¿Qué?: A neural network's assessment of the collocation between null complementizers and Spanish verbs

Isaac Ang & Olivia Bottomley **Instructors: Mohammad Sharafat, Haripriya Mehta**

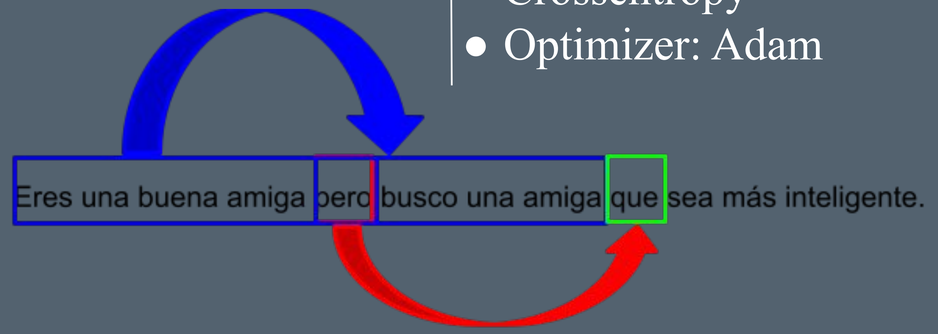
University at Buffalo
The State University of New York

Motivation

- 1350: null complementizer occurs 2% of the time
- Common matrix verbs: null complementizer occurs 90% of the time
  - Interested in seeing if this occurs in Spanish
- Computing power on corpus
- Analyze language patterns with our model
- Predict whether *que* follows a verb

Preprocessing

- Dataset: 1.5 million sentences
- Combined files into Dataframe
- Checked whether *que* exists after a verb in each sentence
- Tokenized the sentences
- 50% of Spanish corpus went to training



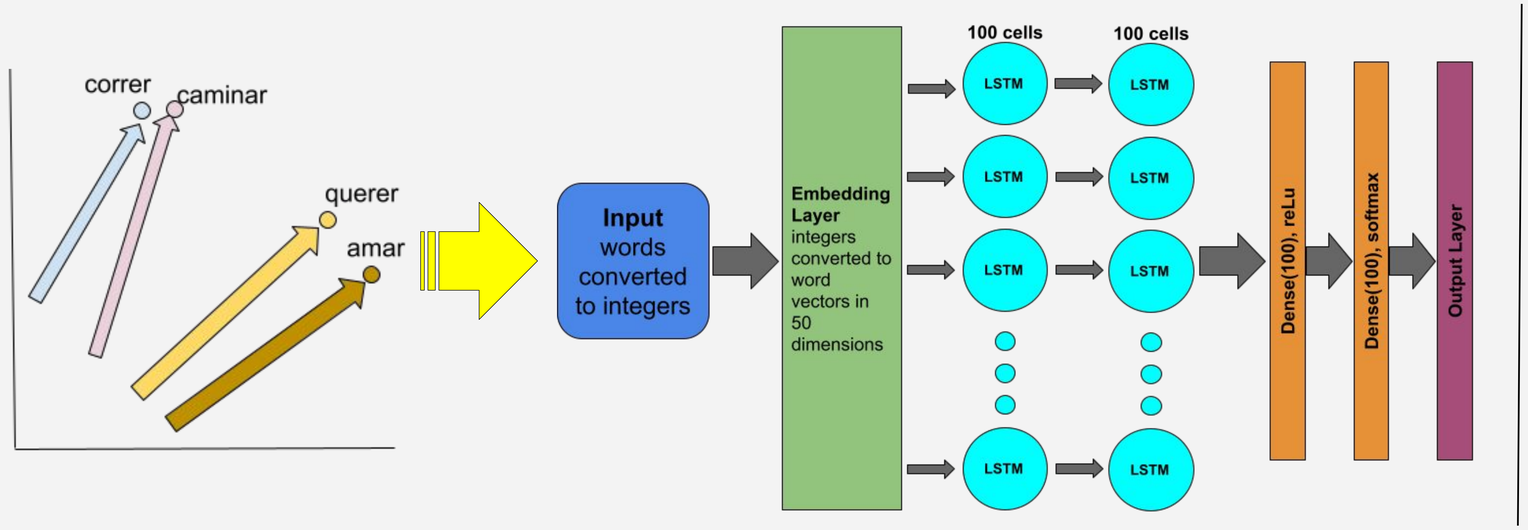
Model

- Trained model on 5-word sequences (4-word feature and 1-word label)
- LSTM predicts word based on a group of word embeddings
- Loss: Sparse Categorical Crossentropy
- Optimizer: Adam

Results

- 1 epoch
- 33 million sequences
- Validated on 1000 sentences

Validation accuracy: 74%
Sensitivity: 47.5%
Specificity: 26.75%
→ Model correctly predicts 'que' more than it correctly predicts that there isn't 'que'



Confusion Matrix

	Pos	Neg
True	190	107
False	89	14

Verb Type	*Que*-Frequency(%)
Epistemic	~21.71
Volitional	~17.90
Stative	~22.74

References

Adrián Rodríguez Riccelli. Espero estén todos: The distribution of the null subordinating complementizer in two varieties of spanish. In Language Variation and Contact-Induced Change, pages 299–333. John Benjamins, 2018.
Jiyoung Yoon. The grammaticalization of the spanish complement-taking verb without a complementizer.Journal of Social Sciences, 11(3):338, 2015.
Sali Tagliamonte and Jennifer Smith. No momentary fancy! the zero 'complementizer' in english dialects.English Language & Linguistics, 9(2):289–309, 2005