

The LEONIDE Corpus

Description of the anonymisation tags

In the LEONIDE Corpus, all words and sequences of words were anonymized that could reveal the identity of the authors of the texts or other people that the text was related to. Public figures (politicians, sportsperson, artists, etc.), and neutral mentions of places and institutions were not included in the process of anonymization.

Names of real individuals, animals, places and schools were substituted with a respective placeholder and additionally annotated with the annotation tag *tran_anonymization*:

- Individuals:
 - o **Forename**: This placeholder is used to anonymize the forename of an existing person.
 - o **Surname**: This placeholder is used to anonymize the forename of an existing person.
- Animals:
 - o **Animalname**: This placeholder is used to anonymize the name of an existing animal (e.g. a dog, a horse) that helps to identify the author or any other existing person.
- Places:
 - o **Placename**: This placeholder is used to anonymize the name of existing cities, villages, streets, squares, bars and any other locations that may reveal the identity of the author or any other existing person.
 - o **Schoolname**: This placeholder is used to anonymize the name of an existing school.

If there are several names of individuals or animals, place names or names of schools used in one text, a counting was added to the placeholder which reflects the order of appearance (e.g. "Surname_1", "Placename_2").