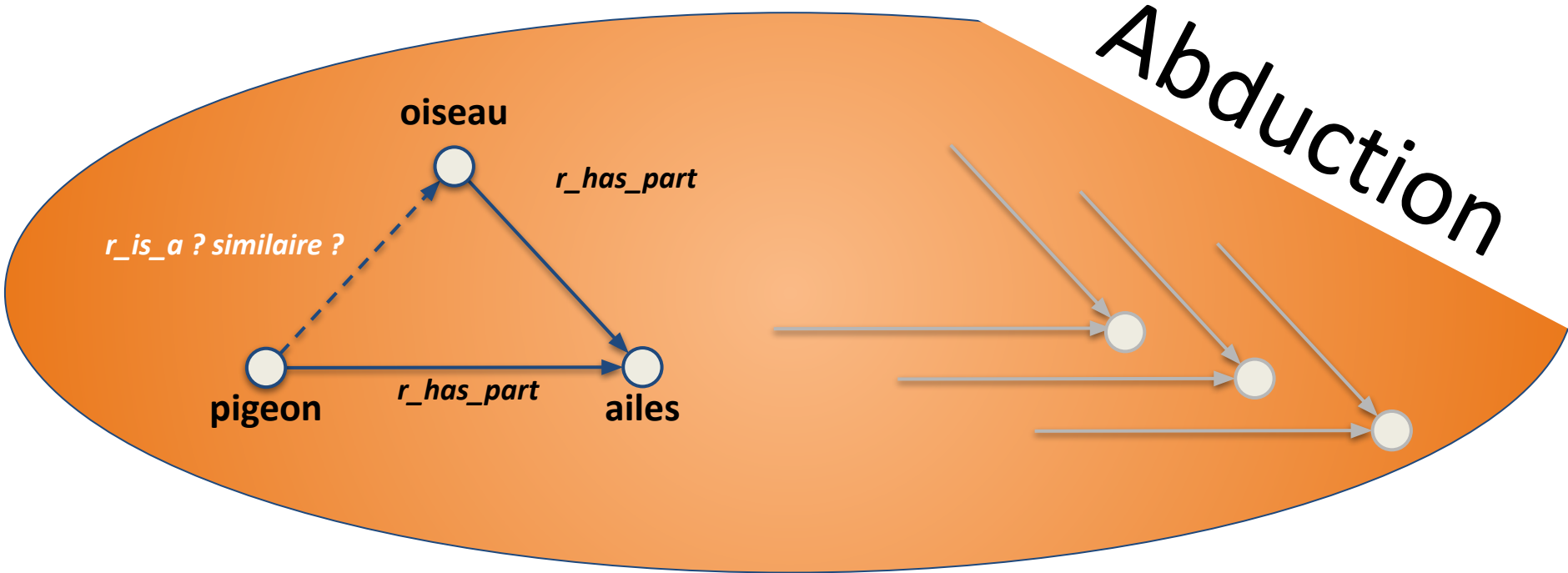
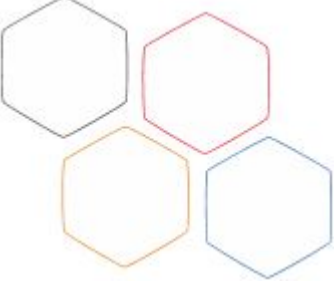
raisonnement du spécifique  
vers le générique



il faut quelques exemples !!

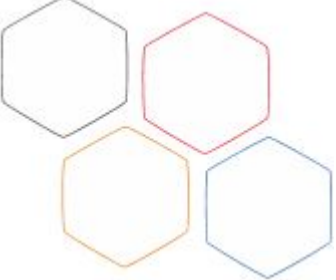
# Des inférences ?

raisonnement à partir de 'choses' qui ressemblent

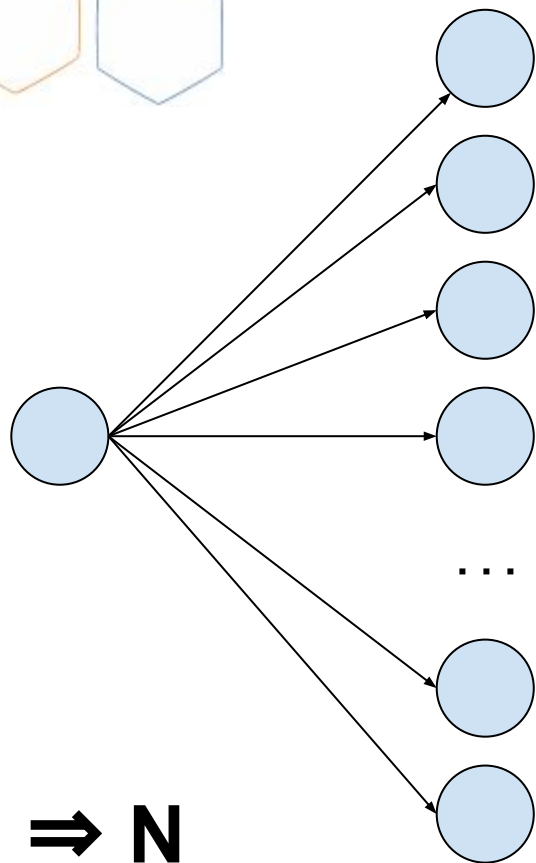


il faut un peu plus d'exemples





# Du réseau sémantique au réseau neuronal



**1 ⇒ N**

chat

⇒

souris ,  
félin ,  
chien ,  
animal ,  
miauler ,  
chaton ,  
chatte ,  
griffe  
ronronner ,  
minou ,  
queue ,  
moustache ,  
animal de compagnie ...

pour aller  
plus loin

# Du réseau sémantique au réseau neuronal

pour aller plus loin

n

Entrée



...



pe1

pen

**N  $\Rightarrow$  P**



ps1

psp

Sortie



...



vent,  
toiture,  
bourrasque

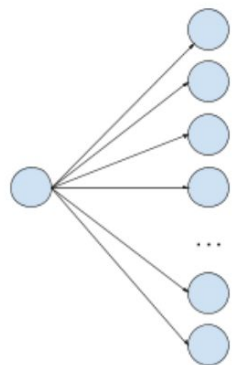
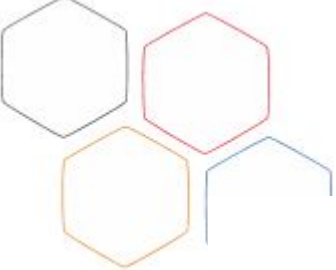
$\Rightarrow$

chute de tuiles,  
assurance,  
tuiles cassées,  
Garantie Tempête,  
...

p

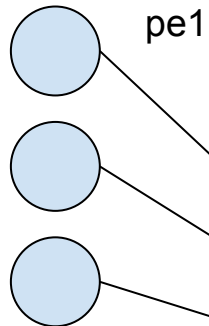
associations contextualisées

# Du réseau sémantique au réseau neuronal



+

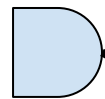
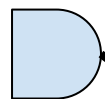
Entrée



pe1

pei

pen



...



x1

x2

psj = xm

Sortie



...



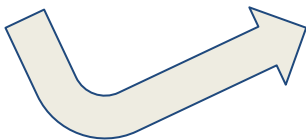
...



psp

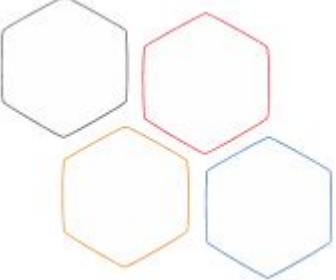
$$(pe_i \times ps_j)^{1/2}$$

$$(n \times \sum_i pe_i \times p \times \sum_j ps_j \times m \times \sum_k x_k)^{1/6}$$



WHAAAAA?!?!?





texte donné

# Extraction thématique

(extrait de l'entrée *Dioxyde de Carbone*)

*À pression atmosphérique, il se sublime à 78,5 °C (passage de l'état solide à l'état gazeux), mais ne fond pas (passage de l'état solide à l'état liquide)*

> 350 000  
neurones  
27 millions de  
synapses  
(juillet 2022)

Termes extraits (ensemble E) : *pression atmosphérique ; se sublimer ; °C ; degré ; passage de l'état solide à l'état gazeux ; état solide ; état gazeux ; fondre ; passage de l'état solide à l'état liquide ; état liquide*

Termes associés (ensemble S) : *dioxyde de carbone \* gaz carbonique \* hydride carbonique \* sublimation \* degré Celsius*

sublimation > physique . état . degré Celsius . sublimer .

anhydride carbonique . gaz carbonique . dioxyde de carbone . liquide .



sublimation . état solide .

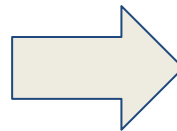
calculé par la machine 20

## Poil de carotte

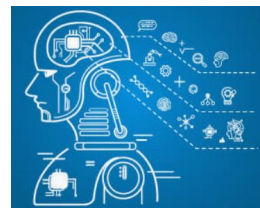
Poil de carotte est un enfant mal aimé, qui, pour lutter contre les humiliations quotidiennes et la haine maternelle, n'a que la ruse, cette arme des faibles.

## Vipère au Poing

Le livre décrit l'enfance et l'adolescence du narrateur, Jean Rezeau, dit Brasse-Bouillon. Ce dernier relate ses rapports avec sa famille, et notamment avec sa mère Paule Pluvignec, dite Folcoche, une marâtre cruelle et peu aimante.



## Enfance malheureuse



calculé par la  
machine

# OACS - c'est quoi une clause abusive ?

Composer

id auteur année nom 30-60 / 1861 date

Clauses (1861)

Poubelle (32)

[3950] auteur : EBT nom : C526-4 dernière modification : 2022-06-08 13:40:48  
domaine : \* \* \* Banque année : 2018  
statut : abusive  
fondement : - info statut : -

Le montant de la remise est porté au crédit du compte du client sous réserve d'encaissement.

Remarques : -



Idéfix

(re)calculer Idéfix

Calculé le 2022-06-08 13:40:48

Termes extraits :

\* porté au crédit d'un compte \* être porté \* être du \* sous réserve \* compte client \* encaissement \* réserve de \* montant d'une remise \* crédit \* compte \*

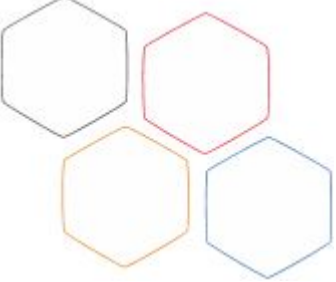
Résultat:

clause abusive • banque • ardoise>tablette • ardoise>Nom: • argent • bancaire • finance • comptabilité • rentrée • dette •



calculé par la  
machine

[3949]	EBT	2018	C526-3 : Des frais de notification sont perçus chaque fois que la banque est contrainte d'intervenir auprès du client pa...	2022-06-08 13:33:05
[3948]	EBT	2018	C526-2NA : La banque fournira au consommateur, avant la souscription du contrat, le texte de la convention, sur sup...	2022-06-08 09:13:39
[3947]	SB	2020	C668-2NA : Les informations que vous fournissez ne peuvent pas être transférées à des entités à travers le monde, ou...	2022-06-08 20:18:12
[3946]	EBT	2018	C526-2 : La banque fournira le texte de la convention sur simple demande du client, sur support papier ou support du...	2022-06-08 09:06:57
[3945]	SB	2020	C668-2 : Toutes les informations que vous fournissez peuvent être transférées à des entités à travers le monde, ou être...	2022-06-08 20:17:31
[3944]	SB	2020	C667-4NA : Nous ne pouvons pas utiliser vos données personnelles pour envoyer des notifications importantes, telles...	2022-06-08 20:13:44
[3943]	SB	2020	C667-4 : Parfois, nous pouvons utiliser vos données personnelles pour envoyer des notifications importantes, telles q...	2022-06-08 20:12:06



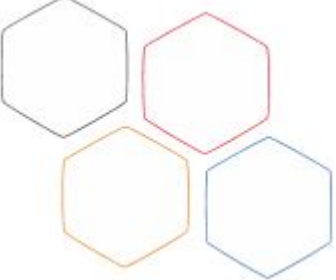
# Extraction thématique

## Applications - analyse sémantique / thématique

- Gestion de débats scolaires/citoyens (projet AREN) - GDN
- Comptes rendus en imagerie / biologie médicale
- Cuisine / nutrition
- Winograd (explications de textes, agents conversationnels)

Points clés de ce que fait la machine (l'IA)

Apprentissage permanent / Inférences / Système explicatif

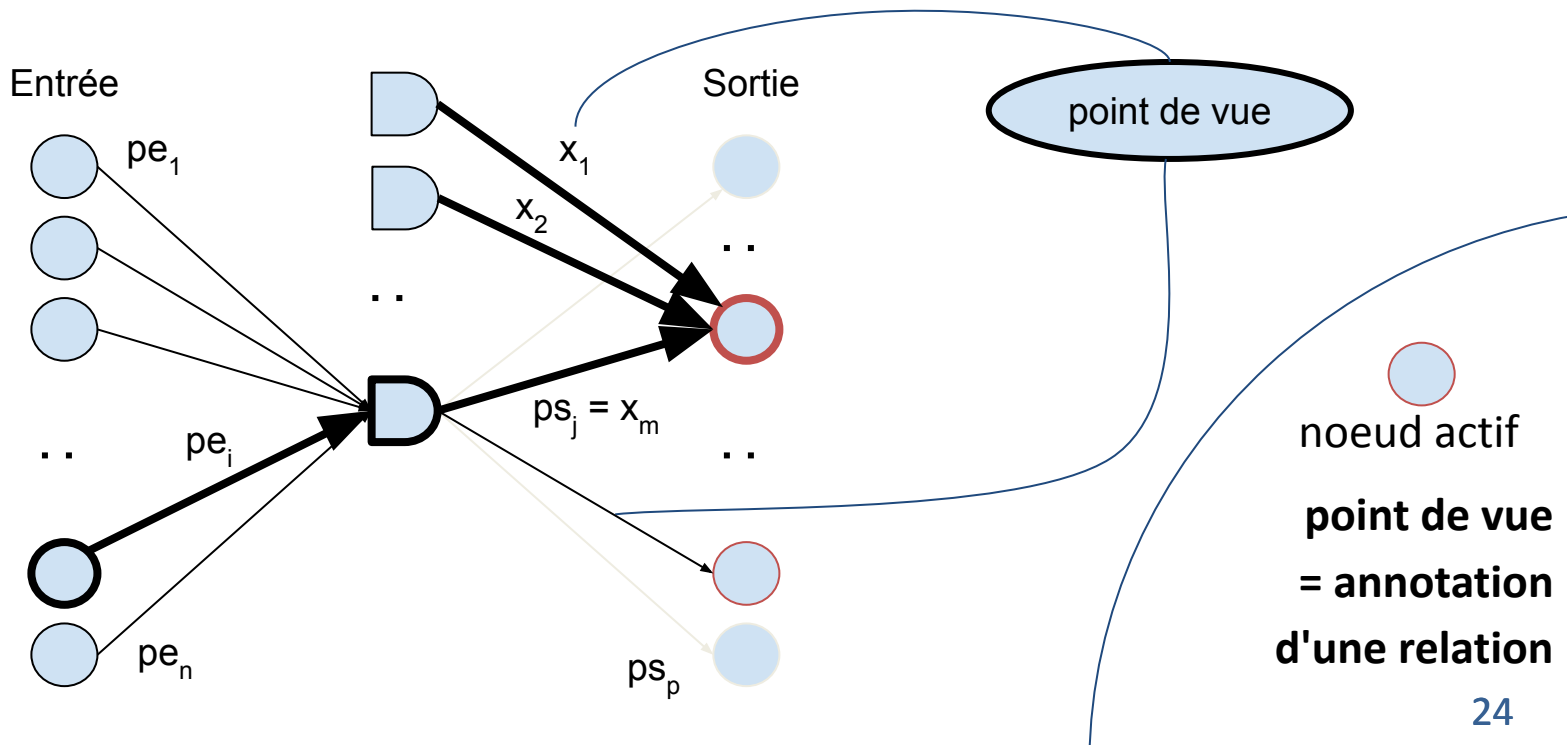


# Du réseau sémantique au réseau neuronal

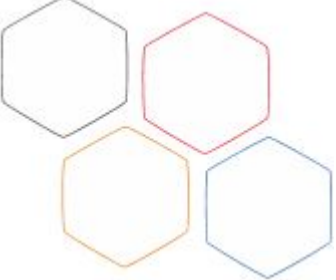
Filtrage par le point de vue (ou la tâche) :

"identification de clauses abusives", "diagnostic orthodontique"

pour aller  
encore plus loin





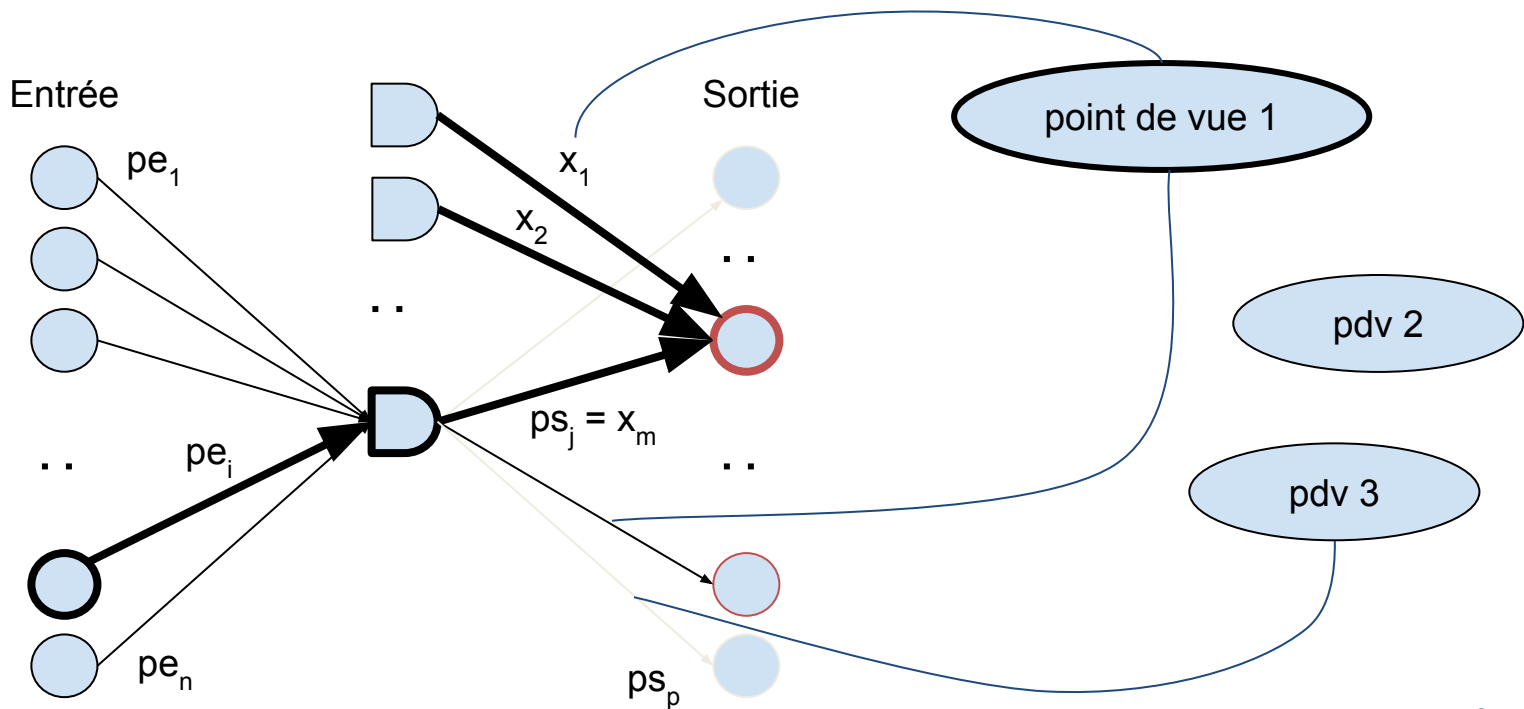


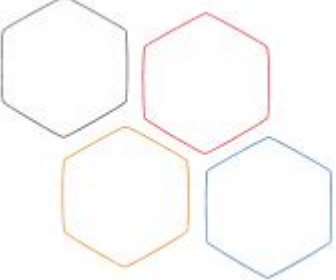
# Du réseau sémantique au réseau neuronal

Filtrage par le point de vue, par exemple

"identification de clauses abusives", "diagnostic orthodontique"

pour aller  
encore plus loin





# Du réseau sémantique au réseau neuronal

Filtrage par le point de vue

" Le voyageur ne peut assurer qu'un retour s'effectuera au même aéroport qu'au départ et ne prendra pas en charge les frais pouvant en résulter "

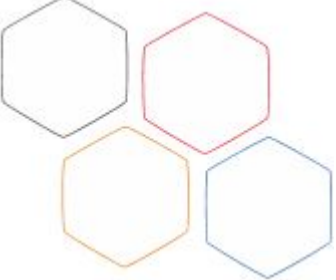
pas de filtrage



filtrage  
identification de  
clauses abusives

aéroport de retour \* voyage de  
retour \* frais \* retour voyageur \*  
voyage \* procéder \* prise en charge des frais  
pouvant en résulter \* prise en charge des frais \* frais  
de retour

clause abusive



# Quoi retenir ?

Pour des signaux faibles la machine  
peut aider l'humain à identifier :

les co-occurrences

? $\Rightarrow$  corélations

? $\Rightarrow$  implications

? $\Rightarrow$  causalités

? $\Rightarrow$  caractéristiques typiques

? $\Rightarrow$  génériques / spécifiques

? ...

mais pour cela la machine à (encore) besoin de l'être humain pour le sens commun  
et pas *forcément* besoin de mégadonnées, on peut apprendre automatiquement avec peu  
d'exemples, voire avec zéro (si si !) exemples..



# Merci pour votre lecture



RezoJDM  
est une ressource librement  
disponible



# Résumé

## **Peut-on faire une IA qui comprend ce qu'on lui dit ? ce qu'on lui écrit ?**

La compréhension automatique du langage naturel se heurte à quelques difficultés, dont les deux principales sont l'ambiguïté et les contenus implicites. L'ambiguïté est quasi omniprésente bien que les locuteurs humains ne la perçoivent pas toujours et la résolve souvent naturellement (l'avocat est véreux - il regarde la fille avec un télescope). L'implicite concerne les informations importantes que le locuteur n'a pas jugé bon d'explicitier (elle lui demande de refaire ses lacets avant d'aller courir dehors). Afin de pouvoir réaliser une machine capable d'interpréter correctement un énoncé, il semble nécessaire de disposer d'un grand nombre de connaissances, à la fois de spécialité mais surtout de sens commun. Ces dernières sont utiles afin de pouvoir raisonner automatiquement afin de découvrir les informations implicites pertinentes. Le projet JeuxDeMots vise à construire une grande base de connaissances à partir de jeux en ligne, mais également d'agents artificiels réalisant un grand nombre de tâches relevant de formes de raisonnement.