

# References and Links

This chapter compiles relevant references and provides links to projects where the lemmatization of Slovene has been developed and applied to Slovene texts.

## **Projects, in which the system has been developed:**

[JOS - Linguistic Annotation of Slovene: Methods and Resources](#)

[Communication in Slovene](#)

[Janes - Resources, Tools and Methods for the Research of Nonstandard Internet Slovene](#)

[Development of Slovene in a Digital Environment](#)

## **Training corpora containing manually revised lemmas:**

Arhar Holdt, Špela; Krek, Simon; Dobrovoljc, Kaja; Erjavec, Tomaž; Gantar, Polona; Čibej, Jaka; Pori, Eva; Terčon, Luka; Munda, Tina; Žitnik, Slavko; Robida, Nejc; Blagus, Neli; Može, Sara; Ledinek, Nina; Holz, Nanika; Zupan, Katja; Kuzman, Taja; Kavčič, Teja; Škrjanec, Iza; Marko, Dafne; Jezeršek, Lucija; Zajc, Anja, 2022, Training corpus SUK 1.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1747>.

Lenardič, Jakob; Čibej, Jaka; Arhar Holdt, Špela; Erjavec, Tomaž; Fišer, Darja; Ljubešič, Nikola; Zupan, Katja; Dobrovoljc, Kaja, 2022, CMC training corpus Janes-Tag 3.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1732>.

## **References:**

Erjavec, Tomaž; Fišer, Darja; Krek, Simon in Ledinek, Nina. 2010. The JOS Linguistically Tagged Corpus of Slovene. V: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valeta, Malta, Maj. European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2010/pdf/139\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/139_Paper.pdf) [PDF]

Erjavec, Tomaž. 2012. MULTEXT-East: morphosyntactic resources for Central and Eastern European languages. Language Resources and Evaluation, 46(1): 131–142. DOI: [10.1007/s10579-011-9174-8](https://doi.org/10.1007/s10579-011-9174-8) [PDF]

Krek, Simon; Erjavec, Tomaž; Dobrovoljc, Kaja; Gantar, Polona; Arhar Holdt, Špela; Čibej, Jaka in Brank, Janez. The ssj500k training corpus for Slovene language processing. V: Fišer, D. in Erjavec, T. Jezikovne tehnologije in digitalna humanistika: zbornik konference: 24.-25. september 2020, Ljubljana, Slovenija. Ljubljana: Inštitut za novejšo zgodovino, 2020. Str. 24–33. [http://nl.ijs.si/jtdh20/pdf/JT-DH\\_2020\\_Krek-et-al\\_The-ssj500k-Training-Corpus-for-Slovene-Language-Processing.pdf](http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Krek-et-al_The-ssj500k-Training-Corpus-for-Slovene-Language-Processing.pdf) [PDF]

---

Revision #6  
Created 13 November 2023 13:57:45 by Tina Munda  
Updated 14 November 2023 22:16:12 by Tina Munda