Université de Montréal - Université Côte d'Azur
Colloque Créativités Artificielles
Nice, France

Critique and Formalism

The Role of Formal Sciences for an Internalist Critique of LLMs

Juan Luis Gastaldi

www.giannigastaldi.com

ETH zürich

April 28, 2026

Outline

Introduction: Critique and Formalism

Empiricism and Formalism in Machine Learning

Formal Explainability

Theoretical Interpretability

Conclusion: AI and Structuralism

Outline

Where Art Thou, Critique?

- 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."

Outline

An Empiricist Turn?

Interpretability as a Natural Science

The Structure of Scientific Revolutions by Thomas Kuhn [42] is a classic text on the history and sociology of science. In it, Kuhn distinguishes between "normal science" in which a scientific community has a paradigm, and "extraordinary science" in which a community lacks a paradigm, either because it never had one or because it was weakened by crisis. It's worth noting that "extraordinary science" is not a desirable state: it's a period where researchers struggle to be productive.

Kuhn's description of pre-paradigmatic fields feel eerily reminiscent of interpretability today [9] There isn't consensus on what the objects of study are, what methods we should use to answer them, or how to evaluate research results. To quote a recent interview with Ian Goodfellow: "For interpretability, I don't think we even have the right definitions." [43]

One particularly chal
shared sense of how
for dealing with this,
especially those with
which can evaluate h
HCI background may

But interpretability c
neural networks are
biology. Such work w
be held to the standa

(Olah et al., 2020)

The Empiricization of Computer Science

MANOEL HORTA RIBEIRO
DEC 17, 2025

Yet the nature of computer science has changed tremendously: it is now an empirical science, driven by observation and experiment as much as by theory and construction. If you don't believe me, spend 5 minutes going over the best paper awards for three conferences in different subfields in CS. If you did so in 2025, you might have found papers like "Characterizing and Detecting Propaganda-Spreading
ing Mental Health
s" (CSCW; Human
aching
scribe online user
da, people using
own examination of
al networks.

independently

(Horta Ribeiro, 2025)

Zico Kolter, *Building Safe and Robust AI Systems*, Keynote at ICLR 2025.

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

Empirical evaluation

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

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

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

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

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

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

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

$\ldots$

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

Empirical evaluation

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

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

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

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

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

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

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

$\ldots$

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

$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$

Empirical evaluation

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

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

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

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

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

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

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

$\ldots$

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

$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$

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

Empirical evaluation

$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$
$P' := \lambda r.\lambda s.\lambda f.\lambda x.f(f(f(f(fx))))$

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

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

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

Empirical interpretability

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

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

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

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

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

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

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

...

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

$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$

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

Empirical interpretability

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

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

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

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

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

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

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

$\ldots$

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

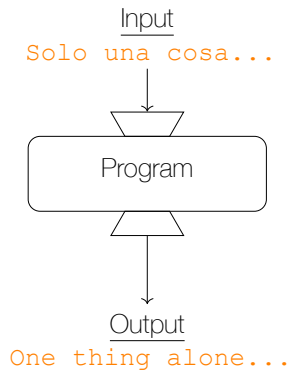
$$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda f.\lambda x.f(fx))(\lambda f.\lambda x.f(f(fx)))$$
$$\lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)(\lambda g.\lambda y.g(gy))(\lambda h.\lambda z.h(h(hz)))$$
$$\lambda n.\lambda f.\lambda x.(\lambda g.\lambda y.g(gy))f(nfx)(\lambda h.\lambda z.h(h(hz)))$$
$$\lambda n.\lambda f.\lambda x.(\lambda g.\lambda y.g(gy))f(nfx)(\lambda h.\lambda z.h(h(hz)))$$
$$\lambda f.\lambda x.(\lambda g.\lambda y.g(gy))f((\lambda h.\lambda z.h(h(hz)))fx)$$
$$\lambda f.\lambda x.(\lambda y.f(fy))((\lambda h.\lambda z.h(h(hz)))fx)$$
$$\lambda f.\lambda x.(\lambda y.f(fy))((\lambda z.f(f(fz)))x)$$
$$\lambda f.\lambda x.(\lambda y.f(fy))(f(f(fx)))$$
$$\lambda f.\lambda x.f(f(f(f(fx))))$$

Empirical interpretability

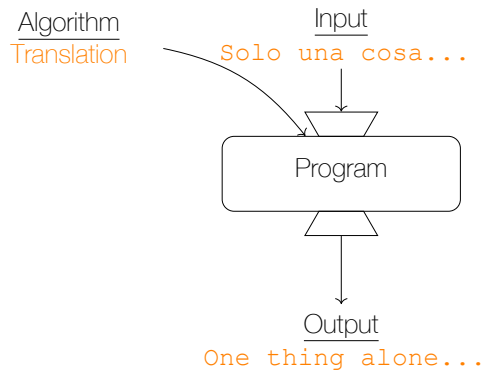
$P := \lambda m.\lambda n.\lambda f.\lambda x.mf(nfx)$

$P'' :=$ λRófÄÒêÑ5È|Àxñ=∞ù ÿmWf286ëy'SÒú>v&ìÂ ¬2 óÉ7öç∞{ã>2flB°µG#À9ÇU ∞btYBô¥ Ù ë‰3;5 å[l–èuô¨Ü ° 7–Ù.λ:^4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv –gÓÿ/ëiijO‡Œfi • J1«€ø˛Ï hãêt‡æY$^6 FïW»RÙKg¢¨.λ‡d¯… D2÷˛ò ° xêÈy. Ó”¢b Bé£NÈ1Ê‡/Û9Ñµ–/JYÇõË9ÿÀÈ.λÁÍ À^öÇ,»fq∞±îˇB5Ì>O˜g™“6Ωe“æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨.λÆà€fUòfEÙ·Í m#„˛4\r√¬÷Îpò»y*vtÄJÃF1ûÁóz«ñM”DjŒ BËèÍT _ Êa‰ÁÇΩ @\Ø^˜]Î›h‡.:^4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv–gÓÿ /ëiijO‡Œfi • J1«€ø˛Ï hãêt‡æY$^6 FïW»RÙKg¢¨ÁÍ À^öÇ,»fq∞±îˇB5Ì>O˜g™“6 Ωe“æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨(‡d¯… D2÷˛ò ° xêÈy. Ó”¢bBé£NÈ1Ê‡/Û9Ñµ–/ JYÇõË9ÿÀÈÁÍ À^öÇ,»fq∞±îˇB5Ì>O˜g™“6Ωe“æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨Æ à€fUòfEÙ·Í m#„˛4\r√¬÷Îpò»y*vtÄJÃF1ûÁóz«ñM”DjŒBËèÍT _ Êa‰ÁÇΩ @\ Ø^˜]Î›h‡)(ËÍÛeí4WµÍ }w„$Ω“K5èÃ¶%3[m ’˜BÃfifÒ; óJçCËîõŸØcB,ñ$Åå}ÓÅØ3¡ ˇ?ö¯oŒ@ fl8ˇR CÆoˇ* &<Ÿ ¬o12À‰ àÓÜ#ï ”,ú¨ «ô˛„∞Îâä“øÃd|^Ñ’Ê yØ;^W ›″wò[}\»ÖÈúw’6<ùˇ=ãÕ ¯Ì‘Dz?2±|é‘ 3Â/rxµ∞µ$ÀëÂ*lf˜îû’+ÍVìy[a]Gæßägô/,uÑ)

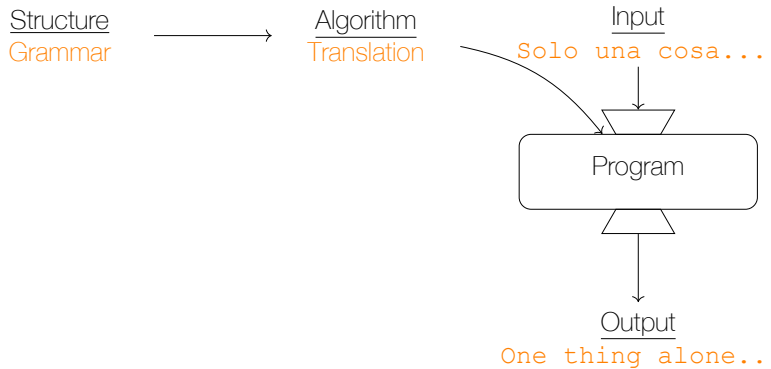
Implicit Structure



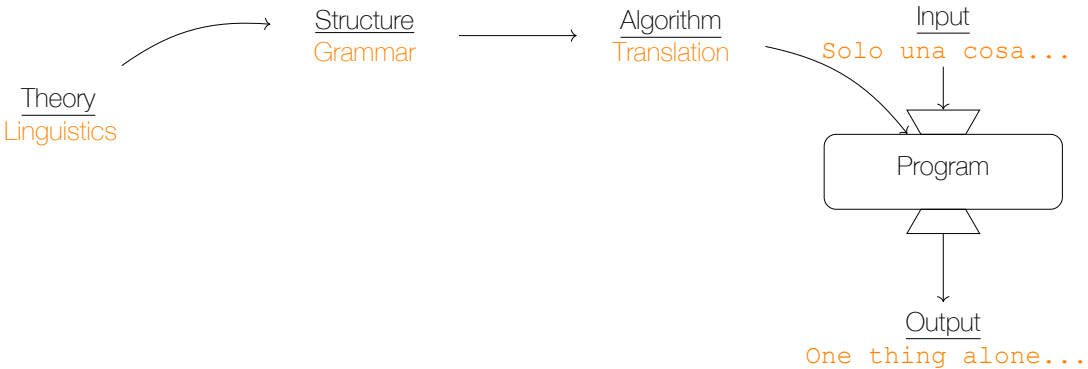
Implicit Structure



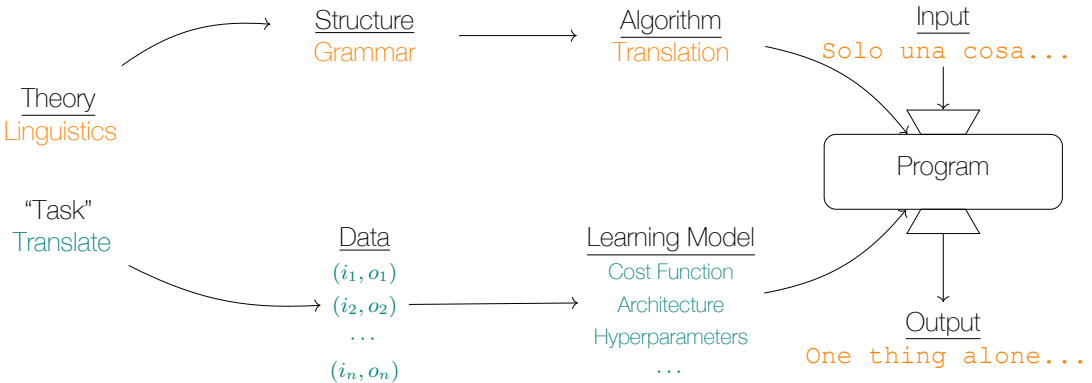
Implicit Structure



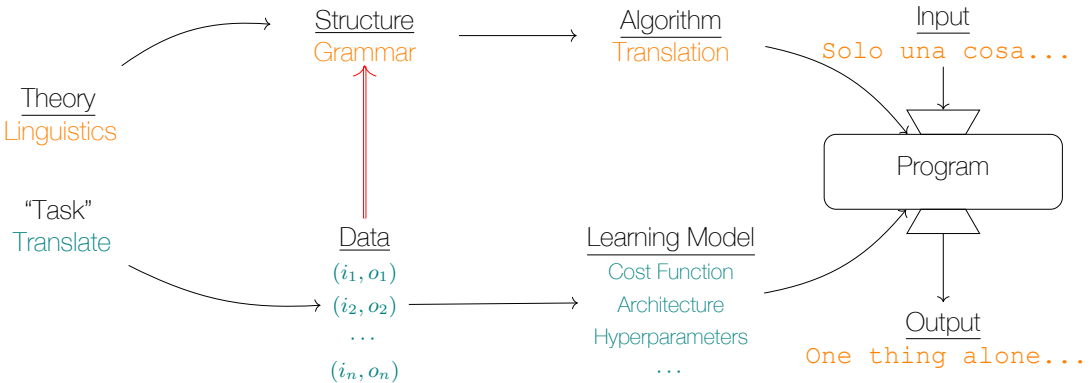
Implicit Structure



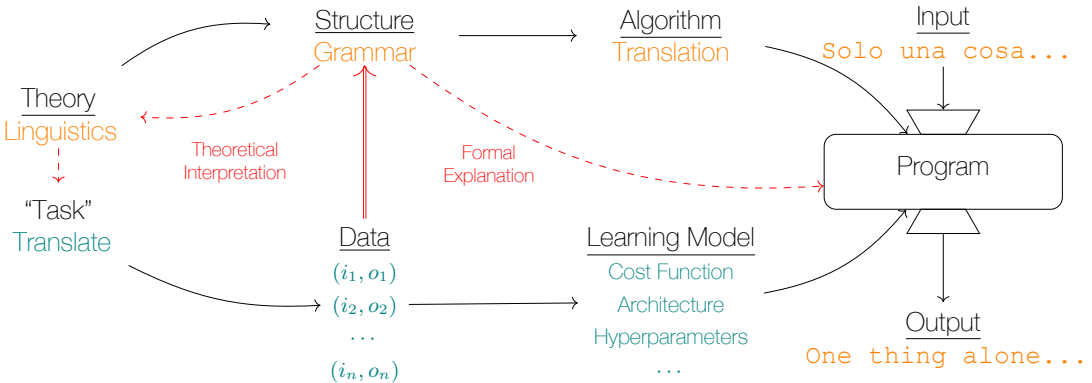
Implicit Structure



Implicit Structure



Making It Explicit



Outline

Embeddings in LLMs

Epistemology of Machine Learning
Distributional Language Models

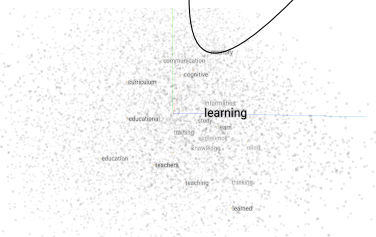
Tokenization
(Sennrich et al., 2016)

Embedding
(Mikolov et al., 2013)

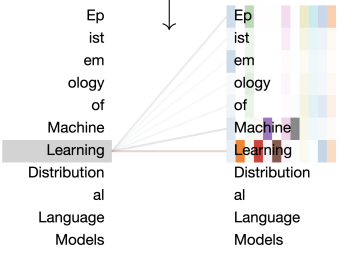
Attention
(Vaswani et al., 2017)

Epistemology of Machine Learning
Distributional Language Models

(https://tiktokenizer.vercel.app)

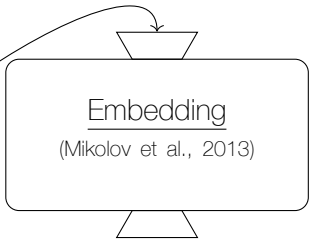


(https://projector.tensorflow.org)



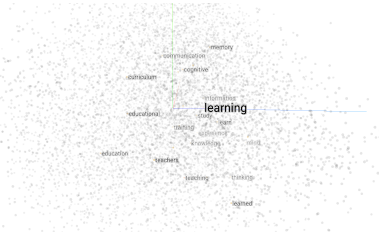
(https://github.com/jessevig/bertviz)

Embeddings in LLMs



Epistemology of Machine Learning
Distributional Language Models

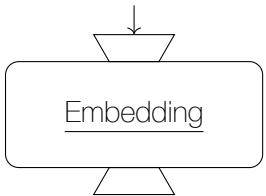
(https://tiktokenizer.vercel.app)



(https://projector.tensorflow.org)

Neural Embeddings

Epistemology of Machine Learning
Distributional Language Models



Neural Embeddings

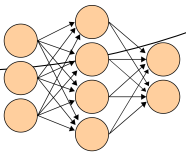
Structure

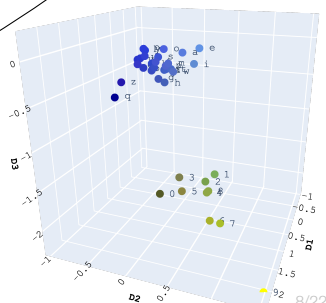
?

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

Embedding

Data





word2vec Explained

(Levy & Goldberg, 2014)

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

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

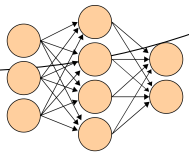
- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.
- The Singular Value Decomposition (SVD) provides an exact solution to this problem.

Embedding structure

Structure

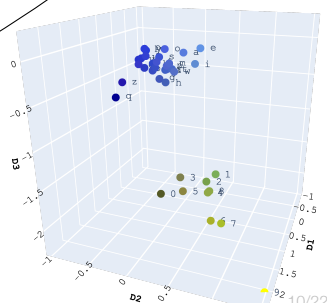
?

Data



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

Embedding



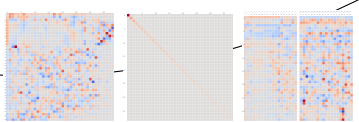
Embedding structure

Structure

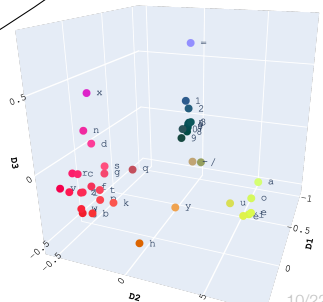
{-, /, 0, 1, 2, . . . , 8, 9, =,
a, b, c, . . . , w, x, y, z, é}

Embedding

Data

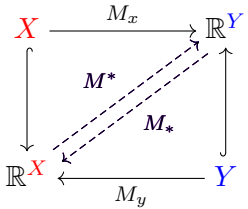


SVD



From matrices to distributional operators

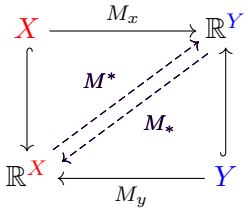
$$M \colon X \times Y \to \mathbb{R}$$
$$(x, y) \mapsto \mathrm{pmi}(x, y)$$



$$M_* M^* \colon \mathbb{R}^X \to \mathbb{R}^X$$
$$M^* M_* \colon \mathbb{R}^Y \to \mathbb{R}^Y$$

From matrices to distributional operators

$$M \colon X \times Y \to \mathbb{R}$$
$$(x, y) \mapsto \mathrm{pmi}(x, y)$$



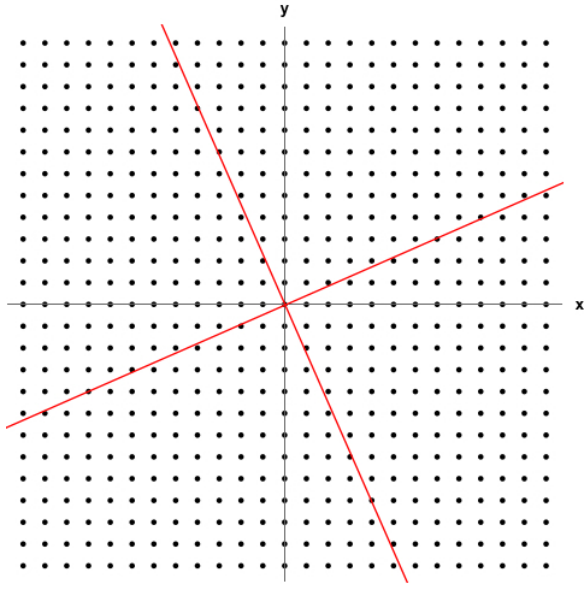
$$M_* M^* \colon \mathbb{R}^X \to \mathbb{R}^X$$
$$M^* M_* \colon \mathbb{R}^Y \to \mathbb{R}^Y$$

$$\{u_1, \ldots, u_m\} \subset \mathbb{R}^X$$
$$\{v_1, \ldots, v_n\} \subset \mathbb{R}^Y$$
$$\{\lambda_1, \ldots, \lambda_{\min(m,n)}, 0, \ldots, 0\}$$

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

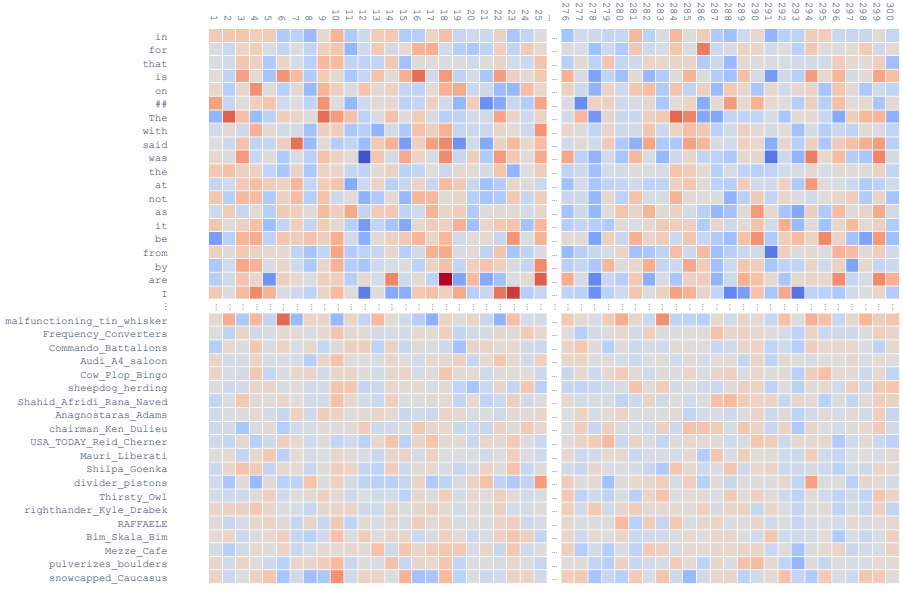
The u_i and v_i are (linear) fixed points!

Eigenvectors

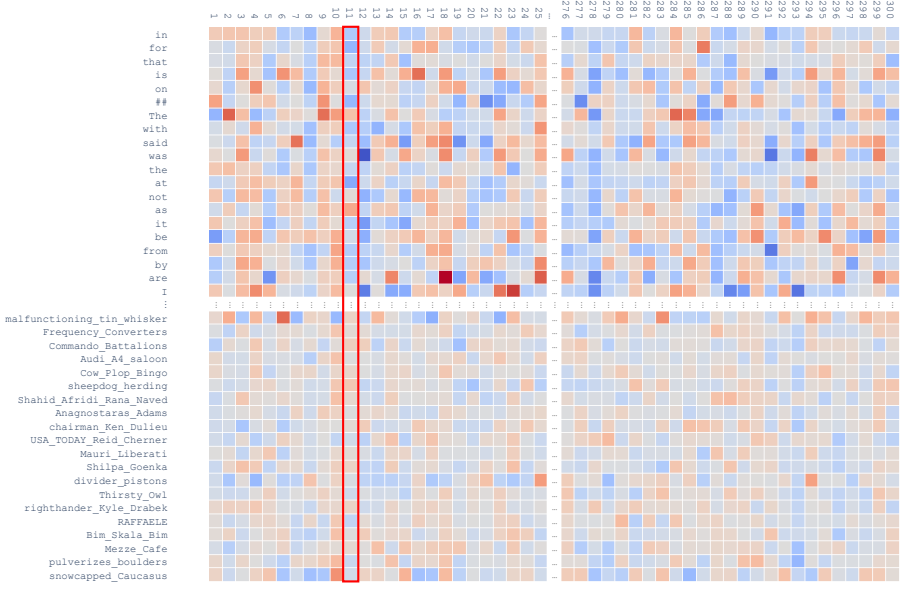


Embedding dimensions as fixed points

Embedding dimensions as fixed points

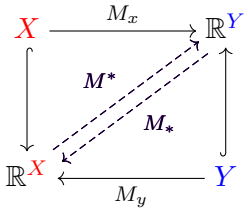


Embedding dimensions as fixed points



Embeddings as Functors Over Categories

$$M : X \times Y \to \mathbb{R}$$
$$(x, y) \mapsto \text{pmi}(x, y)$$



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

$$\{u_1, \ldots, u_m\} \subset \mathbb{R}^X$$
$$\{v_1, \ldots, v_n\} \subset \mathbb{R}^Y$$
$$\{\lambda_1, \ldots, \lambda_{\min(m,n)}, 0, \ldots, 0\}$$

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

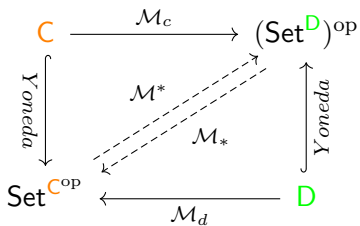
The u_i and v_i are (linear) fixed points!

Embeddings as Functors Over Categories

$$\mathsf{C}^{\mathrm{op}} \times \mathsf{D} \to \mathsf{Set}$$

$$\Downarrow$$

$$\mathcal{M}^* \colon \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \colon \mathcal{M}_*$$



$$\mathcal{M}_*\mathcal{M}^* \colon \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}}$$

$$\mathcal{M}^*\mathcal{M}_* \colon (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \to (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}}$$

$$\mathrm{Fix}(\mathcal{M}_*\mathcal{M}^*) := \{f \in \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} | \mathcal{M}_*\mathcal{M}^*(f) \cong f\}$$

$$\mathrm{Fix}(\mathcal{M}^*\mathcal{M}_*) := \{g \in (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} | \mathcal{M}^*\mathcal{M}_*(g) \cong g\}$$

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

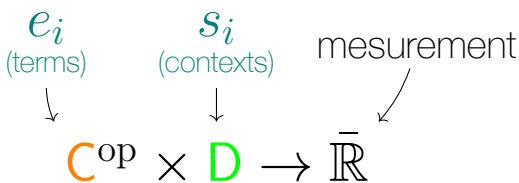
$\mathcal{M}^* f_i \cong g_i$ and $\mathcal{M}_* g_i \cong f_i$

The nucleus is a category complete and cocomplete

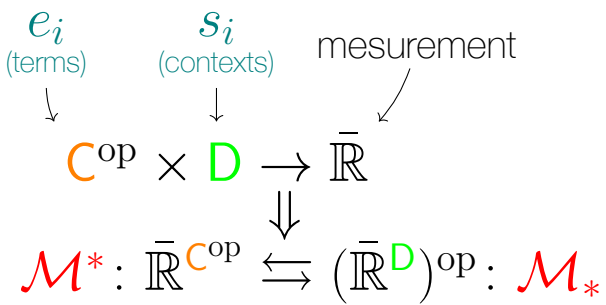
Categories C and D can be enriched!

E.g.: $\mathcal{M}^* \colon \bar{\mathbb{R}}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (\bar{\mathbb{R}}^{\mathsf{D}})^{\mathrm{op}} \colon \mathcal{M}_*$

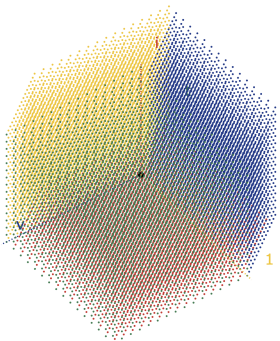
The Nucleus of the Profunctor



The Nucleus of the Profunctor

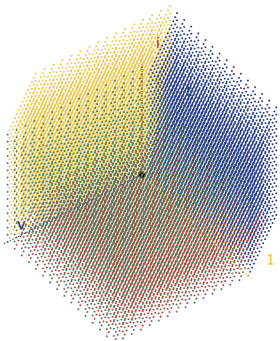


The Nucleus of the Profunctor

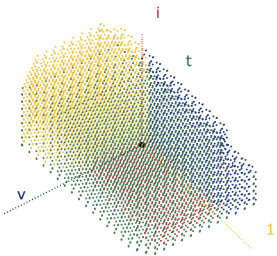


$$\begin{array}{ccc} \mathsf{C}^{\mathrm{op}} \times \mathsf{D} & \to & \bar{\mathbb{R}} \\ & \Downarrow & \\ \mathcal{M}^* \colon \bar{\mathbb{R}}^{\mathsf{C}^{\mathrm{op}}} & \leftrightarrows & \left(\bar{\mathbb{R}}^{\mathsf{D}}\right)^{\mathrm{op}} \colon \mathcal{M}_* \end{array}$$

The Nucleus of the Profunctor

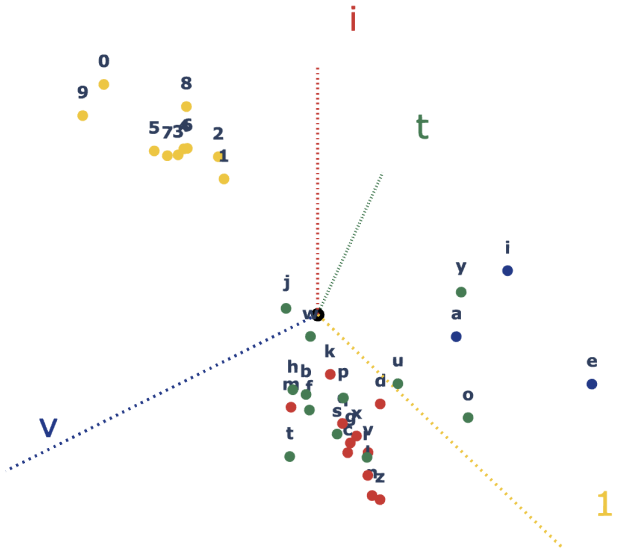


$$\mathcal{M}_*\mathcal{M}^*$$

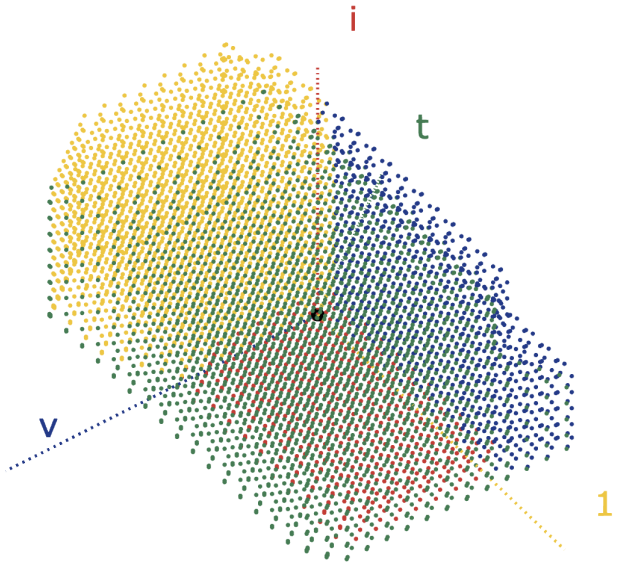


$$\begin{array}{ccc} \mathsf{C}^{\mathrm{op}} \times \mathsf{D} & \to & \bar{\mathbb{R}} \\ & \Downarrow & \\ \mathcal{M}^* \colon \bar{\mathbb{R}}^{\mathsf{C}^{\mathrm{op}}} & \leftrightarrows & \left(\bar{\mathbb{R}}^{\mathsf{D}}\right)^{\mathrm{op}} \colon \mathcal{M}_* \end{array}$$

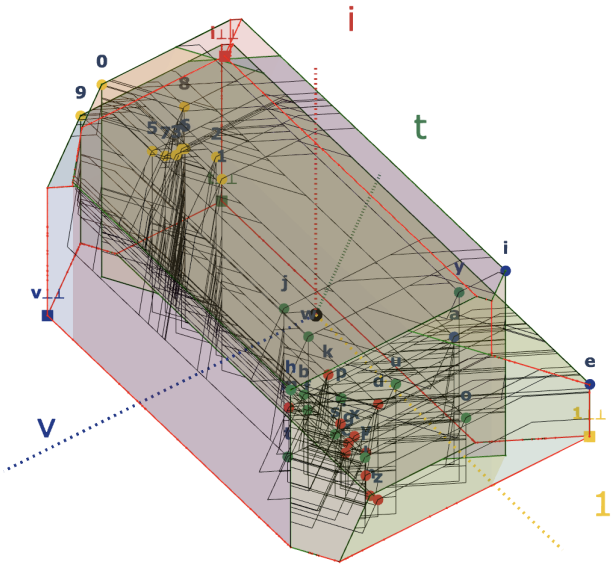
Geometry of the Nucleus



Geometry of the Nucleus



Geometry of the Nucleus



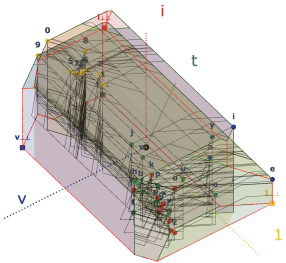
Theory of Computational Types

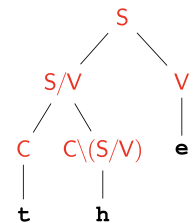
Definition (Polar/Orthogonal - Girard, 2011)

[G]iven a binary operation, noted $a, b \rightsquigarrow \langle a|b\rangle : A \times B \to C$ 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 $\mathrm{Pol}(A) \subset \mathcal{P}(A)$ 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, $X^{\perp\perp\perp} = X^\perp$.





(Gastaldi and Pellissier, 2021)

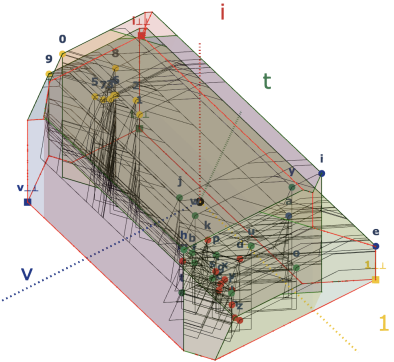
Outline

From the Distributional to the Structuralist Hypothesis

Theory
"Task"

?

Structure



From the Distributional to the Structuralist Hypothesis

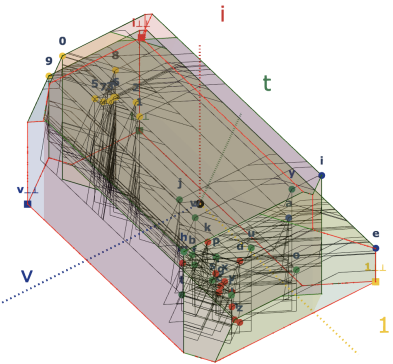
$$C^{op} \times D \rightarrow \bar{\mathbb{R}}$$

Distributional Hypothesis

The content of linguistic units is determined by their *distribution* in a corpus.

Theory
"Task"

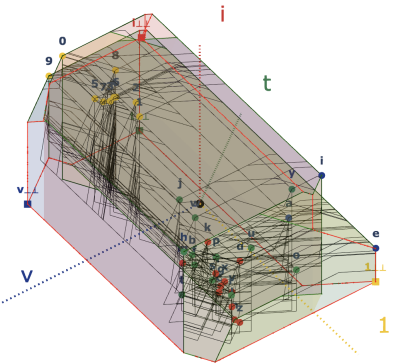
Structure



From the Distributional to the Structuralist Hypothesis

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

Structure



Distributional Hypothesis

The content of linguistic units is determined by their ***distribution*** in a corpus.

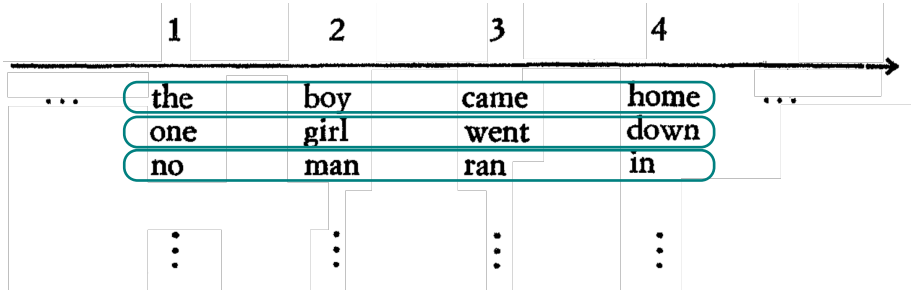
Theory
"Task"

Structuralist Hypothesis

Linguistic content is the effect of a virtual ***structure*** underlying linguistic practices within a community

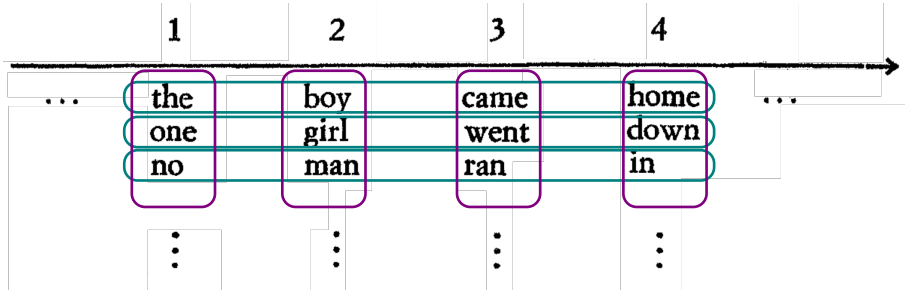
$$\bar{\mathbb{R}}^{C^{\mathrm{op}}} \leftrightarrows (\bar{\mathbb{R}}^{D})^{\mathrm{op}}$$

Syntagmas and Paradigmes



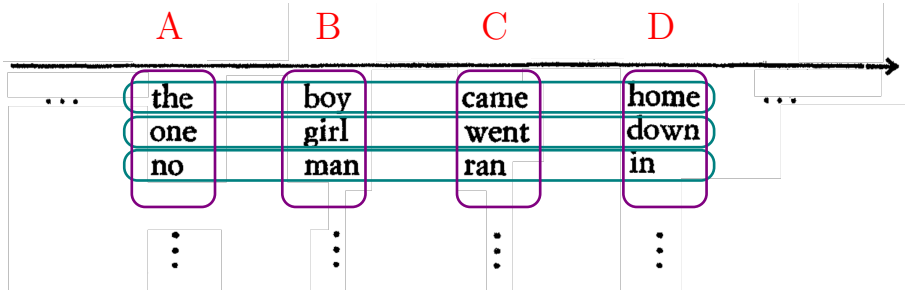
(Hjelmslev, 1971a)

Syntagmas and Paradigmes



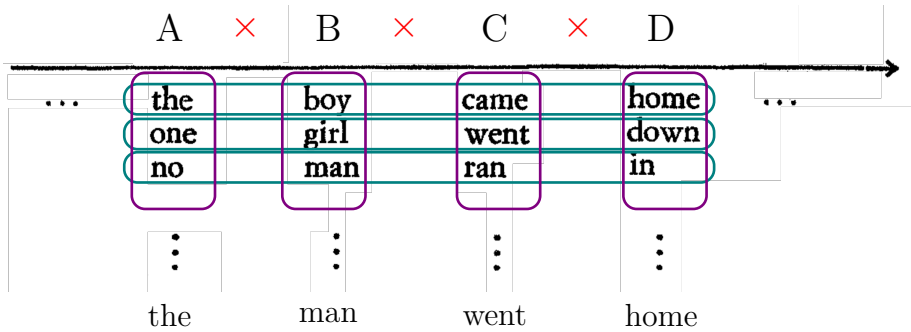
(Hjelmslev, 1971a)

Syntagmas and Paradigmes



(Hjelmslev, 1971a)

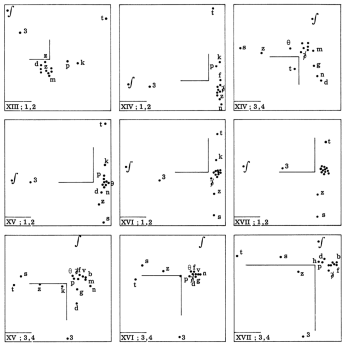
Syntagmas and Paradigmes



(Hjelmslev, 1971a)

Outline

Structuralist Formalism



v. p.	1°	2°	3°	4°	5°
XII	,99	,98	,97	,92	,86
XV	,71	,41	,37	,37	,19

v. p.	1°	2°	3°	4°	5°
XIII	,94	,92	,86	,83	,75
XVI	,55	,29	,25	,19	,04

v. p.	1°	2°	3°	4°	5°
XIV	,87	,62	,49	,47	,45
XVII	,44	,23	,17	,09	,02

(benzécri1976codage)

1) une voyelle neutre (amorphe), caractérisée par l'absence de chacune des propriétés β, φ, χ, λ : $[\circ]$;

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

$$[\varepsilon] = \varphi\,,\ [\mathrm{a}] = \beta\,,\ [\mathrm{y}] = \chi\,,\ [\mathrm{\alpha}] = \lambda$$

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

$$[\theta] = \varphi\beta\,,\ [\mathrm{i}] = \chi\varphi\,,\ [\mathrm{u}] = \chi\beta\,,\ [\mathrm{e}] = \lambda\varphi\,,\ [\mathrm{o}] = \lambda\beta\,,\ [\mathrm{\ddot{e}}] = \chi\lambda\,;$$

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

$$[\mathrm{\varepsilon}] = \chi\lambda\varphi\,,\ [\mathrm{\varrho}] = \chi\lambda\beta\,,\ [\mathrm{\ddot{a}}] = \varphi\beta\chi\,,\ [\mathrm{\delta}] = \varphi\beta\lambda\,;$$

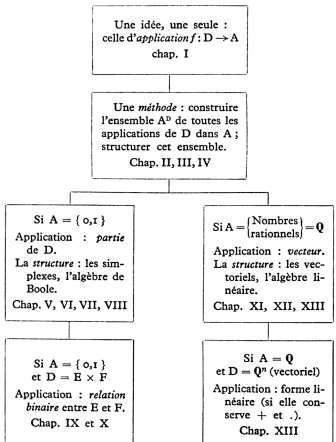
5) une voyelle polymorphe, caractérisée par les quatre propriétés considérées : la voyelle russe $[\mathrm{\ddot{\imath}}] = \varphi\beta\chi\lambda$.

On obtient alors le diagramme de la figure 2.

Fig. 2.

(Marcus, 1967)

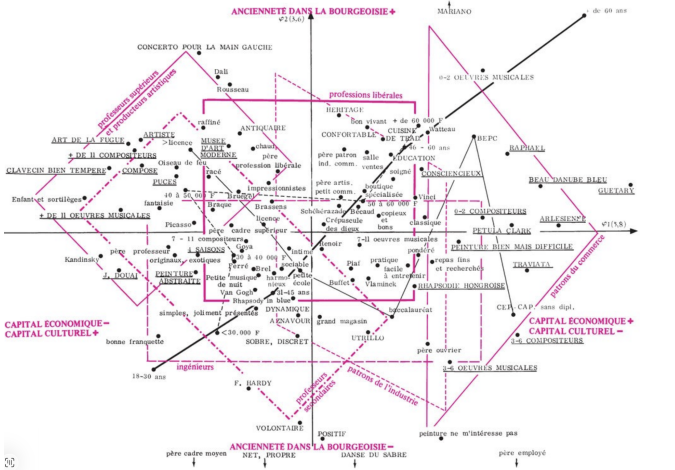
Organigramme



(Barbut1967_T1)

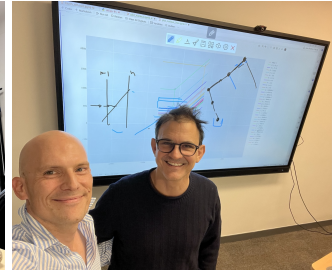
Structuralist Social Sciences

"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)

Collaborations



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

Reference Papers

- 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. (2021).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*

- Gastaldi, J. L., Jarvis, S., Seiller, T., & Terilla, J. (2026, January). *Geometric Structures in $\mathbb{R}$-enriched adjunctions* [preprint under review]. https://hal.science/hal-05452748

References I

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021).On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bourdieu, P. (1979). *La distinction: Critique sociale du jugement*. Éditions de Minuit.

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*.

Chomsky, N. (1992, November). Language and the "cognitive revolutions" [Delivered November 23–27, 1992].

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., Jarvis, S., Seiller, T., & Terilla, J. (2026, January). *Geometric Structures in $\mathbb{R}$-enriched adjunctions* [preprint under review]. https://hal.science/hal-05452748

Gastaldi, J. L., & Pellissier, L. (2021).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

Girard, J.-Y. (2011, September). *The blind spot*. European Mathematical Society.

Hjelmslev, L. (1971a). La structure fondamentale du langage. In *Prolégomènes à une théorie du langage* [Prolégomènes à une theorie du langage] (pp. 177–231). Éditions de Minuit.

Hjelmslev, L. (1971b). *Prolégomènes à une théorie du langage*. Éditions de Minuit.

Kirschenbaum, M. (2023). *Again theory: A forum on language, meaning, and intent in the time of stochastic parrot*. https://critinq.wordpress.com/2023/06/26/again-theory-a-forum-on-language-meaning-and-intent-in-the-time-of-stochastic-parrots/

References II

Levy, O., & Goldberg, Y. (2014).Neural word embedding as implicit matrix factorization. *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, 2177–2185.

Marcus, S. (1967). *Introduction mathématique à la linguistique structurale*. Dunod.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013).Distributed representations of words and phrases and their compositionality. *CoRR*, *abs/1310.4546*.

Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S. (2020).Zoom in: An introduction to circuits [https://distill.pub/2020/circuits/zoom-in]. *Distill*. https://doi.org/10.23915/distill.00024.001

Sennrich, R., Haddow, B., & Birch, A. (2016).Neural machine translation of rare words with subword units. *Proceedings of the 54th Annual Meeting of the ACL*, 1715–1725.

Underwood, T. (2023, October 15). *The empirical triumph of theory* [Accessed: 2023-10-15]. https://critinq.wordpress.com/2023/06/29/the-empirical-triumph-of-theory/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Université de Montréal - Université Côte d'Azur
Colloque Créativités Artificielles
Nice, France

Critique and Formalism

The Role of Formal Sciences for an Internalist Critique of LLMs

Juan Luis Gastaldi

www.giannigastaldi.com

ETH zürich

April 28, 2026