Description of the anonymisation tags in the DiDi Corpus

In the DiDi Corpus, all words and sequences of word were anonymised that could reveal the identity of the authors of the text 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 anonymisation.

The following anonymisation tags were used:

- **_PersNE_**: Names of people and pets (first names, surnames, nicknames, profile names of Facebook, etc.):
  - note:
    - Compounds: If a _PersNE_ is part of a compound, only the proper name itself is anonymised, the other part of the compound remains original. In accordance with the original, the spelling of the compound can be with or without hyphen, e.g.:
      - _PersNE_-Witz ("_PersNE_ joke")
      - _PersNE_kuschlobnt ("an evening to cuddle with _PersNE_")
    - Case marking: The case-marking suffixes (genitiv) are maintained. In accordance with the original, the spelling varies with the original language (with or without apostrophes, e.g.:
      - _PersNE_‘s Privatsphäreneinstellung (’_PersNE_‘s privacy setting’)
      - _PersNE_‘s Popile (’_PersNE_‘s baby’)
      - _PersNE_‘s mother
  - @mentions: @mentions also follow the anonymisation regularities, e.g.:
    - @_PersNE_

- **_GruppeNN_**: Collective tag for people and group of people (inhabitants and other collective terms that include individual-related characteristics of any kind, such as place of origin, characteristic preferences, profession, occupation, e.g. Bozner, Boznerin

- **_GeoNE_**: Topographic names, specific domiciles and addresses of the authors and their friends, localities where the authors and their friends can be usually met (names of cities and villages, mountains, valleys, regions, streets, public places):
  - Compounds: If a _GeoNE_ is part of a compound, only the proper name itself is anonymised, the other part of the compound remains original. In accordance with the original, the spelling of the compound can be with or without hyphen, e.g.:
    - _GeoNE_aufentholt ("stay at _GeoNE_")
    - _GeoNE_City
  - @mentions: @mentions also follow the anonymisation regularities, e.g.:
    - @_GeoNE_

- **_GeoADJA_**: Adjectival derivations of _GeoNE_ (Bozner, bolzanino)
- **_InstNE_:** Names of non-topographic places that could reveal the identity of the authors and their friends (clubs and bars, hotels, restaurants, discotheques, sports clubs, huts, bands, schools, etc.)
  - **Compounds:** If an _InstNE_ is part of a compound, only the proper name itself is anonymised, the other part of the compound remains original. In accordance with the original, the spelling of the compound can be with or without hyphen, e.g.:
    - _InstNE_ SchülerInnen, (“pupils from _InstNE_”)
    - _InstNE_‐Zeltlager (’_InstNE_ tent camp’)
  - **@mentions:** @mentions also follow the anonymisation regularities, e.g.:
    - @_InstNE_

- **_XXX_:** Quotations of other Facebook users that for which we do not have a consent, numbers of bank accounts and servers, passwords, postal codes, street number, etc.
- **_link_:** Links to webpages
- **_mail_:** e-mail addresses
- **_tel_:** phone numbers