INSTITUTIONEN FÖR
SVENSKA SPRÅKET

GU-ISS-2021-02

SweLL pseudonymization guidelines

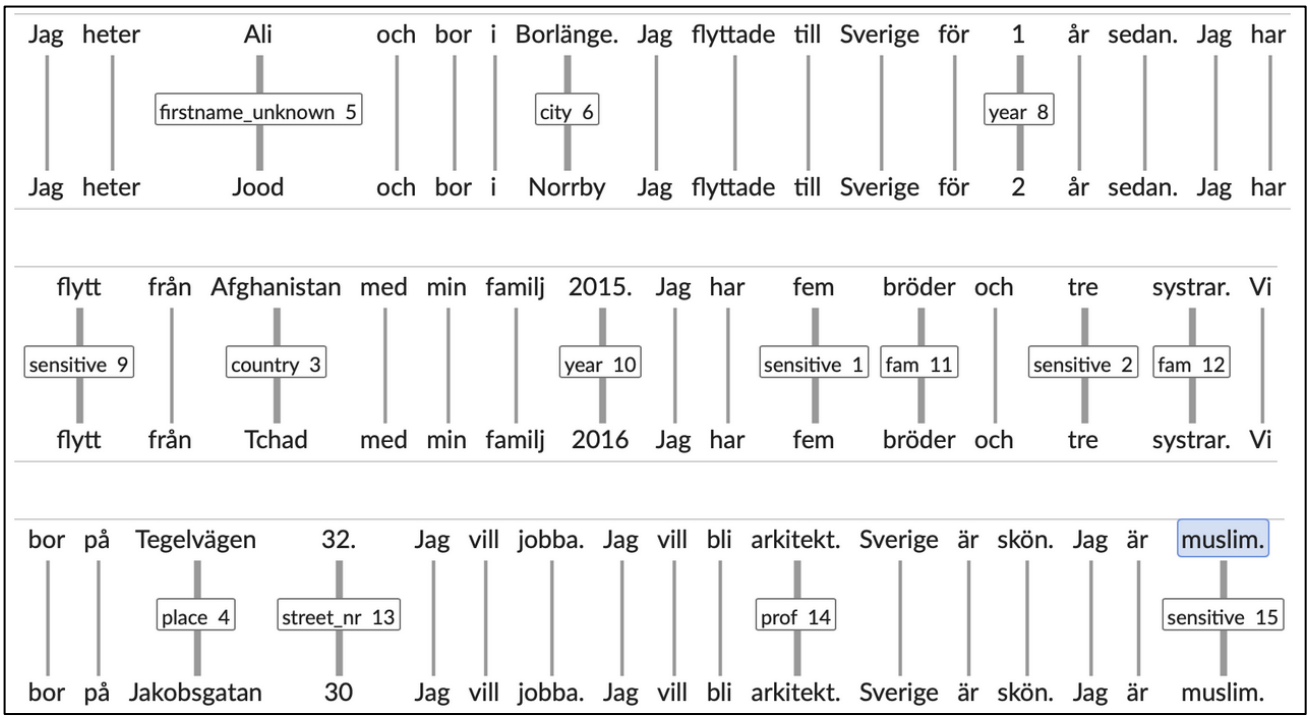
Beáta Megyesi, Lisa Rudebeck, Elena Volodina

Forskningsrapporter från institutionen för svenska språket, Göteborgs universitet
Research Reports from the Department of Swedish

ISSN 1401-5919

SweLL
Pseudonymization guidelines

by Beáta Megyesi, Lisa Rudebeck and Elena Volodina



August 2021

The SweLL guideline series:

SweLL Transcription guidelines
by Elena Volodina and Beáta Megyesi

SweLL Pseudonymization guidelines
by Beáta Megyesi, Lisa Rudebeck and Elena Volodina

SweLL Normalization guidelines
by Lisa Rudebeck, Gunlög Sundberg and Mats Wirén

SweLL Correction annotation guidelines
by Lisa Rudebeck and Gunlög Sundberg

Preface

by Elena Volodina, Lena Granstedt, Beáta Megyesi, Yousuf (Samir) Ali Mohammed, Julia Prentice, Lisa Rudebeck, Gunlög Sundberg and Mats Wirén

During years starting 2017-2021 we have been working on setting up the main building blocks for empirically based research on Swedish as a second language which we release under the name of the *SweLL infrastructure*. This work entailed collecting and manually annotating learner written essays, which we refer to as *SweLL-gold corpus*. However, this process turned out to be highly versatile and involved a lot of work "behind the scene". 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) and a taxonomy of personally identifiable information (PII, for successful pseudonymization). Second, to make sure that the manual annotation is as consistent as possible, we developed a set of tools to support the annotation itself and the management of the annotation process. Third, to make sure the resulting collection of essays can reach the intended user, we worked on legal aspects of access to the material as well as on visualization of the corpus so that it may be browsed and analyzed statistically, from the point of textual, educational and linguistic characteristics.

The current document is a part of the SweLL guidelines series consisting of four parts which aim to report how we have worked on the material and which decisions we have made. Guidelines are available for each step in the manual annotation process, including:

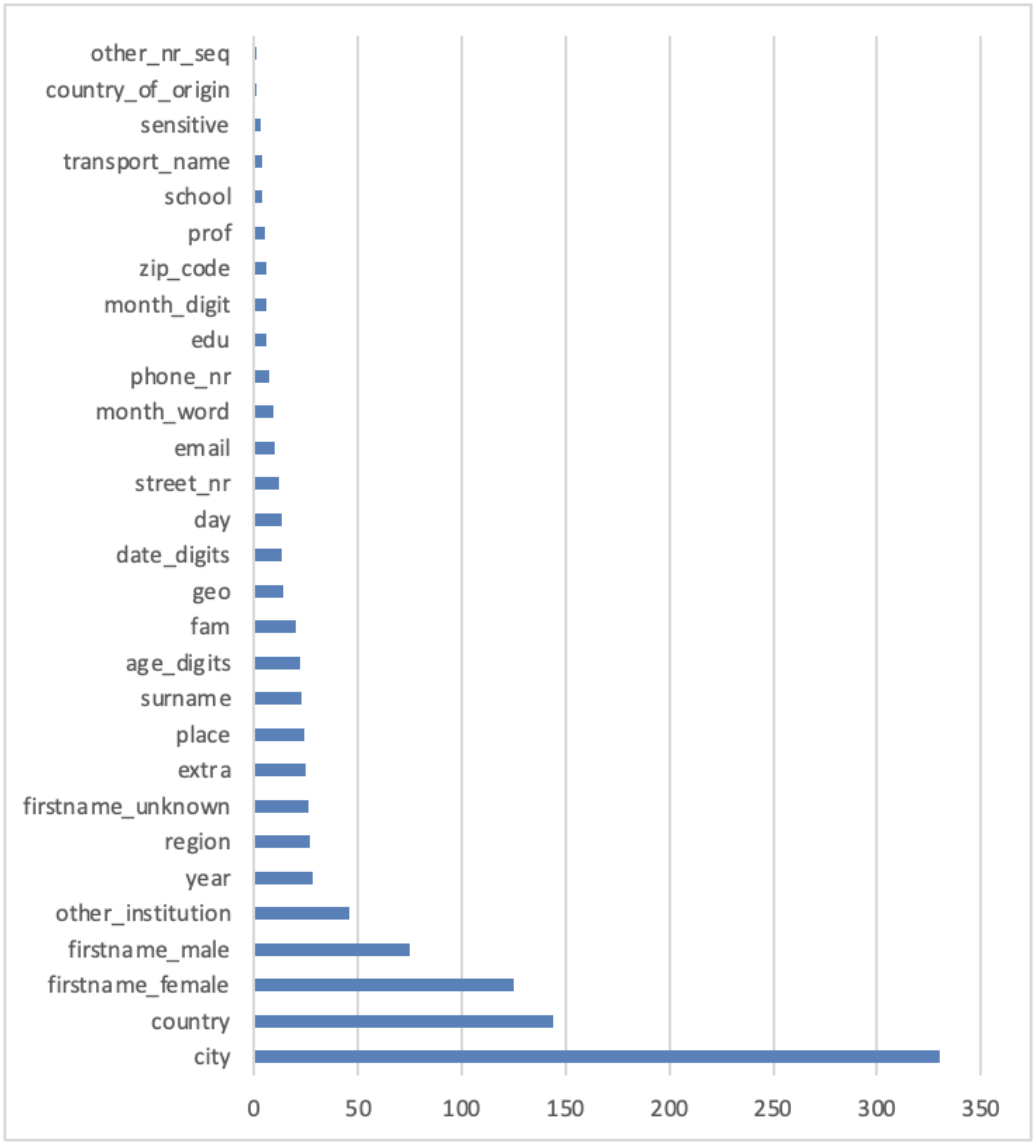
- Transcription guidelines
- Pseudonymization guidelines
- Normalization guidelines
- Correction annotation guidelines

We specifically described all processes in English to make sure our principles and experience can be of help to people working on other learner infrastructure projects independent of the language.

The pseudonymization work started with the definition of a taxonomy of Personally Identifiable Information (PII), summarized on the right in 19 descriptive categories.

Hard replacement (9 categories)
Age; Bank account; License numbers; Dates; E-mail; Phone number; Personal identity number; Url; Zip code
Placeholder (3 head categories)
Geo-data: country, region, city, street, area Institution: school, work, institution, other Personal names: female, male, neutral, surname
Sensitive markup (7 categories)
Education; Profession; Family members (Ethnical info; Political views; Religious views; Sexual info)

To give you an overview of the SweLL pseudonymization categories in the actual data, have a look at the graph below with PII tags and their frequency in the SweLL-gold data.



More information about the metadata used in the corpus and an overview of the taxonomies can be found here: https://spraakbanken.github.io/swell-release-v1/Metadata-SweLL

A short introduction to the SweLL project

SweLL - Swedish Learner Language – is a research infrastructure for Swedish as a second language. It was funded by Riksbankens Jubileumsfond 2017-2020 (IN16-0464:1), and had four participating universities: University of Gothenburg (project leadership), Stockholm University, Uppsala University and Umeå University.

The SweLL infrastructure project had as an aim to lay the fundament for digital Second Language Acquisition research by:
(1) collecting and manually annotating learner essays written by learners of Swedish at different levels of development
(2) developing well-functioning annotation principles, tagsets and processes, and thoroughly describing them
(3) developing and documenting digital tools for processing and storing of learner essays
(4) making the data and tools available through a portal developed for digital resources and tools for second language acquisition research of Swedish

The learner corpus infrastructure SweLL includes:

(1) The SweLL portal that is used for collection, storage and versioning of essays, administration of the annotation process, statistical overview, inter-annotator agreement, import and export of the data.
(2) The SweLL portal hosts a collection of more than 680 essays that have been digitized and manually transcribed from handwritten samples during the course of this project. All essays were pseudonymized to protect the privacy of each individual learner. A larger portion of the essays – 502 texts, the so-called SweLL-gold corpus – were normalized, i.e. re-written in order to fit the norms of standard Swedish by correcting erroneous and deviant language, and each correction was assigned a correction label describing the difference between the learner's version (source text) and the corrected version (target text).
(3) Several other tools are available for future users of the infrastructure:

- SVALA annotation tool for performing manual annotation steps (pseudonymization, normalization, correction annotation) (Wirén et al. 2019)
- Automatic pseudonymizer service (included as a part of the SVALA tool, and available through github for potential extensions or re-use in other projects) (Volodina et al. 2020)

(4) Extensive work was done to document how the learner data were processed, which includes

- selection and documentation of associated metadata (corpus-related, student-related, task-related, school-related and essay-related)
- taxonomies for pseudonymization and correction annotation, and
- guidelines for all (manual) annotation steps (transcription, pseudonymization, normalization and correction annotation)

(5) Thorough work has been carried out to make sure that the GDPR guidelines and ethical principles are followed. In consultation with the university lawyers at the University of Gothenburg, the access principles have been defined and legal basis double-checked. Access to essays can be granted following an application. As of 2021, according to the GDPR, users outside Europe cannot get immediate access to the data in its entirety. Their applications need to be processed by the

university lawyers on a case-to-case basis. Applicants inside EU can get access to the full dataset provided their intended use targets L2-oriented research, development or pedagogical applications.
(6) The data can be browsed through corpus search interface Korp (https://spraakbanken.gu.se/korp/) with specific solutions for L2-material facilitating filtering for e.g. texts written by writers of a certain age, gender, mother tongue, or writers at a certain proficiency level or course, a certain text type – all with a possibility for full-text view.

More information about the project and tools are available at the project page:
https://spraakbanken.gu.se/projekt/swell

Acknowledgments

Our gratitude goes to teachers and assessors who supported us during the essay collection stage. They were many, and without their enthusiasm we would not have been able to build any infrastructure today. We are grateful to learners who were positive to allow their texts to be used for teaching, research and development.

We would like to acknowledge a group of advisors, assistants and developers who are not listed as co-authors of this preface, but who have been involved during the different periods of the SweLL project:

- University of Gothenburg: Arild Matsson, Ildikó Pilán, Monica Reichenberg, Dan Rosén, Carl Johan Schenström
- Stockholm University: Sofia Brusling, Sofia Johansson, Miku Westerholm

We would also love to extend our gratitude for the generous financial support provided by the main funder Riksbankens Jubileumsfond, as well as for the indispensable financial support during the last months of finalizing the infrastructure from Språkbanken Text and Swe-Clarin at the University of Gothenburg as well as by the Department of Swedish Language and Multilingualism at Stockholm University.

August 2021

Elena Volodina, University of Gothenburg
Lena Granstedt, Umeå university
Beáta Megyesi, Uppsala university
Yousuf (Samir) Ali Mohammed, University of Gothenburg
Julia Prentice, University of Gothenburg
Lisa Rudebeck, Stockholm University
Gunlög Sundberg, Stockholm university
Mats Wirén, Stockholm university

swell-project

Pseudonymization guidelines

Beáta Megyesi, Lisa Rudebeck, Elena Volodina (June, 2018 – May 2019)

Online version of this document: https://spraakbanken.github.io/swell-project/Pseudonymization_guidelines

Contents

The purpose of pseudonymization is to de-identify all information that can reveal the identity of the person who wrote the text. This information can include person names, age, addresses and phone numbers, city names and other geographical names, etc.

On top of this, some information is also marked as “potentially sensitive” during the pseudonymization process. This is information which does not in itself disclose the identity of the writer, but which would be particularly harming to reveal were the identity of the writer to be disclosed in spite of the de-identifying efforts. Sensitive information is for instance information on political or religious views of the writer. The information marked as potentially sensitive will be reviewed before publication of the corpus to evaluate whether it needs to be hidden or not.

Your task as an assistant is 1) to identify all information that can relate to the specific person who wrote the text, and categorize what type of information it is so that the person can be de-identified by changing/hiding the specific information, and 2) to mark potentially sensitive information related to the writer. The replacement of the personal information is performed automaticaly given the assigned label.

This document contains instructions for how to proceed.

1. Basic principles

1. Remove/change the information that can reveal a person behind the essay(s), yet

keep to the *minimal change* rule. The data should be usable in research scenarios.

2. Data on *deviations from standard Swedish will be lost* for the pseudonymized strings (e.g. mis-spellings etc.). This also holds for text segments the form of which is dependent on the pseudonymized string (for instance prepositions preceeding pseudonymized city and country names, e.g. in Germany -> in Cuba).

3. Annotators have to make the *assessment of the risks and needs* for pseudonymization (an element of subjectivity).

4. Tokens should not be pseudonymized solely on the basis of them belonging to a specific category listed among the pseudonymization categories, but on the basis of them *potentially revealing the identity of the writer*. For instance, not all country or city names are pseudonymized, but only those which, **together with the context**, 1) may be connected to the writer (e.g. because the city may be identified as the writer's home town), and 2) reveal information which is specific enough to be used to identify the writer. Accordingly, in a text where Istanbul is mentioned as a city where the writer has lived or as a city where a family member of the writer lives (etc.), *Istanbul* should be pseudonymized. But not so in a text providing general information about Istanbul. And while the information that the writer stems from the Baltic countries may be reason to pseudonymize *Baltikum* (as a region), the information that the writer stems from Europe does not necessitate pseudonymization, since Europe is such a large region which may be assumed to be the home region for a large number of potential writers.

5. Keep *track* of whether the token is "original" or "masked". (This is done automatically by the annotating tool.)

6. Categories that need to be *marked in the texts, but not necessarily replaced*. An assessment should be made later when enough statistics is collected over the learners behind the essays , as well as the assembled texts and metadata on each particular writer:
    - country: the same pseudonymization tag, < country >, is used for:
        - country of origin (*Jag kommer från Syrien* versus *Jag kommer från Luxembourg*) - depending upon how many subjects in the database are from the named countries
        - country of "intermediate" residence (*Vi har stannat en månad i Turkiet*)
        - **Note:** Mentions of *Sweden* as a country of origin or residence are not marked.
    - number of family members (*Jag har fem bröder och fyra systrar* –> Eng. "I have five brothers and four sisters") - an estimation is necessary to see whether it is a normal pattern in many essays. If yes - no masking/suppression

is necessary
- professions (*Jag är webbutvikler*)
- education

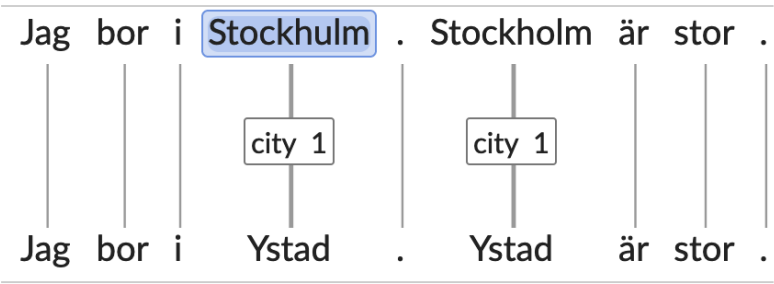
7. Categories that can be used for discrimination, such as political views, religious convictions or sexual orientation, should also be marked (with the tag < sensitive >) without being masked right away. A decision needs to be made later in the process, before publication. E.g. *I en dag såg vi en stor demstration det var för mycket människor vill inte Turkiets statsminister Ardogan och vi kände mycket glad för att det var första dag ser vi en fri demstration*.
8. Although information about languages spoken by the writer may help identifying the writer, such information is not pseudonymized, since this information is nevertheless included in the metadata which will be available for the corpus users.

2. Supra-categories

May be applied on top of other categories, as (extra)linguistic information.

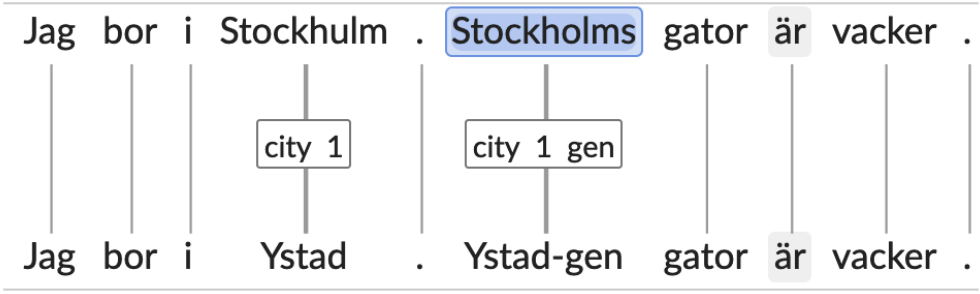
2.1 Running numbers

Applies to all *personally identifiable information (PII)* and their *@placeholders*. Each unique PII type (e.g.name) should get its own running number, starting with 1 in each individual essay. If the same PII is repeated in the text, the same running number is assigned to it. This is done automatically, but the automatically assigned running number may be changed manually. A manual change of the running number is necessary when the same PII (for instance the same city) is referred to by non-identical strings (for instance due to mis-spelling).



2.2 Morphology

- Case: < gen > , e.g. Volvo**s**
- Definiteness: < def > , e.g. Stadsbiliotek**et**s
- Number: < pl > , e.g. Mölndalsb**or**
- Only marked forms are tagged, i.e. genitive case is marked, whereas nominative case is not (by default everything is assumed to have nominative case).



3. Pseudonymization

3.1 Personal Names

- Types: < firstname_male > , < firstname_female > , < firstname_unknown > , < initials > , < middlename > , < surname >
- < firstname > vs < surname > are sometimes difficult to distinguish between. In uncertain cases, follow the standard Swedish order: < firstname > < middlename > < surname >
- Pseudonymization: (suggested source is from a national statistical agency: https://svn.spraakbanken.gu.se/sb-arkiv/lexikon/scb-namn)
  - Provide a list with first names, male, female and gender neutral (incl. international).
  - For surnames, gender-specific types, when unclear use gender-neutral names
  - Provide a list with surnames (incl. international)
  - Middlenames: Replace with an initial “A”

    - Initial: Replace by “A”, keep delimiters
- To consider:
    - allow cross-reference/anaphora resolution, i.e. allow to keep track of the PIIs that the L2 learner refers to, e.g. if more than one unique name occurs in the text, each unique name shall be replaced by a unique pseudonym. This is handled automatically, through the procedure with running numbers (see above).
    - random substitution for each unique name in the text given a list of names or
    - select a few names (while keeping the gender and cross-reference info) and use these names - throughout all texts

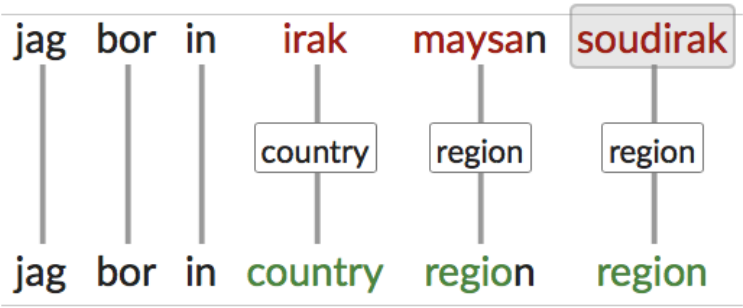
3.2 Geographic data

- Types: < foreign >, < area > , < city > , < geo > , < country > , < place > , < region > , < street_nr >, < zip_code >
    - < foreign > : this tag is combined with the following tags when applied to places *outside of Sweden*: < region > , < city >, < place > , < geo >. For places in Sweden the same four tags are used *without* the additional tag < foreign >.
    - < country > : except Sweden
    - < zip_code > : zip/area code
    - < region > : geographical/political unit larger than a city but not equivalent to a country. E.g. *Sörmland*, *Stockholms län*, *Region Blekinge*, *Svealand*, *Sydirak*, *Toscana*, *Baltikum*. Regions outside of Sweden are marked with the additional tag < foreign >.
    - < city > : the tag < city > is used for a large category of populated areas, including cities, urban districts both larger and smaller than cities, and villages. Cities outside Sweden are marked with the additional tag < foreign >.
    - < area > : part of a city, e.g. *Vesterbro*, *Greenwich village*. The tag < area > is only used for non-Swedish settings, and is always (reduntantly) combined with the tag < foreign >. (For Swedish settings the tag < city > is used also for city parts.)
    - < place > : specific place, street, square, bridge, name of a bus/tram/metro stop. Places outside Sweden are marked with the additional tag < foreign >.
    - < geo > : This tag is used for any additional type of geographic name not among the other categories, for instance forests, lakes, mountains, etc. Geographical names referring to entities outside Sweden are marked with the

additional tag < foreign >.

- < street_nr > : street or place number

- Uncertain categorization:
  - In some cases local knowledge or extensive research is needed in order to determine the most suitable category for a geographic name: It may be hard to figure out whether a name rather refers to a region or to a city, to a city or to an area, etc. Such extensive research is not motivated by the objective of the pseudonymization; what's important is that information potentially revealing the identity of the writer is pseudonymized, and that the categorization (and subsequent replacement) of the pseudonymized entities in the text are consistent with the rest of the information given in the text.



- Pseudonymization: (suggested source: http://www.geonames.org)
  - Random substitution given a list of named entities of various attributes for each attribute, except for Sweden
  - Two approaches are possible: (a) to substitute a PII with another PII of the same category (e.g. *Barcelona* for *Reykjavik*), or (b) to substitute with a dummy name of the form *A-city*, (e.g. *A-city* for *Reykjavik*). Strategy (a) gives better readability to the text, while strategy (b) helps avoid accidental semantic or grammatical errors, e.g. *I live in Barcelona where I can ski all year round* where *Barcelona* is an automatic replacement of *Reyklavik*. In SweLL-gold.v1 strategy (b) is applied.
  - < zip_code >: ABCDEF alt Replace letters with ABC and each number with 0 (ABC 0000), keep the delimiter

3.3 Institutions

- < school > , < work > , < other_institution >
- The institution tags are used to pseudonymize institutions mentioned in the texts which may be used to identify the writer, such as the school, work or sport's team

of the writer (or a person related to the writer).
    - < school > is used for all education-providing institutions (primary school, secondary school, university, etc.)
    - < work > is used for an institution which is revealed as the writer's working place (or the working place of a person related to the writer). When an institution is identified as a working place, the tag < work > is applied instead of other tags which may otherwise be applied. For instance: If a text reveals that the writer works in a named school, the tag < work >, rather than the tag < school >, is used to pseudonymize the name of the school.
    - < other_institution > is used for all other institutions in need for pseudonymization, such as a sport's team or an NGO
- Pseudonymization:
    - Replace from a list of school names and companies (e.g. from Yellow pages) alternatively use *A-school, B-workplace, C-institution*. In SweLL-gold.v1 the second alternaative is used.

3.4 Transportation

- < transport_name >, < transport_nr >
- < transport_name >: used for transport lines or transport systems with specific names, e.g. *gröna linjen, Lidingöbanan, Pågatågen*
    - *Tvärbanan*, and similar words which may be interpreted both as type nouns and as names for specific lines, are pseudonymized with this tag.
    - Words which clearly refer to transportation types rather than to specific lines, such as *tunnelbana, buss, pendeltåg* etc. should not be pseudonymized, although some of them reveal a city or limit the number of possible cities.
- < transport_nr > : used for the number of a specific line, e.g. "buss *528*", "linje *3*". The tag should be placed only on the *number*.
- Pseudonymization:
    - < transport_name > : Replace with *A-linjen, B-linjen* etc. to avoid adding inconcistensies into the text
    - < transport_nr > : Replace actual number with 1, in case of several numbers in the same text, enumerate (1, 2, 3...)
- **Note**: Names of stations and stops, such as *Mariatorget, Centralen*, are pseudonymized with the tag < place > in the *geographic data* group.

3.5 Age

- < age_digits >, < age_string >
- Person's age (e.g. 18 years old)
- Pseudonymization:
  - Change the year within the range of numbers in 5-year interval. If an author writes 18 y.o., provide a number from a range of numbers < age > (+ - 2) - > e.g. 16-20. (This is done automatically by the annotation tool.)
  - The same as above applies to < age_string > , rendered in strings.
  - (There is a complication, though: if for example age is written in letters (and also misspelled, like "niotton" or "slxtton"), then automatic replacement becomes nontrivial. There is a need of an option to add "pseudonimyzation" manually directly in the tool by rewriting the target token. At the moment this is not possible. Another issue with this is that misspelling can be pretty bad and there is a need for "interpretation" by an assistant, e.g. "åttonde" år (elder sister) versus "tionde" år (little sister). Issues of this kind are handled individually on a case-to-case basis.

3.6 Dates

- Types: < date_digits > , < day > , < month_digit >, < month_word >, < year > .
- Pseudonymization: Ideally - keep the delimiters as in the original (, . - /)
  - < day > - > random number between 1-28
  - < month_digit > - > random replace 1-12
  - < month_word > - > random replace: januari, februari, ...
  - < year > - > 5-year interval: e.g. 2013 is replaced by a random number from a range of numbers (+ - 2), i.e. 2011-2015
  - < date_digits > - used for all dates that are written as a sequence of numbers with delimiters. Replace all numbers with "1" to a standard from 11-11-1111, (keeping delimiters not implemented) e.g.
    - 2018-12-01 –> @date_digits –> 1111-11-11
    - 18/01/12 –> @date_digits –> 11-11-1111
    - 180112 –> @date_digits –> 11-11-1111
    - 18.01.12 –> @date_digits –> 11-11-1111

        - 01/12 –> @date_digits –> 11-11-1111

3.7 Phone numbers

- < phone_nr >
- Pseudonymization:
    - Replace each number with a “0” in the sequence (e.g. 0000-000000) (and keep the delimiter)

3.8 Email addresses

- < email >
- Pseudonymization:
    - One single for all: email@dot.com

3.9 Web pages

- < url >, applies to personal webpages or webpages that can disclose some information about the person such as link to a workplace.
- Pseudonymization:
    - Replace all with: url.com

3.10 Social security numbers

- < personid_nr >
- Pseudonymization:
    - Replace each number with: 123456-0000 (and keep the delimiter (-))

3.11 Account numbers

- < account_nr >
- Pseudonymization:
  - Replace each number with 0 (and keep the delimiter(s))

3.12 Certificate, licence numbers

- < license_nr > (e.g. vehicle)
- Pseudonymization:
  - Replace letters with ABC and each number with 0 (ABC 0000)

3.13 Other sequence of numbers

- < other_nr_seq >
- Pseudonymization:
  - Replace each number with 0 (and keep delimiters)

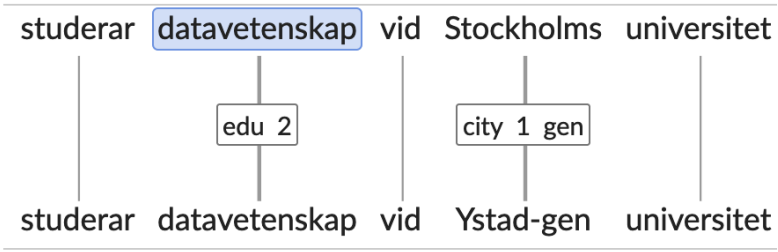
3.14 Extra (something else, not covered in the previous categories)

- < extra >
- When the pseudonymizer comes across some kind of information in a text which may potentially be used to identify the writer, but which is not covered by any of the other pseudonymization categories, the tag < extra > is used.
- The < extra > tag may for instance be used to mark information about very specific events in the writer's life.
- By default we consider all "Extra" tags as obligatory to pseudonymize. However, there is a need to re-evaluate the category after the initial pseudonymization and see whether there is a need to separate between obligatory and non-obligatory pseudonymization of "extras".

<!–//: # (In that case, the following could apply:)

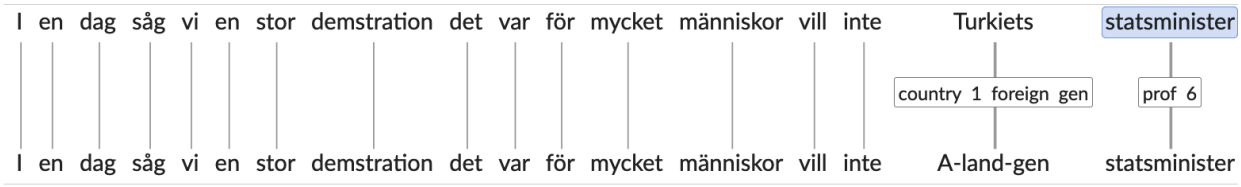
3.15 Mark up but do not pseudonymize

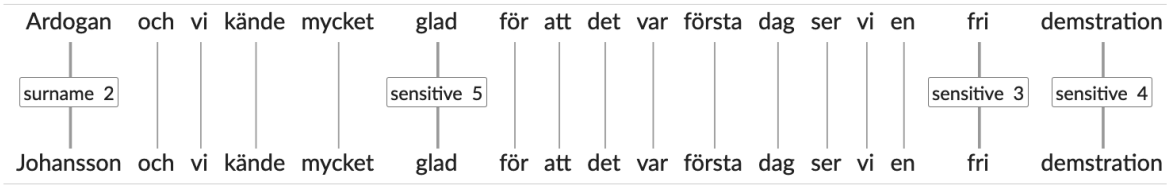
- < prof >, professions, e.g. *webbutvecklare*
- < edu >, education: Use for degrees etc., e.g. *datavetare*, "jag har en examen i *kemi*"
- < fam>, family members: Use for words for family members, friends, etc., e.g. *mamma, farfar, son, kusin, kompis, vän*.



- < sensitive >, sensitive information:
  - Markup: assign a "sensitive" @placeholder to at least one token per sentence. When deciding which and how many tokens to mark with the < sensitive > label, a guiding principle is that pseudomization of the marked tokens could potentially suffice. However, the whole sentence needs to be reviewed later on before final decisions about these pseudonymizations are made, and fewer rather than more tokens should be marked. A possible solution for the example below is to mark the tokens "glad", "fri" and "demstration".
  - **Note**: Sensitive information which could be covered by other pseudonymization categories should be assigned these other labels, e.g. "Turkiet" and "Ardogan" in the example below.

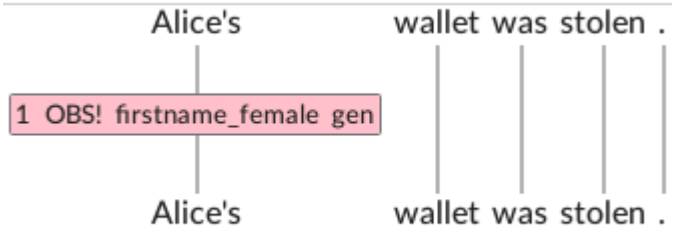
  Example:

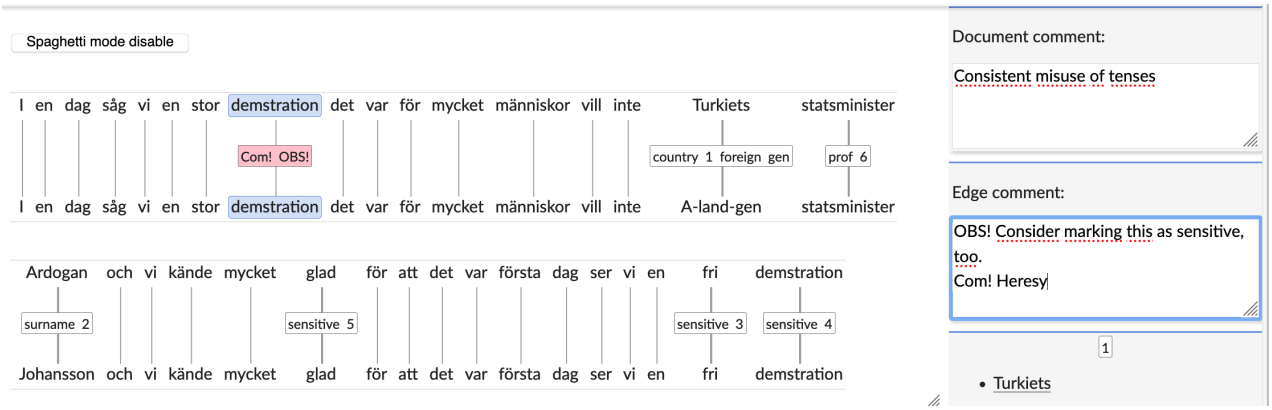




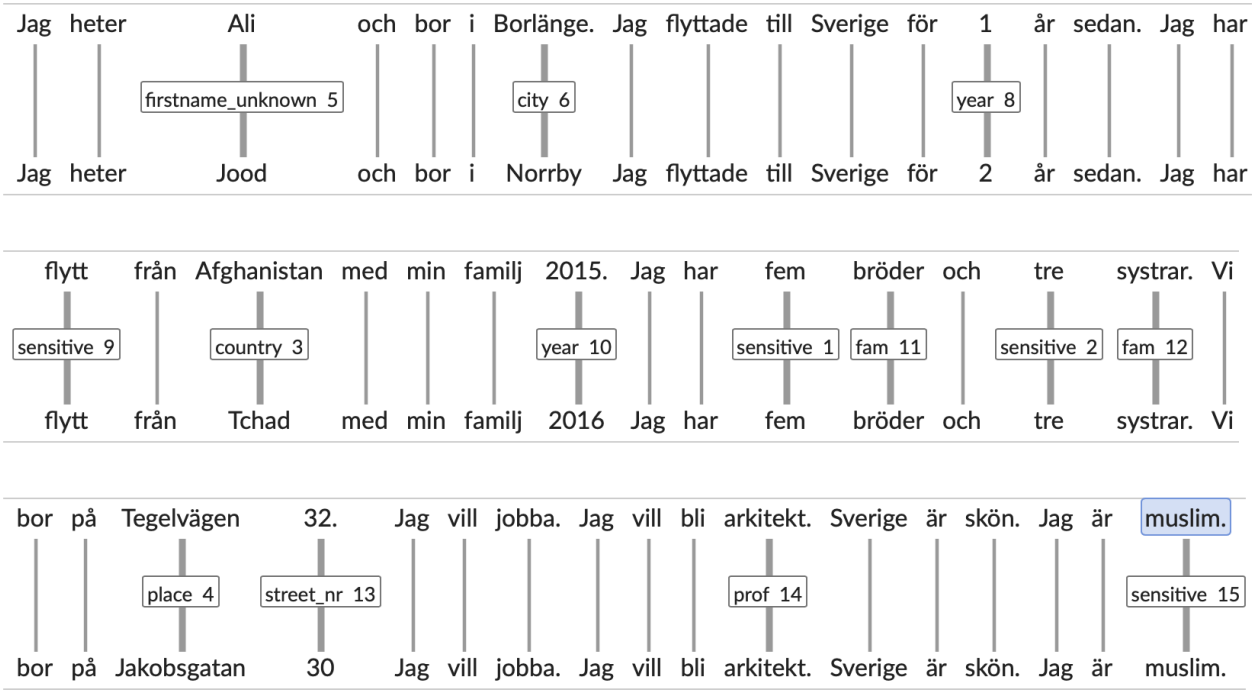
3.16 Comments

- < OBS! >, < Com! >, document comments
- The < OBS! > tag is used for marking a place to return to or for making comments which may be useful for later stages in the work with making the text ready for the corpus release (i.e. normalization and correction annotation).
- The < Com! > tag is used for marking specific text sequences in need of comments which are judged to be useful for the future users of the corpus and which are thus intended to be kept in the published corpus.
- Both the < OBS! > tag and the < Com! > tag are connected to an “edge comment” field where notes may be made. The label(s) get red-pink background for easier identification when there is a need to return to it/them.
- There is also a *document comment* field in which notes about the text as a whole may be made. These notes may provide essential information for later stages (normalization, correction annotation), or in some cases information which is meant for the corpus user and which is thus meant to be kept in the published corpus.





4. Text example



5. Resources (lists) for pseudonyms and automatic pseudonymization scripts

Resources and lists are collected in a private repository here: https://github.com/SamirYousuf/LR_project

6. SweLL annotation tool

SVALA is used for both manual annotation and for supportive automatic pseudonymization. A demo-version of the tool can be found here: https://spraakbanken.gu.se/swell/dev/#

7. SweLL publications on the topic

- Beáta Megyesi, Sofia Johansson, Dan Rosén, Carl-Johan Schenström, Gunlög Sundberg, Mats Wirén & Elena Volodina. (2018). Learner Corpus Anonymization in the Age of GDPR: Insights from the Creation of a Learner Corpus of Swedish. Proceedings of the 7th NLP4CALL workshop. (https://ep.liu.se/ecp/152/006/ecp18152006.pdf)[https://ep.liu.se/ecp/152/006/ecp18152006.pdf]
- Elena Volodina, Yousuf Ali Mohammed, Arild Matsson, Sandra Derbring, Beatá Megyesi. (2020). Towards Privacy by Design in Learner Corpora Research: A Case of On-the-fly Pseudonymization of Swedish Learner Essays. COLING-2020. https://aclanthology.org/2020.coling-main.32.pdf
- Wirén Mats, Arild Matsson, Dan Rosén, Elena Volodina. 2019. SVALA: Annotation of Second-Language Learner Text Based on Mostly Automatic Alignment of Parallel Corpora. CLARIN-2018 post-conference volume. LiUP Press. https://ep.liu.se/ecp/159/023/ecp18159023.pdf

GU-ISS, Forskningsrapporter från Institutionen för svenska språket, är en oregelbundet utkommande serie, som i enkel form möjliggör spridning av institutionens skriftliga produktion. Det främsta syftet med serien är att fungera som en kanal för preliminära texter som kan bearbetas vidare för en slutgiltig publicering. Varje enskild författare ansvarar för sitt bidrag.

GU-ISS, Research reports from the Department of Swedish, is an irregular report series intended as a rapid preliminary publication forum for research results which may later be published in fuller form elsewhere. The sole responsibility for the content and form of each text rests with its author.