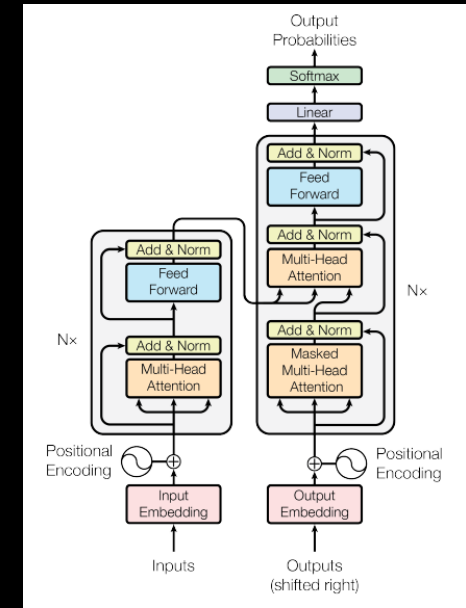




# Une date à retenir

« Attention is all you need »



N-gram

Tf-idf

Sac de mots

Plongement  
de mots

LSTM

RNN

BARThez

CamemBERT

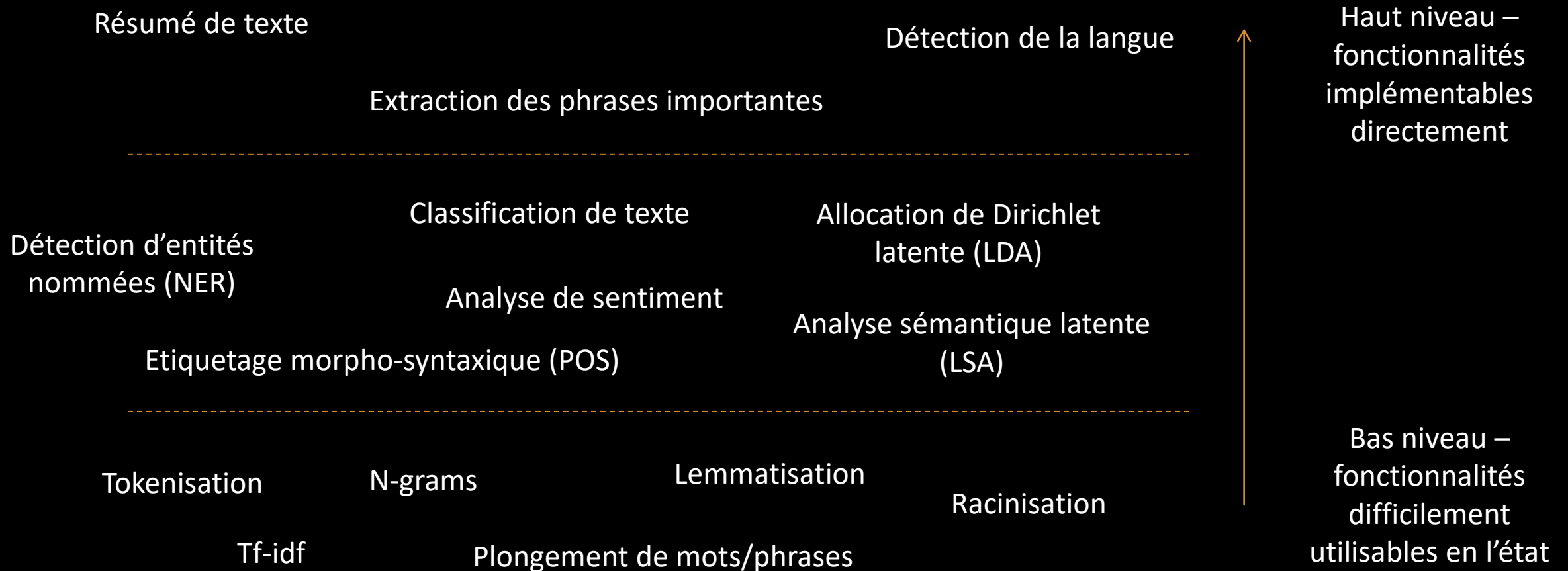
FlauBERT

2019

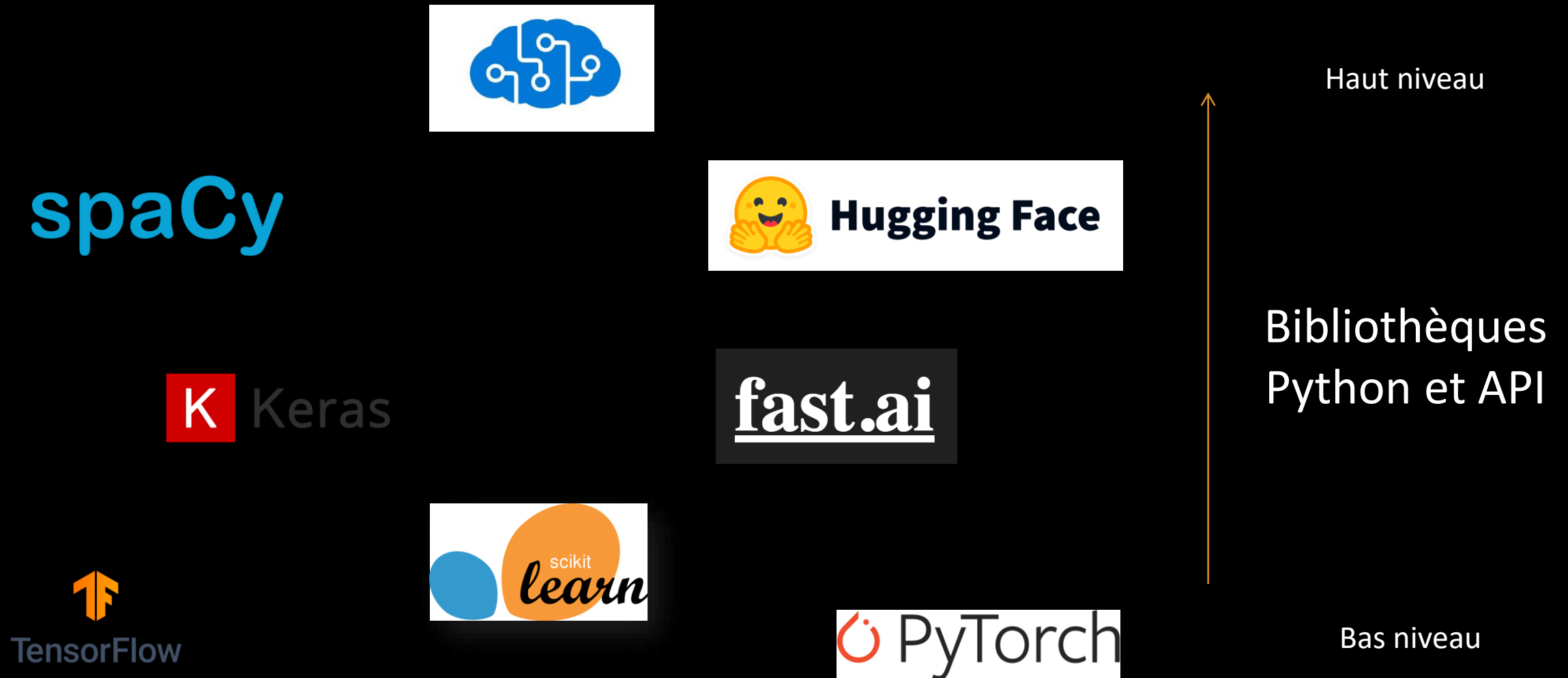
2020

Juin 2017

# Du bas niveau au haut niveau



# Du bas niveau au haut niveau



# Détection d'entités nommées

Named Entity Recognition (NER)



```
import spacy
nlp = spacy.load("fr_core_news_sm")
doc = nlp("Les meetups python Grenoble sont une excellente idée - il faut en faire
davantage !")
for ent in doc.ents:
    print(ent.text, ent.label_, ent.start_char, ent.end_char, spacy.explain(ent.label_))
```

Text	Label	Start char	End char	Explanation
Grenoble	LOC	19	27	Non-GPE locations, mountain ranges, bodies of water

# Etiquetage morpho-syntaxique



Part-Of-Speech (POS) tagging

```
import spacy
nlp = spacy.load("fr_core_news_sm")
doc = nlp("La clarté est préférable à l'opacité.")
for token in doc:
    print(token.text, token.lemma_, token.pos_, token.tag_, token.dep_,
          token.shape_, token.is_alpha, token.is_stop)
```

Text	Lemma	POS	Tag	Dep	Shape	alpha	stop
La	le	DET	DET	det	Xx	True	True
clarté	clarté	NOUN	NOUN	nsubj	xxxx	True	False
est	être	AUX	AUX	cop	xxx	True	True
préférable	préférable	ADJ	ADJ	ROOT	xxxx	True	False
à	à	ADP	ADP	case	x	True	True
l'	le	DET	DET	det	x'	False	True
opacité	opacité	NOUN	NOUN	obl:arg	xxxx	True	False
.	.	PUNCT	PUNCT	punct	.	False	False

# Analyse de sentiment

Sentiment analysis



**Hugging Face**

```
import torch
from transformers import AutoTokenizer, AutoModelForSequenceClassification
tokenizer = AutoTokenizer.from_pretrained("moussaKam/barthez-sentiment-classification")
model = AutoModelForSequenceClassification.from_pretrained("moussaKam/barthez-sentiment-classification")
tok = tokenizer.encode("Les meetups Python sont une excellente idée - il faut en faire davantage !")
print("Positif" if torch.argmax(model(torch.tensor([tok]))[0]).item() == 1 else "Négatif")
```

Les meetups Python sont une excellente idée - il faut en faire davantage !



POSITIF

Les meetups Python ne sont pas une excellente idée - il faut en faire davantage !



NEGATIF



Image par [Gerd Altmann](#) de [Pixabay](#)