

Looking for **linguistic** markers using **deep learning** methods  
**Example with political discourse learning**

Laurent Vanni

UMR 7320: Bases, Corpus, Langage - CNRS Engineer - Computer Science



**Linguistics & Computer science**  
**Real interdisciplinary cooperation ?**

Some kind of **black box**



Are **Digital Humanities** a new way to **understand** neuronal networks ?  
Are **neuronal networks** a new way to **understand** linguistics ?

## Linguistic discussions

Corneille versus Molière

- Who wrote "tartuffe" ?



Who is E. Macron ?

- Left-right political spectrum



Are **neural networks** a new way to **understand** linguistics ?

# About the presentation



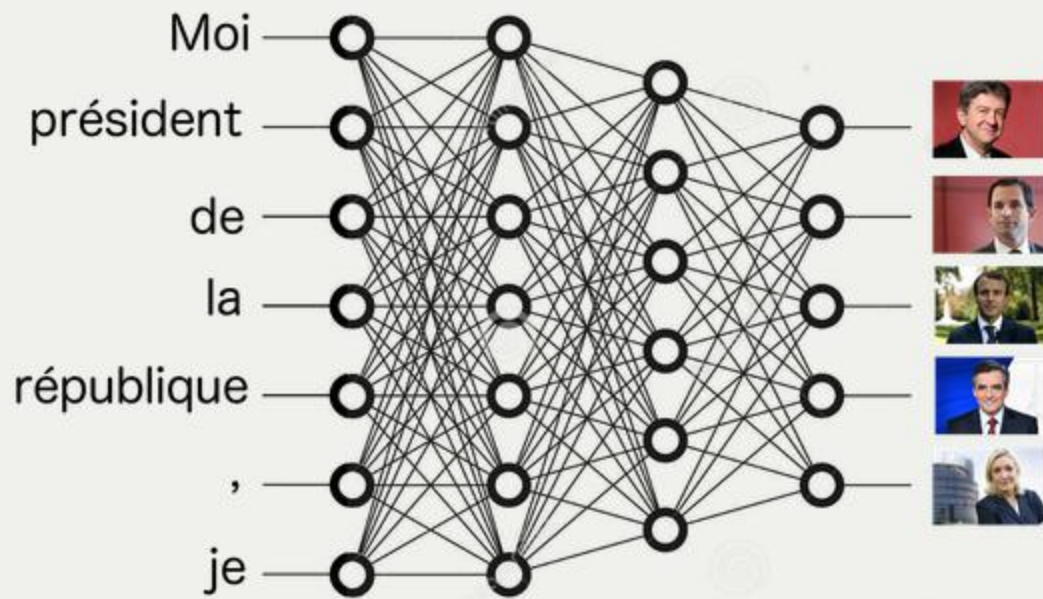
**First step**

**Political discourse learning**

**IDEX JEDI project - Académie 5**

# Objectives

## Authorship attribution



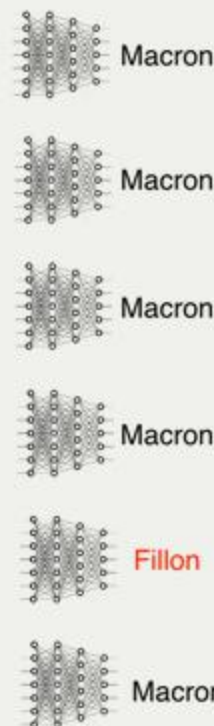
What can Artificial Intelligence learn about political discourse ?



# Attribution level

## New metrics

Aloes que l'année s'achève, je suis heureux de vous retrouver pour vous présenter pour la première fois mes vœux pour l'année 2018. Je vous espère en famille, au milieu de vos proches, de celles et ceux qui vous aiment. Je sais aussi que certains d'entre vous sont aujourd'hui au travail parce qu'ils font partie des forces armées ou des forces de l'ordre, parce qu'ils sont médecins ou personnels soignants, parce qu'ils sont en charge des transports ou de la continuité des services publics. Je veux ce soir les remercier pour cet engagement. Je sais aussi que plusieurs d'entre vous ce soir sont seuls, souffrent ou sont malades et je sais que dans ces moments de fête et de retrouvailles, cette solitude et cette souffrance sont plus dures encore à supporter. Alors à nos concitoyens qui sont dans cette situation, je veux dire qu'ils appartiennent à une grande Nation et que les fils tendus qui nous tiennent, sont plus forts que leur solitude et je leur adresse une pensée fraternelle. L'année 2017 s'achève et je ne veux pas passer trop de temps à revenir sur celle-ci, ce fut l'année du choix : le choix du peuple français, votre choix par lequel vous m'avez attribué votre confiance et avec elle votre impatience, vos exigences, vos attentes ; j'en suis pleinement conscient. Aussi, depuis mon élection en mai dernier, je me suis attaché à simplement faire ce à quoi je m'étais engagé durant la campagne présidentielle. Le Premier ministre, son gouvernement sont à la tâche depuis le mois de mai dernier avec un Parlement profondément renouvelé, pour mettre en œuvre ses engagements. Par vos choix durant l'année 2017, vous avez profondément renouvelé notre vie politique et vous avez permis qu'une transformation en profondeur de notre pays advienne : à l'école pour nos enfants, au travail pour l'ensemble de nos concitoyens, pour le climat, pour le quotidien de chacune et chacun d'entre vous. Ces transformations profondes ont commencé et se poursuivront avec la même force, le même rythme, la même intensité pour l'année 2018. L'année qui s'ouvre en effet est celle de nombreux défis et nous construisons là une bonne part de notre avenir. Pour nos territoires ruraux et nous devons construire l'accès à la téléphonie mobile et au numérique, aux transports et permettre plus d'innovations économiques et sociales, pour nos quartiers populaires, en permettant la mobilité économique et sociale et en luttant contre les discriminations ; pour nos agriculteurs en leur permettant de vivre dignement du prix payé ; pour nos Outre-mer qui ont beaucoup souffert ces derniers mois et auxquels je veux adresser un salut tout particulier ; en adaptant nos règles et en construisant des filières économiques fortes qui permettront davantage d'autonomie énergétique et de créations d'emplois ; pour l'égalité entre les hommes et les femmes qui nécessitera la aussi des changements de loi et un ressaisissement de toute notre société ; pour les indépendants et les entrepreneurs, avec des règles simplifiées ; un droit à l'erreur enfin réalisé ; pour les salariés, en permettant la formation tout au long de la vie et des sécurités nouvelles ; et pour nos fonctionnaires en clarifiant leurs missions et nos attentes et en récompensant leurs efforts. Vous le voyez, tous ces chantiers jalonnent l'année 2018 et impliqueront un engagement plein et entier du Premier ministre et de son gouvernement. Je continuerai à faire ce pourquoi vous m'avez élu : rendre la France plus forte et plus juste ; permettre, non pas d'adapter notre pays aux changements du monde, mais lui permettre d'être ce qu'il est : un pays fort avec une exigence universelle qui, parce qu'il est plus fort, produit davantage, peut justement assurer la solidarité sur le sol national et avoir une exigence humaniste à l'international. Je sais que plusieurs d'entre vous ne partagent pas la politique qui est conduite par le gouvernement aujourd'hui ; je les respecte et je les écouterai toujours ; je m'assurerais que tous les débats soient conduits et que toutes les voix y compris d'opposants soient entendues mais pour autant, je n'arrêterai pas d'agir. Toujours j'écouterai, j'expliquerai notre situation et la réalité de celle-ci ; je respecterai et toujours à la fin, je ferai car c'est ce dont notre pays a besoin et c'est ce que



Macron  
64.02%



TEXT

DEEPLARNING

NEW METRICS



# Achievements

- Political observatory : <http://mesure-du-discours.unice.fr>
- hundreds visits / day
  - Twitter : [@MesureDiscours](https://twitter.com/MesureDiscours)

- Improvement of [Hyperbase](#) web
- New features based on deeplearning
  - New political databases



elysee



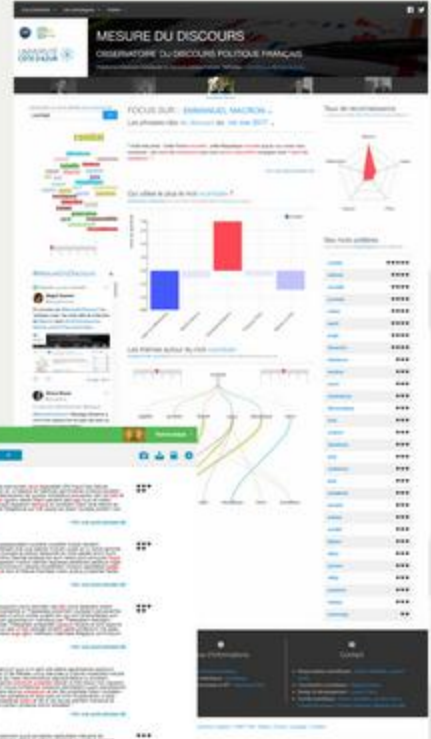
Campagne2017



Campagne2012



Campagne2007



# MESURE-DU-DISOURS

## CAMPAGNE PRÉSIDENTIELLE 2017

Observatoire du discours politique français. Méthodes [logométriques](#) & [deep learning](#)



Jean-Luc Mélenchon



Benoît Hamon



Emmanuel Macron



François Fillon



Marine Le Pen

FOCUS SUR LE DISCOURS : [EMMANUEL MACRON](#) ▾ 1ER MAI 2017 ▾ : [VOIR LE DISCOURS COMPLET](#)

### Cooccurrences

Rechercher un mot et afficher [ses cooccurents](#)



combat

réponse  
valeur transformation  
soir bataille égalité  
savant république  
ami partie lutte  
monde travail journée

### Passages-clés

Extraits statistiques basés sur le vocabulaire spécifique du discours

“ peur pour leurs enfants et nous devons aussi l' entendre . Le 7 mai prochain , Ce sera donc l' heure du choix . Entre Madame Le Pen l' héritière . Ne la sifflez pas , Combattez -la allez convaincre , Faites voter , faites -la perdre dimanche prochain . ”

### Qui utilise le plus le mot «[combat](#)» ?

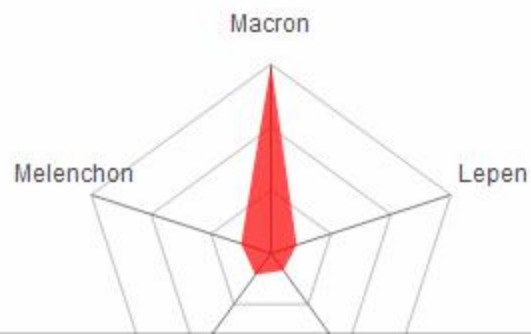
Distribution statistique du mot chez l'ensemble des locuteurs du corpus.

[Voir une autre phrase-clé](#)

### Taux de reconnaissance

Le discours reflète [64.02%](#) de Emmanuel

Macron



# Publications

## Book :

D. Mayaffre, C. Bouzereau, M. Ducoffe, M. Guaresi, F. Precioso, L. Vanni, "Les mots des candidats" , in P. Perrineau, *Le vote disruptif. Les élections présidentielles et législatives de 2017*, pp. 18-56

## Article

M. Guaresi et D. Mayaffre, "L'impossible réparation : l'image brisée de François Fillon durant la campagne présidentielle française", *Langage et Société*, 164/2, 2018, p. 39-56

## Presentations :

- M. Ducoffe, F. Precioso : **UCA Deep Learning School** (avec NVIDIA)
- L. Vanni : **Société Française de Statistique** (SFdS)
- D. Mayaffre : **Société d'Etude des Langages du Politique** (SELP)





# Recognition

Press :

- Paris Match : "Présidentielle 2017 : à chacun son vocabulaire"
- Nice Matin : "Ces chercheurs décryptent les discours des candidats"
- 20 minutes : "Des chercheurs dans le secret des discours des candidats"
- Le JDD : "Au débat, Marine Le Pen a commis une triple erreur"
- Le Figaro : "Emmanuel Macron fait un rêve comme Martin Luther King"
- Others : [leParisien](#), [LesEchos](#), [BFMTV](#), [LaDepeche](#), ...

Radio :

- France Inter : [Comme un bruit qui court](#)
- France Bleue

TV :

- M6 : [le 12-45](#)
- France 3 : [le 19/20](#)
- [Public Senat](#)

Academy :

- [UCA NEWS](#)
- [CNRS Hebdo](#)



## To sum up

About deep learning in the Human Sciences

Efficient methods for text classification  
Useless for linguistics without explanation



How does it work ?

**Second step**

**DeepText**

**IDEX JEDI project - Académie 1 & 5**

## Objectives

Open the black boxes



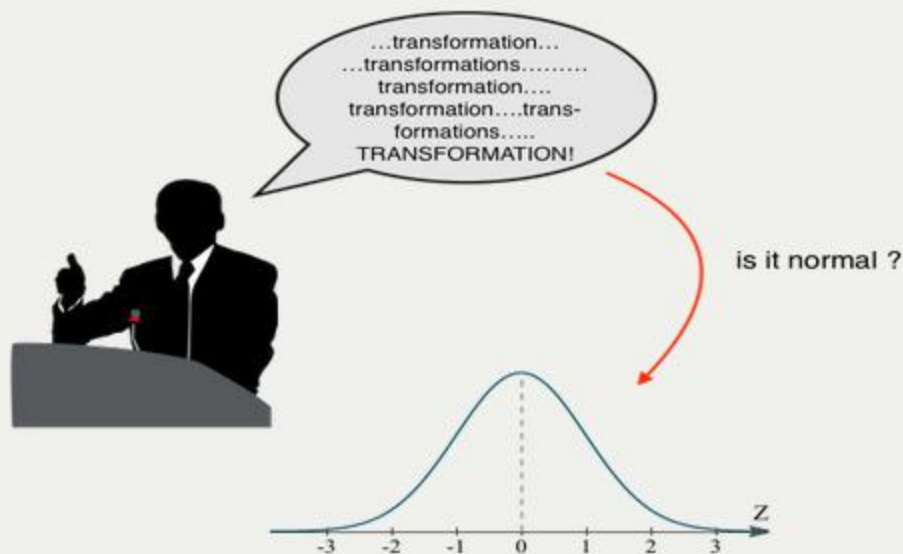
Both **linguistics** and **deep learning**



Texts are made up of **words**  
words can be **specific** to texts

Metrics :

- absolute frequency
- relative frequency
- **z-score**



## Benchmark

**Linguistics** can identify authors using **statistics**

|               | z-score         |         |                 |                             |
|---------------|-----------------|---------|-----------------|-----------------------------|
|               | without filters | $z > 0$ | filter by nouns | filter by nouns and $z > 0$ |
| <b>French</b> | 58%             | 52%     | 88%             | 89%                         |
| <b>Latin</b>  | 69%             | 62%     | 84%             | 82%                         |

"ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables" (L. Vanni, D. Mayaffre and D. Longrée), in D. Iezzi et al. (dir.) JADT' 2018, UniverItalia, Rome, 2018, pp. 459-466. [[hal-01823560](#)]

# Benchmark

## So can **deep learning**...

|               | z-score         |       |                 |                           | deep learning |
|---------------|-----------------|-------|-----------------|---------------------------|---------------|
|               | without filters | z > 0 | filter by nouns | filter by nouns and z > 0 |               |
| <b>French</b> | 58%             | 52%   | 88%             | 89%                       | 96%           |
| <b>Latin</b>  | 69%             | 62%   | 84%             | 82%                       | 91%           |

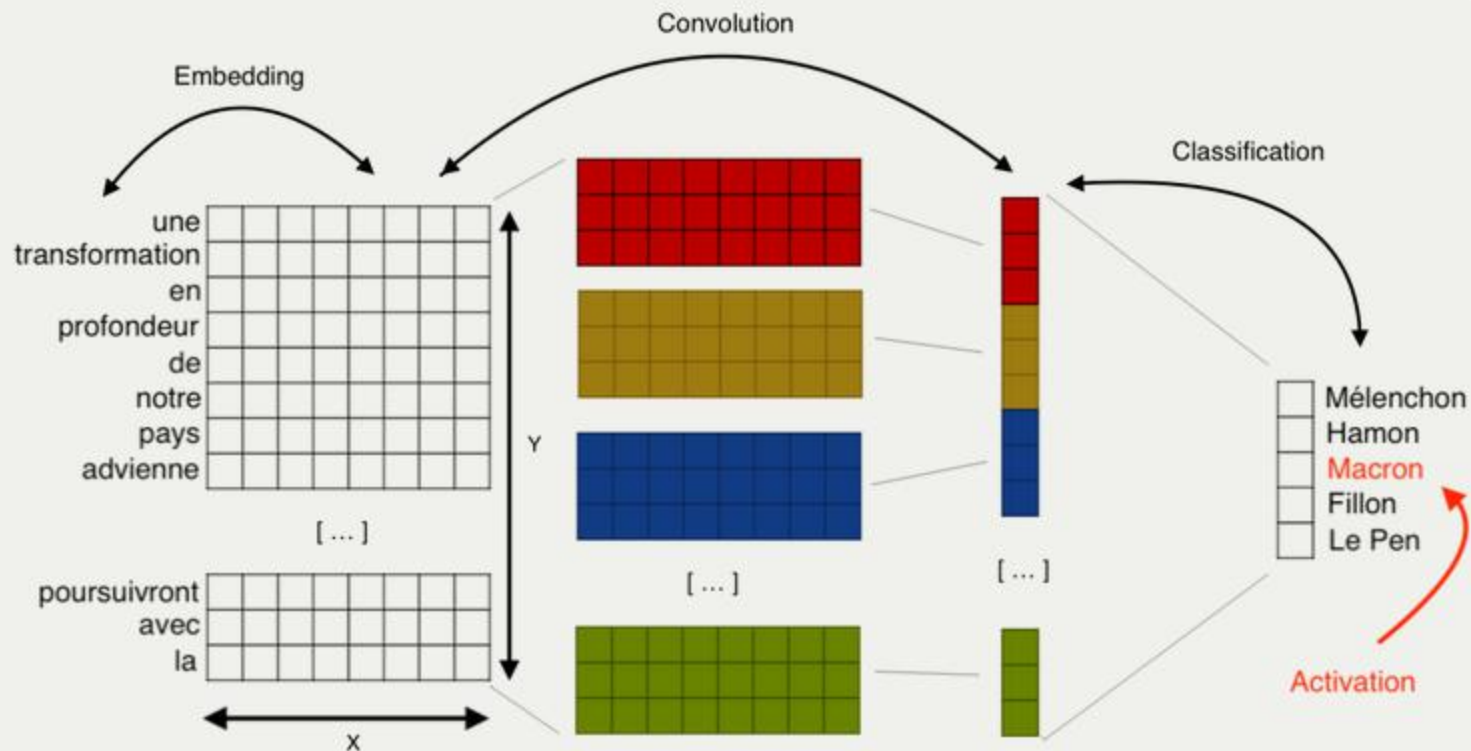
"ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables" (L. Vanni, D. Mayaffre and D. Longrée), in D. Iezzi et al. (dir.) JADT' 2018, UniverItalia, Rome, 2018, pp. 459-466. [[hal-01823560](#)]

**Deep learning**

**A deep method for a deep text...**

**How does it work?**

# Convolutional Neural Networks



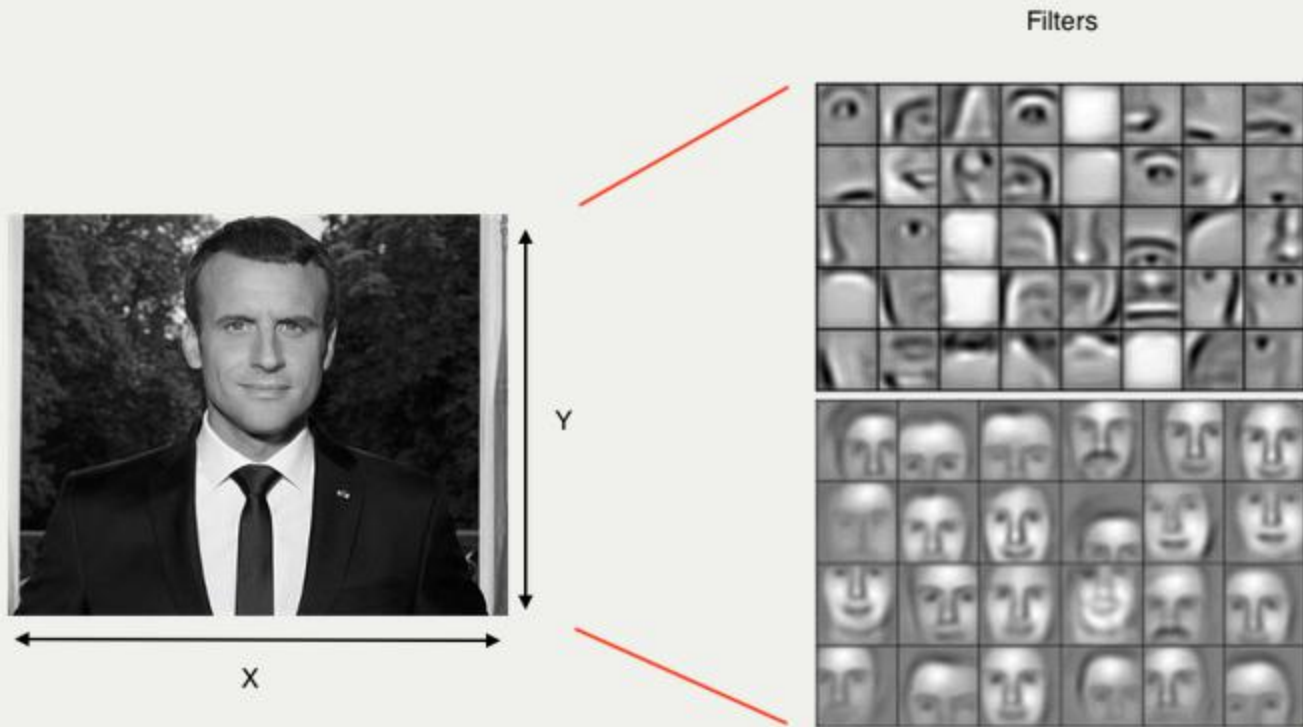
## Convolution

Learning an *abstraction* of the data



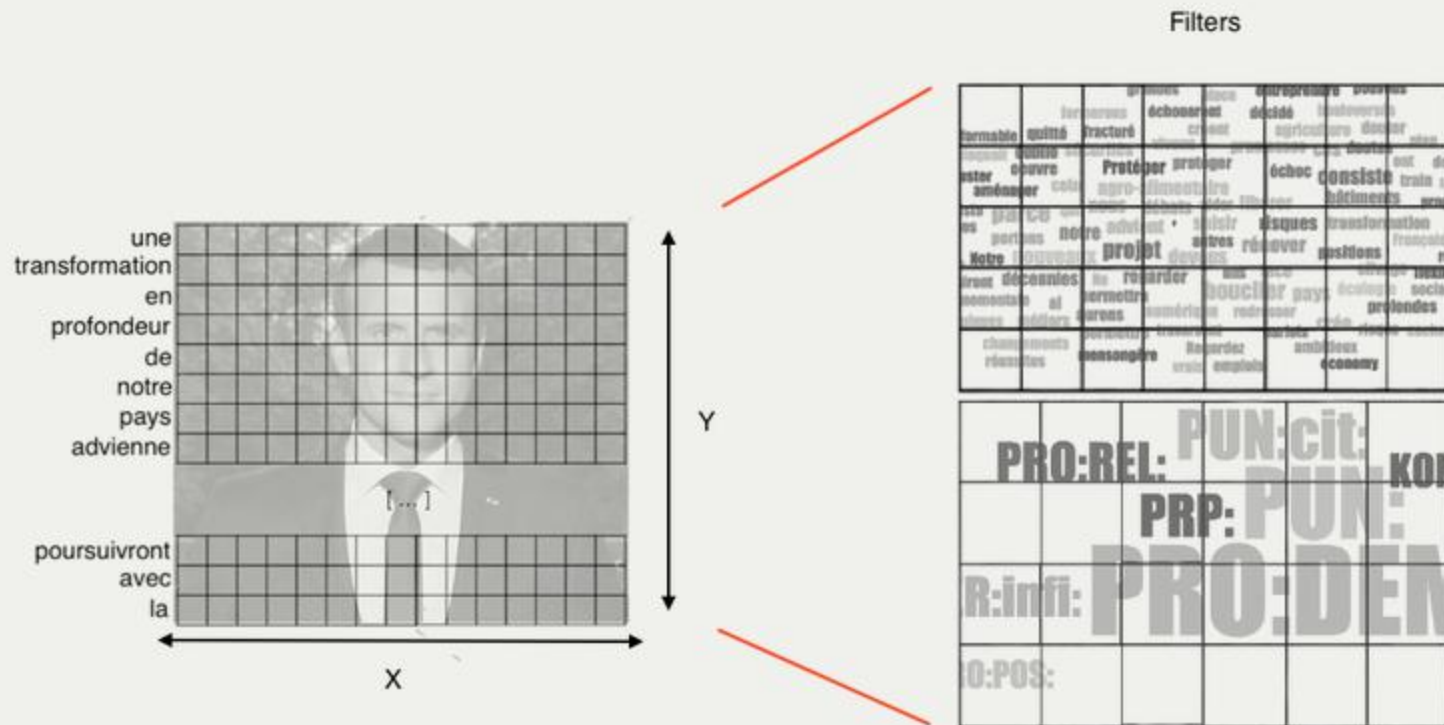
# Convolution

Learning an abstraction of Macron...



# Convolution

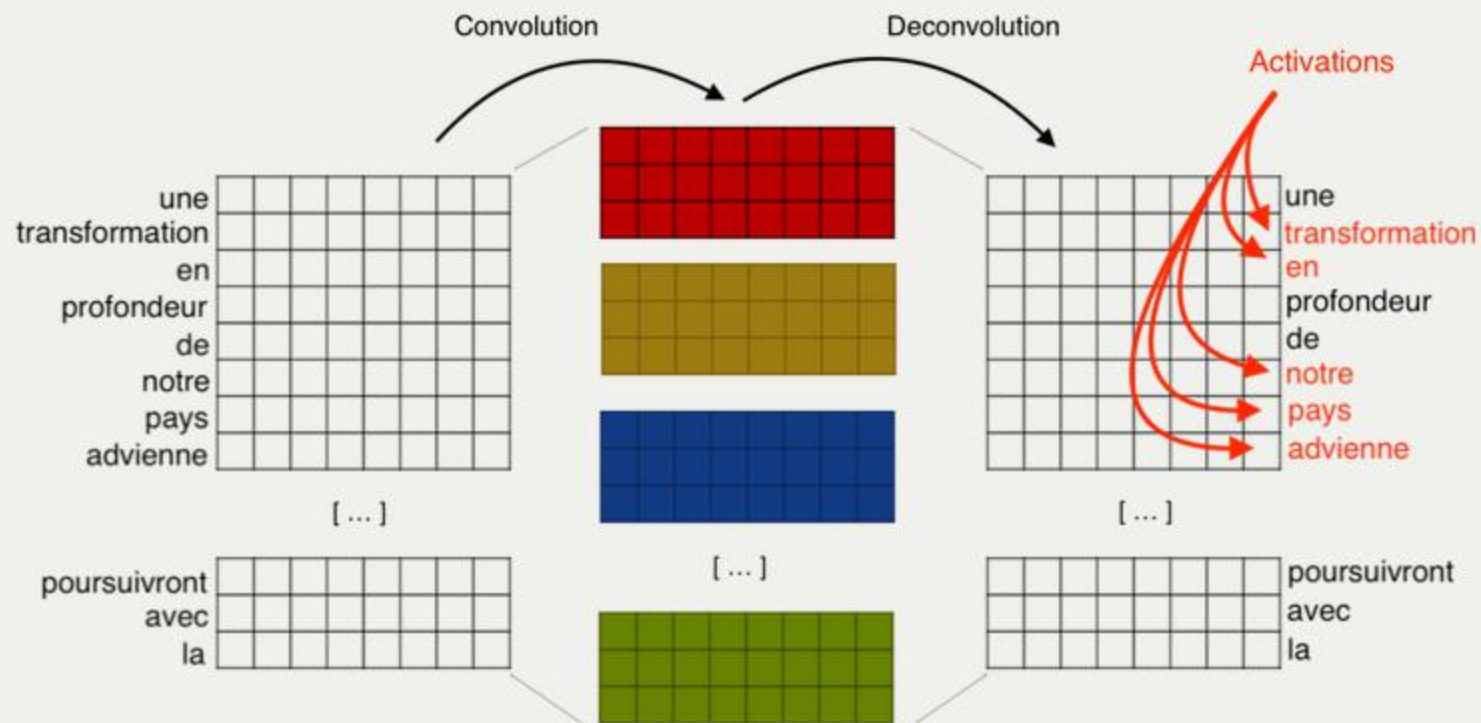
Learning another abstraction of Macron...



**can we (humans) read these filters ?**

# Deconvolution

Open the black box...





# Hyperbase

Web Edition



New project



Existing project



Library

## Lastest databases visited



elysee



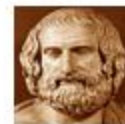
gauche\_droite



poetae



tragiques



euripide



latin

## Publications

**ACL 2018** - "Text Deconvolution Saliency (TDS): a deep tool box for linguistic analysis" (L. Vanni, M. Ducoffe and al.), 56th Annual Meeting of the Association for Computational Linguistics, Jul 2018, Melbourne [[hal-01804310](#)]

**JADT 2018** - "ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables" (L. Vanni, D. Mayaffre and D. Longrée), in D. Iezzi et al. (dir.) JADT' 2018, UniverItalia, Rome, 2018, pp. 459-466. [[hal-01823560](#)]

**Next step**

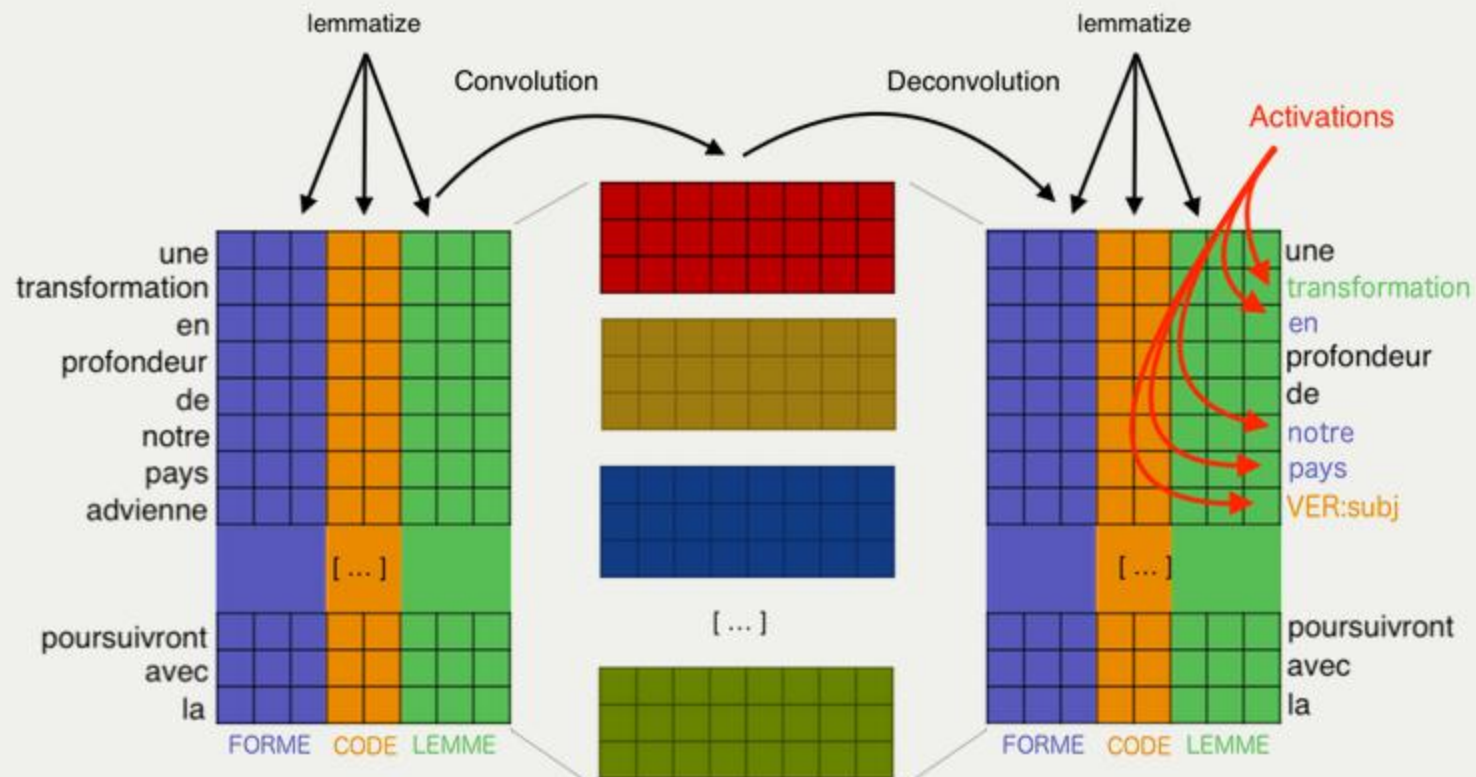
**Going deeper**

**Lemmatize the text**



# Deconvolution

Linguistic pattern detection ?



# Hyperbase

Web Edition



New project



Existing project



Library

Lastest databases visited



elysee



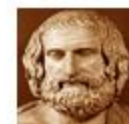
gauche\_droite



poetae



tragiques



euripide



latin

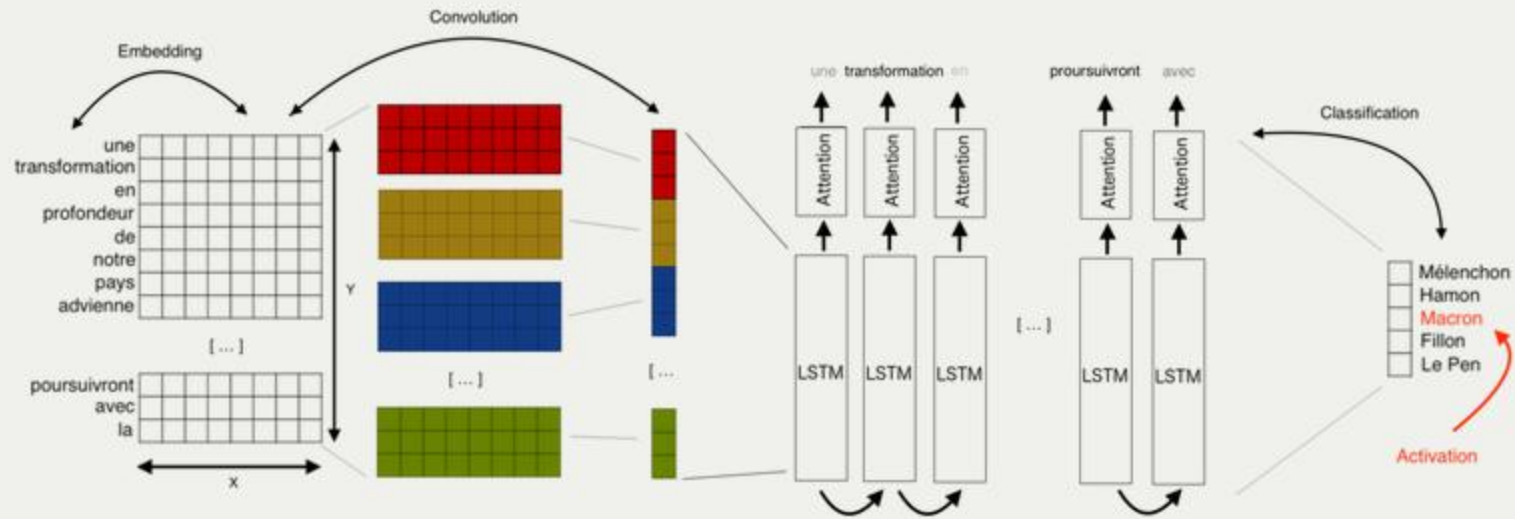


**One more step**

**Going deeper and deeper**

**LSTM + attention layer on the text**

# LSTM + Attention





About



New project



Existing project



Library

Lastest databases visited



elysee



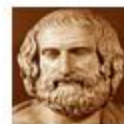
gauche\_droite



poetae



tragiques



euripide



latin



Looking for **linguistic** markers using **deep learning** methods

## Example with political discourse learning

---

See also :

- **ACL 2018** - "Textual Deconvolution Saliency (TDS) : a deep tool box for linguistic analysis" L. Vanni, M. Ducoffe, D. Mayaffre, F. Precioso et Al. In Proceedings Of *The 56th Annual Meeting of the Association for Computational Linguistics* [**hal-01804310**]
  - Technical demonstration : <http://deeptext.unice.fr>
  - Slides : <http://deeptext.unice.fr/DSmeetup>
- 

[laurent.vanni@unice.fr](mailto:laurent.vanni@unice.fr) - [damon.mayaffre@unice.fr](mailto:damon.mayaffre@unice.fr) - [frederic.precioso@unice.fr](mailto:frederic.precioso@unice.fr)



# DEEP TEXT

IDEX | TECHNICAL DEMONSTRATION POWERED BY [HYPERBASE.UNICE.FR](#) LINGUISTICS WEB TOOL

ABOUT

ENGLISH

LATIN

FRENCH

PUBLICATIONS

CONTACT

Using deep learning to do identification tasks on textual data (authorship detection, sentiment analysis, ...) is a well-known application. Neuronal networks are efficient on those tasks but black box oriented. The goal of this project is to open the box and gives us as many tools as possible to understand the underlying mechanisms such as linguistic pattern detection, occurrences, ...

We provide here three different languages ([English](#), [Latin](#), [French](#)), where you can observe how AI makes its decisions.





Looking for **linguistic** markers using **deep learning** methods

## Example with political discourse learning

---

See also :

- **ACL 2018** - "Textual Deconvolution Saliency (TDS) : a deep tool box for linguistic analysis" L. Vanni, M. Ducoffe, D. Mayaffre, F. Precioso et Al. In Proceedings Of *The 56th Annual Meeting of the Association for Computational Linguistics* [[hal-01804310](#)]
  - Technical demonstration : <http://deeptext.unice.fr>
  - Slides : <http://deeptext.unice.fr/DSmeetup>
- 

[laurent.vanni@unice.fr](mailto:laurent.vanni@unice.fr) - [damon.mayaffre@unice.fr](mailto:damon.mayaffre@unice.fr) - [frederic.precioso@unice.fr](mailto:frederic.precioso@unice.fr)