

CNRS - Concours chercheurs 2025
DR Section 53 - Concours n° 53/01

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

Juan Luis Gastaldi

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



Parcours et travaux

| | |
|--------------|---|
| 1997-2007 | Recherche pré-doctorale Argentine, France (UNR, ENS, Paris 1, UPMC) |
| 2008-2014 | Recherche doctorale France (Bordeaux Montaigne) |
| 2015-2022 | Recherche post-doctorale France, Suisse, Tchéquie, USA (ETH, MSCA, CUNY, CMU) |
| 2023-Présent | Nouvelle recherche doctorale Suisse (ETH Zurich) |

Sciences Politiques

Philosophie du langage et pensée formelle

Comment une science de l'homme est-elle possible?

- Héritage **critique** kantien contre le positivisme
- Dépassement du partage **anthropologie/ontologie**
- Anthropologie ou théorie de l'**expression**
(Foucault, 1954; Deleuze, 1954)
- Alliance entre **savoirs formels** et philosophie critique contemporaine

- ◊ Gastaldi (2009, 2010, 2011, 2015, 2016b, 2019)

Parcours et travaux

1997-2007 Recherche pré-doctorale
Argentine, France
(UNR, ENS, Paris 1, UPMC)

2008-2014 Recherche doctorale
France
(Bordeaux Montaigne)

2015-2022 Recherche post-doctorale
France, Suisse, Tchéquie, USA
(ETH, MSCA, CUNY, CMU)

2023-Présent Nouvelle recherche doctorale
Suisse
(ETH Zurich)

Philosophie et histoire des sciences formelles

Quelle critique pour les sciences formelles?

- Épistémologie historique
- La notion de *formel* a elle-même une histoire (récente)
- En devenant formelle, la mathématisation n'entraîne pas la naturalisation

- ◊ Gastaldi (2014, 2016a, 2022, 2024a, 2024b, 2024d)
- ◊ Gastaldi (ed.) (2024a, 2024b)

Parcours et travaux

1997-2007 Recherche pré-doctorale
Argentine, France
(UNR, ENS, Paris 1, UPMC)

2008-2014 Recherche doctorale
France
(Bordeaux Montaigne)

2015-2022 Recherche post-doctorale
France, Suisse, Tchéquie, USA
(ETH, MSCA, CUNY, CMU)

2023-Présent Nouvelle recherche doctorale
Suisse
(ETH Zurich)

Epistémologie des modèles de langage par apprentissage machine

Quelles pratiques formelles pour une théorie critique du langage?

- Analyse computationnelle de corpus textuels
- Modèles **distributionnels** de langage

- ◊ Gastaldi (2021, 2024c), Gastaldi, Moot, and Rétoré (2024), and Gastaldi and Pellissier (2021)

Parcours et travaux

1997-2007 Recherche pré-doctorale
Argentine, France
(UNR, ENS, Paris 1, UPMC)

2008-2014 Recherche doctorale
France
(Bordeaux Montaigne)

2015-2022 Recherche post-doctorale
France, Suisse, Tchéquie, USA
(ETH, MSCA, CUNY, CMU)

2023-Présent Nouvelle recherche doctorale
Suisse
(ETH Zurich)

Fondements formels des modèles de language par apprentissage machine

Quels fondements pour une science formelle du langage?

- Fondements **formels** des pratiques computationnelles
(catégories, types)
- Fondements **structuraux** des principes distributionnels
(segmentation, structure)

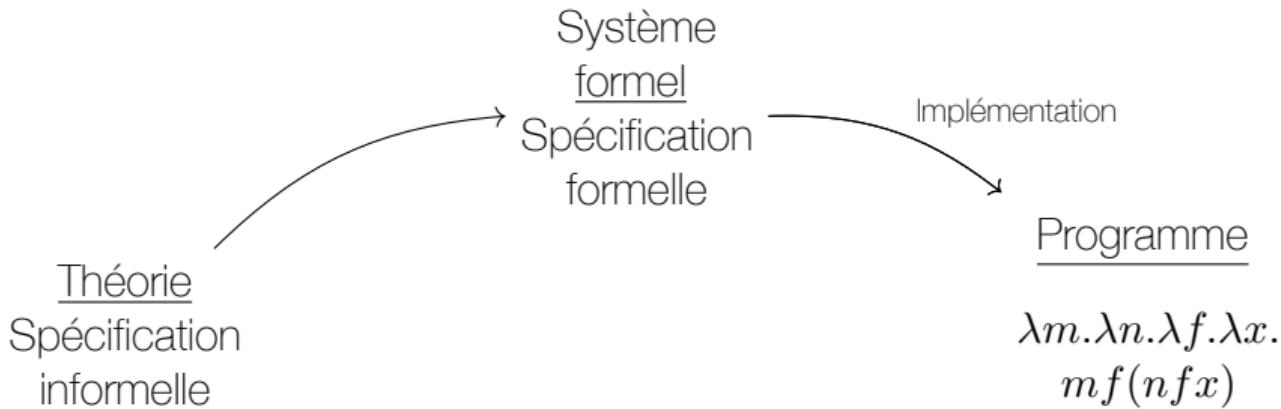
- ◊ Bradley et al. (2024), Gastaldi, Terilla, et al. (2024),
Giulianelli et al. (2024), Vieira et al. (2024), and Zouhar
et al. (2023a, 2023b)

La structure implicite des données

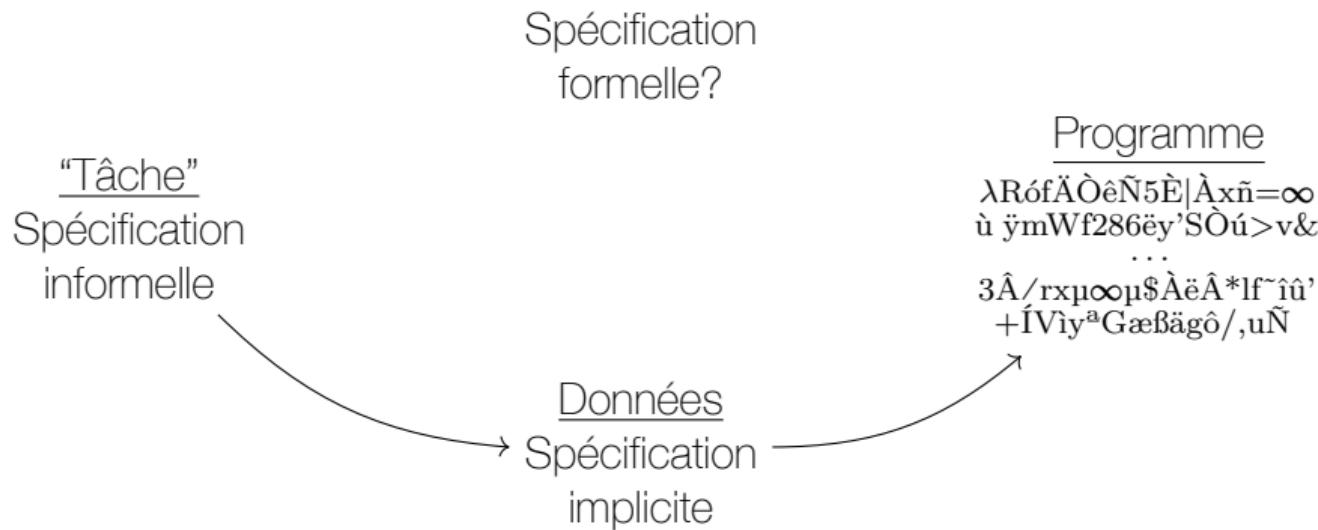
Programme

$$\begin{aligned} &\lambda m. \lambda n. \lambda f. \lambda x. \\ &mf(nfx) \end{aligned}$$

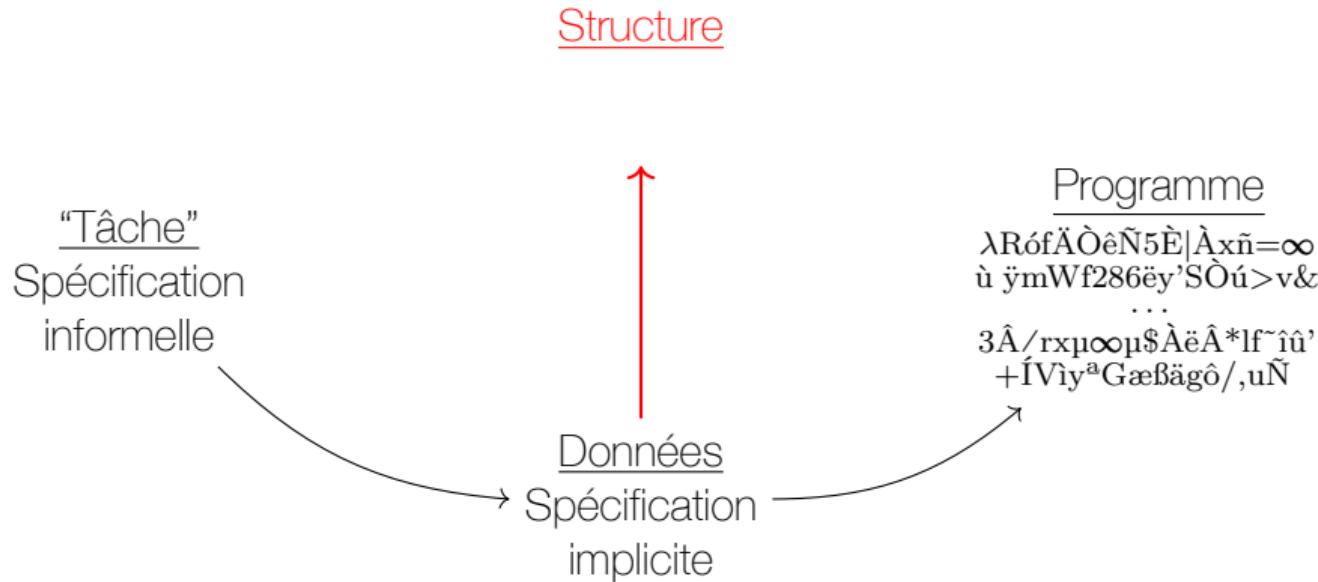
La structure implicite des données



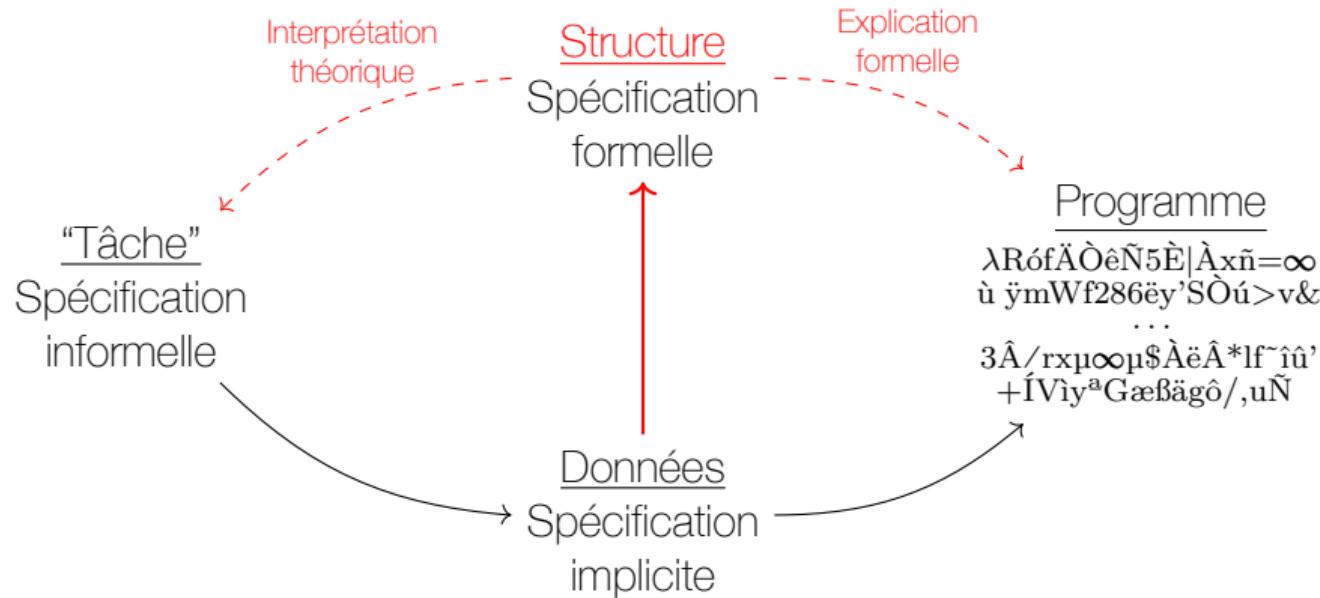
La structure implicite des données



La structure implicite des données



La structure implicite des données



Axe 1: Explicabilité formelle

Tokenisation

(Sennrich et al., 2016)

Embedding

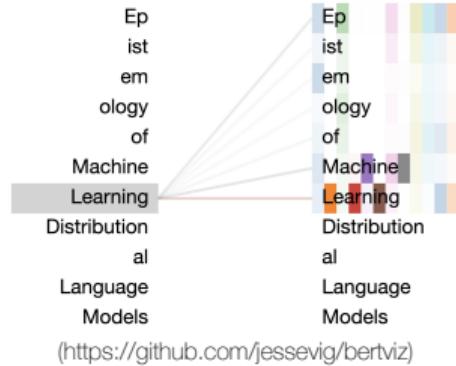
(Mikolov et al., 2013)

Attention

(Vaswani et al., 2017)

**Epistemology of Machine Learning
Distributional Language Models**

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



Axe 1: Explicabilité formelle

Tokenisation

(Sennrich et al., 2016)

Embedding

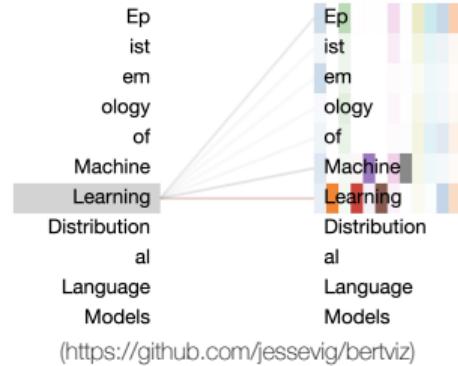
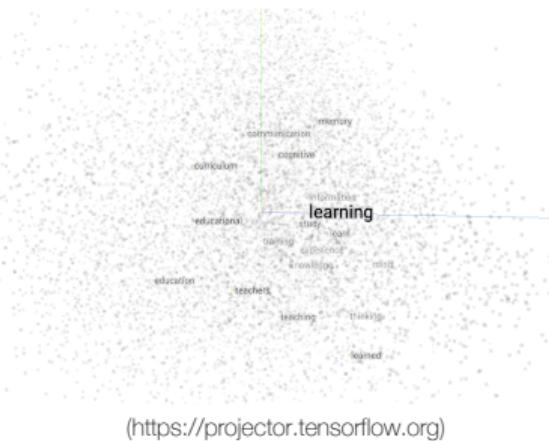
(Mikolov et al., 2013)

Attention

(Vaswani et al., 2017)

**Epistemology of Machine Learning
Distributional Language Models**

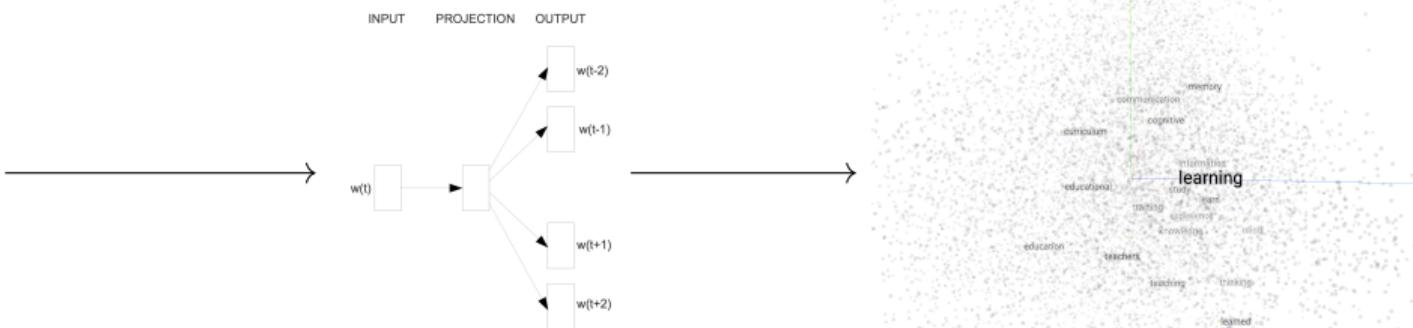
(<https://tiktokizer.vercel.app>)



Embedding et SVD

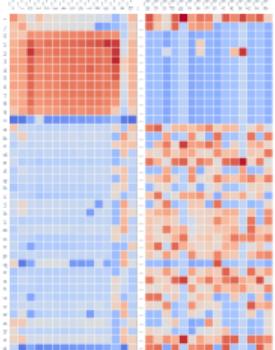


Embedding et SVD

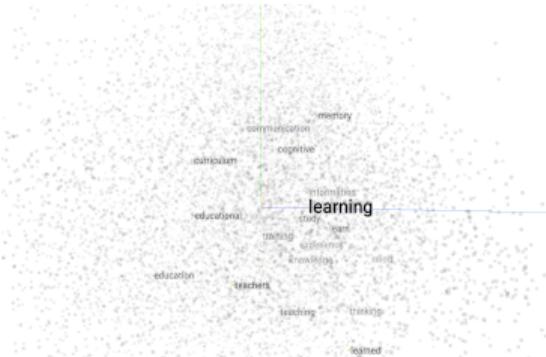
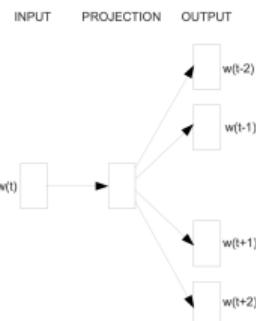


Embedding et SVD

Matrice PMI

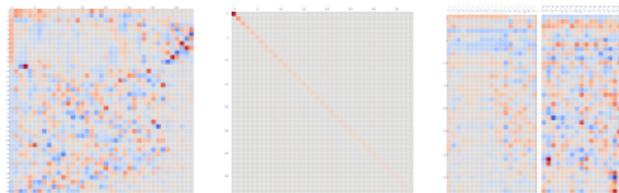
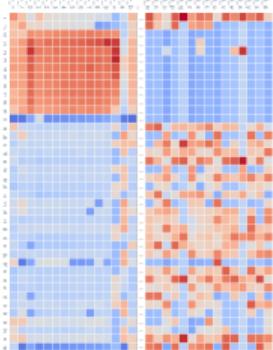


WIKIPEDIA
The Free Encyclopedia



Embedding et SVD

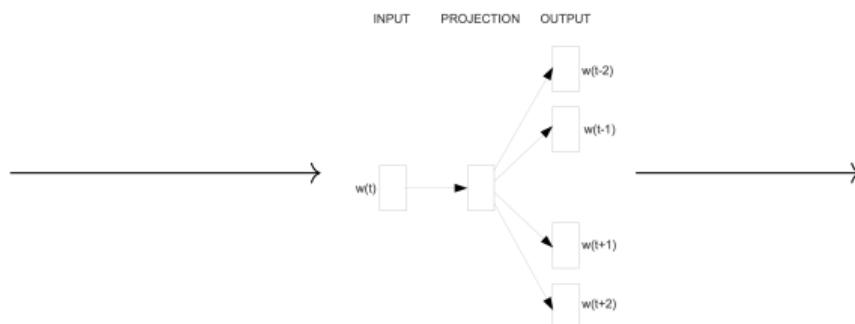
Matrice PMI



SVD

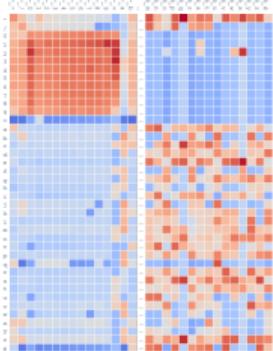


WIKIPEDIA
The Free Encyclopedia

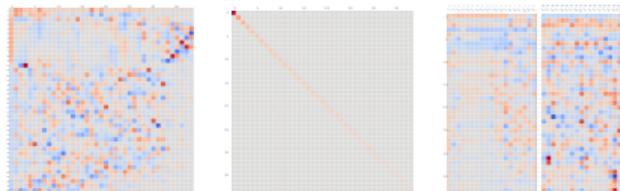


Embedding et SVD

Matrice PMI



Vecteurs singuliers

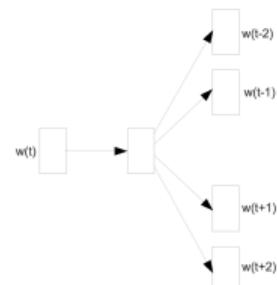


SVD



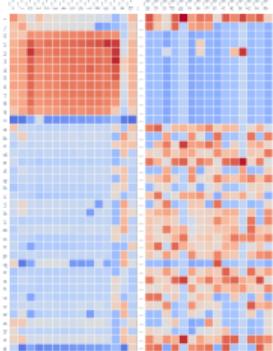
WIKIPEDIA
The Free Encyclopedia

INPUT PROJECTION OUTPUT

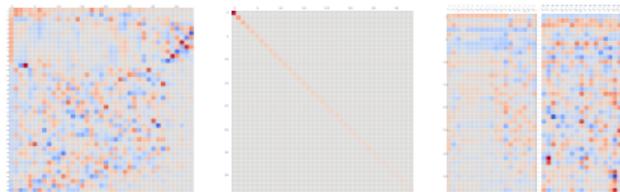
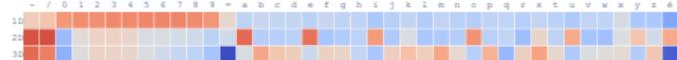


Embedding et SVD

Matrice PMI

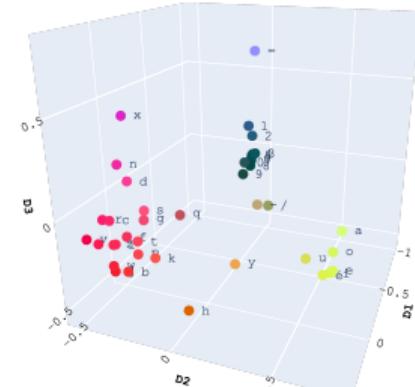


Vecteurs singuliers



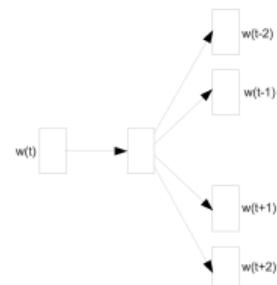
SVD

Embedding SVD



WIKIPEDIA
The Free Encyclopedia

INPUT PROJECTION OUTPUT



De l'algèbre linéaire aux catégories

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

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

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

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

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

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

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

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

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

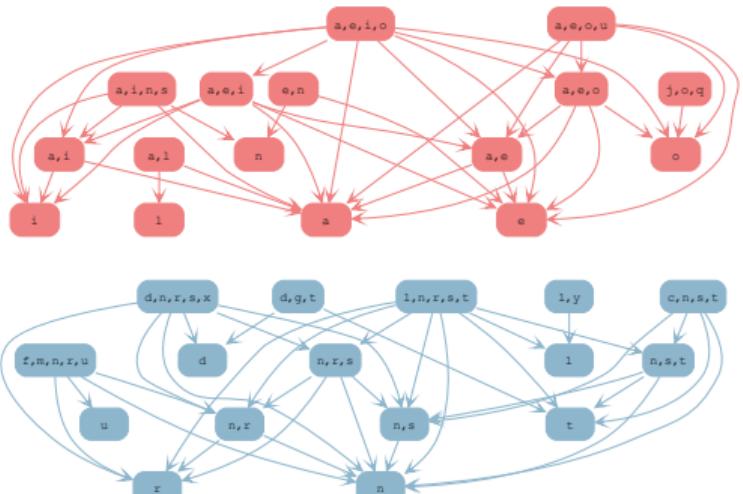
$$\text{Fix}(\mathcal{M}^* \mathcal{M}_*) := \{g \in (\mathbf{Set}^{\mathbf{D}})^{\text{op}} \mid \mathcal{M}^* \mathcal{M}_*(g) \cong g\}$$

Structures catégoriques

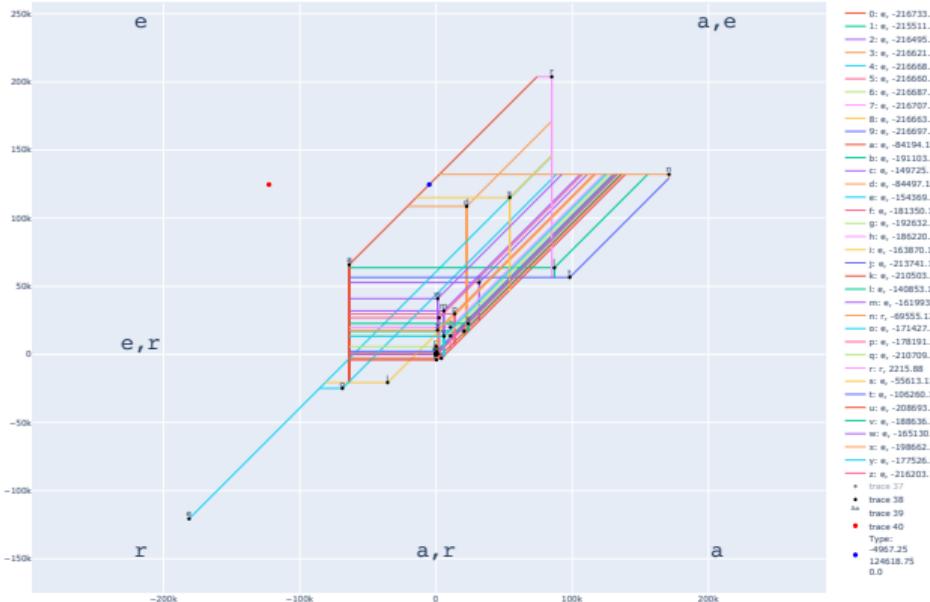
Algèbre linéaire

| - / 0 1 2 3 4 5 6 7 8 9 = | a b c d e f g h i j k l m n o p q r s t u v w x y z é |
|---------------------------|---|
| 1D | ... |
| 2D | ... |
| 3D | ... |

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



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



Axe 2: Interprétabilité théorique

Hypothèse distributionnelle

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

Hypothèse structurale

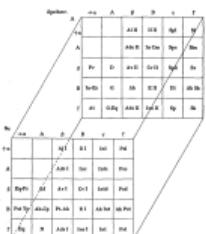
Le contenu linguistique est l'effet d'une structure virtuelle dérivée des pratiques linguistiques dans une communauté.



Axe 2: Interprétabilité théorique

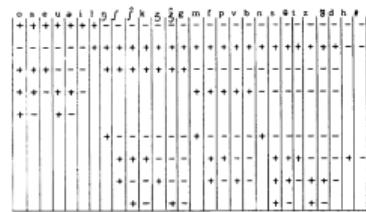
Hypothèse distributionnelle

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

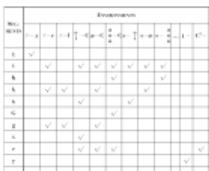


(Hjelmslev, 1935)

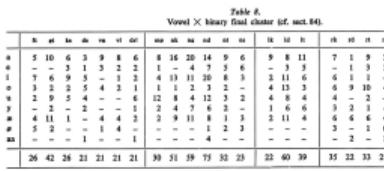
1. Vocalic/Non-vocalic
2. Consonantal/Non-consonantal
3. Compact/Diffuse
4. Grave/Acute
5. Flat/Plain
6. Nasal/Oral
7. Tense/Lax
8. Continuant/interrupted
9. Strident/Mellow



(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.



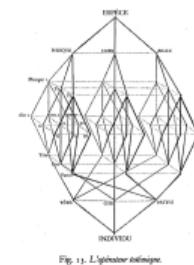
(Bourdieu, 1979)



(Foucault, 1966)

| Répondant | État + élargir | éliminer tout à la fois | Famille |
|------------------------------|----------------|-------------------------|---------|
| Prix | + | + | - |
| Photos intimes | + | + | - |
| Séances | + | + | + 10 |
| Choses de la nature | + | + | + |
| Signes de la famille | + | + | + |
| Signes de la maison | + | + | + |
| Signes de l'habitat | + | + | + |
| Chômage | + | + | + |
| Pratiques de l'habitat | + | + | + |
| Mobilier | + | + | + |
| Murs et cloisons intérieures | + | + | + |
| Meubles | + | + | + |
| Objets de l'habitat | + | + | + |
| Pratiques de l'habitat | + | + | + |
| Pratiques de l'habitat | + | - | - |

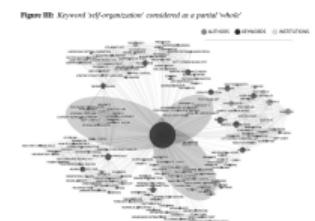
(Lévi-Strauss, 1949)



(Lévi-Strauss, 1962)



(Bourdieu, 1994)



(Latour et al., 2012)

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

Publications en cours

- ◊ “Content from expressions” (accepté à *Synthèse*)
- ◊ “LMs over Canonical BPEs” (soumis à ICML)
- ◊ “Distributional typing” (en cours)
- ◊ Chapitre sur LLMs (en cours, Oxford Handbook)

Collaborations et projets

- ◊ Cluster “Foundations of AI”
(CUNY, Simons Foundation)
- ◊ Projet “Human Forms”
(soumis à E. Schmidt Foundation)

Conférences à venir

- ◊ ICLR 2025 (Singapour, 24-28/04)
- ◊ Séminaire “IA et créativité” (Strasbourg, 13/05)
- ◊ Symposium “Chat Token Vector” (Venise, 11-13/06 - Keynote)
- ◊ Séminaire “Linguistics and Language Models” (Dagstuhl, 21-25/07)

Références I

- 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.
- 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>
- Deleuze, G. (1954). Jean Hyppolite, *Logique et existence*. In D. Lapoujade (Ed.), *L'île déserte et autres textes. textes et entretiens 1953-1974* (pp. 18–23). Éditions de Minuit.
- Foucault, M. (1954). Introduction. (J. Verdeaux, Trans.). In L. Binswanger (Ed.), *Le Rêve et l'Existence* (pp. 9–128). Desclée de Brouwer.
- Foucault, M. (1966). *Les mots et les choses : Une archéologie des sciences humaines*. Gallimard.
- Gastaldi, J. L. (2009). La politique avant l'être. deleuze, ontologie et politique. *Cités*, (40), 59–73.
<http://www.jstor.org/stable/40599521>
- Gastaldi, J. L. (2010). Qu'est-ce qu'une figure? Lyotard et le problème du fondement d'une théorie de l'expression. In C. Pagès (Ed.), *Lyotard à nanterre*. Klincksieck.
- Gastaldi, J. L. (2011). L'esthétique au sein des mots: Discours, figure, ou le renouvellement du projet critique. In P. Maniglier (Ed.), *Le moment philosophique des années 1960 en france* (pp. 537–556). Presses Universitaires de France.
- 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. (2015). Le sens d'une Logique du Sens: Deleuze, Frege et le rendez-vous manqué. In A. Jdey (Ed.), *Gilles deleuze: Politiques de la philosophie* (pp. 205–227). Métis Press.

Références II

- Gastaldi, J. L. (2016a). Frege's *Habilitationsschrift*: Magnitude, Number and the Problems of Computability. In F. Gadducci & M. Tavosanis (Eds.), *History and philosophy of computing* (pp. 168–185). Springer International Publishing.
- Gastaldi, J. L. (2016b). Par-delà métaphore et littéralité. Le statut des mathématiques dans l'œuvre de Deleuze. *Implications Philosophiques*.
<http://www.implications-philosophiques.org/actualite/une/par-dela-metaphore-et-litteralite/>
- Gastaldi, J. L. (2019). L'archéologie à l'épreuve des savoirs formels. Mathématiques et formalisation dans le projet d'une archéologie des savoirs. In J.-F. Braunstein, I. M. Diez, & M. Vagelli (Eds.), *L'épistémologie historique. histoire et méthodes*. Éditions de la Sorbonne.
- 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. (2022). Boole's Untruth Tables: The Formal Conditions of Meaning Before the Emergence of Propositional Logic. In J.-Y. Béziau, J.-P. Desclés, A. Moktefi, & A. C. Pascu (Eds.), *Logic in question: Talks from the annual sorbonne logic workshop (2011- 2019)* (pp. 119–149). Springer International Publishing.
https://doi.org/10.1007/978-3-030-94452-0_7
- Gastaldi, J. L. (2024a). Computing Cultures: Historical and Philosophical Perspectives. *Minds and Machines*, 34(1), 1–10. <https://doi.org/10.1007/s11023-023-09653-x>
- Gastaldi, J. L. (2024b). De Morgan's De Morgan's Laws Duality in the Emergence of Formal Logic. In R. Krömer & E. Haffner (Eds.), *Duality in 19th and 20th century mathematical thinking* (pp. 61–99). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-59797-8_3

Références III

- Gastaldi, J. L. (2024c). How to Do Maths with Words: Neural Machine Learning Applications to Mathematics and Their Philosophical Significance. In B. Sriraman (Ed.), *Handbook of the history and philosophy of mathematical practice* (pp. 3191–3226). Springer International Publishing.
https://doi.org/10.1007/978-3-031-40846-5_142
- Gastaldi, J. L. (2024d). Introduction to the Semiology of Mathematical Practice. In B. Sriraman (Ed.), *Handbook of the history and philosophy of mathematical practice* (pp. 2955–2970). Springer International Publishing.
https://doi.org/10.1007/978-3-031-40846-5_130
- Gastaldi, J. L. (Forthcoming 2024c). Content from Expressions. The Place of Textuality in Deep Learning Approaches to Mathematics. *Synthese (under review)*.
- Gastaldi, J. L., Moot, R., & Rétoré, C. (2024). Le contexte en traitement automatique des langues. In G. Hassler (Ed.), *Le contexte en question*. Iste.
- 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>
- Gastaldi (ed.), J. L. (Ed.). (2024a). Computing Cultures: Historical and Philosophical Perspectives [Special Issue]. *Minds and Machines*, 34(1). Special Issue. <https://link.springer.com/collections/dgcffbghba>
- Gastaldi (ed.), J. L. (2024b). Section: Semiology of Mathematical Practice. In B. Sriraman (Ed.), *Handbook of the History and Philosophy of Mathematical Practice* (pp. 2955–3226). Springer International Publishing.
<https://doi.org/10.1007/978-3-031-40846-5>
- Girard, J.-Y. (2006). *Le point aveugle: Cours de logique. vers la perfection*. Editions Hermann.

Références IV

- 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. (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.
- 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/3f5ee243547dee91fb0d053c1c4a845aa-Paper.pdf

Références V

- Vieira, T., LeBrun, B., Giulianelli, 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
DR Section 53 - Concours n° 53/01

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

Juan Luis Gastaldi

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



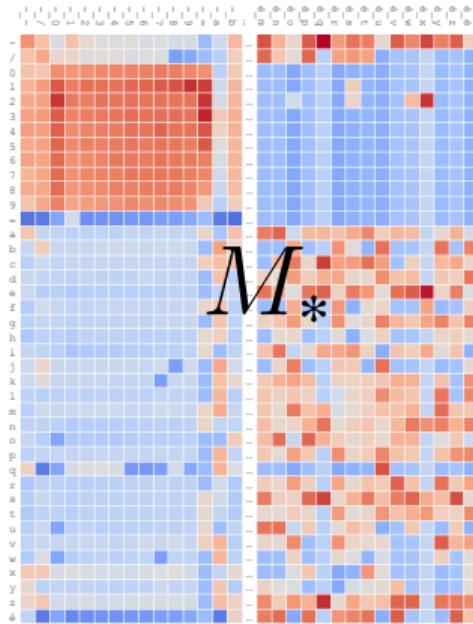
Axe 1: Objectifs

- ◊ 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.

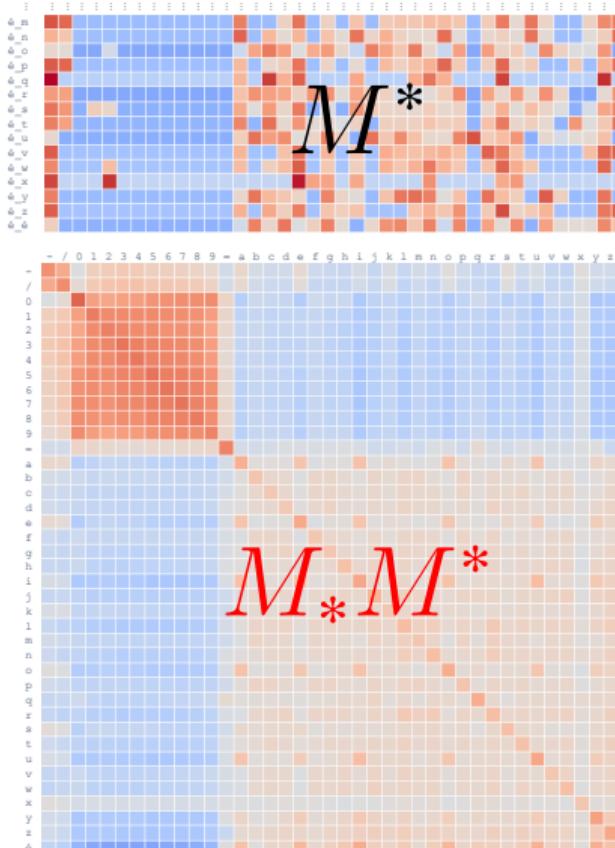
Axe 2: Objectifs

- ◊ 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



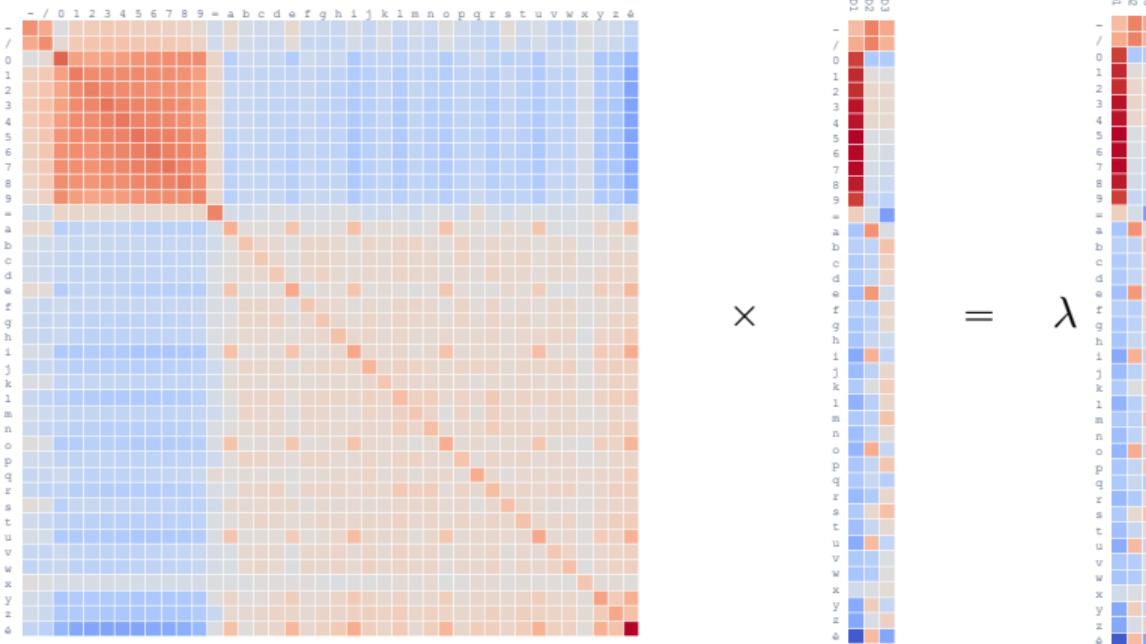
M_*



$M_* M^*$

Vecteurs propres comme points fixes

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

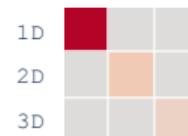


Traits structuraux

Eigenvectors of $M_* M^*$:



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



Eigenvectors of $M^* M_*$:



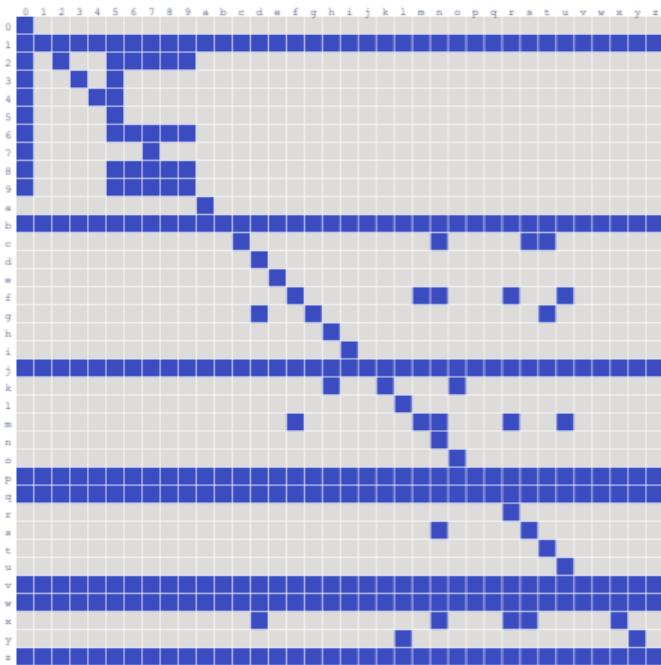
Mots

| | -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

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



★



? \equiv



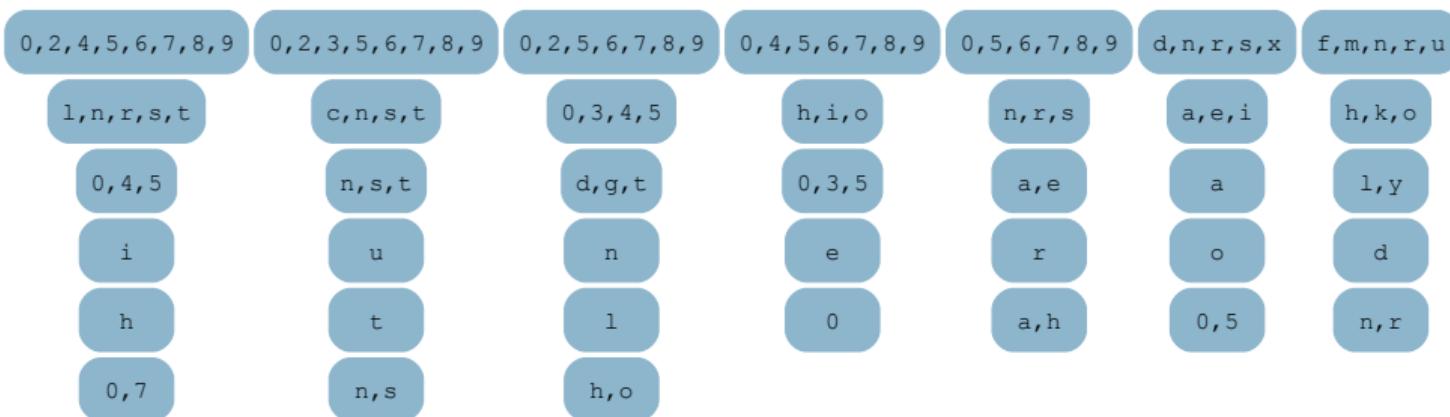
“Eigensets”

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

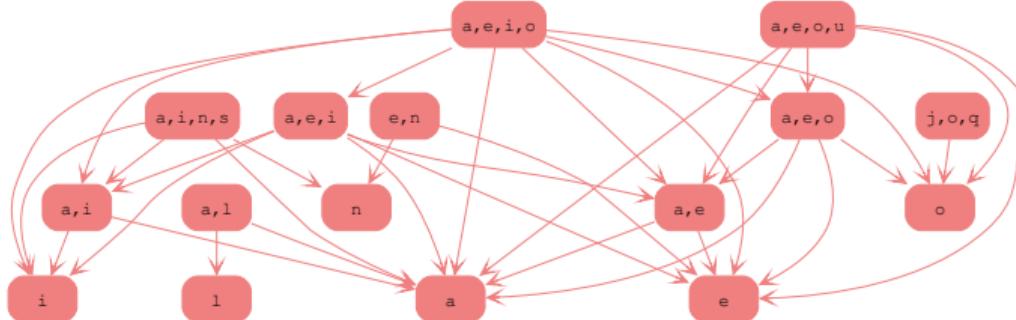
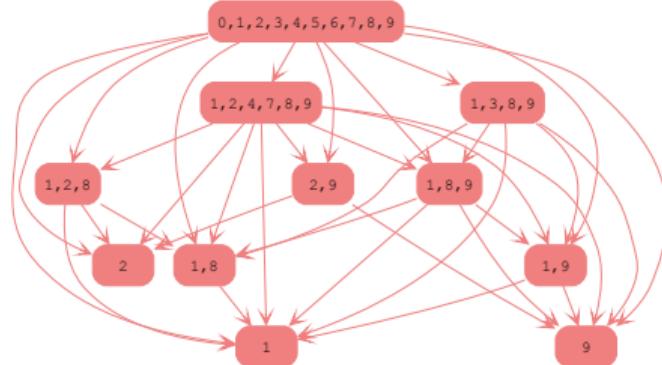
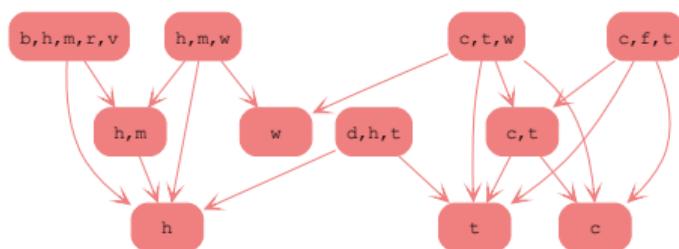
| | | | | | | |
|-----------------------|---------------|-------------|-----------|-----------|-----------|-----------|
| $0,1,2,3,4,5,6,7,8,9$ | $1,2,4,7,8,9$ | b,h,m,r,v | a,e,i,o | a,e,o,u | a,i,n,s | $1,3,8,9$ |
| $1,2,8$ | h,m,w | $1,8,9$ | d,h,t | j,o,q | c,f,t | c,t,w |
| a,e,o | a,e,i | h,m |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

“Eigensets”

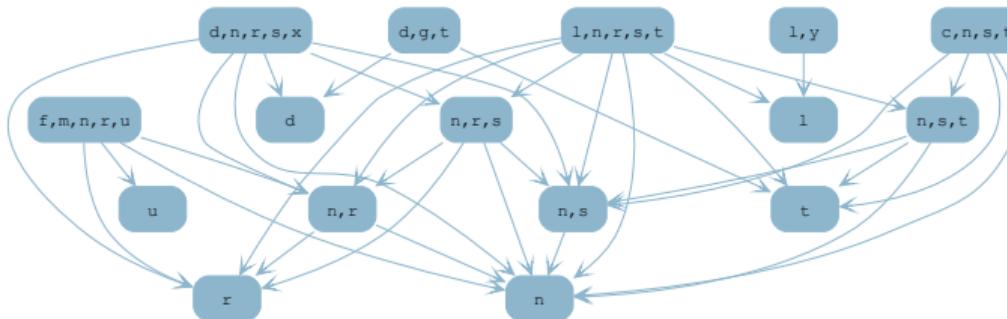
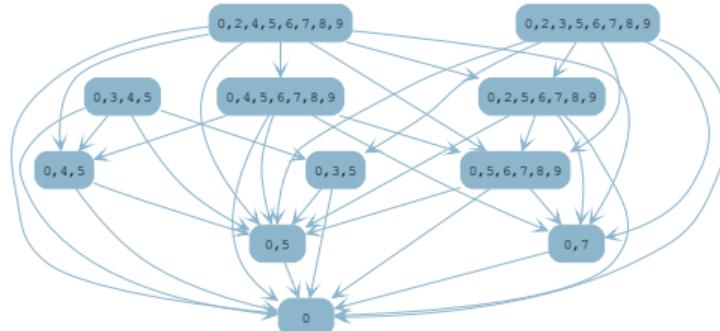
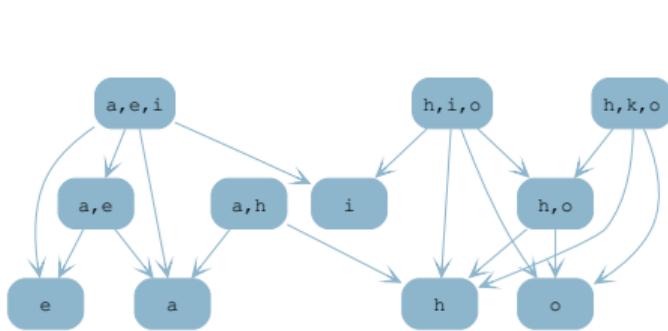
$$M_i^* M_*^i \mathbf{d} = \mathbf{d}$$



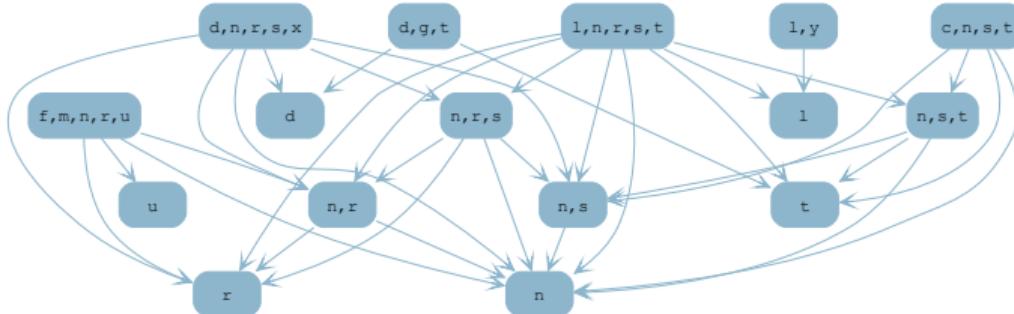
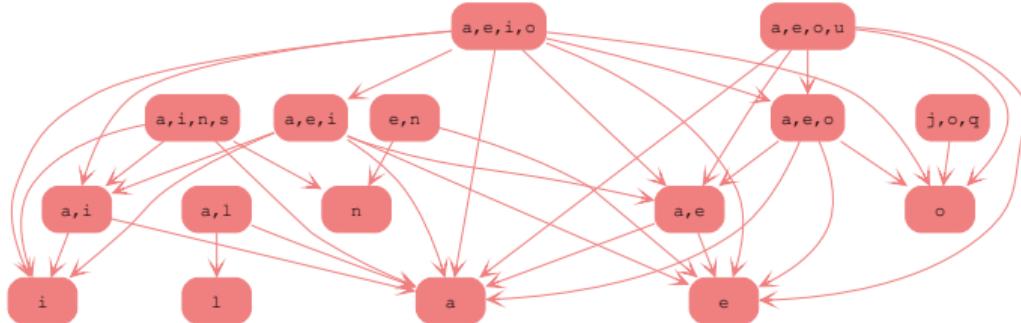
Quelle Structure?



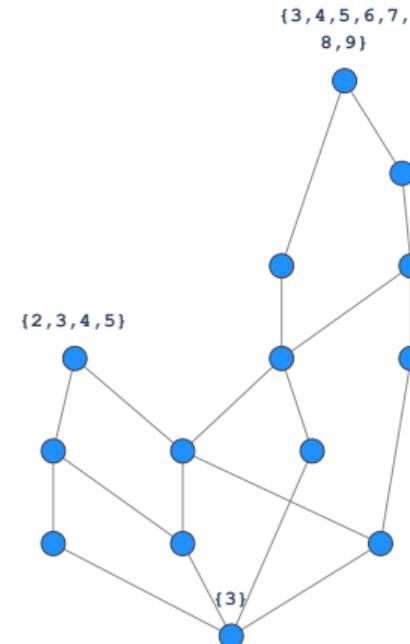
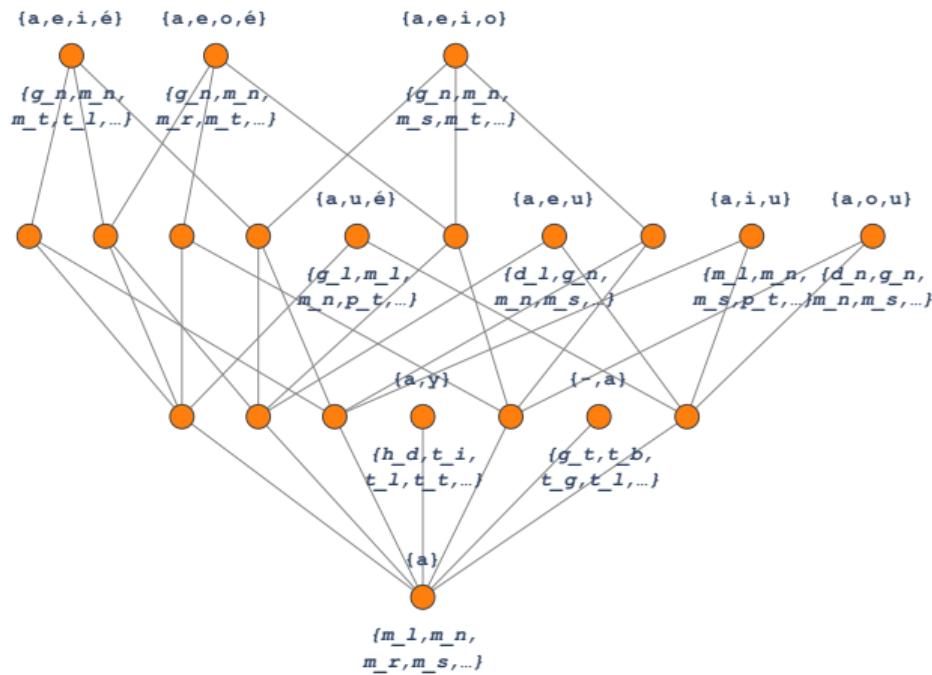
Quelle Structure?



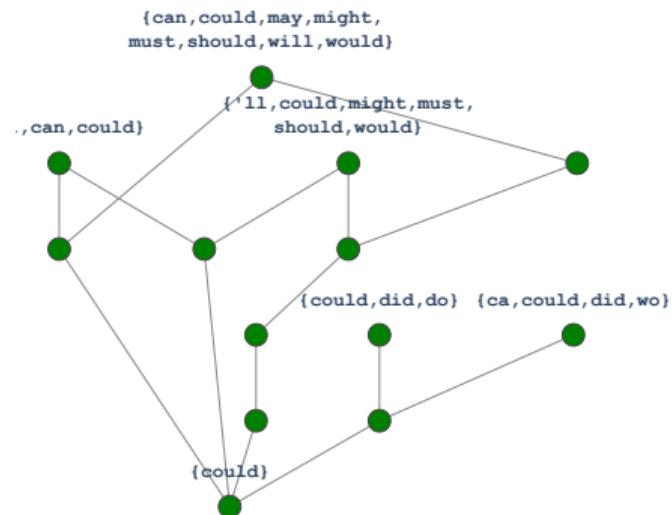
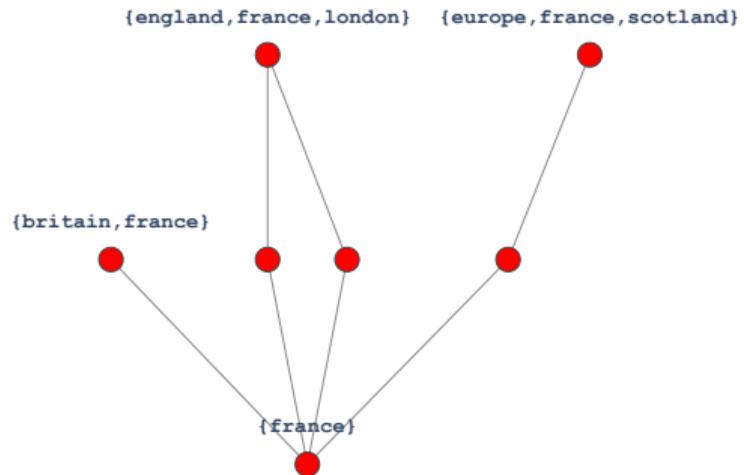
Quelle Structure?



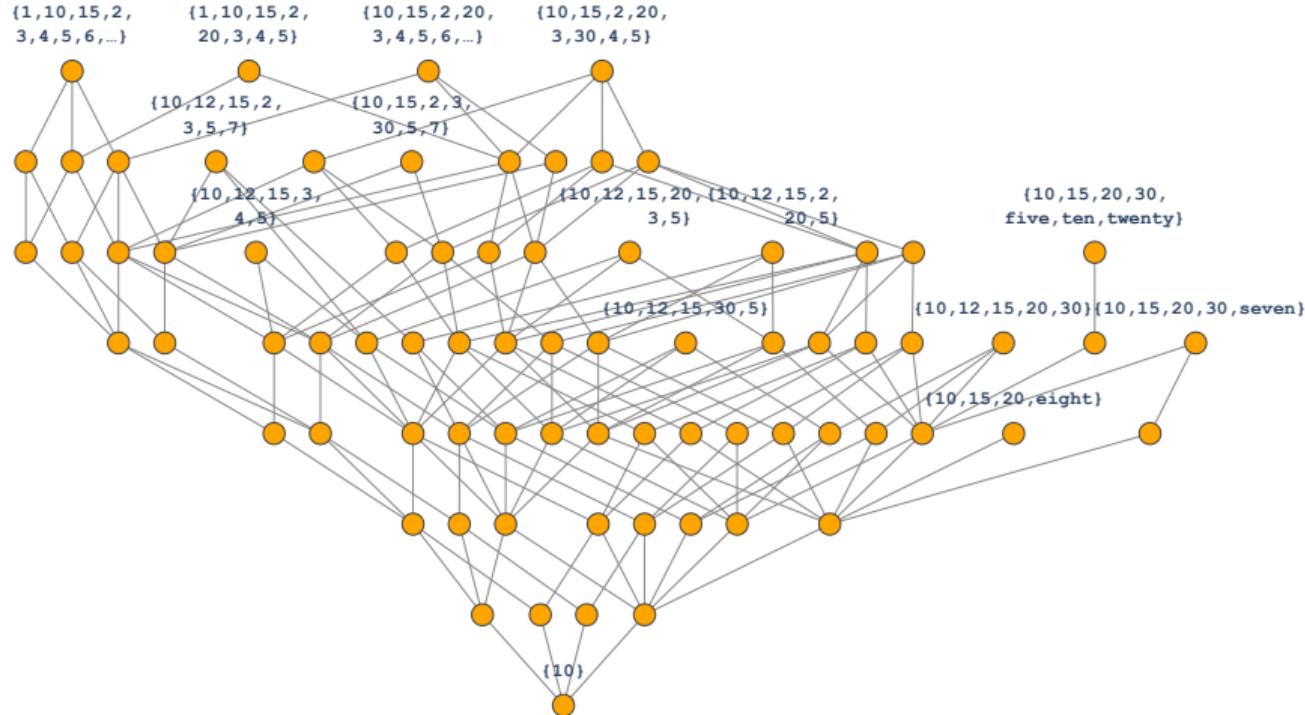
Concepts formels



Concepts formels (mots)

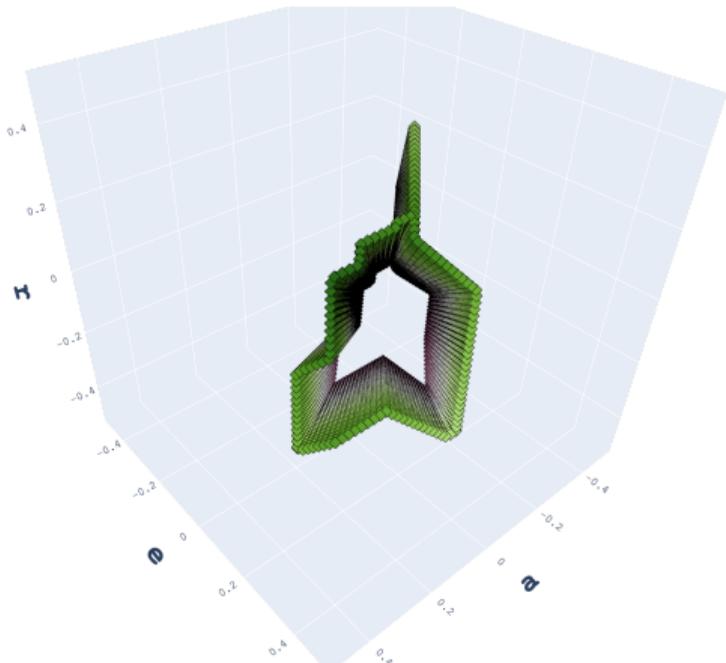
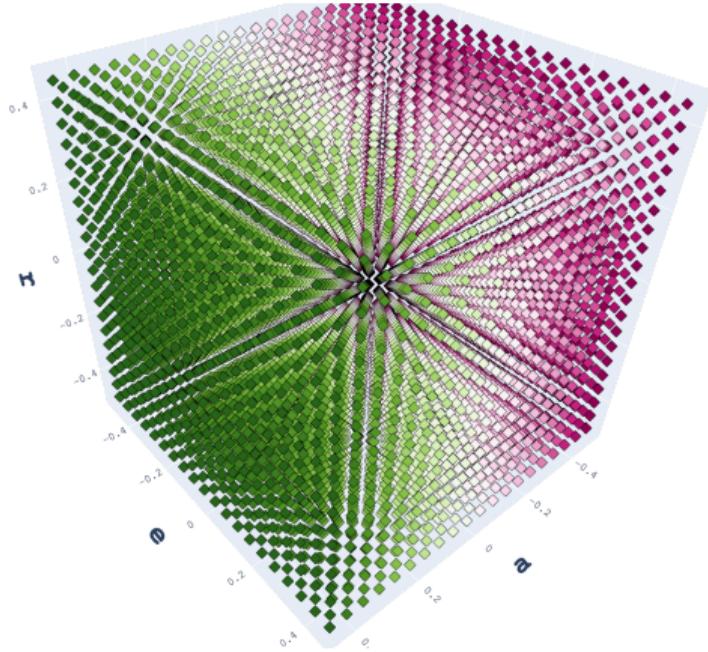


Concepts formels (mots)

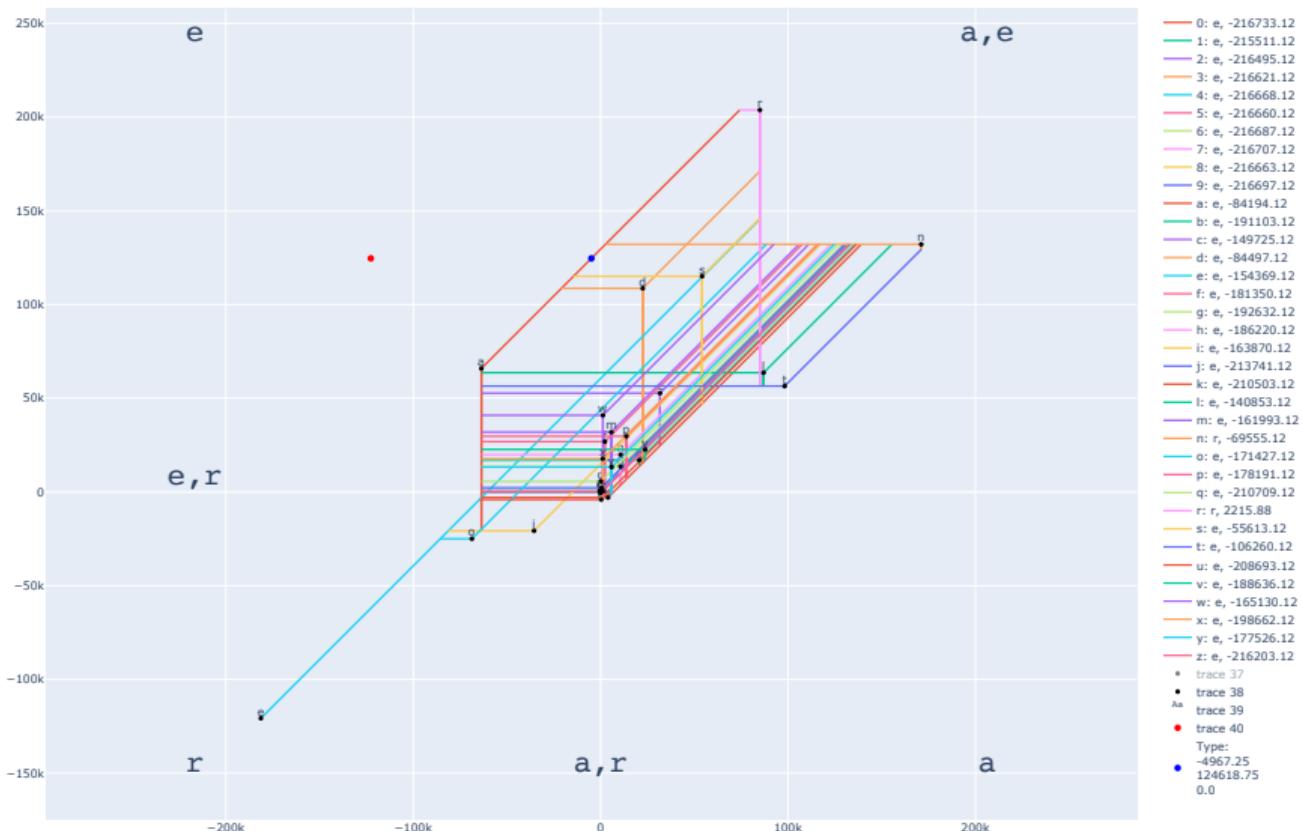


Noyau (nucleus)

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



Structure interne du noyau



Théorie des types computationnels

Definition (Polaire/Orthogonal - Girard, 2006)

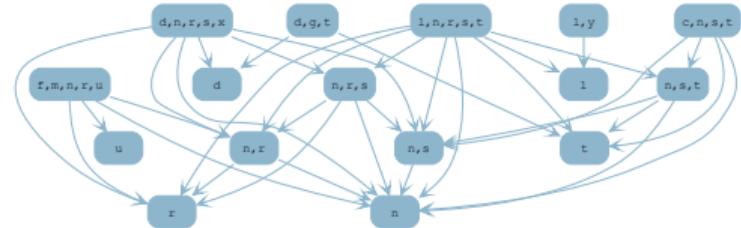
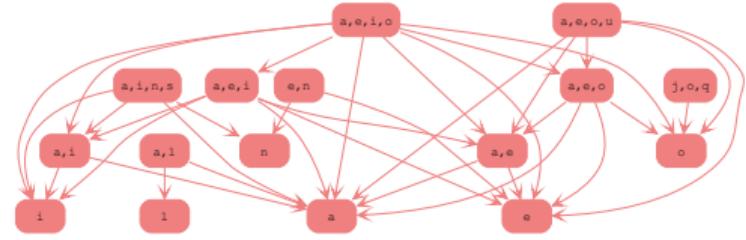
[É]tant donnée une fonction binaire

$a, b \rightsquigarrow \langle a|b \rangle : A \times B \rightarrow C$ 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, $X^{\perp\perp\perp} = X^\perp$.



Théorie des types computationnels

Definition (Polaire/Orthogonal - Girard, 2006)

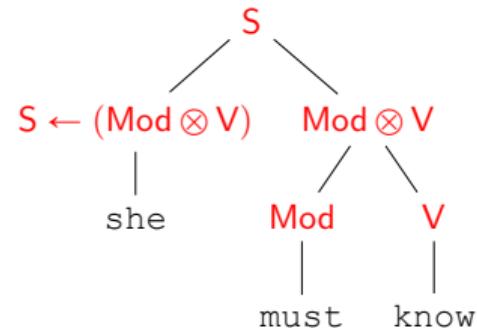
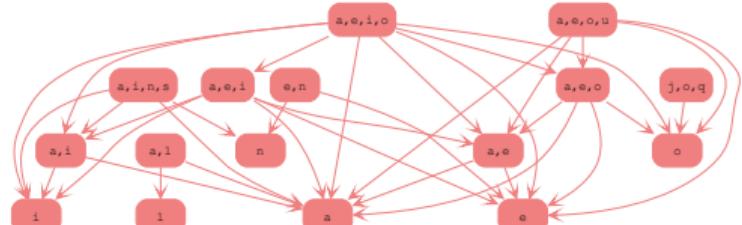
[É]tant donnée une fonction binaire

$a, b \rightsquigarrow \langle a|b \rangle : A \times B \rightarrow C$ 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, $X^{\perp\perp\perp} = X^\perp$.



(Gastaldi and Pellissier, 2021)

Matrice et analogie

a = your
c = my

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

your : house
my : apartment

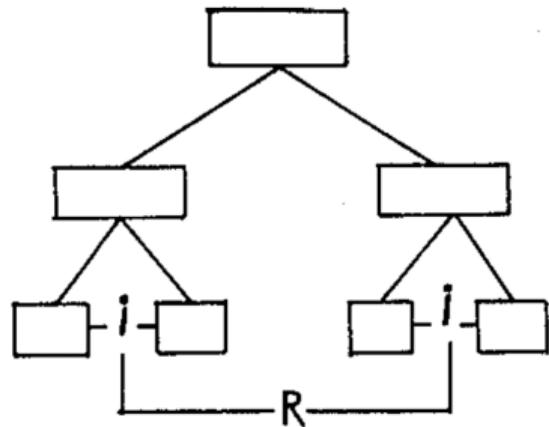
| | ... | w | x | y | z | ... |
|-----|-----|---|---|---|---|-----|
| ... | ... | 0 | 0 | 0 | 0 | ... |
| a | ... | 0 | 1 | 1 | 0 | ... |
| b | ... | 0 | 0 | 1 | 1 | ... |
| c | ... | 1 | 0 | 0 | 1 | ... |
| ... | ... | 0 | 0 | 0 | 0 | ... |

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.

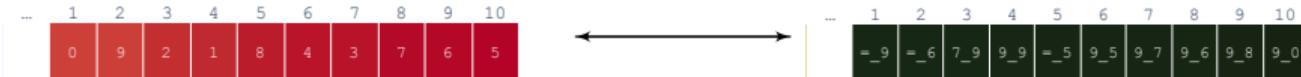
(Hjelmslev, 1975, Df. 24)

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.

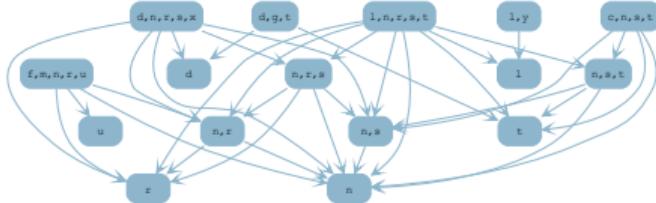
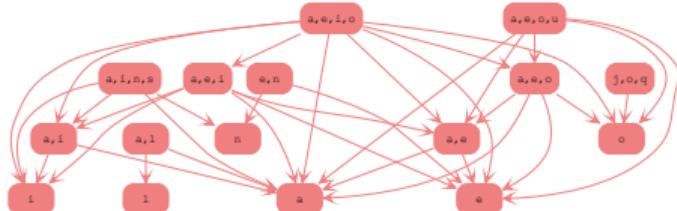
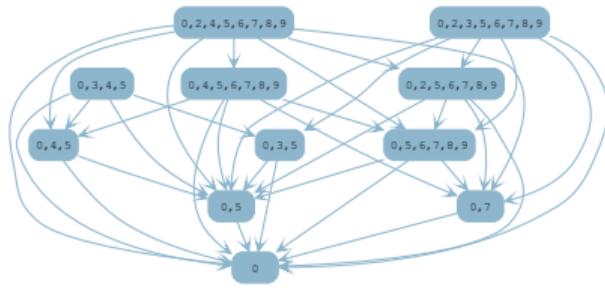
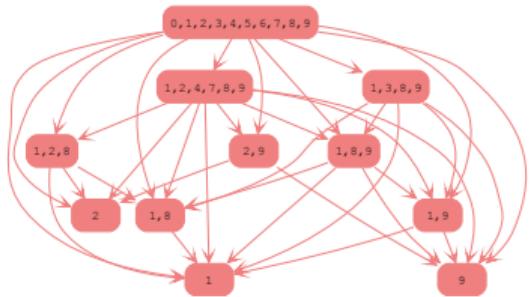
(Hjelmslev, 1975, Df. 23)



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)

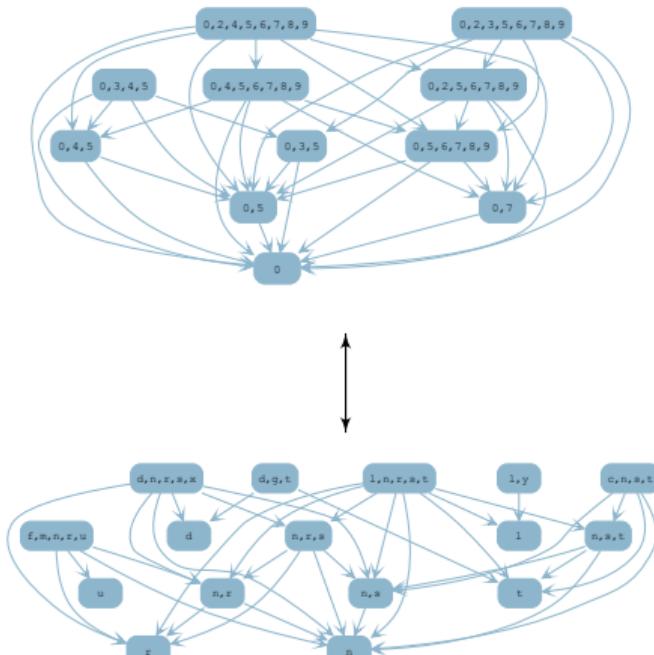
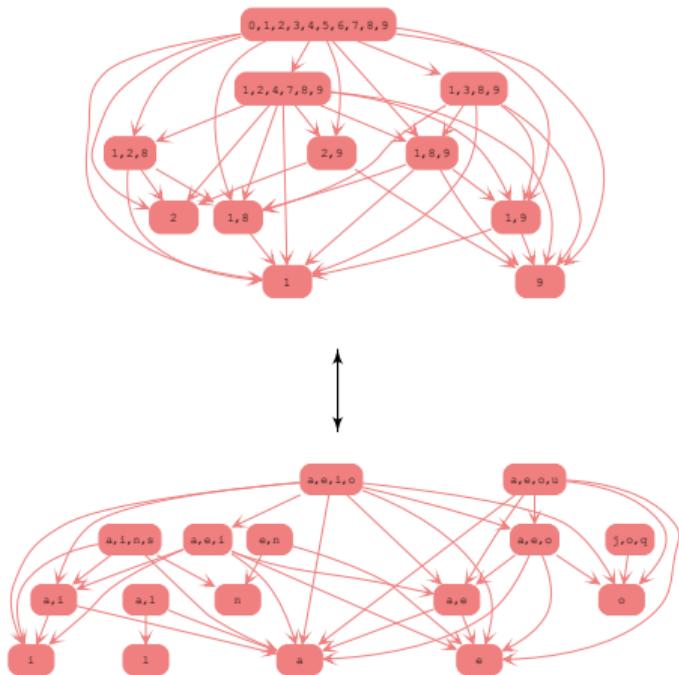


Illustration du contenu formel

(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