

CNRS - Concours chercheurs 2025
CR Section 53 - Concours n° 53/03

*Épistémologie des modèles distributionnels de langage
par apprentissage machine*
Explicabilité formelle et interprétabilité théorique

Juan Luis Gastaldi

http://www.jlgastaldi.com/assets/gastaldi_cnrs_cr.pdf



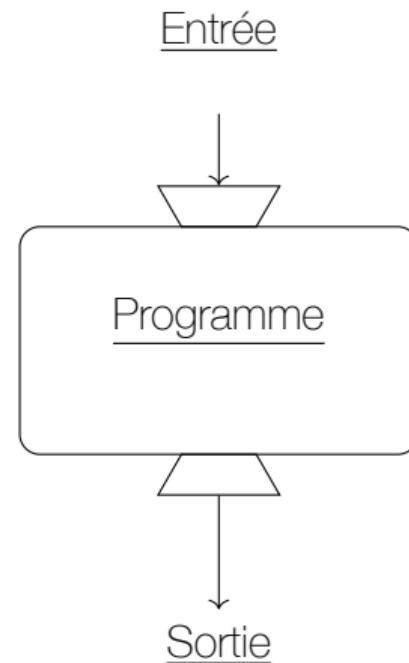
1997-2007	Recherche pré-doctorale Sciences Po, Philosophie, Maths Argentine, France (UNR, ENS, Paris 1, UPMC)	2023-Présent	Nouvelle recherche doctorale Informatique (ML, TAL) Suisse, USA (ETH Zurich)
2008-2014	Recherche doctorale Philo et Hist des Sciences France (Bordeaux Montaigne)		
2015-2022	Professeur d'Ens. Artistique Philo et Hist des Idées France (MO.CO.ESBA)		
2015-2022	Recherche post-doctorale Philo et Informatique France, Suisse, Tchéquie, USA (ETH, MSCA, CUNY, CMU)		

1997-2007	Recherche pré-doctorale Sciences Po, Philosophie, Maths Argentine, France (UNR, ENS, Paris 1, UPMC)
2008-2014	Recherche doctorale Philo et Hist des Sciences France (Bordeaux Montaigne)
2015-2022	Professeur d'Ens. Artistique Philo et Hist des Idées France (MO.CO.ESBA)
2015-2022	Recherche post-doctorale Philo et Informatique France, Suisse, Tchéquie, USA (ETH, MSCA, CUNY, CMU)

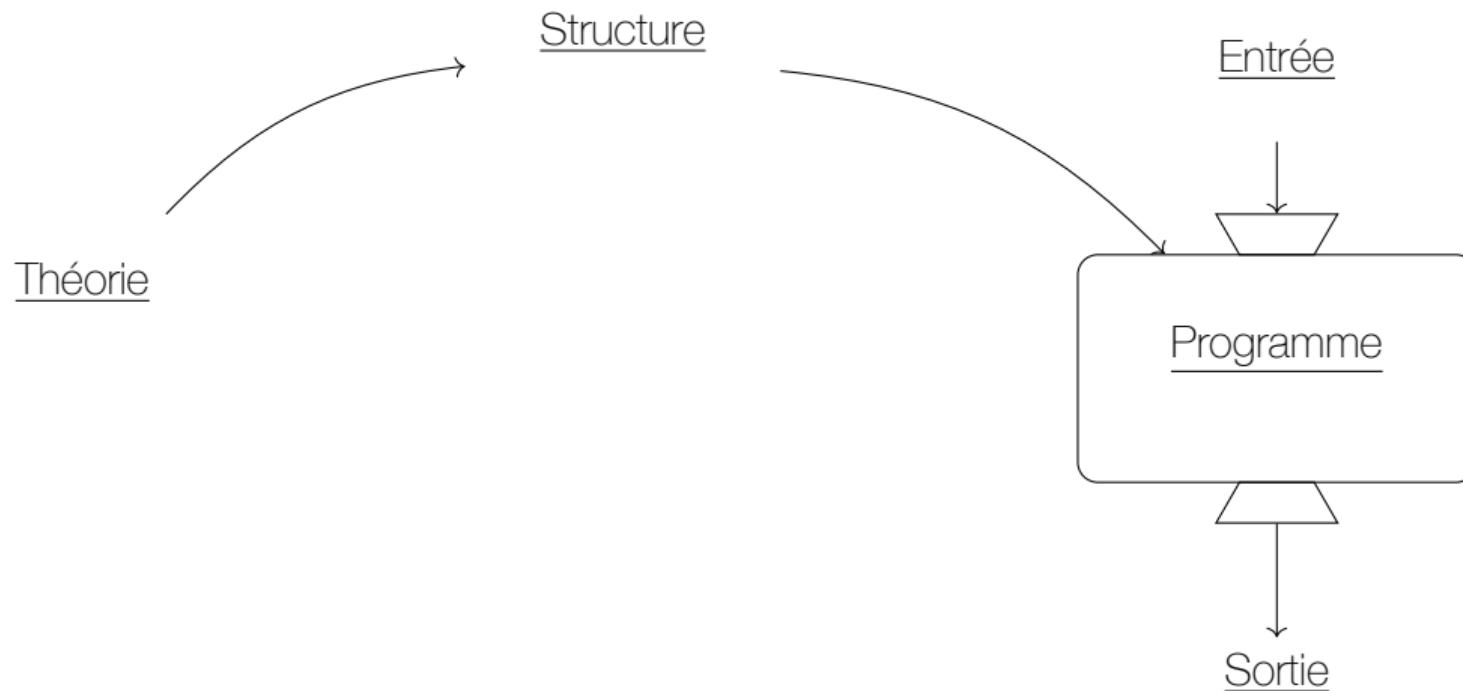
2023-Présent Nouvelle recherche doctorale
Informatique (ML, TAL)
Suisse, USA
(ETH Zurich)

Recherche
<ul style="list-style-type: none">– Formalisme critique: Alliance entre les humanités critiques et les sciences formelles– Approche philosophique, historique, théorique et technique– Thèse en Philosophie: Philosophie et histoire de la mathématisation de la logique (Gastaldi, 2014)– Thèse en Informatique: Tokenisation en TAL (Gastaldi et al., 2024; Julianelli et al., 2024; Vieira et al., 2024; Zouhar et al., 2023a, 2023b)

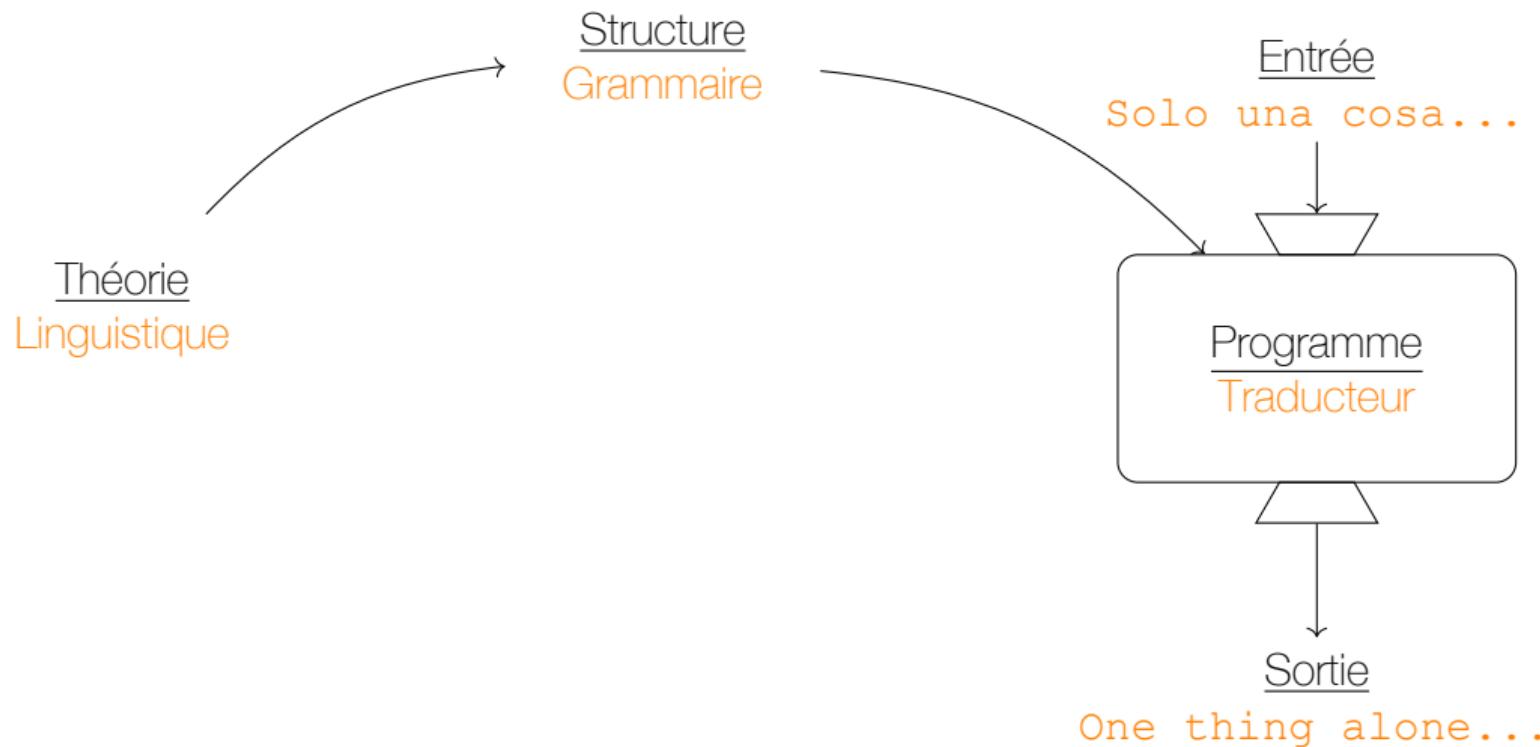
La structure implicite des données



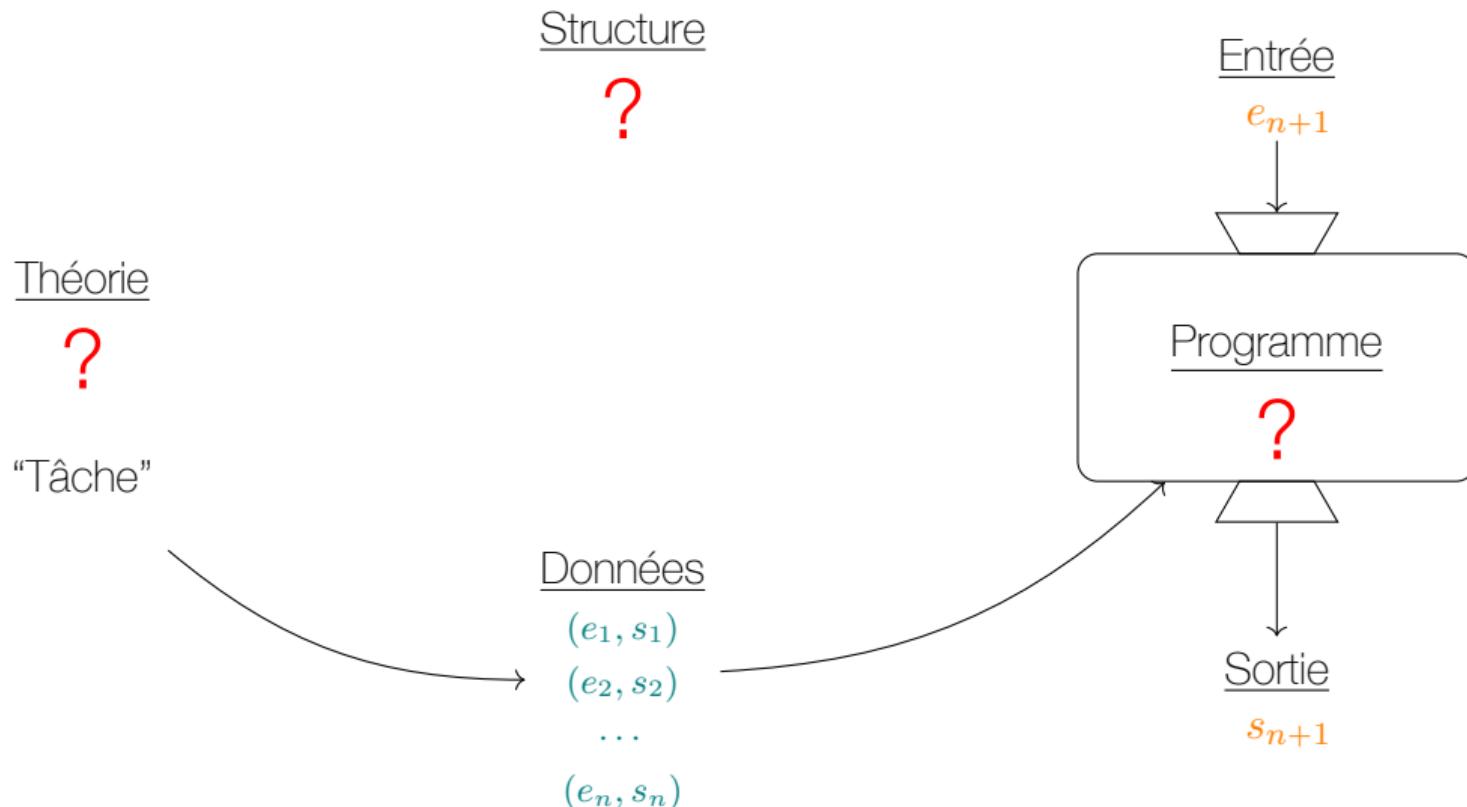
La structure implicite des données



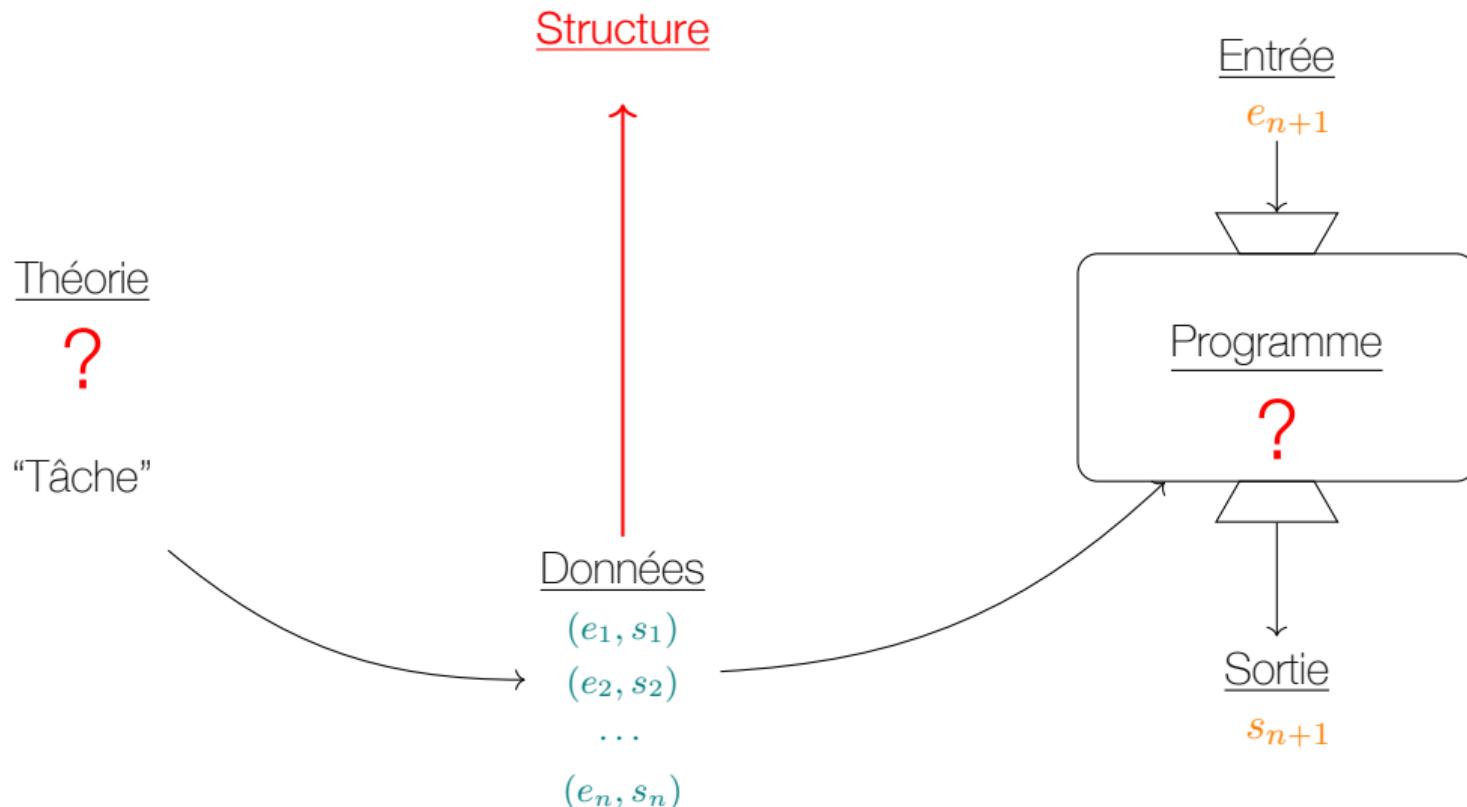
La structure implicite des données



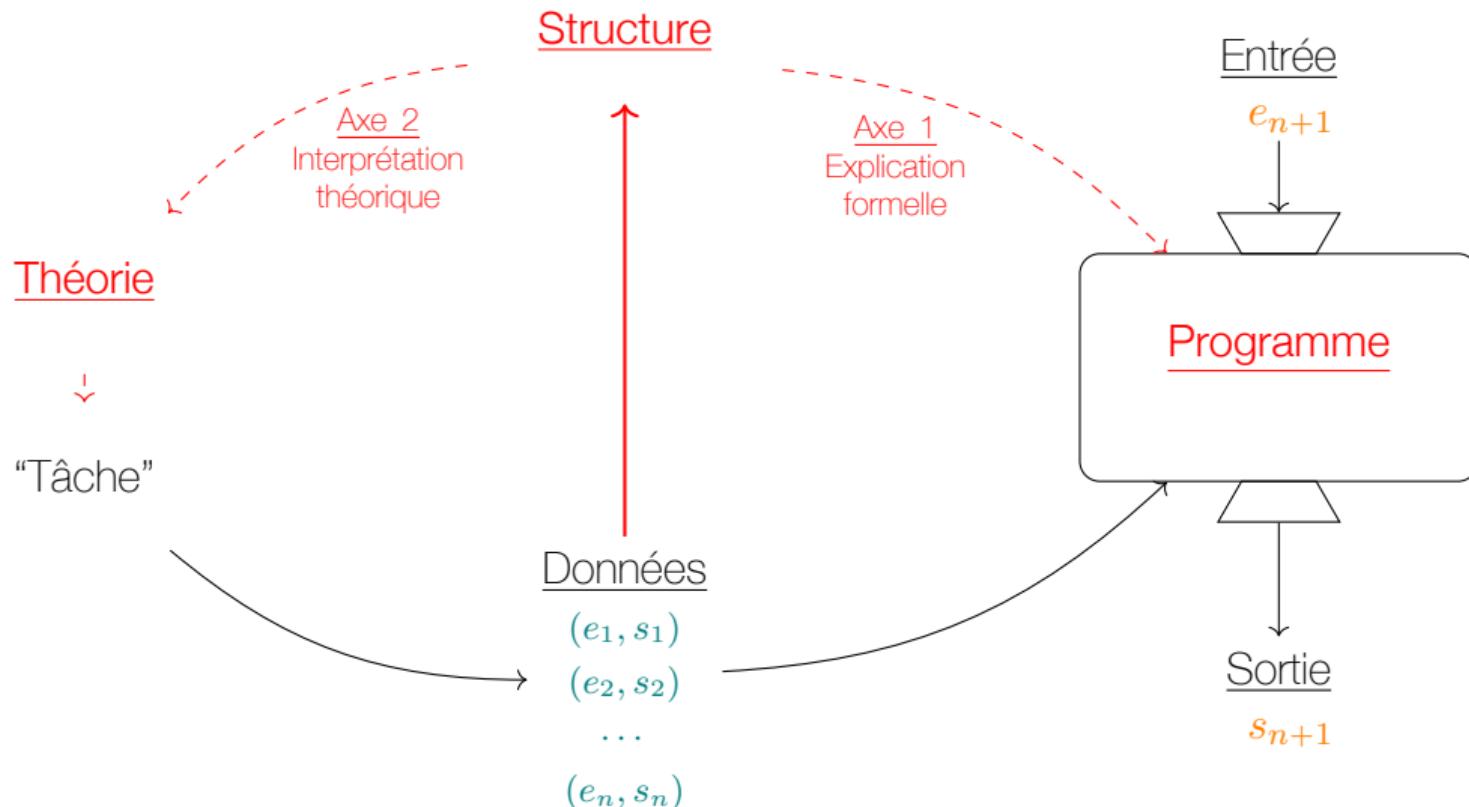
La structure implicite des données



La structure implicite des données

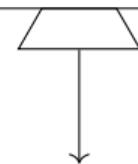
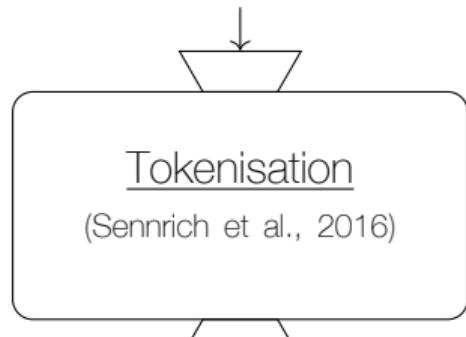


La structure implicite des données



Axe 1: Explicabilité formelle

Epistemology of Machine Learning
Distributional Language Models

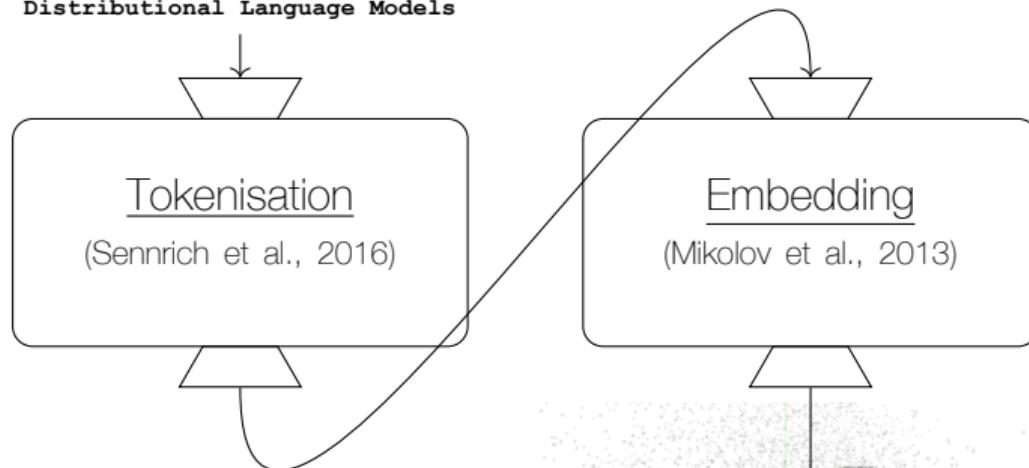


Epistemology of Machine Learning
Distributional Language Models

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

Axe 1: Explicabilité formelle

**Epistemology of Machine Learning
Distributional Language Models**



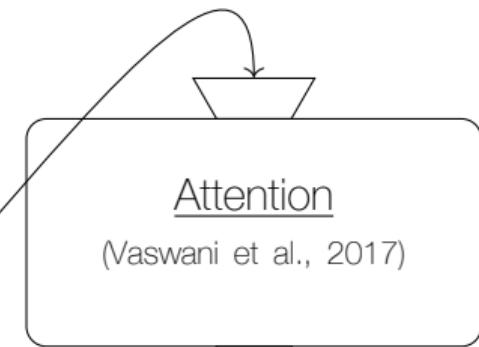
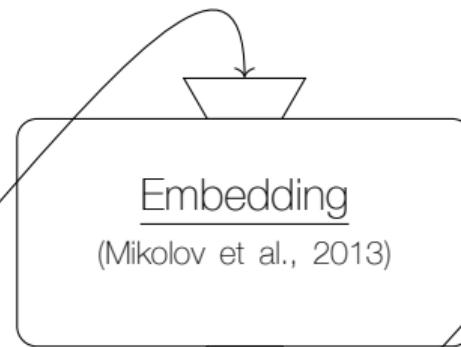
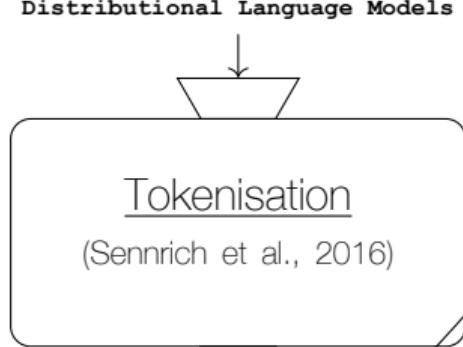
**Epistemology of Machine Learning
Distributional Language Models**

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

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

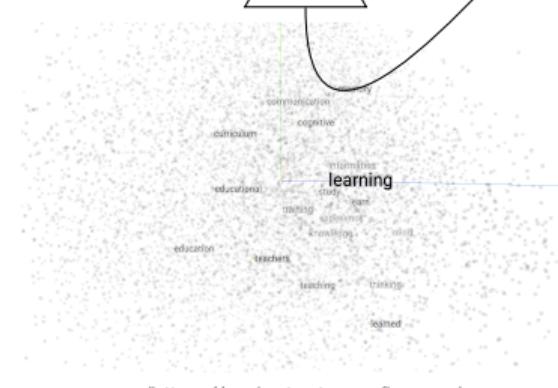
Axe 1: Explicabilité formelle

Epistemology of Machine Learning
Distributional Language Models



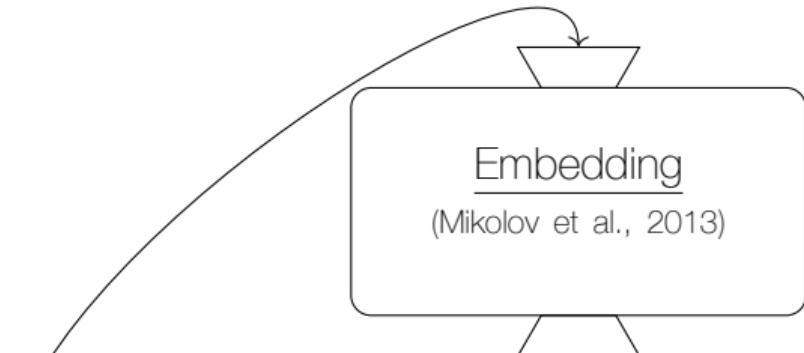
Epistemology of Machine Learning
Distributional Language Models

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



Ep
ist
em
ology
of
Machine
Learn
ing
Distribution
al
Language
Models
(<https://github.com/jessevieg/bertviz>)

Axe 1: Explicabilité formelle



Epistemology of Machine Learning
Distributional Language Models

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



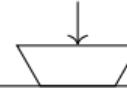
(<https://projector.tensorflow.org>)

La structure des embeddings

Structure

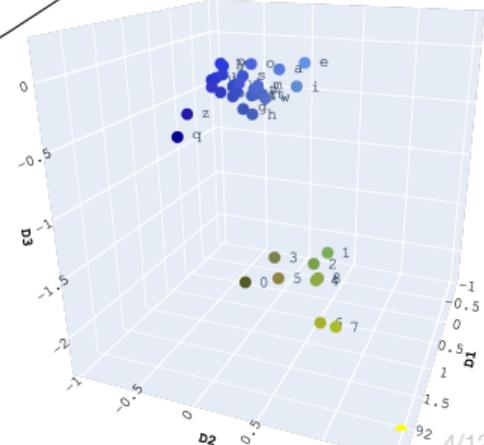
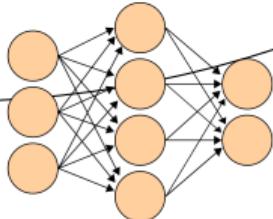
?

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



Embedding

Données

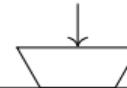


La structure des embeddings

Structure

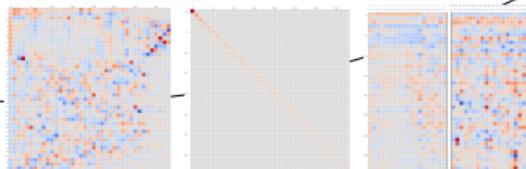
?

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

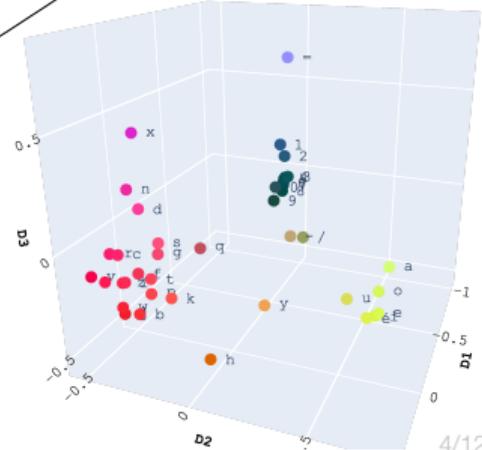


Embedding

Données



SVD

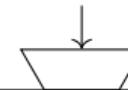


La structure des embeddings

Structure

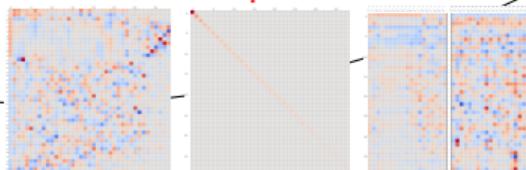


$\{-, /, 0, 1, 2, \dots, 8, 9, =,$
 $a, b, c, \dots, w, x, y, z, \acute{e}\}$

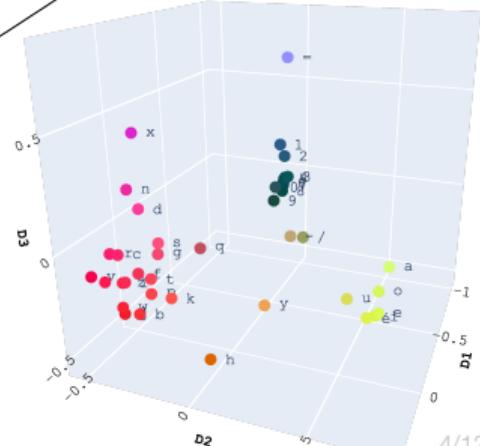


Embedding

Données



SVD

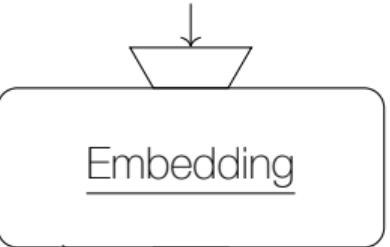


La structure des embeddings

Structure

?

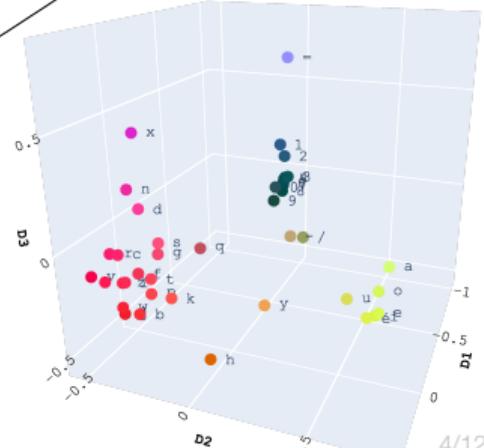
{-, /, 0, 1, 2, ..., 8, 9, =,
a, b, c, ..., w, x, y, z, é}



Données

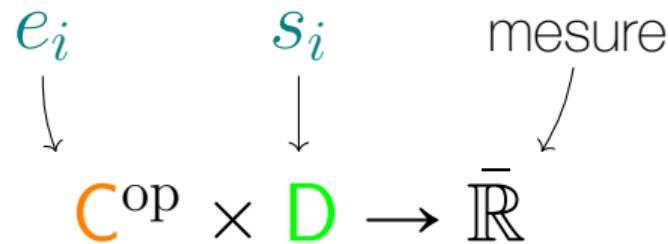


$$C^{\text{op}} \times D \rightarrow \bar{\mathbb{R}}$$



Structure

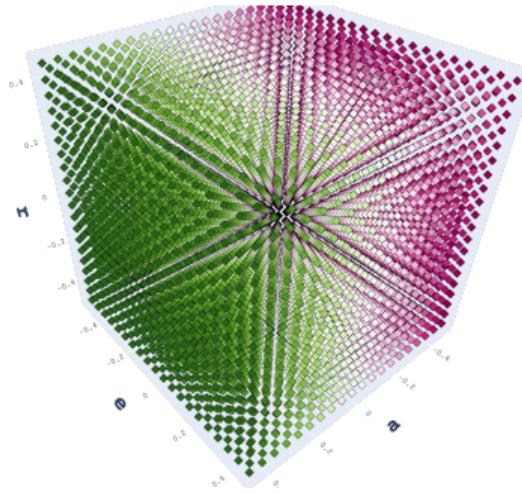
?



Structure

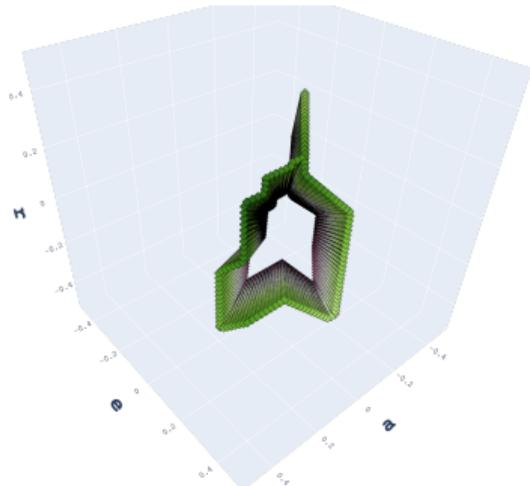
?

$$\begin{array}{ccc} e_i & s_i & \text{mesure} \\ \downarrow & \downarrow & \swarrow \\ C^{\text{op}} \times D & \rightarrow & \bar{\mathbb{R}} \\ & \Downarrow & \\ M^*: \bar{\mathbb{R}}^{C^{\text{op}}} & \leftrightarrows & (\bar{\mathbb{R}}^D)^{\text{op}}: M_* \end{array}$$

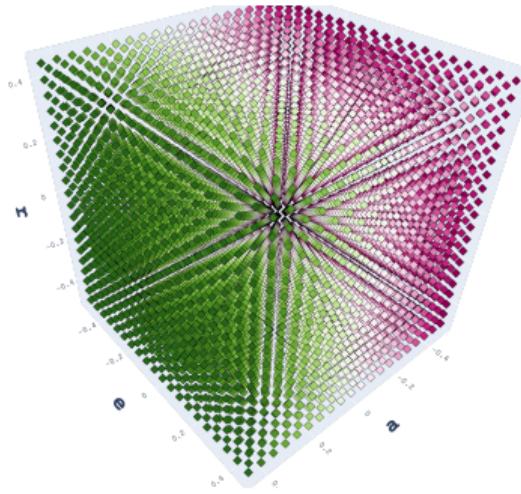
Structure

$$\begin{array}{c} \textcolor{orange}{C}^{\text{op}} \times \textcolor{green}{D} \rightarrow \bar{\mathbb{R}} \\ \Downarrow \\ \mathcal{M}^*: \bar{\mathbb{R}}^{\textcolor{orange}{C}^{\text{op}}} \rightleftarrows (\bar{\mathbb{R}}^{\textcolor{green}{D}})^{\text{op}}: \mathcal{M}_* \end{array}$$

Structure



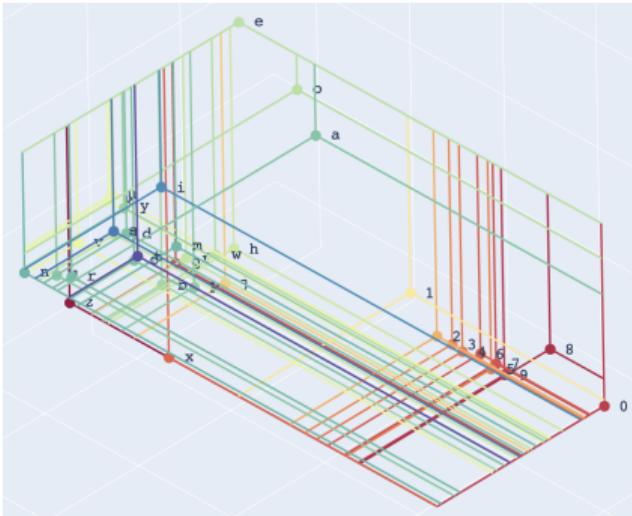
$$\mathcal{M}_* \mathcal{M}^*$$



$$\begin{array}{c}
 \textcolor{orange}{C}^{\text{op}} \times \textcolor{green}{D} \rightarrow \bar{\mathbb{R}} \\
 \Downarrow \\
 \mathcal{M}^*: \bar{\mathbb{R}}^{\textcolor{orange}{C}^{\text{op}}} \rightleftarrows (\bar{\mathbb{R}}^{\textcolor{green}{D}})^{\text{op}}: \mathcal{M}_*
 \end{array}$$

Noyau et Types

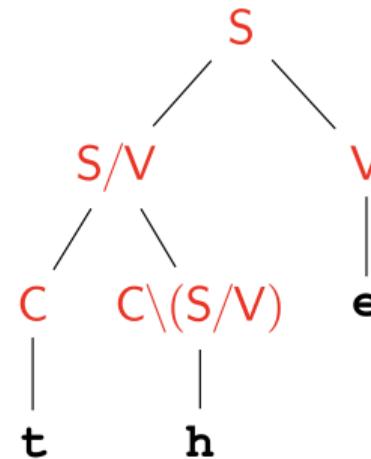
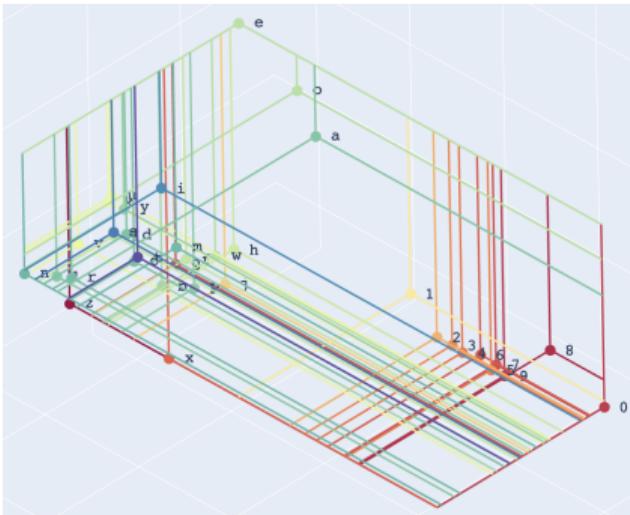
Structure



$$\begin{array}{c} \textcolor{orange}{C}^{\text{op}} \times \textcolor{green}{D} \rightarrow \bar{\mathbb{R}} \\ \Downarrow \\ \mathcal{M}^*: \bar{\mathbb{R}}^{\textcolor{orange}{C}^{\text{op}}} \rightleftarrows (\bar{\mathbb{R}}^{\textcolor{green}{D}})^{\text{op}}: \mathcal{M}_* \end{array}$$

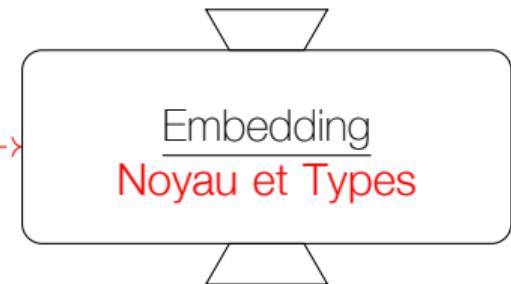
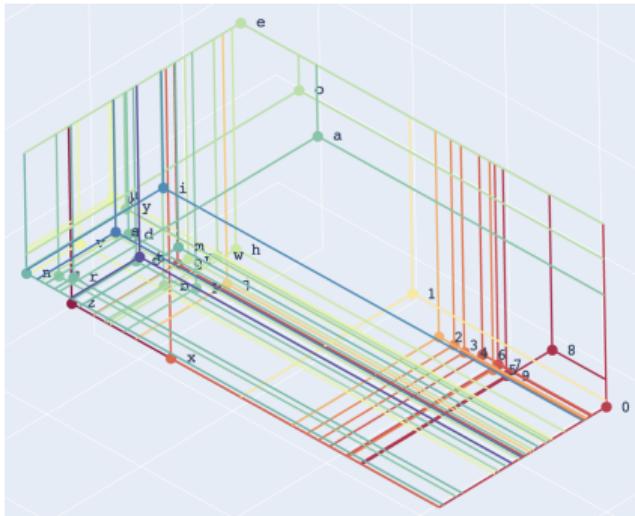
Noyau et Types

Structure

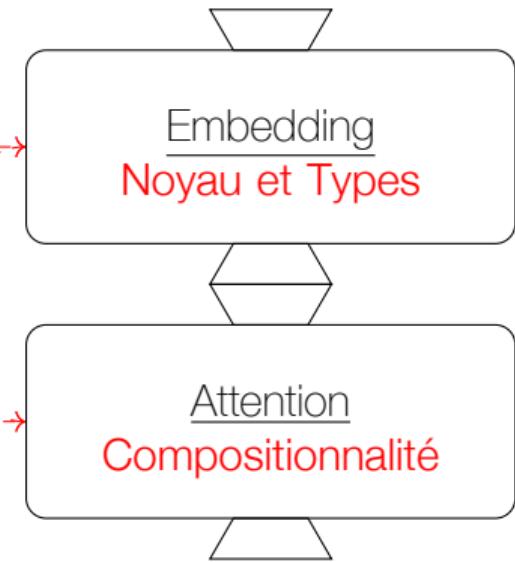
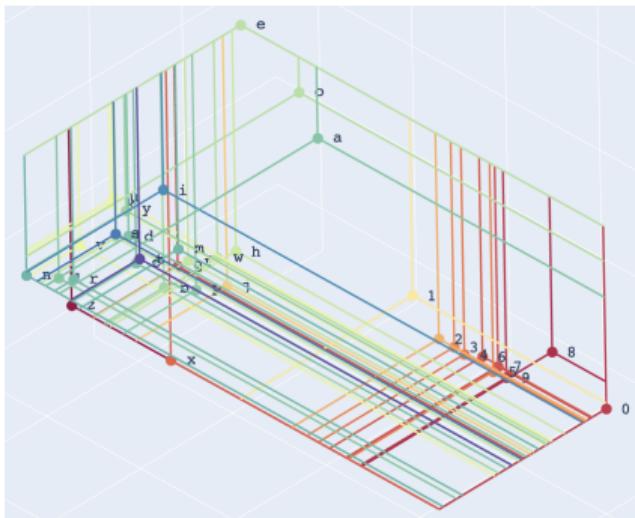


$$\begin{array}{c} \textcolor{orange}{C}^{\text{op}} \times \textcolor{green}{D} \rightarrow \bar{\mathbb{R}} \\ \Downarrow \\ \mathcal{M}^*: \bar{\mathbb{R}}^{\textcolor{orange}{C}^{\text{op}}} \rightleftarrows (\bar{\mathbb{R}}^{\textcolor{green}{D}})^{\text{op}}: \mathcal{M}_* \end{array}$$

Structure

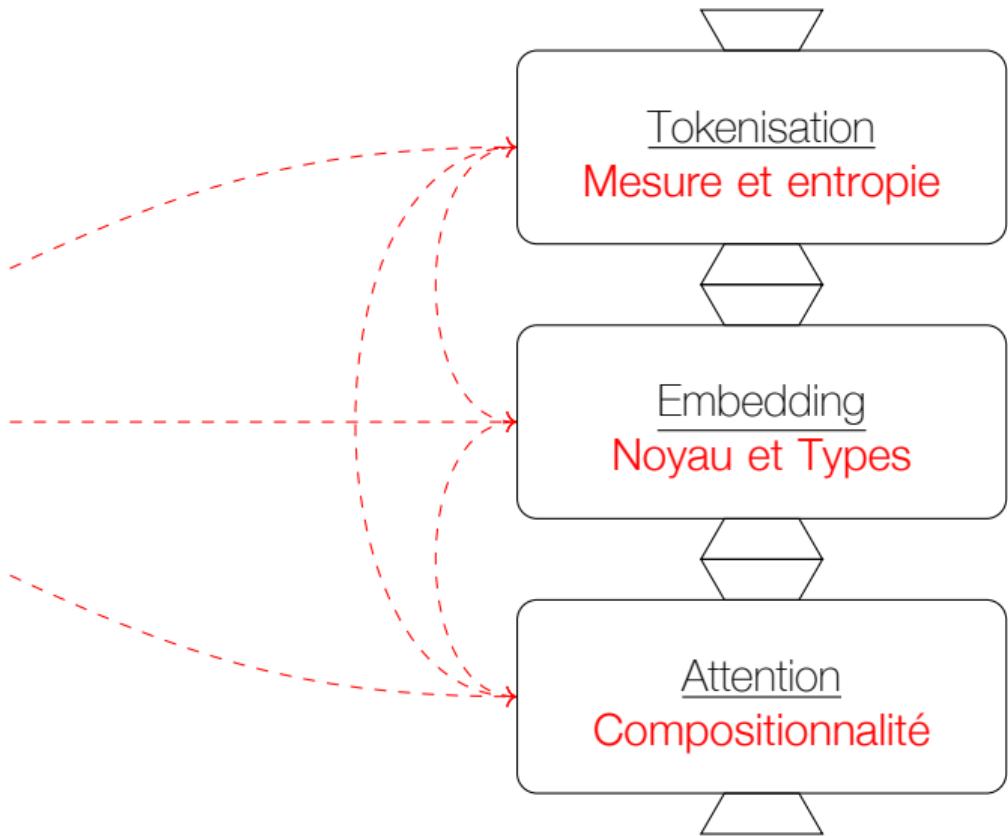
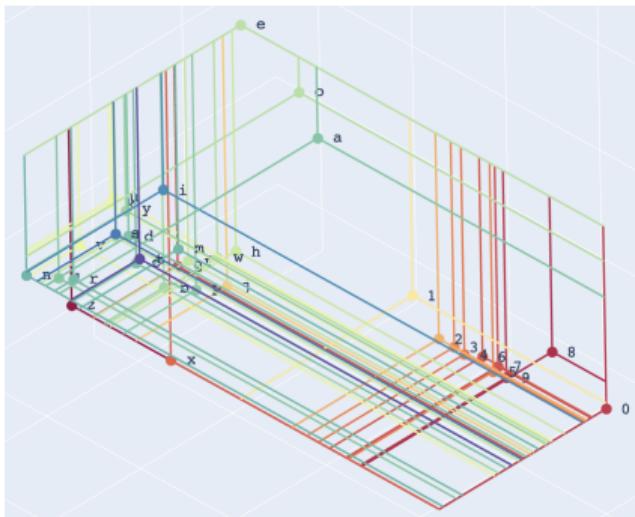


Structure



Axe 1: Objectifs

Structure



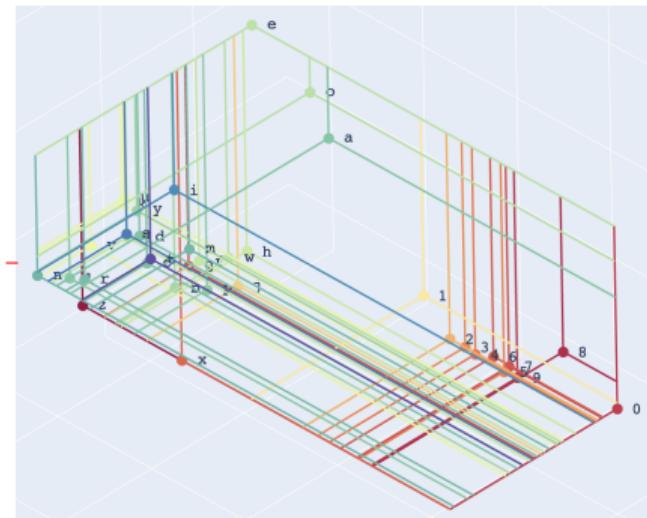
Axe 2: Interprétabilité théorique

Théorie
"Tâche"

?



Structure



Axe 2: Interprétabilité théorique

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

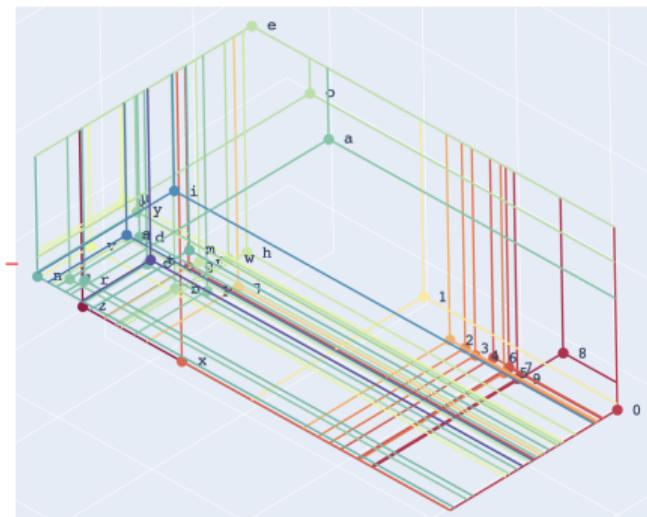
Hypothèse distributionnelle

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

Théorie
"Tâche"



Structure



Axe 2: Interprétabilité théorique

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

Hypothèse distributionnelle

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

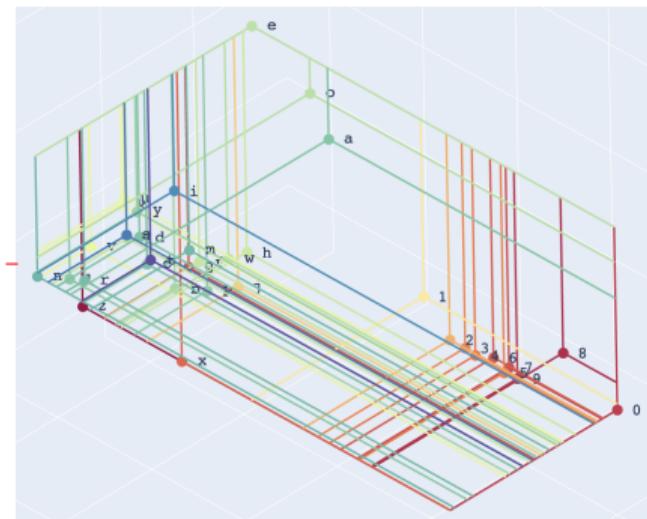
Théorie
"Tâche"



Hypothèse structurale

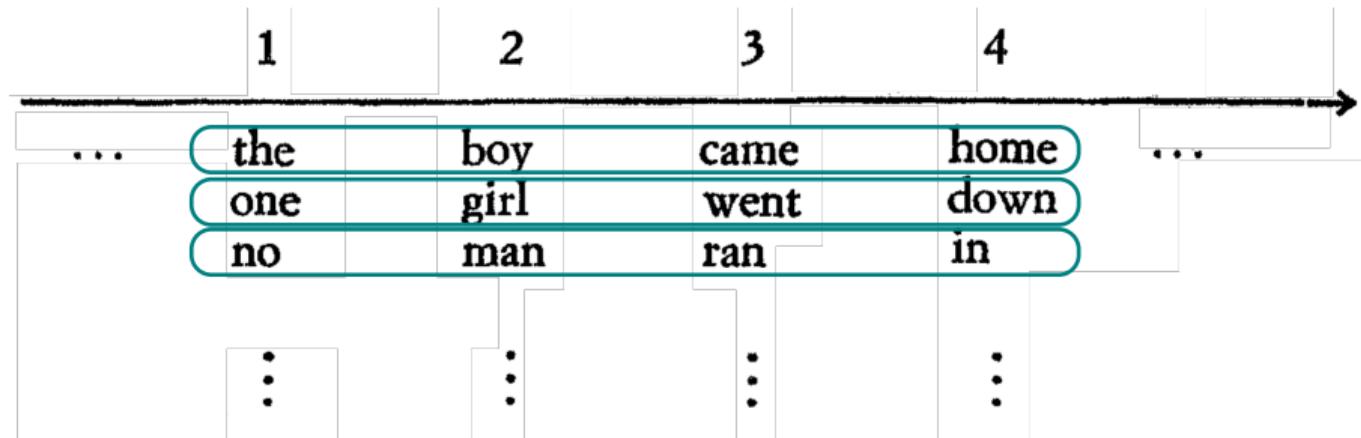
Le contenu linguistique est l'effet d'une structure virtuelle dérivée des pratiques linguistiques dans une communauté.

Structure



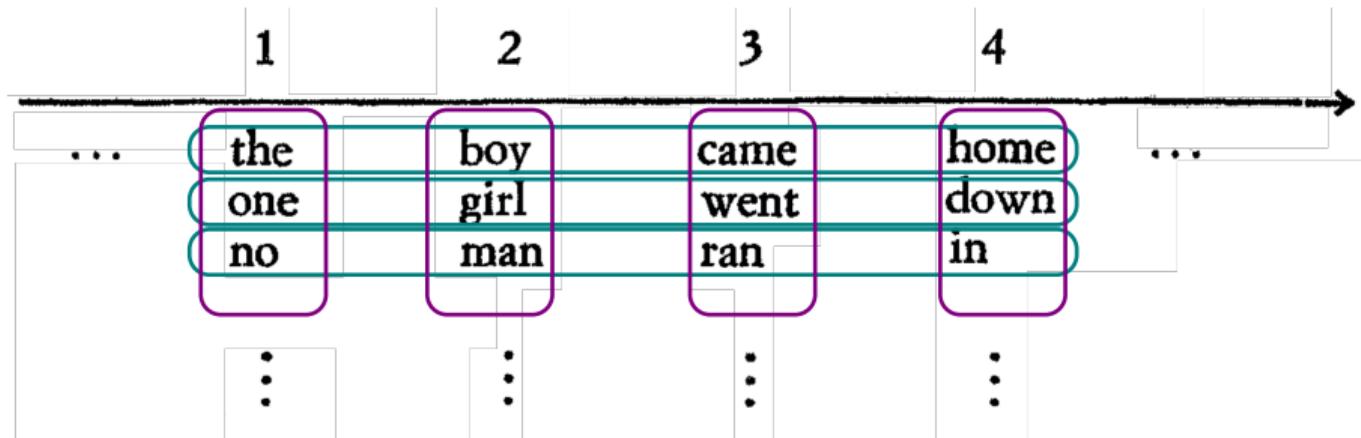
$$\bar{\mathbb{R}}^{\text{C}^{\text{op}}} \leftrightarrow (\bar{\mathbb{R}}^{\text{D}})^{\text{op}}$$

Syntagmes et Paradigmes



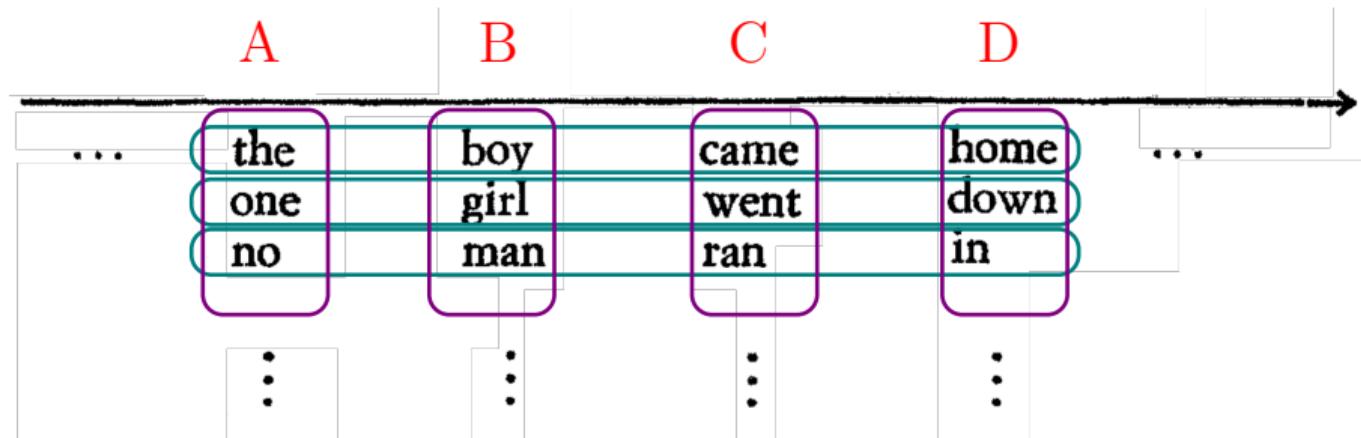
(Hjelmslev, 1971)

Syntagmes et Paradigmes



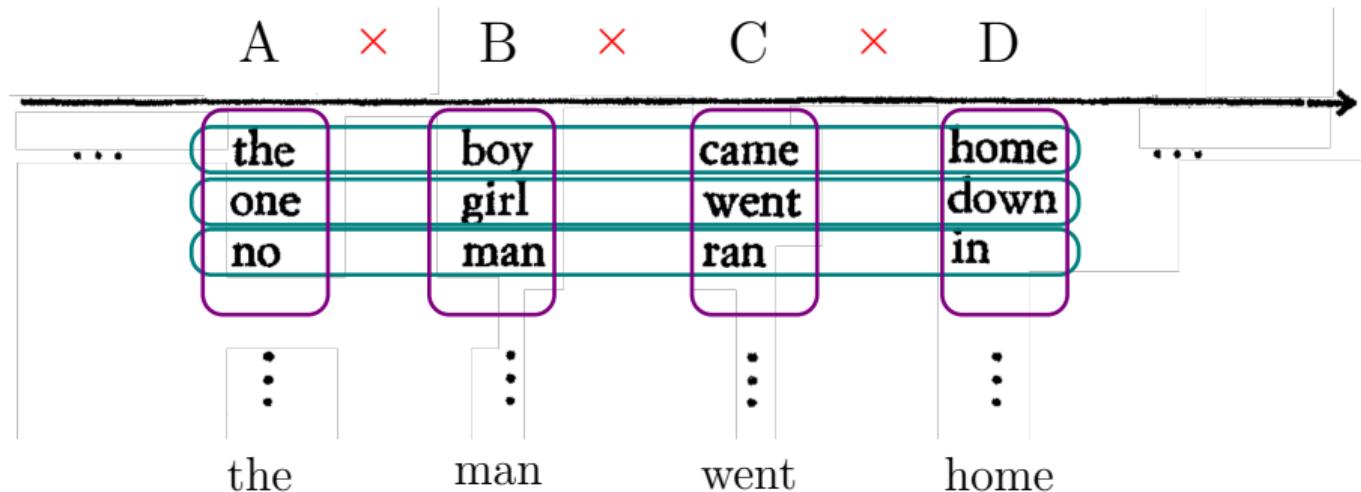
(Hjelmslev, 1971)

Syntagmes et Paradigmes

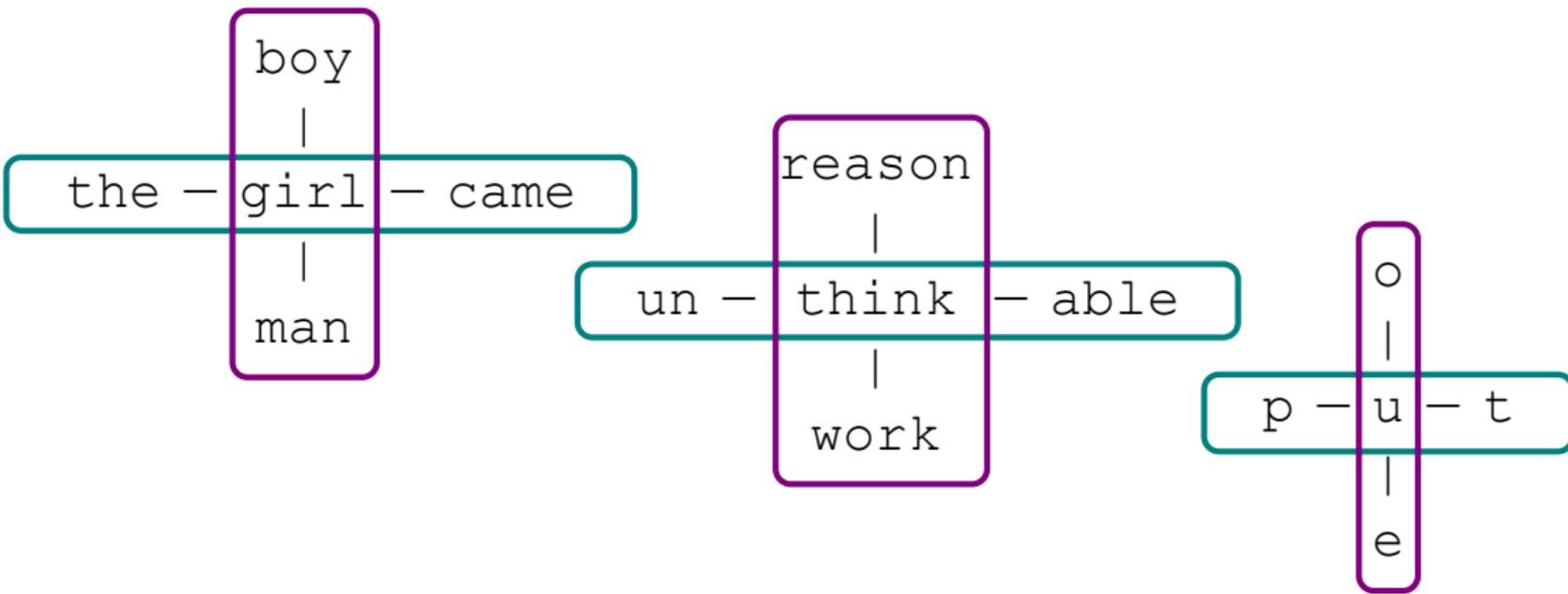


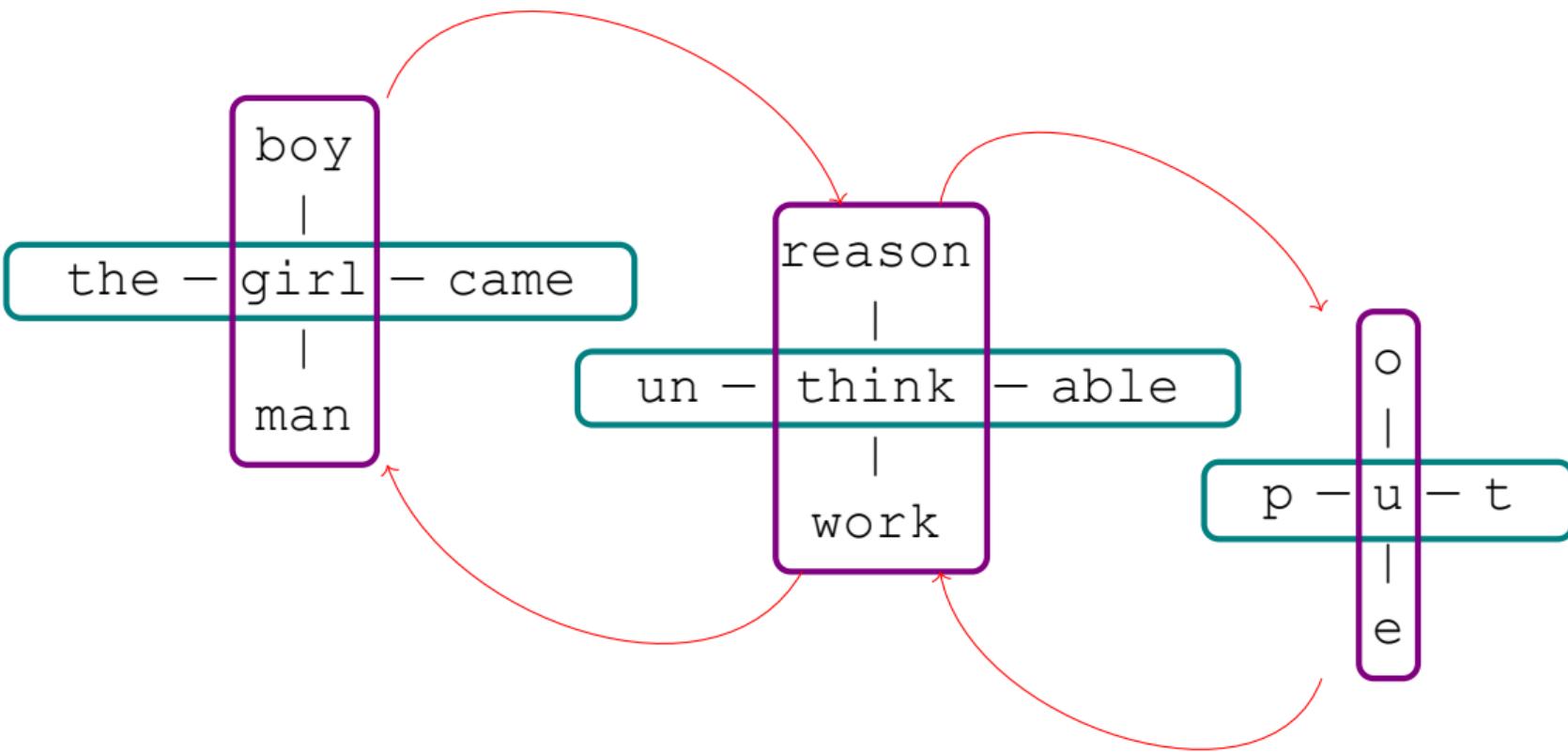
(Hjelmslev, 1971)

Syntagmes et Paradigmes



(Hjelmslev, 1971)





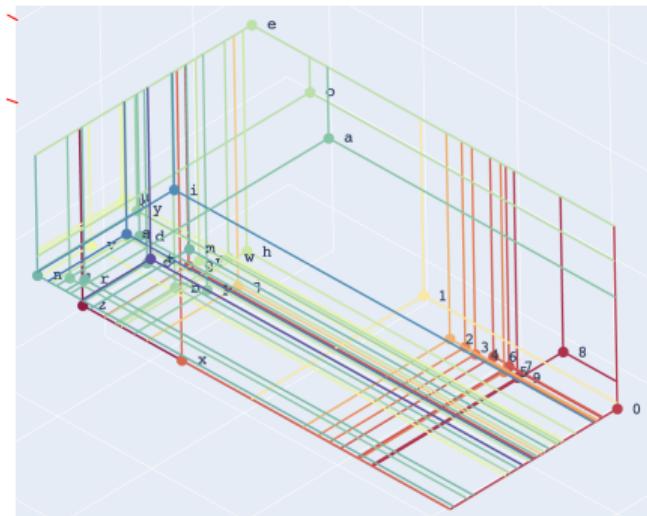
Axe 2: Objectifs

Unités
Classes
Relations

Sémantique
Syntaxe
Morphologie
Phonologie



Structure



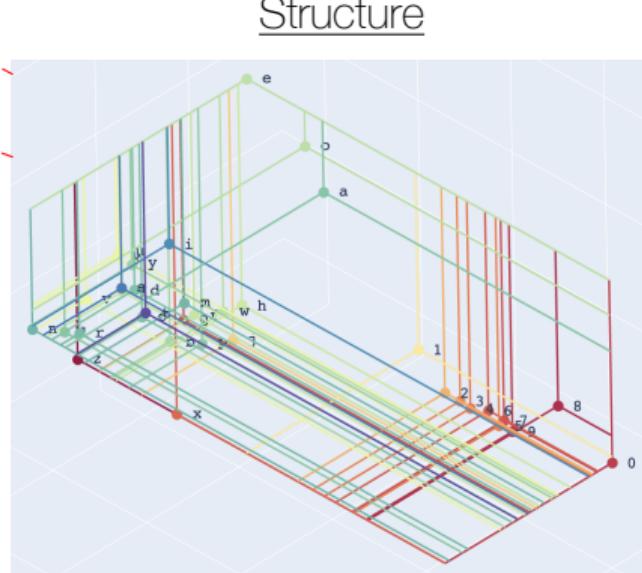
Axe 2: Objectifs

Unités
Classes
Relations

Sémantique
Syntaxe
Morphologie
Phonologie

1. Vocalic/Non-vocalic	o a e u ə i ɪ ɔ ʌ ʃ k ʒ ʒ 2 g m f p v b n s ə t z ʒ d h #
2. Consonantal/Non-consonantal	- - - - - + - - -
3. Compact/Diffuse	+ + + - - - + + + + + + + + + + + + + + + + + + + - - -
4. Grave/Acute	+ + - + + - + + + + + + + + + + + + + + + + + + - - -
5. Flat/Plain	+ - - + - - + - - + - - + - - + - - + - - + - - + - -
6. Nasal/Oral	+ - - - - - + - - - - + - - - + - - + - - + - - + - -
7. Tense/Lax	+ + + - - - + + - - + + - - + + + - - + + + - - + - -
8. Continuant/Interrupted	+ - - + - - + - + - + + + - + + + + + - - + - - + - -
9. Strident/Mellow	+ - - + - - + - + - + - + - + - + - + - + - + - + - -

(Jakobson et al., 1952)



Compétences interdisciplinaires

- ◊ Doctorats en Philo et en Info (en cours)
- ◊ Publications en Philo, Info et Maths
(Phil&Tech, Minds and Machines, Synthese)
(ACL, ICLR, ICML, AMS)
- ◊ Conférences invitées
Montréal, NYC, Cambridge, Montpellier,
Singapour (6 derniers mois)
Venice (keynote), Dagstuhl (2 mois prochains)

Gestion de la recherche

- ◊ Directeur du Dép. de Rech. (MO.CO.ESBA)
- ◊ Directeur executif du Turing Center (ETH)
- ◊ Marie Skłodowska-Curie Fellow
- ◊ Projet "Human Forms"
(soumis à E. Schmidt Foundation)
- ◊ Cluster "Foundations of AI"
(CUNY, Simons Foundation)

Participation dans la communauté scientifique

- ◊ (Vice)-Président de HaPoC
- ◊ Évaluateur pour Horizon Europe (MSCA)
- ◊ Reviewer (Nature SR, Phil&Tech, HSSC, ACL)

Enseignement et encadrement

- ◊ Enseignement interdisciplinaire international
(Argentine, France, Tchéquie, Suisse)
- ◊ Encadrement d'étudiants
en Philosophie, Informatique, Art (L, M)

LIPN, UMR 7030 (Paris)

- ◊ Équipe LoCal (Logique et Calcul)
- ◊ Accent sur les **fondements**
(théorie des types, théorie de catégories, TAL)
- ◊ Rapprochement de différentes équipes
(eg. axe "Sc. des données")
- ◊ Forte interdisciplinarité
(santé, linguistique, physique, philosophie)

LIRMM, UMR 5506 (Montpellier)

- ◊ Équipe **TEXTE** (Exploration et exploitation de données textuelles)
- ◊ Accent sur les **applications**
(grammaires catégorielles, TAL, th. des types)
- ◊ Activités **transversales**
(eg. axe "IA et Sc. des données")
- ◊ Forte interdisciplinarité
(projet Muse: "Nourrir, Soigner, Protéger")

Dans les deux cas

- ◊ Collaboration et contact avec des membres et la direction
- ◊ Intégration des aspects **épistémologiques** et **sociétaux** dans la recherche
- ◊ Présentation de mon travail aux équipes

Références |

- Bourdieu, P. (1979). *La distinction: Critique sociale du jugement*. Éditions de Minuit.
- Bourdieu, P. (1994). *Raisons pratiques: Sur la théorie de l'action*. Éditions du Seuil.
- Foucault, M. (1966). *Les mots et les choses : Une archéologie des sciences humaines*. Gallimard.
- Gastaldi, J. L. (2014, September). *Une archéologie de la logique du sens : arithmétique et contenu dans le processus de mathématisation de la logique au XIXe siècle* (Publication No. 2014BOR30035) [Theses]. Université Michel de Montaigne - Bordeaux III. <https://tel.archives-ouvertes.fr/tel-01174485>
- 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>
- Gastaldi, J. L., Terilla, J., Malagutti, L., DuSell, B., Vieira, T., & Cotterell, R. (2024). The Foundations of Tokenization: Statistical and Computational Concerns. <https://arxiv.org/abs/2407.11606>
- Girard, J.-Y. (2006). *Le point aveugle: Cours de logique. vers la perfection*. Editions Hermann.
- Giulianelli, M., Malagutti, L., Gastaldi, J. L., DuSell, B., Vieira, T., & Cotterell, R. (2024). On the Proper Treatment of Tokenization in Psycholinguistics [To appear in the Proceedings of EMNLP 2024].
<https://arxiv.org/abs/2410.02691>
- Harris, Z. (1960). *Structural linguistics*. University of Chicago Press.
- Hjelmslev, L. (1935). *La catégorie des cas*. Wilhelm Fink Verlag.
- Hjelmslev, L. (1971). 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. (1975). *Résumé of a Theory of Language*. Nordisk Sprog-og Kulturforlag.
- Jakobson, R., Fant, G. M., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. MIT Press.

Références II

- Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). 'The whole is always smaller than its parts' - a digital test of Gabriel Tardes' monads. *The British Journal of Sociology*, 63(4), 590–615.
- Lévi-Strauss, C. (1949). *Les structures élémentaires de la parenté*. Presses Universitaires de France.
- Lévi-Strauss, C. (1962). *La pensée sauvage*. Plon.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546.
- 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.
- Spang-Hanssen, H. (1959). *Probability and structural classification in language description*. Rosenkilde; Bagger.
- 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/3f5ee243547dee91fb053c1c4a845aa-Paper.pdf
- Vieira, T., LeBrun, B., Julianelli, M., Gastaldi, J. L., DuSell, B., Terilla, J., O'Donnell, T. J., & Cotterell, R. (2024). From language models over tokens to language models over characters. <https://arxiv.org/abs/2412.03719>
- Zouhar, V., Meister, C., Gastaldi, J. L., Du, L., Sachan, M., & Cotterell, R. (2023b). Tokenization and the Noiseless Channel. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 5184–5207. <https://doi.org/10.18653/v1/2023.acl-long.284>
- Zouhar, V., Meister, C., Gastaldi, J. L., Du, L., Vieira, T., Sachan, M., & Cotterell, R. (2023a). A Formal Perspective on Byte-Pair Encoding. *Findings of the Association for Computational Linguistics: ACL 2023*, 598–614. <https://doi.org/10.18653/v1/2023.findings-acl.38>

CNRS - Concours chercheurs 2025
CR Section 53 - Concours n° 53/03

*Épistémologie des modèles distributionnels de langage
par apprentissage machine*
Explicabilité formelle et interprétabilité théorique

Juan Luis Gastaldi

http://www.jlgastaldi.com/assets/gastaldi_cnrs_cr.pdf



De l'algèbre linéaire aux catégories

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

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

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

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

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

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

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

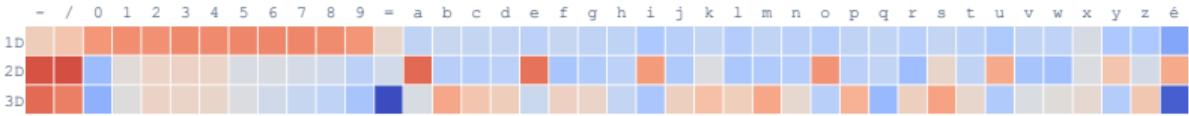
$$\mathcal{M}^* \mathcal{M}_*: (\text{Set}^D)^{\text{op}} \rightarrow (\text{Set}^D)^{\text{op}}$$

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

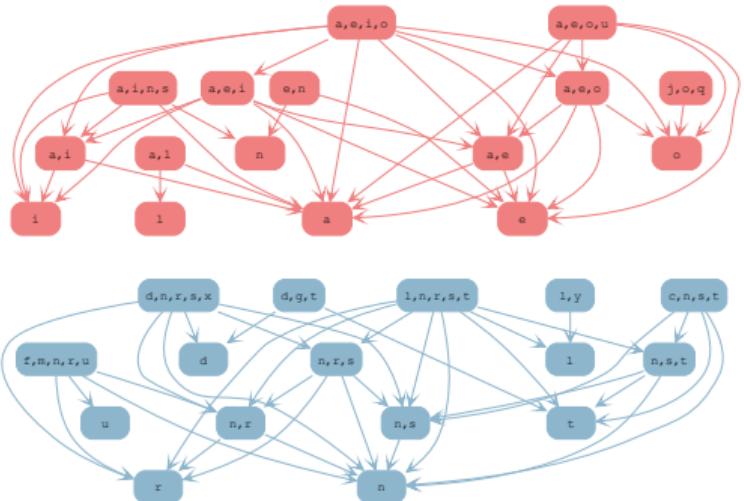
$$\text{Fix}(\mathcal{M}^* \mathcal{M}_*) := \{g \in (\text{Set}^D)^{\text{op}} \mid \mathcal{M}^* \mathcal{M}_*(g) \cong g\}$$

Structures catégoriques

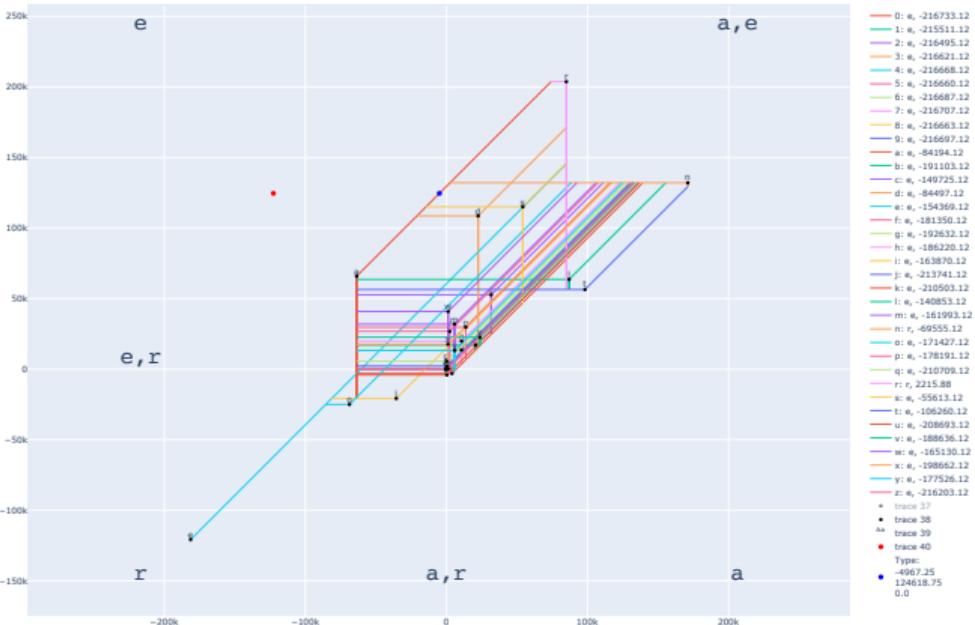
Algèbre linéaire



$$\mathcal{M}^*: \textcolor{red}{2^C}^{\text{op}} \leftrightarrows (\textcolor{red}{2^D})^{\text{op}}: \mathcal{M}_*$$



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



Théorie des types computationnels

Definition (Polaire/Orthogonal - Girard, 2006)

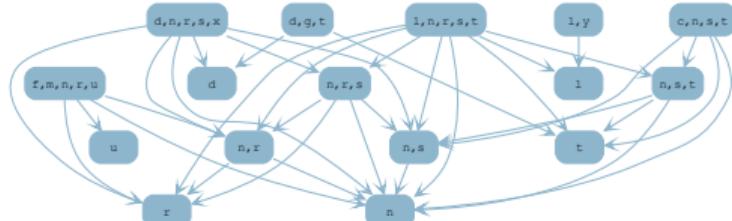
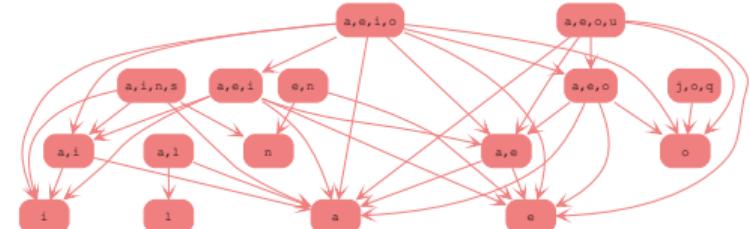
[É]tant donnée une fonction binaire

$a, b \rightsquigarrow \langle a|b \rangle : A \times B \rightarrow C$ et un sous-ensemble $P \subset C$ (le « pôle »), on peut définir le *polaire* $X^\perp \subset B$ d'un sous-ensemble $X \subset A$ (resp. $Y^\perp \subset A$ d'un sous-ensemble $Y \subset B$) par :

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

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

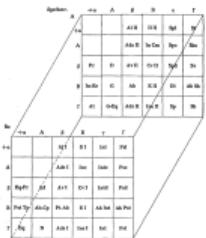
- ◊ L'application « polaire » est décroissante: $X \subset X' \Rightarrow X'^\perp \subset X^\perp$.
- ◊ L'ensemble $\text{Pol}(A) \subset \mathcal{P}(A)$ des ensembles *polaires*, i.e., de la forme Y^\perp , est stable par intersections arbitraires. En particulier, A est polaire et $X^{\perp\perp}$ est le plus petit ensemble polaire contenant X .
- ◊ En conséquence, $X^{\perp\perp\perp} = X^\perp$.



Axe 2: Interprétabilité théorique

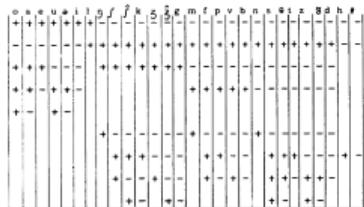
Hypothèse distributionnelle

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



(Hjelmslev, 1935)

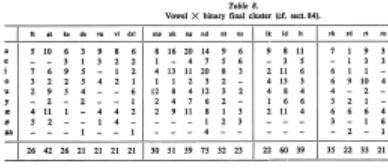
1. Vocalic/Non-vocalic
2. Consonantal/Non-consonantal
3. Compact/Diffuse
4. Grave/Acute
5. Flat/Plain
6. Nasal/Oral
7. Tense/Lax
8. Continuant/interrupted
9. Strident/Mellow



(Jakobson et al., 1952)



(Harris, 1960)



(Spang-Hanssen, 1959)

Hypothèse structurale

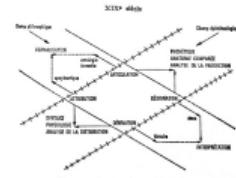
Le contenu linguistique est l'effet d'une structure virtuelle dérivée des pratiques linguistiques dans une communauté.

Répondeur	Filtre	Échelle	Dimension	Faculté
Prix	+	+	—	
Prix, infant	—	+	—	
Beauté	—	+	—	—
Prise de la parole	—	+	—	
Signification de l'acte	—	+	—	
Acte	—	+	—	
Clef	—	+	—	
Progrès	—	+	—	
Signification de l'acte	—	+	—	
Mérite à des fins extrinsèques	—	+	—	
Autonomie	—	+	—	
Obéissance	—	+	—	
Signification de l'acte	—	+	—	
Progrès de l'individu	—	+	—	
Vérité	—	+	—	

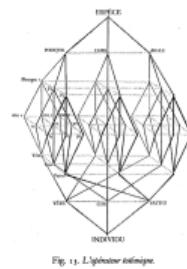
(Lévi-Strauss, 1949)



(Bourdieu, 1979)



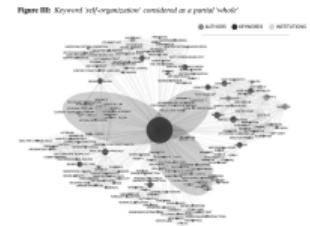
(Foucault, 1966)



(Lévi-Strauss, 1962)



(Bourdieu, 1994)



(Latour et al., 2012)

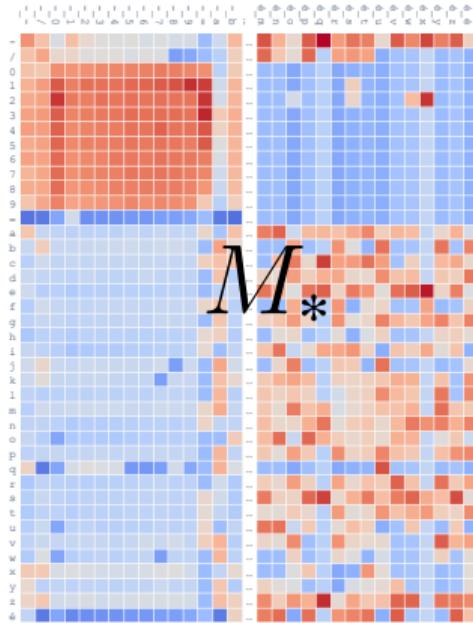
Axe 1: Objectifs

- ◊ Short Term
 - ◊ Systematizing, Formalizing, and Testing
 - ◊ Enriching Over $\bar{\mathbb{R}}$
 - ◊ Defining a Product over the Nucleus
- ◊ Medium Term
 - ◊ Characterizing Tokenization from a Structural Standpoint
 - ◊ Scaling for Real World Data
- ◊ Long Term
 - ◊ Assessment of state-of-the-art DNN language models in the light of the proposed formal framework (performance and efficiency comparison, limit properties, verification guarantees).
 - ◊ Exploration of other base categories and measures.
 - ◊ Further optimization.
 - ◊ Development of software packages and a usable integrated software framework for the scientific community.

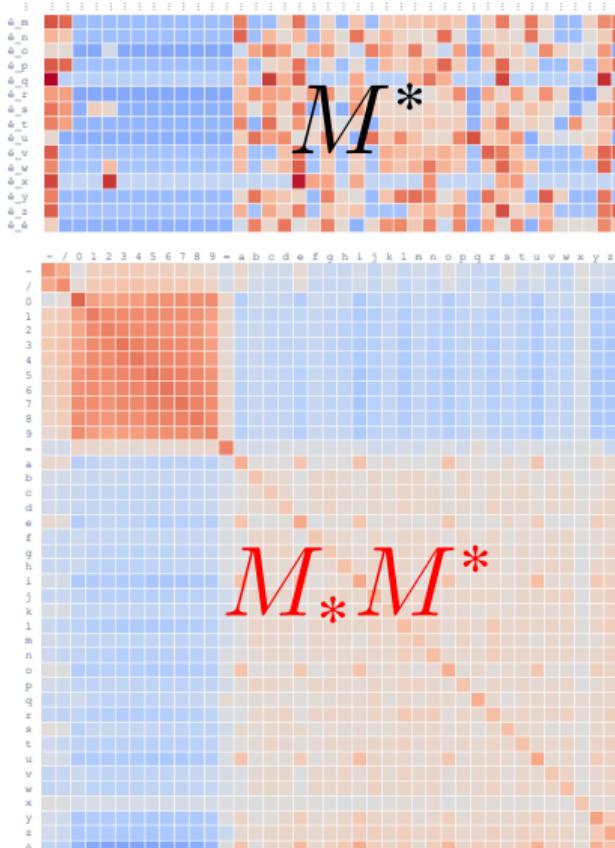
Axe 2: Objectifs

- ◊ Short Term
 - ◊ Characterizing Epistemological Stakes through a Historical Lens
 - ◊ Interpreting Paradigms as Types
 - ◊ Addressing Semantic Aspects
- ◊ Medium Term
 - ◊ Revisiting the Theory of Distinctive Features and Exploring Possible Extensions
 - ◊ Studying the Syntactic properties of the Nucleus' compositional Structure
- ◊ Long Term
 - ◊ The study of pragmatic limits to the formal content captured by the set of types definable through a structural analysis of linguistic data.
 - ◊ The evaluation of the effects on ML models of the bias towards written language.
 - ◊ The assessment of other classically structuralist principles such as segmentation and double patterning.
 - ◊ A critique of the Symbolist-Connectionist debate in light of the results established in this program.
 - ◊ The exploration of possible applications of the framework proposed to non-linguistic data (sounds, images, formal languages, etc.)
 - ◊ The study of the conditions for the generalization of the use of the model proposed across scientific disciplines, specially in the Social Sciences and the Humanities.

$M_* M^*$ comme matrice de covariance



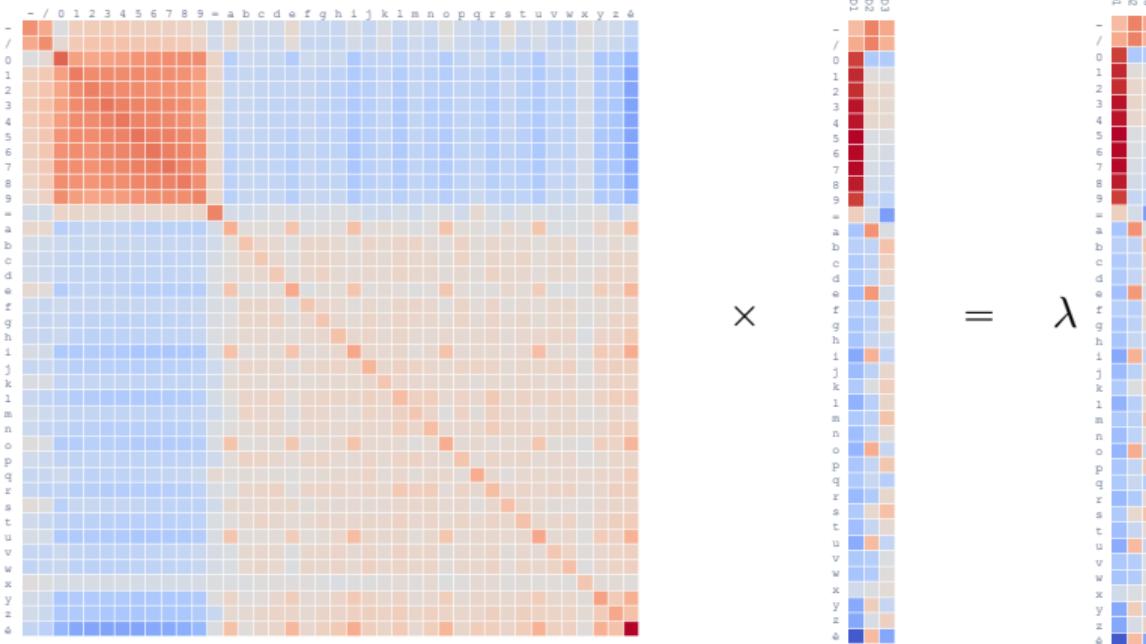
M_*



$M_* M^*$

Vecteurs propres comme points fixes

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

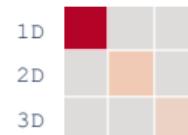


Traits structuraux

Eigenvectors of $M_* M^*$:



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

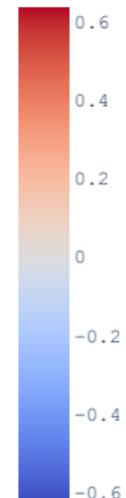


Eigenvectors of $M^* M_*$:



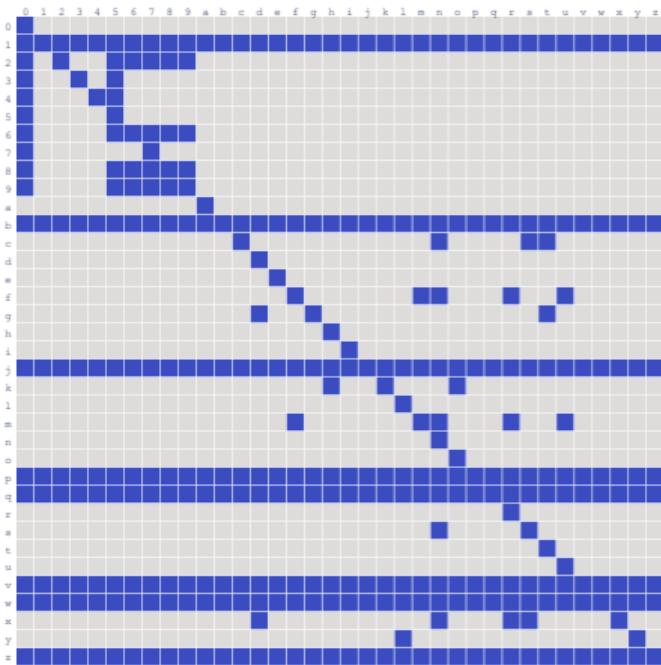
Mots

	-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



Points fixes booléens

$$M_i^* M_*^i \textcolor{blue}{d} = d$$



★

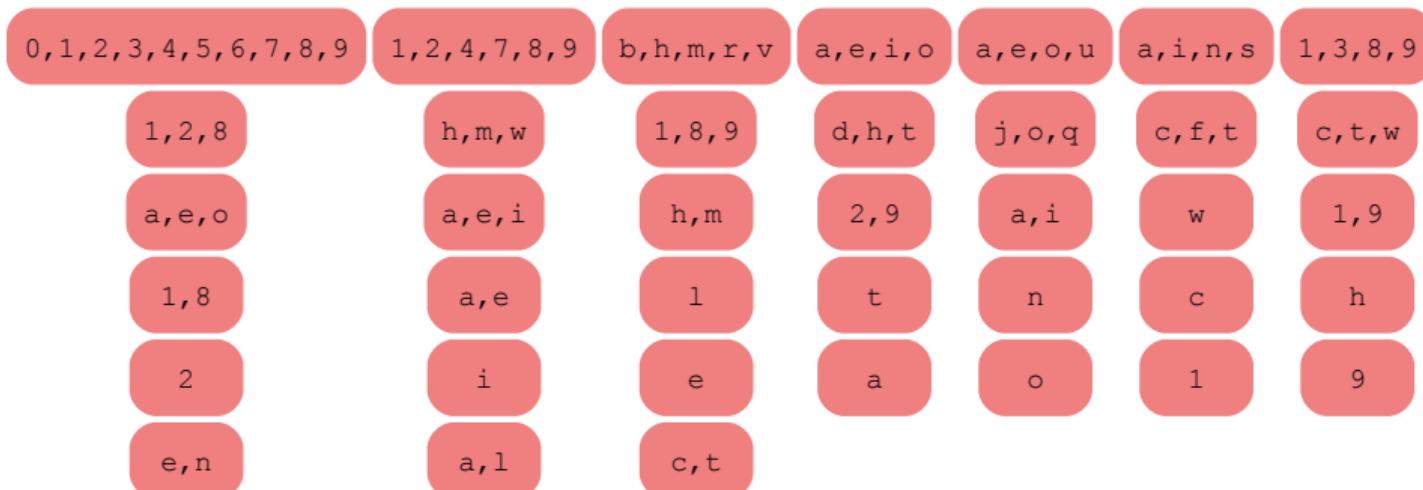


?



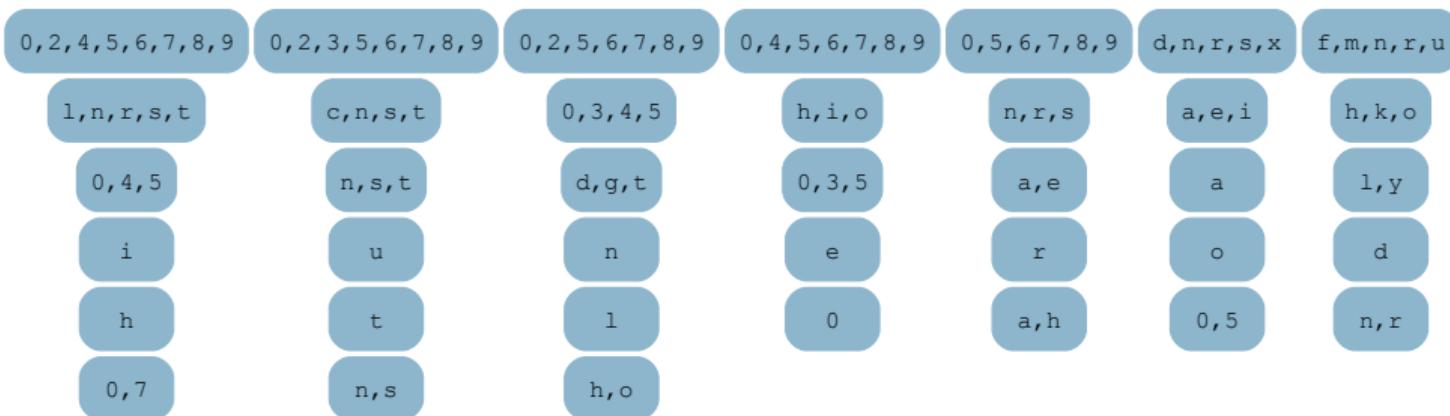
“Eigensets”

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

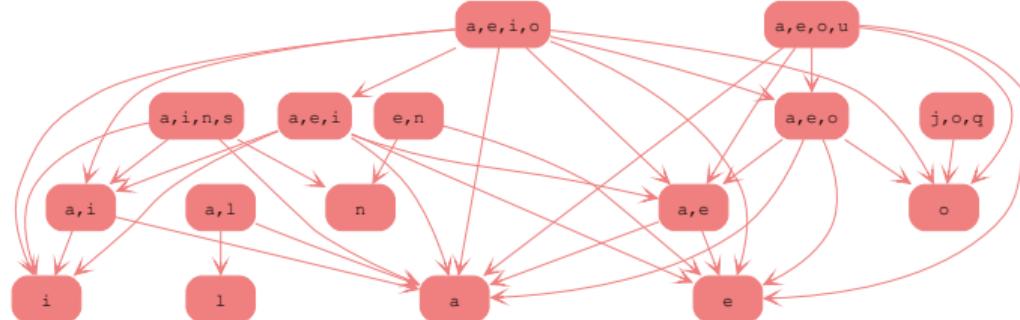
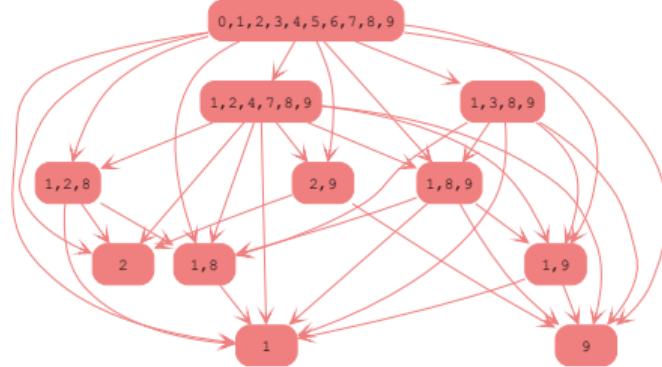
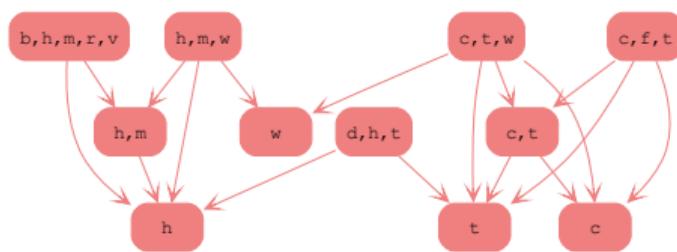


“Eigensets”

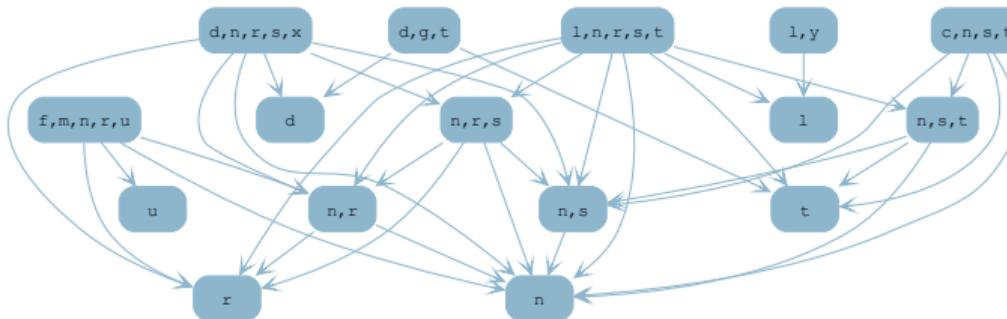
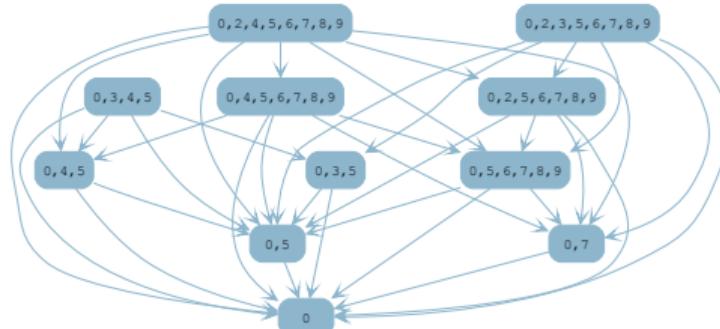
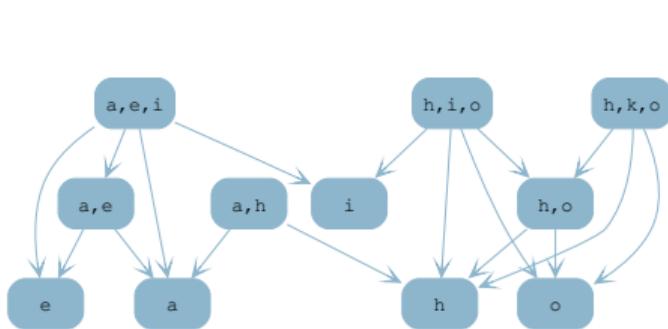
$$M_i^* M_*^i \mathbf{d} = \mathbf{d}$$



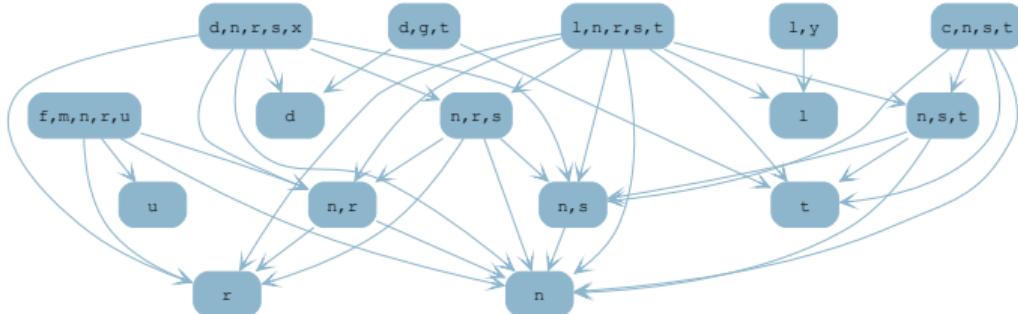
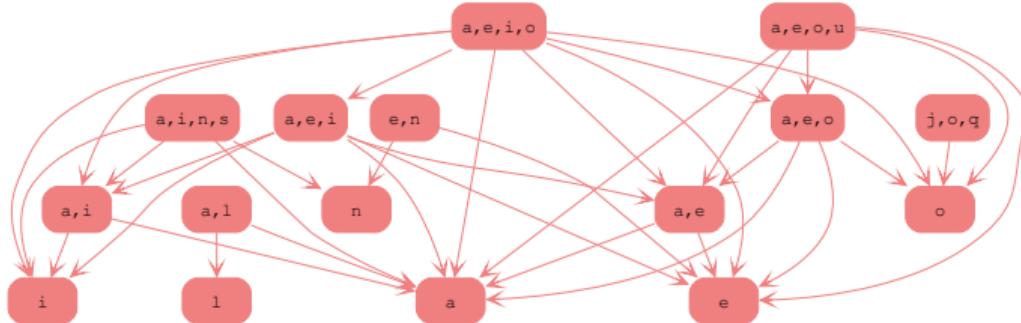
Quelle Structure?



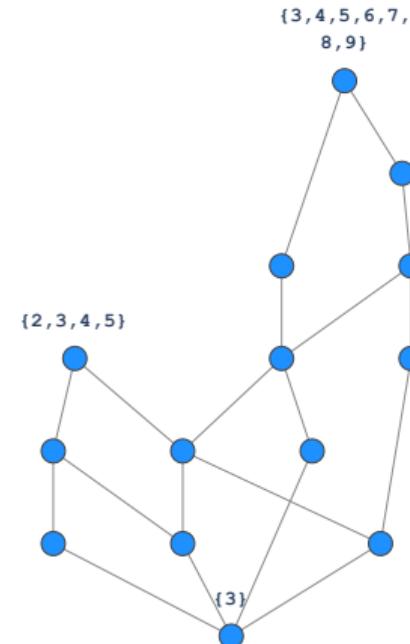
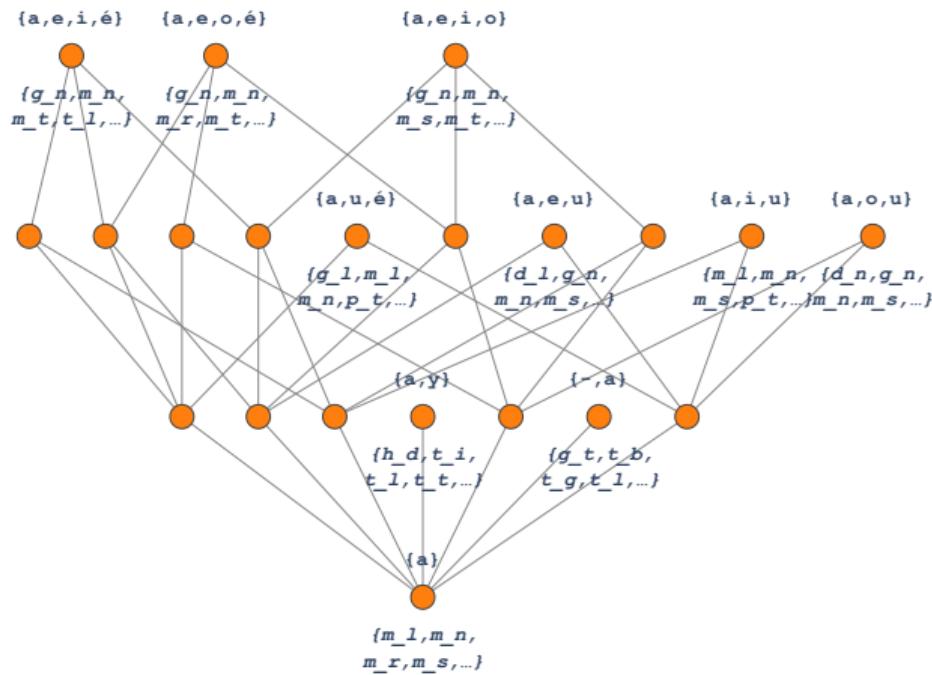
Quelle Structure?



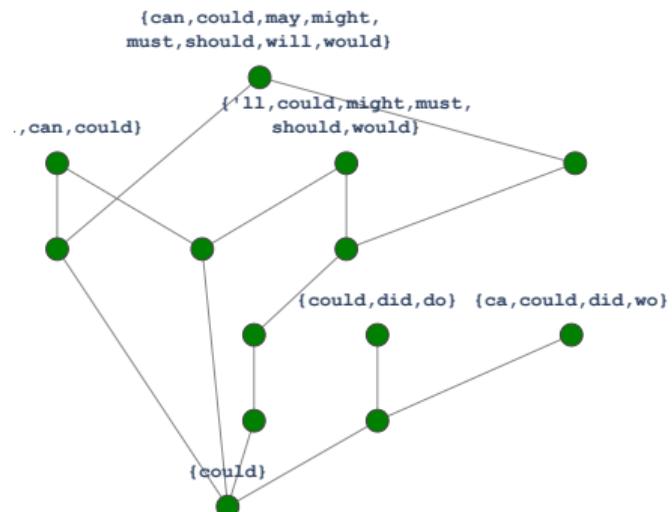
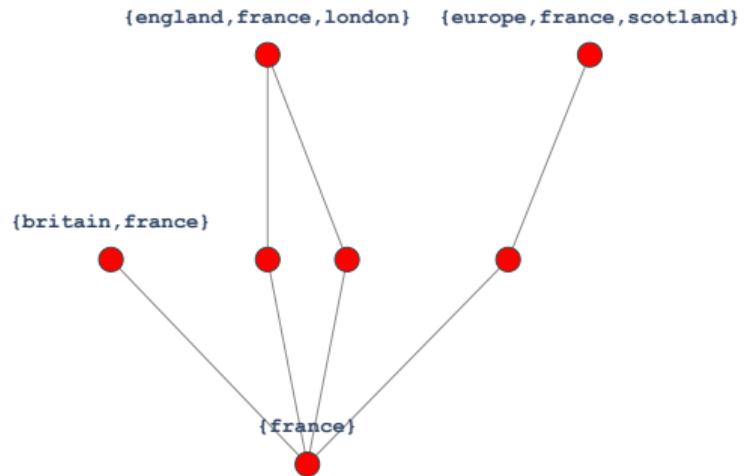
Quelle Structure?



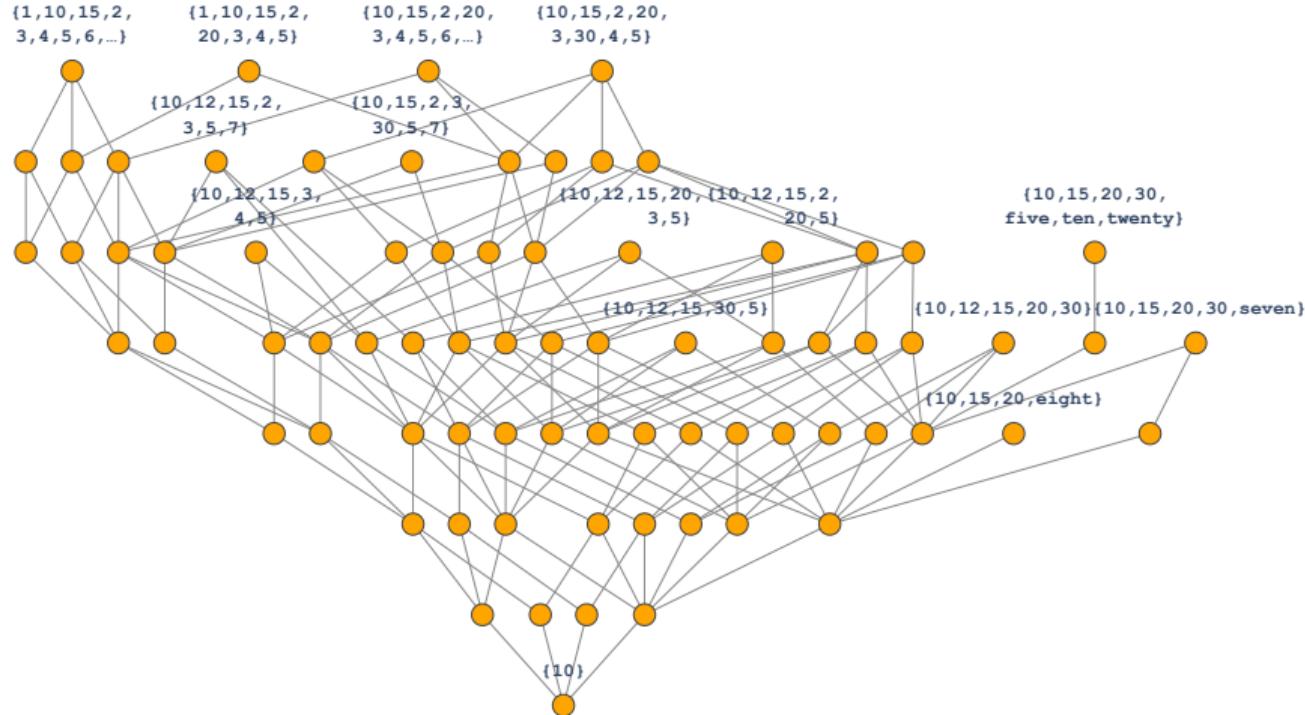
Concepts formels



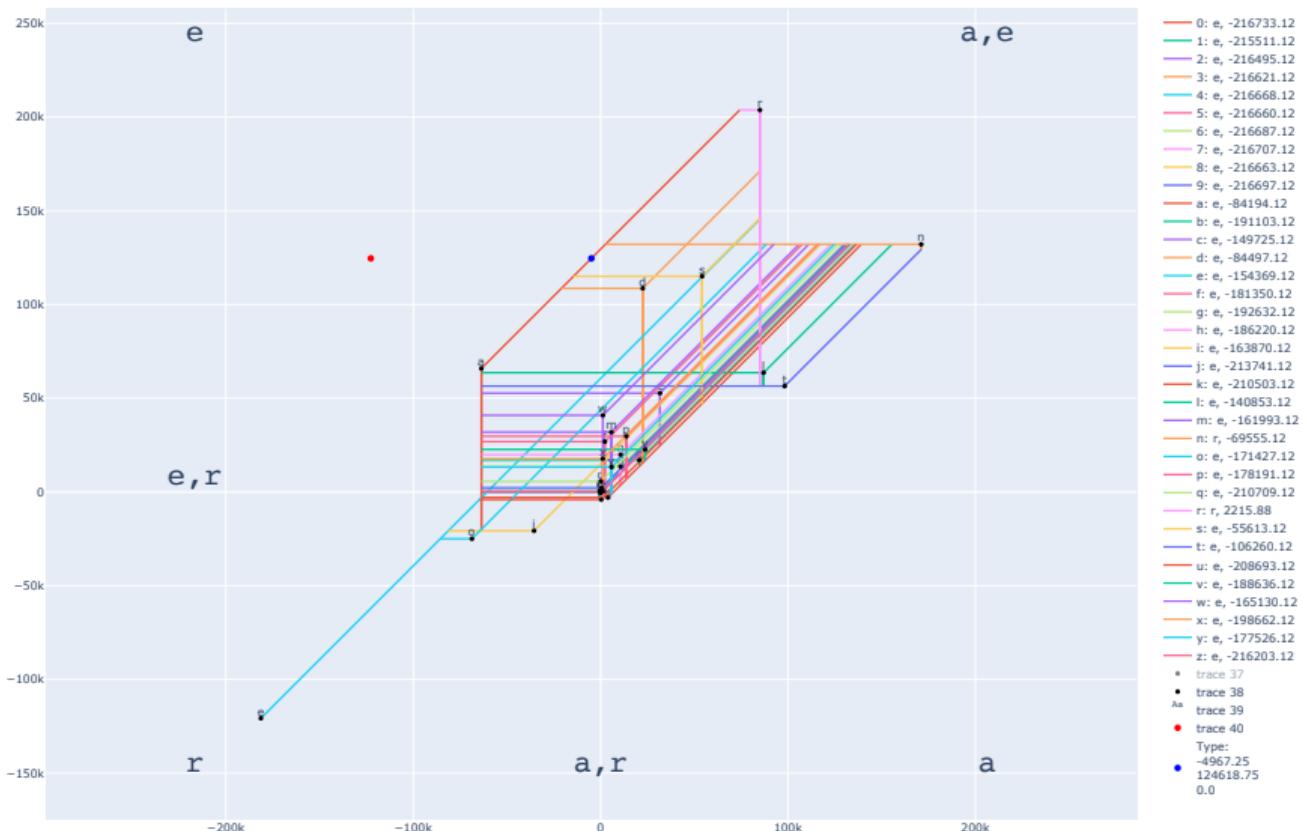
Concepts formels (mots)



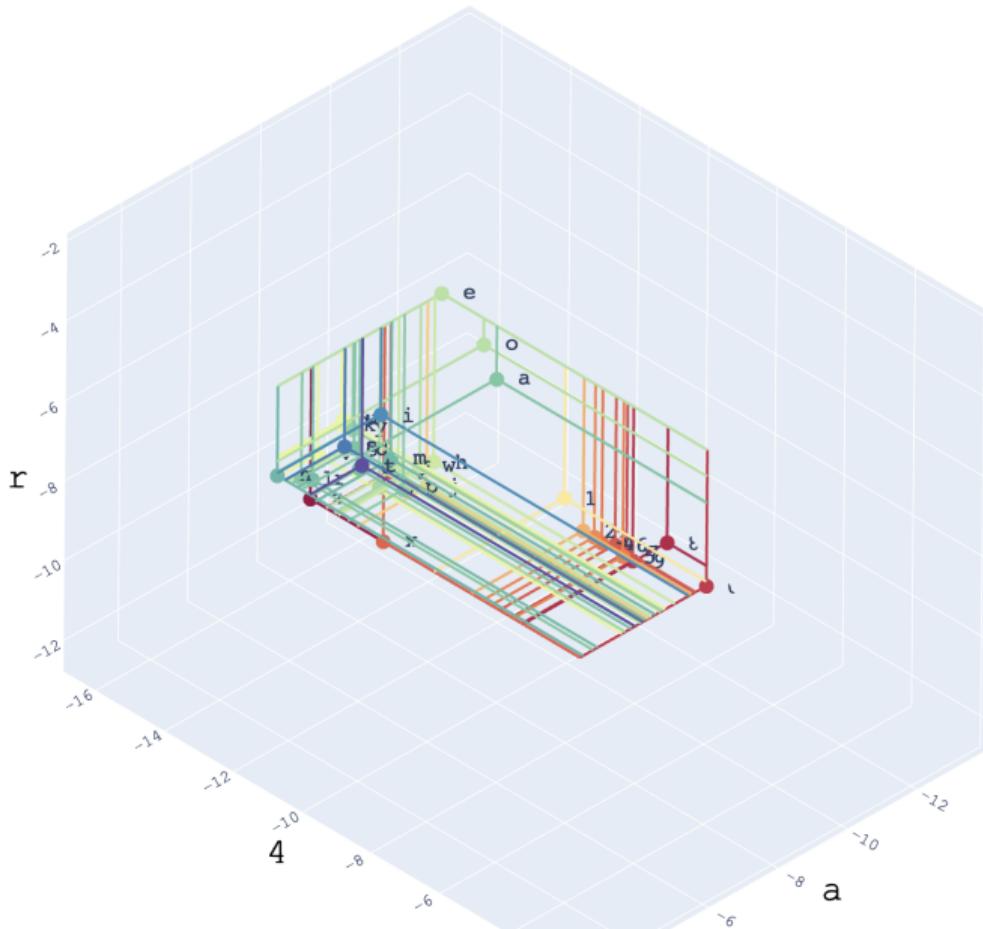
Concepts formels (mots)



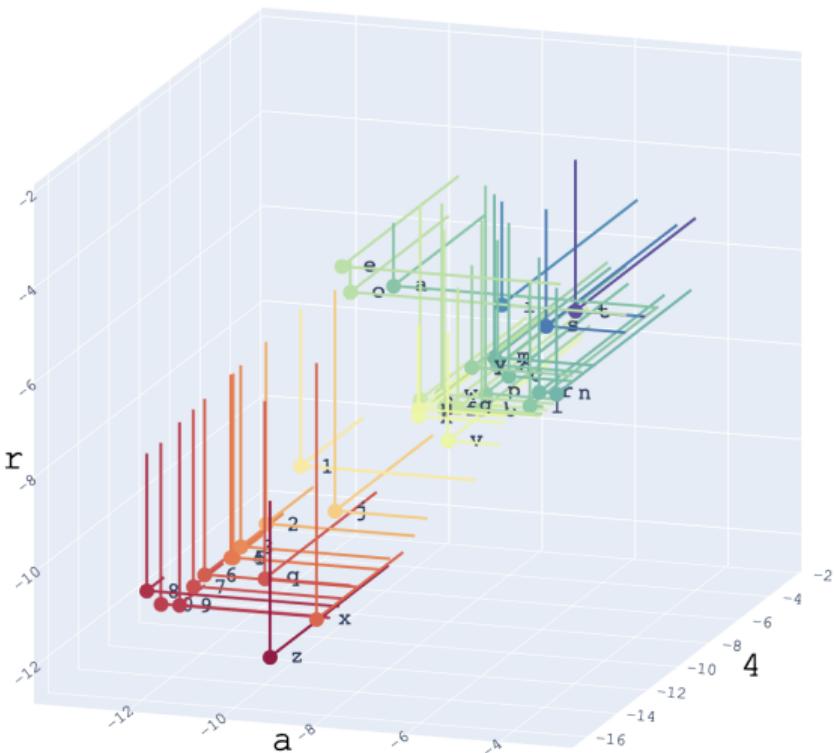
Structure interne du noyau



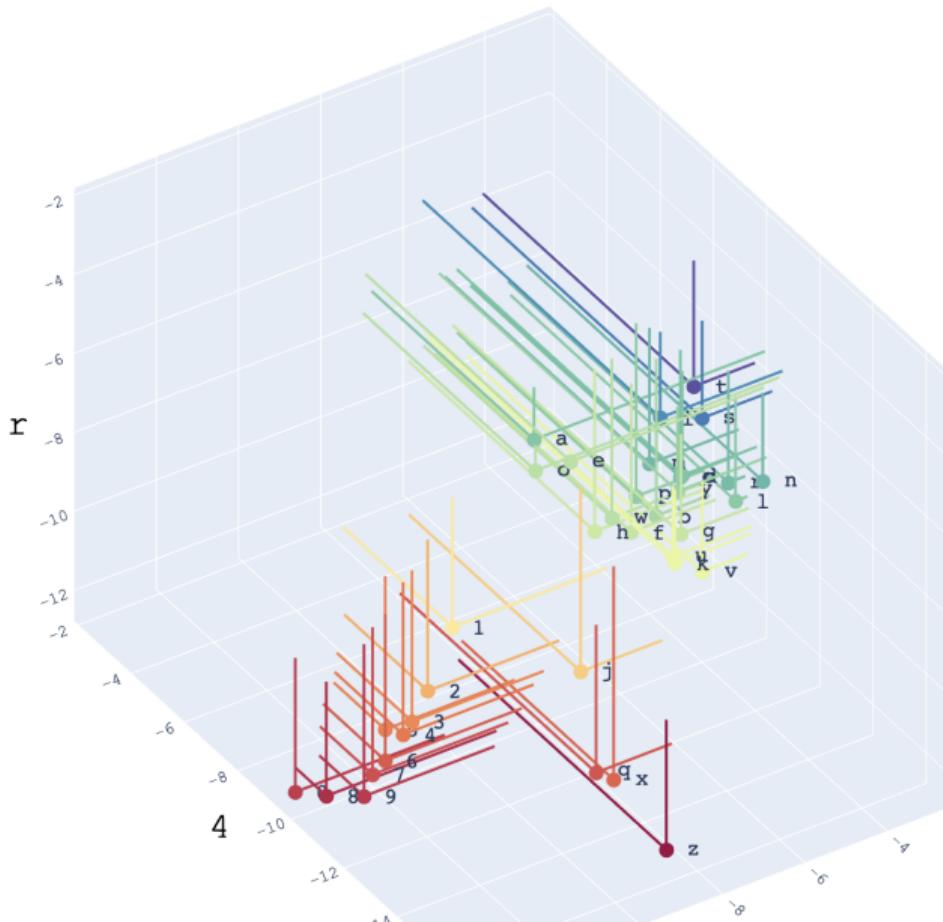
Structure interne du noyau



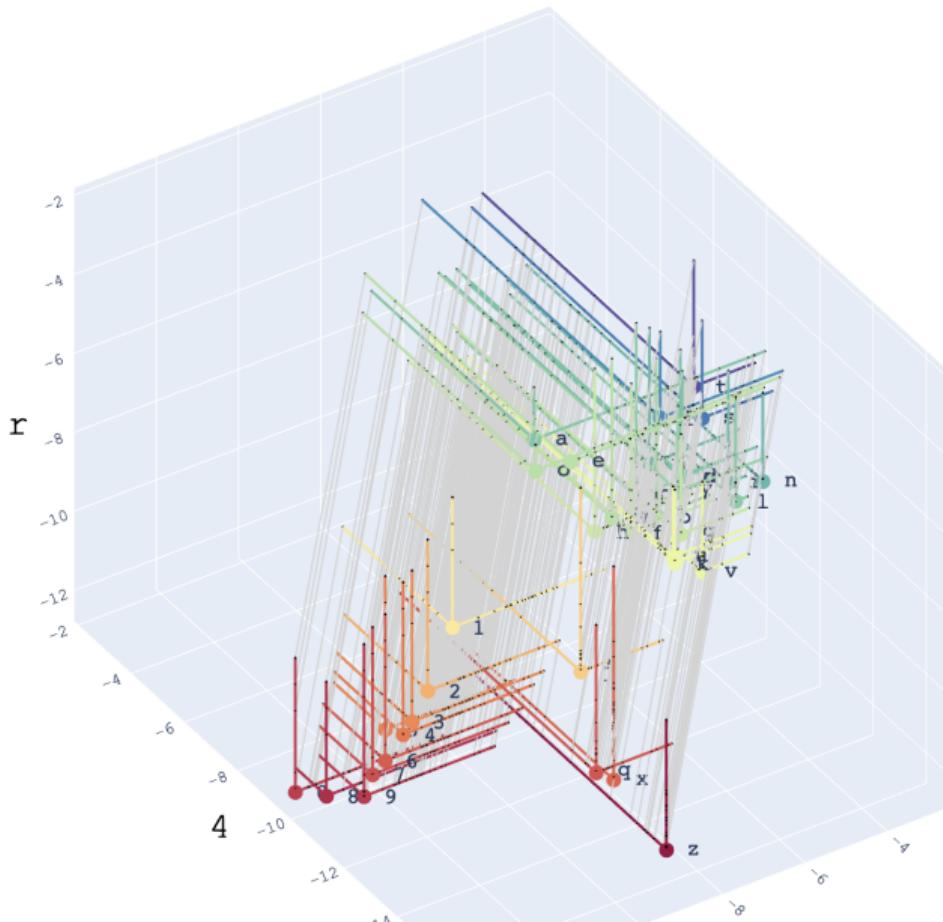
Structure interne du noyau



Structure interne du noyau



Structure interne du noyau



Matrice et analogie

a = your
c = my

w = apartment
x = house
y = chair
z = stool

your : house
my : apartment

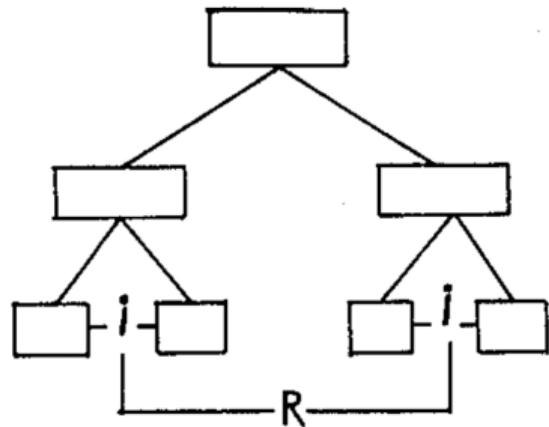
	...	w	x	y	z	...
...	...	0	0	0	0	...
a	...	0	1	1	0	...
b	...	0	0	1	1	...
c	...	1	0	0	1	...
...	...	0	0	0	0	...

Une **sémiotique** [...] est une hiérarchie dont chacune des composantes admet une analyse ultérieure en classes définies par relation mutuelle, de telle sorte que chacune de ces classes admette une analyse en dérivés définis par mutation mutuelle.

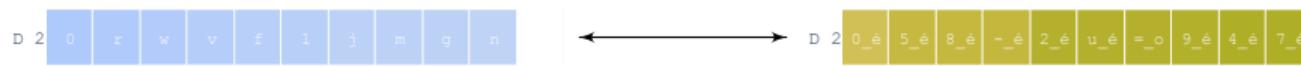
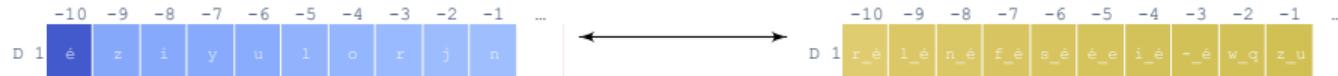
(Hjelmslev, 1975, Df. 24)

Une **mutation** [...] est une fonction existant entre des dérivés du premier degré d'une seule et même classe, une fonction qui a une relation à une fonction entre d'autres dérivés de premier degré d'une seule et même classe et appartenant au même rang.

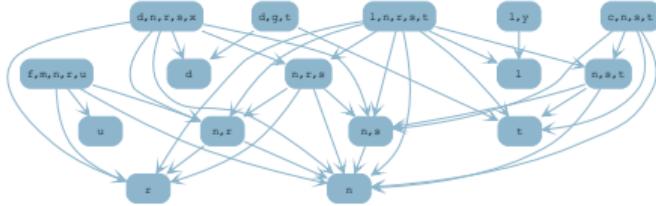
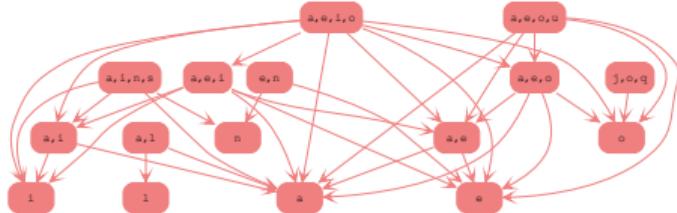
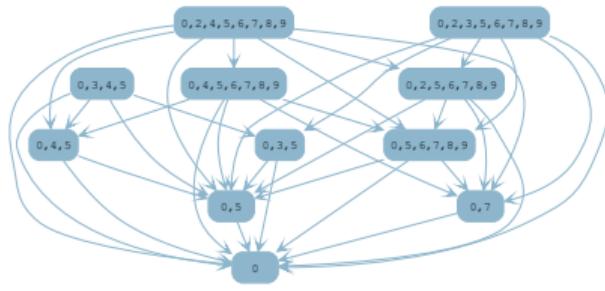
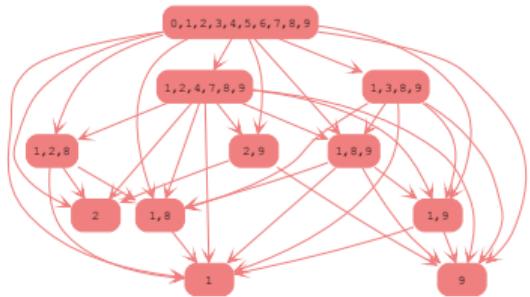
(Hjelmslev, 1975, Df. 23)



Syntagmatique et Texte (Vecteurs)



Syntagmatique et Texte (Noyau/Types)



Paradigmatique et Langue (Vecteurs)

-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	...
D 1	é	z	i	y	u	l	o	r	j	n
...	1	2	3	4	5	6	7	8	9	10
	0	9	2	1	8	4	3	7	6	5

-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	...
D 1	x_é	l_é	n_é	f_é	s_é	e_é	i_é	u_é	w_q	z_u
...	1	2	3	4	5	6	7	8	9	10
	=_9	=_6	7_-9	9_-9	=_5	9_-5	9_-7	9_-6	9_-8	9_0

0	r	w	v	f	l	j	m	g	n	...
D 2	3	y	u	é	i	o	e	a	-	/
	z	p	f	m	g	t_g	é_m	z_m	z_g	z_q

0_é	5_é	8_é	-_é	2_é	u_é	=_o	9_é	4_é	7_é	...
D 2	d_m	z_p	z_f	k_m	r_g	t_g	é_m	z_m	z_g	z_q
	z	p	f	m	g	t_g	é_m	z_m	z_g	z_q

Paradigmatique et Langue (Noyau/Types)

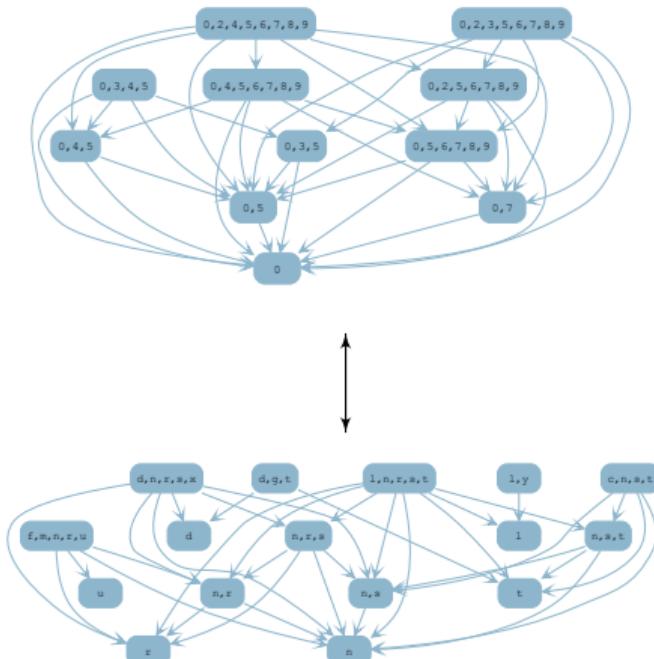
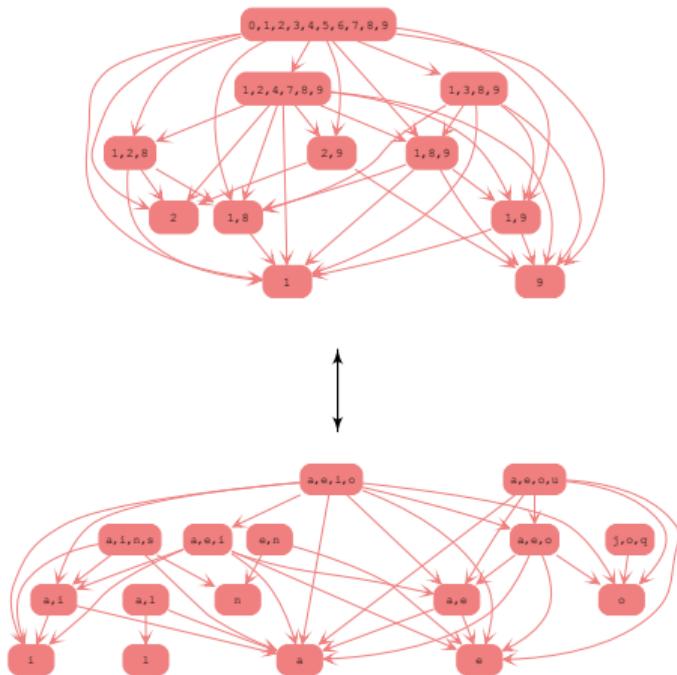


Illustration du contenu formel

(Gastaldi and Pellissier, 2021)

Characteristic Content

```
{cat, dog, spider,  
gavagai}
```

Atomic Type

Syntactic Content

"the gavagai is on the
mat"

Profunctor Nucleus

Informational Content

```
{cat:0.059%,  
dog:0.012%,  
spider:0.009%  
gavagai:0.000%}
```

Probability Distribution