

Giving Dutch Some Style

Data and Models for Text Formality Transfer in Dutch

Marieke Weultjes
mariekeweultjes@hotmail.com

Huiyuan Lai
h.lai@rug.nl

Malvina Nissim
m.nissim@rug.nl



**university of
groningen**



GroNLP

Style in Image

Original Photo



The Starry Night by
Vincent van Gogh, 1889



The Shipwreck of the
Minotaur by J.M.W.
Turner, 1805



Der Schrei by Edvard
Munch, 1893



Formality Style in Language

“Style is an intuitive notion involving the manner in which something is said”

- McDonald and Pustejovsky. EACL 1985

“The most important dimension of variation between styles or registers”

- Heylighen and Dewaele, 1999

Formality Style in Language

“Style is an intuitive notion involving the manner in which something is said”

- McDonald and Pustejovsky. EACL 1985

“The most important dimension of variation between styles or registers”

- Heylighen and Dewaele, 1999

Formality Transfer

Informal

this screen is top, just the right size...

Formal

I like this screen, it is just the right size.

Formality Transfer

Informal

this screen is top, just the right size...

Formal

I like this screen, it is just the right size.

TASK: given some input, generate a well-formed text in the desired formality style while preserving the original content

Formality Transfer

Informal

this screen is top, just the right size...

Formal

I like this screen, it is just the right size.

TASK: given some input, generate a well-formed text in the desired formality style while preserving the original content

Fluency

Style Strength

Content
Preservation

Formality Transfer

GYAFC [Rao & Tetreault; 2018]

A large-scale formality dataset containing parallel informal-formal pairs for both training and evaluation in English

Formality Transfer

GYAFC [Rao & Tetreault; 2018]

A large-scale formality dataset containing parallel informal-formal pairs for both training and evaluation in English

XFORMAL [Briakou et al; 2021]

A multilingual formality dataset containing manually crafted test data for Italian, French, and Portuguese

The training pairs are automatically translated from GYAF

Formality Transfer in Dutch

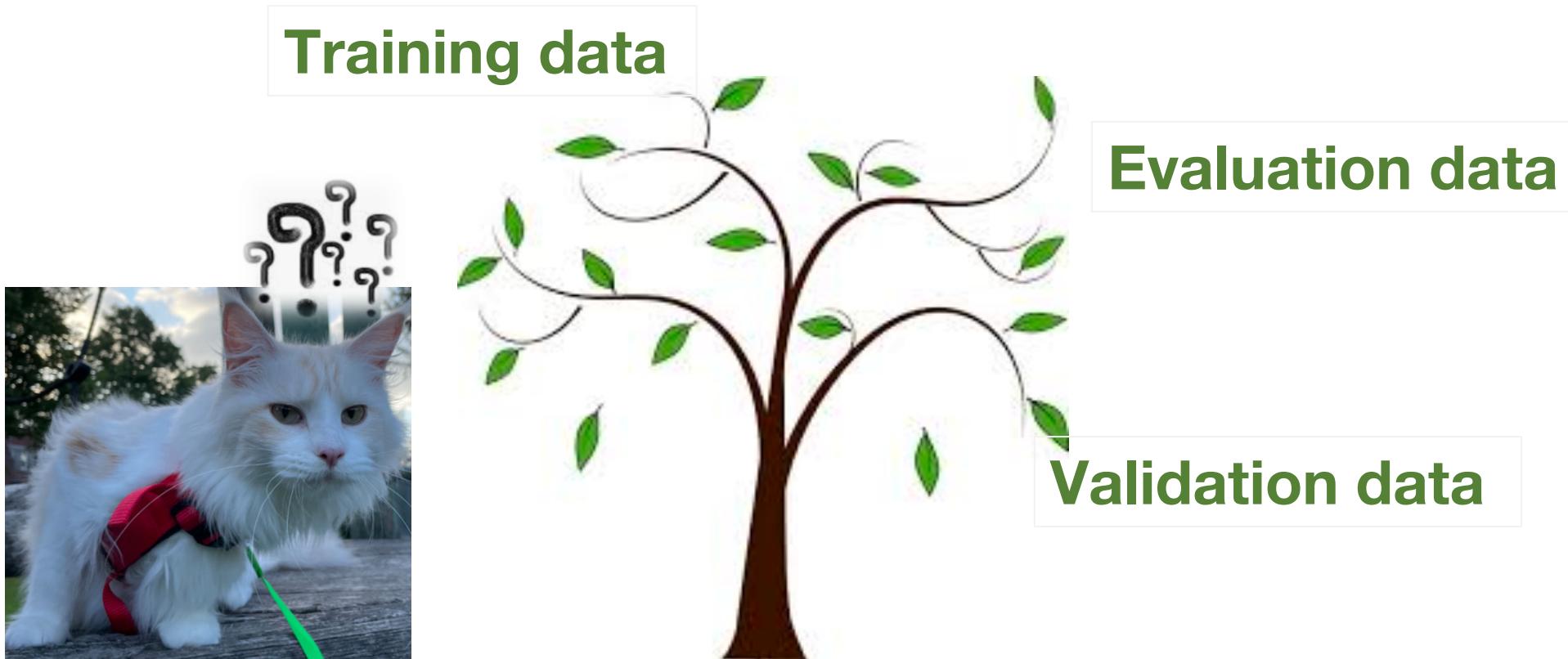
GYAFC [Rao & Tetreault; 2018] – English

XFORMAL [Briakou et al; 2021] – Italian, French, and Portuguese



How about formality transfer in Dutch?

Formality Transfer in Dutch



Formality Transfer in Dutch: Data

Training Data

Machine translated pairs from English GYAFC

Formality Transfer in Dutch: Data

Training Data

Machine translated pairs from English GYAFC

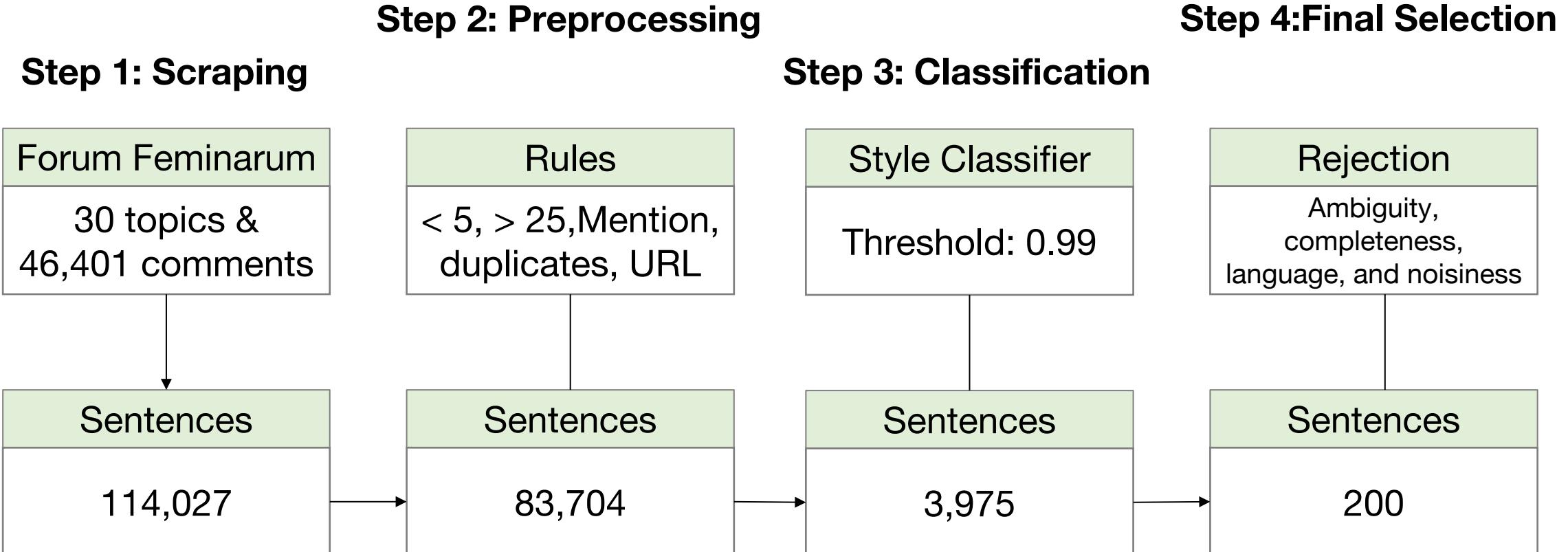
Validation Data

Informal texts paired with formal rewrites

Evaluation Data

Informal texts paired with formal rewrites

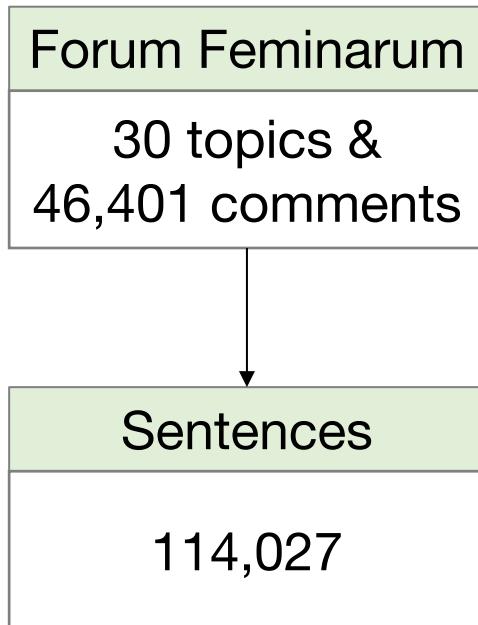
Formality Transfer in Dutch: Data



An overview of the informal data collection process

Formality Transfer in Dutch: Data

Step 1: Scraping



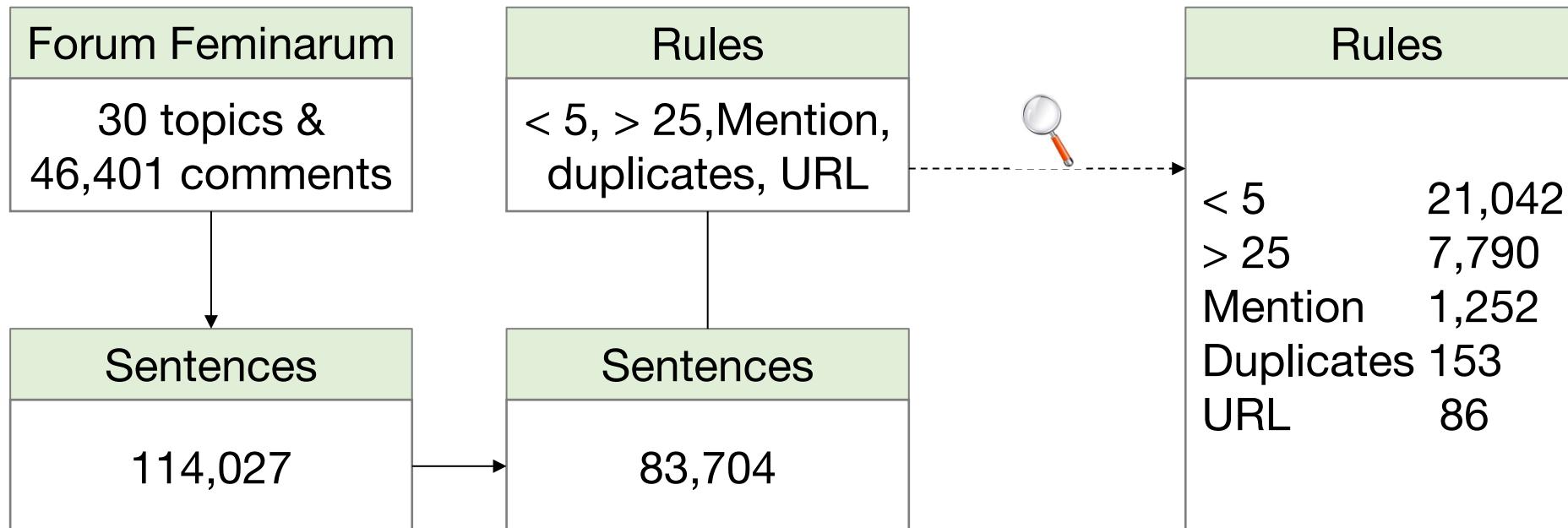
Forum Feminarum [1]: a medium for online discussions among Dutch users.

Entertainment domain: discussions about movies, music, and everything seen on television or online sources such as YouTube

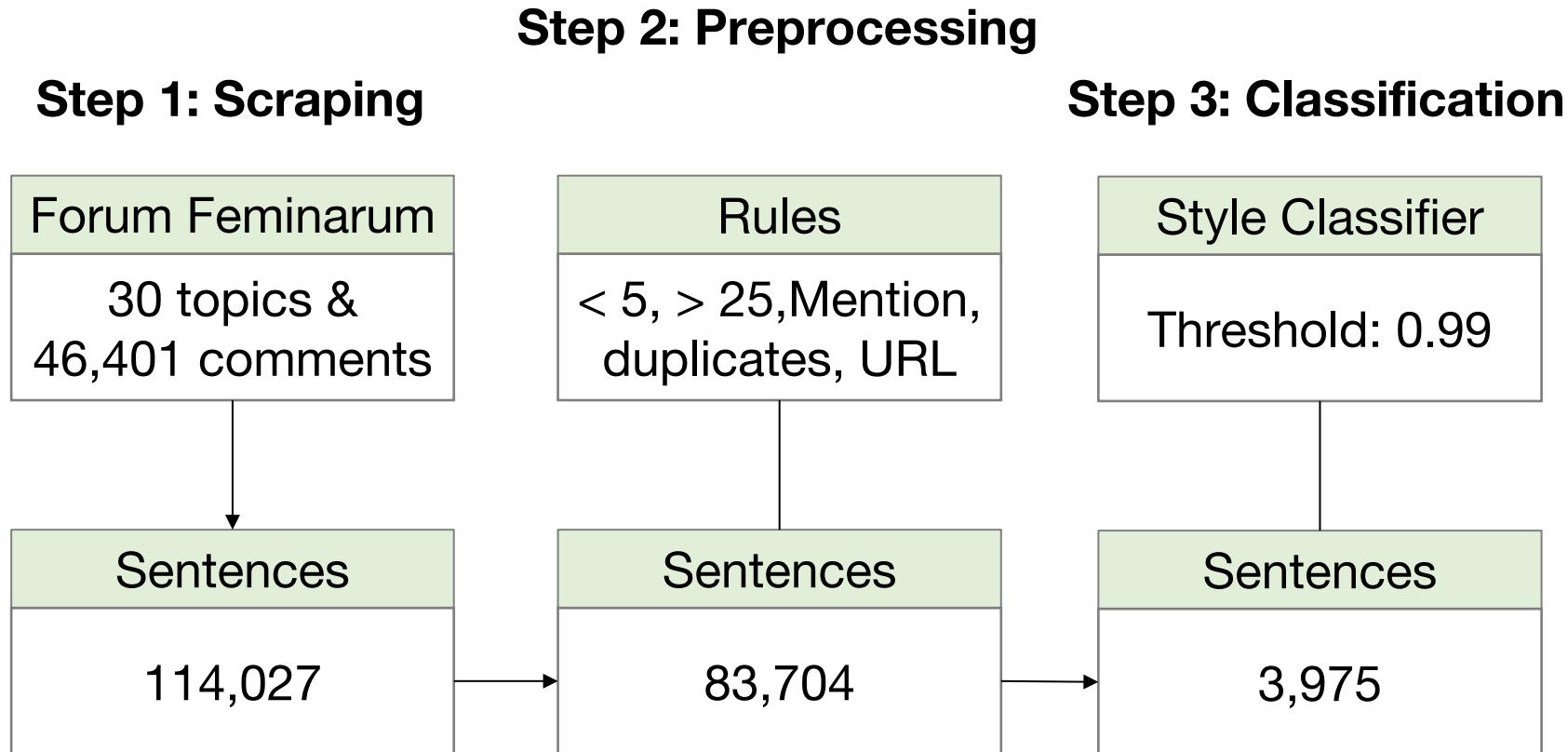
Formality Transfer in Dutch: Data

Step 2: Preprocessing

Step 1: Scraping

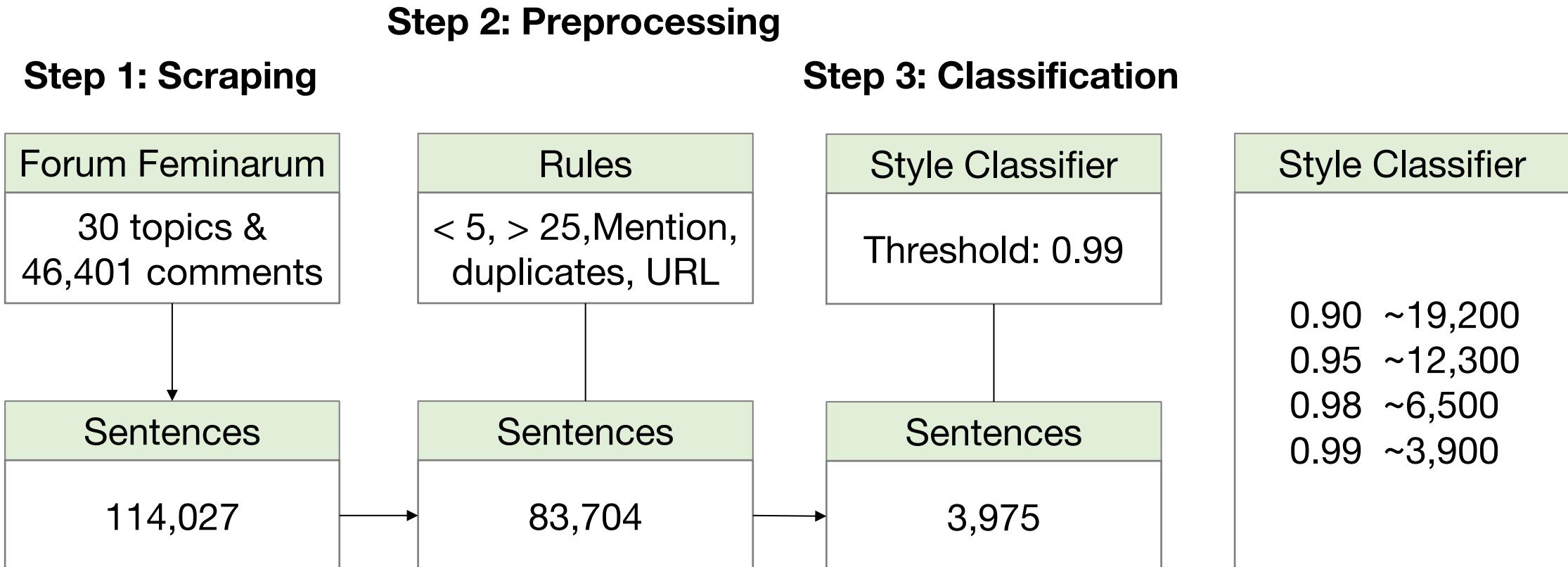


Formality Transfer in Dutch: Data

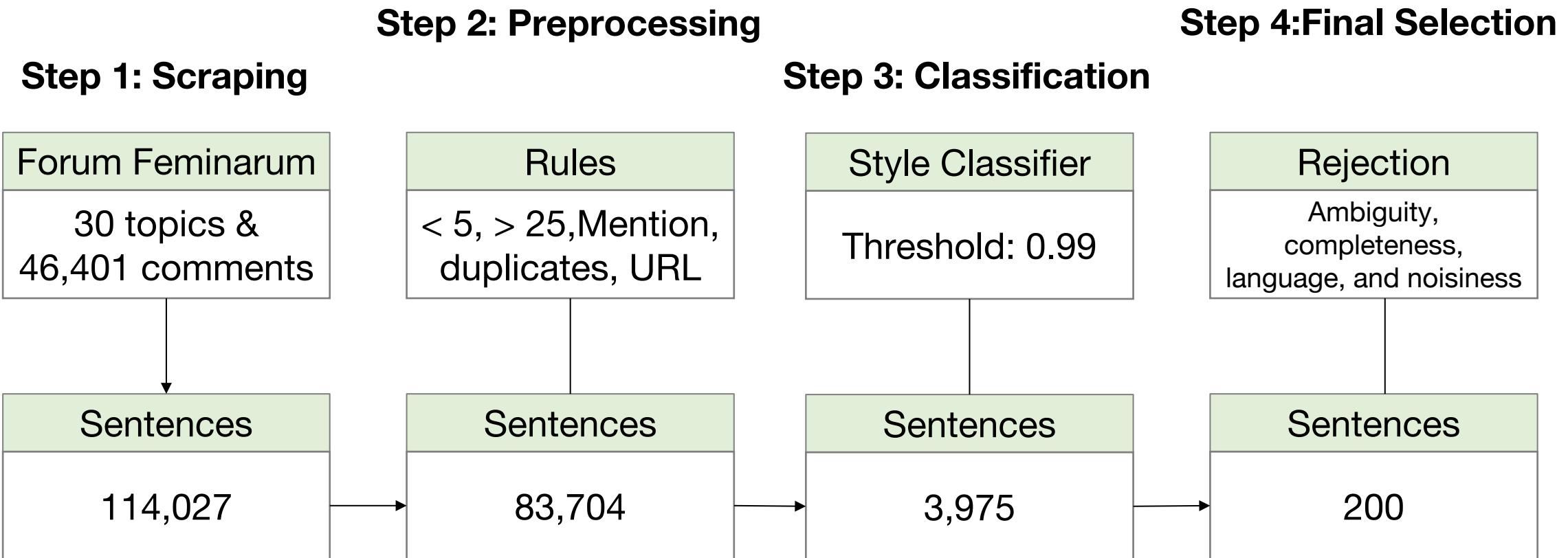


A mBERT based classifier (informal vs formal) trained on the GYAFc dataset to learn the formality classification.

Formality Transfer in Dutch: Data



Formality Transfer in Dutch: Data



Formality Transfer in Dutch: Data

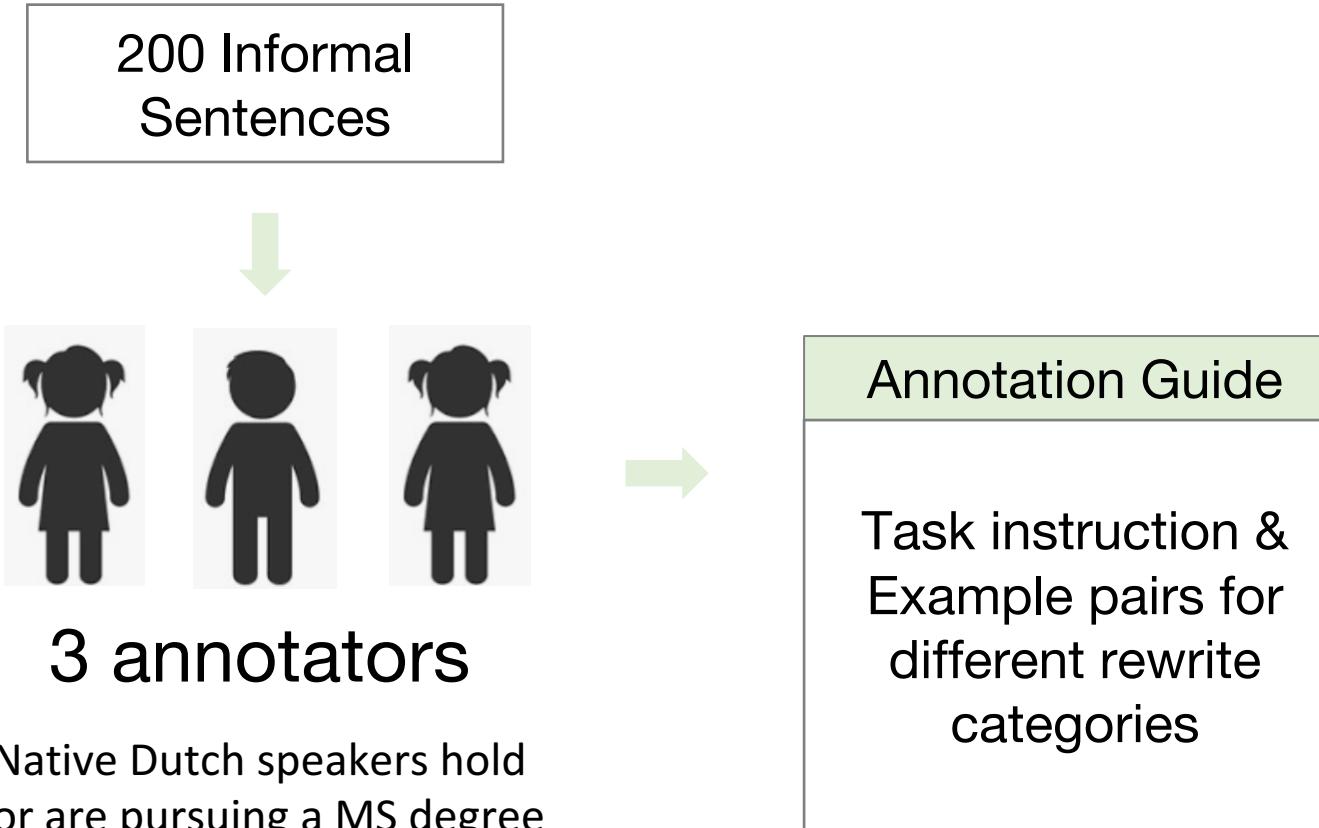
200 Informal
Sentences



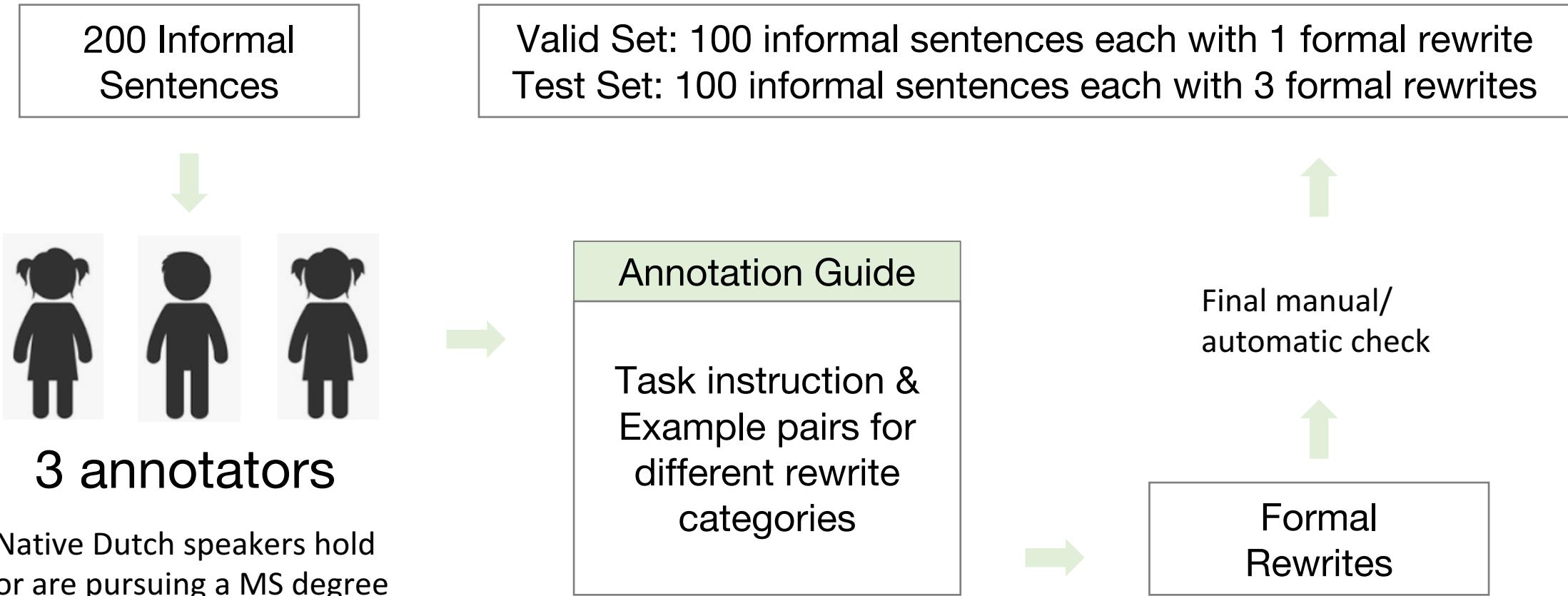
3 annotators

Native Dutch speakers hold
or are pursuing a MS degree

Formality Transfer in Dutch: Data



Formality Transfer in Dutch: Data



Formality Transfer in Dutch: Data

| Category | N | % | Informal Original | Formal Rewrite |
|----------------|----|-----|--|---|
| Paraphrase | 46 | 92% | Maar oh oh wat gek dat mensen daar wat van vinden | Het is toch niet gek dat mensen daar een mening over hebben. |
| Capitalization | 33 | 66% | wie worden er nou echt beter van? | Wie worden er nou echt beter van? |
| Punctuation | 39 | 78% | mgoed niet dat die kinderen verder een heel normaal en veilig leven leiden | Maar goed, die kinderen leiden verder geen heel normaal en veilig leven . |
| Delete Fillers | 23 | 46% | hahahaah hoe salty is deze krant | Hoe geïrriteerd is deze krant? |
| Completion | 8 | 16% | niet voorstellen dat deze look serieus was | Ik kan me niet voorstellen dat die outfit serieus was bedoeld. |
| Spelling | 2 | 4% | Doe dan Jan Kooiman aub!! | Doe dan Jan Kooijman alsjeblieft! |
| Contraction | 2 | 4% | mgoed niet dat die kinderen verder een heel normaal en veilig leven leiden | Maar goed , die kinderen leiden verder geen heel normaal en veilig leven. |
| Normalization | 12 | 24% | Doe dan Jan Kooiman aub!! | Doe dan Jan Kooijman alsjeblieft! |
| Lowercase | 2 | 4% | Maar hoe DOM ben je dat je denkt dat dit niet gaat uitkomen?? | Hij is vrij dom als hij denkt dat dit niet gaat uitkomen. |
| Split Sentence | 7 | 14% | dat ook ja, ik vind alles vreemd aan deze winactie | Dat is ook zo. Ik vind alles vreemd aan deze winactie. |
| Repetition | 2 | 4% | Ojaaa dat maakt wel verschil idd | Oja , dat maakt inderdaad een verschil. |
| English | 15 | 30% | hahahaah hoe salty is deze krant | Hoe geïrriteerd is deze krant? |
| Swearwords | 4 | 8% | och jezus wat verschrikkelijk dat laatste | Jemig , dat laatste deel is verschrikkelijk. |

Types of formality edits and their frequency over a sample of 50 instances

Formality Transfer in Dutch: Data

| | Formality | Source | BLEU | | |
|-------------|-----------|--------|-------------|-------------|-------------|
| | | | Reference 1 | Reference 2 | Reference 3 |
| Source | 0.5 | | 19.4 | 7.1 | 11.9 |
| Reference 1 | 78.6 | 79.2 | | 15.3 | 21.9 |
| Reference 2 | 89.7 | 71.9 | 77.3 | | 16.2 |
| Reference 3 | 85.9 | 75.1 | 80.5 | 78.0 | |
| | | | COMET | | |

Evaluation metrics to gain insight into the style and similarity of our dataset.

Formality Transfer in Dutch: Data

| | Formality | Source | BLEU | | |
|-------------|-----------|--------|-------------|-------------|-------------|
| Source | 0.5 | | Reference 1 | Reference 2 | Reference 3 |
| Reference 1 | 78.6 | 79.2 | | 15.3 | 21.9 |
| Reference 2 | 89.7 | 71.9 | 77.3 | | 16.2 |
| Reference 3 | 85.9 | 75.1 | 80.5 | 78.0 | |
| | | COMET | | | |

Evaluation metrics to gain insight into the style and similarity of our dataset.

Formality Transfer in Dutch: Data

| | Formality | Source | BLEU | | |
|-------------|-----------|--------|-------------|-------------|-------------|
| | | | Reference 1 | Reference 2 | Reference 3 |
| Source | 0.5 | | 19.4 | 7.1 | 11.9 |
| Reference 1 | 78.6 | 79.2 | | 15.3 | 21.9 |
| Reference 2 | 89.7 | 71.9 | 77.3 | | 16.2 |
| Reference 3 | 85.9 | 75.1 | 80.5 | 78.0 | |
| | | | COMET | | |

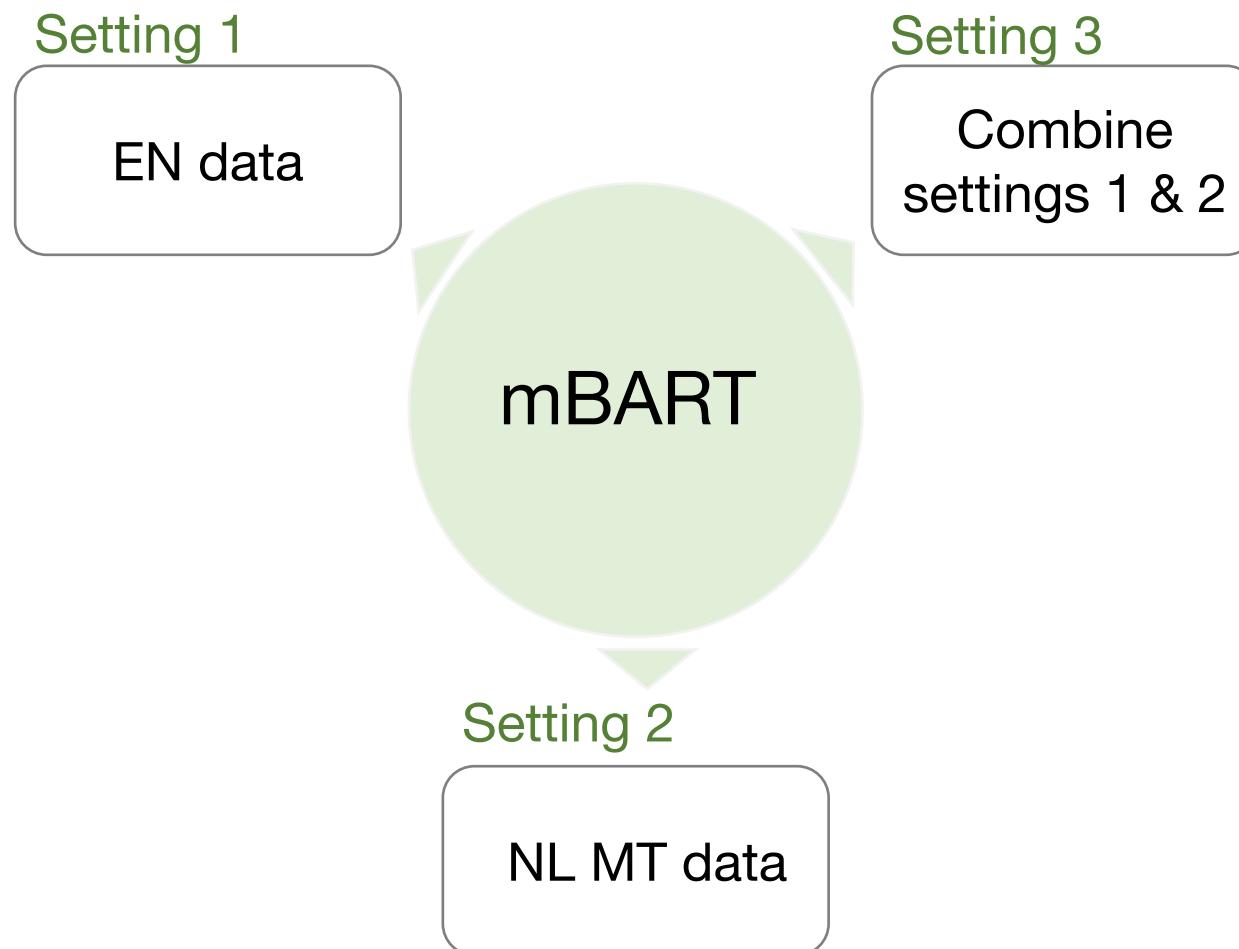
Evaluation metrics to gain insight into the style and similarity of our dataset.

Formality Transfer in Dutch: Data

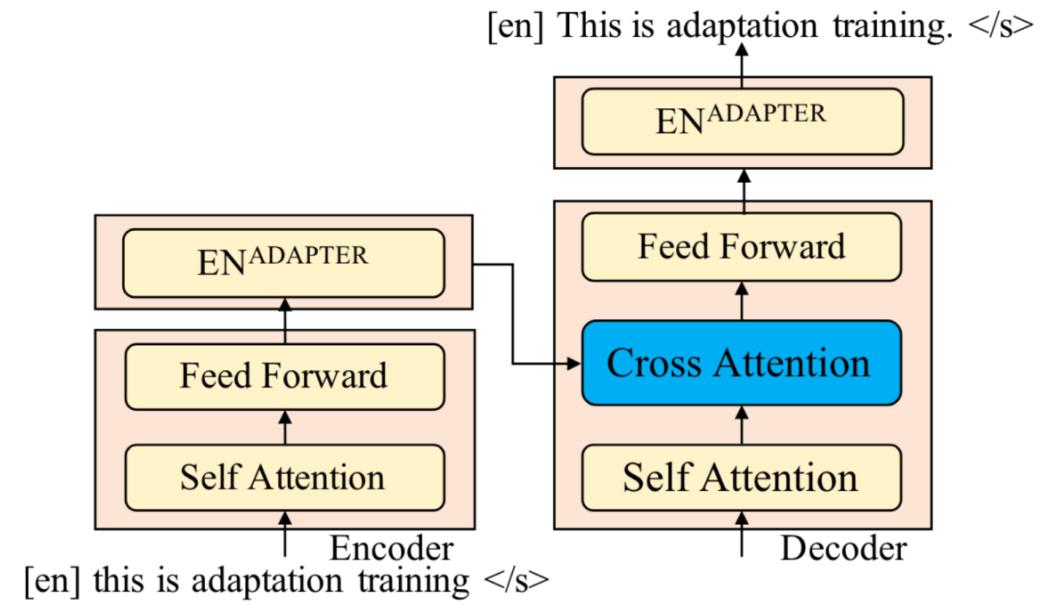
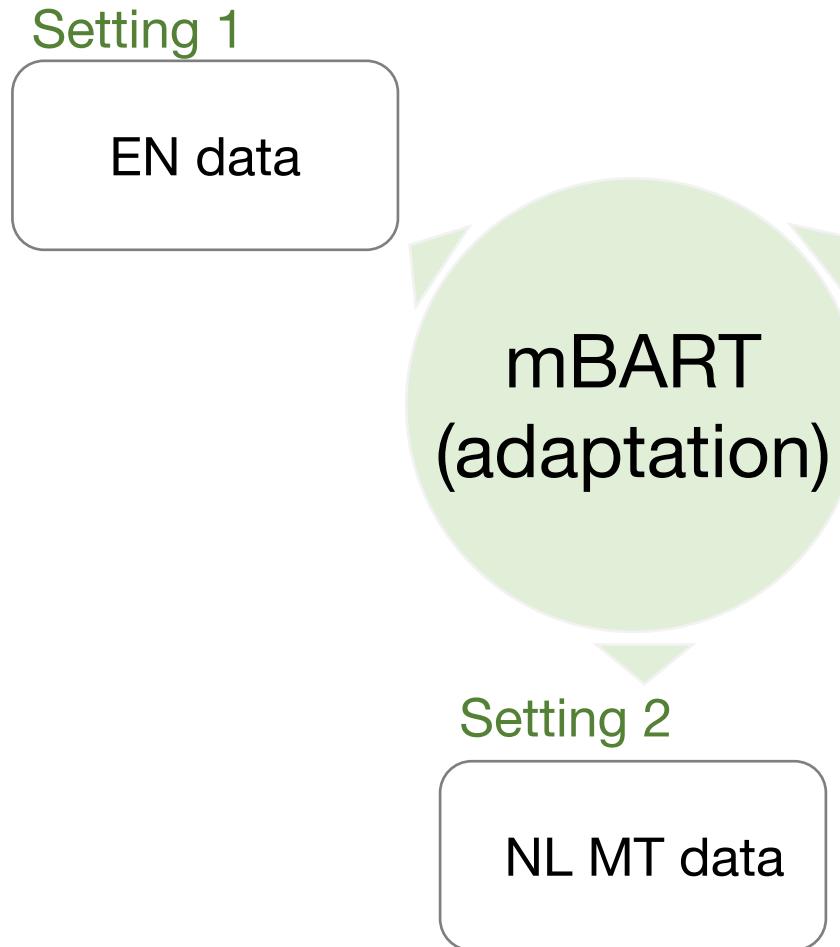
| | Formality | Source | BLEU | | |
|-------------|-----------|--------|-------------|-------------|-------------|
| | | | Reference 1 | Reference 2 | Reference 3 |
| Source | 0.5 | | 19.4 | 7.1 | 11.9 |
| Reference 1 | 78.6 | 79.2 | | 15.3 | 21.9 |
| Reference 2 | 89.7 | 71.9 | 77.3 | | 16.2 |
| Reference 3 | 85.9 | 75.1 | 80.5 | 78.0 | |
| | | | COMET | | |

Evaluation metrics to gain insight into the style and similarity of our dataset.

Formality Transfer in Dutch: Models



Formality Transfer in Dutch: Models



Adaptation training

Formality Transfer in Dutch: Results

| | | Formality | BLEU | COMET | HM |
|-----------------------------------|------------|-------------|-------------|-------------|-------------|
| Setting 1 (EN data) | Vanilla | 86.7 | 5.8 | 22.9 | 10.9 |
| | Adaptation | 37.6 | 30.8 | 26.9 | 33.9 |
| Setting 2 (NL MT data) | Vanilla | 71.6 | 33.8 | 27.2 | 45.9 |
| | Adaptation | 77.4 | 32.1 | 26.8 | 45.4 |
| Setting 3 (1 & 2) | Vanilla | 75.5 | 37.1 | 27.1 | 49.8 |
| | Adaptation | 67.2 | 36.3 | 27.1 | 47.1 |

Automatic evaluation results. HM is harmonic mean of formality accuracy and BLEU.

Formality Transfer in Dutch: Results

| | | Formality | BLEU | COMET | HM |
|-----------------------------------|------------|-------------|-------------|-------------|-------------|
| Setting 1 (EN data) | Vanilla | 86.7 | 5.8 | 22.9 | 10.9 |
| | Adaptation | 37.6 | 30.8 | 26.9 | 33.9 |
| Setting 2 (NL MT data) | Vanilla | 71.6 | 33.8 | 27.2 | 45.9 |
| | Adaptation | 77.4 | 32.1 | 26.8 | 45.4 |
| Setting 3 (1 & 2) | Vanilla | 75.5 | 37.1 | 27.1 | 49.8 |
| | Adaptation | 67.2 | 36.3 | 27.1 | 47.1 |

Automatic evaluation results. HM is harmonic mean of formality accuracy and BLEU.

Formality Transfer in Dutch: Results

| | | Formality | BLEU | COMET | HM |
|-----------------------------------|------------|-------------|-------------|-------------|-------------|
| Setting 1 (EN data) | Vanilla | 86.7 | 5.8 | 22.9 | 10.9 |
| | Adaptation | 37.6 | 30.8 | 26.9 | 33.9 |
| Setting 2 (NL MT data) | Vanilla | 71.6 | 33.8 | 27.2 | 45.9 |
| | Adaptation | 77.4 | 32.1 | 26.8 | 45.4 |
| Setting 3 (1 & 2) | Vanilla | 75.5 | 37.1 | 27.1 | 49.8 |
| | Adaptation | 67.2 | 36.3 | 27.1 | 47.1 |

Automatic evaluation results. HM is harmonic mean of formality accuracy and BLEU.

Formality Transfer in Dutch: Results

| | | Formality | BLEU | COMET | HM |
|-----------------------------------|------------|-------------|-------------|-------------|-------------|
| Setting 1 (EN data) | Vanilla | 86.7 | 5.8 | 22.9 | 10.9 |
| | Adaptation | 37.6 | 30.8 | 26.9 | 33.9 |
| Setting 2 (NL MT data) | Vanilla | 71.6 | 33.8 | 27.2 | 45.9 |
| | Adaptation | 77.4 | 32.1 | 26.8 | 45.4 |
| Setting 3 (1 & 2) | Vanilla | 75.5 | 37.1 | 27.1 | 49.8 |
| | Adaptation | 67.2 | 36.3 | 27.1 | 47.1 |

Automatic evaluation results. HM is harmonic mean of formality accuracy and BLEU.

Formality Transfer in Dutch: Results

| | | |
|--------------------|-------------------|--|
| Source | | Zo'n lekker nummer dat ook btw aanrader voor iedereen |
| Reference 1 | | Zo een leuk liedje trouwens, dat kan ik iedereen aanraden. |
| Reference 2 | | Dit is een goed nummer dat ik aan iedereen aan kan raden. |
| Reference 3 | | Dit is een zeer goed nummer. Ik raad het iedereen aan om er naar te luisteren. |
| Setting 1 | Vanilla | I love this song. It is a great song. It is also a great song. I recommend it to everyone. |
| | Adaptation | Zo'n lekker nummer is een aanrader voor iedereen. |
| Setting 2 | Vanilla | Het is een goed nummer, dat is ook een aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |
| Setting 3 | Vanilla | Het is een erg goed nummer. Het is een echte aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |

Example outputs generated by different models.

Formality Transfer in Dutch: Results

| | | |
|--------------------|-------------------|--|
| Source | | Zo'n lekker nummer dat ook btw aanrader voor iedereen |
| Reference 1 | | Zo een leuk liedje trouwens, dat kan ik iedereen aanraden. |
| Reference 2 | | Dit is een goed nummer dat ik aan iedereen aan kan raden. |
| Reference 3 | | Dit is een zeer goed nummer. Ik raad het iedereen aan om er naar te luisteren. |
| Setting 1 | Vanilla | I love this song. It is a great song. It is also a great song. I recommend it to everyone. |
| | Adaptation | Zo'n lekker nummer is een aanrader voor iedereen. |
| Setting 2 | Vanilla | Het is een goed nummer, dat is ook een aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |
| Setting 3 | Vanilla | Het is een erg goed nummer. Het is een echte aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |

Example outputs generated by different models.

Formality Transfer in Dutch: Results

| | | |
|--------------------|-------------------|--|
| Source | | Zo'n lekker nummer dat ook btw aanrader voor iedereen |
| Reference 1 | | Zo een leuk liedje trouwens, dat kan ik iedereen aanraden. |
| Reference 2 | | Dit is een goed nummer dat ik aan iedereen aan kan raden. |
| Reference 3 | | Dit is een zeer goed nummer. Ik raad het iedereen aan om er naar te luisteren. |
| Setting 1 | Vanilla | I love this song. It is a great song. It is also a great song. I recommend it to everyone. |
| | Adaptation | Zo'n lekker nummer is een aanrader voor iedereen. |
| Setting 2 | Vanilla | Het is een goed nummer, dat is ook een aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |
| Setting 3 | Vanilla | Het is een erg goed nummer. Het is een echte aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |

Example outputs generated by different models.

Formality Transfer in Dutch: Results

| | | |
|--------------------|-------------------|--|
| Source | | Zo'n lekker nummer dat ook btw aanrader voor iedereen |
| Reference 1 | | Zo een leuk liedje trouwens, dat kan ik iedereen aanraden. |
| Reference 2 | | Dit is een goed nummer dat ik aan iedereen aan kan raden. |
| Reference 3 | | Dit is een zeer goed nummer. Ik raad het iedereen aan om er naar te luisteren. |
| Setting 1 | Vanilla | I love this song. It is a great song. It is also a great song. I recommend it to everyone. |
| | Adaptation | Zo'n lekker nummer is een aanrader voor iedereen. |
| Setting 2 | Vanilla | Het is een goed nummer, dat is ook een aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |
| Setting 3 | Vanilla | Het is een erg goed nummer. Het is een echte aanrader voor iedereen. |
| | Adaptation | Zo'n lekker nummer is ook een aanrader voor iedereen. |

Example outputs generated by different models.

Summary

Data

- Formality transfer dataset in Dutch (manually for testing, and automatically translated for training)

Models

- mBART based baselines (adaptation training strategy).
- Relying solely on English training data yields poor results, while using NL machine-translated data produces considerably better outcomes



@HuiyuanLai @MalvinaNissim @GroNLP



<https://github.com/MariekeWeultjes/dutch-style-transfer>