Annotation Management Tool: A Requirement for Corpus Construction

Yousuf Ali Mohammed, Arild Matsson, Elena Volodina
University of Gothenburg, Sweden
name.surname1.surname2@svenska.gu.se

Abstract

We present an annotation management tool, SweLL portal, that has been developed for the purposes of the SweLL infrastructure project for building a learner corpus of Swedish (Volodina et al., 2019). The SweLL portal has been used for supervised access to the database, data versioning, import and export of data and metadata, statistical overview, administration of annotation tasks, monitoring of annotation tasks and reliability controls. The development of the portal was driven by visions of longitudinal sustainable data storage and was partially shaped by situational needs reported by portal users, including project managers, researchers, and annotators.

1 Introduction

During 2017–2021, we were setting up the foundation for empirically based research on Swedish as a second language. The results were released in 2021 under the name of SweLL infrastructure, as a part of Nationella Språkbanken and Swedish CLARIN.[1] The core work entailed collecting and manually annotating learner written essays, the SweLL-gold corpus (Volodina et al., 2019). However, this process turned out to be more complex and involved a lot of work "behind the scenes". *First*, to make sure the annotations are reliable, we invested extensive work into developing and documenting a taxonomy of corrections (or errors, a more traditional term used in other projects) (Rudebeck and Sundberg, 2021) and a taxonomy of personally identifiable information for successful pseudonymisation (Megyesi et al., 2021). *Second*, to ensure the consistency of the manual annotation, we developed a tool to support the annotation itself, namely the Svala annotation tool (Wirén et al., 2019) and tool for the management of the annotation process, the SweLL portal. *Third*, to make sure the resulting collection of essays can reach the intended user, we worked on the legal aspects of access to the material as well as on the visualisation of the corpus so that it may be browsed and analyzed statistically based on textual, educational and linguistic characteristics.

From the above follows that an infrastructure project dealing with the construction and annotation of an electronic learner corpus entails the collection of data and metadata, followed by the meticulous selection of essays for manual annotation to ensure the balance and representativity of various metadata (e.g. the balance between texts of different genres and topics, between the writer's gender and education level, etc.) and the annotation itself. There are four pillars that tend to be named in connection to digital infrastructures: data, tools for data annotation, tools for data exploration, and expertise (Volodina et al., 2016).[2] What is usually overlooked is some project management environment.

Fort (2016) and (Hovy and Lavid, 2010) emphasise the need for an annotation management software that would ensure the reliability of manual annotations. There are two main reasons for that: *First*, a corpus of good quality must boast representativeness of the language it embodies and balance of the samples that characterise the language. This requires monitoring the collected text instances with regards to the various types of metadata. *Second*, the data as such is only the first step, the most interesting

[1] https://spraakbanken.gu.se/en/projects/swell

[2] https://spraakbanken.gu.se/projekt/swell

Yousuf Ali Mohammed, Arild Matsson and Elena Volodina 2022. Annotation Management Tool: A Requirement for Corpus Construction. *Selected papers from the CLARIN Annual Conference 2021. Ed. by Monica Monachini and Maria Eskevich.* Linköping Electronic Conference Proceedings 189, pp. 101–108. DOI: https://doi.org/10.3384/9789179294441

research can be done when the data is annotated for one or another text- or language-related feature, and this annotation should be reliable. 'Tools decay, data stay' is only true when data is reliably annotated.

The debate on the quality of data annotation often goes in the direction of (1) *tag sets* – their size and ambiguity, (2) *guidelines* – their clarity and degree of detail, and (3) *tools used for annotation* – their user-friendliness and support in annotation. The annotation management as such – database handling, statistical overviews, inter-annotator agreement controls, etc. – is often overlooked or simply not considered in time. This, then, results in an annotation project being managed using Excel files, which leads to errors, imbalance, loss of annotation or information and ultimately to the reduction of annotation quality (Stemle et al., 2019).

A number of data management tools have been developed in different projects. Most of them, however, were initially developed to support manual annotation, but with time added some functionality for database communication and versioning. Some examples of those are `TEITOK` (Janssen, 2016), `WebAnno` (de Castilho et al., 2014), the UAM Corpus Tool (O'Donnell, 2008). Such tools combine data annotation with data management, which makes them task-oriented and reduces their flexibility in the choice of an annotation paradigm. For example, in the case of `TEITOK`, the annotation is performed using xml TEI format, which may or may not be an optimal format, even though the data management functionality might satisfy a new project. Creating a universal tool that would satisfy any project (i.e. 'one size fits all') is no simple task. Due to the outlined considerations, we opted to develop our own tools, separating data management from data annotation.

The *SweLL-gold corpus* that we have been constructing over the past several years is aimed at researchers, developers and teachers to promote the fields of Second Language Acquisition (SLA), Language Assessment (LA), Intelligent Computer-Assisted Language Learning (ICALL) and Language Technology approaches to those – predominantly within CLARIN and other European (due to the GDPR restrictions) user groups. Due to this, a high standard of annotation is required. The `SweLL portal` is one of the steps to ensure those standards. Looking back at our experiences and analyzing the benefits of an annotation management tool, we can say that its use has helped in more ways than just corpus preparation. Among others, we have tested uploading other (bonus) learner corpora to the portal, and exporting them from the portal applying a unified set of metadata attributes and values (using 'N/A' as a value for absent attributes). This step has helped us make several Swedish learner corpora interoperable with each other, *interoperability* being a known challenge in CLARIN-related context (König et al., 2021; Stemle et al., 2019; Volodina et al., 2018).

The `SweLL portal` is deployed on the university servers at Språkbanken Text, Sweden, and only permits the storage of psudonymised text data according to GDPR regulations. Due to this fact, we are restrictive about allowing free access to the `SweLL portal` for other users who may incidentally upload personally identifiable text data. Only approved users are added to the portal. The code for the portal is available at a GitHub repository[3] for users in need of a data management tool.

Below, we describe the architecture of the `SweLL portal`, from data management to data import and export, and outline some current developments and future plans.

2 Data Management

The `SweLL portal` is a user-friendly tool for metadata and data collection and for annotation management. The three modules (*datacollection, task_manager, annotation*) in Figure 1 are loosely dependent on one another, so that another hypothetical project might replace only the *datacollection* and/or *annotation* modules with their custom implementations.

The *datacollection* module contains the SweLL metadata model. It is an interface that communicates with the database, where one can store, access and manage *metadata* about learners, tasks, schools and individual texts. New metadata records can be created following a standard form and stored in the metadata container. The metadata container consists of four objects:

1. *Source* stores information about the school where the essays have been collected. A *Source* is represented by a school ID, the type of education and the course type.

[3] https://github.com/spraakbanken/swell-portal

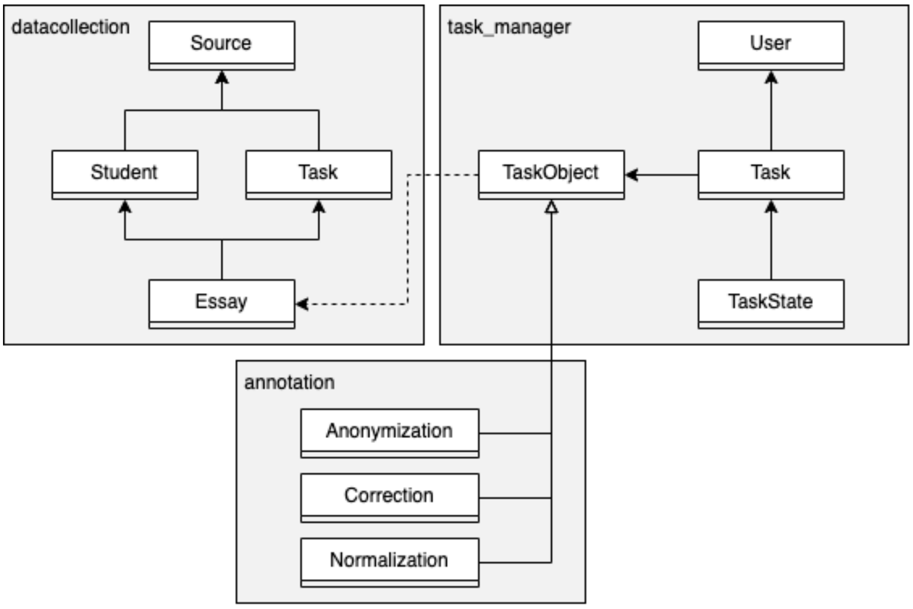
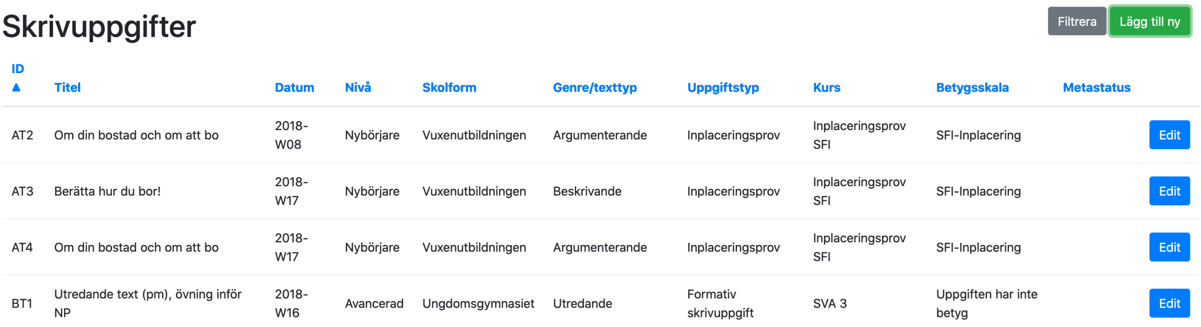


Figure 1: A partial class diagram of the application data model.

2. *Student* stores information about students, including student IDs, and their structured socio-demographic information about gender, mother tongue(s), education, etc.

3. *Task* stores information about the task learners have received for essay writing, including descriptive information about the genre, topic, grading system, allowed time, etc.

4. *Essay* metadata is a record created as a response by a *Student* to a *Task*, which also stores information about the individual performance on an essay. (Essay texts are not stored here, but in *TaskObject* and *TaskState* objects, see below.)



Skrivuppgifter

ID ▲	Titel	Datum	Nivå	Skolform	Genre/texttyp	Uppgiftstyp	Kurs	Betygsskala	Metastatus	
AT2	Om din bostad och om att bo	2018-W08	Nybörjare	Vuxenutbildningen	Argumenterande	Inplaceringsprov	Inplaceringsprov SFI	SFI-Inplacering		Edit
AT3	Berätta hur du bor!	2018-W17	Nybörjare	Vuxenutbildningen	Beskrivande	Inplaceringsprov	Inplaceringsprov SFI	SFI-Inplacering		Edit
AT4	Om din bostad och om att bo	2018-W17	Nybörjare	Vuxenutbildningen	Argumenterande	Inplaceringsprov	Inplaceringsprov SFI	SFI-Inplacering		Edit
BT1	Utredande text (pm), övning inför NP	2018-W16	Avancerad	Ungdomsgymnasiet	Utredande	Formativ skrivuppgift	SVA 3	Uppgiften har inte betyg		Edit

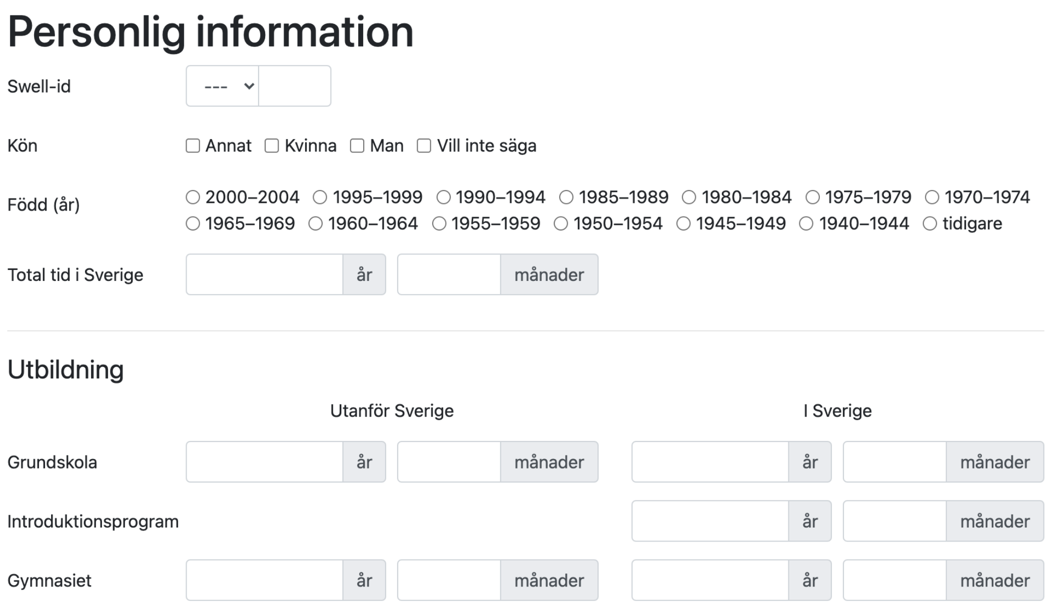
Figure 2: A list of *Task* metadata records.

In the user interface, under the *Metadata* tab, it is possible to get an overview of all items in each of the four objects mentioned above (e.g. Figure 2[4]). For each object, there exists an option to filter the metadata based on different attributes, to open existing records for editing and to add new records (e.g. Figure 3).

3 Annotation Task Management

The annotation task management is based on two modules in Figure 1, namely the *annotation* module and the *task manager* module.

[4] Text in the Figures is predominantly in Swedish.



Personlig information

Swell-id --- ▾

Kön □ Annat □ Kvinna □ Man □ Vill inte säga

Född (år) ○ 2000–2004 ○ 1995–1999 ○ 1990–1994 ○ 1985–1989 ○ 1980–1984 ○ 1975–1979 ○ 1970–1974 ○ 1965–1969 ○ 1960–1964 ○ 1955–1959 ○ 1950–1954 ○ 1945–1949 ○ 1940–1944 ○ tidigare

Total tid i Sverige år månader

Utbildning

	Utanför Sverige		I Sverige	
Grundskola	år	månader	år	månader
Introduktionsprogram			år	månader
Gymnasiet	år	månader	år	månader

Figure 3: A record for *Student* personal metadata.

The *annotation* module consists of three tasks: Anonymisation,[5] Normalisation and Correction Annotation. Each task type is defined through the workflow control that opens them in an external annotation tool, SVALA (Wirén et al., 2019), which is a stand-alone application. The SVALA code for all three tasks can be adapted to annotation of other languages than Swedish.

The *task_manager* module allows superusers to create, assign and manage annotation tasks using the *TaskObject* shown in Figure 1. A *TaskObject* pairs a task type with a specific essay, e.g. *anonymisation of essay A1AT2*. It is implemented with a generic reference to the essay metadata object, ensuring a loose module dependency. The *TaskObject* is a collection of three objects, *User, Task* and *TaskState*.

1. *User* is a record associated with an annotation expert performing the task. The record consists of an ID of a portal user who has been assigned the annotation task.

2. *Task* represents which annotation task is being performed (anonymisation, normalisation, correction annotation) and tracks a specific user's work on a *TaskObject*. If more than one user work on the same annotation task, they each have a separate *Task* with separate progress.

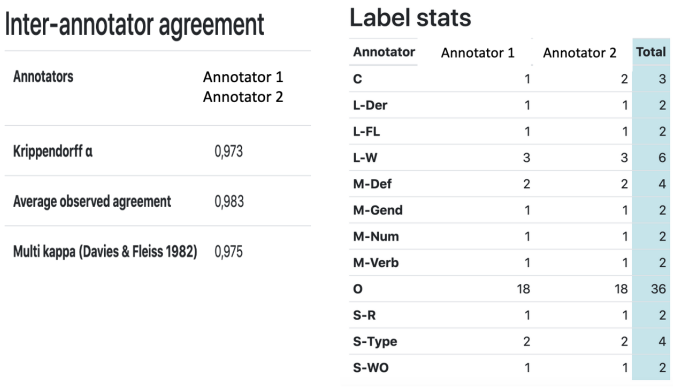
3. *TaskState* shows three states of work on a *Task*: assigned, started and completed, each of which can have a Boolean `yes/no` flag. Work in the annotation tool generates a sequence of *TaskStates*, each a snapshot version of the text plus annotations.

When a *User* starts an annotation *Task*, the essay opens in the external annotation tool SVALA (Wirén et al., 2019)[6]. A unique version of the essay is saved in the *TaskState* on every introduced change by the annotator in the SVALA tool. Once the annotation task is completed, the task can be marked as *Done*. This updates the status of the *TaskState* in the `SweLL portal`.

The functionality of the portal allows the superuser to assign the same Correction Annotation task to several users. When two or more versions of the same correction annotation task are completed, it is possible to measure Inter-Annotator Agreement (Figure 4). There is a possibility to click on the EssayID on the Annotations page to view the full text and to monitor the progress of the annotation.

[5] Later we switched from using the term *Anonymisation* to use the term *Pseudonymisation*.

[6] SVALA demo version: `https://spraakbanken.gu.se/swell/dev/`



Inter-annotator agreement

Annotators	Annotator 1 Annotator 2
Krippendorff α	0,973
Average observed agreement	0,983
Multi kappa (Davies & Fleiss 1982)	0,975

Label stats

Annotator	Annotator 1	Annotator 2	Total
C	1	2	3
L-Der	1	1	2
L-FL	1	1	2
L-W	3	3	6
M-Def	2	2	4
M-Gend	1	1	2
M-Num	1	1	2
M-Verb	1	1	2
O	18	18	36
S-R	1	1	2
S-Type	2	2	4
S-WO	1	1	2

Figure 4: Inter-annotator agreement for a particular essay and a pair of annotators.

4 Annotation Tool and its Demo Version

The SVALA tool (Wirén et al., 2019) is a stand-alone annotation tool that was developed with the aim of supporting the manual annotation work on learner essays in a user-friendly way. Annotators can use this tool for different annotation tasks such as pseudonymisation, normalisation and correction annotation. The essays are shown in the original and target versions, and as a ’spaghetti’ version. The spaghetti format maps the source text (*written by learners*) to the target text (*normalised by the annotator*) token by token and allows the annotators to add the labels (*pseudonymisation or correction*) to each edge in the graph. This tool also comes with an automatic pseudonymisation pipeline that can de-identify the personal information in an essay using rule-based methods.

SVALA demo version[7] is a copy of the SVALA tool (Wirén et al., 2019) that is publically available for anyone interested in testing the annotation of their datasets. This version is not connected to the `SweLL portal` or any database, and is used for demo-purposes and for viewing the full content in the essays through the Korp corpus search tool (Ahlberg et al., 2013). Full texts that are opened in the demo version of SVALA can be modified without any risk of sabotaging the annotations on the server, since these changes are not saved to the database.

5 Statistics

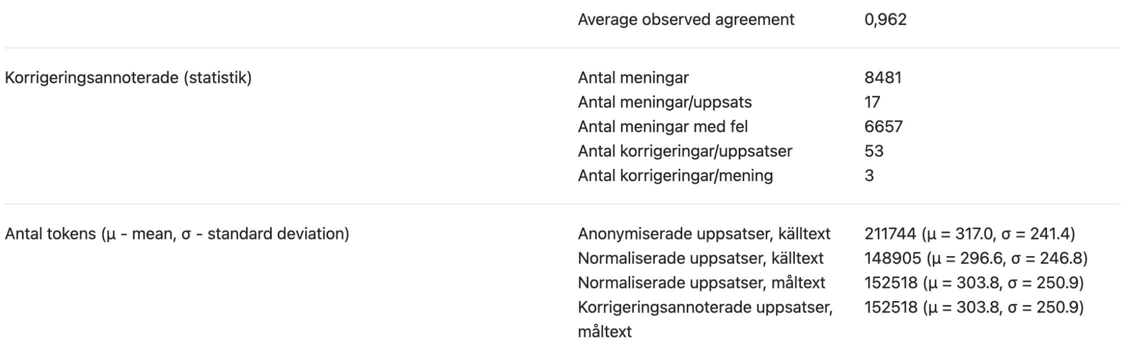
The statistics section shows an overview of the metadata and its frequencies, as well as frequencies over tokens and sentences in the *SweLL-gold corpus*. Described below are two ways of viewing the statistics:

1. *Statistics*: On this page, one can view the statistics of the pseudonymisation and correction labels together with the number of tokens, correct sentences and incorrect ones in the *SweLL-gold corpus* as shown in Figure 5. The statistics for attributes in student, task and essay metadata are also given on this page. There is an option to download the statistics as a CSV file.

2. *Summary*: Metadata for students, tasks and essays as well as annotation data can be filtered and viewed in a table format as shown in Figure 6. To have a better understanding of the data one can even view the tables summarised into graphs. The statistics can be downloaded for further work either in CSV, plain text or JSON formats. The summary page also has an option to visualise and monitor the progress of the annotation work for the project.

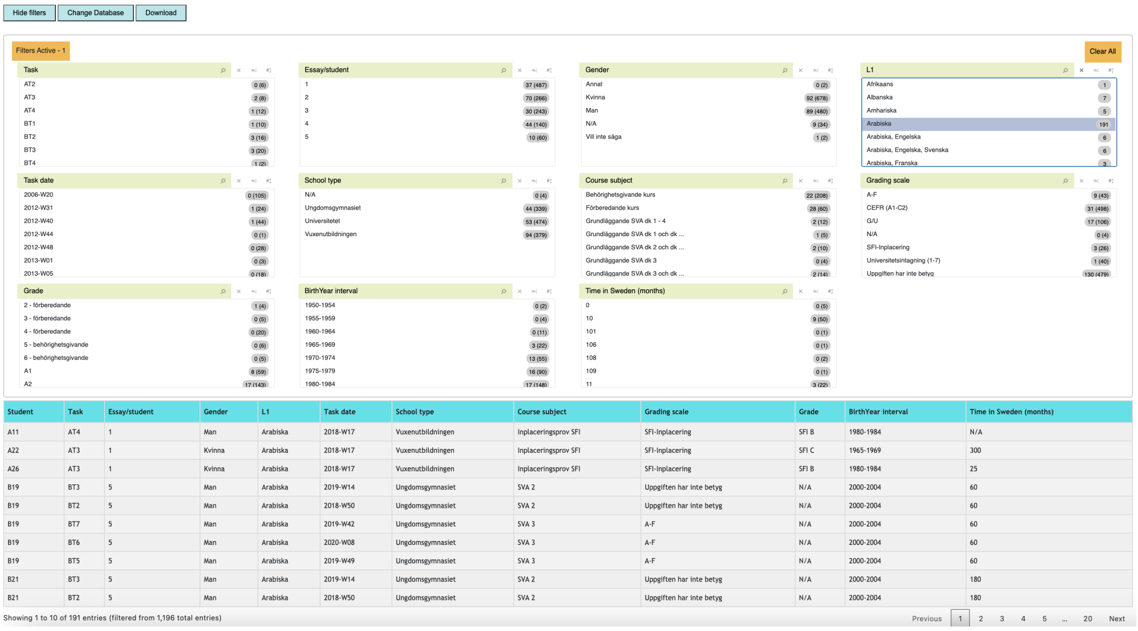
A decision has been made in the project not to implement advanced search and filters. Instead, a possibility is provided to download files with statistics that can be opened using Excel or processed through other programs that offer advanced filtering.

[7] SVALA demo version: https://spraakbanken.gu.se/swell/dev/



	Average observed agreement	0,962
Korrigeringsannoterade (statistik)	Antal meningar Antal meningar/uppsats Antal meningar med fel Antal korrigeringar/uppsatser Antal korrigeringar/mening	8481 17 6657 53 3
Antal tokens (µ - mean, σ - standard deviation)	Anonymiserade uppsatser, källtext Normaliserade uppsatser, källtext Normaliserade uppsatser, måltext Korrigeringsannoterade uppsatser, måltext	211744 (µ = 317.0, σ = 241.4) 148905 (µ = 296.6, σ = 246.8) 152518 (µ = 303.8, σ = 250.9) 152518 (µ = 303.8, σ = 250.9)

Figure 5: Running statistics over all material in the portal (excerpt).



Task	
AT2	0 (6)
AT3	2 (8)
AT4	1 (12)
BT1	1 (10)
BT2	3 (16)
BT3	3 (20)
BT4	1 (2)

Essay/student	
1	37 (487)
2	70 (266)
3	30 (243)
4	44 (140)
5	10 (60)

Gender	
Annat	0 (2)
Kvinna	92 (678)
Man	89 (480)
N/A	9 (34)
Vill inte säga	1 (2)

L1	
Afrikaans	1
Albanska	7
Amhariska	5
Arabiska	191
Arabiska, Engelska	6
Arabiska, Engelska, Svenska	6
Arabiska, Franska	3

Task date	
2006-W20	0 (105)
2012-W31	1 (24)
2012-W40	1 (44)
2012-W44	0 (1)
2012-W48	0 (28)
2013-W01	0 (3)
2013-W05	0 (18)

School type	
N/A	0 (4)
Ungdomsgymnasiet	44 (339)
Universitetet	53 (474)
Vuxenutbildningen	94 (379)

Course subject	
Behörighetsgivande kurs	22 (208)
Förberedande kurs	28 (60)
Grundläggande SVA dk 1 - 4	2 (12)
Grundläggande SVA dk 1 och dk ...	1 (5)
Grundläggande SVA dk 2 och dk ...	2 (10)
Grundläggande SVA dk 3	0 (4)
Grundläggande SVA dk 3 och dk ...	2 (14)

Grading scale	
A-F	9 (43)
CEFR (A1-C2)	31 (498)
G/U	17 (106)
N/A	0 (4)
SFI-Inplacering	3 (26)
Universitetsintagning (1-7)	1 (40)
Uppgiften har inte betyg	130 (479)

Grade	
2 - förberedande	1 (4)
3 - förberedande	0 (5)
4 - förberedande	0 (20)
5 - behörighetsgivande	0 (6)
6 - behörighetsgivande	0 (5)
A1	8 (59)
A2	17 (143)

BirthYear interval	
1950-1954	0 (2)
1955-1959	0 (4)
1960-1964	0 (11)
1965-1969	3 (22)
1970-1974	13 (55)
1975-1979	16 (90)
1980-1984	17 (148)

Time in Sweden (months)	
0	0 (5)
10	9 (50)
101	0 (1)
106	0 (1)
108	0 (2)
109	0 (1)
11	3 (22)

Student	Task	Essay/student	Gender	L1	Task date	School type	Course subject	Grading scale	Grade	BirthYear interval	Time in Sweden (months)
A11	AT4	1	Man	Arabiska	2018-W17	Vuxenutbildningen	Inplaceringsprov SFI	SFI-Inplacering	SFI B	1980-1984	N/A
A22	AT3	1	Kvinna	Arabiska	2018-W17	Vuxenutbildningen	Inplaceringsprov SFI	SFI-Inplacering	SFI C	1965-1969	300
A26	AT3	1	Kvinna	Arabiska	2018-W17	Vuxenutbildningen	Inplaceringsprov SFI	SFI-Inplacering	SFI B	1980-1984	25
B19	BT3	5	Man	Arabiska	2019-W14	Ungdomsgymnasiet	SVA 2	Uppgiften har inte betyg	N/A	2000-2004	60
B19	BT2	5	Man	Arabiska	2018-W50	Ungdomsgymnasiet	SVA 2	Uppgiften har inte betyg	N/A	2000-2004	60
B19	BT7	5	Man	Arabiska	2019-W42	Ungdomsgymnasiet	SVA 3	A-F	N/A	2000-2004	60
B19	BT6	5	Man	Arabiska	2020-W08	Ungdomsgymnasiet	SVA 3	A-F	N/A	2000-2004	60
B19	BT5	5	Man	Arabiska	2019-W49	Ungdomsgymnasiet	SVA 3	A-F	N/A	2000-2004	60
B21	BT3	5	Man	Arabiska	2019-W14	Ungdomsgymnasiet	SVA 2	Uppgiften har inte betyg	N/A	2000-2004	180
B21	BT2	5	Man	Arabiska	2018-W50	Ungdomsgymnasiet	SVA 2	Uppgiften har inte betyg	N/A	2000-2004	180

Showing 1 to 10 of 191 entries (filtered from 1,196 total entries)

Figure 6: Statistics with a filter function.

6 Data Import and Export

In SweLL portal, there are two different methods to import the essays into the database, namely, an XML-import and a raw text import. The XML import has an additional functionality to automatically create and insert the metadata for a *Student, Task*, and *Essay* on importing an XML file. The metadata attributes are part of the header tag in the XML file. In raw text import, the metadata for *Student, Task*, and *Essay* should be created manually in advance to correctly store the raw text. On completion of the import, successfully imported essays are ready for annotation in the task manager module.

The data export functionality gives users the ability to export the data based on a variety of selections as shown below.

- Annotation type - normalisation or correction annotation
- School - school ID (A,B,C...)
- Mother tongue - Arabic, English etc.
- Status - complete, incomplete or both

- Data type - source or target
- File type - XML, JSON or plain text.

There is a script for converting the working format based on JSON into XML that can be used for import of the data into the Korp search tool (Ahlberg et al., 2013).

Access to the `SweLL portal` is password-protected and only users with proper access rights can import and export the data. The work on documenting the `SweLL portal` is ongoing, with some documentation available from the SweLL project webpage.[8]

7 Future

The `SweLL portal` at the moment of release contained 502 essays with correction annotation and approx. 700 essays without correction annotation. In the past six months, the collection has grown with approximately 50 more essays which are in the process of manual normalisation and correction annotation. We expect it to grow further since we are currently collaborating with two international teams who are reusing our tools: a Slovenian team and a French team. Apart from enriching our data, the collaboration brings into spotlight aspects that we had not previously considered, including the expansion of metadata types, e.g. to cover individual differences (modern aptitude tests), more refined taxonomy for tasks (e.g. what linguistic parameters they stimulate, as a support for language assessment and cross-comparison of tasks), new taxonomy for content-based feedback and feedback on linguistic features/errors; etc.

We are also planning to provide a possibility for teachers/researchers to visualise the progress of a subcorpus (e.g. a Class) based on error correction labels. This will help them to identify the necessary focus for the learners, or to see whether or not the learners are making progress from one written task to the next.

Our further plans include:

1. Making the data statistics available for non-login users
2. Adding new types of users, e.g. teachers and classes
3. Creating subfolders for the data inside a project so different groups can work and collaborate
4. Creating a possibility for different projects to work independently from each other
5. Visualising learner/groups/class progress over time.

8 Concluding Remarks

In the current project, the aspects of data storage and annotation management have been taken seriously following the arguments outlined in (Fort, 2016) and (Hovy and Lavid, 2010) that stable annotation management is one of the important prerequisites for the creation of well-balanced and reliably annotated corpora. The portal development was incremental, with changes introduced in response to the needs of the project. Overall, both project researchers and project assistants were aided in their work through the `SweLL portal` functionalities. The `SweLL portal` will continue to be used for new learner corpus annotation projects as well as for statistical exploration of the material as a part of a newly developed SweLL infrastructure for second language research.

Given the architechture of the main components in the portal (Figure 1), some further level of abstraction can be added to it, so that the approach and the framework could be applied to a broader scope of corpus annotation projects. That would entail, for example, scenarios such as an abstracting database/*datacollection* module, a module for *annotation* types/states and a *task_manager* on the one hand; and adding flexibility for the integration of an external annotation tool, on the other; with an active visualisation module and statistics tool.

[8] https://spraakbanken.gu.se/en/projects/swell/swell-docs

Acknowledgements

This work has been supported by a research grant from the Swedish Riksbankens Jubileumsfond *SweLL - research infrastructure for Swedish as a second language*, IN16-0464:1, and by *Nationella språkbanken*, funded jointly by the Swedish Research Council (2018–2024, contract 2017-00626) and the 10 participating partner institutions. We also wish to thank the anonymous reviewers for their valuable comments on a previous version.

References

Ahlberg, M., Borin, L., Forsberg, M., Hammarstedt, M., Olsson, L.-J., Olsson, O., Roxendal, J., and Uppström, J. 2013. *Korp and Karp - a bestiary of language resources: the research infrastructure of Språkbanken*, 429–433. Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013).

de Castilho, R. E., Biemann, C., Gurevych, I., and Yimam, S. M. 2014. *WebAnno: a flexible, web-based annotation tool for CLARIN.* Proceedings of the CLARIN Annual Conference (CAC). Citeseer.

Fort, K. 2016. *Collaborative Annotation for Reliable Natural Language Processing: Technical and Sociological Aspects.* John Wiley & Sons.

Hovy, E. and Lavid, J. 2010. *Towards a 'science'of corpus annotation: a new methodological challenge for corpus linguistics*, 22(1):13–36. International journal of translation.

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

König, A., Frey, J.-C., and Stemle, E. W. 2021. *Exploring Reusability and Reproducibility for a Research Infrastructure for L1 and L2 Learner Corpora*, 12(5):199. Information. Multidisciplinary Digital Publishing Institute.

Megyesi, B., Rudebeck, L., and Volodina, E. 2021. *SweLL pseudonymization guidelines.* Technical report. GU-ISS Forskningsrapporter från Institutionen för svenska språket (2011-), University of Gothenburg.

O'Donnell, M. 2008. *Demonstration of the UAM CorpusTool for text and image annotation*, 13–16. Proceedings of the ACL-08: HLT Demo Session.

Rudebeck, L. and Sundberg, G. 2021. *SweLL correction annotation guidelines.* Technical report. GU-ISS Forskningsrapporter från Institutionen för svenska språket (2011-), University of Gothenburg.

Stemle, E. W., Boyd, A., Jansen, M., Lindström Tiedemann, T., Mikelić Preradović, N., Rosen, A., Rosén, D., Volodina, E., et al. 2019. *Working together towards an ideal infrastructure for language learner corpora.* Presses universitaires de Louvain.

Volodina, E., Megyesi, B., Wirén, M., Granstedt, L., Prentice, J., Reichenberg, M., and Sundberg, G. 2016. *A Friend in Need? Research agenda for electronic Second Language infrastructure.* Proceedings of Swedish Language Technology Conference (SLTC) 2016, Umeå, Sweden.

Volodina, E., Jansen, M., Stemle, E. W., Lindström Tiedemann, T., Mikelić Preradović, N., Ragnhildstveit, S. K., Tenfjord, K., and Koenraad, D. 2018. *Interoperability of Second Language Resources and Tools*, 90–94. Proceedings of the CLARIN Annual Conference 2018, Pisa, Italy.

Volodina, E., Granstedt, L., Matsson, A., Megyesi, B., Pilán, I., Prentice, J., Rosén, D., Rudebeck, L., Schenström, C.-J., Sundberg, G., and Wirén, M. 2019. *The SweLL Language Learner Corpus: From Design to Annotation.* Northern European Journal of Language Technology, Special Issue.

Wirén, M., Matsson, A., Rosén, D., and Volodina, E. 2019. *SVALA: Annotation of Second-Language Learner Text Based on Mostly Automatic Alignment of Parallel Corpora.* Linköping University Electronic Press. Proceedings of CLARIN 2018.