Neural Factor Graph Models for Cross-lingual Morphological Tagging

Chaitanya Malaviya, Matthew R. Gormley, Graham Neubig

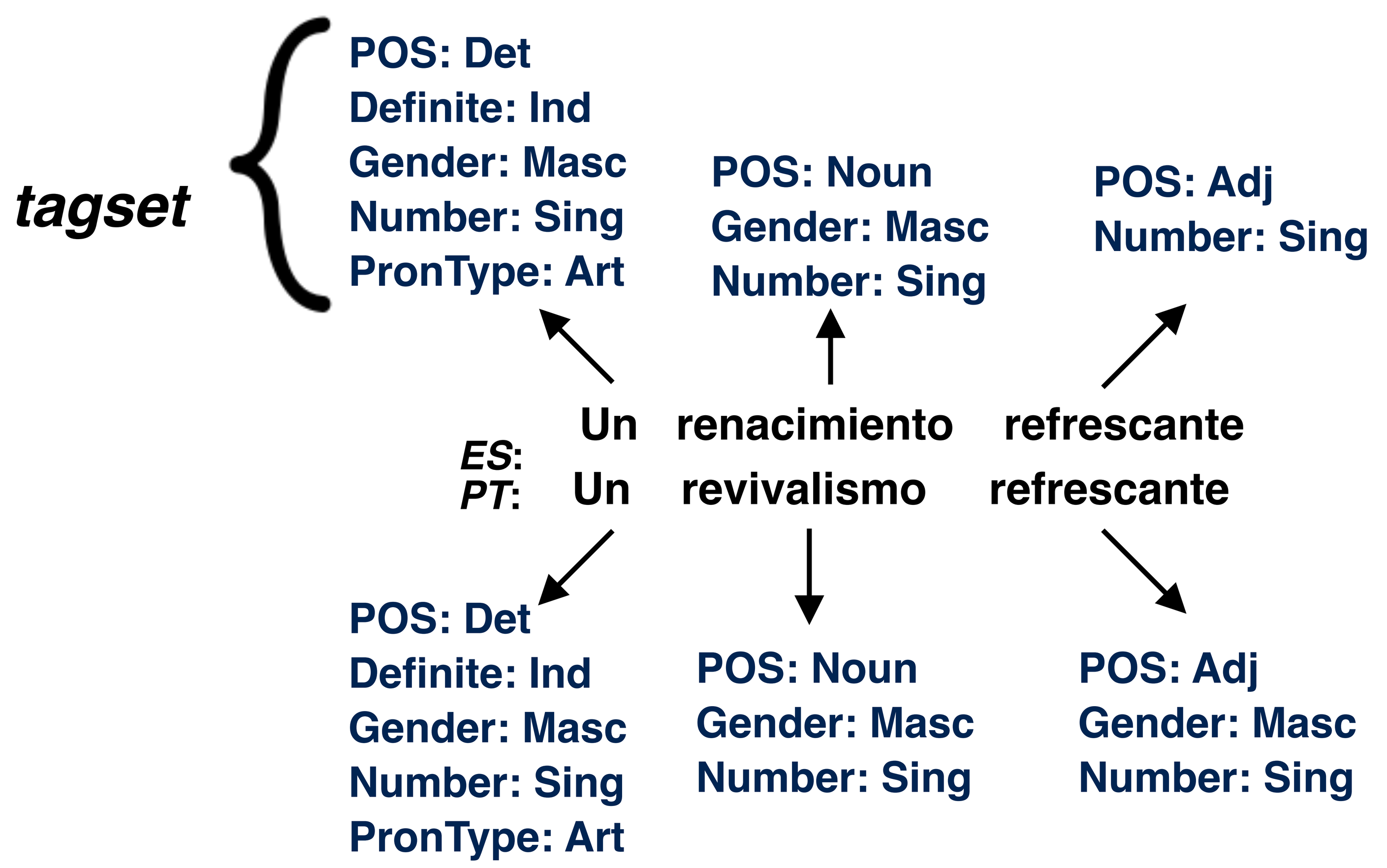
Language Technologies Institute

NEULAB

Code: github.com/chaitanyamalaviya/NeuralFactorGraph

Morphological Tagging

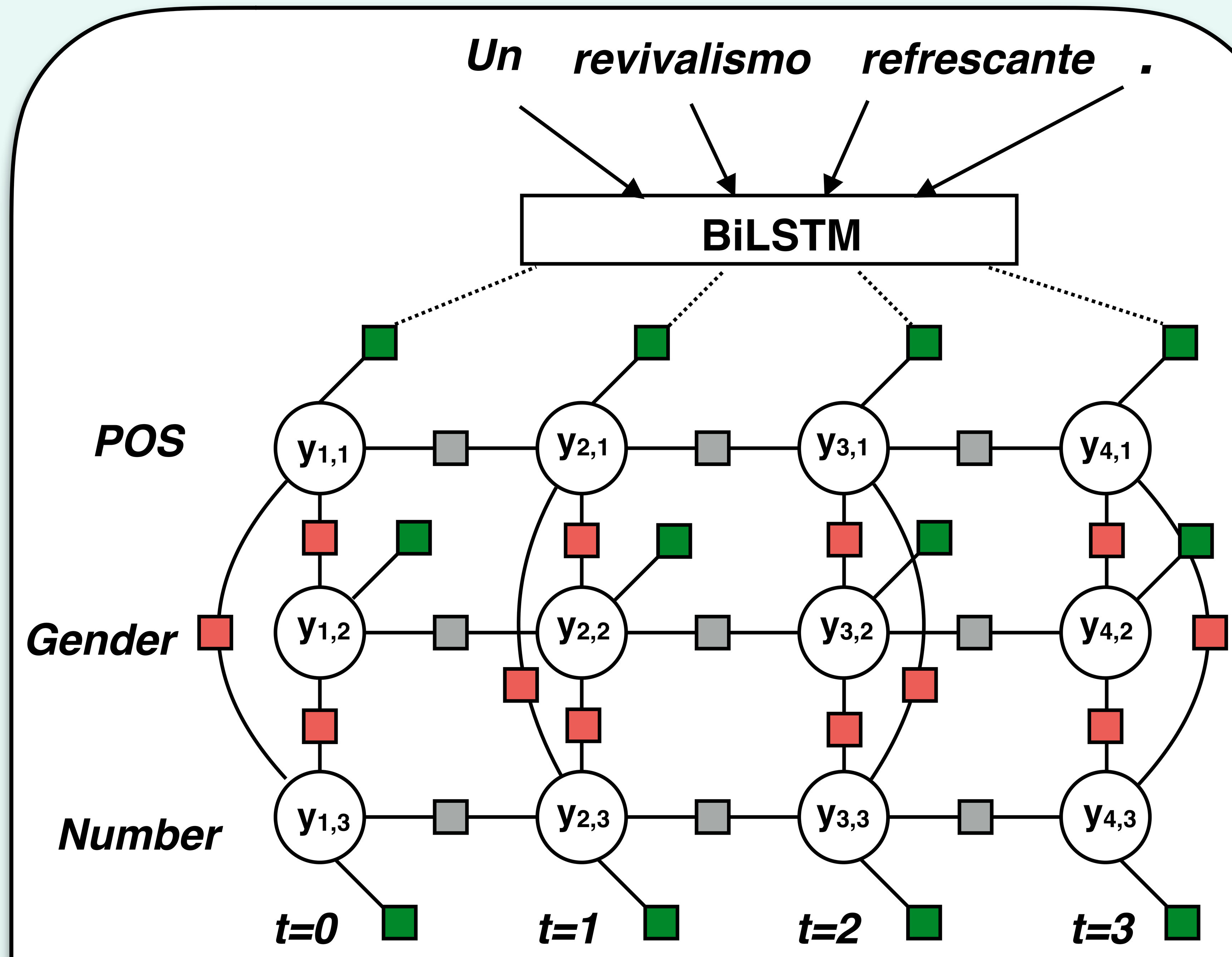
- **Morphological Analysis**: Prediction of syntactic properties of words. Useful in downstream tasks like parsing and translation.



- **Low-Resource Tagging:** Limited annotated data available for most world's languages. Cross-lingual training with HRL partner boosts performance.
- **Baseline Model (Cotterell and Heigold, 2017)**:
  - ✦ BiLSTM model with prediction over tagsets.
  - ✦ Word representations obtained from char BiLSTM.

Motivation

- **Problems** with Baseline Model:
  - ✦ Output space: All tagsets seen in training data.
  - ✦ Tagsets don't exactly overlap between HRL & LRL.
- **Proposed Model : Hybrid Graphical Model + NN** architecture with following advantages:
  - ✦ Predict tags individually; generates arbitrary sets.
  - ✦ Model inter-time tag dependencies.
    {Gender: Masc}$_t$ => {Gender: Masc}$_{t+1}$
  - ✦ Model pairwise dependencies between tags.
    {Verbform: Infinitive}$_t$ => {Tense: NULL}$_t$



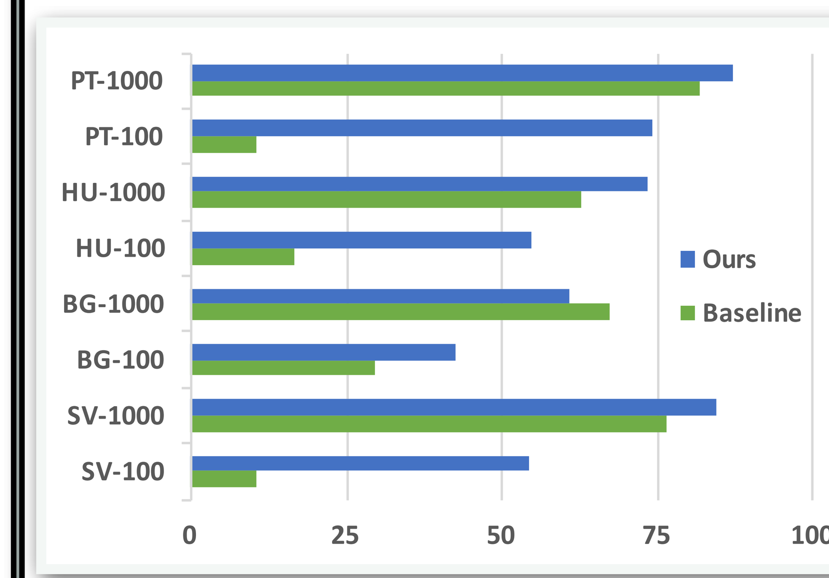
Model Architecture

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{t=1}^{T} \prod_{\alpha \in \mathcal{C}} \psi_\alpha(\mathbf{y}_\alpha, \mathbf{x}, t)$$

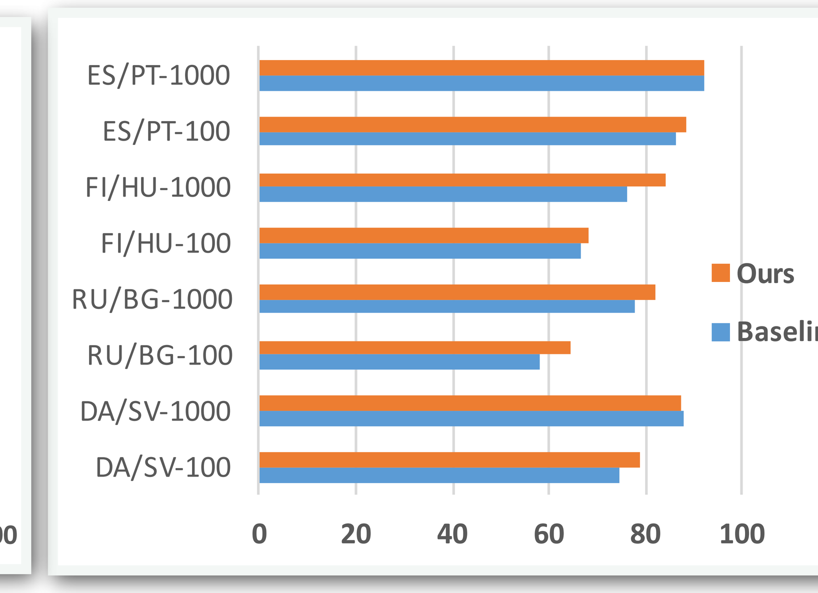
- **Architecture**:
  - ✦ Factorial CRF + unary potentials from BiLSTM.
- **Factor Types**:
  - ✦ **Transition:** Capture information about neighboring tags.
  - ✦ **Pairwise:** Model correlations between tags.
  - ✦ **Neural:** BiLSTM predicts label scores for all tags.
- **Inference on factorial CRF**:
  - ✦ Exact Inference is hard; we resort to Loopy Belief Propagation.
- **Decoding**:
  - ✦ Minimum Bayes Risk Decoding for Hamming Loss.
- **Language-specific weights**:
  - ✦ Learn language-specific weights for each language & generic weights for trends across both languages. For example, $\lambda_T = \lambda_{T,\,gen} + \lambda_{T,\,lang}$

Experiments & Results

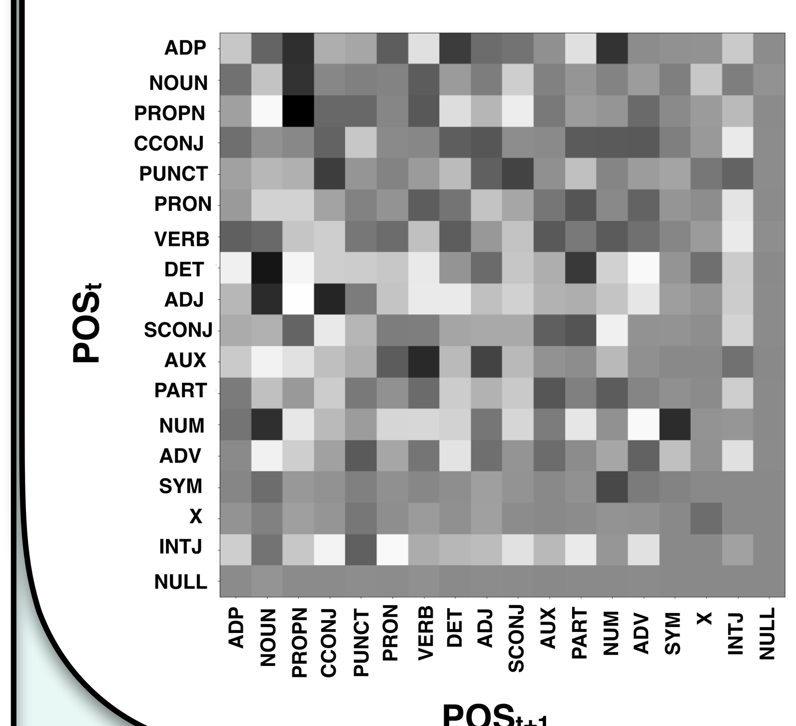
- Pick HRL /LRL partners from 4 language families.
- Limit target language dataset size to 100 or 1000 sentences.

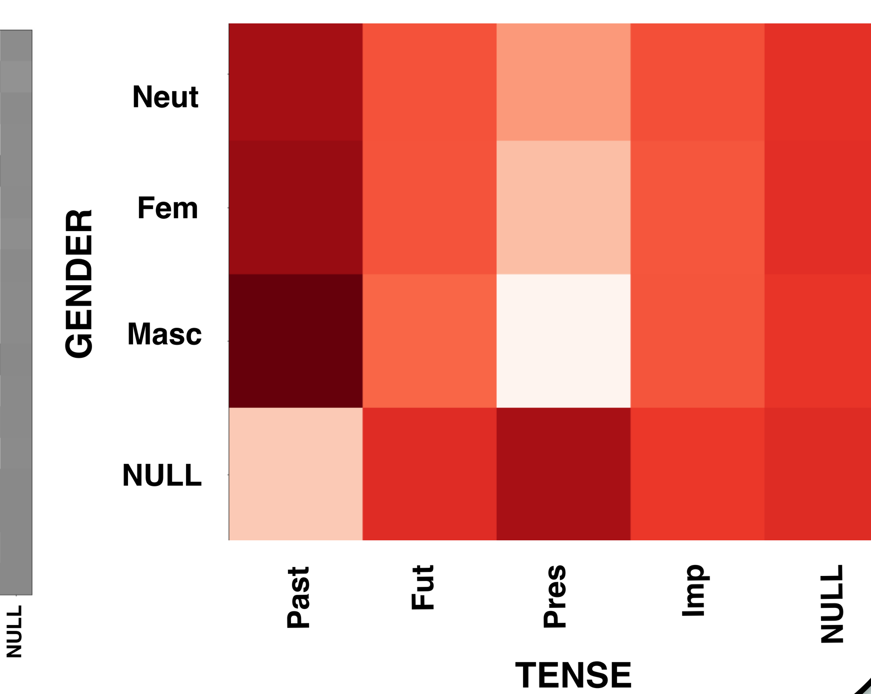


Monolingual F1 scores



Cross-lingual F1 scores





Conclusions

- **Novel framework for sequence tagging**: hybrid model of factorial CRFs with unary potentials from biLSTM.
- Utilize **expressiveness of NN representations** while graphical model approach ensures:
  1) **model interpretability**
  2) **exploiting variable dependencies**
- Empirically strong performances on morph tagging.
- Can be extended to multiple languages and other sequence labeling tasks.