TAGGING THE SWITCH

Exploring POS & LID Tagging in English-Spanish Tweets

Rocky Klopfenstein
rklopfenstein27@amherst.edu

Christian Manzi
cmanzi25@amherst.edu

Gretta Ineza
gineza26@amherst.edu

COSC-243-2425S
Professor Wein

01 | BACKGROUND

- **Language Identification** (LID) identifies the language of each word in multilingual text
- **Parts of Speech Tagging** (POS) assigns grammatical categories (e.g., noun, verb) to each word
  - *Both have been studied in code-switched text, though rarely in combination*
- **Motivation:** As a group of bilingual student researchers, we were drawn to multilingual conversational texts. We aim to improve translational technologies through enhanced LID & POS tagging to support linguistic equity and inclusion

02 | METHODOLOGY

- **Data:** 'Bangor-Miami' corpus containing English-Spanish code-switched tweets (Deuchar et al., 2009)
- **Models:** XLM-R and RemBERT, with BERT as a baseline
- **Finetuning:** Various methods of continued pretraining:
  - Single Task (LID or POS independently)
  - Sequential (e.g., LID → POS)
  - Joint Multitask (LID + POS)
  - Joint Multitask w/ Loss Weighting
- **Evaluation:** Measure performance using F1 score, accuracy, and qualitative analysis of tagged outputs

03 | RESULTS

- Single-task and sequential models performed similarly overall; (POS - LID) showed slight **improvements in F1 and accuracy**, while task-specific shifts suggest task interference
- Unweighted multitask models underperformed, with F1 dropping 4-7%, likely due to task imbalance, which degraded joint performance
- Applying **uncertainty-based loss weighing** (Kendall et al., 2018) improved task balance and reduced task interference, leading to slight improvements in the combined F1 score
- Accuracy scores outperformed F1 scores in multitask models (≥ 97.5 ± 0.2%), indicating class imbalance and overprediction of majority classes
- **RemBERT** outperformed all other models, confirming its strength in multilingual learning
- We argue that sequential learning offers *practical advantages*, and loss-weighted joint multitask training shows clear, scalable benefits for future work

# word	lid	pos
pero	spa	CONJ
viste	spa	VERB
las	spa	DET
cositas	spa	NOUN
que	spa	PRON
compraron	spa	VERB
para	spa	ADP
los	spa	DET
speed	eng	NOUN
bumps	eng	NOUN
?	eng&spa	PUNCT

Sample sentence from data

Model	Single Task (LID)	Single Task (POS)	Sequential (POS - LID)	Sequential (LID - POS)	JMT - Unweighted	JMT - Weighted
XLM-R	0.930	0.966	0.931	0.966	0.892	0.893
RemBERT	0.935	0.968	0.937	0.966	0.886	0.887
mBERT	0.927	0.966	0.928	0.9657	0.894	0.895
Average	0.931	0.967	0.932	0.966	0.891	0.892

Table 1. Macro-averaged F1 scores

04 | FUTURE WORK

- Expand to **additional language pairs**, such as Hindi-English or Arabic-French, to assess generalizability
- Investigate new data sources beyond tweets, such as **formal literature**, to explore how style and text formality affect LID/POS tagging
- Incorporate additional **morphological features** to improve accuracy