

# CM 6 ET 7 - L'INTELLIGENCE ARTIFICIELLE ET LES HUMANITÉS

---

Marie Puren

27 novembre et 4 décembre 2020

UVSQ | M1 ACPCI, ECMAH, EMAS, HCS, RCL

# PLAN

Définition

Un peu d'histoire

Les principes de l'intelligence artificielle

IA pour les humanités : un exemple

Traiter un fichier nominatif de manière automatisée

La traduction automatique

# DÉFINITION

---

## SELON UN PIONNIER DE L'IA

Définition formulée par Marvin Lee Minsky :

"Construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisantes par des êtres humains car elles demandent des processus mentaux de haut niveau tels que l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique."

## UN PEU D'HISTOIRE

---

# HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE : LES DÉBUTS

- 1950 : Alan Turing, "Computing Machinery and Intelligence" dans la revue "Mind". S'intéresse à la question de l'intelligence pour les machines et se demande comment qualifier une machine de "consciente"
- 1956 : terme "Intelligence artificielle (IA)" officiellement utilisé durant un atelier organisé à Dartmouth College par Marvin Lee Minsky, Marvin McCarthy, Allen Newell, Claude Shannon... => tous les pionniers de l'IA.

# HISTOIRE DE L'IA : LES DÉBUTS

- 1959 : Hebert Simon et Allen Newell élaborent le GPR (General Problem Resolver) = programme informatique d'IA capable de résoudre tous types de problèmes géométriques et mathématiques.
- Ambition de créer le premier programme capable de reproduire la capacité humaine à résoudre des problèmes
- Echec du GPR abandonné en 1967

## HISTOIRE DE L'IA : LES DÉBUTS

1964-1966 : programme **ELIZA** développé au MIT. Programme informatique capable de dialoguer en anglais, en incarnant le rôle d'un psychothérapeute rogérien : reformulation des affirmations des patients en questions.

- Fonctionnement par reconnaissance des formes. Exemple : affirmation "A" => "Pourquoi dîtes-vous A?" Capable de poser des questions plus larges comme "Parlez-moi de votre famille" à partir d'affirmations comportant les termes "mère", "père", "fille", "fils"...
- Echec d'ELIZA qui ne permettait pas de répondre à son interlocuteur(trice) de façon satisfaisante. Seulement capable de le ou la relancer.

# HISTOIRE DE L'IA : LE CREUX DE LA VAGUE (1960-1980)

- 1966 : Rapport **ALPAC** du gouvernement américain pessimiste vis à vis de l'IA => suppression des subventions aux recherches sur la traduction automatique
- 1972 : Hubert Dreyfus dans "What Computers Can't Do"
  - Principal problème de l'IA : l'incapacité à imiter la capacité de l'homme à utiliser le contexte pour déterminer le sens des mots et des phrases.
  - Prédit que les avancées dans le domaine de l'IA stagneront tant que ce problème ne sera pas résolu.

- 1974 : avènement des premiers systèmes experts (= reproduction des mécanismes cognitifs d'un expert).
- Le plus célèbre : **Mycin** conçu dans sa thèse par Edward H. Shortliffe, pour l'aide au diagnostic et au traitement de maladies bactériennes du sang.
- 1979 : Mycin considéré par le "Journal of American Medical Association" comme aussi bon que les experts médicaux.

# HISTOIRE DE L'IA : VERS LA RECONNAISSANCE (1980-AUJOURD'HUI)

- 1983 : création du principe de l'apprentissage par renforcement  
= une pénalité diminue / une récompense augmente la probabilité d'effectuer une action dans une certaine situation.
- Nouvelles branches de l'IA comme la vision de la machine (computer vision). Travaux sur l'utilisation d'une caméra reliée à un ordinateur pour améliorer le contrôle qualité (reconnaissance des formes, par exemple).

# HISTOIRE DE L'IA : VERS LA RECONNAISANCE (1980-AUJOURD'HUI)

- Début des années 90 (guerre du Golfe) : investissement des militaires dans le domaine de la recherche en IA notamment pour l'amélioration de systèmes d'aide à la décision et la création de systèmes autonomes (drones).
- Industrie commence à utiliser l'IA plus couramment
- 1996 : le champion du monde d'échec Garry Kasparov est battu par le logiciel Deep Blue d'IBM. Démontre que l'IA est plus performante que l'homme dans certains domaines précis.
- Depuis les années 2000, recherche scientifique très active.

# LES PRINCIPES DE L'INTELLIGENCE ARTIFICIELLE

---

# LES DEUX VOIES PRISES PAR L'INTELLIGENCE ARTIFICIELLE

Deux voies différentes mais complémentaires : l'IA dite "forte" et l'IA dite "faible"

# L'INTELLIGENCE ARTIFICIELLE "FORTE"

- Fait référence à une machine qui soit capable :
  - d'éprouver une réelle conscience de soi,
  - de ressentir de vrais sentiments,
  - de comprendre ce qui la pousse à faire telle ou telle action.
- Fonctionnement identique à celui du cerveau humain : possible d'apprendre, c'est-à-dire d'emmagasiner de nouvelles informations et de modifier son comportement si nécessaire, en fonction de ces informations.
- Expérience facile pour un humain, très compliquée pour l'IA

# L'INTELLIGENCE ARTIFICIELLE "FORTE"

Importantes réserves sur la création d'IA "forte" :

- Pas possible de construire une IA forte avec les ordinateurs que nous avons aujourd'hui
- Pour certains, possible de construire une IA "forte" dans le futur, grâce à un matériel et les programmes adéquats.

# L'INTELLIGENCE ARTIFICIELLE "FAIBLE"

- Moins ambitieuse et plus réalisable
- Programme capable de raisonner, d'apprendre et même de résoudre des problèmes, mais cette fois, le programme simule seulement l'intelligence.
- Pour voler, tentative d'imiter le vol des oiseaux (construction d'ailes que l'on agite) => Echec. Aujourd'hui, on peut voler "comme les oiseaux", mais en utilisant des méthodes différentes pour effectuer une tâche similaire
- Même chose pour IA : au lieu de tenter de reproduire une conscience humaine, on peut chercher à donner des comportements intelligents à des systèmes de calcul.

## L'INTELLIGENCE ARTIFICIELLE "FAIBLE"

- Pas de pensée originale pour l'IA faible, exécution d'une tâche qui lui a été incomptée d'accomplir.
- Analyse de la situation pour tenter de résoudre le problème.  
L'IA va transcrire le problème dans son propre "langage" informatique afin d'exécuter la tâche demandée.

- Comment créer une IA?
- Nécessite des procédés spécifiques relevant de la programmation
- 2 possibilité : les systèmes-experts et le réseau de neurones formel

- Logiciel capable de **simuler le comportement d'un expert humain dans un domaine précis**
- Mettre en mémoire les connaissances théoriques de l'expert.
- Possible de pouvoir les retraduire sous forme d'aide au diagnostic, une sorte de questions/réponses avec l'utilisateur
- Exemple : un système-expert médical qui aide un médecin à poser un diagnostic au vu des observations faites par le médecin humain.

- Base de données de plusieurs milliers d'entrée (plusieurs milliers de symptômes par exemple) : garantit précision de l'analyse (un médecin pourrait omettre des détails importants, pas un système-expert)
- Mais pas d'inaffabilité de la machine qui peut elle aussi faire des erreurs (bugs informatiques, problèmes techniques)

# LE RÉSEAU DE NEURONES FORMEL

- Réseau de neurones formel : IA de type "forte"
- Reproduire la cognition humaine pour faire "penser" la machine.

# LE RÉSEAU DE NEURONES FORMEL

- Neurone formel = artificiel : représentation mathématique et informatique du neurone biologique.
- Reproduit certaines caractéristiques biologiques au moyen de fonctions et de valeurs numériques.

- Machine learning ou apprentissage machine ou apprentissage automatique : étude et entraînement d'algorithmes<sup>1</sup> pour que les ordinateurs apprennent à apprendre et fassent des prédictions sur de vastes quantités de données
- Deux sortes : **apprentissage supervisé et non supervisé**. Même objectif = extraire des connaissances en explorant des corpus de données
  - Apprentissage supervisé : entraînement préalable de l'algorithme à partir de données d'entraînement, fournies par l'humain.
    - On applique ensuite l'algorithme pour fouiller des données inconnues.
    - La robustesse de l'algorithme dépend donc de la qualité de son entraînement.
  - Apprentissage non-supervisé : pas d'intervention humaine
    - L'algorithme apprend de lui-même en cherchant des corrélations.
    - L'algorithme va alors regrouper et classer les données dans des groupes qui lui semblent homogènes.

---

1. Suites d'opérations et d'instructions qui permettent de résoudre un problème donné. Concrètement, une recette de cuisine est un algorithme.

# L'APPRENTISSAGE PROFOND OU DEEP LEARNING

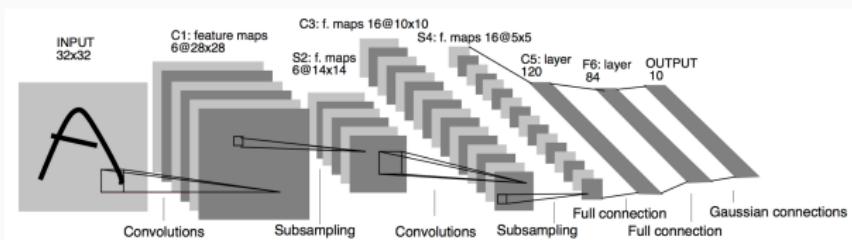
- Deep learning = réseau de neurones profonds = réseau avec plusieurs couches de neurones
- Succès du deep learning aujourd'hui dû :
  - Entraînement de plusieurs couches de neurones qui apprennent à apprendre => prendre des décisions
  - Développement facilité par de grandes quantités de données d'entraînement disponibles (provenant du Web)
  - Augmentation de capacité de calcul des ordinateurs

# L'APPRENTISSAGE PROFOND OU DEEP LEARNING

- Machine apprend par elle-même en utilisant un réseau de neurones artificiels s'inspirant du cerveau humain
- Réseau composé de dizaines voire centaines de "couches" de neurones
  - Chacune reçoit et interprète les informations de la couche précédente
  - Ex. : reconnaissance des lettres avant de reconnaître les mots; reconnaissance d'un visage avant de chercher l'identité de la personne

# LES RÉSEAUX DE NEURONES CONVOLUTIFS

- Années 1990 : Yann Le Cun propose les réseau de neurones convolutifs (CNNs ou Convolutional Neural Networks)



**FIGURE :** Architecture d'un réseau de neurones convolutifs

- CNNs permettent notamment des applications en vision par ordinateur
- 1ères applications industrielles de l'apprentissage profond
  - Lecture des chèques et codes postaux grâce aux CNNs

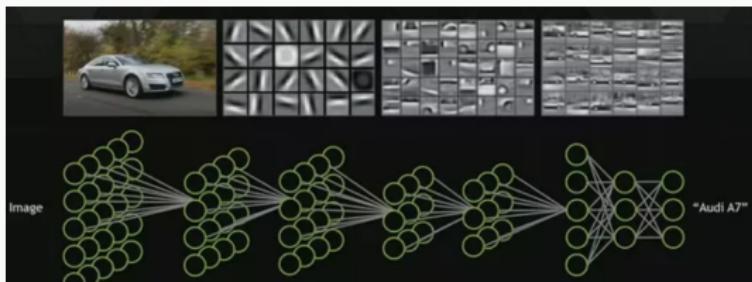
# LES RÉSEAUX DE NEURONES CONVOLUTIFS

- 2012 : AlexNet, réseau de neurones profond, révolutionne la vision par ordinateur
- Obtient des performances exceptionnelles sur ImageNet, une grande base de données d'images disponible en ligne.

=> CNNs reviennent sur le devant de la scène.

# POURQUOI PLUSIEURS COUCHES ?

- Chaque couche reçoit et interprète les informations de la couche précédente



**FIGURE :** Fonctionnement d'un réseau de neurones convolutifs

- Avec cette image, les neurones :
  - De la première couche détectent des contours
  - De la deuxième couche identifient des parties d'une voiture
  - De la dernière couche extraient les caractéristiques d'une marque
- Au fur et à mesure : ajout d'informations jusqu'à obtenir une réponse

## POURQUOI PLUSIEURS COUCHES ?

- Utilisation de neurones spécialisés dans chaque couche
- Par exemple un neurone détectera les contours verticaux

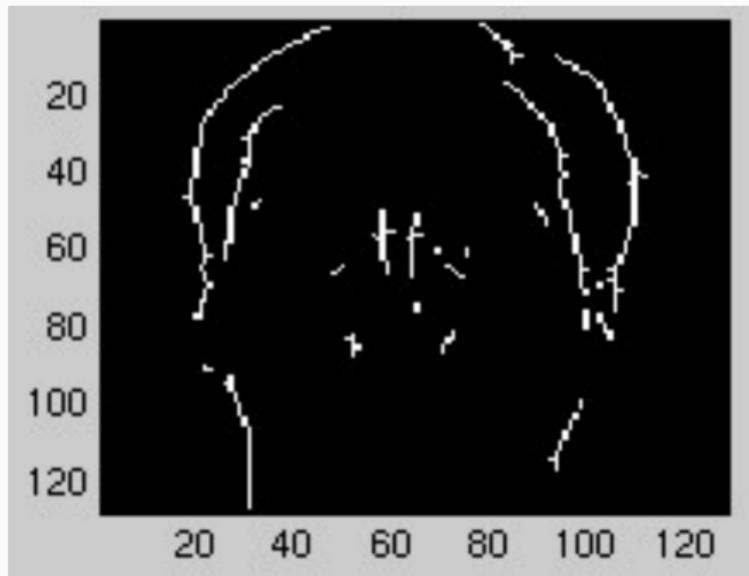


FIGURE : Détection des contours verticaux

## POURQUOI PLUSIEURS COUCHES ?

- Un autre neurone détectera les contours horizontaux

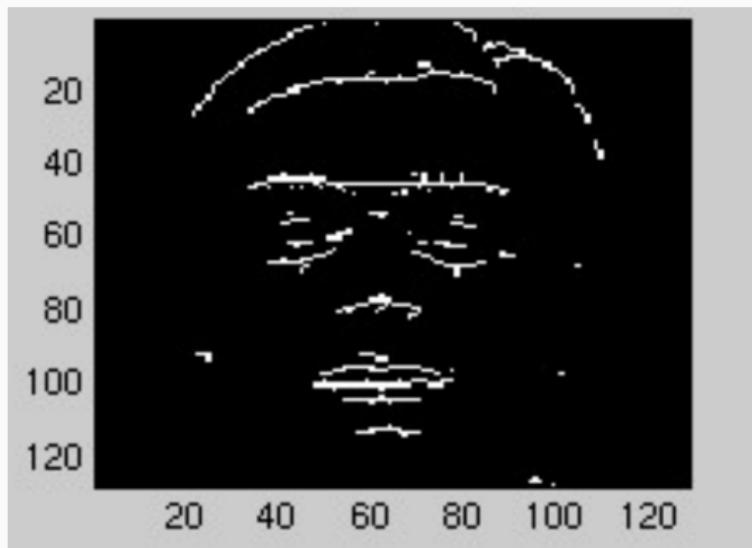
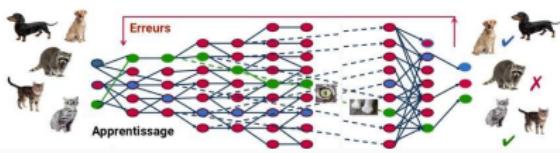


FIGURE : Détection des contours horizontaux

- On obtient enfin un visage quand on superpose les contours verticaux et horizontaux

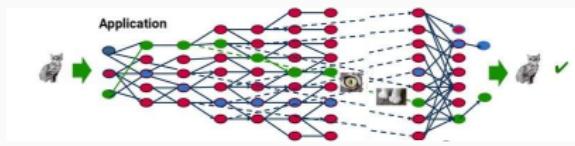
# UN PROCESSUS D'AUTOAPPRENTISSAGE



Processus  
d'autoapprentissage par  
essai-erreur

- Connaît ce que devrait être la sortie des neurones de la dernière couche (non-cachée)
- Entraînement du réseau de neurones jusqu'à ce que l'ordinateur donne la bonne réponse
- A chaque itération : "mauvaises" réponses éliminées et renvoyées vers les couches précédentes pour améliorer le modèle mathématique

# UN PROCESSUS D'AUTOAPPRENTISSAGE



Application à des données  
inconnues

- Quand appliqué à d'autres cas, le modèle capable de reconnaître un chat, même si le concept de chat lui est inconnu
- Données de départ (d'entraînement) = essentielles. De la variété des données dépend la performance finale.

## DES EXEMPLES D'APPLICATION

- La reconnaissance de l'écriture
- La reconnaissance vocale
- La reconnaissance des visages
- Les jeux-vidéos (simuler un comportement intelligent)
- Les robots : les robots-aspirateur par exemple

D'autres exemples sur [Wikipédia](#)

## IA POUR LES HUMANITÉS : UN EXEMPLE

---

## A QUOI SERT L'INTELLIGENCE ARTIFICIELLE ?

- Numérisation en masse de textes, d'images etc. permet d'avoir accès à une masse immense de données
- Problème : comment les traiter ?
- Dans la plupart des cas, travailler sur ces immenses corpus demanderait un temps déraisonnable.
- Face à ces situations, intérêt de l'intelligence artificielle.

Benjamin Charles Germain Lee, "Machine learning, template matching, and the International Tracing Service digital archive : Automating the retrieval of death certificate reference cards from 40 million document scans", in Digital Scholarship in the Humanities, 2018, [en ligne](#)

- Retrouver de la façon la plus exhaustive et efficace possible, les informations sur les personnes décédées en camp de travail ou de concentration sous le régime nazi
- A partir d'une copie numérique des [Archives Arolsen](#) : retrouver les noms de celles et ceux qui sont morts en camp de travail et de concentration

Informatique (ici IA) va aider le centre d'archives à remplir ses fonctions :

- Donner des informations sur la destinée de personnes persécutées par le régime nazi
- Rechercher des parents proches
- Délivrer des informations aux survivants et aux familles des victimes du nazisme
- 2007 : Archives Arolsen dupliquent ses archives numérisées
- Copie de ces archives numérisées disponibles via **US Holocaust Memorial Museum**, pour en permettre l'accès aux familles et aux chercheurs.

## LES SOURCES : LES ARCHIVES AROLSEN

- Existence d'un **Central Name Index** (CNI) qui rassemble des cartes avec des informations sur des personnes déplacées après leur libération des camps.
- Fichier de près de 46 millions de cartes concernant 17,5 millions de personnes

# EXEMPLES DE FICHES

(g)

Name: KONECZNY, Josef		Age: 43,40
File No.: 2-16285		
Sex: Male		
B.D.: 20.11.1905		
B.P.: X Ref.		
Address: 11, Bismarckstrasse, Berlin		
Occupation: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Last known residence: 11, Bismarckstrasse, Berlin		
Date of birth: 20.11.1905		
Place of birth: 11, Bismarckstrasse, Berlin		
Relatives: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Address: 11, Bismarckstrasse, Berlin		
Relation: Son		
Remarks: 11, Bismarckstrasse, Berlin		

Inquiry Card, Variant 1 (CNI card of Josef Koneczny, 0.1/28005270/ITS Digital Archive, USHMM)

(h)

Name: KONECZNY, Josef		Age: 43,40
File No.: 2-16285		
Sex: Male		
B.D.: 20.11.1905		
B.P.: X Ref.		
Address: 11, Bismarckstrasse, Berlin		
Occupation: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Last known residence: 11, Bismarckstrasse, Berlin		
Date of birth: 20.11.1905		
Place of birth: 11, Bismarckstrasse, Berlin		
Relatives: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Address: 11, Bismarckstrasse, Berlin		
Relation: Son		
Remarks: 11, Bismarckstrasse, Berlin		

Inquiry Card, Variant 2 (CNI card of Josef Koneczny, 0.1/28005232/ITS Digital Archive, USHMM)

(i)

Name: KONECZNY, Josef		Age: 43,40
File No.: 2-16285		
Sex: Male		
B.D.: 20.11.1905		
B.P.: X Ref.		
Address: 11, Bismarckstrasse, Berlin		
Occupation: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Last known residence: 11, Bismarckstrasse, Berlin		
Date of birth: 20.11.1905		
Place of birth: 11, Bismarckstrasse, Berlin		
Relatives: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Address: 11, Bismarckstrasse, Berlin		
Relation: Son		
Remarks: 11, Bismarckstrasse, Berlin		

Zentralnamenskartei (ZNK) (Original Card Type 2) (CNI card of Josef Koneczny, 0.1/28005219/ITS Digital Archive, USHMM)

(j)

Name: KONECZNY, Josef		Age: 43,40
File No.: 2-16285		
Sex: Male		
B.D.: 20.11.1905		
B.P.: X Ref.		
Address: 11, Bismarckstrasse, Berlin		
Occupation: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Last known residence: 11, Bismarckstrasse, Berlin		
Date of birth: 20.11.1905		
Place of birth: 11, Bismarckstrasse, Berlin		
Relatives: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Address: 11, Bismarckstrasse, Berlin		
Relation: Son		
Remarks: 11, Bismarckstrasse, Berlin		

Württemberg-Baden Card (Original Card Type 2) (CNI card of Josef Koneczny, 0.1/28005119/ITS Digital Archive, USHMM)

(k)

Name: KONECZNY, Josef		Age: 43,40
File No.: 2-16285		
Sex: Male		
B.D.: 20.11.1905		
B.P.: X Ref.		
Address: 11, Bismarckstrasse, Berlin		
Occupation: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Last known residence: 11, Bismarckstrasse, Berlin		
Date of birth: 20.11.1905		
Place of birth: 11, Bismarckstrasse, Berlin		
Relatives: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Address: 11, Bismarckstrasse, Berlin		
Relation: Son		
Remarks: 11, Bismarckstrasse, Berlin		

Allied Expeditionary Force (AEF) Registration Card (Original Card Type 2) (CNI card of Josef Koneczny, 0.1/28005297/ITS Digital Archive, USHMM)

(l)

Name: KONIECZNY, Josef		Age: 43,40
File No.: 2-16285		
Sex: Male		
B.D.: 20.11.1905		
B.P.: X Ref.		
Address: 11, Bismarckstrasse, Berlin		
Occupation: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Last known residence: 11, Bismarckstrasse, Berlin		
Date of birth: 20.11.1905		
Place of birth: 11, Bismarckstrasse, Berlin		
Relatives: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Address: 11, Bismarckstrasse, Berlin		
Relation: Son		
Remarks: 11, Bismarckstrasse, Berlin		

Care and Maintenance CM/1 Card (Reference Card) (CNI card of Josef Konierzny, 0.1/28005160/ITS Digital Archive, USHMM)

(m)

Name: KONECZNY, Stanislaw		Age: 43,40
File No.: 2-16285		
Sex: Male		
B.D.: 20.11.1905		
B.P.: X Ref.		
Address: 11, Bismarckstrasse, Berlin		
Occupation: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Last known residence: 11, Bismarckstrasse, Berlin		
Date of birth: 20.11.1905		
Place of birth: 11, Bismarckstrasse, Berlin		
Relatives: 1. Befehlshaber der 1. Kompanie, 1. Bataillon, 1. Infanterie-Division, Preuss.		
Address: 11, Bismarckstrasse, Berlin		
Relation: Son		
Remarks: 11, Bismarckstrasse, Berlin		

Multi-Line Card (Reference Card) (CNI card of Stanislaw Koneczny, 0.1/28003281/ITS Digital Archive, USHMM)

# OBJECTIF

- Cartes indexées par noms de personnes, et pas par types.
- Comment repérer facilement et rapidement les certificats de décès parmi ces 46 millions de fiches ?

- OCR<sup>2</sup> non applicable en raison de la piète qualité de la numérisation
- Utilisation du “template matching” en apprentissage supervisé : combinaison de vision de la machine et de l’apprentissage machine
  - Entraînement réalisé sur des portions d’image (ici, des mentions apparaissant sur les fiches), étiquetées manuellement de la façon suivante :
    - Carte de certificat de décès (2 lignes en Allemand)
    - Carte de référence pour les certificats de décès (2 lignes en anglais)
    - Carte de référence pour les certificats de décès (3 lignes, Allemand)
    - Miscellanées (pas certificats de décès - le reste)
  - Echantillon d’entraînement de 10698 fiches
  - Entraînement de la machine à reconnaître, parmi les millions de fiches, celles qui correspondent à des certificats de décès

---

2. Reconnaissance optique de caractères : transformation automatique d’une image en texte exploitable via un traitement de texte par exemple

# RÉSULTATS

- Après entraînement, application de l'algorithme sur plus de 39 millions de cartes numérisées.
- 312183 identifications en moins de 14 jours, sur un bon ordinateur de bureau => près de 100% de réussite
- Pourrait facilement être appliqué à d'autres archives

## UN AUTRE EXEMPLE : LA TRADUCTION AUTOMATIQUE

**Ulrike Wuttke**  
@UWuttke

Auf YouTube gibt es viele gute Yogavideos für verschiedene Stufen zum zu Hause üben. Meine liebsten #Yoga Channels sind von Mady Morrison (dt.) und Yoga with Kassandra (engl.). Vielleicht bietet Eurer lokales Yogastudio online Klassen an? Bleibt gesund & Namaste! 🙏

À l'origine en allemand et traduit par Google

Il existe de nombreuses bonnes vidéos de yoga pour différents niveaux à pratiquer à la maison sur YouTube. Ma chérie #Yoga Les chaînes sont de Mady Morrison (allemand) et Yoga with Kassandra (anglais). Peut-être que votre studio de yoga local propose des cours en ligne? Restez en bonne santé et Namaste! 🙏

7:19 AM · 23 mars 2020 · [TweetDeck](#)

**FIGURE :** Traduction automatique d'un tweet en allemand

# UN AUTRE EXEMPLE : LA TRADUCTION AUTOMATIQUE

Environ 80 800 résultats (0,46 secondes)

[en.wikipedia.org › wiki › Maya\\_Jasanoff](#) ▾ [Traduire cette page](#)

## Maya Jasanoff - Wikipedia

Maya R. Jasanoff is an American academic. She serves as Coolidge Professor of History at Harvard University, where she focuses on the history of Britain and ...  
Education: [Harvard College; Cambridge Univ...](#) Parents: [Jay Jasanoff](#) (father); [Sheila Jasanoff](#) ...  
[Career](#) · [Books](#) · [Awards](#) · [Bibliography](#)

[history.fas.harvard.edu › people › maya-jasanoff](#) ▾ [Traduire cette page](#)

## Maya Jasanoff | Harvard University | History Department

Maya Jasanoff's teaching and research extend from the history of the British Empire to global history. She is the author of three prize-winning books. The Dawn ...

[ces.fas.harvard.edu › People](#) ▾ [Traduire cette page](#)

## Maya Jasanoff | Center for European Studies at Harvard ...

Maya Jasanoff is Coolidge Professor of History at Harvard University and Resident Faculty at the Minda de Gunzburg Center for European Studies. She ...

[www.fnac.com › Maya-Jasanoff](#) ▾

## Maya Jasanoff : tous les produits | fnac



## Maya Jasanoff

Professeure

Traduit de l'anglais - Maya R. Jasanoff est une universitaire américaine. Elle est professeur d'histoire Coolidge à l'Université de Harvard, où elle se concentre sur l'histoire de la Grande-Bretagne et de l'Empire britannique. [Wikipédia \(anglais\)](#)

[Afficher la description d'origine](#) ▾

**Date de naissance :** 1974 (Âge: 46 ans)

**FIGURE :** Traduction automatique du début d'un article Wikipédia (résultats Google)

- Dans les années 30 : premiers brevets pour des "machines de traduction". Par exemple : celle de Georges Artsrouni ou Peter Trojanskij
- 7 janvier 1954 : "Georgetown-IBM experiment" présente la première machine à traduire à IBM (New York) => possible seulement de traduire 250 mots et 49 phrases soigneusement choisis du russe vers l'anglais

- Effets positifs : stimulation de la recherche et des financements aux US et dans le monde
  - US dans les années 60 : besoin de traduire massivement des textes en russe
  - URSS : même besoin, pour traduire des textes en anglais
- 1968 : création par le gouvernement américain de la société Systran (System Translation)

## UNE DEMANDE QUI EXPLOSE

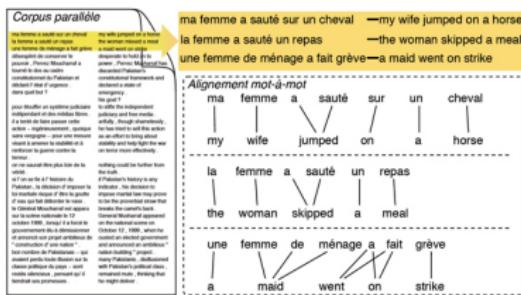
- Influence de la construction européenne : service de traduction devenu de plus en plus important au fur et à mesure des élargissements de l'Union Européenne + de plus en plus de traducteurs professionnels qui s'aident de la machine
- Demande croissante avec l'explosion du Web dans les années 90 : de plus en plus de textes dans de plus en plus de langues différentes
- Demande encore plus importante avec l'explosion des réseaux sociaux : traduction instantanée de tweets, de posts, impossible à réaliser par des humains au vu du volume des données à traiter

## LE SERVICE BABEL FISH DANS LES ANNÉES 90-2000

- Décembre 1997 : lancement de AltaVista Translation Service par Digital Equipment Corporation (DEC) et Systran à l'adresse "babelfish.altavista.com"
- Babel Fish fait référence au "Poisson Babel" imaginé par Douglas Adams dans *Le Guide du voyageur galactique* (1978) : espèce de petit poisson extraterrestre qui, enfoncé dans l'oreille, permet de comprendre toutes les langues de l'univers
- Alimenté par des dictionnaires multilingues
- Performances peuvent laisser à désirer, mais permet de comprendre des documents en langue étrangère
- Mai 2012 : remplacé par Bing Translator, jusqu'à son arrêt en juin 2013

# L'APPARITION DE LA TRADUCTION STATISTIQUE DANS LES ANNÉES 90

- Apparition à la fin des années 80 de grands corpus bilingues "alignés" comme le **Hansard** canadien (transcriptions officielles des débats parlementaires canadiens en français et en anglais)
- Traduction statistique repose sur un corpus parallèle constitués de textes dans différentes langues et qui sont alignés = alignment met en relation des unités (paragraphes, phrases, mots)



- Permet de déterminer traductions futures à partir de traduction existantes
- 2007 : approche utilisée par **Google Traduction**

## GÉNÉRATION DE SCORES DE TRADUCTION

la	the	0.6
la	this	0.4

- Chaque ligne = une règle de traduction
  - 1ère colonne = mot ou séquence de mots dans la langue source
  - 2ème colonne = mot ou séquence de mots dans la langue cible
  - 3ème colonne = score. Ici score du modèle de traduction c'est-à-dire la probabilité que ce mot soit traduit de cette manière
    - "La" a une probabilité de 0.6 (60% des cas) d'être traduit par "The"
    - "La a une probabilité de 0.4 (40% des cas) d'être traduit par "This"
- Phrase à traduire = segmentée en unités (mots, séquences de mots et signes de ponctuation)
- Production d'hypothèses de traduction en utilisant les règles présentes dans le dictionnaire bilingue pour traduire les unités reconnues
- Chaque hypothèse associée à un score de traduction : la traduction finale retournée = celle avec le plus haut score

- Apparition dans le milieu des années 2010 de l'apprentissage profond (Deep Learning)
- Avantage de l'apprentissage profond sur l'approche statistique : considère la phrase dans son entier c'est-à-dire que l'on ne traduit pas la phrase par fragments que l'on assemble ensuite.
- 2016 : Approche généralisée à l'ensemble des acteurs de la traduction automatique lorsque Google a annoncé le remplacement de son système de traduction basé sur approche statistique, par un modèle d'apprentissage profond

=> Apprentissage profond immédiatement imposé au vu de l'amélioration des résultats qu'il a apporté au service de traduction de Google

## OBJECTIF : COMPRENDRE LE SENS DE LA PHRASE

- Assez rapidement (dès années 60), est apparue la nécessité de comprendre le sens de la phrase pour la traduire correctement
- Comprendre le sens de la phrase en contexte : même phrase pas exactement le même sens en fonction des autres phrases qui l'entourent
- Autrement dit : pas possible de faire de la traduction mot à mot

## COMPRENDRE LE SENS DE LA PHRASE : LE MOT "TERRE" EN ANGLAIS

En français : terre = sol, planète, la côte vue de la mer ou encore une propriété agricole.

En anglais, traduction varie en fonction du contexte :

- La terre du jardin : "soil"
- Astronaute qui voit la Terre : "Earth!"
- Le terrain de tennis de Roland-Garros (terre battue) : "clay tennis court"
- Un marin qui voit la côte : "Land!"

# L'APPLICATION À LA TRADUCTION AUTOMATIQUE

1. Procéder à une analyse contextuelle globale => sens des mots représentés par leur entourage
2. Regrouper les mots en ensembles larges, sémantiquement homogènes baptisés "plongements de mots" ("word embeddings" en anglais)

Avantage : on tient compte du contexte du mot considéré et du contexte des mots qui possèdent un sens proche => améliore considérablement la traduction des mots, notamment les mots rares

- Analyse hiérarchique de la phrase : on prend en compte le contexte de chaque mot, puis de chaque groupe de mots, puis de chaque phrase complète
- Représentation contextuelle dynamique car elle change pour chaque unité linguistique
- Succès de la méthode car se rapproche de la manière dont le cerveau humain comprend les mots et les phrases => le système infère les connaissances en faisant des rapprochements sémantiques entre les mots. Capable donc de produire des traductions plus pertinentes selon le contexte

Enfin on génère la phrase dans la langue cible.

# LE LANCEMENT DE DEEPL TRANSLATOR

- 2017 : lancement de DeepL Translator, basé sur intelligence artificielle
- Société qui existe depuis 2009 : Linguee = premier moteur de recherche de traduction sur Internet
  - Outil très populaire (en 2017, plus de 10 milliards de demandes, plus d'un milliard d'utilisateurs)
  - Ne traduit pas un mot par son équivalent dans la langue cible mais fournit un mot dans son contexte avec des exemples
  - Données proviennent de sites Web bilingues (par ex. brevets, textes du parlement européen, Unesco)

▲ Sources externes (non révisées)

[...] commercialisant de la boue, du gravier, du sable et de la terre, j'imagine qu'il pourra aménager une carrière de gravier.	If one can make a living out of marketing mud, gravel, sand and dirt, then I suppose one could create a gravel pit. <a href="http://www2.parf.gc.ca">www2.parf.gc.ca</a>
J'ai été chanceux, mais Dame Fortune est bien inconstante, et je ne m'assais pas ma terre sur elle.	But Lady Luck is a fickle mistress and I certainly wouldn't bet the farm on her. <a href="http://aca-fcc.ca">aca-fcc.ca</a>
[...] améliorations sont donc nécessaires pour rétablir la fonction du drain et maintenir la productivité de la terre agricole.	Thus maintenance, repairs or improvements are required to restore the drain's function in order to maintain the productivity of the agricultural land. <a href="http://ofo-mpo.gc.ca">ofo-mpo.gc.ca</a>
[...] étang pour laquiner le saumon ou la truite est un réel plaisir pour bien des pêcheurs à la ligne de Terre Neuve et du Labrador.	Getting out on a river or pond in pursuit of a salmon or trout is a highlight for many anglers in Newfoundland and Labrador. <a href="http://ofo-mpo.gc.ca">ofo-mpo.gc.ca</a>

FIGURE : Exemples de traduction en contexte (Linguee)

# LE LANCEMENT DE DEEPL TRANSLATOR

- Linguee : données d'entraînements pour DeepL
- Réseau neuronal de DeepL entraîné sur ces données grâce à un superordinateur installé en Islande
- Très grande puissance de calcul : 5,1 milliards d'opérations par seconde
- Considéré comme meilleur système de traduction automatique actuellement

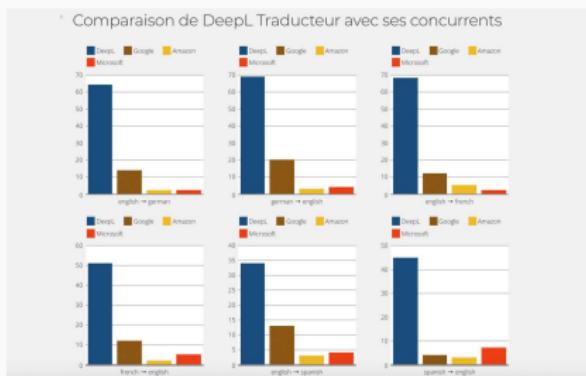


FIGURE : Evaluation par des traducteurs professionnels (fév. 2020)

- Technologie très gourmande en données => besoin de beaucoup de données pour obtenir de bons résultats
- Domination de l'anglais
  - Corpus parallèles de plusieurs millions de mots : peu probable que l'on dispose de corpus bilingues suffisants au delà d'une dizaine / quinzaine de langues
  - Plus facile de traduire un texte vers l'anglais : peu de variation des mots en anglais (genre, nombre voire temps verbal peu marqué)
  - Plus compliqué en allemand ou en russe par exemple, car beaucoup d'informations attachées aux mots