Wolf, Sebastian Johannes; Henrich, Andreas; Blank, Daniel

Characterization of toponym usages in texts

Date of secondary publication: 17.02.2025

Accepted Manuscript (Postprint), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-1064841

Primary publication

Wolf, Sebastian Johannes; Henrich, Andreas; Blank, Daniel (2014): Characterization of toponym usages in texts, in: Chris Jones und Ross Purves (Ed.), GIR '14 : Proceedings of the 8th Workshop on Geographic Information Retrieval, New York: ACM, doi: 10.1145/2675354.2675703.

Characterization of Toponym Usages in Texts

Sebastian Johannes Wolf
SAP SE, Walldorf, Germany
sebastian.wolf@sap.com

Andreas Henrich
University of Bamberg, Germany
andreas.henrich@uni-bamberg.de

Daniel Blank
University of Bamberg, Germany
daniel.blank@uni-bamberg.de

ABSTRACT

Toponyms in texts and search queries are often used figuratively and do not directly refer to the locations they reference in their literal sense. Different usage kinds and stylistic devices characterize toponym usages in texts. It is thus crucial for a Geographic Information Retrieval (GIR) system to precisely distinguish these different toponym usages at indexing and at query time in order to best address a given *information need* and the *geospatial footprint* of a document.

For that purpose, we analyze which of the classic stylistic devices such as allegories, metaphors, or metonymies are used together with toponyms. We use these categories as a foundation for a systematic approach towards the *characterization of toponym usages in texts* which we believe is necessary to further boost retrieval effectiveness of future GIR systems. A prototype implements this characterization exemplary for texts written in German. We evaluate the effectiveness of our approach against a reference corpus to show the general feasibility. Our approach provides a basis for a wide range of more sophisticated applications such as for example text genre detection.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Linguistic processing*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Selection process*

General Terms

Experimentation, Languages, Measurement

Keywords

Computational Linguistics, Linguistic Devices, Geographical Information Retrieval, Geospatial Grounding

1. INTRODUCTION

A central challenge to any information retrieval system is to fulfill the *information needs* that its users have when they perform a search. Even if the system can correctly determine the context of a particular query, it is still difficult to obtain the *relevant* documents and return them to the user. This is especially true when the information need contains a geographical reference. Both, search queries and indexed documents can contain geospatial information explicitly and implicitly, for example concrete location names as *toponyms*, the user's current location, or the location where a document was written.

Different challenges arise not only in the task of toponym detection, that is, to differentiate between location references and other terms. Also the resolution of toponyms, that is, the allocation of concrete coordinates to toponyms, has for a long time been a research topic and progress in finding effective solutions has been made. However, as a next step for further improving the effectiveness of GIR applications, the analysis on *how* a certain toponym is used has to be addressed. This has so far only been covered by a small amount of publications. The *characterization of toponym usages* can for example be of vital importance to a correct spatial grounding, when certain location names are not used literally but in a figurative sense referencing for example a government or an institution. One can then even argue if such a location reference still qualifies as a toponym or not. Also the pure grammatical context of a toponym, that is, if it is used as a reference or an attribute, can be crucial for the detection of the user's information need or the *locations* of a document.

Consequently, we present YGRIEGA (YGRIEGA Georeferenced Rhetorical figures Identifier, Extractor, and Grammar Analyzer) that is able to distinguish both, the different toponym *usage kinds* and the particular *stylistic devices* a location reference may be used in. With respect to its usage kinds, we distinguish three different classes. Concerning the stylistic devices, we are using a classic schema that originates in the Latin rhetoric. Though this paper discusses the toponym characterization in texts written in German, the used concepts, methodologies, and characteristics can easily be applied to other languages, especially Germanic and Romance ones such as English or Spanish.

We start by outlining the related work in toponym characterization and computational linguistics in section 2. In section 3, we describe which linguistic devices can be used in the context of toponyms and how they can be differentiated against each other. YGRIEGA itself, its components and our developments on top of them are depicted in section 4. The results that YGRIEGA achieves when it annotates the different linguistic devices are outlined in section 5. Finally, the conclusion in section 6 summarizes the accomplishments made so far and gives an outlook on possible future enhancements and new developments in the context of this topic.

2. RELATED WORK

Independent from the concrete characterization of location references, many authors have dealt with the identification of rhetorical figures and the structuring of sentences and documents. In many cases, their approaches can be transferred to the characterization of toponym usages. Kelly et al. [8] for example already define a comprehensive ontology of rhetorical figures which are classified according to several dimensions. Besides the pure theory, Daniel Marcu [11] already presents some concrete steps towards the analysis of sentences and their structure in so-called discourse trees. Jakub Gawryjolek [5] developed a tool for the annotation and visualization of rhetorical figures in English texts.

Markert and Hahn [12] concentrate in their analysis and interpretation of stylistic devices in the English language on the metonymy. However, they use this category as an umbrella term for several figures. Therefore, it is not directly comparable to the differentiation we use for the development of YGRIEGA which is described in section 3. Besides that, Markert and Hahn [12] clearly oppose the idea of previous works and consider the usage of terms in a literal and figurative sense to be of equal importance. Before, terms that are used figuratively were clearly of minor importance to researchers. Consequently, Markert and Hahn [12] analyze and interpret the different categories independently from each other in the same profoundness. For that purpose, they use different linguistic models for example to classify part-whole relationships of terms.

In another work, Markert and Nissim [16] publish an algorithm to detect metonymies based on supervised artificial learning. Here, the authors concretize their usage of different subkinds of metonymies by several examples in detail. Markert and Nissim [13] pick up this categorization in their solution to the SemEval-2007 task on metonymy resolution. Moreover, they evaluate different approaches for the detection of rhetorical figures and use among other things for example the information extraction tool ANNIE[1] that is part of the *General Architecture for Text Engineering* (GATE)[2] language analysis framework to detect toponyms and names of organizations.

Some other works also address especially the characterization of toponym usages, however, like the aforementioned ones, only for the English language. Buscaldi and Rosso [2] concentrate on part-whole relationships of locations and use the lexical-semantic database WordNet. Their central goal is a better toponym resolution in case of geo-geo ambiguities. However, according to them, WordNet is only of limited use for their purpose. Leveling and Hartrumpf present in [10] the details on a classifier for metonymically used toponyms. Those authors use the categories that Markert and Hahn outline in [12] and point out that it is of vital importance for GIR to distinguish literal and figurative toponym usages.

[1] *http://gate.ac.uk/ie/annie.html*, all URLs in this paper last visited: 29.9.14
[2] *http://gate.ac.uk/*

Leveling and Hartrumpf argue that in case of metonymic usages of toponyms, the location names do not really reference concrete locations. Based on these results, Nastase et al. [14] additionally use WikiNet[3] to detect metonymically used toponyms for improving the detection rates even more. Moreover, Nastase et al. state that first experiments with unsupervised learning algorithms achieve promising results in the area of toponym characterization.

Based on the related work, the present paper introduces a new toponym characterization with a fine-grained distinction of usage kinds (defined for the purposes of this paper) and stylistic devices (as presented in [9]) specialized to toponym usages. The main contribution is a feasibility study demonstrating that usage kinds and stylistic devices relevant to GIR can be detected with reasonable accuracy.

3. LINGUISTIC DEVICES

3.1 Usage Kinds

In addition to the classic characterization of text and speech in stylistic devices such as the metaphor or the allegory, it can be useful to distinguish several simple usage kinds that can be derived directly from the sentence structure. They can then form the foundation for additional analyses and the detection of the actual stylistic devices. Moreover, the usage kinds can already be applied to reveal some valuable characteristics of a text, for example if certain kinds are being used more often in one text genre than in another. The identified three usage kinds are as follows:

Reference

If toponyms are used together with *adpositions* such as *in*, *um*, or *bei*, which mean *in*, *around*, or *at* in English, the author often wants to establish a connection for example between an action and a certain location—in other words the locations are used as references. So, if toponyms are used together with pre- or postpositions, we call the corresponding usage kind *reference*. The following example sentence illustrates this with the toponym in **bold** and the preposition in SMALL CAPS:

Preposition: IN **Ukraine** the fear of instability grows.

Attribute

In other cases, toponyms are not used individually in a sentence, but as descriptions to other proper nouns, for example as an adjective or as a noun in possessive case (genitive). Moreover, also a partitive which is semantically a genitive usage is included here. We call this usage kind *attribute*; the following sentences show some examples for that:

Toponym as Adjective:
The **Würzburg** RESIDENCE is a Baroque palace.

Possessive Case:
Iceland's POLICE denies its responsibility.

Partitive (Semantic Genitive):
The EMPEROR of **China** resigns.

[3] *http://www.h-its.org/english/research/nlp/download/wikinet.php#WikiNet*

Standard

All other toponym usages are collected in the third kind called *standard*. Most of these location references are used individually or are proper nouns that are qualified by an adjective that is not a toponym itself.

Without Attributes:
Bamberg was built up on seven hills.

Toponym with Adjective:
The BEAUTIFUL **Spain** is always worth a visit.

3.2 Stylistic Devices

Figures

Today's definitions of the different stylistic devices have often their origins in ancient Rome or Greece. Already in the early times of democracy, politicians needed to convince others simply by means of speech. Many orators therefore used the way they tried to make a good case as a *weapon*, as Kolmer and Rob-Santer [9] argue in their book about rhetoric. However, though or even because this discipline has undergone centuries of research, there is still no complete categorization on the different figures that all researchers agree on. So, to be able to analyze the ways a toponym can be used in a text, we choose the definitions outlined in [9] and build our prototype around them. Of course, we only choose those stylistic devices that are relevant in the context of toponym usages.

Stylistic devices are classified in two different categories, *figures* and *tropes*. *Figures* are those elements of a text, which "are being replaced by a less common one" [9]. In these cases, the original sense of such a word or a group of words remains the same. We will present now the most relevant figures that may occur in the context of toponym usages.

Alliteration: If two or more consecutive words share the same start phoneme, especially if these are consonants, this is called an *alliteration*.

Example: **B**amberg **b**eats any other city by its beauty.

Geminatio: In contrast to the alliteration, a *geminatio* is characterized by the duplication of a complete word or a group of words.

Example: **Denmark**, **Denmark** is the place to go!

Anaphora: The *anaphora* is again a figure of word duplication. However, in this case the word or group of words must be the same at the beginning of consecutive sentences or sentence parts.

Example: **France** will grow, **France** will recover.

Epiphora: An *epiphora* is very similar to an anaphora. The duplication of words or word groups occurs at the end of consecutive sentences or sentence parts.

Example: Europe looks on **Russia**. No, the whole world looks on **Russia**.

Anadiplosis: An *anadiplosis* is the combination of an anaphora with an epiphora. In this figure, the last word or word group in a sentence or a sentence part is repeated at the beginning of the next one.

Example: That can only happen in **Austria**. **Austria** has always been special.

Comparison: In many cases, location references are compared with other ones for example to illustrate similarities or differences better. This is called a *comparison* or *similitudo* in Latin.

Example: The devastated area is THREE TIMES BIGGER THAN the size of the **Saarland**.

Tropes

In contrast to normal figures, *tropes* identify those words or groups which are used in a meaning other than their literal one. The words are used figuratively, for example to better illustrate the content or to challenge the reader in order to decrypt the actual meaning of the text. In the context of toponyms, several *tropes* can be used:

Synecdoche: If a term is replaced by another one with a broader or narrower meaning than the original one, this is called *synecdoche*. Normally, such stylistic devices are more widely referred to as *Pars pro toto* (Part for a whole) and *Totum pro parte* (Whole for a part). With respect to toponyms, *synecdoches* are often used to refer to a government of a certain country, a city, or a region by using its capital.

Example: **Berlin** decided to raise taxes on tobacco.

Antonomasia: An *antonomasia* describes the usage of a word or a group of words that include a characteristic attribute of the originally referred term. Many locations have certain characteristics that describe the city, country, or region very well and can therefore be used as a replacement for the name itself. The city of Bamberg for example is located in Franconia and was built on top of seven hills—like Rome. Using an *antonomasia*, Bamberg can be replaced by the "Franconian Rome".

Example: **The Franconian Rome** has a long history.

Metonymy: Compared to a synecdoche, a *metonymy* changes the original meaning of the referred term even more so that no more but a loose logical, spatial, or temporal relationship to the original word or group of words remains. In the context of toponyms, a metonymy describes for example the usage of a location name when only its inhabitants are meant.

Example: **France** resisted the Nazis.

Metaphor: A *metaphor* is often used if the author wants to illustrate an abstract or incomprehensible term to the readers. A word or group of words is called a *metaphor* if the replaced and the used term have no relationship to each other with respect to their literal meaning. In the context of toponyms, metaphors are for example used if the referenced area is not contoured exactly, but if the user can deduct it by the used term(s).

Example: Visit the **Cradle of Humankind**!

Allegory: The *allegory* is a stylistic device that uses a person with all its characteristic traits as a symbol for the original term. In other words, in the context of toponyms, a whole nation, a city, or a region is personified by a person or an icon.

Example: **Uncle Sam** sends his troops to Afghanistan.

4. TOPONYM CHARACTERIZATION

To be able to characterize toponyms correctly in German texts, several different types of software need to be aligned. First and foremost, we need a system that is able to do a profound *toponym recognition* as this constitutes the foundation of any further analyses on usage kinds and other stylistic devices. Moreover, we also need tools which can be applied for computational linguistics, especially for texts written in German, and an environment that can be used to put the different pieces of software together. Additionally, the tool shall make it possible to write own extensions to the existing solutions. With respect to these requirements, we chose to use GATE [3] for this purpose. The overall architecture of our solution with GATE as foundation, several other tools as plugins, and our own analysis component on top of them will now be outlined in detail.

4.1 Existing GATE Plugins

GATE is already equipped with a standard tokenizer and sentence splitter for several languages. Moreover, the standard shipment offers a reference application that includes a well-defined pipeline with these components and the part-of-speech (POS) tagger *TreeTagger* that has proven to provide excellent results in German texts [18, 19]. In addition to that, we used the existing named entity recognition capabilities of the *German* plugin of the GATE Developer bundle for an initial toponym recognition.

However, the final *toponym characterization* requires additional, more sophisticated textual analysis tools. To correctly determine the usage kinds and especially the stylistic devices, we need detailed information about the structure of a sentence and morphological information about all parts of it. A GATE integration of the *Stanford Natural Language Parser (NLP)* for a sentence dependency analysis is available, but it currently only supports English and Chinese[4]. So there is no benefit in using the Stanford NLP for our purpose. On the other hand, there is a solution for the morphological analysis of German texts provided as GATE plugins by the Semantic Software Lab of the Concordia University in Montréal (Canada) that relies on an existing *TreeTagger* infrastructure[5]. Thus, we use this *Durm German Lemmatization System* [17] as well in our implementation.

4.2 Plugged-in External Resources

The existing morphological analysis with the Durm German Lemmantization System already returns some valuable information about certain sentence parts, for example which of them serve as subject, predicate, or object. However, without the aforementioned dependency analysis of all sentence parts, it is for example hardly possible to analyze the sometimes quite complex German sentence structure, especially when one or more subordinate clauses are used. As the GATE Developer distribution does not include capabilities to perform such a grammatical dependency analysis for German texts, an external tool providing these features needs to be connected to our application pipeline. Due to its availability as free software and some successful initial tests, we choose the Zurich Dependency Parser for German (ParZu) for this purpose [20]. Figure 1 shows the dependencies within an example sentence ("The Olympic Winter Games 2014 in Sochi have begun."). The visualization spans up a tree starting from the auxiliary verb "haben" to the full verb "begonnen" and the subject "Winterspiele" with all its attributes. This tree-like dependency information is vital for a proper detection of certain usage kinds and stylistic devices. ParZu's command-line interface makes it possible to integrate it as an additional *processing resource* (PR) using GATE's *TaggerFramework*. As a result, the dependency informations provided by ParZu become additional features for the already existing *Token* annotations provided by the standard GATE plugins.

Moreover, several initial tests of GATE's built-in toponym recognition capabilities revealed that though the precision of the recognized location references is comparatively good, the recall is not satisfactory as many toponyms are not detected. To get a better foundation for the following toponym characterization, we integrate other resources in our application pipeline to improve the toponym detection rate. First, we choose the Named Entity Recognizer (NER) [4] by the Stanford Natural Language Processing Group[6] which can also be integrated in the existing application pipeline as additional processing resource using GATE's *TaggerFramework*.

However, even with the combination of both NERs, several toponyms—especially smaller populated places or less known locations—are still not detected. As we use German texts for our prototype, we decide to integrate the location database OpenGeoDB[7] which concentrates on toponyms in the German speaking areas of central Europe into our application pipeline. Since the built-in NER of GATE works with a combination of gazetteers and rules, we extend these with the data extracted from OpenGeoDB instead of building a new processing resource. To do so, we take all location names listed in the OpenGeoDB database, remove the ones that share their names with the 10.000 most common German words[8] to avoid geo/non-geo ambiguities and add the remaining entries to a single new gazetteer with a new corresponding rule.

Using the OpenGeoDB data in our application pipeline, we still face issues with rarely used toponyms originating outside of the German speaking areas of central Europe. Consequently, we integrate GeoNames[9] as an additional toponym database into the application pipeline. However, due to its vast amount of data, we do not integrate this database the same way as the entries of OpenGeoDB, but build an own GATE processing resource that uses the GeoNames Web Service API[10]. This *GeoNames.org Toponym Resolver* processing resource becomes part of the YGRIEGA GATE plugin. It exposes some of the configuration options that are available in order to restrict the search results for example to populated places or to specify a minimum number of inhabitants. Moreover, the processing resource also does not return any occurrences that have the same name as one of the 10.000 most used German words to avoid ambiguities also here.

[4] *http://nlp.stanford.edu/downloads/parser-faq.shtml#c*
[5] *http://www.semanticsoftware.info/durm-german-lemmatizer*
[6] *http://nlp.stanford.edu/software/CRF-NER.shtml*
[7] *http://www.opengeodb.org/wiki/OpenGeoDB*
[8] *http://wortschatz.uni-leipzig.de/Papers/top10000de.txt*
[9] *http://www.geonames.org*
[10] *http://www.geonames.org/source-code/*

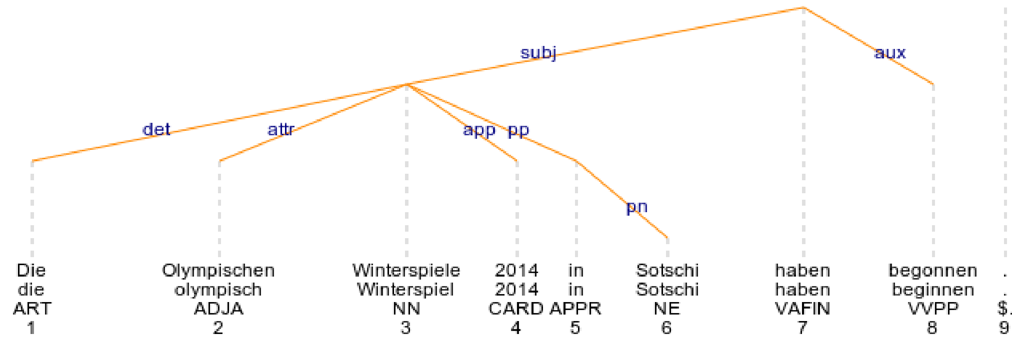

Figure 1: ParZu: Sentence dependency analysis

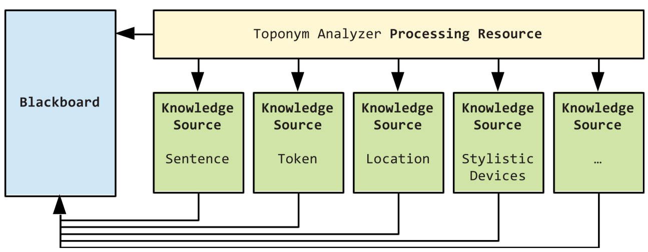

Figure 2: YGRIEGA: Blackboard architecture

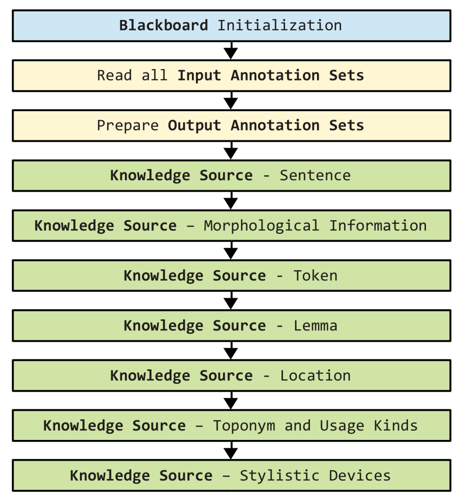

Figure 3: YGRIEGA: Component initialization

4.3 The YGRIEGA Toponym Analyzer

Having the built-in and plugged-in resources in place, the prototype is able to provide solid information about sentence structure, morphological information, and existing toponyms in German texts. This forms the foundation for the central component of the application pipeline which analyzes the usage kinds and stylistic devices a toponym is used in. The *Toponym Analyzer* processing resource annotates found toponyms with their respective usage kind and possibly existing stylistic devices using simple rules based on the sentence structure and morphological information. Moreover, certain rules also make use of *GermaNet* [6], a lexical-semantic database with information about a huge amount of German words.

Looking at the variety of different resources that all provide a small part to the final problem solving process, the *Toponym Analyzer* needs an appropriate architecture that is able to make use of the available information. Transferred to real life, many parallels to existing problem solving strategies can be found, for example when a *blackboard* is used to find dependencies, connections between different artifacts, or simply to get an overview over the current state of available information. In software development, such a *blackboard architecture* was already described and put to practical use in the 1970s [15]. Following GATE with the different resources as *knowledge sources*, the framework with the configured application pipeline as *control unit*, and the document with its annotations as the *blackboard*, the *Toponym Analyzer* implements such a blackboard architecture itself (see figure 2).

The GATE processing resources that are part of the configured application pipeline store the gathered information in annotation types with different features. The *Token* annotation for example stores data on the *lemma* of a token or its respective parent token based on ParZu's dependency analysis (see section 4.2). Additionally, there are *Location* annotations for identified toponyms, *Sentence* annotations for whole sentences containing references to all containing *Tokens*, and many more. For each of the already existing annotation types, a dedicated *knowledge source* retrieves the individual annotations using the GATE API and stores them on an own blackboard using facades as tailored data objects that facilitate the actual analysis process.

These tailored facades can even more contribute to the toponym characterization if they reflect the explicit and implicit dependencies between the different annotation types. Starting from the sentence representation which should allow to easily retrieve all contained tokens, these token facades should also allow a direct access to their corresponding lemmas or the morphological information. To be able to link all facades efficiently to each other, the different annotation types are read by their corresponding knowledge sources in a well-defined order which is outlined in figure 3. Afterwards, the knowledge sources that are responsible for the determination of the toponym usage kinds and stylistic devices can make use of the already initialized facades.

The newly identified toponym characteristics are stored in a new annotation type *Toponym*. The main reason for that is

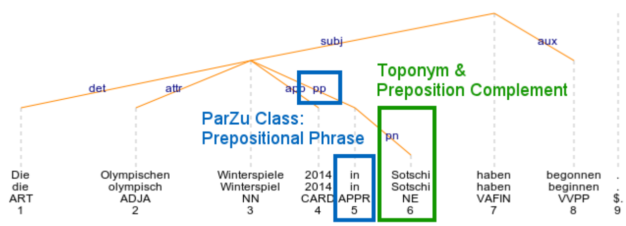

Figure 4: Rule for usage kind *reference*

an easier differentiation between the results of the toponym recognition, which are written to the *Location* annotation by the internal and external GATE plugins, and the actual toponym characterization.

4.4 Methodology

All usage kinds and most of the stylistic devices are detected using grammatical rules implemented in the *Toponym Analyzer*. German, being a comparatively complex language from a grammatical point of view, makes it often difficult to cover all possible cases of a certain usage kind or stylistic device. However, many detection rules can be formulated using only a few conditions, whereas some can become quite extensive. We will now outline the applied methodology only for a few selected examples. Of course, our tool supports all usage kinds and stylistic devices described in section 3.

The usage kind *reference* is characterized by the combination of a toponym with an assigned pre- or postposition. Therefore, the *Toponym Analyzer* needs to check all detected toponyms if they are the complement of an existing adposition. This can mainly be done by an analysis of the POS tags retrieved by TreeTagger (Tag[11] `APPR`) and the *Location* annotations returned by the toponym recognition resources. In this case, the algorithm also needs to be aware of the fact that for example articles and adjectives may stand between a preposition and a toponym. However, if the sentence dependency analysis detects a prepositional phrase (ParZu class `pp`) that is linked to a toponym tagged as preposition complement (class `pn`), the usage kind *reference* can properly be identified (see figure 4).

The rhetorical figure *comparison* can be detected in a similar way. The POS tag `KOKOM` for a certain word already reveals the existence of a comparison in a sentence. However, the *Toponym Analyzer* would face the same issue as before with respect to the connection between the conjunction and the toponym if no dependency analysis is used. Therefore, also in this case the rule for the stylistic device *comparison* makes use of the analysis done by ParZu and by the toponym recognition resources. Only if a toponym is used as conjoined element (class `cj`) to a comparative conjunction (class `kom`), a *comparison* is identified (see figure 5).

All other *usage kinds* and *rhetorical figures* outlined in section 3 are detected in a similar way, that is, using POS tagging, the dependency analysis by ParZu, and the toponym recognition resources. However, the detection of *tropes* is more difficult and needs additional or other sources of information. If a toponym is used as *synecdoche*, it is not the actual location which is referred to, but for example an institution that is situated in the particular city or country. In many of these cases, the toponym is the subject of a sentence

[11] *http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf*

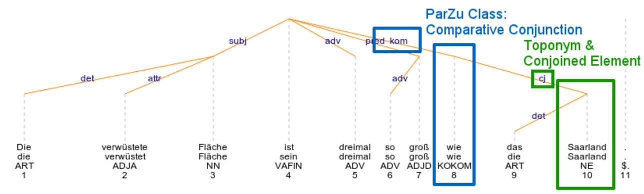

Figure 5: Rule for stylistic device *comparison*

	precision	recall
GATE pure	87%	65%
GATE extended	85%	70%
GATE ext., Stanford	83%	85%
GATE ext., Stanford, GeoNames	67%	95%

Table 1: Toponym recognition effectiveness

and the predicate contains a verb that does not make sense if the toponym is used in its literal sense. Actually, only a small set of verbs such as *to be founded* or *to be located* are used together with toponyms, if the respective location is being referred to.

The detection algorithm for a *synecdoche* uses ParZu and the Durm German Lemmatizer plugins to determine the subject (the morphological information will contain *nominative* as case) and the *lemma* of the actual predicate that belongs to the subject. Moreover, the identified toponyms are checked if they have the usage kind *standard* assigned to them, mainly to avoid false positives if also other words with a toponym as *attribute* are used as subject. This lemma of the predicate must be categorized by *GermaNet* as belonging to the word classes of *creation* ("Schöpfung"), *change* ("Veränderung"), *location* ("Lokation"), or *place* ("Ort"). If this is not the case, we consider the toponym to be used as a *synecdoche*.

For the other *tropes*, we could not identify general rules that are able to properly detect the usage of an *antonomasia*, a *metonymy*, a *metaphor*, or an *allegory*. Moreover, *metonymies* cannot be distinguished easily from *synecdoches*. Therefore, actual *metonymies* are generally identified as *synecdoches* in our prototype. However, for the other *tropes*, we use dedicated gazetteers, as most of the words or word groups which make up a toponym that is used as such a trope are not used in other contexts. Consequently, all occurrences of for example *Down Under*, *Franconian Rome*, or *Holy Land* are considered to be an *antonomasia*, usages of *Uncle Sam* or *Britannia* are treated as *allegories*, and *Cradle of Humankind* is identified as a *metaphor*. All gazetteers are plugged in using dedicated rules extending the existing ones for location detection in the German plugin of GATE Developer.

5. RESULTS AND EVALUATION

In section 2, we outline that to our knowledge there is currently no related work that has already done a toponym characterization with the fine-grained distinction of usage kinds and stylistic devices that we apply—especially for texts written in German. As a consequence, we are also not aware of any *reference corpus* that we could use to evaluate the performance of our approach. Therefore, we compile a set of 15 texts from various sources such as news or travel reports. In these texts, one person manually annotated all toponyms,

	corp.	det.	match	prec.	recall
usage kinds	475	546	427	78.2%	89.9%
reference	226	207	197	95.2%	87.2%
attribute	53	101	49	48.5%	92.5%
standard	196	238	181	76.1%	92.3%
stylistic devices	129	110	96	87.3%	74.4%
figures	53	51	48	94.1%	90.6%
alliteration	17	18	17	94.4%	100.0%
geminatio	8	8	8	100.0%	100.0%
anaphora	12	12	12	100.0%	100.0%
epiphora	6	4	4	100.0%	66.7%
anadiplosis	4	4	4	100.0%	100.0%
comparison	6	5	5	100.0%	83.3%
tropes	76	59	48	81.4%	63.2%
synecdoche	56	45	34	75.6%	60.7%
antonomasia	9	9	9	100.0%	100.0%
metonymy	6	0	0	n/a	0.0%
metaphor	2	2	2	100.0%	100.0%
allegory	3	3	3	100.0%	100.0%

Table 2: Toponym characterization effectiveness in terms of precision (prec.) and recall based on the number of occurrences in the corpus (corp.), the number of detected occurrences (det.), and the number of correctly detected occurrences (match).

their usage kinds and stylistic devices. We use these annotations as a *gold standard* for the evaluation of our application pipeline. In the texts, there are 475 toponyms in total, 129 of them used as stylistic devices.

Our evaluation is based on the metrics *recall* and *precision*. We apply the formulas used by GATE in its performance evaluation tool with some minor simplifications here so that the metrics do not reference the occurrences of complete documents, but single annotations or even *features* of annotations. As the *toponym characterization* is built upon the results of a previously performed *toponym recognition*, the results for the characterization can at most be as good as for the recognition. To be able to correlate both results, we analyze the effectiveness of the recognition and the characterization in different measurements.

An *occurrence* in the sense of the used formulas therefore relates to a *Toponym* annotation for the evaluation of the *toponym recognition*, or to the usage kind or stylistic device feature for the *toponym characterization*. If the application pipeline annotates a word or a word group with a usage kind or a stylistic device, this is called a *detected occurrence*; if this detected occurrence matches the one stored in the reference corpus, we call it a *match*. The formulas are as follows:

$$Precision = \frac{\#\text{ matches}}{\#\text{ detected occurrences}} \tag{1}$$

$$Recall = \frac{\#\text{ matches}}{\#\text{ occurrences in corpus}} \tag{2}$$

To be able to determine which application pipeline fits best, we evaluate four different combinations of the recognition resources that are outlined in section 4.1 and 4.2. The first application pipeline *GATE pure* only includes the recognition features available with the original distribution of GATE Developer. The second pipeline *GATE extended* includes the original features and the OpenGeoDB-based gazetteer. The third application pipeline is based on the second one, but also includes the Stanford NER, whereas the connection to GeoNames is only used in the 4th pipeline.

Table 1 shows the results for the toponym recognition. Taking both recall and precision into consideration, only the extended GATE together with the Stanford NER yields overall results with both numbers clearly above 80%. If GeoNames is additionally used, the recall is by far the best compared to the other pipelines. However, this setting produces a comparatively large amount of false positives which leads to a lower precision. The individual results differ considerably for certain texts. Texts that do not contain rarely used toponyms but for example only bigger cities in Germany, are already well covered by the original GATE distribution. Other ones, such as a short excerpt from a tour guide through Iceland can only be handled effectively by the complete pipeline including the GeoNames connector. Due to these differences, we chose to use the respective best-fitting recognition resources for each individual text for the toponym characterization based on the usual F-measure values to consider recall and precision equally. Out of the 15 texts in our evaluation corpus this meant that *GATE pure* and *GATE extended* were used for two texts, each. *GATE ext., Stanford* was used for 7 texts and *GATE ext., Stanford, GeoNames* for 4 texts. We used the best-fitting recognition resources to minimize the potential influence of erroneously tagged toponyms on the characterization effectiveness.

The results for the usage kinds and the stylistic devices are outlined in table 2. If we compare these results with the numbers from the depending toponym recognition, the toponym characterization results do not stay far behind. A detailed look reveals that especially the detection of the different usage kinds does only produce a small amount of false positives for the previously detected toponyms. The errors are mostly due to not detected attributes such as the prefixes "EU-" or "US-" and enumerations where all elements should be annotated with the same usage kind as the first one. In both cases, only the *standard* usage kind is assigned.

The *Toponym Analyzer* also correctly detects most of the stylistic devices in the reference corpus. However, the results are not as good as the ones for the usage kinds. The performance is also not equally distributed over the different figures and tropes. Most of the figures are detected properly. The undetected comparison figure in table 2 is due to the fact that the comparison is used together with an enumeration. Here, we face the same issue as with the aforementioned usage kinds. Another point worth mentioning is that only four of six epiphora are correctly identified. In concrete this is caused by the fact that "holy land" and "promised land" are not recognized as synonyms. This shows two aspects. First, it might be worth a discussion if synonyms can form epiphora, and, second, the quality of the detection of some stylistic devices obviously directly correlates with the completeness and correctness of the respective word lists.

If the application pipeline needs to recognize certain kinds of tropes, it also has several drawbacks, though still the majority of toponyms is annotated correctly. Particularly challenging are certain kinds of synecdoches, mainly when they are not the subject of a sentence or if either the morphological or the dependency information does not contain proper information about the correct subject. Moreover, certain synecdoches are used together with a preposition which is hard to detect as the system cannot deduce the *intention* from the sentence's predicate. Finally, the corpus contains some metonymies which are annotated as synecdoches as expected.

6. CONCLUSIONS AND OUTLOOK

The experimental results presented in section 5 show that the prototype is able to properly recognize and characterize toponyms in German texts. Especially with respect to usage kinds and most occurrences of rhetoric figures, the error rate remains at a low level. However, especially certain kinds of tropes, such as synecdoches and metonymies, challenge the *Toponym Analyzer* considerably.

Other components also face certain difficulties, mainly because of the complex German grammar, especially with respect to the sentence structure. As a result, the morphological information and also partly the ParZu dependency information is sometimes misleading, but only in few cases. Moreover, though the modular architecture facilitates the usage of multiple components, their individual results are in some cases inconsistent to each other. Therefore, a more detailed evaluation and a tailored integration of these components with respect to their advantages and disadvantages could also improve the overall performance considerably in the future.

Independently from the concrete results and the actual analyzed language, this work proposes a new approach for the characterization of toponym usages. In contrast to already existing categories such as *place-for-people* or *place-for-event*, our categories are based on the classic definitions of rhetorical figures and tropes. Moreover, we propose the distinction of three different toponym usage kinds that can facilitate the subsequent analysis substantially. The gained results can for example leverage the toponym characterization based on machine learning algorithms by using the *Location* and *Toponym* annotations or their features.

There are also indications that suggest that the individual occurrence frequency of toponym usage kinds and assigned stylistic devices can be used to detect the actual genre of a text. A conceptually similar approach is described in [1] for movie analysis. Similarly, it is conceivable that for example texts containing a higher than average amount of toponyms used as synecdoches are more likely news than other ones.

Consequently, we believe that this work provides a valuable contribution to the development of GIR systems, especially in the area of detecting the users' information needs and determining geographic footprints of documents [7]. The proposed categorization of toponym usages and its implementation in a prototype for German texts can be the foundation for additional analyses or for improvements to challenges such as finding the actual location of a document by distinguishing literally used toponyms from figuratively referenced ones.

7. REFERENCES

[1] D. Bamman, B. O'Connor, and N. A. Smith. Learning Latent Personas of Film Characters. In *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 352–361, Sofia, Bulgaria, 2013. ACL.

[2] D. Buscaldi and P. Rosso. A conceptual density-based approach for the disambiguation of toponyms. *Intl. Journal of Geographical Information Science*, 22(3):301–313, Mar. 2008.

[3] H. Cunningham and D. Maynard. GATE: an architecture for development of robust HLT applications. In *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 168–175, Philadelphia, PA, USA, 2002. ACL.

[4] J. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proc. of the 43nd Annual Meeting of the Association for Computational Linguistics*, pages 363–370, Ann Arbor, MI, USA, 2005. ACL.

[5] J. Gawryjolek, C. Dimarco, and R. Harris. An Annotation Tool for Automatically Detecting Rhetorical Figures. In *Proc. of the IJCAI-09 Workshop on Computational Models of Natural Argument*, *http://www.cmna.info/CMNA9/proceedings/CMNA9-Gawryjolek%20et%20al.pdf*, last visit: 25.8.14, Pasadena, CA, USA, 2009.

[6] B. Hamp and H. Feldweg. GermaNet - A Lexical-Semantic Net for German. In *Proc. of ACL Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15, Madrid, Spain, 1997.

[7] A. Henrich, V. Lüdecke, and D. Blank. Approaches for determining the geographic footprint of arbitrary terms for retrieval and visualization. In *Proceedings of the 16th ACM SIGSPATIAL Intl. Conf. on Advances in Geographic Information Systems*, GIS '08, pages 1–4, New York, NY, USA, 2008. ACM.

[8] A. R. Kelly, N. A. Abbott, R. A. Harris, C. DiMarco, and D. R. Cheriton. Toward an ontology of rhetorical figures. In *Proc. of the 28th Intl. Conf. on Design of Communication*, pages 123–130, Sao Carlos, Sao Paulo, Brazil, 2010. ACM.

[9] L. Kolmer and C. Rob-Santer. *Textbook Rhetoric (in German)*. Verlag Ferdinand Schöningh, Paderborn, 2002.

[10] J. Leveling and S. Hartrumpf. On metonymy recognition for geographic information retrieval. *Intl. Journal of Geographical Information Science*, 22(3):289–299, Mar. 2008.

[11] D. Marcu. The rhetorical parsing of natural language texts. In *Proc. of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 96–103, Madrid, Spain, 1997. ACL.

[12] K. Markert and U. Hahn. Understanding metonymies in discourse. *Artificial Intelligence*, 135(1-2):145–198, Feb. 2002.

[13] K. Markert and M. Nissim. Data and models for metonymy resolution. *Language Resources and Evaluation*, 43(2):123–138, Feb. 2009.

[14] V. Nastase, A. Judea, K. Markert, and M. Strube. Local and global context for supervised and unsupervised metonymy resolution. In *Proc. of the Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 183–193, Jeju Island, Korea, 2012. ACL.

[15] H. P. Nii. Blackboard Systems. Technical report, Stanford University, CA, USA, CS-TR-86-1123, 1986.

[16] M. Nissim and K. Markert. Syntactic features and word similarity for supervised metonymy resolution. In *Proc. of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 56–63, Sapporo, Japan, 2003. ACL.

[17] P. Perera and R. Witte. A Self-Learning Context-Aware Lemmatizer for German. In *Proc. of the Conf. on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 636–643, Vancouver, BC, Canada, 2005. ACL.

[18] H. Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proc. of the Intl. Conf. on New Methods in Language Processing*, Manchester, UK, 1994.

[19] H. Schmid. Improvements in part-of-speech tagging with an application to German. In *Proc. of the SIGDAT-Workshop*, Dublin, Ireland, 1995. ACL.

[20] R. Sennrich, G. Schneider, M. Volk, and M. Warin. A new hybrid dependency parser for German. In *Proc. of the Biannual Meeting of the German Society for Computational Linguistics and Language Technology*, pages 115–124, Potsdam, Germany, 2009. GSCL.