Input? Output? What does it do? How is evaluated? How does it relate to other work in the field) followed by leading a critical discussion on key positive and negative aspects, availability of code and datasets

Positives

Negatives

Evaluation

What baseline?

Is lexical best?

Comparison to previous work or state of art?

Choosing most likely class of token

What about lemma/stem baseline

Statistical significance

Why tagging intead of parsing?

Joint inference with dependency parsing

Feature selection by exhaustive search

On what data?

Tuning of SVM

On what data?

Lots of parameter tuning

More examples of how corpus is tagged

Breakdown of tags

Unbalanced data

Reproducibility

Requires parsing, not necessarily scalable

Impact

Clarity

Originality / Innovativeness

Soundness / Correctness

References / Meaningful Comparison

Impact of Ideas or Results

Impact of Resources

Clarity?

Reproducibility?

Correctness?

Methodology?

Evaluation?

Impact?

Start with examples

Belief

Desire

intention

Mention orthogonality

Focus on belief

Annotation

CB

Doesn’t account for degrees of belief

NCB

NA

O

Lexical

Extracted without parsing

syntactic