

Annotation Guidelines: Early New High German Mining Texts

Elisabeth Gruber-Tokić¹, Milena Peralta Friedburg², Gerhard Rampl³

“Text Mining Medieval Mining Texts (T.M.M.M.T.)”⁴
University of Innsbruck

2022

DOI: 10.5281/zenodo.6275197



ÖAW
ÖSTERREICHISCHE
AKADEMIE DER
WISSENSCHAFTEN

 universität
innsbruck
Institut für
Sprachwissenschaft

¹ ORCID: <https://orcid.org/0000-0003-3387-3097>

² ORCID: <https://orcid.org/0000-0002-9367-4623>

³ ORCID: <https://orcid.org/0000-0001-9333-922X>

⁴ Official website, <https://miningtext.at/>, 29.12.2021.

Introduction

The project “Text Mining Medieval Mining Texts” processes two historical mining sources: “Verleihbuch der Rattenberger Bergrichter”⁵ (TLA⁶ Hs. 37, 1460-1463 AD) and “Schwazer Berglehenbuch”⁷ (TLA Hs. 1587, approx. 1515 AD) stored by the Tyrolean Regional Archive, Innsbruck (Austria).

Both historical documents are of central importance for late medieval mining in the mountain regions of Schwaz and Rattenberg in Tyrol (Austria) because they contain data regarding the official claims to mine ores and minerals in the mining district of Brixlegg-Rattenberg and Falkenstein (Tyrol, Austria).

T.M.M.M.T.’s main objective is the extraction of information related to legal relationships between people, claims (grant of exploitation rights), mines and the localisation of mines in the landscape. Place names, person names and names of tunnels, shafts and mining areas within the historical sources allow the representation of mutual relationships and act as connecting links. The digitised and transcribed versions are available on T.M.M.M.T.’s read and search platform “Mining Hub”⁸.

Both late medieval mining documents are annotated to create a gold standard usable to develop an automated NER workflow of late Middle High German/ Early New High German texts on the one hand. On the other hand, the created metadata is also further processed and finally imported in a knowledge graph. Annotation is being conducted using the basic tag set of Transkribus (Read Coop 2021).

Transkribus was chosen as a basic annotation tool for different reasons: First, the platform has been used right from the beginning, which means both documents have been digitised with the help of a ScanTent and the DocScan App and further processed with Transkribus. Second, a basic tag set is available which can be customized. The tags comprise properties (attributes) to specify the tag and add further information. Third, the platform offers the possibility to export the corrected transcripts as well as the tags using different formats, e.g. TEI compliant or Excel export. The T.M.M.M.T. project uses the Excel export function of the tags for integration into a Postgres database (PostgreSQL 2021). (Read Coop 2021).

It must be outlined that basic knowledge of Transkribus is required to use the guidelines (tag creation, property creation, annotation). Otherwise, we recommend the How-to Guides on the website of Transkribus⁹.

T.M.M.M.T.’s annotation guidelines are based on the articles “Impresso Named Entity Annotation Guidelines (Version 2.1)” (Ehrmann et al. 2019), “Anleitung zur Erstellung von Annotationsrichtlinien” (Reiter 2020: 193ff.) and “How To enrich transcribed documents with mark-up” (Transkribus 2019).

With respect to the research objectives, only words and phrases are annotated that comprise the following entities: person, location (place and mine) and date (to be specified in the next section).

During a first annotation turn in December 2020 we stumbled across certain annotation difficulties that need further discussion to create a balanced and useful annotation of the historical documents. These guidelines are aimed at the other members of the T.M.M.M.T. research group and all persons in general who annotate late medieval German texts.

⁵ Complete transcript available. DOI: 10.5281/zenodo.6274562

⁶ Tiroler Landesarchiv (=Tyrolean Regional Archive, Innsbruck, Austria)

⁷ Complete transcript available. DOI: 10.5281/zenodo.6274928

⁸ Mining Hub, <https://transkribus.eu/r/mining-hub/#/>, 29.12.2021.

⁹ Transkribus ‘How to guides’, <https://readcoop.eu/transkribus/resources/how-to-guides/>, 29.12.2021.

Annotation, Entity Types and subtypes

As already mentioned, the fundamental research objective of T.M.M.M.T. is the extraction and representation of the legal relationships between people, claims and mines over space and time. Therefore, only words and phrases are annotated that comprise the following entities: (detailed explanation follows below)

- person (standard Transkribus tag)
- place (standard Transkribus tag)
- mine (custom tag)
- date (standard Transkribus tag)

Further standard tags were used to mark information regarding the text itself: “sic”, “unclear”. The standard Transkribus tag set contains further tags, which were not used in the T.M.M.M.T. project and therefore not further mentioned in these guidelines.

It is also possible to create tags and/or add properties (=attributes) to existing tags if needed. In this case the tag for the entity “mine” was created, as well as properties for the entities “person” and “date”. (Detailed explanation follows below).

Annotation

In Transkribus, the process of annotation must be conducted in the transcript editor (see fig. 1). Words and phrases must be selected, and the correct tag has to be assigned by right-click or customized shortcuts. There is an overview of the tags of the current transcript, the tag set and the attributes (in Transkribus called “properties”) of the currently selected tag on the left-hand side.

The properties must be completed manually after assigning the tag (via copy/paste). The completion of the tag attributes is optional, the tag export is always possible. T.M.M.M.T. decided to complete the tag attributes as far as possible for further processing reasons. Any extra information gathered while the annotation process supports the development of the NER tool, the creation of gazetteers (place or person gazetteers) and semantic representation.

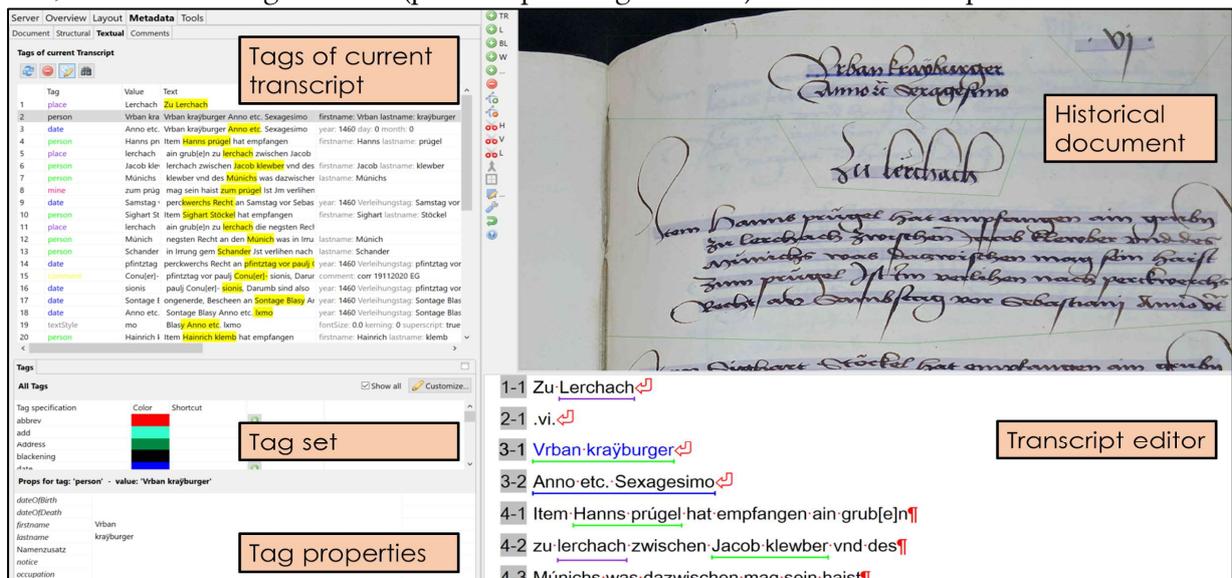


Fig. 1. Screenshot of Transkribus with added explanations.

Entities (tags)

This section focuses on the description of the annotated entities. The tags and their related attributes are further specified. The chosen examples are meant to guide and give support while annotating difficult text passages.

1.1 Person

Attributes (in Transkribus called “properties”)

- *firstname* First name
- *lastname* Last name
- *descriptor* (custom attribute) comprises a descriptive phrase added to the first name of a person that is used to identify individuals e.g., *Michel ab dem Hochhaws* (first name *Michel* + *descriptor*)
- *occupation* Occupation
- *title* (custom attribute) Honorary title
- *dateOfBirth* Date of Birth
- *dateOfDeath* Date of Death

Coverage of the type Person

- Considered as Person: real persons e.g., *vrban krayburger*
 - also: the mine owner’s name when referring to his property e.g., *krayburgers schacht* (*krayburgers* is tagged as ‘person’)
- Not considered as person: expressions which do not contain a proper name, isolated functions not attached to a person name (cf. Ehrmann et al. 2019), names of saints which form part of a name of a mine or a date. Those first names occur in combination with different orthographic variations of *sankt* ‘saint, holy, sacred’

Person Components

- Titles: were not tagged separately but added as separate attribute within the person-tag (e.g., *Freiherr*)
- Functions: were not tagged but added within the occupation slot in the person-tag, when mentioned within the text (e.g., *perckrichter*, *dischlermaister*, *huetman*, *doctor*)
exception: within the following phrases *sneider* and *lechner* could be either a person’s last name or occupation: *Sneider von púchling*, *Lechner ab dem Lehen*. In this case the phrase receives a person tag and *sneider* is added to the attributes descriptor and occupation (see examples below)
- Descriptor: was tagged within the person tag and added to the attributes. In late medieval mining texts descriptive terms occur in combination with proper person **names**. These terms may designate personal attributes, location of origin, occupation or parentage. Although these phrases belong to the person’s name and were used to identify individuals, they are not yet to be categorized as last names. Nevertheless, these descriptors often developed into last names as we know them today (e.g., *Vlrich perckgesell von Swatz*, *Michel ab dem Hochhaws*, *Niclas ratperger von Swatz*). Anyway, the use of last name and descriptor is not stable within the historical documents which proves the current change in the use of person names. The same person may be mentioned in different ways, two examples:
1) *Christan schuster von lisfeld* is also mentioned as *Cristan von Liesfeld*.
2) *Vll weber von slitters* is also mentioned as *Vll weber*.

The first example shows that the word *schuster* is used as description of Cristan but not as a last name since in the other reference the word *schuster* is not mentioned.

The second example suggests another situation. While *weber von slitters* is used to describe the person, the second mention only shows *weber*. That way, it is unclear if *weber* is used as description or a last name. These cases must be discussed separately in the course of annotation.

Another challenge are references to persons by using other persons names, for example *Hanns meixners son* “son of Hanns Meixner” or *Jórig Rentleins gesell* “Jórig Rentleins assistant”. From a linguistic point view both examples only *Hanns Meixner* and *Jórig Rentlein* have to be annotated as names. But since the current annotation is also used for semantic representation purposes, there was an exception to be made. Therefore, both examples were annotated as persons and received the attribute “descriptor”.

Table 1 Examples of annotated persons

Entity in text	Annotation
<i>Cristoffen Kirchpichler</i>	Completely annotated First name: <i>Cristoffen</i> Last name: <i>Kirchpichler</i>
<i>Bernhart Planck von Landshut</i>	Completely annotated First name: <i>Bernhart</i> Last name: <i>Planck</i> Descriptor: <i>von Landshut</i> + Place tag: <i>Landshut</i>
<i>Michel ab dem Hochhaws</i>	Completely annotated First name: <i>Michel</i> Descriptor: <i>ab dem Hochhaws</i>
<i>wolfgang von grásperg</i>	Completely annotated First name: <i>wolfgang</i> Descriptor: <i>von grásperg</i> + Place-tag: <i>grásperg</i>
<i>Annthoni vom Roß</i>	Completely annotated First name: <i>Annthoni</i> Descriptor: <i>vom Roß</i>
<i>Caspar pegk aus der grúben</i>	Completely annotated First name: <i>Caspar</i> Last name: <i>pegk</i> Descriptor: <i>aus der grúben</i>
<i>Hánnsel des friedleins knecht</i>	Completely annotated First name: <i>Hánnsel</i> Descriptor: <i>des friedleins knecht</i> Occupation: <i>knecht</i>

<i>herren wilhalmen freyherren zu wolckhenstain</i>	Completely annotated First name: <i>wilhalmen</i> Title: <i>freyherren zu wolckhenstain</i> + Place-tag: <i>wolckhenstain</i>
<i>Maximilian von gotes gnaden Erwelter Romischer Kayser</i>	Completely annotated First name: <i>Maximilian</i> Title: <i>Erwelter Romischer Kayser</i>
<i>Sneider von púchling</i>	Completely annotated Occupation: <i>Sneider</i> Descriptor: <i>Sneider von púchling</i> + Place-tag: <i>púchling</i>
<i>Vlrich perckgesell von Swatz</i>	Completely annotated First name: <i>Vlrich</i> Last name: <i>perckgesell</i> Descriptor: <i>von Swatz</i> Occupation: <i>perckgesell</i> + Place-tag: <i>Swatz</i>
<i>Cúntz Artzknapp von prichslegk</i>	Completely annotated First name: <i>Cúntz</i> Last name: <i>Artzknapp</i> Descriptor: <i>von prichslegk</i> Occupation: <i>Artzknapp</i> + Place-tag: <i>prichslegk</i>
<i>ob des fridleins grubn</i>	Only <i>fridleins</i> annotated Last name: <i>fridleins</i>
<i>verlegen paw vnderhalb Oswalds SpringJnclee gruben</i>	Only <i>Oswalds SpringJnclee</i> annotated First name: <i>Oswalds</i> Last name: <i>SpringJnclee</i>
<i>Hans englhart munßmaister</i>	Completely annotated First name: <i>Hans</i> Last name: <i>englhart</i> Occupation: <i>munßmaister</i>
<i>crstoff dorner Canntzler</i>	Completely annotated First name: <i>crstoff</i> Last name: <i>dorner</i> Occupation: <i>Canntzler</i>
<i>Hannsen des pluterers Sun</i>	Completely annotated First name: <i>Hannsen</i> Descriptor: <i>des pluterers Sun</i>
<i>Hanns meixners son</i>	Completely annotated

	Descriptor: <i>Hanns meixners son</i>
<i>Bruder gilg</i>	Completely annotated Title: <i>Bruder</i> First name: <i>Gilg</i>
<i>den alten niderist</i>	Completely annotated Last name: <i>niderist</i> Descriptor: <i>alten</i>

1.2 Place

Attributes

- *country* Country

Coverage of the type place

- Places include cities and smaller units (e.g., country, village, hamlet, locality, micro toponym)
- The line between place name and appellative is quite narrow, and sometimes the decision if a word needs a tag is very difficult. Each case must be decided separately. Worst case, we opted for annotation and a later in group discussion

Place Components

- Places mentioned within the pit name were additionally tagged with a place-tag

Table 2 Examples of annotated places

Entity in text	Annotation
<i>valckhenstain</i>	Completely annotated
<i>Swatz</i>	Completely annotated
<i>Rottenstain</i>	Completely annotated
<i>gruben am Anlaß</i>	Only <i>Anlaß</i> annotated
<i>Ynnsprugg</i>	Completely annotated
<i>Ertzhertzog Sigmundt von Österreich</i>	Only <i>Österreich</i> annotated as place Country: <i>Österreich</i> + Person-tag: <i>Ertzhertzog Sigmundt von Österreich</i>
<i>Haintz aus der Huben</i>	Only <i>Huben</i> annotated as place + Person-tag: <i>Haintz aus der Huben</i>
<i>Sant Simon Judas in Torntall</i>	Only <i>Torntall</i> annotated as place + Mine-tag: <i>Sant Simon Judas in Torntall</i>

1.3 Mine

Attributes

- *name* Mine name

Coverage of the type mine

- Mines include all types of mines (e.g., pit, gallerie, mine shaft)

Mine components

- Pit names mostly consist of different orthographic variations of *sankt* ‘saint, holy, sacred’ + name of a saint, e.g. *sand oswalt*; note: *oswalt* is NOT tagged as person
- Within the historical texts pit names mostly follow the preposition *zu* ‘to’ or the phrase *haist zu* ‘is called’. If present, the preposition *zu* is always annotated together with the mine: *zu dem wolgemut*

Entity boundaries are not always clear-cut e.g., *Zu Sannt Wolfgang ob der Fasnacht*, *Zu Sannt Lienhart Jm Prant*: In the historical document TLA Hs. 1587, pit names may contain descriptions of the location, which are either to be understood as an addition to the name of the mine or separately as a description of the location. This description may contain information about the location in order to avoid ambiguity.

Different ways to describe the location of a pit:

1. Location of the pit in relation to another pit: *Zu Sant Kathrein auf der Laimzech*
2. Pit name contains place name: *Sant Veit im Nassntal*
3. Pit name contains locative appellatives: *Zu Sannt Michel aufm weg*

Table 3 Examples of annotated mines

Entity in text	Annotation
<i>Zu Sannt Wolfgang ob der Fasnacht</i>	Completely annotated
<i>vns[er] frawn gruben</i>	Completely annotated
<i>Zu der Fronica im Schontall</i>	Completely annotated + Place-tag: <i>Schontall</i>
<i>zu Sand oswalt</i>	Completely annotated
<i>Zu sant Gilg im Tuller Gassl</i>	Completely annotated + Place-tag: <i>Tuller Gassl</i>
<i>Zu Sant Jacoben Trenckhpach</i>	Completely annotated + Place-tag: <i>Trenckhpach</i>
<i>zu Hodritschner</i>	Completely annotated
<i>zum gotberat</i>	Completely annotated
<i>zu dem wolgemut</i>	Completely annotated

1.4 Date

Attributes

- *day* day (standard attribute: only numbers allowed)
- *month* month (standard attribute: only numbers allowed)
- *year* year (standard attribute: only numbers allowed)
- *datePhrase* date phrase, (custom attribute). Allows the caption of the whole phrases that refer to the liturgical feast, e.g., *sambstag nach letare*

Coverage of the type date

- Dates comprise either the word *anno* or *Jar* ‘year’
- all dates consisting of Latin numbers if existent e.g., *1497 Jar*, *1517 Jar*
- all dates consisting of roman numerals, e.g., *lxmo*, *lxiido*
- all dates consisting of combinations of roman numerals and numerals written in words, e.g., *lxsecundo*
- all phrases that start with a day of the week and end either with *Jar*, *lx.*, *Anno etc.*: *Eritag vor Egidi Anno etc. lxiido* ‘tuesday before Egidi Anno etc. 62’, *Sambstag nach Michaelis, Anno etc. lxsecundo*
- dates are usually introduced by a preposition *am* or *an*

Date components

The method of dating in the historical documents is by naming the festive day or the day of the week, indicating the festive day before or after that day: e.g., *pfintztag vor Jnuocauit Jm 1496 Jar*, *mitag vor Cantate Jm 1497 Jar*, *sambstag nach letare Im 1497 Jar*. It is possible to convert all the dates manually to day/month/year using, for example, Grotefeld¹⁰, which is not really effective.

Therefore, to capture the date phrases and analyze them, the property ‘*datePhrase*’ was added to the tag. This way, the phrases are captured and may be converted automatically in another moment.

Table 4 Examples of annotated dates

Entity in text	Annotation
<i>pfintztag vor Jnuocauit Jm 1496 Jar</i>	Completely annotated Year: 1496 <i>DatePhrase: pfintztag vor Jnuocauit Jm 1496 Jar</i>
<i>Sand pauls tag Conu[er]sionis Anno domi[ni] etc. lxprimo</i>	Completely annotated Year: 1461 <i>datePhrase: Sand pauls tag Conu[er]sionis Anno domi[ni] etc. lxprimo</i>
<i>mitag vor Cantate Jm 1497 Jar</i>	Completely annotated Year: 1497 <i>DatePhrase: mitag vor Cantate Jm 1497 Jar</i>

¹⁰ Grotefeld website, <http://bilder.manuscripta-mediaevalia.de/gaeste//grotefeld/grotefeld.htm>, 29.12.2021.

<i>sambstag nach letare Im 1497 Jar</i>	Completely annotated Year: 1497 <i>DatePhrase: sambstag nach letare Im 1497 Jar</i>
<i>1510 Jar</i>	Completely annotated Year: 1510

Tag export

As already mentioned, Transkribus offers the export of tags in TEI, alto/page or XLS format which allows further processing of the data.

T.M.M.M.T. Generated data

You may also be interested in further datasets and transcripts generated by the research team of T.M.M.M.T.:

1. Hs. 37 TEI Datasets, DOI: 10.5281/zenodo.6276620
2. Hs. 1587 TEI Datasets, DOI: 10.5281/zenodo.6276685
3. Historical Place Gazetteer and Historical Mine register, DOI: 10.5281/zenodo.6275763
4. Historical Person Register, DOI: 10.5281/zenodo.6276521
5. Citeable Transcripts on Zenodo: Hs. 37, DOI: 10.5281/zenodo.6274562
6. Citeable Transcripts on Zenodo: Hs. 1587, DOI: 10.5281/zenodo.6274928
7. POS-tagged corpus "Montanistika" is available on CQP Workbench CQP
<http://sprawi-cqpweb.uibk.ac.at/CQPweb/>
8. Knowledge Graph TMMMT, DOI: 10.5281/zenodo.6276586
9. Animated Maps, ArcGIS Insight:
<https://insights.arcgis.com/#/view/8ad6a04ce90d4c129f24e0a223e43496>, 25/02/2022

References

- Ehrmann, Watter, Romanello, & Clematide. (2020). Impresso Named Entity Annotation Guidelines (Version 2.2). Zenodo. <http://doi.org/10.5281/zenodo.3585750>, 29.12.2021.
- Ehrmann M., Romanello M., Flückiger A., Clematide S. (2020) Overview of CLEF HIPE 2020: Named Entity Recognition and Linking on Historical Newspapers. In: Arampatzis A. et al. (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2020. Lecture Notes in Computer Science, vol 12260. Springer, Cham. https://doi.org/10.1007/978-3-030-58219-7_21, 29.12.2021.
- Gruber-Tokić, Elisabeth (2022): Verleihbuch der Rattenberger Bergrichter, 1460-1463 (Tiroler Landesarchiv Hs. 37), University of Innsbruck. DOI: 10.5281/zenodo.6274562
- Gruber-Tokić, Elisabeth (2022): Schwazer Berglehenbuch, approx. 1515 (Tiroler Landesarchiv Hs. 1587), University of Innsbruck. DOI: 10.5281/zenodo.6274928
- PostgreSQL: <https://www.postgresql.org/>, 29.12.2021.
- Reiter Nils (2020): Anleitung zur Erstellung von Annotationsrichtlinien. In: Riter Nils, Pichler Axel, Kuhn Jonas (eds.) Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt. Walter de Gruyter GmbH, Berlin/Boston. 193-202. DOI <https://doi.org/10.1515/9783110693973>
- Read Coop: <https://readcoop.eu/transkribus/>, 29.12.2021.
- Transkribus: "How To enrich transcribed documents with mark-up" Version v1.9.1 Last update of this guideline: 2/12/2019.