Annotated Data, version 1

Deliverable 3.3.1

Version FINAL

Authors: Agata Cybulska[1], Marieke van Erp[1], Sara Tonelli[2], Rachele Sprugnoli[2], Manuela Speranza[2], Anne-Lyse Minard[1], Ruben Urizar[3], Begoña Altuna[3], Piek Vossen[1], Antske Fokkens[1]

Affiliation: VUA[1], FBK[2], EHU[3]

Building structured event indexes of large volumes of financial and economic data for decision making

ICT 316404

Grant Agreement No.	316404
Project Acronym	NEWSREADER
Project Full Title	Building structured event indexes of large volumes of financial and economic data for decision making.
Funding Scheme	FP7-ICT-2011-8
Project Website	http://www.newsreader-project.eu/
Project Coordinator	Prof. dr. Piek T.J.M. Vossen VU University Amsterdam Tel. + 31 (0) 20 5986466 Fax. + 31 (0) 20 5986500 Email: piek.vossen@vu.nl
Document Number	Deliverable 3.3.1
Status & Version	FINAL
Contractual Date of Delivery	3 February 2014
Actual Date of Delivery	February 3, 2014
Type	report
Security (distribution level)	Public
Number of Pages	34
WP Contributing to the Deliverable	3
WP Responsible	FBK
EC Project Officer	Susan Fraser

Authors: Agata Cybulska[1], Marieke van Erp[1], Sara Tonelli[2], Rachele Sprugnoli[2], Manuela Speranza[2], Anne-Lyse Minard[1], Ruben Urizar[3], Begoña Altuna[3], Piek Vossen[1], Antske Fokkens[1]

Keywords: annotation, data

Abstract: This deliverable describes the annotation efforts of year 1. We present the setup of the annotation task, the guidelines used and the two resulting datasets. For the first year, we decided to split the annotation effort into two sub-tasks, the first focuses on the mention level and annotates new texts selected from Wikinews, the second task extends the EventCorefBank to make a first step towards benchmarking cross-document annotation in NewsReader.

Table of Revisions

Version	Date	Description and reason	By	Affected sections
0.1	December 2013	Structure of the deliverable set	Marieke van Erp	All
0.2	January 2014	First draft of the document	All Authors	All
0.3	January 2014	Revised document	All Authors	All
0.4	February 2014	Added IAA and Discussion	Anne-Lyse Minard, Agata Cybulska, Sara Tonelli, Marieke van Erp	3.1.1.9,4

Executive Summary

This deliverable describes the annotation efforts of year 1 of the NewsReader project. To reconstruct story lines from news, first events and participants are extracted through an NLP pipeline. The results of this are input to a semantic layer where contradictions and complementary information are reconciled and are ultimately stored in a knowledge base. To measure the performance of the automatic event extraction, benchmark datasets need to be created, which is the focus of this deliverable.

With respect to the project goals, a number of NLP tasks must be performed to generate a semantic representation of text. We decided to create two data sets within Newsreader. One specifically designed for development and testing of approaches to event coreference resolution and another one - a dataset annotated with a much broader scope of information that can be used for research of diverse aspects of language and development of multiple kinds of tools. The decision to dedicate a separate annotation task to event coreference resolution was motivated by the fact that event extraction and resolution of event coreference are crucial for all further processing steps within Newsreader. Furthermore, dataset for experiments on event coreference must be organized explicitly around a number of seminal events in order to capture cross-document coreference.

In this document, we describe the two datasets that have been annotated, as well as the guidelines for annotating them and the general setup of the annotation tasks. Together these tasks cover a large part of the types of information that NewsReader aims to extract and evaluate.

Contents

List of Tables

1 Introduction

In this deliverable, the annotation efforts of year 1 of the NewsReader project are described. The goal of the NewsReader project[1] is to reconstruct event story lines from the news by automatically processing daily news streams. For this purpose, an NLP pipeline has been constructed that extracts mentions of events, locations, dates, and participants (see WP02). The results of the extraction phase serve as input to a semantic layer where contradictions and complementary information are reconciled (see WP05) and are ultimately stored in a knowledge base (see WP06). To measure the performance of the automatic event extraction, benchmark datasets need to be developed, which is the focus of WP03.

As the scope of the NewsReader project encompasses both natural language processing and semantic web, any benchmark creation efforts need to take both domains into account. We therefore spent considerable time and effort developing annotation guidelines that transcend customary annotations in either domains. This deliverable focuses on the in text annotation that is more customary in the (computational) linguistics domain.

Within Newsreader, we decided to perform two annotation tasks resulting in creation of two datasets, each annotated with different applications in mind.

The first annotation task focuses on annotating a broad scope of information within single document using CAT guidelines (as described in Deliverable 3.1 "Annotation Module"). This extensively annotated dataset will have multiple applications due to the different grammatical and semantic aspects that are captured in it. Within NWR, it will serve as training and test set for linguistic research and development of diverse NLP tools.

A big step towards creation of a semantic layer over textual data is solving coreference between mentions of events. Event coreference resolution is the difficult task of determining whether two (or more) event descriptions refer to the same event. As this task is crucial for the construction of a semantic representation of text, we decided to have a separate annotation task that is specifically geared towards semantic-based event coreference resolution as delineated in D.5.1.1 "Event Narrative Module". One of the datasets commonly used in coreference resolution experiments is the EventCorefBank which contains 482 texts capturing inter- and intra-document descriptions of 43 seminal events [Bejan and Harabagiu, 2010]. But as Cybulska and Vossen show [Cybulska and Vossen, 2014b] the ECB corpus is not representative of large volumes of news as those to be processed within the Newsreader project. This is the reason why we decided to extend the ECB to become ECB+. The second annotation task focuses on marking mentions of events and their times and entities in the ECB+ corpus as well as on annotating the coreference relation between mentions of an event component within and across documents (as described in Technical Report NWR-2014-1).

This deliverable is structured as follows. In Section 2 the setup of the two annotation tasks are described, here we describe the number of annotators, the work plan and we give a summary of the guidelines used in each task. In Section 3, we describe the datasets that were used in both tasks. We conclude with a discussion and outlook on further annotation

[1] `http://www.newsreader-project.eu`

efforts within the project in Section 4.

2 Overview of the Annotation Task

2.1 NewsReader intra-document annotation task

Three partner institutions were involved in the NewsReader intra-document annotation task for English: FBK, VUA and EHU. As the leading institution, FBK produced the annotation guidelines with the collaboration of the other partners ([Tonelli *et al.*, NWR-2014-2]). After a training phase, in which FBK guided VUA and EHU in using the annotation tool (CAT, the CELCT Anotation Tool [Moretti and Sprugnoli, NWR-2014-5]), the annotation of English data (consisting of 20 documents selected from Wikinews) started with a discussion phase in which two annotators at each site annotated the first part (five sentences) of the same documents following the instructions given in the guidelines. The doubts aroused at this stage were discussed in a series of Skype meetings between VUA, EHU and FBK in order to reach an agreement in the annotation criteria. The remaining documents were also annotated by all involved partners and were used for the computation of inter-annotator agreement.

Annotation at document level consists of two main tasks: the detection and annotation of markables (i.e. entities, events, temporal expressions, numerical expressions, temporal signals, and causal signals) and the detection and annotation of relational information between markables (i.e. coreference, participant roles, causal links, temporal links, subordinating links and grammatical links), as explained in detail in the annotation guidelines ([Tonelli *et al.*, NWR-2014-2]).

An important feature of the NWR annotation is the distinction between instances and mentions to handle the annotation of the textual realization of entities and events (i.e. mentions) as well as the coreference chains that link different mentions to the same instance. The annotation of entities is based on ACE [Linguistic Data Consortium, 2008], the annotation of events is inspired by ISO-TimeML [ISO TimeML Working Group, 2008].

2.1.1 Entity instances and entity mentions

An entity (or entity instance) is an object or set of objects in the world or the mental representation of an object. Each entity is described through an empty element tag (i.e. no text-consuming) with the entity_type and tag descriptor attributes plus the entity class and external reference optional attributes.

The entity type attribute specifies the entity type from a semantic perspective. Its possible values are: i. PERSON (PER), for each distinct person or set of people mentioned in a document; ii. LOCATION (LOC), for two different types of location entities, i.e. those which can be defined on a geographical or astronomical basis (e.g. bodies of water, celestial bodies, addresses, etc.) and those (composite locations) which constitute a political entity and are comprised of a physical location, a population and a government (e.g. nations, states, provinces, population centers, etc.); iii. ORGANIZATION (ORG), for corporations,

agencies, and other groups of people defined by an established organizational structure (e.g. commercial organizations, educational organizations, religious organizations, sports organizations, etc.); iv. PRODUCT (PRO) for anything that can be offered to a market including, among others, facilities, vehicles, and food; v. FINANCIAL (FIN), for the entities belonging to the financial domain which are not included in one of the entity types described above; vi. MIXED (MIX), for conjunctions of entities belonging to different entity types.

The tag descriptor is a human-friendly identifier of the entity (for instance its name).

The entity class attribute expresses the definiteness of the entity instance. Its possible values are Specific Referential (when the entity being referred to is a particular, unique object or set of objects), Generic Referential (for entities not referring to a particular, unique object or set of objects, but to a kind or type of entity), Under-specified Referential (if it is impossible to determine its referent), and Negatively Quantified (when the entity has been quantified such that it refers to the empty set of the type of object mentioned).

The external reference attribute should contain the DBpedia URI identifying the specific entity instance.

An entity mention is the textual realization of an entity, that is the portion of text in which an entity is referenced within a text. In the annotation process, each entity mention is described through that portion of text (extent) and the following optional attributes: syntactic head and syntactic type (determined by the syntactic category of its syntactic head).

2.1.2 Event instances and event mentions

An event (or event instance) is the mental representation of the event to which various types of linguistic elements (e.g. nouns, verbs, pronouns) refer within a text. Each event is described through an empty element tag (i.e. no text-consuming) with the event class and tag descriptor attributes plus the external reference optional attribute.

The event class attribute has three possible values: i. SPEECH_COGNITIVE, for speech acts and cognitive events (in particular events that describe the action of declaring/narrating something and mental states and mental acts that involve mental or cognitive processes); ii. GRAMMATICAL, for events that are semantically dependent on a content verb/noun or on another event (e.g. light verbs or copula verbs, aspectual verbs and nouns signalling the initiation, continuation, or termination of another event, verbs and nouns expressing causal and motivational relations, verbs and nouns expressing occurrence, etc.); iii. OTHER, for all the events not covered by the previous classes.

The tag descriptor is a human-friendly description of the event instance.

The external reference attribute should contain the DBpedia URI identifying the specific event instance.

An event mention is the textual realization of an event, that is the portion of text in which an event is referenced within a text. Syntactically, the linguistic elements which may realize an event are verbs in finite or non-finite form, nouns (which can realize eventualities through a nominalization process from verbs or can have an eventive meaning in their

lexical properties or due to the co-text of their occurrence), pronouns (whose annotations are crucial to identify event co-reference), adjectives, and prepositional constructions.

In the annotation process, each event mention is described through that portion of text (extent) with the factuality and certainty attributes plus the following optional attributes: predicate (corresponding to the lemma of the token describing the event), pos (specifying the grammatical category which may realize an event), tense (capturing standard distinctions in the grammatical category of verbal tense), aspect (capturing standard distinctions in the grammatical category of semantic aspect), polarity (distinguishing affirmative and negative statements), and modality (conveying different degrees of modality of an event).

The factuality attributes conveys whether an event mention is presented as corresponding to a real situation in the world (FACTUAL), to a situation that has not happened (COUNTERFACTUAL) or has a speculative status (NON-FACTUAL); the certainty attribute expresses the binary distinction between certain and uncertain event mentions.

2.1.3 Temporal expressions

The annotation of temporal expressions, based on the ISO-TimeML guidelines [ISO TimeML Working Group, 2008]) and includes durations (i.e. periods of time), expressions describing calendar dates (with a granularity equal or greater than day), time, points or intervals of time smaller than a day (e.g. clock times), and sets of times (i.e. reoccurring time expressions). The following attributes are annotated: value (a normalized value based on the ISO-8601 standard); type (indicating whether it is a DATE, TIME, DURATION, or SET); functionInDocument (for the document creation date, used to indicate its function as a temporal anchor for other temporal expressions); anchorTimeID (in the case of anchored temporal expressions, annotated to show how the value is determined); beginPoint and endPoint (annotated to strengthen the annotation of durations).

Following the ISO-TimeML guidelines, we allow the annotation of empty, non-text consuming temporal expressions whenever a temporal expressions can be inferred from a text-consuming one. For example, in the case of anchored durations (i.e. expressions in which a duration is anchored to a further temporal reference, such as *two months ago*), we annotate a text-consuming DURATION TIMEX3 (whose value is *P2M*, i.e. a period of two months, in out example) and a DATE empty tag, whose value is the date it refers to. Similarly, in the case of range expressions, i.e. TIMEX3 made of two temporal expressions denoting the begin and end points of an implicit duration (e.g. *from 2005 to 2008*), we annotate two text consuming DATE temporal expressions and an empty DURATION temporal expression.

2.1.4 Numerical expressions

The annotation of numerical expressions includes both amounts and percentages. In the annotation process, each numerical expression is described through its textual extent and the type attribute, which has three possible values: PERCENT, MONEY (used for amounts described in terms of currencies), and QUANTITY (used for general amounts).

2.1.5 Temporal signals

Temporal signals (or SIGNALs), inherited from ISO-TimeML, are all those textual elements which make explicit a temporal relation between two event mentions, two temporal expressions, or an event mention and a temporal expression. The range of linguistic expressions which are to be marked as signals is restricted to temporal prepositions, temporal conjunctions, temporal adverbs, and some special characters, such as - and /, in temporal expressions denoting ranges.

2.1.6 Causal signals

Causal signals (or C-SIGNALs) are textual elements that indicate the presence of a causal relation between two events. We annotate as causal signals all causal uses of prepositions (e.g. *because of, on account of, as a result of, due to*, conjunctions (e.g. *because, since, so that, hence, thereby, by*), adverbial connectors (e.g. *as a result, so, therefore*), and clause-integrated expressions (e.g. *the result is, the reason why*).

2.1.7 Relations

The second main task is the detection and annotation of relations, which consists in the annotation of different types of links between markables to model the following phenomena: coreference, participant roles, causality, temporal ordering, subordination, dependency of grammatical events on governing events.

The REFERS_TO relation is used to represent the reference of an entity mention to an entity instance and the reference of an event mention to an event instance. As two or more entity/event mentions can refer to the same entity or event (in this case these mentions corefer) REFERS_TO is a many-to-one relation. According to our guidelines two event mentions corefer if their discourse elements (e.g. agents, location, and time) are identical in all respects, as far as one can tell from their occurrence in the text [Hovy *et al.*, 2013], i.e there is no semantic difference between them and it is possible to replace one mention with the other, sometimes with just some small syntactic modifications, without any semantic change.

The HAS_PARTICIPANT relation is a one-to-one relation linking an event mention to an entity mention or a to numerical expression which plays a role in the event (the participant).

In the current version of the NewsReader guidelines, PropBank used as the reference framework for the assignment of semantic role labels, which are encoded through the semantic role attribute. In PropBank there are five numbered arguments (corresponding either to the required arguments of a predicate or to those arguments that occur with high-frequency in actual usage) which correspond to the following semantic roles [Bonial *et al.*, 2010]: ARG0: agent; ARG1: patient; ARG2: instrument, benefactive, attribute; ARG3: starting point, benefactive, attribute; ARG4: ending point. As for modifiers (i.e. ARGM in PropBank), we use the value ARGM-LOC for locative modifiers and the value

ARGM-OTHER for all other modifiers (including for example comitatives, goal, and extent modifiers).

In NewsReader, we annotate causal relations between causes and effects denoted by event mentions (CLINKs) taking into consideration three basic categories of causation: cause, enable, and prevent. We will annotate all three types of causation but only if there is an explicit causal constructions between two event mentions. If a relation is signalled by the presence of causative conjunctions and prepositions (annotated as C-SIGNALs) they are reported in the c-signalID attribute of the CLINK.

The annotation of reported speech in NewsReader leans on the TimeML approach, reducing the scope of SLINKs to the annotation of subordinating relations between a SPEECH_ COGNITIVE event mentions and the events denoting its complement and expressing the message of reported or direct utterance/thought.

Grammatical relations (GLINKs) are used to link a mention of an event of type GRAMMATICAL (the source of the relation) to the mention of the event encoding its governing content verb or noun (the target). For example, this relation holds between an aspectual verb or noun and its event argument, a verb or a noun expressing occurrence and the occurred event.

Temporal relations (TLINKs) are used to link two event mentions, two temporal expressions or an event mention and a temporal expression. In order to create story lines, it is important to link each event to (at least) one other event in the text.

The annotation of temporal relations includes the following attributes: reltype, indicating how the two elements are temporally related, and signalID, which represents the ID of the SIGNAL that explicitly indicates the presence of a TLINK.

The possible values of the `relType` attribute (some of which are binary, one being the inverse of the other) are: BEFORE (an event/timex occurs before another), AFTER (the inverse of BEFORE), INCLUDES (one event/timex includes the other), IS_INCLUDED (the inverse of INCLUDES), MEASURE (used to connect an event and a DURATION TIMEX which provides information on the duration of the related event), SIMULTANEOUS (two events happen at the same time, or an event is perceived as happening at a moment in time), IMMEDIATELY BEFORE (one event/timex occurs immediately before the other), IMMEDIATELY AFTER (the opposite of IMMEDIATELY BEFORE), BEGINS (a timex or an event marks the beginning of another timex or event), BEGUN_BY (the inverse of BEGINS), ENDS (a timex or an event marks the ending of another event or timex), and ENDED_BY (the inverse of ENDS).

The annotation of temporal relations is divided into five subtasks: TLINKs between event mentions and the date creation time, TLINKs between main event mentions (the main event mentions correspond to the root element of the parsed sentence), TLINKs between main event mention and subordinated event mention in the same sentence (the subordinated event is identified on the basis of syntactic relations of dependencies), TLINKs between event mentions and temporal expressions in the same sentence, and TLINKs between temporal expessions (typically created when two temporal expressions in the same sentence are connected by a signal).

2.1.8 Updates with respect to Deliverable D3.1

With reference to the annotation guidelines presented in the NewsReader Deliverable D3.1, the most important changes are:

1. in the classification of event instances, the class CONTEXTUAL (meant to identify events relevant for the financial domain) were substituted by the more general class OTHER, which all the events not covered by the SPEECH_COGNITIVE and GRAMMATICAL classes;

2. in the classification of entity instances, entities of type MIX (for conjunctions of entities of different types) were introduced;

3. four attributes (functionInDocument, anchorTimeID, beginPoint and endPoint) were added to the annotation of temporal expressions;

4. the annotation empty temporal expressions was introduced;

5. some attributes for entities, entity mentions, events, and events mentions were made optional (i.e., i. the entity class and the external reference attributes for entities, ii. the syntactic type attribute for entity mentions, iii. the external reference attribute for events, iv. the predicate, part of speech, tense, aspect, polarity, and modality attribute);

6. the selection of PropBank as the framework for the assignment of semantic role labels to the HAS_PARTICIPANT relation and the subsequent introduction of the sem_role attribute led to the elimination of the ENG_dep attribute;

7. considerable effort has been done to define the guidelines for the annotation of CLINKs and GRAMMATICAL events.

2.2 ECB+ Annotation Task

The ECB+ annotation task focuses on annotation of mentions of events with their times and entities as well as coreference between them in text. Having in mind a specific application of event coreference resolution as well as conditioned by the need to guide the work of annotators with a different profile than those working on the Newsreader CAT annotation task (i.e. having extensive background in linguistics and NLP), it was necessary to devise different annotation guidelines for the purpose of ECB+ annotation. The major differences between the ECB+ and the NewsReader guidelines presented in Subsection 2.1 are (1) in the ECB+ annotation scheme we distinguish more entity and event categories, following guidelines from [Linguistic Data Consortium, 2008], [Pustejovsky *et al.*, 2003] and [Saurí *et al.*, 2005], and (2) fewer relations are marked as the focus is on coreference rather than complete annotation of event relations. In the remainder of this section, we will detail the most important aspects of the ECB+ guidelines. The complete guidelines can be found in [Cybulska and Vossen, 2014a].

1. action		*checked into, crash*
2. time		*on Monday*
3. location		*rehab in Malibu, California*
4. participant	human	*Linsay Lohan*
	non-human	*Car*

Table 1: Event Components

2.2.1 Event model

In the annotation guidelines of the Automatic Content Extraction program (ACE), an "event" is defined as a specific occurrence of something that happens, often a change of state, involving participants [Linguistic Data Consortium, 2008]. In the TimeML specification, "events" are described as "situations that happen or occur" that can be punctual or durational, as well as stative predicates describing "states or circumstances in which something obtains or holds true" [Pustejovsky *et al.*, 2003].

Expanding the above definitions, in the ECB+ annotation task we model events from news data as a combination of four components:

1. an event action component describing what happens or holds true
2. an event time slot anchoring an action in time describing when something happens or holds true
3. an event location component specifying where something happens or holds true
4. a participant component that gives the answer to the question: who or what is involved with, undergoes change as a result of or facilitates an event or a state; we divide event participants into **human participants** and **non-human participants**.

This annotation task requires the annotators to annotate event actions, locations, times and participants in text. For example in the sentence:

On Monday Lindsay Lohan checked into rehab in Malibu, California after a car crash.

Lindsay Lohan is a human participant involved with the event, *car* is a non-human participant, *On Monday* tells us when the event happened, *rehab in Malibu, California* is the place where the event happened and *checked into* and *crash* constitute actions.

In the ECB+ annotation specification we make a distinction between mentions (descriptions) of events in text and what they refer to, that is, their denotation (e.g. *World War II*, *WWII* and *Second World War* all refer to a global war between 1939 and 1945).

If an event is described more than once in one or in multiple texts, we say that its descriptions are *coreferent*. The second part of this annotation task consists of marking the inter- and intra-document coreference relation between mentions of actions and participants, times and locations. Consider the following sentences:

Lindsay Lohan checked into rehab.

Ms. Lohan entered a rehab facility.

These two sentences might refer to the same event, although as Ms. Lohan has been to rehab multiple times, it may also refer to two different instances. If one can determine based on the context that two event instances refer to the same real world event, they should be annotated as coreferent. If not, the actions should not be made coreferent. But the human participants from our example sentences should be marked as coreferent either way, as they refer to the same person. One would also need to figure out, whether *rehab* and *rehab facility* refer to the same facility and annotate it accordingly.

2.2.2 Event centric annotation

ECB+ annotation was performed with focus on the future task of semantic-based coreference resolution (D.5.1.1). With that in mind the ECB+ annotation specification was designed to be **event centric**. We annotate mentions of event components in text from the point of view of an event action, marking:

1. participants involved with an action as opposed to any participant mention occurring in a sentence

2. time when an action happened as opposed to any time expression mentioned in text

3. location in which the action was performed in contrast to a locational expression that does not refer to the place where an action happened.

For example *her father* in the sentence *Her father told ABC News he had no idea what exactly was going to happen* refers to the only human participant of the reporting action described in the sentence namely the father. The denotation of *her* does not refer to a participant of the reporting action hence we would leave *her* un-annotated. On the other hand *her* in the sentence *Her stay in rehab is over* does denote a human participant of action *stay*. Similarly *Mondays* in *I hate Mondays* does not refer to the time when the state holds true but in this sentence it should be annotated as a non-human participant. Event centric thinking was applied throughout the whole annotation effort and it guided the decision making process with regards to annotation of particular linguistic phenomena.

We will now take a look at three aspects of the ECB+ annotation guidelines: (1) the extent of a component mention in text, (2) the form in which a component mention can be expressed in language and (3) typology of entities and action classes applied for the purpose of this annotation task.

2.2.3 Mention extent

With respect to times and locations in ECB+, we annotate whole expressions, not only the head of a phrase as *two years ago*, *3 days later*, or *in the capital of Turkmenistan*, in *southern Iraq*. In the case of participants we annotate only the head of a phrase. By "head" we mean either the pronoun or, for NPs, the nominal part of the NP that is not used as a modifier and that expresses the most specific meaning (*The President of the U.S.*

Barack Obama; *David Cameron, the Prime Minister of UK*). For instance in the case of the NP *the US soldiers* only *soldiers* should be marked as the head of the NP and in the case of *the deceased man*, *man* should be annotated as a human participant and *deceased* as an action. Note that the head might consist of more than one word, in the case of proper names (e.g. *Barack Obama*).

With exception of locations and times, we do not annotate whole NPs but only their heads and we do not annotate markables within the extent of a bigger markable for instance a participant mention within the extent of a longer participant mention (*U.S. Secretary of State John Kerry*). Entity types (corresponding to annotation tags) are always to be assigned to the head of an entity phrase so for instance *the US soldiers* would get the entity type found appropriate for its head *soldiers* (we do not annotate US and its type).

Whether an action is verbal (like *the earth quaked*) or nominal (like *the earthquake*), we always annotate the word that is the strongest carrier of the action meaning that is the head of an action phrase (for example *People would rather hear the positive things being talked about than the negatives* or *This terrible war could have ended in a month*). In verbal action phrases, the "auxiliary" verbs are not annotated. The same holds for polarity markers applying to actions (e.g. negation words like *not*). Negation might be attached to different event components. We decided to always indicate negation as a property of an action (a number of negated action annotation tags will mark negated events). Besides auxiliary verbs, all main verbs including aspectuals (like *start*, *stop*, *continue*) and causative verbs (like *cause*) should be annotated as separate actions.

Some events, because of their historical significance, have been given their own name. People tend to refer to these events not in a descriptive way, but instead use those names. Examples include *9/11*, *September 11* or *World War II*. These event descriptions are annotated with ALL their elements. Phrasal verbs and idioms are also annotated including all elements.

We annotate the nominal, adjectival or adverbial part of a predicative phrase as action of the class "state" (*Game Five hero David Ross was happy just to be here*). At the same time, if part of a predicative phrase is a location, time or participant mention it should also be tagged as such.

In case of actions constituted by a combination of a verb and a noun (including light verb combinations), both parts of the action phrase are to be annotated separately from each other; i.e. the verb as an action and the noun as a participant. If the noun refers to an action, it is also to be annotated as an action (*Russia has made an offer to Syria*).

2.2.4 Mention part of speech

Event components can by expressed by means of different **parts of speech**.

We annotate actions that are expressed by:

- verbs
- nouns and proper names

- present- or past- participles also in modifier position
- predicative phrases expressed by adjectives, pronouns and nouns (also in noun phrases or prepositional phrases)
- and pronouns.

We annotate locations and times expressed by proper names, common nouns and adverbs. We annotate human and non-human participant entities expressed by proper names, common nouns and pronouns.

Note that actions, locations, times and participants can occur in text as modifiers of heads of nominal phrases as in *Connecticut school shooting*, *the deceased men*, *Tuesday's meeting*. If modifiers refer to a separate event component than the one referred to by the head of the nominal phrase they must be annotated as well.

2.2.5 Annotation tags

We decided to make an explicit distinction between specific **entity types**: human event participants, non-human participants, times, and locations (and a number of more specific subtypes amongst them e.g. HUMAN_PART_PER for participants of subtype individual person) as well as between a set of **action classes**.

We annotate event times following the types from the TIMEX3 specification (Pustejovsky et al., 2003). When annotating time expressions, the annotators shall specify one of the four subtypes: DATE (ECB+ tag TIME_DATE), TIME (TIME_OF_THE_DAY), DURATION (TIME_DURATION) and SET (TIME_REPETITION).

We annotate participants and locations expanding on the ACE entity subtypes [Linguistic Data Consortium, 2008].

We define event locations in line with ACE's general PLACE attribute, corresponding to entity types GPE, LOC or FAC referring to a physical location. Three tags are meant for event location annotation: (1) LOC_GEO corresponding to both, ACE's geo-political entities as well as ACE's location entities, (2) LOC_FAC meant for facility entities. The intention is that mentions tagged as both (1) and (2) reference in a sentence where an action happened. We also defined a third location tag: (3) LOC_OTHER – for any remaining type of event locations encountered in text.

We define human event participants similarly to ACE's event participants of entity type PER (ECB+ tag HUMAN_PART_PER), ORG (HUMAN_PART_ORG) but also metonymically used GPE (HUMAN_PART_GPE), FAC (HUMAN_PART_FAC) and VEH (HUMAN_PART_VEH) when referring to a population or a government (or its representatives). Besides these five subtypes we also distinguish two additional ones: HUMAN_PART_MET – for any remaining metonymically expressed human participants of events (*He has sworn loyalty to the flag* or *The crown gave its approval*) as well as HUMAN_PART_GENERIC for generic mentions referring to a class or a kind of human participants or their typical representative without pointing to any specific individual or individuals of a class [Linguistic Data Consortium, 2008], for instance generic *you* or *one* as event participants.

Next to locations, times and human participants we recognize a fourth entity type – NON_HUMAN_PART – for ALL remaining entity mentions – that is, besides human participants of events, event times and locations – that contribute to the meaning of an event action. These will often be artefacts expressed as a (direct or prepositional) object of a sentence or as PP phrases not in object position such as instrument phrases. Within the NON_HUMAN_PART type we distinguish a special sub-tag for generic entities: NON_HUMAN_PART_GENERIC for generic mentions referring to a class or a kind of non human entities or their typical representative without pointing to any specific individual object or objects of a class [Linguistic Data Consortium, 2008] for instance in the sentence: *Linda loves cats*.

We annotate actions with a limited set of classes from the whole set defined in the *TimeML Annotation Guidelines 1.2.1* [Saurí *et al.*, 2005]. We take over five event classes from the TimeML specification [Pustejovsky *et al.*, 2003]: occurrence (ECB+ tag ACTION-_OCCURRENCE), perception (ACTION_PERCEPTION), reporting (ACTION_REPORTING), aspectual (ACTION_ASPECTUAL) and state (ACTION_STATE). Additionally we employ two more action classes, one for causal events (ACTION_CAUSATIVE) and one for generic actions (ACTION_GENERIC). These seven classes have seven equivalents, to indicate polarity of the event. Polarity provides insight into whether the event did or did not happen. Negation of events can be expressed in different ways, including the use of negative particles (like *not*, *neither*), other verbs (like *deny*, *avoid*, *be unable*), or by negation of participants involved with an event as in *No soldier went home*. We will annotate negation as an action property by means of a set of action classes based on the seven non-negated action classes but with indication of negation through addition of a negation tag (NEG_) in front of an action class tag.

2.2.6 Relation annotation

Within this annotation task we annotate both, **inter-** and **intra-document** coreference relations.

Coreference relations can be established through mentions of:

- actions
- human participants
- non-human participants
- locations
- times

Two or more time expressions, location or participant mentions corefer with each other if they refer respectively to the same time, place or participants. Two action mentions corefer if they refer to the same instance of an action i.e. an action that happens or holds

true: (1) in the same time, (2) in the same place and (3) with the same participants involved.

Coreference should never be assigned between an action and an entity. Coreference shall not be established between entities belonging to different entity types for example between a location and a participant.

Anaphoric coreference is annotated.

In text one often comes across copular constructions with verbs like *be*, *appear*, *feel*, *look*, *seem*, *remain*, *stay*, *become*, *end up*, *get* (copular verbs list taken from OntoNotes annotation guidelines, 2007) as in:

(1) *This boy is James.*

If the subject (*this boy* referring specifically to this particular boy and not any other) and its complement (*James*) both refer to the same entity in the world, which in this case is *James*, coreference between the two should be annotated.

If however, the reference of the sentence subject and of the subject complement is not EXACTLY the same as in:

(2) *James is just a little boy.*

coreference should NOT be marked. In example (2) *James* refers to a particular boy called *James* but the phrase *a little boy* is indefinite and might refer to any little boy in the world, not necessarily to *James*. *James* in this case is just one element of the whole set, hence the reference of the two is not identical.

2.2.7 Setup of the ECB+ Annotation Task

The ECB+ annotation task is divided into three subtasks:

- Intra-document annotation of the newly created ECB+ corpus component [Cybulska and Vossen, 2014b]
- Modification of the ECB 0.1 annotation [Lee *et al.*, 2012; Recasens, 2011] of the EventCorefBank [Bejan and Harabagiu, 2010]
- Annotation of cross-document coreference.

Two student assistants were hired for a period of three months to perform the annotation. They were paid for their work. Both of them are native speakers of English pursuing a degree at VU University Amsterdam (one of them was an exchange student from the UK, both are British nationals). After the annotators were trained, we moved on to the first stage of the annotation.

Firstly, a newly created **ECB+ corpus component** of 502 news articles was annotated. The annotators were given the task to annotate mentions of event actions together with mentions of their participants, times and locations and intra-document coreference between them in the new ECB+ corpus component. The first topic of the new ECB+ component was annotated as burn in by both annotators. The next three topics were also annotated by both annotators (in total 55 texts per person annotated by both) and the

remainder of the corpus (447 texts) was divided between the two student assistants and annotated once.

In the second stage of the annotation process, adjustments were made to the ECB 0.1 annotation [Lee *et al.*, 2012; Recasens, 2011] of the **EventCorefBank** [Bejan and Harabagiu, 2010] (480 texts) to ensure compatibility of annotations of both corpus components. Each annotator worked on half of the data. In the ECB+ annotation scheme we make an explicit distinction between action classes and between a number of entity types. We re-annotated the ECB 0.1 annotation so that we not only have event actions and entities annotated (ECB 0.1. distinguishes between two tags: ACTION and ENTITY), but can also know precisely whether an entity is a location, time expression or participant. The same applies to actions that were re-annotated with specific action classes. Wherever necessary, adjustments were made with regards to mention extent. For human and non-human participant entities annotated in the ECB 0.1 corpus we made sure that only the head of a mention was explicitly annotated. With regards to times and locations we marked the whole phrase if not already done so. Regarding action annotation wherever necessary we additionally annotated light verbs and adjectival predicates. Finally adjustments were made to ensure that ECB 0.1 is compatible with the event centric annotation of the new corpus component.

The re-annotation efforts were focused on sentences that were selected during annotation of ECB 0.1. This allowed us to speed up the re-annotation process significantly. In principle we took over coreference relations established in ECB 0.1 but wherever needed we added new chains or adjusted the existing ones.

The third and final step in the ECB+ annotation process is to establish **cross-document coreference** relations between actions, times, locations and participants. We plan to move to this final phase at the beginning of February 2014.

The intra-document annotation in the first two stages of the ECB+ annotation process was performed by means of CAT - Content Annotation Tool[Bartalesi Lenzi *et al.*, 2012][2]. In the final stage for annotation of cross-document coreference relations we will use a tool called CROMER (CRoss-document Main Event and entity Recognition). CROMER is a NewsReader extension of a multi-user web interface [Bentivogli *et al.*, 2008] designed within the Ontotext project (`http://ontotext.fbk.eu/`).

3 Data

3.1 NewsReader Data

In order to make the annotated data of the NewsReader intra-document annotation task available not only to the project partners, but also to the wider audience of NLP researchers, the manual annotation is performed on a selection of *Wikinews*[3].

[2]previously known as CELCT Annotation Tool, `http://www.celct.it/projects/CAT.php`

[3]`http://en.wikinews.org/wiki/Main_Page`

Wikinews is a collection of multilingual online news article written collaboratively in a wiki-like manner, and it has been chosen for several reasons: first, because the data are freely available and can be re-distributed under Creative Commons license. Second, because it covers several languages, in particular English (the most represented one), Spanish, Italian and Dutch. Dumps in different languages referred to specific days can be downloaded in XML format. For NewsReader, we downloaded the dumps of August 2013. The English version includes 19,385 news[4] , the Italian one around 8,247 news[5], the Spanish 7,779[6] and the Dutch ∼1083[7]. The average length of an article is 250 words.

Some articles in different languages are comparable because editors can put the reference link to the page in another language, from which the article was created. For instance the English-Italian subset found in our dump includes 620 news. Besides, additional metadata have been added manually, and may be relevant to NewsReader activities. In particular, a category (e.g. Law, Economy and business, Disasters and accidents, etc.) is assigned to each news. Besides, named entities inside each document are linked to the corresponding Wikipedia page, and the link to the original source of the news article is also reported.

For all the above mentioned reasons, the NewsReader intra-document annotation task has been performed on a selection of *Wikinews*. In the first project year, annotation has focused on the English data.

From the full dump of news articles, we extracted those where "Apple" was mentioned in the title (69 files). Then, we manually selected 20 of them for the first annotation activities (training, inter-annotator agreement). Articles were selected in order to ensure that different articles deal with the same topic over time (e.g. launch of a new product, life of Steve Jobs), such that events and entities would co-refer in this small corpus and a sort of story line could be built based on the article publication dates and content.

An example of Wikinews is reported below. Note that for our annotation task we only consider the title, the date and the content.

```
<file id="2898">
  <head>NewsReader intra-document annotation task
    <url>8573</url>
    <pagesize>3412</pagesize>
    <dtime>Mon Oct 07 15:51:46 CEST 2013</dtime>
    <content-type>text</content-type>
    <encoding>UTF8</encoding>
    <username>Brian McNeil</username>
    <title>Sprint/RealNetworks to provide cell phone Internet radio and podcasts in US</
        title>
    <date>September 19, 2005</date>
    <keywords>Sprint Nextel Corporation, RealNetworks, KCRW, Santa Monica, National Public
```

[4] http://dumps.wikimedia.org/enwikinews/20130814/enwikinews-20130814-pages-articles.xml.bz2

[5] http://dumps.wikimedia.org/itwikinews/20130808/itwikinews-20130808-pages-articles.xml.bz2

[6] http://dumps.wikimedia.org/eswikinews/20130820/eswikinews-20130820-pages-articles.xml.bz2

[7] http://dumps.wikimedia.org/nlwikinews/20130817/nlwikinews-20130817-pages-articles.xml.bz2

```
        Radio, California, Sanyo, Samsung, LG Group, Apple Computer, Motorola, ITunes,
       Cingular Wireless</keywords>
    <categories>United States, North America, Internet, Economy and business</categories>
    <wordnum>242</wordnum>
    <charnum>1555</charnum>
  </head>
  <content>Wireless provider Sprint announced today, that they would be working with
      RealNetworks to launch a streaming music service, for Sprint wireless customers,
      called Rhapsody Radio. The service will allow users to listen to podcasts and six
      streaming radio stations (rap, hip-hop hits, alternative, country, '70s, and music
      news) from their mobile phone. Users will be able to listen to live broadcasts of
      those stations over their cellular connection. For a monthly subscription fee of
      \$6.95, users will also have unlimited access to commercial-free radio broadcasts.
      Selected streaming podcasts from KCRW, Santa Monica 89.9 FM, and National Public
      Radio's Southern California station will also be available. A special Beats N Breaks
       stream is also available. Beats N Breaks provides background music from hip-hop
      songs for users to rap with.
</content>
</file>
```

3.1.1 Inter-Annotator Agreement

We measure inter-annotator agreement on both markables and relations with the Dice's coefficient ([Dice, 1945]). In the case of many-to-one relation, we also compute the Krippendorff's alpha coefficient ([Krippendorff, 1980]).

Three annotators have annotated the three same files. We first compute the agreement scores by pairs of annotators and then the macro-average on the pairwise scores.

Markables The agreement on markables is evaluated with the Dice's coefficient. The Dice's coefficient is computed by multiplying by two the number of annotated markables whose extent matches exactly and then dividing it by sum of annotated markables in the two files.

The scores are given in Table 2.

In addition to this, we also evaluate agreement on non-optional attributes. In this case, we compute the accuracy of attribute value by taking into consideration only matched markables. Event mentions have two non-optional attributes: `factuality` and `certainty`; temporal expressions (TIMEX3) have three non-optional attributes: `type`, `value` and `functionInDocument` (functInDoc); numerical expressions (VALUE) have one non-optional attribute: `type`.

As shown in Table 2, results are satisfactory, with an agreement macro-average above 0.77 for all markables except for entity mentions.

We do not provide data about inter-annotator agreement for entity and event instances because by definition they have no textual extents on which to base the comparison and no discriminating attributes. It is evaluated indirectly through the computation of the agreement obtained on the annotation of REFERS_To relations as each entity/event instance is the target of one and only one REFERS_TO relation.

	Dice's coefficient	Accuracy for attributes		
ENTITY_MENTION	0.599			
EVENT_MENTION	0.794	factuality 0.796	certainty 0.940	
SIGNAL	0.933			
TIMEX3	0.856	type 0.957	value 0.813	functionInDocument 1
VALUE	0.778	type 0.722		

Table 2: Agreements for the markables

Relations We define a relation match as a pair of relations with matching source and matching target. For each relation type we measure two Dice's coefficients. In the first case the Dice's coefficient is computed by multiplying by two the number of relation matches and then dividing it by the number of relations involving as source and target only markables annotated in both files. In the second case the Dice's coefficient is computed by multiplying by two the number of relation matches and then dividing it by the total number of annotated relations in both files.

As for markables, we also indicate the accuracy for the non-optional attributes of relations.

The scores are given in Table 3.

	Dice's coefficient (markables match)	Dice's coefficient (all relations)	Accuracy for attributes
HAS_PARTICIPANT	0.709	0.395	sem_role 0.811
SLINK	0.745	0.631	
GLINK	0.889	0.727	
TLINK	0.552	0.473	type 0.645
REFERS_TO	0.704	0.534	

Table 3: Agreements for the relations

The agreement is generally good, especially if we consider only the relations involving markables annotated in both files (first column). The Dice's coefficient is above 0.7 for all relations except TLINKs.

Corpus component	Topic-Nr	Text-Nr
ECB 0.1	43	480
New component ECB+	43	502

Table 4: ECB+ statistics

In the specific case of REFERS_TO relations, which have (non text consuming) instances as targets, we define a relation match as a pair of relations in which the sources are the same and the attributes of the targets (`type` for entities and `class` for events) match.

For the REFERS_TO relation we compute also the Krippendorff's alpha, which is often used for coreference annotation agreement. We measure it only on the source cluster, i.e. without considering the target. Unlike the Dice's coefficient the alpha takes into account the units inside the source cluster. The macro-average of the alpha obtained for each annotator pair is 0.850.

3.2 ECB+ Data

The ECB+ corpus consists of two corpus components: (1) texts from the EventCoref-Bank (ECB) (Bejan and Harabagiu, 2010) and (2) the newly created component of ECB+ [Cybulska and Vossen, 2014b]. Table 4 presents some corpus statistics.

The **ECB corpus** consists of 43 topics (each corresponding to a seminal event), which in total contain 482 texts from GoogleNews archive (`http://news.google.com`), selectively annotated (amongst other relations) with within- and cross-document event coreference. Events were annotated in accordance with the TimeML specification [Pustejovsky *et al.*, 2003].

The annotation of the ECB was extended by [Lee *et al.*, 2012], following the OntoNotes annotation guidelines [Pradhan *et al.*, 2007]. This version of the corpus consists of 480 texts (text 4 from topic 7 and text 13 from topic 19 from the ECB were missing from the copy which we managed to find on the web). The re-annotation process resulted in fully annotated sentences and annotation of NP coreference relations (no specific annotation of entity or action types was performed).

The validity of corpus based studies depends on the notion of **representativeness** of a corpus. If a corpus is not representative of the sampled language population, one cannot be sure that the results of experiments obtained on it can be generalized onto the intended language population [Sinclair, 2004]. The ECB corpus is a resource often used in event coreference experiments. While containing multiple documents describing particular real world events, in most cases, ECB captures only single instances of each particular event type. For instance the seminal event from ECB topic one, that is Tara Reid's check-in into rehab in 2008, constitutes the only rehab-related event coreference chain in the corpus; and so the only instance of a rehab check-in event captured by the corpus. As shown in [Cybulska and Vossen, 2014b] the distribution of event descriptions in the ECB is not representative of the news that one can find on the web. Because the number of event instances per topic is limited (in most cases referring to only one real world event of an

event type, with exception of few topics like earthquake, acquisition, death and fire), event descriptions from a particular topic tend to share their entities (for a specific overview of main coreference chains in the ECB+ see table below). By that the event coreference task becomes simplified to topic classification. With the objective to make the ECB corpus more representative of large volume streams of news, we augmented the topics of the ECB corpus with 502 texts reporting different instances of event types provided in the ECB to increase the representativity of the corpus. The first ECB topic consists of texts outlining Tara Reid's check-in to rehab in 2008. We created an extension to topic number one of the ECB (called "1plus"), that is constituted by a collection of texts describing another real world event of the same type, namely Lindsay Lohan going into a rehab facility in 2013. **ECB+** is a new resource for evaluation of approaches to event coreference resolution that is more representative of large volume streams of news, published over a longer period of time [Cybulska and Vossen, 2014b].

ECB+ texts were collected through Google News search. On average we gathered roughly eleven texts per topic.

Table 5 specifies events captured in both components of the corpus, the original ECB and in the new component of ECB+. Column 1 "T" specifies the topic number. Column 2 provides a brief description of seminal events in ECB. Column 3 lists the seminal events of the new corpus component. Columns 4 ("Tn^1") and 5 ("Tn^2") indicate the number of texts collected per topic (column 4 refers to the ECB and column 5 to the new ECB+ corpus component).

T	Seminal event ECB	Seminal event ECB+	Tn^1	Tn^2
1	T. Reid checks into rehab in 2008	L. Lohan checks into rehab in 2013	18	21
2	H. Jackman announced as next Oscar host 2010	E. Degeneres announced as next Oscar host 2014	10	11
3	Courthouse escape Brian Nicols Atlanta 2008	Prison escape A.J. Corneaux Jr. Texas 2009	9	11
4	B. Page dies in LA 2008	E. Williams dies in LA 2013	14	11
5	Philadelphia 76ers fires M.Cheeks 2008	Philadelphia 76ers fires J. O'Brien 2005	13	10
6	"Hunger Games" sequel negotiations C.Weitz 2008	"Hunger Games" sequel negotiations G.Ross 2012	9	11
7	W. Klitchko defended IBF, IBO, WBO titles from H. Rahman 2008	W. Klitchko defended IBF, IBO, WBO titles from T.Thompson 2012	11 -1	11
8	bank explosion Oregon 2008	bank explosion Athens 2012	8	11
9	Bush changes ESA 2008	Obama changes ESA 2009	10	13
10	Angels made an eight year offer to M. Teixeira 2008	Red Socks made an eight year offer to M. Teixeira 2008	8	13
11	parliamentary election in Turkmenistan 2008	parliamentary election in Turkmenistan 2013	11	5
12	Indian Navy prevents a pirate attack on an Ethiopian vessel Gulf of Aden 2008	Indian Navy prevents a pirate attack on merchant vessels Gulf of Aden 2011	19	11

13	Wassila Bible Church fire in Alaska 2008	Mat-Maid Dairy fire in Alaska 2012	22	11
14	Waitrose supermarket fire in Banstead, Surrey 2008	Waitrose supermarket fire in Wellington 2013	10	11
16	Avenues Gang assassination of J.A. Escalante Cypress Park 2008	Hawaiian Gardens assassination of sheriff's deputy J.Ortiz Hawaiian Gardens 2005	3	11
18	deadly office shooting Vancouver 2008	deadly office shooting Michigan 2007	16	11
19	riots in Greece over teenagers death 2008	riots in Brooklyn over teenagers death 2013	15 -1	11
20	Qeshm island earthquake 2008	Qeshm island earthquake 2005	5	11
21	Bloomington hit and run 2008	Queens hit and run 2013	12	14
22	S.D. Crawford Smith accused of killing co-workers Staunton 2008	Y.Hiller accused of killing co-workers Philly 2010	9	14
23	M.Vinar dies in a climbing accident on Mount Cook 2008	R.Buckley, D.Rait dies in climbing accidents on Mount Cook 2013	10	11
24	4 robbers in drag steal jewelry in Paris 2008	4 robbers steal jewelry in Paris 2013	15	11
25	the Saints put R.Bush on injured reserve 2008	the Saints put P.Thomas on injured reserve 2011	15	11
26	Mafia member G.L.Presti dies in prison Sicily 2008	Mafia member V.Gigante dies in prison Montana 2005	13	11
27	Microsoft releases an IE patch 2008	Microsoft releases an IE patch 2013	17	11
28	Mark Felt dies in CA 2008	Fred LaRue dies in Miss. 2004	13	11
29	Colts beat Jaguars, secure no. 5 seed in the playoffs Fla. 2008	Colts beat Chiefs, secure no. 5 seed in the playoffs Missouri 2012	11	13
30	France Telecom cable disruption in the Mediterranean 2008	Seacom cable disruption Egypt 2011	14	13
31	T.Hansbrough becomes all-time leading scorer N.C. 2008	D.McDermott becomes all-time leading scorer Missouri 2013	14	11
32	Gary Gomes double murder New Bedford 2009	John Jenkin double murder Cumbria 2013	8	11
33	J.Timmons on trial for stray bullet killing of a 10 year old girl Albany, N.Y. 2008	A.Lopez on trial for stray bullet killing of Z.Horton Brooklyn 2011	5	11
34	Sanjay Gupta nominated for U.S. Surgeon General 2009	Regina Benjamin nominated for U.S. Surgeon General 2013	16	12
35	V.Jackson arrested under DUI in San Diego 2009	J.Williams arrested under DUI in San Diego 2009	10	11
36	W.Blackmore, J.Oler polygamy trial Canada 2009	Jeff Warren polygamy trial Texas 2011	9	11
37	6.1 earthquake Indonesia 2009	6.1 earthquake Indonesia 2013	7	14
38	small earthquake in Sonoma County 2009	small earthquake in Sonoma County 2013	4	11

39	Matt Smith role take over "Doctor Who" 2009	Peter Capaldi role take over "Doctor Who" 2013	14	11
40	Apple announces new MacBook Pro CA 2009	Apple announces new MacBook Pro CA 2012	10	11
41	Israel bombs Jabaliya camp 2009	Sudan bombs Yida camp 2011	9	11
42	T-Mobile USA adds new BlackBerry model to portfolio 2009	T-Mobile USA adds new BlackBerry model to portfolio 2012	13	10
43	AMD acquires ATI 2006	AMD acquires Seamicro 2012	8	15
44	Hewlett-Packard acquires EDS 2008	Hewlett-Packard acquires EYP 2007	7	15
45	S.Peterson found guilty of killing pregnant wife L.Peterson CA 2004	C.K.Simpson found guilty of killing pregnant girlfriend K.M.Flynn Mississippi 2013	8	12

Table 5: Topic specification of ECB+ components

3.2.1 Inter-Annotator Agreement

We calculated the preliminary inter-annotator agreement scores on topic 1 of the new ECB+ corpus component which contains 21 texts. We focused on measuring how much agreement there is on the assignment of event component tags per token of a mention, leaving the calculation of agreement on coreference to the cross-document annotation phase. For the purpose of this calculation, a number of sentences describing the seminal event of topic 1 was preselected. Both annotators were asked to annotate the same sentences in all 21 texts of topic 1. To measure the inter-annotator agreement between the annotators we used Cohen's Kappa [Cohen, 1960], a measurement that considers chance agreement. We calculated Cohen's Kappa when distinguishing all 30 annotation tags and also when looking at the main components that is grouping the specific tags into 5 categories: ACTION, LOC, TIME, HUMAN_PARTICIPANT and NON_HUMAN_PARTICIPANT. On the first topic our two coders reached Cohen's Kappa of 0.76 when assigning all 30 tags. This score can be interpreted as substantial agreement [Landis and Koch, 1977]. The inter-annotator agreement on the five main event component tags reached agreement level perfect: 0.82 Cohen's Kappa, although note that in these calculations untagged tokens were considered (for which we automatically assigned tag UNTAGGED). The confusion matrix below shows the distribution of the five main tags in the first topic of the corpus component as coded by the annotators.

An analysis of the confusion matrix revealed that the annotators mainly struggled with the definition of mention extents, annotating whole mention phrases while the guideline specified otherwise that is to only annotate the head (or the other way around). After additional training we continued with annotation of the remaining topics of the new ECB+ component.

	Action	Time	Location	Human participant	Non-human participant	Untagged
Action	449	1	6	4	2	31
Time	1	329	1	0	0	47
Location	14	0	419	3	5	29
Human participant	7	0	5	361	0	24
Non-human participant	6	0	0	2	7	1
Untagged	120	27	57	42	7	1546

Table 6: Confusion matrix ECB+ topic 1; five component annotation by two coders

4 Discussion and Outlook

In this deliverable, we have described the status of the annotation efforts in year 1 of the NewsReader project. As the differences between the first version of the annotation guidelines (see Deliverable 3.1) and the current annotation guidelines show, event annotation is a difficult task and we do not exclude further updates as our annotation efforts and discussions with our colleagues in the field progress. Compared to the first guidelines version, however, we have already taken into account the remarks and discussions emerged during the 1^{st} Workshop on EVENTS[8], since the NewsReader consortium was present in the program committee and participated to the working sessions. This second version of the guidelines is also more aware of the requirements of WP4 and WP6.

Both the NewsReader and ECB+ annotation guidelines annotate mentions of events, times and entities in text and relations between them. Their manner in achieving this however is slightly different as they have different applications in mind and what comes with it, focus on diverse annotation aspects and are meant for annotation of data, selection of which was performed in different ways. While the ECB+ corpus is organized around a number of seminal events, capturing within- and cross-document coreference chains, the CAT annotation was applied to news dealing with the same topic over time so that a storyline is captured.

For NewsReader, it is very important to look beyond events and form story lines. The CROMER annotation tool has been set up in such a manner that it can be extended to not only be used for cross-document coreference, but also to relate events to story lines. We expect to have this functionality in place sometime in mid-2014.

Since all annotation activities are planned to end in month 24, in the second project year FBK, EHU and VUA will all be fully involved in this task, concluding first the annotation of English documents and then performing language-specific annotations. Although defining the guidelines, training the annotators and implementing the annotation tools took longer than expected, we are confident that the benchmarks in all project languages will be made available in compliance with the planned milestones.

[8] `https://sites.google.com/site/cfpwsevents/`

References

[Bartalesi Lenzi *et al.*, 2012] Valentina Bartalesi Lenzi, Giovanni Moretti, and Rachele Sprugnoli. CAT: the CELCT Annotation Tool. In *Proceedings of LREC 2012*, 2012.

[Bejan and Harabagiu, 2010] Cosmin Bejan and Sanda Harabagiu. Unsupervised event coreference resolution with rich linguistic features. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1412–1422, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

[Bentivogli *et al.*, 2008] Luisa Bentivogli, Christian Girardi, and Emanuele Pianta. Creating a gold standard for person crossdocument coreference resolution in italian news. In *Proceedings of the LREC 2008 Workshop on Resources and Evaluation for Identity Matching, Entity Resolution and Entity Management*, 2008.

[Bonial *et al.*, 2010] Claire Bonial, Olga Babko-Malaya, Jinho D. Choi, Jena Hwang, and Martha Palmer. Propbank annotation guidelines, version 3.0. Technical report, Center for Computational Language and Education Research, Institute of Cognitive Science, University of Colorado at Boulder, Pisa, Italy, 2010. `http://clear.colorado.edu/compsem/documents/propbank_guidelines.pdf`.

[Cohen, 1960] J. Cohen. The coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 21(1):37–46, 1960.

[Cybulska and Vossen, 2014a] Agata Cybulska and Piek Vossen. Guidelines for ecb+ annotation of events and their coreference. Technical Report NWR-2014-1, VU University Amsterdam, 2014.

[Cybulska and Vossen, 2014b] Agata Cybulska and Piek Vossen. Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution. In *Proceedings of LREC 2014*, 2014.

[Dice, 1945] Lee Raymond Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, July 1945.

[Hovy *et al.*, 2013] Eduard Hovy, Teruko Mitamura, Felisa Verdejo, Jun Araki, and Andrew Philpot. Events are not simple: Identity, non-identity, and quasi-identity. *NAACL HLT 2013*, page 21, 2013.

[ISO TimeML Working Group, 2008] ISO TimeML Working Group. ISO TC37 draft international standard DIS 24617-1, August 14 2008. `http://semantic-annotation.uvt.nl/ISO-TimeML-08-13-2008-vankiyong.pdf`.

[Krippendorff, 1980] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology*. Sage Publications, 1980.

[Landis and Koch, 1977] J. R. Landis and G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[Lee *et al.*, 2012] Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. Joint entity and event coreference resolution across documents. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 489–500, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[Linguistic Data Consortium, 2008] Linguistic Data Consortium. Ace (automatic content extraction) english annotation guidelines for entities, version 6.6 2008.06.13. Technical report, Linguistic Data Consortium, June 2008. `http://projects.ldc.upenn.edu/ace/docs/English-Entities-Guidelines_v6.6.pdf`.

[Moretti and Sprugnoli, NWR-2014-5] Giovanni Moretti and Rachele Sprugnoli. CAT User Manual for the NewsReader EU Project. Technical report, Fondazione Bruno Kessler, NWR-2014-5.

[Pradhan *et al.*, 2007] Sameer Pradhan, Lance Ramshaw, Ralph Weischedel, Jessica MacBride, and Linnea Micciulla. Unrestricted coreference: Indentifying entities and events in ontonotes. In *Proceedings of the IEEE International Conference on Semantic Computing (ICSC)*, 2007.

[Pustejovsky *et al.*, 2003] J. Pustejovsky, J. Castano, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer, G. Katz, and D. Radev. Timeml: Robust specification of event and temporal expressions in text. In *IWCS-5 Fifth International Workshop on Computational Semantics*, 2003.

[Recasens, 2011] Marta Recasens. Annotation guide- lines for entity and event coreference. `http://www.bbn.com/NLP/OntoNotes`, 2011.

[Saurí *et al.*, 2005] Roser Saurí, Jessica Littman, Robert Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. Timeml 1.2.1 annotation guidelines, October 2005. `http://timeml.org/site/publications/timeMLdocs/annguide_1.2.1.pdf`.

[Sinclair, 2004] John Sinclair. Developing linguistic corpora: a guide to good practice. `http://ota.ahds.ac.uk/documents/creating/dlc/chapter1.htm#section4`, 2004.

[Tonelli *et al.*, NWR-2014-2] Sara Tonelli, Rachele Sprugnoli, and Manuela Speranza. NewsReader Guidelines for Annotation at Document Level. Technical report, Fondazione Bruno Kessler, NWR-2014-2.