Geotagging a Diachronic Corpus of Alpine Texts: Comparing Distinct Approaches to Toponym Recognition

Tannon Kew, Anastassia Shaitarova, Isabel Meraner, Janis Goldzycher
Simon Clematide and Martin Volk

Institute of Computational Linguistics, University of Zurich, Switzerland
{*tannon.kew, anastassia.shaitarova, isabel.meraner, janis.goldzycher*}*@uzh.ch*,
{*simon.clematide, volk*}*@cl.uzh.ch*

Abstract

Geotagging historic and cultural texts provides valuable access to heritage data, enabling location-based searching and new geographically related discoveries. In this paper, we describe two distinct approaches to geotagging a variety of fine-grained toponyms in a diachronic corpus of alpine texts. By applying a traditional gazetteer-based approach, aided by a few simple heuristics, we attain strong high-precision annotations. Using the output of this earlier system, we adopt a state-of-the-art neural approach in order to facilitate the detection of new toponyms on the basis of context. Additionally, we present the results of preliminary experiments on integrating a small amount of crowdsourced annotations to improve overall performance of toponym recognition in our heritage corpus.

1 Introduction

Identifying spatial information in cultural and historical corpora is a crucial step in putting these texts on the map. Geotagging describes the task of establishing the connection between textual mentions of geographic locations, also known as toponyms, with geographic information systems (Amitay et al., 2004; Lieberman et al., 2010). This natural language processing (NLP) task provides the essential information required for location-based search queries (Moncla et al., 2014) and has thus found many uses in diverse fields, such as geography, question answering, bio-medicine and digital humanities, among others.

Geotagging consists of two main tasks. First, toponym mentions must be identified in text. This step, known as toponym recognition, can be seen as a location-oriented subtask of named entity recognition (NER), which is more broadly interested in identifying and classifying textual mentions of various entities such as people, locations and organisations. Second, an identified toponym needs to be linked to its true geographic referent via a unique identifier stored in an external knowledge base. This step, which we refer to as toponym resolution, often requires disambiguating identified toponyms in order to establish that link.[1]

When it comes to handling historical and cultural documents, lexical and orthographic shifts in language as well as political and administrative changes make geotagging, and NER in general, particularly challenging. In this paper, we describe our heritage corpus of alpine and mountaineering texts (Section 2), which poses unique challenges due to its domain and diachronicity. We detail two distinct efforts in geotagging a variety of fine-grained toponyms in this corpus, focusing primarily on German[2]. First, we give an overview of an earlier gazetteer-based approach oriented towards high-precision annotations (Section 3). Then, we present a current state-of-the-art neural approach which takes advantage of recent advances in sequence labelling techniques, the high-precision gazetteer-based output and a small number of

---

[1]In the relevant literature, both toponym recognition and toponym resolution have many guises which are used somewhat interchangeably (cf. Moncla et al., 2014; Gritta et al., 2018). We follow the naming conventions provided by Leidner (2007) and Lieberman et al. (2010) and adopt the term 'geotagging' to refer to the combined task of toponym recognition and resolution.

[2]Neural experiments involving French are currently underway.

crowdsourced annotations (Section 4). Finally, we evaluate and compare these two approaches and provide results from preliminary experiments on our annotation platform which aims to bring humans into the loop for geotagging heritage data (Section 5).

2 A Heritage Corpus of Alpine Texts

Our corpus consists of over 150 years of alpine and mountaineering articles. The majority of the corpus comes from the yearbooks published by the Swiss Alpine Club (SAC) between 1864 and 2015. Also included in the corpus are the articles from the journal of the British Alpine Club from 1969 to 2008. As such, it is a largely multilingual corpus containing texts in German, French, English, Italian and Romansh. Table 1 provides an overview of general corpus statistics for each language. The corpus is rich in geographic references and is a valuable resource detailing many aspects of life in the mountains. Topics covered range from mountaineering and hiking expeditions, flora and fauna, geography, and geological changes to social, cultural and linguistic diversity in alpine regions.

Being a diarchronic heritage corpus, its development has posed a number of challenges regarding digitisation and linguistic annotation. Numerous experiments have been undertaken to semantically enrich this corpus as both a historic and a linguistic resource. These include, but are not limited to, a novel approach to correcting optical character recognition (OCR) errors (Clematide et al., 2016); gazetteer and rule-based NER for the annotation of personal names, toponyms, organisations and time expressions (Ebling et al., 2011); improved lemmatisation for German separable prefix verbs and elliptical compound nouns (Volk et al., 2016); innovative techniques for sentence alignment in parallel texts (Sennrich and Volk, 2010); and the creation of a manually annotated parallel treebank with more than 1000 sentences in French and German for the purpose of assisting statistical machine translation (Göhring and Volk, 2011).

3 A Gazetteer-Based Approach to Geotagging a Heritage Corpus

Earlier work in geotagging our heritage corpus has relied largely on a gazetteer-based approach, targeting fine-grained geographic categories, such

Figure 1: A page from the 1906 SAC yearbook with the article "On the research of mountain names".

as towns, mountains, lakes, glaciers, valleys and mountain cabins in order to enable location-based searching. Gazetteer-based approaches make use of curated lists of known geographic locations along with their relevant metadata (e.g. longitude and latitude, population, elevation, etc.).

Relying on gazetteers poses two major challenges. On the one hand, simple string matching with gazetteer entries can result in a high count of false positives due to the fact that toponyms often overlap with common nouns (e.g. Bath) and personal names (e.g. Washington). On the other hand, gazetteers are inherently incomplete and thus suffer from a lack of coverage, resulting in a high count of false negatives (Davari et al., 2019; Magge et al., 2018). However, gazetteers simplify the task of linking toponym mentions to their real-world referents. As such, they are typically unavoidable for automatically assigning geographic coordinates to a given toponym and have been a popular choice in geotagging historical and toponym-dense corpora (Won et al., 2018;

language	texts	tokens	token types	lemma types
German	12.8k	23.4m	769k	325k
French	12.8k	22.3m	418k	96k
Italian	0.16k	0.32m	39k	18k
Romansh	0.01k	0.014m	4k	0.2k
English	1.5k	6.5m	181k	60k

Table 1: Overview of our heritage corpus of alpine texts.

Moncla et al., 2014).

Our gazetteers were sourced from the Swiss Federal Office of Topography (SwissTopo)[3] and the community-based resource GeoNames[4]. We relied primarily on SwissTopo for identifying locations in Switzerland, while GeoNames was used to account for names of foreign mountain ranges and peaks that also frequently occur in the corpus.

Due to Switzerland's multilingual landscape, the same geographic entity often has different names in the local languages and dialects. For example, the *Matterhorn*, which straddles Switzerland and Italy, is also commonly referred to by its Italian name *Cervino* or the French name *Cervin*. In SwissTopo, however, names are listed only in the official language of the region, emphasising the lack of coverage associated with gazetteer-based approaches. As such, we have adopted numerous techniques to supplement our gazetteers.

Using the corpus itself as a resource, we extracted all words with suffixes typically denoting certain toponyms. For example, in German, mountain names often end with *-horn*, *-grat* and *-stock*. We then filtered out those mentions which are homographs of common nouns using a list collected from an online German dictionary[5]. The remaining unmatched words were then manually validated before being subsequently added to the gazetteer. Similarly, words which often serve as the first component of a multiword toponym in other languages, such as French *Cabane* and Italian *Rifugio* denoting cabins and French *Aiguille* and Romansh *Piz* for mountain peaks, were also leveraged to extract additional toponym candidates to extend gazetteer coverage. Furthermore, the most important exonyms (e.g. *Genfersee* for *Lac Léman*) and

known spelling variants, such as hyphenated forms (e.g. *Monte-Rosa* for *Monte Rosa*) were also added.

Aside from additions, certain toponyms were also removed from the gazetteers in order to reduce the number of false positives. Since common nouns are capitalised in German, we deleted some toponyms which are homographs of frequent common nouns (e.g. *Nase* (nose)) as well as a handful of generic nouns denoting places (e.g. *Alptal* (alpine valley)).

Lastly, we extended toponym recognition beyond simple string matching to account for adjectives and prepositions that commonly occur in compound toponyms and can be inflected or even abbreviated (e.g. *gross* (big) in *Grosser Mythen*). Normalisation, lemmatization and decompounding were applied to address other inflections, such as genitive (*Bern - Berns*) and plural (*Fergenhorn - Fergenhörner*) forms (Volk et al., 2009).

3.1 Heuristics for Toponym Disambiguation

Despite the cautious crafting of gazetteers, not all ambiguities could be avoided. As such, we complemented our gazetteer-based approach with some simple heuristics to leverage internal contextual clues in an attempt to resolve both geo/non-geo and geo/geo ambiguities.

Geo/non-geo ambiguity occurs when a toponym has the same surface form as a non-toponym in the language (Amitay et al., 2004). For example, *Mönch* can refer to a mountain in Switzerland and the German word for 'monk'. A simple heuristic that helps to disambiguate between a toponym and a common noun is the occurrence of a preceding indefinite article. If the German article *ein* precedes *Mönch* in the same noun phrase, then it can only refer to the inhabitant of a monastery since it is not possible to use an indefinite article with a unique (i.e. definite) toponym in German.

Geo/geo ambiguities arise when a single name has numerous real-world referents, for example, there are 17 different mountain peaks with the name *Schwarzhorn* in Switzerland alone. In general, an effective method for determining the correct referent of an ambiguous toponym is to use simple external heuristics based on prominence measures (e.g. population count) (Leidner and Lieberman, 2011), however, when dealing with fine-grained toponyms, such as mountain cabins or hiking trails (see Moncla et al., 2014), such measures are not suitable. In mountaineering and alpine texts, however, certain types of toponyms are often followed by their elevation (e.g. *Schwarzhorn, 3207 m*). Using this information, we resolved ambiguous toponyms by selecting the candidate from our gazetteer that has the closest elevation measurement. We restricted candidate selection to consider only the referents where the elevation does not deviate more than 50 metres from the measurement specified in the text. This prevents assigning the wrong geolocation given potentially erroneous information in the text. For those toponyms that could not be disambiguated with this approach, we assigned a placeholder referent ID. Toponyms listed in our gazetteer without a predefined ID from SwissTopo or Geonames received a 'null' ID. Table 2 displays the distribution of detected toponyms throughout the entire corpus. Here, linked IDs account for disambiguated toponyms while distinct types indicate the number of unique toponym IDs.

	total linked	total ambig.	total null	distinct types
cabin	20k	–	9k	0.3k
city	212k	20k	–	2k
glacier	15k	2k	10k	0.4k
lake	7k	0.7k	5k	0.3k
mountain	234k	72k	59k	8k
valley	28k	2k	19k	0.6k

Table 2: Corpus-wide toponym recognition and resolution ID counts for the gazetteer-based approach.

This gazetteer-based approach to toponym recognition, assisted by simple heuristics for toponym resolution, scores relatively high precision among most toponym types, yet suffers from low recall, particularly among frequent categories (see Table 4). Manual inspection reveals that prevalent historical spelling variations (e.g. *Fiesch* and pre-1905 *Viesch*) and the extensive use of endonyms for places in and around Switzerland (e.g. German: *Etsch* and Italian: *Adige*) remain a major challenge in identifying all toponyms in this corpus.

4 A Neural Approach to Geotagging a Heritage Corpus

In an attempt to improve upon the previous gazetteer-based approach, we adopt a neural approach to toponym recognition, inspired by current state-of-the-art techniques in NER. Similar to Huang et al. (2015); Ma and Hovy (2016), we apply a bidirectional recurrent neural network architecture (BiLSTMs) followed by a conditional random field (CRF) layer for the detection of toponyms in our corpus. This approach incorporates contextual string embeddings, as introduced by Akbik et al. (2018) in the Flair framework. Contextual string embeddings have been shown to perform well in downstream sequence-labelling tasks, such as NER. Due to the fact that they comprise character-level information, these embeddings are particularly well suited to modelling an open vocabulary.

We use a stacked embedding architecture (Akbik et al., 2018) to concatenate Flair's character-based forward embeddings trained on our corpus with general-purpose fastText word embeddings pre-trained on web data (Grave et al., 2018). The idea here is that the general-purpose embeddings provide sufficient global syntactic knowledge, while the in-domain Flair embeddings, which are trained on our own corpus, capture high-level contextual and orthographic idiosyncrasies typical for historical mountaineering reports.

4.1 Exploiting Gazetteer-based Labels as Silver Standard Training Data

Typically, the major challenge in machine learning and neural approaches to toponym recognition is acquiring domain-specific labelled training data (Davari et al., 2019). Therefore, we exploit the output of the gazetteer-based system described in Section 3, which acts as a low-effort silver standard and avoids having to rely on
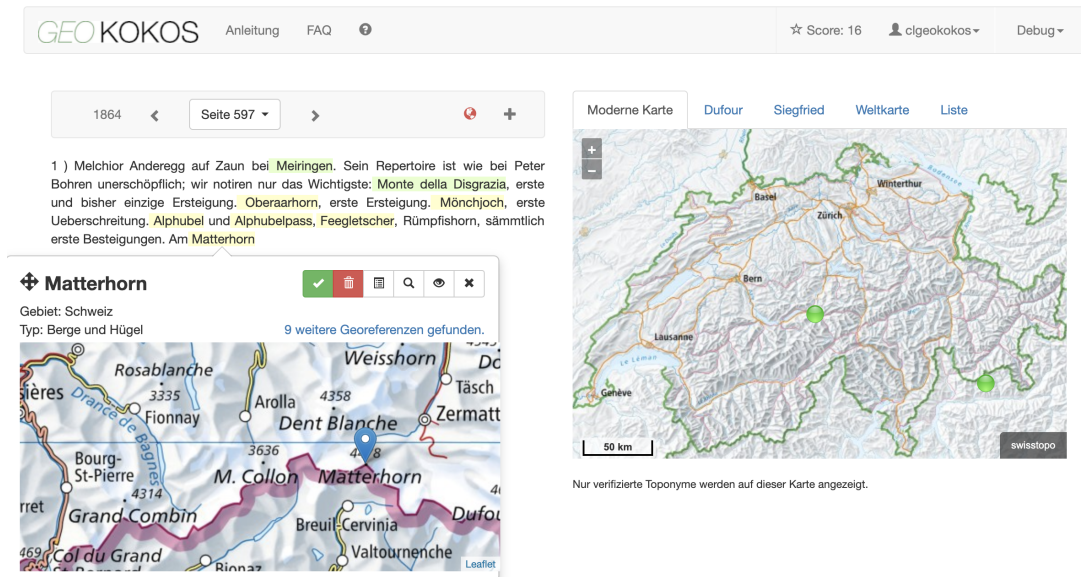
Figure 2: Screenshot of the GeoKokos citizen science web application.

time-consuming manual annotations. We extract sentences with at least one toponym from eight nineteenth-century yearbooks as the basis of our training data.

To increase the overall quality of the training material, we also integrate 1000 manually annotated sentences from two twentieth-century yearbooks. Finally, we extend the gazetteer-based annotations in the training set to cover toponym categories for rivers and regions, which had previously been ignored in our earlier approach, due to the fact that these entities are associated with multiple geographic coordinates.

The result of these steps is a hybrid 'silver-gold', toponym-dense training set consisting of approximately 15,000 sentences with more than 28,000 high-precision labelled toponyms. Using this initial training set, we establish a robust neural baseline model (see Table 4). For model tuning, we extract approximately 2,000 sentences, each with at least one toponym, from an additional yearbook as a silver development set.

4.2 Turning Silver to Gold with GeoKokos

In order to improve our neural model for toponym recognition, we aim to enhance the initial silver-gold training data by incorporating crowdsourced human annotations collected through the GeoKokos platform[6]. GeoKokos is a citizen science initiative designed to bring

humans into the loop in order to resolve toponym recognition and resolution challenges that are problematic for fully automated techniques.

With GeoKokos we adopt a similar technique to Clematide et al. (2016), which made use of crowdsourced corrections for OCR errors. However, instead of relying entirely on human correction of the system output as done by Clematide et al. (2016), here, we iteratively collect new human annotations and use these to re-train our neural model. Then, updated model predictions are fed back into the system which, in turn, further assist users with their annotations.

GeoKokos provides an easy-to-use web interface (see Figure 2) that gives people the opportunity to select pages from certain yearbooks for reading and annotating. After registering an account, users have the option to delete, add, update and verify toponym annotations. Updating allows the user to alter the toponym category of an existing annotation, while interactive maps enable the user to verify and mark a toponym's location. For each correction action performed, users are awarded points. As noted by Clematide et al. (2016), even such a simple point-based statistic provides an aspect of gamification in linguistic annotations (Chamberlain et al., 2013) and, of course, friendly competition amongst users, which can be a strong motivator for those interested in the task or the content itself.

Through this process, an ever-growing number

---

[6] https://geokokos.ch

of crowd corrections will gradually turn our initial silver standard training set to gold, providing improved training data and hopefully resulting in ever better models for toponym recognition in our heritage alpine corpus.

4.3 Assisting Crowdsourced Toponym Resolution

For toponym resolution, we rely solely on the crowdsourced verification actions, which provide us with reliable geotagged annotations for those yearbooks available on GeoKokos. However, to assist users, we also incorporate a map-based approach (Buscaldi, 2011) for ranking potential geographic candidates. This algorithm assumes that textual proximity implies geographic proximity and consists of two main steps: candidate extraction and candidate ranking. First, we normalise all predicted toponyms in the text and extract a list of potential candidates from a database using string matches. Then, each candidate pair is scored and ranked according to their relative geographical proximity. Given all the toponym mentions on a single page, the geographically closest candidates receive the highest scores and are suggested to the user. This simple technique thus offers users a shortcut for verifying a toponym annotation, reducing the number of required clicks.

4.4 A Pilot Experiment with Crowd-Corrected Training Data

Since the ultimate goal of this approach is to iteratively learn and to benefit from the crowd-corrected annotations as early as possible, we conduct a pilot experiment based on a small amount of human annotations collected from the GeoKokos platform so far. Here, we expect the neural model to exhibit a high adaptive capacity in terms of recall, even when dealing with a small amount of crowdsourced annotations.

Table 3 displays the distribution of crowd correction actions performed on the two yearbooks chosen for this experiment. In total, we collected 831 corrections. Since our baseline training data consists of approximately 15,000 toponym-dense sentences sampled from the yearbooks available on GeoKokos, only 291 corrections were incorporated into the training data for this experiment.[7] Using these same

action	1864	1874	total
untouched	2,467	4,185	**6,652**
others	39,317	57,826	**97,143**
verified	310	223	**533**
added	54	66	**120**
updated	45	79	**124**
deleted	35	19	**54**
total	**444**	**387**	**831**

Table 3: Crowd corrections for the SAC yearbooks from 1864 and 1874. 'Untouched' refers to silver standard toponym labels not changed by a user, while 'others' accounts for the remaining unlabelled tokens in these yearbooks.

sentences for training ensures comparability between the two models and allows us to assess the influence of these crowd-corrections.

5 Evaluation

Using a manually annotated gold standard of approximately 1,300 sentences created specifically for this task, we evaluate the two distinct approaches to toponym recognition in our heritage corpus of alpine texts. A direct comparison between the earlier gazetteer-based approach and our recent experiments in applying state-of-the-art neural models shows a promising increase in overall performance. Table 4 shows the evaluation scores per toponym label for the gazetteer-based approach, the neural baseline approach and our pilot experiment, which incorporates a small number of crowdsourced corrections.

The gazetteer-based approach shows strong performance for CABIN, LAKE and GLACIER, categories that are infrequent and tend to have rather regular and unambiguous suffixes in German (e.g. *-hütte* (cabin), *-see* (lake), *-gletscher* (glacier)). However, this approach suffers from low recall among more frequent categories where toponyms are generally more diverse. Comparing these results to the neural baseline, we see an increase in recall for the most frequent toponym categories MOUNTAIN (+6 p.p.) and CITY (+2 p.p.). This underlines the generalisation capacity of the neural approach to

---

[7]Since annotations are stored in IOB format, a correction

implies that an O (outside) label was changed to a B (beginning) or I (inside) label via an addition action, or a B or I label was changed to an O label via a deletion action.

		Gazetteer-based			Neural Baseline			Neural+annotations		
Category	Freq.	Prec	Rec	F_1	Prec	Rec	F_1	Prec	Rec	F_1
CABIN	3	1.0	1.0	1.0	0.50	0.34	0.40	0.75	1.0	0.86
CITY	89	0.74	0.61	0.67	0.78	0.63	0.70	0.77	0.68	0.72
GLACIER	20	0.95	0.80	0.87	0.95	0.80	0.87	0.95	0.85	0.90
LAKE	6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
MOUNTAIN	225	0.92	0.62	0.74	0.93	0.68	0.79	0.88	0.75	0.81
REGION*	49	0.89	0.47	0.62	0.85	0.45	0.59	0.88	0.45	0.60
RIVER*	9	0.50	0.12	0.19	1.0	0.23	0.37	1.0	0.45	0.62
VALLEY	47	0.95	0.79	0.87	0.98	0.79	0.88	0.96	0.83	0.89
Micro average		**0.88**	**0.63**	**0.73**	**0.90**	**0.66**	**0.76**	**0.87**	**0.71**	**0.79**

Table 4: Precision, recall, and F_1-score for all toponym categories. *For the purpose of this evaluation, annotations for REGION and RIVER have been added to the original gazetteer-based approach using a gazetteer-lookup on the basis of normalised string matches.

toponym recognition in combination with context-sensitive string embeddings (Akbik et al., 2018) given sufficient training data.

Inspecting the results of our pilot experiment shows that the neural model is highly sensitive, even to a small number of integrated crowdsourced corrections. Taking into account the *n*-best predictions allows the model to output an alternative toponym label, which may not have received the highest probability, thus making it possible to flexibly increase the recall. Setting a probability threshold of 0.7 yielded the best results for this experiment. Here, we observe a noticeable improvement in recall across the board, resulting in a promising increase in the overall micro F_1-score on top of the neural baseline (+3 p.p.) and the gazetteer-based approach (+6 p.p.). These results indicate that a small amount of human-corrected annotations incorporated into the training set has a positive effect on the model's ability to identify new toponyms in the corpus. We expect that including more sentences containing crowd-corrected annotations in the training data will further improve the neural model.

6 Conclusion

In this paper we have presented two distinct approaches to geotagging a range of fine-grained toponyms in a heritage corpus of alpine texts. The

goal of the earlier gazetteer-based approach was to achieve high precision for the purpose of reliable location-based searching. Consequently, this method also exhibited high counts of false negatives, resulting in low recall scores. This highlights the major shortcomings of relying on a finite list of known entities. Our new and ongoing approach relies largely on the previous work done in geotagging our corpus, while attempting to address the problem of low recall. Applying state-of-the-art contextual string embeddings and a neural model for toponym recognition allows us to attain a flexible system capable of predicting whether a given word constitutes a toponym based on the surrounding context, rather than relying solely on a list of previously known entities. Comparing these two approaches shows that the neural model outperforms the gazetteer-based system. Additionally, we have shown that the neural model responds well to even a small amount of crowd-corrected annotations. The GeoKokos platform enables us to efficiently bring humans into the loop to gradually turn our silver-standard training data to gold through an iterative learning process. As a result, it brings humans and machine learning techniques together, working hand-in-hand to improve geotagging in our heritage corpus.

References

Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual String Embeddings for Sequence Labeling. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*. Santa Fe, NM, USA, pages 1638–1649.

Einat Amitay, Nadav Har'El, Ron Sivan, and Aya Soffer. 2004. Web-a-Where: Geotagging Web Content. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Sheffield, UK, pages 273–280.

Davide Buscaldi. 2011. Approaches to disambiguating toponyms. *SIGSPATIAL Special* 3:16–19.

John Chamberlain, Karën Fort, Ugo Kruschwitz, Mathieu Lafourcade, and Massimo Poesio. 2013. Using Games to Create Language Resources: Successes and Limitations of the Approach. In Iryna Gurevych and Jungi Kim, editors, *Theory and Applications of Natural Language Processing*, Springer, page 42.

Simon Clematide, Lenz Furrer, and Martin Volk. 2016. Crowdsourcing an OCR Gold Standard for a German and French Heritage Corpus. In *Proceedings of the 10th Language Resources and Evaluation Conference (LREC)*. Portorož, Slovenia.

MohammadReza Davari, Leila Kosseim, and Tien D. Bui. 2019. Toponym Identification in Epidemiology Articles - A Deep Learning Approach. *arXiv preprint* arXiv:1904.11018.

Sarah Ebling, Rico Sennrich, David Klaper, and Martin Volk. 2011. Digging for Names in the Mountains: Combined Person Name Recognition and Reference Resolution for German Alpine Texts. In *Proceedings of the 5th Language & Technology Conference (LTC)*. Poznan, Poland.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning Word Vectors for 157 Languages. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*. Miyazaki, Japan.

Milan Gritta, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. 2018. What's Missing in Geographical Parsing? *Language Resources and Evaluation* 52(2):603–623.

Ann Göhring and Martin Volk. 2011. The Text+Berg Corpus: An Alpine French-German Parallel Resource. In *Proceedings of the 18th Traitement Automatique des Langues Naturelles Conference (TALN)*. Montpellier, France.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv preprint* arXiv:1508.01991.

Jochen L. Leidner. 2007. *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names*. Ph.D. thesis, University of Edinburgh, Edinburgh, UK.

Jochen L. Leidner and Michael D. Lieberman. 2011. Detecting Geographical References in the Form of Place Names and Associated Spatial Natural Language. *SIGSPATIAL Special* 3(2):5–11.

Michael D. Lieberman, Hanan Samet, and Jagan Sankaranarayanan. 2010. Geotagging with Local Lexicons to Build Indexes for Textually-Specified Spatial Data. In *Proceedings of the 26th International Conference on Data Engineering (ICDE)*. pages 201–212.

Xuezhe Ma and Eduard Hovy. 2016. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany, pages 1064–1074.

Arjun Magge, Davy Weissenbacher, Abeed Sarker, Matthew Scotch, and Graciela Gonzalez-Hernandez. 2018. Deep Neural Networks and Distant Supervision for Geographic Location Mention Extraction. *Bioinformatics* 34(13):i565–i573.

Ludovic Moncla, Walter Renteria-Agualimpia, Javier Nogueras-Iso, and Mauro Gaio. 2014. Geocoding for Texts with Fine-grain Toponyms: An Experiment on a Geoparsed Hiking Descriptions Corpus. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Dallas, Texas, pages 183–192.

Rico Sennrich and Martin Volk. 2010. MT-Based Sentence Alignment for OCR-Generated Parallel Texts. In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas (AMTA)*. Denver, Colorado.

Martin Volk, Noah Bubenhofer, Adrian Althaus, and Maya Bangerter. 2009. Classifying Named Entities in an Alpine Heritage Corpus. *Künstliche Intelligenz (KI) 4* pages 40–43.

Martin Volk, Simon Clematide, Johannes Graën, and Phillip Ströbel. 2016. Bi-Particle Adverbs, PoS-Tagging and the Recognition of German Separable Prefix Verbs. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS)*. Bochum, Germany.

Miguel Won, Patricia Murrieta-Flores, and Bruno Martins. 2018. Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora. *Frontiers in Digital Humanities* 5:2.