Vernacular AI
Digital Theory Lab, NYU
New York, NY, USA

The Structure, Not the Prompt

For a Critical Formalism

Juan Luis Gastaldi

ETH zürich

February 7, 2025

Outline

Outline

Introduction

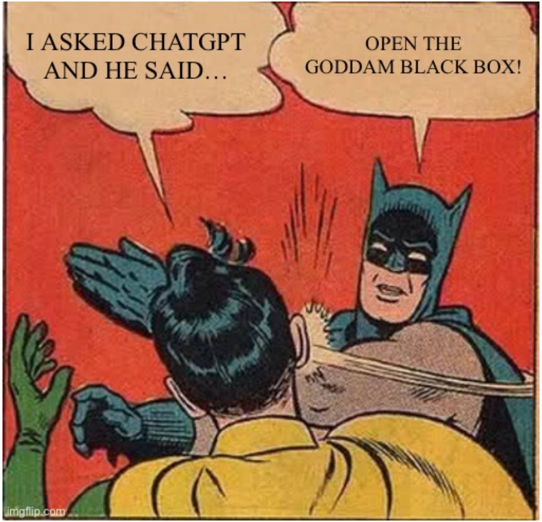
NLMs as Formal Objects

The Structure(s) of the Embeddings

- The Algebra Behind the Embeddings
- The Structure Behind the Algebra
- The Categories Behind the Structure

Conclusion

...Not the Prompt



Where Art Thou, Critique?

Where Art Thou, Critique?

- Good "externalist" critique

- Poor "internalist" critique
  - The main "critical" reference remains the "Stochastic Parrots" approach (Bender & Koller, 2020; Bender et al., 2021)

  - 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."

The Critical Argumentative Matrix

Knowledge depends on language

↓

The relation between language and the world is essentially arbitrary

↓

Any regularity in language/knowledge is not natural but cultural/social/political

↓

We should resist existing regularities and create new ones

The Critical Argumentative Matrix

Knowledge depends on language
(Epistemological)

↓

The relation between language and the world is essentially arbitrary

↓

Any regularity in language/knowledge is not natural but cultural/social/political
(Political)

↓

We should resist existing regularities and create new ones
(Aesthetic)

The Critical Argumentative Matrix

Knowledge depends on language
(Epistemological)

[The relation between language and the world is essentially arbitrary?]

Any regularity in language/knowledge is not natural but cultural/social/political
(Political)

↓

We should resist existing regularities and create new ones
(Aesthetic)

Critique and Formalism

- At the source of this situation is the new foundational role played by formal sciences in the 20th century
  - For a theory of language: Carnap, Gödel, Turing, Shannon, Harris, Chomsky...

Critique and Formalism

- At the source of this situation is the new foundational role played by formal sciences in the 20th century
  - For a theory of language: Carnap, Gödel, Turing, Shannon, Harris, Chomsky...
- The critical tradition has either withdrawn from the areas conquered by formal approaches, or made formal approaches the target of criticism

Critique and Formalism

- At the source of this situation is the new foundational role played by formal sciences in the 20th century
  – For a theory of language: Carnap, Gödel, Turing, Shannon, Harris, Chomsky...
- The critical tradition has either withdrawn from the areas conquered by formal approaches, or made formal approaches the target of criticism
- We need a new strategy: Elaborate a critical formalism

Critique and Formalism

- At the source of this situation is the new foundational role played by formal sciences in the 20th century
  - For a theory of language: Carnap, Gödel, Turing, Shannon, Harris, Chomsky...
- The critical tradition has either withdrawn from the areas conquered by formal approaches, or made formal approaches the target of criticism
- We need a new strategy: Elaborate a critical formalism
- In the case of AI, a critical formalism can provide:
  - New epistemological tools countering dogmatic perspectives stemming from within the field
  - New theoretical tools contributing to the non-dogmatic positive production of knowledge

Outline

Introduction

NLMs as Formal Objects

The Structure(s) of the Embeddings

- The Algebra Behind the Embeddings
- The Structure Behind the Algebra
- The Categories Behind the Structure

Conclusion

Neural LMs as Computable Functions

Neural LM

$\downarrow$

?

Neural LMs as Computable Functions

Neural LM

↓

🧠 ❓

Neural LMs as Computable Functions

Neural LM

↓

🤖 ❓

Neural LMs as Computable Functions

Neural LM

↓

🦜❓

Neural LMs as Computable Functions

Neural LM

Neural LMs as Computable Functions

Neural LM

Neural LMs as Computable Functions

$$\begin{array}{c}\text{Neural LM}\\ \downarrow \\ f\ \color{red}{!}\end{array}$$

Neural LMs as Computable Functions

Neural LM

$$\downarrow$$

Function

$$f : \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

Neural LMs as Computable Functions

Neural LM

$\downarrow$

Computable

Function

$$f : \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

Neural LMs as Computable Functions

Neural LM

$\downarrow$

Computable
Function

$$f : \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

$\downarrow$

Turing
Machine

(Turing, 1937)

Neural LMs as Computable Functions

Neural LM

$\downarrow$

Computable

Function

$$f : \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

Recursive Function	Turing Machine	Lambda Term
(Gödel, 1986 (1934))	(Turing, 1937)	(Church, 1936)

Neural LMs as Computable Functions

Neural LM

$\downarrow$

Computable

Function

$$f : \mathbb{R}^{n} \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \dots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^{m}$$

Recursive Function (Gödel, 1986 (1934)) $\Longleftrightarrow$ Turing Machine (Turing, 1937) $\Longleftrightarrow$ Lambda Term (Church, 1936)

Neural LMs as Computable Functions

Neural LM

$\downarrow$

Computable Function

$$f : \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

Recursive Function (Gödel, 1986 (1934))	$\Longleftrightarrow$	Turing Machine (Turing, 1937)	$\Longleftrightarrow$	Lambda Term (Church, 1936)

Neural LMs as Computable Functions

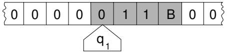
Neural LM

$\downarrow$

Computable Function

$$f : \mathbb{R}^n \xrightarrow{f_1} \mathbb{R}^{n_1} \xrightarrow{f_2} \cdots \xrightarrow{f_K} \mathbb{R}^{n_K} \xrightarrow{g} \mathbb{R}^m$$

Recursive Function	$\Longleftrightarrow$	Turing Machine	$\Longleftrightarrow$	Lambda Term
(Gödel, 1986 (1934))		(Turing, 1937)		(Church, 1936)



credit: Nynexman4464

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

Empirical Evaluation

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

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

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

$\dots$

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

Empirical Evaluation

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

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

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

Empirical Evaluation

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

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

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

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

$$\rightsquigarrow$$

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

Empirical Evaluation

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

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

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

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

$\dots$

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

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

$$\rightsquigarrow \rightsquigarrow \rightsquigarrow \rightsquigarrow \rightsquigarrow \rightsquigarrow \rightsquigarrow$$

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

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

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

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

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:ˆ4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv
–gÓÿ/ëiijO‡Œﬁ • J1«€ø¸Ï hãêt‡æY$ˆ6 FïW»RÙKg¢¨.λ‡d¯… D2÷˛ò ˚ xêÈy. Ó"¢b
Bé£NÈ1Ê‡/Û9Ñµ–/JYÇõË9ÿÀÈ.λÁÍ ÀˆöÇ,»fq∞±î˘B5Ì>O˜g™"6Ωe"æëC/ã… Ö
· fÓ Å]ÑåyÊN°Ê ›¨.λÆà€fUòfEÙ˙Í m#„¸4\r√¬÷Îpò»y*vtÄJÃF1ûÁóz«ñM"DjŒ
BËèÍT _ Êa‰ÁÇΩ @\Øˆ˜]Î›h‡.:ˆ4mÓØ¥´è—+ÌsÖ,$+gï„B™÷o–#ïŸê Ûv–gÓÿ
/ëiijO‡Œﬁ • J1«€ø¸Ï hãêt‡æY$ˆ6 FïW»RÙKg¢¨ÁÍ ÀˆöÇ,»fq∞±î˘B5Ì>O˜g™"6
Ωe"æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨(‡d¯… D2÷˛ò ˚ xêÈy. Ó"¢bBé£NÈ1Ê‡/Û9Ñµ–/
JYÇõË9ÿÀÈÁÍ ÀˆöÇ,»fq∞±î˘B5Ì>O˜g™"6Ωe"æëC/ã… Ö · fÓ Å]ÑåyÊN°Ê ›¨Æ
à€fUòfEÙ˙Í m#„¸4\r√¬÷Îpò»y*vtÄJÃF1ûÁóz«ñM"DjŒBËèÍT _ Êa‰ÁÇΩ @\
Øˆ˜]Î›h‡)(ËÎÛeí4WµÍ }w„$Ω"K5èÃ¶%3[m '˘BÃﬁfÒ; óJçCËîõŸØcB,ñ$Ãå}ÓÅØ3¡
˘?ö¯oŒ@ ﬂ8˘R CÆo˘* &<Ÿ ¬o12À‰ àÓÜ#ï ˝,ú¨ «ô¸„∞Îâä"øÃd|ˆÑ'Ê yØ;ˆW
›˝wò[}\»ÖÈúw'6<ù˘=ãÕ ¯Ì'Dz?2±|é' 3Â/rxµ∞µ$ÀëÂ*lf˜îû'+ÍVìyªGæßägô/,uÑ)

Making it Explicit

$$\frac{\text{Theory}}{\text{Informal Specification}}$$

Making it Explicit

$$\frac{\text{Theory}}{\text{Informal Specification}} \longrightarrow \frac{\text{Formal System}}{\text{Formal Specification}}$$

Making it Explicit

$$\frac{\text{Theory}}{\text{Informal Specification}} \longrightarrow \frac{\text{Formal System}}{\text{Formal Specification}} \xrightarrow{\text{Implementation}} \underline{\text{Function}}$$

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

Making it Explicit

Formal
Specification

"Task"
Informal
Specification

Data
Implicit
Specification

Function

λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy'SÒú>v&
. . .
3Â/rxμ∞μ$ÀëÂ*lf˜îû'
+ÍVìyªGæßägô/,uÑ

Making it Explicit

Structure

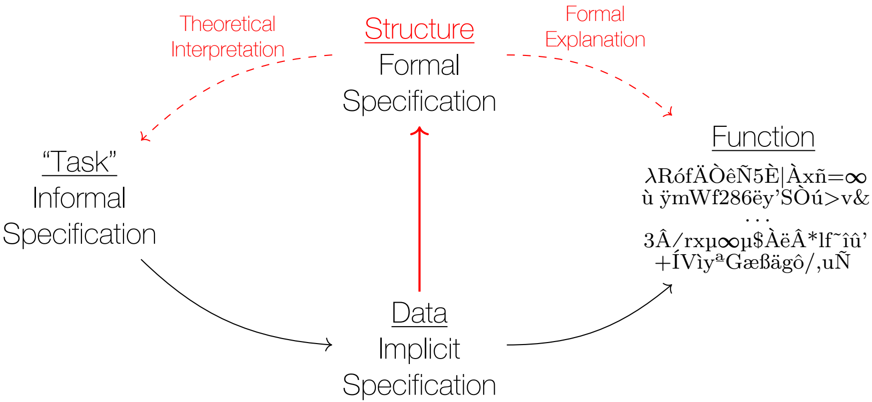
"Task"
Informal
Specification

Data
Implicit
Specification

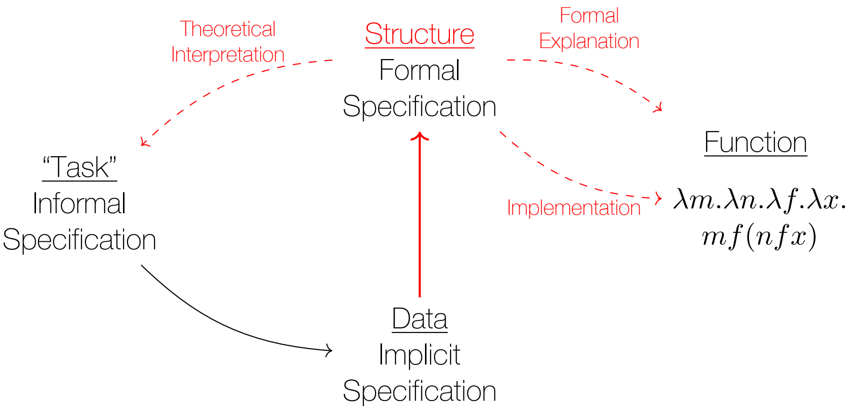
Function

λRófÄÒêÑ5È|Àxñ=∞
ù ÿmWf286ëy'SÒú>v&
...
3Â/rxµ∞µ$ÀëÂ*lf˜îû'
+ÍVìyªGæßägô/,uÑ

Making it Explicit



Making it Explicit



Outline

Introduction

NLMs as Formal Objects

The Structure(s) of the Embeddings

- The Algebra Behind the Embeddings
- The Structure Behind the Algebra
- The Categories Behind the Structure

Conclusion

Outline

Introduction

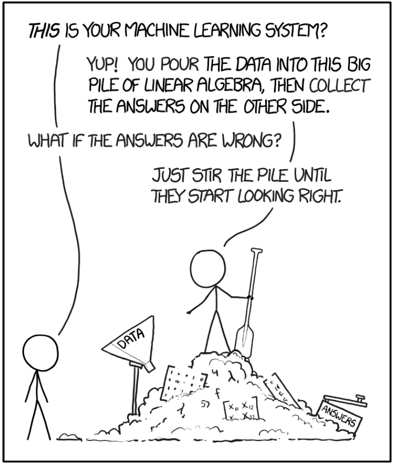
NLMs as Formal Objects

The Structure(s) of the Embeddings

- The Algebra Behind the Embeddings
- The Structure Behind the Algebra
- The Categories Behind the Structure

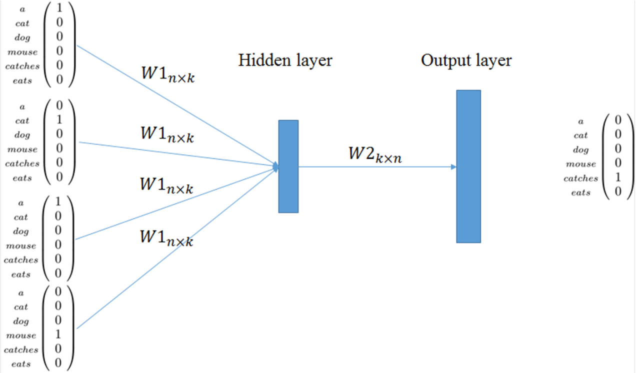
Conclusion

Machine-Learning the Embedding Space



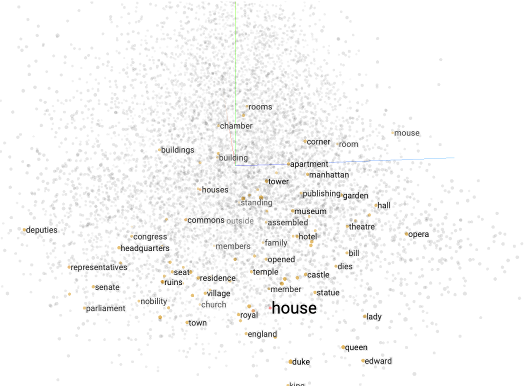
Credit: xkcd.com

a cat catches a mouse

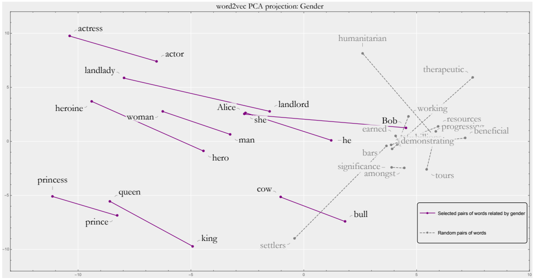


Credit: Ferrone et al., 2017

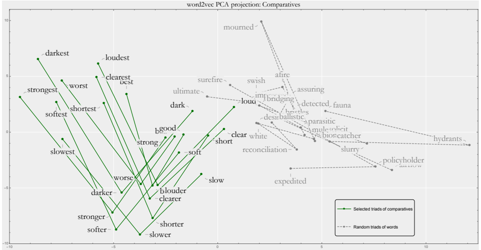
Embedding Space: Similarity and Analogy

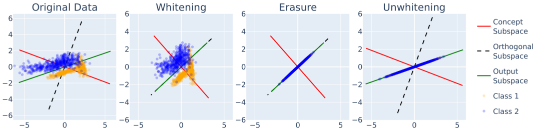


(https://projector.tensorflow.org)

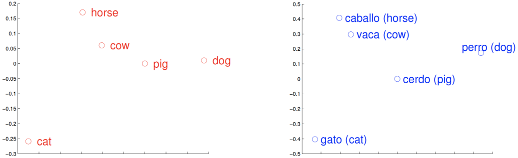


Embedding Space: Other Applications

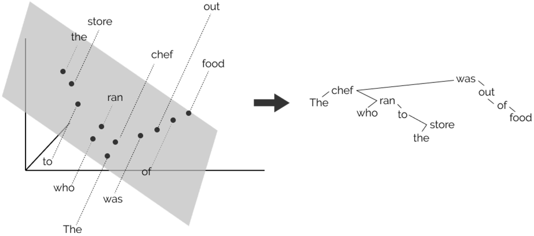




(Belrose et al., 2024)



(Mikolov et al., 2013)



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

word2vec Explained

(Levy and Goldberg, 2014)

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

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

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

word2vec Explained

(Levy and Goldberg, 2014)

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

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

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

word2vec Explained

(Levy and Goldberg, 2014)

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

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

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

word2vec Explained

(Levy and Goldberg, 2014)

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

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

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

word2vec Explained

(Levy and Goldberg, 2014)

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

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

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

word2vec Explained

(Levy and Goldberg, 2014)

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

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

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

word2vec Explained

(Levy and Goldberg, 2014)

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

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

- Word2vec performs an implicit, low-dimensional factorization of a pointwise mutual information (pmi), word-context matrix.

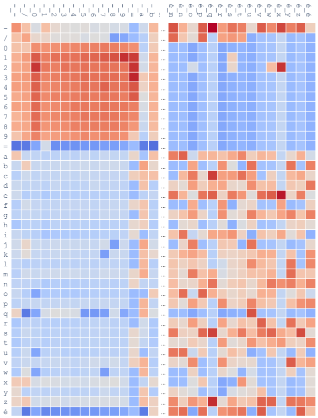
- The Singular Value Decomposition (SVD) provides an exact solution to this problem.

Example: Characters in Wikipedia

$$\textcolor{red}{W} = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

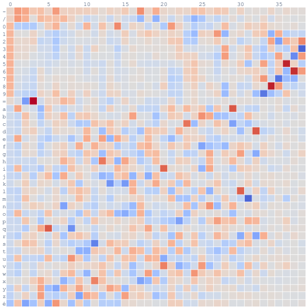
$$\textcolor{blue}{C} = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

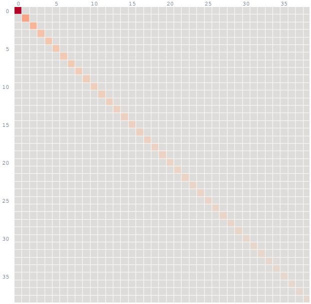
$$\begin{aligned} M_{\textcolor{red}{w}\textcolor{blue}{c}} &= \operatorname{pmi}(\textcolor{red}{w}, \textcolor{blue}{c}) \\ &= \log \frac{p(\textcolor{red}{w}, \textcolor{blue}{c})}{p(\textcolor{red}{w})p(\textcolor{blue}{c})} \end{aligned}$$

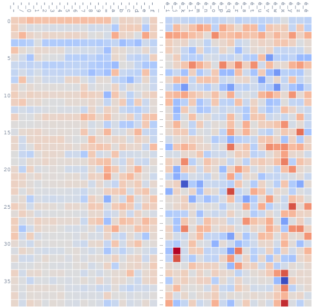


SVD of Wikipedia Character PMI Matrix

$$U \qquad \Sigma \qquad V^T$$

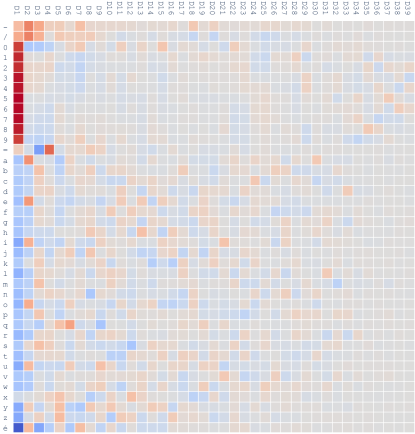






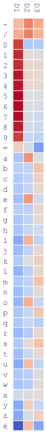
Truncate

$$U \times \Sigma$$



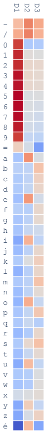
Truncate

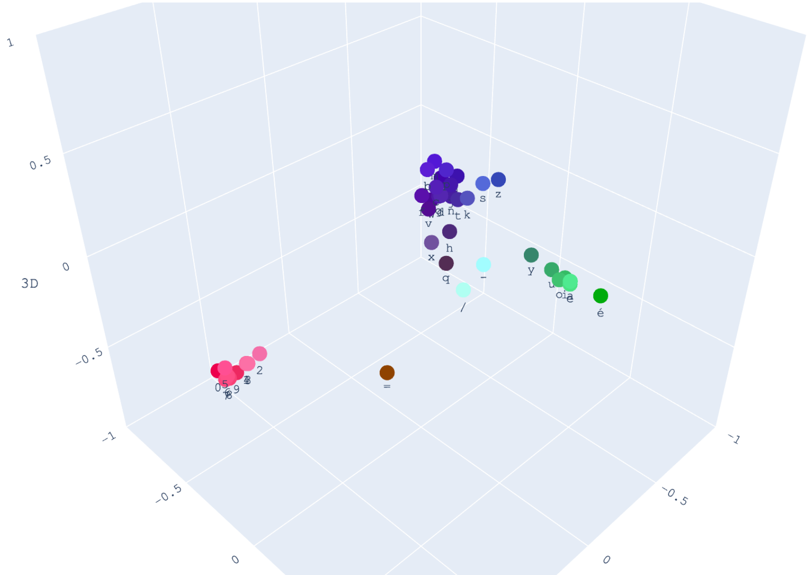
$$\hat{U} \times \hat{\Sigma}$$



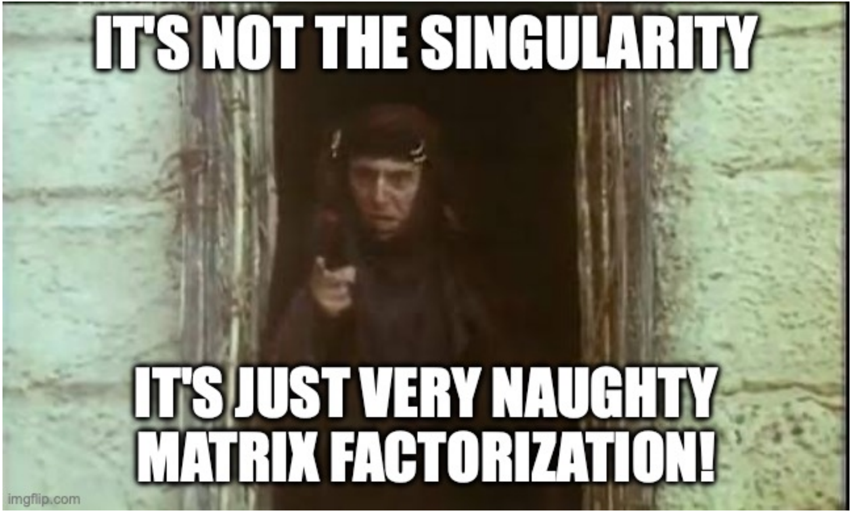
Plot

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





What to conclude?



But Why?

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)

Outline

Introduction

NLMs as Formal Objects

The Structure(s) of the Embeddings

- The Algebra Behind the Embeddings
- The Structure Behind the Algebra
- The Categories Behind the Structure

Conclusion

Embeddings as Functions Over Sets

$$\color{red}{X} = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$\color{blue}{Y} = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

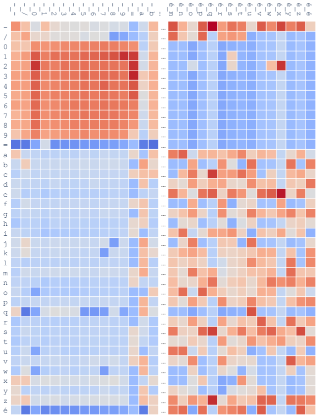
Embeddings as Functions Over Sets

$$\color{red}{X} = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$\color{blue}{Y} = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

$$M : \color{red}{X} \times \color{blue}{Y} \to \mathbb{R}$$

$$(\color{red}{x}, \color{blue}{y}) \mapsto \mathrm{pmi}(\color{red}{x}, \color{blue}{y})$$



Embeddings as Functions Over Sets

$$\color{red}{X} \color{black}{= \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}}$$

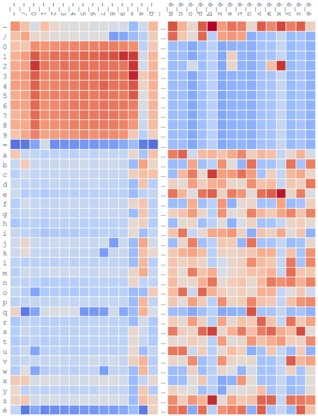
$$\color{blue}{Y} \color{black}{= X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}}$$

$$M \colon \color{red}{X} \color{black}{\times} \color{blue}{Y} \color{black}{\to \mathbb{R}}$$

$$(\color{red}{x}\color{black}{,} \color{blue}{y}\color{black}{) \mapsto \mathrm{pmi}(}\color{red}{x}\color{black}{,} \color{blue}{y}\color{black}{)}$$

$$M_x \colon \color{red}{X} \color{black}{\to \mathbb{R}^{\color{blue}{Y}}}$$

$$\color{red}{x} \color{black}{\mapsto} \color{blue}{M(x, -)}$$



Embeddings as Functions Over Sets

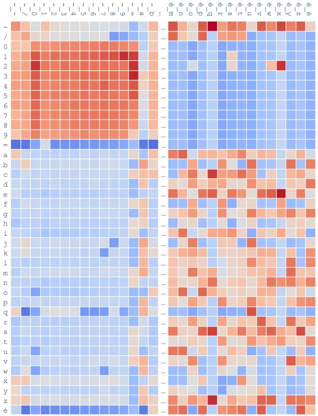
$$X = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$Y = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

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

$$M_x \colon X \to \mathbb{R}^Y$$
$$x \mapsto M(x, -)$$

$$M_y \colon Y \to \mathbb{R}^X$$
$$y \mapsto M(-, y)$$



Embeddings as Functions Over Sets

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

$$Y = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

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

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

$$\begin{aligned} M_x \colon X &\to \mathbb{R}^Y \\ x &\mapsto M(x, -) \end{aligned}$$

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

$$\begin{aligned} M_y \colon Y &\to \mathbb{R}^X \\ y &\mapsto M(-, y) \end{aligned}$$

Embeddings as Functions Over Sets

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

$$Y = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

$$\begin{aligned} M \colon X \times Y &\to \mathbb{R} \\ (x, y) &\mapsto \operatorname{pmi}(x, y) \end{aligned}$$

$$\begin{aligned} M_x \colon X &\to \mathbb{R}^Y \\ x &\mapsto M(x, -) \end{aligned}$$

$$\begin{aligned} M_y \colon Y &\to \mathbb{R}^X \\ y &\mapsto M(-, y) \end{aligned}$$

$$\begin{tikzcd} X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \\ \mathbb{R}^X & Y \arrow[l, "M_y"] \arrow[u, hook] \end{tikzcd}$$

Embeddings as Functions Over Sets

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

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

$$\begin{aligned} M \colon X \times Y &\to \mathbb{R} \\ (x, y) &\mapsto \operatorname{pmi}(x, y) \end{aligned}$$

$$\begin{aligned} M_x \colon X &\to \mathbb{R}^Y \\ x &\mapsto M(x, -) \end{aligned}$$

$$\begin{aligned} M_y \colon Y &\to \mathbb{R}^X \\ y &\mapsto M(-, y) \end{aligned}$$

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[dl, dashed, shift left, "M_*"] \\
\mathbb{R}^X \arrow[ur, dashed, shift left, "M^*"] & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

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

Embeddings as Functions Over Sets

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

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^{Y} \arrow[dl, dashed, shift left, "M_*"] \\
\mathbb{R}^{X} \arrow[ur, dashed, shift left, "M^*"] & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

Embeddings as Functions Over Sets

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

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

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^{Y} \arrow[dl, dashed, shift left, "M_*"] \\
\mathbb{R}^{X} \arrow[ur, dashed, shift left, "M^*"] & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

Embeddings as Functions Over Sets

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

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

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

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

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

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[dl, dashed, shift left, "M_*"] \\
\mathbb{R}^X \arrow[ur, dashed, shift left, "M^*"] & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

Embeddings as Functions Over Sets

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

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

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[dl, dashed, bend left, "M_*"] \\
\mathbb{R}^X \arrow[ur, dashed, "M^*"] & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

$$M = U\Sigma V^T \qquad
\begin{aligned}
U &:= [u_1, \ldots, u_m] \\
V &:= [v_1, \ldots, v_n] \\
\Sigma &:= \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_r} \end{bmatrix}
\end{aligned}$$

Embeddings as Functions Over Sets

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

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

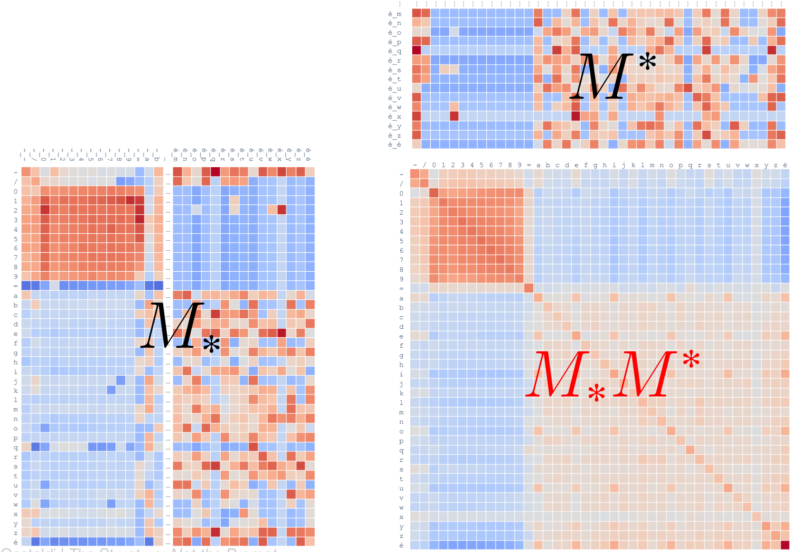
$$M_*M^* u_i = \lambda_i u_i$$
$$M^*M_* v_i = \lambda_i v_i$$

The u_i and v_i are (linear)
fixed points!

$$\begin{tikzcd}
X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[dl, dashed, shift left, "M_*"] \\
\mathbb{R}^X \arrow[ur, dashed, shift left, "M^*"] & Y \arrow[l, "M_y"] \arrow[u, hook]
\end{tikzcd}$$

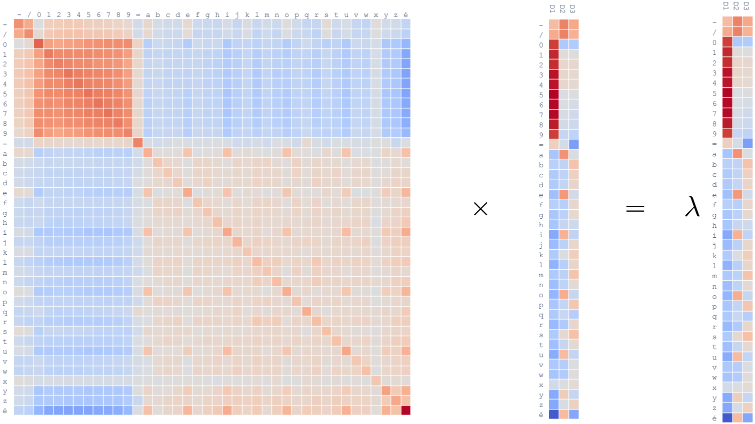
$$M = U \Sigma V^T \qquad U := [u_1, \dots, u_m] \qquad V := [v_1, \dots, v_n] \qquad \Sigma := \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_r} \end{bmatrix}$$

M_*M^* as a Covariance Matrix

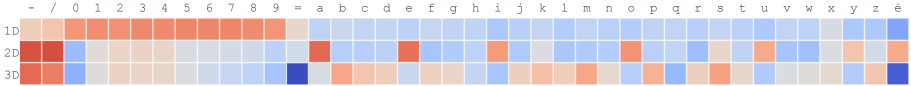


Eigenvectors as Fixed Points

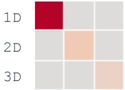
$$M_* M^* u = \lambda u$$



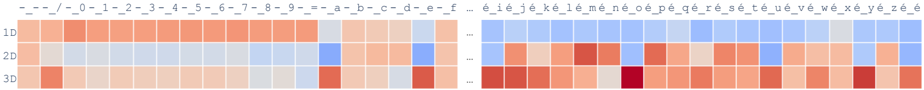
Eigenvectors of $M_* M^*$:



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

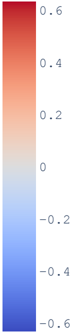


Eigenvectors of $M^* M_*$:



Words

	-5	-4	-3	-2	-1	0	1	2	3	4	5
D 1	church	university	field	house	centre	...	held	used	offered	found	made
D 2	use	leave	keep	buy	meet	...	boy	club	sun	uk	hotel
D 3	show	boy	project	move	play	...	production	size	interests	activities	nature
D 4	used	expected	made	considered	allowed	...	london	europe	scotland	france	england
D 5	used	expected	food	water	england	...	during	couple	under	series	lot
D 6	perhaps	indeed	under	during	in	...	cup	bit	series	couple	lot
D 7	difficult	hard	easy	necessary	close	...	won	gave	started	saw	took
D 8	europe	scotland	england	france	lot	...	middle	want	needs	army	could
D 9	wish	tried	seem	seemed	began	...	received	established	won	published	produced
D 10	10	15	20	30	3	...	from	on	black	into	through



Outline

Introduction

NLMs as Formal Objects

The Structure(s) of the Embeddings

- The Algebra Behind the Embeddings
- The Structure Behind the Algebra
- The Categories Behind the Structure

Conclusion

Embeddings as Functors Over Categories

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

$$Y = X \times X = \{\texttt{(-,-)}, \texttt{(-,/)}, \texttt{(-,0)}, \ldots, \texttt{(é,z)}, \texttt{(é,é)}\}$$

$$\begin{aligned} M \colon X \times Y &\to \mathbb{R} \\ (x, y) &\mapsto \operatorname{pmi}(x, y) \end{aligned}$$

$$\begin{aligned} M_x \colon X &\to \mathbb{R}^Y \\ x &\mapsto M(x, -) \end{aligned}$$

$$\begin{aligned} M_y \colon Y &\to \mathbb{R}^X \\ y &\mapsto M(-, y) \end{aligned}$$

$$\begin{tikzcd} X \arrow[r, "M_x"] \arrow[d, hook] & \mathbb{R}^Y \arrow[dl, dashed, bend left=10, "M_*"] \\ \mathbb{R}^X \arrow[ur, dashed, bend left=10, "M^*"] & Y \arrow[l, "M_y"] \arrow[u, hook] \end{tikzcd}$$

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

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

Embeddings as Functors Over Categories

$$\mathsf{C} = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

$$\mathsf{D} = \mathsf{C} = \{\texttt{-}, \texttt{/}, \texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{=}, \texttt{a}, \texttt{b}, \texttt{c}, \ldots, \texttt{w}, \texttt{x}, \texttt{y}, \texttt{z}, \texttt{é}\}$$

Profunctor

$$\begin{aligned} \mathcal{M}\colon \mathsf{C}^{\mathrm{op}} \times \mathsf{D} &\to \mathsf{Set} \\ (c, d) &\mapsto \mathcal{M}(c, d) \end{aligned}$$

$$\begin{aligned} \mathcal{M}_c\colon \mathsf{C} &\to \left(\mathsf{Set}^{\mathsf{D}}\right)^{\mathrm{op}} \\ c &\mapsto \mathcal{M}(c, -) \end{aligned}$$

$$\begin{aligned} \mathcal{M}_d\colon \mathsf{D} &\to \mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \\ d &\mapsto \mathcal{M}(-, d) \end{aligned}$$

$$\begin{tikzcd}
\mathsf{C} \arrow[r, "\mathcal{M}_c"] \arrow[d, hook, "Yoneda"'] & (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \arrow[dl, dashed, shift left, "\mathcal{M}_*"] \\
\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \arrow[ur, dashed, shift left, "\mathcal{M}^*"] & \mathsf{D} \arrow[l, "\mathcal{M}_d"] \arrow[u, hook, "Yoneda"']
\end{tikzcd}$$

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

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

Embeddings as Functors Over Categories

Isbell Adjunction

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

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

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

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

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

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

$$\begin{tikzcd}
\mathsf{C} \arrow[r, "\mathcal{M}_c"] \arrow[d, hook, "Yoneda"'] & (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \arrow[dl, dashed, shift left, "\mathcal{M}_*"] \\
\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \arrow[ur, dashed, shift left, "\mathcal{M}^*"] & \mathsf{D} \arrow[l, "\mathcal{M}_d"] \arrow[u, hook, "Yoneda"']
\end{tikzcd}$$

Categories C and D can be enriched!

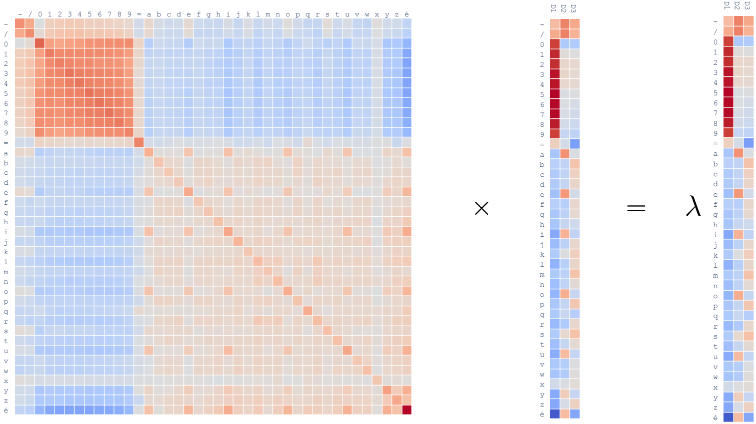
E.g.:

$$\mathcal{M}^*: 2^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (2^{\mathsf{D}})^{\mathrm{op}} : \mathcal{M}_*$$

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

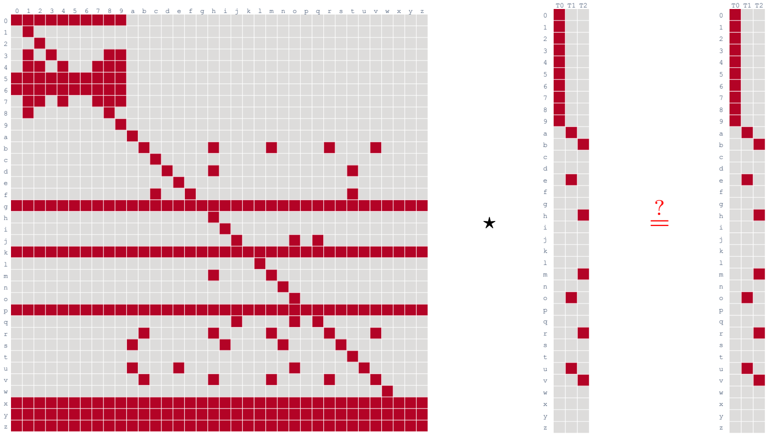
Binary Fixed Points

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



Binary Fixed Points

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

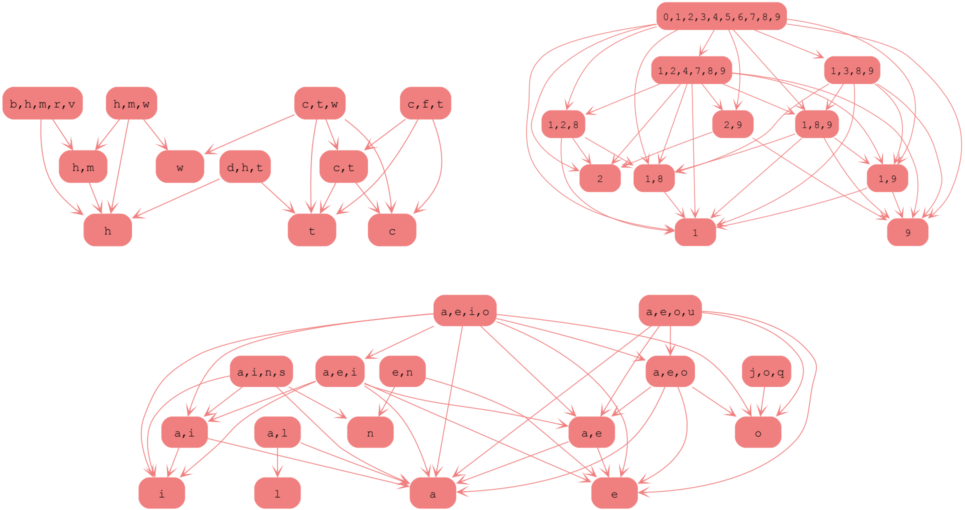


“Eigensets”

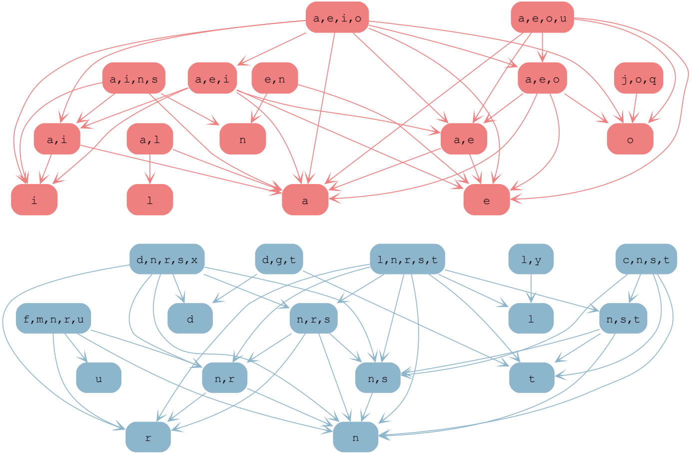
$$\mathcal{M}_*\mathcal{M}^*f = f$$

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	1	9
e,n	a,l	c,t				

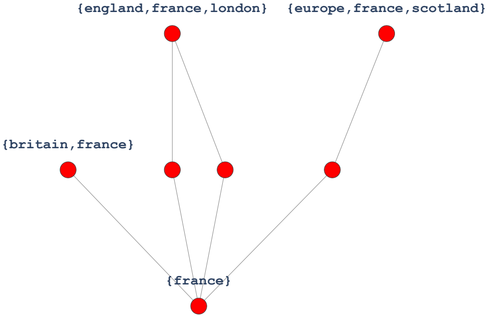
Which Structure?

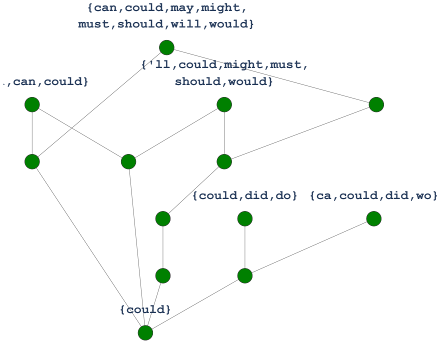


Which Structure?



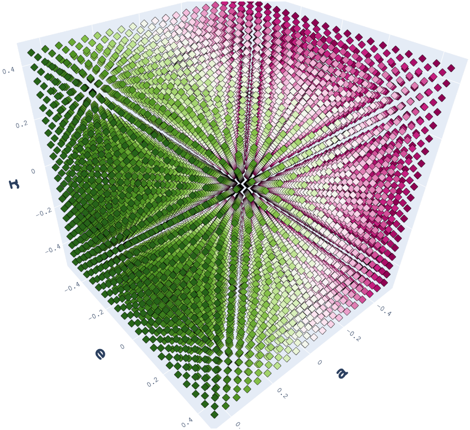
Formal Concepts (Words)

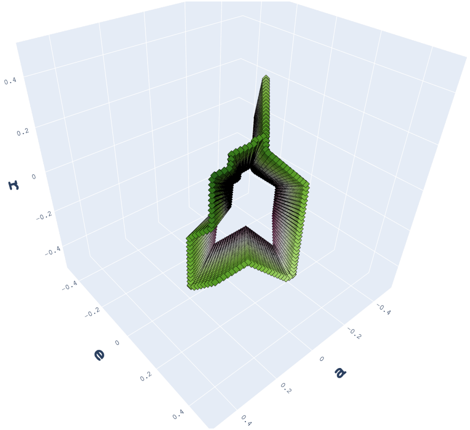




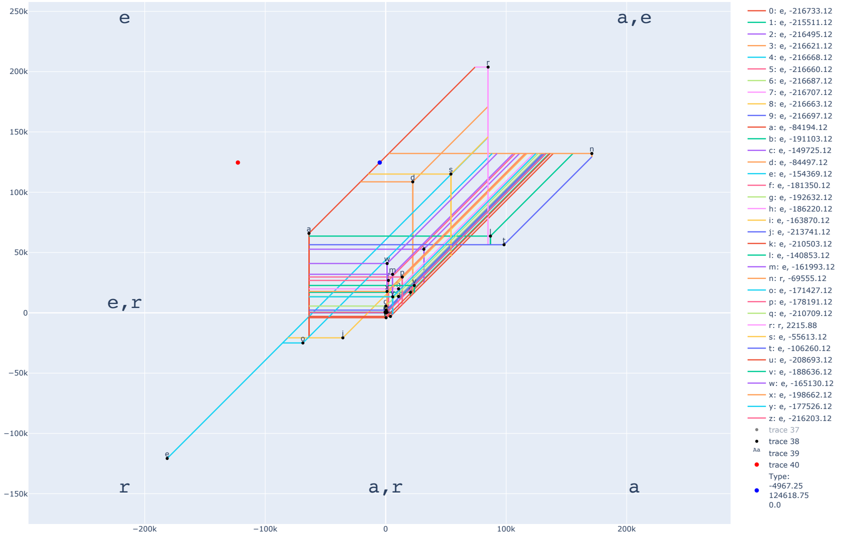
Nucleus

$$\bar{\mathbb{R}}^{\{a,e,r\}} \xrightarrow{\mathcal{M}_*\mathcal{M}^*} \bar{\mathbb{R}}^{\{a,e,r\}}$$





Internal Structure of the Nucleus

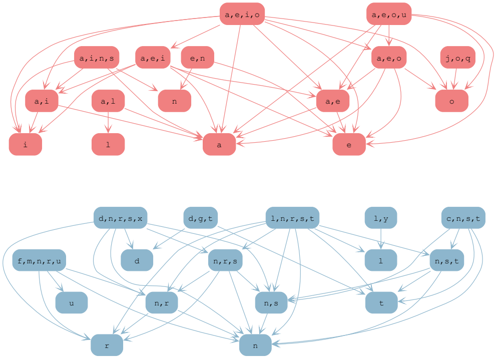


Definition (Polar/Orthogonal - Girard, 2011)

[G]iven a binary operation, noted
$a, b \rightsquigarrow \langle a|b\rangle \colon A \times B \to C$ 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 :

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

- The map 'polar' is decreasing:
  $X \subset X' \Rightarrow X'^{\perp} \subset X^{\perp}$.
- The set 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, $X^{\perp\perp\perp} = X^{\perp}$.



Theory of Computational Types

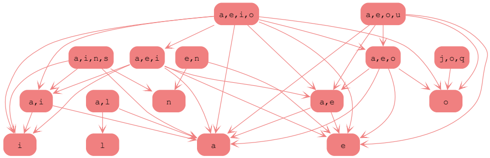
Definition (Polar/Orthogonal - Girard, 2011)

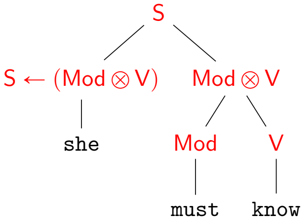
[G]iven a binary operation, noted $a, b \rightsquigarrow \langle a|b\rangle \colon A \times B \to C$ and a subset $P \subset C$ (the 'pole') one can define the *polar* $X^\perp \subset B$ of a subset $X \subset A$ (resp. $Y^\perp \subset A$ of a subset $Y \subset B$) by :

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

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

- The map 'polar' is decreasing: $X \subset X' \Rightarrow X'^\perp \subset X^\perp$.
- The set 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, $X^{\perp\perp\perp} = X^\perp$.





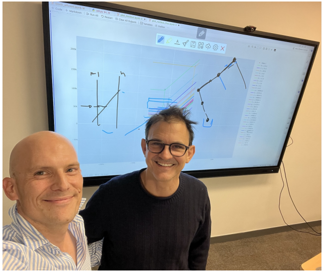
(Gastaldi and Pellissier, 2021)

Outline

Conclusion: For a Critical Formalism

- It is urgent to address of the epistemological dimension of the critical project in its own terms
- This requires to develop a critical approach within formal sciences where formalization is not assumed to lead to naturalization
  - The new role of data within formal sciences is crucial in this sense
- A critical formalism will be incomplete if it remains disconnected from the political, and even the artistic dimension of the critical program
  - We need a new alliance between the formal sciences, the human sciences, and the arts.

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*. 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

References I

Belrose, N., Schneider-Joseph, D., Ravfogel, S., Cotterell, R., Raff, E., & Biderman, S. (2024).Leace: Perfect linear concept erasure in closed form. *Proceedings of the 37th International Conference on Neural Information Processing Systems*.

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

Bender, E. M., & Koller, A. (2020).Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. https://doi.org/10.18653/v1/2020.acl-main.463

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

Church, A. (1936).An unsolvable problem of elementary number theory. *American Journal of Mathematics*, *58*(2), 345–363.

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*. https://doi.org/10.1080/03080188.2021.1890484

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

References II

Gödel, K. (1986 (1934)). On undecidable propositions of formal mathematical systems. In *Collected works* (pp. 346–371). Clarendon Press Oxford University Press.

Goldberg, Y., & Levy, O. (2014).Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *CoRR*, *abs/1402.3722*.

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/

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.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., Le, Q., & Strohmann, T. (2013). *Learning representations of text using neural networks. NIPS deep learning workshop 2013 slides*.

Nietzsche, F. (1873). On truth and lying in a non-moral sense [Originally unpublished; written in 1873.]. (R. Speirs, Trans.). In R. Geuss & R. Speirs (Eds.), *The birth of tragedy and other writings* (pp. 141–153). Cambridge University Press.

Turing, A. (1937).On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230

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/

Vernacular AI
Digital Theory Lab, NYU
New York, NY, USA

The Structure, Not the Prompt

For a Critical Formalism

Juan Luis Gastaldi

ETH zürich

February 7, 2025