# A comparison of sequential and combined approaches for named entity recognition in a corpus of handwritten medieval charters

Emanuela Boroş*, Verónica Romero†, Martin Maarand*, Kateřina Zenklová‡,
Jitka Křečková‡, Enrique Vidal†, Dominique Stutzmann§, Christopher Kermorvant*
*TEKLIA, Paris, France
†PRHLT Research Center, Universitat Politècnica de València, Spain
‡Národní archiv, Prague, Czech Republic
§IRHT-CNRS, Paris, France

***Abstract*—This paper introduces a new corpus of multilingual medieval handwritten charter images, annotated with full transcription and named entities. The corpus is used to compare two approaches for named entity recognition in historical document images in several languages: on the one hand, a sequential approach, more commonly used, that sequentially applies handwritten text recognition (HTR) and named entity recognition (NER), on the other hand, a combined approach that simultaneously transcribes the image text line and extracts the entities. Experiments conducted on the charter corpus in Latin, early new high German and old Czech for name, date and location recognition demonstrate a superior performance of the combined approach.**

***Index Terms*—Named entity recognition, Handwritten Text Recognition, historical document processing, multilingualism.**

## I. Introduction

For centuries, historians have read historical texts and sources in the search of information, focusing on dates, and on place, person, and organization names. Their scholarly editions generally provide indexes of names. The correct identification of names is a dynamic process of textual and source criticism (basis and goal at the same time) and nurture all Humanities fields (historical sociology, demographics, art history, linguistics, etc.). Collections of gazetteers and authority data have been published as early as the 18th c. to help this work (e.g. the 35-volumes strong collection of *Dictionnaire topographique de la France* launched in 1859[1]). In historical, especially medieval, documents, the identification and disambiguation of named entities are complex, for person names as well as for locations. In Europe, people used to have only a given name. Family names or surnames started to become common in the 13th c., but in some regions or social classes much later (17th c. for Welsh people), many people shared the same name, the spelling was diverse in Vernacular and Latin, but also within one language (e.g. Guillelmus, Willelmus, William, Wilhelm, Guillaume, etc.). Locations may have disappeared or changed completely, and are very ambiguous and have also very different spellings, which make their identification very difficult, and even more to automate.

[1]https://dicotopo.cths.fr/about

Fig. 1. An example of text from a charter in Czech in the corpus (Prag, National Archive, Archives of Czech monasteries abolished by Joseph II, No. 1555). Issued in Český Krumlov on 11 October 1405. Jan, a parish priest in the town of Velešín at the Rosenberg family's domain, gives his right to receive taxes to the monastery of the Order of Saint Clare in Český Krumlov.

In this paper, we introduce a new corpus of handwritten medieval document images, manually transcribed and annotated with named entities. An example is shown in Figure 1. This corpus is used to compare two approaches for named entity recognition (NER): a sequential approach which first produces a transcription of the document images using handwritten text recognition (HTR) techniques and then detects named entities, compared to a combined approach which produces the transcription and the entities simultaneously. Experiments are presented for the extraction of person names, locations, and dates in three languages: Latin, early new high German and old Czech.

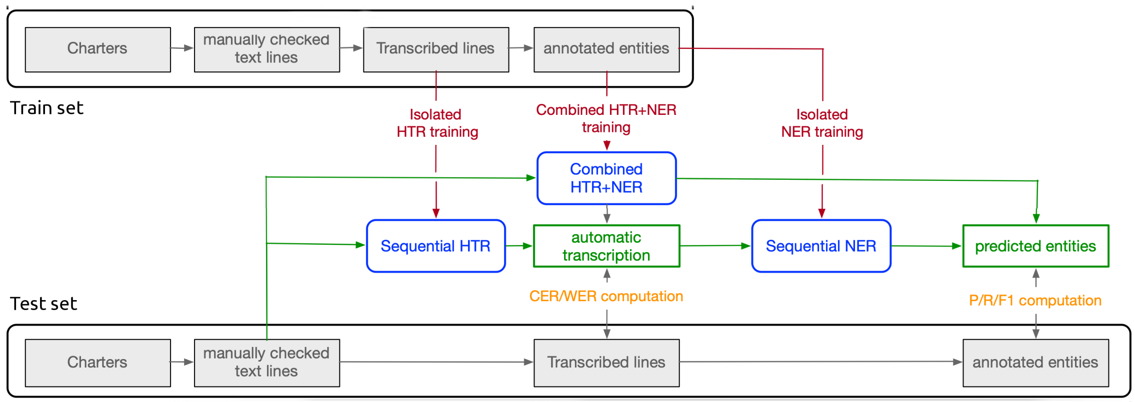



Fig. 2. Overview of the training and testing of the two approaches: sequential or combined approaches of handwritten text recognition (HTR) and named entity recognition (NER).

## II. The Annotated Handwritten Charters Corpus

The corpus of the National Archives consists of 499 charters from the Archives of the Bohemian Crown and archives of monasteries, chosen for their historic and linguistic significance for history of Central Europe. The corpus contains documents legally safeguarding the privileges of the Bohemian kingdom treaties and pledges of allegiance, marriage contracts, church administration, inheritance matters and property transfers. As a basis of the National Archives' corpus older editions were used [Friedrich, 1993], [Emler, 1892], [Hrubý, 1928]. Charters date from 1145 to 1491, with a strong focus on the reign of the Luxembourg dynasty in the 14th c. Earlier charters are included mainly from older editions, with a complete revision according to modern editorial rules following [Uhde and Hirsch, 2009] (German), [Ryba, 2009] (Latin), [Daňhelka, 1985] (Czech). We decided to focus on Vernacular, German and esp. Czech texts, that are relatively rare and less studied on an international level, and selected 124 Latin, 173 German, and 202 Czech charters, the latter ranging from the end of the 14th c. to 1491.

The charters are digitised and available via *Monasterium.net*[2]. Transkribus [Kahle et al., 2017] was used to connect images with transcriptions. Transliteration was used as the main editing method. Locations, personal names and the name of God were transcribed with capital initial letters, all other words in minuscules. Word division, punctuation, and vowels like *u/v, i/j*, were normalised, letter *w* transcribed according to the original text. Abbreviations were extended without brackets.

Transkribus was also used to tag named entities such as personal names, locations, and dates, and record the position and length on the line, and if they overlap on the next line (hereafter: "continuation"). Personal names were tagged to a full extent including first names, surnames, predicates, and titles. They often include a location, which is tagged within the personal names (they are so called "nested entities"). Identification of locations and linking them to corresponding entries in the GeoNames gazetteer is currently in progress. This task is difficult, as charters include mentions of extinct places or names that can refer to several homonymous locations. The selected corpus is only a small part of the 43,000 digitised charters from the National Archives and other archives in the Czech Republic in *Monasterium.net*, where research possibilities are limited due to a lack of full-text search options, indexes, and transcriptions. The current research could offer new and more effective research possibilities.

## III. Related Work

Statistical methods for named entity recognition, based on Conditional Random Fields, such as Stanford CoreNLP [Manning et al., 2015], are still commonly used in Digital Humanities, with great success [Torres Aguilar, 2019], [Erdmann et al., 2019]. Recent models based on neural networks [Ma and Hovy, 2016], [Lample et al., 2016], [Yadav and Bethard, 2018] are gaining popularity on historical texts [Won et al., 2018]. However, reaching high performance with neural networks requires a large amount of data which is rarely available for historical languages. Combining external resources such as gazetteers and using active learning strategies [Erdmann et al., 2019] allows reaching a good performance without manual annotation of large corpora.

NER is very sensible to text quality and the amount of training data. Measures on the NER as a downstream task after applied an Optical character recognition (OCR) tool, based on an extrinsic evaluation focused on English data with different quality levels of OCR, show the impact of OCR quality on the task itself [Hamdi et al., 2019]. The authors show that, in a sequential model, the performance of a NER can drop with at least 30 percentage points, when the character error rate (CER) of the OCR process increases from 1% up to 7%. NER's precision can lose 25% with word error rate (WER) $>$ 20% or, in this instance, CER between 3.6 and 6.3%. [Kettunen et al., 2017] show a F-score for NER between 30% and 60% for an OCR with WER $=$ 27%. [Torres Aguilar, 2019] evaluates the size of the needed training data and the proposed system achieves a precision of

[2]https://icar-us.eu/cooperation/online-portals/monasterium-net/

0.9 for the exact match of the inside tokens of an entity of type Person (I-PERS, conform with the IOB format) and 0.8 for place names (I-PLACE) on clean lemmatised Latin, with an amount of 5300 documents, charters but the performance drops to 0.72 for person names and 0.6 for locations with a model based on fewer (500) documents. In the case of sequential models, lemmatisation and stemming approaches have shown to improve the performance of NER systems for highly inflectional languages [Konkol and Konopík, 2014], but fewer studies have been made regarding the effectiveness of these methods on text produced by an OCR system, where the performance of a NER can be affected by the extreme sensitivity to OCR errors particularly in the suffix parts of words [Kettunen et al., 2017].

To reduce the impact of cascading error between HTR and NER, the current trend is to develop systems that can be trained end-to-end using fully neural architectures [Carbonell et al., 2018], [Ghannay et al., 2019]. In this paper, we evaluate both approaches on the medieval handwritten charters corpus.

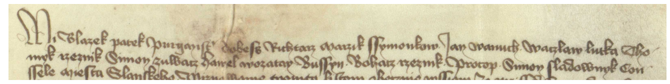
## IV. Description of the systems

In this section, the two approaches to NER studied in this paper are described. The first one consists of a sequential approach where, in a first step, the text line images are transcribed with an HTR system and then, in a second step, named entities are detected. The second one is a combined approach, in which both the transcription and the named entities are recognised simultaneously. Figure 2 presents the training and testing procedures for the two approaches. For both systems, the input is the image of a text line. For each charter in the corpus, the text lines were automatically detected using Transkribus and manually corrected.

### A. Sequential approach

*1) Handwritten text recognition:* The handwritten text recognition model is built using the Kaldi library [Arora et al., 2019]. The model consists of two main parts: the optical model and the language model. The optical model is a hybrid Deep Neural Network-Hidden Markov Model (DNN-HMM) model composed of six convolutional layers, four layers of time-delay neural networks and an output layer with softmax activation. The language model is an $n$-gram model trained on subwords generated by Byte Pair Encoding method [Sennrich et al., 2015].

*2) Named entity recognition:* We used the model proposed by [Lample et al., 2016], a deep neural network architecture for sequence labeling. This model uses character-based and word-based representations (embeddings). Additionally, we use capitalisation features that can be among these four cases: all characters in a word are lowercase, all are are capitalised, the first letter is capitalised or the word contains a capitalised letter. The character embeddings are passed through a bidirectional LSTM (BiLSTM) [Graves et al., 2009], and the output is concatenated with the word and capitalisation embeddings and fed into another BiLSTM. A Conditional random fields (CRF) is used on top to jointly decode labels for the whole sequence of words. A more detailed description of the model can be found in [Lample et al., 2016].



```
Mi <persName> Blazek Patek, purgmistr </persName>,
<persName> Dobess, richtarz </persName>, <persName>
Marzik Ssymonkow </persName>, <persName> Jan Warn
ich </persName>, <persName> Waczlaw Lutka
</persName>, <persName> Tho¬ </persName>

<persName> myk, rzeznik </persName>, <persName>
Simon Zuwacz </persName>, <persName> Hawel, woza
tay </persName>, <persName> Bussyn Bohacz, rzeznik
</persName>, <persName> Procop </persName>,
<persName> Simon, sladownyk, con¬ </persName>
```

Fig. 3. Example of the tagged text used in training in the combined approach

### B. Combined approach

This approach simultaneously transcribes the text line images and extracts the corresponding NEs. In this approach, both NE information and transcripts are used to train the system and, as can be seen in Figure 2, the output of the system has dual information: the hypothesised transcripts and the corresponding NE tags. This approach follows the ideas introduced in [Carbonell et al., 2018], [Quirós et al., 2018].

For every NE category considered in the collection (person name, location, and date) two tags have been defined: one tag to indicate that the NE starts and another one to close it (<persName>, </persName>, <placeName>, </placeName>, <date>, </date>). Then, this tagged transcription has been used to train both the optical model and the language model used during the handwritten text recognition process. Figure 3 presents an example of two tagged text lines used to train both types of models accordingly.

Formally speaking, given a handwritten line image represented by a feature vector sequence, $\mathbf{x} = x_1 x_2 \dots x_m$, the HTR problem can be formulated as the problem of finding the most likely word sequence: $\widehat{\mathbf{w}} = \widehat{w}_1 \widehat{w}_2 \dots \widehat{w}_l$,

$$\widehat{\mathbf{w}} = \arg\max_{\mathbf{w}} p(\mathbf{w} \mid \mathbf{x}) \tag{1}$$

By further considering the sequence of NE tags, $\mathbf{t} = t_1\ t_2\ \dots\ t_l$, associated to the word sequence as a hidden variable in equation (1), using the Bayes' rule to decompose the probability into optical knowledge and syntactic knowledge, and approximating the sum by the dominating term, we can rewrite the previous equation as:

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} \sum_{\mathbf{t}} p(\mathbf{w}, \mathbf{t} \mid \mathbf{x}) \tag{2}$$

$$\approx \arg\max_{\mathbf{w}} \max_{\mathbf{t}} p(\mathbf{x} \mid \mathbf{w}, \mathbf{t}) \cdot p(\mathbf{w}, \mathbf{t}) \tag{3}$$

Finally, from the decoding process, we can obtain not only the best word sequence hypothesis, but also the best sequence of NE tags of the most probable sentence:

$$(\hat{\mathbf{t}}, \hat{\mathbf{w}}) \approx \arg\max_{\mathbf{t},\mathbf{w}} p(\mathbf{x} \mid \mathbf{w}, \mathbf{t}) \cdot p(\mathbf{w}, \mathbf{t}) \tag{4}$$

The combined HTR+NER model is based on a deep neural network model that consists of a stack of several convolutional layers followed by one or more BiLSTMs. Finally, a softmax activated output layer computes an estimate for the probabilities of each character in the training alphabet, plus a special non-character symbol. The overall architecture is often referred to as *Convolutional-Recurrent Neural Networks* (CRNN) [Shi et al., 2015].

The tagging process was directly modeled at the optical level, that is, the CRNN was trained with line images and the corresponding raw tagged transcripts, where each tag is considered as one more symbol in the transcript. In this way, the number of optical units of the CRNN is increased with two extra symbols for every NE.

For modelling the syntactic and semantic information, $p(\mathbf{w}, \mathbf{t})$, a stochastic finite-state transducer $n$-gram has been used, whose input is a sequence of characters and the output is the tagged sequence. This transducer is trained with exactly the same tagged transcripts that the optical models. In order to apply the trained finite-state transducer contextual constraints to the CRNN output character probabilities, we follow one of the approaches presented in [Bluche, 2015]. The edge probabilities of the transducer are adequately combined with the CRNN output character posteriors, suitably scaled with character priors. The resulting stochastic transducer, along with the classical Viterbi decoding algorithm, is used to obtain an optimal tagging and transcription hypothesis of the original input line image.

## V. Experiments

### A. Dataset

For the following experiments in this work, we split the 499 charters in training, validation, and testing sets. In order to simplify the models and the evaluation procedure, we took into consideration neither entities that were split into different lines, nor entities that were inside other entities (nested entities), i.e. if a location was included in an entity of type person, we discarded the location annotation.

A description of the different sets in terms of the number of charters, the number of transcribed lines and the number of tokens (words) is given in Table I for the three languages. A statistical description of the entities is given in table II. It should be noted that dates are particularly long, with an average of nine words, which is unusual.

TABLE I
Dataset splits (number of pages, lines, and tokens (words))

| | Number | ***Czech*** | ***German*** | ***Latin*** | ***All*** |
|---|---|---|---|---|---|
| train | pages | 161 | 138 | 99 | 398 |
| | lines | 2,905 | 2,556 | 1,585 | 7,046 |
| | tokens | 52,708 | 60,427 | 28,815 | 141,950 |
| validation | pages | 21 | 18 | 12 | 51 |
| | lines | 300 | 252 | 150 | 702 |
| | tokens | 5,997 | 5,841 | 2,467 | 14,305 |
| test | pages | 20 | 17 | 13 | 50 |
| | lines | 381 | 388 | 229 | 998 |
| | tokens | 6,891 | 9,843 | 3,995 | 20,729 |

TABLE II
Average entity lengths in number of tokens (words) characters (chars), and entities for each set and language (nested entities and continuation ignored). ***PER=Person***, ***LOC=Location***, ***DAT=Date***

| | avg. length | ***Czech*** | ***German*** | ***Latin*** | ***All*** |
|---|---|---|---|---|---|
| ***PER*** | tokens | 2.7 | 3.4 | 2.1 | 2.7 |
| | chars | 18.1 | 23.1 | 15.7 | 18.7 |
| | count | 1,385 | 800 | 951 | 3136 |
| ***LOC*** | tokens | 1.2 | 1.0 | 1.1 | 1.1 |
| | chars | 9.2 | 7.4 | 8.3 | 8.2 |
| | count | 1,054 | 1,540 | 890 | 3,484 |
| ***DAT*** | tokens | 8.3 | 10.8 | 6.7 | 9.0 |
| | chars | 59.3 | 67.8 | 52.7 | 61.4 |
| | count | 313 | 304 | 150 | 767 |

### B. Experimental details

*1) Sequential approach:* For HTR, all the line images were grayscaled and scaled to a height of 40 pixels without applying any other preprocessing.

The hyperparameters for the BiLSTM-based model are the following: the dimensionality of the word embeddings is 300, of the character embeddings is 25, and of the capitalisation embeddings is 1. All embeddings are initialised based on a normal distribution by default, leaving the opportunity that the embeddings are learned on the task. We also experimented with pre-trained embeddings (fastText [Bojanowski et al., 2017]) for German, Latin, and Czech, but they did not bring any improvements in performance. We consider that this is due to the lack of pre-trained embeddings for historical variations of words in these low-resource languages. For regularisation, we apply a dropout of $0.5$ on the concatenation of all embeddings.

*2) Combined approach:* For carrying out the experiments, all the line images were scaled to a height of 64 pixels and contrast enhancement and noise removal was applied as in [Villegas et al., 2015]. Then, the text line images of the training partitions were used to train the corresponding CRNN. The combined system is *Laia Toolkit* [Puigcerver, 2017], a deep learning framework especially built for the task of HTR and tasks that require an HTR model. The hyperparameters for the CRNN are detailed in [Puigcerver, 2017], with the difference that, instead of using five convolutional blocks, we use just four, where the number of filters at the $n$-th convolutional layer is equal to $16n$.

The transcripts of the training (including tags) were also used to train a character-based 8-gram model with Kneser-Ney back-off smoothing. This language model, along with the trained CRNN, was subsequently used with the Kaldi [Povey et al., 2011] decoder to produce the results for all the test set image text lines.

### C. Evaluation of the handwriting recognition

To asses the quality of the transcription obtained with both approaches, we compute the character error rate (CER) and

TABLE III
AUTOMATIC HANDWRITING RECOGNITION ERROR RATES ON THE TEST SET FOR THE SEQUENTIAL AND COMBINED APPROACHES, IN CHARACTER (CER) AND WORD (WER) ERROR RATE.

| | **CER** | **WER** |
|---|---|---|
| Sequential model | 8.9 | 29.3 |
| Combined model | 8.0 | 26.8 |

TABLE IV
COMPARISON OF DIFFERENT APPROACHES AT LINE-LEVEL AND PAGE-LEVEL: AUTOMATIC NAMED ENTITY RECOGNITION ON MANUALLY TRANSCRIBED TEXT (CLEAN TEXT+NER), AUTOMATIC TRANSCRIPTION WITH THE SEQUENTIAL APPROACH (SEQUENTIAL HTR+NER) OR THE COMBINED APPROACH (COMBINED HTR+NER)

| | **Text line-level** | | | | **Page-level** | | | |
|---|---|---|---|---|---|---|---|---|
| | ***PER*** | ***LOC*** | ***DAT*** | ***All*** | ***PER*** | ***LOC*** | ***DAT*** | ***All*** |
| | **Clean text + NER** | | | | | | | |
| P | 70.45 | 86.03 | 65.96 | 77.36 | 71.39 | 86.95 | 65.96 | 78.22 |
| R | 65.72 | 53.0 | 80.52 | 59.11 | 66.61 | 57.19 | 80.52 | 62.03 |
| F1 | 68.01 | 65.59 | 72.5 | 67.01 | 68.92 | 69.0 | 72.5 | 69.19 |
| | **Sequential HTR + NER** | | | | | | | |
| P | 19.07 | 58.58 | 13.16 | 38.74 | 20.7 | 62.21 | 13.16 | 41.26 |
| R | 14.51 | 34.97 | 12.98 | 26.25 | 15.67 | 39.37 | 12.98 | 28.84 |
| F1 | 16.48 | 43.79 | 13.07 | 31.29 | 17.84 | 48.22 | 13.07 | 33.96 |
| | **Combined HTR + NER** | | | | | | | |
| P | 33.23 | 63.12 | 21.21 | 49.25 | 34.12 | 69.58 | 22.73 | 53.33 |
| R | 20.14 | 55.33 | 18.18 | 37.08 | 20.86 | 59.41 | 59.41 | 39.76 |
| F1 | 25.08 | 58.97 | 19.58 | 42.3 | 25.89 | 64.1 | 20.83 | 45.55 |

the word error rate (WER) on test text lines. The results are presented in table III. The two models yield similar performance, but the combined model is slightly better (26.8% WER versus 29.3%). Since the architecture of the two models is not identical, we can not conclude that the advantage is due to the combined training, more experiments should be done.

### *D. Evaluation of named entity recognition*

To evaluate the extraction of the NEs, we compute precision (P), recall (R), and F1 scores, micro averaged, in a coarse manner. The coarse-grained metrics are based on the exact match of the entities, disconsidering the position of the entities in text, the entities strings and types are considered correct if they match the entities in the ground truth. We distinguish between two levels of evaluation: text line-level, where the metrics are applied by comparing corresponding text lines, and page-level (a page represents a charter), where the metrics are applied on the entire corresponding pages. Due to the fact that the use of digital libraries requires easy access to documents and that named entities are among the most relevant information extents that help to properly index digital documents, we are interested only in the presence of entities in a document, more than their exact position in the text.

Table IV shows the results considering the identified entities from the test set, at text line-level and page-level. The results for the entities of type *PER*, *LOC*, and *DAT* are also detailed.

The first conclusion that we can draw from the Table IV is that the high error rate of the HTR systems has an important impact on NER: globally, F1 measure drops from 69.2 on clean text to 33.9 for sequential approach and 45.5 for the combined approach. However, the combined approach seems to be more robust to HTR errors: its WER is 2.5 percentage points lower and its F1 is 11.6 percentage points better. Two main explanations can be suggested: first, the sequential approach suffers from error propagation between HTR and NER. Second, the combined approach is trained in a multi-task setting where the same model must at the same time transcribe the text and detect the entities. Multi-task training is known for increasing the generalization power of models, especially in the case of having limited annotated data, because learning just one task has the risk of overfitting, while learning two tasks jointly enables the model to obtain a better representation through averaging the noise patterns.

TABLE V
COMPARISON OF THE DIFFERENT APPROACHES AT TEXT LINE-LEVEL AND PAGE-LEVEL PER LANGUAGE

| | **Text line-level** | | | **Page-level** | | |
|---|---|---|---|---|---|---|
| | ***Czech*** | ***German*** | ***Latin*** | ***Czech*** | ***German*** | ***Latin*** |
| | **Clean text + NER** | | | | | |
| Precision | 69.91 | 84.95 | 59.29 | 70.82 | 85.22 | 59.68 |
| Recall | 50.0 | 65.64 | 43.6 | 50.71 | 67.93 | 44.35 |
| F1 | 58.3 | 74.06 | 50.24 | 59.09 | 75.6 | 50.88 |
| | **Sequential HTR + NER** | | | | | |
| Precision | 27.51 | 52.66 | 20.74 | 29.04 | 54.0 | 21.81 |
| Recall | 13.73 | 35.74 | 11.37 | 14.54 | 37.91 | 11.99 |
| F1 | 18.31 | 42.58 | 14.69 | 19.37 | 44.55 | 15.47 |
| | **Combined HTR + NER** | | | | | |
| Precision | 41.56 | 56.26 | 42.93 | 41.99 | 61.76 | 39.79 |
| Recall | 26.09 | 48.32 | 30.15 | 26.36 | 51.81 | 27.54 |
| F1 | 32.05 | 51.99 | 35.42 | 32.39 | 56.35 | 32.55 |

Regarding the text line and page-level evaluations, one can notice that the latter is always better, this being easily explained by our evaluation scheme: an entity is considered correct if at least one of its occurrences is correct. Therefore, at the page-level, the repetition of entities improves the detection rate. As explained before, the evaluation at the page-level is more relevant from an application point of view since we aim at attaching the extracted information to the charter (page).

Table V presents the results per language. The extraction results are in general better for German than for Czech, and Latin. The limited amount of data in Latin may explain the lowest score in this language. Both systems perform better on German, most probably because, comparing with Czech and Latin, it is the least inflected language in the corpus. Comparisons with NER on modern, not inflected languages (especially English) or with medieval inflected, but lemmatised Latin show comparable results.

We also notice that the performance of both models decreases considerably in the case of the entities of type ***DAT***, at text line-level, for around 20 percentage points, and around 10 percentage points, and at page-level, for more than 50 percentage points, for the sequential model, and the combined model, respectively. We consider that this is due to the fact that, since the average length of the ***DAT*** entities oscillates between 6 and 10 words, while the other types of entities have between one and three words (see Table II), and the evaluation is based on strict string match between ground-truth and the results, the probability of an incorrect transcribed word inside an entity in extremely high, and thus, the decrease in performance.

## VI. Conclusion and future work

We presented a comparison of sequential and combined approaches for named entity recognition in a corpus of handwritten medieval charter images automatically transcribed. The latter, trained end-to-end, showed the best performances for the three languages and for all types of entities. To address a multilingual, linguistically challenging corpus, the implementation of a language detection module and a lemmatisation step for Czech and Latin will have to be tested. As demonstrated by the related work, different strategies can also be implemented to enhance the results, such as reducing the WER and increasing the training data for named entities. Furthermore, continued entities (the entities that are split into two or more text lines) should obviously be taken into account. Nested entities can create an issue with double labelling and the evaluation, but they are clearly needed for the understanding of the documents. For this, in turn, we consider that the most important necessity is the identification of named entities and their linking to gazetteers and dictionaries of people.

## Acknowledgments

This work is part of the *HOME History of Medieval Europe* research project, supported by the European JPI Cultural Heritage and Global Change (Grant agreements No. ANR-17-JPCH-0006 for France, MSMT-2647/2018-2/2, id. 8F18002 for Czech Republic and PEICTI Ref. PCI2018-093122 for Spain).

## References

[Arora et al., 2019] Arora, A., Chang, C. C., Rekabdar, B., BabaAli, B., Povey, D., Etter, D., Raj, D., Hadian, H., Trmal, J., Garcia, P., et al. (2019). Using asr methods for ocr. In *International Conference on Document Analysis and Recognition*.

[Bluche, 2015] Bluche, T. (2015). *Deep Neural Networks for Large Vocabulary Handwritten Text Recognition*. PhD thesis, Ecole Doctorale Informatique de Paris-Sud.

[Bojanowski et al., 2017] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*.

[Carbonell et al., 2018] Carbonell, M., Villegas, M., Fornés, A., and Lladós, J. (2018). Joint recognition of handwritten text and named entities with a neural end-to-end model. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*.

[Daňhelka, 1985] Daňhelka, J. (1985). Směrnice pro vydávání starších českých textů. *Husitský Tábor*, 8.

[Emler, 1892] Emler, J. (1892). *Regesta diplomatica nec non epistolaria Bohemiae et Moraviae*, volume IV.

[Erdmann et al., 2019] Erdmann, A., Wrisley, D. J., Allen, B., Brown, C., Cohen-Bodénès, S., Elsner, M., Feng, Y., Joseph, B., Joyeux-Prunel, B., and de Marneffe, M.-C. (2019). Practical, Efficient, and Customizable Active Learning for Named Entity Recognition in the Digital Humanities. In *Proceedings of NAACL-HLT 2019*.

[Friedrich, 1993] Friedrich, G. (1904–1993). *Codex diplomaticus et epistolaris (epistolarius) regni Bohemiae*.

[Ghannay et al., 2019] Ghannay, S., Caubriere, A., Esteve, Y., Camelin, N., Simonnet, E., Laurent, A., and Morin, E. (2019). End-To-End Named Entity and Semantic Concept Extraction from Speech. In *IEEE Spoken Language Technology Workshop*.

[Graves et al., 2009] Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., and Schmidhuber, J. (2009). A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Transaction on PAMI*, 31(5).

[Hamdi et al., 2019] Hamdi, A., Jean-Caurant, A., Sidere, N., Coustaty, M., and Doucet, A. (2019). An Analysis of the Performance of Named Entity Recognition over OCRed Documents. In *Joint Conference on Digital Libraries*.

[Hrubý, 1928] Hrubý, V. (1928). *Archivum coronae regni Bohemiae*, volume II, Inde ab a. MCCCXLVI usque ad a. MCCCLV.

[Kahle et al., 2017] Kahle, P., Colutto, S., Hackl, G., and Mühlberger, G. (2017). Transkribus - a service platform for transcription, recognition and retrieval of historical documents. In *International Conference on Document Analysis and Recognition*.

[Kettunen et al., 2017] Kettunen, K., Mäkelä, E., Kuokkala, J., Ruokolainen, T., and Niemi, J. (2017). Modern tools for old content-in search of named entities in a finnish ocred historical newspaper collection 1771-1910. *DHQ: Digital Humanities Quarterly*, 11(3).

[Konkol and Konopík, 2014] Konkol, M. and Konopík, M. (2014). Named entity recognition for highly inflectional languages: effects of various lemmatization and stemming approaches. In *International Conference on Text, Speech, and Dialogue*.

[Lample et al., 2016] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

[Ma and Hovy, 2016] Ma, X. and Hovy, E. (2016). End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *54th Annual Meeting of the Association for Computational Linguistics*.

[Manning et al., 2015] Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2015). The Stanford CoreNLP Natural Language Processing Toolkit. In *Annual Meeting of the Association for Computational Linguistics*.

[Povey et al., 2011] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., and Vesely, K. (2011). The Kaldi speech recognition toolkit. In *ASRU*.

[Puigcerver, 2017] Puigcerver, J. (2017). Are multidimensional recurrent layers really necessary for handwritten text recognition? In *International Conference on Document Analysis and Recognition*.

[Quirós et al., 2018] Quirós, L., Bosch, V., Serrano, L., Toselli, A. H., and Vidal, E. (2018). From HMMs to RNNs: Computer-assisted transcription of a handwritten notarial records collection. In *Int. Conf. on Frontiers in Handwriting Recognition*.

[Ryba, 2009] Ryba, K. (2009). Pravidla pro transkripci latinských literárních rukopisných textů. Technical report, Masaryk University, Brno.

[Sennrich et al., 2015] Sennrich, R., Haddow, B., and Birch, A. (2015). Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.

[Shi et al., 2015] Shi, B., Bai, X., and Yao, C. (2015). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *CoRR*, abs/1507.05717.

[Torres Aguilar, 2019] Torres Aguilar, S. (2019). *A model of automatic named entities recognition and textual structures for Latin medieval diplomatic corpora.* Theses, Université Paris-Saclay.

[Uhde and Hirsch, 2009] Uhde, K. and Hirsch, V. (2009). Grundsätze für die textbearbeitung im fachbereich historische hilfswissenschaften. Technical report, Archivschule Marburg.

[Villegas et al., 2015] Villegas, M., Romero, V., and Sánchez, J. A. (2015). On the modification of binarization algorithms to retain grayscale information for handwritten text recognition. In *Iberian Conference on Pattern Recognition and Image Analysis*.

[Won et al., 2018] Won, M., Murrieta-Flores, P., and Martins, B. (2018). Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora. *Frontiers in Digital Humanities*, 5(March).

[Yadav and Bethard, 2018] Yadav, V. and Bethard, S. (2018). A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. In *Proceedings of the 27th International Conference on Computational Linguistics*.