

# Multilingual Sentiment Valence

Final Project for the Course

„Deep Learning for Natural Language Processing“ (Dr. Ben Roth)  
CIS, LMU Munich

Simon Preissner  
January 31 2018

# The Task at Hand

Sentiment Valence = Force of Sentiment → Regression Task

Multilingual Word Embeddings (WEs)

- Will multilingual WEs outperform monolingual WEs?
- Can multilingual WEs help with monolingual data?

# Resources

Data: Twitter data from SemEval2018, subtask 3

Embeddings: [http://www.cs.cmu.edu/~afm/projects/multilingual\\_embeddings.html](http://www.cs.cmu.edu/~afm/projects/multilingual_embeddings.html)

Languages: English (EN) and Spanish (ES)

EN + ES = MULTI

# Model

Keras, Sequential API

Bidirectional LSTM

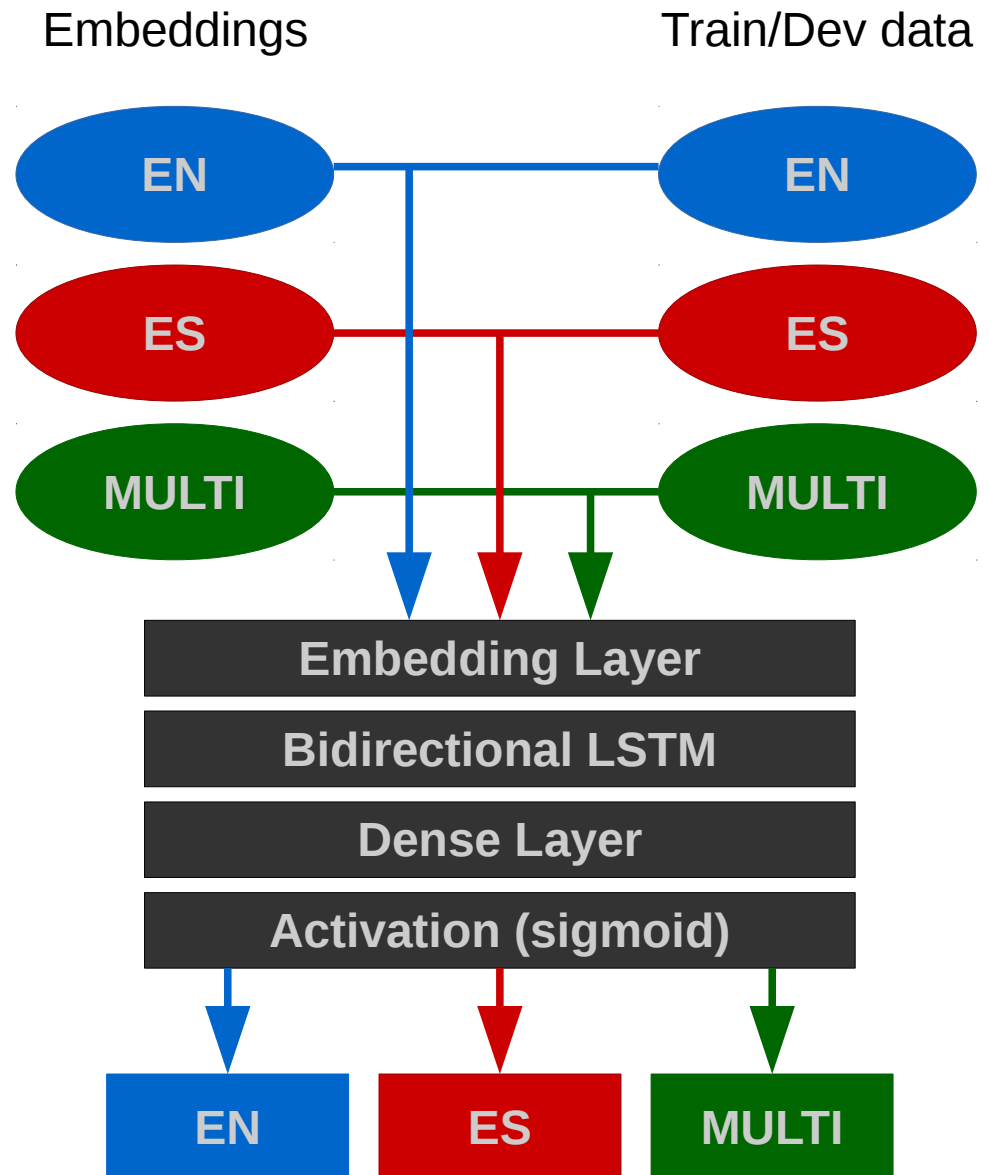
Optimization with  
hyperopt.py

3 models:

EN

ES

MULTI



# Results

- Training on all available data (80/20 split)  
EN: 950      ES: 1250      MULTI: 2200
- Tests on optimized models
- MSE for evaluation (in the table:  $\sqrt{\text{mse}}$ )

Test Data	EN	ES	MULTI
English	<b>0.2319</b>	0.3085	<b>0.1754</b>
Spanish	0.2382	<b>0.1850</b>	<b>0.1755</b>
both	0.2347	0.2746	<b>0.1755</b>

MULTI > EN on English data

MULTI > ES on Spanish data

↑  
EN underperforms

↑  
Solid performance in both languages

**multiple languages → more training data → better training?**

# What I Learned

- Most programming work: preprocessing (70%)
- Most processing time needed: optimization (75%)
- Most stressful time: 3½h of uploads on Sunday at 8pm
- Most relaxed time: during optimization (→ 2h coffee breaks, yay!)