Toponym Disambiguation Using Events

Kirk Roberts, Cosmin Adrian Bejan, and Sanda Harabagiu
Human Language Technology Research Institute
University of Texas at Dallas
Richardson TX 75080
{kirk,ady,sanda}@hlt.utdallas.edu

Abstract

Spatial information that grounds events geographically is often ambiguous, mainly because the same location name can be used in different states, countries, or continents. Spatial mentions, known as *toponyms*, must be disambiguated in order to understand many spatial relations within a document. Previous methods have utilized both "flat" and ontology-based ranking techniques to identify the correct reference. We argue that the use of location ontologies alone is not sufficient. Since toponyms are used in documents that refer to events grounded geographically, additional pragmatic knowledge can thus be used. To be able to identify the correct reference we enhanced ontology-based methods previously reported with techniques that consider the participants in events including people, organizations, and locations. Disambiguating geographical names over an ontology is cast as a probabilistic problem resolved by logistic regression. Our experimental results on the SpatialML corpus (Mani et al. 2008) indicate that event structures do indeed play an important role in understanding toponyms.

1. Introduction

Toponym disambiguation is the task of grounding spatial locations in text (toponyms) by normalizing them to some structured representation (e.g., geo-coordinates, database entry, or location within a geographic ontology). This task proves to be quite difficult for some highly ambiguous locations. For example, there are over one thousand cities named San Jose (as well as some states and counties). In a short case study, while over 80% of location names are globally unique, 83% of names actually used in text are ambiguous, over 60% of which have more than 5 possible resolutions. More details on this case study can be seen in Table 1.

Globally ambiguous names		
Duplicates	Entries	Percent
1	2,150,855	80.2%
2+	531,550	19.8%
5+	86,493	3.2%
10+	30,759	1.1%
50+	2,294	0.086%
100+	617	0.023%
1000+	5	0.0002%

Ambiguous names in corpus		
Duplicates	Names	Percent
1	119	16.6%
2+	596	83.4%
5+	438	61.3%
10+	310	43.3%
50+	83	11.6%
100+	16	2.2%

Table 1: Case study on ambiguous names. Globally ambiguous names collected using USGS and NGA gazetteers. Ambiguous names in corpus collected on 715 unique names in 438 documents from SpatialML (Mani et al. 2008).

Understanding the spatial aspects of events in documents has always required the grounding of the spatial locations within *event structures*. Yet, we believe that the reverse is true as well: the event structure contributes to the understanding of the spatial grounding. Consider the following sentence: "*President* **Obama** *returned to* **Washington** *on Sunday.*" Disambiguating the toponym "*Washington*" requires the use of contextual cues we argue are best provided by the event structure. After identifying "*Obama*" as a co-participant in the "*returned*" event with "*Washington*", we know that some spatial relationship exists between them. While there are numerous potential spatial relationships between participants in events, often defining the exact relationship requires the locations to already be spatially grounded. But for the purpose of resolving toponyms, we only need to provide the disambiguation algorithm with enough information to choose the most likely candidate (i.e., "*Washington, D.C.*" over "*Washington State*") using a shallow event structure.

The rest of this paper is organized as follows. Section 2 discusses related work in toponym resolution and event detection. Section 3 outlines the approach for our ontology and use of events and their participants. Section 4 proposes four hypotheses to show how events can contribute information to a disambiguation system. Section 5 details our experiments and discusses the results. Finally, Section 6 summarizes our conclusions and identifies areas of future research.

2. Related Work

Much of the early work in toponym resolution emerged from the study of Geographic Information Systems (GIS) and the need to assign a location mention with a latitude and lon-

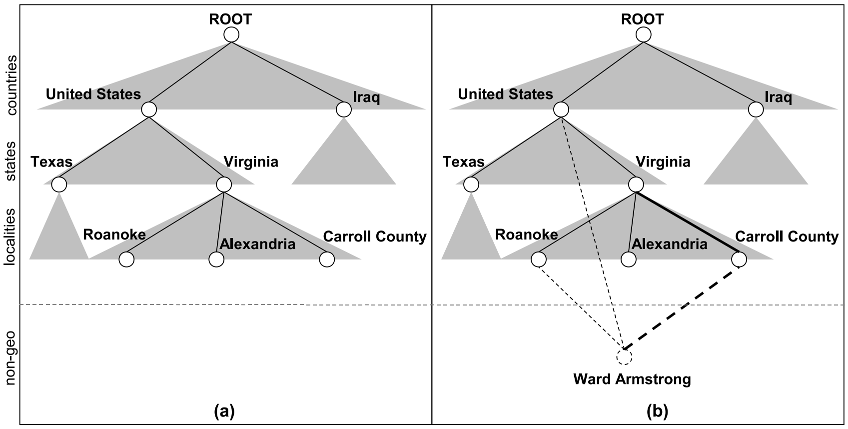

Figure 1: (a) a sub-section of our geographic ontology, (b) the same sub-section with a non-geographic node (*Ward Armstrong*) inserted and connected in an event with *Carroll County* and *Virginia* (event lines in bold).

gitude for referencing on a map. (Smith and Crane 2001) disambiguates toponyms for the purpose of visualizing information from a digital library. Their algorithm uses contextual heuristics derived using an expanding window technique. (Leidner, Sinclair, and Webber 2003) utilized a "one sense per discourse" rule as well as a "bounding box" distance heuristic.

(Buscaldi and Rosso 2008) and (Volz, Kleb, and Mueller 2007) both employ ontologies to aid in the disambiguation process, but in much different ways. (Buscaldi and Rosso 2008) utilizes the WordNet hierarchy directly and employs a "conceptual density" approach based on a WordNet distance. (Volz, Kleb, and Mueller 2007), however, builds an ontology from two large gazetteers and sets class weights to rank candidates.

Toponym disambiguation may be viewed as a specialized form of entity linking (such as the Knowledge Base Population (KBP) task (McNamee and Dang 2009)) in which only locations are considered. A further generalization is word sense disambiguation (WSD). For a survey of WSD methods, see (Navigli 2009).

The SpatialML corpus (Mani et al. 2008) contains several types of spatial information, including toponyms disambiguated into publicly available databases. They describe a statistical ranking method which we use as the baseline for our approach. Unfortunately, the numbers in that paper are not easily comparable to our own, as we only intend to disambiguate a subset of the resolved toponyms in SpatialML, as described below.

3. The Framework

While others have used ontologies to disambiguate geo-locations, we intentionally simplify our ontology to facilitate learning. These pure ontological approaches, however, have traditionally under-performed simple statistical ranking methods, which we validate with our own experiments. We believe this under-performance exists for two reasons. First, ontologies are difficult to disambiguate over due to their graphical structure. Learning requires generalization, but proper geo-ontologies are difficult to generalize as geo-political regions are organized according to local custom (e.g., New Zealand has no state/province structure, only counties; the Vatican is a city-state; while Russia has federal divisions, each of which may contain states (oblasts), semi-autonomous Republics, federal cities, as well as other distinct types of first-order administrative divisions each with different types of sub-divisions). Second, ontologies alone do not model much of the linguistic information that can be used to determine the correct geo-location. To overcome these difficulties, we simplify our ontological structure and introduce a novel approach by adding event information to the ontology.

3.1 The Geo-Ontology

We have built an ontology of almost every geo-political entity (GPE) on the planet. We define a GPE similar to the ACE specification[1], only omitting continents and arbitrary regions. This ontology has a simple tree structure with four levels: a root (i.e., Earth), countries, states, and localities. We choose this structure for its strong consistency: rarely do cities cross state/provincial borders, yet this is often the case with counties (e.g., each of New York City's five Boroughs is itself a county).

Figure 1(a) illustrates a small part of our ontological structure. To populate this ontology, we have used the freely available gazetteer Geonames[2]. We chose Geonames because it is actually a combination of the two most commonly used gazetteers (GEOnet Names Server (GNS)[3] provided by the US National Geospatial-Intelligence Agency and the Geographic Names Information System (GNIS)[4] provided by the US Geological Survey) as well as dozens of other resources. The version of Geonames we have used contains 6,912,700 entries (2,720,984 of which we categorize as GPEs), each of which contain geo-coordinates, population data, and alternate names including acronyms ("*USA*" for "*United States*"), nicknames ("*Big Apple*" for "*New York City*"), and non-English spellings.

[1] http://www.itl.nist.gov/iad/mig/tests/ace/

[2] http://www.geonames.org/

[3] http://earth-info.nga.mil/gns/html/

[4] http://geonames.usgs.gov/domestic/download_data.htm

3.2 Event Structures

An *event* mention in a natural language context is a span of text that refers to a real-world event. Events need not be limited to verbs. The TimeBank corpus (Pustejovsky et al. 2003) specifies three syntactic classes for events: (i) tensed verbs ("*has left*", "*was captured*", "*will resign*"), (ii) stative adjectives ("*sunken*", "*stalled*", "*on board*"), and (iii) event nominals ("*merger*", "*Military Operation*", "*Gulf War*"). In addition to the event reference, each event has a set of syntactically dependent participants (e.g., "*John left the company*"), many of which include GPEs (e.g., "*The Pope returned to Rome*").

While most ontological disambiguation methods use other location candidates in the document to help resolve a given instance, we propose adding knowledge of event structures in addition to document-level knowledge. Evidence that two entities participate in the same event is a strong indicator of a locative relation between the participants. Alternative strategies to find these relations without event structures are (1) using a sentence as a single event, and (2) accounting for token distance between entities. But beyond the linguistic theory, we believe that strategy (1) produces too many incorrect relations, while (2) unnecessarily penalizes syntactically complex constructions. Results in (Buscaldi and Rosso 2008) suggest document context outperforms sentence context. Combined with our results from Section 5, we believe this confirms our belief in the importance of event context.

3.3 Linking Event Participants

We consider three types of entity participants in events: people, organizations, and GPE locations. We start with the base ontology constructed using locations from Geonames. We then use links in Wikipedia to connect people and organizations to the GPEs that they are related to. Figure 1(b) illustrates how one of these additional entities is connected to our ontology in an event. In the example, the politician *Ward Armstrong* is connected to *Roanoke* (where he was born), *United States*, *Collinsville* (where he lives), and *Carroll County* (the district he represents). The bold lines between *Ward Armstrong*, *Carroll County*, and *Virginia* indicate an event involving all three as participants. The details behind selecting edges for a particular event are explained in Section 4.

When mapping persons and organizations into Wikipedia, we assume disambiguation is not required. If Wikipedia indicates that the string *Robert Burns* predominately refers to the Scottish poet, our algorithm assumes that is what the document is referring to. On the other hand, if Wikipedia uses a disambiguation page for entities such as *Mark Wilson*, we ignore the entity. Strategies for disambiguating individuals and organizations can be found in the KBP task referenced earlier. Additionally, no coreference is used to extend the number of events an entity participates in. Both these tasks are left to future work.

Locations are mapped into Wikipedia using a simple heuristic. For countries and U.S. states, the name is assumed to be the exact Wikipedia article title. For cities in the U.S., the pattern [*city-name, state-name*] is used. For everything else, [*name, country-name*] is used.

Alternation Method (with examples)	Strategy
None	BASIC
Case: Alters capitalization. "*us*" to "*US*" and "*Us*"	CASELESS
Type: Removes geo-type. "*New York City*" to "*New York*"	SAFE
Direction: Removes direction. "*North Baghdad*" to "*Baghdad*"	
Abbreviation: Expands ISO-3166 abbreviations and normalizes. "*AF*" to "*Afghanistan*", "*U.S.A.*" to "*USA*"	
Wikipedia Redirect: Uses Wikipedia redirect links. "*Myanmar*" to "*Burma*"	
Demonym: Maps gentilics to locations. "*Texan*" to "*Texas*".	
Wikipedia Suggest: Matches "For . . . , see XXX" pattern at top of Wikipedia pages. "*Washington*" to "*Washington, D.C.*"	MODERATE
Comma Split: Takes first name in compounds. "*Atlanta, Georgia*" to "*Atlanta*"	
Wikipedia Disambiguation: Results from Wikipedia disambiguation pages. "*Lincoln*" to "*Lincoln, Nebraska*"	AGGRESSIVE

Table 2: Gazetteer retrieval alternation modules and their respective strategies (all higher level strategies inherit alternation modules from lower level strategies).

4. Identification of Geo-Locations

Detecting the correct geo-location for a mention in text proceeds in three steps: (1) finding location mentions, (2) retrieval of candidate gazetteer entries, and (3) disambiguating the gazetteer entries to determine the correct, normalized geo-location. In this paper, we assume the mentions are given as the output of a named entity system and thus limit our scope to the last two steps.

4.1 Retrieval of Gazetteer Data

Our database contains over 2.7 million geo-political locations with over 1.8 million alternate names. However, the alternate names provided by Geonames are often not enough to find the correct database entry. For example, the Geonames data contains "*USA*" as an alternate for "*United States*", but not "*U.S.A.*", nor does it contain demonyms like "*New Yorker*", "*Scottish*", or "*Senegalese*". We therefore utilize a sequence of alternation modules to expand the number of candidates. Table 2 shows the five alternation strategies employed with the corresponding alternation modules for each strategy. The choice of a strategy depends on a trade-off between precision and recall: more aggressive strategies are more likely to retrieve the correct gazetteer entry, but also more likely to retrieve irrelevant candidates, which may be incorrectly chosen. Our experiments with these retrieval strategies are detailed in Section 5.

4.2 Disambiguation of Gazetteer Entries

Given a document with n toponyms, the retrieval step will generate n candidate sets. Let C_i represent the set of candidates for the i^{th} toponym mention in the document. The goal of the disambiguation step is then to rank all $c \in C_i$ such that the correct resolution, $\hat{c}$, is ranked in the top position. In order to compare the performance of the event-based system along with our geographical ontology, we used four hypotheses for testing:

Hypothesis 1 (H_1): "Flat" We propose a statistical ranking model that chooses the best toponym resolution. Specifically, given a set of target candidates C_i, we rank $c \in C_i$ by probabilistic function H_1:

$$H_1(c) = \frac{1}{1 + e^{-w^T x_c}} \quad (1)$$

where w is our learned weight vector and x_c is the feature vector that corresponds to candidate c. H_1 intentionally conforms with the classic sigmoid function, $\frac{1}{1+e^{-z}}$, in order to train a logistic regression classifier to maximize the probability that the candidate with the highest H_1 score is the correct resolution. We refer to this method as a *flat* disambiguator because it requires no ontological information.

Hypothesis 2 (H_2): Ontology Transition Probability We desire an ontological ranking mechanism that determines the most likely paths through the ontology to disambiguate toponym candidates. Given n unique toponym mentions, there are $\prod_{i=1}^{n} |C_i|$ possible combinations for resolving these mentions. Note that the unique toponym mention requirement structurally enforces a one sense per document requirement on the system. Let Ψ be our ontology graph. For each possible combination of candidates we create an *assignment tree* A that is the subgraph of Ψ containing (i) one candidate node for each toponym mention from the document, and (ii) the minimum set of parent nodes and edges of Ψ such that the nodes in A form a complete tree. Both person and organization nodes may have more than one "parent" in the ontology. When this happens, simple edge distance is used to find the closest node in the assignment tree. Note that it is quite possible to have an intractable number of potential assignment trees. We discuss our handling of this in Section 5.1.

We propose a statistical ranking model for finding the best the assignment tree $\hat{A}$ that assigns every candidate to the correct entry in Ψ. Let an edge $e_{x,y}$ from parent x to child y in A represent the probability that a node within the subtree rooted by y contains a correct toponym resolution. Thus, the weight on $e_{x,y}$ is the transition probability from x to y. As above, we define the probability as a sigmoidal function of the weights and features:

$$P(e_{x,y}) = \frac{1}{1 + e^{-w^T x_e}} \quad (2)$$

where w again is the learned weights, but now x_e is the feature vector for the edge $e_{x,y}$. We then define the probabilistic function H_2 to be the product of the probabilities of the edges in assignment tree A:

$$H_2(A) = \prod_{e \in A} P(e) \quad (3)$$

Note that our training method does not allow for direct training on the assignment trees. Rather we attempt to indirectly rank better assignment trees higher by adjusting edge weights.

Hypothesis 3 (H_3): Ontology Transition and Node Probability By limiting itself to the product of edge probabilities, Hypothesis 2 presents some severe theoretical limitations. Most important among these is that no candidate will ever be selected if one of its ancestors is also a candidate (e.g., the state "*New York*" will always be selected over the city when both are candidates) because the assignment tree with the child node is guaranteed to have a lower probability. We therefore propose a statistical model that combines both edge probabilities and node probabilities. Since Hypothesis 1 attempts to estimate node probabilities, we can alter H_1 from (1) to work on assignment trees:

$$H_1(A) = \prod_{v \in A} H_1(v) \quad (4)$$

where v is both a vertex in A and a candidate for a mention. Now, H_3 is simply the product of H_1 and H_2:

$$H_3(A) = H_1(A) H_2(A) \quad (5)$$

If H_1 and H_2 are viewed as probabilistic models, then H_3 is a linear combination of two conditionally dependent functions. However, this model does allow the disambiguator to overcome the *New York* problem from above if the H_1 probability of *New York City* is sufficiently high.

Hypothesis 4 (H_4): Ontology Event and Node Probability We require an ontological model that is able to account for the event structure within a document. To accomplish this, we convert a given assignment tree into an *assignment forest* by only considering the edge probabilities between events. Each event in the document represented by the assignment tree A is then extracted into its own assignment tree, A_e. We then define H_4 similarly to H_3:

$$H_4(A) = H_1(A) \prod_{A_e \in A} H_2(A_e) \quad (6)$$

Note that H_4 is capable of traversing the same edge probability multiple times if more than one event covers a node. This occurrence is common in natural language documents and has a desirable consequence: highly unlikely event assignment trees penalize H_4 multiple times. Because H_4 makes no attempt to analyze the complete assignment tree, this model ignores document-wide context. Event context is assumed to be sufficient information to resolve mentions. Candidates that do not take part in an event are evaluated on their H_1 score alone.

Features While there are numerous possible features for toponym resolution (see (Leidner 2006)), experimental results identified just a few important features for both H_1 (5 features) and H_2 (6 features). In contrast to the complex heuristics mentioned in Section 2, our features are quite simple. Most of the more complicated features mentioned in those papers were largely ignored by the learning system, favoring instead features that provided a simple glimpse of the location's profile. In our experiments, we used the following features:

◁ LOG(POPULATION) ▷ The population of the target node was the dominant indicator (a logarithmic function was used for smoothing);
◁ STRINGMATCH ▷ Exact string matches;
◁ SUBSTRINGMATCH ▷ Sub-string matches were used to help indicate spurious candidates;
◁ ADMINFEATURE ▷ Assigns a value of 0.5 for state capitals and 1.0 for national capitals to give preference to administratively important locations;
◁ TYPEFEATURE ▷ Actually three boolean features to indicate the node is a country, state, or city;
◁ EDGEFEATURE ▷ While the logistic regression intercept has no impact on flat classification, it is used to weight the default edge probability, thus giving relative favor to assignment trees with fewer edges and also accounting for how persons and organizations relate to the geographical nodes in the ontology.

5. Experimental Results

5.1 The Data

The SpatialML corpus (Mani et al. 2008) consists of 428 documents manually tagged with numerous spatial information, including 6,337 PLACE (geo-location) annotations. Of interest to us are the 5,573 PLACE annotations that we consider to be GPEs. We have discarded 48 documents that contain no GPEs and 11 that contain only one, leaving 369 documents with 5,562 GPE annotations. Each PLACE annotation is then mapped to the gold Geonames ID for evaluation purposes.

Many SpatialML documents contain far too many ambiguous annotations to generate all potential assignment trees. Indeed, using the MODERATE retrieval strategy we found 17 documents that would generate more than 2^{31} assignment trees. To provide a more computationally tractable dataset, we used the flat disambiguator to limit each candidate set to 10 candidates. This alone would not be enough, for there would still be 10^n possible assignment trees for a document with n mentions. We therefore used the flat disambiguator again to choose the top 500 candidate combinations for each document. Limited testing indicates this actually improved results by a narrow margin.

For detecting people and organizations, we use the freely available BIOS named entity tagger[5]. It has an F-measure of 85.49 on person names and 92.2 on organization names on the CoNLL test set. For detecting events, we used the system described in (Bejan 2007) based on the TimeBank 1.2 corpus (Pustejovsky et al. 2003). It achieves an F-measure of 82.94 using 5-fold cross validation.

5.2 Retrieval Experiments

For each of the 5,562 toponyms in SpatialML, we computed precision, recall, and F-measure for each of the five strategies listed in Table 2. The results are listed in Table 3. Precision is measured over all candidates, of which there can be at most 1 correct. This can be thought of as the precision of a disambiguation system that chooses a candidate at random.

[5]http://www.surdeanu.name/mihai/bios/

Strategy	Candidates	Precision	Recall	F-measure
BASIC	58,560	5.97	62.67	10.89
CASELESS	67,154	5.71	68.78	10.54
SAFE	83,800	6.31	94.81	11.82
MODERATE	85,764	6.31	97.01	11.37
AGGRESSIVE	617,995	0.88	97.36	1.74

Table 3: Result of gazetteer retrieval stage for the five alternation strategies.

	Accuracy	Ontology Score
H_1	92.00	93.27
H_2	87.85	92.08
H_3	92.35	94.72
H_4	93.57	94.83

Table 4: Accuracy scores across 5,562 instances for the four hypotheses discussed in Section 4.

While the SAFE strategy achieved the highest F-measure, the retrieval module cannot be looked at as a complete system: the quality of the disambiguation system affects the choice of strategy. While higher precision certainly helps the downstream disambiguator, the retrieval recall fundamentally limits the final accuracy of the system. The MODERATE strategy actually performs best as input to the disambiguator. The additional incorrect locations are likely outliers that the disambiguator can easily ignore. The AGGRESSIVE strategy, however, is far too imprecise. It often returns common, populous locations as erroneous candidates, and thus confuses the disambiguator heavily. But to confirm the point made above, the AGGRESSIVE strategy well outperforms both the BASIC and CASELESS strategies in limited end-to-end testing.

5.3 Disambiguation Experiments

We evaluated all four hypotheses on the SpatialML data. We use two metrics to evaluate the accuracy of our system. The first is simple accuracy, or the number of correct toponym resolutions ranked in the top position. The second is an ontology-inspired metric that assumes wrong answers that at least share a parent of the correct answer are better than ones that do not. For instance, if the correct resolution for "*Henry*" is "*Henry County, Virginia, USA*", while both "*Henry, Georgia, USA*" and "*Henry, Zambia*" are incorrect, the former is sightly more correct than the latter. We consider this is a valid approach for a secondary metric because many reasoning approaches might only need the state, country or distance between two locations, so the "closer" location, while still wrong, is better. Since our ontology has three levels of content, we award a full point for an exact match, 0.66 points for matching the correct state, and 0.33 points for matching the correct country. This metric gives us an idea of the type of errors each hypothesis makes by comparing the ontology score to the simple accuracy. The results on the four hypotheses using 5-fold cross validation are shown in Table 4.

H_4 outperformed all other hypotheses in both accuracy and ontology score, revealing that participants in events are more tightly linked than a document-wide approach such as

N	Recall	Ontology Recall
1	93.57	94.83
2	94.64	95.67
3	95.32	96.10
4	95.48	96.21
5	95.68	96.33
6	95.74	96.39
7	95.74	96.39
8	95.78	96.40
9	95.81	96.43
10	95.84	96.46

Table 5: Recall of H_4 when selecting the top N candidates ranked by the disambiguator.

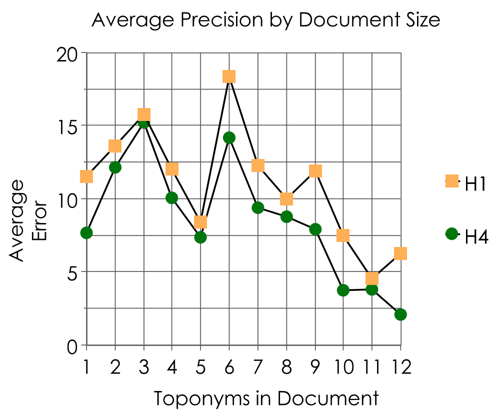

Figure 2: Average error of H_1 and H_4 by number of toponyms in the document.

H_3. Additionally, while H_2 is by far the worst system, the fact that its ontology score is close to H_1 relative to their respective accuracies suggests that it does select ontologically more likely candidates, but has difficulty selecting the best candidate. In examining the errors that H_2 makes, it commonly selects small towns closer to the other nodes in the ontology instead of selecting large, well-known cities. By combining H_1 with H_2 to form H_3, H_2 contributes little to the normal accuracy, but clearly raises the ontology score.

Table 5 shows our evaluation on using H_4 as a ranking method. Only slight improvements are made by choosing more than one candidate in an attempt to maximize recall. Given our choice of the MODERATE retrieval strategy, the upper bound for recall is 97.01, suggesting that about 1% of the toponyms are extremely difficult to disambiguate, not even being selected in the top 10 ranked candidates.

Ontological methods are designed to incorporate context. To test this we experimented with disambiguator performance as a function of document size. Figure 2 illustrates the average precision of H_4 as compared to H_1, our non-ontological model. Many of the larger documents proved easier for both hypotheses, though H_4 consistently performs better for $n > 6$. While documents with large numbers of toponyms do not necessarily contain large numbers of events, there is a correlation. The average document contains 1.2 events with more than one participant, while documents with at least 6 toponyms contain an average of 2.7 such events.

6. Conclusion

Our work has demonstrated the importance of event structures when disambiguating toponyms. While previous models have utilized ontologies with document-wide information, a probabilistic model that only estimates probabilities of events outperformed a document-wide approach. In the future we plan to study in further detail how individual events can affect the spatial grounding. For instance, an event such as "*John swam from San Francisco to Oakland*" suggests a tighter relation than "*John flew from Los Angeles to Tokyo*". Utilizing this kind of linguistic information should allow us to disambiguate even more ambiguous toponyms.

References

Bejan, C. A. 2007. Deriving Chronological Information from Texts through a Graph-based Algorithm. In *Proceedings of the 20th Florida Artificial Intelligence Research Society International Conference (FLAIRS'07)*.

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

Leidner, J.; Sinclair, G.; and Webber, B. 2003. Grounding Spatial Named Entities for Information Extraction and Question Answering. In *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References*, 31–38.

Leidner, J. 2006. Toponym Resolution: A First Large-Scale Comparative Evaluation. Technical report.

Mani, I.; Hitzeman, J.; Richer, J.; Harris, D.; Quimby, R.; and Wellner, B. 2008. SpatialML: Annotation Scheme, Corpora, and Tools. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC-08)*.

McNamee, P., and Dang, H. T. 2009. Overview of the TAC 2009 Knowledge Base Population Track. In *Proceedings of the 2009 Text Analysis Conference*.

Navigli, R. 2009. Word sense disambiguation: A survey. In *ACM Computing Surveys*, volume 41, 1–69.

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

Smith, D. A., and Crane, G. 2001. Disambiguating Geographic Names in a Historical Digital Library. 127–136.

Volz, R.; Kleb, J.; and Mueller, W. 2007. Towards ontology-based disambiguation of geographical identifiers. In *The 16th International World Wide Web Conference (WWW-2007), Banff, Alberta, Canada*.