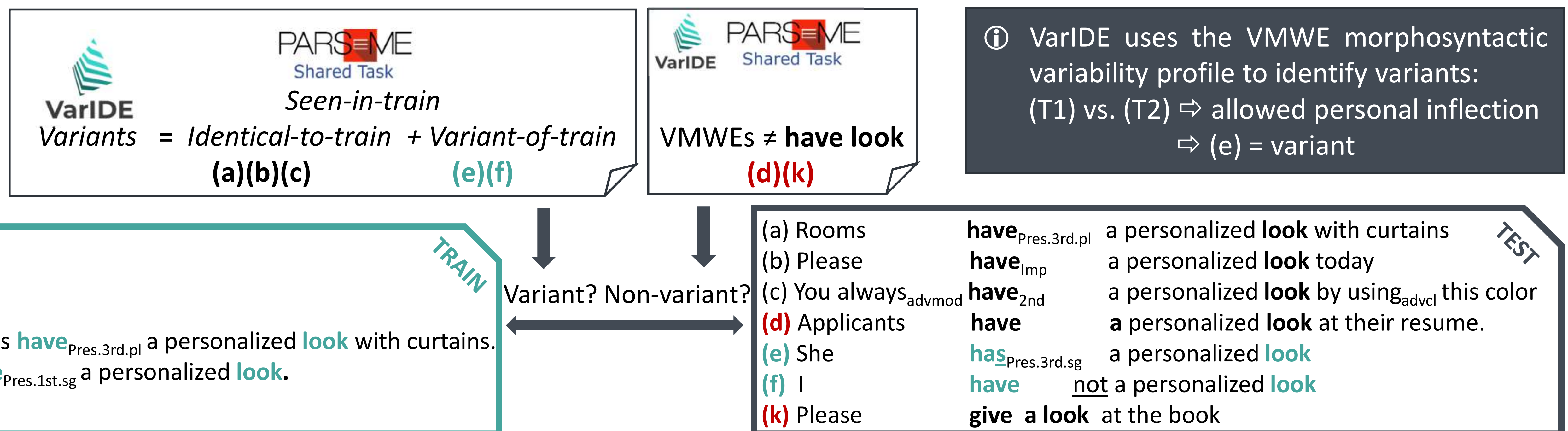
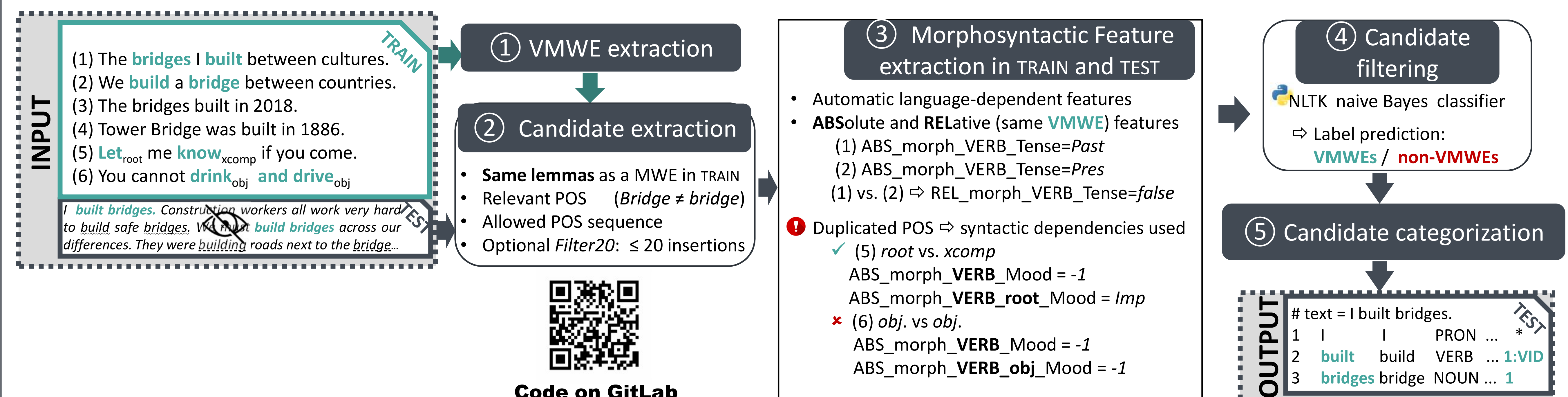


Verbal MWEs (VMWEs) Variant IDentification based on previous work^[1]

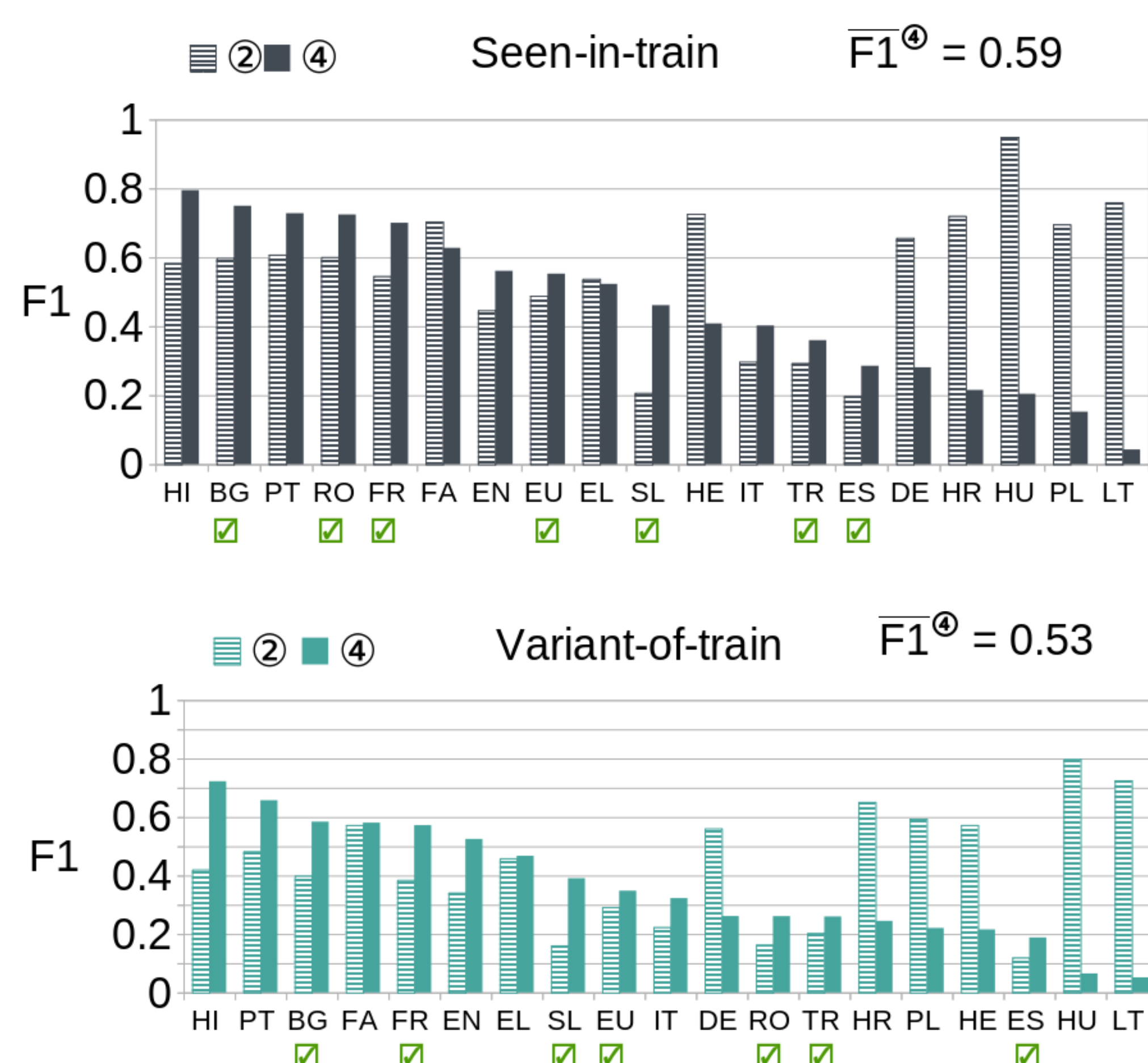


VarIDE system for 19 languages



Results

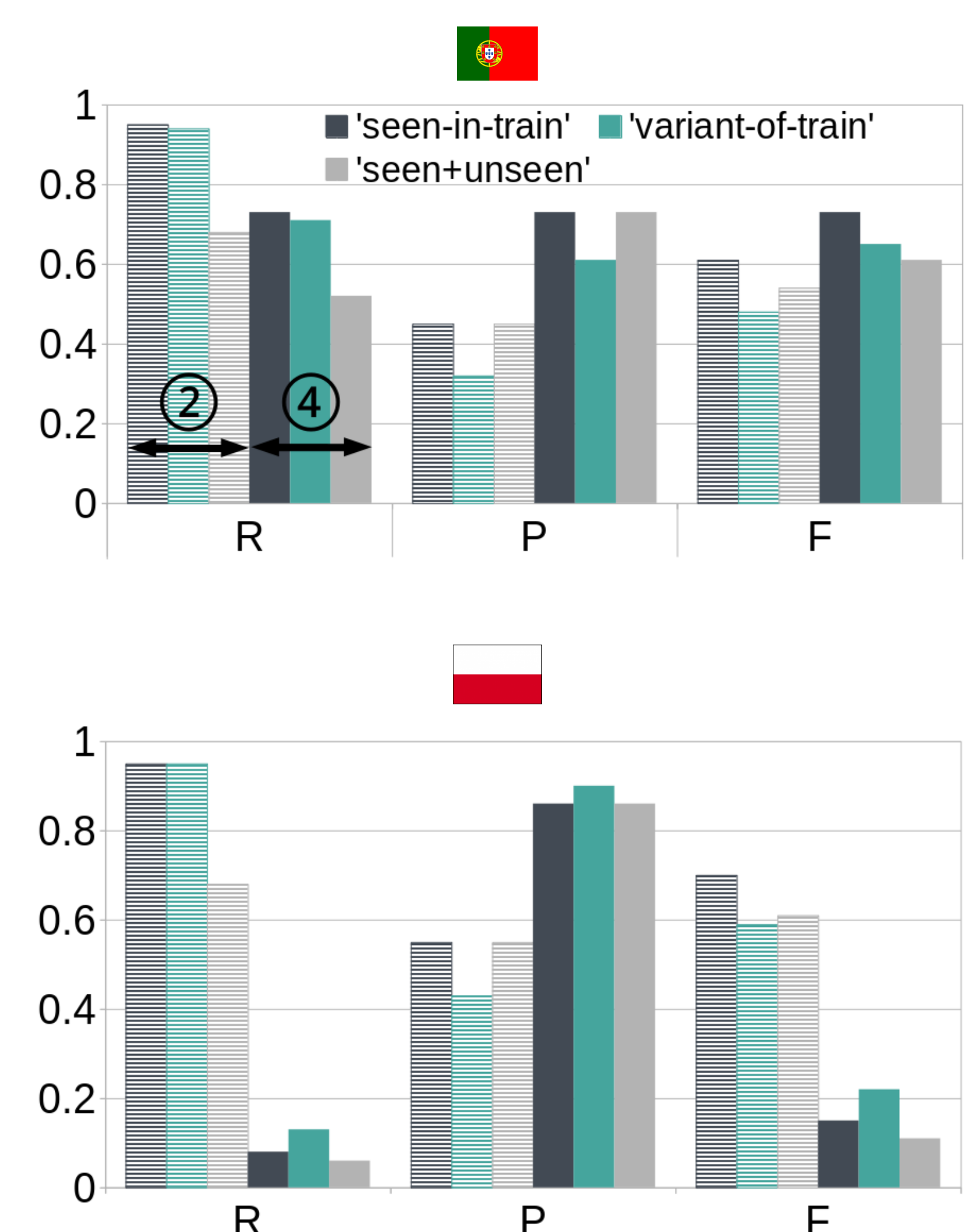
- Recall for **Extraction**^② and **Filtering**^④ (✓ when R > 0.8)



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 🇵🇹 vs. 🇵🇱



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