

Séminaire SPHePS
Paris, France

Structuralisme et Intelligence Artificielle

Juan Luis Gastaldi

ETH zürich

24 Mai, 2023



This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730

Structuralisme et IA

DNNs et NLP

Word Embeddings

Exemple

La Structure Derrière Les Embeddings

Conclusions et Défis

Structuralisme et IA

DNNs et NLP

Word Embeddings

Exemple

La Structure Derrière Les Embeddings

Conclusions et Défis

L'Hypothèse Distributionnelle

- ◇ “You shall know a word by the **company** it keeps!” (Firth, 1957)
- ◇ “Words which are similar in meaning occur in similar **contexts**” (Rubenstein & Goodenough 1965)
- ◇ “Words with similar meanings will occur with similar **neighbors** if enough text material is available” (Schütze & Pedersen 1995)
- ◇ “A representation that captures much of how words are used in natural **context** will capture much of what we mean by meaning” (Landauer & Dumais 1997)
- ◇ “Words that occur in the same **contexts** tend to have similar meanings” (Pantel 2005)
- ◇ “The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic **contexts** in which A and B can appear” (Lenci, 2010)

Structuralisme et IA



Structuralisme et IA



Structuralisme et IA

1. Dans quelle mesure les modèles actuels de réseaux neuronaux profonds (DNN), en particulier pour le traitement du langage naturel (NLP), peuvent-ils être **considérés comme structuralistes** ?

Structuralisme et IA

1. Dans quelle mesure les modèles actuels de réseaux neuronaux profonds (DNN), en particulier pour le traitement du langage naturel (NLP), peuvent-ils être **considérés comme structuralistes** ?
2. Quelles peuvent être les **conséquences** de la promotion d'une compréhension structuraliste de ces modèles ?

Structuralisme et IA

1. Dans quelle mesure les modèles actuels de réseaux neuronaux profonds (DNN), en particulier pour le traitement du langage naturel (NLP), peuvent-ils être **considérés comme structuralistes** ?
2. Quelles peuvent être les **conséquences** de la promotion d'une compréhension structuraliste de ces modèles ?
3. Quelles sont les **conditions** pour que ces conséquences se produisent **effectivement** ?

L'Hypothèse Structurale

“Il semble légitime en tous cas de poser a priori l'hypothèse qu'**à tout processus répond un système** qui permette de l'analyser et de le décrire au moyen d'**un nombre restreint de prémisses**. Il doit être possible de considérer tout processus comme composé d'**un nombre limité d'éléments qui réapparaissent constamment dans de nouvelles combinaisons**. On devrait pouvoir, en se fondant sur l'analyse du processus, regrouper ces éléments en classes, chaque classe étant définie par l'homogénéité de ses possibilités combinatoires, et pouvoir, à partir de ce classement préalable, établir **un calcul général exhaustif des combinaisons possibles**.”

(Hjelmslev, *Prolégomènes à une théorie du langage*, p. 16)

Structuralisme et IA

DNNs et NLP

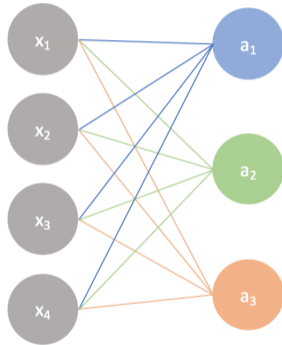
Word Embeddings

Exemple

La Structure Derrière Les Embeddings

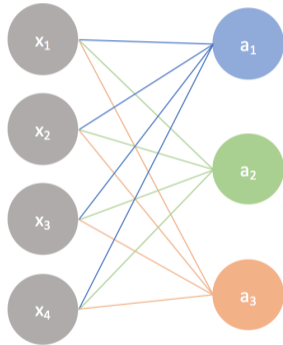
Conclusions et Défis

Réseaux de Neurones



Credit: Jeremy Jordan
<https://www.jeremyjordan.me/intro-to-neural-networks/>

Réseaux de Neurones



$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow{\text{activation}} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Credit: Jeremy Jordan
<https://www.jeremyjordan.me/intro-to-neural-networks/>

Produit de Matrices

The diagram illustrates the calculation of an element in the product of two matrices. It shows three matrices: L_i , C_j , and a third matrix representing the result of the dot product.

- L_i is a matrix with rows $a_{i1}, \dots, a_{ik}, \dots, a_{in}$. The row $a_{i1}, \dots, a_{ik}, \dots, a_{in}$ is highlighted in pink.
- C_j is a matrix with columns $b_{11}, \dots, b_{1j}, \dots, b_{1n}$, $b_{k1}, \dots, b_{kj}, \dots, b_{kn}$, $b_{n1}, \dots, b_{nj}, \dots, b_{nn}$. The column $b_{1j}, \dots, b_{kj}, \dots, b_{nj}$ is highlighted in pink.
- The resulting matrix has a dot product $\dots \cdot C_{ij}$ in its i -th row and j -th column.

A curved arrow points from the pink row of L_i and the pink column of C_j to the dot product $\dots \cdot C_{ij}$ in the resulting matrix.

$$L_i \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1j} & \dots & b_{1n} \\ \vdots & & \vdots & & \vdots \\ \vdots & & b_{kj} & & \vdots \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \dots & b_{nj} & \dots & b_{nn} \end{pmatrix} \begin{pmatrix} \vdots \\ \vdots \\ \dots \cdot C_{ij} \\ \vdots \end{pmatrix}$$

Produit de Matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

Produit de Matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 19 \end{pmatrix}$$

Produit de Matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 19 \end{pmatrix}$$

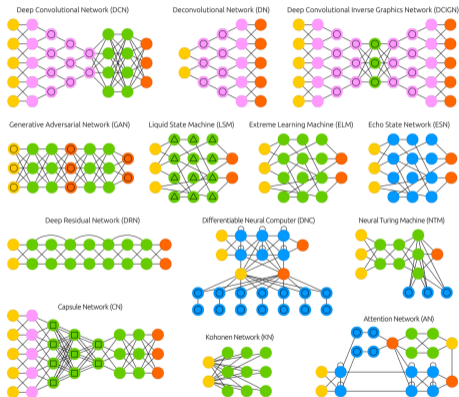
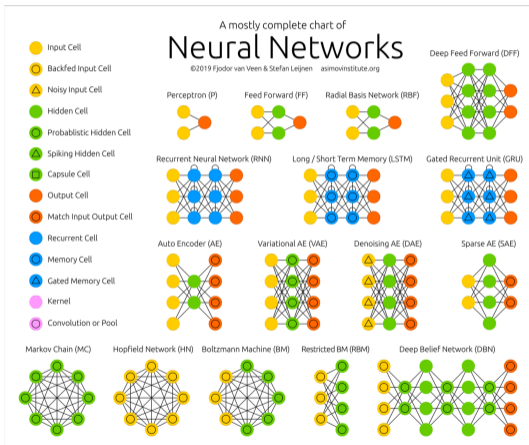
Produit de Matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 19 & & \\ & 40 & \end{pmatrix}$$

Produit de Matrices

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \\ 29 & 40 & 51 \end{pmatrix}$$

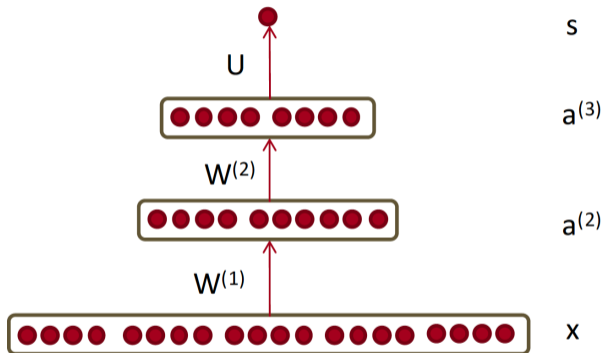
Réseaux de Neurones Profonds (DNNs)



Source: <https://www.asimovinstitute.org/neural-network-zoo/>

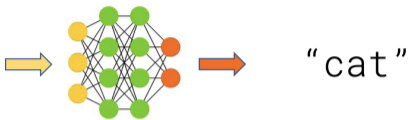
Traitement Automatique des Langues (NLP)

$$\begin{aligned}x &= z^{(1)} = a^{(1)} \\z^{(2)} &= W^{(1)}x + b^{(1)} \\a^{(2)} &= f\left(z^{(2)}\right) \\z^{(3)} &= W^{(2)}a^{(2)} + b^{(2)} \\a^{(3)} &= f\left(z^{(3)}\right) \\s &= U^T a^{(3)}\end{aligned}$$



Credit: Manning & Socher, Stanford CS224n course, 2017

Traitement Automatique des Langues (NLP)



"Ceci est une phrase"



"What's a sentence?"



Trois Composantes Principales des Modèles de NLP

Subword Tokenization
(Sennrich et al., 2016)

Word Embeddings
(Mikolov, Sutskever, Chen, Corrado, and Dean, 2013)

Self-Attention
(Vaswani et al., 2017)

Trois Composantes Principales des Modèles de NLP

Subword Tokenization
(Sennrich et al., 2016)

Word Embeddings
(Mikolov, Sutskever, Chen, Corrado, and Dean, 2013)

Self-Attention
(Vaswani et al., 2017)

Structuralisme et IA

DNNs et NLP

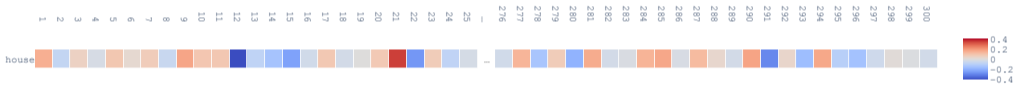
Word Embeddings

Exemple

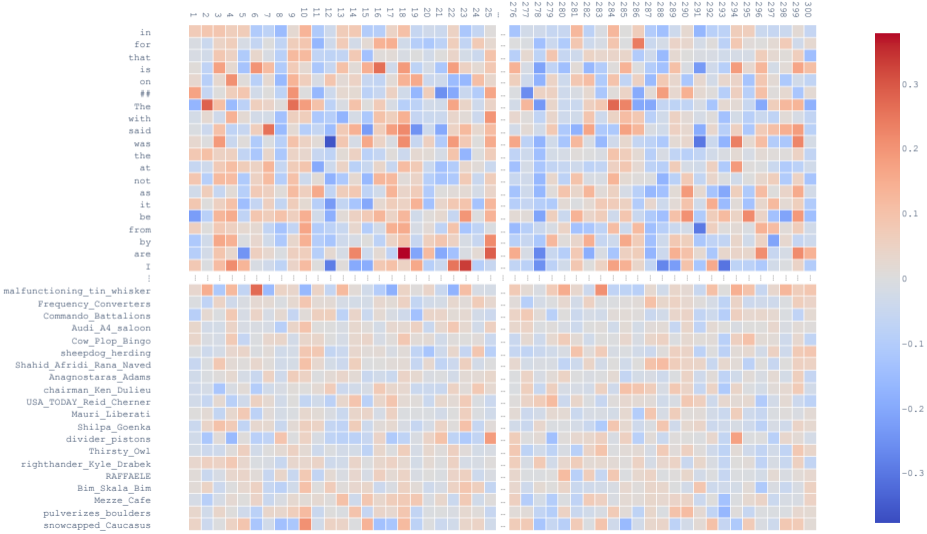
La Structure Derrière Les Embeddings

Conclusions et Défis

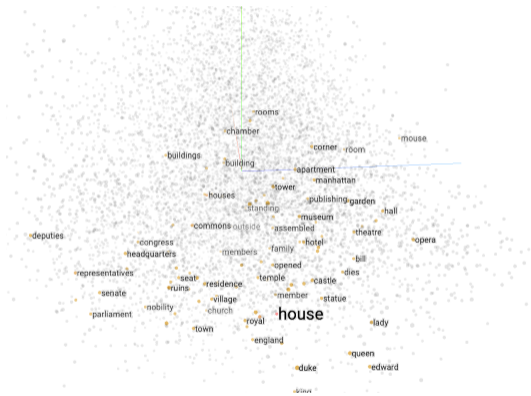
Word Embeddings: Vector



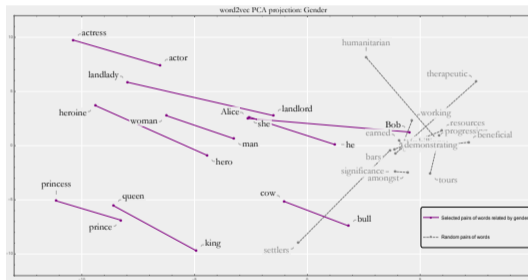
Word Embeddings: Matrix



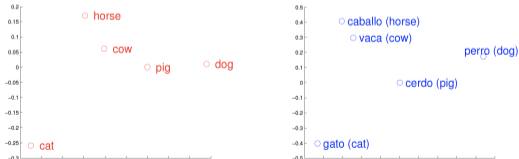
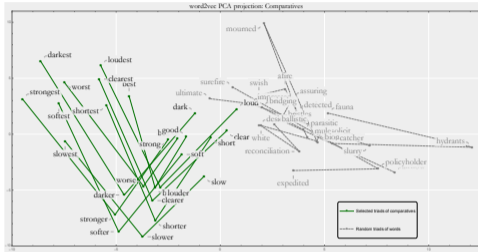
Embedding Space: Similarity and Analogy



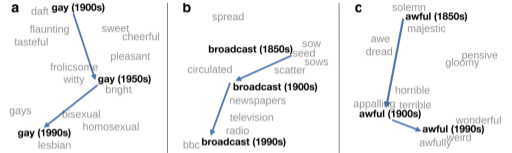
(<https://projector.tensorflow.org>)



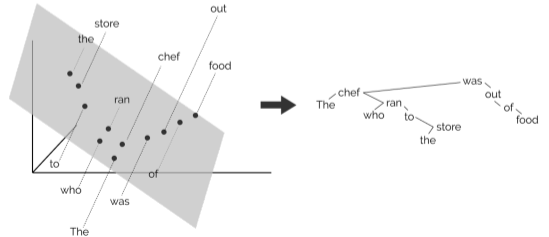
Embedding Space: Other Applications



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

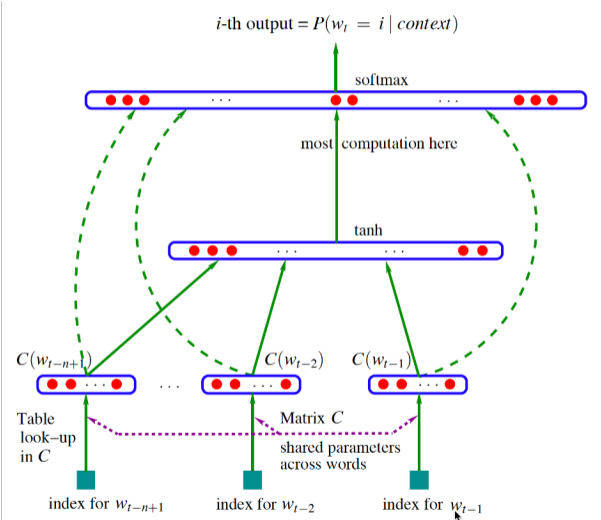


(Hamilton et al., 2016)



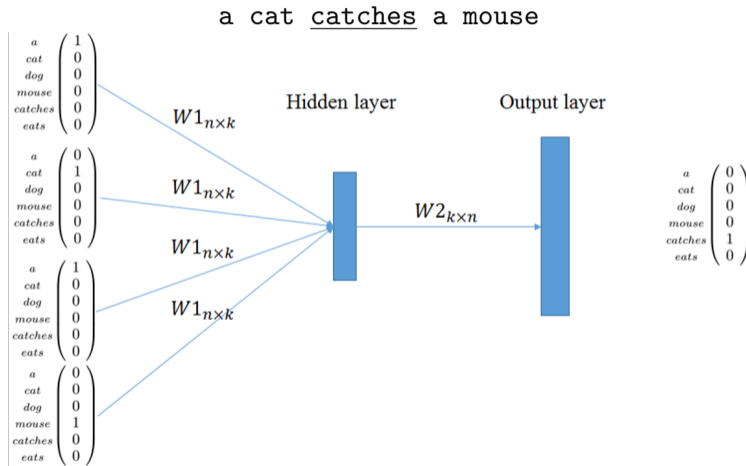
(<https://nlp.stanford.edu/~johnhew/structural-probe.html>)

Premiers Modèles NLP Neuronaux



(Bengio et al., 2003)

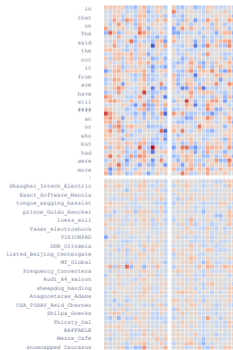
word2vec



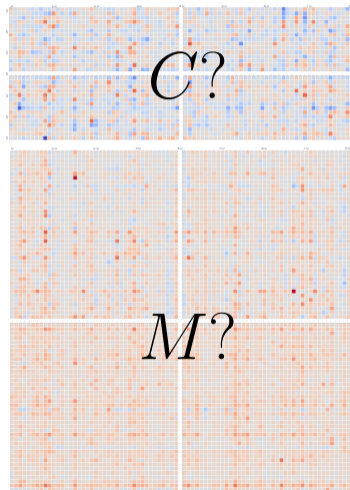
Credit: Ferrone et al., 2017

word2vec Comme Factorisation Implicite de Matrice

(Levy and Goldberg, 2014)



word2vec Comme Factorisation Implicite de Matrice (Levy and Goldberg, 2014)



$$W \times C \approx M$$

word2vec Expliqué

(Levy and Goldberg, 2014)

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

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

word2vec Expliqué (Levy and Goldberg, 2014)

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

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

Trois résultats:

- ◇ $M = PMI(w, c) - \log k$ (Pointwise Mutual Information)

word2vec Expliqué (Levy and Goldberg, 2014)

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

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

Trois résultats:

- ◇ $M = PMI(w, c) - \log k$ (Pointwise Mutual Information)
- ◇ W est de **base dimensionnalité**

word2vec Expliqué

(Levy and Goldberg, 2014)

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

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

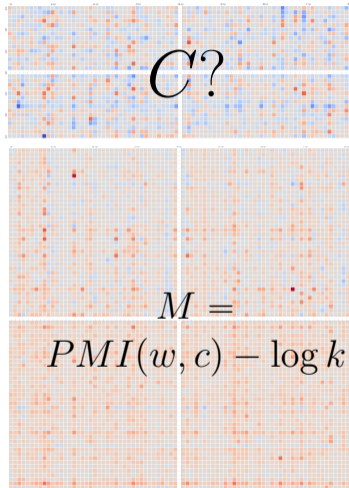
Trois résultats:

- ◇ $M = PMI(w, c) - \log k$ (Pointwise Mutual Information)
- ◇ W est de **base dimensionnalité**
- ◇ La **Décomposition en Valeurs Singulières (SVD)** offre une **solution exacte** pour trouver W

Information Mutuelle Ponctuelle (PMI)



W



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

Décomposition en Valeurs Singulières (SVD)

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

Où:

M = matrice $m \times n$ (réelle or complexe)

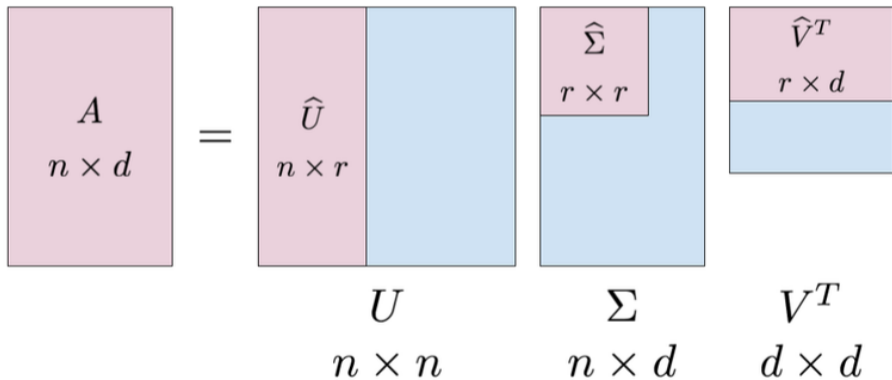
U = matrice unitaire $m \times m$

Σ = matrice $m \times n$ positive réelle rectangulaire diagonale

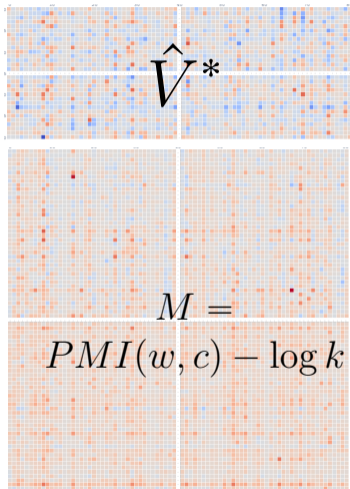
V^* = matrice adjointe de V , matrice unitaire $n \times n$

SVD Tronquée

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



Embeddings en tant que SVD tronquée



$$M \approx \hat{U} \times \hat{\Sigma} \times \hat{V}^*$$

Structuralisme et IA

DNNs et NLP

Word Embeddings

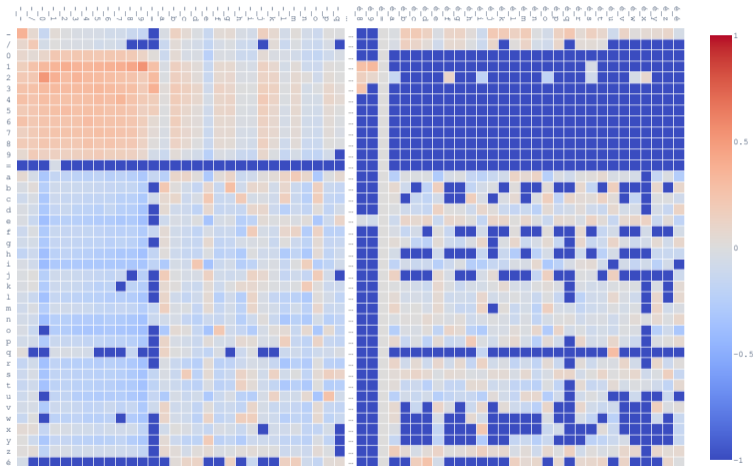
Exemple

La Structure Derrière Les Embeddings

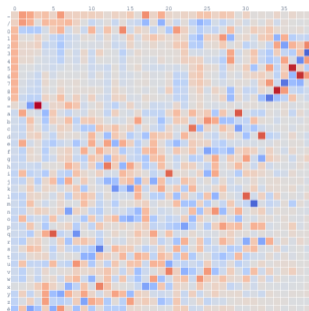
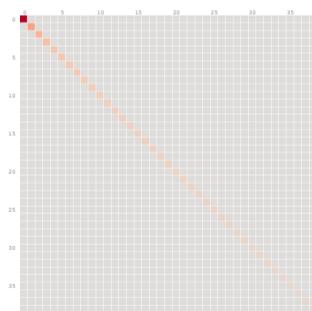
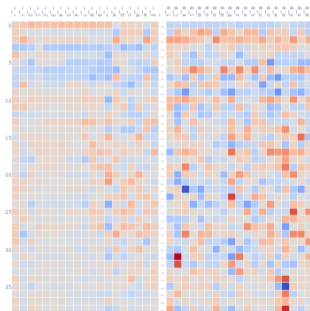
Conclusions et Défis

Exemple: Caractères dans Wikipedia

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

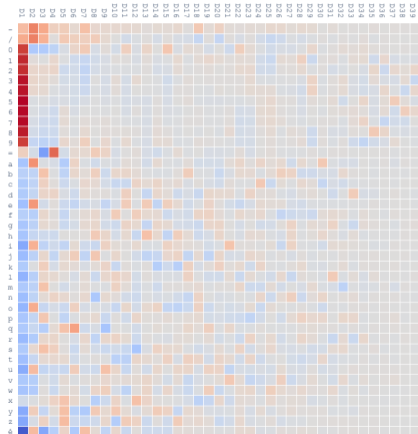


SVD d'Une Matrice PMI des Caractères dans Wikipedia

 U  Σ  V^* 

Tronquer et Plonger

$$U \times \Sigma$$



Tronquer et Plonger

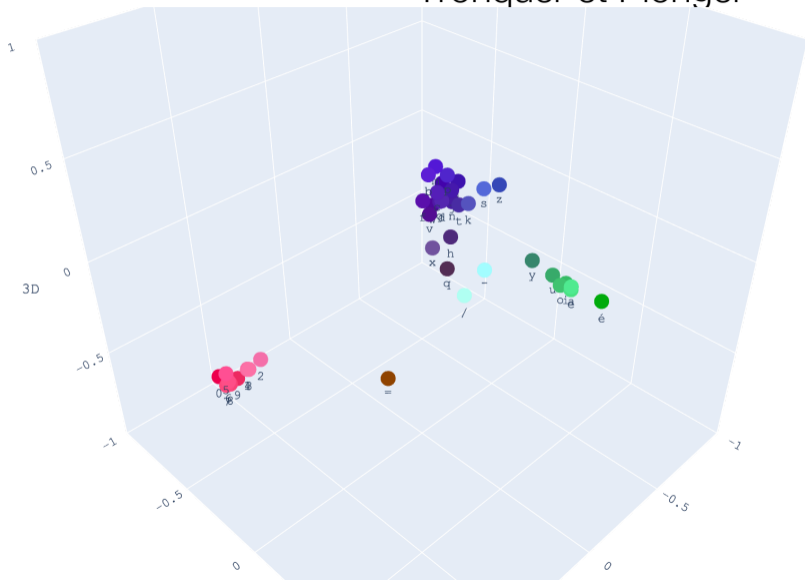
$$\hat{U} \times \hat{\Sigma}$$



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



Tronquer et Plonger



Structuralisme et IA

DNNs et NLP

Word Embeddings

Exemple

La Structure Derrière Les Embeddings

Conclusions et Défis

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)

Décomposition en Valeurs Singulières (SVD)

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

Où:

M = matrice $m \times n$ (réelle or complexe)

U = matrice unitaire $m \times m$

Σ = matrice $m \times n$ positive réelle rectangulaire diagonale

V^* = matrice adjointe de V , matrice unitaire $n \times n$

Décomposition en Valeurs Singulières (SVD)

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

Où:

M = matrice $m \times n$ (réelle or complexe)

U = matrice unitaire $m \times m$

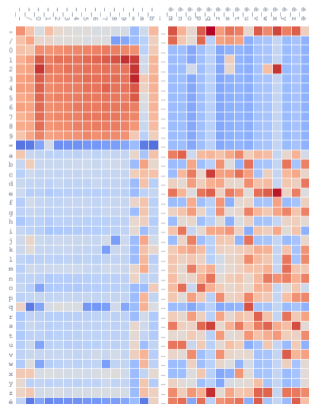
Σ = matrice $m \times n$ positive réelle rectangulaire diagonale

V^* = matrice adjointe de V , matrice unitaire $n \times n$

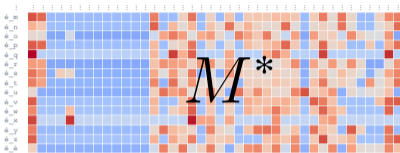
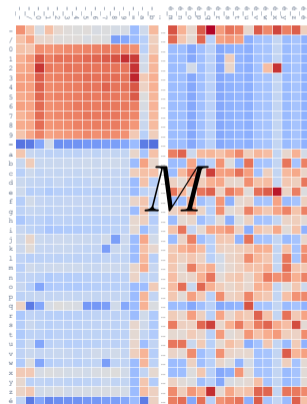
En particulier:

- ◊ Les colonnes de U (vecteurs sing. gauche) sont les **vecteurs propres de MM^***
- ◊ Les lignes de V^* (vecteurs sing. droits) sont les **vecteurs propres de M^*M**
- ◊ Les éléments non nuls de Σ (valeurs singulières non nulles) sont les racines carrés des **valeurs singulières** non nulles **de MM^* ou M^*M**

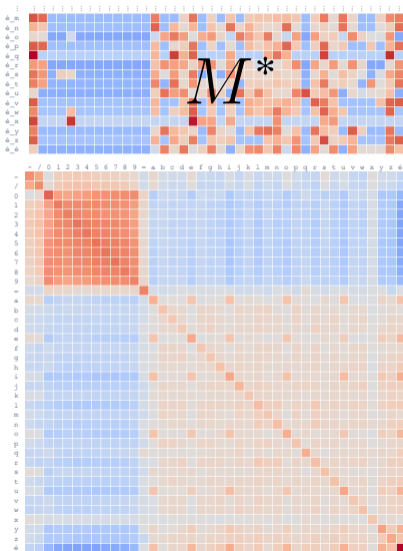
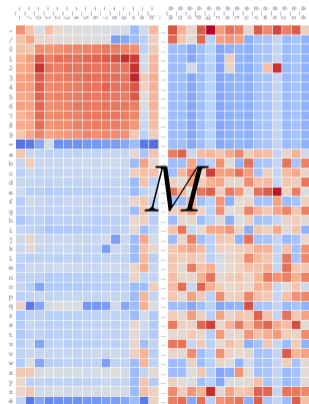
$M \times M^*$ Comme Matrice de Covariance



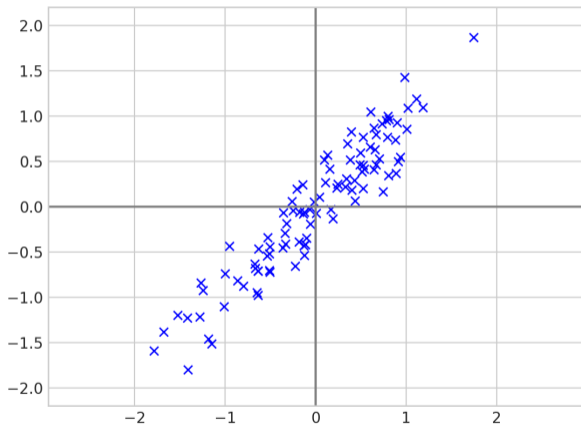
$M \times M^*$ Comme Matrice de Covariance



$M \times M^*$ Comme Matrice de Covariance

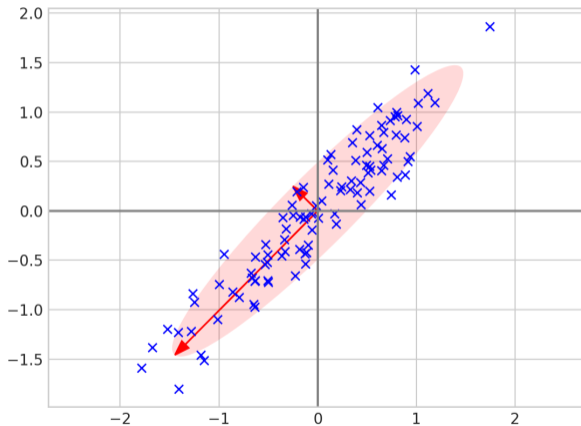


Vecteurs Propres et Valeurs Propres



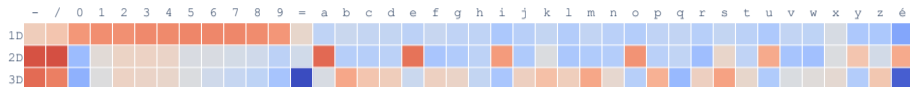
Credit: Joel Laity

Vecteurs Propres et Valeurs Propres



“Structure Propre”

Vecteurs propres de $M \times M^*$:



Commutation



“Structure Propre”

Vecteurs propres de $M \times M^*$:

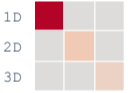


“Structure Propre”

Vecteurs propres de $M \times M^*$:



Valeurs propres de $M \times M^*$:

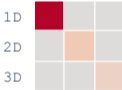


“Structure Propre”

Vecteurs propres de $M \times M^*$:



Valeurs propres de $M \times M^*$:



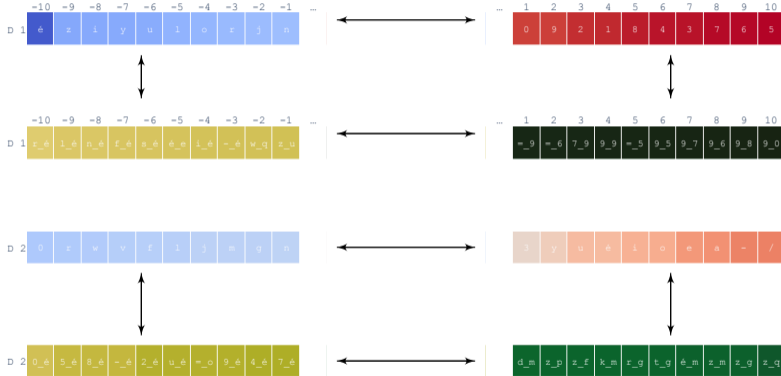
Vecteurs propres de $M^* \times M$:



Commutation



Commutation

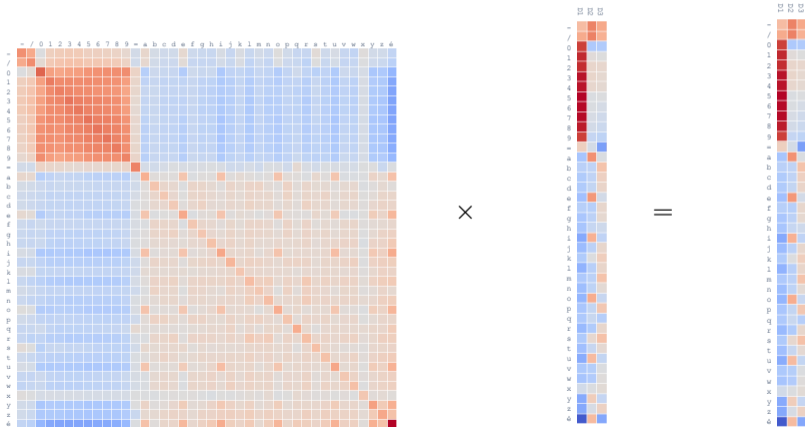


Vecteurs Propres Comme Points Fixes

$$(M \times M^*)v = \lambda v$$

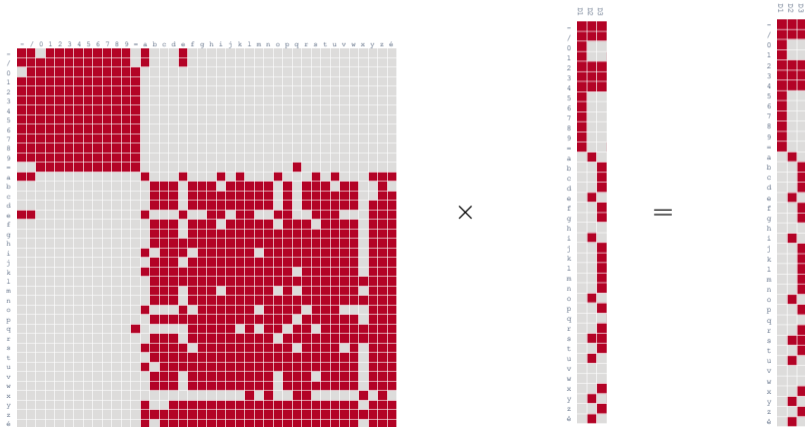
Vecteurs Propres Comme Points Fixes

$$(M \times M^*)v = \lambda v$$

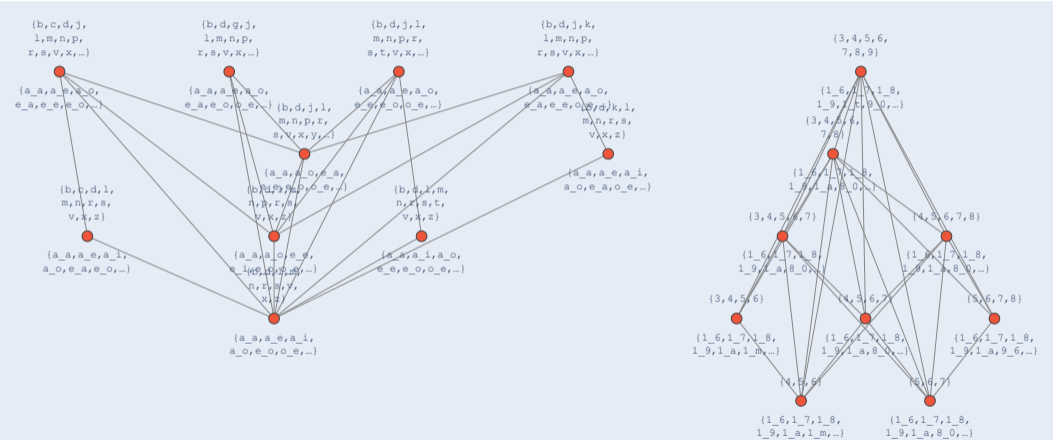


Binarisation: Concepts Formels

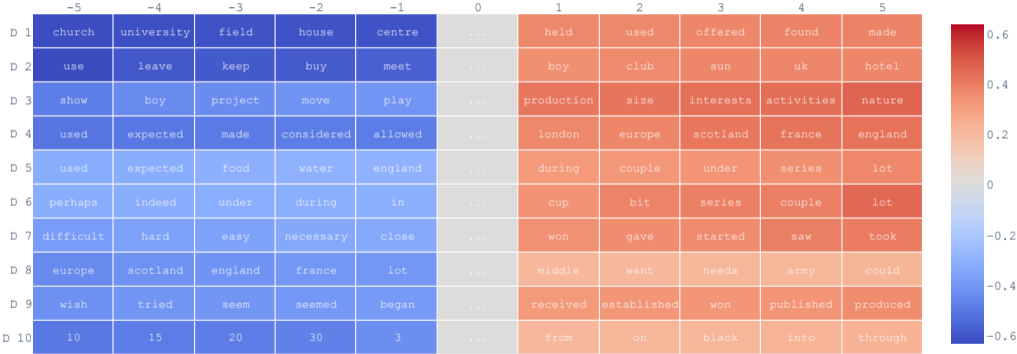
$$(M \times M^*)v = \lambda v$$



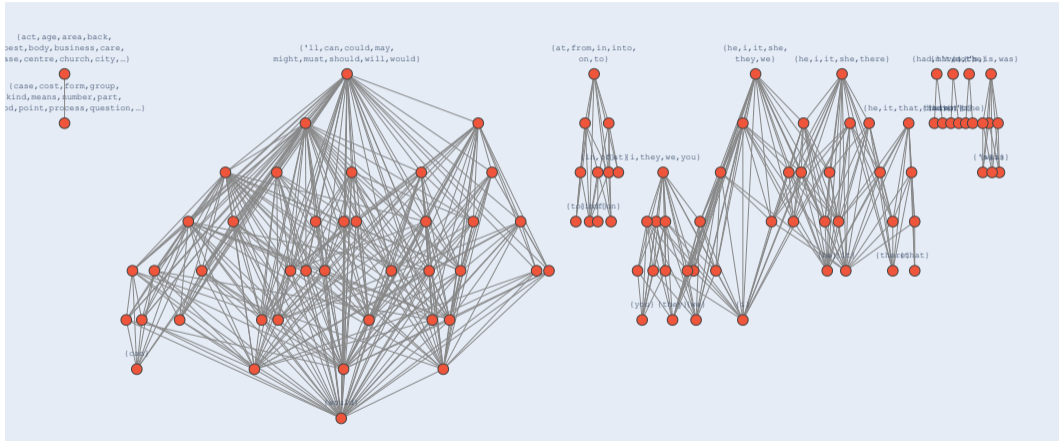
Concepts Formels



Mots



Concepts Formels Lexicaux



Structuralisme et IA

DNNs et NLP

Word Embeddings

Exemple

La Structure Derrière Les Embeddings

Conclusions et Défis

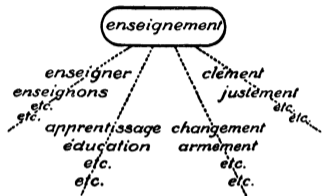
Conclusions

- ◇ Les modèles de langage neuronaux (MLN) sont la mise en œuvre implicite d'une théorie du langage

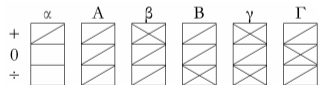
Conclusions

- ◇ Les modèles de langage neuronaux (MLN) sont la mise en œuvre implicite d'une théorie du langage
- ◇ A partir d'un traitement opérationnel des relations syntagmatiques, les MLN déduisent les relations paradigmatisques et explorent les propriétés de commutation pour identifier des unités paradigmatisques et syntagmatiques à différents niveaux, et des dépendances entre elles.

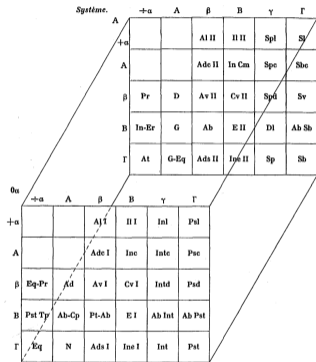
Outils Structuralistes



(Saussure, 1980)



(Hjelmslev, 1975)



(Hjelmslev, 1935)

SEG- MENTS	ENVIRONNEMENTS											
	#-r	#-r	#-l	c i	-Cæ-C	a o u	-C s	-c i	s-æ	s- o u	... l	C ³ -
l	✓											
t		✓			✓	✓	✓	✓	✓	✓		
K						✓				✓		
k		✓	✓		✓				✓			
K					✓		✓					
G						✓						
g		✓	✓		✓							
G					✓							
r					✓	✓	✓					✓
r												✓

(Harris, 1960)

Outils Structuralistes

	a	b	d	e	f	g	h	i
a	aa	ab	ad		af	ag	ah	
b	ba							bi
d	da			de				di
e		eb	ed			eg		
f				fe				
g								gi
h	ha							hi
i			id				ih	i

Diagram 1.

	b	d	f	g	h	a	e	i
f						fa	fe	
h						ha		hi
g						ga	ge	gi
b						ba	be	bi
d						da	de	di
a	ab	ad	af	ag	ah	aa		
e	eb	ed	ef	eg				
i	ib	id		ig	ih			i

Diagram 3.

	I	II	III	IV			
	p r s t	i o u	y	&			
I	p r s t	-	+	+	+	-	
II	i o u	+	+	+			-
III	y	-	+	+			
IV	&						+

Diagram 2.

(SpangHanssen1959)

Outils Structuralistes

Table 8.
Vowel × binary final cluster (cf. sect. 84).

	ft	gt	ks	ds	vn	vl	drl	mp	nk	ng	nd	nt	ns	lk	ld	lt	rk	rd	rt	rn	S	T	iC	
a	5	10	6	3	9	8	6	8	16	20	14	9	6	9	8	11	7	1	9	3	168	281	3	a
e	-	-	3	1	3	2	2	1	-	4	7	5	6	-	3	5	-	1	3	3	49	95	33	e
i	7	6	9	5	-	1	2	4	13	11	20	8	3	2	11	6	6	1	1	-	116	171	-	i
o	3	2	2	5	4	2	1	1	1	2	3	2	-	4	13	3	6	9	10	4	77	120	-	o
u	2	9	5	4	-	-	6	12	8	4	12	3	2	4	8	4	4	-	2	-	89	143	-	u
y	-	2	-	2	-	-	1	2	4	7	6	2	-	1	6	6	3	2	1	-	45	56	-	y
æ	4	11	1	-	4	4	2	2	9	11	8	1	3	2	11	4	6	6	6	4	99	145	-	æ
ø	5	2	-	-	1	4	-	-	-	-	1	2	3	-	-	-	3	-	1	6	28	47	10	ø
aa	-	-	-	1	-	-	1	-	-	-	4	-	-	-	-	-	-	2	-	1	9	11	-	aa
	26	42	26	21	21	21	21	30	51	59	75	32	23	22	60	39	35	22	33	21	680	1069	46	

(SpangHanssen1959)

Conclusions

- ◊ Les modèles de langage neuronaux (MLN) sont la mise en œuvre implicite d'une théorie du langage
- ◊ A partir d'un traitement opérationnel des relations syntagmatiques, les MLN déduisent les relations paradigmatiques et explorent les propriétés de commutation pour identifier des unités paradigmatiques et syntagmatiques à différents niveaux, et des dépendances entre elles.

Conclusions

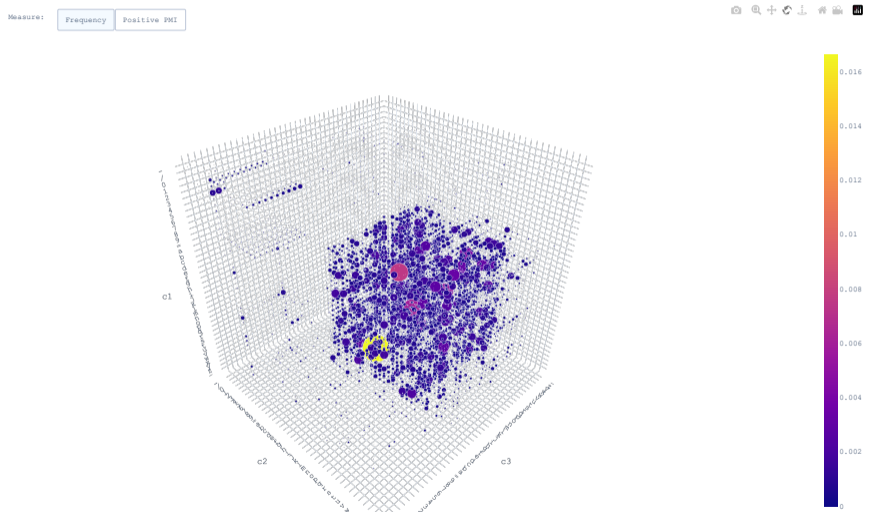
- ◇ Les modèles de langage neuronaux (MLN) sont la mise en œuvre implicite d'une théorie du langage
- ◇ A partir d'un traitement opérationnel des relations syntagmatiques, les MLN déduisent les relations paradigmatiques et explorent les propriétés de commutation pour identifier des unités paradigmatiques et syntagmatiques à différents niveaux, et des dépendances entre elles.
- ◇ Si les MLN peuvent rendre compte du sens, ce n'est pas parce qu'ils sont intelligents, mais parce que le sens est l'effet de la structure (linguistique) et que les MLN effectuent une analyse structurale du langage.

Conclusions

- ◇ Les modèles de langage neuronaux (MLN) sont la mise en œuvre implicite d'une théorie du langage
- ◇ A partir d'un traitement opérationnel des relations syntagmatiques, les MLN déduisent les relations paradigmatiques et explorent les propriétés de commutation pour identifier des unités paradigmatiques et syntagmatiques à différents niveaux, et des dépendances entre elles.
- ◇ Si les MLN peuvent rendre compte du sens, ce n'est pas parce qu'ils sont intelligents, mais parce que le sens est l'effet de la structure (linguistique) et que les MLN effectuent une analyse structurale du langage.
- ◇ Restituer la base structuraliste implicite peut fournir de l'interprétabilité et une réorientation du champ de recherche.

- ◇ Politique du corpus
- ◇ Théorie et philosophie du langage non cognitive
- ◇ Traitement conjoint et intégré de la tokenization, plongement et attention
- ◇ Connexion entre des propriétés distributionnelles et structurales
- ◇ Traitement de dépendance de long terme
- ◇ Calculabilité, tractabilité
- ◇ Généralisation vers des corpus non linguistiques (sémiologie)

Généralisation



Articles de Référence

- ◇ J. L. Gastaldi. **Why Can Computers Understand Natural Language?**
In: *Philosophy & Technology* 34.1 (2021), pp. 149–214.
- ◇ J. L. Gastaldi and L. Pellissier. **The calculus of language: explicit representation of emergent linguistic structure through type-theoretical paradigms**
In: *Interdisciplinary Science Reviews* 46.4 (2021), pp. 569–590.
- ◇ T.-D. Bradley, J. L. Gastaldi, J. Terilla, **The Structure of Meaning in Language: Moving from Linear Algebra to Category Theory**
Under review.

Références I

- Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *J. Mach. Learn. Res.*, 3, 1137–1155.
- Goldberg, Y., & Levy, O. (2014). Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *CoRR*, [abs/1402.3722](https://arxiv.org/abs/1402.3722).
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. *CoRR*, [abs/1605.09096](https://arxiv.org/abs/1605.09096).
- Harris, Z. (1960). *Structural linguistics*. University of Chicago Press.
- Hjelmslev, L. (1935). *La catégorie des cas*. Wilhelm Fink Verlag.
- Hjelmslev, L. (1971). *Prolégomènes à une théorie du langage*. Éditions de Minuit.
- Hjelmslev, L. (1975). *Résumé of a Theory of Language*. Nordisk Sprog-og Kulturforlag.
- 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*.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *CoRR*, [abs/1310.4546](https://arxiv.org/abs/1310.4546).
- Saussure. (1980). *Cours de linguistique générale*. Payot.
- 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.

Références II

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.

Séminaire SPHePS
Paris, France

Structuralisme et Intelligence Artificielle

Juan Luis Gastaldi

ETH zürich

24 Mai, 2023



This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730