Building a Gold Standard for Temporal Entity Extraction from Medieval German Texts

Natalia Korchagina

1.1. Introduction. Motivation.

‣ Importance of temporal dimension:

   ‣ Undiscovered facts about historical events/personalities

‣ Manual extraction is time-consuming

‣ Needed: a tool for automatic temporal entity extraction

‣ Application examples of temporal tagging:

   ‣ **information extraction:**

      ‣ events are summarized and chronically ordered

   ‣ **information retrieval:**

      ‣ time as a query topic

1.1. Introduction. Motivation.

‣ Specificity of temporal expressions in historical texts: high spelling and lexical variation
  ‣ Use of tools for modern texts doesn't lead to good results

expression in old German	meaning in modern German
Xbris	Dezember
sant Conrats tag anno etc. lxxxxiij	26 November 1493

A temporal entity extraction system adapted for historical corpora is need.

Why do we need a Gold Standard?
‣ Base for quality estimation of the temporal extraction system
‣ No corpora of temporal annotations for historical German available

1.2. Introduction. Data.

‣ The project is funded and the data is provided by the Swiss Law Sources Foundation

‣ Historical Swiss legal texts

‣ 118 volumes of critical editions available at

http://ssrq-sds-fds.ch/online/

‣ Available for digital processing:

28 volumes

‣ **~7 million tokens** of historical data

‣ 23,5 of 28 available volumes are in German

‣ Text creation time: 10th-18th centuries

‣ Multilinguality: texts in German, French, Latin, Romansh, Italian.

1.3. Introduction. Task.

▸ **Goal:** build a corpus of gold standard annotation of temporal entities in historical German texts.

Workflow:

1. Select data to be manually annotated.
2. Adapt annotation guidelines.
3. Analyse the annotated data.
4. Use the resulting corpus as a base for experiments

2.1. Creation of a Gold Standard Corpus. Dataset.

‣ *Number of articles:* 50

‣ *Volumes*: 9 volumes from 5 cantons

‣ *Size:* about 32,000 tokens

‣ *Domain:* legal

‣ *Types of documents:* legal cases and transactional documents

2.1. Creation of a Gold Standard Corpus. Dataset.

Period: 1450-1550 (Early New High German) : abundance of texts, old yet readable

year 1425, 1278 articles

Number of articles

Year of issue

29 September 2016

JTDH 2016
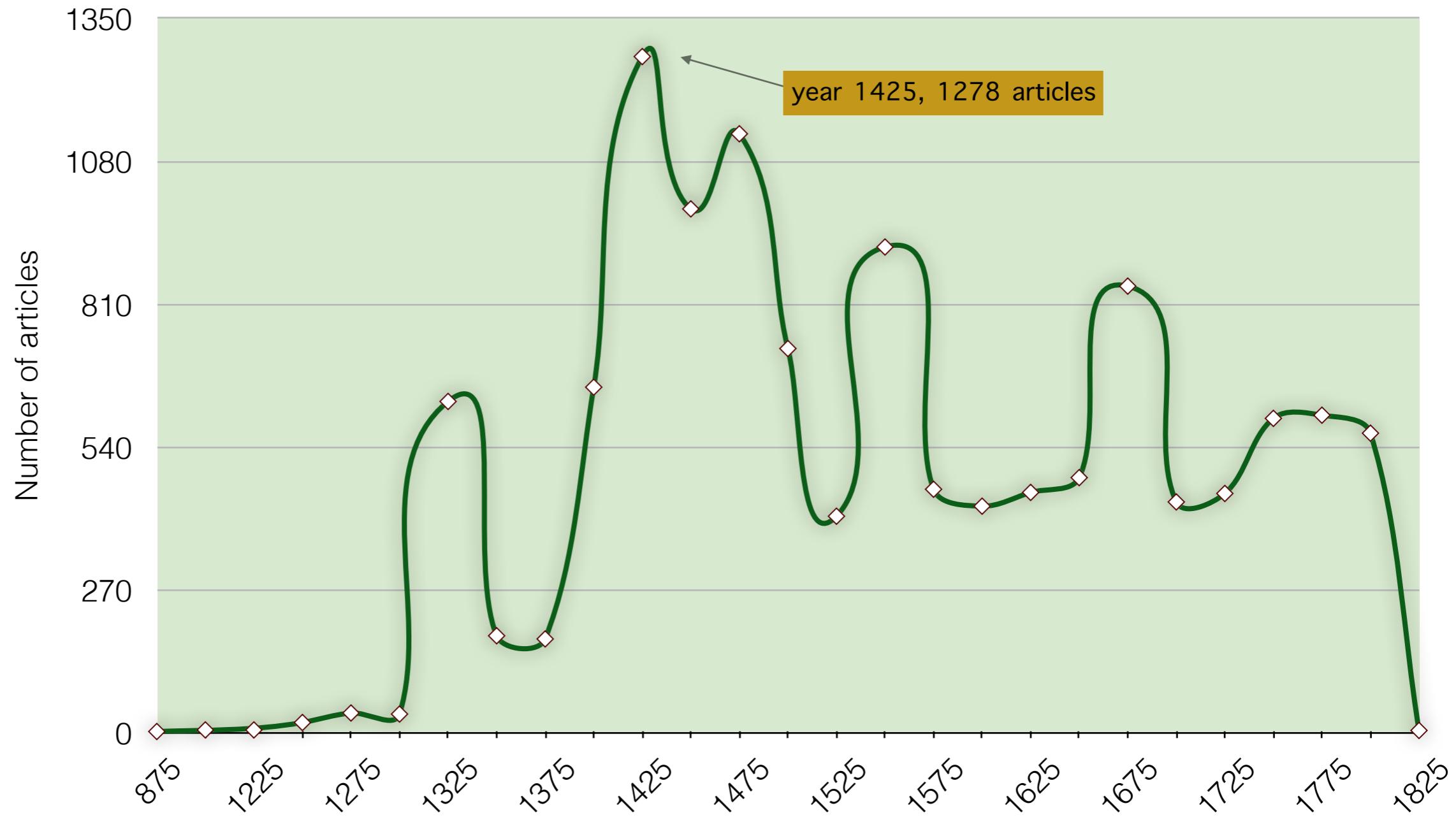
2.2. Creation of a Gold Standard Corpus. Annotation settings.

Annotators: 2

Time allocated: 40 hours

Corpus given: set of articles automatically annotated with a rule-based temporal tagger HeidelTime (Strötgen und Gertz, 2013) adapted for Text+Berg corpus (Volk et al., 2010), (Rettich, 2013).

200 text segments were identified initially by HeidelTime as temp.expressions

Annotation framework: Oxygen XML Author with a CSS and an XML-schema adapted for the task.

2.2. Creation of a Gold Standard Corpus. Annotation settings.

2.3. Domain-specific annotation guidelines.

There are general guidelines for temporal and event annotations (Sauri et al., 2006)

Can't be entirely reused:

- too general, don't reflect particularities of historical corpora:

 developed for annotating instances of an English dataset of the TempEval 2010,

 don't have old German-specific examples, e.g. with saint days.

- too detailed:

 describe the use of more attributes then we need for the task, "mod", "temporalFunction"

What to annotate?

How to annotate?

2.3. Domain-specific annotation guidelines.

What to annotate?

Different kinds of temporal expressions to be considered:

explicit: 25. December 1490

implicit: Christmas

relative: Friday after Christmas

---

markable: today

(can be situated on a timeline)

non-markable: soon

(cannot be situated on a timeline)

2.3. Domain-specific annotation guidelines.

What to annotate?

TIMEX3 Extent:

1) General rule of span-economy: the extent of a tag should be as small as possible:

<p style="color:red"><TIMEX3>am 3. Oktober 1999</TIMEX3></p>

vs.

<p style="color:green">am <TIMEX3>3.Oktober 1999</TIMEX3></p>

2) Grammatical categories:

- Noun, e.g.   *Montag*

- Noun Phrase (NP), e.g.   *suntag judica in der vasten*

- Adjective, e.g.   *jährliche <…>*

- Adverb, e.g.   *nun*

- Adjective or adverb phrases, e.g.   *alle nacht*

2.3. Domain-specific annotation guidelines.

How to annotate?

Task: check the automatically produced annotation,
add new tags when necessary

Annotation format: subset of the **TimeML** l-ge (Pustejovsky et al, 2003):

‣ *TIMEX3* tags for temporal exp. and *SIGNAL* tags
‣ Attributes: *type* and *value:*
Types: DATE/ TIME/ SET/ DURATION
Value: ISO 8601 format for date and time

```
<TIMEX3 tid="t695" type="DATE" value="1493-07-05">fritag</TIMEX3>

<SIGNAL sid="s15">nach</SIGNAL>

<TIMEX3 tid="t1064" type="DATE" value="1493-07-04">Ulrici</TIMEX3>

<TIMEX3 tid="t1065" type="DURATION" value="P1D" beginPoint="t695" endPoint="t1064"/>
```

2.4. Creation of a Gold Standard. Resulting corpus.

‣ **Inter-annotator agreement:**

	Detection	Classification
Average observed agreement	0.75	0.89
Cohen's Kappa	0.74	0.76

‣ After adjudication performed with MAI adjudication tool (Stubbs, 2011):

867 tags:

- 807 temporal expressions
- 60 signals

2.4. Creation of a Gold Standard. Resulting corpus.

Comparison between the Gold Standard and its predecessors, i.e. annotations

produced:

1) automatically by a rule-based temporal tagger HeidelTime (HT)

2) by human annotators A1 and A2

	Recall	Precision	Correctly classified	Correctly normalised
HT	0.26	0.96	96%	81%
A1	0.93	0.95	91%	84%
A2	0.88	0.90	95%	79%

2.4. Creation of a Gold Standard. Resulting corpus.

A rule-based temporal tagger cannot handle (so far):

‣ spelling and lexical **variation**, e.g. expressions with the meaning "evening":

Abend	aubend	aebents	stübglogen
abende	aubends	abentz	zenacht gessen
abends	abentt	abendes	zenacht essen
abendes	abentts	stübglogge	znacht essen
äbend	abent	stübgloggen	Nacht essen
aubent	abents	Stübgloggen	nachtessen
aubents	aebent	stübgloge	schlaff trunck

‣ complex temporal expressions with **Saints' days**:

e.g. *fritag nechst nach sanntt Gallen tag* -> 22 Oktober

(en: Friday after St. Galls day)

3.1. Experiments based on the Gold Standard.

Spelling normalisation to facilitate the use of existing temporal taggers:

1) Replace historical variants of temporal expressions with modern ones

2) Pass the **normalised corpus** to a temporal tagger

<...> viertzehenhundert und nuin und funffzig jare. <...>

viertzehenhundert ~ *vierzehnhundert*
nuin ~ *neun*
funffzig ~ *fünfzig*
jare ~ *Jahre*

<...> vierzehnhundert und neun und fünfzig Jahre. <...>

Temporal tagger

<...><TIMEX3>vierzehnhundert und neun und fünfzig Jahre</TIMEX3>. <...>
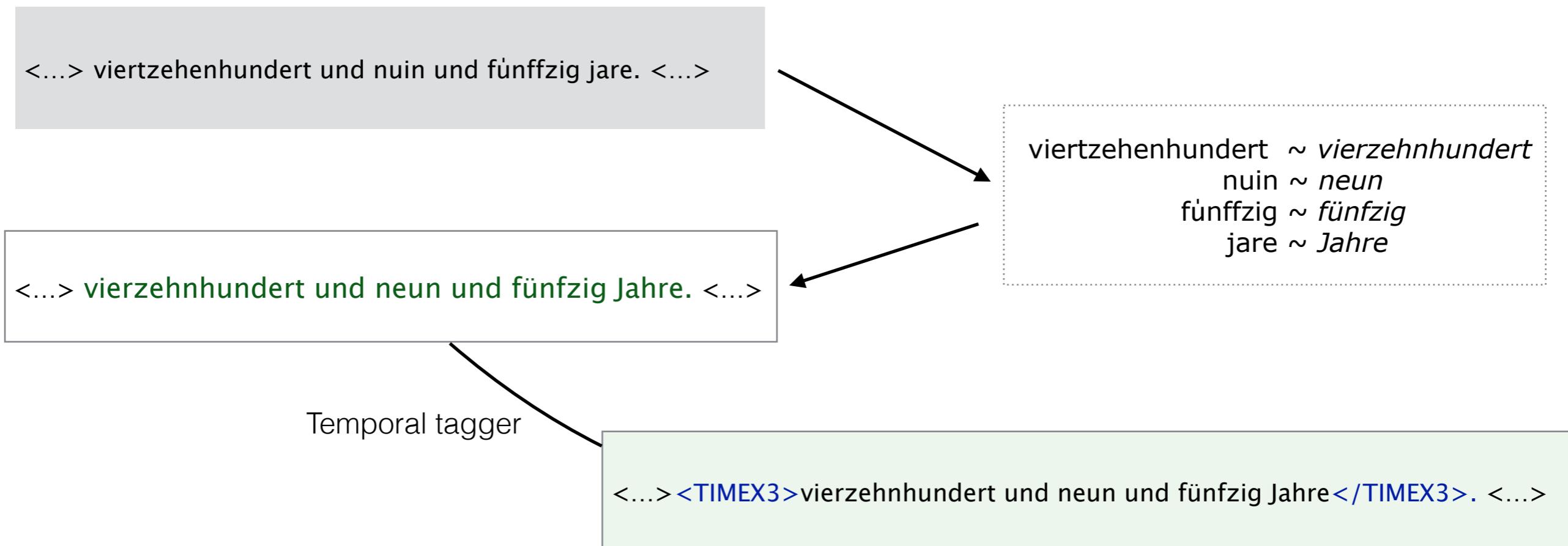
3.1. Experiments based on the Gold Standard.

Spelling normalisation to facilitate the use of existing temporal taggers:

1) Replace historical variants of temporal expressions with modern ones

2) Pass the **normalised corpus** to a temporal tagger

3) Some improvement in temporal entity extraction should be observed, but perfect results are not expected:

- single temporal expressions will be matched, but not if they are part of more complicated ones, for

which **the extraction rule is absent**

4. März 1647 can be recognised ,
1647 März 4 cannot be recognised

- there is always a possibility that a test set will contain some expressions that **do not have similar**

variants in resources

3.1. Experiments based on the Gold Standard.

Successful cases of applying normalisation techniques for similar tasks:

Verb identification in Early Modern Swedish texts.

Method used: edit-based.

Improvement: 64.2% (unnorm) to 86.2% (Pettersson et al., 2013)

Method used: character-based statistical machine translation.

Improvement: 64.2% to 87.7% (Pettersson et al., 2014)

POS-tagging of Early New High German texts

Method used: edit-based

Improvement: by 2% (Logăcev et al., 2014)

3.1. Experiments based on the Gold Standard.

To what extent can be improved the performance of a temporal tagger developed for modern texts by using normalisation as a pre-processing step?

Trial of spelling normalisation methods.

1. **Levenshtein-based** approach with context-sensitive weights (Pettersson et al., 2013)

   Experiment with the resources for German provided in the default version:

   GerManC corpus (1650-1780) (Scheible et al., 2011)

   Normalisation accuracy: 66% [1]

Evaluation of the temporal annotation produced by HeidelTime before (HT) and after (HT_norm_Lev) normalisation:

	Recall	Precision	Correctly classified	Correctly normalised
HT	0.26	0.96	96%	81%
HT_norm_Lev	0.27	0.89	96%	85%

[1] Calculated for a manually normalised part (10%) of the Gold Standard

4. Future work

‣   Further evaluation of various normalisation techniques;

‣   After the best-performing normalisation method is established:

    1) manually post-process a significant portion of the normalised output;
    2) apply a modern temporal tagger and evaluate the improvement of tagging a
       normalised corpus.

For expressions, not recognised as temporal even after normalisation:

‣   Try machine learning, e.g.

character-based classifiers learning the "shape of the word"

(Klein et al., 2003) and (Qi et al., 2014)

References

1. D. Klein, J. Smarr, H. Nguyen, and C. D. Manning. 2003. Named entity recognition with character level models. In: *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003* -Volume 4, CONLL '03, pages 180–183, Stroudsburg, PA, USA. Association for Computational Linguistics.

2. P. Logăcev, K. Goldschmidt, and U. Demske. 2014. Pos-tagging historical corpora: The case of early new high german. In: *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories*, Tübingen, Germany.

3. E. Pettersson, B. Megyesi, and J. Nivre. 2013. Normalization of historical text using context-sensitive weighted levenshtein distance and compound splitting. In: *Proceedings of the 19th Nordic Conference on Computational Linguistics*.

4. E. Pettersson, B. Megyesi and J. Nivre. 2014. Verb Phrase Extraction in a Historical Context. In: *The First Swedish National SWE-CLARIN Workshop at the Swedish Language Technology Conference (SLTC)*.

5. J. Pustejovsky, J. Castaño, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer, and G.Katz. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. IWCS-5, Fifth International Workshop on Computational Semantics.

6. Y. Qi, S. G. Das, R. Collobert, and J. Weston. 2014. Deep learning for character-based information extraction. In: *Advances in Information Retrieval -36th European Conference on IR Research,* ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings, pages 668–674.

7. K. M. Rettich. 2013. Automatische Annotation von deutschen und französischen temporalen Ausdrücken im Text+Berg-Korpus Zusammenfassung. Master thesis, University of Zurich.

8. R. Saurí, J.Littman, R. Gaizauskas, A. Setzer, and J.Pustejovsky. 2006. TimeML annotation guidelines, version 1.2.1.

9. S. Scheible, R. J. Whitt, M. Durrell, and P. Bennett. 2011. Evaluating an 'off-the-shelf' pos-tagger on early modern german text. In: *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities,* LaTeCH '11, pages 19–23, Stroudsburg, PA, USA. Association for Computational Linguistics.

10. J. Strötgen and M. Gertz. Multilingual and Cross-domain Temporal Tagging. Language Resources and Evaluation, 47(2):269–298, 2013.

11. A.Stubbs. 2011. MAE and MAI: Lightweight Annotation and Adjudication Tools. In: *Proceedings of the Linguistic Annotation Workshop V*, Association of Computational Linguistics, Portland, Oregon.

12. M. Volk, N. Bubenhofer, A. Althaus, M. Bangerter, L. Furrer, and B. Ruef. 2010. Challenges in building a multilingual alpine heritage corpus. In: *Proceedings of LREC,* Valetta, Malta.

Thank you for your attention!
Questions?