Determination of event coreference

Peter Francis North
B.A. (Mod.) Computer Science, Linguistics and a Language
Final Year Project, April 2015
Supervisor: Dr. Carl Vogel

School of Computer Science and Statistics
O’Reilly Institute, Trinity College, Dublin 2, Ireland
DECLARATION

I hereby declare that this project is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university

______________________________________________
Peter Francis North

______________________________________________
April 7th, 2015
Acknowledgements

I would like to thank Carl Vogel for his never ending support and commitment to CSLL and the students therein. Carl was always there to lend a hand, and was always helpful and encouraging no matter the problem we faced at any given time. Through his sound advice, Carl was able to ease the pressure of the final year project and final year in general.

I would also like to thank my sister, Nadine, for keeping a smile on my face, and my brother, John, for his moral support and advice throughout the year, both of whom helped to keep me on my toes and cheer me up when the workload was difficult.

Most importantly, I’d like to thank my parents for giving me the opportunity to attend University, and for their unwavering support from the beginning and throughout. With their guidance, I was able to get through the hardest of times.

Thank you, everybody.
Abstract

A wide variety of corpora covering news topics are used today in order to determine event and entity coreference resolution. Some considered to be seminal corpora, such as the ACE and the ecb (Eventcorefbank), have been named as such. Over time, the use of these corpora with evaluation metrics and changing conditions have showcased varying results in coreference resolution over the years. In 2014, an extended version of the ecb corpus, renamed ecb+, was freely released to the public to provide a more precise and “event-centric” corpus with which the determinacy of event coreference resolution could be focused more on the seminal events and their event mentions. With the help of correlation between the seminal events and the event mentions relating directly to the event in question, the ecb+ provides, in theory, a way to determine more accurately the event coreference relation. The question posed in this paper challenges the validity of the corpus with reference to the annotation guidelines, and whether or not the extended corpus will be a valuable source for the determination of events in future. Through the evaluation of the annotation technique, and re-annotation performed on selected texts from the corpus, it was determined that agreement could be found between the annotators of the ecb+ and this study, therefore answering the question that it should prove to be a seminal source for coreference resolution in the future. A number of tests were performed on the texts provided by the corpus, and by studying the annotation guidelines, re-annotating and comparing to the original annotations, this study finds itself in significant agreement with the annotators.
Chapter 1

Introduction

Have you ever read something online that led to the feeling of déjà vu? Did you get that gut feeling, a voice telling you that what you were reading about had happened before? Sometimes, reading about an event in the news can remind us of things which have happened in the past. At times these events may be different, other times they may refer to the exact same occurrence. How can we tell, however, if they are the same? For the human mind, it is simple. Our minds are able to comprehend whether two events are the same based on our understanding of the language used. Language provides us with the context we need in order to understand a situation more clearly. The context of an event can be vague or underwhelming at times, but it can also range to crystal clarity through an abundance of context behind the statement. It is often the case that we would use a search engine to find a specific event from the past. The more we know about an event, the more context we can provide the engine, which in turn leads to more accurate findings. For example, if we know the name of an actor or actress who won an Oscar in a certain year, using a search engine such as google and entering these parameters would most likely lead us to the results we wish to find. If, however, this actor or actress won multiple Oscars on more than one occasion, and the only parameters we mentioned were the name and the word “Oscar”, our results would likely be less accurate, providing us with results from every year an Oscar was won by this person.

Such notions are taken into consideration in this dissertation. By looking at a number of texts in the ecb+ corpus (Cybulska & Vossen, 2014) taken from google news streams, which contain similar seminal events, and analysing the context within, it is the goal of this dissertation to theoretically determine whether or not the event mentions in question are truly coreferent or merely similar in terms of the context provided, and to decide whether or not this corpus is an improvement on previous corpora, making it a better source for those wishing to more clearly determine event coreference. In chapter 2, event and entity coreference will be explained, as well as how they relate to each other. An idea of the previous work in this area with the predecessor corpus (Eventcorefbank) will also be discussed. An attempt will be made to examine and understand the annotation guidelines provided by Cybulska and Vossen (2014) in chapter 3. These annotations will be further studied through analysing the corpus at the level of the XML files contained within. Through analysing the granularity of the corpus, annotations in practice and in theory, a greater understanding of how the “event-centric” approach of the ecb+ aims to be a seminal source for future event coreference resolution will be obtained. The coreference between events and event mentions will be analysed in terms of how they were annotated and the agreement of that with the re-annotations provided in this study. An attempt will be made to also analyse cross-document and within-document coreference with
relation to the event mentions. This will be done in chapter 4. By analysing the annotation guidelines of the ECB+ corpus, and by re-annotating the documents contained within the corpus, we will compare the results of this dissertation to the annotations provided by the annotators of the ECB+ corpus in order to see if we agree. 30 documents will be selected across all topic folders of the ecb+ and re-annotated according to the provided guidelines, and then compared to the actual annotations that were used for the released ecb+ corpus. As well as determining agreement between the annotations of component parts, this will allow for a further test to determine agreement in relation to within-document coreference of events and event mentions. A further number documents will be analysed within one topic folder with the goal of determining how much this dissertation agrees with the cross-document coreferencing of the ecb+. The agreement level will be determined by applying Cohen’s Kappa statistic (Cohen, 1960), with attention being paid mostly to the P(A) coefficient taken directly from Kappa. This coefficient will highlight the percentage to which we agree with the annotators of the ecb+ (Carletta, 1996). In the final chapter, chapter 5, a discussion of what the results mean will be included. The potential agreement in all tests will lead us to decide whether or not the corpus in question can indeed become a seminal source for event coreference. Ideas for work in the future relating to ecb+ and event coreference will be suggested.
Chapter 2

Event and Entity Coreference

2.1. Introduction

As this study focuses on event coreference and the annotations of an “event-centric” corpus, as Cybulska and Vossen describe the ECB+ (2014), it may be prudent to start by defining a number of factors and then bringing them together. A good starting point would be to define exactly what an event is. In order to fully understand the event mentions and their relations in the corpus, providing a definition for entities would also be a good place to start. Throughout the study of event coreference, entity coreference has also been discussed. Therefore, providing a background and definition of both entity and event coreference can provide us with further clarity. Note: the Eventcorefbank+ can be downloaded freely from this source: http://www.newsreader-project.eu/results/data/

2.2. What is an event?

Events have been defined in a number of ways, but there is rarely any real disagreement between the definitions. According to the annotation guidelines provided by the Automatic Content Extraction program, abbreviated as the ACE, one can define an event as a distinct occurrence of something that happens. The occurrence usually involves participants and a change of state (LDC, 2005, cited Cybulska & Vossen, 2014). According to the TimeML specification, an event refers to something that happens or occurs which may be “punctual or last for a period of time”, and which also may describe “states or circumstances in which something obtains or holds true” (Pustejovsky et al., 2003). In essence then, one might say that an event can refer to anything that happens around us which involves human or non-human participants, in which something occurs to change the state of the situation as it stood before the events occurrence.

As events seem to refer to a change of state, or a specific happening, it seems reasonably logical to deduce from this that an event might correspond to verbs or verb within sentences. For example, in “I crashed the car” the event in question is the crashing of a car, and therefore “crashing” seems like the logical choice. This is not an incorrect conclusion to make. However, as pointed out by Webber in 1988 (cited Lee et al., 2012), noun phrases or “NPs can also refer to events”. Looking at the previous example from a different point of view, it is clear to see that this is true. If one was to enquire about what had occurred, the answer describing
the event could be described by the simple noun phrase, “a car crash”. Lee et al. explain in their work (2012) that focusing on NPs is a good way to “restrict the challenging problem of coreference resolution”, but it narrows our view, thus we miss certain relations between noun phrases and verbs. An example they provided refers to the suicide of a suspected Mafia boss, Lo Presti, and shows the event relationship between *hanged* in (a) and *his suicide* in (b):

(a) Police said Lo Presti had hanged himself.

(b) His suicide appeared to be related to clan feuds.

If someone were to examine the above sentences separately, it might be difficult to deduce the relation between the verb in (a) and the noun phrase in (b). However, with context provided, if the two sentences occur in the same news article regarding the same seminal event, it is simple for us to comprehend that both parts of speech refer to the same event.

2.3. What is an entity?

It is common, when investigating coreference resolution, to come across entity coreference resolution more often than event coreference. More often than not, you will be able to find texts which focus on the two together rather than separately. But what exactly is an entity? We have defined what an event is, so how do entities fit into our discussion? Simply put, an entity refers to any non-event constituent in a sentence. Usually, this is limited to nouns, and is never a verb (unless of course the verb is nominalised). Taking the example we used above to describe what an event is, the entities in question would be *Lo Presti*, and *Police*. To be more precise, an entity is “a real-world person, place, organization or object”, (Rao et al., 2010) for example the suspected Mafia boss mentioned here. A narrower definition was provided by Haghighi and Klein (2010), that an entity can be defined specifically as an individual or an object. There is still the question, however, of what entities have to do with events. As I mentioned above, occurrences can be categorised as referring to the same event through context. It is largely through entities that the context is provided for us to determine that similarity. If a real world object can influence event coreference, then it must be relevant. This is something I will explain in the following section.
2.4. Event Coreference and Entity Coreference

In order to explain why entities are relevant to this study, we need first to look at the close ties between event coreference and entity coreference. Although there have been separate studies done on both, it is often difficult to have one without the other to a certain degree, even if the other is renamed or disguised as something else. As definitions for events and entities have now been provided, the next step is to clearly explain what event coreference and entity coreference is. Essentially, one might say that the two are rather similar, as both are concerned with how certain constituents refer to each other, but there is a slight difference that distinguishes one from the other. In a study performed with entity coreference resolution as the task at hand, the coreference could be defined as the identification and connection of all mentions of entities in the text that refer to the same entity (Rao et al., 2010). This is slightly different to event coreference, which has more to do with the Quinean theory, proposed by Quine (1985), that two events are the same if they refer to the same object, “which is well defined in space and time” (Bejan, Harabagiu, 2010). Therefore, the main difference between event coreference and entity coreference is exactly what each refers to. Entity coreference is concerned with the same real world object, place or thing, whereas event coreference is concerned with the same occurrence or happening. This leads to the question: why is entity coreference relevant to this study, if the main focus is on a corpus that was specifically designed to be “event-centric” (Cybulska & Vossen, 2014) and a source for those interested in event coreference?

Let’s take, as another example, something that was provided by Cybulska and Vossen (2014). Although geared more towards event coreference, this example is also adequate to clearly explain the entities, and how those entities relate to the events.

(a) On Monday Linsay Lohan checked into rehab in Malibu, California after a car crash.

(b) Linsay Lohan checked into rehab.

(c) Ms. Lohan entered a rehab facility.

This is where entities and entity coreference become relevant to this dissertation, and to event coreference. As stated in their work, event coreference resolution is the aim of determining whether or not two event descriptions refer to the same event (Cybulska & Vossen, 2014). The event descriptions are another way to refer to the entities involved with the event, and are also referred to as “event mentions”. Therefore, through entities, the context and extra information behind determining the event coreference is provided. In the ECB+ corpus examined in this dissertation, any entity that refers to the seminal event is annotated in a certain way, and deemed then to be an event mention. If we break down Cybulska and Vossen’s example shown
above in part (a), and then compare it to the sentences (b) and (c) provided afterward, we can easily show how the entities (which we will refer to as event mentions from now on), play a role in event coreference resolution. The event mentions also cover parts of speech which aren’t defined as entities, for example time. Time is not an entity, as it is not a real world object. As the analysis of the corpus is one of the aims of this project, something which will be discussed extensively later on, it is best to avoid going into too much detail at present. However, some detail is required here in order to explain the relation between event and entity coreference in the example.

Each of the sentences above can be split into events and event mentions, depending on parts of speech. Focus on example (b) and (c) first. The seminal event in question is seen clearly in these parts, but no other context is given. It is clear here, that the action of checking into rehab is the event in question, and it involves at least one entity which is a human participant, that participant being Linsay Lohan. The other potential event mention is that of the rehab facility. However, as there is no extra context given, it is difficult to deduce whether or not from rehab in (b) is the same as rehab facility in (c). It is entirely probable that they may refer to two different rehab facilities in different locations, but it is also possible that rehab in (b) is referring to a state of being rather than a location of a facility. One more probability is vaguely apparent, but not clearly present, is the fact that the actress in question has been to rehab more than one time. Therefore, in the event that all the event mentions do indeed refer to the same facility, the event may still be in question, as they could refer to different times (Cybulska & Vossen, 2014). This is why (a) provides us with the clearest reason why entity coreference, and entities event mentions) in general, are helpful to those wishing to determine event coreference. The breakdown of the components in (a) is shown below in table 1. This is how Cybulska and Vossen (2014) explained how sentence components should be broken down in the ecb+ guidelines. While the issue of agreement will be under discussion and up for debate later on in this dissertation, it is difficult to argue with the breakdown of this particular sentence.

<table>
<thead>
<tr>
<th>action</th>
<th>checked into, crash</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>On Monday</td>
</tr>
<tr>
<td>location</td>
<td>rehab in Malibu, California</td>
</tr>
<tr>
<td>human participant</td>
<td>Linsay Lohan</td>
</tr>
<tr>
<td>non-human participant</td>
<td>car</td>
</tr>
</tbody>
</table>

Table 1: A breakdown of event components, taken from ecb+ (Cybulska & Vossen, 2014)

The action components correspond with the event in this case. The events that occurred in this example involve Linsay Lohan “checking into” rehab, and the car “crash”. The entities associated with this event include the location “rehab in Malibu, California” and both participants, the “car” and “Linsay Lohan” herself. The time component is not an entity in this
example, however it nevertheless falls into the category of event mentions, as it refers to the time in which the event “checked into” rehab takes place.

2.5. Previous Work

The corpus that will be evaluated in this report, as mentioned above, is the ecb+ corpus. Ecb+ is the abbreviation given for Eventcorefbank+. The ecb+ corpus is a version of the original ecb corpus which was extended by Cybulska and Vossen in 2014. Cybulska and Vossen (2014) explain in their research how they wished to create a corpus that was more representative of events, in order to lead to more accurate results on the determinance of event coreference. Coining the term “event-centric”, they extended the ecb corpus and, with the help of annotators, modified the corpus to be more about the seminal event in question. As mentioned above, anything that directly refers to the event was annotated, defined as an event mention, and was pivotal in providing a degree of context so that coreference between events could be more easily determined.

Before the ecb+ is discussed, it would be useful to take a look at some of the previous work in event and entity coreference resolution that was done in relation to the corpus in its earlier form. The original ecb corpus can be traced back to 2008, in a paper called “A Linguistic Resource for Discovering Event Structures and Resolving Event Coreference” (Bejan & Harabagiu, 2008). In this paper, their entire process for annotating the ecb in its original form is described (Bejan & Harabagiu, 2010). The original ecb was put to use first in the paper by Bejan and Harabagiu (2010). Their approach involved applying a new class of non-parametric Bayesian models to an open domain task of event coreference, with respect to how effectively it can be performed (Bejan & Harabagiu, 2010). Their design approach was based on the potential for nearly infinite features and forms and categorical outcomes, and as such they designed the project to cluster these items together in their appropriate fields. Their approach is also congruent to the Quinean theory that “two event mentions are coreferential if they have the same event properties and share the same event participants” (Bejan & Harabagiu, 2010). The ecb was used in their experiments for this task. They designed the ecb to increase event type diversity, in part due to what they considered the ACE corpus’s limitation in solving event coreference as its event types are restricted, and as it is suitable only for within-document event coreference resolution. They also wanted to be able to use the ecb as a source that was useful for evaluating their models for cross document and within document coreference resolution, something that was under examination in this task (Bejan & Harabagiu, 2010). Using the ecb to determine error analysis, they were able to compute the coreference between
events with respect to the event mentions and their referents contained within the corpus. In conclusion, their study produced a number of non-parametric Bayesian models which were designed for tedious problems involving the clustering of similar components, such as seminal event document clusters, and the items of the same characteristics. Their results were valuable to event coreference, as they proved that their work was adequate for the task of solving real data applications in which the features are considered to be free parameters (Bejan & Harabagiu, 2010).

Event coreference was looked at again in 2012 with relation to the ecb corpus (Lee et al., 2012). Lee et al. wanted to analyse entities and events jointly, and thus introduced a system to do so. Their system produced clusters of event and entity mentions and used linear regression to represent the clusters. Their system allowed for meaningful information to flow between entities and events, and semantic role dependencies were applied to the clusters as this information flowed (Lee et al., 2012). They designed their system to handle noun events and verbal events, as well as entities within documents. Similar to what will be seen later in the ecb+, the events and entities within the ecb work both ways to help the determinance of event coreference and entity coreference i.e. the events provide context for the entities and vice versa. Theoretically, this allows for a more accurate result which will better represent the coreference that is in question. This study used a training set and test data set from documents which were derived from the ecb, but not the ecb itself. They were interested in the seminal event structure of the ecb, and wanted to expand on the work performed by Bejan and Harabagiu. Bejan and Harabagiu used the ecb to determine event coreference, but Lee et al., had goals of seeing how it could work with relation to entity coreference as well as event coreference. The original ecb corpus was extended for this purpose, including “fully annotated sentences” and newly annotated “entity coreference relations” (Lee et al., 2012). In order to make the corpus more specific to event and entity coreference, any relations other than coreference relations were removed entirely. Using five coreference evaluation metrics, the results of this experiment showed that the model proposed by Lee et al. performed better, “almost unanimously”, than those it compared to. They also concluded that their model allows for events to help entity coreference and entities to help event coreference, providing a system and a source that would be useful to the field of coreference resolution.

Table 2 shows the statistics of the ecb corpus before it was extended to the ecb+ by Cybulska and Vossen (2014).
<table>
<thead>
<tr>
<th>Corpus Components</th>
<th>Number total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topics</td>
<td>43</td>
</tr>
<tr>
<td>Documents</td>
<td>482</td>
</tr>
<tr>
<td>Entities</td>
<td>1068</td>
</tr>
<tr>
<td>Entity Mentions</td>
<td>5447</td>
</tr>
<tr>
<td>Events</td>
<td>774</td>
</tr>
<tr>
<td>Event Mentions</td>
<td>2533</td>
</tr>
</tbody>
</table>

*Table 2: Eventcorefbank statistics (Lee et al., 2012)*.

2.6. Conclusion

Events and entities have now been discussed in detail. It is easy to see understand the definitions behind them both, as long as one has a certain understanding of the grammar. It is interesting to note that it may be difficult to efficiently represent event coreference without the involvement of entities, and sometimes vice versa. This has been shown in the work by Lee et al., as well as its potential for even greater results. Although hidden in most works relating to event coreference, entities and other event mentions still play a big part in providing the context needed to correctly identify events and the events that they corefer with. The next chapter will take us to the breakdown of the ecb+ corpus. Through the analysis of the corpus, the goal of this study will be to determine whether or not the annotation guidelines provided by Cybulska and Vossen (2014) for the ecb+ accurately depict an "event-centric" representation of a corpus, and through this it should also become clear as to why the entities and entity coreference is relevant. Practically speaking, the corpus will be broken down to its file level and the structure contained within, written in XML, will help to provide a realisation of the guidelines in practice.
Chapter 3

The Eventcorefbank+ Corpus

3.1. Introduction

A discussion was provided in chapter two which detailed accounts of events and entities and their coreference. An explanation was also provided which discussed the predecessor versions of the ecb+ corpus and how they were used in event coreference studies. These discussions lead now to the corpus which is under evaluation in this dissertation. The ecb+ corpus was extended by Cybulska and Vossen (2014) in order to provide a more “event-centric” source for research in the area of event coreference. The questions asked by this dissertation are as follows: does this corpus improve on the prior versions? Is it truly representative of being “event-centric”? How exactly are the events and event mentions annotated? Finally, does this study agree with how the corpus was annotated, and what does this tell us about the potential of the corpus for future studies?

In order to answer these questions, a complete breakdown of the ecb+ corpus is required. In this chapter, the background and granularity of the corpus will be provided. The annotation guidelines of the ecb+ corpus (Cybulska & Vossen, 2014) will be analysed and discussed, and examples of the annotation process and the annotation tags will be explained in full. An effort will also be made to breakdown one of the files of the ecb+ corpus in detail, including any other annotation tools that were used in the process of re-annotation. By breaking down the files, which are written in XML, within the corpus, the relations that are defined on this level can be fully fleshed out and understood. Following this long process of explanation in the first half of the chapter, the annotations of the corpus and agreement with the annotators will be questioned. As well as looking at a number of fully annotated files by the annotators, this study will attempt to perform a blind annotation of another group of files within the corpus, following the ecb+ guidelines. The goal here will be to determine if there is agreement between the annotations performed by this study and those performed by the annotators. Examples will be provided to compare and contrast the finished annotations provided by both parties. Will there be agreement? Will this study disagree with the annotators? Will the results be mixed, varied or ambiguous? What will this tell us about the corpus?
3.2. Annotation Guidelines Breakdown

The first step in fully understanding this corpus is to provide an overview of its background and its features. The corpus has been described in chapter 2 under the section covering previous work, but this section will build on what information has already been provided. As stated above, it must be pointed out that the ecb+ corpus is very new, and as such has not been used as a source in any works published to determine event coreference as of this date. This is one reason why an extensive analysis of the corpus is interesting to this study, an examination of its potential before it was used by any of the evaluation metrics to determine event coreference. It was extended by Cybulska and Vossen (2014) so that it would be able to represent events more efficiently. The goal of this was in the annotation guidelines which deemed that the newly released corpus should be tagged in a way which made it “event-centric”, a corpus of events and event mentions which related exclusively to the events which were up for potential coreference (Cybulska & Vossen, 2014). Before the annotation guidelines are discussed, the granularity of the corpus will be explained.

In the original ecb corpus, there were 482 documents with a total of 43 topics, however it is noted in that two texts were not included in the downloaded ecb corpus, making the true number of documents 480 (Bejan & Harabagiu, 2010). These topics were separated by folder, and in each of these folders “a topic is described as a seminal event” (Lee et al., 2012). The topics were taken from google news.

When the ecb corpus was extended to the ecb+, the number of documents was increased to 982 due to the inclusion and annotation of new 502 new corpus components (Cybulska & Vossen, 2014). In an attempt to make the ecb+ more representative of news streams in larger volumes, the corpus was also augmented to include different instances of event types in the corpus. For example, in the first topic of the ecb, the seminal event described is that of Tara Reid entering rehab in 2008. The augmentation to include similar event types which were not quite coreferent, but contained coreferent event mentions is shown in the addition of a second seminal event in topic one, that of Linsay Lohan entering rehab in 2013 (Cybulska & Vossen, 2014). The entry of more than one seminal event was done to each of the 43 folders in the corpus, so that each folder contained two seminal events that were closely related, but not quite event coreferent. This meant that there were 43 seminal events derived from the original ecb, and 43 more events entered in the ecb+. The action that takes place in each of the seminal events is coreferent as shown above (involving an entity checking into rehab at a certain time). The entities are not coreferent, therefore it is determined that the events are not coreferent. It should be noted here that folder 15 and 17 were missing from the downloaded ecb+ corpus, however the number of topics remains 43 with the inclusion of folders 44 and 45. Table 3 is shown below in order to give a sample overview of the comparison of topics and the different seminal events in each folder between topic 18 and 24. This table
shows the similarity of the events and the influence the entities and event mentions have to determine that the two seminal events in each folder are not coreferent.

<table>
<thead>
<tr>
<th>Topic Number</th>
<th>ecb seminal event</th>
<th>ecb+ seminal event</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Deadly office shooting Vancouver 2008</td>
<td>deadly office shooting Michigan 2007</td>
</tr>
<tr>
<td>19</td>
<td>Riots in Greece over teenagers death 2008</td>
<td>riots in Brooklyn over teenagers death 2013</td>
</tr>
<tr>
<td>20</td>
<td>Qeshm island earthquake 2008</td>
<td>Qeshm island earthquake 2005</td>
</tr>
<tr>
<td>21</td>
<td>Bloomington hit and run 2008</td>
<td>Queens hit and run 2013</td>
</tr>
<tr>
<td>22</td>
<td>S.D. Crawford Smith accused of killing co-workers Staunton 2008</td>
<td>Y. Hiller accused of killing co-workers Philly 2010</td>
</tr>
<tr>
<td>23</td>
<td>M. Vinar dies in a climbing accident on Mount Cook 2008</td>
<td>R. Buckley, D. Rait die in climbing accidents on Mount Cook 2013</td>
</tr>
<tr>
<td>24</td>
<td>4 robbers in drag steal jewellery in Paris 2008</td>
<td>4 robbers steal jewellery in Paris 2013</td>
</tr>
</tbody>
</table>

*Table 3: An overview of the different seminal events in each ecb+ topic folder (Cybulska & Vossen, 2014).*

In order to fully comprehend the granularity of the corpus, it might also be prudent to provide a table which explains its statistics after re-annotation. In order to do this properly, however, the annotation process behind the corpus must be explained in detail. In chapter 2, the annotation components were mentioned briefly in the example about Linsay Lohan checking into a rehab facility in Malibu. The annotations are far more complex than the singular tags given in this example. In the annotation guidelines for the ecb+ (Cybulska & Vossen, 2014), the components annotated are largely based on the prior annotation specifications. For example, the action classes in the ecb+, those referring largely to verbs and some noun phrases, are based on the TimeML specifications which were defined in the TimeML Annotation Guidelines 1.2.1 (Sauri et al., 2006). Others are based on various entity and event annotation specifications, all of which will be covered as the explanation of annotation is provided. As they have already been mentioned, and because it makes sense to start the definitions where events are mostly associated, the starting point will be the action classes. For a number of the tags, example words are given to illustrate what should be annotated a certain way. Sometimes, these words are given without sentences. This dissertation will provide example sentences for words in each of the classes in order to make sure there is clarity.
From the analysis done by this study, and by simply looking at the ecb+ annotation guidelines, the action component refers to the verb or noun phrase that most adequately describes the event. The action components are split according to the TimeML Specifications (Sauri et al., 2006) into five action classes: OCCURRENCE, PERCEPTION, ASPECTUAL, REPORTING and STATE (Pustejovsky et al., 2003, cited Cybulska & Vossen, 2014). A further two classes are also specified, those of CAUSATIVE and GENERIC. These seven classes are provided to clarify the polarity of events, giving context and allowing users to understand whether or not an event happened (Cybulska & Vossen, 2014). For those versed in these class definitions, applying them to action mentions might seem a trivial matter that is easily understood. For those who are not familiar, these action classes will be described in detail now.

(a) ACTION_OCCURRENCE:

This tag, in the opinion of this study, is defined in a very similar way to that of an event. The ACTION_OCCURRENCE tag is suitably used for most news events, and perhaps the tag used to describe the seminal events in the topic folders of the ecb+. It is specifically used to describe “something that happens or occurs in the world”, and applies to the most immediate verbs that describe an event. Words given as examples by Cybulaska and Vossen (2014) include verbs such as crash, build, die, and nouns like eruption and explosion. Nominalisation of verbs and vice versa is also allowed here. The tags are highlighted in bold.

→ There was a car crash on 5th avenue this morning.
→ I crashed the car earlier on my way to work.
→ The library was built on the outskirts of the town rather than in the centre.
→ When the volcano erupted, the populace vacated the town.
→ The President of the United States of America plays golf in his free time.

(b) ACTION_PERCEPTION:

The perception tag in relation to actions is related to the occurrence tag, however it refers to words which describe the witnessing of an action occurrence rather than the occurrence itself. It covers words that make clear that an action has been perceived to have occurred, or words which describe any perception to any occurrence. Examples provided by Cybulaska and Vossen (2014) list words such as see, feel, glimpse, and would also extend to words such as overhear, smell and notice.
→ I saw the car crash on 5th avenue this morning.
→ Did you hear the explosion earlier?
→ I couldn’t help but overhear that the president plays golf here regularly.
→ It’s easy to notice the library outside the town.
→ I beheld the flames of the volcano from perilously close.

(c) ACTION_REPORTING:
This tag is concerned with parts of speech that tell of an occurrence. If someone or something states that something has happened, this tag is used to annotate the words that describe the telling, i.e. the narration of an event, reporting the news, telling a story, explaining how something happened etc. (Cybulska & Vossen, 2014).

→ The people were told of the exploding volcano.
→ You need to report this to the police.
→ The narrator described the car crash with an air of sorrow.
→ I witnessed a robbery this afternoon.
→ John says that the library is not in the town centre.

(d) ACTION_ASPECTUAL:
According to the TimeML specifications (Sauri et al., 2006), the aspectual class refers to five facets of event history which are related to an event (cited Cybulska & Vossen, 2014). These facets are defined as “Initiation, reinitiation, termination, culmination and continuation”. Simply put, the aspectual tag in the action class is used to highlight words which signify the change in state of events or situations.

→ She ended our relationship when I crashed her car on 5th avenue.
→ The volcano eruption began only at 6 am and finished hours later.
→ The library doesn’t open on Sundays.
World War II ended after Japan's surrender.

The thief continued looting after the witness noticed him.

(e) ACTION_STATE:
In logic, one can describe a state as something which holds true. A simple proof can be set up in propositional calculus in which if X holds of Y, then Z holds true. A basic example in natural language could be defined as such: Premise X states that “if the heater is on, then the room is warm”. Premise Y states that “The heater is on”. Therefore, from these premises a conclusion can be deduced and assigned to Z which states that “the room is warm”. The tag in question here, ACTION_STATE, is similar but not limited to logic such as this. It applies to states of being, the state of the moment, how the state holds, and so on. Examples which could be annotated as such in the corpus include hope, being in love, whether or not someone is or is not a fireman, someone’s temperature, and how best one can describe the shape of a situation, whether amicable or explosive. It is usually “assigned to the non-verbal part of predicative phrases” (Cybulska & Vossen, 2014).

I was not too excited to be in the library.

My Mother was always caring.

The explosion was loud, and the lava flowed quicker than I would have expected.

She was angry when I crashed her car, and not so gently told me to pack my things.

The hungry thief vowed to stop when he had earned enough money.

(f) ACTION_CAUSATIVE:
This tag is used to annotate those phrases which describe the causing of something. As the tag clearly states, anything that is causative can be annotated with this tag. The field of words and phrases covered by this tag includes mentions like caused, led to, make, generate, produce, and bring about (Cybulska & Vossen, 2014).

I made less money this year.

After he crashed his girlfriend’s car, her reaction led to their separation.

The government brought about changes that led to a decline in support.
When it **came to light** that the president plays golf in this facility, publicity for the company was **generated** and membership applications poured in.

**Mixing red and blue** **produces** a purple shade.

**(g) ACTION GENERIC:**

If any event in the corpus is not anchored in space or time, then it is annotated with the generic tag (Cybulska & Vossen, 2014). A generic event is therefore something that is entirely unspecific, and cannot be specifically defined. These actions need to be tagged, nevertheless. According to the TimeML specifications (Sauri et al., 2006, cited Cybulska & Vossen, 2014), words and phrases which emerge as generic action mentions include use (of), travel, emerging and keeping track.

- **Are you keeping track** of the score?
- **After the car crash,** Ethan **emerged** unscathed.
- John **made good use** of the library’s faculties.
- **Should we make our way** to the market?
- Religious **travelling** is not my cup of tea.

These seven classes can also be expressed a negative annotation tag. In languages, negation can be expressed in a number of ways. In morphology, negative prefixes can be used to denote negation in action class words and phrases as shown above. For example, if someone describes a state of safeness, the negation of such a state would require the negative prefix “un-” to make unsafeness (Schmid, 2011). The same applies to negative particles such as not and neither, or to verbs like deny and avoid, and also if the participants of a sentence are negated, “no soldier went home” (Cybulska & Vossen, 2014). As such, any action tagged word or phrase which has been discussed above in the seven classes must be annotated with a negative tag if it is preceded by negation. As this study feels that a sufficient explanation is not provided in the guidelines, a number of contrasting examples are shown below to reinforce the tagging process of negation. In each of the examples, the (i) repeats one of the sentences from above in each of the seven classes, and (ii) provides the negation which determines the negative tag. It must also be noted here that it is not the negative particle or verb etc. that is annotated with the negative tag. The word that would be annotated with the ACTION_OCCURRENCE tag, for example, would be annotated with the NEG_ACTION_OCCURRENCE tag if and only if there is negation of the action.
(a) ACTION_OCCURRENCE VS NEG_ACTION_OCCURRENCE:

(i) There was a car crash on 5th avenue this morning.

(ii) There was not a car crash on 5th avenue this morning.

(b) ACTION_PERCEPTION VS NEG_ACTION_PERCEPTION:

(i) I saw the car accident on 5th avenue this morning.

(ii) I didn't see the car accident on 5th avenue this morning.

(c) ACTION_REPORTING VS NEG_ACTION_REPORTING:

(i) You need to report this to the police.

(ii) You don't need to report this to the police.

(d) ACTION_ASPECTUAL VS NEG_ACTION_ASPECTUAL:

(i) The thief continued looting after the witness noticed him.

(ii) The thief discontinued looting after the witness noticed him.

(e) ACTION_STATE VS NEG_ACTION_STATE:

(i) My Mother was always caring.

(ii) My Mother was always uncaring.

(f) ACTION_CAUSATIVE VS NEG_ACTION_CAUSATIVE:

(i) The government brought about changes that led to a decline in support.

(ii) The government brought about no changes that led to a decline in support. (Negation here on “brought about”, no negation on “led to”, thus “led to would be tagged normally as ACTION_CAUSATIVE).
(g) ACTION_GENERIC VS NEG_ACTION_GENERIC:

(i) John made good use of the library’s faculties.

(ii) John made no good use of the library’s faculties.

The ecb+ guidelines explain the annotation process of time components in a very clear manner. The examples provided are trivial and very easy to understand. This dissertation provided example sentences for the action classes above as they were largely absent from the guidelines. The parts of speech that are annotated with time tags come with their own examples, however they are not included in sentences. It is understandable, due to the triviality in determining whether or not a phrase refers to a time. This study will cover the time tags nevertheless, and place the phrases in example sentences. The tags defined here are also taken from the TimeML specifications, and have four major annotation types including duration, date, time and set (Pustejovsky et al., 2003).

(a) TIME_DATE: This tag is very self-explanatory. It refers to a space in time that is specific to a date that is covered in by calendar time (Cybulska & Vossen, 2014). The most common example might be someone’s date of births, an abstract term denoting a number of months or a season, or a particular week or day.

→ It was my birthday yesterday.
→ Japan surrendered on the 15th of August, 1945, which was a Wednesday.
→ Fiona is travelling to Budapest in April.
→ Between this week and last week I have been so busy.
→ The weather is lovely during spring.

(b) TIME_OF_THE_DAY:

Again, the tag suggests that annotation of phrases with this tag is also trivial. This tag corresponds to the guidelines laid out in the TimeML specifications (Pustejovsky et al., 2003, cited Vossen & Cybulksa, 2014), and is very specific to components which refer to a particular time during a one day period. This of course encompasses expressions of hours, minutes and seconds. Interestingly, this tag can also extend to components which would usually be covered in the TIME_DATE tag above, but with the inclusion of minutes, hours or seconds, even if it refers to a specific calendar date, it would be annotated with the TIME_OF_THE_DAY tag instead. However, the time must be specific, i.e. not durational (which will be discussed next).
Class ends in five minutes.
I trimmed the hedges this morning at 10 am.
If John doesn’t finish his homework by half past 6, he can forget about the football.
The mass murderer was pronounced dead at 27 seconds past 5 pm.
We had a late evening lunch.

(c) TIME_DURATION:
Any time components which represent a time period that can be measured between one point and another can be annotated with the TIME_DURATION tag. This tag is used for less specific time points, but can be utilised for duration in time that can be measured from seconds onwards.

Class only lasted five minutes today.
I trimmed the hedges this morning between 10 am and 12 pm.
We have two weeks until our lease is up.
The convicted murderer spent 7 years on death row.
I spent yesterday evening in the chapel.

(d) TIME_REPETITION:
In the TimeML specifications, the SET tag is defined as a set of times, or a repetition of times (Sauri et al., 2006). In other words, this tag is used in conjunction with words which explain that something occurs more than once.

I regularly visit my Grandfather in the home.
It is important to brush your teeth twice a day.
Every Wednesday I have a scheduled meeting with my supervisor.
Monthly payments are mandatory for club members.
The locals based their religion on the *periodical* movements of the skies.

Location components are explained next in Cybulska and Vossen’s ecb+ guidelines (2014). In the opinion of this study, this component is also discussed very well in the guidelines, however it is not as trivial as the time component. Again, there are a number of examples given which would be deduced as location components, but a number of the examples will be extended upon again here, in order to provide greater clarity. Anything in the corpus which pertains to a location, whether a specific geographical location, a vaguely abstract location, or simply a building with no further context, must be annotated with a location tag. These tags are defined based on the ACE’s PLACE attribute, which refers to a physical location (LDC, 2008, cited Cybulska & Vossen, 2014). There are three separate location tags to differentiate between.

(a) *LOC_GEO*:

This tag is perhaps the easiest to define. It refers to a physical location which has an actual geographical location in the world (or elsewhere), as long as that location is exactly specific. A way to describe this is to imagine a map. If you can point out a place on the map, whether to a field, a building, or a small lake, you would be pointing at something that would be annotated as a geographical location. For example, a greatly specific location such as “the headquarters of the CIA in Langley, Virginia”, or something discrete and yet specific such as “the tree on the hill near Scotshouse village” would both be tagged as such. Also described here is the example “around the world”. While not specific, this phrase might stand for a number of geological locations. Whether or not this study agrees with this phrase being tagged as such is relevant, but it will be discussed later at the end of this chapter. This tag corresponds to both ACE’s GPE, which covers geo-political entities, and ACE’s LOC, which covers specific locations (Cybulska & Vossen, 2014).

The war in Iraq claimed many lives.

We travelled across America.

The plane crash landed on the river Danube.

Many people visit Trinity College in Dublin during the year.

The oval office of the white house in Washington D.C. is where the president spends a great portion of his time.
(b) LOC_FAC:

This particular tag is less specific. It has less to do with an actual geographical location, and more to do with an entity which represents a building or structure in which an event occurred (Cybulski & Vossen, 2014). Oftentimes it may seem like facilities are covered in the previous tag (for example, the white house is a facility), however those locations that are not geographically specific should be covered by this tag. If a person tells another to meet them at their house, but no address is given, the TIME_FAC tag would be an appropriate annotation. Alternately, with an address provided, geographically cementing the entity that is the house, the LOC_GEO tag would be applied instead.

→ The crowds exceeded the number of people in a shopping mall that I was used to.
→ He grew up on the streets and based his music on those times.
→ She checked into rehab last night.
→ I don’t go to church very often.
→ My brother’s friend has it hard, and sleeps in the shed these days.

(c) LOC_OTHER:

For any remaining type of location entity or event mention in a text that is not covered by LOC_FAC or LOC_GEO, the LOC_OTHER tag is applied. This is essentially the “remainder” tag for those parts of speech which would be left un-annotated if there were only two location classes. A number of location entities that this tag can be applied to include locations that are generally arbitrary, such as “on the table”, or “at the window” etc. This study concludes that it can refer to specific locations within unspecific locations.

→ I dropped the book on the table for you.
→ Darth Vader paused at the window to observe the chaos.
→ Max lobbed the ball at the hoop.
→ The badly behaved child was told to sit in the corner.
→ Sara lost her bag under the benches.

Human participant components are discussed next in the guidelines (2014). These participants range to cover all human type entities, from singular persons to groups run by persons.
(Cybulska & Vossen, 2014). Cybulska and Vossen based their human component annotation guide on previous work shown in the ACE participant tags PER and ORG when annotating persons and person organisations, and the ACE participant tags GPE, FAC and VEH to cover human participants such as geo-political entities, human run facilities, populations and governments (LDC, 2005, 2008, cited Cybulska & Vossen, 2014). Aside from the action components, human participants are perhaps the most common components in the ecb+ corpus. The events reported in each folder are often related to at least one, if not more than one, human participant.

(a) **HUMAN_PART_PER:**

This tag refers to the human entity or event mention that most closely resembles a single human or a group of humans. It can include a person’s name, their honorific or title, a familial grouping, and personal pronouns that relate back to either. Cybulska and Vossen describe this tag simply as a reference to individuals, or a group of individuals (2014). It should be noted that certain components take precedence over others. To bring the discussion back briefly, if there is a text about the production manager in a factory being angry, then “production manager” is the human participant. If, however, the text states explicitly that “the production manager, Jerry, is angry”, then “production manager” is no longer the human participant. Instead, that role would be applied to “Jerry”. This is, again, as the tag in question must be applied to that component which stands irrevocably for the human participant.

→ *The gardener watered the flowers.*
→ *The gardener, Sean, watered the flowers.*
→ *The actress was sentence to community service.*
→ *My family will not be joining us this evening.*
→ *Although Mary was disappointed, she didn’t let the grief get to her.*

(b) **HUMAN_PART_ORG:**

This tag initially seemed ambiguously defined when paired against the previous tag. This ambiguity will be discussed later in the agreement section. Any human organisation entities can be annotated with this tag, according to the ecb+ guidelines. This includes, but is not limited to, “corporations, agencies and other groups of people defined by an established organisational structure” (Cybulska & Vossen, 2014). The tag in question is perhaps less arbitrary than the grouping of individuals in the previous tag, and more specific to an organised structure of some calibre.
→ The Military has a concentrated presence in the western region.
→ Sometimes the media tell people what they want to hear.
→ The press were denied entry.
→ McDonalds are opening a new branch downtown.
→ Representatives of rival software companies Microsoft and Sony were both present today to talk about the upcoming consoles.

(c) HUMAN_PART_GPE:
Geo-political entities are defined by this tag, as well as any social groups that refer to a government or a settlement/city (Cybulska & Vossen, 2014). Countries, cities, towns, occupied lands, and anything defined by borders which refer to their inhabitants are included in this list, but only if they refer to inhabitants.

→ Russia and the United States held talks today in D.C.s capitol building.
→ Greenbay lost to the Vikings today.
→ Liverpool were disappointed today after their loss in Old Trafford.
→ Hollywood is preparing itself for the award show later this week.
→ Many people are expected at the sermon on Easter Friday, the Vatican told sources.
(d) **HUMAN\_PART\_FAC:**

The human participant facility tag refers to event mentions of facilities (Cybulska & Vossen, 2014), however there is a difference between this tag and the LOC\_FAC tag discussed earlier. The LOC\_FAC tag refers to the entity itself, that of a facility of some sort that is located in an unspecific place. The HUMAN\_PART\_FAC tag refers to this entity by referring to the people who are using the facilities. Therefore, when something is annotated with this tag it describes the facility not as a location, but as a group of people who represent the facility. For example, “The school is in Belfast” describes a facility location in Belfast, whereas “the school says it is in Belfast” refers to the people representing the school, who have declared on behalf of the school, where they are situated.

→ The School pushed for stricter rules and regulations following the incident.
→ Langley is withdrawing its support for the drone program.
→ Asda decided to lower their prices when they noticed the price drops in Lidl.
→ The prices were lowered in Asda after Lidl said they would be dropping theirs.
→ Washington hopes to achieve more in its efforts to put people to work in 2016.

(e) **HUMAN\_PART\_VEH:**

Taken from the ACE tag VEH when referring to a “population or a government usually occurring with geo adjectives”, this tag is used to refer to vehicle entities under control by said geo-political factions (LDC, 2005, 2008, cited Cybulska & Vossen, 2014). Anything that can be defined as a mode of transport that can bring people from point A to point B is annotated with this tag, as long as it is used in reference to the people behind a geo-political entity.

→ German tanks moved into the region to provide much needed support.
→ British ships may be patrolling the length of the British channel.
→ US fighter jets received orders to remove insurgents at the following coordinates.
→ Government issued sedans showed up to escort the ambassador to the airport.
→ The Hells Angels’ bikes caused quite a stir as they roared up Main Street.

(f) **HUMAN\_PART\_MET:**
This tag is used for any metonymically expressed human participant mentions that are left over. In other words, if a human participant event mention is referred to through some medium that has not been covered by the human participant tags above, then it will likely fall into this category.

→ The recruit swore loyalty to the crown when he signed up with the British Forces.

→ “I pledge allegiance to the flag…”

→ At least 5 homes are reporting the same problem.

→ Our eyes in the field are telling us that we should move, now.

→ I have a hand who helps with all the chores.

(g) HUMAN_PART_GENERIC:

This tag refers to generic event mentions of human participants. It is usually applied to everything else that is not covered by the other human participant component tags, in other words, not to any particular individual or group of individuals. The generic tag is the opposite of everything else, which is specific, and it refers to those human participants that cannot be referred to in any other way than generically (LDC, 2008, cited Cybulska & Vossen, 2014).

→ One wishes it were so.

→ Have you met up with the others yet?

→ You wouldn’t wish such a fate on anyone.

→ “Oh my god! They killed Kenny!” (When Kenny’s death is no one’s fault)

→ All who stand in my way will pay.

The final component part that is explained in the ecb+ annotation guidelines is that of non-human participant event mentions. The previous component covers a vast amount of entities pertaining to humans. This component is designed to cover every other participant that is left over i.e. any participants which are not human participants, locations or time event mentions must be tagged as non-human participants (Cybulska & Vossen, 2014). Intuitively, the entities which come to mind are likely nouns which are not action annotated. According to the ecb+ guidelines, they “will often be artefacts expressed as a (direct or prepositional) object of a
sentence or as “prepositional phrases such as instrument phrases and others which are not in the position of the object (2014).

(a) NON_HUMAN_PART:
All event mentions which are not annotated under any other tags are to be labelled as such.

- Look at the coat.
- Sharpen a pencil with the knife (Cybulska & Vossen, 2014).
- I drove the car to the river.
- The footballer refused to sign the contract due to unmet demands.
- Toilets are gross.

(b) NON_HUMAN_PART_GENERIC:
Another generic tag is defined within the non-human participant section. Again, this refers to unspecific event mentions that have not been tagged under any other heading that has been discussed above, including the traditional NON_HUMAN_PART tag. The example given by Cybulska and Vossen in the guidelines (2014) shows that this tag points to certain entities or classes of entities “or their typical representative” without actually referring to whatever specific non-human entity or individual class object or objects that that might be (LDC, 2008, cited Cybulska & Vossen, 2014).

- Linda loves cats (Cybulska & Vossen, 2014).
- There are a few things lying around.
- I have stuff to take care of.
- Cows don’t eat beef.
- Michael is generally not very fond of hamsters, but he loves his hamster anyway.

This concludes the annotation guidelines for the ecb+. At first glance, it can be a lot to take in. At the same time, however, as the annotation guidelines are based on extensive study and work that came before, the specificity of the tagging system of this corpus makes incredible sense in theory. One of the questions of this study asks if it also makes sense in practice. The
expected answer, due to the work that led to this corpus, is of course yes. However, it will be discussed later. Before this can be done, an examination of the XML files contained within the corpus must be undertaken.

### 3.3. XML File Breakdown

In this section, we will attempt to break down the files contained within the ecb+ corpus. This will allow for a greater understanding of the annotation process in practice, as well as reinforcing the theory which was covered in the previous section. The events and event mentions of each text are written in XML. The breakdown of these files is not covered in the work by Cybulska and Vossen. It is of great interest to this study to therefore examine the files more closely, as the breakdown will help in the later discussion of agreement. XML is an abbreviation of Extensible Markup Language, and it was designed to be easily readable by both humans and machines. Used to represent character data, entities and parsed data, XML is primarily used to represent text files and to partially describe the function of programming languages. It is largely designed to be easy to create, easy to read, and usually leaves no room for ambiguity in its design, resulting in a solid representation of whatever document it describes. One of its design goals was compatibility with programming languages, allowing for the files to be easily read and parsed by programmers (Bray, et al., 2006).

This study will focus extensively on the breakdown of one XML file. The structure of the file will be analysed, and tokenization, token marking, and coreference relation definitions will be discussed. Afterwards, the part of speech tokens which were tagged according to the guidelines specified by Cybulska and Vossen (2014) will be analysed. The example this study will use is the very first file (sorted alphabetically) in topic 1. This document happens to also be the first file in the original ecb corpus rather than one of the newly added component parts of the ecb+. The seminal events in this folder are:

- T. Reid checks into rehab in 2008 (ecb seminal event)
- L. Lohan checks into rehab in 2013 (ecb+ seminal event)

As the file under examination involves the ecb seminal event, it is naturally about Tara Reid. As the XML files involve texts taken from news streams, the first step any human would take is reading the article. Ignoring factors such as syntax for now, the human mind (if the human understands English) can easily understand the context of a news article, and therefore deduce the seminal event from the entities and event mentions involved, simply by reading the words. The text in question is neither too long nor too short. For the purpose of this example, it is probably an appropriate size. The text is highlighted in bold below.
“Perennial party girl Tara Reid checked herself into promises treatment centre, her rep told people. “We appreciate your respect to her and her family's privacy at this time," the 33-year-old actress's rep Jack Ketsoyan told the magazine exclusively for their Web site. Ketsoyan would not say what the "American Pie" star was being treated for. A friend of the actress told people she went to Promises on Tuesday and that her friends and family supported her decision. Reid launched a new fashion line called Mantra in December.”

(Taken from ecb+ corpus, topic folder 1, 1_10ecb.xml, Cybulska & Vossen, 2014)

The full XML file can be found in the appendix (document A.1.). After reading this text, one might be able to think of the headline very easily. It is clear to this study at least, that the seminal event outlined in the corpus for ecb topic 1 holds true, and therefore the headline would indeed be something like “T. Reid checks into rehab”. The first step in the XML file is tokenization. Each word and punctuation mark in the text is assigned a token ID (t_id). Some parts of speech are tokenized together, for example if a single entity is represented by two words joined by a hyphen. Such an example is not seen in this file, however. As the sentences in the files are tokenized, they are also given a sentence number. Like in many programing languages, the starting index begins at 0 and is incremented after each sentence. The token IDs begin at 1, however. Table 4 demonstrates pictorially the tokenization performed on the first sentence (sentence = 0) in 1_10ecb.xml. See appendix for the entire XML file.

<table>
<thead>
<tr>
<th>t_id</th>
<th>token</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Perennial</td>
</tr>
<tr>
<td>2</td>
<td>party</td>
</tr>
<tr>
<td>3</td>
<td>girl</td>
</tr>
<tr>
<td>4</td>
<td>Tara</td>
</tr>
<tr>
<td>5</td>
<td>Reid</td>
</tr>
<tr>
<td>6</td>
<td>checked</td>
</tr>
<tr>
<td>7</td>
<td>herself</td>
</tr>
<tr>
<td>8</td>
<td>into</td>
</tr>
<tr>
<td>9</td>
<td>Promises</td>
</tr>
<tr>
<td>10</td>
<td>Treatment</td>
</tr>
<tr>
<td>11</td>
<td>Centre</td>
</tr>
<tr>
<td>12</td>
<td>,</td>
</tr>
<tr>
<td>13</td>
<td>her</td>
</tr>
<tr>
<td>14</td>
<td>rep</td>
</tr>
<tr>
<td>15</td>
<td>told</td>
</tr>
<tr>
<td>16</td>
<td>people</td>
</tr>
<tr>
<td>17</td>
<td>.</td>
</tr>
</tbody>
</table>

Table 4: Initial sentence tokenization in 1_10ecb.xml
The annotation tagging begins after the tokenization process. Any token that is tagged is referred to in the markables category of the XML file. In this category, the token ID is assigned a markable, an annotation tag based on the ecb+ guidelines which were discussed in the previous section of this report (i.e. ACTION_OCCURRENCE, HUMAN_PART_PER, etc.). The token ID is also assigned a markable ID (m_id) which refers to the annotation tag. If one does not have a copy of the ecb+ guidelines, but wishes to see which tokens have been tagged, using the file itself and the tags within can be a tedious process. Looking at the markables category, one must find the t_id. This refers them to the token. As each t_id has been assigned a markable, or annotation tag, one can see which tag has been assigned to the t_id, and from there, which token the markable defines. Without writing a simple program to extract the tokens and their corresponding ids and annotation tags, the process is arduous, and requires the reader to scroll down to markables and back up to tokens for each annotation they wish to find out. Although it is potentially trivial to understand the process, table 5 below extends table 4 to highlight the markable IDs assigned to the token IDs, and by extension the tokens themselves.

<table>
<thead>
<tr>
<th>t_id</th>
<th>token</th>
<th>m_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Perennial</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>party</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>girl</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tara</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>Reid</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>checked</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>herself</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>into</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>Promises</td>
<td>36</td>
</tr>
<tr>
<td>10</td>
<td>Treatment</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>Centre</td>
<td>36</td>
</tr>
<tr>
<td>12</td>
<td>,</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>her</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>rep</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>told</td>
<td>34</td>
</tr>
<tr>
<td>16</td>
<td>people</td>
<td>39</td>
</tr>
<tr>
<td>17</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

*Table 5: Markable IDs assigned to tokens and token IDs*

As was described in the previous section, event mentions and entities are annotated with tags most suitable to them, as decreed by the guidelines provided by Cybulska and Vossen (2014). In table 5, multiple m_ids can be applied to the same t_ids. The m_ids are not exclusive to any one token, although they can be. It depends on the part of speech that is being tagged, and the tag which is being assigned. For example, the tokens “Tara” and “Reid” are assigned t_ids 4 and 5
respectively. Both tokens are, however, assigned the same m_id, 20. This is because both tokens together refer to a single entity, a human participant. Therefore, according to the guidelines of the ecb+, and under the annotation provided by the annotators, “Tara Reid” is assigned the markable HUMAN_PART_PER, the m_id of which is 20. It must also be noted that corresponding tokens need not necessarily follow each other in chronological order to be annotated as the same event mention. The tokens 6 and 8 are assigned the m_id 32, and are therefore the same event mention. Table 6 below provides a clearer representation of the annotation tags assigned to the tokens in this sentence, and condenses the tagged tokens denoted as markables into one string.

<table>
<thead>
<tr>
<th>m_id</th>
<th>marked tokens</th>
<th>Annotation tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>checked into</td>
<td>ACTION_OCCURRENCE</td>
</tr>
<tr>
<td>34</td>
<td>told</td>
<td>ACTION_REPORTING</td>
</tr>
<tr>
<td>20</td>
<td>Tara Reid</td>
<td>HUMAN_PART_PER</td>
</tr>
<tr>
<td>21</td>
<td>herself</td>
<td>HUMAN_PART_PER</td>
</tr>
<tr>
<td>26</td>
<td>rep</td>
<td>HUMAN_PART_PER</td>
</tr>
<tr>
<td>36</td>
<td>Promises Treatment Centre</td>
<td>LOC_FAC</td>
</tr>
<tr>
<td>39</td>
<td>people</td>
<td>HUMAN_PART_ORG</td>
</tr>
</tbody>
</table>

Table 6: Markables, markable IDs and tokens marked.

There are further observations one could make to the data above. For example, there are two HUMAN_PART_PER components which refer to the same entity. In the sentence, “herself” refers back to “Tara Reid”, and despite being the same entity, they are assigned different m_ids. The reason behind this is simple. Although the same entity, they are in fact two separate event mentions. In order to be assigned the same m_id, they would have to go together as a singular mention of the event. As they are, however, the same entity, a coreference relation must be defined between them. This is also covered in the XML file under the markables category, and also under the relations category.

It should be noted here that a number of tools were used by the annotators in determining cross document and within document coreference relations (Cybulska & Vossen, 2014). One of the tools used was the Content Annotation Tool, or CAT (Bartalesi Lenzi et al., 2012). This tool was used to annotate mentions of actions, times, participants and locations within the text, and the annotators used it for within document coreference. If there are coreference relations within the same file, this tool was used to define those relations. The other tool which was utilised by the annotators was the CROMER tool (Bentivogli et al., 2008, Girardi et al., 2014), the Cross-document Main event and entity recognition tool. This tool was used to mark cross document relations. The cross document coreference relations exist between event mentions in files that fall within the boundaries of the same seminal event. In relation to table 5, cross document and within document coreference can be easily seen by looking at the cross document relations defined in the relations category, and then at the instances of coreference that are defined at the end of the markables category. To show one
example of this coreference, this study refers you to the relations category of 1_10ecb.xml in the appendix.

Directly before the closing bracket of the markables category, instance_id tags are applied to markables. These IDs are created by whoever uses the CROMER tool, in this case the annotators, and they represent instances of cross document coreference between event mentions (Cybulska & Vossen, 2014). Notice the HUMAN_PART_PER with an m_id of 43. The descriptor of this component is “t1b_tara_reid”, and is assigned an instance_id “HUM16236184328979740”. The m_id that this instance is tagged with is known as the target markable ID. The instance_id created here refers to a single event mention, the entity that is “Tara Reid”. In every other file that is bound by the seminal event “T. Reid checks into rehab in 2008”, any ecb component file in topic folder 1, this instance_id will always refer to the entity, Tara Reid. As such, cross document coreference is always marked for event mentions through their instance_id. The target m_id is not shared across documents, and is reserved for within document coreference.

In order to explain the within document coreference, this study again refers you to the Relations category in the 1_10ecb.xml file in the appendix. As was just discussed, the target markable ID (target m_id) of the event mention “Tara Reid” is 43. In the third cross document coreference relation under the relations category, the instance_id (or note, in this case) is “HUM16236184328979740”. As shown earlier, this also refers to Tara Reid. Within this category there are also source m_ids 21, 22, 23, and 20. These source m_ids are, in fact, all markable IDs which refer to a markable that refers to the event mention of “Tara Reid”. In the sentence this study analysed earlier, and shown in table 5, m_id 20 refers to the tokens that make the string “Tara Reid”, and m_id 21 refers to the token “herself”, which within the context of the sentence is a personal pronoun which refers to the same entity. Therefore, all source m_ids which have the same target m_id are coreferent within the document. Table 7 shown below shows the instances of cross document and within document coreference in topic 1 of the ecb+ corpus. The underlined source m_ids refer to markables that were covered in the example sentence. Those not underlined refer to the remainder of markables in the file which can be seen in the appendix.

<table>
<thead>
<tr>
<th>source m_id</th>
<th>target m_id</th>
<th>instance_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>41</td>
<td>ACT16235311629112331</td>
</tr>
<tr>
<td>36, 37</td>
<td>42</td>
<td>LOC16235213289813758</td>
</tr>
<tr>
<td>21, 22, 23, 20</td>
<td>43</td>
<td>HUM16236184328979740</td>
</tr>
<tr>
<td>39, 40</td>
<td>45</td>
<td>HUM16236907954762763</td>
</tr>
<tr>
<td>26</td>
<td>46</td>
<td>HUM16284637796168708</td>
</tr>
<tr>
<td>32, 33, 38</td>
<td>48</td>
<td>ACT16236402809085484</td>
</tr>
</tbody>
</table>
Table 7: Source m_ids that share a target m_id are coreferent within the document. The target m_id is exclusive to each document, and the instance_id is common across multiple documents which contain the same seminal event and event mentions, thus the instance_id highlights cross document coreference.

3.4. Conclusion

A discussion has now been provided of the granularity of the ecb+ corpus, as well as a detailed explanation of the annotation guidelines set up by Cybulska and Vossen (2014). This leads the study to the questions that were asked at the beginning. Is there agreement to be found between the annotators of the ecb+ and this study? Can we annotate the ecb+ corpus blindly, without the influence of the annotations that the annotators agreed upon? Do we agree with the annotation guidelines? What does this mean for the corpus? Did the “event-centric” notion hold up, and does it define the corpus as a potentially seminal source for future event coreference resolution? Finally, using Cohen’s Kappa statistic, what is the exact agreement that can be deduced between the annotations provided in 2014 vs the annotations provided by this study? These questions will be addressed in Chapter 5.
Chapter 4

The Experiments

4.1. Introduction

This chapter will detail the questions that this study has asked. Before the statistical results are revealed, details involving the experiment itself must be covered. A critical analysis will first be performed on the guidelines of the ecb+, in order to ascertain the opinion of this study on its usefulness for future work in event coreference resolution. Afterwards, testing will be performed on the annotated XML files. In these tests, this study will attempt to re-annotate a number of the documents in the corpus, following the guidelines set out by Cybulska and Vossen (2014). These new annotations will be compared to the annotations of the corpus, and agreement will be calculated using Cohen’s Kappa statistic (Cohen, 1960). This interrater agreement test will be performed on multiple levels of the corpus: firstly, a single document will be analysed and the Kappa statistic performed based on the annotations within that single file. Secondly, a group of documents will be analysed and re-annotated from across the entire corpus. The Kappa statistic will be performed again, but this time it will determine the level of agreement between the documents were there is agreement and the documents where there is not.

4.2. Experiments and Results

The first step in the experiment was begun after a wide understanding of the annotation guidelines had been obtained. With this knowledge, and with a copy of the guidelines nearby for reference, the re-annotation process could begin. A number of documents from the corpus were preselected and the texts within were re-annotated according to the ecb+ guidelines. This first test was done in order to determine our agreement with the annotators tagging of the event mentions within documents. A biased approach was avoided by not looking at the annotations that were defined by the annotators. The results that were obtained from this step were interesting, as they ranged from total agreement within the documents to broad disagreement. At times during the experiment, re-reading the guidelines after annotation, or simply comparing the results of our annotations to the annotations of the ecb+, triggers of doubt about what we had achieved went off. This often led to a change in opinion, as the comparison allowed opened us up to new perspectives that were not thought of initially. At other times, the comparison would lead to bafflement and total disagreement. More often than
not however, this study found the re-annotation results to be in agreement with the annotations of the ecb+.

As a first example, after re-annotating a file in topic 1 (the first folder), 1_8ecb.xml, we found that our result was almost 100% in agreement, but not quite. This file will be listed in the appendix (document A.2). The disagreement was in the following sentence:

“After telling people in October that she didn’t need to do any of that anymore...”

We had re-annotated the underlined word with the annotation tag ACTION_OCCURRENCE. Upon closer inspection, and by contrasting this annotation with the annotation provided in the file (see document A.2, m_id 10, t_id 17), our opinion changed and we retagged the word as an ACTION_REPORTING component. The first impression was that “telling” was an action taking place, and therefore was the occurrence of an action. However, “telling” is in fact an action that is reporting another action, the action which is held under the word “that”, which is an ACTION_OCCURRENCE component. After changing our re-annotation, we found ourselves 100% in agreement with the annotation of components in this file. Calculating the Kappa statistic on this file would be irrelevant and impractical, as a result of our agreement.

There are a number of other files we chose to analyse and annotate at the base level. Also from the first topic was the file 1_16ecbplus.xml. The text in the file was re-annotated and compared to the annotations in the ecb+ corpus. Our results were largely different than that of the previous test, and this is mainly due to disagreement on one word which was deemed to be an ACTION_OCCURRENCE component, and another component which was tagged as LOC_FAC which this study felt should have been annotated with the LOC_GEO tag. A sentence is shown below:

“Lindsay Lohan checks into rehab in Newport Beach”

The annotators of the ecb+ concluded that “rehab” in the sentence above is an event mention that is an action occurrence component. This contradicts the example we provided at the beginning of this report, in which “checking into” was the action that was occurring. In this example, “check into” is also tagged as an ACTION_OCCURRENCE component, which this study agrees with. Following this logically, should a deduction not be made that “checks into rehab” is a single event mention of type ACTION_OCCURRENCE? According to the file, both “checks into” and “rehab” are action occurrence components. However, they are annotated separately, and are therefore two separate event mentions. If “rehab” was a word that was used to signify a
state of action, this dissertation would be inclined to agree. However, it is also the opinion of this study that it is most grammatical if “rehab” was annotated as a location facility in conjunction with Newport Beach. This leads us, however, to our second disagreement.

“In Newport Beach” was annotated in the ecb+ as LOC_FAC. This study immediately disagreed. According to the guidelines, any location that can be defined as being part of a geographical location should be annotated with the LOC_GEO tag. One of the examples provided in the guidelines is “A 7.2 magnitude earthquake hit in southern California this afternoon”. Southern California is, of course, a geographical location. Using a replacement test, and due to the fact that Newport Beach is also a geographical location as a city in Orange County, wouldn’t the sentence “A 7.2 magnitude earthquake hit in Newport Beach this afternoon” also make sense? For this reason, this study disputes the annotation of “In Newport Beach” in the ecb+ as a LOC_FAC component, and holds fast to the re-annotation of this component as LOC_GEO. A Kappa statistic was performed on this file as a result of this dispute, and a further number of disputes later. There is a slight problem however when calculating the Kappa statistic. As the annotations have already been finalised and agreed upon by the annotators of the ecb+, performing Cohen’s Kappa statistic will always result in the agreement rating of 0. This is due to the fact that it would never make sense to say that the annotators disagree with data that they had finalised and settled on. They will always say yes. There are, of course, instances where this study will disagree with the annotators, as has been shown above. The Kappa statistic would normally be calculated in the following way:

\[ K = \frac{(P(A) - P(E))}{(1 - P(E))} \]

K stands for the Kappa coefficient (Carletta, 1996), and ranges in value between 0 and 1. A Kappa coefficient of 1 signifies complete agreement between raters, and 0 signifies that there is no agreement between the raters, other than what would be expected by chance. P(A) is the coefficient which represents the proportion of times that the coders agree, and P(E) represents the proportion of times that it would be expected for coders to agree by chance (Carletta, 1996). As the Kappa statistic will always be 0, as the annotators will always say yes, and never no, we are more concerned in representing the P(A) value, the proportion of times we agree with the annotations given in the ecb+. In file 1_16ecbplus.xml there were 26 marked event mentions or components. Of these 26, this study agreed 22 times. Our P(A) for the marking of event mentions in this file is therefore approximately 84.5%. In spite of the disagreements that we made above, this is still a substantial agreement value of the event mention annotations within this text. Table 8 shows some event mentions from 1_16ecbplus.xml and how they were annotated by the annotators of the ecb+, and how this study re-annotated the event mentions.
Another example of this sort was carried out on file 38_9ecbplus.xml. This is in topic folder 38, of which the seminal event is “Small earthquake in Sonoma County 2008” (Cybulska & Vossen, 2013). Of the 10 components annotated in this file, this study agreed 10 times, resulting in a P(A) value of 100%. In the 23rd topic folder, a partial analysis of file 23_3ecb.xml was performed. The initial sentence was re-annotated. There were 6 annotated components in this file, and after re-annotation this study agreed with 5 out of 6, resulting in a P(A) coefficient value of 83%. This process was performed on 30 documents contained within the ecb+ corpus from most of the topic folders. Although there was not 100% agreement between the re-annotation of event mentions in this study and the annotations of the ecb+, the P(A) score was significantly high in every test. None were lower than 75%, and more than most were scored at 100% or slightly less. Table 9 gives a detailed account of the agreeability, represented by the P(A) coefficient rounded off to the nearest percentage, that was determined between this dissertation and the annotators, with reference to the annotations of the event mentions in individual documents. The reasons why this dissertation may have disagreed in certain areas will be explained in the error analysis section.

<table>
<thead>
<tr>
<th>Event Mention Component</th>
<th>ecb+ annotation</th>
<th>Re-annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 2, 2013 1:35pm</td>
<td>TIME_OF_THE_DAY</td>
<td>TIME_OF_THE_DAY</td>
</tr>
<tr>
<td>in Newport Beach</td>
<td>LOC_FAC</td>
<td>LOC_GEO</td>
</tr>
<tr>
<td>rehab in a New York facility</td>
<td>LOC_FAC</td>
<td>LOC_FAC</td>
</tr>
<tr>
<td>L.A.</td>
<td>LOC_GEO</td>
<td>LOC_GEO</td>
</tr>
<tr>
<td>prosecutors</td>
<td>HUMAN_PART_PER</td>
<td>HUMAN_PART_PER</td>
</tr>
</tbody>
</table>

Table 8: Contrast shown in re-annotation

<table>
<thead>
<tr>
<th>Document</th>
<th>P(A) / Agreement Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_18ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>1_8ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>1_16ecb.xml</td>
<td>92%</td>
</tr>
<tr>
<td>2_11ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>3_1ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>4_10ecb.xml</td>
<td>94%</td>
</tr>
<tr>
<td>5_10ecbplus.xml</td>
<td>100%</td>
</tr>
<tr>
<td>6_11ecbplus.xml</td>
<td>90%</td>
</tr>
<tr>
<td>6_2ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>7_9ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>8_7ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>9_7ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>10_17ecbplus.xml</td>
<td>100%</td>
</tr>
<tr>
<td>11_1ecbplus.xml</td>
<td>97%</td>
</tr>
</tbody>
</table>

36
The next step in the study was to find a broader reading of agreeability between the re-annotations of event mentions and their annotations in the ecb+. This step involved a binary answer: a yes or no question. Taking the 20 randomly selected documents from across the ecb+ corpus, how many of those documents did this study agree with in full, and how many did not agree in full? This question was again in relation to the event mention components and their annotations within the document. If there was any disagreement in a document between the re-annotations provided by this study and the original annotations of the ecb+, the document would be marked under “disagreed”. If there is 100% agreement, then the document would be marked as “agreed”. This stage of the experiment in admittedly a little harsh, but it allows for a wider range of our agreement vs theirs to be represented, rather than the limited degree that can be represented through single files. As it resulted, however, the range of agreement was still sufficiently high. The results can be roughly deduced from table 9 above. Of the 30 documents that were selected and re-annotated, this study agreed with the prior annotations of event mentions for 20 of those documents, resulting in an agreement percentage of 66.6%. This result was not very satisfactory, in spite of the reasonably high agreement ratio. This study felt that the test might misrepresent the true agreement between all components annotated. In order to represent the second test more clearly, it was decided that another test should be performed that was more precise in representing the agreement. In total, 447 event mentions or components were annotated by this dissertation in the re-annotation process. Of those 447 components, in comparison to the same components and their annotations in the ecb+ corpus, the number of times there was event mention agreement was 437. This results in a P(A) of

Table 9: This table highlights the P(A), the percentage of agreement, between the event mention annotations of the ecb+ and the re-annotations of event mentions by this dissertation.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>12_2ecbplus.xml</td>
<td>100% (*96%)</td>
</tr>
<tr>
<td>13_14ecb.xml</td>
<td>75%</td>
</tr>
<tr>
<td>14_5ecb.xml</td>
<td>90%</td>
</tr>
<tr>
<td>16_11ecbplus.xml</td>
<td>100%</td>
</tr>
<tr>
<td>18_15ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>19_9ecbplus.xml</td>
<td>100%</td>
</tr>
<tr>
<td>19_6ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>20_3ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>23_3ecb.xml</td>
<td>83%</td>
</tr>
<tr>
<td>23_10ecb.xml</td>
<td>88%</td>
</tr>
<tr>
<td>26_10ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>33_2ecb.xml</td>
<td>97%</td>
</tr>
<tr>
<td>33_8ecbplus.xml</td>
<td>100%</td>
</tr>
<tr>
<td>38_9ecbplus.xml</td>
<td>100%</td>
</tr>
<tr>
<td>40_1ecb.xml</td>
<td>100%</td>
</tr>
<tr>
<td>43_4ecb.xml</td>
<td>100%</td>
</tr>
</tbody>
</table>
97.7%, a significantly higher percentage of agreement than the rough test produced, meaning that the total event mention annotation agreement between the efforts of this study and the original annotations are over 30% higher than the original second test would have suggested. These results are very promising for the ecb+.

The third experiment utilised the 30 selected documents as well. This test focused on how much this dissertation agrees with the annotations representing within-document coreference of event mentions. In other words, how much do we agree that the within-document annotations of the ecb+ hold true? Any within-document event mentions which corefer means, as was discussed in previous sections that they refer to the same entity. For example, as a re-clarification, in the sentence “Sean washes his car”, there are two coreferent event mentions which would be annotated with the HUMAN_PART_PER tag, “Sean” and “his”, as they refer to the same entity. As this also involves the event mentions within the 30 documents, and as the results of the second test concluded a very high ration of agreement between the re-annotations and original annotations, the expected within-document coreference relations agreement is also very high. It would make no sense, therefore, if these results were low. A test was performed anyway for the sake of understanding the practice of within-document coreference. We will focus on two files from the selected 30, one with 100% agreement proportion and the file with a lower P(A) value. Therefore, we will look at files 43_4ecb.xml and 14_5ecb.xml respectively, with reference to the event mentions and the agreement on their coreference within the files. Topic 43 in the ecb component revolves around the seminal event “AMD acquires ATI 2006”, as such 43_4ecb.xml contains a news item pertaining to this event. As the P(A) calculated for this file was 100%, it follows that the agreement between within document event mention coreference should also be 100%, as both parties are in total agreement about how the components were marked. Performing a calculation on this file is therefore irrelevant. Table 10 shows the components which are coreferent within this document, as annotated by the annotators of the ecb+ and then agreed by the re-annotations performed by this dissertation.

<table>
<thead>
<tr>
<th>Source event mentions</th>
<th>Target event mention</th>
</tr>
</thead>
<tbody>
<tr>
<td>“devices”, “graphics technology”</td>
<td>NON_HUMAN_PART (marked unknown)</td>
</tr>
<tr>
<td>“Advanced Micro Devices Inc.”, “AMD”</td>
<td>HUMAN_PART_ORG</td>
</tr>
<tr>
<td>“company”</td>
<td></td>
</tr>
<tr>
<td>“Supplier”, “intel corp.”, “intel”</td>
<td>HUMAN_PART_ORG</td>
</tr>
</tbody>
</table>

Table 10: An example of within-document coreference contained within 43_4ecb.xml. This dissertation agrees in full with the coreferring event mentions above.

In 14_5ecb.xml, the P(A) coefficient was calculated as 90%. Therefore there was some disagreement. Upon re-analysis, it emerges that there is also disagreement between this
dissertation and the annotators. Table 11 below shows how the annotators defined the within-document coreference between a few event mentions in 14_ecb.xml. Refer to the appendix for 14_5ecb.xml (document A.3). As written in the corpus, “the scene” and “a Waitrose store in high street, Banstead” are defined as coreferring event mentions. This dissertation agrees. However it does not agree with the other within-document coreference mentioned here. The spokeswoman mentioned in table below, deduced from the context of the file, is a spokeswoman for the Surrey Police. However, the annotation guidelines state that individual persons must be annotated with the HUMAN_PART_PER tag (Cybulska & Vossen, 2014). Despite being affiliated with an human organisation, that of Surrey Police, this dissertation disagrees that this should determine the annotation tag delivered below to the spokeswoman. After all, the definition for HUMAN_PART_ORG mentions nothing about individual persons, and in fact refers only to groups in an organised manner. “HUMAN_PART_ORG tag denotes organization entities limited to corporations, agencies and other groups of people defined by an established organizational structure” (Cybulska & Vossen, 2014).

<table>
<thead>
<tr>
<th>Source event mentions</th>
<th>Target event mention</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Surrey Police”, “spokeswoman”</td>
<td>HUMAN_PART_ORG</td>
</tr>
<tr>
<td>“the scene”, “a Waitrose store in high street, Banstead”</td>
<td>LOC_FAC</td>
</tr>
</tbody>
</table>

*Table 11: Within-document coreference in 14_5ecb.xml as defined in the ecb+. This dissertation disagrees with the coreferring event mentions in the first row.*

The final test could not be completed properly, due to time constraints. The last experiment involves the level of agreement between this dissertation and the cross-document event mention coreference of the ecb+. As was the case in the previous test, it may not be entirely necessary to perform a test of this calibre as the determined P(A) between all of the components annotated was 97.7%. Again, this determines that, within the documents that were re-annotated, the level of agreement will not possibly be low, and will always be significantly high. However, a small focus test was performed nevertheless. Taking a number of documents from the same folder, topic one, which details the seminal event “T. Reid check into rehab in 2008”, of which we had already re-annotated and compared two files, 1_18ecb.xml and 1_8ecb.xml, we compared the cross-document coreference instances that were defined by the annotators. The instances, which were set up using the CROMER tool, allowed for cross-document relations to be formed between event mentions i.e. a “Tara Reid” mention in 1_8ecb.xml would be coreferent to a “Tara Reid” mention in 1_18ecb.xml. Again, due to the lack of time remaining at the end of the experimentation process, this notion was only narrowly explored. However, the definitions provided by the annotators were very clear, and the instances of cross-document coreference were very easy to follow. As such, this dissertation
agrees with the cross-document coreference between these two files 100%. The event mentions included instances of “Tara Reid” and instances of “Promises Treatment Centre”, the rehabilitation facility in Malibu. It was difficult to contradict the coreference across these documents, in no small part due to the agreement of the event components shown in the first experiment.

4.3. Error Analysis

Before moving on to the conclusion, a small discussion of the errors encountered during the re-annotation process will be discussed. The errors vary, from difference in annotation to re-annotation of components that were not relevant, and from small errors noticed within the ecb+ as well as change in opinion through comparison. In the first experiment, there were a total of 10 event mention parts that remained re-annotated differently. One such example was discussed in section 4.2 in relation to the annotation of an individual as a HUMAN_PART_ORG instead of a _HUMAN_PART_PER tag. Similar disagreements exist between this dissertation and the annotators. Another disagreement was discussed earlier as well, in the annotation of “rehab” in the sentence “Lindsay Lohan checks into rehab in Newport Beach” as an ACTION_OCCURRENCE mention. This was in document 1_16ecbplus.xml. This study feels that “rehab in Newport Beach” should have been annotated under the LOC_FAC tag, rather than being two separate event mentions. There was also disagreement in 4_10ecb.xml. As explained, TIME_DURATION tags refer to a time space between two points that can be measured (Cybulska & Vossen, 2014). In this file, the TIME_DURATION file was applied to the sentence “…more than a half of inspiration for everybody from Madonna to the suicide girls…” to the word “half”. This dissertation disagreed with this tag initially, however upon further contemplation, changed opinion. Madonna and the suicide girls are HUMAN_PART_PER, human entities, but they are specific to two different times. In other words, there is a generation gap. So the annotation of “half” as TIME_DURATION may in fact be accurate, due to the vague time duration reference between two human entities. Another example of disagreement concerns re-annotation where there was no prior annotation. In 12_2ecbplus.xml, there were 23 components annotated in the ecb+ corpus. When the re-annotation experiment was performed, this study found that there were 24 event mention components instead. “A statement from the defence PRO says the incident happened yesterday…” In the sentence here, “says” was re-annotated by this study under the ACTION_REPORTING tag, following the guidelines that the reporting of an action occurrence or event should be annotated under it. However, this was unannotated in the ecb+. This study feels that this should have been marked as an event mention component. Finally, a small error was encountered in the original annotations of 14_5ecb.xml. An event mention “fire” was somehow split into two tokens, “fir” and “e”, but still marked as the same event mention (Cybulska & Vossen, 2014).
Ultimately, there was not much disagreement between re-annotation and the original annotation. The level of agreement was very high in all tests, thus the results are promising for the potential of the ecb+ corpus. The experience of the annotators versus the inexperience of annotation for the re-annotation process might have been an influence in the correction of some tagging, however the fact that there was some disagreement at all, no matter how little, might be something worth looking at.

4.4. Conclusion

This chapter focused on the main problem of this dissertation, the determination of event and event mention coreference agreement between this study and the annotators of the ecb+. Four experiments were performed, two large and two small. The results deduced suggest that this dissertation agrees a great deal with the annotators of the ecb+, and a P(A) score of 97.7% agreement was calculated. The difference in opinion was discussed, as well as how the inexperience of the re-annotation process in this case may have led to a reversal in opinion in times of contrast. In the concluding chapter, a discussion will be provided as to what these results mean for the ecb+ and event coreference in the future.
Chapter 5
Conclusions

5.1. Introduction

Due to restraints in time, a larger scale analysis of the corpus could not be completed. A detailed analysis of the component annotation was carried out, as well as the re-annotation of texts while following the ecb+ guidelines set out by Cybulska and Vossen (2014) in order to determine the percentage of agreement between this dissertation and the original annotators. Tests to determine the degree of within-document and cross-document coreference among event mentions was attempted at a lesser degree, also due to time restraints, however these final tests may not have been entirely necessary due to the overall level of agreement that was calculated between 437 out of the 447 event mentions that were analysed. Due to this high percentage, the expected agreement of within- and cross-document coreference could also only be determined to be significantly large. However in narrowing the testing down to a focused set of 30 documents, congruent to approximately 3% of the corpus, and considering that there were close to 1000 documents in total, the experiments that this dissertation undertook demonstrates in some capacity the usefulness of the ecb+ in relation to studies relating to event coreference. This chapter will attempt to draw conclusions from the readings and results which were shown in chapter 4. This will help to answer the questions that were asked in the beginning.

5.2. What do the results mean?

As all results deduced from experiments were very high, and thus agreement was calculated as 97.7%, the conclusion of this study is that the Eventcorefbank+ is a very efficient corpus. The guidelines defined by Cybulska and Vossen (2014), influenced by earlier efforts, provided a clear and precise basis on which to work to achieve an annotation of texts which are at their core, “event-centric”. It is the opinion of this dissertation that the annotation guidelines did indeed help to define a corpus that will be very useful in the field of event coreference in future. The questions that were asked in the beginning set out a task to see if this corpus would sufficiently represent the event mentions within. The answer concluded by this dissertation, due to the results, is yes. This corpus should be considered by anyone who wishes to approach event coreference in their own way. Where the annotation guidelines useful and appropriate? This dissertation says that they were. The annotation guidelines provided were less than trivial, and
very detailed. In spite of this, they were easy to understand and apply after having been studied. There are a number of cases where the annotation guidelines were vague or unspecific, as was discussed above, however for the most part they split all entities and event mentions into neatly defined categories which were easy to separate and thus made them a valuable source for the annotation of texts, in particular, news items. This is further reinforced again by the agreement between the annotators of the ecb+ and the re-annotation performed in this study. This brings the final point. As there was a significant level of agreement between the annotators in 2014 (Cybulska & Vossen 2014), and again significant agreement here, the future of this corpus looks very promising.

5.3. Future Work

As of this paper, the ecb+ corpus has not been used in a study to determine event coreference. It would be interesting to see research into further event coreference with this corpus, and the use of evaluation metrics such as the MUC (Vilain et al., 1995) being applied to it. This would allow for the difference in calculation of event coreference with reference to this corpus to be compared fairly with event coreference of the past that was performed on various other corpora.

It would also be interesting to see a modification of the corpus in question to make it more entity-centric rather than event specific. From the way the corpus is annotated, this study assumed that it is already very much geared for that possibility. The event mentions contained within are largely references to entities, in any case.

Finally, a professional re-annotation of the corpus could be performed to get rid of any small errors which are in the corpus. There are not many, perhaps the errors deduced by this study are not errors at all, but in order to really solidify this corpus as the future seminal source for event and entity coreference, a new set of hands could really up the rating. It may be improbable, but the re-clarification of the ecb+ guidelines to eliminate ambiguity once and for all would definitely improve on the corpus and on the annotation process for annotators. As discussed above, the guidelines are incredibly resourceful and this study endorses them, however the termination of all doubt by rewriting the guidelines would be an interesting approach to take towards the pinnacle of quality.
References


Appendix

**Document A.1. 1_10ecb.xml (Cybulska & Vossen, 2014):**

```xml
<Document doc_name="1_10ecb.xml" doc_id="DOC15653231646358310">
  <token t_id="1" sentence="0" number="0">Perennial</token>
  <token t_id="2" sentence="0" number="1">party</token>
  <token t_id="3" sentence="0" number="2">girl</token>
  <token t_id="4" sentence="0" number="3">Tara</token>
  <token t_id="5" sentence="0" number="4">Reid</token>
  <token t_id="6" sentence="0" number="5">checked</token>
  <token t_id="7" sentence="0" number="6">herself</token>
  <token t_id="8" sentence="0" number="7">into</token>
  <token t_id="9" sentence="0" number="8">Promises</token>
  <token t_id="10" sentence="0" number="9">Treatment</token>
  <token t_id="11" sentence="0" number="10">Center</token>
  <token t_id="12" sentence="0" number="11">,</token>
  <token t_id="13" sentence="0" number="12">her</token>
  <token t_id="14" sentence="0" number="13">rep</token>
  <token t_id="15" sentence="0" number="14">told</token>
  <token t_id="16" sentence="0" number="15">People</token>
  <token t_id="17" sentence="0" number="16">.</token>
  <token t_id="18" sentence="1" number="0">``</token>
  <token t_id="19" sentence="1" number="1">We</token>
  <token t_id="20" sentence="1" number="2">appreciate</token>
  <token t_id="21" sentence="1" number="3">your</token>
</Document>
```
respect to her and her family's privacy at this time, '' the 33-year-old actress's rep Jack Ketsoyan told the magazine exclusively for their Web
Ketsoyan would not say what the "American Pie" star was being treated for.

A friend of the actress told "People" she went...
to Promises on Tuesday and that her friends and her family supported her decision.

Reid launched a new fashion line called Mantra in December.
</ACTION_OCCURRENCE>

<ACTION_OCCURRENCE m_id="32">
  <token_anchor t_id="6"/>
  <token_anchor t_id="8"/>
</ACTION_OCCURRENCE>

<ACTION_OCCURRENCE m_id="33">
  <token_anchor t_id="75"/>
</ACTION_OCCURRENCE>

<ACTION_OCCURRENCE m_id="38">
  <token_anchor t_id="88"/>
</ACTION_OCCURRENCE>

<ACTION_REPORTING m_id="34">
  <token_anchor t_id="15"/>
</ACTION_REPORTING>

<ACTION_REPORTING m_id="35">
  <token_anchor t_id="72"/>
</ACTION_REPORTING>

<HUMAN_PART_PER m_id="28">
  <token_anchor t_id="85"/>
</HUMAN_PART_PER>

<HUMAN_PART_PER m_id="29">
  <token_anchor t_id="68"/>
</HUMAN_PART_PER>

<HUMAN_PART_PER m_id="20">
  <token_anchor t_id="4"/>
  <token_anchor t_id="5"/>
  50
</HUMAN_PART_PER>
<HUMAN_PART_PER m_id="21">
<token_anchor t_id="7"/>
</HUMAN_PART_PER>
<HUMAN_PART_PER m_id="22">
<token_anchor t_id="74"/>
</HUMAN_PART_PER>
<HUMAN_PART_PER m_id="23">
<token_anchor t_id="87"/>
</HUMAN_PART_PER>
<HUMAN_PART_PER m_id="26">
<token_anchor t_id="14"/>
</HUMAN_PART_PER>
<HUMAN_PART_PER m_id="27">
<token_anchor t_id="83"/>
</HUMAN_PART_PER>
<LOC_FAC m_id="36">
<token_anchor t_id="9"/>
<token_anchor t_id="10"/>
<token_anchor t_id="11"/>
</LOC_FAC>
<LOC_FAC m_id="37">
<token_anchor t_id="77"/>
</LOC_FAC>
<TIME_DATE m_id="30">
<token_anchor t_id="78"/>
<token_anchor t_id="79"/>
51
Document A.2. 1_8ecb.xml (Cybulska & Vossen, 2014):

Oh,
it's been so long since we were on the rehab beat.
After telling People in October that she didn't "need to do any of that anymore," American Pie actress and former
Tara Reid checked into Promises Treatment Center on Friday for an unspecified problem.

Promises is a celeb favorite, favored by Britney Spears and Lindsay.
Lohan

---

it

's

also

where

Mel

Gibson

retired

to

after

his

2006

accident

and

anti-Semitic

outburst

Not

quite

the

Christmas

alternative

we

're

're

looking

for
Get well soon, Tara!
Document A.3. 14_5ecb.xml (Cybulska & Vossen, 2014):

<Ten fire engines were tackling the</Ten>
last of the blaze which began on the roof of the store in Banstead, Surrey, before 9.

Residents nearby were evacuated as a precaution.
A Surrey Police spokeswoman said tonight:
``Surrey and Surrey Fire and Rescue are currently at the scene of a large fire at..."
a Waitrose Store in High Street, which is being treated as potentially suspicious.

Police were alerted to a fire on the
The fire has progressed rapidly through the store which is almost totally destroyed.

Local road closures close to...
the high street have been put in and residents living close to the fire are currently being evacuated. Around 10 fire service pumps are.
in attendance at the scene.

No one is believed to be injured at this time.

''People are asked to avoid the high street''
area if at all possible.

''Anyone with information is asked to call police on 0845 125 2222 or Crimestoppers anonymously on 0800 555 111.'
<Markables>

<ACTION_OCCURRENCE m_id="26">
  <token_anchor t_id="73"/>
</ACTION_OCCURRENCE>

<ACTION_OCCURRENCE m_id="24">
  <token_anchor t_id="58"/>
</ACTION_OCCURRENCE>

<ACTION_REPORTING m_id="19">
  <token_anchor t_id="39"/>
</ACTION_REPORTING>

<HUMAN_PART_ORG m_id="21">
  <token_anchor t_id="43"/>
  <token_anchor t_id="44"/>
</HUMAN_PART_ORG>

<HUMAN_PART_ORG m_id="22">
  <token_anchor t_id="46"/>
  <token_anchor t_id="47"/>
  <token_anchor t_id="48"/>
  <token_anchor t_id="49"/>
</HUMAN_PART_ORG>

<LOC_FAC m_id="23">
  <token_anchor t_id="53"/>
  <token_anchor t_id="54"/>
</LOC_FAC>

<LOC_FAC m_id="25">
  <token_anchor t_id="61"/>
  <token_anchor t_id="62"/>
</LOC_FAC>
</Markables>
<ACTION_OCCURRENCE m_id="28" RELATED_TO=""
TAG_DESCRIPTOR="t14b_fire" instance_id="ACT17478306085573007"/>
</Markables>
</Relations>

<CROSS_DOC_COREF r_id="33682" note="ACT17478306085573007">
<source m_id="24"/>
<target m_id="28"/>
</CROSS_DOC_COREF>

<CROSS_DOC_COREF r_id="33683" note="ACT17478359686333015">
</CROSS_DOC_COREF>
<source m_id="26"/>
<target m_id="29"/>
</CROSS_DOC_COREF>

<CROSS_DOC_COREF r_id="33684" note="LOC17478339342882751">
  <source m_id="25"/>
  <source m_id="23"/>
  <target m_id="30"/>
</CROSS_DOC_COREF>

<CROSS_DOC_COREF r_id="33685" note="HUM17478326731930159">
  <source m_id="21"/>
  <source m_id="27"/>
  <target m_id="31"/>
</CROSS_DOC_COREF>

</Relations>

</Document>