eurac research

Transc&Anno

Nadezda Okinina, Lionel Nicolas
Eurac Research

nadezda.okinina@eurac.edu

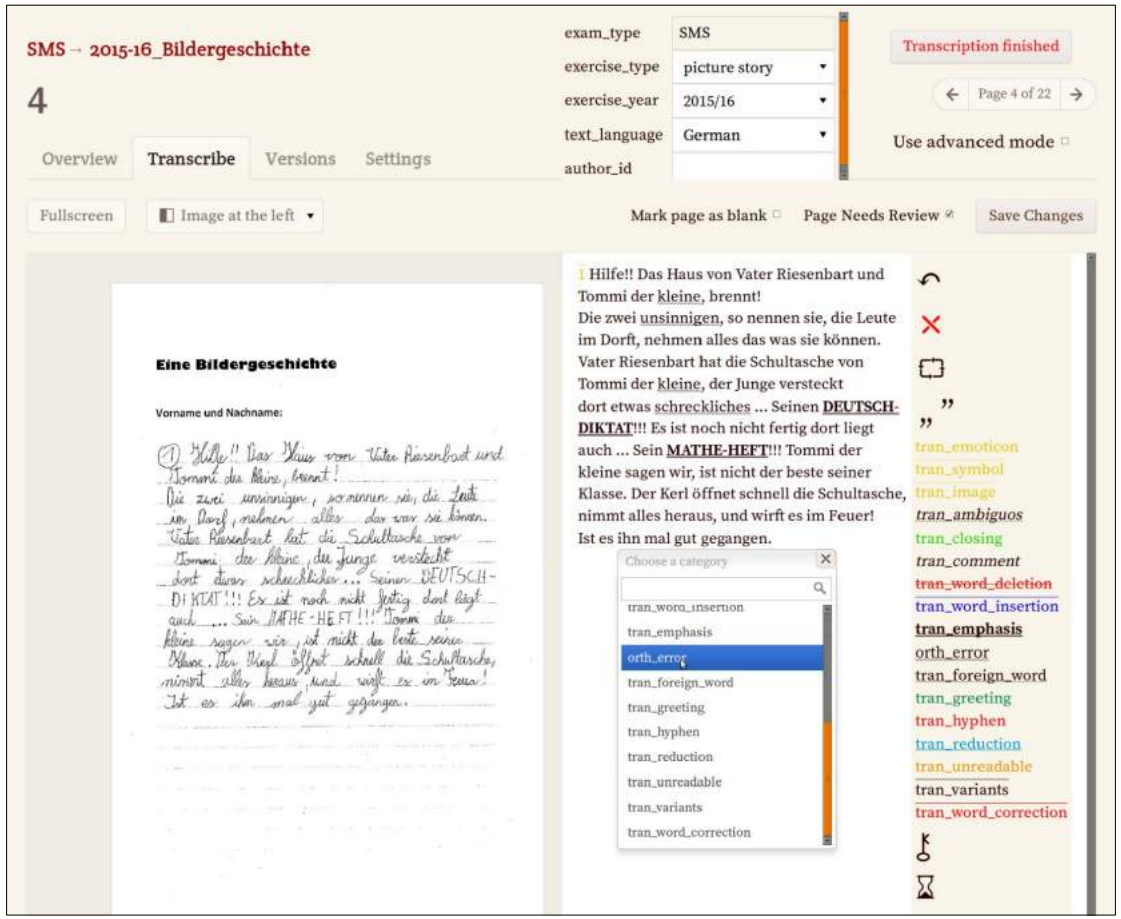
Transc&Anno

- **transcription** of text images
- their shallow on-the-fly **annotation**

- Freely available on https://github.com/commul/transcanno

- Created by adapting FromThePage https://github.com/benwbrum/fromthepage

Transc&Anno



Has been used

at the Institute for Applied Linguistics

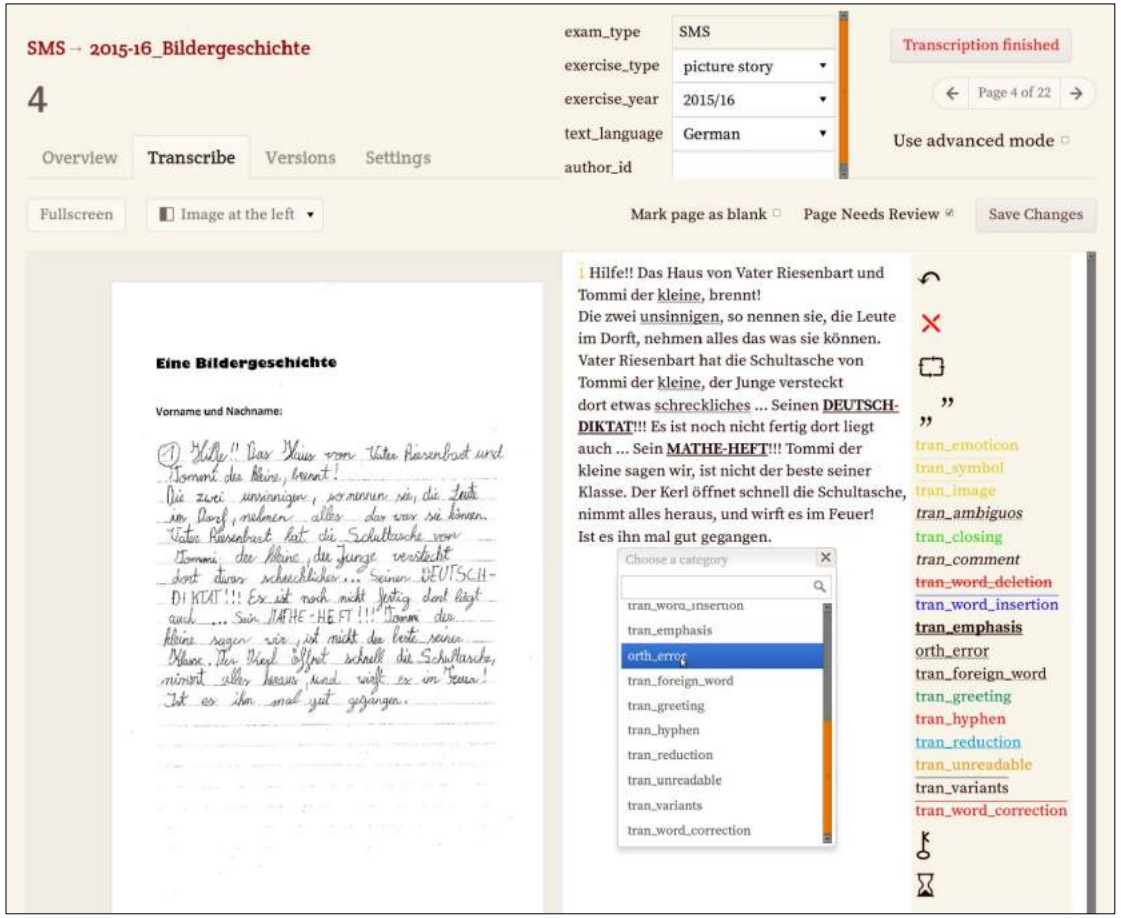
at Eurac Research for the transcription of

2 learner corpora:

- SMS (English, German, Italian)
- Matura (German)

The usage revealed some imperfections of the program that didn’t impede the successful accomplishment of the transcription and annotation of the 2 collections.

Transc&Anno



On-line tool working in the Firefox browser:

- Transcribers can work from distance.
- Collection owners can monitor their progress from distance.

Installation requires help from IT specialists.

Transc&Anno

1 Today we will all together:

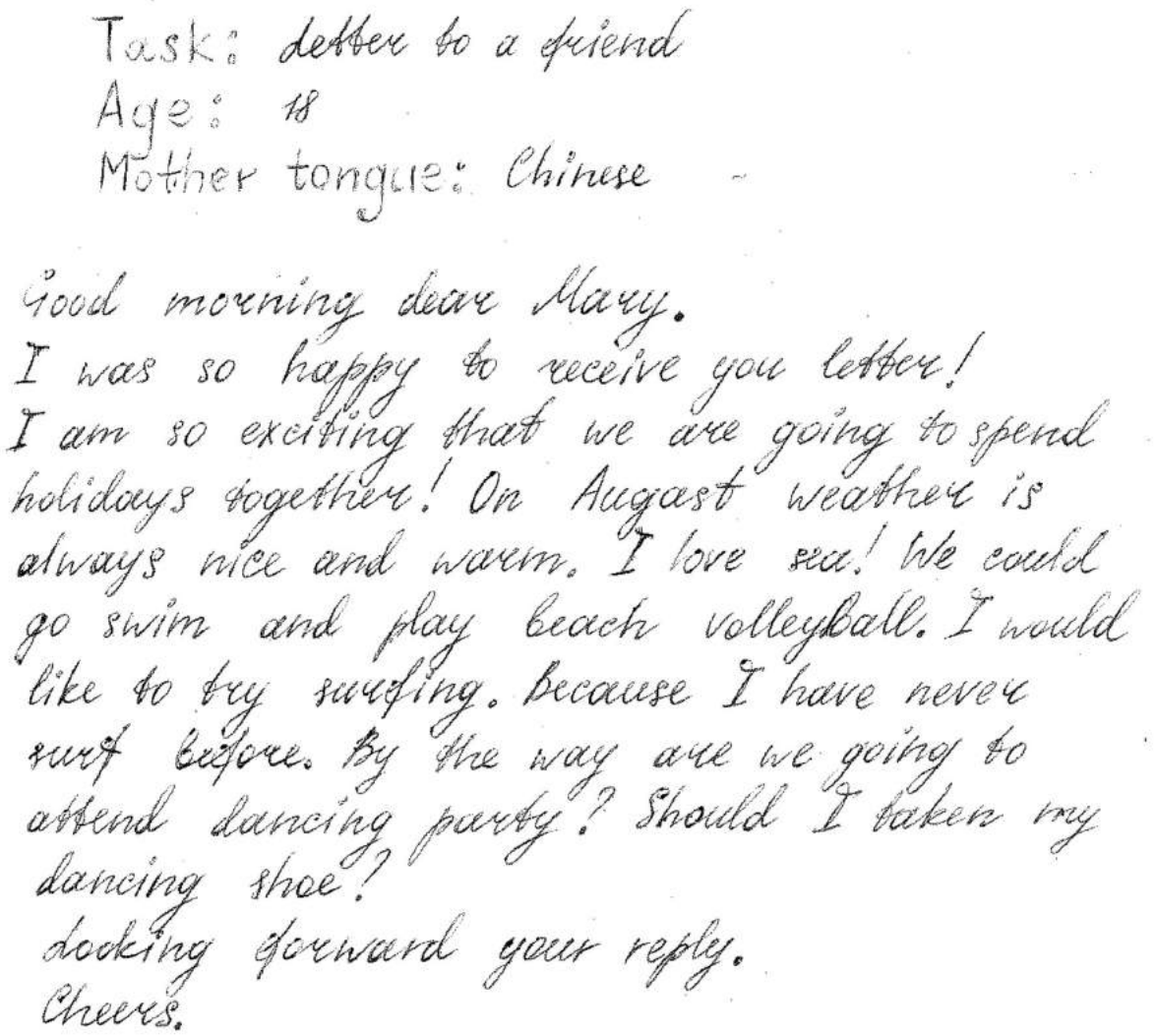
- Read a fake handwritten learner text
- Decide how we want to annotate it
- Create a new text collection in Transc&Anno installed on a Eurac server and accessible through the Firefox browser
- Upload the learner text into Transc&Anno
- Create annotation categories in Transc&Anno
- Transcribe and annotate the learner text
- Export our transcription into XHTML and TEI XML formats

2 Each of you on his own (and with my help) will:

- Create a collection for his own texts in Transc&Anno (1/2 texts)
- Upload his own texts (1/2 texts)
- Create annotation categories for his own texts
- Transcribe his own texts
- Export his own texts into XHTML and TEI XML formats

Note that your texts will be saved on the Eurac server and deleted after. Don’t upload sensitive information.

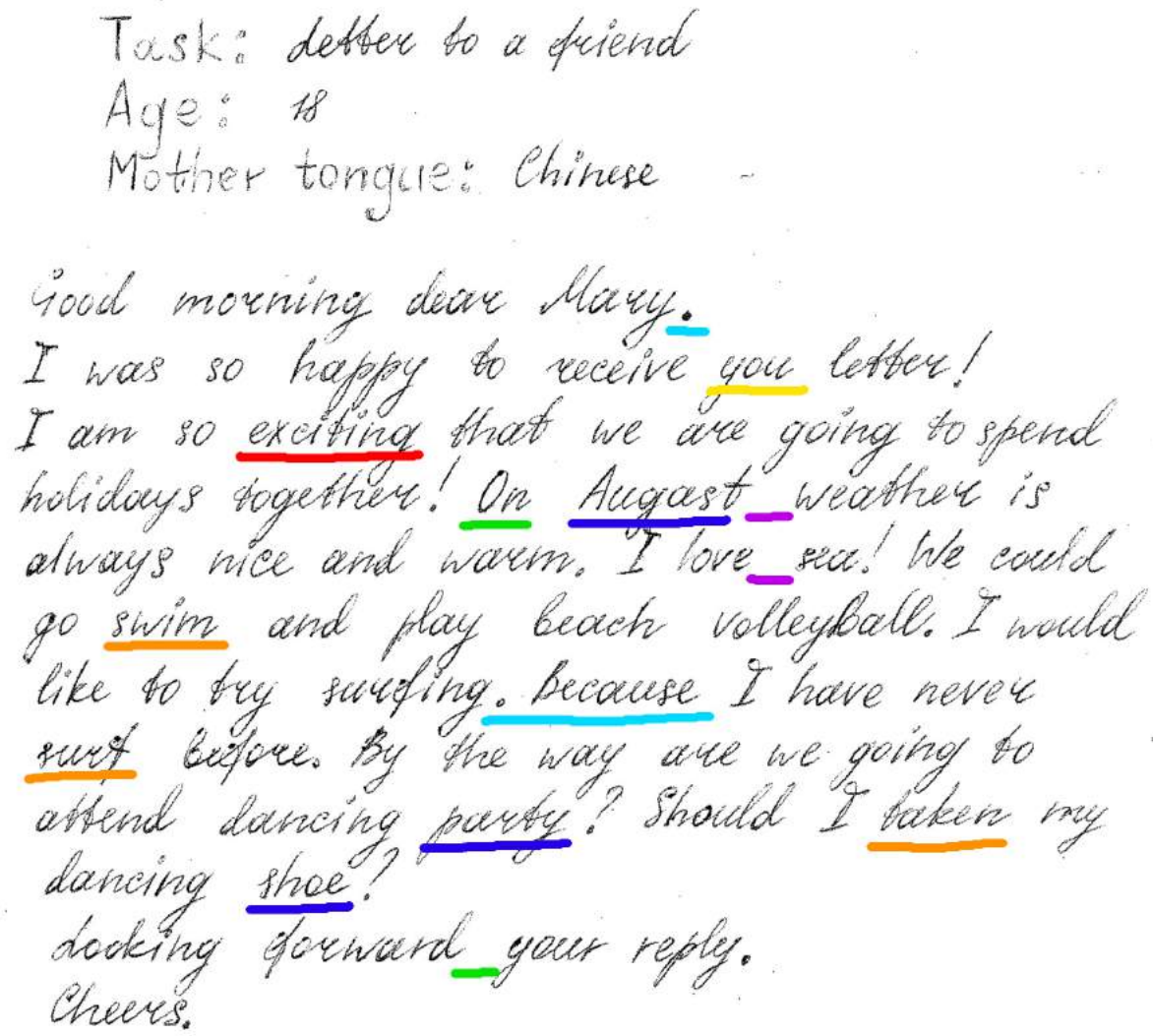
Transc&Anno



Task: letter to a friend
Age: 18
Mother tongue: Chinese

Good morning dear Mary.
I was so happy to receive you letter!
I am so exciting that we are going to spend holidays together! On August weather is always nice and warm. I love sea! We could go swim and play beach volleyball. I would like to try surfing. Because I have never surf before. By the way are we going to attend dancing party? Should I taken my dancing shoe?
looking forward your reply.
Cheers.

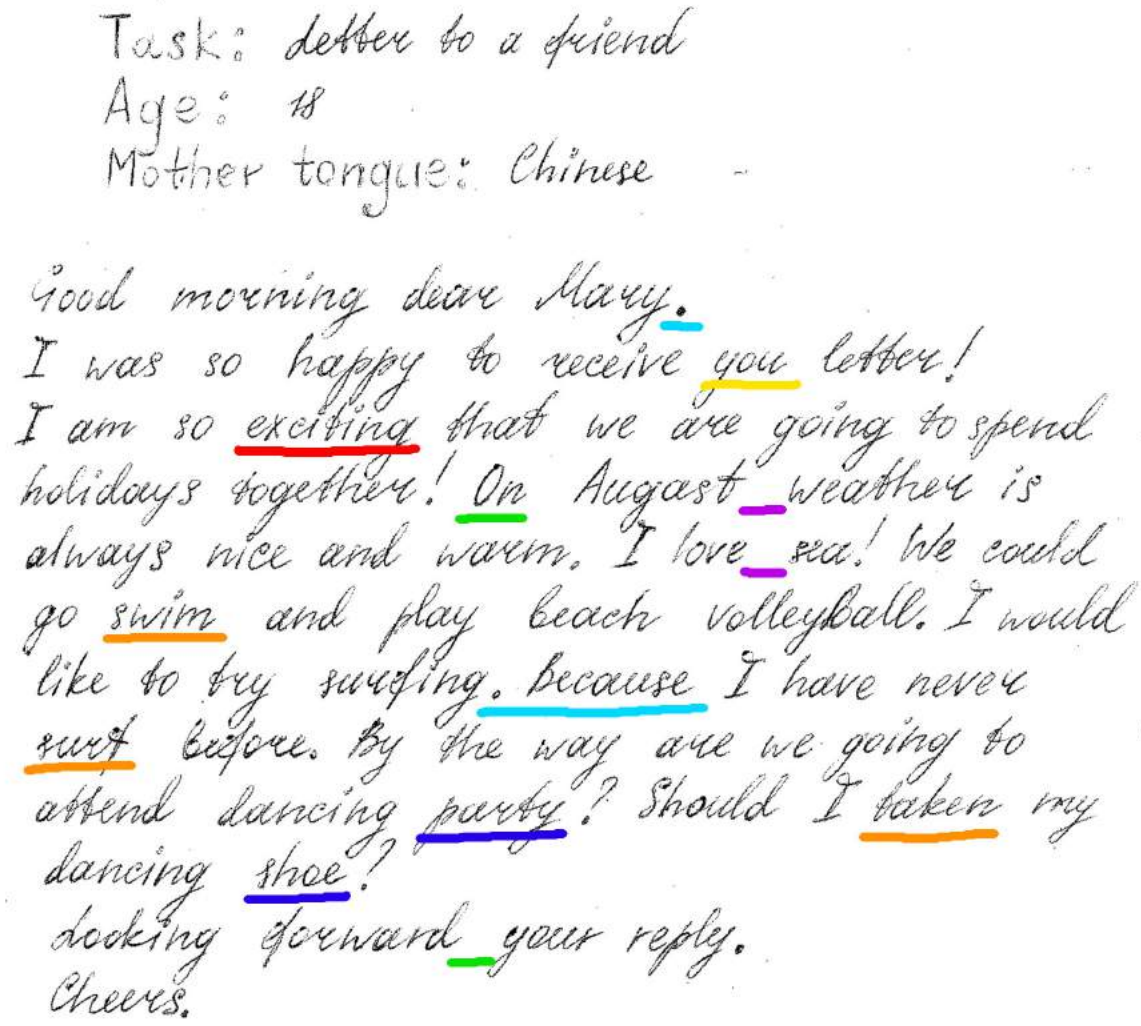
Transc&Anno



Task: letter to a friend
Age: 18
Mother tongue: Chinese

Good morning dear Mary.
I was so happy to receive you letter!
I am so exciting that we are going to spend holidays together! On August weather is always nice and warm. I love sea! We could go swim and play beach volleyball. I would like to try surfing. because I have never surf before. By the way are we going to attend dancing party? Should I taken my dancing shoe?
looking forward your reply.
Cheers.

Transc&Anno



Task: letter to a friend
Age: 18
Mother tongue: Chinese

Good morning dear Mary.
I was so happy to receive you letter!
I am so exciting that we are going to spend holidays together! On August weather is always nice and warm. I love sea! We could go swim and play beach volleyball. I would like to try surfing. because I have never surf before. By the way are we going to attend dancing party? Should I taken my dancing shoe?
looking forward your reply.
Cheers.

Header categories describing the whole text:

- Task
- Age
- L1

Annotation categories:

- Punctuation error
- Verb error
- Article error
- Preposition error
- Noun error
- Pronoun error
- Adjective error

Each error has:

- a target (a correction)
- a type (delete, add or replace)

Transc&Anno

Firefox browser

1 Please **sign in**.

Login and password are the same: your family name in small letters

(without umlaut, stress or other marks, only in Latin characters) followed by **CLARIN**.

Example: Joseph König has a login **konigCLARIN** and a password **konigCLARIN**

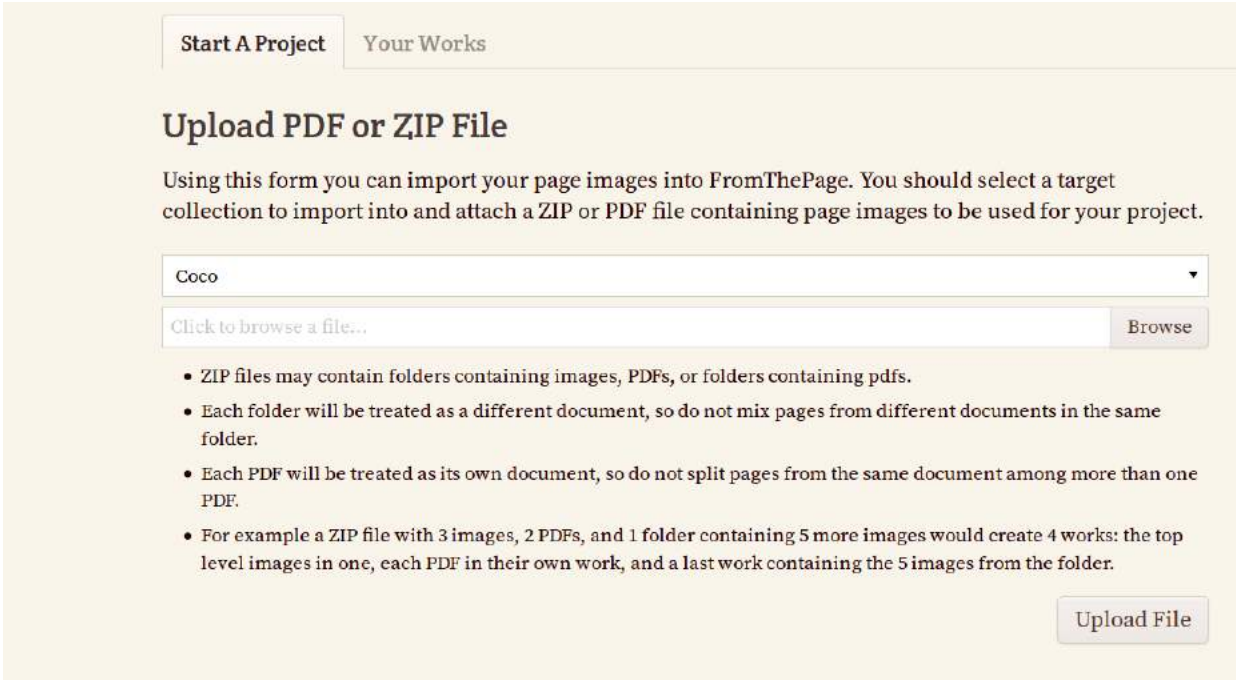
2 Please go to the collection ***Learner_Text*** where I uploaded our learner text and created the necessary header and annotation categories. Observe.

Please, **don’t change anything in the transcription**.

Transc&Anno

3 Create a new collection with your family name as title.

4 Upload our learner text into the new collection.



Transc&Anno

5 Go to the **Subjects** tab and **create header categories**.

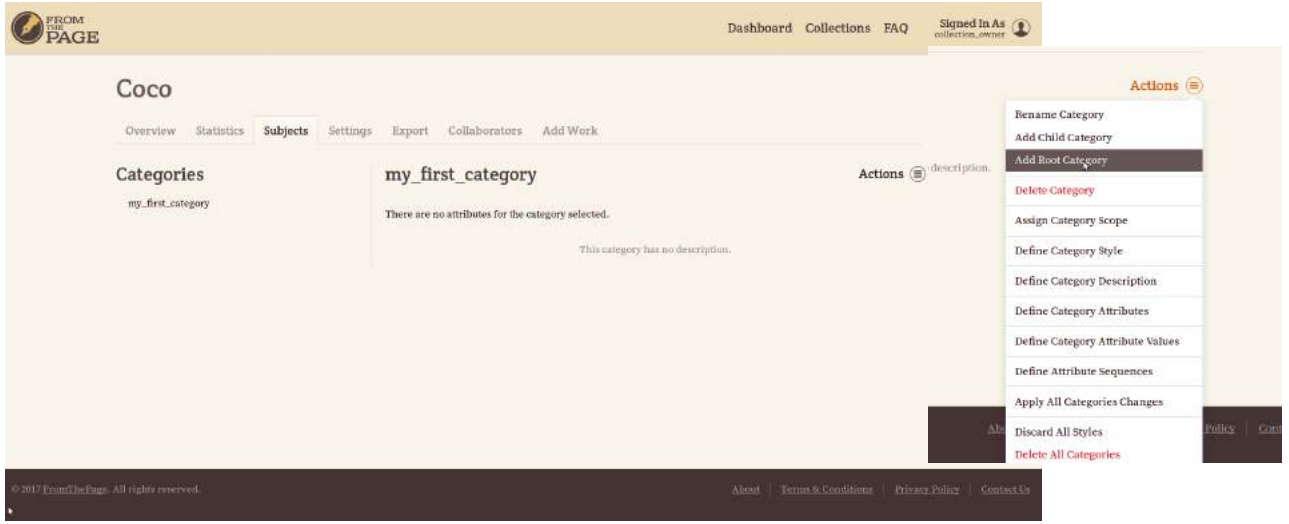
Don’t forget to add a description to each category.

**Refer to the user guide for clarifications.*

6 Create **annotation categories**.

Don’t forget to add a:

- Description
- Style
- *Target* attribute
- *Type* attribute
- Allow user input for the *target* attribute
- Define possible values of the *type* attribute



Transc&Anno

7 Go to the **Overview** tab, then click on the name of the file to transcribe and **transcribe the learner text** using the annotation categories you have just created.
Try to use **hot keys** while inserting annotation tags.

**Refer to the user guide for clarifications.*

8 When you have finished transcribing push the **Transcription finished** button in order to delete previous versions of the transcription and save space in the database.

9 Click on the collection name and go to the **Export** tab.
Click on **Export all works**. Observe the result.
Click on **TEI**. Observe the result.
Note that export into the CSV format does not work. If you get an error page, just go back.

Transc&Anno

10 Now that you are familiar with the tool, you can

- **create a new collection for your own texts** and
- repeat all the steps once again.

CONTACT US

Eurac Research
Institute for Applied Linguistics

Bozen/Bolzano Italy

nadezda.okinina@eurac.edu

www.eurac.edu

eurac
research