

Prédire l'aspect linguistique en anglais au moyen de *transformers*

Aspect linguistique : *télicité* et *durée*

Aspect : Propriété temporelle des actions, des événements et des états.

Télicité : Une action qui a un point final est **télique**.

The soup cooled in an hour. I ate an apple.

Un état ou une action qui n'a pas de point final est

atélique. *The soup cooled for an hour. I eat apples.*

Durée : Une action **durative** a une durée perceptible, avec ou sans point final.

The snow melts every spring. I slept all morning.

Une action **stative** n'a pas de durée perceptible.

Your idea sounds great. I like chocolate.

Modèles *transformers*

Architecture *encoder* avec un mécanisme de *self-attention* à (12/24) *couches* et (12/16) *têtes* :

BERT, RoBERTa, ALBERT, XLNet.

Jeu de données

1. **Friedrich and Gateva (2017)** : 6354 phrases annotées pour la *télicité* ou la *durée*.

Aspect	Étiquette	Phrases	Train	Validation	Test
Télicité	télique	3,220	5,083	635	636
	atélique	3,134			
Durée	stative	1,861	4,095	512	512
	dynamique	3,258			

2. **Paires minimales de test** : jeu de données que nous avons créé nous-même, composé de 80 phrases annotées pour la *télicité* et 40 phrases annotées pour la *durée*.

Notre question de recherche

Les modèles *transformers* sont-ils sensibles à l'aspect verbal ?

Méthodologie

modèles pré-entraînés

+ de **phrases annotées** pour la *télicité* / la *durée*
(+ de **vecteurs token_type_ids** (position du verbe))



Finetuning pour la classification binaire de la *télicité* / la *durée* des phrases



Évaluation **quantitative** et **qualitative**

Vecteur token_type_ids

tokens	He	worked	well	and	earned	much	.	[SEP]
token_type_ids	0	1	0	0	0	0	0	0

Modèles utilisés & précision

Modèle	Télicité	Durée
bert-base-uncased	0.72 / 0.68	0.84 / 0.74
bert-base-cased	0.79 / 0.67	0.85 / 0.73
bert-large-uncased	0.72 / 0.68	0.79 / 0.73
bert-large-cased	0.79 / 0.66	0.74 / 0.72
roberta-base	0.65	0.72
roberta-large	0.67	0.64
xlnet-base-cased	0.62 / 0.64	0.71 / 0.73
xlnet-large-cased	0.51 / 0.64	0.64 / 0.64
albert-base-v2	0.51 / 0.62	0.82 / 0.71
albert-large-v2	0.51 / 0.51	0.64 / 0.64
CNN (50 ep., sans lemmat.)	0.60	0.65
Régression logistique	0.53	0.64

(avec/sans le vecteur token_type_ids)

Résultats quantitatifs

Les modèles *base* sont plus performants que les modèles *large*.

Les modèles *cased* sont (un peu) plus performants que les modèles *uncased*.

Les modèles qui ont des représentations plus complexes (sub-words, dépendances) sont moins performants.

Le jeu de données est trop petit pour le *finetuning*.

Télicité:

- ▶ xlnet-large-cased, albert-base-v2, albert-large-v2 sont peu performants
- ▶ token_type_ids fait augmenter la précision (<13%)

Durée:

- ▶ xlnet-large-cased, roberta-base, albert-large-v2 sont peu performants
- ▶ token_type_ids fait augmenter la précision (<12%)

Résultats qualitatifs

Télicité:

Bonne précision, mais des erreurs sur les phrases avec un verbe qui est (a)télique dans un contexte opposé:

I eat_[+télique] a fish for lunch on Fridays_[-télique].
→ classée télique

Les paires minimales: des cas plus complexes

The PM made that declaration yesterday_[+télique].
The PM made that declaration for months_[-télique].
→ classées téliques

Durée:

Bonne précision, mais quelques erreurs de classification dans tous les modèles:

Do you hear music?_[+stative] → classée durative