# Towards an Ideal Tool for Learner Error Annotation

**Špela Arhar Holdt**[1,2], **Tomaž Erjavec**[3], **Iztok Kosem**[1,2], **Elena Volodina**[4]
[1]Faculty of Arts, University of Ljubljana, Slovenia
[2]Faculty of Computer and Information Science, University of Ljubljana, Slovenia
[3]Dept. of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia
[4]Språkbanken Text, University of Gothenburg, Sweden
{spela.arharholdt, iztok.kosem}@ff.uni-lj.si
tomaz.erjavec@ijs.si, elena.volodina@svenska.gu.se

**Abstract**

Annotation and analysis of corrections in learner corpora have always presented technical challenges, mainly on account of the fact that until now there has not been any standard tool available, and that original and corrected versions of texts have been mostly stored together rather than treated as individual texts. In this paper, we present CJVT Svala 1.0, the Slovene version of the SVALA tool, which was originally used for the annotation of Swedish learner language. The localisation into Slovene resulted in the development of several new features in SVALA such as the support for multiple annotation systems, localisation into other languages, and the support for more complex annotation systems. Adopting the parallel aligned approach to text visualisation and annotation, as well as storing the data, combined with the tool supporting this, i.e. SVALA, are proposed as new standards in Learner Corpus Research.

**Keywords:** learner corpora, corpus building, Svala

## 1. Introduction

Learner corpora are samples of texts produced by learners of a language, e.g. native, second or foreign language, and usually contain so-called error annotation, sometimes referred to as 'correction annotation'.

As such, correction of errors produces a second version of the original essays – naturally, a corrected one. However, very few projects have acknowledged that fact by treating error-corrected learner texts as aligned parallel versions of the original texts, some rare examples being Reznicek et al. (2012); Rosen et al. (2014); Boyd et al. (2014); Volodina et al. (2019). Most widely-used approach until today consists in adding corrections directly to the erroneous token within the same original sentence (e.g. Tenfjord et al., 2006; Granger, 2008; Mendes et al., 2016; Glisic and Ingason, 2022).

While discussions are still ongoing as to the 'ideal' approaches to learner corpus annotation (e.g. Stemle et al., 2019), we argue in this paper that original and corrected versions of learner texts should be treated as individual texts, i.e. as translations from one to another, and should be represented as aligned parallel texts – partly due to the user-friendliness of this approach, and partly due to the fact that this way we give credit to the corrected version as an independent interpretation of a learner text. We also argue that this approach requires reconsidering a number of choices: among others, the choice of a tool for manual annotation, the choice of data structures for representation of linked parallel data and – probably, most challenging – the adaptation of parallel data structures to fit into the widely accepted corpus search interfaces based on corpus workbench principles. Finally, we propose the solutions developed in our project as new standards for the field of Learner Corpus Research.

Below, we describe the context in which the work has been carried out with respect to annotation tools (Section 2); introduce the SVALA tool (Wirén et al., 2019) developed in the SweLL project (Volodina et al., 2019) and the way we have modified it to the needs of our project (Section 3); present the data format and the technological solution for the search environment (Section 4); summarize the user experience and value for the community (Section 5); and outline future steps (Section 6).

## 2. Background and Related Work

Learner corpora development is a highly cross-disciplinary enterprise, and requires attention to the demands of the involved disciplines. For one, we expect the corpus to be manually annotated by linguists or researchers from the field of Second Language Acquisition (SLA) – which sets demands on the user-friendliness of the tool to ensure consistent annotations, ultimately making it faster and more streamlined for annotators to complete their tasks. Second, the output of the work in step one should be compatible with search interfaces that systems engineers are dealing with – which requires the data format to be technologically sound. Third, the end-users from humanities should be able to interact with the search interfaces in a user-friendly way without any requirements of extensive



*LREC-COLING 2024*, pages 16392–16398

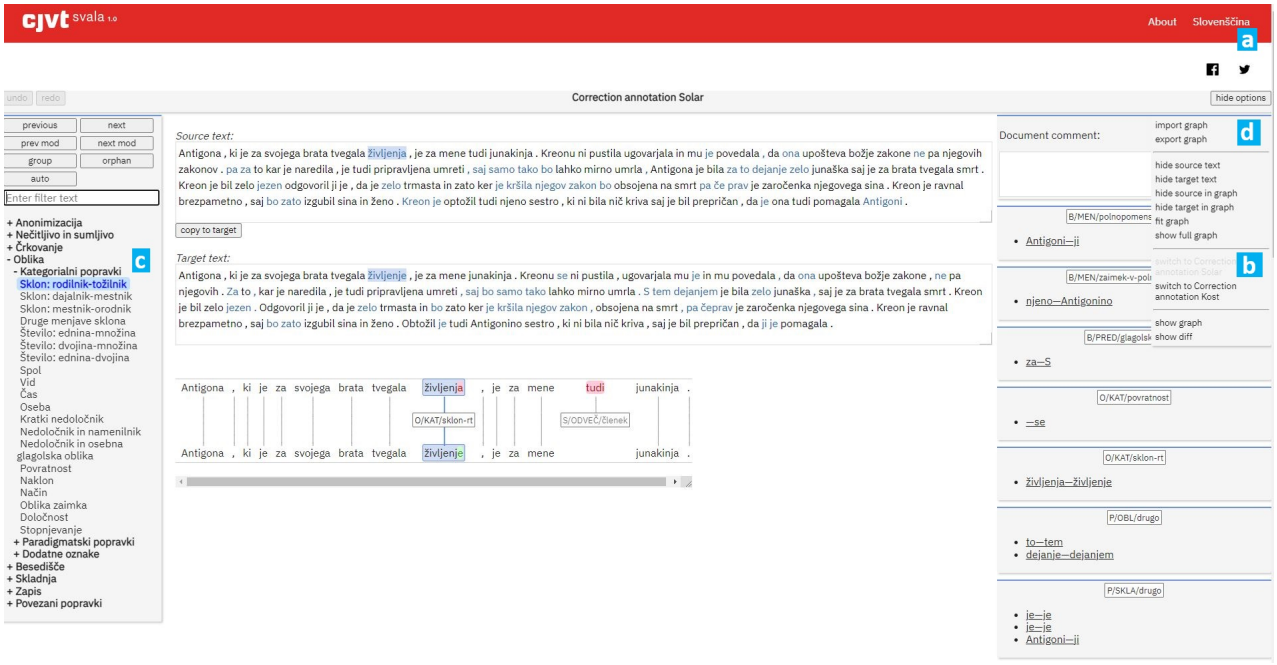


Figure 1: The interface of CJVT Svala 1.0 with highlighted novelties.

technological awareness or competence – here, visualization of the data is a priority alongside reliable conversion between data formats. Finally, the end-users from Natural Language Processing or other technology-oriented fields should be able to use the data in machine-intensive approaches – which, again, sets demands on the data formats and, specifically, on the consistency of annotations.

In an ideal world, all of these demands should be given equal attention and fair treatment, starting from the annotation tool. In the real world, however, we frequently see examples where user-friendliness of an annotation tool is overlooked for the sake of retaining formats that systems engineers can easily work with (e.g. Tenfjord et al., 2006; Lüdeling et al., 2005; Glisic and Ingason, 2022). One such example is the adapted version of the Sketch Engine tool, which has been used for the analysis of Cambridge Learner Corpus[1] and the first two versions of the Slovenian corpus of student writing Šolar. Over the years, user-friendliness of XML-based annotation tools has been improved (e.g. Obeid et al., 2013; Janssen, 2016), however, the standard assumption that the corrections are a feature of the original text remained.

The conceptual shift towards viewing the corrected version as an individual text in its own right came first in projects FALKO (Reznicek et al., 2012), MERLIN (Boyd et al., 2014), CzeSL (Rosen et al., 2014), later followed in SweLL (Volodina et al., 2019) and LaVa (Darģis et al., 2020). Of these projects, two have implemented annotation tools that offer a way to visualize word order changes in a user-friendly way, namely the tool `feat` (Hana et al., 2014) used for annotation of CzeSL, Czech learner language, and `SVALA` (Wirén et al., 2019) used for annotation of SweLL, Swedish learner language.

Given our intention to work with the learner corpus of Slovene in a parallel fashion, we focused in particular on annotation tools aimed at parallel representation in learner corpus context, of which `SVALA` has proven to be the best one. `SVALA` offers both a possibility to work in a text environment in a usual way – which has been highly appreciated by the linguists, - as well as automatically builds aligment on a token level between the two versions, saving a lot of manual effort (Rosén et al., 2018; Wirén et al., 2019). Besides, the `SVALA` output format is JSON, which is easy to work with in automatic applications. However, JSON requires conversion to representations acceptable in corpus search interfaces, which is discussed in Section 4.

## 3. CJVT SVALA 1.0

For Slovene, three resources incorporating language corrections can be mentioned: the Šolar corpus, comprising texts written by primary and secondary school students together with teacher corrections (Kosem et al., 2016); the Lektor corpus, comprising texts by adult native speakers together with proofreader corrections (Popič, 2014); and the KOST corpus with texts by learners of Slovene as

[1] https://www.cambridge.org/sketch/help/userguides/Using%20the%20Learner%20Corpus%201.1.pdf

a second/foreign language and corrections made by the corpus creators (Stritar Kučuk, 2022). Until recently, assembling these corpora proved to be a rather intricate and time-consuming process. For example, (Arhar Holdt and Kosem, 2023) illustrate the array of methodological challenges faced during the development of the Šolar corpus from version 1.0 to 3.0.

The opportunity to refine the corpus building methodology presented itself within the framework of the national project "Development of Slovene in a Digital Environment" (2020–23),[2] where our objective was to streamline the development of crucial language resources for contemporary Slovene. Acknowledging the SVALA tool as a state-of-the-art solution due to its open accessibility, rich functionalities, and user-friendly design, we were keen to integrate it into our workflows. Choosing to customize the tool to meet Slovene requirements, we incorporated the modules for transcription, basic anonymization, and error/correction annotation. The automated anonymization and annotation workflow management were deemed less urgent for our current needs and postponed for future endeavours.

The result is CJVT SVALA that can be accessed online [3], and the upgraded programme code was added to the SVALA fork on GitHub under an open licence (Anonymised Github). While adapting SVALA for the needs of the Slovene community, we also kept in mind the wider use of the tool. The new features that can be benefited from, are:

a) **Multiple languages:** The programme now has the option to support localisation to multiple languages. For CJVT SVALA, we chose to make the interface available both in Slovene and English. We added Slovene translations, but other languages can be easily introduced.

b) **Multiple annotation systems:** The programme now has the option to include multiple annotation systems and the ability to switch between them through the interface. For the annotation of corrections in different Slovene corpora, different annotation systems are used. Currently, CJVT SVALA supports two annotation systems (for corpora Šolar and KOST), while new ones can be easily added.

c) **Complex annotation systems:** We introduced the option to group correction labels together and to navigate through them in the part of the interface where the appropriate correction label is selected and assigned. This supports a user-friendly use of complex annotation systems, such as the annotation system of the corpus Šolar that comprises 180 correction labels, arranged into three hierarchical levels (Arhar Holdt et al., 2022b). Moreover, the labels in the menu can now be presented with longer, more intuitive names, which eases their search and selection.

d) **Added file saving:** We transitioned from backend connections to a server to saving and loading files on a local computer. For this feature, we also included the display of the text name in the interface as well as in the exported file, supporting the annotation workflows and versioning of the annotated data.

e) **Other modifications and simplifications:** CJVT SVALA is adapted so that it can be used independently of the infrastructure that was otherwise available for Swedish (namely the Swell portal). Certain additional simplifications were introduced, for example, we re-arranged the "show options" dropdown, removed interactive links to graphs, and downgraded the interactive user help to a regular, descriptive one.

Figure 1 showcases the CJVT SVALA interface, highlighting the aforementioned enhancements with corresponding alphabet letters. In the central part of the interface, two paragraphs of transcribed text are presented – the *Source text*, composed by a student, and the *Target text*, incorporating language corrections. Beneath the texts lies a graph connecting both versions, utilizing colors for clarity: errors in the source text are marked in red, while corrections are displayed in green. Clicking on a link between the original and corrected word or phrase allows the user to assign a label to the specific language correction. Labels can be selected from the left-side menu or via the search box above it. Positioned above the menu are buttons offering navigation to the previous/next link, the previous/next modification, and for manually grouping or ungrouping linked words.

## 4. Output Format

As mentioned, SVALA outputs the results of the annotation in JSON, in particular, as three JSON files, one containing the original text, the other the edited text, and the third one links between the tokens or segments of the first two files with each link accompanied by the code of the type of the correction performed.

To enable further processing of corpora (in particular Šolar (Arhar Holdt et al., 2022a) and KOST (Stritar Kučuk et al., 2023)) and the use of corpus data within the tools at our disposal, we also developed a conversion procedure that transforms the JSON format into XML, using the Text Encoding Initiative (TEI) Guidelines[4] with the TEI customisation

[2] https://rsdo.slovenscina.eu/en

[3] https://orodja.cjvt.si/svala/

[4] https://tei-c.org/release/doc/tei-p5-doc/en/html

as recommended by CLARIN.SI.[5] The corpus is encoded as one XML document composed of the overall TEI root containing the TEI header giving the (mostly manually inserted but some automatically produced) corpus meta-data and XIncludes of the three parts of the corpus, i.e. as in JSON, the file with original text, the file with corrected text and the file with links (expressed as TEI link groups) between the two together with the correction codes. The files with the original and corrected text also contain the metadata of the individual text. These two files were then also automatically annotated with the CLASSLA-Stanza annotation pipeline (Ljubešić and Dobrovoljc, 2019), adding information about each token's lemma, MULTEXT-East morphosyntactic description (Erjavec, 2012), Universal Dependencies part-of-speech and morphological features (de Marneffe et al., 2021), and the syntactic dependencies according to the Slovenian JOS formalism of syntactic dependencies (Erjavec et al., 2010). In the Appendix, an example of a TEI encoded sentence is presented.

## 5. Value for the Community and Current Limitations

The value of the CJVT SVALA tool for users, particularly annotators, has been underscored through an evaluation conducted as part of preparing the KOST Slovene learner corpus (Stritar Kučuk, 2022). This assessment involved 39 third-year Slovene university students enrolled in the elective course Slovene as a Second and Foreign Language at the Faculty of Arts, University of Ljubljana. The aim was to determine the tool's intuitiveness and its effectiveness in facilitating language annotation with minimal prior training. Participants, after a brief training session, annotated 172 texts. The findings revealed that a negligible fraction of annotation errors were attributed to the SVALA application itself, highlighting its user-friendly design (Stritar Kučuk, 2023). The participants furthermore appreciated the opportunity for a practical application of their linguistic knowledge in a technologically accessible manner. This feedback attests to SVALA's intuitive design, affirming its role as a beneficial tool for both novice and seasoned annotators.

The SVALA tool's open license and adaptable code can be leveraged for different languages and corpora. First, one needs to translate the interface to the desired language. Next, the annotation system (or multiple systems) that will be used for the specific project needs to be integrated into the code. Once the tool has been adapted for a new language or corpus and published online, it supports user-friendly annotation, with the capability to export annotated data. These can then be linguistically tagged with selected tools and, at the end, formatted to the proposed XML TEI format, ensuring that the resulting corpora are ready for research and other purposes.

By following these steps, the SVALA tool can be utilized for annotating corpora in various languages and for different purposes, enriching the fields of linguistics, natural language processing, and other related disciplines with valuable data on language corrections. However, there are identifiable shortcomings that, once addressed, could enhance the international applicability. Firstly, the current setup process, although transparent, requires a degree of technical proficiency that might deter potential users lacking in programming expertise. Streamlining this aspect through more automated setup procedures or detailed, step-by-step guidance could lower the entry barrier for a wider user base. Another challenge lies in the tool's untested compatibility with non-Latin scripts, a crucial area for expansion to ensure true global utility. Addressing this would necessitate testing across a variety of writing systems and possible technical adaptations to guarantee seamless functionality. Additionally, the process of localizing the interface and integrating new annotation systems, while feasible, demands certain effort and time. Simplifying these tasks through predefined (interface) templates could expedite the adaptation process for new languages.

## 6. Conclusions

We have presented the Slovene localised version of the SVALA tool, which had been originally developed for the annotation of Swedish learner language. The Slovene version has confirmed not only the usefulness of SVALA as an annotation tool for learner language but also the methodological approach to data storage supported by the tool. The benefits of such international collaboration between research teams with similar needs and challenges are also evidenced by the fact that the Slovene version has introduced several new features that will benefit the future users of SVALA.

Based on our past experience with annotation of language corrections, and non-existence of similar open-source tools for annotation, we propose that this approach of treating original learner texts and their corrected versions as aligned parallel texts should become a standard in Learner Corpus Research, and in other corpora containing corrections. In that vein, we would urge the Learner Corpus community to adopt SVALA as the tool used for annotation and analysis, and help expand its functionalities rather than continue developing tailor-made (in-house) solutions.

[5] https://github.com/clarinsi/TEI-schema.

The top priority of our current work is to enhance the existing pipeline with a specialized corpus concordancer that can fully leverage the new structure and the rich metadata and annotations present in corpora featuring error annotations. While the new format has made integration into popular concordancers like noSketch Engine (Rychlỳ, 2007) and KonText (Machálek, 2020) easier, the search and visualization of corpus data remain constrained to separate views, either in the original texts or the corrected versions. We are thus developing a user-friendly tool that comprehensively presents both text versions, enables searches for both the original and corrected forms, and provides advanced statistics on language corrections. Such a tool will be of immense value to researchers, as well as textbook authors, curriculum designers, and educators, ultimately leading to the creation of empirically-based teaching materials and improvements in the language learning process.

## 7. Acknowledgements

Work on the Swedish SVALA has been supported by a research grant from the Swedish Riksbankens Jubileumsfond '*SweLL – research infrastructure for Swedish as a second language*', grant IN16-0464:1. Work on the article for the last author has been supported by Nationella språkbanken and HUMINFRA, both funded by the Swedish Research Council (contracts 2017-00626 and 2021-00176) and their participating partner institutions. For the Slovene team, the authors acknowledge the financial support of the Slovenian Research and Innovation Agency through the project *Empirical foundations for digitally-supported development of writing skills*(J7-3159), the programme *Language Resources and Technologies for Slovene* (P6-0411) and the research infrastructure *CLARIN.SI*.

## 8. Bibliographical References

Špela Arhar Holdt and Iztok Kosem. 2023. Šolar, the developmental corpus of slovene. Preprint, doi: 10.21203/rs.3.rs-3274669/v1.

Špela Arhar Holdt, Tadeja Rozman, Mojca Stritar Kučuk, Simon Krek, Irena Krapš Vodopivec, Marko Stabej, Eva Pori, Teja Goli, Polona Lavrič, Cyprian Laskowski, Polonca Kocjančič, Bojan Klemenc, Luka Krsnik, and Iztok Kosem. 2022a. *Developmental corpus Šolar 3.0*. Centre for Language Resources and Technologies, University of Ljubljana / Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1589.

Špela Arhar Holdt, Polona Lavrič, Robert Roblek, and Teja Goli. 2022b. Categorizing teachers' corrections: Guidelines for annotating the Šolar corpus. version 1. Technical report, Project "Development of Slovene in a Digital Environment". Retrieved October 10th, 2023, from https://wiki.cjvt.si/books/11-developmental-corpus-solar/page/annotation-guidelines.

Adriane Boyd, Jirka Hana, Lionel Nicolas, Detmar Meurers, Katrin Wisniewski, Andrea Abel, Karin Schöne, Barbora Štindlová, and Chiara Vettori. 2014. The MERLIN corpus: Learner Language and the CEFR. In *LREC'14*, Reykjavik, Iceland. European Language Resources Association (ELRA).

Roberts Darģis, Ilze Auzi na, Kristine Levane-Petrova, and Inga Kaija. 2020. Detailed Error Annotation for Morphologically Rich Languages: Latvian Use Case. In *Human Language Technologies – The Baltic Perspective*, pages 241–244. IOS Press.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Tomaž Erjavec. 2012. MULTEXT-East: Morphosyntactic Resources for Central and Eastern European Languages. *Language Resources and Evaluation*, 46(1):131–142.

Tomaž Erjavec, Darja Fišer, Simon Krek, and Nina Ledinek. 2010. The JOS Linguistically Tagged Corpus of Slovene. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Isidora Glisic and Anton Karl Ingason. 2022. The nature of Icelandic as a second language: An insight from the learner error corpus for Icelandic. In *CLARIN Annual Conference*, pages 23–33.

Sylviane Granger. 2008. Learner Corpora. In Anke Lüdeling and Merja Kytö, editors, *Corpus Linguistics. An International Handbook*, volume 1, chapter 15, pages 259–275. Mouton de Gruyter, Berlin.

Jirka Hana, Alexandr Rosen, Barbora Štindlová, and Jan Štěpánek. 2014. Building a learner corpus. *Language resources and evaluation*, 48:741–752.

Maarten Janssen. 2016. TEITOK: Text-faithful annotated corpora. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4037–4043.

Iztok Kosem, Tadeja Rozman, Špela Arhar Holdt, Polonca Kocjančič, and Cyprian Laskowski. 2016. Šolar 2.0: nadgradnja korpusa šolskih pisnih izdelkov. In *Zbornik konference Jezikovne tehnologije in digitalna humanistika*, pages 95–100.

Nikola Ljubešić and Kaja Dobrovoljc. 2019. What does Neural Bring? Analysing Improvements in Morphosyntactic Annotation and Lemmatisation of Slovenian, Croatian and Serbian. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 29–34, Florence, Italy. Association for Computational Linguistics.

Anke Lüdeling, Maik Walter, Emil Kroymann, and Peter Adolphs. 2005. Multi-level error annotation in learner corpora. In *Proceedings of corpus linguistics*, volume 1, pages 14–17. Citeseer.

Tomáš Machálek. 2020. Kontext: Advanced and flexible corpus query interface. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 7003–7008.

Amália Mendes, Sandra Antunes, Maarten Janssen, and Anabela Gonçalves. 2016. The COPLE2 corpus: a learner corpus for Portuguese. In *LREC'16*.

Ossama Obeid, Wajdi Zaghouani, Behrang Mohit, Nizar Habash, Kemal Oflazer, and Nadi Tomeh. 2013. A web-based annotation framework for large-scale text correction. In *The Companion Volume of the Proceedings of IJCNLP 2013: System Demonstrations*, pages 1–4.

Damjan Popič. 2014. Revising translation revision in Slovenia. *New Horizons in Translation Research and Education 2*, page 72.

M. Reznicek, A. Lüdeling, C. Krummes, and F. Schwantuschke. 2012. *Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.0*. Humboldt-Universität zu Berlin, Berlin, Germany.

Alexandr Rosen, Jirka Hana, Barbora Štindlová, and Anna Feldman. 2014. Evaluating and Automating the Annotation of a Learner Corpus. *Lang. Resour. Eval.*, 48(1):65–92.

Dan Rosén, Mats Wirén, and Elena Volodina. 2018. Error Coding of Second-Language Learner Texts Based on Mostly Automatic Alignment of Parallel Corpora. In *CLARIN Annual Conference 2018, Pisa, Italy, 8–10 October, 2018*, pages 181–184.

Pavel Rychlỳ. 2007. Manatee/bonito-a modular corpus manager. In *RASLAN*, pages 65–70.

Egon W Stemle, Adriane Boyd, Maarten Jansen, Therese Lindström Tiedemann, Nives Mikelić Preradović, Alexandr Rosen, Dan Rosén, and Elena Volodina. 2019. Working together towards an ideal infrastructure for language learner corpora. *Widening the Scope of Learner Corpus Research*.

Mojca Stritar Kučuk. 2022. Kost med korpusi usvajanja tujega jezika. *Na stičišču svetov*, pages 323–334.

Mojca Stritar Kučuk, Helena Šter, Staša Pisek, Ivana Petric Lasnik, Jana Kete Matičič, Nataša Pirih Svetina, Daniela Preglau, Špela Arhar Holdt, Luka Krsnik, Tomaž Erjavec, Jasmina Pegan, and Damjan Huber. 2023. *Slovene learner corpus KOST 2.0*. Centre for Language Resources and Technologies, University of Ljubljana / Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1887.

Mojca Stritar Kučuk. 2023. Error annotation in Slovene learner corpus KOST - why L1 students can(not) do the job. In *Abstracts from CLARC 2023–Language and Language Data International Scientific Conference, Rijeka, Croatia, 28-30 September, 2023*. Faculty of Humanities and Social Sciences University of Rijeka.

Kari Tenfjord, Paul Meurer, and Knut Hofland. 2006. The ASK corpus: A language learner corpus of Norwegian as a second language. In *LREC'06*, pages 1821–1824.

Elena Volodina, Lena Granstedt, Arild Matsson, Beáta Megyesi, Ildikó Pilán, Julia Prentice, Dan Rosén, Lisa Rudebeck, Carl-Johan Schenström, Gunlög Sundberg, et al. 2019. The SweLL language learner corpus: From design to annotation. *Northern European Journal of Language Technology (NEJLT)*, 6:67–104.

Mats Wirén, Arild Matsson, Dan Rosén, and Elena Volodina. 2019. Svala: Annotation of second-language learner text based on mostly automatic alignment of parallel corpora. In *CLARIN Annual Conference, Pisa, Italy, 8-10 October, 2018*, pages 222–234. Linköping University Electronic Press.

## Appendix: TEI Encoding Example

This appendix gives an example of a TEI encoded sentence snippet from the Šolar 3.0 corpus (Arhar Holdt et al., 2022a), which was converted from the SVALA JSON format.

The teacher crossed out a stretch of two words (where the error types were classified as "superfluous adverbs" and "pleonasm") in the sentence below. Units that translate from single words into phrases are indicated by square brackets.

Seveda je smešen tudi ta ~~težko oz.~~ skoraj neverjeten razplet sodbe. (Slovene)

'[Of course] [it is] funny also this ~~difficult i.e.~~ almost incredible finale [of the verdict].'

```
<s xml:id="solar8s.8.3">
  <w ana="mte:L" msd="UPosTag=PART"
     lemma="seveda" xml:id="solar8s.8.3.1">Seveda</w>
  ...
  <w ana="mte:O" msd="UPosTag=X|Abbr=Yes"
     lemma="oz."
     xml:id="solar8s.8.3.7">oz.</w>
  <w ana="mte:L" msd="UPosTag=PART"
     lemma="skoraj"
     xml:id="solar8s.8.3.8">skoraj</w>
  <w ana="mte:Ppnmein" msd="UPosTag=ADJ|Case=Nom|..."
     lemma="neverjeten"
     xml:id="solar8s.8.3.9">neverjeten</w>
  <w ana="mte:Somei" msd="UPosTag=NOUN|Case=Nom|..."
     lemma="razplet"
     xml:id="solar8s.8.3.10">razplet</w>
  <w ana="mte:Sozer" msd="UPosTag=NOUN|Case=Gen|..."
     lemma="sodba"
     xml:id="solar8s.8.3.11"
     join="right">sodbe</w>
  <pc ana="mte:U" msd="UPosTag=PUNCT"
      xml:id="solar8s.8.3.12">.</pc>
  <linkGrp corresp="#solar8s.8.3"
   targFunc="head argument" type="JOS-SYN">
    <link ana="jos-syn:modra"
     target="#solar8s.8.3 #solar8s.8.3.1"/>
    ...
    <link ana="jos-syn:modra"
     target="#solar8s.8.3 #solar8s.8.3.12"/>
  </linkGrp>
</s>
...
<s xml:id="solar8t.8.3">
  <w ana="mte:L" msd="UPosTag=PART"
     lemma="seveda"
     xml:id="solar8t.8.3.1">Seveda</w>
  ...
  <w ana="mte:L" msd="UPosTag=PART"
     lemma="skoraj"
     xml:id="solar8t.8.3.6">skoraj</w>
  <w ana="mte:Ppnmein" msd="UPosTag=ADJ|Case=Nom|..."
     lemma="neverjeten"
     xml:id="solar8t.8.3.7">neverjeten</w>
  <w ana="mte:Somei" msd="UPosTag=NOUN|Case=Nom|..."
     lemma="razplet"
     xml:id="solar8t.8.3.8">razplet</w>
  <w ana="mte:Sozer" msd="UPosTag=NOUN|Case=Gen|..."
     lemma="sodba"
     xml:id="solar8t.8.3.9"
     join="right">sodbe</w>
  <pc ana="mte:U" msd="UPosTag=PUNCT"
      xml:id="solar8t.8.3.10">.</pc>
  <linkGrp corresp="#solar8t.8.3"
   targFunc="head argument" type="JOS-SYN">
    <link ana="jos-syn:modra"
     target="#solar8t.8.3 #solar8t.8.3.1"/>
    ...
  </linkGrp>
</s>

<linkGrp type="CORR" targFunc="orig corr"
  corresp="#solar8s.8.3 #solar8t.8.3">
  <link type="ID"
   target="#solar8s.8.3.1 #solar8t.8.3.1"/>
  <link type="ID"
   target="#solar8s.8.3.2 #solar8t.8.3.2"/>
  ...
  <link
   type="S/DOD/pleonazem|S/ODVEČ/prislov-drugo"
   target="#solar8s.8.3.6 #solar8s.8.3.7"/>
  ...
  <link type="ID"
   target="#solar8s.8.3.12 #solar8t.8.3.10"/>
</linkGrp>
```