

# Annotated Data, version 2

## Deliverable D3.3.2

Version FINAL

**Authors:** Marieke van Erp<sup>1</sup>, Piek Vossen<sup>1</sup>, Rodrigo Agerri<sup>3</sup>, Anne-Lyse Minard<sup>2</sup>,  
Manuela Speranza<sup>2</sup>, Ruben Urizar<sup>3</sup>, Egoitz Laparra<sup>3</sup>, Itziar Aldabe<sup>3</sup>,  
German Rigau<sup>3</sup>

**Affiliation:** (1) VUA, (2) FBK, (3) EHU



BUILDING STRUCTURED EVENT INDEXES OF LARGE VOLUMES OF FINANCIAL AND ECONOMIC  
DATA FOR DECISION MAKING  
ICT 316404

<b>Grant Agreement No.</b>	316404
<b>Project Acronym</b>	NEWSREADER
<b>Project Full Title</b>	Building structured event indexes of large volumes of financial and economic data for decision making.
<b>Funding Scheme</b>	FP7-ICT-2011-8
<b>Project Website</b>	<a href="http://www.newsreader-project.eu/">http://www.newsreader-project.eu/</a>
<b>Project Coordinator</b>	Prof. dr. Piek T.J.M. Vossen VU University Amsterdam Tel. + 31 (0) 20 5986466 Fax. + 31 (0) 20 5986500 Email: <a href="mailto:piek.vossen@vu.nl">piek.vossen@vu.nl</a>
<b>Document Number</b>	Deliverable D3.3.2
<b>Status &amp; Version</b>	FINAL
<b>Contractual Date of Delivery</b>	January 2015
<b>Actual Date of Delivery</b>	January 31, 2015
<b>Type</b>	Report
<b>Security (distribution level)</b>	Public
<b>Number of Pages</b>	40
<b>WP Contributing to the Deliverable</b>	WP3
<b>WP Responsible</b>	FBK
<b>EC Project Officer</b>	Susan Fraser
<b>Authors:</b>	Marieke van Erp <sup>1</sup> , Piek Vossen <sup>1</sup> , Rodrigo Agerri <sup>3</sup> , Anne-Lyse Minard <sup>2</sup> , Manuela Speranza <sup>2</sup> , Ruben Urizar <sup>3</sup> , Egoitz Laparra <sup>3</sup> , Itziar Aldabe <sup>3</sup> , German Rigau <sup>3</sup>
<b>Keywords:</b>	annotation, data, benchmarking
<b>Abstract:</b>	This deliverable describes the annotation efforts of year 2. We present the setup of the annotation tasks (both intra-document and cross-document), the guidelines used (for all four project languages), the resulting gold standard datasets and an evaluation of the current NewsReader system on the manually annotated gold standard datasets. We also present the annotation guidelines and baseline evaluation for the creation of cross-document timelines, for which the NewsReader team is organising a shared task in the SemEval competition.

## Table of Revisions

Version	Date	Description and reason	By	Affected sections
0.0	1 December 2014	Init	Marieke van Erp	all
0.1	15 December 2014	Added timelines description	Marieke van Erp	5
0.2	17 December 2014	Added Event coreference Evaluation	Piek Vossen	??
0.3	22 December 2014	Added NERC benchmark English	Rodrigo Agerri	??
0.4	26 December 2014	Added cross-document annotation guidelines and IAA	Anne-Lyse Minard	3, 4
0.5	19 January 2015	Added intra-doc annotation guidelines for Italian and Spanish	Manuela Speranza, Ruben Urizar, Anne-Lyse Minard	2
0.6	21 January 2015	Added temporal processing benchmark English	Anne-Lyse Minard	??
0.7	23 January 2015	Added English nominal coreference intra-document evaluation data	Rodrigo Agerri	??
0.8	29 January 2015	Added English NED intra-document evaluation	Itziar Aldabe	??
0.9	29 January 2015	Internal Review	Egoitz Laparra, Ruben Urizar, Itziar Aldabe	All
0.9	31 January 2015	Check by coordinator	VUA	-
1.1	19 June 2015	Update of the description of the datasets	Anne-Lyse Minard	4
1.1	20 July 2015	Check by coordinator	VUA	-



## Executive Summary

This deliverable describes the annotation efforts of the second year of the NewsReader project. This deliverable provides an update to Deliverable D3.3.1 Annotated Data v1. This update entails adaptations of the guidelines, in particular those of the attribution element (which was previously called factuality), the expansion to Spanish, Italian and Dutch and their accompanying datasets as well as our progress on cross-document annotation, both for events as well as for timelines.

The intra-document guidelines that were created for English (NewsReader technical report NWR-2014-2. Version 4.1 (Feb 2014)) were used as a mold for the Spanish, Italian and Dutch guidelines (and published as technical reports NWR-2014-6, NWR-2014-7 and NWR-2014-8 respectively). This deliverable will only highlight the changes particular to each language with respect to the English guidelines.

This deliverable also describes the data that was annotated for the four project languages. In order to create a balanced corpus and to facilitate cross-lingual benchmarking, the decision was taken to translate the 120 English Wikinews articles that were chosen for the English benchmarking effort into the other three project languages. These were then annotated and aligned to the English text.

As part of the cross-document annotation effort, we have gone one step beyond cross-document event annotation, as presented in Deliverable 3.3.1, and have added timeline annotations to the NewsReader corpus. The timelines that were generated from the 120 Wikinews articles served as gold standard for the SemEval-2015 Task 4: TimeLine: Cross-Document Event Ordering (pilot task)<sup>1</sup>.

The three steps in our annotation effort (intra-document event annotation, cross-document event annotation and cross-document timeline annotation) are necessary steps to structure and organise data in such a way that story lines can be distilled from the sources.

---

<sup>1</sup><http://alt.qcri.org/semEval2015/task4/>



# Contents

<b>Table of Revisions</b>	<b>3</b>
<b>1 Introduction</b>	<b>9</b>
<b>2 Intra-document Annotation</b>	<b>10</b>
2.1 Updates of the intra-document guidelines for English . . . . .	10
2.1.1 Headlines . . . . .	10
2.1.2 Entity type PRODUCT . . . . .	10
2.1.3 Factuality (Attribution) . . . . .	11
2.1.3.1 The <i>certainty</i> attribute . . . . .	11
2.1.3.2 The <i>polarity</i> attribute . . . . .	11
2.1.3.3 The <i>time</i> attribute . . . . .	11
2.1.3.4 The <i>special_cases</i> attribute . . . . .	12
2.1.3.5 Examples of attribution values annotation . . . . .	12
2.1.3.6 No attribution annotation . . . . .	12
2.2 Italian and Spanish annotation task . . . . .	13
2.3 Intra-document Annotation guidelines for Italian and Spanish . . . . .	15
2.3.1 Contractions of prepositions and definite articles (Italian’s articulated prepositions) . . . . .	15
2.3.2 Modal verbs . . . . .	16
2.3.3 Clitics . . . . .	16
2.3.4 Null subject . . . . .	17
2.4 Alignment guidelines for Italian and Spanish . . . . .	18
2.5 Intra-document Annotation guidelines for Dutch . . . . .	20
<b>3 Cross-document Annotation Task</b>	<b>20</b>
3.1 Entities . . . . .	21
3.1.1 Entity mentions . . . . .	21
3.1.2 Entity instances . . . . .	21
3.2 Events . . . . .	21
3.2.1 Event mentions . . . . .	21
3.2.2 Event instances . . . . .	22
3.3 Annotation task for English . . . . .	23
3.3.1 Phase I: Entity annotation . . . . .	23
3.3.2 Phase II: Event annotation . . . . .	24
3.4 Annotation task for Dutch, Italian and Spanish . . . . .	24
<b>4 Data</b>	<b>25</b>
4.1 Overview of the data . . . . .	25
4.2 Inter-Annotator Agreement in Intra-Document Annotation . . . . .	26
4.2.1 English . . . . .	26

---

4.2.2	Italian . . . . .	26
4.2.3	Spanish . . . . .	26
4.2.4	Dutch . . . . .	26
4.3	Inter-Annotator Agreement in cross-document annotation for English . . .	27
4.4	Description of the datasets . . . . .	28
4.4.1	English . . . . .	30
4.4.2	Dutch . . . . .	31
4.4.3	Italian . . . . .	32
4.4.4	Spanish . . . . .	33
4.4.5	Cross-lingual data . . . . .	33
<b>5</b>	<b>Timelines</b>	<b>34</b>
5.1	SemEval-2015 Task 4. TimeLine: Cross-Document Event . . . . .	34
5.2	NWR Timelines Dataset . . . . .	36
5.3	Outcomes . . . . .	36
<b>6</b>	<b>Conclusions and Future Work</b>	<b>37</b>



# 1 Introduction

This deliverable describes the annotation efforts of the second year of the NewsReader project. The goal of the NewsReader project<sup>2</sup> 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 WP04). 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. This deliverable is an update of Deliverable D3.2.1 Annotated Data v1. This update entails adaptations of the guidelines, in particular those of the attribution element (which was previously called factuality), the expansion to Spanish, Italian and Dutch and their accompanying datasets as well as our progress on cross-document annotation, both for events as well as for timelines.

In Y1, a core dataset of 120 English Wikinews articles was defined for the creation of the NewsReader gold standard annotated data set and guidelines for both intra-document and cross-document event annotation were defined. In Y2, the English gold standard annotation of the 120 articles was completed, and the articles were translated by professionals to the other three project languages, Spanish, Italian and Dutch. This ensured access to non-copyrighted articles in all project languages on the same topics, and even the option to compare the results of the NewsReader pipeline in the different languages at a finegrained level.

As part of the benchmarking effort, an evaluation of the NewsReader NLP pipeline has been undertaken and is presented in the deliverable D4.2.3.

In Year 1, we experimented with cross-document event annotation on the ECB+ (Cybulska and Vossen, 2014). The results from these experiments led the project to take the cross-document annotation one step further in Y2 and attempt cross-document timeline annotation. This annotation was done in the framework of the SemEval challenge<sup>3</sup>, providing the project with an opportunity to share the created datasets directly with the research community and obtain feedback on them. As the SemEval campaign is still ongoing, the results were presented at NAACL-HLT<sup>4</sup> in June.

This deliverable is structured as follows. In Section 2, we describe the updates to the intra-document annotation guidelines (NewsReader technical report NWR-2014-2. Version 4.1 (Feb 2014)). Followed by the update to the cross-document annotation guidelines in Section 3. In Section 4, we describe the dataset that was used in the annotation tasks as well as an overview of the annotation efforts into the different project languages. In Section 5, the timeline annotation for SemEval-2015 Task 4: TimeLine: Cross-Documents Event Ordering (pilot task) is described.

This deliverable rounds off with a conclusion and pointers for future work in Section 6.

---

<sup>2</sup><http://www.newsreader-project.eu>

<sup>3</sup><http://alt.qcri.org/semeval2015/>

<sup>4</sup><http://naacl.org/naacl-hlt-2015/>

## 2 Intra-document Annotation

In this section, we detail the updates to the intra-document guidelines for English as put forward in D3.3.1, as well as the intra-document guidelines for the other project languages.

### 2.1 Updates of the intra-document guidelines for English

During the annotation process, some issues were encountered that led us to update the intra-document guidelines for English. The two main changes concern the definition of the entity class and the annotation of attribution (previously called factuality). The changes are incorporated in an update of the annotation guidelines, published as technical report: Sara Tonelli, Rachele Sprugnoli, Manuela Speranza and Anne-Lyse Minard (2014) *News-Reader Guidelines for Annotation at Document Level*. NWR-2014-2-2. Version FINAL (Aug 2014). Fondazione Bruno Kessler.

#### 2.1.1 Headlines

In addition to the annotation of the first 5 sentences of each document, we decided to also annotate the headlines. The headline annotation gives information about the main event of the news (or main topic). The headlines should be annotated with event and entity mentions, `has_participant` and `refers_to` relations. The temporal relations do not need to be annotated because they can be obtained through the events in the document corefering to the events of the headline.

ex: *[Apple]*<sub>ENTITY</sub> *[unveils]*<sub>EVENT</sub> *[iPod nano]*<sub>ENTITY</sub>

#### 2.1.2 Entity type PRODUCT

We defined a new entity type, PRODUCT, which substituted ARTIFACT to include a wider spectrum of entities.

Product is anything that can be offered to a market that might satisfy a want or need<sup>5</sup>. This includes facilities (i.e. buildings, airports, highways, bridges, etc. as well as other structures and real estate improvements), vehicles (i.e. physical devices primarily designed to move an object from one location to another), weapons (i.e. physical devices primarily used as instruments for physically harming or destroying other entities), food (both human-made and produced by plants), products (including also abstract products such as software), functionalities (or features) of products, services, and trademarks (i.e. elements used for the public recognition of a company, for example logo).

Examples: vehicles, browser, internet access, trademark

---

<sup>5</sup>Definition taken from Wikipedia: [http://en.wikipedia.org/wiki/Product\\_\(business\)](http://en.wikipedia.org/wiki/Product_(business)).

### 2.1.3 Factuality (Attribution)

A profound study on factuality in text carried out at VUA led to new insights into how factuality related values should be annotated. The term factuality was replaced by the term attribution values since we attribute statements to sources in NewsReader and do not make any claims about their factual status. Attribution values of an event include the time it took place, the certainty of the source about it, and whether it is confirmed or denied (polarity). The adaptation of the scheme was led by VUA with the contribution of FBK and EHU. The first observations and insights into how to annotate attribution values were published in van Son *et al.* (2014). This work also presents the first steps towards representing world views by combining attribution and opinions.

In this section we describe the new defined attributes and we give examples of special cases.

**2.1.3.1 The *certainty* attribute** It expresses how certain the source about an event is: **certain**, **probable** and **possible**. Probable and possible events are typically marked in the text by the presence of modal verbs or modal adverbs:

Markers of probability: *probably, likely, it's probable, it's likely*

Markers of possibility: *possibly, it's possible, maybe, perhaps, may, might, could*

The certainty of events is based on textual properties. We follow the guidelines from FactBank<sup>6</sup> to distinguish between POSSIBLE and PROBABLE events. The idea behind the distinction is that an event can be possibly true or possibly not true at the same time, but something cannot be probably true and probably not true at the same time.

**2.1.3.2 The *polarity* attribute** It captures the distinction between affirmative and negative statements. Its values are POS for events with positive meaning (i.e. in most of the affirmative sentences), NEG for events with negative meaning (i.e. in most of the negative sentences), and UNDERSPECIFIED when it is not possible to specify the polarity of an event.

**2.1.3.3 The *time* attribute** It specifies the time an event took place or will take place, i.e. the semantic temporal value of an event. Its values are NON\_FUTURE for present and past events, FUTURE for events that will take place and UNDERSPECIFIED when the time of an event cannot be deduced.

In the specific case of reported speech, the value of the time attribute is always related to the time of utterance and not to the time of writing (i.e. when the utterance is reported). For instance, *leave* in “*John said he would leave for Scotland*” is annotated as FUTURE (because John made a statement about the future) even if, at the time of writing, the leaving might have already taken place.

---

<sup>6</sup>[http://www.cs.brandeis.edu/~roser/pubs/fb\\_annotGuidelines.pdf](http://www.cs.brandeis.edu/~roser/pubs/fb_annotGuidelines.pdf)

**2.1.3.4 The *special\_cases* attribute** It captures if the statement has some special status that influences its attribution: general statement (GEN), main clause of a conditional construction (COND\_MAIN\_CLAUSE) or if clause of a conditional construction (COND\_IF\_CLAUSE). The default value of this attribute is NONE.

Events that are properties should be marked as general statement. Properties should be distinguished from events that are true in the present but have a time span that covers also some portion of the past and of the future.

**2.1.3.5 Examples of attribution values annotation** We call *attribution values* of an event the information concerning when it took place, the certainty of the source about it, and whether it is confirmed or denied. The *attribution values* consist of the value of attributes certainty, polarity, time and special\_cases.

*The president forgot to inform the cabinet.*

predicate	certainty	polarity	time	special_cases
<i>forgot</i>	CERTAIN	POS	NON_FUTURE	NONE
<i>inform</i>	CERTAIN	NEG	NON_FUTURE	NONE

*I don't remember, maybe Obama was born in 1961.*

predicate	certainty	polarity	time	special_cases
<i>remember</i>	CERTAIN	NEG	NON_FUTURE	NONE
<i>born</i>	POSSIBLE	POS	NON_FUTURE	NONE

*John does not know whether Mary came.*

predicate	certainty	polarity	time	special_cases
<i>know</i>	CERTAIN	NEG	NON_FUTURE	NONE
<i>came</i>	POSSIBLE	UNDERSPECIFIED	NON_FUTURE	NONE

*If we pollute our planet, future generation will suffer*

predicate	certainty	polarity	time	special_cases
<i>pollute</i>	UNDERSPECIFIED	POS	FUTURE	COND_IF_CLAUSE
<i>suffer</i>	CERTAIN	POS	FUTURE	COND_MAIN_CLAUSE

**2.1.3.6 No attribution annotation** For event mentions referring to actions that are not really used as events in the text (i.e. they do not refer to a specific event and they are not anchored in time), attribution should not be annotated.

*Volkswagen did not say how much the XL1 costs to build.*

predicate	certainty	polarity	time	special_cases	comment
<i>say</i>	CERTAIN	NEG	NON_FUTURE	NONE	
<i>costs</i>	-	-	-	-	no attribution annotation
<i>build</i>	-	-	-	-	no attribution annotation

## 2.2 Italian and Spanish annotation task

The corpora used to build a benchmark for Italian and Spanish (as well as Dutch) consist of translations of the English corpus. The alignment between the source corpus (in English) and the corpus in the target language (Italian, Spanish and Dutch) has been done at the sentence level.

We took advantage of the alignment with English corpus and experimented on projecting the English intra-document annotation to the other languages. We provided annotators with files containing both sentences in Italian or Spanish aligned to the sentences in English. The annotation tool used (CAT) allows to visualize the annotation done for English when annotating the Italian or Spanish corpora. Figure 1 shows the annotation task in CAT. In the text panel of the interface, each English sentence (1) is followed by its translation in Italian or in Spanish (2). The instances that had been annotated previously within the English annotation task are displayed in the interface (4).

The annotation of Italian or Spanish sentences consists of four main steps:

- identification and annotation of the extent of a markable (as described in the NewsReader Guidelines);
- alignment of a markable to the corresponding markable in the original English text. This is performed through an attribute of type reference link, the `markId_English` attribute (3), that has been added to all text consuming markables in Italian or Spanish. The value of the `markId_English` attribute is filled through drag and drop of the corresponding markable annotated in English. If there is no equivalence of a ES/IT markable in English, annotators must create also the needed instances and relations (as for the English intra-doc annotation task);
- check of the relations (REFERS\_TO, TLINK, CLINK, HAS\_PARTICIPANT, SLINK), the instances (EVENT, ENTITY and empty TIMEX3) and the attributes of the markables (certainty, polarity, timex type, timex value, etc.) imported automatically into Italian or Spanish thanks to the reference links between ES/IT and English markables;
- annotation of missing relations.

Using this method we speed up the annotation task for Italian and Spanish, making around 3 times faster than the English annotation task.

File Task Markable Relation Statistics Help

Corpus Confirm Markable Delete Selection Task Selection ITNewsReader: EVENT\_MENTION

Markables Empty Tags

4

ENTITY

- corporate stake in GM Europe
- United States
- Barack Obama
- General Motors
- GM creditors
- money
- debt
- Canada
- Ontario
- Canada + Ontario
- Canada + Ontario + US
- Germany
- GM Europe
- Shareholders

EVENT

- do
- speech
- present
- reorganization
- plan
- announcement
- file
- bankruptcy
- protection
- describe
- be 1
- make
- explain
- propose
- state
- intent
- be 2

Markable Attributes

Current Extent: Casa Bianca

Markable type: ENTITY\_MENTION

3 markId\_English: the White House

lang: IT

head:

syntactic\_type:

comment:

mark\_fix\_id:

\*unsaved values

Save

1 2

Barack\_Obama\_presenta\_piano\_di\_salvataggio\_dopo\_la\_dichi

S0 Barack Obama presents rescue plan after GM declaration of bankruptcy

S1 Barack Obama presenta piano di salvataggio dopo la dichiarazione di fallimento di GM

S2 June 01, 2009

S3 1 giugno 2009

In a televised speech from the White House at 16:00 UTC today, President of the United States Barack Obama presented a reorganization plan following the 12:00 UTC announcement by General Motors that it had filed for bankruptcy and Chapter 11 protection from its creditors, the largest bankruptcy of a U.S. manufacturing company.

S4

In un discorso televisivo dalla Casa Bianca alle 16:00 UTC di oggi, il Presidente degli Stati Uniti Barack Obama ha presentato un piano di riorganizzazione a seguito dell'annuncio, avvenuto alle ore 12:00 UTC, da parte di General Motors riguardo all'istanza di fallimento presentata e alla richiesta di tutela dai creditori prevista dal Chapter 11, il più grande fallimento di una società di produzione statunitense.

S5

Describing the problem with the company as one that had been "decades in the making."

S6 Obama explained the rationale behind his proposed reorganization plan for General Motors.

Nel descrivere il problema della società come un guaio " che si preparava da decenni",

S7 Obama ha spiegato le ragioni alla base del piano di riorganizzazione proposto per General Motors.

He stated that his intent was not to "perpetuate" the bad business decisions of the past.

Figure 1: Visualization of the annotation task in CAT

## 2.3 Intra-document Annotation guidelines for Italian and Spanish

For the annotation of the Spanish and Italian corpora we adopted the NewsReader guidelines defined for English (Tonelli *et al.* (NWR2014-2-2)).

In this section we describe only the extensions needed to adapt them to the specific morpho-syntactic features of Italian and Spanish. The revision and adaptation of the annotation guidelines for events is based on the It-TimeML guidelines (Caselli *et al.* (2011)) and on the Spanish TimeML guidelines (Saurí *et al.* (2009, 2010); Saurí (2010)), while the revision and adaptation of the annotation guidelines for entities is based on the I-CAB guidelines (Magnini *et al.* (2006)).

### 2.3.1 Contractions of prepositions and definite articles (Italian’s articulated prepositions)

In the annotation of entity mentions and time expressions in English, prepositions are excluded from the extent while articles are included (e.g. *to [the family]*, *in [the next months]*). This is problematic for Italian and Spanish which, unlike English, have contractions of simple prepositions and definite articles. This phenomenon, which is common to many prepositions in Italian (e.g. *di*, *a*, *da*, *in*, *su*) and includes both singular and plural (e.g. *al* vs. *agli*) and both masculine and feminine (e.g. *al* vs. *alla*), is limited in Spanish to two contractions, *al* and *del*, in which the prepositions *a* or *de* respectively merge with the masculine singular definite article *el*.

Ex-IT: ***al*** *governo degli Stati Uniti* ‘to the US government’

Ex-ES: ***al*** *gobierno de los Estados Unidos* ‘to the US government’

Ex-IT: ***dal*** *5 novembre* ***al*** *10 dicembre* ‘from November 5 to December 10’

Ex-ES: ***del*** *5 de noviembre* ***al*** *10 de diciembre* ‘from November 5 to December 10’

Based on the above mentioned related work, we decided that these contractions should not be split but treated as single units in the annotation process. In particular:

- ENTITY MENTIONS: following the I-CAB guidelines, they should be included in the extent;
- TIMEX3s: following It-TimeML and Spanish TimeML, they should not be included in the extent; when a time expression is introduced by a contraction, this is usually to be marked as a temporal SIGNAL.

Ex-IT: ENTITY MENTION [***al*** *governo degli Stati Uniti*] ‘(to) the US government’

Ex-ES: ENTITY MENTION [***al*** *gobierno de los Estados Unidos*] ‘(to) the US government’

Ex-IT: SIGNAL+TIME EXPRESSION [***dal***] [*5 novembre*] [***al***] [*10 dicembre*] ‘from November 5 to December 10’

Ex-ES: SIGNAL+TIME EXPRESSION [*del*] [*5 de noviembre*] [*al*] [*10 de diciembre*] ‘from November 5 to December 10’

### 2.3.2 Modal verbs

According to the NewsReader guidelines for English (Tonelli *et al.* (NWR2014-2-2)), which are based on TimeML (Pustejovsky *et al.* (2003)), modal verbs are not annotated as events and the `modality` attribute is associated to the main verb (the value of the attribute is the token corresponding to the modal verb). On the other hand, the annotation of modals in NewsReader for Italian and Spanish follows It-TimeML and Spanish TimeML respectively: verbs expressing modality are themselves annotated as events (in particular, in the case of NewsReader, as events of type GRAMMATICAL); in addition, a GLINK (grammatical link) is created between the modal (source) and the main (target) verb (the `modality` attribute associated to the main verb is optional).

For instance, in the Spanish sentence *podemos jugar* ‘we can play’, two events must be annotated. Both verbs *podemos* ‘we can’ and *jugar* ‘to play’ are annotated as events, the verb conveying modality (*podemos*) being marked as an event of type GRAMMATICAL. Then a grammatical link is created between it and the verb *jugar* ‘to play’.

Ex-ES: [*podemos*] [*jugar*] ‘we can play’

Ex-IT: [*possiamo*] [*giocare*] ‘we can play’

Ex-ES: [*tendrán*] *que* [*mejorar*] ‘they will have to improve’

Ex-IT: [*dovranno*] [*migliorare*] ‘they will have to improve’

Ex-ES: [*podrías*] [*descansar*] ‘you could / might take a rest’

Ex-IT: [*potresti*] [*risposare*] ‘you could / might take a rest’

### 2.3.3 Clitics

For Spanish and Italian, we have devised specific guidelines to handle clitics, which do not exist in English.

Ex-IT: *Aveva già deciso di **parlargli*** ‘He had decided to talk to him’

Ex-ES: *Había decidido **hablarle*** ‘He had decided to talk to him’

As with contractions of prepositions and definite articles, we have decided to leave the annotation at token level in the case of clitics. In particular, in the case of a token composed of a verb (i.e. an event) and a clitic (i.e. a pronominal mention of an entity), the whole token will be annotated both as an entity and as an event. As it is important to distinguish the two annotated elements, the `head` attribute of the entity mention (see NewsReader Guidelines, section 3.2) is not optional for clitics as it is for all other types of entity mentions, and the `pred` attribute of the event mention (see NewsReader Guidelines,



section 5.2.1) is not optional either.

Ex-IT: EVENT MENTION: [parlargli], pred “parlare”

Ex-IT: ENTITY MENTION: [parlargli], head “gli”

Ex-ES: EVENT MENTION: [hablarle], pred “hablar”

Ex-ES: ENTITY MENTION: [hablarle], head “le”

As far as clitics in pronominal verbs are concerned, we have created specific guidelines for the different classes. Truly reflexive (the object of the action is the same as the subject) and reciprocal pronouns (expressing mutual action or relationship among the referents of a plural subject) are annotated as entities.

Ex-IT: [***Mi***] *sono ferito in montagna* ‘I hurt myself in the mountains’

Ex-ES: [***Me***] *lastimé en la montaña* ‘I hurt myself in the mountains’

Ex-IT: *Quelle due persone [**si**] amano* ‘Those two people love each other’

Ex-ES: *Esas dos personas [**se**] aman* ‘Those two people love each other’

In the case of benefactive (the focus refers to the person or thing an action is being done for) and pseudo-reflexive pronouns (which occur with intransitive pronominal verbs), we have no entity annotation.

Ex-IT: ***Mi** sono lavato le mani* ‘I washed my hands’

Ex-ES: ***Me** lavé las manos* ‘I washed my hands’

Ex-IT: ***Si** è mosso troppo velocemente* ‘He moved too fast’

Ex-ES: ***Se** movía demasiado deprisa* ‘He moved too fast’

When the Spanish “se” and the Italian “si” are used as impersonal pronouns (which corresponds to ‘one’, ‘you’, ‘we’, or ‘they’ in English) and as passive pronouns, they are not annotated.

Ex-IT: ***Si** dice che sia molto intelligente* ‘they/people say he is very smart’

Ex-ES: ***Se** dice que es muy inteligente* ‘they/people say he is very smart’

Ex-IT: *Da qui **si** vede il lago* ‘from here the lake can be seen’

Ex-ES: *Desde aquí **se** ve el lago* ‘from here the lake can be seen’

### 2.3.4 Null subject

As Italian and Spanish, unlike English, are null-subject languages (e.g. languages whose grammar permits to have clauses lacking an explicit subject), we devised specific guidelines for the annotation of null subjects. Follow four examples of null subjects in Spanish and in Italian.

Ex 1: Barack Obama gave a speech. He said that [...]  
Ex-ES: Barack Obama dio un discurso. [Ø] Dijo que [...]  
Ex-IT: Barack Obama fece un discorso. [Ø] Disse che [...]

Ex 2: We can [...]  
Ex-ES: [Ø] Podemos [...]  
Ex-IT: [Ø] Possiamo [...]

Ex 3: It rains [...]  
Ex-ES: [Ø] Lueve [...]  
Ex-IT: [Ø] Piove [...]

Ex 4: The explosion occurred at 9.30 PM. It resulted in eight injured.  
Ex-ES: La explosión ocurrió a la 9.30 pm. [Ø] Causó ocho heridos.  
Ex-IT: L'esplosione è avvenuta alle 21.30. [Ø] Ha provocato otto feriti.

Null subjects have finite verb forms as predicates. They should only be annotated if they refer to an existing entity instance (see examples 1 and 2 above). So, for instance, they should not be annotated in the case of impersonal constructions (see example 3) or if they refer to event instances (see example 4). Annotators should mark null subjects through the creation of an empty, non text-consuming ENTITY\_MENTION tag. The empty entity mention is linked through a REFERS\_TO relation to the corresponding entity instance (also all other types of relation can be created, following the guidelines for regular text consuming entity mentions). In addition, annotators should fill the free-text attribute “tag descriptor” with a human friendly name (such as a name or a pronoun) and the number of the sentence in which their predicate appears (e.g. *ObamaS6*).

As the NewsReader annotation of Italian and Spanish texts is done automatically by projecting the English annotation through manual entity mention alignment, it is not necessary to create any REFERS\_TO (or other) relations for null subjects (marked as empty entity mentions) when they have been aligned to English entity mentions. Due to translation divergences, there may be null subjects that only occur in Spanish or in Italian but not in English. In these cases, the REFERS\_TO relations and the other relations will have to be annotated manually.

## 2.4 Alignment guidelines for Italian and Spanish

Although the texts in the Spanish and Italian corpora were translations of the English ones, obviously word-to-word translation was not always feasible. Due to these translation divergences, it was not always possible to make a one-to-one alignment of Spanish or Italian markables with their corresponding English ones.

The translation divergences are countless, but next we give an account of some of the most frequent types of translation divergences that caused misalignments and we present

the main rule that the annotators should follow.

On the one hand, some English constructions may be translated into Spanish and Italian with a single word. For instance, several light verb constructions and verb idioms may have a one-word equivalent in Spanish or Italian.

Ex-ES: *Apple **has plans** to import* ⇒ *Apple **planea** importar* (lit. 'Apple plans to import')

Ex-IT: *Apple **has plans** to import* ⇒ *Apple **intende** importare* (lit. 'Apple plans to import')

Ex-ES: ***bringing** production [...] **to an end*** ⇒ *y **acabó** la producción [...]* (lit. 'and production was finished')

Ex-IT: ***bringing** production [...] **to an end*** ⇒ *y **interruppe** la produzione [...]* (lit. 'and production was finished')

Also, many verb-particle construction (phrasal verbs) are translated with a single word.

Ex-ES: *he **put off** his appearance before the press for an hour* ⇒ ***retrasó** en una hora su comparecencia ante la prensa*

Ex-IT: *he **put off** his appearance before the press for an hour* ⇒ ***posticipó** di un'ora la sua apparizione di fronta alla stampa*

Other English constructions may also have a one-word equivalent.

Ex-ES: *it **is believed to be worth...*** ⇒ *se **estima** en...* (lit. 'it is estimated in')

Ex-IT: *it **is believed to be worth...*** ⇒ *si **stima** in...* (lit. 'it is estimated in')

On the other hand, sometimes the translation for some English one-word verbs is also a multiword construction, for example, light verb constructions.

Ex-ES: *they **used*** ⇒ ***hicieron uso** de* (lit. 'they made use of')

Ex-IT: *they **used*** ⇒ ***fecero uso** di* (lit. 'they made use of')

Furthermore, some multiword expressions may be translated through verbal periphrases in Spanish or Italian, for example, the phrasal verb *call back* may be translated in Spanish with the periphrasis *volver a llamar* (lit. 'return to call') comprising two event verbs while there was a single one in English. Equally, the adverb *again* may also be translated with the periphrasis 'volver a + infinitive' (lit: 'return + infinitive').

Ex-ES: *Boeing allowed to **bid** for [...] contracts **again*** ⇒ *Boeing puede **volver a presentarse** a contratos de [...] (lit: 'Boeing can return to bid contracts [...])'*

In all these divergences, the alignment between source and target markables is made between the elements semantically closer (e.g. English noun *plans* with Spanish verb

*planea*) rather than between elements belonging to the same part of speech (e.g. a verb with a verb).

## 2.5 Intra-document Annotation guidelines for Dutch

With Dutch being a sister language of English, no major changes were necessary to adapt the intra-document annotation guidelines from English to Dutch, save for a section devoted to the Dutch adverb ‘er’. The word can carry a variety of meaning that can be classified into four types of use, namely locative, presentative, prepositional and quantitative (Bennis, 1986). For the NewsReader annotation, only the locative use is deemed important, for example in the sentence “Hij woont er al jaren” (He has lived there for years) ‘er’ is to be annotated as a mention of an entity of class LOCATION.

Furthermore, in the Dutch language compounding is more prevalent, which led to changes in some of the examples and difficulties of applying the word count rule to determine whether an entity mention is a NAM (proper name) or a NOM (common noun). Lastly, the annotators were made aware that the Dutch language contains more discontinuous predicates, which affects the event mention annotation layer.

## 3 Cross-document Annotation Task

Three partner institutions were involved in the NewsReader cross-document annotation task for English: FBK, EHU and VUA. As the leading institution for the annotation effort, FBK produced the annotation guidelines with the collaboration of other partners (Speranza and Minard (NWR2014-9)). After a training phase, in which FBK guided VUA and EHU in using the annotation tool (CROMER Girardi *et al.* (2014)), the annotation of English started with an agreement phase.

Annotation at corpus level consists of two main steps:

- cross-document entity coreference annotation of all entities annotated in the first 5 sentences and the headline of each file;
- cross-document entity and event coreference annotation starting from a set of seed entities. All the mentions that corefer to the seed entities should be annotated as well as the events in which the seed entities are participants.

For the identification of entity mentions, for the definition of their extent, and for the annotation of coreference, we use the NewsReader intra-document annotation guidelines (Tonelli *et al.* (NWR2014-2-2)).

Sections 3.1 and 3.2 describe the annotation of entity mentions and instances and the annotation of event mentions and instances.

In Section 3.3, we detail the annotation task for the English corpora, i.e. the different steps of the cross-document annotation, and in Section 3.4 we present the specific annotation task for the Dutch, Italian and Spanish corpora.

## 3.1 Entities

### 3.1.1 Entity mentions

As far as the extent of entity mentions is concerned, annotators should apply the same guidelines provided for the intra-document annotation (for entity mention extent, see section 3.1 of NWR-2014-2-2). The only exception is that CONJ-mentions (i.e. entity mentions connected by a coordinating conjunction) will not be annotated.

### 3.1.2 Entity instances

Each entity instance has the following attributes:

- class (compulsory): to be filled according to the guidelines provided for the intra-document annotation (see section 2.1 of NWR-2014-2-2 on Entity types);
- name (compulsory): to be filled according to the guidelines provided for the intra-document annotation (see section 2.4 of NWR-2014-2-2 on Tag Descriptor);
- short description (compulsory): a short description of the entity instance, whose aim is to distinguish it from other entity instances with similar names;
- external reference (compulsory): to be filled according to the guidelines provided for the intra-document annotation (see section 2.3 of NWR-2014-2-2);
- comment (optional).

## 3.2 Events

### 3.2.1 Event mentions

The annotation of event mentions is based on the intra-document annotation guidelines (for event mention extent, see section 5.1 of NWR-2014-2-2).

Event is used as a cover term to identify “something that can be said to obtain or hold true, to happen or to occur” (ISO TimeML Working Group, 2008). This notion can also be referred to as eventuality (Bach, 1986) including all types of actions (punctual or durative) and states as well.

Some events annotated following the NewsReader guidelines could not go on a timeline, for example because they didn’t happen (counter-factual events) or they are uncertain. In order to annotate only events potentially candidates to participate in a timeline, we have defined criteria based on the intra-document annotation Guidelines.

We annotate verbs —except if they are modified by a modal word—, nouns and pronouns. Adjectives generally express a property or attribute of an entity, and anchoring them in time is not simple. So adjectival events will not be annotated.

Events are classified according to semantic features. Those classified as “grammatical” are dependent on a content verb/noun and don’t have a time span, so they will not be

annotated. We have also decided to leave out cognitive events (i.e. events that describe mental states or mental acts).

The last criterion is based on the factuality and certainty of events. Counter-factual events will not be part of a timeline because they did not take place. Non-factual events are speculative events, so we do not know if they happen or not. If it is certain that they will happen (e.g. “the conference will take place on Monday”), they will be annotated. But if they are uncertain (e.g. “the conference may take place later”), we will not annotate them.

### 3.2.2 Event instances

CROMER event instances have the following attributes:

- class (compulsory): to be filled according to the guidelines provided for the intra-document annotation (see section 4.2 of NWR-2014-2-2);
- name (compulsory): to be filled according to the guidelines provided for the intra-document annotation (see section 4.1 of NWR-2014-2-2 on Tag Descriptor for Events);
- short description (compulsory): a short description of the event instance, whose aim is to distinguish it from other event instances with similar names;
- time (compulsory for punctual events): a date (maximum granularity day), following the TIMEX3 format (see Section 6.2.2 of NWR-2014-2-2); if not known, add the values XXXX-XX-XX, XXXX-XX or XXXX depending on granularity;
- begin (compulsory for durative events): the starting date of the event (maximum granularity day), following the TIMEX3 format (see Section 6.2.2 of NWR-2014-2-2); if not known, add the values XXXX-XX-XX, XXXX-XX or XXXX depending on granularity;
- end (if known, compulsory for durative events): the ending date (maximum granularity day), following the TIMEX3 format (see Section 6.2.2 of NWR-2014-2-2);
- external reference (compulsory): to be filled according to the guidelines provided for the intra-document annotation (see section 4.3 of NWR-2014-2-2);
- comment (optional).

In case of repeated events or grouped events the following rules should be applied to fill the time attributes:

- repeated events (e.g. “I go to work every morning”)
  - time: if known add the value of the set temporal expression (following the TIMEX3 format);

- begin (compulsory for repeated events): add the date of the first time the repeated event happened; if not known, add the values XXXX-XX-XX, XXXX-XX or XXXX depending on granularity;
- end: if known, add the date of the last time the repeated event happened;
- grouped events (e.g. "the explosions caused huge damage") are considered as if they were durations
  - begin (compulsory for grouped events): add the date of the event that happened chronologically first; if not known, assign the values XXXX-XX-XX, XXXX-XX or XXXX depending on granularity;
  - end: if known, put the date of the event that happened chronologically last;

### 3.3 Annotation task for English

The cross-document annotation is carried out using a tool called CROMER (CRoss-document Main Event and entity Recognition) (Girardi *et al.* (2014)).

#### 3.3.1 Phase I: Entity annotation

For each entity annotated in the first 6 sentences (included the headline) using CAT, the annotation task consists of the following steps:

- check if an entity instance already exists in CROMER and if not create it;
- assign the mention chains to the entity instance (if annotators find annotation errors in the import from CAT, they can correct them);
- find other mentions of the entity in the corpus (only in the first 6 sentences of each document annotated with CAT) and assign them to the entity instance.

We manually selected a set of relevant target entities that appeared in at least two different documents and were involved in more than two events. Each partner institution receives the list of the target entities to be annotated in each corpus.

For each selected entity the annotation task consists of the following steps:

- check if an entity instance already exists in CROMER and if not create it;
- assign the mention chains imported from CAT (which refer only to the first 6 sentences of a document) to the entity instance (if annotators find annotation errors in the import from CAT, they can correct them);
- find all other mentions of the entity in the corpus (i.e. in the remaining part of each document), annotate them with the correct extent (no attributes have to be annotated) and assign them to the entity instance.

### 3.3.2 Phase II: Event annotation

Annotators should annotate all events having as participant one of the seed entities annotated in phase I.

More specifically, the annotation task consists of the following steps:

- For each selected entity, get the list of all documents in which that entity is mentioned.
- For each document, identify all event mentions having the annotated entity as participant and annotate them as follows:
  - check if the event instance to which it refers already exists; if it does not exist, create it;
  - create a HAS-PARTICIPANT relation between the event instance (source) and the entity instance (target);
  - if the event mention is one imported from CAT, assign it to the event instance it refers to; otherwise annotate the mention with the correct extent (no attributes have to be annotated) and assign it to the event instance it refers to.

## 3.4 Annotation task for Dutch, Italian and Spanish

The cross-document annotation task for Dutch, Italian and Spanish is carried out semi-automatically profiting from the fact that texts are aligned at sentence level with English texts.

For Italian and Spanish, in addition to the alignment of mentions done within the intra-document annotation task, we have annotated and aligned the mentions annotated in English as part of the cross-document annotation task in the full document.

For Dutch the mentions annotated in the first six sentences have been aligned with their corresponding English mentions. In addition, the mentions annotated in English in the rest of the document have been annotated in Dutch, if present, and aligned.

Based on this alignment with English mentions, we have imported automatically the cross-document annotation to Italian, Dutch and Spanish texts.

In the corpora annotated by projection of the English annotation, some mentions are missing due to translation divergences. This requires a manual verification and completion of the cross-document annotation.

The alignment enables us to have a cross-lingual annotated corpus, where all the occurrences of one instance are linked in the 4 languages through a unique identifier.

For example, the entity “Alan Mulally” (with the id PER37406424840192289) is annotated in the corpus about “GM, Chrysler and Ford” in the 4 languages. The id of the instance makes it possible to access six mentions in English, six in Dutch, five in Italian and five in Spanish.

English: [...] *said* [[*the firm’s chief executive officer*], [*Alan Mulally*]].

Dutch: [...] *zei* [[*Alan Mulally*], *de chief executive officer van het bedrijf*]].



Italian: [...] *ha ribadito* [[l'amministratore delegato della società] [Alan Mulally]].  
Spanish: [...] *dijo* [[Alan Mulally], [director ejecutivo de la empresa]].

Similarly, the event of the announcement of the release of Mac OS X Leopard by Steve Jobs is annotated in the corpus about “Apple Inc.” in the 4 languages, and is accessible through the id SPE57304114754781218.

English: *Steve **announced** that Mac OS X Leopard would be released in 2007*  
Dutch: *Steve **meldde** dat Mac OS X Leopard wordt uitgebracht in 2007*  
Italian: *Steve ha **annunciato** il rilascio di Mac OS X Leopard nel 2007*  
Spanish: *Steve **annunció** que Mac OS X Leopard se lanzará al mercado en 2007*

## 4 Data

In this section we give an overview of the data annotated within the project. In Section 4.2 and 4.3 we describe the inter-annotator agreements computed at the beginning of the different annotation tasks. Afterwards an analysis of the annotated corpora in 4 languages is detailed in Section 4.4.

### 4.1 Overview of the data

As was mentioned in D3.3.1, it is important to the NewsReader project to be able 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. Therefore, we chose Wikinews<sup>7</sup> as our core corpus for the annotation effort.

Next to the 20 articles concerning Apple Inc. that were used for defining and finetuning the annotation guidelines in Y1, 10 more articles on the same topic were selected, as well as 30 articles concerning the Airbus and Boeing companies, 30 articles concerning three automobile manufacturers (General Motors, Chrysler and Ford) and 30 articles concerning the stock market published mainly during the 2008 financial crisis. As in Y1, the articles were selected in such a way that the corpus contains different articles that deal with the same topic over time (e.g. launch of a new product, discussion of the same financial indexes). This enables us to benchmark our cross-document event coreference modules, as well as to build cross-document time and story lines.

Since Wikinews does not contain enough overlapping articles between the four project languages to create balanced corpora around the same topics in the different project languages, the decision was made to translate the originally selected English Wikinews articles into Spanish, Italian and Dutch. For the translations, professional translation companies were hired that translated the articles in a sentence-by-sentence manner. The Dutch translations were checked by a NewsReader team member fluent in both English and the target language.

---

<sup>7</sup>[http://en.wikinews.org/wiki/Main\\_Page](http://en.wikinews.org/wiki/Main_Page)

## 4.2 Inter-Annotator Agreement in Intra-Document Annotation

### 4.2.1 English

In deliverable D3.3.1, Section 3.1.1, we reported on the inter-annotator agreement for all aspects (all markables, attributes and relations) of the intra-document annotation. In this section, we provide specific data on the inter-annotator agreement on attribution (previously called factuality), for which the guidelines have been updated.

We measured the inter-annotator agreement with the Dice’s coefficient on 97 event mentions (referring to 83 distinct event instances) annotated in 7 files. Each event mention was annotated with attribution values by two annotators.

The results are presented in Table 1. The agreement is over 0.90 for `certainty`, `time` and `polarity` attributes. For the `special_cases` attribute, the agreement is over 0.80, with main disagreements on “general statement” events.

	<code>certainty</code>	<code>time</code>	<code>special_cases</code>	<code>polarity</code>
Dice’s coefficient	0.94	0.90	0.84	0.94

Table 1: Inter-annotator agreement for attribution value annotation

### 4.2.2 Italian

The annotation of the Italian translation of the corpus through the alignment procedure was performed by an expert annotator who is a native speaker of Italian. As no other Italian speaker annotator was part of the consortium during the annotation phase it was impossible to compute data about inter-annotator agreement for Italian. However the data on inter-annotator agreement provided for English can be used as a reference as the guidelines followed for Italian do not differ from those for English, except for a small number of linguistic phenomena that are not present in English (see Section 2.3).

### 4.2.3 Spanish

The Spanish intra-document annotation was carried out by two native speakers of Spanish. One of the annotators had already taken part in the annotation of English documents and was in charge of training the new annotator as well as reviewing the latter’s annotation task. Besides, since Spanish annotation was mostly based on the English guidelines (see Section 2.3), inter-annotator agreement was not measured.

### 4.2.4 Dutch

The Dutch intra-document annotation effort was split over 9 different annotators each of which only took care of one layer at a time in the annotations.<sup>8</sup> This way, the annotators

<sup>8</sup>Some of the annotators took on more than one task

could focus on one part of the annotation without having to concern themselves with the entire annotation guidelines, speeding up training.

The tasks took place in two phases and were divided as follows:

#### **Phase 1: November 2014**

- Entity mentions and instances
- Event mentions and instances
- Temporal expressions
- Numeric expressions, signals and c-signals

#### **Phase 2: December-January 2014**

- Has participant relationships
- Attribution
- Temporal links
- Subordinated links, grammatical links and causal links

The event and entity mentions and instances were shared between two annotators, each annotating the entities and events in 60 articles. For all other tasks, the annotators worked on all 120 articles.

All annotators were asked to annotate three documents from the Apple corpus to compute the inter-annotator agreement against the gold standard created by one of the Dutch trainers. The results are presented in Table 2. Most reported scores are Dice's coefficient, except for the many to one relations in the last row, for which we report the global alpha score. As with the English annotation, the TLINKs provided the greatest challenge for the annotators. On the other markables and relations a high inter-annotator agreement is achieved.

### **4.3 Inter-Annotator Agreement in cross-document annotation for English**

We measured inter-annotator agreement on both mention extents and instances with the Dice's coefficient.

Three annotators each annotated a corpus of 30 documents starting from one seed entity, i.e. they annotated entity coreferences referring to the seed entity and the events in which the seed entity was a participant. This annotation was done in the full text. The corpus was composed of articles from WikiNews about *Apple Inc.* and the seed entity was *iPhone 4*. We first computed the agreement scores by pairs of annotators and then the macro-average on the pairwise scores.

Markables		
	macro-average (markable)	macro-average (token)
SIGNAL	0.67	0.67
VALUE	0.83	0.83
C-SIGNAL	1.00	1.00
TIMEX3	0.97	1.00
EVENT_MENTION	0.78	0.92
ENTITY_MENTION	0.82	0.85
Relations one to one		
	macro-average	macro-average
TLINK	0.39	0.32
CLINK	1.00	1.00
HAS_PARTICIPANT	0.77	0.61
GLINK	1.00	0.33
SLINK	1.00	0.70
Relations many to one		
	global alpha	
REFERS_TO	0.45	

Table 2: Inter-annotator agreement Dutch gold standard annotation

The scores are given in Table 3. The results are satisfactory, with the agreement macro-average above 0.80 for entity mentions and coreference relations (REFERS\_TO). For event mentions and event instances annotation the agreement is above 0.65. One reason for this difference is that for entity mentions and entity coreferences annotation the Guidelines are similar to the one used for intra-document annotation, while for event mentions and instances annotation the guidelines are specific to the cross-document annotation task.

	macro-average
ENTITY (product)	0.81
EVENT (speech-cognitive or other)	0.66
REFERS_TO	0.84
EVENT INSTANCES	0.68

Table 3: Inter-annotator agreement on the cross-document annotation task

#### 4.4 Description of the datasets

An overview of the annotated corpora and the different levels of annotation is represented in Figure 2.

In this section we present a quantitative description of each corpus for each language.

In Table 4, Table 6, Table 8 and Table 10 statistics about the intra-document annotation are presented. The intra-document annotation was done on the first six sentences of each

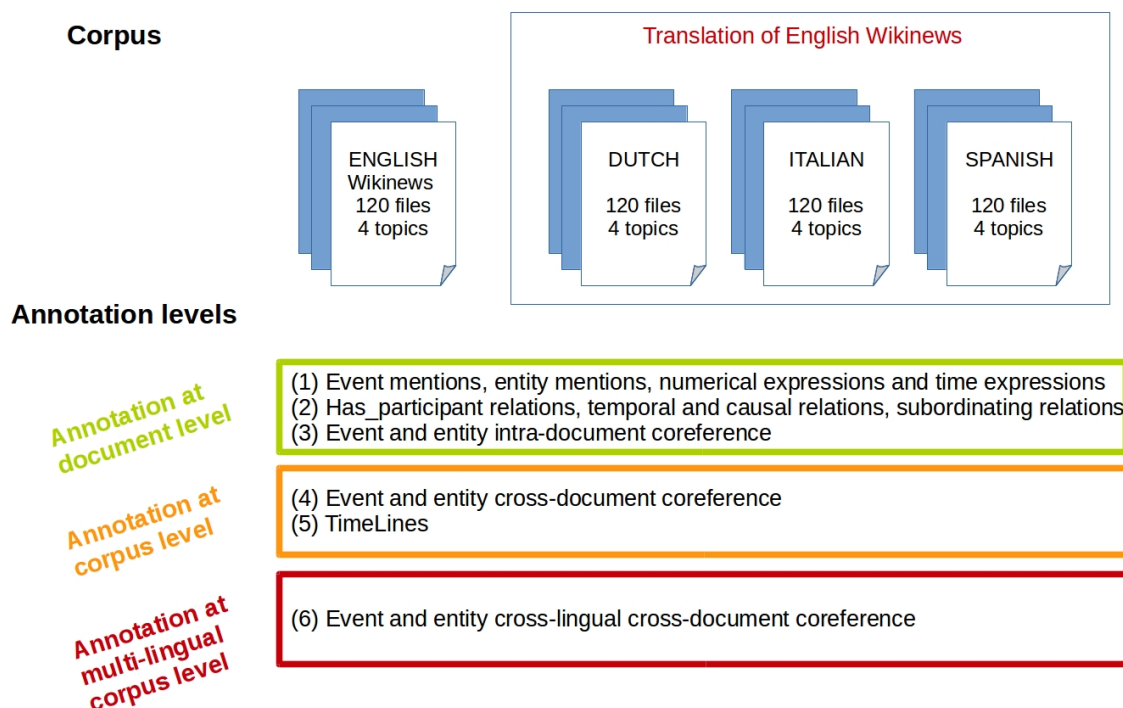


Figure 2: Overview of the NewsReader annotated corpora

file, including the headline and the document creation time. For each language, we first give some general figures about the corpora, namely, the number of files, the number of sentences and the number of tokens. Then we indicate the number of markables of each type (entity and event mentions, entity and event instances, time expressions, etc.) and the number of relations (temporal links, causal links, has participant, etc.). For Italian and Spanish, we also specify the number of empty entity mentions (i.e. null subjects) that have been annotated.

In Table 5, Table 7, Table 9 and Table 11, statistics are given about the cross-document annotation of our data. The cross-document annotation was done on the same files as intra-document annotation, starting from the existing annotation, but on full texts.

The cross-document annotation task was composed of two subtasks:

- cross-document coreference annotation of all the entities annotated in the first six sentences during the intra-document annotation task;
- cross-document annotation of a set of pre-selected target entities in full texts and of the events in which these entities were involved.

For the first subtask we give the number of entity instances and for the second subtask we provide the number of target entities, the number of event instances and the number of intra-document and cross-document corefering events.

Finally, in Table 12 we present quantitative data about the cross-lingual annotation level of our corpora.

#### 4.4.1 English

	Stock	GM	Airbus	Apple	Total	
# files	30	30	30	30	120	
# sentences	180	178	180	179	717	
# tokens	3,332	3,636	3,590	3,423	13,981	
markables	EVENT_MENTION	521	577	519	479	2,096
	EVENT	455	413	447	402	1,717
	ENTITY_MENTION	454	773	742	821	2,790
	ENTITY	285	286	367	354	1,292
	TIMEX3	164	144	101	116	525
	VALUE	144	129	97	48	418
	SIGNAL	93	74	64	60	291
	C-SIGNAL	9	3	12	5	29
relations	REFERS_TO	735	696	810	742	2,983
	TLINK	373	529	408	479	1,789
	CLINK	22	4	14	10	50
	GLINK	42	75	55	37	209
	HAS_PARTICIPANT	364	507	604	503	1,978
	SLINK	25	105	68	41	239

Table 4: Intra-document annotation in the Wikinews corpus in English

	Stock	GM	Airbus	Apple	Total
# files	30	30	30	30	120
# sentences	459	429	446	463	1797
# tokens	9,916	10,063	9,909	10,343	40,231
# entity instances	139	180	230	236	785
# target entities	13	12	13	6	44
# event instances	215	237	350	156	958
# corefering events	28	40	77	17	162
# cross-doc corefering events	7	7	14	7	35

Table 5: Cross-document annotation in the Wikinews corpus in English

## 4.4.2 Dutch

	Stock	GM	Airbus	Apple	Total	
# files	30	30	30	30	120	
# sentences	180	178	180	179	717	
# tokens	3,582	3,691	3,815	3,559	14,647	
markables	EVENT_MENTION	372	332	425	381	1,510
	EVENT	306	286	328	290	1,210
	ENTITY_MENTION	530	677	771	751	2,729
	ENTITY	350	333	364	278	1,325
	TIMEX3	156	125	89	110	480
	VALUE	148	126	99	39	412
	SIGNAL	99	67	65	60	291
	C-SIGNAL	21	11	19	10	61
relations	REFERS_TO	651	612	690	563	2,516
	TLINK	433	407	358	318	1,516
	CLINK	12	8	18	10	48
	GLINK	33	11	43	35	122
	HAS_PARTICIPANT	510	512	399	509	1,930
	SLINK	28	68	62	53	211

Table 6: Intra-document annotation in the Wikinews corpus in Dutch

	Stock	GM	Airbus	Apple	Total
# files	30	30	30	30	120
# sentences	459	429	446	461	1,795
# tokens	10,502	10,159	10,554	10,756	41,971
# entity instances	119	136	196	160	602
# target entities	13	12	13	6	44
# event instances	194	209	312	139	840
# corefering events	30	36	70	16	150
# cross-doc corefering events	4	3	13	7	27

Table 7: Cross-document annotation in the Wikinews corpus in Dutch

## 4.4.3 Italian

	Stock	GM	Airbus	Apple	Total	
# files	30	30	30	30	120	
# sentences	180	178	180	179	717	
# tokens	3,856	4,003	4,096	3,721	15,676	
markables	EVENT_MENTION	555	597	547	507	2,206
	EVENT	507	469	468	448	1,892
	ENTITY_MENTION	465	672	756	813	2,706
	ENTITY	282	295	392	370	1,339
	TIMEX3	161	131	101	114	507
	VALUE	143	125	92	54	414
	SIGNAL	86	65	50	50	251
	C-SIGNAL	11	3	14	7	35
relations	REFERS_TO	752	718	811	777	3,058
	TLINK	366	480	387	464	1,697
	CLINK	27	3	19	12	61
	GLINK	59	104	77	62	302
	HAS_PARTICIPANT	353	421	576	501	1,851
	SLINK	27	90	64	39	220
# empty mentions	11	11	13	21	56	

Table 8: Intra-document annotation in the Wikinews corpus in Italian

	Stock	GM	Airbus	Apple	Total
# files	30	30	30	30	120
# sentences	473	445	450	477	1845
# tokens	11,045	11,041	11,180	11,274	44,540
# entity instances	134	156	217	226	733
# target entities	13	12	13	6	44
# event instances	201	197	341	141	880
# corefering events	25	43	74	19	161
# cross-doc corefering events	5	6	14	7	32

Table 9: Cross-document annotation in the Wikinews corpus in Italian



#### 4.4.4 Spanish

	Stock	GM	Airbus	Apple	Total	
# files	30	30	30	30	120	
# sentences	180	178	180	179	717	
# tokens	3,929	4,087	3,997	3,830	15,843	
markables	EVENT_MENTION	560	598	571	519	2,248
	EVENT	503	598	535	452	2,088
	ENTITY_MENTION	509	738	794	810	2,851
	ENTITY	305	291	389	361	1,346
	TIMEX3	152	131	92	111	486
	VALUE	144	121	92	47	404
	SIGNAL	114	60	57	49	280
	C-SIGNAL	18	3	10	9	40
relations	REFERS_TO	755	723	837	758	3,073
	TLINK	532	583	512	559	2,186
	CLINK	32	4	14	11	61
	GLINK	70	80	84	76	310
	HAS_PARTICIPANT	503	557	561	532	2,153
	SLINK	33	110	58	37	238
# empty mentions	21	44	23	26	114	

Table 10: Intra-document annotation in the Wikinews corpus in Spanish

	Stock	GM	Airbus	Apple	Total
# files	30	30	30	30	120
# sentences	469	431	445	475	1820
# tokens	11,339	11,059	10,986	11,404	44,788
# entity instances	136	169	202	208	715
# target entities	13	12	13	6	44
# event instances	207	208	340	144	899
# corefering events	23	40	52	19	134
# cross-doc corefering events	6	7	11	7	31

Table 11: Cross-document annotation in the Wikinews corpus in Spanish

#### 4.4.5 Cross-lingual data

In Table 12 we give some statistics about the cross-lingual aspect of our data. We computed the number of event and entity instances annotated and linked in the 4 languages through a unique identifier. For example, in the corpus about “Stock Market” there are 174 event instances linked in the 4 languages and 118 entity instances.

In the second part of the table, we provide the number of shared event instances by language pairs. To help understanding these numbers we also indicate the number of event instances annotated in English. For example, in the corpus about “Airbus and Boeing”, 298 event instances are both annotated in English and Dutch, and 318 are shared in Italian and Spanish.

	Stock	GM	Airbus	Apple	Total
# event instances cross-lingual	174	171	289	119	742
# entity instances cross-lingual	118	129	192	155	585
# event instances EN-NL	194	209	312	139	840
# event instances EN-IT	201	197	341	141	880
# event instances EN-ES	199	208	326	144	877
# event instances NL-IT	185	177	306	125	780
# event instances NL-ES	181	185	294	130	778
# event instances IT-ES	189	190	318	133	830
# event instances EN	215	237	350	156	958

Table 12: Cross-lingual annotation in the Wikinews corpus (EN: English, NL: Dutch, IT: Italian, ES: Spanish)

## 5 Timelines

As part of going beyond document-based evaluations, the NewsReader team set up a Timeline evaluation in the context of the SemEval-2015: Semantic Evaluation Exercises.<sup>9</sup> The task, “TimeLine: Cross-Document Event Ordering”, was accepted as a pilot task in order to gauge the state-of-the-art in cross-document timeline creation. In this section, we detail the task description, explain the annotation steps, the resulting corpus and the outcomes of the SemEval timeline task.

### 5.1 SemEval-2015 Task 4. TimeLine: Cross-Document Event

The TimeLine task revolves around ordering events across documents in a timeline about a particular entity. For this task, the English WikiNews articles that were already annotated in the intra- and cross-document annotation task are utilised.

For each of the sub-corpora (Apple, Airbus-Boeing, GM-Chrysler-Ford, Stockmarket), up to 15 entities that are central to the corpus are defined. These are entities that occur in multiple documents and which play a role in different events. For each of the entities, the events in which the entity is a participant in the Arg0 or Arg1 propbank roles (often agent and patient, respectively) are selected and ordered chronologically. In case the event cannot be anchored to a particular date, the date is left blank.

<sup>9</sup><http://alt.qcri.org/semeval2015/>

The timelines are represented in a tab separated format in which the first column denotes the ordering, the second one the time anchor, and all following columns are co-referring events. In the events, the documentID and the sentence number are encoded for easy retrieval of the event in text. A timeline may thus look like the following:

**iTunes**

1	2003	11778-3-launch	11778-4-launch
2	2007	11778-4-pass	
3	2008-01	11778-7-hold	
4	2008-02	11778-2-pass	11778-5-pass
4	2008-02	11778-3-accounts_for	

The timelines were generated semi-automatically from the cross-document manually annotated text and verified by NewsReader team members.

As this is a new task, the threshold to participate in the task is kept low by offering participants different levels of difficulty in creating the timelines, varying from a subtrack in which the participants are provided with the event mentions and only need to order the events (without temporal anchoring) and a subtrack in which only raw text is provided in which events need to be detected, ordered and anchored temporally. The four tracks offered have the following setup:

**Track A (main track):**

input data: raw text

output: full TimeLines (ordering of events and assignment of time anchors)

**Subtrack A:**

input data: raw text

output: TimeLines consist of just ordered events (no assignment of time anchors)

**Track B:**

input data: texts with manual annotation of event mentions

output: full TimeLines (ordering of events and assignment of time anchors)

**Subtrack B:**

input data: texts with manual annotation of event mentions

output: TimeLines consist of just ordered events (no assignment of time anchors)

The full task description and annotation guidelines can be found at <http://www.newsreader-project.eu/publications/technical-reports/> as the following techreports:

- Anne-Lyse Minard, Manuela Speranza, Bernardo Magnini, Marieke van Erp, Itziar Aldabe, Ruben Urizar, Eneko Agirre and German Rigau. *TimeLine: Cross-Document Event Ordering. SemEval 2015 – Task 4*. NWR-2014-10. Fondazione Bruno Kessler.
- Anne-Lyse Minard, Alessandro Marchetti, Manuela Speranza, Bernardo Magnini, Marieke van Erp, Itziar Aldabe, Ruben Urizar, Eneko Agirre and German Rigau. *TimeLine: Cross-Document Event Ordering. SemEval 2015 – Task 4. Annotation Guidelines*. NWR-2014-11. Fondazione Bruno Kessler.

## 5.2 NWR Timelines Dataset

We used as dataset the Wikinews corpus annotated as part of the project. The subcorpus about “Apple Inc.” had been used as trial data and the other 3 subcorpora as evaluation data. Based on the cross-document annotation done in these 4 subcorpora, we have built timelines about seed entities. The timelines were automatically built using the cross-document annotation for events, as well as time anchor attributes of event instances and the has\_participant relations. Afterwards the timelines were manually corrected.

In Table 13 we describe the two datasets. The data used for the evaluation consists of 90 documents and 38 timelines.

Dataset	Trial corpus	Evaluation corpora			
	Apple	Airbus-Boeing	GM-Chrysler-Ford	Stock-market	Total eval dataset
# documents	30	30	30	30	90
# sentences	464	446	430	459	1,335
# timelines	6	13	12	13	38
length of timelines	29.3	21.1	20.5	16.9	20.0
# unique event mentions	188	331	305	264	900
average # docs by timeline	5.8	6.2	5.7	9.1	7.1

Table 13: Description of the Timelines Dataset

## 5.3 Outcomes

Whilst 29 teams had signed up for the TimeLines shared task, only 4 teams submitted results in the end. A paper detailing the results of this effort is in preparation, and the official results of the challenge are presented in Table 14.

Participant	CORPUS1	CORPUS2	CORPUS3	F1 score	TOTAL	
	F1	F1	F1		Precision	Recall
Track A						
SPINOZAVU_1	4.07	5.31	0.42	3.15	7.95	1.96
SPINOZAVU_2	2.67	0.62	0.00	1.05	8.16	0.56
WHUNLP_1	8.31	6.01	6.86	<b>7.28</b>	14.10	4.90
SubTrack A						
SPINOZAVU_1	1.20	1.70	2.08	<b>1.69</b>	6.70	0.97
SPINOZAVU_2	0.00	0.92	0.00	0.27	13.04	0.14
Track B						
GPLSIUA_1	22.35	19.28	33.59	<b>25.36</b>	21.73	30.46
GPLSIUA_2	20.47	16.17	29.90	22.66	20.08	26.00
HeidelToul_1	19.62	7.25	20.37	17.03	20.11	14.76
HeidelToul_2	16.50	10.94	25.89	18.34	13.58	28.23
SubTrack B						
GPLSIUA_1	18.35	20.48	32.08	<b>23.15</b>	18.90	29.85
GPLSIUA_2	15.93	14.44	27.48	19.18	16.19	23.52
HeidelToul_1	12.23	14.78	16.11	14.42	19.58	11.42
HeidelToul_2	13.24	15.88	21.99	16.67	12.18	26.41

Table 14: Official results of TimeLine task

## 6 Conclusions and Future Work

In this deliverable, we described the annotation efforts of the second year of the NewsReader project. There are three main parts to this deliverable, namely updates of the English intra-document guidelines and translations of the guidelines for the other three project languages, Spanish, Italian and Dutch. Then we described the data as well as the intra- and cross-document annotation task and results. In the third part, we described the TimeLine SemEval shared task that the NewsReader team is organising as part of furthering the state-of-the-art in cross-document information extraction and towards cross-document storyline extraction.

This year’s efforts further consolidated the English annotation guidelines, and we branched out to the other project languages. The annotation effort for English is completed save for a few documents in the cross-document annotation effort. As we took on the TimeLine SemEval task, we needed to allocate time to annotate cross-document time lines, which was an additional task that was not foreseen originally. However, through the CROMER tool, this could be done in a semi-automatic fashion, which sped up the task greatly. The TimeLine task provided us with insights toward creating storylines, as well as a timelines dataset with 38 cross-document timelines that we intend to use in further research as well as share with the research community.

Furthermore, for Dutch, Italian and Spanish, the intra-document annotation effort is completed, delivering a four-language aligned corpus of 120 news articles with a rich

linguistic annotation.

In year 3, we will also perform a cross-lingual evaluation effort, to investigate in detail whether the NLP modules in the different languages extract the same information from the texts, and also to see if they can aid each other. We intend to organise another shared task concerning cross-lingual extraction, as well as around story line extraction for which we build upon the timeline dataset. Furthermore, we will release the annotated datasets, which we intend to accompany with a journal paper submission.

## References

- Emmon Bach. The algebra of events. *Linguistics and Philosophy*, 9:5–16, 1986.
- Hans Bennis. *Gaps and Dummies*. Amsterdam University Press, 1986. 2005 reprint.
- Tommaso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta, and Irina Prodanof. Annotating Events, Temporal Expressions and Relations in Italian: the It-TimeML Experience for the Ita-TimeBank. In *Linguistic Annotation Workshop*, pages 143–151, 2011.
- Agata Cybulska and Piek Vossen. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26-31 2014.
- Christian Girardi, Manuela Speranza, Rachele Sprugnoli, and Sara Tonelli. Cromer: A tool for cross-document event and entity coreference. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26-31 2014.
- 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>.
- Bernardo Magnini, Emanuele Pianta, Christian Girardi, Matteo Negri, Lorenza Romano, Manuela Speranza, Valentina Bartalesi Lenzi, and Rachele Sprugnoli. I-CAB: the Italian Content Annotation Bank. In *Proceedings of LREC 2006 - 5th Conference on Language Resources and Evaluation*, 2006.
- James Pustejovsky, José M. Castaño, Robert Ingria, Roser Saurí, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *New Directions in Question Answering*, pages 28–34, 2003.
- R Saurí, O Batiukova, and J Pustejovsky. Annotating events in spanish. timeml annotation guidelines (version tempeval-2010). Technical report, Barcelona Media. Technical Report BM 2009-01, 2009.
- R Saurí, E Saquete, and J Pustejovsky. Annotating time expressions in spanish. timeml annotation guidelines (version tempeval-2010). Technical report, Barcelona Media. Technical Report BM 2010-02, 2010.
- Roser Saurí. Annotating temporal relations in catalan and spanish. timeml annotation guidelines. (version tempeval-2010). Technical report, Barcelona Media. Technical Report BM 2010-04, 2010.
- Manuela Speranza and Anne-Lyse Minard. NewsReader Guidelines for Cross-Document Annotation. Technical report, Fondazione Bruno Kessler, NWR2014-9.

Sara Tonelli, Rachele Sprugnoli, Manuela Speranza, and Anne-Lyse Minard. NewsReader Guidelines for Annotation at Document Level. Technical report, Fondazione Bruno Kessler, NWR2014-2-2.

Chantal van Son, Marieke van Erp, Antske Fokkens, and Piek Vossen. Hope and Fear: Interpreting Perspectives by Integrating Sentiment and Event Factuality. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26-31 2014.