



Brève introduction à l'apprentissage automatique

par **Julien Velcin**

Université Lumière Lyon 2, laboratoire ERIC

<http://mediamining.univ-lyon2.fr/velcin>

DIU Enseigner l'informatique au lycée

juin 2020

Plan du cours

- Objectifs de l'apprentissage automatique
- Exemples d'application
- Cas d'étude 1 : classification de documents
- Cas d'étude 2 : traduction automatique
- L'apprentissage automatique en pratique
- Références



Objectifs du « machine learning »

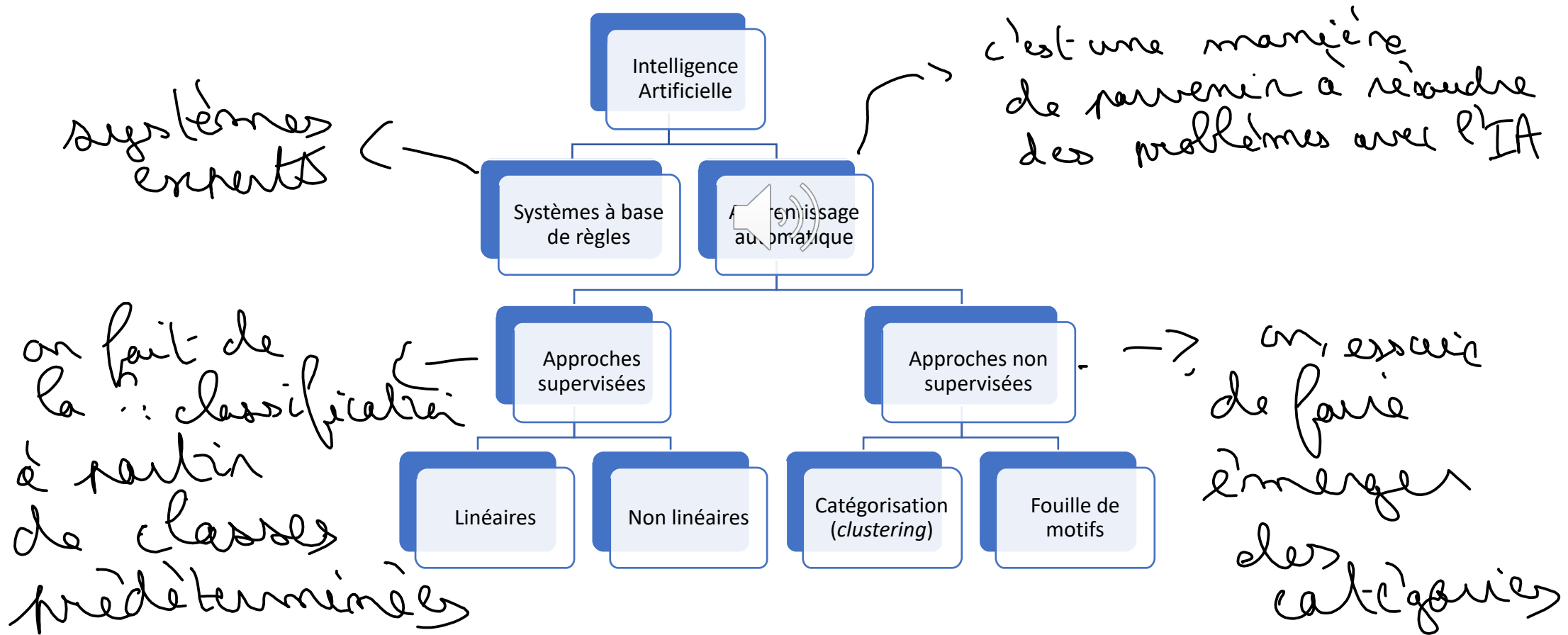
- L'objectif principal consiste à apprendre automatiquement à **généraliser** à partir d'exemples observés afin de pouvoir faire de l'inférence sur de **nouveaux exemples** jamais été observé auparavant (principe inductif)



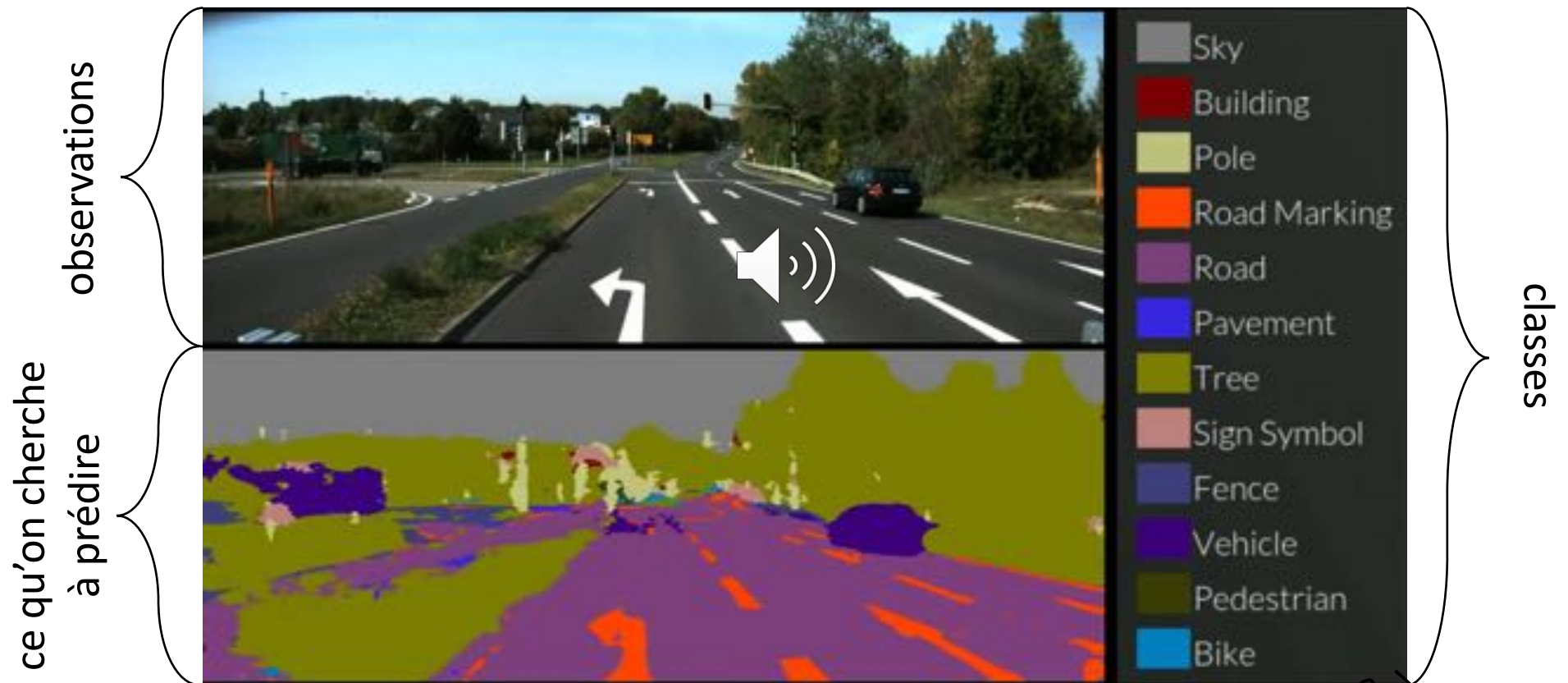
- Différentes familles d'algorithmes :

- apprentissage par coeur
- apprentissage par cas => exemple de l' algorithme des KNN
- apprentissage par renforcement
- classification non supervisée / catégorisation (*clustering*)
- classification supervisée, régression
- algorithmes génératifs

Positionnement vis-à-vis de l'IA

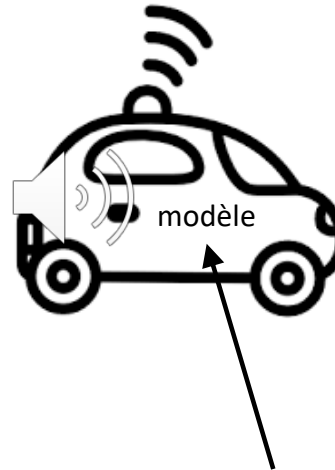
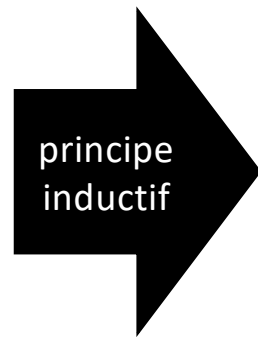


Exemple de la classification pour la reconnaissance d'objets dans des images

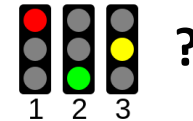


on veut d'une part segmenter l'image
mais classer ces parties de l'image

L'apprentissage automatique (ici, classification)



nouvelle situation



3 classes de comportements pour la voiture

attention : il s'agit d'un modèle *mathématique* !

Quelle est la classe la plus appropriée pour cette nouvelle situation



Exemples d'application

Julien Velcin

DIU EIL, 2019-2020

Jeux

Apprendre à jouer
aux jeux vidéos

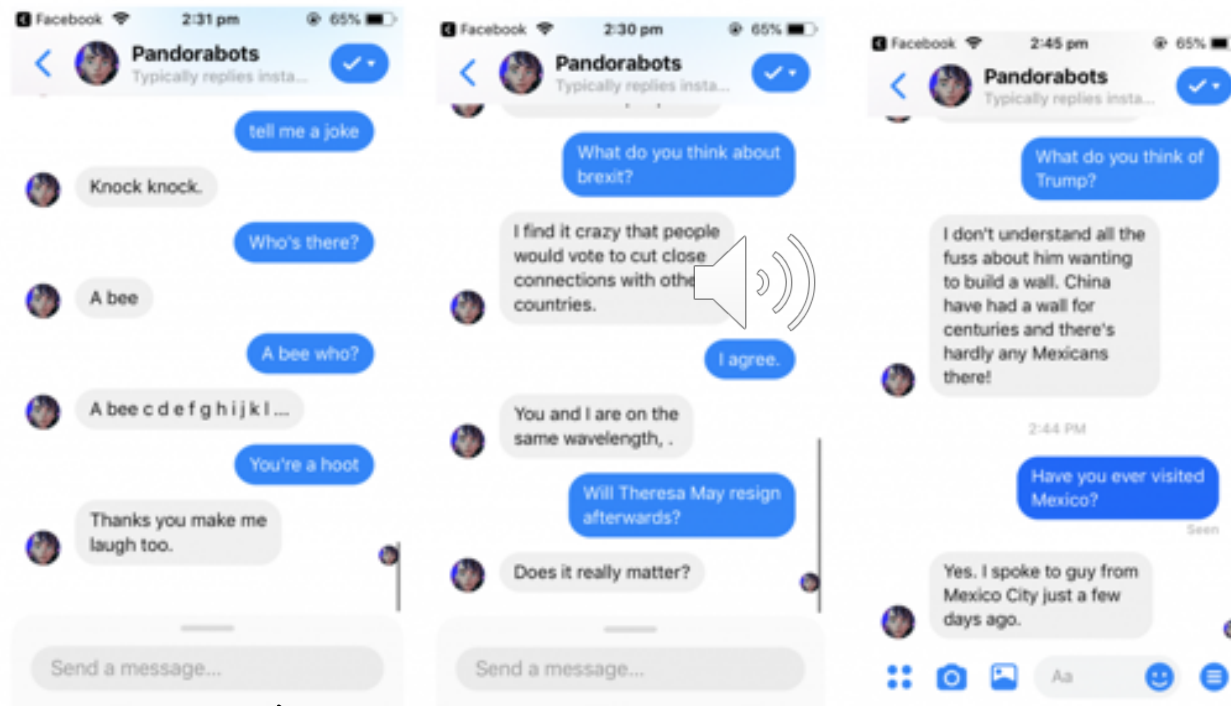
On essaie de
développer des
systèmes qui
apprennent à
jouer aux jeux
vidéos en jouant
contre elle-même, des
humains ou non.

On aimerait aussi que



Les images du jeu à plusieurs
seuls

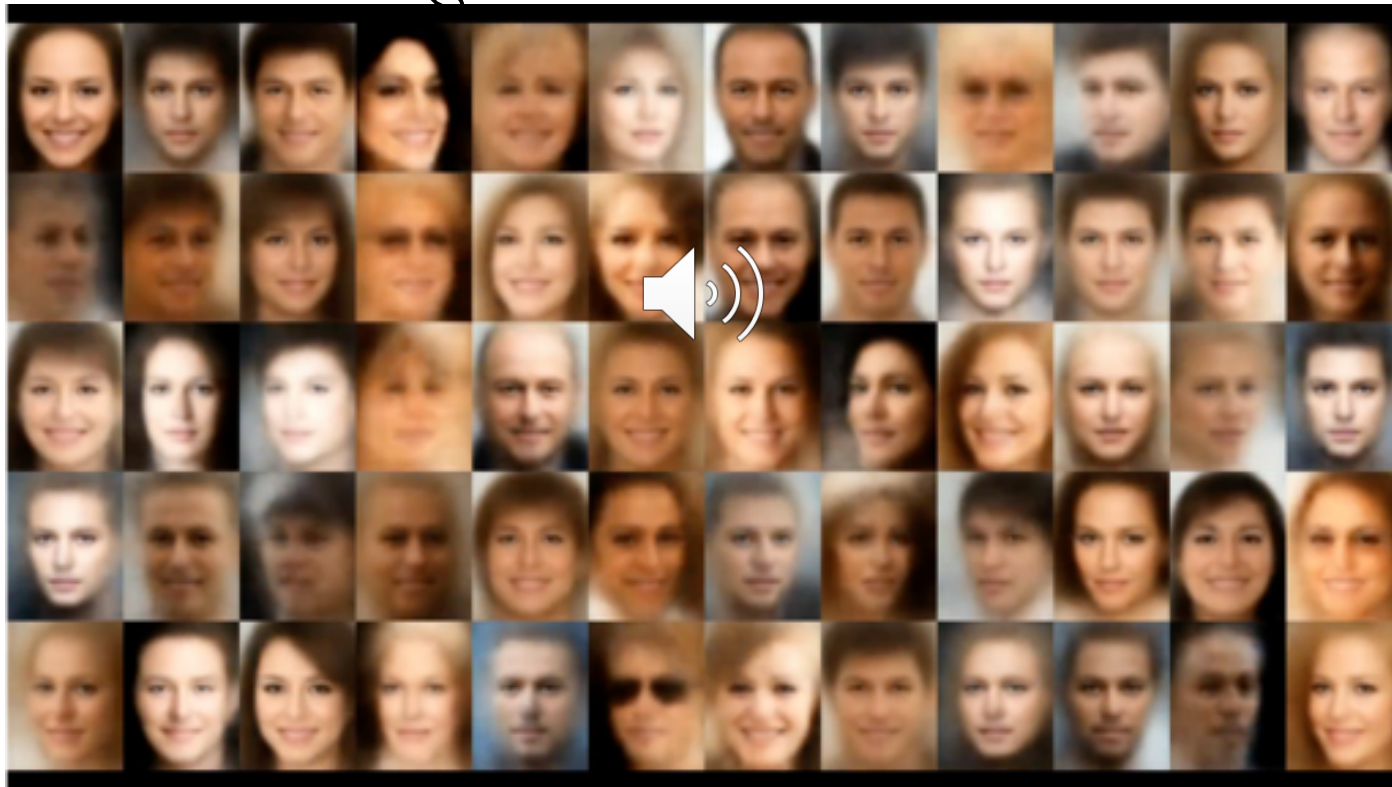
Robots conversationnels



Dans les années 60, les chat bot étaient
basés sur des règles. Maintenant
on apprend aux chat bot à partir de logs de
conversations

<https://www.pandorabots.com/mitsuku/>

Génération de visages avec des techniques
d'apprentissage profond.



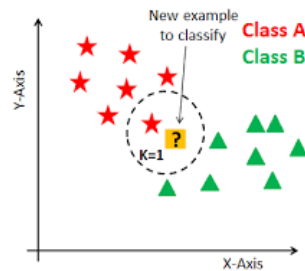
Cas d'étude 1 : classification de documents

Julien Velcin

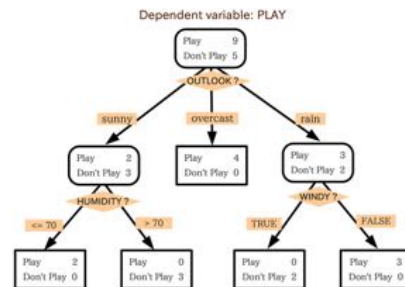
DIU EIL, 2019-2020

Classer des documents

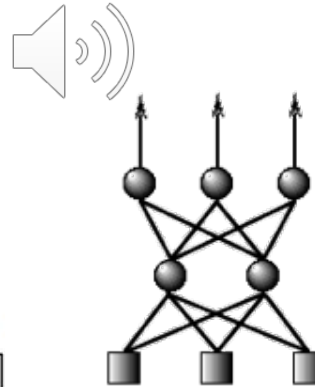
- Par exemple détecter la **polarité** d'un tweet ou classer un article dans une **thématique** déterminée (ex. sport ou économie)
- De (très) nombreux algorithmes existent :



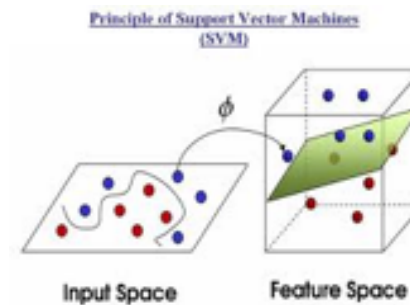
KNN



arbre de
décision



réseaux de
neurones
artificiels

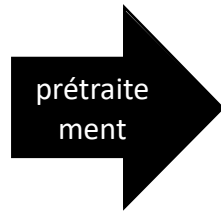


SVM

Par exemple la classification d'un article
selon la thématique

Procédure de traitement des données

matrice document-terme
 qui compte les occurrences des mots dans
 termes décrivant les documents (vocabulaire)
 par ex. T1 = « covid-19 », T2 = « football » et T3 = « masques »



| | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 |
|------|----|----|----|----|----|----|----|----|
| Doc1 | 0 | 0 | 4 | 3 | 0 | 1 | 0 | 2 |
| Doc2 | 0 | 2 | 4 | 0 | 2 | 3 | 0 | 0 |
| Doc3 | 4 | 0 | 1 | 3 | 0 | 1 | 0 | 1 |
| Doc4 | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 0 |
| Doc5 | 0 | 0 | 2 | 0 | 0 | 4 | 0 | 0 |
| Doc6 | 1 | 1 | 0 | 2 | 0 | 1 | 1 | 3 |
| Doc7 | 2 | 1 | 3 | 4 | 0 | 2 | 0 | 2 |

| |
|---|
| A |
| B |
| A |
| B |
| C |
| A |
| A |

classes

par ex. A = « covid-19 »
 B = « sport »
 C = « plongée »



modèle pour la classification
 d'un **nouveau** document Doc_n

traitement
 automatique
 des langues,
 feuille textuelle

Questions :

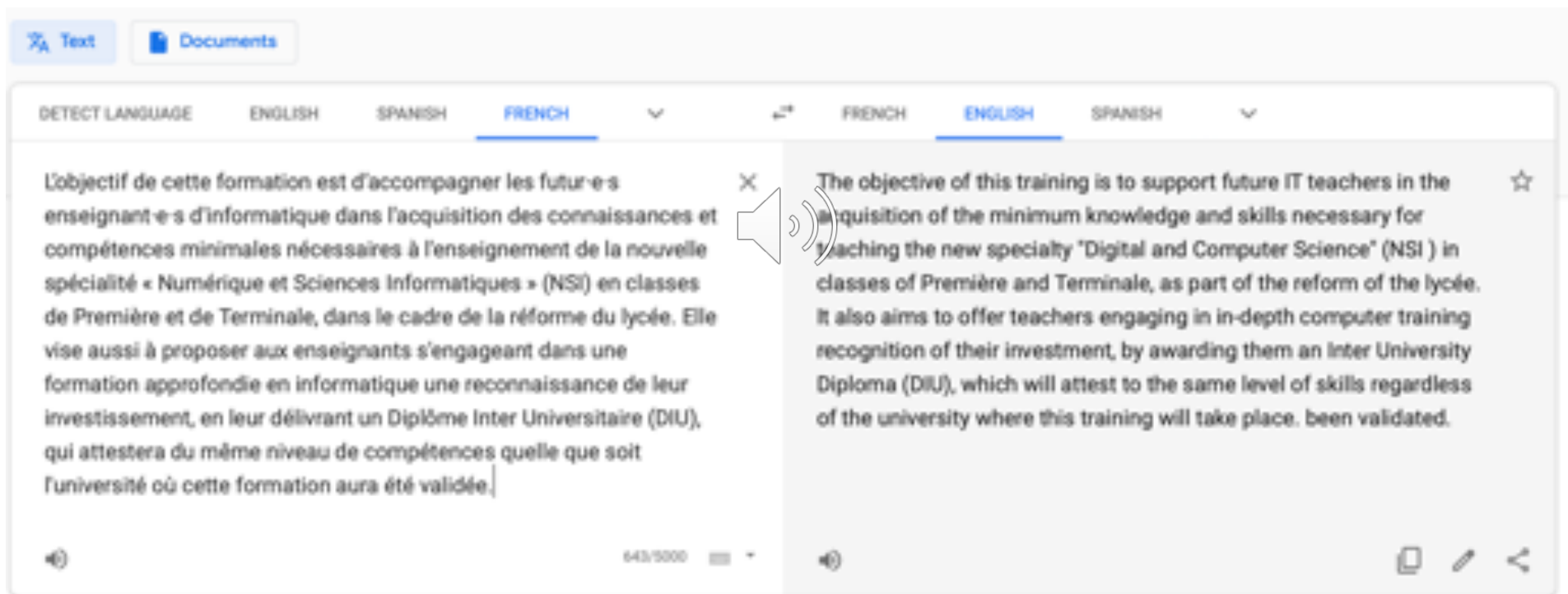
• Est-ce que connaître les
 mots rares est-
 intéressant ? les mots
 plus fréquents

Cas d'étude 2 : traduction automatique

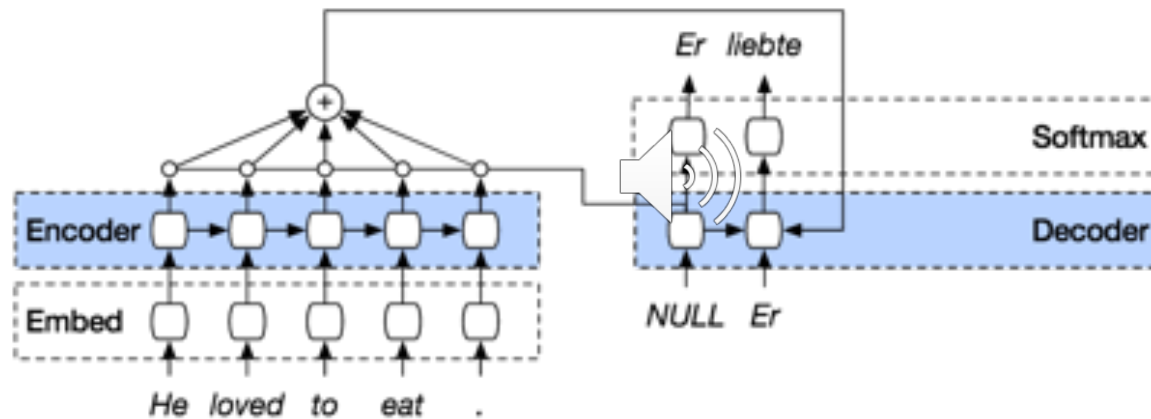
Julien Velcin

DIU EIL, 2019-2020

Traduire une langue en une autre



Réseaux de neurones profonds (*deep learning*)
prendre en compte l'ordre des mots, les
relations entre eux...



entrée : phrase
mots insérés dans le réseau de neurones
par une première couche puis vont être encodés
numériquement pour
Machine Learning car il faut fournir des exemples

L'apprentissage automatique

... en pratique



Julien Velcin

DIU EIL, 2019-2020

Quelques précautions à prendre

Important

- Bien définir la tâche qu'on souhaite résoudre

- Identifier les données qui vont permettre à la machine d'apprendre (attention aux biais !)

- Préparer les données à l'apprentissage

- Apprendre à bien généraliser : méthodologie de l'apprentissage automatique et sur-apprentissage

- Souvent plusieurs critères : précision des résultats, interprétabilité, consommation et empreinte écologique...

prendre des données collectées depuis une source unique \Rightarrow biais

Il faut préparer les données pour les élèves

Classification ?
supervisée
Régression pour prédire une valeur ?

(Prétraitement des données \Rightarrow demande du temps)



* risque d'overfitting
• méthodologie: séparer les données
d'apprentissage des données de test.

* * sensibiliser au coût écologique des
calculs
• réseaux de neurones artificiels
L> Boîtes noires



Quelques références

Julien Velcin

DIU EIL, 2019-2020

Ouvrages

