

Technologies de l'esprit. Machines à co-habiter
LLCP - Université Paris 8 | La Générale
Paris, France

Ce que les mathématiques peuvent apporter à la critique des LLMs

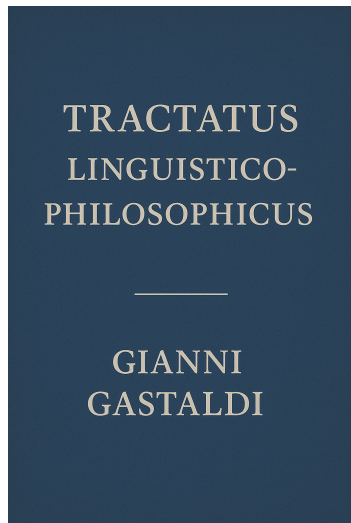
Éléments pour un formalisme critique

Juan Luis Gastaldi

www.giannigastaldi.com

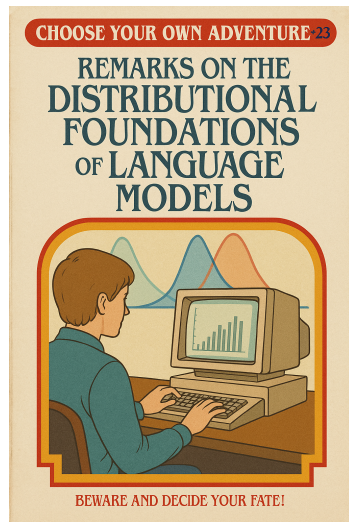
ETH zürich

11 mai, 2026



TRACTATUS
LINGUISTICO-
PHILOSOPHICUS

GIANNI
GASTALDI

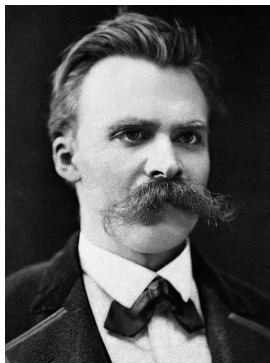


◇ Kirschenbaum (2023):

Bender et al.'s (2021) paper "offers a **disarmingly linear account of how language, communication, intention, and meaning work**, one that would seem to sidestep decades of scholarship around these same issues in literary theory [...] the passage would be **red meat for a graduate critical-theory seminar**."

◇ Underwood (2023):

"The beautiful **irony** of this situation [...] is that a generation of **humanists trained on Foucault** have now rallied around "On the Dangers of Stochastic Parrots" to **oppose a theory of language that their own disciplines invented**, just at the moment when computer scientists are reluctantly beginning to accept it."



“Dans quelque coin reculé de l’univers ruisselant du scintillement d’innombrables systèmes solaires, il y eut un jour un astre sur lequel **des animaux intelligents inventèrent le connaître**. Ce fut la minute la plus **orgueilleuse** et la plus **menteuse** de l’« histoire universelle »...”

De la vérité et du mensonge au sens extra-moral
(Nietzsche, 1873)

La **connaissance** dépend du **langage**

La **connaissance** dépend du **langage**



La relation entre le langage et le monde est **essentiellement arbitraire**

La **connaissance** dépend du **langage**



La relation entre le langage et le monde est **essentiellement arbitraire**



Les régularités dans le langage/la connaissance **ne sont pas naturelles**,
mais **culturelles/sociales/politiques**

La **connaissance** dépend du **langage**



La relation entre le langage et le monde est **essentiellement arbitraire**



Les régularités dans le langage/la connaissance **ne sont pas naturelles**,
mais **culturelles/sociales/politiques**



Nous devons **résister** aux régularités existantes et en **créer** de nouvelles

La connaissance dépend du langage
(Épistémologie)



La relation entre le langage et le monde est essentiellement arbitraire



Les régularités dans le langage/la connaissance ne sont pas naturelles,
mais culturelles/sociales/politiques
(Politique)



Nous devons résister aux régularités existantes et en créer de nouvelles
(Esthétique)

La connaissance dépend du langage
(Épistémologie)



[La relation entre le langage et le monde est essentiellement arbitraire?]



Les régularités dans le langage/la connaissance ne sont pas naturelles,
mais culturelles/sociales/politiques
(Politique)



Nous devons résister aux régularités existantes et en créer de nouvelles
(Esthétique)

1.1



1.21



1.22



1.23



1.2

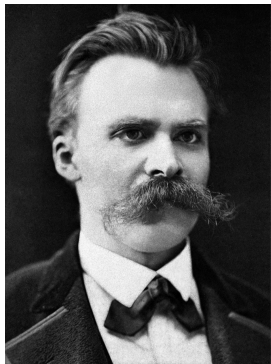


1.2 *Les limites de la critique dans la production du savoir tiennent à la place des savoirs formels

- 1.21 Les savoirs formels sont une construction récente
- 1.22 Ils ont été mobilisés pour fonder des épistémologies dogmatiques
- 1.23 La tradition critique a fait du formalisme une cible

1.31





“...tout peuple possède donc un ciel conceptuel semblable [mathématiquement divisé], et qui le surplombe; [...] On peut bien sur ce point admirer l’homme pour le puissant génie de l’architecture qu’il est: il réussit à ériger une cathédrale conceptuelle infiniment compliquée sur des fondations mouvantes, en quelque sorte sur de l’eau courante. À vrai dire, pour trouver un point d’appui sur de telles fondations, il ne peut s’agir que d’une construction semblable à une toile d’araignée, si fine qu’elle peut suivre le courant du flot qui l’emporte, si résistante qu’elle ne peut être dispersée au gré du vent.

(Nietzsche, 1873)

1.32



1.3



1.3 *Une nouvelle alliance entre pensée critique et formalisme
est nécessaire

1.31 Le formalisme n'est pas un naturalisme

1.32 Un formalisme critique est possible



1

*Nous avons besoin d'un *formalisme critique*

1.1 *La critique de l'IA est à court de carburant

1.2 *Les limites de la critique dans la production du savoir
tiennent à la place des savoirs formels

1.3 *Une nouvelle alliance entre pensée critique et formalisme est
nécessaire

2.11



2.12



$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$0: \lambda f. \lambda x. x$$

$$1: \lambda f. \lambda x. f x$$

$$2: \lambda f. \lambda x. f (f x)$$

$$3: \lambda f. \lambda x. f (f (f x))$$

$$4: \lambda f. \lambda x. f (f (f (f x)))$$

$$5: \lambda f. \lambda x. f (f (f (f (f x))))$$

...

$$n: \lambda f. \lambda x. \underbrace{f(\dots (f x) \dots)}_{n \text{ times}}$$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$0: \lambda f. \lambda x. x$$

$$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$$

$$1: \lambda f. \lambda x. f x$$

$$2: \lambda f. \lambda x. f (f x)$$

$$3: \lambda f. \lambda x. f (f (f x))$$

$$4: \lambda f. \lambda x. f (f (f (f x)))$$

$$5: \lambda f. \lambda x. f (f (f (f (f x))))$$

...

$$n: \lambda f. \lambda x. \underbrace{f(\dots (f x) \dots)}_{n \text{ times}}$$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

0: $\lambda f. \lambda x. x$	$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$
1: $\lambda f. \lambda x. f x$	$\Downarrow$
2: $\lambda f. \lambda x. f (f x)$	$\Downarrow$
3: $\lambda f. \lambda x. f (f (f x))$	$\Downarrow$
4: $\lambda f. \lambda x. f (f (f (f x)))$	$\Downarrow$
5: $\lambda f. \lambda x. f (f (f (f (f x))))$	$\Downarrow$
...	$\Downarrow$
$n: \lambda f. \lambda x. \underbrace{f(\dots(f x)\dots)}_{n \text{ times}}$	$\Downarrow$
	$\lambda f. \lambda x. f (f (f (f (f x))))$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$P' := \lambda r. \lambda s. \lambda f. \lambda x. f(f(f(f(fx))))$$

$$0: \lambda f. \lambda x. x$$

$$\lambda r. \lambda s. \lambda f. \lambda x. f(f(f(f(fx)))) (\lambda f. \lambda x. f(fx)) (\lambda f. \lambda x. f(f(fx)))$$

$$1: \lambda f. \lambda x. f x$$

$$\Downarrow$$

$$2: \lambda f. \lambda x. f(fx)$$

$$\Downarrow$$

$$3: \lambda f. \lambda x. f(f(fx))$$

$$\Downarrow$$

$$4: \lambda f. \lambda x. f(f(f(fx)))$$

$$\Downarrow$$

$$5: \lambda f. \lambda x. f(f(f(f(fx))))$$

$$\Downarrow$$

$$\dots$$

$$\Downarrow$$

$$n: \lambda f. \lambda x. \underbrace{f(\dots(fx)\dots)}_{n \text{ times}}$$

$$\Downarrow$$

$$\lambda f. \lambda x. f(f(f(f(fx))))$$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

0:	$\lambda f. \lambda x. x$	$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$
1:	$\lambda f. \lambda x. f x$	$\Downarrow$
2:	$\lambda f. \lambda x. f (f x)$	$\Downarrow$
3:	$\lambda f. \lambda x. f (f (f x))$	$\Downarrow$
4:	$\lambda f. \lambda x. f (f (f (f x)))$	$\Downarrow$
5:	$\lambda f. \lambda x. f (f (f (f (f x))))$	$\Downarrow$
...		$\Downarrow$
n:	$\lambda f. \lambda x. \underbrace{f (\dots (f x) \dots)}_{n \text{ times}}$	$\Downarrow$
		$\lambda f. \lambda x. f (f (f (f (f x))))$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

0: $\lambda f. \lambda x. x$	$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda f. \lambda x. f (f x)) (\lambda f. \lambda x. f (f (f x)))$
1: $\lambda f. \lambda x. f x$	$\lambda m. \lambda n. \lambda f. \lambda x. m f (n f x) (\lambda g. \lambda y. g (g y)) (\lambda h. \lambda z. h (h (h z)))$
2: $\lambda f. \lambda x. f (f x)$	$\lambda n. \lambda f. \lambda x. (\lambda g. \lambda y. g (g y)) f (n f x) (\lambda h. \lambda z. h (h (h z)))$
3: $\lambda f. \lambda x. f (f (f x))$	$\lambda n. \lambda f. \lambda x. (\lambda g. \lambda y. g (g y)) f (n f x) (\lambda h. \lambda z. h (h (h z)))$
4: $\lambda f. \lambda x. f (f (f (f x)))$	$\lambda f. \lambda x. (\lambda g. \lambda y. g (g y)) f ((\lambda h. \lambda z. h (h (h z))) f x)$
5: $\lambda f. \lambda x. f (f (f (f (f x))))$	$\lambda f. \lambda x. (\lambda y. f (f y)) ((\lambda h. \lambda z. h (h (h z))) f x)$
...	$\lambda f. \lambda x. (\lambda y. f (f y)) ((\lambda z. f (f (f z))) x)$
n : $\lambda f. \lambda x. \underbrace{f (\dots (f x) \dots)}_{n \text{ times}}$	$\lambda f. \lambda x. (\lambda y. f (f y)) (f (f (f x)))$
	$\lambda f. \lambda x. f (f (f (f (f x))))$

$$P := \lambda m. \lambda n. \lambda f. \lambda x. m f (n f x)$$

$$P'' := \lambda R \acute{o} f \ddot{A} \ddot{O} \hat{e} \tilde{N} 5 \ddot{E} | \grave{A} x \tilde{n} = \infty \grave{u} \grave{y} m W f 286 \ddot{e} y' S \ddot{O} \acute{u} > v \& \grave{i} \hat{A} - 2 \acute{o} \acute{E} 7 \ddot{o} \zeta \infty \{ \tilde{a} > 2 f l B^{\circ} \mu G \# \grave{A} 9 \zeta U$$

$$\infty b t Y B \hat{o} \pounds \ddot{U} \ddot{e} \%_{00} 3; 5 \grave{a} [l - \acute{e} u \hat{o} \ddot{U}^{\circ} 7 - \ddot{U}. \lambda: ^4 m \acute{O} \emptyset \pounds ' \acute{e} - + \grave{I} s \ddot{O}, \$ + g \ddot{i}, B^{TM} \div o - \# \grave{i} \ddot{Y} \hat{e} \hat{U} v$$

$$- g \acute{O} \ddot{y} / \ddot{e} i i j O \ddot{f} C e f i \bullet J 1 \ll \epsilon \emptyset, \grave{I} h \hat{e} t \ddot{f} \acute{a} e Y \$ ^6 F \ddot{i} W \gg R \grave{U} K g c \ddot{r}. \lambda \ddot{f} d^- \dots D 2 \div \zeta \grave{o}^{\circ} x \hat{e} \ddot{E} y. \acute{O} \ddot{r} c b$$

$$B \acute{e} \pounds N \grave{E} 1 \hat{E} \ddot{f} / \hat{U} 9 \tilde{N} \mu - / J Y \zeta \ddot{o} \ddot{E} 9 \ddot{y} \grave{A} \grave{E} \lambda \acute{A} \acute{I} \grave{A}^{\circ} \ddot{O} \zeta, \gg f q \infty \pm \hat{i} \sim B 5 \grave{I} > O \sim g^{TM} \text{“} 6 \Omega e \text{“} \acute{a} \ddot{e} C / \tilde{a} \dots \ddot{O}$$

$$\cdot f \acute{O} \grave{A}] \tilde{N} \acute{a} y \hat{E} N^{\circ} \hat{E} \succ \ddot{r}. \lambda \pounds \grave{a} \epsilon f U \grave{o} f E \ddot{U} \acute{I} m \# \ddot{r}, 4 \backslash r \sqrt{-} \div \hat{I} p \grave{o} \gg y^* v t \ddot{A} J \ddot{A} F 1 \hat{u} \acute{A} \acute{o} z \ll \tilde{n} M \text{”} D j C E$$

$$B \ddot{E} \grave{e} \acute{I} T _ \hat{E} a \%_{00} \acute{A} \zeta \Omega @ \backslash \emptyset^{\sim}] \hat{I} \succ h \ddot{f} : ^4 m \acute{O} \emptyset \pounds ' \acute{e} - + \grave{I} s \ddot{O}, \$ + g \ddot{i}, B^{TM} \div o - \# \grave{i} \ddot{Y} \hat{e} \hat{U} v - g \acute{O} \ddot{y}$$

$$/ \ddot{e} i i j O \ddot{f} C e f i \bullet J 1 \ll \epsilon \emptyset, \grave{I} h \hat{e} t \ddot{f} \acute{a} e Y \$ ^6 F \ddot{i} W \gg R \grave{U} K g c \ddot{r} \acute{A} \acute{I} \grave{A}^{\circ} \ddot{O} \zeta, \gg f q \infty \pm \hat{i} \sim B 5 \grave{I} > O \sim g^{TM} \text{“} 6$$

$$\Omega e \text{“} \acute{a} \ddot{e} C / \tilde{a} \dots \ddot{O} \cdot f \acute{O} \grave{A}] \tilde{N} \acute{a} y \hat{E} N^{\circ} \hat{E} \succ \ddot{r} (\ddot{f} d^- \dots D 2 \div \zeta \grave{o}^{\circ} x \hat{e} \ddot{E} y. \acute{O} \ddot{r} c b B \acute{e} \pounds N \grave{E} 1 \hat{E} \ddot{f} / \hat{U} 9 \tilde{N} \mu - /$$

$$J Y \zeta \ddot{o} \ddot{E} 9 \ddot{y} \grave{A} \grave{E} \acute{A} \acute{I} \grave{A}^{\circ} \ddot{O} \zeta, \gg f q \infty \pm \hat{i} \sim B 5 \grave{I} > O \sim g^{TM} \text{“} 6 \Omega e \text{“} \acute{a} \ddot{e} C / \tilde{a} \dots \ddot{O} \cdot f \acute{O} \grave{A}] \tilde{N} \acute{a} y \hat{E} N^{\circ} \hat{E} \succ \ddot{r} \pounds$$

$$\grave{a} \epsilon f U \grave{o} f E \ddot{U} \acute{I} m \# \ddot{r}, 4 \backslash r \sqrt{-} \div \hat{I} p \grave{o} \gg y^* v t \ddot{A} J \ddot{A} F 1 \hat{u} \acute{A} \acute{o} z \ll \tilde{n} M \text{”} D j C E B \ddot{E} \grave{e} \acute{I} T _ \hat{E} a \%_{00} \acute{A} \zeta \Omega @ \backslash$$

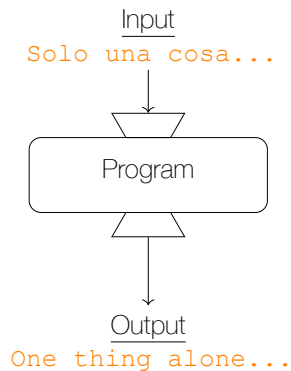
$$\emptyset^{\sim}] \hat{I} \succ h \ddot{f}) (\grave{E} \hat{I} \hat{U} e i 4 W \mu \acute{I} \} w \ddot{r}, \$ \Omega \text{“} K 5 \grave{e} \hat{A} \P \%_3 [m \succ B \acute{A} f i f \ddot{O}; \acute{o} J \zeta C \ddot{E} \hat{i} \ddot{O} \ddot{Y} \emptyset c B, \tilde{n} \$ \hat{A} \hat{a} \} \acute{O} \hat{A} \emptyset 3 j$$

$$\succ ? \ddot{o} - o C E @ f l 8 \sim R C \pounds o \sim ^* \& < \ddot{Y} - o 1 2 \hat{A} \%_{00} \grave{a} \acute{O} \ddot{U} \# \grave{i} \ddot{r}, \acute{u} \ddot{r} \ll \hat{o} \ddot{r}, \infty \hat{I} \hat{a} \hat{a} \text{“} \emptyset \hat{A} d | \sim \tilde{N} ' \hat{E} y \emptyset; ^{\circ} W$$

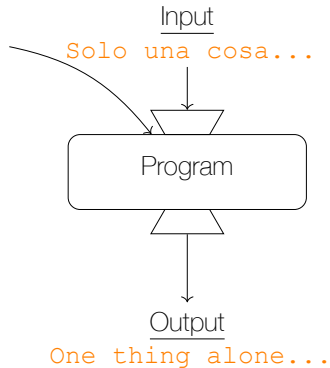
$$\succ \text{”} w \grave{o} [] \backslash \gg \ddot{O} \ddot{E} \acute{u} w ' 6 < \grave{u} \sim = \tilde{a} \ddot{O} _ \hat{I} ' D z ? 2 \pm | \acute{e} ' 3 \hat{A} / r x \mu \infty \mu \$ \hat{A} \ddot{e} \hat{A} * l f \sim \hat{i} \hat{u} ' + \acute{I} V \grave{i} y^a G \acute{a} \pounds \grave{a} g \hat{o} / , u \tilde{N})$$

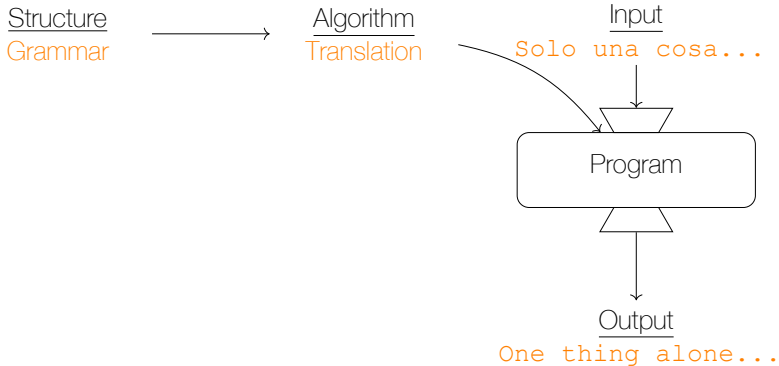
2.13

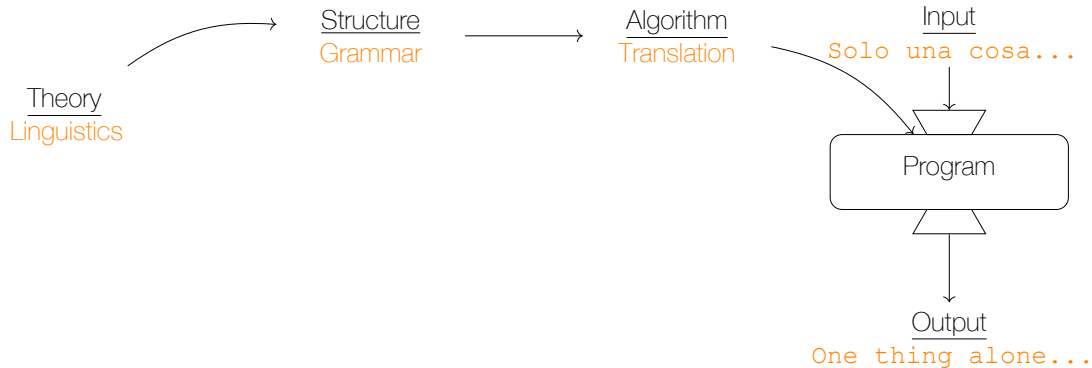


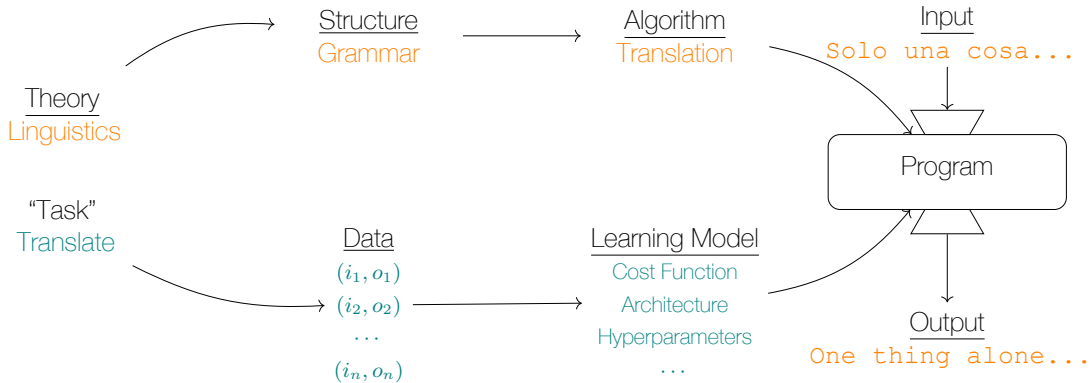


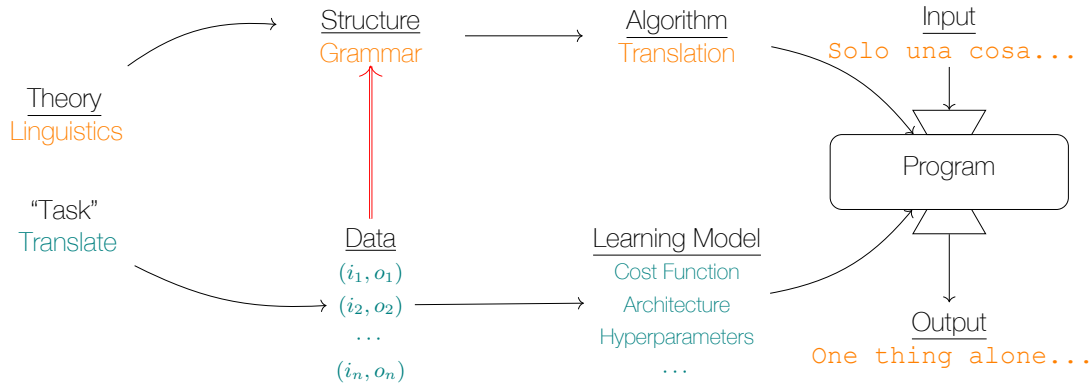
Algorithm
Translation

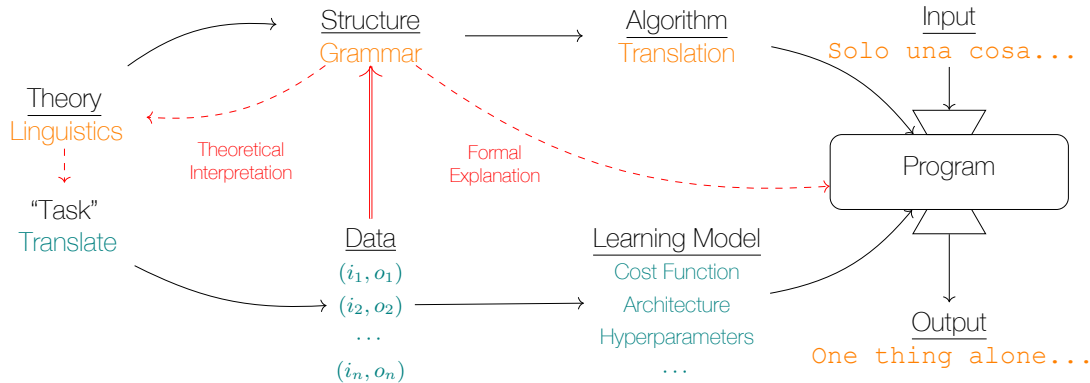












2.14



2.1



2.1 *L'étude empirique des LLMs n'a pas de fondement épistémologique

- 2.11 L'informatique traverse un tournant empirique autour des LLMs
- 2.12 Mais les LLMs ne sont que des fonctions calculables
- 2.13 Il n'existe pas de moyen empirique de savoir ce qu'une fonction calculable fait
- 2.14 *La seule question épistémologique valide est: de quoi cette fonction est-elle l'implémentation?

2.21



2.22



2.23



2.2



2.2

*Les LLMs n'ont aucune portée cognitive a priori

- 2.21 La portée cognitive des modèles de langage computationnels n'est pas inconditionnelle
- 2.22 La condition épistémologique assurant un tel lien ne s'applique pas aux LLMs
- 2.23 L'absence de portée cognitive n'empêche pas les LLMs d'être des modèles du langage

2



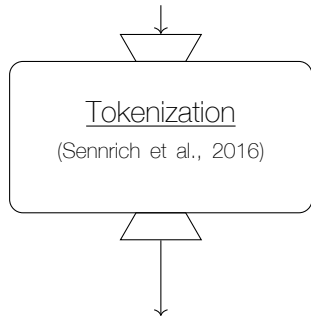
*Un formalisme critique habilite une *critique épistémologique* de l'IA

- 2.1 *L'étude empirique des LLMs n'a pas de fondement épistémologique
- 2.2 *Les LLMs n'ont aucune portée cognitive a priori

3.1

Dans la boîte noire

Epistemology of Machine Learning
Distributional Language Models



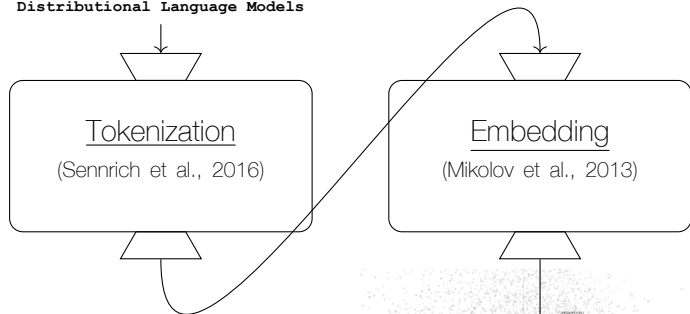
Epistemology of Machine Learning
Distributional Language Models

(<https://tiktokenizer.vercel.app>)

3.1

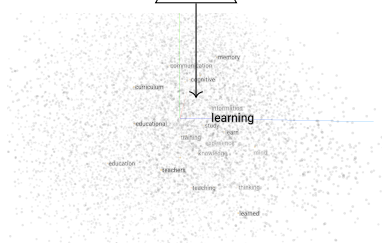
Dans la boîte noire

Epistemology of Machine Learning
Distributional Language Models



Epistemology of Machine Learning
Distributional Language Models

(<https://tiktokenizer.vercel.app>)

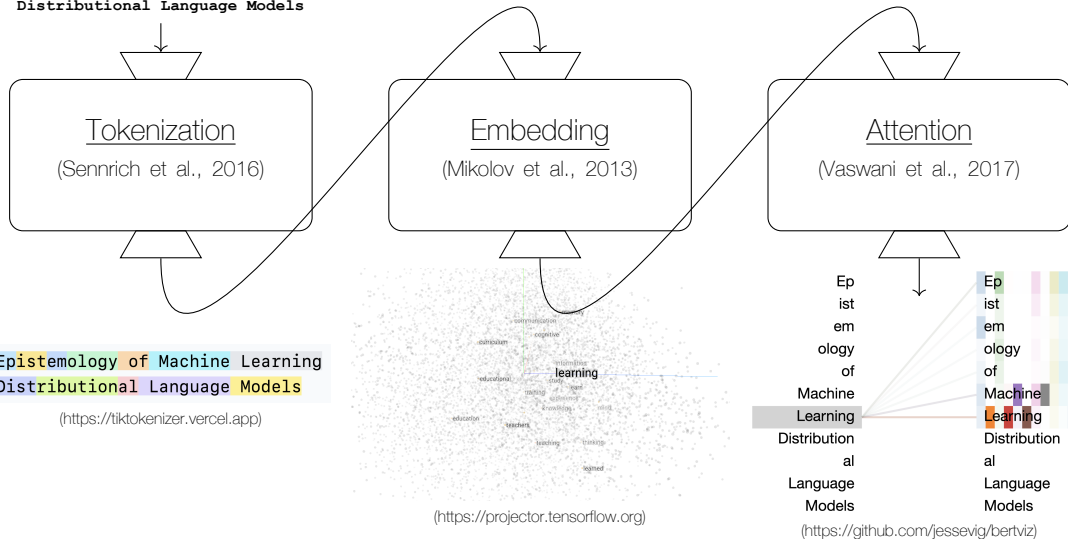


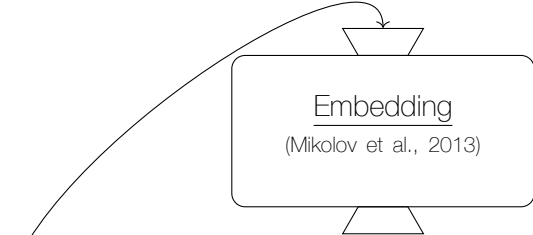
(<https://projector.tensorflow.org>)

3.1

Dans la boîte noire

Epistemology of Machine Learning
Distributional Language Models

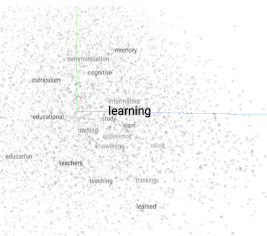




Embedding
(Mikolov et al., 2013)

Epistemology of Machine Learning
Distributional Language Models

(<https://tiktokenizer.vercel.app>)



(<https://projector.tensorflow.org>)

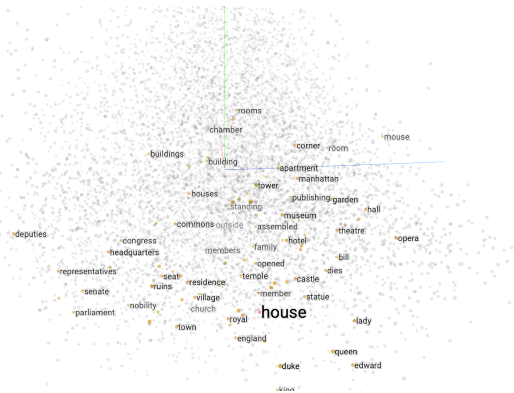
3.1

Qu'est-ce qu'un embedding?

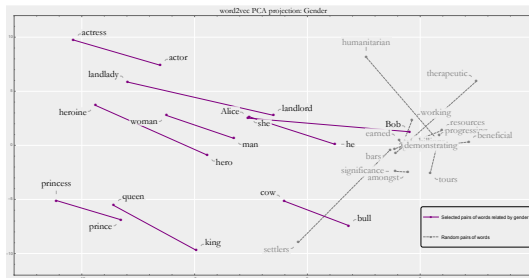


3.1

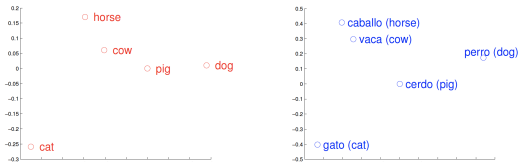
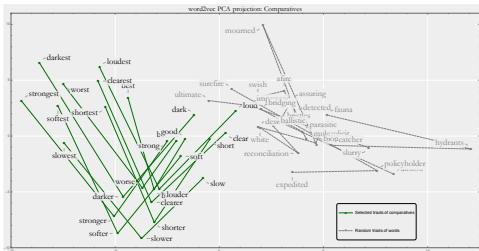
Embeddings: Similarity and Analogy



(<https://projector.tensorflow.org>)

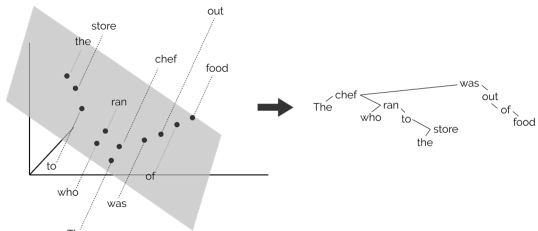
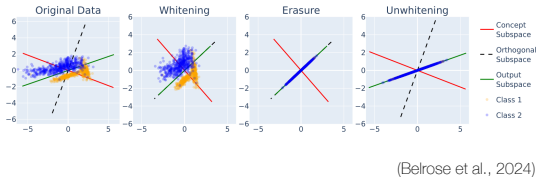


3.1



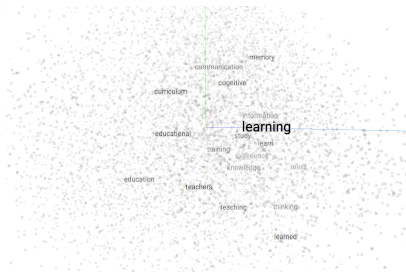
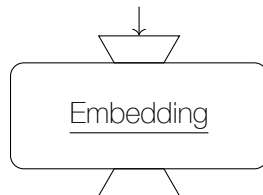
(Mikolov, Sutskever, Chen, Corrado, Dean, et al., 2013)

Embeddings: Other Applications



(<https://nlp.stanford.edu/~johnhew/structural-probe.html>)

Epistemology of Machine Learning
Distributional Language Models



3.1

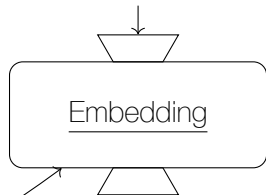
Structure



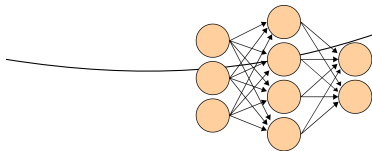
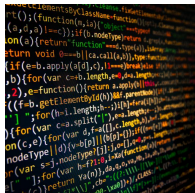
L'espace latent

Epistemology of Machine Learning

Distributional Language Models



Data



3.1

Structure

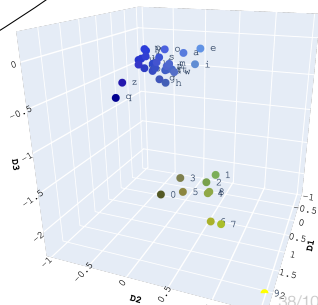
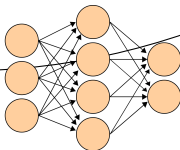
?

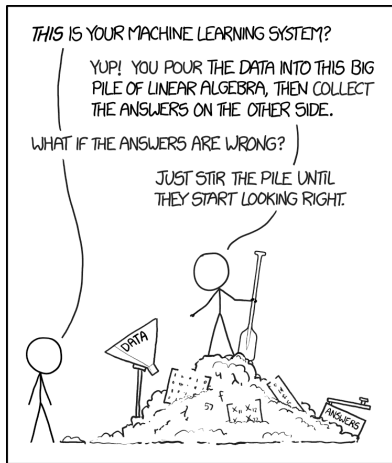
L'espace latent

{-, /, 0, 1, 2, ..., 8, 9, =,
a, b, c, ..., w, x, y, z, é}

Embedding

Data

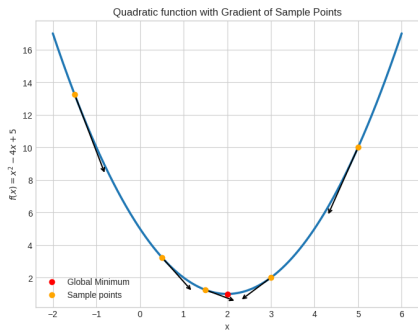




Credit: xkcd.com

3.1

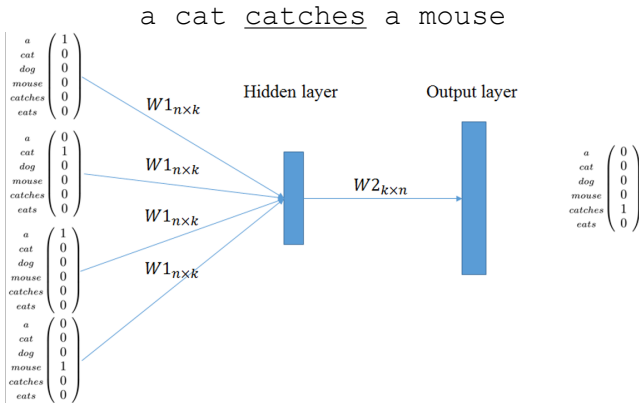
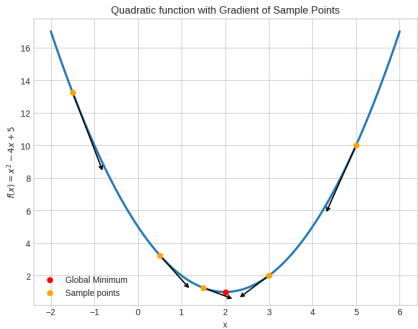
word2vec



Credit: CodeSignal

3.1

word2vec



Credit: CodeSignal

Credit: Ferrone et al., 2017

3.1



$$\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)]$$

Where:

$\vec{w}$ = vector representation for word w

$\vec{c}$ = vector representation for context c

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

k = number of “negative” (arbitrary) samples

c_N = arbitrary context sampled from P_D

$P_D(c)$ = empirical unigram distribution of c in the data D , i.e. $\frac{\#(c)}{|D|}$

3.2

word2vec expliqué (Levy and Goldberg, 2014)

$$\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)]$$

$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w, c) (\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)])$$

$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w, c) (\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)])$$

$$\frac{\partial \ell}{\partial (\vec{w} \cdot \vec{c})} = 0$$

$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w, c) (\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)])$$

$$\frac{\partial \ell}{\partial (\vec{w} \cdot \vec{c})} = 0 \quad \text{when} \quad \vec{w} \cdot \vec{c} = \log \left(\frac{\#(w, c) \cdot |D|}{\#(w) \cdot \#(c)} \right) - \log k$$

$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w, c) (\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)])$$

$$\begin{aligned} \frac{\partial \ell}{\partial (\vec{w} \cdot \vec{c})} = 0 \quad \text{when} \quad \vec{w} \cdot \vec{c} &= \log \left(\frac{\#(w, c) \cdot |D|}{\#(w) \cdot \#(c)} \right) - \log k \\ &= \textcolor{red}{PMI}(w, c) - \log k \end{aligned}$$

$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w, c) (\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)])$$

$$\begin{aligned} \frac{\partial \ell}{\partial (\vec{w} \cdot \vec{c})} = 0 \quad \text{when} \quad \vec{w} \cdot \vec{c} &= \log \left(\frac{\#(w, c) \cdot |D|}{\#(w) \cdot \#(c)} \right) - \log k \\ &= \textcolor{red}{PMI}(w, c) - \log k \end{aligned}$$

Additional constraint: $\vec{w}$ and $\vec{c}$ should be **low dimensional**

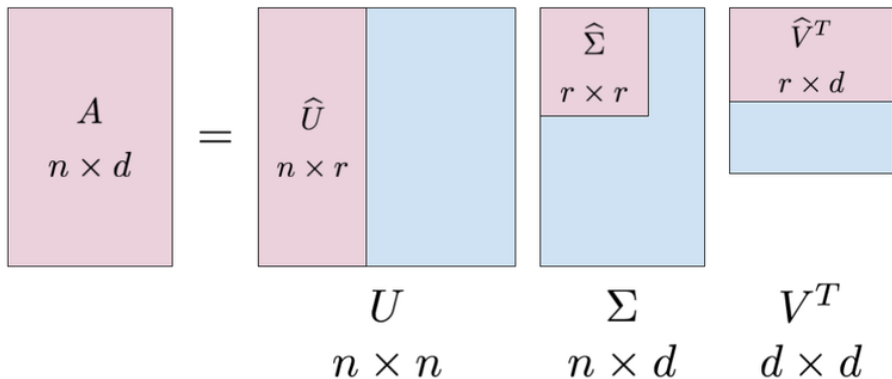
$$\ell = \sum_{w \in V_w} \sum_{c \in V_c} \#(w, c) (\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} [\log \sigma(-\vec{w} \cdot \vec{c}_N)])$$

$$\begin{aligned} \frac{\partial \ell}{\partial (\vec{w} \cdot \vec{c})} = 0 \quad \text{when} \quad \vec{w} \cdot \vec{c} &= \log \left(\frac{\#(w, c) \cdot |D|}{\#(w) \cdot \#(c)} \right) - \log k \\ &= \textcolor{red}{PMI}(w, c) - \log k \end{aligned}$$

Additional constraint: $\vec{w}$ and $\vec{c}$ should be **low dimensional**

There exists an **exact solution** ...

$$M = U\Sigma V^*$$



3.2

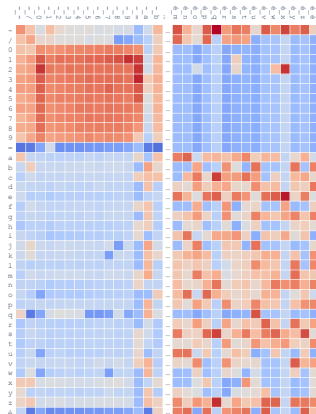
Exemple: Caractères dans Wikipedia

$W = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$

$C = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$

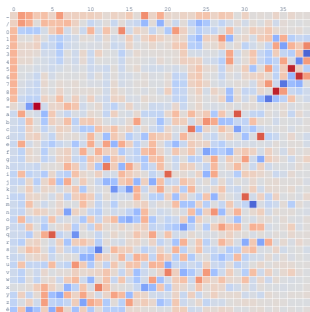
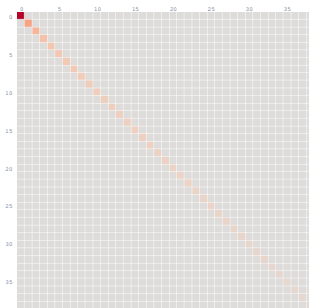
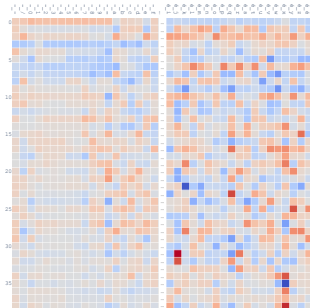
$$M_{wc} = \text{pmi}(w, c)$$

$$= \log \frac{p(w, c)}{p(w)p(c)}$$



3.2

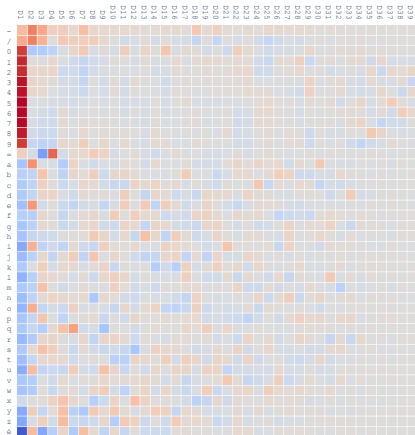
SVD sur une matrice PMI de caractères dans Wikipedia

 U  Σ  V^T 

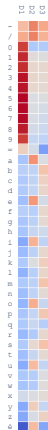
3.2

Tronquer

$$U \times \Sigma$$

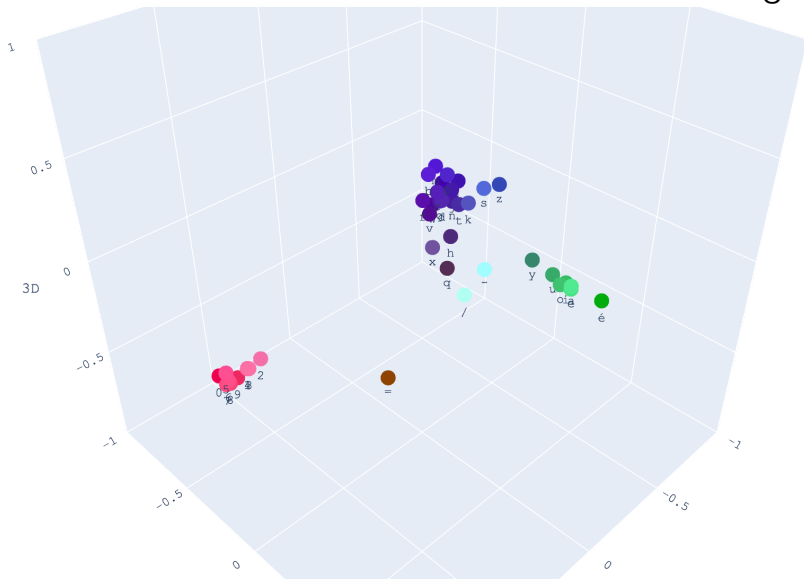
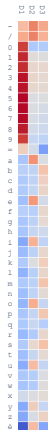


$$\hat{U} \times \hat{\Sigma}$$



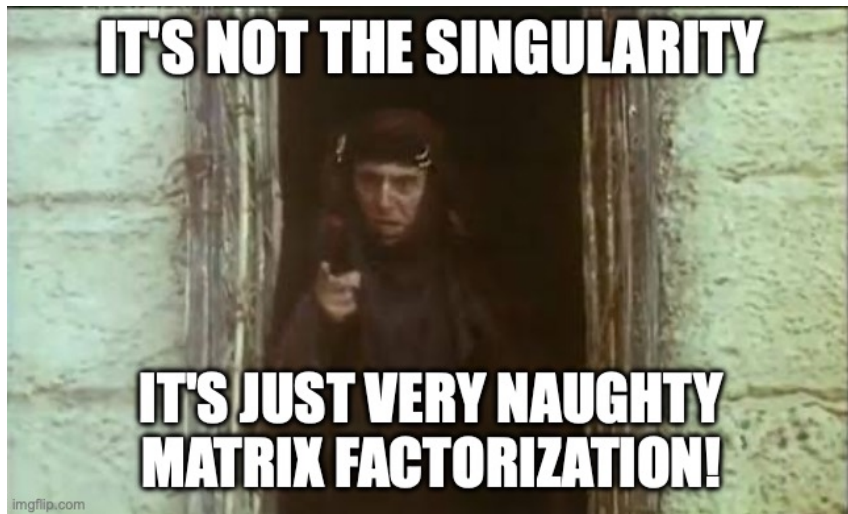
3.2

$$\hat{U} \times \hat{\Sigma}$$



3.2





3.3

Structure de l'espace latent

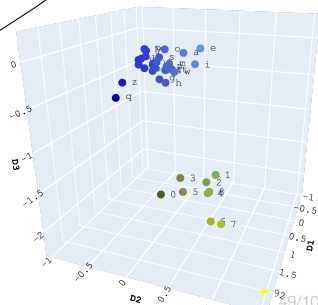
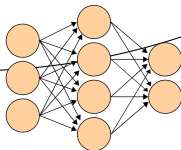
Structure

?

$\{-, /, 0, 1, 2, \dots, 8, 9, =,$
 $a, b, c, \dots, w, x, y, z, \acute{e}\}$

Embedding

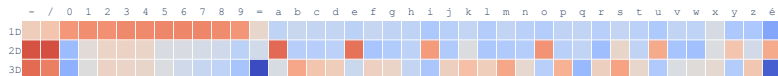
Data



3.3

Structure de l'espace latent

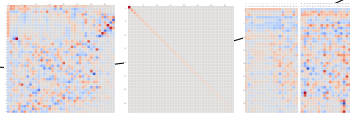
Structure



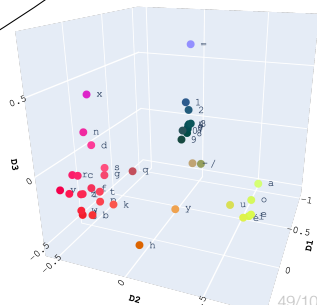
$\{-, /, 0, 1, 2, \dots, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$

Embedding

Data



SVD



4 Why does this produce good word representations?

Good question. We don't really know.

The distributional hypothesis states that words in similar contexts have similar meanings. The objective above clearly tries to increase the quantity $v_w \cdot v_c$ for good word-context pairs, and decrease it for bad ones. Intuitively, this means that words that share many contexts will be similar to each other (note also that contexts sharing many words will also be similar to each other). This is, however, very hand-wavy.

Can we make this intuition more precise? We'd really like to see something more formal.

(Goldberg and Levy, 2014)

3.3

Une matrice peut être comprise comme une fonction

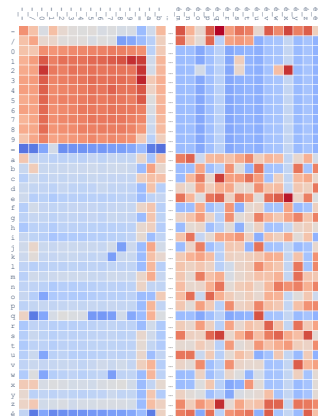
$$M: X \times Y \rightarrow \mathbb{R}$$

$$X = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

$$Y = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$$

$$M: X \times Y \rightarrow \mathbb{R}$$

$$(x, y) \mapsto \text{pmi}(x, y)$$



3.3

Une matrice peut être comprise comme une fonction

$$M: X \times Y \rightarrow \mathbb{R}$$

$$X = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

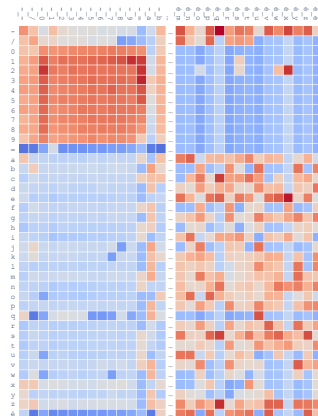
$$Y = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$$

$$M: X \times Y \rightarrow \mathbb{R}$$

$$(x, y) \mapsto \text{pmi}(x, y)$$

$$M_x: X \rightarrow \mathbb{R}^Y$$

$$x \mapsto M(x, -)$$



3.3

Une matrice peut être comprise comme une fonction

$$M: X \times Y \rightarrow \mathbb{R}$$

$$X = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

$$Y = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$$

$$M: X \times Y \rightarrow \mathbb{R}$$

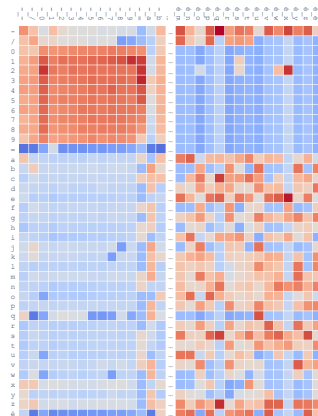
$$(x, y) \mapsto \text{pmi}(x, y)$$

$$M_x: X \rightarrow \mathbb{R}^Y$$

$$x \mapsto M(x, -)$$

$$M_y: Y \rightarrow \mathbb{R}^X$$

$$y \mapsto M(-, y)$$



3.3

Une matrice peut être comprise comme une fonction

$$M: X \times Y \rightarrow \mathbb{R}$$

$$X = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

$$Y = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$$

$$M: X \times Y \rightarrow \mathbb{R}$$

$$(x, y) \mapsto \text{pmi}(x, y)$$

$$X \xrightarrow{M_x} \mathbb{R}^Y$$

$$M_x: X \rightarrow \mathbb{R}^Y$$

$$x \mapsto M(x, -)$$

$$\mathbb{R}^X \xleftarrow{M_y} Y$$

$$M_y: Y \rightarrow \mathbb{R}^X$$

$$y \mapsto M(-, y)$$

3.3

Une matrice peut être comprise comme une fonction

$$M: X \times Y \rightarrow \mathbb{R}$$

$$X = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

$$Y = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$$

$$M: X \times Y \rightarrow \mathbb{R}$$

$$(x, y) \mapsto \text{pmi}(x, y)$$

$$M_x: X \rightarrow \mathbb{R}^Y$$

$$x \mapsto M(x, -)$$

$$M_y: Y \rightarrow \mathbb{R}^X$$

$$y \mapsto M(-, y)$$

$$\begin{array}{ccc} X & \xrightarrow{M_x} & \mathbb{R}^Y \\ \downarrow & & \uparrow \\ \mathbb{R}^X & \xleftarrow{M_y} & Y \end{array}$$

3.3

Une matrice peut être comprise comme une fonction

$$M: X \times Y \rightarrow \mathbb{R}$$

$$X = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

$$Y = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$$

$$M: X \times Y \rightarrow \mathbb{R}$$

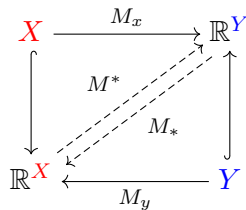
$$(x, y) \mapsto \text{pmi}(x, y)$$

$$M_x: X \rightarrow \mathbb{R}^Y$$

$$x \mapsto M(x, -)$$

$$M_y: Y \rightarrow \mathbb{R}^X$$

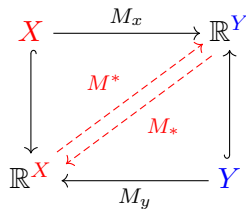
$$y \mapsto M(-, y)$$



$$M^*: \mathbb{R}^X \rightarrow \mathbb{R}^Y$$

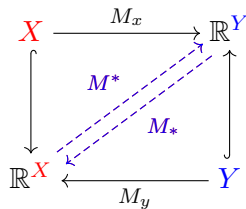
$$M_*: \mathbb{R}^Y \rightarrow \mathbb{R}^X$$

$$M_* M^* : \mathbb{R}^X \rightarrow \mathbb{R}^X$$



$$M_* M^* : \mathbb{R}^X \rightarrow \mathbb{R}^X$$

$$M^* M_* : \mathbb{R}^Y \rightarrow \mathbb{R}^Y$$



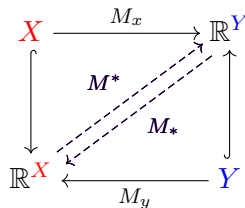
$$M_* M^* : \mathbb{R}^X \rightarrow \mathbb{R}^X$$

$$M^* M_* : \mathbb{R}^Y \rightarrow \mathbb{R}^Y$$

$$\{u_1, \dots, u_m\} \subset \mathbb{R}^X$$

$$\{v_1, \dots, v_n\} \subset \mathbb{R}^Y$$

$$\{\lambda_1, \dots, \lambda_{\min(m,n)}, 0, \dots, 0\}$$



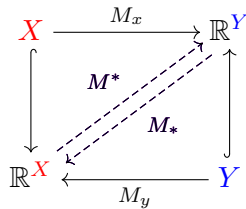
$$M_* M^* : \mathbb{R}^X \rightarrow \mathbb{R}^X$$

$$M^* M_* : \mathbb{R}^Y \rightarrow \mathbb{R}^Y$$

$$\{u_1, \dots, u_m\} \subset \mathbb{R}^X$$

$$\{v_1, \dots, v_n\} \subset \mathbb{R}^Y$$

$$\{\lambda_1, \dots, \lambda_{\min(m,n)}, 0, \dots, 0\}$$



$$U := [u_1, \dots, u_m]$$

$$M = U \Sigma V^T \quad V := [v_1, \dots, v_n]$$

$$\Sigma := \begin{bmatrix} \sqrt{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\lambda_r} \end{bmatrix}$$

$$M_* M^* : \mathbb{R}^X \rightarrow \mathbb{R}^X$$

$$M^* M_* : \mathbb{R}^Y \rightarrow \mathbb{R}^Y$$

$$\{u_1, \dots, u_m\} \subset \mathbb{R}^X$$

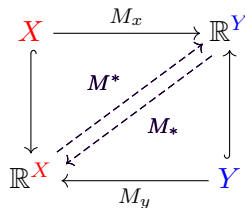
$$\{v_1, \dots, v_n\} \subset \mathbb{R}^Y$$

$$\{\lambda_1, \dots, \lambda_{\min(m,n)}, 0, \dots, 0\}$$

$$M_* M^* u_i = \lambda_i u_i$$

$$M^* M_* v_i = \lambda_i v_i$$

The u_i and v_i are (linear)
fixed points!



$$M = U \Sigma V^T$$

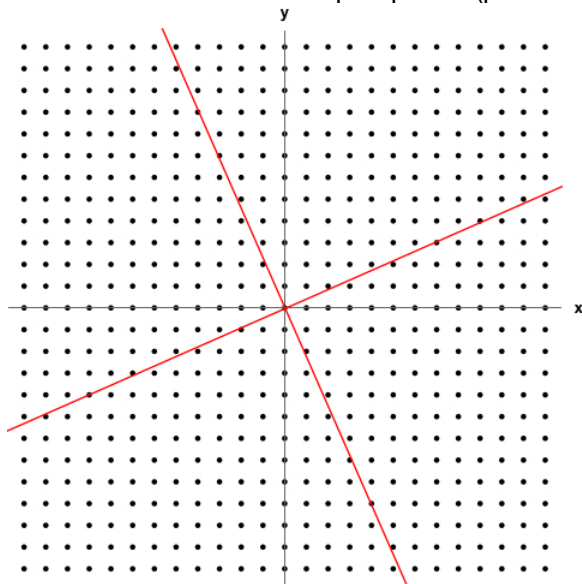
$$U := [u_1, \dots, u_m]$$

$$V := [v_1, \dots, v_n]$$

$$\Sigma := \begin{bmatrix} \sqrt{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\lambda_r} \end{bmatrix}$$

3.3

Vecteurs propres (points fixes linéaires)



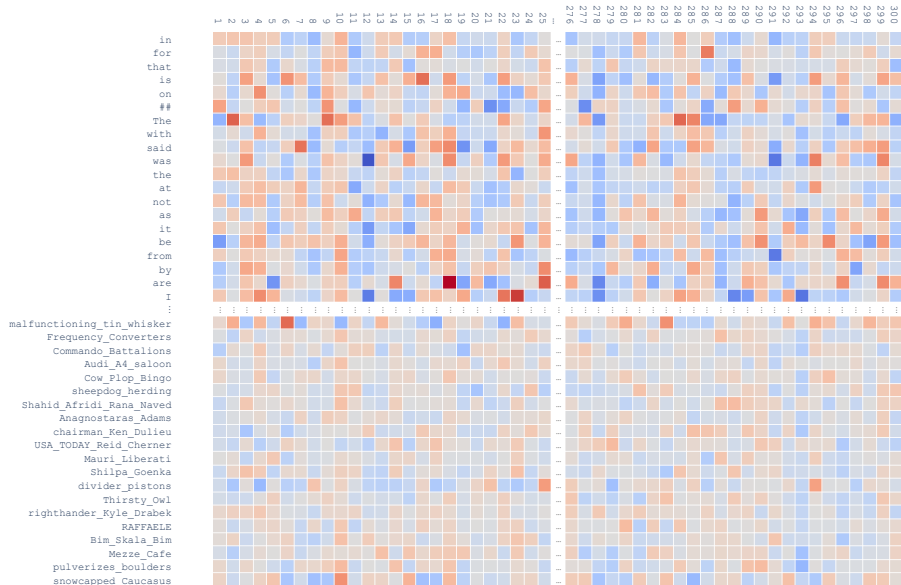
3.3

Embeddings en tant que points fixes



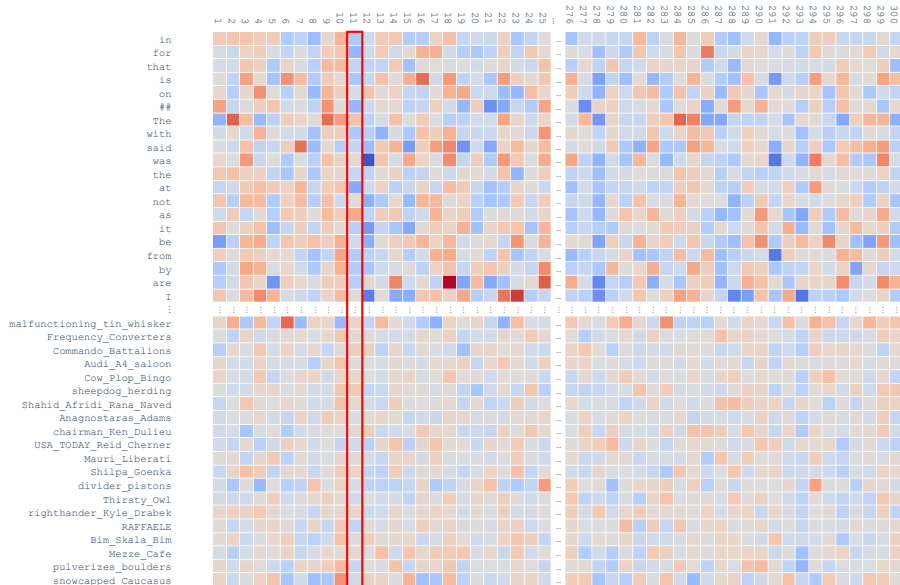
3.3

Embeddings en tant que points fixes



3.3

Embeddings en tant que points fixes

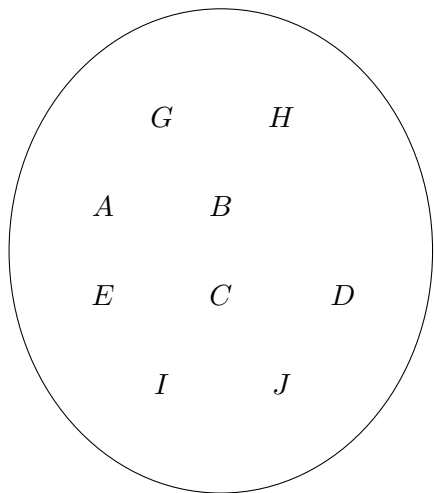


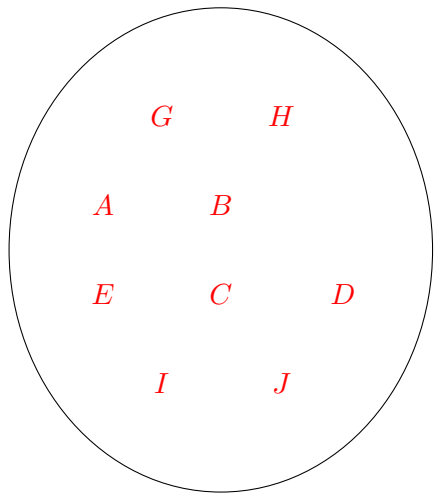
3.3



3.411

*Une catégorie est comme un ensemble muni d'une structure

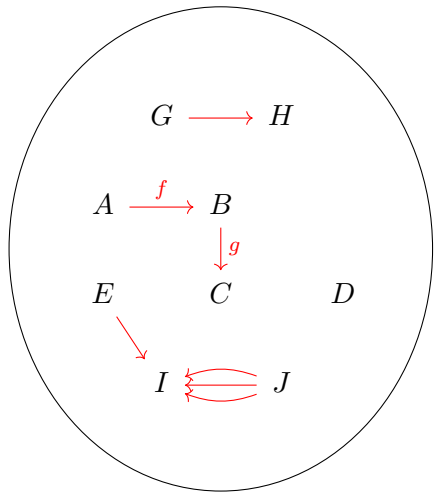




Definition (Category – Awodey, 2010)

Data:

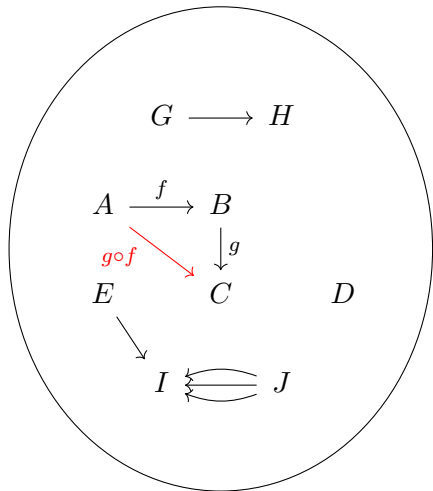
- ◇ Objects: *A, B, C, ...*



Definition (Category – Awodey, 2010)

Data:

- ◇ Objects: $A, B, C, \dots$
- ◇ Arrows: $f, g, \dots$

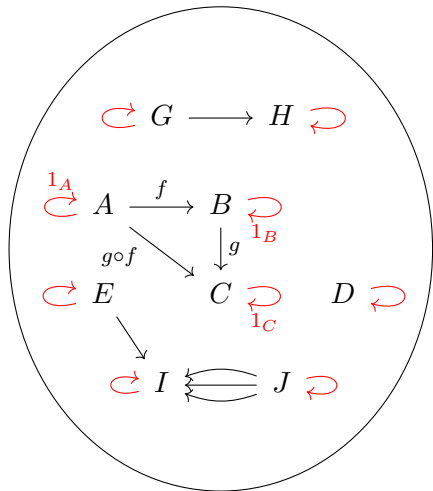


Definition (Category – Awodey, 2010)

Data:

- ◇ Objects: $A, B, C, \dots$
- ◇ Arrows: $f, g, \dots$
- ◇ Composition: Given $f : A \rightarrow B$ and $g : B \rightarrow C$, there is given an arrow

$$g \circ f : A \rightarrow C$$



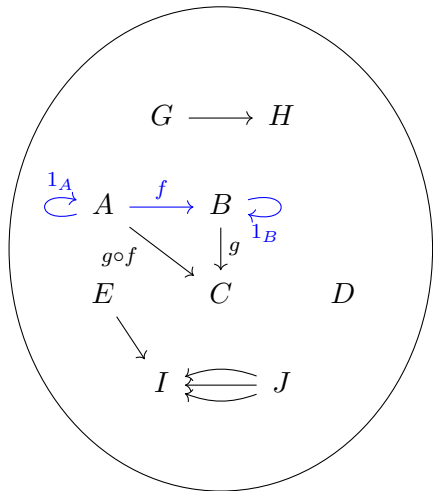
Definition (Category – Awodey, 2010)

Data:

- ◇ Objects: $A, B, C, \dots$
- ◇ Arrows: $f, g, \dots$
- ◇ Composition: Given $f : A \rightarrow B$ and $g : B \rightarrow C$, there is given an arrow

$$g \circ f : A \rightarrow C$$

- ◇ Identity: For each A , there is $1_A : A \rightarrow A$



Definition (Category – Awodey, 2010)

Data:

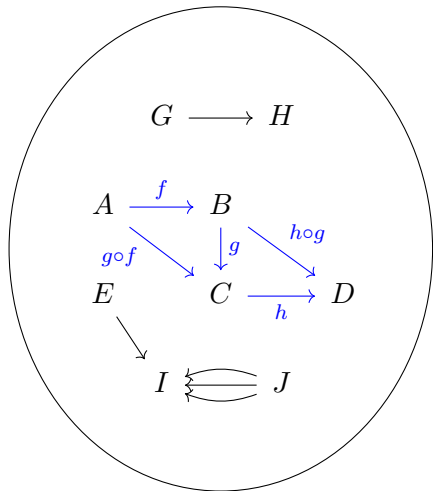
- ◇ Objects: $A, B, C, \dots$
- ◇ Arrows: $f, g, \dots$
- ◇ Composition: Given $f : A \rightarrow B$ and $g : B \rightarrow C$, there is given an arrow

$$g \circ f : A \rightarrow C$$

- ◇ Identity: For each A , there is $1_A : A \rightarrow A$

Laws:

- ◇ Unit: $f \circ 1_A = f = 1_B \circ f$



Definition (Category – Awodey, 2010)

Data:

- ◇ Objects: $A, B, C, \dots$
- ◇ Arrows: $f, g, \dots$
- ◇ Composition: Given $f : A \rightarrow B$ and $g : B \rightarrow C$, there is given an arrow

$$g \circ f : A \rightarrow C$$

- ◇ Identity: For each A , there is $1_A : A \rightarrow A$

Laws:

- ◇ Unit: $f \circ 1_A = f = 1_B \circ f$
- ◇ Associativity: $f \circ (g \circ h) = (f \circ g) \circ h$

3.411

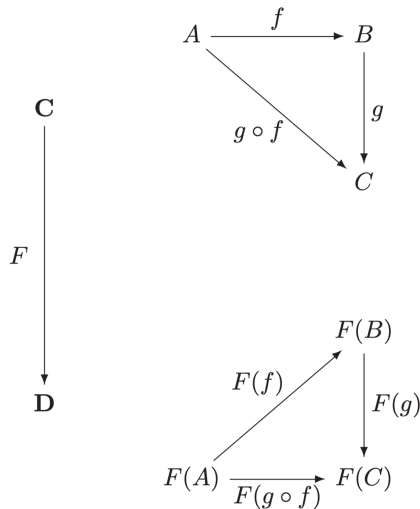


Definition (Functor – Awodey, 2010)

A functor

$$F: \mathbf{C} \rightarrow \mathbf{D}$$

between categories $\mathbf{C}$ and $\mathbf{D}$ is a mapping of objects to objects and arrows to arrows, in such a way that



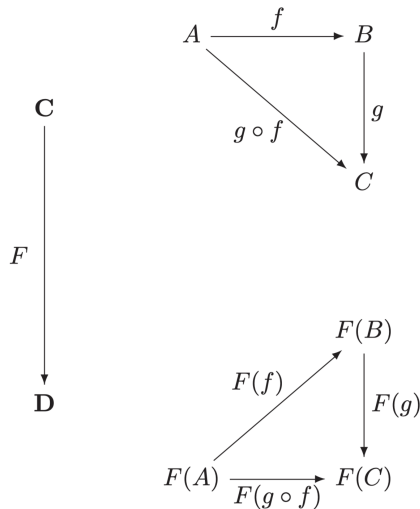
Definition (Functor – Awodey, 2010)

A functor

$$F: \mathbf{C} \rightarrow \mathbf{D}$$

between categories $\mathbf{C}$ and $\mathbf{D}$ is a mapping of objects to objects and arrows to arrows, in such a way that

$$(a) \quad F(f : A \rightarrow B) = F(f) : F(A) \rightarrow F(B)$$



Definition (Functor – Awodey, 2010)

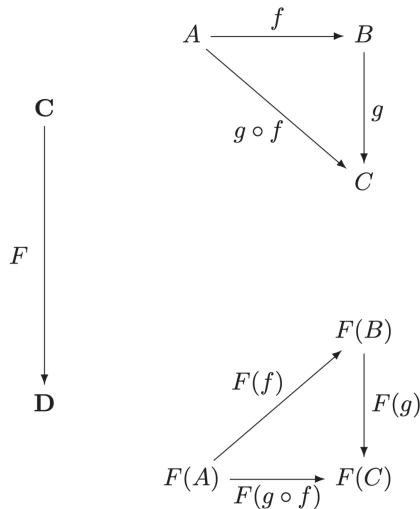
A functor

$$F: \mathbf{C} \rightarrow \mathbf{D}$$

between categories $\mathbf{C}$ and $\mathbf{D}$ is a mapping of objects to objects and arrows to arrows, in such a way that

$$(a) \quad F(f : A \rightarrow B) = F(f) : F(A) \rightarrow F(B)$$

$$(b) \quad F(1_A) = 1_{F(A)}$$



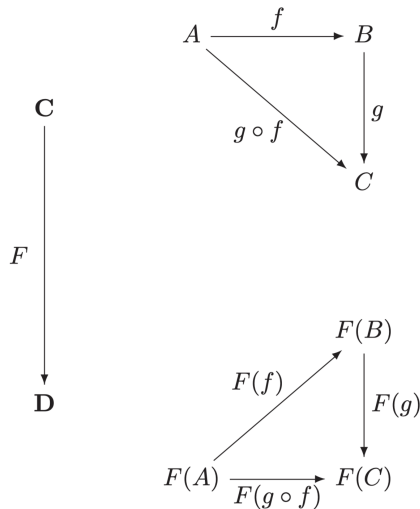
Definition (Functor – Awodey, 2010)

A functor

$$F: \mathbf{C} \rightarrow \mathbf{D}$$

between categories $\mathbf{C}$ and $\mathbf{D}$ is a mapping of objects to objects and arrows to arrows, in such a way that

- (a) $F(f : A \rightarrow B) = F(f) : F(A) \rightarrow F(B)$
- (b) $F(1_A) = 1_{F(A)}$
- (c) $F(g \circ f) = F(g) \circ F(f)$



3.412



Definition 2.15. In any category $\mathbf{C}$, a *product diagram* for the objects A and B consists of an object P and arrows

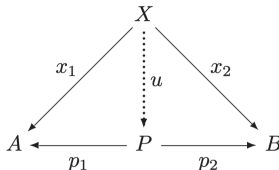
$$A \xleftarrow{p_1} P \xrightarrow{p_2} B$$

satisfying the following UMP:

Given any diagram of the form

$$A \xleftarrow{x_1} X \xrightarrow{x_2} B$$

there exists a unique $u : X \rightarrow P$, making the diagram



commute, that is, such that $x_1 = p_1 u$ and $x_2 = p_2 u$.

(Awodey, 2010)

3.413

Un profoncteur est un foncteur du produit de deux catégories quelconques vers la catégorie Set

$$\begin{array}{ccc} \textit{term}_i & \textit{context}_i & \textit{measure} \\ \downarrow & \downarrow & \swarrow \\ \text{C}^{\text{op}} \times \text{D} \rightarrow \text{Set} \end{array}$$

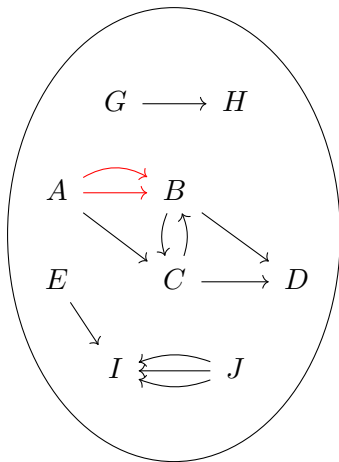
3.413



3.414

Une catégorie enrichie sur $\mathcal{V}$ est une catégorie dont les flèches entre deux objets sont des valeurs dans $\mathcal{V}$

$$\text{hom}(A, B) \\ \mathbf{C}(A, B) = \{f \in \mathbf{C} \mid f : A \rightarrow B\}$$

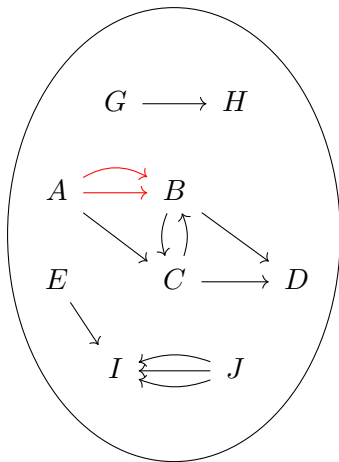


3.414

Une catégorie enrichie sur $\mathcal{V}$ est une catégorie dont les flèches entre deux objets sont des valeurs dans $\mathcal{V}$

$$\text{hom}(A, B) \\ \mathbf{C}(A, B) = \{f \in \mathbf{C} \mid f : A \rightarrow B\}$$

$$\mathbf{C}(A, B) \in \mathbf{Set}$$



3.414 Une catégorie enrichie sur $\mathcal{V}$ est une catégorie dont les flèches entre deux objets sont des valeurs dans $\mathcal{V}$

$$\begin{aligned} \text{hom}(A, B) \\ \mathbf{C}(A, B) &= \{f \in \mathbf{C} \mid f : A \rightarrow B\} \\ \mathbf{C}(A, B) &\in \mathbf{Set} \end{aligned}$$

Enrichment over $\mathcal{V}$

$\mathbf{C}(A, B) \in \mathcal{V}$,
where $\mathcal{V}$ is a “nice” (monoidal) category

3.414



3.415 Un foncteur entre les catégories enrichies $D \rightarrow C$ induit un
profoncteur $C^{\text{op}} \times D \rightarrow \mathcal{V}$

$$\begin{array}{ccc} \textit{term}_i & \textit{context}_i & \textit{measure} \\ \downarrow & \downarrow & \swarrow \\ C^{\text{op}} & \times D & \rightarrow \textit{Set} \end{array}$$

3.415 Un foncteur entre les catégories enrichies $D \rightarrow C$ induit un
profoncteur $C^{\text{op}} \times D \rightarrow \mathcal{V}$

$$\begin{array}{ccc} \textit{term}_i & \textit{context}_i & \textit{measure} \\ \downarrow & \downarrow & \swarrow \\ C^{\text{op}} & \times D & \rightarrow \mathcal{V} \end{array}$$

3.415 Un foncteur entre les catégories enrichies $D \rightarrow C$ induit un
profoncteur $C^{\text{op}} \times D \rightarrow \mathcal{V}$

The diagram illustrates the mapping of semantic concepts to the components of the enriched profunctor equation. It features three teal labels at the top: $term_i$, $context_i$, and $measure$. Below them, the equation $C^{\text{op}} \times D \rightarrow 2$ is displayed, with C^{op} in orange, D in green, and 2 in red. Arrows indicate the following mappings: a straight arrow from $term_i$ to C^{op} , a straight arrow from $context_i$ to D , and a curved arrow from $measure$ to 2 .

$$\begin{array}{ccc} term_i & context_i & measure \\ \downarrow & \downarrow & \searrow \\ C^{\text{op}} & \times D & \rightarrow 2 \end{array}$$

3.415 Un foncteur entre les catégories enrichies $D \rightarrow C$ induit un
profoncteur $C^{\text{op}} \times D \rightarrow \mathcal{V}$

$$\begin{array}{ccc} \textit{term}_i & \textit{context}_i & \textit{measure} \\ \downarrow & \downarrow & \swarrow \\ C^{\text{op}} & \times D & \rightarrow \bar{\mathbb{R}} \end{array}$$

3.415



3.41



3.41

*De l'algèbre linéaire à la *théorie des catégories*

- 3.411 *Une catégorie est comme un ensemble muni d'une structure
- 3.412 Un foncteur est une application entre catégories
- 3.413 Un profoncteur est un foncteur du produit de deux catégories quelconques vers la catégorie **Set**
- 3.414 Une catégorie enrichie sur $\mathcal{V}$ est une catégorie dont les flèches entre deux objets sont des valeurs dans $\mathcal{V}$
- 3.415 Un foncteur entre les catégories enrichies $\mathbf{D} \rightarrow \mathbf{C}$ induit un profoncteur $\mathbf{C}^{\text{op}} \times \mathbf{D} \rightarrow \mathcal{V}$

3.42 Embeddings en tant que foncteurs sur des catégories

$$X = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

$$Y = X \times X = \{(-, -), (-, /), (-, 0), \dots, (\acute{e}, z), (\acute{e}, \acute{e})\}$$

$$M: X \times Y \rightarrow \mathbb{R}$$

$$(x, y) \mapsto \text{pmi}(x, y)$$

$$M_x: X \rightarrow \mathbb{R}^Y$$

$$x \mapsto M(x, -)$$

$$M_y: Y \rightarrow \mathbb{R}^X$$

$$y \mapsto M(-, y)$$

$$\begin{array}{ccc}
 X & \xrightarrow{M_x} & \mathbb{R}^Y \\
 \downarrow & \nearrow M^* & \uparrow \\
 \mathbb{R}^X & \xleftarrow{M_y} & Y \\
 & \nwarrow M_* &
 \end{array}$$

$$M^*: \mathbb{R}^X \rightarrow \mathbb{R}^Y$$

$$M_*: \mathbb{R}^Y \rightarrow \mathbb{R}^X$$

3.42 Embeddings en tant que foncteurs sur des catégories

$$\mathbf{C} = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

$$\mathbf{D} = \mathbf{C} = \{-, /, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, =, a, b, c, \dots, w, x, y, z, \acute{e}\}$$

Profunctor

$$\mathcal{M}: \mathbf{C}^{\text{op}} \times \mathbf{D} \rightarrow \mathbf{Set}$$

$$(c, d) \mapsto \mathcal{M}(c, d)$$

$$\mathcal{M}_c: \mathbf{C} \rightarrow (\mathbf{Set}^{\mathbf{D}})^{\text{op}}$$

$$c \mapsto \mathcal{M}(c, -)$$

$$\mathcal{M}_d: \mathbf{D} \rightarrow \mathbf{Set}^{\mathbf{C}^{\text{op}}}$$

$$d \mapsto \mathcal{M}(-, d)$$

$$\begin{array}{ccc}
 \mathbf{C} & \xrightarrow{\mathcal{M}_c} & (\mathbf{Set}^{\mathbf{D}})^{\text{op}} \\
 \text{Yoneda} \downarrow & \nearrow \mathcal{M}^* & \uparrow \text{Yoneda} \\
 \mathbf{Set}^{\mathbf{C}^{\text{op}}} & \xleftarrow{\mathcal{M}_d} & \mathbf{D}
 \end{array}$$

$\mathcal{M}_*$ (dashed arrow from $\mathbf{D}$ to $\mathbf{Set}^{\mathbf{C}^{\text{op}}}$)

$$\mathcal{M}^*: \mathbf{Set}^{\mathbf{C}^{\text{op}}} \rightarrow (\mathbf{Set}^{\mathbf{D}})^{\text{op}}$$

$$\mathcal{M}_*: (\mathbf{Set}^{\mathbf{D}})^{\text{op}} \rightarrow \mathbf{Set}^{\mathbf{C}^{\text{op}}}$$

3.42 Embeddings en tant que foncteurs sur des catégories

Adjonction d'Ishbell

$$\mathcal{M}^*: \mathbf{Set}^{\mathbf{C}^{\text{op}}} \rightleftarrows (\mathbf{Set}^{\mathbf{D}})^{\text{op}} : \mathcal{M}_*$$

$$\mathcal{M}_* \mathcal{M}^*: \mathbf{Set}^{\mathbf{C}^{\text{op}}} \rightarrow \mathbf{Set}^{\mathbf{C}^{\text{op}}}$$

$$\mathcal{M}^* \mathcal{M}_*: (\mathbf{Set}^{\mathbf{D}})^{\text{op}} \rightarrow (\mathbf{Set}^{\mathbf{D}})^{\text{op}}$$

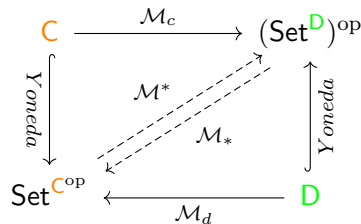
$$\text{Fix}(\mathcal{M}_* \mathcal{M}^*) := \{f \in \mathbf{Set}^{\mathbf{C}^{\text{op}}} \mid \mathcal{M}_* \mathcal{M}^*(f) \cong f\}$$

$$\text{Fix}(\mathcal{M}^* \mathcal{M}_*) := \{g \in (\mathbf{Set}^{\mathbf{D}})^{\text{op}} \mid \mathcal{M}^* \mathcal{M}_*(g) \cong g\}$$

Nucleus of $\mathcal{M} = \{(f_i, g_i)\}$, such that:

$$\mathcal{M}^* f_i \cong g_i \text{ and } \mathcal{M}_* g_i \cong f_i$$

Le **noyau** (nucleus) est une **catégorie complète** et **cocomplète**



Les catégories **C** et **D** peuvent être enrichies!

E.g.:

$$\mathcal{M}^*: \mathbf{2}^{\mathbf{C}^{\text{op}}} \rightleftarrows (\mathbf{2}^{\mathbf{D}})^{\text{op}} : \mathcal{M}_*$$

$$\mathcal{M}^*: \bar{\mathbf{R}}^{\mathbf{C}^{\text{op}}} \rightleftarrows (\bar{\mathbf{R}}^{\mathbf{D}})^{\text{op}} : \mathcal{M}_*$$

3.42



3.4



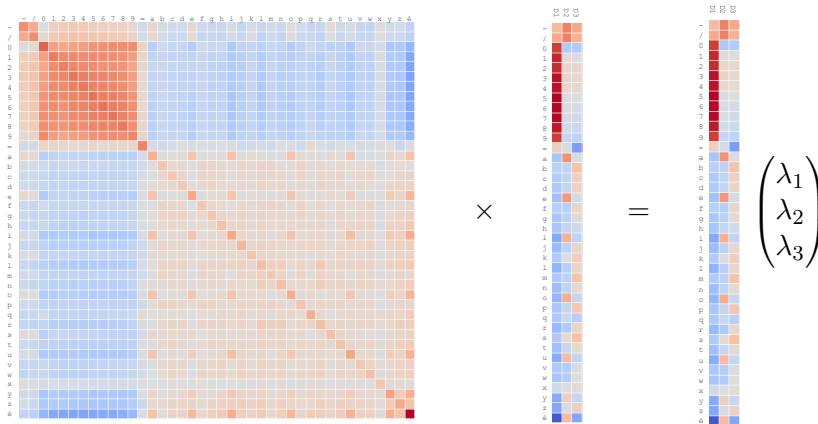
3.4

*Il est possible de généraliser ce résultat

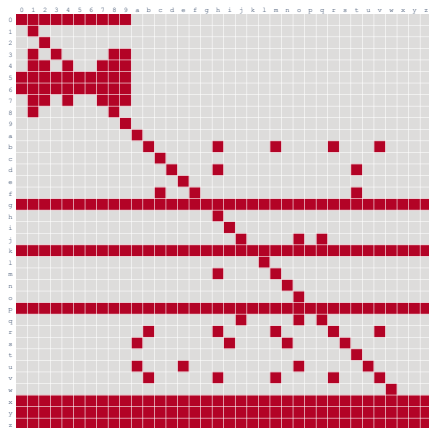
3.41 *De l'algèbre linéaire à la *théorie des catégories*

3.42 *Il existe un parallèle profond entre des opérateurs linéaires et catégoriques

$$M_* M^* u = \lambda u$$



$$\mathcal{M}_* \mathcal{M}^* f = f$$



★

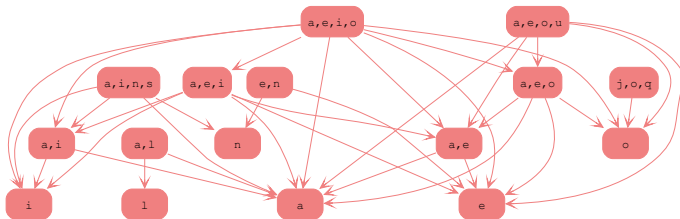
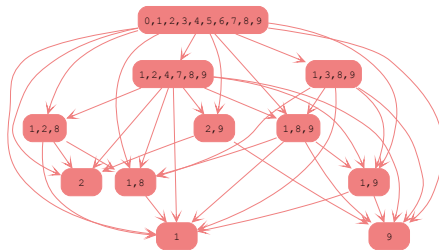
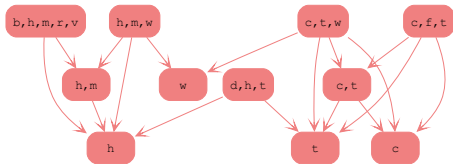


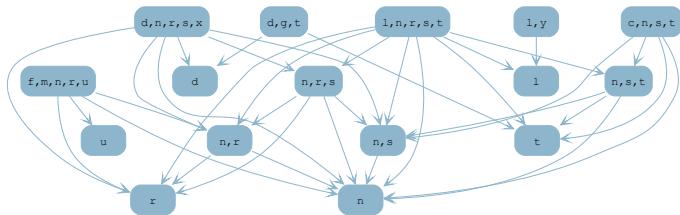
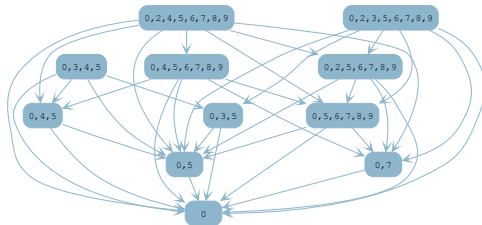
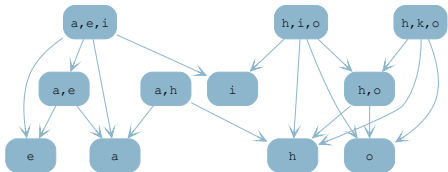
?

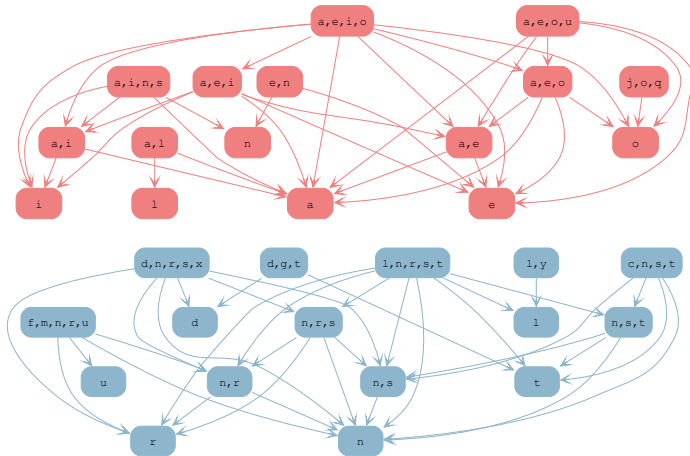


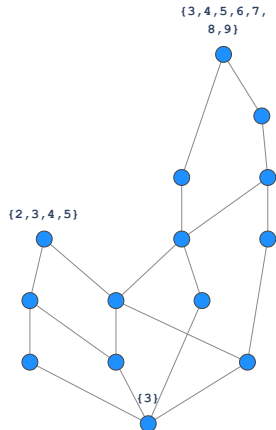
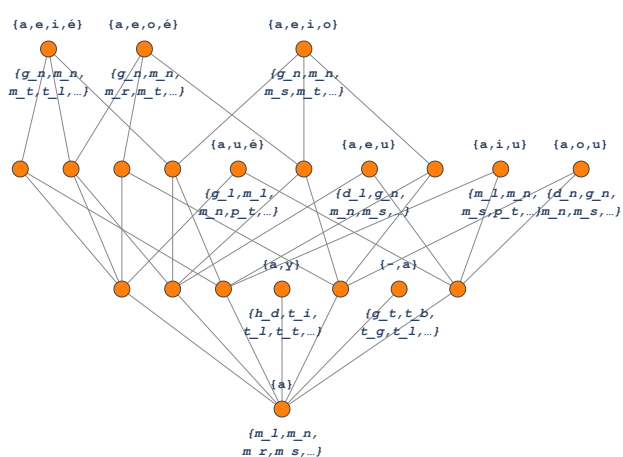
$$\mathcal{M}_* \mathcal{M}^* f = f$$

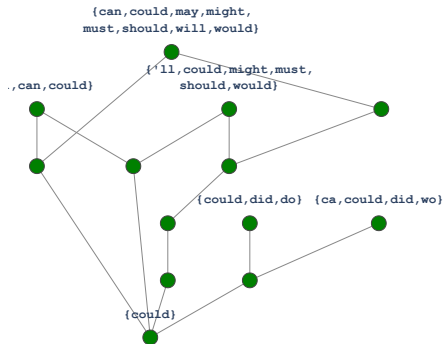
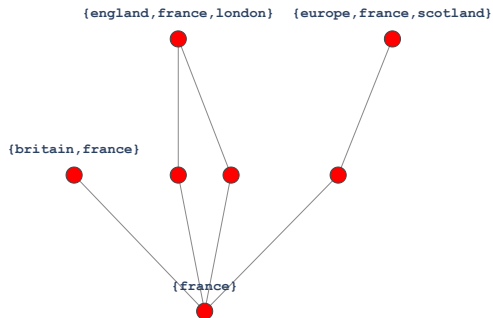
0,1,2,3,4,5,6,7,8,9	1,2,4,7,8,9	b,h,m,r,v	a,e,i,o	a,e,o,u	a,i,n,s	1,3,8,9
1,2,8	h,m,w	1,8,9	d,h,t	j,o,q	c,f,t	c,t,w
a,e,o	a,e,i	h,m	2,9	a,i	w	1,9
1,8	a,e	l	t	n	c	h
2	i	e	a	o	l	9
e,n	a,l	c,t				

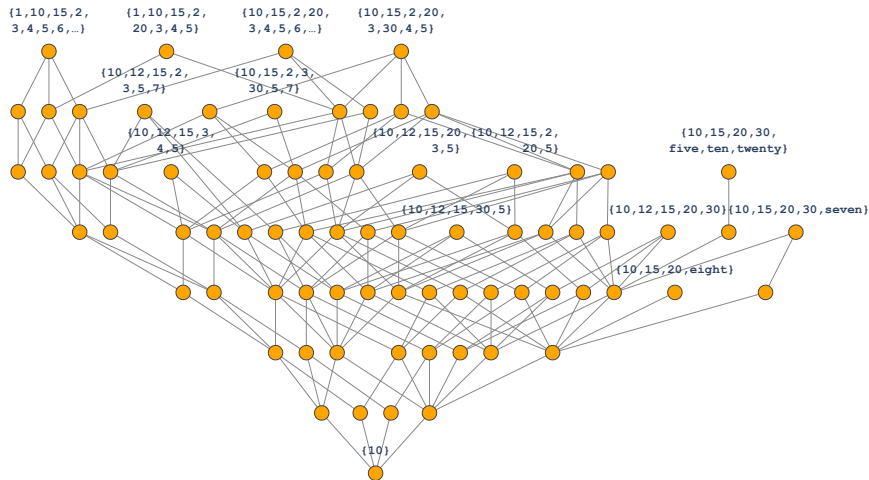










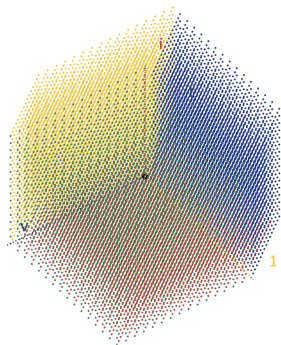


3.51

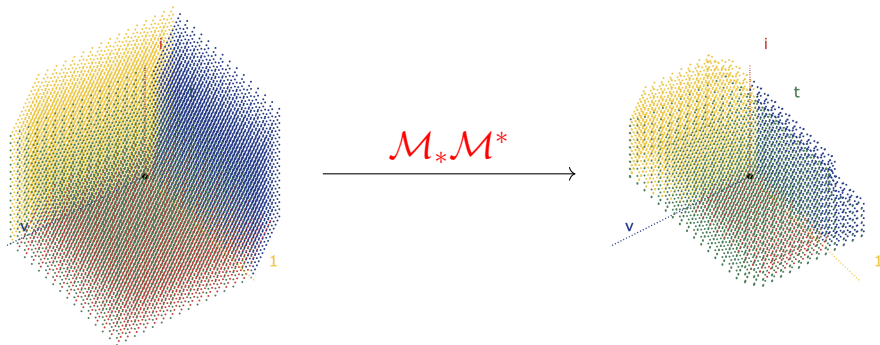


$$\begin{array}{ccc} e_i & s_i & \text{mesurement} \\ \text{(terms)} & \text{(contexts)} & \\ \downarrow & \downarrow & \swarrow \\ \mathbf{C}^{\text{op}} \times \mathbf{D} \rightarrow \bar{\mathbb{R}} \end{array}$$

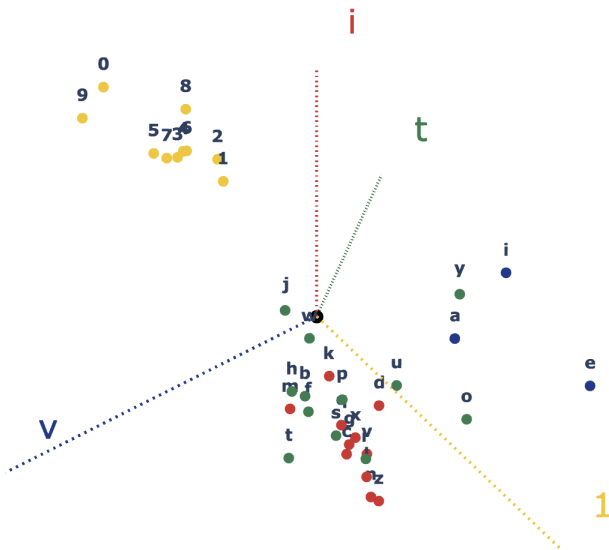
$$\begin{array}{ccc}
 \begin{array}{c} e_i \\ \text{(terms)} \end{array} & \begin{array}{c} s_i \\ \text{(contexts)} \end{array} & \text{mesurement} \\
 \downarrow & \downarrow & \swarrow \\
 \mathbf{C}^{\text{op}} \times \mathbf{D} & \rightarrow & \bar{\mathbb{R}} \\
 \Downarrow & & \\
 \mathcal{M}^* : \bar{\mathbb{R}}^{\mathbf{C}^{\text{op}}} & \rightleftharpoons & (\bar{\mathbb{R}}^{\mathbf{D}})^{\text{op}} : \mathcal{M}_*
 \end{array}$$

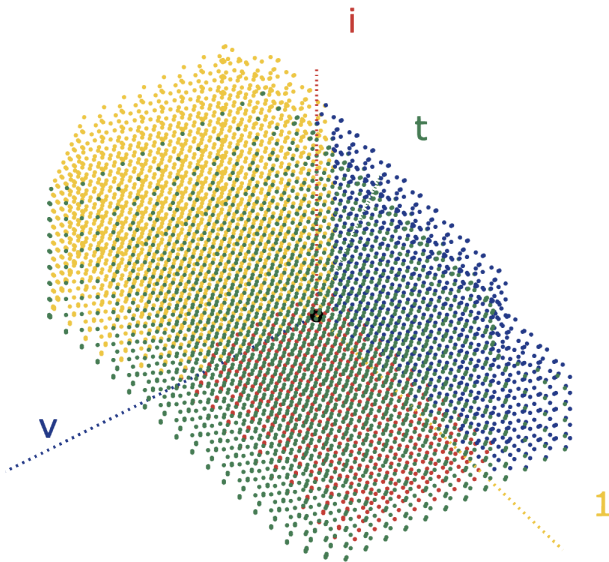


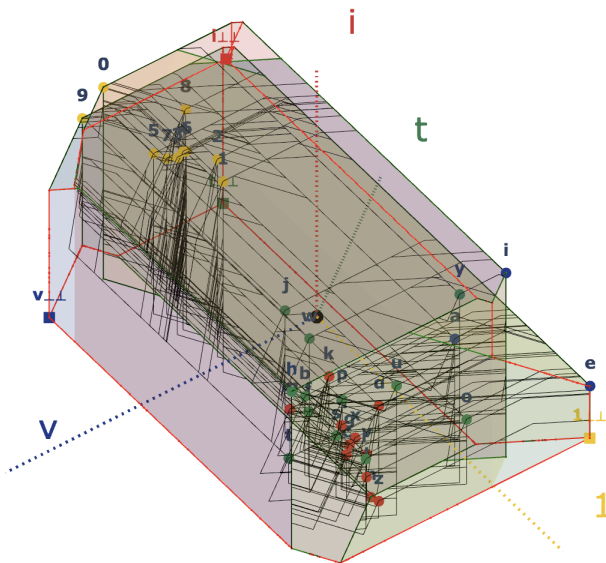
$$\begin{array}{ccc}
 \mathbf{C}^{\text{op}} \times \mathbf{D} & \rightarrow & \bar{\mathbb{R}} \\
 \Downarrow & & \\
 \mathcal{M}^* : \bar{\mathbb{R}}^{\mathbf{C}^{\text{op}}} & \rightleftharpoons & (\bar{\mathbb{R}}^{\mathbf{D}})^{\text{op}} : \mathcal{M}_*
 \end{array}$$



$$\begin{array}{ccc}
 \mathbf{C}^{\text{op}} \times \mathbf{D} & \rightarrow & \bar{\mathbb{R}} \\
 \Downarrow & & \\
 \mathcal{M}^* : \bar{\mathbb{R}}^{\mathbf{C}^{\text{op}}} & \rightleftarrows & (\bar{\mathbb{R}}^{\mathbf{D}})^{\text{op}} : \mathcal{M}_*
 \end{array}$$







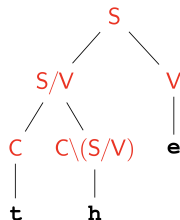
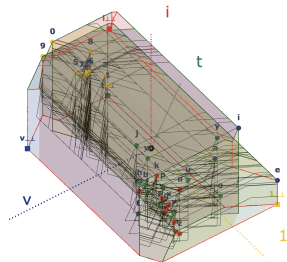
Definition (Polar/Orthogonal - Girard, 2011)

[G]iven a binary operation, noted $a, b \rightsquigarrow \langle a|b \rangle : A \times B \rightarrow C$ and a subset $P \subset C$ (the 'pole') one can define the *polar* $X^\perp \subset B$ of a subset $X \subset A$ (resp. $Y^\perp \subset A$ of a subset $Y \subset B$) by :

$$X^\perp := \{y \in B : \forall x \in X, \langle a|b \rangle \in P\}$$

$$Y^\perp := \{x \in A : \forall y \in Y, \langle a|b \rangle \in P\}$$

- ◇ The map 'polar' is decreasing:
 $X \subset X' \Rightarrow X'^\perp \subset X^\perp$.
- ◇ The set $\text{Pol}(A) \subset \mathcal{P}(A)$ of *polar* sets, i.e., of the form $Y^\perp$, is closed under arbitrary intersections. In particular, A is polar and $X^{\perp\perp}$ is the smallest polar set containing X .
- ◇ As a consequence, $X^{\perp\perp\perp} = X^\perp$.



3.52



3.5



3.5 *Cette généralisation permet de révéler beaucoup plus de structure

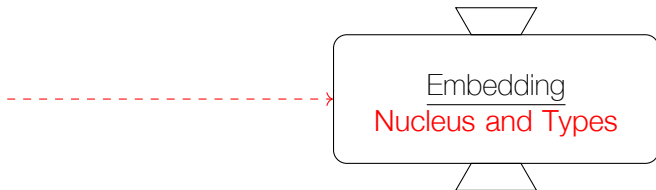
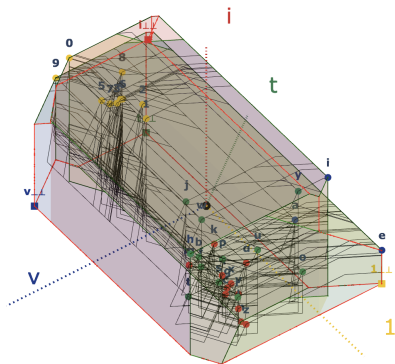
3.51 *Enrichissement sur **2**: Concepts formels

3.52 Enrichissement sur $\bar{\mathbb{R}}$

3.6

Tokenisation, embeddings, attention

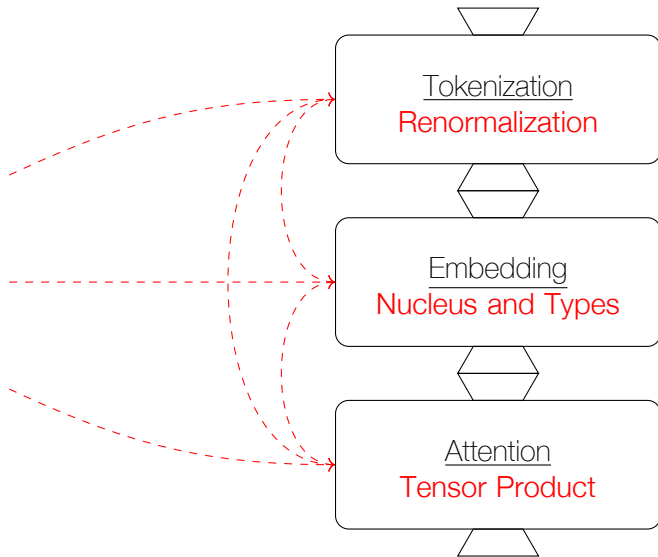
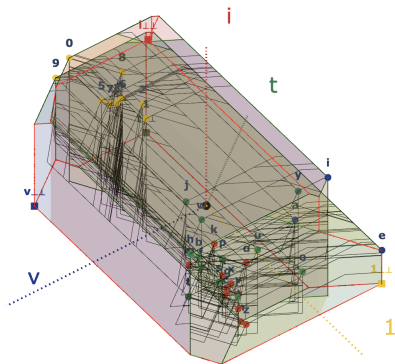
Structure



3.6

Tokenisation, embeddings, attention

Structure



3.6

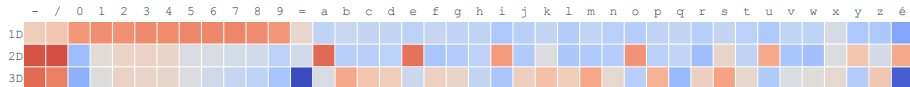


3

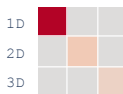


- 3 *Cette critique épistémologique offre les fondements d'une *explicabilité formelle* des LLMs
 - 3.1 *La clé formelle des LLMs réside dans les *embeddings*
 - 3.2 *SVD d'une matrice PMI fournit l'explication formelle des embeddings
 - 3.3 *Ce résultat a d'importantes conséquences pour l'explicabilité
 - 3.4 *Il est possible de généraliser ce résultat
 - 3.5 *Cette généralisation permet de révéler beaucoup plus de structure
 - 3.6 Le noyau du profoncteur pourrait permettre d'étudier la tokenisation, les embeddings et l'attention de manière formellement unifiée

Eigenvectors of M_*M^* :

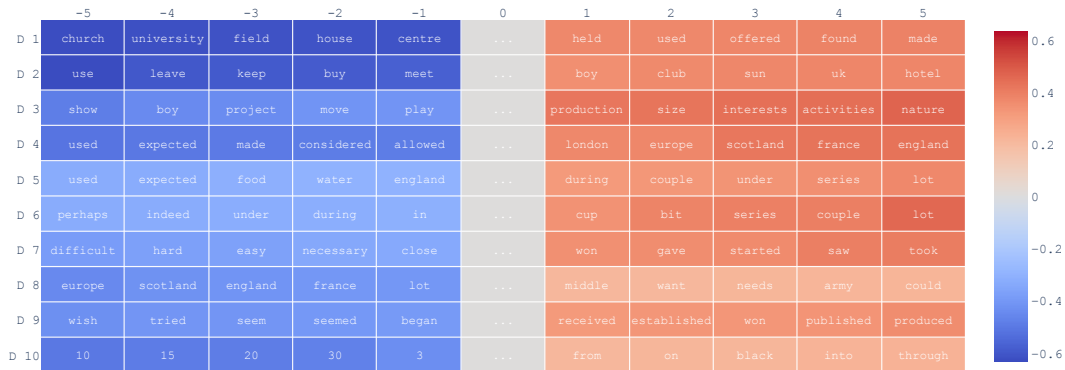


Eigenvalues of M_*M^* and M^*M_* :



Eigenvectors of M^*M_* :





4.1



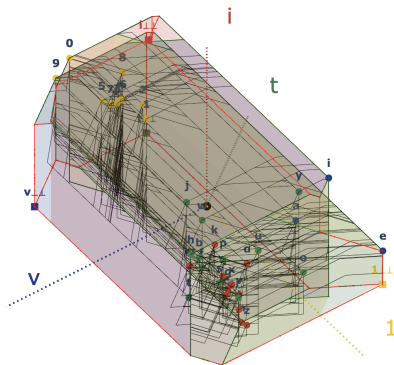
4.2 De l'hypothèse distributionnelle à l'hypothèse structuraliste

Theory
"Task"

?



Structure



4.2 De l'hypothèse distributionnelle à l'hypothèse structuraliste

$$C^{\text{op}} \times D \rightarrow \bar{\mathbb{R}}$$

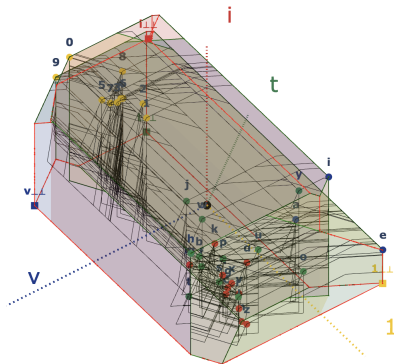
Hypothèse distributionnelle

Le contenu des unités linguistiques est déterminé par leur *distribution* dans un corpus.

Theory
"Task"



Structure



4.2 De l'hypothèse distributionnelle à l'hypothèse structuraliste

$$C^{\text{op}} \times D \rightarrow \bar{\mathbb{R}}$$

Hypothèse distributionnelle

Le contenu des unités linguistiques est déterminé par leur *distribution* dans un corpus.

Theory
"Task"

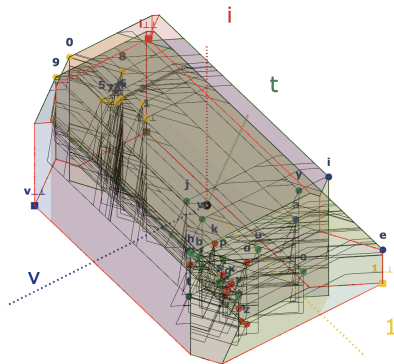


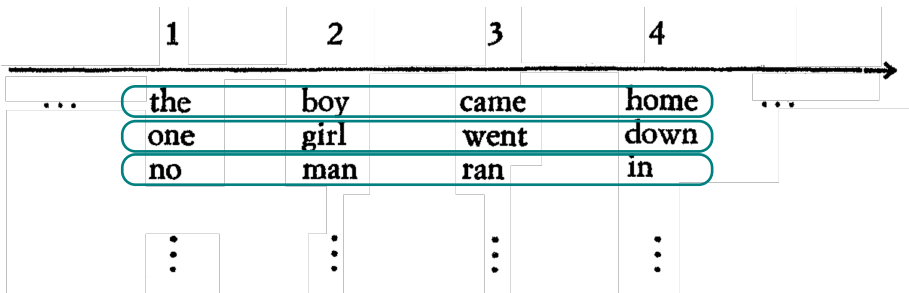
Hypothèse structurale

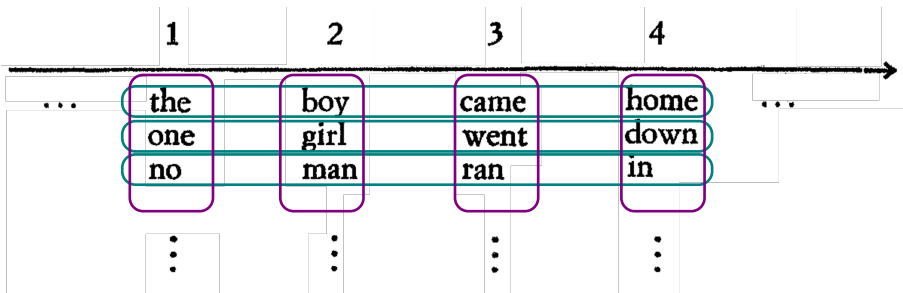
Le contenu linguistique est l'effet d'une **structure** virtuelle dérivée des pratiques linguistiques dans une communauté.

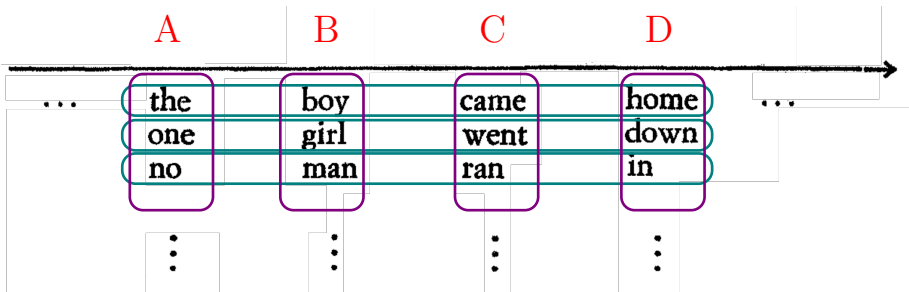
$$\bar{\mathbb{R}}^{C^{\text{op}}} \Leftrightarrow (\bar{\mathbb{R}}^D)^{\text{op}}$$

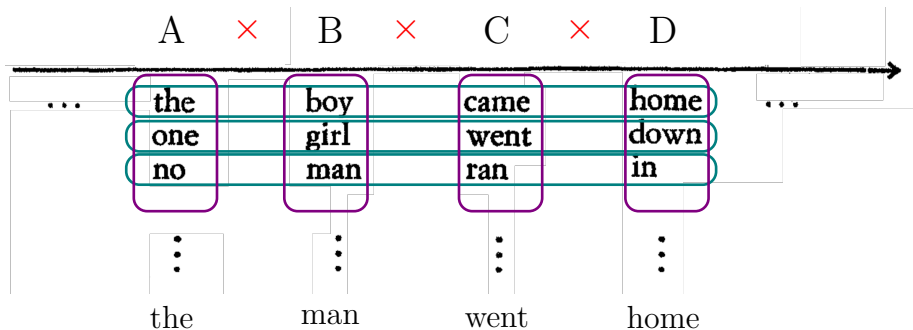
Structure



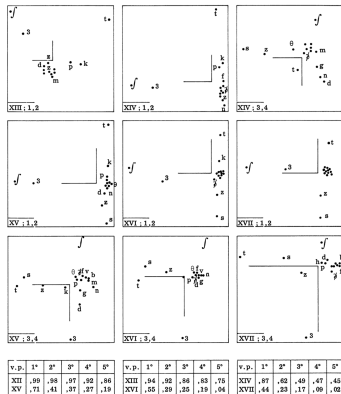








(Hjelmslev, 1971)



(benzécri1976codage)

1) une voyelle neutre (amorphe), caractérisée par l'absence de chacune des propriétés $\beta, \phi, \chi, \lambda$: [a] ;

2) quatre types élémentaires de voyelles, chacun caractérisé par une seule propriété :

$$[e] = \phi, [a] = \beta, [v] = \chi, [z] = \lambda$$

3) six voyelles distinctes, chacune caractérisée par deux propriétés :

$$[\theta] = \phi\beta, [i] = \chi\phi, [u] = \chi\beta, [e] = \lambda\phi, [o] = \lambda\beta, [\partial] = \chi\lambda;$$

4) quatre voyelles combinées, chacune caractérisée par trois propriétés :

$$[\epsilon] = \chi\lambda\phi, [\sigma] = \chi\lambda\beta, [\alpha] = \phi\beta\chi, [\partial] = \phi\beta\lambda;$$

5) une voyelle polymorphe, caractérisée par les quatre propriétés considérées : la voyelle russe [u] = $\phi\beta\chi\lambda$.

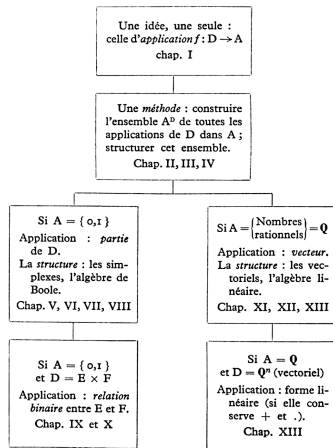
On obtient alors le diagramme de la figure 2.



FIG. 2.

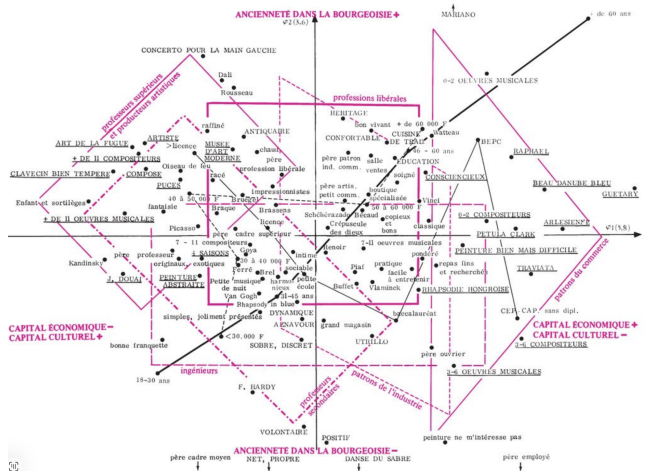
(Marcus, 1967)

Organigramme



(Barbut1967_T1)

“S’il est vrai que [...] la **classe dominante** constitue un espace relativement autonome dont la **structure** est définie par la **distribution** entre ses membres des différentes espèces de capital, [...] on doit **retrouver ces structures** dans l’espace des styles de vie [...]. C’est ce que l’on a essayé d’établir en soumettant à l’analyse des correspondances l’ensemble des données recueillies.”



(Bourdieu, 1979)

4.2



4



4 L'explicabilité formelle ouvre la voie vers une *interprétabilité théorique* des données

- 4.1 Les points fixes linéaires présentent des caractéristiques interprétables
- 4.2 Nous devons passer de l'hypothèse distributionnelle à l'*hypothèse structuraliste*

5



- 1 *Nous avons besoin d'un *formalisme critique*
- 2 *Un formalisme critique habilite une *critique épistémologique* de l'IA
- 3 *Cette critique épistémologique offre les fondements d'une *explicabilité formelle* des LLMs
- 4 L'explicabilité formelle ouvre la voie vers une *interprétabilité théorique* des données
- 5 L'IA n'est rien d'autre que des *sciences humaines* déguisées

Collaborations



J. Terilla (CUNY), T.-D. Bradley (SandboxAQ), L. Pellissier (Paris-Est Créteil), Th. Seiller (CNRS), S. Jarvis (CUNY)

Articles de référence

- ◇ Gastaldi, J. L. (2021). Why Can Computers Understand Natural Language? *Philosophy & Technology*, 34(1), 149–214. <https://doi.org/10.1007/s13347-020-00393-9>
- ◇ Gastaldi, J. L., & Pellissier, L. (2021b). The calculus of language: explicit representation of emergent linguistic structure through type-theoretical paradigms. *Interdisciplinary Science Reviews*, 46(4), 569–590. <https://doi.org/10.1080/03080188.2021.1890484>
- ◇ Bradley, T.-D., Gastaldi, J. L., & Terilla, J. (2024). The structure of meaning in language: Parallel narratives in linear algebra and category theory. *Notices of the American Mathematical Society*. <https://api.semanticscholar.org/CorpusID:263613625>

Argument complet I

- 1 *Nous avons besoin d'un *formalisme critique*
- 1.1 *La critique de l'IA est à court de carburant
- 1.2 *Les limites de la critique dans la production du savoir tiennent à la place des savoirs formels
 - 1.21 Les savoirs formels sont une construction récente
 - 1.22 Ils ont été mobilisés pour fonder des épistémologies dogmatiques
 - 1.23 La tradition critique a fait du formalisme une cible
- 1.3 *Une nouvelle alliance entre pensée critique et formalisme est nécessaire
 - 1.31 Le formalisme n'est pas un naturalisme
 - 1.32 Un formalisme critique est possible
- 2 *Un formalisme critique habilite une *critique épistémologique* de l'IA
 - 2.1 *L'étude empirique des LLMs n'a pas de fondement épistémologique
 - 2.11 L'informatique traverse un tournant empirique autour des LLMs
 - 2.12 Mais les LLMs ne sont que des fonctions calculables

Argument complet II

- 2.13 Il n'existe pas de moyen empirique de savoir ce qu'une fonction calculable fait
- 2.14 *La seule question épistémologique valide est: de quoi cette fonction est-elle l'implémentation?
- 2.2 *Les LLMs n'ont aucune portée cognitive a priori
- 2.21 La portée cognitive des modèles de langage computationnels n'est pas inconditionnelle
- 2.22 La condition épistémologique assurant un tel lien ne s'applique pas aux LLMs
- 2.23 L'absence de portée cognitive n'empêche pas les LLMs d'être des modèles du langage
- 3 *Cette critique épistémologique offre les fondements d'une *explicabilité formelle* des LLMs
 - 3.1 *La clé formelle des LLMs réside dans les *embeddings*
 - 3.2 *SVD d'une matrice PMI fournit l'explication formelle des embeddings

Argument complet III

- 3.3 *Ce résultat a d'importantes conséquences pour l'explicabilité
- 3.4 *Il est possible de généraliser ce résultat
- 3.41 *De l'algèbre linéaire à la *théorie des catégories*
- 3.411 *Une catégorie est comme un ensemble muni d'une structure
- 3.412 Un foncteur est une application entre catégories
- 3.413 Un profoncteur est un foncteur du produit de deux catégories quelconques vers la catégorie **Set**
- 3.414 Une catégorie enrichie sur $\mathcal{V}$ est une catégorie dont les flèches entre deux objets sont des valeurs dans $\mathcal{V}$
- 3.415 Un foncteur entre les catégories enrichies $\mathbf{D} \rightarrow \mathbf{C}$ induit un profoncteur $\mathbf{C}^{\text{op}} \times \mathbf{D} \rightarrow \mathcal{V}$
- 3.42 *Il existe un parallèle profond entre des opérateurs linéaires et catégoriques
- 3.5 *Cette généralisation permet de révéler beaucoup plus de structure

Argument complet IV

- 3.51 *Enrichissement sur **2**: Concepts formels
- 3.52 Enrichissement sur $\bar{\mathbb{R}}$
- 3.6 Le noyau du profoncteur pourrait permettre d'étudier la tokenisation, les embeddings et l'attention de manière formellement unifiée
- 4 L'explicabilité formelle ouvre la voie vers une *interprétabilité théorique* des données
 - 4.1 Les points fixes linéaires présentent des caractéristiques interprétables
 - 4.2 Nous devons passer de l'hypothèse distributionnelle à l'*hypothèse structuraliste*
- 5 L'IA n'est rien d'autre que des *sciences humaines* déguisées

Technologies de l'esprit. Machines à co-habiter
LLCP - Université Paris 8 | La Générale
Paris, France

*Ce que les mathématiques peuvent apporter
à la critique des LLMs*

Éléments pour un formalisme critique

Juan Luis Gastaldi

www.giannigastaldi.com

ETH zürich

11 mai, 2026

- 1 *Nous avons besoin d'un *formalisme critique*
- 2 *Un formalisme critique habilite une *critique épistémologique* de l'IA
- 3 *Cette critique épistémologique offre les fondements d'une *explicabilité formelle* des LLMs
- 4 L'explicabilité formelle ouvre la voie vers une *interprétabilité théorique* des données
- 5 L'IA n'est rien d'autre que des *sciences humaines* déguisées