Corpus annotation with a linguistic analysis of the associations between event mentions and spatial expressions

Jin-Woo Chung **Jinseon You** **Jong C. Park***

School of Computing
Korea Advanced Institute of Science and Technology
291 Daehak-ro, Daejeon, Republic of Korea
{jwchung, jsyou, park}@nlp.kaist.ac.kr

Abstract

Recognizing spatial information associated with events expressed in natural language text is essential for the proper interpretation of such events. However, the associations between events and spatial information found throughout the text have been much less studied than other types of spatial association as looked into in SpatialML and ISO-Space. In this paper, we present an annotation framework for the linguistic analysis of the associations between event mentions and spatial expressions in broadcast news articles. Based on the corpus annotation and analysis, we discuss which information should be included in the guidelines and what makes it difficult to achieve a high inter-annotator agreement. We also discuss possible improvements on the current corpus and annotation framework for insights into developing an automated system.

1 Introduction

Every event is situated within some real-world space and textual descriptions that refer to events in documents also convey such spatial information. Such information is important not only for the interpretation of single events but also for the understanding of the relations among them. Spatial information can be used for various applications such as information extraction, textual entailment, and question answering. For instance, if we want to answer the question "*Where did traffic accidents happen most frequently in 2014?,*" we would need a method to access and collect spatial information associated with all traffic accidents from the relevant textual descriptions. However, such information is usually not provided explicitly in text since humans can intuitively understand from the context where each event occurs.

In general, two factors make it difficult to automatically recognize the location of events in text. First, there are usually more event mentions in text than expressions containing information about the location of events. A system must thus choose the most appropriate spatial expression for a given event mention. Second, such expressions are not always syntactically close to event mentions, which may make it less obvious to recognize their semantic association. The following three sentences exemplify different levels of difficulties in determining whether particular event mentions and spatial expressions are associated, i.e., whether a spatial expression refers to the space where an event occurred.

(1) A fire [broke out]$_{\text{EVENT}}$ at [a refrigerated warehouse]$_{\text{SPACE}}$ yesterday.
(2) A North Korean fishing vessel intruded 10 miles across [the Northern Limit Line]$_{\text{SPACE}}$ and South Korea's Navy [fired]$_{\text{EVENT}}$ 6 warning shots.
(3) He searched all over [the room]$_{\text{SPACE}}$ for his [missing]$_{\text{EVENT}}$ ring.

Sentence (1) shows that the event mention and the spatial expression are syntactically connected in a single clause, which can probably be identified in a straightforward manner with a conventional semantic role labeler. Sentence (2) shows that they exist in the same sentence but not in the same clause, and that there must be some inference in order to

* Corresponding author

29th Pacific Asia Conference on Language, Information and Computation pages 535 - 543
Shanghai, China, October 30 - November 1, 2015

find out that *fired* is likely to occur around *the Northern Limit Line*; for example, *intruded* and *fired* can occur in a similar place and the time interval between them may not be too long. Sentence (3) shows that, even though an event mention and a strong candidate for its spatial expression exist in the same sentence, their association may or may not hold depending on the context; there may be another place where the *missing* event actually happened. In this case, the system may have to search the text backwards to find out where *missing* or other relevant events are mentioned. Such information may, however, not have been stated at all in the available text.

In this paper, we present a linguistic analysis of how event mentions and spatial expressions are associated in text with respect to a corpus annotation process. More specifically, we discuss the following four issues:

- which information should be included in the guidelines in order to recognize spatial information about events in text,
- what kind of difficulties and issues arise during the annotation process,
- what trends are found in the corpus, and
- which factors could be of help to achieve a high inter-annotator agreement and to build an automated system.

The rest of this paper is organized as follows. Section 2 presents previous work on analyzing properties of events in text. Section 3 shows the proposed annotation framework for creating a corpus. Section 4 gives an analysis of the corpus and disagreements between annotators. Section 5 discusses issues on improving the proposed annotation framework, with concluding remarks.

2 Related Work

Research on analyzing aspects of events or relations among them has dealt mainly with temporal aspects and temporal relations. Much effort has been made to establish a specification for describing temporal properties of events in text and to create the labeled data, especially through the TimeML annotation standard and the TimeBank corpus (Pustejovsky et al., 2003a; Pustejovsky et al., 2003b). The availability of the standard and corpus has promoted further studies on extracting temporal information associated with events from text (Lapata and Lascarides, 2006; Mani et al., 2006; Yoshikawa et al., 2009; Mirza and Tonelli, 2014), including the TempEval challenges (Verhagen et al., 2009; Verhagen et al., 2010; UzZaman et al., 2013).

In contrast, analyzing spatial properties of events has received less attention than temporal analysis, though in recent years a few studies attempt to tackle relevant problems. SpatialML (Mani et al., 2008) presents an annotation specification for describing expressions that refer to geographic regions in a way similar to TimeML, but it deals only with spatial relations between non-event entities that are explicitly expressed in a single sentence. In a similar line, Spatiotemporal Markup Language (STML, Pustejovsky and Moszkowicz, 2008) was designed to annotate both the temporal and spatial properties of entities. While it includes the specification for spatial entities associated with events, it focuses primarily on associating motion events with motion-specific arguments, and does not deal with other types of event and other non-argument spatial entities found throughout the text. ISO-Space (Pustejovsky et al., 2011a; Pustejovsky et al., 2011b) addresses the integration of SpatialML and STML to establish the annotation standard. It considers events as a type of spatial entity and allows them to participate in spatial relations. However, these events are annotated only when the spatial relationship is explicitly stated in a single sentence. In particular, it does not consider *implicit* associations between general events and their spatial entities that are found across the text.

Another line of work would be spatial role labeling (Kordjamshidi et al., 2010), which addresses the task of identifying the location of objects and their spatial relations triggered by spatial indicators such as *on*, *at*, and *in*. However, it does not cover the location of general types of event, though a recent series of the SemEval challenges on this task (Kordjamshidi et al., 2012; Kolomiyets et al., 2013) discuss annotating motions.

Blanco and Vempala (2015) propose a method to infer temporally-anchored spatial knowledge from semantic roles. Their goal is to determine whether a certain argument of the verb is located in one of the locative arguments found in the same sentence and to temporally anchor their spatial relationship with respect to the duration of the target event. For example, given the sentence "*John was incarcerated at Shawshank prison*" and its PropBank-style semantic role annotation, they attempt to find out that *John* has been located at *Shawshank prison* **during** event *incarcerated*, but neither before nor after that event. This work makes use of properties of events to infer spatial knowledge, but does not handle the spatial relationship between events and locations outside the sentence.

Unlike the work mentioned above, work on analyzing event-centric spatial relations has not received much attention. The most relevant existing work would be the annotation and recognition of spatial containment relations between event mentions (Roberts et al., 2012; Roberts et al., 2013). They aim at inferring that the spatial boundary of a particular event contains that of another event. For instance, given the sentence "*The bombing victim died immediately*," they infer that the *bombing* event is likely to spatially contain the *died* event. Their work is closely related to ours since it attempts to analyze the spatial aspects of events. However, it does not deal with directly linking event mentions to spatial expressions in a document although they utilize spatial expressions as one of the features for recognizing spatial relations between event mentions. Instead, they put more emphasis on what they call "implicit relation features", suggesting that the spatial containment relations could be recognized based on event semantic properties without relying heavily on contextual clues; we can see, for example, from the example above that the *bombing* and *died* events have some degree of semantic correlation. Their task does not necessarily aim at the recognition of such spatial expressions for events.

To the best of our knowledge, none of the previous studies address the associations between event mentions and spatial expressions found across the entire text.

3 Annotation Framework

In this section, we introduce our framework for annotating spatial expressions for given event mentions. We first describe the data we used for annotation and then present the definition of event mentions and spatial expressions to be annotated in a given document, together with the guidelines for selecting and labeling spatial expressions for given event mentions. We then present the overall annotation process and the corpus statistics.

3.1 Data

We chose to use texts in the broadcast news domain for our corpus as they contain various spatially bound events that happen in the real world as compared to texts in the newswire domain which usually include many editorials and opinions.

We used the data from the OntoNotes project (Hovy et al., 2006) in order to access diverse layers of linguistic annotations during our annotation process, such as part-of-speech tags, parse trees, named entities, and coreferences. We selected 48 documents from the collection of CNN broadcast news in OntoNotes Release 5.0 and used them as our corpus. Table 1 shows the statistics of the selected document collection. The figures in the table suggest that the corpus contains a varying number of words and sentences across the documents.

Measure	Figure
Total number of documents	48
Total number of sentences	416
Total number of words	7,810
Average number of words per sentence	18.8 (std. dev. 10.6)
Average number of sentences per document	8.7 (std. dev. 7.6)
Average number of words per document	162.7 (std. dev. 153.6)

Table 1: Statistics of the data in our corpus

The corpus also includes documents from various topics such as social issues, accidents, politics, finance, sports, and international news. We annotated the associations between event mentions and spatial expressions on top of these documents.

3.2 Annotation guidelines

Event mentions

There is no *de facto* standard definition of event mentions, and researchers usually adopt their own definition that fits into the goal of their work. One of the most widely used definitions would be the one in the TimeML schema. It regards an event mention as "a cover term for situations that happen or occur" (Pustejovsky et al., 2003a). A range of verbs that exhibit changes in the state of the world usually belong to this category. However, we do not restrict event mentions to a certain category of verbs. Instead, we regard almost all verbs as event mentions whether or not they refer to a situation that actually happened or that can be clearly anchored in a timeline. This is because we assume that any situations referred to by verbs including actions and states can be situated within a particular scope of space in the real world where it happens or takes effect. One of the goals of this work is to see if it is possible to pick out expressions that refer to such space for an event mention arbitrarily chosen within a given document.

We consider as event mentions all single word tokens labeled with part-of-speech tags that correspond to base verbs, inflected verbs, gerunds, and participles in the Penn Treebank parse tree of the OntoNotes annotations. These tags include VB, VBD, VBG, VBN, VBP, and VBZ. When gerunds and participles were found, we excluded *be*-verbs and auxiliary verbs used with them and annotated only those gerunds and participles. We also excluded verbs in some patterns that act as auxiliary verbs such as *be going to* and *have to*. For example, given the sentence "*It has not been undertaken but will have to be considered*," we annotated only *undertaken* and *considered* as event mentions to be associated with spatial expressions. Unlike TimeML, we did not consider noun phases as candidate event mentions since it is not clear which type of noun phrase can refer to spatially bound situations. Analyzing spatial aspects of noun phrases would be another interesting line of future work.

Spatial expressions

A spatial expression is either a single word or a sequence of words that refer to a particular space in which the situation or the state referred to by a given event mention happens or takes effect. More specifically, spatial expression S is said to be associated with event mention E if S refers to the space that encloses the spatial bounds of the event referred to by E while it happens or takes effect.

We did not restrict spatial expressions to certain semantic classes as in other studies, such as geographic and geopolitical places (Pustejovsky et al., 2011a; Roberts et al., 2012), locative arguments (Blanco and Vempala, 2015), and entities with spatial indicators (Kolomiyets et al., 2013). We instead asked our annotators to choose any word or phrase that they think provides some information about the spatial bounds of events even though they are not clearly grounded in physical and geographic space, such as *meeting*, *parliament*, *clashes*, *scene*, *demonstration*, *interview*, *network television*, and *political life*.

The annotation of spatial expressions relies largely on annotators' intuitive understanding of the text. In order to enable consistent annotation, we asked the annotators to stick to the following rules which are central to our annotation process, among others.

Rule 1: A spatial expression is either a noun phrase or an adverbial phrase. Our pilot annotation suggests that noun or adverbial phrases are sufficient enough to represent the space associated with events in text. However, we acknowledge one exception to this: adjectival forms of place names and their demonymic equivalents such as *Canadian*, *South American*, and *Northern Irish* can be annotated separately as a spatial expression even though they exist within a longer noun phrase, as shown in the example below.

(4) The [Yugoslav]$_{\text{SPACE}}$ Election Commission claims he did not [win]$_{\text{EVENT}}$ more than 50 % of the vote.

The annotators can choose *Yugoslav* as a spatial expression for event *win* if they consider it to be spatially bound in Yugoslavia. We found that the broadcast news exhibits this pattern frequently; such adjectival and demonymic forms themselves suggest a particular place for events when its nominal forms are not mentioned at all.

Rule 2: If the annotators choose a certain word to be included in a spatial expression for a given event mention, they must annotate the longest noun phrase or adverbial phrase that contains it as a head word. These phrases can contain any kind of modifier such as a relative clause and another nested adverbial phrase, as shown the example below.

(5) Students at a middle school in Calaveras County, California, are [getting]$_{\text{EVENT}}$ an unwanted lesson in entomology.

Here, if the annotators choose *school* as a head noun of a spatial expression for event *getting*, they must annotate "*at a middle school in Calaveras County, California*" as its spatial expression, which is the longest adverbial phrase containing *school* as a head noun, according to Rule 2. However, if they choose *County* as a head noun, they must annotate "*in Calaveras County, California*" as a spatial expression. Our intention behind this is to include as much information as possible in spatial expressions by annotating the longest span of expressions.

Rule 3: If there is more than one expression that refers to the space enclosing the spatial bounds of a given event, the annotators choose the one that refers to the narrowest space. For example, for event *getting* in example (5), we choose "*at a middle school in Calaveras County, California*" as its spatial expression instead of "*in Calaveras County, California*" since the former refers to narrower space than the latter.

The intuition is that narrow space conveys more information than broad space; knowing in example (5) that *getting* is associated with *a middle school* would be more informative than knowing that it is associated with *Calaveras County* because the former is less vague than the latter.

Rule 4: If there is still more than one expression that is not distinguished by Rules 1-3 above, the annotators choose the one that is closest to the event mention. The distance here is measured by the number of sentences between the event mention and its candidate spatial expression. If two equally qualified candidate expressions are found before and after the event mention, respectively, at an equal distance, then the annotators choose the one that appears before the event mention. If two such expressions are found in the same sentence, the annotators choose the one that is syntactically closer to the event mention.

Rule 5: For event mentions referring to a motion that creates a path, the annotators choose three distinct spatial expressions that refer to the beginning, intermediate, and end of the path, respectively, if they exist. When choosing a spatial expression for each of these components of the path, the annotators follow Rules 1-4 above. Such motion event mentions include *arrive*, *leave*, *travel*, and *return*. The following example shows that two motion event mentions appear in a single sentence.

(6) Finally, U.S. Marines [arrived]$_{E1}$ [at the hospital]$_{S1}$ to [take]$_{E2}$ him [to Kuwait and to a specialist burns unit]$_{S2}$.

Here, for event mention E1 (*arrive*), spatial expression S1 (*at the hospital*) can be chosen as the end of its path. For event mention E2 (*take*), spatial expressions S2 (*to Kuwait and to a specialist burns unit*) and S1 (*at the hospital*) can be chosen as the end and the beginning of the path, respectively. Note that associating E2 (*take*) and S1 (*at the hospital*) may require some inference; for instance, E2 may happen shortly after E1 happens.

Possible world analysis: As in the case of example (6), we always interpret the spatial bounds of events under the *possible worlds* assumption; even though they had not occurred or their occurrence is not clear, we estimate their spatial boundary by assuming the situation where they had already occurred. In this way, we can infer that event E2 (*take*) in example (6) has occurred in the space referred to by S1 (*at the hospital*). This type of interpretation can be applied to other similar constructions such as negation, condition, opinion, supposition, and conjecture.

Distinction between definite and plausible associations: Since the spatial information about events is highly implicit in text as discussed in Roberts et al. (2012), in most cases, it would not be possible to annotate spatial expressions with 100% confidence. Certain types of association may be more difficult to justify than others. For this reason, we introduce an additional label to distinguish between definite and plausible associations.

We consider associations to be *definite* if they can be reasonably inferred with common knowledge of the real world. In contrast, if the association cannot be inferred in such a way but is still presumed to exist in certain circumstances, we consider them to be *plausible*. The following example shows a sentence that contains both types of association.

(7) For the second day in a row, Lieutenant General Jay Garner was [mobbed]$_{E1}$ by friendly crowds after [touring]$_{E2}$ [a Kurdish school in the northern [Iraqi]$_{S1}$ city of Irbil]$_{S2}$.

Here, the association between event mention E2 (*touring*) and spatial expression S2 (*a Kurdish school in the northern Iraqi city of Irbil*) is considered to be definite since they are syntactically connected. On the other hand, it would be difficult to be fully confident that event mention E1 (*mobbed*) can be associated with S2. This is probably due to the existence of temporal relation indicator *after*. It suggests that there might be some time interval between E2 and E1, leading the annotators to believe that their locations might be different. In this case, their association is considered to be plausible. If *when* is used instead of *after* in this example, the association could be considered definite. The annotators are allowed to choose only one spatial expression for each of the two types of association for a given event mention; in other words, for general event mentions, the annotators choose at most two spatial expressions: one for the definite association and the other for the plausible association.

3.3 Annotation process

The process of annotating event mentions was fully automated because identifying them relies only on part-of-speech tags which are already provided by the OntoNotes annotations. Spatial expressions are annotated by our annotators, but only when at least one event mention is associated with it; in other words, we did not allow 'dangling' spatial expressions to be included in our corpus, as explained in the annotation guidelines. If the annotators fail to find a spatial expression for a given event mention, they just left it unassociated.

Two annotators participated in the annotation process. They were first provided with the documents that have been pre-annotated with event mentions. They then went over each document and for each event mention chose the text span of spatial expressions that are best associated with the event mention based on the guidelines. After that, they put a labeled link between the event mention and the spatial expression. One of the six labels can be attached to a single link: three labels for the definite associations and three labels for the plausible associations. We also provided the annotators with the web-based annotation tool to facilitate this process. The annotators are also allowed to consult information in the OntoNotes annotation files if necessary for more informed decisions.

Multi-phase annotation: In order to resolve disagreements and to improve the quality of the annotated data, we divided the entire annotation process into four phases and held a meeting whenever the annotation in each phase was completed. In each meeting, the annotators measured the inter-annotator agreement (IAA), analyzed and resolved disagreements, and revised the guidelines if necessary.

4 Corpus Analysis

4.1 Statistics

Table 2 shows the statistics of the annotated data in each annotation phase. On average, 85% (721/846) of the event mentions are associated with at least one spatial expression. This means that for most of the events it is possible to find descriptions of spatial information within a single document. Each spatial expression is associated with 2.5 (337/846) event mentions on average. Each document contains an average of 7.0 spatial expressions with a standard deviation of 5.4 and an average of 15.0 associations with a standard deviation of 14.5.

Statistics \ Phase	1	2	3	4	Total
# documents	5	15	10	18	48
# sentences	45	139	72	160	416
# words	898	2694	1554	2664	7810
# event mentions	95	319	160	272	846
# spatial expressions	31	121	69	116	337
# associations	85	270	140	226	721

Table 2: Statistics of the annotated data

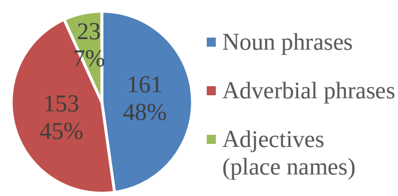

Figure 1. Distribution of phrase types of the annotated spatial expressions

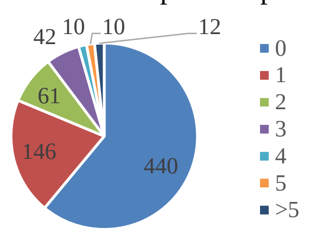

Figure 2. Distribution of the distances between event mentions and spatial expressions

Figure 1 shows the distribution of phrase types of the annotated spatial expressions. The phrase types are obtained from corresponding phrase tags in the OntoNotes parse tree annotations. The figure suggests that adverbial phrases such as locative prepositional phrases take up only less than half the whole spatial expressions and that it is also important to consider noun phrases as candidate spatial expressions.

Figure 2 shows the distribution of distances between event mentions and spatial expressions in the corpus. The distance here is the number of sentences between them. If the distance is zero, it means that they exist in the same sentence. The figure suggests that in many cases, spatial information about events can be found in local context; 61% (440/721) of them are found in the same sentence. The other associations, however, would require some degree of inference.

4.2 Disagreement analysis

Inter-annotator agreement: Conventional IAA measures such as Cohen's Kappa are not applicable to our task because we are not dealing with the data from a fixed set of categories; for each event mention, the annotators must choose the text span of spatial expressions from the entire text. In this work, we address IAAs by calculating the ratio of event mentions for which the two annotators agree. In order to make comparisons for different levels of strictness, we consider the following four cases of agreements for each event mention in documents.

(a) SIMPLE MATCHING: The two annotators both agree or disagree that the given event mention is associated with some spatial expression.
(b) SPAN OVERLAPPING: One of the spatial expressions annotated by one annotator overlaps one of the spatial expressions annotated by the other annotator. This corresponds to a loose measure for comparing two associations.
(c) SPAN MATCHING: *Each* spatial expression annotated by one annotator exactly matches with one of the spatial expression annotated by the other annotator, and vice versa.
(d) SPAN AND LABEL MATCHING: The text span and label of *each* spatial expression annotated by one annotator matches with those of one of the spatial expressions annotated by the other annotator, and vice versa. This corresponds to a strict measure for comparing two associations.

Measure \ Phase	1	2	3	4	Avg.
SIMPLE MATCHING	0.79	0.78	0.80	0.86	0.81
SPAN OVERLAPPING	0.58	0.63	0.49	0.63	0.60
SPAN MATCHING	0.48	0.53	0.39	0.44	0.47
SPAN AND LABEL MATCHING	0.48	0.49	0.36	0.42	0.44

Table 3: Inter-annotator agreements

Table 3 shows inter-annotator agreements for each phase and the entire corpus. Although the overall agreements are not very high, we believe that this is due to the highly implicit nature of spatial information in discourse as discussed in Roberts et al. (2012). The task requires the combination of contextual clues and the world knowledge, and relies heavily on the annotators' intuition and interpretation of implicit information. The annotators sometimes have different interpretations of definite and plausible associations, though they reached an agreement in the discussion after each annotation phase. Near-perfect agreement would thus not be a practical goal in this task.

Simple mistakes: Aside from the disagreements caused by the implicit nature of the task, the annotation within the current framework also produced a number of mismatches in choosing the exact span of spatial expressions even though the annotators correctly chose their head word. They also sometimes made a mistake in choosing the longest phrase by dropping modifiers. Another type of mistake is not to choose the closest one. This case happens when different expressions that refer to the same place are mentioned in a single document. The annotators often missed a closer expression and chose the distant one that refers to the same place, which are not actually genuine errors. We found that more than 40% of the disagreements in phases 3 and 4 are due to these types of mistake.

Although it is difficult to clearly classify the type of disagreements other than the mistakes above, we found that there are some frequent cases of disagreements as shown below.

Remote agents: In some cases, it is not clear whether the agent is remotely involved in a given event. This may cause disagreements between the annotators, as shown below

(8) Kostunica says he won't [turn]$_{\text{EVENT}}$ Milosevic over to a tribunal [in The Netherlands]$_{\text{SPACE}}$ where he was indicted as a war criminal.

One of the annotators was confused whether event *turn* can be associated with spatial expression *in The Netherlands*. The discussion led them to agree that event *turn* does not necessarily imply its agent being located in the remote place if there is no further contextual information that supports it.

Abstract events: It is often difficult to identify the spatial bounds of events because of their vague interpretation.

(9) After today's air strikes, 13 Iraqi soldiers [abandoned]$_{\text{EVENT}}$ [their posts]$_{\text{SPACE}}$ and [surrendered]$_{\text{EVENT}}$ to Kurdish fighters.

Here, while it is clear that event *abandoned* and spatial expression *their posts* are associated, it might not be so clear whether *surrendered* and *their posts* can be associated with each other. One annotator considered *surrendered* as a kind of declaration and associated it with *their posts*, but the other annotator considered that *surrendered* involves the location change of its agents into the place where the entities to which they surrender are located, i.e., the location of *Kurdish fighters*. For this disagreement, the annotators agree that there is a plausible association between *their posts* and *surrendered*.

Containment relations among spatial entities: Some documents contain several expressions that refer to geographic regions in spatial containment relations. For example, one of the documents in our corpus has seven candidate expressions that spatially contain one another, as shown in Figure 3. This made it difficult for the annotators to determine which must be chosen as a spatial expression for a given event mention, especially when its spatial boundary is not clear.

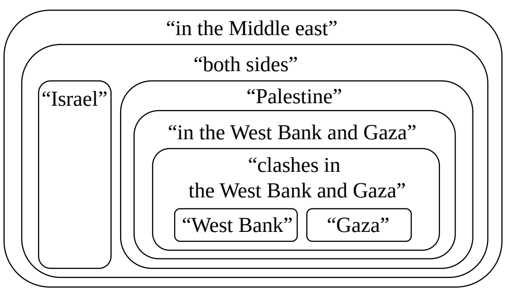

Figure 3. Relative containment relations among seven spatial expressions

Cascading disagreement: Disagreement that arises in a particular event mention often propagates through other neighboring mentions, especially when a set of related events that occur in a short time is mentioned in consecutive sentences. This is because the annotators usually try to cluster similar events first, and then associate them with a particular special expression at the same time.

5 Discussion and Concluding Remarks

In this work, we proposed our framework for annotating associations between event mentions and spatial expressions to analyze spatial information about events in text. Although the highly implicit nature of spatial information makes it difficult to achieve consistent annotation, we see that further improvements can be made on our current framework.

One of them is to restrict event mentions and spatial expressions to a certain category of words in order to remove cases where their spatial boundaries are too implicit. For instance, we could annotate only the event mentions referring to the situation that can be temporally anchored as in TimeML, or could restrict spatial expressions to geographical entities as in SpatialML and ISO-Space.

In order to avoid disagreements raised by mistakes in choosing an exact text span of spatial expressions, we may allow for annotating only head words and let the annotation tool automatically choose the longest phrases using the parse tree of the OntoNotes annotation because our goal is not to identify the exact boundary of such phrases.

Another possible improvement is to augment the current annotation to incorporate further linguistic information in order to facilitate the annotation process and to enable more practical evaluation. The most important one would be to annotate the spatial containment and coreference relations among spatial expressions. As discussed in our disagreement analysis, the annotators often make mistakes or disagree when choosing among spatial expressions that refer to highly overlapping regions, as in Figure 3. It may not be practical to make a sharp distinction among them. The current IAA measures in our framework do not consider the possibility of 'partial matches': for example, "*in the West Bank and Gaza*" and "*clashes in the West Bank and Gaza*". In order to assess the performance of an automated recognition system, there should also be a proper evaluation metric that compensates for these cases, such as CEAF in coreference resolution (Luo, 2005).

Future work also includes increasing the size of the present corpus and augmenting it with other layers of linguistic information such as event coreference. We also plan to build an automated system to recognize the associations with various linguistically motivated features. Our corpus is publicly available at http://nlp.kaist.ac.kr/resources.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2014R1A2A1A11052310).

References

Blanco, Eduardo and Alakananda Vempala, 2015. Inferring Temporally-Anchored Spatial Knowledge from Semantic Roles. Proceedings of the 2015 Annual Conference of the North American Chapter of the ACL, 452-461.

Hovy, Eduard, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: the 90% Solution. Proceedings of the Human Language Technology Conference of the NAACL, 57–60.

Kolomiyets, Oleksandr, Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2013. SemEval-2013 Task 3: Spatial role labeling, 7th International Workshop on Semantic Evaluation.

Kordjamshidi, Parisa, Steven Bethard, and Marie-Francine Moens. 2012. SemEval-2012 Task 3: Spatial role labeling. 6th International Workshop on Semantic Evaluation.

Kordjamshidi, Parisa, Martijn van Otterlo, and Marie-Francine Moens. 2010. Spatial role labeling: Task definition and annotation scheme. 7th International Conference on Language Resources and Evaluation.

Lapata, Mirella and Alex Lascarides. 2006. Learning sentence-internal temporal relations. Journal of Artificial Intelligence Research, 27:85-117.

Luo, Xiaoqiang. 2005. On coreference resolution performance metrics. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, 25-32.

Mani, Inderjeet, Janet Hitzeman, Justin Richer, Dave Harris, Rob Quimby, and Ben Wellner. 2008. SpatialML: Annotation scheme, corpora, and tools. 6th International Conference on Language Resources and Evaluation.

Mani, Inderjeet, Marc Verhagen, Ben Wellner, Chong Min Lee, and James Pustejovsky. 2006. Machine learning of temporal relations. Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, 753-760.

Mirza, Paramita and Sara Tonelli. 2014. An analysis of causality between events and its relation to temporal information. Proceedings of the 25th International Conference on Computational Linguistics, 2097-2106.

Pustejovsky, James, José Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003a. TimeML: Robust specification of event and temporal expressions in text. 5th International Workshop on Computational Semantics.

Pustejovsky, James, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. 2003b. The TIMEBANK corpus. Proceedings of the Corpus Linguistics 2003 conference, 647-656.

Pustejovsky, James and Jessica Moszkowicz. 2008. Integrating motion predicate classes with spatial and temporal annotations. Proceedings of the 22nd International Conference on Computational Linguistics, 95-98.

Pustejovsky, James, Jessica Moszkowicz, and Marc Verhagen. 2011a. ISO-Space: The annotation of spatial information in language. Proceedings of the 6th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation, 1-9.

Pustejovsky, James, Jessica Moszkowicz, and Marc Verhagen. 2011b. Using ISO-Space for annotating spatial information, 10th International Conference on Spatial Information Theory.

Roberts, Kirk, Travis Goodwin, and Sanda Harabagiu. 2012. Annotating spatial containment relations between events. 8th International Conference on Language Resources and Evaluation.

Roberts, Kirk, Michael A. Skinner, and Sanda M. Harabagiu. 2013. Recognizing spatial containment relations between event mentions. 10th International Conference on Computational Semantics.

UzZaman, Naushad, Hector Llorens, Leon Derczynski, Marc Verhagen, James Allen, and James Pustejovsky. 2013. SemEval-2013 Task 1: TempEval-3: Evaluating events, time expressions, and temporal relations. 7th International Workshop on Semantic Evaluation.

Verhagen, Marc, Robert Gaizauskas, Frank Schilder, Mark Hepple, Jessica Moszkowicz, and James Pustejovsky. 2009. The TempEval challenge: Identifying temporal relations in text. Language Resources and Evaluation, 43(2):161–179.

Verhagen, Marc, Roser Saurí, Tommaso Caselli, and James Pustejovsky. 2010. SemEval-2010 Task 13: TempEval-2. 5th International Workshop on Semantic Evaluation.

Yoshikawa, Katsumasa, Sebastian Riedel, Masayuki Asahara, and Yuji Matsumoto. 2009. Jointly identifying temporal relations with markov logic. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 405-413.