Biel Bellavista, Laura Boltà, Sonia Espinilla, and Rafael Servent

Group 10

Negation and Uncertainty Detection using Classical, Machine Learning, and Deep Learning Techniques

Course 2024/2025
Fundamentals of Natural Language
Degree in Artifical Intelligence
Universitat Autònoma de Barcelona

OBJECTIVE

To correctly detect negation and uncertainty cues and their scopes
in clinical texts written in Catalan and Spanish

THE DATA

The full clinical text

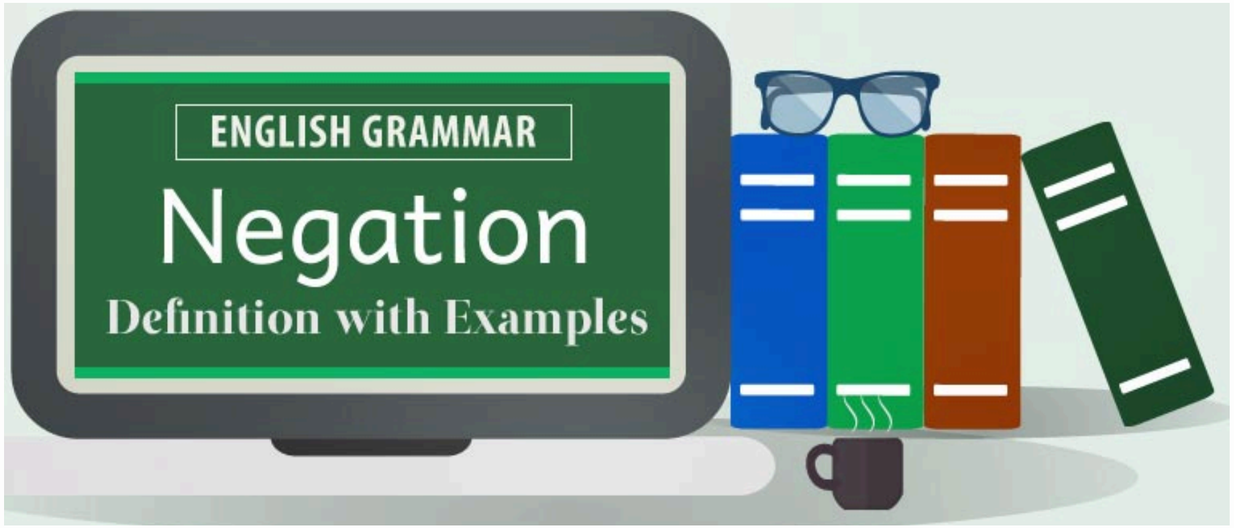
- Catalan and Spanish

Labeled cue words (e.g., no, descarta) as NEG or UNC

- Class imbalance

Scope spans as NSCO (negation scope) or USCO (uncertainty scope)

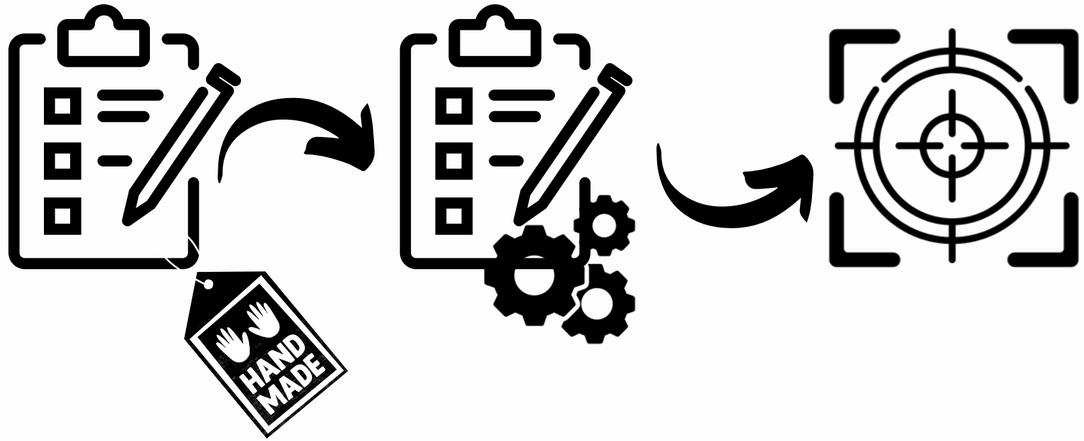
RULES BASED

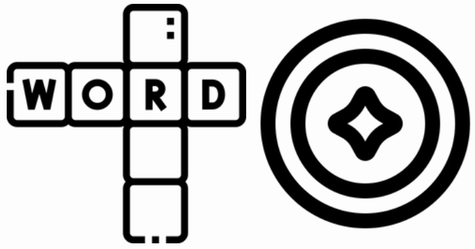


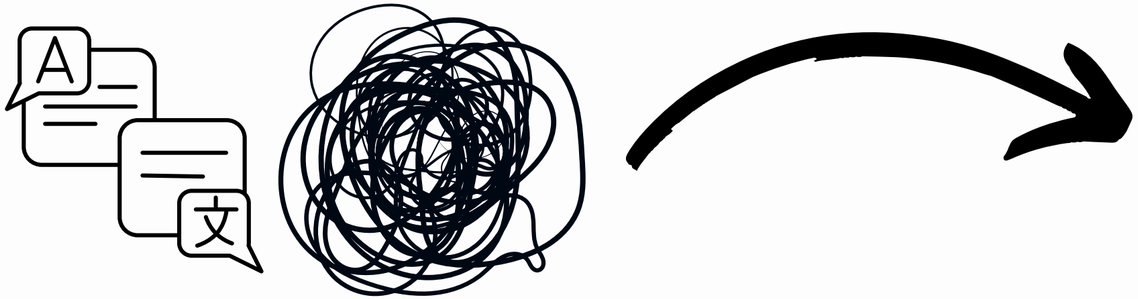
UMLS

RULES BASED

What we tried...



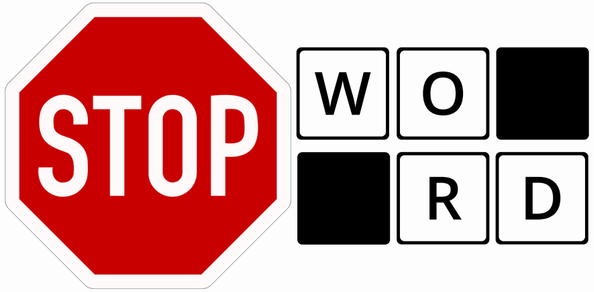




RULES BASED

What worked!

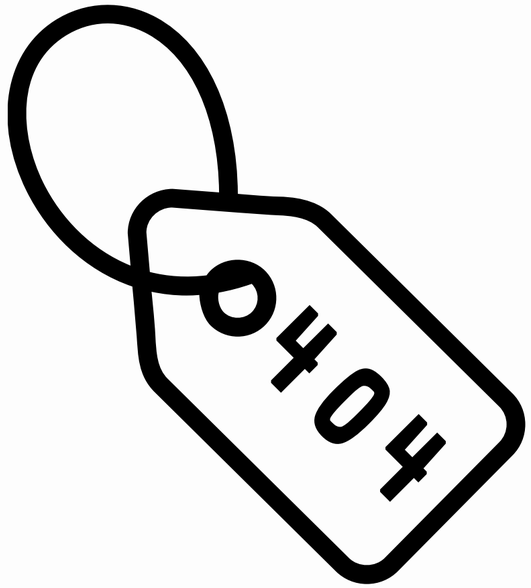
se, le, li



CAT ESP

RULES BASED RESULTS

RULES BASED RESULTS



RULES BASED

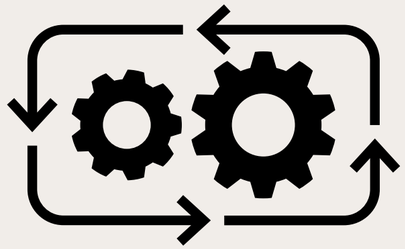
Results

- Average accuracy: 77.6%
- Range of performance: 42% to 95% depending on the text
- Instead of requiring an exact match, we allow partial credit.
- Strong on clear cues like “no”, “sin”
- Weak on uncertain expressions (e.g., “podría”, “es posible”)

Example

If the correct scope is:
“*tomar alcohol por las mañanas, tardes o especialmente noches*”
...but the system only detects up to “*tardes*”, it still gets partial score.

MACHINE LEARNING MODEL

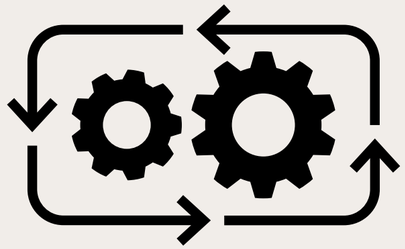


1 fastText

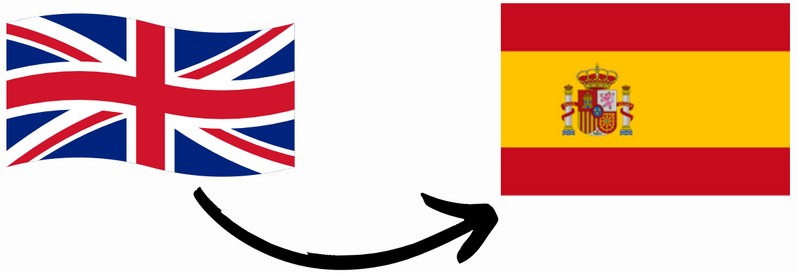
(15%)

I-NEG recall: 0.51 → 0.60 (+0.09)
I-UNC recall: 0.64 → 0.66 (+0.02)
Macro-F_1 (non-O labels): 0.80 → 0.81 (+0.01)

MACHINE LEARNING MODEL

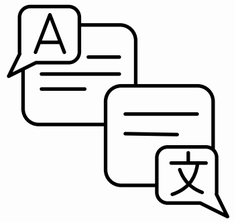


1 fastText



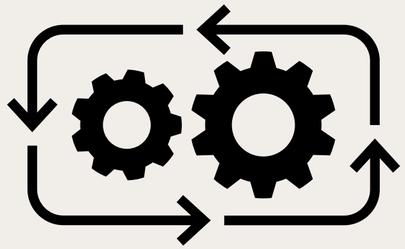
(15%)

2 K-Means

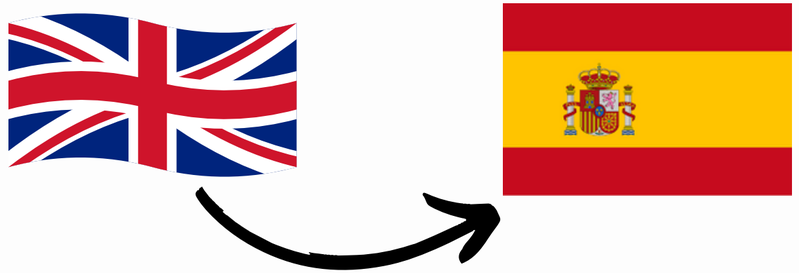


I-NEG recall: 0.60 → 0.62 (+0.02)
I-UNC recall: 0.66 → 0.69 (+0.03)
Macro-F_1: 0.81 → 0.84 (+0.03)

MACHINE LEARNING MODEL

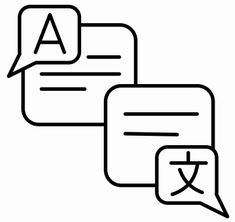


1. fastText



(15%)

2. K-Means



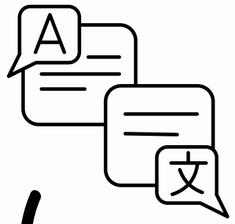
3. Feature Extraction

MACHINE LEARNING MODEL

1. fastText

(15%)

2. K-Means



3. Feature Extraction

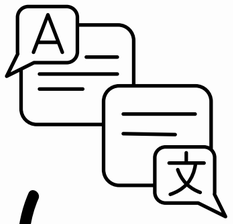
Character-Level and Lexical Shape Features

MACHINE LEARNING MODEL

1 fastText

(15%)

2 K-Means



3 Feature Extraction

Character-Level and Lexical Shape Features

Morpho-Syntactic Features (spaCy)

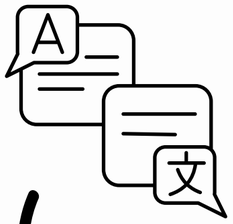
I-NEG recall: 0.62 → 0.66 (+0.04)
I-UNC recall: 0.69 → 0.73 (+0.04)
Macro-F_1: 0.84 → 0.86 (+0.02)

MACHINE LEARNING MODEL

1 fastText

(15%)

2 K-Means



3 Feature Extraction

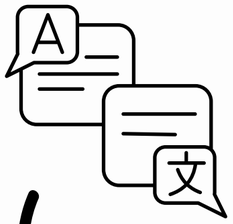
- Character-Level and Lexical Shape Features
- Morpho-Syntactic Features (spaCy)
- Context-Window Features (±2 tokens)

MACHINE LEARNING MODEL

1 fastText

(15%)

2 K-Means



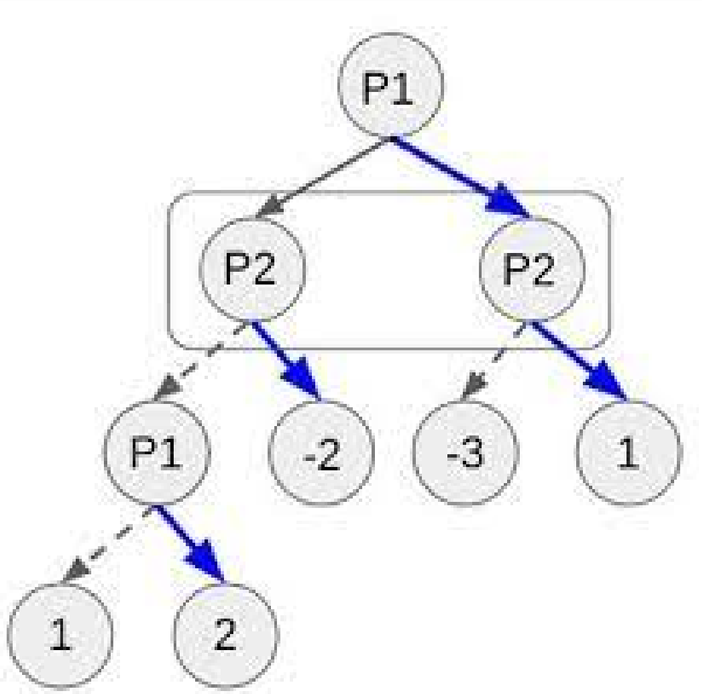
3 Feature Extraction

- **Character-Level and Lexical Shape Features**
- **Morpho-Syntactic Features (spaCy)**
- **Context-Window Features (±2 tokens)**
- **Cluster-Based Semantic Classes**

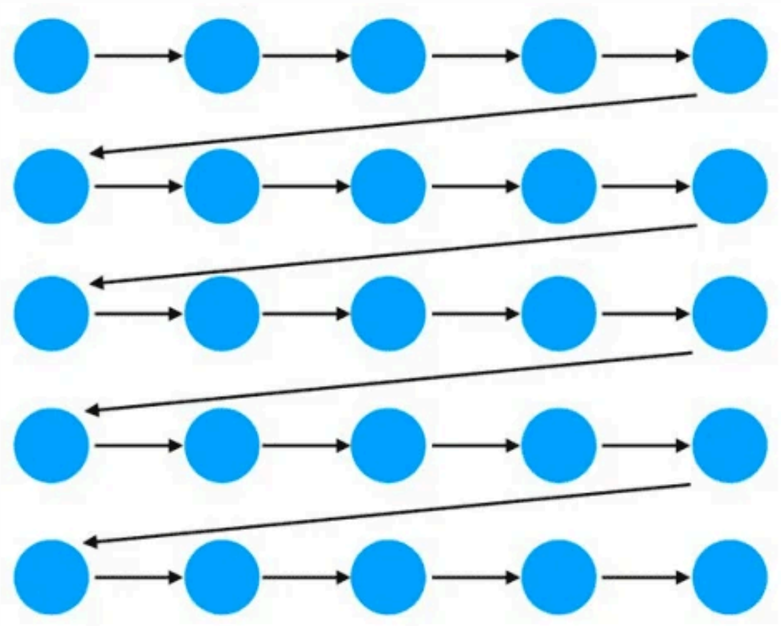
I-NEG recall: 0.66 → 0.68 (+0.02)
I-UNC recall: 0.73 → 0.76 (+0.03)
Macro-F_1: 0.86 → 0.88 (+0.02)

MACHINE LEARNING MODEL

2-Stage CFR



Hyperparameters



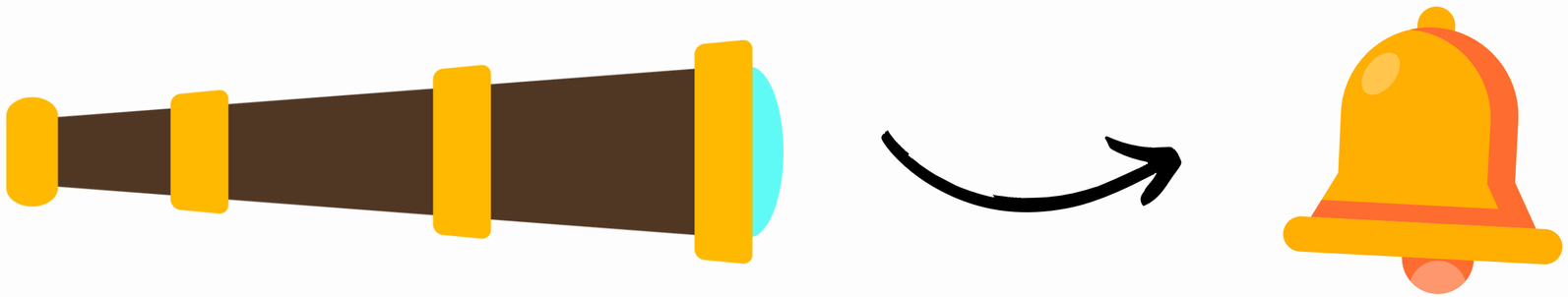
I-CUE recall: 0.68 (steady at strong level)
I-SCOPE recall: 0.76 → 0.77 (+0.01)
Macro-F_1: 0.88 → 0.90 (+0.02)

MACHINE LEARNING MODEL

Post-processing heuristics & merging

1. Lexicon membership
2. CRF boundary predictions
3. Nearest-cue propagation

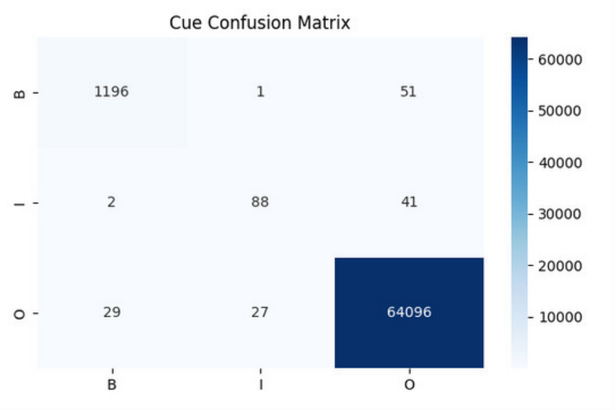
I-NEG recall (end-to-end): 0.68 → 0.77 (+0.09)
I-UNC recall (end-to-end): 0.76 → 0.80 (+0.04)
End-to-end 5-class F_1: ~0.90 → ~0.91 (+0.01)



MACHINE LEARNING MODEL

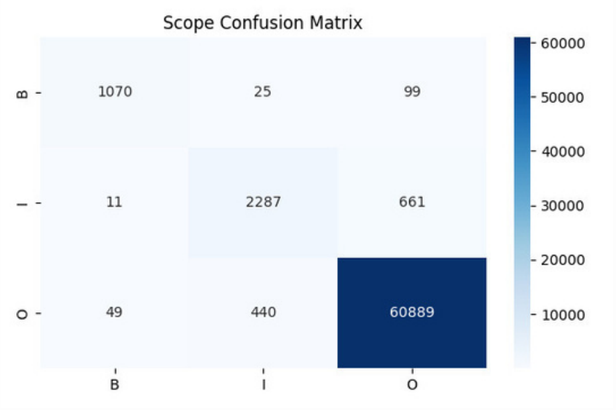
Results

- High accuracy overall (98%)
- Lower performance on continuation tags
- Really good classification of non-cue tokens
- As expected, the model predicts O tokens with very high accuracy



Label	Precision	Recall	F1-score	Support
B-CUE	0.97	0.96	0.97	1248
I-CUE	0.76	0.67	0.71	131
O	1.00	1.00	1.00	64152
Accuracy			1.00	65531
Macro average	0.91	0.88	0.89	65531
Weighted average	1.00	1.00	1.00	65531

Cue



Label	Precision	Recall	F1-score	Support
B-SCOPE	0.95	0.90	0.92	1194
I-SCOPE	0.83	0.77	0.80	2959
O	0.99	0.99	0.99	61378
Accuracy			0.98	65531
Macro average	0.92	0.89	0.90	65531
Weighted average	0.98	0.98	0.98	65531

Scope

DEEP LEARNING MODEL

Word Embeddings → [batch, seq_len, 300]

Using Fasttext 300 dimensions

CharCNNEncoder → [batch, seq_len, 30]

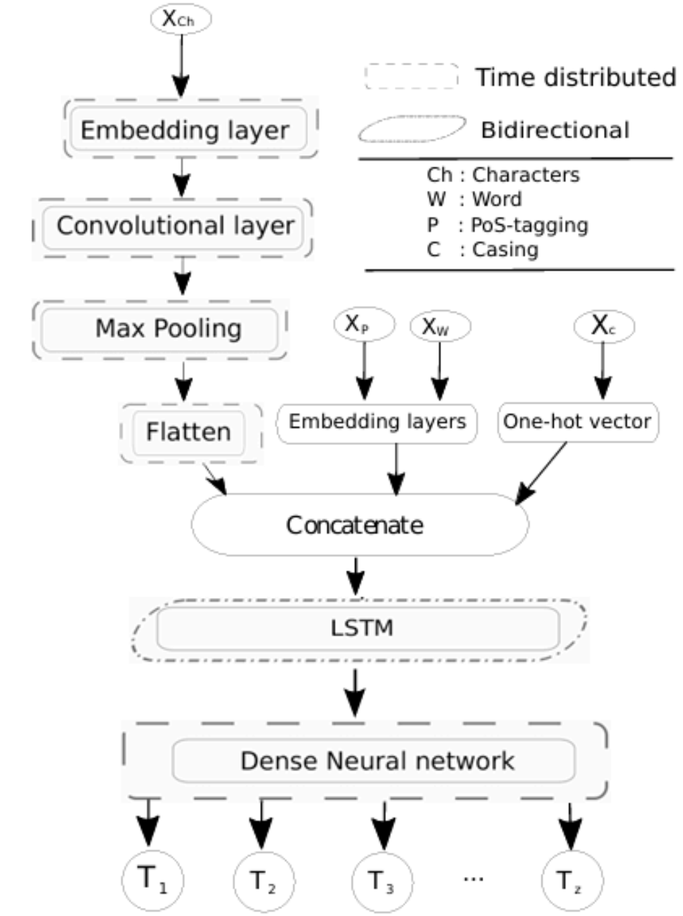
Captures morphological patterns like prefixes or suffixes.

PoS tagging → [batch, seq, 50]

We use the spacy to know the index of each tagging (det, adj, verb, ...) and we get the Pos Tagging of each token.

Casing in One Hot Vector → [batch, len, 8]

Returns an 8-dim one-hot vector for casing feature of a token.



Fabregat, H., Duque, A., Martinez-Romo, J., & Araujo, L. (2021). Extending a Deep Learning Approach for Negation Cues Detection in Spanish. In Proceedings of IberLEF 2021.

CHAR CNN

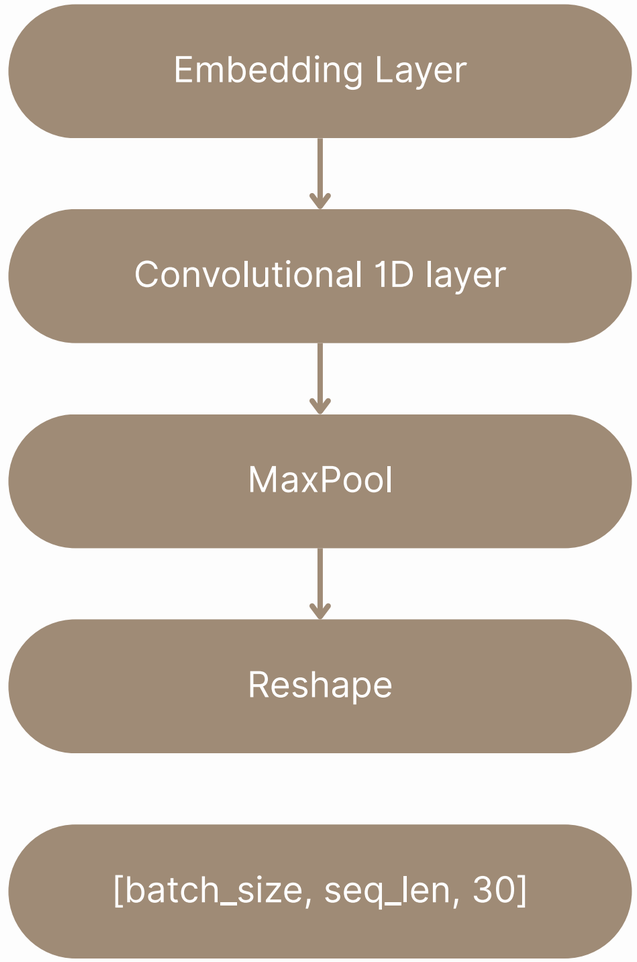
CharCNNEncoder → [batch, seq_len, 30]

The CharCNNEncoder turns each word's character sequence into a fixed-size vector using:

1. Character Embedding:
Each character in a word is mapped to a dense vector (50-dim) using an embedding layer.

2. 1D Convolution:
A 1D CNN slides over the character embeddings to capture local patterns (like prefixes/suffixes). Filter size = 30

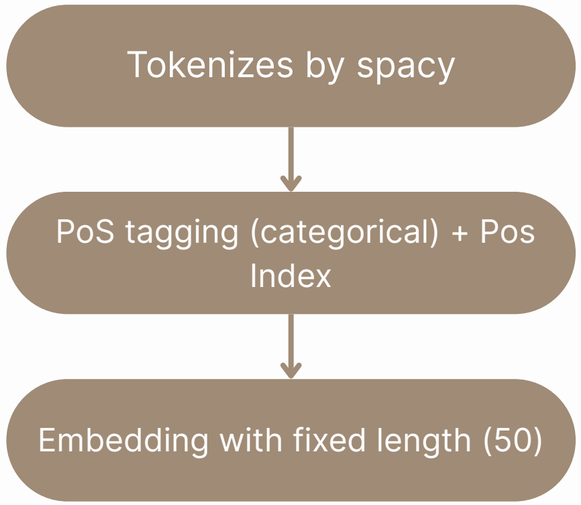
3. Max Pooling:
It takes the most important feature from the convolution output, producing a single 30-dim vector per word.



POS AND CASING

PoS Tagging

Spacy first tokenizes the text and assigns each token a Part-of-Speech (PoS) tag based on its categorical label. These PoS tags are mapped to numeric indices and passed to embedding layer that converts each PoS tag index into a fixed-length dense vector.



Casing with one hot vector

Categories:

0: all lowercase
1: all uppercase
2: initial uppercase (title case)
3: contains digit(s)
4: contains hyphen(s)
5: other (mixed case)
6: all punctuation
7: numeric (all digits)

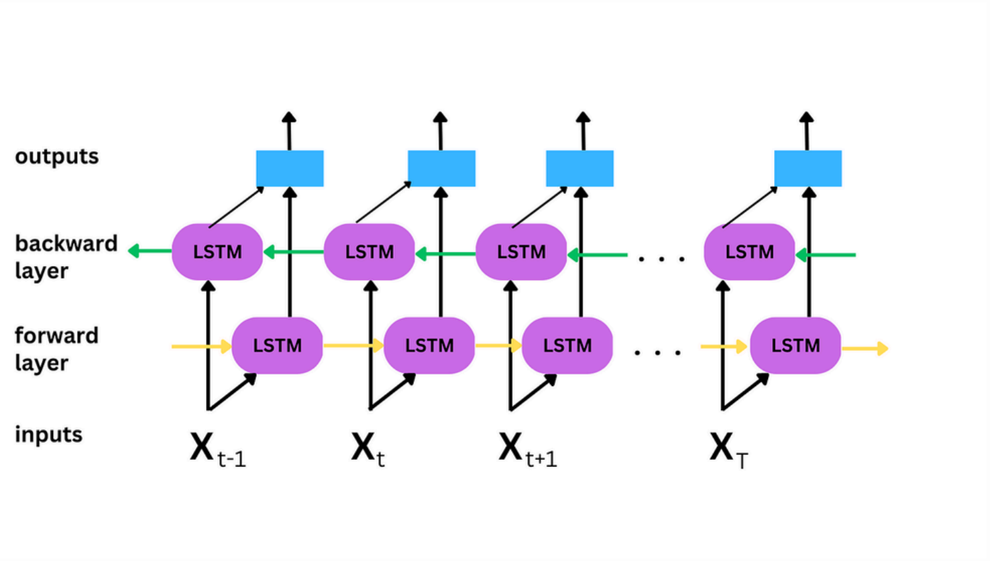
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0] ⟶ “24”

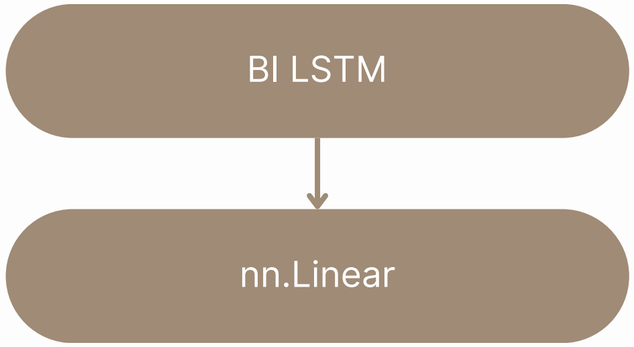
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0] ⟶ ","

BI LSTM

[word vector | char CNN | PoS embedding | casing]
(300) (30) (50) (8)
→ total: 388-dim input per token

```
nn.LSTM(
    lstm_input_dim, hidden_dim = 150,
    num_layers=lstm_layers, → 388
    bidirectional=True,
  )
```





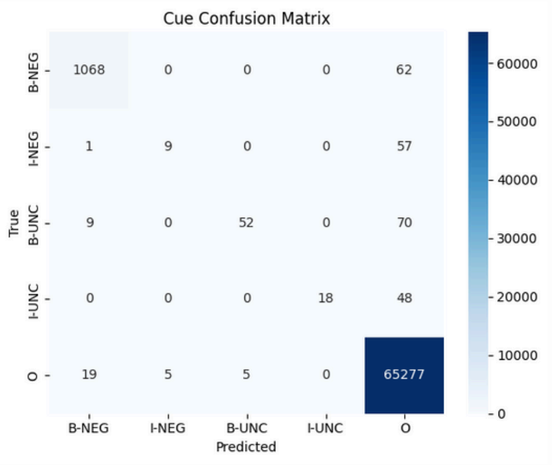
['O', 'B-NEG', 'I-NEG', 'B-UNC', 'I-UNC', 'B-NSCO', 'I-NSCO', 'B-USCO', 'I-USCO']

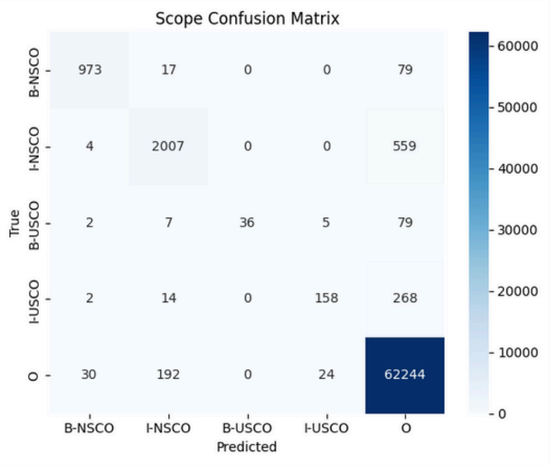
[1.2, -0.3, 0.1, 0.5, 2.1, -1.0, 0.7, 0.0, -0.5]

DEEP LEARNING MODEL

Results

- High accuracy overall (98%)
- Strong detection of cue/scope beginnings
- Imbalanced data challenges
- The model still generalizes well thanks to the BiLSTM, FastText embeddings, character-level CNN, and PoS features.





Label	Precision	Recall	F1-score	Support
B-NEG	0.9736	0.9451	0.9591	1130
I-NEG	0.6429	0.1343	0.2222	67
B-UNC	0.9123	0.3969	0.5532	131
I-UNC	1.0000	0.2727	0.4286	66
O	0.9964	0.9996	0.9980	65306
Accuracy			**0.9959**	66700
Macro avg	0.9050	0.5497	0.6322	66700
Weighted avg	0.9955	0.9959	0.9951	66700

Cue

Label	Precision	Recall	F1-score	Support
B-NSCO	0.9624	0.9102	0.9356	1069
I-NSCO	0.8972	0.7809	0.8350	2570
B-USCO	1.0000	0.2791	0.4364	129
I-USCO	0.8449	0.3575	0.5024	442
O	0.9844	0.9961	0.9902	62490
Accuracy			**0.9808**	66700
Macro avg	0.9378	0.6647	0.7399	66700
Weighted avg	0.9798	0.9808	0.9790	66700

Scope

CONCLUSIONS

Comparison

Rule-Based System

- Transparency and simplicity
- Less accurate
- No training data

Machine Learning Model

- Less Interpretable
- Better performance
- Benefits from training

Deep Learning Model

- More complex, black box
- Better performance with long, irregular scopes
- More computational resources needed

CONCLUSIONS

Future Work

Grammar to Improve Rules

- Right now it uses fixed word lists
- We could make these rules smarter using PoS tagging

Combine the Models

- Rule-based approach for easy cases
- CRF for syntactically structured patterns
- Deep learning for the most ambiguous / context-dependent cases

Bigger Dataset

- For the Machine learning and Deep learning approaches, quantity and quality of the dataset matters

CONCLUSIONS

- Could grow into a more complete tool
- Could build systems that are both accurate and adaptable
- Specially important for under-represented languages like Catalan and Spanish, where tools for medical language are still limited but very much needed

Thank You

Questions are Welcome

UAB

Course 2024/2025
Fundamentals of Natural Language
Degree in Artifical Intelligence
Universitat Autònoma de Barcelona

Biel Bellavista, Laura Boltà, Sonia Espinilla, and Rafael Servent

Group 10