



HAL
open science

VarIDE at PARSEME Shared Task 2018

Caroline Pasquer, Agata Savary, Jean-Yves Antoine, Carlos Ramisch

► **To cite this version:**

Caroline Pasquer, Agata Savary, Jean-Yves Antoine, Carlos Ramisch. VarIDE at PARSEME Shared Task 2018. COLING Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018), Aug 2018, Santa Fe, United States. hal-01866366

HAL Id: hal-01866366

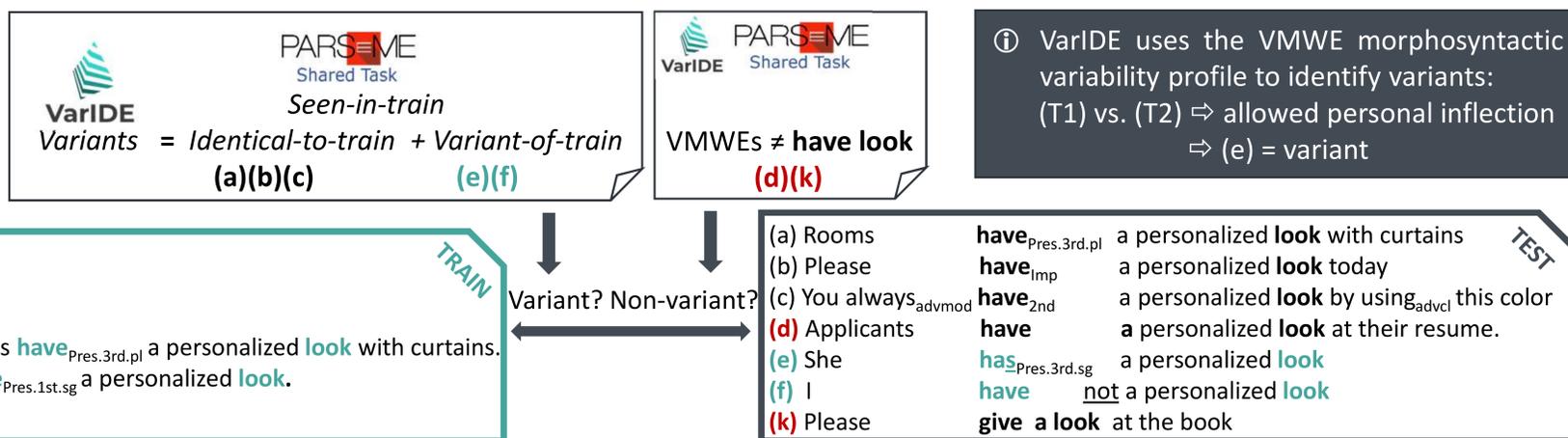
<https://hal.science/hal-01866366>

Submitted on 3 Sep 2018

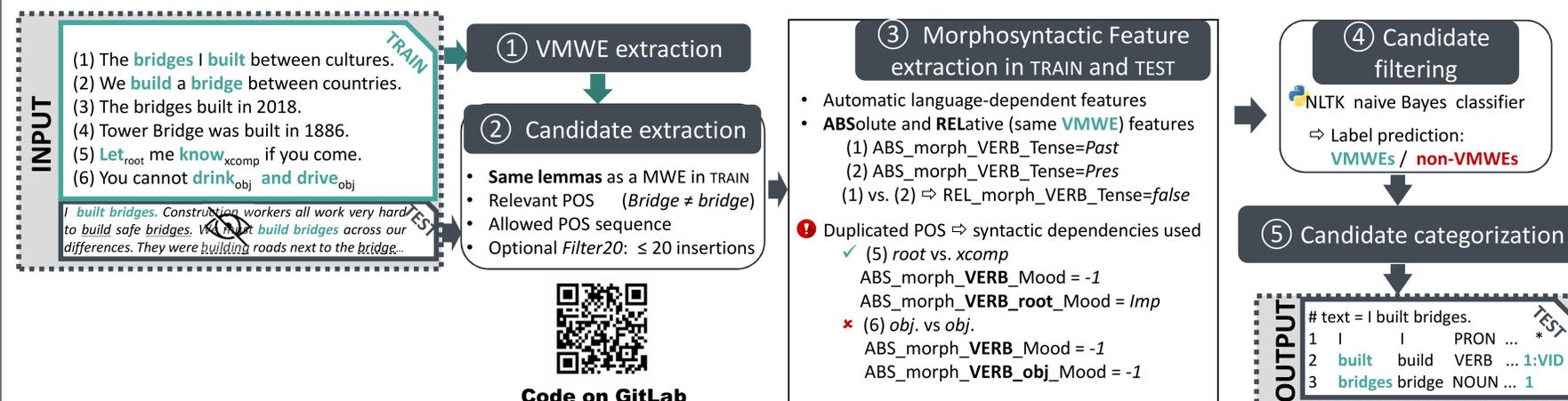
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Verbal MWEs (VMWEs) Variant IDENTification based on previous work^[1]

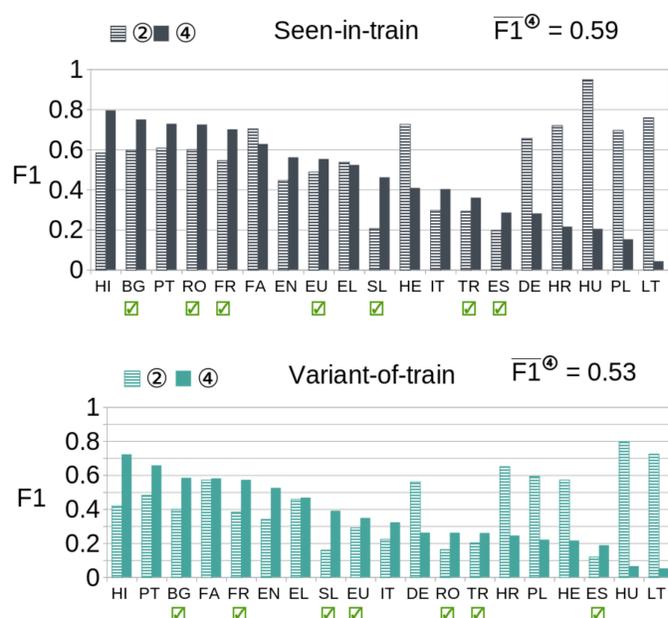


VarIDE system for 19 languages



Results

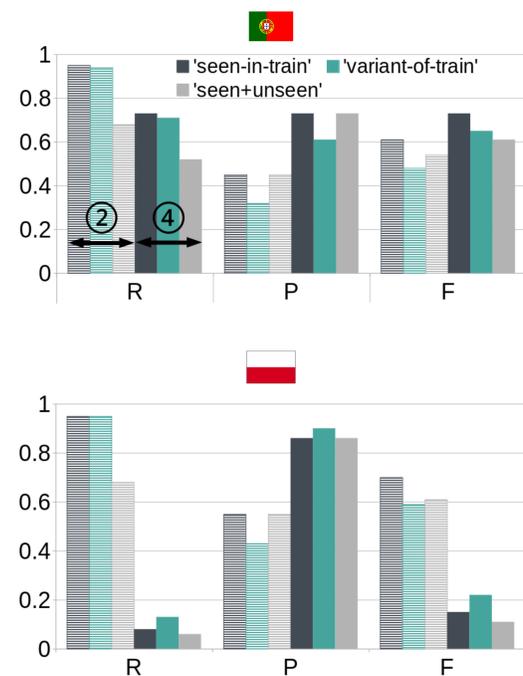
- Recall for **Extraction**^② and **Filtering**^④ (✓ when R > 0.8)



Problems & remedies

- Feature extraction^③ sensitive to TRAIN spurious/unbalanced data
- Non-UD tagsets ⇒ adjustments when possible
- Missing (TR), inconsistent (FR) lemmatization ⇒ handled for FR

Results Portugal vs. Poland



Filtering^④ ↑F1 in PT ↓F1 in PL
⇒ Tagset size matters?

Conclusions and Perspectives

- global F1 = 0.46
- F1(Discontinuity) = 0.37
- Detailed error analysis ⇒ Similar tendency per language family?
- Other classifiers: linear SVM, multilayer perceptron
- Correlation ABSolute/RELative features vs. linguistic justifications?

References, acknowledgements

[1] PASQUER et al., *If you've seen some, you've seen them all: Identifying variants of multiword expressions*, COLING 2018.

Work funded by the French PARSEME-FR grant