

## MMS Dataset Citations

*Citations for the MMS datasets*

### Citations

#### Dataset id: ar\_arsentdl

- Domain: **social\_media**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@InProceedings{dataset_ar_arsentdl,  
  author = {Ramy Baly and  
            Alaa Khaddaj and  
            Hazem M. Hajj and  
            Wassim El{-}Hajj and  
            Khaled Bashir Shaban},  
  title = {{ArSentD-LEV: A Multi-Topic Corpus for Target-based Sentiment Analysis in Arabic  
booktitle = {Proceedings of the Eleventh International Conference on Language Resources an  
  year = {2018},  
  month = {may},  
  date = {7-12},  
  location = {Miyazaki, Japan},  
  editor = {Hend Al-Khalifa and King Saud University and KSA Walid Magdy and University of E  
  publisher = {European Language Resources Association (ELRA)},  
  address = {Paris, France},  
  isbn = {979-10-95546-25-2},  
  language = {english}}
```

```
}
```

## Dataset id: ar\_astd

- Domain: **social\_media**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_ar_astd,  
  title = "{ASTD}: {A}rabic Sentiment Tweets Dataset",  
  author = "Nabil, Mahmoud  and  
    Aly, Mohamed  and  
    Atiya, Amir",  
  booktitle = "Proceedings of the 2015 Conference on Empirical Methods in Natural Language P  
  month = sep,  
  year = "2015",  
  address = "Lisbon, Portugal",  
  publisher = "Association for Computational Linguistics",  
  url = "https://aclanthology.org/D15-1299",  
  doi = "10.18653/v1/D15-1299",  
  pages = "2515--2519",  
}
```

## Dataset id: ar\_bbn

- Domain: **social\_media**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**

- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_ar_bbn,
  title = "Sentiment after Translation: A Case-Study on {A}rabic Social Media Posts",
  author = "Salameh, Mohammad and
    Mohammad, Saif and
    Kiritchenko, Svetlana",
  booktitle = "Proceedings of the 2015 Conference of the North {A}merican Chapter of the Ass
  month = may # "{--}" # jun,
  year = "2015",
  address = "Denver, Colorado",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/N15-1078",
  doi = "10.3115/v1/N15-1078",
  pages = "767--777",
}
```

## Dataset id: ar\_brad

- Domain: **reviews**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@INPROCEEDINGS{dataset_ar_brad,
  author={Elnagar, Ashraf and Einea, Omar},
  booktitle={2016 IEEE/ACS 13th International Conference of Computer Systems and Application
  title={{BRAD} 1.0: Book reviews in Arabic dataset},
  year={2016},
  volume={},
  number={},
  pages={1-8},
  doi={10.1109/AICCSA.2016.7945800}
}
```

## Dataset id: ar\_hard

- Domain: **reviews**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@Book{dataset_ar_hard,  
  author="Elnagar, Ashraf  
and Khalifa, Yasmin S.  
and Einea, Anas",  
  title={Hotel Arabic-Reviews Dataset Construction for Sentiment Analysis Applications},  
  bookTitle="Intelligent Natural Language Processing: Trends and Applications",  
  year="2018",  
  publisher="Springer International Publishing",  
  address="Cham",  
  pages="35--52",  
  isbn="978-3-319-67056-0",  
  doi="10.1007/978-3-319-67056-0_3",  
  url="https://doi.org/10.1007/978-3-319-67056-0_3"  
}
```

## Dataset id: ar\_labr

- Domain: **reviews**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**



```
@inproceedings{dataset_ar_labr,
  title = "{LABR}: A Large Scale {A}rabic Book Reviews Dataset",
  author = "Aly, Mohamed and
    Atiya, Amir",
  booktitle = "Proceedings of the 51st Annual Meeting of the Association for Computational L
  month = aug,
  year = "2013",
  address = "Sofia, Bulgaria",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/P13-2088",
  pages = "494--498",
}
```

## Dataset id: ar\_oclar

- Domain: **reviews**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_ar_oclar,
  author={Al Omari, Marwan and Al-Hajj, Moustafa and Hammami, Nacereddine and Sabra, Amani},
  booktitle={2019 International Conference on Computer and Information Sciences (ICCIS)},
  title={Sentiment Classifier: Logistic Regression for Arabic Services' Reviews in Lebanon},
  year={2019},
  volume={},
  number={},
  pages={1-5},
  doi={10.1109/ICCISci.2019.8716394}
}
```

## Dataset id: ar\_semeval\_2017

- Domain: **mixed**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**

- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_semeval_2017,
  title = "{S}em{E}val-2017 Task 4: Sentiment Analysis in {T}witter",
  author = "Rosenthal, Sara and
    Farra, Noura and
    Nakov, Preslav",
  booktitle = "Proceedings of the 11th International Workshop on Semantic Evaluation ({S}em{E}val-2017)",
  month = aug,
  year = "2017",
  address = "Vancouver, Canada",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/S17-2088",
  doi = "10.18653/v1/S17-2088",
  pages = "502--518",
  abstract = "This paper describes the fifth year of the Sentiment Analysis in Twitter task."
}
```

## Dataset id: ar\_syria\_corpus

- Domain: **social\_media**
- Language: **ar**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_ar_bbn,
  title = "Sentiment after Translation: A Case-Study on {A}rabic Social Media Posts",
  author = "Salameh, Mohammad and
    Mohammad, Saif and
    ..."
```

```

    Kiritchenko, Svetlana",
    booktitle = "Proceedings of the 2015 Conference of the North {A}merican Chapter of the Ass
    month = may # "{--}" # jun,
    year = "2015",
    address = "Denver, Colorado",
    publisher = "Association for Computational Linguistics",
    url = "https://aclanthology.org/N15-1078",
    doi = "10.3115/v1/N15-1078",
    pages = "767--777",
}

```

## Dataset id: bg\_twitter\_sentiment

- Domain: **social\_media**
- Language: **bg**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **no article**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```

@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}

```

## Dataset id: bs\_twitter\_sentiment

- Domain: **social\_media**
- Language: **bs**
- Language family: **Indo-European**
- Genus: **Slavic**

- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **5**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **other**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}
```

## Dataset id: cs\_facebook

- Domain: **social\_media**
- Language: **cs**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative affix**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_cs_social_media,
  title = "Sentiment Analysis in {C}zech Social Media Using Supervised Machine Learning",
  author = "Habernal, Ivan and
    Pt{\`a}{\v{c}}ek, Tom{\`a}{\v{s}} and
    Steinberger, Josef",
  booktitle = "Proceedings of the 4th Workshop on Computational Approaches to Subjectivity",
  month = jun,
```

```

year = "2013",
address = "Atlanta, Georgia",
publisher = "Association for Computational Linguistics",
url = "https://aclanthology.org/W13-1609",
pages = "65--74",
}

```

## Dataset id: cs\_mall\_product\_reviews

- Domain: **reviews**
- Language: **cs**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative affix**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```

@inproceedings{dataset_cs_social_media,
  title = "Sentiment Analysis in {C}zech Social Media Using Supervised Machine Learning",
  author = "Habernal, Ivan and
    Pt{\`a}{\v{c}}ek, Tom{\`a}{\v{s}} and
    Steinberger, Josef",
  booktitle = "Proceedings of the 4th Workshop on Computational Approaches to Subjectivity",
  month = jun,
  year = "2013",
  address = "Atlanta, Georgia",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/W13-1609",
  pages = "65--74",
}

```

## Dataset id: cs\_movie\_reviews

- Domain: **reviews**
- Language: **cs**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**

- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative affix**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_cs_social_media,
  title = "Sentiment Analysis in {C}zech Social Media Using Supervised Machine Learning",
  author = "Habernal, Ivan  and
    Pt{\`a}{\v{c}}ek, Tom{\`a}{\v{s}}  and
    Steinberger, Josef",
  booktitle = "Proceedings of the 4th Workshop on Computational Approaches to Subjectivity",
  month = jun,
  year = "2013",
  address = "Atlanta, Georgia",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/W13-1609",
  pages = "65--74",
}
```

## Dataset id: cs\_news\_stance

- Domain: **social\_media**
- Language: **cs**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative affix**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_cs_social_media,
  title = "Sentiment Analysis in {C}zech Social Media Using Supervised Machine Learning",
  author = "Habernal, Ivan  and
    Pt{\`a}{\v{c}}ek, Tom{\`a}{\v{s}}  and
    Steinberger, Josef",
  booktitle = "Proceedings of the 4th Workshop on Computational Approaches to Subjectivity",
  month = jun,
  year = "2013",
  address = "Atlanta, Georgia",
```

```
publisher = "Association for Computational Linguistics",
url = "https://aclanthology.org/W13-1609",
pages = "65--74",
}
```

## Dataset id: de\_dai\_labor

- Domain: **social\_media**
- Language: **de**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **4**
- Order of subject, object, verb: **no dominant order**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **more than one position**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_dai_labor,
  author = {Narr, Sascha and Michael Hülfehaus and Albayrak, Sahin},
  title = {Language-Independent Twitter Sentiment Analysis},
  booktitle = {Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012)},
  year = {2012},
  location = {Dortmund, Germany},
}
```

## Dataset id: de\_ifeel

- Domain: **social\_media**
- Language: **de**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **4**
- Order of subject, object, verb: **no dominant order**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **more than one position**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**

- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_dai_labor,
  author = {Narr, Sascha and Michael Hülfehaus and Albayrak, Sahin},
  title = {Language-Independent Twitter Sentiment Analysis},
  booktitle = {Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012)},
  year = {2012},
  location = {Dortmund, Germany},
}
```

## Dataset id: de\_multilan\_amazon

- Domain: **reviews**
- Language: **de**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **4**
- Order of subject, object, verb: **no dominant order**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **more than one position**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_multilan_amazon,
  title = "The Multilingual {A}mazon Reviews Corpus",
  author = {Keung, Phillip and
    Lu, Yichao and
    Szarvas, Gy{"o"}rgy and
    Smith, Noah A.},
  booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language P",
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.emnlp-main.369",
  doi = "10.18653/v1/2020.emnlp-main.369",
  pages = "4563--4568",
}
```

## Dataset id: de\_omp

- Domain: **social\_media**
- Language: **de**



- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **4**
- Order of subject, object, verb: **no dominant order**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **more than one position**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_de_omp,
  title = "Academic-Industrial Perspective on the Development and Deployment of a Moderation",
  author = "Schabus, Dietmar and Skowron, Marcin",
  booktitle = "Proceedings of the Eleventh International Conference on Language Resources and Evaluation",
  month = may,
  year = "2018",
  address = "Miyazaki, Japan",
  publisher = "European Language Resources Association (ELRA)",
  url = "https://aclanthology.org/L18-1253",
}
```

## Dataset id: de\_sb10k

- Domain: **social\_media**
- Language: **de**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **4**
- Order of subject, object, verb: **no dominant order**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **more than one position**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_de_sb10k,
  title = "A {T}witter Corpus and Benchmark Resources for {G}erman Sentiment Analysis",
  author = "Cieliebak, Mark and Deriu, Jan Milan and Egger, Dominic and Uzdilli, Fatih",
}
```

```

booktitle = "Proceedings of the Fifth International Workshop on Natural Language Processin
month = apr,
year = "2017",
address = "Valencia, Spain",
publisher = "Association for Computational Linguistics",
url = "https://aclanthology.org/W17-1106",
doi = "10.18653/v1/W17-1106",
pages = "45--51",
abstract = "In this paper we present SB10k, a new corpus for sentiment analysis with appro
}

```

## Dataset id: de\_twitter\_sentiment

- Domain: **social\_media**
- Language: **de**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **4**
- Order of subject, object, verb: **no dominant order**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **more than one position**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```

@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}

```

## Dataset id: en\_amazon

- Domain: **reviews**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**

- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_en_amazon,
  title = "Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects",
  author = "Ni, Jianmo and
    Li, Jiacheng and
    McAuley, Julian",
  booktitle = "Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing",
  month = nov,
  year = "2019",
  address = "Hong Kong, China",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/D19-1018",
  doi = "10.18653/v1/D19-1018",
  pages = "188--197",
}
```

## Dataset id: en\_dai\_labor

- Domain: **social\_media**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_dai_labor,
  author = {Narr, Sascha and Michael Hülfehaus and Albayrak, Sahin},
  title = {Language-Independent Twitter Sentiment Analysis},
  booktitle = {Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012)},
  year = {2012},
}
```

```
location = {Dortmund, Germany},  
}
```

## Dataset id: en\_financial\_phrasebank\_sentences\_75agree

- Domain: **news**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@article{dataset_en_financial_phrasebank_sentences_75agree,  
  author = {Malo, Pekka and Sinha, Ankur and Korhonen, Pekka and Wallenius, Jyrki and Takala},  
  title = {Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts},  
  year = {2014},  
  issue_date = {April 2014},  
  publisher = {John Wiley & Sons, Inc.},  
  address = {USA},  
  volume = {65},  
  number = {4},  
  issn = {2330-1635},  
  url = {https://doi.org/10.1002/asi.23062},  
  doi = {10.1002/asi.23062},  
  journal = {Journal of the Association for Information Science and Technology},  
  month = {apr},  
  pages = {782-796},  
  numpages = {15},  
  keywords = {economics, automatic classification, linguistic analysis}  
}
```

## Dataset id: en\_multilan\_amazon

- Domain: **reviews**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**

- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_multilan_amazon,
  title = "The Multilingual {A}mazon Reviews Corpus",
  author = {Keung, Phillip and
    Lu, Yichao and
    Szarvas, Gy{\o}rgy and
    Smith, Noah A.},
  booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language P
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.emnlp-main.369",
  doi = "10.18653/v1/2020.emnlp-main.369",
  pages = "4563--4568",
}
```

## Dataset id: en\_per\_sent

- Domain: **news**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_en_per_sent,
  title = "Author{'}'s Sentiment Prediction",
  author = "Bastan, Mohaddeseh and
    Koupaei, Mahnaz and
    Son, Youngseo and
    Sicoli, Richard and
```

```

        Balasubramanian, Niranjan",
booktitle = "Proceedings of the 28th International Conference on Computational Linguistics
month = dec,
year = "2020",
address = "Barcelona, Spain (Online)",
publisher = "International Committee on Computational Linguistics",
url = "https://aclanthology.org/2020.coling-main.52",
doi = "10.18653/v1/2020.coling-main.52",
pages = "604--615",
}

```

## Dataset id: en\_poem\_sentiment

- Domain: **poems**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```

@inproceedings{dataset_en_poem_sentiment,
  title = "Investigating Societal Biases in a Poetry Composition System",
  author = "Sheng, Emily  and
    Uthus, David",
  booktitle = "Proceedings of the Second Workshop on Gender Bias in Natural Language Process
  month = dec,
  year = "2020",
  address = "Barcelona, Spain (Online)",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.gebnlp-1.9",
  pages = "93--106",
}

```

## Dataset id: en\_emeval\_2017

- Domain: **mixed**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**

- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_semeval_2017,
  title = "{S}em{E}val-2017 Task 4: Sentiment Analysis in {T}witter",
  author = "Rosenthal, Sara and
    Farra, Noura and
    Nakov, Preslav",
  booktitle = "Proceedings of the 11th International Workshop on Semantic Evaluation ({S}em{
  month = aug,
  year = "2017",
  address = "Vancouver, Canada",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/S17-2088",
  doi = "10.18653/v1/S17-2088",
  pages = "502--518",
  abstract = "This paper describes the fifth year of the Sentiment Analysis in Twitter task.
}
```

## Dataset id: en\_sentistrength

- Domain: **social\_media**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@article{dataset_en_sentistrength,
  author = {Thelwall, Mike and Buckley, Kevan and Paltoglou, Georgios},
  title = {Sentiment Strength Detection for the Social Web},
  year = {2012},
```

```

issue_date = {January 2012},
publisher = {John Wiley & Sons, Inc.},
address = {USA},
volume = {63},
number = {1},
issn = {1532-2882},
url = {https://doi.org/10.1002/asi.21662},
doi = {10.1002/asi.21662},
abstract = {Sentiment analysis is concerned with the automatic extraction of sentiment-relevant information from text. Although most sentiment analysis addresses commercial tasks, such as extracting opinions from product reviews, there is increasing interest in the affective dimension of the social web, and Twitter in particular. Most sentiment analysis algorithms are not ideally suited to this task because they exploit indirect indicators of sentiment that can reflect genre or topic instead. Hence, such algorithms used to process social web texts can identify spurious sentiment patterns caused by topics rather than affective phenomena. This article assesses an improved version of the algorithm SentiStrength for sentiment strength detection across the social web that primarily uses direct indications of sentiment. The results from six diverse social web data sets (MySpace, Twitter, YouTube, Digg, RunnersWorld, BBCForums) indicate that SentiStrength 2 is successful in the sense of performing better than a baseline approach for all data sets in both supervised and unsupervised cases. SentiStrength is not always better than machine-learning approaches that exploit indirect indicators of sentiment, however, and is particularly weaker for positive sentiment in news-related discussions. Overall, the results suggest that, even unsupervised, SentiStrength is robust enough to be applied to a wide variety of different social web contexts.},
journal = {J. Am. Soc. Inf. Sci. Technol.},
month = jan,
pages = {163-173},
numpages = {11}
}

```

## Dataset id: en\_silicone\_meld\_s

- Domain: **chats**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```

@inproceedings{dataset_en_silicone,
  title = "Hierarchical Pre-training for Sequence Labelling in Spoken Dialog",

```



```

author = "Chapuis, Emile and
         Colombo, Pierre and
         Manica, Matteo and
         Labeau, Matthieu and
         Clavel, Chlo{'e}",
booktitle = "Findings of the Association for Computational Linguistics: EMNLP 2020",
month = nov,
year = "2020",
address = "Online",
publisher = "Association for Computational Linguistics",
url = "https://aclanthology.org/2020.findings-emnlp.239",
doi = "10.18653/v1/2020.findings-emnlp.239",
pages = "2636--2648",
}

```

## Dataset id: en\_silicone\_sem

- Domain: **chats**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```

@inproceedings{dataset_en_silicone,
  title = "Hierarchical Pre-training for Sequence Labelling in Spoken Dialog",
  author = "Chapuis, Emile and
           Colombo, Pierre and
           Manica, Matteo and
           Labeau, Matthieu and
           Clavel, Chlo{'e}",
  booktitle = "Findings of the Association for Computational Linguistics: EMNLP 2020",
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.findings-emnlp.239",
  doi = "10.18653/v1/2020.findings-emnlp.239",
  pages = "2636--2648",
}

```

## Dataset id: en\_tweet\_airlines

- Domain: **social\_media**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@misc{dataset_en_tweet_airlines,  
  url={https://www.kaggle.com/crowdflower/twitter-airline-sentiment},  
  author={Crowdflower Inc.},  
  title={Twitter US Airline Sentiment},  
  year={2015}  
}
```

## Dataset id: en\_tweets\_sanders

- Domain: **social\_media**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@article{dataset_en_tweets_sanders,  
  title={{Sanders-Twitter Sentiment Corpus}},  
  author={Sanders, Niek J},  
  journal={Sanders Analytics LLC},  
  year={2011}  
}
```

## Dataset id: en\_twitter\_sentiment

- Domain: **social\_media**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@article{dataset_twitter_sentiment,  
  doi = {10.1371/journal.pone.0155036},  
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},  
  journal = {PLOS ONE},  
  publisher = {Public Library of Science},  
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},  
  year = {2016},  
  month = {05},  
  volume = {11},  
  url = {https://doi.org/10.1371/journal.pone.0155036},  
  pages = {1-26},  
  number = {5},  
}
```

## Dataset id: en\_vader\_amazon

- Domain: **reviews**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_en_vader,  
  title={{VADER}: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text},  
  author={Clayton J. Hutto and Eric Gilbert},  
  booktitle={Proceedings of the International AAAI Conference on Web and Social Media},  
  year={2014},  
  url={https://ojs.aaai.org/index.php/ICWSM/article/view/14550},  
  month={May},  
  pages={216-225},  
  volume=8,  
}
```

## Dataset id: en\_vader\_movie\_reviews

- Domain: **reviews**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_en_vader,  
  title={{VADER}: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text},  
  author={Clayton J. Hutto and Eric Gilbert},  
  booktitle={Proceedings of the International AAAI Conference on Web and Social Media},  
  year={2014},  
  url={https://ojs.aaai.org/index.php/ICWSM/article/view/14550},  
  month={May},  
  pages={216-225},  
  volume=8,  
}
```

## Dataset id: en\_vader\_nyt

- Domain: **news**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**

- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_en_vader,
  title={{VADER}: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text},
  author={Clayton J. Hutto and Eric Gilbert},
  booktitle={Proceedings of the International AAAI Conference on Web and Social Media},
  year={2014},
  url={https://ojs.aaai.org/index.php/ICWSM/article/view/14550},
  month={May},
  pages={216-225},
  volume=8,
}
```

## Dataset id: en\_vader\_twitter

- Domain: **social\_media**
- Language: **en**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_en_vader,
  title={{VADER}: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text},
  author={Clayton J. Hutto and Eric Gilbert},
  booktitle={Proceedings of the International AAAI Conference on Web and Social Media},
  year={2014},
  url={https://ojs.aaai.org/index.php/ICWSM/article/view/14550},
  month={May},
  pages={216-225},
  volume=8,
}
```

```
}
```

## Dataset id: es\_muchocine

- Domain: **reviews**
- Language: **es**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@article{dataset_es_muchocine,  
  title={Experiments in sentiment classification of movie reviews in Spanish},  
  author={Cruz, Fermin L and Troyano, Jose A and Enriquez, Fernando and Ortega, Javier},  
  journal={Procesamiento del Lenguaje Natural},  
  volume={41},  
  pages={73--80},  
  year={2008},  
  publisher={SOC ESPANOLA PROCESAMIENTO LENGUAJE NATURAL-SEPLN DEPT LENGUAJES \& SISTEMAS~...}  
}
```

## Dataset id: es\_multilan\_amazon

- Domain: **reviews**
- Language: **es**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_multilan_amazon,
  title = "The Multilingual {A}mazon Reviews Corpus",
  author = {Keung, Phillip and
    Lu, Yichao and
    Szarvas, Gy{\o}rgy and
    Smith, Noah A.},
  booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language P
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.emnlp-main.369",
  doi = "10.18653/v1/2020.emnlp-main.369",
  pages = "4563--4568",
}
```

## Dataset id: es\_paper\_reviews

- Domain: **reviews**
- Language: **es**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@article{dataset_es_paper_reviews,
  author = {Keith Norambuena, Brian and Lettura, Exequiel and Villegas, Claudio},
  year = {2019},
  month = {02},
  pages = {191-214},
  title = {Sentiment analysis and opinion mining applied to scientific paper reviews},
  volume = {23},
  journal = {Intelligent Data Analysis},
  doi = {10.3233/IDA-173807}
}
```

## Dataset id: es\_semeval2020

- Domain: **social\_media**
- Language: **es**

- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_ semeval_2020,
  title = "{S}em{E}val-2020 Task 9: Overview of Sentiment Analysis of Code-Mixed Tweets",
  author = {Patwa, Parth and
    Aguilar, Gustavo and
    Kar, Sudipta and
    Pandey, Suraj and
    PYKL, Srinivas and
    Gamb{"a"}ck, Bj{"o"}rn and
    Chakraborty, Tanmoy and
    Solorio, Tamar and
    Das, Amitava},
  booktitle = "Proceedings of the Fourteenth Workshop on Semantic Evaluation",
  month = dec,
  year = "2020",
  address = "Barcelona (online)",
  publisher = "International Committee for Computational Linguistics",
  url = "https://aclanthology.org/2020.semeval-1.100",
  doi = "10.18653/v1/2020.semeval-1.100",
  pages = "774--790",
}
```

## Dataset id: es\_twitter\_sentiment

- Domain: **social\_media**
- Language: **es**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**



- Grammatical genders: **masculine, feminine**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}
```

## Dataset id: fa\_sentipers

- Domain: **reviews**
- Language: **fa**
- Language family: **Indo-European**
- Genus: **Iranian**
- Definite articles: **no article**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **2**
- Order of subject, object, verb: **SOV**
- Negative morphemes: **negative affix**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@article{dataset_fa_sentipers,
  author = {Pedram Hosseini and
    Ali Ahmadian Ramaki and
    Hassan Maleki and
    Mansoureh Anvari and
    Seyed Abolghasem Mirroshandel},
  title = {{SentiPers}: {A} Sentiment Analysis Corpus for Persian},
  journal = {Computing Research Repository},
  volume = {arXiv:1801.07737},
  note = {Version 2},
  year = {2018},
  url = {http://arxiv.org/abs/1801.07737},
  eprinttype = {arXiv},
  eprint = {1801.07737},
  timestamp = {Mon, 13 Aug 2018 16:47:47 +0200},
  biburl = {https://dblp.org/rec/journals/corr/abs-1801-07737.bib},
  bibsource = {dblp computer science bibliography, https://dblp.org}
```

```
}
```

## Dataset id: fr\_dai\_labor

- Domain: **social\_media**
- Language: **fr**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **OptDoubleNeg**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_dai_labor,  
  author = {Narr, Sascha and Michael Hülfenhaus and Albayrak, Sahin},  
  title = {Language-Independent Twitter Sentiment Analysis},  
  booktitle = {Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012)},  
  year = {2012},  
  location = {Dortmund, Germany},  
}
```

## Dataset id: fr\_ifeel

- Domain: **social\_media**
- Language: **fr**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **OptDoubleNeg**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_dai_labor,  
  author = {Narr, Sascha and Michael Hülfenhaus and Albayrak, Sahin},
```

```

title = {Language-Independent Twitter Sentiment Analysis},
booktitle = {Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012)}
year = {2012},
location = {Dortmund, Germany},
}

```

## Dataset id: fr\_multilan\_amazon

- Domain: **reviews**
- Language: **fr**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **OptDoubleNeg**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```

@inproceedings{dataset_multilan_amazon,
  title = "The Multilingual {A}mazon Reviews Corpus",
  author = {Keung, Phillip and
    Lu, Yichao and
    Szarvas, Gy{"o"}rgy and
    Smith, Noah A.},
  booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language P
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.emnlp-main.369",
  doi = "10.18653/v1/2020.emnlp-main.369",
  pages = "4563--4568",
}

```

## Dataset id: he\_hebrew\_sentiment

- Domain: **social\_media**
- Language: **he**
- Language family: **Afro-Asiatic**
- Genus: **Semitic**
- Definite articles: **definite affix**
- Indefinite articles: **indefinite word same as one**

- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_he_hebrew_sentiment,
  title = "Representations and Architectures in Neural Sentiment Analysis for Morphologically
  author = "Amram, Adam  and
    Ben David, Anat  and
    Tsarfaty, Reut",
  booktitle = "Proceedings of the 27th International Conference on Computational Linguistics
  month = aug,
  year = "2018",
  address = "Santa Fe, New Mexico, USA",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/C18-1190",
  pages = "2242--2252",
  abstract = "This paper empirically studies the effects of representation choices on neural
}
```

## Dataset id: hi\_semeval2020

- Domain: **social\_media**
- Language: **hi**
- Language family: **Indo-European**
- Genus: **Indic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **3**
- Order of subject, object, verb: **SOV**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SONegV**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_semeval_2020,
  title = "{S}em{E}val-2020 Task 9: Overview of Sentiment Analysis of Code-Mixed Tweets",
  author = {Patwa, Parth  and
    Aguilar, Gustavo  and
    Kar, Sudipta  and
    Pandey, Suraj  and
    PYKL, Srinivas  and
```

```

    Gamb{"a}ck, Bj{"o}rn and
    Chakraborty, Tanmoy and
    Solorio, Tamar and
    Das, Amitava},
    booktitle = "Proceedings of the Fourteenth Workshop on Semantic Evaluation",
    month = dec,
    year = "2020",
    address = "Barcelona (online)",
    publisher = "International Committee for Computational Linguistics",
    url = "https://aclanthology.org/2020.semeval-1.100",
    doi = "10.18653/v1/2020.semeval-1.100",
    pages = "774--790",
}

```

## Dataset id: hr\_sentiment\_news\_document

- Domain: **news**
- Language: **hr**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **5**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **other**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```

@Article{dataset_hr_sentiment_news_document,
  AUTHOR = {Pelicon, Andraž and Pranjić, Marko and Miljković, Dragana and Škrlić, Blaž and Po
  TITLE = {Zero-Shot Learning for Cross-Lingual News Sentiment Classification},
  JOURNAL = {Applied Sciences},
  VOLUME = {10},
  YEAR = {2020},
  NUMBER = {17},
  ARTICLE-NUMBER = {5993},
  URL = {https://www.mdpi.com/2076-3417/10/17/5993},
  ISSN = {2076-3417},
  ABSTRACT = {In this paper, we address the task of zero-shot cross-lingual news sentiment c
  DOI = {10.3390/app10175993}
}

```

## Dataset id: hr\_twitter\_sentiment

- Domain: **social\_media**

- Language: **hr**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **5**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **other**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}
```

## Dataset id: hu\_twitter\_sentiment

- Domain: **social\_media**
- Language: **hu**
- Language family: **Uralic**
- Genus: **Ugric**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **10 or more**
- Order of subject, object, verb: **no dominant order**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
```

```

journal = {PLOS ONE},
publisher = {Public Library of Science},
title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
year = {2016},
month = {05},
volume = {11},
url = {https://doi.org/10.1371/journal.pone.0155036},
pages = {1-26},
number = {5},
}

```

## Dataset id: it\_evalita2016

- Domain: **social\_media**
- Language: **it**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```

@inproceedings{dataset_it_evalita2016,
  TITLE = {{Overview of the Evalita 2016 SENTiment POLarity Classification Task}},
  AUTHOR = {Barbieri, Francesco and Basile, Valerio and Croce, Danilo and Nissim, Malvina and
  URL = {https://hal.inria.fr/hal-01414731},
  BOOKTITLE = {{Proceedings of Third Italian Conference on Computational Linguistics (CLiC-i
  ADDRESS = {Naples, Italy},
  YEAR = {2016},
  MONTH = Dec,
  KEYWORDS = {Natural language processing and web ; Social media analysis ; Sentiment analys
  PDF = {https://hal.inria.fr/hal-01414731/file/paper_026.pdf},
  HAL_ID = {hal-01414731},
  HAL_VERSION = {v1},
}

```

## Dataset id: it\_multiemotions

- Domain: **social\_media**
- Language: **it**
- Language family: **Indo-European**
- Genus: **Romance**

- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative intonation only**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_it_multiemotions,
  author = {Sprugnoli, Rachele},
  year = {2020},
  month = {12},
  pages = {},
  title = {MultiEmotions-It: a New Dataset for Opinion Polarity and Emotion Analysis for Ita
  booktitle = {Proceedings of the Seventh Italian Conference on Computational Linguistics},
}
```

## Dataset id: ja\_multilan\_amazon

- Domain: **reviews**
- Language: **ja**
- Language family: **Japanese**
- Genus: **Japanese**
- Definite articles: **no article**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **8-9**
- Order of subject, object, verb: **SOV**
- Negative morphemes: **negative affix**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **no grammatical gender**

```
@inproceedings{dataset_multilan_amazon,
  title = "The Multilingual {A}mazon Reviews Corpus",
  author = {Keung, Phillip and
    Lu, Yichao and
    Szarvas, Gy{"o"}rgy and
    Smith, Noah A.},
  booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language P
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
}
```



```

url = "https://aclanthology.org/2020.emnlp-main.369",
doi = "10.18653/v1/2020.emnlp-main.369",
pages = "4563--4568",
}

```

## Dataset id: lv\_ltec\_sentiment

- Domain: **social\_media**
- Language: **lv**
- Language family: **Indo-European**
- Genus: **Baltic**
- Definite articles: **demonstrative word used as definite article**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **5**

◀ Order of subject, object, verb: **SVO** ▶

- Negative morphemes: **negative affix**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```

@article{dataset_lv_ltec_sentiment,
  author    = {Uga Sprogis and
              Matiss Rikters},
  title     = {What Can We Learn From Almost a Decade of Food Tweets},
  journal   = {Computing Research Repository},
  volume    = {arXiv:2007.05194},
  note      = {Version 2},
  year      = {2020},
  url       = {https://arxiv.org/abs/2007.05194},
  eprinttype = {arXiv},
  eprint    = {2007.05194},
  timestamp = {Mon, 20 Jul 2020 14:20:39 +0200},
  biburl    = {https://dblp.org/rec/journals/corr/abs-2007-05194.bib},
  bibsource = {dblp computer science bibliography, https://dblp.org}
}

```

## Dataset id: pl\_klej\_allegro\_reviews

- Domain: **reviews**
- Language: **pl**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**

- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_pl_klej_allegro_reviews,
  title = "{KLEJ}: Comprehensive Benchmark for {P}olish Language Understanding",
  author = "Rybak, Piotr and
    Mroczkowski, Robert and
    Tracz, Janusz and
    Gawlik, Ireneusz",
  booktitle = "Proceedings of the 58th Annual Meeting of the Association for Computational L
  month = jul,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.acl-main.111",
  doi = "10.18653/v1/2020.acl-main.111",
  pages = "1191--1201",
}
```

## Dataset id: pl\_opi\_lil\_2012

- Domain: **social\_media**
- Language: **pl**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_pl_opi_lil_2012,
  author = {Pawel Sobkowicz and Antoni Sobkowicz},
  title = {Two-Year Study of Emotion and Communication Patterns in a Highly Polarized Politic
  journal = {Social Science Computer Review},
  volume = {30},
  number = {4},
  pages = {448-469},
}
```

```
year = {2012},
doi = {10.1177/0894439312436512}
}
```

## Dataset id: pl\_polemo

- Domain: **reviews**
- Language: **pl**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_pl_polemo,
  title = "Multi-Level Sentiment Analysis of {P}ol{E}mo 2.0: Extended Corpus of Multi-Domain",
  author = "Koco{'n}, Jan and
    Mi{'l}kowski, Piotr and
    Za{'s}ko-Zieli{'n}ska, Monika",
  booktitle = "Proceedings of the 23rd Conference on Computational Natural Language Learning",
  month = nov,
  year = "2019",
  address = "Hong Kong, China",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/K19-1092",
  doi = "10.18653/v1/K19-1092",
  pages = "980--991"
}
```

## Dataset id: pl\_twitter\_sentiment

- Domain: **social\_media**
- Language: **pl**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**

- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}
```

## Dataset id: pt\_dai\_labor

- Domain: **social\_media**
- Language: **pt**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_dai_labor,
  author = {Narr, Sascha and Michael Hülfenhaus and Albayrak, Sahin},
  title = {Language-Independent Twitter Sentiment Analysis},
  booktitle = {Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012)},
  year = {2012},
  location = {Dortmund, Germany},
}
```

## Dataset id: pt\_ifeel

- Domain: **social\_media**
- Language: **pt**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_dai_labor,
  author = {Narr, Sascha and Michael Hülfehaus and Albayrak, Sahin},
  title = {Language-Independent Twitter Sentiment Analysis},
  booktitle = {Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML-2012)},
  year = {2012},
  location = {Dortmund, Germany},
}
```

## Dataset id: pt\_tweet\_sent\_br

- Domain: **social\_media**
- Language: **pt**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@inproceedings{dataset_pt_tweet_sent_br,
  title = "Building a Sentiment Corpus of Tweets in {B}razilian {P}ortuguese",
  author = "Brum, Henrico and
    Volpe Nunes, Maria das Gra{\c{c}}as",
  booktitle = "Proceedings of the Eleventh International Conference on Language Resources and Evaluation",
  month = may,
  year = "2018",
}
```

```

address = "Miyazaki, Japan",
publisher = "European Language Resources Association (ELRA)",
url = "https://aclanthology.org/L18-1658",
}

```

## Dataset id: pt\_twitter\_sentiment

- Domain: **social\_media**
- Language: **pt**
- Language family: **Indo-European**
- Genus: **Romance**
- Definite articles: **definite word distinct from demonstrative**

- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```

@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}

```

## Dataset id: ru\_sentiment

- Domain: **social\_media**
- Language: **ru**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**

- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_ru_sentiment,
  title = "{R}u{S}entiment: An Enriched Sentiment Analysis Dataset for Social Media in {R}us",
  author = "Rogers, Anna and
    Romanov, Alexey and
    Rumshisky, Anna and
    Volkova, Svitlana and
    Gronas, Mikhail and
    Gribov, Alex",
  booktitle = "Proceedings of the 27th International Conference on Computational Linguistics",
  month = aug,
  year = "2018",
  address = "Santa Fe, New Mexico, USA",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/C18-1064",
  pages = "755--763",
  abstract = "This paper presents RuSentiment, a new dataset for sentiment analysis of social media",
}
```

## Dataset id: ru\_twitter\_sentiment

- Domain: **social\_media**
- Language: **ru**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
}
```

```

    volume = {11},
    url = {https://doi.org/10.1371/journal.pone.0155036},
    pages = {1-26},
    number = {5},
}

```

## Dataset id: sk\_twitter\_sentiment

- Domain: **social\_media**
- Language: **sk**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative affix**
- Polar questions: **interrogative word order**
- Position of negative word wrt SOV: **MorphNeg**
- Prefixing vs suffixing: **weakly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```

@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}

```

## Dataset id: sl\_sentinews

- Domain: **news**
- Language: **sl**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**



- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@Article{Bučar2018,
  author={Bu{\v{c}}ar, Jo{\v{z}}e
    and {\v{Z}}nidar{\v{s}}i{\v{c}}, Martin
    and Povh, Janez},
  title={Annotated news corpora and a lexicon for sentiment analysis in Slovene},
  journal={Language Resources and Evaluation},
  year={2018},
  month={Sep},
  day={01},
  volume={52},
  number={3},
  pages={895-919},
  abstract={In this study, we introduce Slovene web-crawled news corpora with sentiment anno
  issn={1574-0218},
  doi={10.1007/s10579-018-9413-3},
  url={https://doi.org/10.1007/s10579-018-9413-3}
}
```

## Dataset id: **sl\_twitter\_sentiment**

- Domain: **social\_media**
- Language: **sl**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **6-7**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
```

```

    volume = {11},
    url = {https://doi.org/10.1371/journal.pone.0155036},
    pages = {1-26},
    number = {5},
}

```

## Dataset id: sq\_twitter\_sentiment

- Domain: **social\_media**
- Language: **sq**
- Language family: **Indo-European**
- Genus: **Albanian**
- Definite articles: **definite affix**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **4**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```

@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}

```

## Dataset id: sr\_movie\_reviews

- Domain: **reviews**
- Language: **sr**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **5**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**

- Polar questions: **question particle**
- Position of negative word wrt SOV: **other**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@inproceedings{dataset_sr_serb_movie_reviews,
  title = "Reliable Baselines for Sentiment Analysis in Resource-Limited Languages: The {S}e",
  author = "Batanovi{'c}, Vuk and
    Nikoli{'c}, Bo{'v{s}}ko and
    Milosavljevi{'c}, Milan",
  booktitle = "Proceedings of the Tenth International Conference on Language Resources and E",
  month = may,
  year = "2016",
  address = "Portoro{'v{z}}, Slovenia",
  publisher = "European Language Resources Association (ELRA)",
  url = "https://aclanthology.org/L16-1427",
  pages = "2688--2696",
  abstract = "Collecting data for sentiment analysis in resource-limited languages carries a
}
```

## Dataset id: sr\_senticommments

- Domain: **reviews**
- Language: **sr**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **5**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **other**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_sr_senticommments,
  doi = {10.1371/journal.pone.0242050},
  author = {Batanović, Vuk AND Cvetanović, Miloš AND Nikolić, Boško},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {A versatile framework for resource-limited sentiment articulation, annotation, an},
  year = {2020},
  month = {11},
  volume = {15},
  url = {https://doi.org/10.1371/journal.pone.0242050},
  pages = {1-30},
}
```

```
abstract = {Choosing a comprehensive and cost-effective way of articulating and annotating  
number = {11},  
}
```

## Dataset id: sr\_twitter\_sentiment

- Domain: **social\_media**
- Language: **sr**
- Language family: **Indo-European**
- Genus: **Slavic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **5**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **other**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine, neuter**

```
@article{dataset_twitter_sentiment,  
  doi = {10.1371/journal.pone.0155036},  
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},  
  journal = {PLOS ONE},  
  publisher = {Public Library of Science},  
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},  
  year = {2016},  
  month = {05},  
  volume = {11},  
  url = {https://doi.org/10.1371/journal.pone.0155036},  
  pages = {1-26},  
  number = {5},  
}
```

## Dataset id: sv\_twitter\_sentiment

- Domain: **social\_media**
- Language: **sv**
- Language family: **Indo-European**
- Genus: **Germanic**
- Definite articles: **definite affix**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **2**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **interrogative word order**

- Position of negative word wrt SOV: **more than one position**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **common, neuter**

```
@article{dataset_twitter_sentiment,
  doi = {10.1371/journal.pone.0155036},
  author = {Mozetič, Igor AND Grčar, Miha AND Smailović, Jasmina},
  journal = {PLOS ONE},
  publisher = {Public Library of Science},
  title = {Multilingual Twitter Sentiment Classification: The Role of Human Annotators},
  year = {2016},
  month = {05},
  volume = {11},
  url = {https://doi.org/10.1371/journal.pone.0155036},
  pages = {1-26},
  number = {5},
}
```

## Dataset id: th\_wisesight\_sentiment

- Domain: **social\_media**
- Language: **th**
- Language family: **Tai-Kadai**
- Genus: **Kam-Tai**
- Definite articles: **no article**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative auxiliary verb**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **little affixation**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **noun classifiers**

```
@misc{dataset_th_wisesight_sentiment,
  author = {Suriyawongkul, Arthit and
    Chuangsuwanich, Ekapol and
    Chormai, Pattarawat and
    Polpanumas, Charin},
  title = {PyThaiNLP/wisesight-sentiment: First release (v1.0)},
  month = sep,
  year = 2019,
  publisher = {Zenodo},
  version = {v1.0},
  doi = {10.5281/zenodo.3457447},
  url = {https://doi.org/10.5281/zenodo.3457447},
  note = {Zenodo}
```

```
}
```

## Dataset id: th\_wongnai\_reviews

- Domain: **reviews**
- Language: **th**
- Language family: **Tai-Kadai**
- Genus: **Kam-Tai**
- Definite articles: **no article**
- Indefinite articles: **indefinite word distinct from one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative auxiliary verb**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **little affixation**
- Coding of nominal plurality: **mixed morphological plural**
- Grammatical genders: **noun classifiers**

```
@misc{dataset_th_wongnai_reviews,  
  author = {Ekkalak Thongthanomkul and Tanapol Nearunchorn and Yuwat Chuesathuchon},  
  title = {wongnai-corpus},  
  year = {2019},  
  publisher = {GitHub},  
  journal = {GitHub repository},  
  howpublished = {\url{https://github.com/wongnai/wongnai-corpus}}  
}
```

## Dataset id: ur\_roman\_urdu

- Domain: **mixed**
- Language: **ur**
- Language family: **Indo-European**
- Genus: **Indic**
- Definite articles: **no article**
- Indefinite articles: **no article**
- Number of cases: **2**
- Order of subject, object, verb: **SOV**
- Negative morphemes: **negative affix**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SONegV**
- Prefixing vs suffixing: **strongly suffixing**
- Coding of nominal plurality: **plural suffix**
- Grammatical genders: **masculine, feminine**

```
@InProceedings{dataset_ur_roman_urdu,  
  title = "Performing Natural Language Processing on Roman Urdu Datasets",
```

```

author    = "Zareen Sharf and Saif Ur Rahman",
booktitle = "International Journal of Computer Science and Network Security",
volume    = "18",
pages     = "141-148",
year      = "2018",
url       = {http://paper.ijcsns.org/07_book/201801/20180117.pdf}
}

```

## Dataset id: zh\_hotel\_reviews

- Domain: **reviews**
- Language: **zh**
- Language family: **Sino-Tibetan**
- Genus: **Chinese**
- Definite articles: **no article**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**
- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **little affixation**
- Coding of nominal plurality: **no plural**
- Grammatical genders: **noun classifiers**

```

@inproceedings{dataset_zh_hotel_reviews,
  title = "An Empirical Study on Sentiment Classification of {C}hinese Review using Word Emb
  author = "Lin, Yiou  and
           Lei, Hang  and
           Wu, Jia  and
           Li, Xiaoyu",
  booktitle = "Proceedings of the 29th Pacific Asia Conference on Language, Information and
  month = oct,
  year = "2015",
  address = "Shanghai, China",
  url = "https://aclanthology.org/Y15-2030",
  pages = "258--266",
}

```

## Dataset id: zh\_multilan\_amazon

- Domain: **reviews**
- Language: **zh**
- Language family: **Sino-Tibetan**
- Genus: **Chinese**
- Definite articles: **no article**
- Indefinite articles: **indefinite word same as one**
- Number of cases: **no morphological case-making**

- Order of subject, object, verb: **SVO**
- Negative morphemes: **negative particle**
- Polar questions: **question particle**
- Position of negative word wrt SOV: **SNegVO**
- Prefixing vs suffixing: **little affixation**
- Coding of nominal plurality: **no plural**
- Grammatical genders: **noun classifiers**

```
@inproceedings{dataset_multilan_amazon,
  title = "The Multilingual {A}mazon Reviews Corpus",
  author = {Keung, Phillip and
    Lu, Yichao and
    Szarvas, Gy{"o"}rgy and
    Smith, Noah A.},
  booktitle = "Proceedings of the 2020 Conference on Empirical Methods in Natural Language P
  month = nov,
  year = "2020",
  address = "Online",
  publisher = "Association for Computational Linguistics",
  url = "https://aclanthology.org/2020.emnlp-main.369",
  doi = "10.18653/v1/2020.emnlp-main.369",
  pages = "4563--4568",
}
```