

Succès et limites de l'intelligence artificielle.

Un ordinateur peut-il
comprendre nos
propos?



Christian
Retoré
Université de
Montpellier
et LIRMM



Intelligence Artificielle: une machine simule un comportement humain.

- Raisonner
- Voir, reconnaître
- Parler
- Comprendre
- S'adapter, apprendre
- Notamment: acquérir les données nécessaires aux activités ci-dessus.
- L'intelligence artificielle de pallie pas la bêtise naturelle.
(Dans un exercice d'analyse automatique de phrases de Gérard Huet dans les années 1980)



Test de Turing (1952)

- Des personnes dialoguent via un écran et un clavier avec une machine ou un être humain
2/3 des personnes n'arrivent pas à faire la différence
- (1966) Eliza : psychologue virtuel
- Aujourd'hui: Objectif atteint ou pas? Discutable.
- A l'envers tests Captcha : la machine distingue un être humain d'une machine (sélectionner les images de feux tricolores, de ponts, de véhicules,...)



Langue et raisonnement en informatique

- Langue (depuis la 2nde guerre mondiale)
traduction automatique
dès la seconde guerre mondiale
- Raisonnement: années 1970
Systèmes experts
vérifications de circuits
preuves de programmes

Le traitement automatique
du langage naturel
Les outils





La traduction automatique

- The spirit is willing, but the flesh is weak.
The vodka is strong, but the meat is rotten.
- “Out of sight, out of mind.”
“Invisible insanity.”
- A l’inverse: une langue codée et intraduisible:
Enigma
Hopi
- Enjeux: militaires, économiques,



La traduction automatique (à l'ancienne)

1. Enoncé à traduire (nettoyé: je l'appris → je le apprendre 1^{ère} personne passé simple)
2. (analyse syntaxique, sémantique, discursive)
3. Représentation du sens (langage pivot, logique+)
4. (génération dans le langage cible)
5. traduction

Marche très bien lorsque:
domaine restreint
processus interactif



La traduction automatique (moderne)

1. Alignement de corpus bilingue.
2. Traduction de d'expressions.
3. Lissage.
 - *Aujourd'hui, la traduction automatique existe et peut être utile pour traiter de gros volumes de textes standardisés, par [...]*
 - *Machine translation lends itself to high volume traffic of the same kind of text [...]*



Traitement automatique du langage naturel Moteur de recherche

- Tri & pertinence (page ranking)
- Expressions voisines
 - production laitière / production de lait
 - Production de voitures / production automobile
 - production minière / production de mine(s) ??



Traitement automatique du langage naturel

Autres outils

- Interrogation en langage naturel
Quels sont les films des années cinquante qui passent actuellement à Bordeaux?
Les enfants prendront une pizza.
- **Correcteurs orthographiques** (Word / Synapse) :
souligné vert
QuelS livreS crois-tu qu'il sait que je pense que tu as luS ?
- **Résumé automatique**: méthode d'inversion

Le traitement automatique
du langage naturel

▶

Les niveaux d'analyse





Les sons

- Phonétique: **acoustique, traitement du signal**
- Phonologie: **automates**
Bali/Bari indistincts pour un Japonais



Les mots: déclinaisons et conjugaisons, morphologie flexionnelle

- Pluriel
 - langue -> langues
 - cheval → chevaux
 - festival → festivals
- Conjugaison
 - Aller → allons
 - aller → irons
- Automates & transducteurs



Formations des mots: morphologie dérivationnelle

- Forme: **automates**
- Sens: (**fonctions, logique d'ordre supérieur**)
- maison/maisonnette
camion/camionnette
carpe/carpette
- allumer/allumage
témoigner/témoignage
maquiller/maquillage
garer/garage



Parties du discours, catégories grammaticales

- Etiquetage grammatical
- La^{dét} voiture est en panne.
Je la^{pronom} fait réparer.
- Automates
- Modèles de Markov cachés (probabilités)



Grammaire, syntaxe

- ** Réparer fais les la le.
- * Je fais la réparer
- Je la fais réparer
- * [[Pierre [mange une]] pomme]
- Pierre [mange [une pomme]]
- **Grammaires formelles**



Sens des mots, sémantique lexicale

- Livre:
 - Rôle téléique : être lu, informer, cultiver,
 - Rôle constitutif : pages, couvertures
 - Rôle agentif : imprimeur, auteur,
- Blanche a fini un livre.
 - Lire, écrire? Relier??
 - Blanche est une chèvre?
- Apprentissage, vecteurs de mots
- Réseau Lexical construit par des jeux sérieux.
- Fonctions dans les logiques d'ordre supérieur



Sens des phrases, sémantique formelle

- Sens de la phrase (compositionnel, logique, formel)
- Formule logique (logique modale de premier ordre)
- Construit à partir du sens des ingrédients et de la structure grammaticale.
- Logique, lambda calcul



Niveaux d'analyse du langage naturel

- Ambiguïtés
- Il regarde la fille qui se promène avec des jumelles.
Il regarde la fille qui se promène avec chien.
- Les enfants prendront une pizza.
- Un gardien surveille chaque entrée du musée.
- Un étudiant a suivi chaque cours de logique.



Sens en contexte, pragmatique

- Contexte linguistique.
 - Pierre a poussé Paul. Il est tombé.
- Contexte extra linguistique.
 - Allons plutôt dîner dans ce restaurant.



Intelligence artificielle,
Analyse automatique du
langage naturel

Il y a
« comprendre »
et « comprendre ».



Geach était-il l'élève de Wittgenstein?

En 1941, IL épousa la philosophe Elizabeth Anscombe, grâce à LAQUELLE IL entra en contact avec Ludwig Wittgenstein.

Bien qu'IL N'ait JAMAIS suivi l'enseignement académique de CE DERNIER, cependant IL EN éprouva fortement l'influence.

Comprendre?

Calculer une représentation du sens utilisable par une machine

1. Trouver ce paragraphe. Google. Vecteurs de mots.
2. Trouver les relations sémantiques présentes dans ce texte. Repérage, programmation sur du texte, lorsqu'on connaît les relations sémantiques possibles.
3. Construire une représentation logique de ce qui est affirmé, nié, supposé, etc. dans ce texte. Grammaires, logiques.



Vecteurs de mots

1. On se fixe une base de sens
(les mots d'un thésaurus, par exemple)
2. Et grâce à la proximité des mots dans les textes, et à leur présence dans des textes, etc.
3. On associe un vecteur à un mot.
4. Puis en les combinant, à une phrase, à un texte etc.
 - Mesure de similarité thématique: produit scalaire.
 - Calcul très rapide.
 - Nécessite beaucoup de données (Internet)
 - Roi -- Homme → Reine



Apprentissage automatique, machine learning

- On apprend des fonctions de séquence vers des séquences:
- Nécessite
 - de grandes capacités de calcul
 - Beaucoup de données
- Aucune preuve de convergence



Apprentissage automatique, machine learning

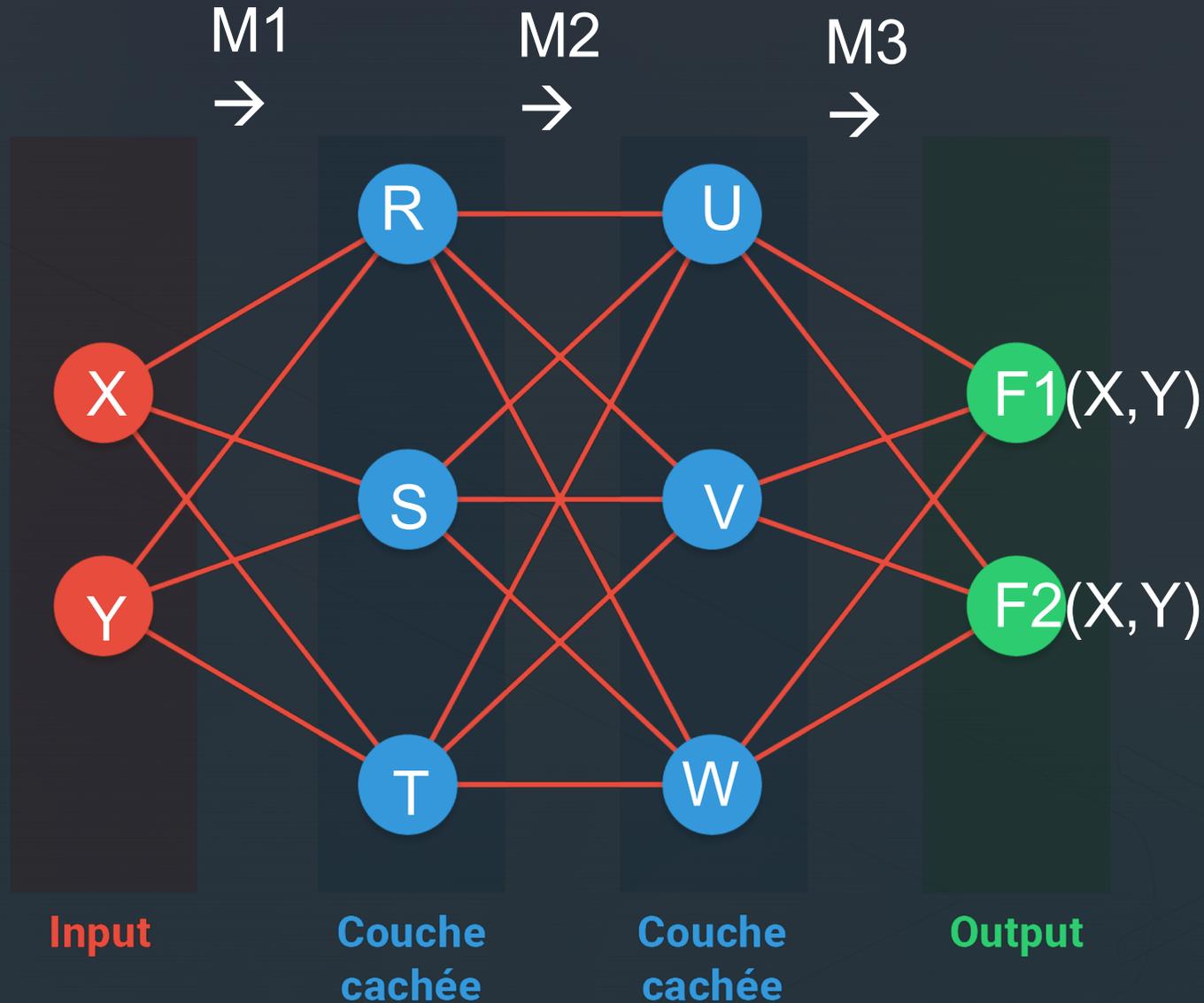
- Apprentissage de séquences vers des séquences:
- Apprentissage par « réglage » de coefficients
- Nécessite
 - de grande capacité de calcul
 - Beaucoup de données
- Aucune preuve de convergence



$(R,S,T)=M1(X,Y)$ Deep learning, réseaux
de neurones

$(U,V,W)=M2(R,S,T)$

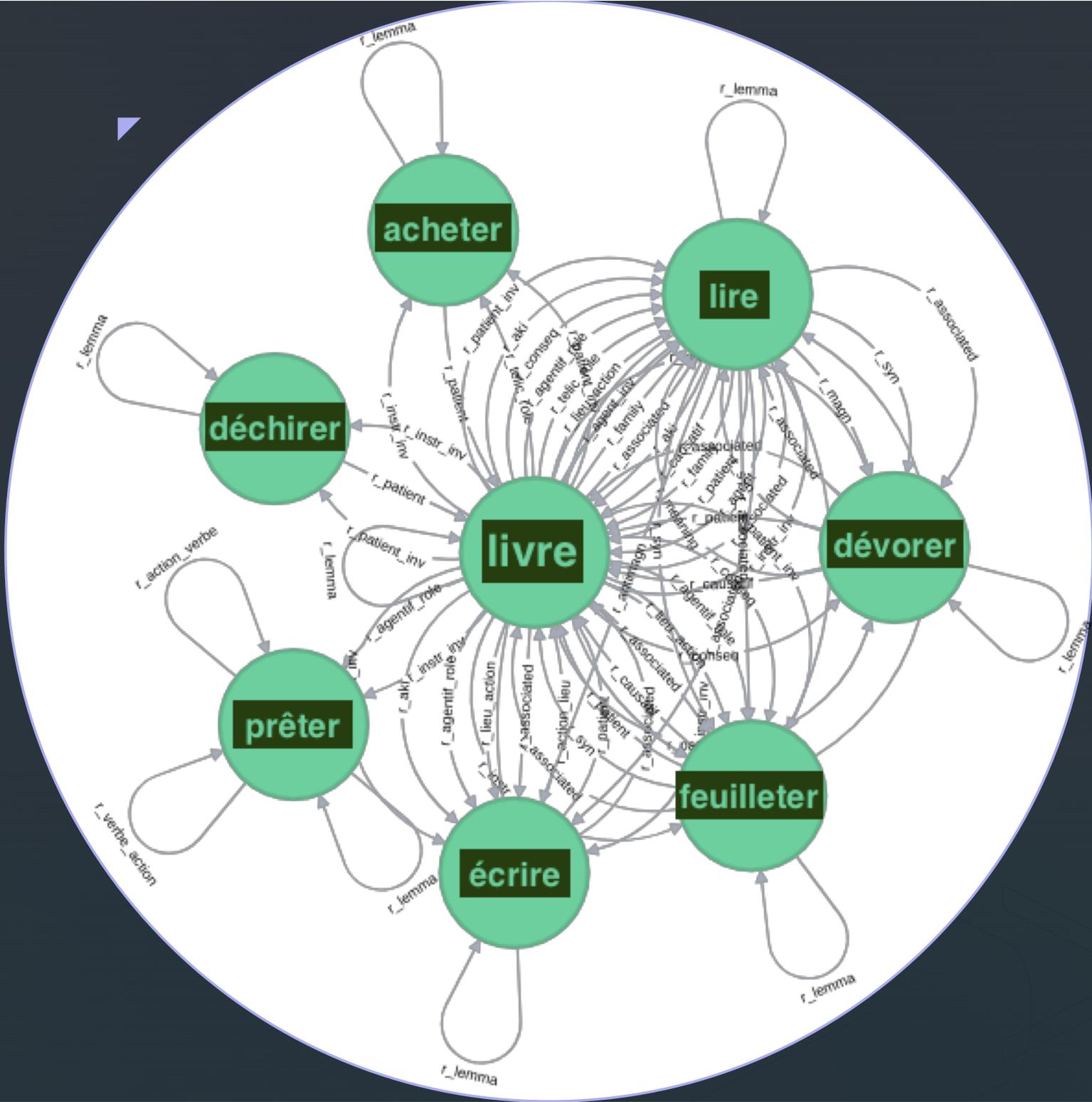
$(F1(X,Y), F2(X,Y))=M3(U,V,W)$





Repérage de relations sémantiques, réseau JeuxDeMots (Lafourcade)

- Jeux sérieux (association d'idée simple, et relations plus complexes)
- $5 \cdot 10^5$ sommets (expressions) et $80 \cdot 10^6$ relations sémantiques de 100 sortes (lieu, agent, objet, ...)
- Ces relations permettent d'identifier le sens du mot pertinent en contexte).
 - Blanche (chèvre) a fini un livre.
 - Une chèvre peut manger du papier.
 - Un livre est fait de papier.
 - Une chèvre peut manger un livre.





Approche logique, Grail (Moot)

- Grammaire acquise sur corpus
- Analyse grammaticale: logique catégorielle
 - Lexique: Mot $M_i \rightarrow$ catégories C_i
 - Supertagging
 - $C_1 \dots C_n \rightarrow C$
la suite de mots $M_1 \dots M_n$ est de catégorie C .
- Analyse grammaticale + sens des mots (calcul)
 \rightarrow formule logique (plusieurs si ambiguïté)
- Prise en compte des glissements de sens
(joker fourni par le lexique)



Approche logique, Grail

- Le caniche a aboyé.
??? La chaise a aboyé.
Le sergent a aboyé après la nouvelle recrue.
- Sortes + glissement de sens +/- compatible.
- Barcelone est un grand port.
Barcelone est un grand port cosmopolite.
Barcelone est un grand port et a remporté 5 ligues des champions.
- Glissements de sens fournis par JeuxDeMots



Approche logique, Grail

- Le caniche a aboyé.
??? La chaise a aboyé.
Le sergent a aboyé après la nouvelle recrue.
- Sortes + glissement de sens +/- compatible.
- Barcelone est un grand port.
Barcelone est un grand port cosmopolite.
Barcelone est un grand port et a remporté 5 ligues des champions.
- Glissements de sens fournis par JeuxDeMots

Etat(s) de l'art

Inférence textuelle





Cette phrase est-elle conséquence de ce texte?

- A company director awarded himself a large payrise.
→ Has a company director awarded and been awarded a large payrise.
- Philip Morris the US food and tobacco group that makes Marlboro, the world's best-selling cigarette, shrugged off strong anti-smoking sentiment in the US.
→ Philip Morris owns the Marlboro brand.



Alternative logique

- Règles grammairales, sens des mots etc.
Analyse de type Grail → formules logiques,
puis démonstrateur automatique logique
- Apprentissage : résultats très supérieurs !!!
- Mais si variation mineure du domaine (vocabulaire)
par rapport au corpus d'apprentissage
Score très mauvais



Argumentation

- Approches statistiques: classe les interventions (messages Twitter ou appréciations sur Amazon) en pour/ contre...
- Mais ignorent les énoncés comportant une négation !!! Car la portée de la négation leur échappe.
- Est-ce un ouragan?
Il y a du vent, mais ce n'est pas un ouragan.
Ce n'est pas une simple tempête mais un ouragan.



Argumentation, débats

- Les méthodes symboliques, exactes sont encore utiles notamment pour des tâches spécifiques nécessitant une analyse profonde et exacte, ainsi que la structure logique: analyse de l'argumentation
- Mais cela nécessite de connaître le domaine précis, la terminologie, les principes de ce domaine.
- Exemple: analyse de débats en ligne d'élèves sur des sujet socio-scientifiques à partir d'un texte.



**Enjeux scientifiques,
économiques et sociaux**

Un succès sans
grandes avancées
scientifiques,
financé par Gafa

Un domaine interdisciplinaire

■ Mathématiques

- Logique
- Théorie des langages formels
- Probabilités et statistiques

En rouge:
une mode
très financée
ANR, CE,

■ Informatique

- Algorithmique
- Génie logiciel

■ Linguistique ?

- Descriptions linguistiques
- Linguistique formelle
- Philosophie du langage

*En fait, pas
grand-chose
de nouveau
sous le soleil.*

La mode intelligence artificielle, machine learning, Big Data

- Pas grand-chose de radicalement nouveau:
 - Statistiques et probabilités anciennes
 - Réseaux de neurones (années 80)
 - Apprentissage
- Mais chacun constate que les outils fonctionnent beaucoup mieux!
 - Enormément de données textuelles disponibles,
 - Ordinateurs bien plus puissants, mémoire de calcul gigantesque

Quelques dangers

- Google et al. vivent des données que nous leurs donnons. En font-ils bon usage?
- Recherche financée par GAFAM... qui ne couvrent pas tous les aspects du domaine.
- Les outils fonctionnent mieux, mais les méthodes sont elles sûres, les résultats sont-ils fiables ? (cf. débat sur les armes autonomes)
- L'apprentissage renforce les solutions « majoritaires »: aucune chance que les suggestions d'Amazon ne vous fasse découvrir:
 - Un auteur méconnu
 - Ni un auteur éloigné de vos goûts habituels.
- Facteur aggravant de disparition des langues peu présentes sur Internet.



Merci de votre attention
Des questions?