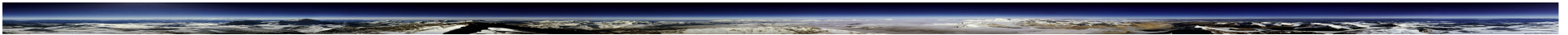


L'Intelligence Artificielle

Quelle perspective sur la/les mémoire(s) ?



Antoine Cornuéjols

AgroParisTech – INRA MIA 518

Plan

1. Bref **historique** de l'Intelligence Artificielle
2. Mémoire, **information** et **calcul**
3. Mémoire et **apprentissage**
4. La mémoire **pour comprendre**
5. Le **big data** et les réseaux de neurones profonds
6. Conclusions : l'IA et **les grandes questions** sur la mémoire

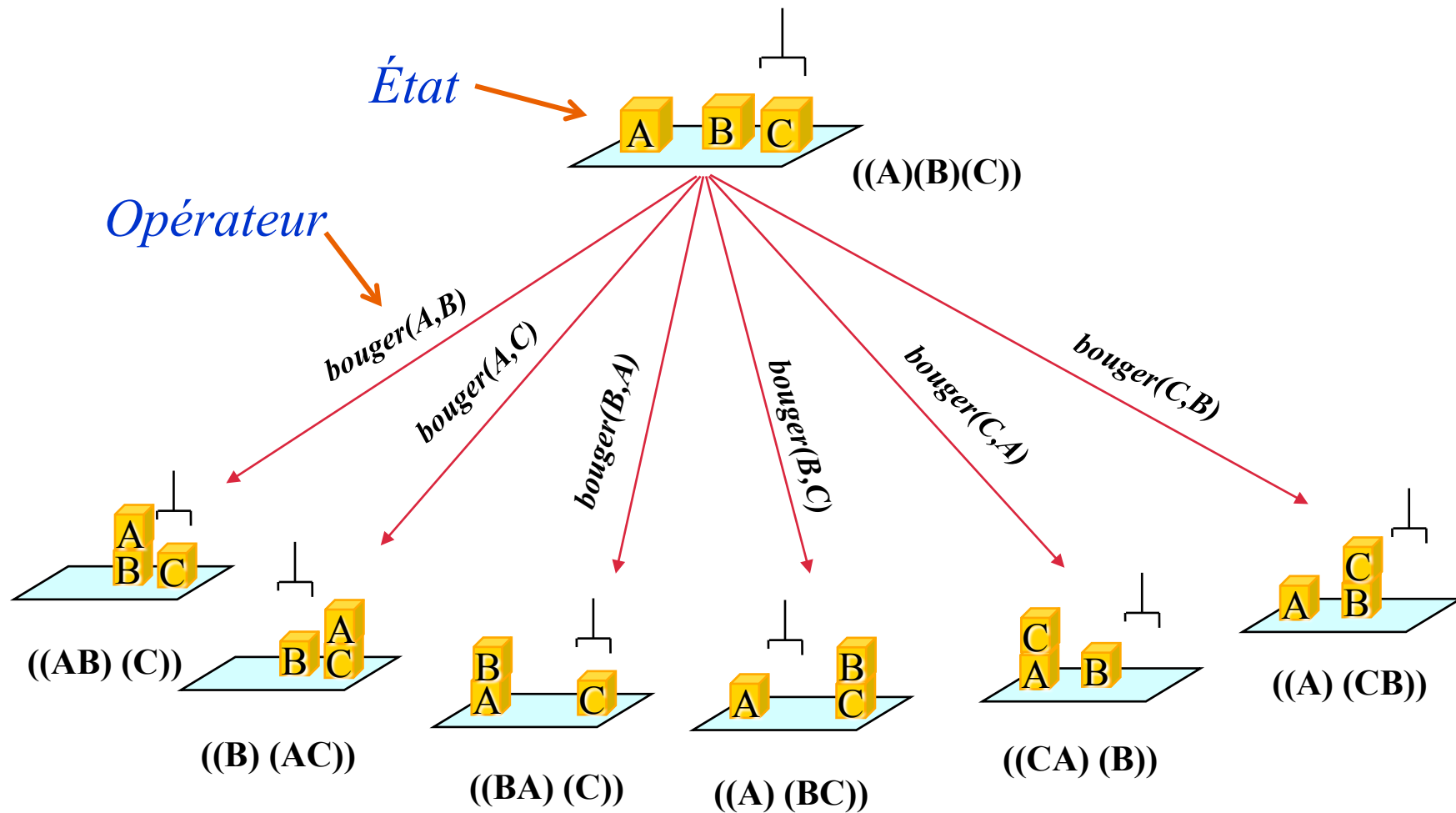
Bref historique de l'IA

IA = méthodes générales de **raisonnement**

1956 - ~ 1970

- **Raisonnement** = manipulation de **représentations** discrètes des connaissances
 - Démonstrateurs de **théorèmes**
 - Résolveurs universels de **problèmes** (GPS)
 - CHECKER : vice-champion du monde au **jeu** de dames (1962)

Raisonnement / résolution de problèmes

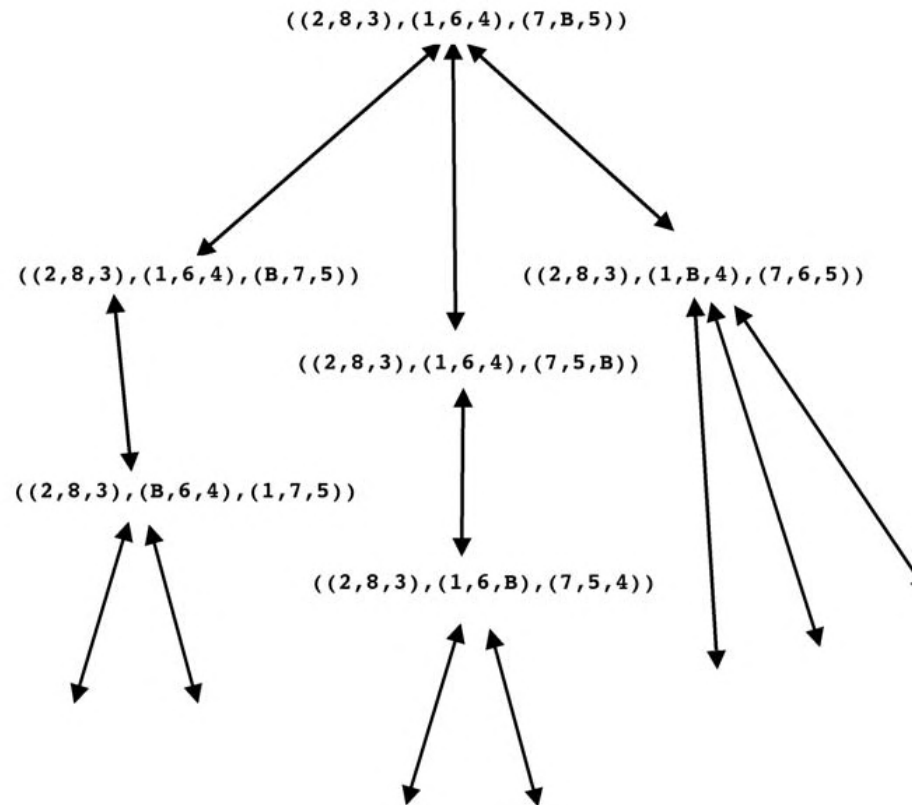


Raisonnement / résolution de problèmes

2	8	3
1	6	4
7		5



1	2	3
8		4
7	6	5



- Méthodes de recherche dans un graphe

IA = méthodes générales de raisonnement

1956 - ~ 1970

- Raisonnement = manipulation de **représentations** discrètes des connaissances
 - Démonstrateurs de théorèmes
 - Résolveurs universels de problèmes (GPS)
 - CHECKER : vice-champion du monde au jeu de dames
- Mais l'intelligence **n'est pas que ça**
- Les experts possèdent un répertoire énorme de **connaissances**

IA : « *Knowledge is power* »

~1970 - ~1985

- Il faut représenter des connaissances multiples et complexes
 1. Représentations structurées : réseaux sémantiques, scripts, schémas, ...
 2. Développement des « systèmes experts »

IA : recours à des méthodes générales d'apprentissage

~1985 - ...

- Développement de nouvelles méthodes d'apprentissage
 - Dont les réseaux de neurones (2^{ème} génération : 1985)
 - Puis les réseaux de neurones profonds (2007 - ...)

Plan

1. Bref historique de l'Intelligence Artificielle
2. Mémoire, information et calcul
3. Mémoire et apprentissage
4. La mémoire pour comprendre
5. Le big data et les réseaux de neurones profonds
6. Conclusions : l'IA et les grandes questions sur la mémoire

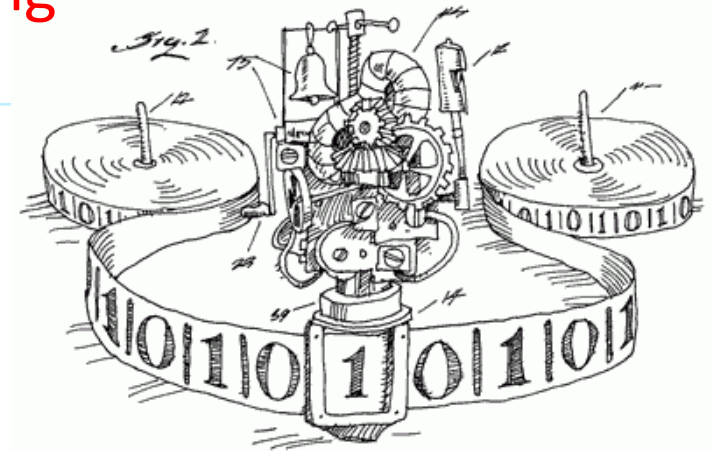
Mémoire, information et calcul

La machine de Turing

Une *machine de Turing* est composée de :

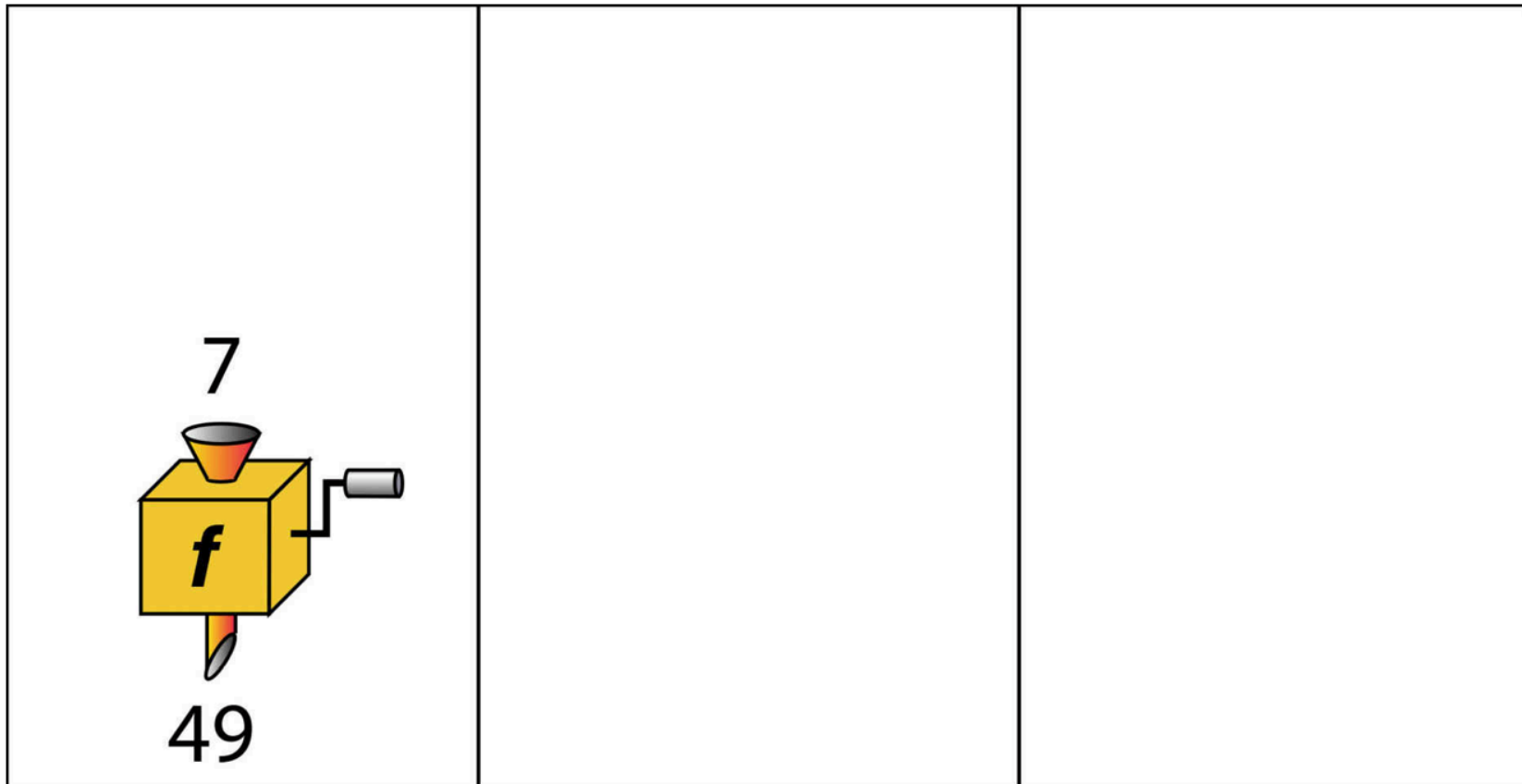
- Un **ruban infini** contient des **cases** sur lesquelles un **symbole** peut être écrit ou lu
- Une **tête de lecture/écriture** peut se déplacer d'une **case** à droite ou à gauche
- Un **ensemble de règles** définissant ce que la machine doit faire quand un symbole est lu en **fonction de l'état courant de la machine**

Le même ruban code les **données** et le **programme** !



Algorithmes ... et ... apprentissage

- La machine de Turing définit tout ce qui est calculable



Plan

1. Bref historique de l'Intelligence Artificielle
2. Mémoire, information et calcul
3. Mémoire et apprentissage
4. La mémoire pour comprendre
5. Le big data et les réseaux de neurones profonds
6. Conclusions : l'IA et les grandes questions sur la mémoire

Mémoire et apprentissage

Le problème de l'induction

- Exemples décrits par :
 - Nombre** (1 ou 2); **taille** (petit ou grand); **forme** (cercle ou carré); **couleur** (rouge ou vert)
- De classe + ou -

Description	Votre prédiction	Vraie classe

Combien de fonctions possibles de X vers Y ?

$$2^{2^4} = 2^{16} = 65,536$$

Combien de fonctions restent après 6 exemples d'apprentissage? $2^{10} = 1024$

Le problème de l'induction

- Exemples décrits par :

Nombre (1 ou 2); **taille** (petit ou grand); **forme** (cercle ou carré); **couleur** (rouge ou vert)

Description	Votre prédiction	Vraie classe
1 grand carré rouge		-
1 grand carré vert		+
2 petits carrés rouges		+
2 grands cercles rouges		-
1 grand cercle vert		+
1 petit cercle rouge		+
1 petit carré vert		-
1 petit r carré ouge		+
2 grands carrés verts		+
2 petits carrés verts		+
2 petits cercles rouges		+
1 petit cercle vert		-
2 grands cercles verts		-
2 petits cercles verts		+
1 grand cercle rouge		-
2 grands carrés rouges	?	

15

Combien de
fonctions
restantes ?

← ?

17 / 76

Le problème de l'induction

- Examples described using:

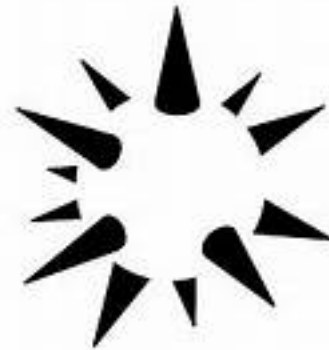
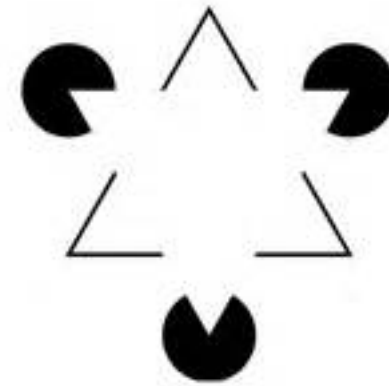
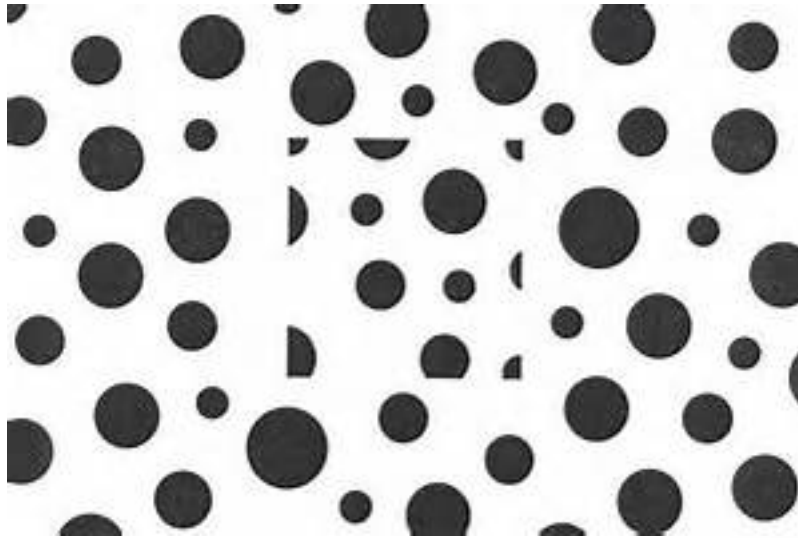
Number (1 or 2); **size** (small or large); **shape** (circle or square); **color** (red or green)

Description	Your prediction	True class
1 grand carré rouge		-
1 grand carré vert		+
2 petits carrés rouges		+
2 grands cercles rouges		-
1 grand cercle vert		+
1 petit cercle rouge		+

Combien de fonctions possibles avec 2 descripteurs de X à Y ? $2^{2^2} = 2^4 = 16$

Combien de fonctions restent après 3 exemples différents ? $2^1 = 2$

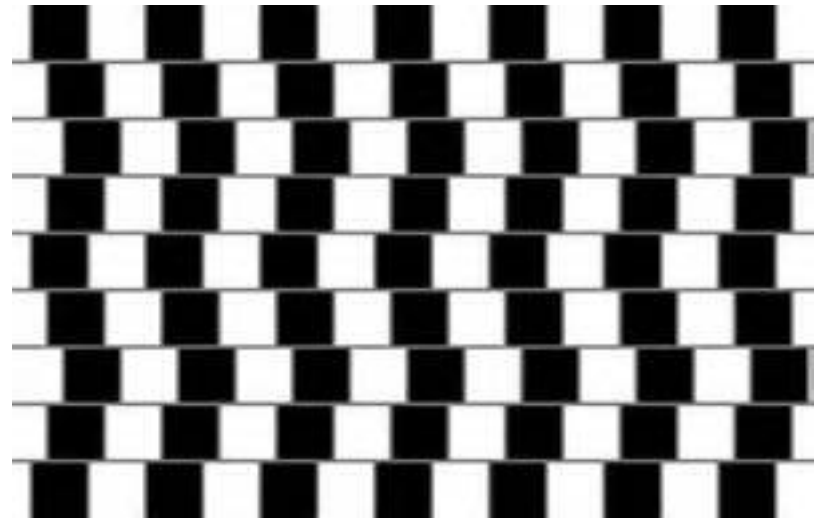
L'apprentissage – une extrapolation



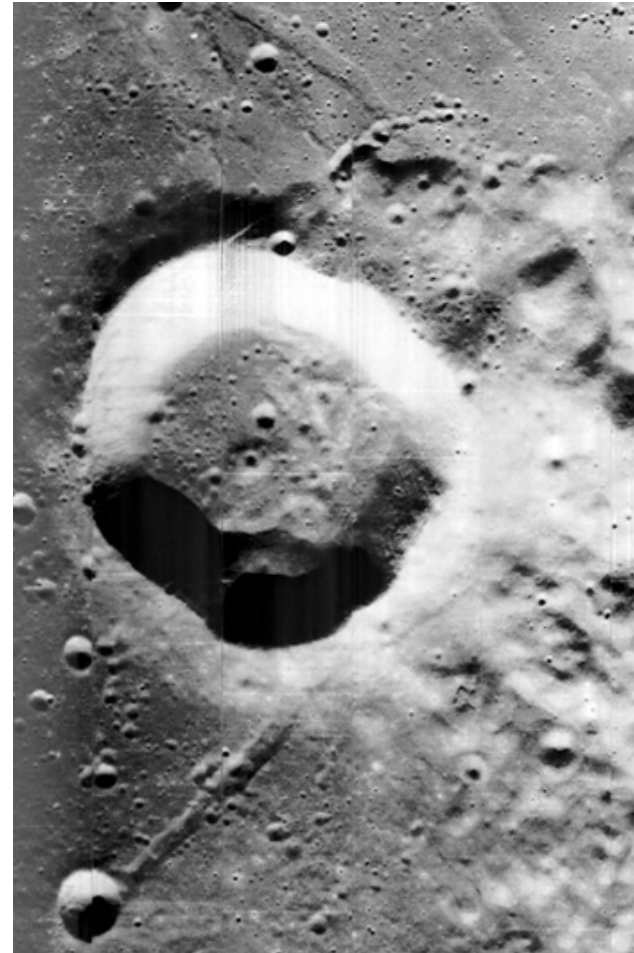
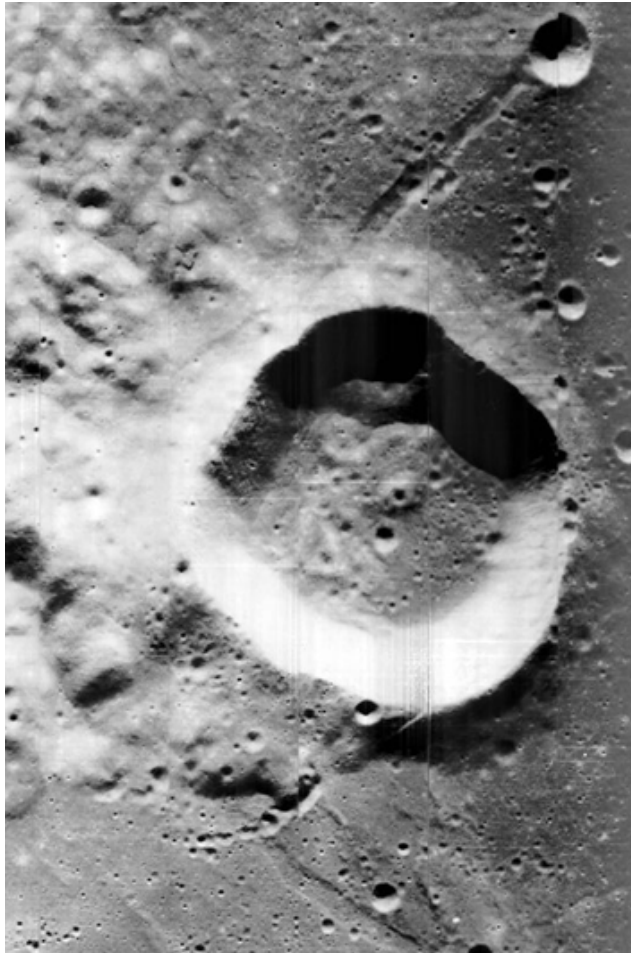
Une extrapolation – soumise à des choix



Des **biais** pouvant conduire à des **illusions**



Induction et illusions

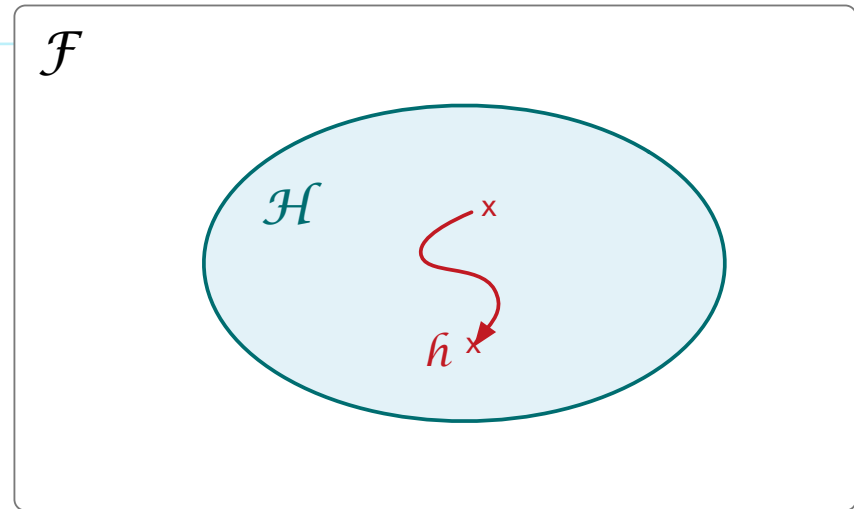


Cratère ou colline ?

Les types de connaissances

1. Les connaissances a priori

- Biais de représentation
- Biais de recherche



2. Les informations sur la fonction apprise (ou considérée)

3. Les données

Plan

1. Bref historique de l'Intelligence Artificielle
2. Mémoire, information et calcul
3. Mémoire et apprentissage
4. La mémoire pour comprendre
5. Le big data et les réseaux de neurones profonds
6. Conclusions : l'IA et les grandes questions sur la mémoire

La mémoire pour comprendre

C'est très (trop) simple

- Les « données » sont **incomplètes** et **ambiguës**

C'est très (trop) simple

- Les « données » sont **incomplètes** et **ambiguës**

1. Interprétation de scènes visuelles

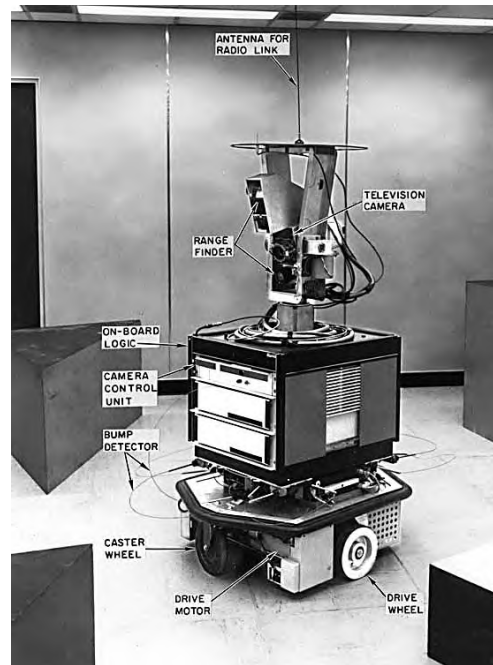


Figure 12.3: Shakey as it existed in November 1968 (with some of its components labeled). (Photograph courtesy of SRI International.)

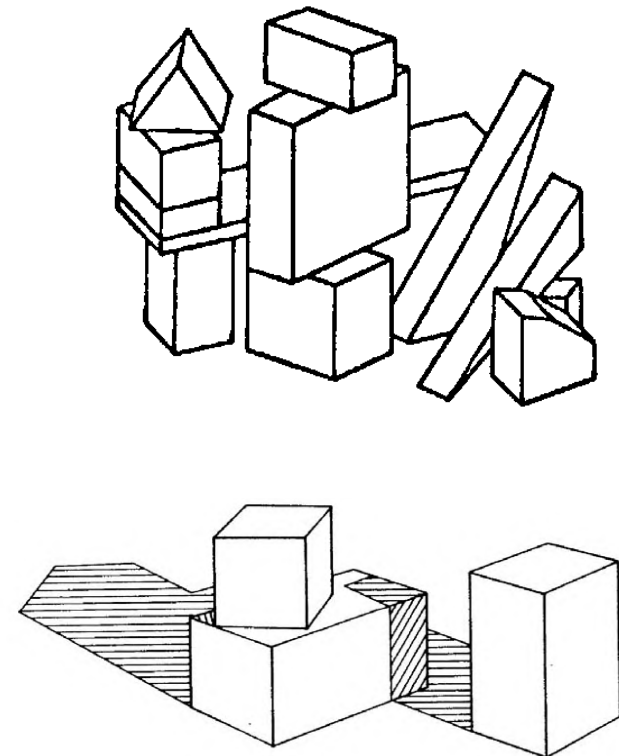


Figure 9.15: A scene with shadows analyzed by Waltz's program. (Illustration used with permission of David Waltz.)

C'est très (trop) simple

2. Interprétation du langage

“Richard hadn't heard from his college roommate Paul for years. Richard had borrowed money from Paul which was never paid back. But now he had no idea where to find his old friend. When a letter finally arrived from San Francisco, Richard was anxious to find out how Paul was.”

C'est très (trop) simple

2. Interprétation du langage

"Richard hadn't heard from his college roommate Paul for years. Richard had borrowed money from Paul which was never paid back. But now he had no idea where to find his old friend. When a letter finally arrived from San Francisco, Richard was anxious to find out how Paul was."

Comment traiter un non événement ?

"John walked in the room and Mary was not there" (John désirait voir Mary et s'attendait à la trouver dans la pièce -> pourquoi n'y est-elle pas ? ...)

Ici, on est prêt à comprendre que Richard se sent une obligation vis-à-vis de Paul. Il faut donc des connaissances sur les relations sociales et sur la psychologie.

Comment traiter "had" (possession ?) et "old" (âge de son ami ? Non. Mais pourquoi et comment?)

Comment infère-t-on que cette lettre vient de Paul alors que ce n'est pas dit explicitement ?

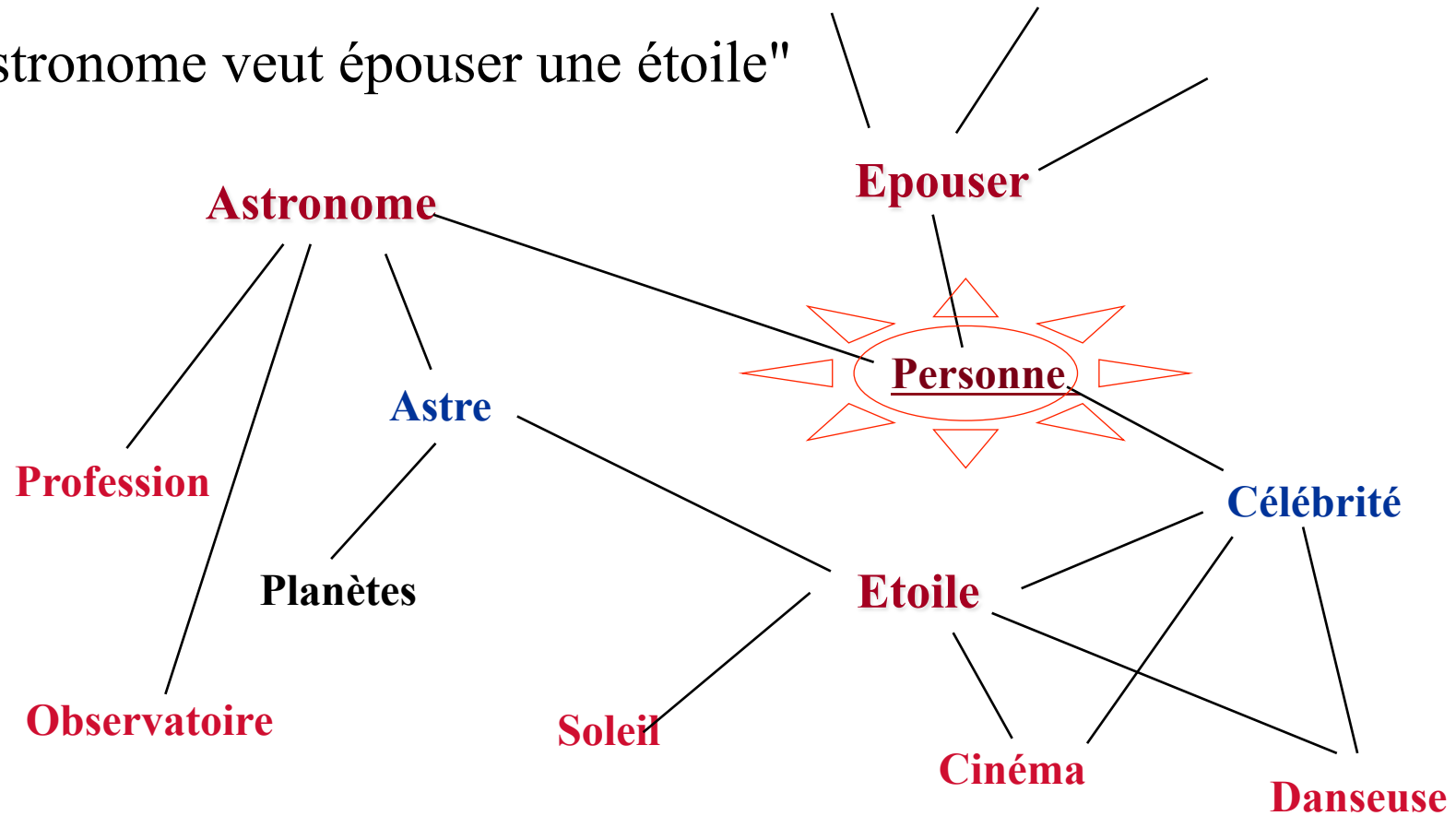
Les réseaux sémantiques

"L'astronome veut épouser une **étoile**"

Etoile ?

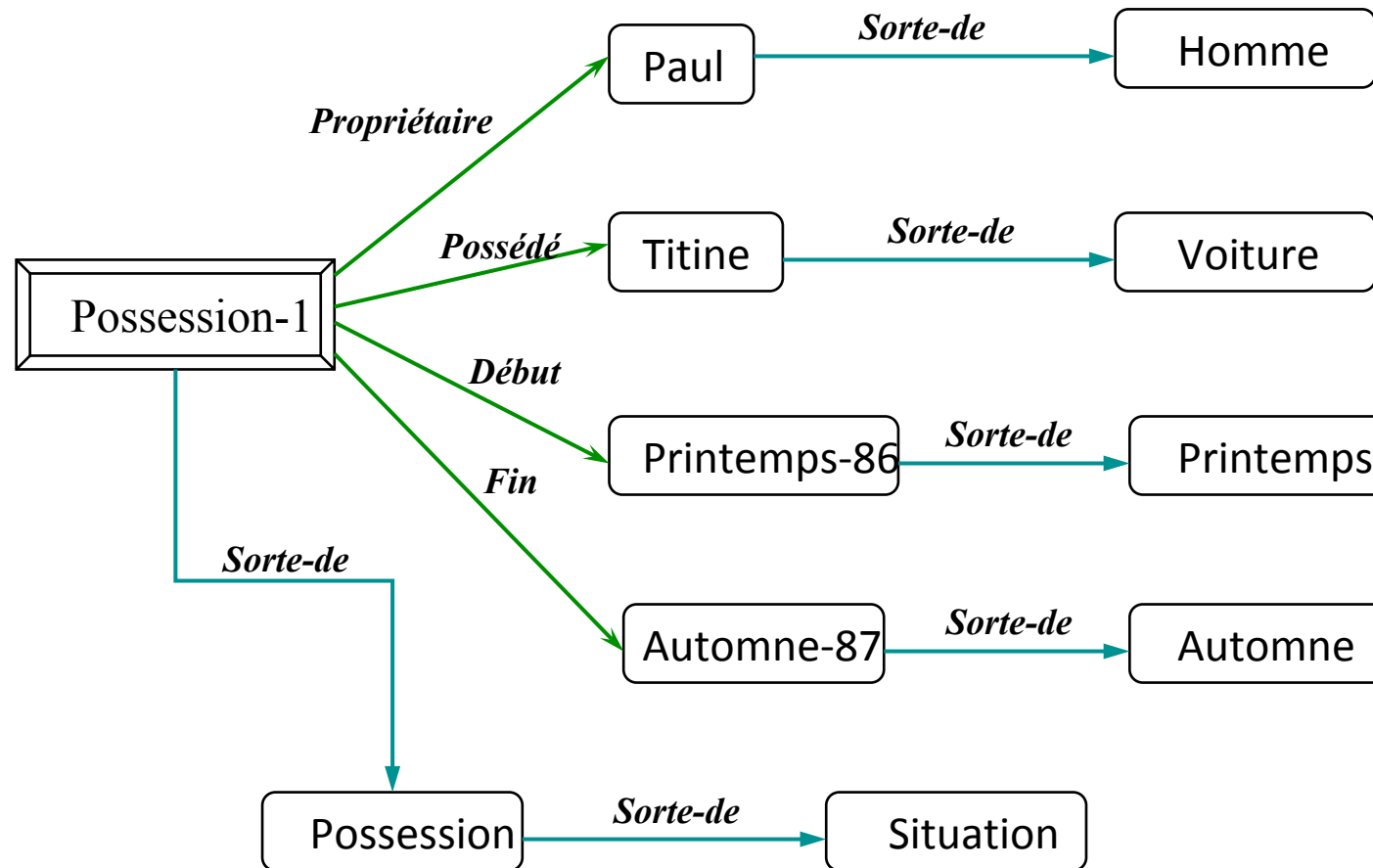
Les réseaux sémantiques

"L'astronome veut épouser une étoile"



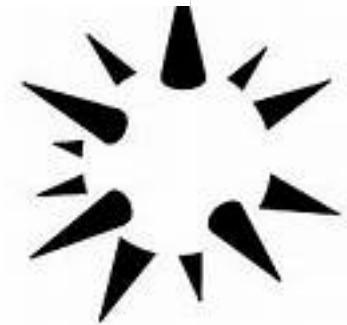
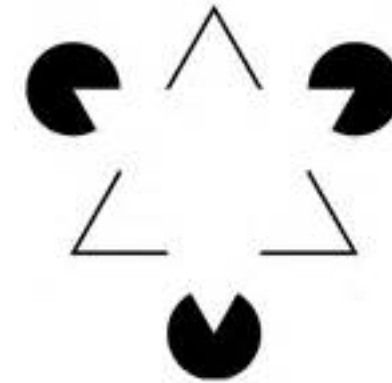
Inférence par propagation d'*activité* ou de *marqueurs* : focalise l'attention

Les réseaux sémantiques



De quoi a-t-on besoin pour interpréter ...

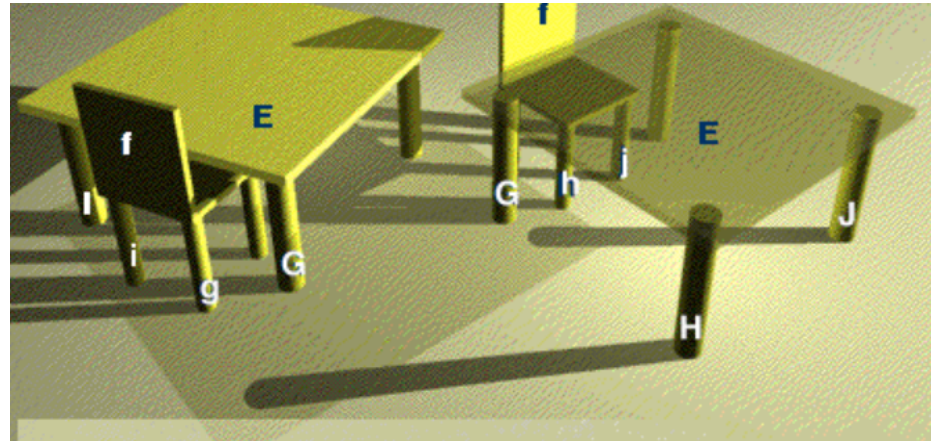
... des entrées **incomplètes / ambiguës** ?



Nous complétons grâce à des **connaissances a priori**

Les « frames » de Marvin Minsky (1974)

Connaissances pour interpréter des scènes visuelles



- When one encounters a **new situation** one **selects from memory a structure** called a *Frame*. This is a remembered *framework to be adapted to fit reality* by changing details as necessary.
- A *frame* is a data-structure for **representing a stereotyped situation**, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about *how to use the frame*. Some is about *what one can expect to happen next*. Some is about *what to do if these expectations are not confirmed*.

Les « frames » de Marvin Minsky (1974)

- A **Frame** is a collection of questions to be asked about a hypothetical situation; it specifies **issues** to be raised and **methods** to be used in dealing with them.
- Questions normales **pour comprendre**
 - *What caused it?* (agent)
 - *What was the purpose?* (intention)
 - *What are the consequences?* (side-effects)
 - *Who does it affect?* (recipient)
 - *How is it done?* (instrument)
- In seeing a room, or understanding a story, **one assembles a network of frames and subframes**. Everything noticed or guessed, rightly or wrongly, is represented in this network.

Les « frames » de Marvin Minsky (1974)

- **EXPECTATION**: How to *select* an initial frame to meet some given conditions.
- **ELABORATION**: How to select and assign subframes to *represent additional details*.
- **ALTERATION**: How to find a frame to *replace* one that does not fit well enough.
- **NOVELTY**: What to do if no acceptable frame can be found.
Can we *modify* an old frame or must we build a new one?
- **LEARNING**: What frames should be *stored*, or *modified*, as a result of the experience?

Les scripts pour comprendre les textes [Schank, 195]

- Script

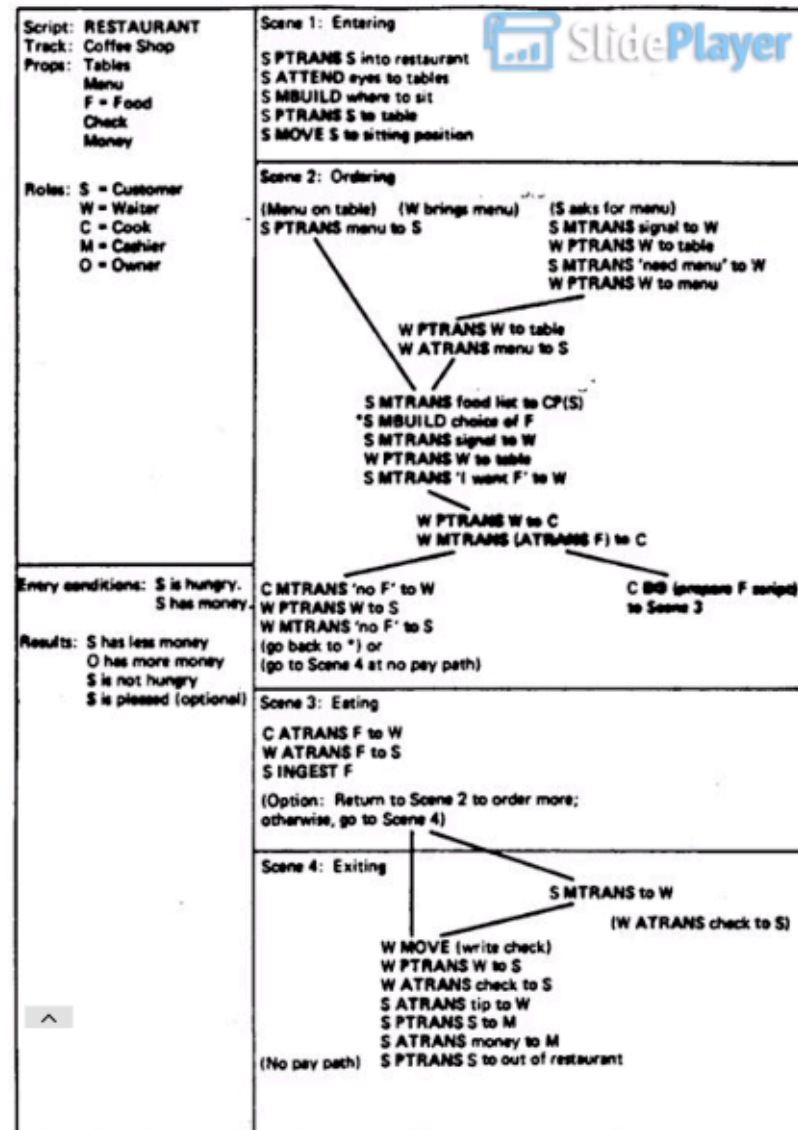
Plans and Scripts

Roger Schank

Example:
Restaurant Script

“John went to a restaurant. He ordered lobster. He paid the bill and left.”

What did John eat?



Types de mémoires

- Mémoire **épisodique**
 - Évènements et histoire des expériences de l'agent
- Mémoire **sémantique**
 - Connaissances conceptuelles

Types de mémoires

- Mémoire **épisodique**
 - Évènements et histoire des expériences de l'agent
- Mémoire **sémantique**
 - Connaissances conceptuelles

– *Que s'est-il passé de notable lors de votre dernier voyage en famille au Grand Canyon ?*

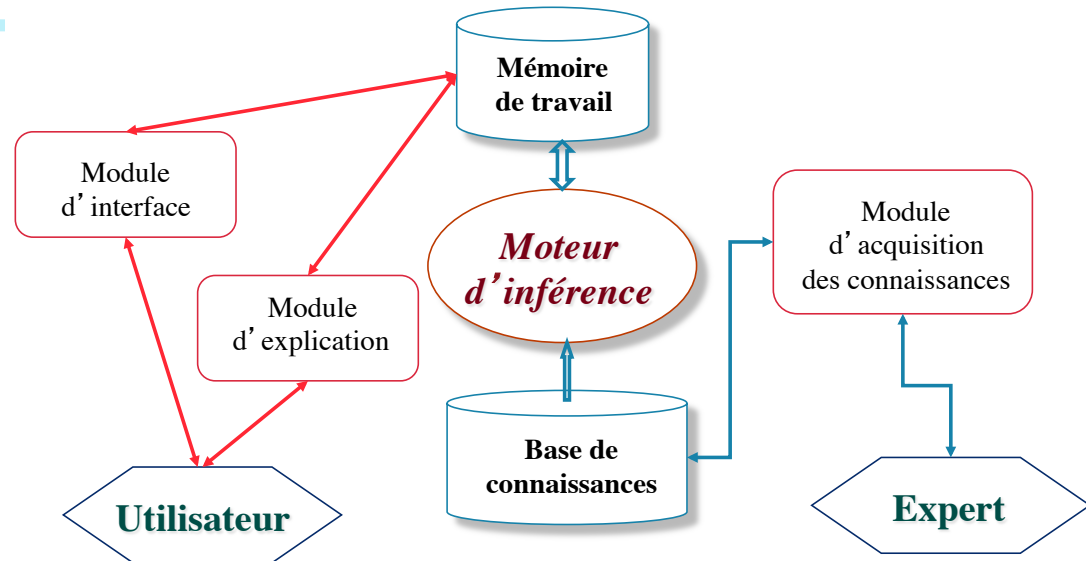
- *Mémoire épisodique*

– *Dans quel pays se situe le Grand Canyon ?*

- *Mémoire sémantique*

Types de mémoires

- Les systèmes experts



- Règle :

Si le spectre de la molécule présente deux pics x_1 et x_2 tels que :

1. $x_1 - x_2 = M + 28$
2. $x_1 - 28$ est un pic élevé
3. $x_2 - 28$ est un pic élevé
4. au moins l'un des pics x_1 et x_2 est élevé

Alors la molécule contient un groupe cétone

Types de mémoires

- Mémoire à **court terme** / Mémoire de **travail**
 - Stocke momentanément les **données** du problème à résoudre et les **pistes envisagées**

Types de mémoires

- Mémoire à **court terme** / Mémoire de **travail**
 - Stocke momentanément les **données** du problème à résoudre et les **pistes envisagées**
- Mémoire à **long terme**
 - Mémoire **déclarative**
 - La grippe se caractérise par ...
 - Mémoire **procédurale**
 - Pour faire un gâteau au chocolat, il faut ...

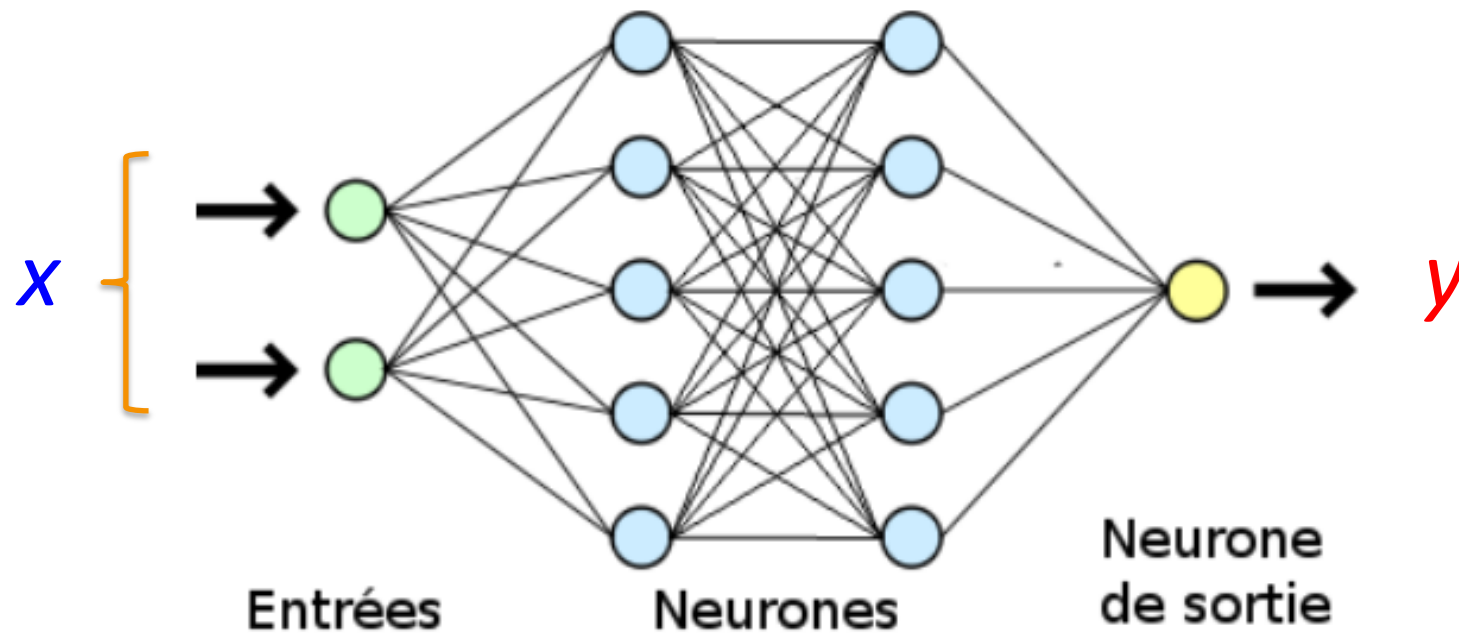
Plan

1. Bref historique de l'Intelligence Artificielle
2. Mémoire, information et calcul
3. Mémoire et apprentissage
4. La mémoire pour comprendre
5. Le big data et les réseaux de neurones profonds
6. Conclusions : l'IA et les grandes questions sur la mémoire

Le Big Data et les réseaux de neurones profonds

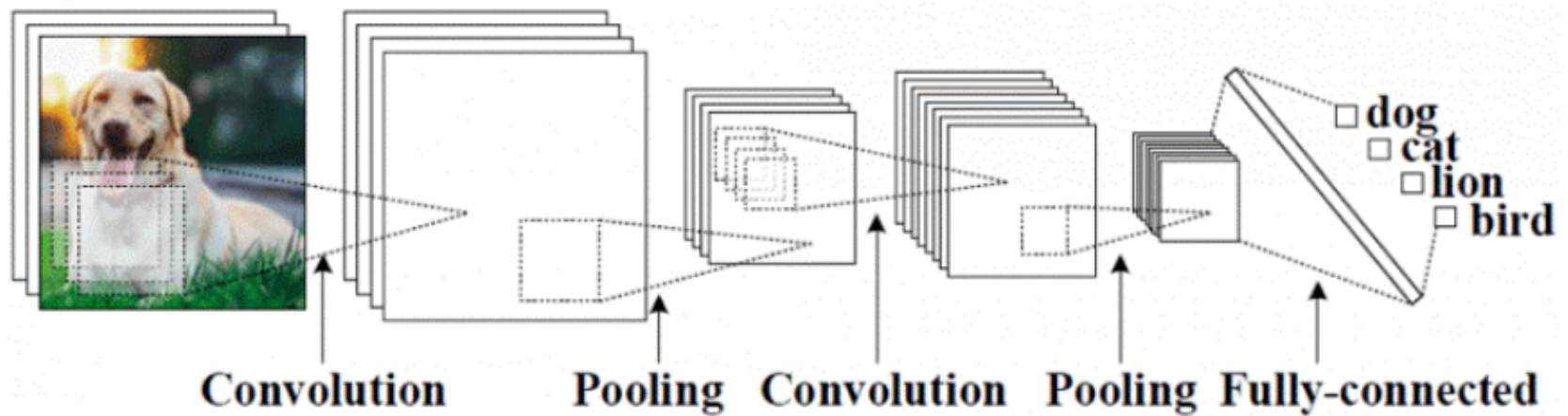
Les réseaux de neurones artificiels

- Réseaux de neurones de type « **Perceptron multi-couche** »



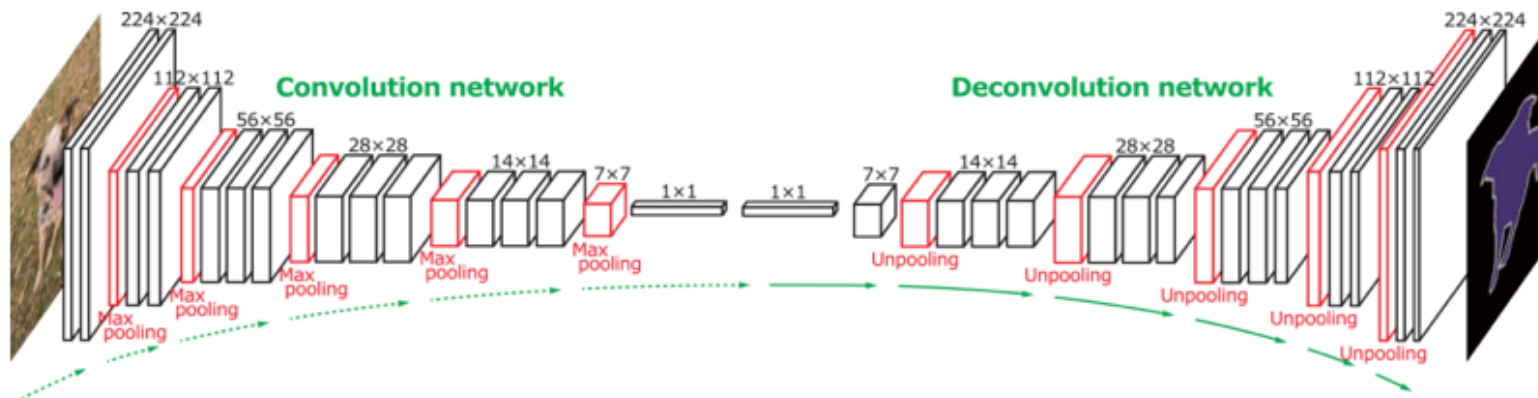
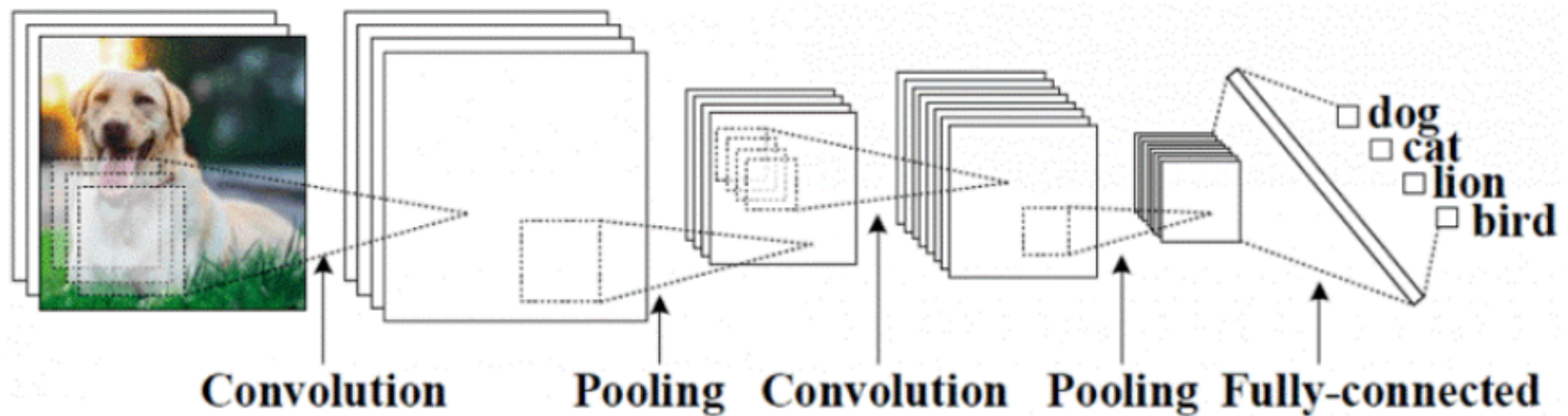
Les réseaux de neurones « profonds »

- Réseaux convolutionnels (Yann LeCun)



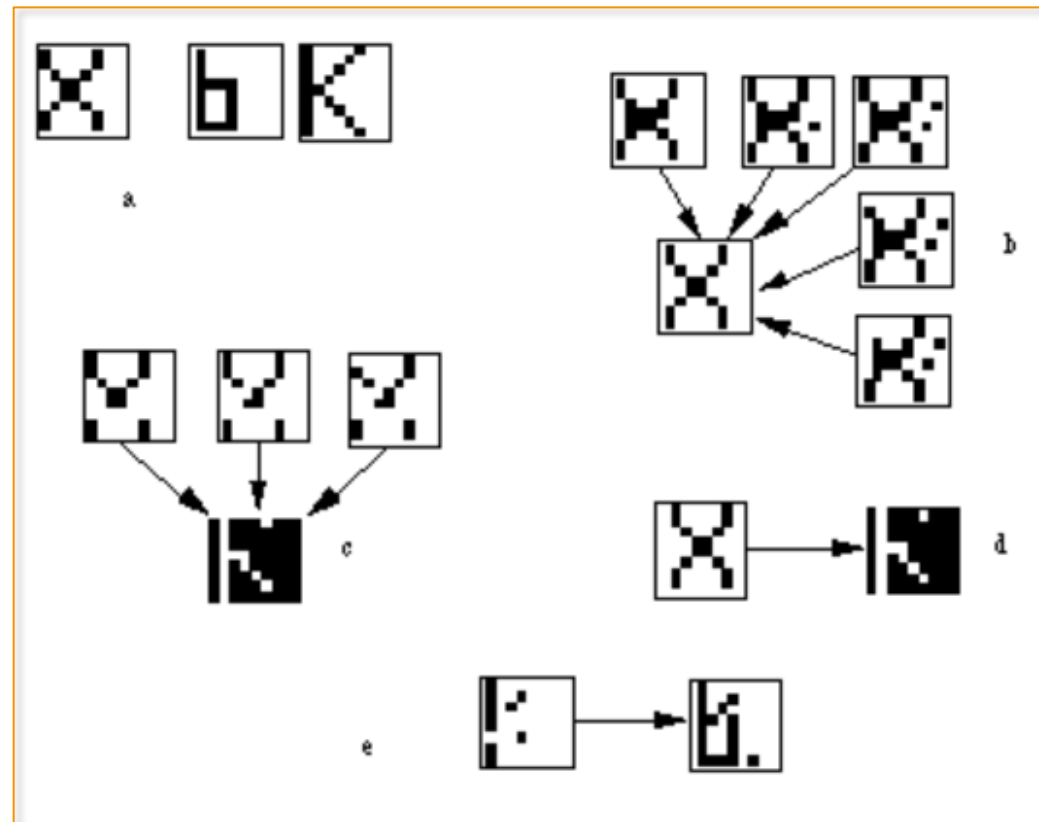
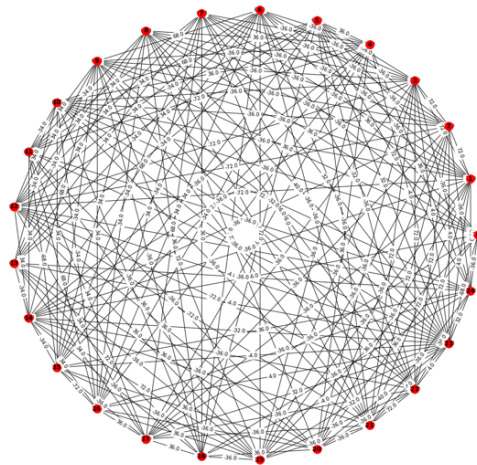
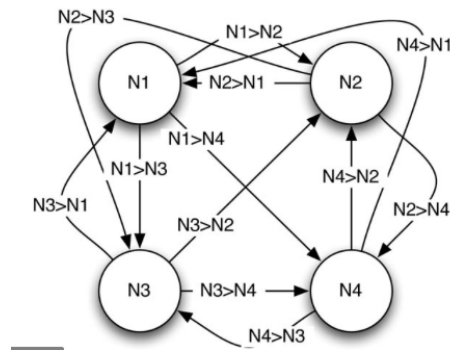
Les réseaux de neurones « profonds »

- Réseaux convolutionnels (Yann LeCun)



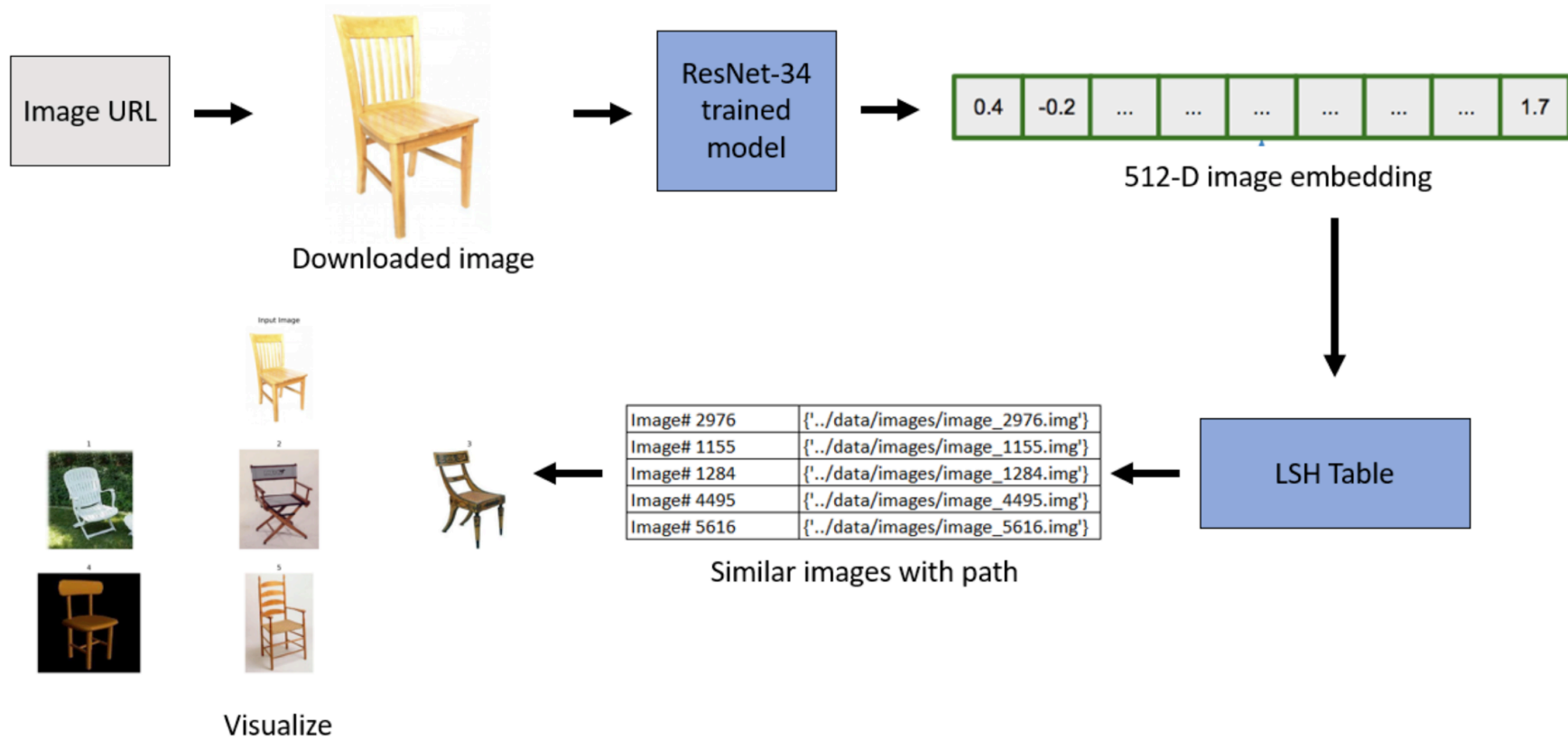
Les mémoires associatives

- Réseaux de neurones de [John Hopfield](#) (1982)

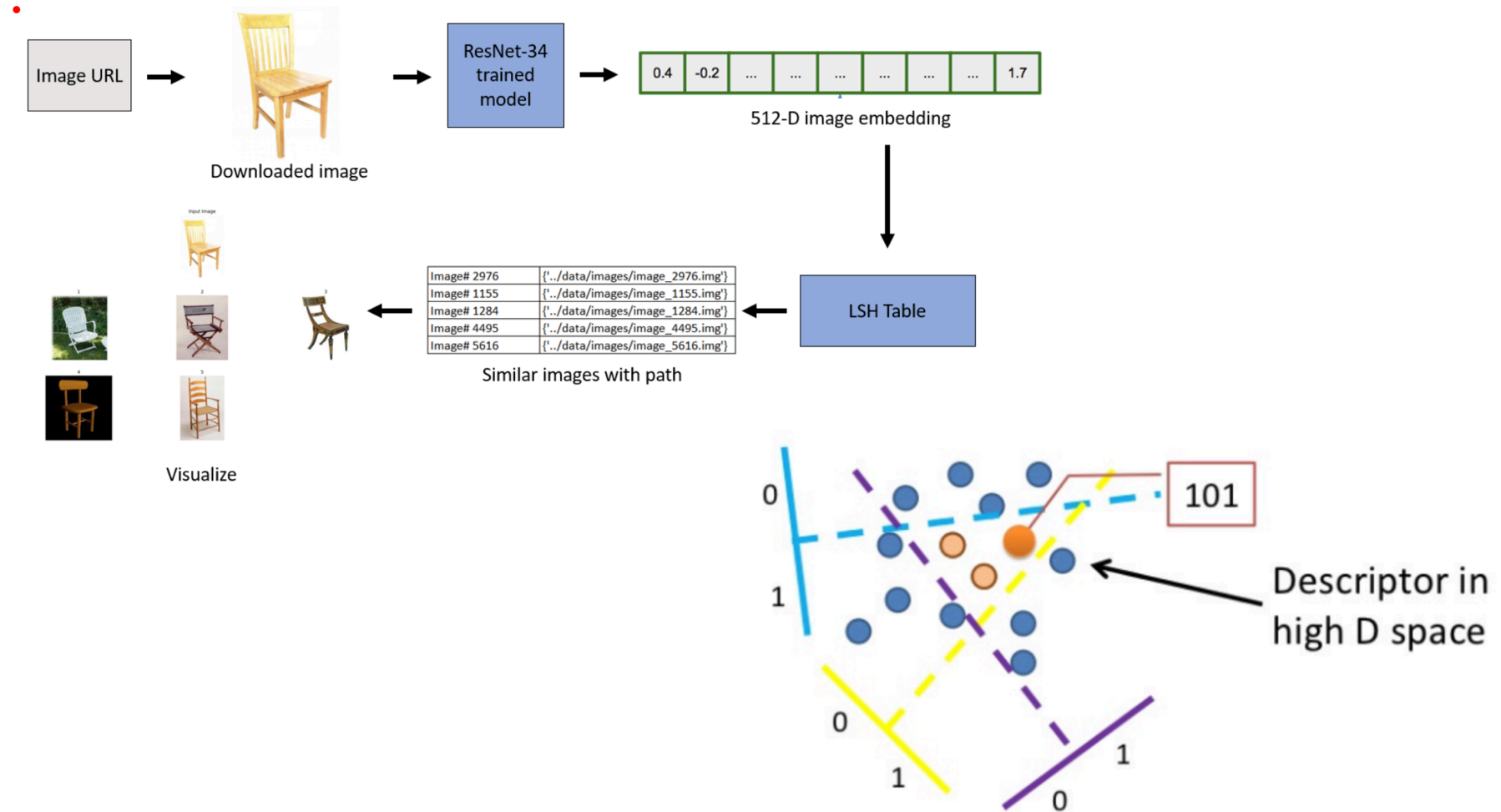


Rappel d'images par Réseau de Neurones Artificiels

- Illustration



Rappel d'images par Réseau de Neurones Artificiels



Un espace sémantique latent

- La question du **rappel** ou de la recherche d'information
 - Requête « *Donne moi des documents portant sur la découverte de connaissances* »
 - On aimerait aussi avoir des documents sur la « *fouille de données* »
 - **Problème** : les mots de la **requête** et des **documents** d'intérêt sont **différents**
- **Idée** : passer par un **espace intermédiaire** dans lequel on **projette les mots** de la requête et des documents et dans lequel **les termes sémantiquement proches sont proches géométriquement**

Un espace sémantique latent

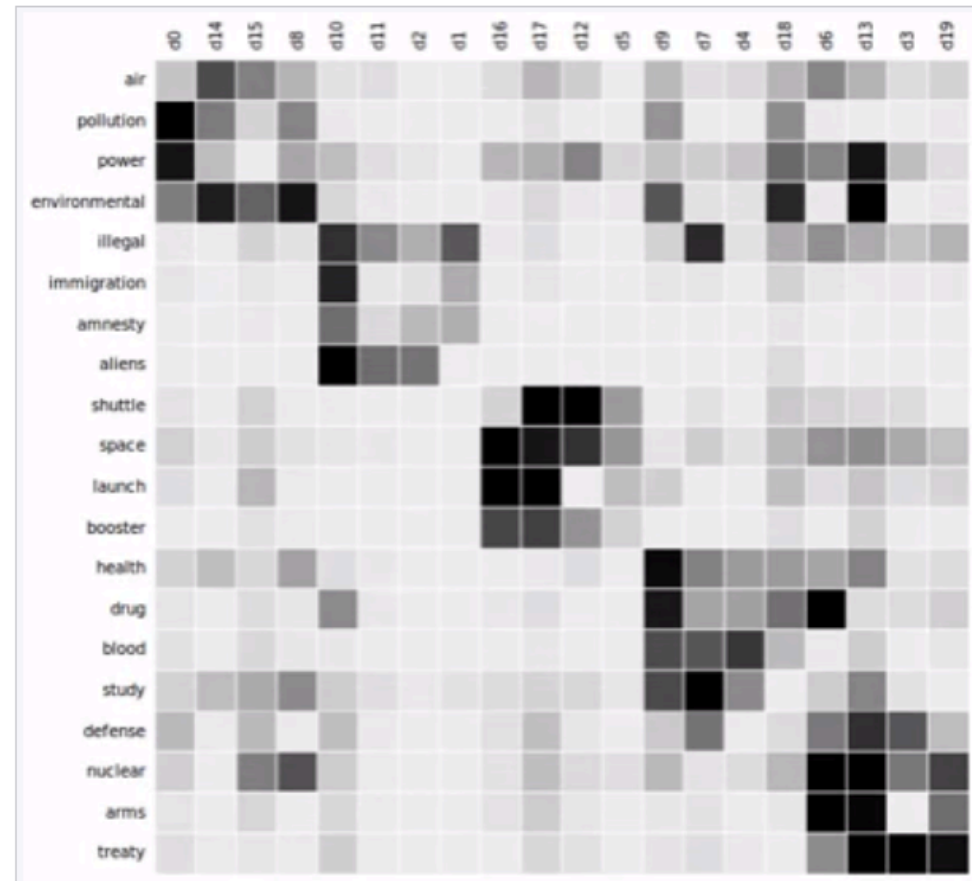
- Hypothèse distributionnelle
 - Les mots utilisés dans des contextes similaires sont proches sémantiquement [Rubenstein & Goodenough, 1965]
 - « to have a *plendid* time in Rome »
 - « to have a *wonderful* time in Rome »
- Espace vectoriel capturant de la sémantique
 - Les **requêtes** et les **documents** sont représentés comme **des vecteurs** dans un espace de termes
(E.g. espace à 3 000 dimensions. Pour chaque mot présent dans le document, on fournit son nombre d'occurrences)

Analyse Sémantique Latente (LSA)

- Les documents sont représentés par des matrices

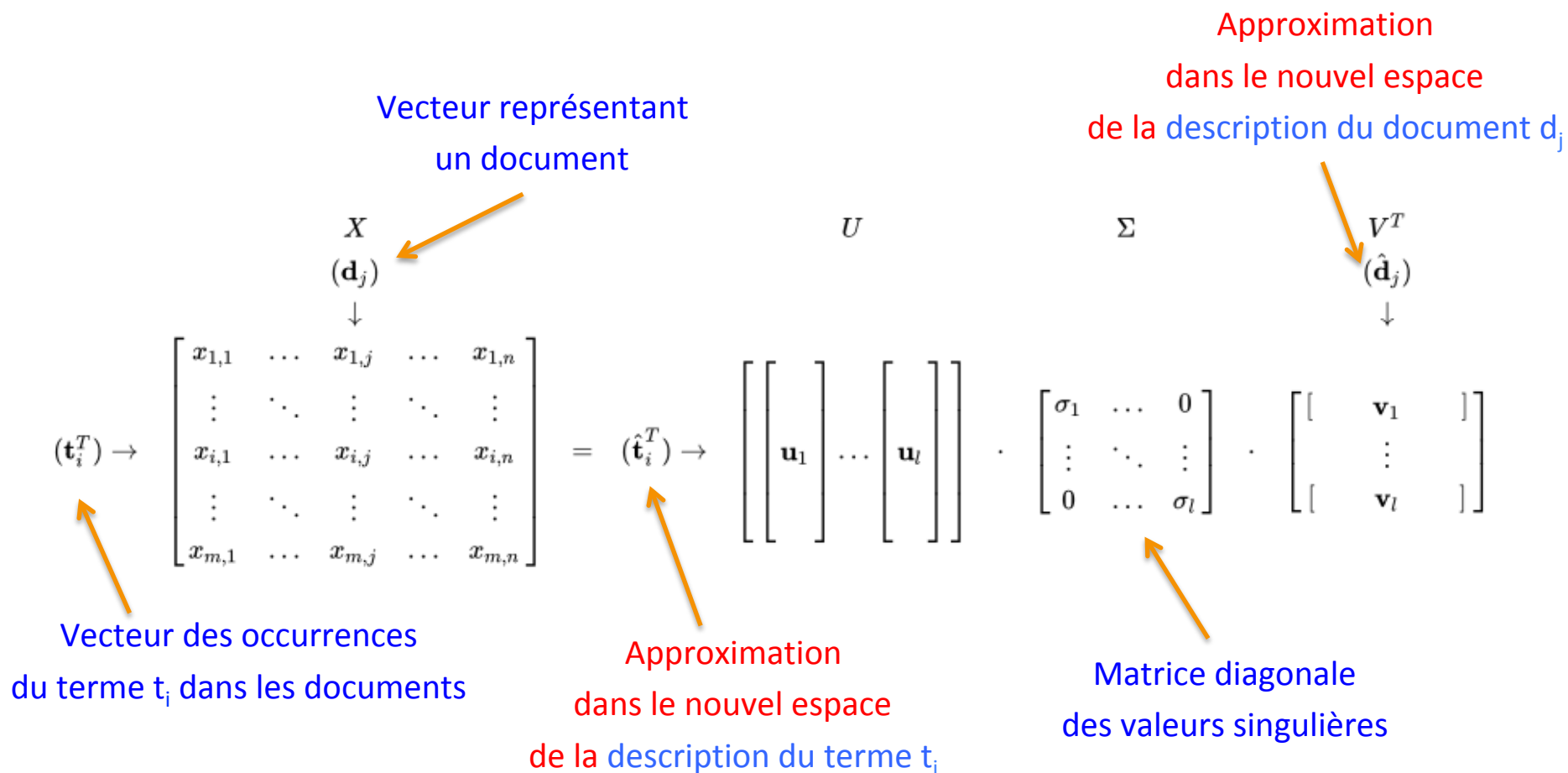
$$\mathbf{t}_i^T \rightarrow \begin{bmatrix} x_{1,1} & \dots & x_{1,j} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \dots & x_{i,j} & \dots & x_{i,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,j} & \dots & x_{m,n} \end{bmatrix}$$

\mathbf{d}_j
↓



Analyse Sémantique Latente (LSA)

- Décomposition en valeurs singulières : compression de l'information

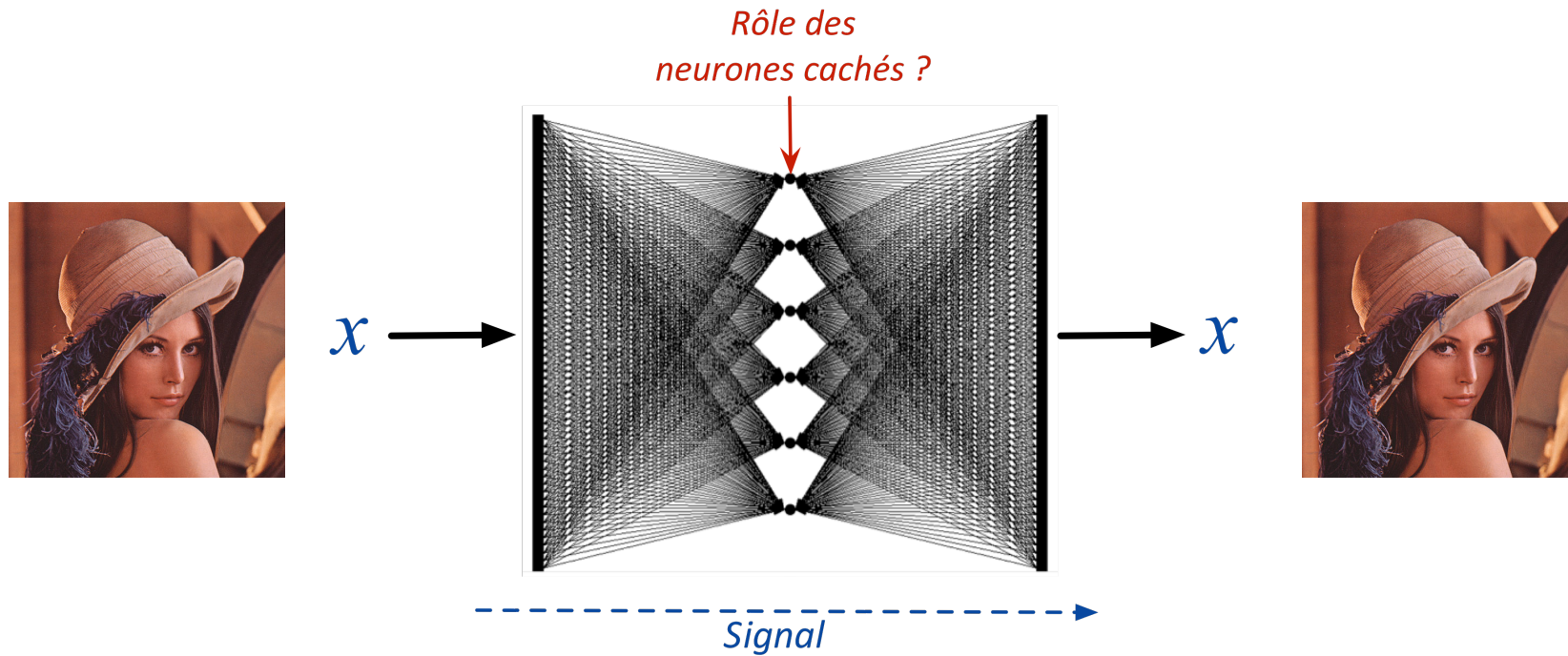


Analyse Sémantique Latente (LSA)

- Permet de :
 1. **Comparer** deux documents en comparant les vecteurs $\sum_k d_j$ et $\sum_k d_i$ dans l'espace latent (e.g. par mesure de similarité cosinus)
 2. **Trouver des relations** entre termes (e.g. synonymie ou antonymie)
 3. Étant donnée une requête (un document), **trouver les documents les plus « proches »**
 4. Identifier des **catégories** (clusters) de documents ou de termes
- Appris à partir de *dizaine de millions de documents*
- Dimension de l'espace latent : 100 à 300

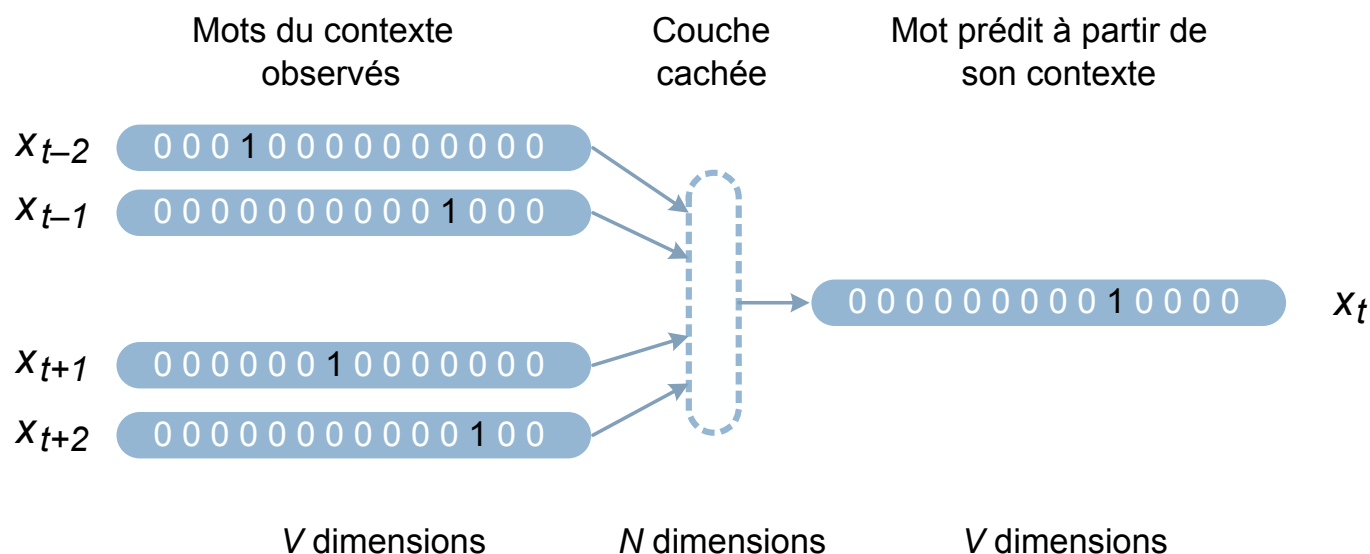
Rôle des couches cachées

- **Quelle représentation** (variables latentes) ?
- Comment choisir l'architecture?



Plongement de mots par réseaux de neurones

- À partir d'un grand corpus de textes, **Word2Vec** associe à **chaque mot** une **représentation dans un espace vectoriel** de dimension 50 à 300 tel que des mots partageant des contextes similaires soient proches.

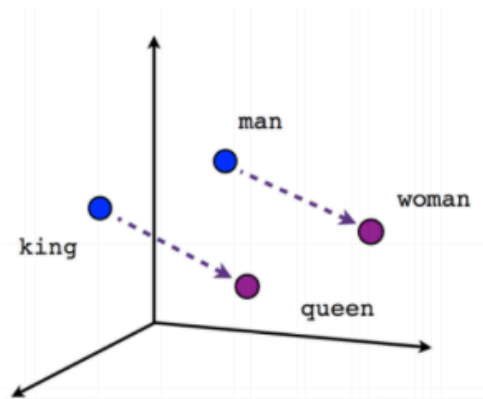


Apprentissage d'une matrice $V \times N$ associant à une entrée mot (dim= V) un vecteur de dimension N

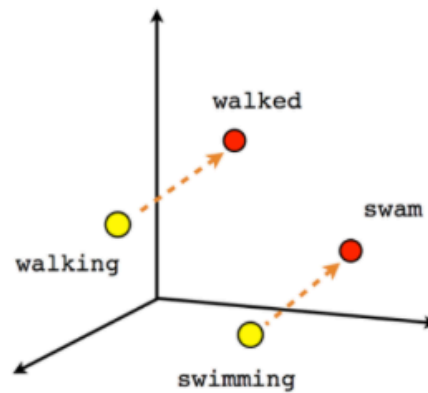
[Mikonov et al. (2013) « Efficient Estimation of Word Representations in Vector Space », *arXiv:1301.3781*]

Plongement de mots par réseaux de neurones

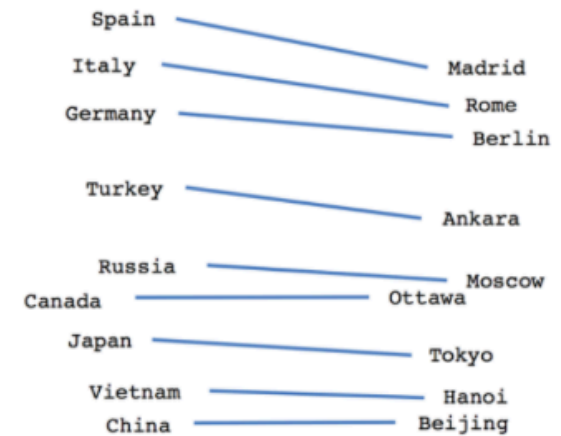
- Illustration



Male-Female



Verb tense



Country-Capital

$$\text{Paris} - \text{France} = x - \text{Pologne}$$

$$x = \text{Paris} - \text{France} + \text{Pologne}$$

Plongement de mots par réseaux de neurones

Applications :

— La **traduction** automatique

- Plongement de mots dans chaque langue
- Plongement de mots bilingues : apprentissage des similarités inter-langages (nécessite des corpus de traductions)

— L'**analyse de sentiments**

- Utilisation du plongement de mot
- Mais avec des techniques spécifiques (e.g. sinon « **bon** » et « **mauvais** » pourraient être proches)

Plan

1. Bref historique de l'Intelligence Artificielle
2. Mémoire, information et calcul
3. Mémoire et apprentissage
4. La mémoire pour comprendre
5. Le big data et les réseaux de neurones profonds
6. Conclusions : l'IA et les grandes questions sur la mémoire

Conclusions

Les grandes questions sur la mémoire

1. Quels sont les grands **types de mémoires** ?

- ✓ **Court** terme (de travail) / **Long** terme
- ✓ **Sémantique** / **Épisodique**
- ✓ **Procédurale** / **Déclarative**
- **Sensorielle** / **Proprioceptive**



Co-développement
IA et sciences cognitives
(70s – 80s)



Notions de
Biais et d'hypothèse

Les grandes questions sur la mémoire

1. Quels sont les grands **types de mémoires** ?

- ✓ **Court** terme (de travail) / **Long** terme
- ✓ **Sémantique** / **Épisodique**
- ✓ **Procédurale** / **Déclarative**
- **Sensorielle** / **Proprioceptive**



Co-développement
IA et sciences cognitives
(70s – 80s)

Autres questions liées

- La **capacité** de la **mémoire de travail** : 7 ± 2
 - Est-ce **optimal** ?
 - Quel **effet** sur les performances / sur la cognition ?
- **Externalisation** de la mémoire
 - **Quel type** de mémoire ?
 - Sémantique ?
 - Épisodique ?
 - **Conséquences** sur la cognition ?



Notions de
Biais et d'hypothèse



Pas de limites en IA
Mais **pas de test systématique**
sur les effets de la taille



Mémoire **non contrôlée**
(e.g. Wikipédia)

Les grandes questions sur la mémoire

2. La **mémorisation** des sensations, évènements, faits, croyances ... ?

- **Mémoriser l'essentiel** et **oublier le superflu**
- Être **fidèle à la réalité** tout en étant **cohérent** avec **notre vision** du monde
- Quel **encodage** ?
 - Localisé ?
 - Distribué ?



Contrôle du **biais**
pour éviter le sur-ajustement



Encodages \neq
pour le **biais** et l'hypothèse



Notion de frames
Épisodique / sémantique



R.N. : encodage **distribué**
optimise la compression de l'information

Les grandes questions sur la mémoire

2. La **mémorisation** des sensations, évènements, faits, croyances ... ?

- **Mémoriser l'essentiel** et **oublier le superflu**
- Être **fidèle à la réalité** tout en étant **cohérent** avec **notre vision** du monde
- Quel **encodage** ?
 - Localisé ?
 - Distribué ?

Autres questions liées

- L'**accès** à la mémoire
 - Comment ?
 - Selon le type de mémoire ?

Contrôle du **biais**
pour éviter le sur-ajustement

Encodages \neq
pour le **biais** et l'hypothèse

Notion de frames
Épisodique / sémantique

R.N. : encodage **distribué**
optimise la compression de l'information

Mémoire **associative** (reconstructive)
Accès par **pattern matching**
Accès par **utilisation d'un index**

Les grandes questions sur la mémoire

3. Règles de la **mémorisation** ?

- Effets de la **répétition**
 - Combien ?
- **Intégration** avec d'autres items mémorisés
- Rôle et causes de **l'oubli**



Big data
vs. Apprentissage à partir
de peu d'exemples



Modèles des 70s – 80s
Architecture blackboard
Scripts, plans, MOPs, ...



Peu de recherches

Les grandes questions sur la mémoire

3. Règles de la **mémorisation** ?

- Effets de la **répétition**
 - Combien ?
- **Intégration** avec d'autres items mémorisés
- Rôle et causes de **l'oubli**

Autres questions liées

- **Accès, consolidation et modification**
 - La mémoire redevient **labile**
 - Effet et optimisation de la **répétition**

Big data

vs. Apprentissage à partir de peu d'exemples

Modèles des 70s – 80s
Architecture blackboard
Scripts, plans, MOPs, ...

Peu de recherches

Recherche sur des modèles de neurosciences
Pas vraiment en IA

Les grandes questions sur la mémoire

4. Mémoire et **attention** ?



Recherche **en cours**

Les grandes questions sur la mémoire

4. Mémoire et **attention** ?



Recherche **en cours**

5. Mémoire et **émotion** ?



?

Les grandes questions sur la mémoire

4. Mémoire et **attention** ?



Recherche **en cours**

5. Mémoire et **émotion** ?



?

6. Mémoire et **sommeil** ?



?

Les grandes questions sur la mémoire

4. Mémoire et **attention** ?



Recherche **en cours**

5. Mémoire et **émotion** ?



?

6. Mémoire et **sommeil** ?



?

7. Nos **défauts** de mémoire :

- des **failles** ?
- des **effets secondaires indésirables** résultant d'optimisation ?



En IA
Soumis à optimisation

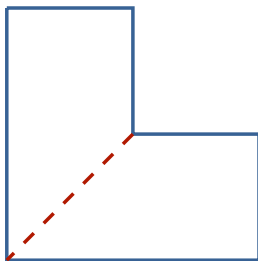
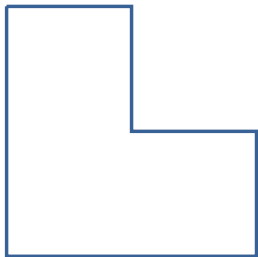
Limites de l'IA

- Les IA **ne savent pas ce qu'elles savent !!**
- Les IA **ne connaissent pas l'ennui** : « *j'ai déjà vu ça 10 fois !* »
- Leur mémoire **ne change pas** lors des accès
- Pas encore de modèle d'apprentissage incorporant **l'attention**
- Modèle d'apprentissage encore limité pour **l'apprentissage en-ligne : pas d'effet de séquence**

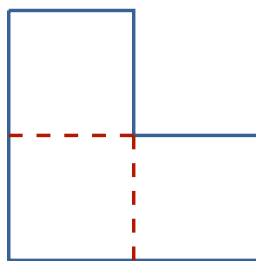
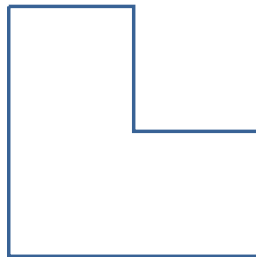
Effets de séquence

Instruction : découper la figure géométrique suivante en n parties **superposables**.

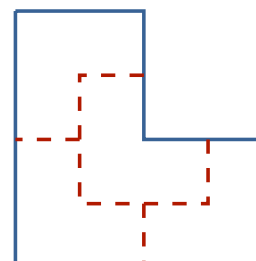
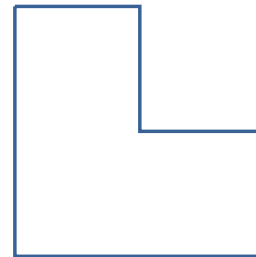
En 2:



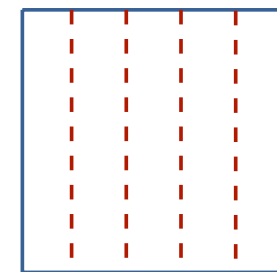
En 3:



En 4:



En 5:



Perspectives

- **Nouveaux scénarios d'apprentissage**
 - Long-life learning et transfert
 - Transfert constructif et transfert négatif
 - Construction de curriculum

*“Another great strength of this book is the way that **ideas build upon one another**. The author has masterfully written a book in which your intuition about **early concepts pave the way for understanding later concepts** even though missing some ideas in the beginning will not cripple you in later chapters.”*

À propos de [Steven Strogatz “Non linear dynamics and chaos”, 2014]

Interrogation

Mémoires et **niveaux de description**

1. **Sciences cognitives** : niveaux des **états mentaux**

- Intentions, buts, plans, croyances

2. Description au **niveau neuronal**

- **Connexions** entre neurones
- Nouveaux **neurones** / neurones détruits
- **Aires** cérébrales



Oscillation de l'IA
entre les deux
niveaux de description