

# Temporal Relations Annotation and Extrapolation Based on Semi-intervals and Bounding Relations

**Alejandro Pimentel**

pimentel@comunidad.unam.mx

**Gemma Bel-Enguix**

gbele@iingen.unam.mx

**Gerardo Sierra**

gsierram@iingen.unam.mx

Universidad Nacional Autónoma de México / México

Instituto de Ingeniería / Ciudad de México

**Azucena Montes**

amr@cenidet.edu.mx

Tecnológico Nacional de México / México

Centro Nacional de Investigación y Desarrollo Tecnológico / Cuernavaca Morelos

## Abstract

The computational treatment of temporal relations is based on the work of Allen, who establishes 13 different types, and Freksa, who designs a cognitive procedure to manage them. Freksa's notation is not widely used because, although it has cognitive and expressive advantages, it is too complex from the computational perspective. This paper proposes a system for the annotation and management of temporal relations that combines the richness and expressiveness of Freksa's approach with the simplicity of Allen's notation. Our method is summarized in the application of bounding relations, thanks to which it is possible to obtain the temporary representation of complete neighborhoods capable of representing vague temporal relations such as those that can be frequently found in a text. Such advantages are obtained without the need to greatly increase the complexity of the labeling process since the markup language is almost the same as TimeML, to which only a second temporary "relType" type label relationship is added. Our experiments show that the temporal relationships that present vagueness are in fact much more common than those in which a single relationship can be established precisely. For these reasons, our new labeling system achieves a more agreeable representation of temporal relations.

## 1 Introduction

The understanding of the sequence in a narrative is a fundamental aspect in the comprehension of a text. News, novels, stories, narratives and many other types of texts contain events that occur in a certain chronological order, whether they express time explicitly or implicitly. The reader needs to obtain the temporal notion of the events that occur within a text in order to understand what is being narrated.

There has been a large number of investigations, mainly theoretical, for the use of temporal relations to establish a chain of reference points, consisting mainly of events anchored to a known point in time or another event. With the help of these chains, timelines can be arranged. This type of approach allows for an abstract representation of time that is similar to the way in which humans communicate, since they rely on relative expressions, often using as reference the moment of speech or creation of the document. Temporal relations can be used similarly (Allen, 1983; Allen and Koomen, 1983; Vilain and Kautz, 1986; Freksa, 1992; Bramsen et al., 2006a; Bramsen et al., 2006b; Chambers et al., 2014).

The identification of temporal relations between pairs of events or time intervals is a widely used way to express time. This type of annotation uses a set of realizations that has been designed as the core to classify and annotate each one of the relations of the entities of a text. The set of representations is very important for annotation and temporal reasoning. Allen (1983) and Freksa (1992) have developed tools for the construction of these representations.

---

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

Knowing how to annotate temporal relations is necessary to deal with them. The annotation of temporally related events is absolutely dependent on the set of temporal relations that is used. Allen and Freksa are the authors of the most widespread theories on the topic. Freksa's temporal neighborhoods are richer than Allen's proposal in the representation of time, but also more complex (31 features vs. 13). Therefore, Allen's relations are usually preferred when labeling a corpus, although they can cause confusion when ambiguous temporal relations are presented in the texts (Derczynski, 2016).

This work presents a variant in Freksa's system of annotation and handling of temporal relations. The goal is to be as rich in the expressiveness of time as Freksa's relations while keeping the simplicity of only using Allen's system. The nucleus of the proposal is the ability to handle ambiguity in temporal relations thanks to the compatibility our system has with Freksa's neighborhoods.

Section 2 shows the theory and sets of temporal relations of both Allen and Freksa, as well as the most important features of these systems our proposal is based on. In Section 3, we analyze aspects regarding expressiveness of annotation systems. In Section 4, we detail our proposal of bounding temporal relations. Section 5 explains an example of the 'relabeling' of a corpus that was already annotated with Allen's set of tags. In Section 6, some results obtained with the 'relabeling' process are shown. The paper closes in Section 7 with conclusions and future work.

## 2 Temporal Relations

Time can be considered as a linear phenomenon. If there are two events or entities that can be assigned a temporary space, it is possible to give them an order or identify the relative position within a general timeline. If there are, for example, a pair of events *A* and *B*, we could order them as: *A* occurs before *B*, *A* occurs after *B*, *A* begins before *B* finishes, etc. The different ways in which events may be related are defined in a set of possibilities that form the different temporal relations.

### 2.1 Allen's proposal

Allen (1983) states that a good model and representation of time must fulfill four main points: 1) It must allow imprecision. Much temporal knowledge is relative (for example, "A was before B") and has little to do with absolute dates. 2) It must allow uncertainty in the information. In many cases, the exact relation between two moments in times is unknown, but one can get clues about how they can be related. The transitivity that can be found in the algebra Allen developed is a clear example of uncertainty. 3) It must allow variation in the dimensionality of reasoning. In the case of history, for example, time can be considered only in terms of days or even years. However, when dealing with phenomena or designs, it is necessary to consider dimensions such as milliseconds or less. 4) It must allow persistence. That is, it must be able to consider inferences such as "If I left the car outside this morning, it should still be there now".

The same author establishes the thirteen relationships that can exist between two events. This is shown in Table 1 in three columns. The left column indicates the annotation for each of the relations. They are paired with the inverse relations, so that in each row of this column two symbols can be observed, one that relates event *X* to event *Y*, and the other for the inverse (event *Y* to event *X*). The central column graphically shows the relation for a pair of *X* and *Y* events. The illustration shows a line that represents the duration of the events in a schematic way. The position of the lines between them expresses the temporal relation that occurs between them. Each illustration corresponds to two relations that are inverses, except for **equal**, whose inverse is the same relation. The column on the right contains the interpretation relation with words.

An important aspect of Allen's work is that having an event related to a second event, and the latter being related to a third one, the first one can also be related to the third one. This author studies this phenomenon and develops an algebra of intervals for the thirteen categories that allows knowing the way in which the events that are indirectly related, are linked. To represent the uncertainty of the concrete relation between a pair of events, he uses logical operations with the possible relations between the pairs.

Relation	Illustration	Interpretation
$X < Y$		$X$ before $Y$
$X > Y$		
$X \text{ m } Y$		$X$ meets $Y$
$Y \text{ mi } X$		
$X \text{ o } Y$		$X$ overlaps $Y$
$Y \text{ oi } X$		
$X \text{ s } Y$		$X$ starts $Y$
$Y \text{ si } X$		
$X \text{ d } Y$		$X$ during $Y$
$Y \text{ di } X$		
$X \text{ f } Y$		$X$ finishes $Y$
$Y \text{ fi } X$		
$X = Y$		$X$ equals $Y$

Table 1: Allen's temporal relations.

## 2.2 Freksa's annotation

Freksa (1992) introduces neighborhoods of temporal relations to deal with the problems of complexity that Allen's algebra has when working with large amounts of indirect relations (Vilain and Kautz, 1986). Freksa retakes the usage of time intervals to develop his work from a cognitive perspective. He takes into account neighborhoods of temporal relationships linked by perception restrictions. Figure 1 shows the scheme representation given by Freksa for Allen's thirteen relationships. The cognitive intuition behind Freksa's representation is that certain temporal relations are more similar between them than others. The closer two relations are, the easier it is to transform them into the other when moving their boundaries. The diagram from Figure 1 arises from this principle. Different ways of manipulating the ends of the events results in different connection paths in the diagram. For example, Figure 2 shows the connections that consider relations adjacent to those that can be transformed directly, that is, if one can convert a temporary relation into another when moving one of the ends of the relation.

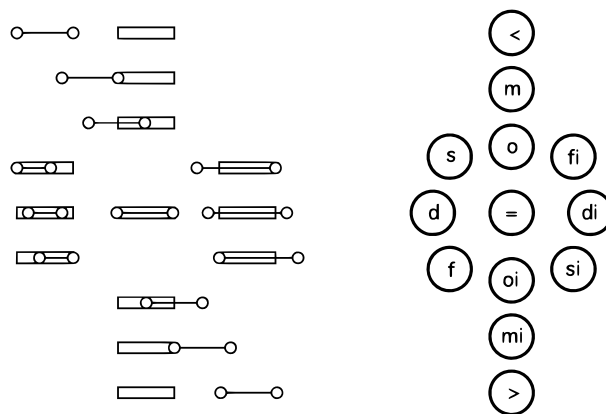


Figure 1: Temporal relations structured on neighborhoods. On the right side, relations of Table 1 are introduced. On the left, schematic representation of the intervals of two events according to the relations on the right.

Within the structure, a set of relations held together through the connections is considered a neighborhood of temporal relations. Neighborhoods can be used to handle temporary information in a simpler way in cases of uncertainty. The algebra of Allen (1983) results in the disjunction of a group of temporal relations. With thirteen different relationships in a disjunction to express ambiguity, there are up to 8,192 different possible disjunctions. Freksa (1992) notes that an improvement in the handling of information can be achieved since the groups of relations acceptable for perception form neighborhoods within their diagram. Around 250 relations are needed to express all the combinations.

As for the transitivity results, all of them are continuous points in the diagram and, moreover, a small fraction of combinations that can be obtained. Frekxa’s system has 31 neighborhoods.

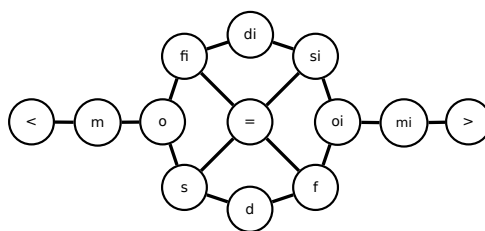


Figure 2: For Frekxa, the connections between relations are given according to the possible transformations between them.

### 3 Temporal relations compromise

To work with temporal relations, some annotation system is needed to allow the connection of two events in a specific way. Therefore, the set of temporal relations becomes very important since it delimits the quantity of labels. Allen’s set of relations includes 13 different labels, while Frekxa’s suggests 31 categories to manage neighborhoods. Having a smaller set of relations represents a great increase in the simplicity of labeling and processing. Many works use a reduced set of labels – between four to six – and get good results (Setzer et al., 2005; Verhagen et al., 2007; Verhagen et al., 2010; Kolomiyets et al., 2012; Chambers et al., 2014). However, it also implies a loss of information. A bit more sophisticated is the annotation system of the 3rd TempEval, that uses the complete set of Allen’s relations. Nonetheless, there are some modifications that should be considered, as Derczynski et al. (2013) report.

Derczynski (2016) describes the sets of temporal relations according to two dimensions which describe the range combinations that they are capable of capturing. The first one is *expressiveness vs. simplicity*, which refers to the range of different combinations that the system can capture, the simpler the system, the fewer combinations that can be expressed; and the second one, *specificity vs. laxness*, which refers to how constrained a time relation is once a label is assigned, or how much information is needed to be able to do the labeling without having to assume information. These dimensions are generally inversely proportional, since the more accurately a representation expresses a temporal relation, the more complex it is. Also, when a representation can specify exactly the temporal relation between two events, the possibility of vagueness is lost.

Allen’s temporal relations are richer than the simple ones (like the ones used on early versions of TempEval) and provide more accurate information about the way events are temporally related. However, they are disadvantageous in that it is necessary to know both intervals completely in order to accurately establish any temporal relation. Such specific knowledge is lost by inferring temporal relations from related ones (see the case of temporal algebra (Allen, 1983)) and difficult to find in general texts. Derczynski (2016) introduces an example called *the opera problem*, that illustrates this phenomenon:

Given the statement *Irene went to the opera today*, interval *O* is the visit to the opera and interval *T* is today. As shown in Figure 3, without knowing when Irene left, one cannot choose a single interval relation out of: a) *O* is included in *T*; b) *O* finishes *T*; c) *O* overlaps *T*.

In annotation systems, it is preferable to have a single temporal relation linking two events. Nevertheless, when cases such as the *opera problem* arise, the annotators find themselves in a confusing situation. This problem is frequent in annotation systems such as TimeML, that results in a reduced agreement between annotators. In the particular case of TimeML, its inter-annotator agreement on relation types was 0.71 *kappa* among experts in this kind of annotation.

Derczynski (2016) experiments with several different types of representations of temporal relations, concluding that although a representation with greater expressiveness is preferable, a greater expressiveness in a representation of temporal relations is inversely proportional to the performance of a system that uses them.

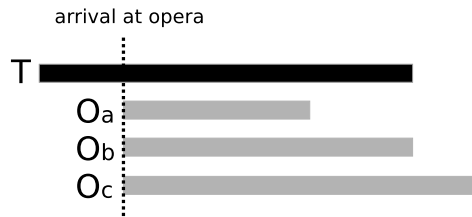


Figure 3: The Opera problem: linguistic ambiguity leads to an inability to choose a single interval relation. Here, the time of arrival at the opera is known, but not the departure, making it hard to relate intervals  $O$  and  $T$ .

#### 4 Our proposal: Bounding temporal relations

We consider that Allen's representation of temporal relations has as much *expressiveness* as possible, that is, we will not be able to find more specific temporal relations. In contrast, we find that the annotation of this kind of relations is deficient in *laxness* due to the constraint of being able to handle only an absolute *specificity*.

On the other hand, the representation of Freksa's temporal relations has a greater degree of *laxness* due to the possibility to deal with several possible temporal relations between a pair of events. Despite this, its *simplicity* decreases considerably because of the many additional tags that the annotator needs to know in order to be able to correctly label the neighborhood (31 vs 13). We assume this is the reason why there are no corpora labeled with this kind of representation.

As stated in the example of the *opera problem* (see Figure 3), a lack of complete temporal information can result in an ambiguous temporal relation that different annotators will easily interpret differently. The objective of this paper is to propose a temporary representation that facilitates dealing with ambiguous temporal relations without greatly increasing the complexity of its annotation. That is to say, we seek to increase the *laxness* of Allen's temporal relations without harming the *specificity* that it is able to capture, in other words, increase its *expressiveness* with a minimal decrease of *simplicity*.

By studying all the neighborhoods suggested by Freksa (1992), it was found that we could take advantage of the way that the relations turn out to be arranged within the neighborhoods. We claim it is possible to identify the relations that are delimited by the neighborhood boundaries. We found that for every neighborhood there is a pair of bounding relations that mark the extremes that encompass it and provide enough information to obtain the complete neighborhood; in other words, no pair of different neighborhoods has the same bounding relations.

In order to get a correct management of temporal information by means of bounding relations, it is necessary to have a way to crowd the full temporal neighborhood from those bounding relations. Therefore, a simple algorithm for filling Freksa's neighborhoods is also proposed. Given that the tagged, or otherwise automatically detected, bounding relations may not be compatible with any of the neighborhoods, the algorithm is responsible for generalizing rules to obtain a plausible one (even if it is not one of Freksa's set) for any pair of given bounding relations.

The procedure to find the neighborhood of relations uses a graph in which temporal relations are nodes connected with edges, as shown in Figure 4. Starting from the pair of bounding relations nodes in this graph, it is possible to get the temporal neighborhood by calculating the union of the shortest paths between them. The image, despite being very similar to that of Figure 2, presents a couple of extra edges that are very important to obtain reliable neighborhoods for any case.

With this, it is possible to obtain the Freksa's neighborhoods that correspond in their extremes with the bounding relations. If a couple of relations without an equivalent in Freksa's neighborhoods are found, the generalized rules will find a compatible neighborhood for the bounding relations.

##### 4.1 Usage of bounding relations

The above refers to a conceptual summary of the modification we are presenting to the temporary representations of Allen and Freksa. However, what implications does this have? What are the specific

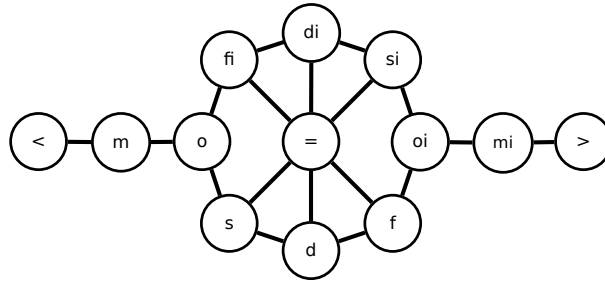


Figure 4: Adjacent temporal relations graph useful for temporal neighborhood crowding.

modifications that should be considered within this new representation?

The proposed changes are based on the TimeML markup language. This system proposes 4 types of tagging structures: EVENT, TIMEX3, SIGNAL, and LINK. EVENT is a structure that denotes the events of the text, TIMEX3 is used to identify those words with explicit temporal expressions while SIGNAL is used to denote the words that indicate how they are related to each other, and finally, the tag of most interest to this paper: LINK. There can be three types of LINKs: SLINK (for subordination), ALINK (which links an aspectual event and its argument event), and TLINK (temporal link). The latter is the one that indicates the type of temporal relationship between two events, this is precisely where the proposed change will reside. TLINK consists of the two temporally related instances and the temporal relation, which can be one of Allen’s 13 temporal relations. We illustrate this with the following example taken from the TimeML annotation guides (Saurí et al., 2006):

- *John taught on monday*

```
<TLINK eventInstanceID="ei1" relatedToTime="t1" signalID="s1"
relType="IS_INCLUDED" \>
```

The previous example shows how the temporal relation from the sentence *John taught on monday* is labeled according to the TimeML annotation guidelines. The example shows only what corresponds to the TLINK tagging; however, we can see that this tag refers to the event (EVENT) ”ei1” and the time (TIMEX3) ”t1”, which corresponds to the words ”taught” and ”monday” respectively; both instances are related by the signal (SIGNAL) ”s1” given by the word ”on”. Finally, we have the temporal relation ”relType” which is the only element that will be affected by our new way of tagging.

What we propose is the addition of a second *relType* tag. With this, the previous example would look like this:

- *John taught on monday*

```
<TLINK eventInstanceID="ei1" relatedToTime="t1" signalID="s1"
relType1="IS_INCLUDED" relType2="IS_INCLUDED" \>
```

Therefore, there are two *relType* tags: *relType1* and *relType2*. We can see that both values have the same label; this will be the case when the annotator has all the necessary information to establish one and only one temporary relation safely.

Considering the opera problem (see Section 3) there can be a confusion for the assignment of the temporal relation between IS\_INCLUDED, FINISHES, and DURING. That would be, in Allen’s annotation, the relations: *oi*, *f*, *d*.

The annotators could have the option of labeling the relations with all the possible temporary relations that they consider relevant. However, this would make the labeling more complex. Furthermore, the system of labels should be able to handle multiple values for a single category. An alternative option is to use Freksa’s annotation, which for the case of the example has the neighborhood: *younger contemporary of*. This single tag is able to express the three possible relations that the temporal relation of the problem of the opera can take. Nonetheless, this labeling is also complex because the annotator would have to handle a set of 31 neighborhoods instead of 13 relationships.

In our system, the labeling is done only for the bounding relations of the neighborhood; no temporal relations are added to the annotation set since there are still 13 categories. The labels receive a second value of type of relation (*relType*); no matter the size of the neighborhood, it will always be a constant of two. For the case of the annotator, the change would be to focus, within the confusing relations, on those that work as limits for the neighborhood without worrying about those intermediate relationships. As it was shown in the example (Figure 3), the bounding relations for this particular case would be: *a* and *c* and the label would be as follows:

- *Irene went to the opera today*

```
<TLINK eventInstanceID="e11" relatedToTime="t1" signalID="s1"
relType1="IS_INCLUDED"
relType2="DURING" \>
```

With the above, it is possible to perfectly retrieve one of the neighborhoods found by Freksa. In Figure 5, this process is exemplified. We have the bounding relations that delimit the possible temporal relations. The figure highlights the *d* and *oi* relations, as well as the shortest route in the graph that allows these nodes to be joined. This shortest route passes through the relation *f*, which is added to the neighborhood once the information of the complete set that expresses the possible temporal relations in the problem of the opera is extrapolated.

In order to locate the shortest paths for the crowding of the temporal relations (any algorithm for routing between two nodes of a graph can be used) it is important to take into account that if there is more than one route, the union of all must be considered.

This change for dealing with temporal relations allows the complexity of the annotation to be marginally increased since the labels remain exactly the same. At the same time, it offers all the flexibility and expressiveness present in the temporary neighborhoods of Freksa's neighborhoods.

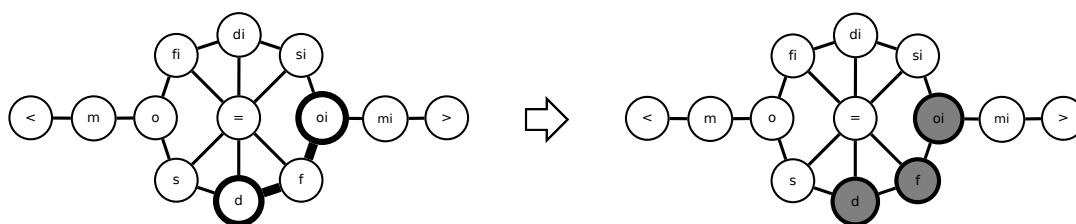


Figure 5: For the opera problem, there is confusion between the relations in the *younger contemporary* of neighborhood. The bounding relations (*d* and *oi*) are enough information for this neighborhood to be found automatically.

## 5 Relabeling a time annotated corpus

In order to test our claims, we organized a relabeling of a time annotated corpus with our new proposed method for temporal relation identification. In this section, we explain the details of the process of relabeling, as well as the results and insights we could observe.

For the relabeling process, we worked over the TimeBank 1.2 Corpus. This corpus has 183 news articles that have been annotated following the TimeML 1.2.1 specification. The taggers were instructed in the possibility of representing vagueness in temporal relations and how to achieve this by identifying the bounding relations. With this information, the taggers were completely free to decide whether the text presented a specific relation, a vague relation, or no relation at all.

We developed a GUI system that reads the original annotated file relation by relation and asks the annotator for the pair of bounding relations. Once the relabeling is done, the result is a file with the same format as the originally annotated corpus but with two labels for the temporal relation instead of one.

The relabeling task was performed independently by two English speakers outside this research who were debriefed of the objectives of temporal annotations and the new method we are proposing, along

with instructions for using as many temporal relations as they understood from the text. Table 2 shows examples of different kind of relations that were found in the tagging process. The resulting corpus after this relabeling method is freely available<sup>1</sup>

Example	Tagger1	Tagger2
Revenue <b>gained</b> 13% to \$77.3 million from <b>\$68.5</b> million.	MEETS- MEETS	MEETS- MEETS
The insurer’s earnings from commercial property/casualty lines <b>fell</b> 59% in the latest quarter, while it <b>lost</b> \$7.2 million in its personal property/casualty business	EQUALS- OVERLAPPED	EQUALS- OVERLAPPED
She said the move would <b>result</b> in a after-tax charge of less than \$4 million to be <b>spread</b> over the next three quarters	OVERLAPS- EQUALS	MEETS- EQUALS
Compaq Computer nose-dived \$8.625 a share, to \$100, and <b>pulled</b> other technology issues lower after <b>reporting</b> lower-than-expected earnings after the stock market closed Wednesday	MET-AFTER	STARTS- AFTER

Table 2: Examples showing the kind of events that were given to the taggers, as well as their respective responses. Temporal relations are not symmetric. In the examples, **green events** are related to **blue events**.

## 6 Results

Although the new relabeled corpus contains two temporal relations for every pair of related events, it is necessary to complete the temporal neighborhood from the pair of bounding relations in order to analyze the proposed method. For this, we developed a system that follows the procedure described in Section 4.1 to get the full elements of the neighborhood expressed as a combination of Allen’s original set of relations. From the relabeling, a result that seems important to mention is the proportion of vagueness that the taggers found in the texts. That is, of all the original TLINKs, how many of them were cataloged as a single well-defined temporary relation, and how many were perceived as a vague relation that could be interpreted as more than one temporary relation.

On the one hand, one of the taggers found temporal relations present in 67.03% of the pairs linked by TLINK, while the other tagger found temporal relations in 69.96% of the pairs of events.

Within those events that they considered with a temporal relation, the first tagger found that in 94.64% of cases the temporal relation could be interpreted in more than one way. For the second tagger, this proportion rose to 99.75% of the cases. This result reflects the ambiguity that is actually present in the temporary expression of events written in natural language. Also, for previous temporary tagging methods, there is a vagueness that could not be represented.

As an insight of the labels that were able to be accurately assigned to a single temporal relation, we find it interesting to highlight that 86.60% of the cases coincided with the EQUALS relation.

With the new temporary labeling system that we are proposing, the categories of temporal relation that may exist between pairs of events are no longer mutually exclusive. That is why the Cohen’s kappa coefficient for annotator agreement measurement cannot be applied. Nevertheless, we can measure the agreement between taggers, taking into consideration the intersection that exists between neighborhoods that results from the tagged bounding relations. Considering full neighborhoods, we note intersection exists between neighborhoods in 75.84% of the first tagger’s cases, and in 79.15% of the second tagger’s cases.

In Figure 6, side by side, the general proportion from both taggers are shown: The internal ring on both sides shows the relation between temporarily related events (orange) *versus* not related events (blue). On the middle ring, we highlighted in yellow the proportion of events tagged as vague, and finally, the

<sup>1</sup><https://gitlab.com/gil.iingen/br-timeml>



external ring shows in green the proportion of relations where we found intersection between the tags of the resulting neighborhoods.

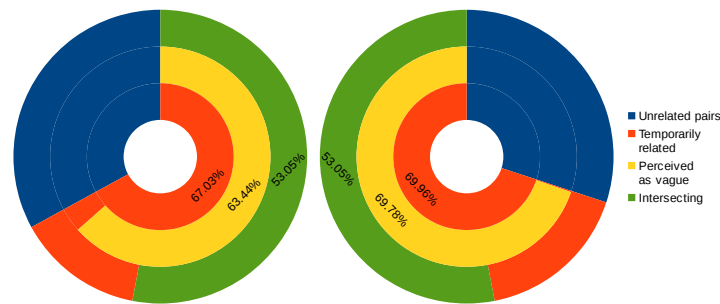


Figure 6: Taggers proportions side by side. On the inner ring, related vs unrelated events. On the middle ring, detected vagueness is highlighted. On the outer ring, the agreement is shown.

We consider a complete intersection when the pair of labels of two taggers match. This happens in 36.84% of the cases where intersection appears. However, we shall also consider that intersection is not limited to these cases since, although the pair of labels do not match, it is possible that some of its neighborhood nodes intersect. Then, we name this kind of cases as partial intersection. On the re-tagged case, we found 63.16% of partial intersection of which 73.08% of the cases intersect in at least half of the nodes. That is another point that reinforces the importance of vagueness.

Figure 7 represents the total intersected tags, differentiating how many of them match completely versus how many of them do partially. In the latter case, we note how many of them match partially in at least half of their nodes. Furthermore, to evaluate the new agreement between the labels and to have a better insight on the accuracy of the method and the vagueness and confusion of the temporal expressions in the texts, a dilation metric was executed. Dilation is a useful tool to know how different the re-taggings were. As mentioned in Section 4.1, our proposed method is the completion of a neighborhood from the bounding relations of the labeling. That said, if we take the labels of both taggers as bounding relations to make a global completion of the neighborhood, we can measure the discrepancy between the two labels by knowing the difference in the number of nodes that occur between the individual neighborhood and the neighborhood formed by both taggers together. This difference in the number of nodes is what we mean by dilation.



Figure 7: Intersection distribution: the proportion of exact matches inside the intersecting neighborhoods, alongside with the proportion of the intersection that covered half or more of the neighborhood.

For the cases where there were intersection between the re-tagged neighborhoods, the average number of nodes of the resulting graph was 2.40 nodes, while the average dilation was of 0.14, while on the other hand, on the cases where there was no intersection, the average number of nodes in the graphs were of 1.92 and the average dilation was of 4.05. Thanks to these data, we can observe that the intersection between the taggers is not due to the choice of very wide neighborhoods but that the temporal notion that they perceive effectively corresponds to a reduced, but variable, range of possibilities. We can also observe

that the dilation metric varies dramatically when taking “positive” and “negative” cases separately since for the cases where there is no intersection, the metric indicates that it is necessary to add on average 4 nodes to be able to connect the assigned tags. It is also important to keep in mind that with this metric, the higher the average number of nodes per graph is, the smaller the dilation should tend to be since in order to join two large sets of nodes, logic indicates that fewer nodes should be necessary.

Taking into account the totality of the data, we will have an average of 2.25 nodes per neighborhood and a dilation of 1.37. We observe that, in general, the temporal relations tags do not usually move away from each other farther than one separation node. This can be contrasted with a baseline “labeling” of temporal relations, where two automatic taggers randomly assigned bounding relations for the corpus. For these cases, by repeating the same comparison procedure between the taggers, we obtained an average of 4.56 nodes per neighborhood and a dilation of 2.70, that is, larger neighborhoods that nevertheless need more nodes to be able to join. A summary of the comparison of average number of nodes and the dilation data are shown in Table 3. These results correspond so well with the observations on (Freksa, 1992) that the difference between temporal relations cannot be considered independently.

Neighborhood category	Avg. num. nodes	Dilation
Complete set	2.25	1.37
Intersected	2.40	0.14
Un-intersected	1.92	4.05
Random	4.56	2.70

Table 3: Average number of nodes and dilation comparison between different temporal neighborhoods.

## 7 Conclusion and future work

We have proposed a modification in the annotation and handling of temporal relations to find an efficient way to increase expressiveness and versatility with minimal increase in the complexity in the annotation.

The adaptations this work suggests intend to improve the type of temporary relation that can be represented by the labels without increasing the complexity. The problem that the annotators have with ambiguous relations has already been exposed. However, not all temporal relations are confusing in the same way since those that are closer are those that are more often misinterpreted (Derczynski, 2016). By modifying the labeling system to allow the capture of such ambiguity, we seek to facilitate the uniformity of the labeling. At the moment, there is no record of any corpus of temporal relations tagged with Freksa’s neighborhood system, and we assume that the reason for this is the complexity that the labeling would carry. We seek to overcome this difficulty with this new system. As future work, it would be advisable to undertake the task of labeling or modifying the existing tagged corpora to obtain a training corpus with an expressiveness equivalent to Freksa’s temporal neighborhoods.

## Acknowledgements

We acknowledge the contribution of Diego Córdova. This work is supported by CONACYT scholarship 483759, PAPIIT-UNAM projects IA401219 and AG400119, and CONACYT CB A1-S-27780.

## References

- James F. Allen and Johannes AGM Koomen. 1983. Planning using a temporal world model. In *IJCAI*, volume 8, pages 711–714.
- James F. Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.
- Philip Bramsen, Pawan Deshpande, Yoong Keok Lee, and Regina Barzilay. 2006a. Finding temporal order in discharge summaries. In *AMIA*, volume 2006, page 81.

- Philip Bramsen, Pawan Deshpande, Yoong Keok Lee, and Regina Barzilay. 2006b. Inducing temporal graphs. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 189–198.
- Nathanael Chambers, Taylor Cassidy, Bill McDowell, and Steven Bethard. 2014. Dense event ordering with a multi-pass architecture. *Transactions of the Association for Computational Linguistics*, 2:273–284.
- Leon Derczynski, Hector Llorens, and Naushad UzZaman. 2013. Timeml-strict: clarifying temporal annotation. *arXiv preprint arXiv:1304.7289*.
- Leon Derczynski. 2016. Representation and learning of temporal relations. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1937–1948.
- Christian Freksa. 1992. Temporal reasoning based on semi-intervals. *Artificial intelligence*, 54(1):199–227.
- Oleksandr Kolomiychuk, Steven Bethard, and Marie-Francine Moens. 2012. Extracting narrative timelines as temporal dependency structures. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 88–97.
- Roser Saurí, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. 2006. Timeml annotation guidelines. *Version*, 1(1):31.
- Andrea Setzer, Robert Gaizauskas, and Mark Hepple. 2005. The role of inference in the temporal annotation and analysis of text. *Language Resources and Evaluation*, 39(2-3):243–265.
- Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 75–80.
- Marc Verhagen, Roser Saurí, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th international workshop on semantic evaluation*, pages 57–62.
- Marc B. Vilain and Henry A. Kautz. 1986. Constraint propagation algorithms for temporal reasoning. In *AAAI*, volume 86, pages 377–382.