

EventDNA: Annotation Guidelines for
Entities and Events in Dutch News Texts
(v1.0)

LT3 Technical Report - LT3 19-01

Camiel Colruyt, Orphée De Clercq and Véronique Hoste

Camiel.Colruyt@UGent.be, Orphee.DeClercq@UGent.be, Veronique.Hoste@UGent.be

LT3 - Language and Translation Technology Team
Department of Translation, Interpreting and Communication
Ghent University

<https://www.lt3.ugent.be/>

28/01/2019

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 3 |
| 1.1 | The annotation process | 4 |
| 1.2 | Using WebAnno | 5 |
| 2 | Annotating entities | 8 |
| 2.1 | Introduction to entities | 8 |
| 2.2 | Mention extent and mention level type | 9 |
| 2.2.1 | Named entity mentions | 9 |
| 2.2.2 | Nominal entity mentions | 11 |
| 2.2.3 | Pronominal entity mentions | 12 |
| 2.3 | Entity type | 13 |
| 2.3.1 | Person entities (PER) | 13 |
| 2.3.2 | Organization entities (ORG) | 14 |
| 2.3.3 | Location entities (LOC) | 14 |
| 2.3.4 | Miscellaneous entities (MISC) | 15 |
| 2.3.5 | Metonymy | 16 |
| 2.4 | Individuality | 16 |
| 2.5 | Entity coreference | 17 |
| 2.6 | Annotating entities in WebAnno | 17 |
| 3 | Annotating events | 21 |
| 3.1 | Events and event mentions | 21 |
| 3.1.1 | Identifying event mentions | 21 |
| 3.2 | Event types | 23 |
| 3.3 | Description of event types | 23 |
| 3.3.1 | Conflict events | 25 |

| | | |
|----------|--|-----------|
| 3.3.2 | Journalism events | 26 |
| 3.3.3 | Politics events | 27 |
| 3.3.4 | Unknown events | 27 |
| 3.4 | Event arguments | 28 |
| 3.5 | Event realis properties | 30 |
| 3.5.1 | Positive/negative | 30 |
| 3.5.2 | Tense | 31 |
| 3.5.3 | Modality | 31 |
| 3.6 | Event prominence | 33 |
| 3.6.1 | Cases with one main event | 34 |
| 3.6.2 | Cases with multiple main events | 35 |
| 3.7 | Annotating events in WebAnno | 36 |
| 4 | Annotating IPTC Media Topics | 40 |
| 4.1 | Choosing the correct topics | 40 |
| 4.2 | Annotating media topics in WebAnno | 42 |
| A | EventDNA event type table | 45 |

Chapter 1

Introduction

Every news article contains mentions of real-world news events, like elections, natural disasters, armed attacks, etc. What we call the *main event* or *foreground event* of an article is the new information which leads to the article being written. It can usually be found in the title or the lead paragraph of the article. For example, in a fictional article entitled “*Russian spies arrested in England*,” the main event would be the arrest of the Russian spies.

The article can refer to other events related to the main event in order to place it in a context. We call these *background events*. Each event description also includes additional pieces of information called *event arguments*. The main category of arguments are the *entities* that participate in the event, like the *Russian spies* above. Besides entities, arguments can include information on time and place (like *in England*).

EventDNA is an annotation project that aims to identify and label foreground events in news text, as well as the entities that participate in those events. The end result of the annotation process is a dataset that can be used to train machine learning systems to automatically extract the same information from new texts. This document introduces and defines the EventDNA schema, and guides the annotator through the annotation process.

The characteristics of the EventDNA annotation task are as follows. We focus on the main and background events in hard news.¹ Our data consists

¹Hard news, in this context, means news of a factual, informative nature concerning world events. It corresponds to the scope of the #NewsDNA project. ‘Hard’ news is not a strict

of Dutch-language newspaper articles acquired from a number of Flemish media companies.

A newspaper consists of a title and a text body. The body is usually structured so that the most important information is described in the first paragraph, which is often called the lead. The rest of the article body contains extra information, background and details. That way, the reader can get the core of the message immediately and read the rest of the article if he or she wants to. An assumption of EventDNA annotation is that the article's main event can always be found in the title or lead paragraph of the article. For this reason, we only present the annotator with the title and lead of each article.

1.1 The annotation process

The concepts of EventDNA annotation (entities, events, coreference etc.) are introduced in the following chapters. This section outlines the order in which the annotator should undertake the different annotation steps. Annotating each layer over the whole document separately from the next minimizes the chance for error: the annotator (1) can focus on one concept at a time and (2) is invited to go over the text a number of times, perhaps finding errors made in previous steps. We therefore encourage annotators to follow the following steps. Each of these steps will be thoroughly explained in the following sections.

1. Annotate entities

- (a) First pass: mark all entities in the document, without assigning features yet. Understand the content of the text.
- (b) Second pass: go over each entity from the beginning and assign its features. Allow yourself to reconsider decisions from the first pass.
- (c) Third pass: perform entity coreference by linking annotations that refer to the same entity.

2. Annotate events

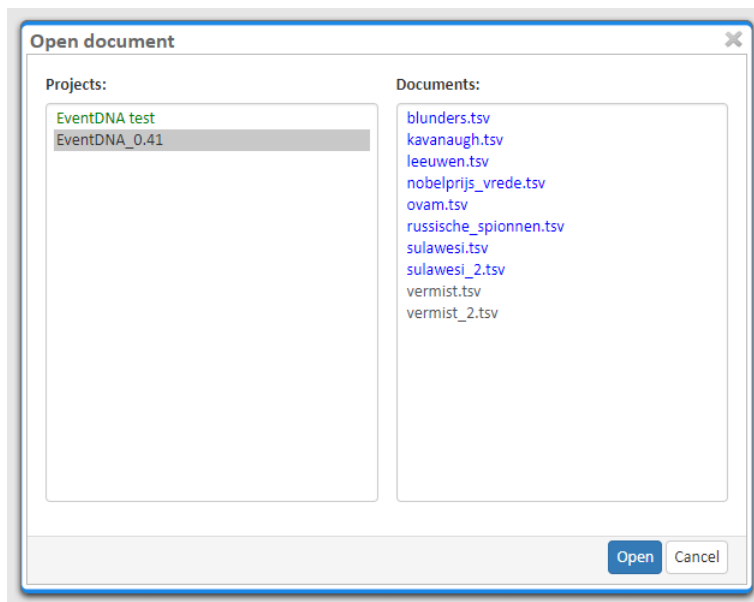
definition but generally excludes news about culture, sport, showbusiness and lifestyle.

- (a) Fourth pass: mark all events in the document, without assigning features yet. At this stage, you decide with events to annotate, without going into detail. It is useful to turn off the visualization of entity annotations.
 - (b) Fifth pass: go over each event and fill in its features.
 - (c) Sixth pass: perform event coreference by linking annotations that refer to the same event.
3. Perform IPTC Media Topic annotation. Going over the list of media topics, select which are appropriate and write out the list in the Media Topic annotation layer.

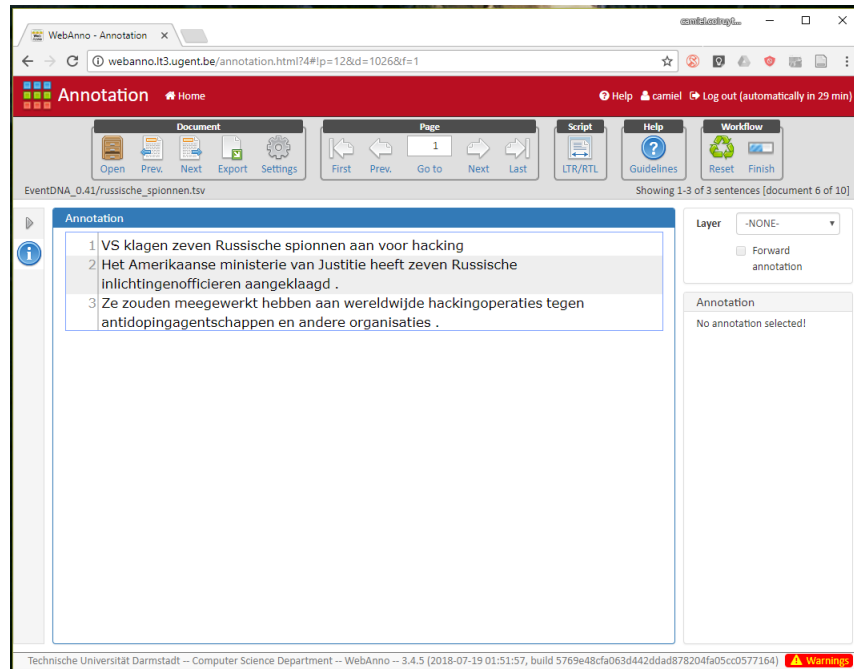
1.2 Using WebAnno

We annotate articles using the WebAnno software. This section provides a practical guide to viewing documents in WebAnno and introduces the concepts of annotation layers. The annotation of entities and events themselves are explained in sections 2.6 for entities and 3.7 for events.

When opening up WebAnno, you are prompted to select a project and a document to annotate.



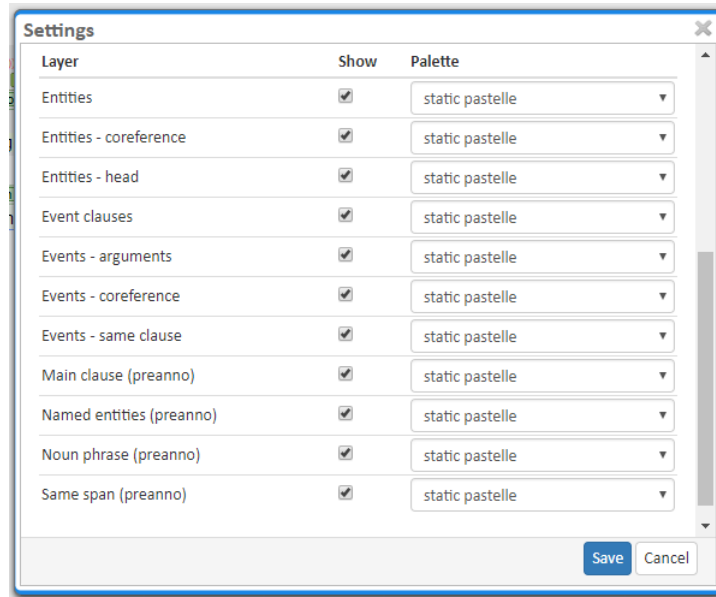
The document annotation screen looks like this. The items on the top menu bar are clear for the most part. The *LTR/RTL* button switches the text to run from left to right, to right to left and vice-versa; we do not need to use it.



In WebAnno, there are two types of annotations: span annotations and relation annotations. **Span annotations** are annotations that span across one or more words of the text. Each type of annotation lives on what WebAnno calls a **layer**. In EventDNA, there is a layer for *entity* annotations, one for *event* annotations and one for IPTC Media Topic annotation - as well as separate layers for entity heads, event arguments and discontinuous event clauses (we return to these three layers later in this document).

Relation annotations are annotations that exist between span annotations. In EventDNA, we use two types of relation annotation: one for coreference between entity annotations and one for coreference between events. Coreference means the two span annotations in the relation refer to the same thing: for example, *Emmanuel Macron* and *the president of France* refer to the same person.

Through the settings menu, you can change some visual options as well as toggle which annotation layers are visible. By default, all layers are visible.



Chapter 2

Annotating entities

2.1 Introduction to entities

All news articles contain mentions of **entities**. What we call entities are objects in the real world that a string of text can refer to, like people, companies, places, etc. For instance, example 1 contains references to two persons, one location (*Amerikaanse*, used as an adjective) and one organization.

- (1) De [*Amerikaanse*]^{LOC-NAM} president [*Donald Trump*]^{PER-NAM} is van plan om volgende week dinsdag toch zijn 'State of the Union' te houden, ondanks het verzoek van de voorzitter van het [*Huis van Afgevaardigden*]^{ORG-NAM}, [*Nancy Pelosi*]^{PER-NAM}, om uitstel wegens de shutdown.

The text strings that refer to entities are called **entity mentions**. Mentions can be long or short and can overlap with one another. Multiple mentions in a sentence can refer to the same entity.

In the EventDNA project, we annotate *specific* mentions that point at three possible entity types: **persons** (abbreviated PER), **organizations or companies** (ORG) and **places or locations** (LOC). Additionally, we assign the label **miscellaneous** (MISC) to entities that are relevant to our task but that we can't easily label PER, LOC or ORG. These mentions can be names (like *Narendra Modi*, *PVDA* or *Vlaanderen*), demonyms and adjectives based on locations (like *Vlaming*, *Vlaams*), but also nominal and pronominal constructions (like *de minister-president van India* or *hij*). In addition to a type,

we also annotate the individuality and coreference features of each entity. These features are described in the following sections. The last section in this chapter will describe how to annotate entities practically in the WebAnno interface.

Specific and non-specific entities

Entity mentions usually refer to specific objects in the world, as in the examples mentioned above. These are **specific entities**. Other mentions refer instead to a class of entities without pointing at real objects. These are **nonspecific entities**, as shown in example 2. In EventDNA, **we only annotate specific entities**.

- (2) a. Vroeger dachten [*mensen*] dat de Aarde rond was.
 - b. [*Men*] zegt veel maar doet weinig.
 - c. [*Sommige bedrijven*] deden het goed tijdens de crisis, [*anderen*] niet.
- The entities in these examples are not specific and are therefore not annotated.

2.2 Mention extent and mention level type

The **mention extent** is the single string of text that indicates an entity. It can not be split up across multiple strings. Entities can be referred to by a name (like *Barack Obama*), a noun phrase (*the 44th president of the US*) or a pronoun (*he, him, his*). This feature is called the **mention level type**.

All examples in this section are marked with their entity type for reference. Type will be discussed in the next section; for now, remember that the possible types are person (PER), organization (ORG) and location (LOC).

2.2.1 Named entity mentions

Named entity mention encompass references made using proper names, abbreviations, acronyms, aliases, etc. The extent of a named mention is only the name. It does not include modifiers like the definite article *the*, titles, or relative clauses. Place names used as a modifier, like the adjectives

Belgische or *Napolitaanse*, are also considered named entity mentions that refer to the location.

- Typical named entity mentions are simple names: *[Albert Einstein]*^{PER-NAM}, *[Credit Suisse]*^{PER-NAM}, *[Zuid-Afrika]*^{LOC-NAM}
- Generational or ordinal suffixes are considered part of the name: *[Robert Downey Jr.]*^{PER-NAM}, *[Leopold III]*^{PER-NAM}
- Articles are not annotated as part of the mention: *de [Sovjet-Unie]*^{LOC-NAM}, *de [VS]*^{LOC-NAM}
- Place names used as adjectives are considered named mentions referring to that location: *[Duits]*^{LOC-NAM}, *[Argentijns]*^{LOC-NAM}
- Coordinated entities are annotated separately: *[Karl Marx]*^{PER-NAM} en *[Trotsky]*^{PER-NAM}, *[Pierre]*^{PER-NAM} en *[Marie Curie]*^{PER-NAM}, *de [VN]*^{PER-NAM} en *de [Verenigde Staten]*^{PER-NAM}

News text often introduces entities with titles and descriptions. These descriptions can often be thought of as nominal entity mentions that are coreferent with the named mention. However, **when a named entity appears alongside a title or a description, we only annotate the named entity mention.**

- (3) a. De stad *[Antalya]*^{LOC-NAM} in *[Turkije]*^{LOC-NAM}
→ *De stad* can be thought of as a nominal mention that is coreferent with *Antalya*, but we do not not annotate it.
- b. De *[Turkse]*^{LOC-NAM} stad *[Antalya]*^{LOC-NAM}
→ The nominal mention *de Turkse stad*, which we do not annotate, contains the named mention *Turkse*, which we do annotate.
- c. Bondskanselier *[Merkel]*^{PER-NAM}
- d. *[Angela Merkel]*^{PER-NAM}, de *[Duitse]*^{LOC-NAM} bondskanselier
- e. Voormalig premier *[Rajoy]*^{PER-NAM}
- f. De *[Popocatépetl]*^{LOC-NAM}, een actieve vulkaan en met 5426 meter hoogte de op een na hoogste berg van *[Mexico]*^{LOC-NAM}.

- g. [*Ralph Vary Chamberlin*]^{PER-NAM}, de beroemde [*Amerikaanse*]^{PER-NAM} entemoloog
- h. Het prachtige [*Barcelona*]^{LOC-NAM}

2.2.2 Nominal entity mentions

Nominal entity mentions refer to an entity through a noun phrase. The extent of a mention like this includes articles and modifiers (adjectives, relative clauses, etc.) on either side of the mention. Apart from the extent, we also annotate the head of nominal mentions. This is the head noun of the construction. If you would remove all words from the mention extent (except the definite article) but kept the head, the construction would still be grammatically correct.

- (4) a. [*De verdachte*]^{PER-NOM}
- b. [*Het gekelderde bedrijf*]^{ORG-NOM}
- c. [*De schrijver van het boek*]^{PER-NOM}
- d. [*De man die me de weg vroeg*]^{PER-NOM}
- e. [*Een 19e-eeuws kasteel bij de [Zweedse] hoofdstad [Stockholm]*]^{LOC-NAM}
- f. [*de [Duitse] bondskanselier*]^{PER-NOM}
- g. [*Een jonge vrouw uit het [West-Vlaamse] [Staden]*]^{PER-NOM} kwam om het leven.
- h. Onder [*de slachtoffers*] waren ook [*vier [Belgen.]*]^{PER-NOM}
- i. [*Twaalf anderen*]^{PER-NOM} raakten gewond.
 → Syntactically, *twaalf* modifies *anderen*, which is therefore the head.

A useful rule of thumb for determining the extent of a nominal mention is that you should be able to replace the marked entity by a pronoun, as in example 5.

- (5) a. [*De voormalige premier van Spanje*]^{PER-NOM} laat weten dat hij afstapt als hoofd van zijn partij.
- b. [*Hij*]^{PER-PRON} laat weten dat hij afstapt als hoofd van zijn partij.

It is often difficult to decide whether a mention is nominal or named. We use the following rule of thumb: capitalized mentions are names, non-capitalized mentions are nominal constructions.

- (6) a. De [*Kamer van Koophandel*]^{ORG-NAM} vs. [*de kamer van koophandel*]^{ORG-NOM}. In the case of de [*kamer van Koophandel*]^{ORG-NAM}, we assume the writer made a mistake and meant to capitalize both words.
- b. De [*Katholieke Kerk*]^{ORG-NAM} vs. [*de Katholieke kerk*]

When a named mention incorporates another name, such that the other name is an official part of the surrounding name, only the surrounding name must be annotated.

- (7) Het [*London Philharmonic Orchestra*]^{ORG-NAM}
- (8) Maar: het [*Russisch*]^{LOC-NAM} **leger**^{ORG-NOM}

Demonyms

Demonyms are annotated twice. They refer once to the people they indicate, and once to the location. The reference to people is a PER-NOM annotation; as a nom, it can be longer than the demonym itself. The reference to the place is a LOC-NAM annotation.

- (9) a. [*De hongerige [Jemenieten]*]^{LOC-NAM}^{PER-NOM}
- b. [*Een [Amerikaan]*]^{LOC-NAM}^{PER-NOM}

2.2.3 Pronominal entity mentions

Entities expressed by a **pronoun** must also be tagged. Pronouns can be anaphoric, meaning they refer to some previous entity in the sentence, or they may stand alone. Remember that we annotate on the sentence level: if a pronoun has an antecedent, it occurs in the same sentence. For us, however, the distinction does not matter: we annotate all pronouns as separate entities, regardless of their anaphoric character. This means that if a pronoun occurs in a sentence alongside its antecedent, and both are the argument of an event, both must be annotated as such. The identity

between them is established by co-reference links in another stage of the annotation process.

- (10) *[President [Trump]]* noemt *[Saipov]* een beest en dreigt ermee *[hem]* naar *[Guantanamo]* te sturen.

We annotate only the personal and possessive Dutch pronouns:

- Persoonlijke voornaamwoorden als onderwerp: *ik, jij, je, u, hij, zij, ze, het, wij, we, jullie, u, zij, ze*
- Persoonlijke voornaamwoorden als voorwerp: *mij, me, jou, je, u, hem, haar, het, ons, jullie, je, u, hen, hun, ze*
- Bezittelijke voornaamwoorden: *mijn, jouw, je, uw, zijn, haar, zijn, ons, onze, jullie, uw, hun*

2.3 Entity type

All entities belong to a certain **type**: PER, ORG, LOC or MISC. We only annotate entities that belong to one of these types.

2.3.1 Person entities (PER)

Person entities are humans or groups of humans referred to by a name, a noun phrase or a pronoun.

- We annotate living and deceased persons: *[Vladimir Poetin]^{PER-NAM}, [Zhou Youguang]^{PER-NAM}, [Einstein]^{PER-NAM}*
- We do not annotate fictional or animal characters: *[Kuifje]* en zijn hond *[Bobby]*, *[Andy Serkis]^{PER-NAM}* werd de stem van kapitein *[Haddock]*
- Nominal and pronominal mentions are also annotated: *[de twee Amerikanen]^{PER-NOM}, [Morgan Freeman]^{PER-NAM} is [[zijn]^{PER-PRON} favoriete acteur]^{PER-NOM}*

2.3.2 Organization entities (ORG)

Organization entities are associations with a certain fixed organizational structure. It is typical of organizations that they can change members without changing identity. Typical examples of organizations are:

- Companies and businesses: *[Lockheed Martin]^{ORG-NAM}, [Nike]^{ORG-NAM}, [Zalando]^{ORG-NAM}, [Google]^{ORG-NAM}, [Pizza Mario]^{ORG-NAM}, [Twitter]^{ORG-NAM}, [Sony]^{ORG-NAM}, restaurant *[De Lange Muur]^{ORG-NAM}**
- Governments, their related institutions, and non-governmental organizations: *de [[Amerikaanse]^{LOC-NAM} regering]^{ORG-NOM}, de [Senaat]^{ORG-NAM} (but: *[de senaat]^{ORG-NOM}*), het *[ministerie van landbouw]^{ORG-NOM}, de [Verenigde Naties]^{ORG-NAM}, dierenrechtenorganisatie [Animal Rights]^{ORG-NAM}, de [NASA]^{ORG-NAM}**
- Schools, universities, hospitals: *de [Universiteit Gent]^{ORG-NAM}, [California Institute of Technology]^{ORG-NAM}, het AZ [Sint-Maria Halle]^{ORG-NAM}*
- Sports teams and music bands: *de [Rode Duivels]^{ORG-NAM}, de [Beatles]^{ORG-NAM}*
- Nominal and pronominal uses are also annotated: *[de grootste supermarkt van [Canada]^{LOC-NAM}]^{ORG-NOM} sluit zeven van [haar]^{ORG-PRON} winkels*

It is often unclear whether a given entity should be tagged as an organization or a group of people. As a rule of thumb, if the mention refers to an organization with a defined structure, whose members can change without changing the identity of the entity, then we annotate it as an organization.

- (11) Organizations: *[de politie]^{ORG-NOM}, [het [Russische]^{LOC-NAM} leger]^{ORG-NOM}, het [London Philharmonic Orchestra], de [raad van bestuur]*
- (12) Groups of persons: *[de leden van de task force]^{PER-NOM}, [de muzikanten]^{PER-NOM}, [de protesteerders]^{PER-NOM}*

2.3.3 Location entities (LOC)

Every name that refers to a place gives rise to a LOC entity. These are geographical entities such as areas, landmasses or bodies of water. Country

names also count as LOC entities unless they are used in a metonymical sense. Cases of metonymy are described in section 2.3.5.

Typical examples of locations are:

- Country and place names: *[België]^{LOC-NAM}, de [Verenigde Staten]^{LOC-NAM}, [Azië]^{LOC-NAM}, het [Midden-Oosten]^{LOC-NAM}, [Barcelona]^{LOC-NAM}, [Siberië]^{LOC-NAM}, de [Atlantische Oceaan]^{LOC-NAM}, de [Stille Zuidzee]^{LOC-NAM}*
- Descriptions of places, borders, lines and points: *[de nieuwe fabriek]^{LOC-NOM}, [LOC-NOM][de grens tussen [België]^{LOC-NAM} en [Frankrijk]^{LOC-NAM}]*
- The adjectives derived from these names (separately from any annotation made on the head noun): *[de [Belgische]^{LOC-NAM} regering]^{ORG-NOM}, [het [Amerikaanse]^{LOC-NAM} leger]^{ORG-NOM}, [Midden-Oosterse] smaken*
- Place names coordinated with a compass direction: *[Oost-Azië]^{LOC-NAM}, [West-Siberië]^{LOC-NAM}*
- When the direction is separate from the place, two annotations must be made, because the two mentions do not refer to the same entity: *[het oosten van [België]^{LOC-NAM}]^{LOC-NOM}, [het midden van [Nieuw-Caledonië]^{LOC-NAM}]^{LOC-NOM}*
- Planets and space objects: *[de maan]^{LOC-NOM}, [de zon]^{LOC-NOM} en de planeet [Aarde]^{LOC-NAM}*
- Demonyms are annotated twice, once as a LOC and once as a PER: *[de jonge [Amerikaan]^{LOC-NAM}]^{PER-NOM}, [een hele bende [Russen]^{LOC-NAM}]^{PER-NOM}.*

2.3.4 Miscellaneous entities (MISC)

The goal of EventDNA entity annotation is to capture all news-relevant entities in an article. An entity is relevant if it plays an important role in the news story. Sometimes, an entity is found that is very relevant to the news story but does not fall cleanly into the PER, LOC, ORG categories. In those cases, the MISC type can be used as a fallback option. The MISC type should be applied sparingly and carefully.

- (13) *[‘Vluchtelingenboot’ Aquarius]^{MISC-NAM} kent moeilijke tocht naar [Spanje]^{LOC-NAM}*
- (14) *Het afkondigen van het ‘[Stockholm Akkoord]^{MISC-NAM}’ is volgens VN-baas [Guterres]^{PER-NAM} een ‘belangrijke stap in het conflict’.*

- (15) Op de *[Kalandeberg]*^{LOC-NAM} in *[Gent]*^{LOC-NAM} wordt momenteel gezocht naar *[De Rechtvaardige Rechters]*^{MISC-NAM}, het beroemde verdwenen paneel van het *[Lam Gods]*^{MISC-NAM}.

2.3.5 Metonymy

A metonymy is a figure of speech in which a part of an object is used to refer to the entire object. For example, *the White House* is often used as a shorthand for *the United States government*. We introduce a rule for such cases, where place names (normally LOC entities) are used instead of an organization name. We consider the mention – *the White House*, in this case – to refer to an ORG entity and not a LOC entity.

To test for these cases, we try to replace the mention with the full ORG entity it refers to. Example 16a demonstrates this. In some cases, as in example 16b, it could be argued the mention refers to the American people rather than the American government, which would point to a PER entity. However, in these cases, we still consider it an ORG entity. We consider the **population of a country** (or of any location) to be a social construct which counts as an organization.

- (16) a. *[Amerika]*^{ORG-NAM} legt nieuwe sancties op.
→ *Amerika* can be substituted with *de Amerikaanse regering* in this case. This is a clear case of metonymy.
- b. *[Amerika]*^{ORG-NAM} heeft een nieuwe leider gekozen.

Another form of metonymy is found when a person stands in for an organization. This is the case in example 17, where the person *Trump* is mentioned on the same level as the ORG entity *Europe*. We do not consider this form of metonymy: *Trump* is still a person.

- (17) *[Europa]*^{ORG-NAM} waarschuwt *[Trump]*^{PER-NAM} voor gevolgen VS-taks op auto's

2.4 Individuality

Individuality denotes whether the mention refers to a single entity (INDIV) or a group of entities (GROUP). In the case of coordinated entities, as in

example 18b, we annotate all entities separately.

- (18) a. *[De group experts]^{GRP}* trok aan de alarmbel.
b. Oliebedrijven *[Shell]^{IND}* en *[Nexon]^{IND}*
→ Note that we annotate coordinated entities separately.
c. *[Ik]^{IND}*, *[Jan]^{IND}*, *[Karel]^{IND}* en *[mijn andere vrienden]^{GRP}*
d. *[Alle drie bedrijven]^{GRP}* zeggen de beslissing aan te vechten.
e. *[De Paus]^{IND}* reisde naar *[Brazilië]^{IND}* en *[Paraguay]^{IND}*.

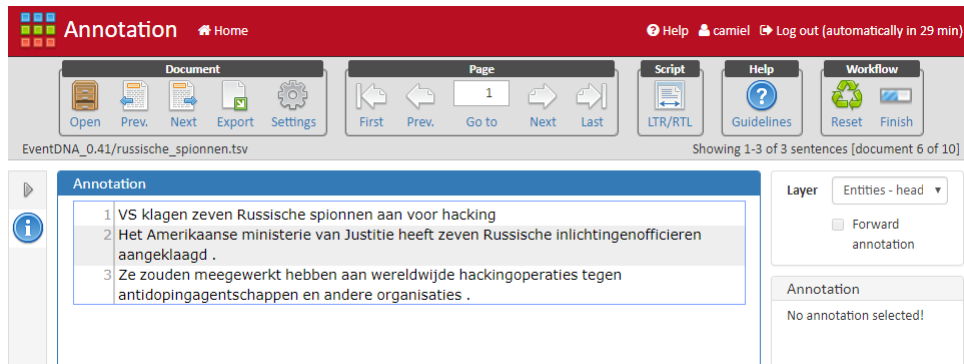
2.5 Entity coreference

Entity mentions can be coreferent, meaning they refer to the same entity in the world. The conditions for coreference are that the entities intuitively refer to the same entity and that they carry the same entity type. Besides this, entity level can (naturally) differ: an entity referred to by name in one part of the document can be referred to with a pronoun in the next. In example 19, Mariano Rajoy is referred to by name three times, by a descriptive noun phrase once, and four times with a pronoun.

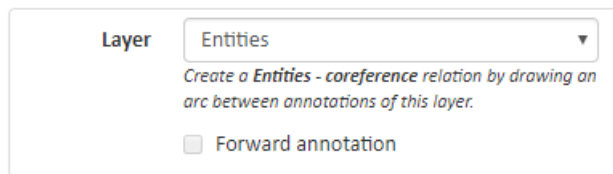
- (19) a. **Rajoy** kondigt "definitief" afscheid aan uit de politiek
b. **Mariano Rajoy**, vorige week nog **premier van Spanje**, liet gisteren al weten dat **hij** opstapt aan het hoofd van **zijn** Partido Popular.
c. Vandaag kondigde **hij zijn** "definitief" afscheid van de politiek aan.
d. Het Spaanse parlement stemde vorige vrijdag de regering van **Rajoy** weg.

2.6 Annotating entities in WebAnno

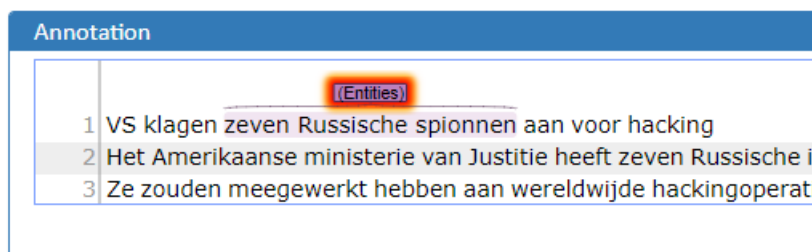
When starting up WebAnno, the screen that greets you looks as follows.



To annotate an entity, select the Entities layer in the drop-down menu on the right side of the screen. This will make that layer active.



Select a span of tokens with the mouse to create an Entity annotation. The selection will automatically snap to cover whole tokens. You can also double click a token to select only that token.



You can enter the entity's feature in the menu box on the right of the screen. WebAnno will warn you if you've forgotten to enter a required feature.

The screenshot shows a form titled "Annotation" with a "Delete" button in red and a "Clear" button. The form contains the following fields:

- Text:** A text input field containing "zeven Russische spionnen".
- Comment:** An empty text input field.
- Entity type:** A dropdown menu.
- Head:** A section containing a "Select role" dropdown menu and an "Add" button.
- Individuality:** A dropdown menu.
- Mention level type:** A dropdown menu.

Entity type, individuality and mention level type can be selected from drop-down menus.

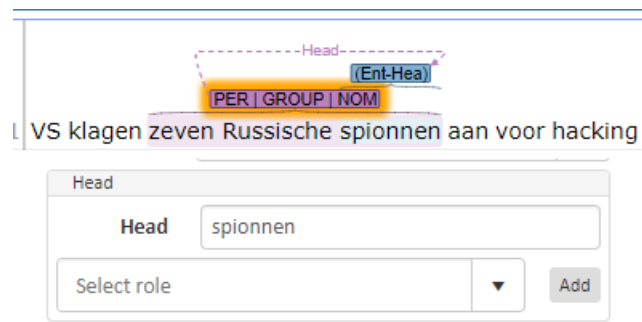
This image shows a close-up of the "Entity type" dropdown menu. The menu is open, displaying the following options:

- LOC
- MISC
- ORG
- PER

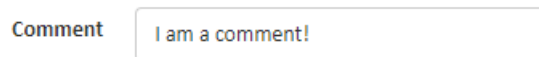
The Head feature is a type of feature called a slot feature. Annotating a slot feature automatically creates an annotation in a different layer: *Entities - Head*. To create a Head annotation, select the Head feature from the drop-down menu. (There is only one option here, but there will be more when we talk about annotating event slot features.) Then click on 'add' to the right of the drop-down menu.

This image shows a close-up of the "Head" dropdown menu. The menu is open, and the "Head" option is selected. The "Add" button is visible to the right of the dropdown.

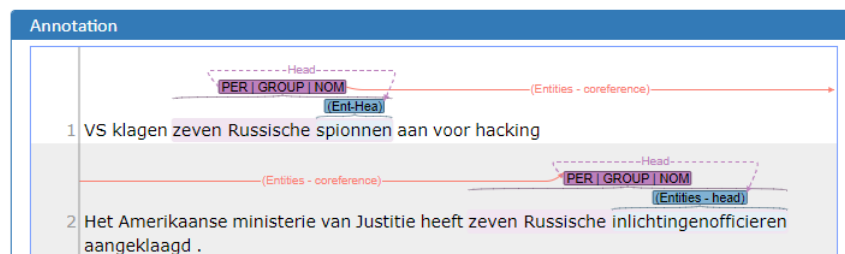
An orange field will appear, telling you to select tokens to fill the slot feature. Select the token that forms the head of the nominal mention. This will create an annotation on that token in the *Entities - Head* layer, and it will create a connection between the entity and head annotations with the feature *Head*. (Choosing 'select role' will create another head annotation, but we can only have one here.)



The **Comment** field can be filled in with free text. You can use this field to provide a comment about your reasoning or indicate your doubts when you find it difficult to determine one of the features of the entity. You are not expected to fill in this field when there are no problems with annotating the entity.



You can indicate **coreference** between two entities by simply drawing an arc between them (clicking and holding the mouse on one of the entity bubbles, then dragging the mouse to the other entity bubble and releasing). This will create an arc labeled *Entities - coreference*.



Chapter 3

Annotating events

3.1 Events and event mentions

3.1.1 Identifying event mentions

Every article we annotate contains mentions of **events**. A single sentence may contain one, several, or none. An event can be described by a syntactic clause, a nominal construction, or a infinitive construction as in the examples below. A clause is a grammatical unit that lies just below the sentence level; it is usually composed of a subject, a verb and an object. A sentence can be composed of one or multiple clauses.

(20) *[De rebellen vielen het fort 's ochtends vroeg aan.]*

(21) *[De aanval] verraste de verdedigers.*

(22) Net over de grens stopte hij om *[pillen te kopen.]*

In EventDNA , we call syntactic clauses, nominal constructions and infinitive constructions that refer to events **event mentions**. An event mention points to an event in the real world in the same way that an entity mention points to an object in the real world. The description of an event is usually accompanied by other pieces of information that answer the *wh*-questions: *what, who, where, when, why, how*. These often correspond to the entities we annotated previously. We call the entities that fill these roles **event arguments**.

We consider an event mention to consist of the sentence clause, nominal construction or infinitive construction in which the event and all its arguments are described. It is important to note that relative clauses that are part of an entity mention are kept as part of the event clause. In other words, the event mention must cover all the entity mentions that serve as arguments to that event.

In the examples below, event mentions are marked [*like this*].

- (23) a. [*Al Qaeda pleegt nieuwe aanslag in Kenia*]
→ A terrorist attack is a straightforward example of an event. The event clause is the entire sentence; it also contains the *who* and *where*, which are event arguments.
- b. [*Een Europees onderzoek detecteerde honderden producten met verboden hoeveelheden chemische stoffen.*]
→ The verb *detecteerde* hints at an event: the discovery of new information after an investigation.
- c. *Na een lange verkiezingsstrijd* [*behoudt de Japanse driepartijen-coalitie haar meerderheid in het Lagerhuis.*]
→ The coalition's victory is the target event of this event clause. *Na een lange verkiezingsstrijd* is a subordinate clause to the event clause; it is not annotated as part of the event clause. *In het lagerhuis* specifies *meerderheid*, which is the object of the event clause, and is therefore also part of the clause.
- d. [*De Japanse driepartijencoalitie, die onder leiding staat van premier Junichiro Koizumi, behoudt haar meerderheid in het Lagerhuis.*]
→ The relative clause beginning with *die* is a relative clause attached to *de Japanse driepartijencoalitie*, and is therefore part of the event clause.

In order to determine the boundaries of event clauses precisely, it is necessary to know what the arguments of the event are. The arguments that an event can take are determined by its type. In the following sections, we will go over the features that event mentions can take: type and subtype, the event arguments, the realis properties and the prominence feature. Then, we will come back on exactly what event mentions must be annotated and how they are separated from the rest of the text.

3.2 Event types

We consider each event to be an instance of a certain event type and subtype. Types are semantic categories. In example 24, three events are mentioned: the attack in Zamboanga, the death of the victims, and a separate attack that happened the day before. Both attacks evoke events of the type `Conflict.Attack`, and the death of the victims is a `Life.Die` event. Today's attack and the death of the victims are related, but they are annotated as separate events because they belong to different types. Event types have a 2-level hierarchy: `Conflict` and `Life` are higher-level types or supertypes, while `Attack` and `Die` are subtypes that further specify their supertype.

- (24) a. *Abu Sayyaf pleegt aanslag op Amerikaanse soldaten*
b. *Bij een bomaanslag in de Filipijnse stad Zamboanga kwamen woensdagavond drie mensen, onder wie een Amerikaanse soldaat om het leven. De aanslag is vermoedelijk het werk van de rebellenbeweging Abu Sayyaf. In de buurt van een christelijke kerk in Zamboanga ontplofte gisteren nog een bom. Bij die aanslag vielen geen slachtoffers.*

One trigger can evoke multiple events of different kinds – in fact, a number of event types frequently go hand in hand. In example 25, the event clause evokes both a `Transaction.TransferMoney` event and the `Transaction.TransferOwnership` of the pills.

- (25) *Net over de grens stopt hij om [een paar kilo pillen te kopen.]*

Image 3.2 shows all event types used in the EventDNA annotation. We also annotate events that we think are newsworthy but that can't easily be brought under a specific type: we use the `Unknown` type for these. A description of each of the types follows.

3.3 Description of event types

This section defines and describes the applicability of each event type used in EventDNA. The EventDNA type taxonomy is based on the Rich ERE taxonomy. For a detailed overview of the Rich ERE types used in EventDNA, we

| Type | Subtype | Type | Subtype |
|--------------------|-------------------|--------------------|-------------------|
| Business | StartOrg | Justice | Acquit |
| | MergeOrg | | Appeal |
| | EndOrg | | ArrestJail |
| Conflict | Attack | | ChargeIndict |
| | Demonstrate | | Convict |
| | Conflict | | Execute |
| Contact | Broadcast | | Extradite |
| | Contact | | Fine |
| | Correspondence | | Pardon |
| | Meet | | ReleasePardon |
| Life | BeBorn | | Sentence |
| | Die | | Sue |
| | Divorce | | TrialHearing |
| | Injure | | Investigation |
| | Marry | | Justice |
| Manufacture | Artifact | Transaction | TransferMoney |
| Movement | TransportArtifact | | TransferOwnership |
| | TransportPerson | Journalism | Publication |
| Personnel | Elect | | Investigation |
| | EndPosition | Politics | Vote |
| | StartPosition | Unknown | Unknown |

refer to [1], p. 34-83. The following section only covers the event types introduced in `EventDNA:Conflict.Conflict`, `Journalism.Publication`, `Journalism.Investigation`, `Politics.Vote` and `Unknown.Unknown`.

For a visual overview of event types, argument slots for each type and possible entity types for each argument, see appendix A.

3.3.1 Conflict events

`Conflict.Conflict`

`Conflict.Conflict` events cover forms of conflicts or warfare without a physical component. It is mostly used for events that are metaphorically described as conflicts, like *handelsoorlog*, or conflicts of a non-physical nature like cyber attacks. It does not cover verbal forms of conflict like insults or criticisms. It does not cover types of conflict situations that fall under other types, like an arrest. Keep in mind the restriction that an event should refer to a specific occurrence.

- Entity (PER, ORG): A PER or ORG entity instigating, defending itself or in any other way involved in the conflict.
- Place (LOC): The place where the event takes place.
- Time (time): The time the event takes place.

Examples:

- In April werd [een cyberaanval tegen de [Organizatie voor het Verbod op Chemische Wapens]^{Entity}]`Conflict.Conflict` vrijdeld.
- [De handelsoorlog tussen de [VS]^{Entity} en [China]^{Entity}]`Conflict.Conflict` loeide deze maand op een laag pitje.
- Beide prijswinnaars worden geëerd voor hun strijd tegen seksueel geweld als oorlogswapen. → *hun strijd...* is not annotated, because it does not refer to a specific occurrence.

3.3.2 Journalism events

Journalism.Publication

Journalism.Publication events cover the publication of reports, articles, books, etc. or the revealing of information from a journalistic source. The publication of a newspaper article is also an instance of this event.

- Publisher (PER, ORG): The entity that publishes the information. Can be the same as the author.
- Author (PER, ORG): The entity that produces the information. Can be the same as the publisher.
- Content (content): The content, subject or theme of the published information.
- Place (LOC): The place where the event takes place.
- Time (time): The time the event takes place.

Journalism.Investigation

Journalism.Investigation events cover investigations made by journalistic sources. Journalistic sources include newspapers, magazines or news websites, including independent news sources (that work outside of the traditional media structure). Investigation events typically cover acts of investigative journalism, and may result in a the publication of a report (which would be a Journalism.Publication event.)

- Investigator (PER, ORG): The entity which carries out the investigation.
- InvestigationTarget (PER, ORG, target): The target or object under investigation. Can be a person, organization or a more general 'target' entity. A target can be a specific entity or event, or a general theme (like *drug smuggling* or *homelessness*).
- Place (LOC): The place where the event takes place.
- Time (time): The time the event takes place.

3.3.3 Politics events

Politics.Vote

A `Politics.Vote` occurs when a group of people vote on an issue. The issue can take many forms, such as an election or a referendum. It usually covers votes made by large groups of people, such as the vote for a new president. The voters can also be a restricted group of people, such as votes made by congressmen in congress. This type does not cover voting events that happen outside the context of politics, like votes taken by executives in companies or organizations.

- Voter (PER, ORG): The entity which votes.
- Topic (topic): The topic or object of the vote, for instance *the presidential election* or the passing of a law.
- Place (LOC): The place where the event takes place.
- Time (time): The time the event takes place.

3.3.4 Unknown events

Unknown.Unknown

The `Unknown.Unknown` type is meant as a fallback option for events that must be annotated but do not clearly belong to a more specific event type. In such cases, the annotator can use the `Unknown` event type. When this happens, the annotator is invited to provide a comment about his choice.

The `Unknown` type should be used carefully and sparingly. The annotator should not take this type as an invitation to annotate everything that looks like an event in the document. When considering using this type, the annotator should first ask him/herself if the event is very relevant to the article, and only use this type if so. He/she should also keep in mind the restriction that an event must point to a specific occurrence

- Participant (Specifier): Any entity involved in the event. Can refer specifically to PER, LOC or ORG entities, but also covers any other relevant entities.

- Place (LOC): The place where the event takes place.
- Time (time): The time the event takes place.

Examples:

- *[[Hongwei Meng]^{Part}, de baas van de internationale organisatie Interpol, is spoorloos verdwenen.]^{Unknown}*
- Nog steeds problemen om voedsel en hulp te verspreiden op Sulawesi na *[de tsunami die [het eiland]^{Part}. [vorige week]^{Time} trof.]^{Unknown}*
→ The tsunami is an important, relevant event. We do not annotate the *problemen...* because it does not point to a specific occurrence.
- De verdachten zouden meegewerkt hebben aan een *[grootschalige hackingoperatie in [Juli vorig jaar.]^{Time}]^{Unknown}*
- *[Nobelprijs voor Vrede gaat naar Congolese arts [Denis Mukwege]^{Part} en Jezidi-activist [Nadia Murad.]^{Part}]^{Unknown}*
- *[[Vijf leeuwen]^{Part} afgemaakt in [Zuid-Afrika]^{Place}]^{Unknown}* → TheLife.Die type only applies to PER entities, and therefore not to animals.

3.4 Event arguments

Each event has a number of corresponding arguments. An argument is a participant in the event or some other piece of information related to it, such as the object of the action or the place where the event takes place. Each event type expects a number of arguments of specific types. The `Conflict.Attack` event type, for instance, asks for an Attacker, a Target, an Instrument, a Place and a Time. The Attacker must be a person, organization, or a location (used metonymically). Arguments are either the entities annotated in the previous annotation step (like Attacker), or it can be an argument filler (like Time).

An event's arguments slots do not have to be filled if the relevant argument is not mentioned. It can happen that an event is mentioned without any arguments.

(26) De terreuraanslag in *[New York]^{LOC-NAM}* maakte *[acht gewonden]^{PER-NOM}*.

- Conflict.Attack
- Attacker = (empty)
- Target = acht mensen
- Instrument = (empty)
- Place = New York
- Time = (empty)

Argument fillers

Most arguments to events are entities which have already been annotated in the previous annotation step. However, many events ask for arguments that are not annotated as entities. We call this type of arguments **argument fillers**. For example, example 27 contains a `Transaction.TransferMoney` event that asks for a Money argument. The string *700 euro* fills this role. However, *700 euro* is not an entity. *700 euro* is annotated as an argument to an event without being an entity with a type, specificity, etc. Time expressions are another very common kind of argument filler.

(27) *Dries gaf [700 euro]^{MONEY} uit aan een tweedehandse moto.*

Multiple arguments the same role

Multiple arguments can take the same role. In example 28, the `Contact.Meet` type asks for an Entity argument, which is a person or organization that attends the meeting. Two are mentioned in this example, so the event carries two Entity arguments. In the same way, there can be more than two Victim arguments to an Attack event, or more than two defendants in a `Justice.Sentence` event. In example 29, we see multiple related entities like *Queens, het westen van New York* and *New York*. All these entities are annotated as arguments to the event.

(28) *[Mark Zuckerberg]^{PER-NAM} heeft een hoorzitting voor [twee senaatscommissies]^{ORG-NOM} in [Washington]^{LOC-NAM} [donderdagochtend]^{ARGFILLER:TIME} al bij al goed doorstaan.*

- Contact.Meet
- Entity = Mark Zuckerberg

- Entity = twee senaatscommissies
- Place = Washington
- Time = donderdagochtend

(29) De terreuraanslag in de wijk *[Queens]^{LOC-NAM}*, in *[het westen van [New York]^{LOC-NAM}]^{LOC-NOM}*, maakte *[acht gewonden]^{PER-NOM}*, onder hen *[twee Belgen]^{PER-NOM}*.

- Conflict.Attack
- Attacker = (empty)
- Target = acht mensen
- Target = twee Belgen
- Instrument = (empty)
- Place = Queens
- Place = het westen van New York
- Place = New York
- Time = (empty)

Where do we look for arguments?

The arguments and argument fillers of an annotated can be found anywhere within the sentence in which the event is annotated. It cannot come from another sentence. As explained previously, an argument to an event must be a part of that event clause.

3.5 Event realis properties

The realis properties of an event are a set of properties that indicate whether the event actually happened or not. These properties are as follows:

3.5.1 Positive/negative

When the writer or speaker explicitly states that the event did not take place, the event is NEGATIVE. This can happen with the use of explicit negation words or by context, as in the examples below. All other events are POSITIVE.

- (30) a. De voorzitter heeft de overeenkomst niet getekend.
 b. De voorzitter weigerde de overeenkomst te tekenen.
 c. Hij probeerde te scoren, maar de bal miste het net.

3.5.2 Tense

TENSE denotes the time at which the event took place — PAST, PRESENT or FUTURE — relative to the textual anchor time. The textual anchor time is the time of publication or broadcast. If the tense cannot be determined, we use the label UNSPECIFIED. This is typically the case for events expressed as a nominal construction or in the telegram-style headers of news articles.

- (31) a. De VN zal het embargo opheffen. → FUTURE
 b. Hun ontmoeting nam gisteren plaats. → PAST
 c. Hij werd gedwongen de boete te betalen. → PAST
 d. De politie is op zoek naar de dader. → PRESENT
 e. De aankoop vindt vandaag plaats. → FUTURE
 → At the time of publication, the purchase had not yet taken place.
 f. Hongaren opnieuw de straat op → UNSPECIFIED
 g. [*Het vrijgeven van de beelden*] veroorzaakte een schokgolf door het parlement. → UNSPECIFIED
 → Note that we refer to the publication of the images as the target event, and not to the effect this caused. The tense of the complete sentence is PAST, but we do not annotate it.

3.5.3 Modality

Modality refers to the level of assertion of an event. An event is ASSERTED if the writer or speaker refers to it as if it were a real event, regardless of tense or positivity.

- (32) a. De aanslag vond gisteren plaats in het centrum van de stad.
 → ASSERTED
 b. De overeenkomst wordt morgen getekend. → ASSERTED

Contrasting with this are events that are hypothetical or not real in some way, again regardless of tense or positivity. An event can be positive or negative (affirmed or denied) but at the same time not asserted. These events fall under the following modalities, though the list is not exhaustive:

- Believed, suspected or rumoured events
- Hypothetical events
- Commanded and requested events
- Threatened, proposed and discussed events
- Desired events
- Promised events
- Otherwise unclear constructions

The following are all examples of OTHER events, with the exception of example 33d. The target events are *[marked]*. This shows an important aspect of events: **the modal verbs, mental verbs and verbs of communication that often accompany other events (like *zouden, willen, denken, zeggen, dreigen, eisen, verbieden*, etc.) do not represent events themselves.**

- (33) a. Zou hij *[president worden]*, de wereld zou er anders uitzien.
→ hypothetical → POS, PRESENT, OTHER
- b. Vroeger wou ik *[astronaut worden]*.
→ desired → POS, PAST, OTHER
- c. Het gerucht van *[zijn arrestatie]* deed de ronde.
→ believed, rumoured → POS, PRESENT, OTHER
- d. De politie zegt *[de verdachte deze morgen te hebben aangehouden]*.
→ asserted → POS, PAST, ASSERTED
→ In this case, the speaker (the police) asserts the event has taken place.
- e. Iedereen dacht dat *[het schip gezonken was]*.
→ believed → POS, PRESENT, OTHER

- f. Ik wou dat [*ik mijn portefeuille niet kwijt was geraakt.*]
→ desired → NEG, PAST, OTHER
- g. De soldaten dreigden ermee [*het huis in brand te steken.*]
→ threatened → POS, FUTURE, OTHER
- h. Nu dat [*de samenvoeging*] waarschijnlijk niet meer zal plaatsvinden...
→ NEG, FUTURE, OTHER
- i. Men maakt zich ongerust dat [*zo'n ramp nog eens zou kunnen gebeuren*].
→ POS, FUTURE, OTHER
- j. De oliebedrijven eisen [*compensatie van de Nederlandse regering*].
→ requested → POS, FUTURE, OTHER
- k. De regeringspartij heeft de oppositie verboden nog meer *protesten te organiseren*.
→ POS, FUTURE, OTHER
→ The event we are focussing on is the protests organized by the opposition, not the act of forbidding these protests.

3.6 Event prominence

A paragraph of news text can mention several different events. In EventDNA, we presume that one of the events described in a piece of news text is the **main event**. The main event is the event that caused the reporter to write the article. Contrasting with this, the **background events** give background or context around the main event. We annotate both main and background events, and annotate them as such. This feature is called **prominence**.

The main event can usually be found in the title of the article and/or in the first sentence of the lead paragraph. Identifying the main event can be straightforward if it is a specific occurrence; it is more difficult when the article discusses multiple events that are closely related. An article can have multiple main events. The following sections discuss the main and background events of a number of real newspaper articles. Every time,

the first sentence given is the title of the article and the others represent the lead paragraph.

3.6.1 Cases with one main event

- (34) a. *[Cabine ontploft aan Flageyplein]*^{Main}
→ The main event is explicit: the explosion of a high-voltage cabin.
- b. *[Ter hoogte van het Flageyplein in Brussel is gisteren een hoogspanningscabine ontploft.]*^{Main} *[Gewonden vielen er niet.]*^{Backgr.}
→ The first sentence repeats the main event, with more detail. The second sentence refer to a (negative) wounding event. This is extra information given alongside the main event. We therefore annotate it as a background event.
- (35) a. *[Kenia berecht daders dubbele aanslag]*^{Main}
- b. *[Vier Kenianen worden sinds gisteren berecht wegens hun mogelijke betrokkenheid bij een dubbele anti-Israëliische aanslag in november 2002.]*^{Main} *[De dubbele aanslagen vonden plaats in het Keniaanse Mombasa]*^{Backgr.} en *[werden opgeëist door het terreurnetwerk Al Qaeda]*^{Backgr.}. *[Het streng beveiligde proces wordt gehouden in de Keniaanse hoofdstad Nairobi]*^{Main}. *[De vier aangeklaagden pleiten onschuldig.]*^{Backgr.}
→ The main event is the trial of the four suspects. It is stated explicitly in the title and first sentence of the lead paragraph. The actual attacks are given as background information in the following sentence. The third sentence refers to the trial again and provides a location for it. The fourth sentence provides information related to the main event but doesn't refer to it explicitly.
- (36) a. Vladimir Poetin verstevigt positie na verblijf in het Kremlin
- b. *[De Russische regeringspartij Verenigd Rusland kroont Vladimir Poetin vandaag tot haar voorzitter.]*^{Main} Tenminste als de aftredende president ingaat op de talloze smeekbedes van de partijtop. Maar daar twijfelt wellicht niemand aan.
→ The title is somewhat vague and doesn't refer to a tangible

occurrence. The first sentence of the lead paragraph is more explicit: the main event is the selection of Vladimir Putin as head of his party. The following sentences nuance the first: the decision is not yet final. We do not find events specific enough to annotate in these sentences.

3.6.2 Cases with multiple main events

Difficult situations can arise when two events happen nearly at the same time or are given the same importance by the writer, or when it is not possible to state that one event gives context to the other. In these cases, the annotators should assess what they think are foreground events and what events give context to the story, keeping in mind that they can annotate more than one event as the main event.

- (37) a. Politieke aardbeving in Israël
b. *[De Israëlische premier Ariel Sharon heeft zijn lidmaatschap van de Likoedpartij opgezegd]^{Main 1} en [het ontslag van zijn regering aangeboden]^{Main 2} bij de president. [Sharon gaat nu een nieuwe partij vormen.]^{Backgr.}*
→ Two events are given equal weight as potential main events in the lead sentence: Ariel Sharon's leaving the Likud party and the resignation of his government. One event is not clearly the cause of the other, and it is not clear one happens before the other. We annotate both events as main events.
- (38) a. *[Abu Sayyaf pleegt aanslag op Amerikaanse soldaten]^{Main 1}*
b. *[Bij een bomaanslag in de Filipijnse stad Zamboanga kwamen woensdagavond drie mensen, onder wie een Amerikaanse soldaat om het leven.]^{Main 2} De aanslag is vermoedelijk het werk van de rebelbeweging Abu Sayyaf. [In de buurt van een christelijke kerk in Zamboanga ontplofte gisteren nog een bom]^{Backgr.}. [Bij die aanslag vielen geen slachtoffers.]^{Backgr.}*
→ We find two candidates for main event: the attack in Zamboanga and the death of the victims. We consider that these events are both main events, since the writer gives each item the same weight in the text.

- (39) a. *[Syrië verdacht van moord op Libanees politicus]*^{Main 1}
- b. *[De Libanese politicus en journalist Gibran Tueni is gisteren omgekomen bij [een aanslag.]*^{Backgr.}*]*^{Main 2} *[Het Syrische regime is de hoofdverdachte van de moord.]*^{Main 1} Damascus kreeg net gisteren ook kritiek van de Verenigde Naties. Volgens de VN werken de Syriërs onvoldoende mee met het onderzoek naar de moord op de Libanese ex-premier Rafik Hariri.
- Three events share the foreground of the article: the death of Gibran Tueni, the terrorist attack which led to his death, and the suspicion towards Syria for the murder. The terrorist attack is not elaborated upon, so we consider it a background event even though it happened at the same time as the death of Gibran Tueni. The other two events are both main events.

3.7 Annotating events in WebAnno

After having created an annotation, the annotator fills in the event features in the menu on the right side of the screen. The features are shown in alphabetical order, but it is wise to annotate the event type before the subtype and the subtype before the arguments.

Annotation
Delete
Clear

Text VS klagen zeven Russische spionnen aan voor hacking

Arguments ⓘ

▼
Add

Comment

Event subtype ⓘ ▼

Event type ▼

Modality ▼

Positive/negative ▼

Prominence ▼

Same clause

▼
Add

Tense ▼

The Modality, Positive/negative, Prominence and Tense features are all filled in through drop-down menus. The Type and Subtype also work that way. However, when a type has been selected, the appropriate subtypes will be highlighted in the Subtype drop-down menu.

Event subtype

Event type

Modality

Positive/negative

Prominence

Same clause

Select role

▼

ArrestJail

ReleaseParole

TrialHearing

Sentence

Fine

ChargeIndict

Sue

▼

Add

Then, when a subtype has been selected, the possible argument roles for that subtype will be selected automatically, ready to be activated.

Annotation Delete Clear

Text VS klagen zeven Russische spionnen aan voor hacking

Arguments

Plaintiff <Click to activate>

Adjudicator <Click to activate>

Defendant <Click to activate>

crime <Click to activate>

Place <Click to activate>

Time <Click to activate>

Select role ▼ Add

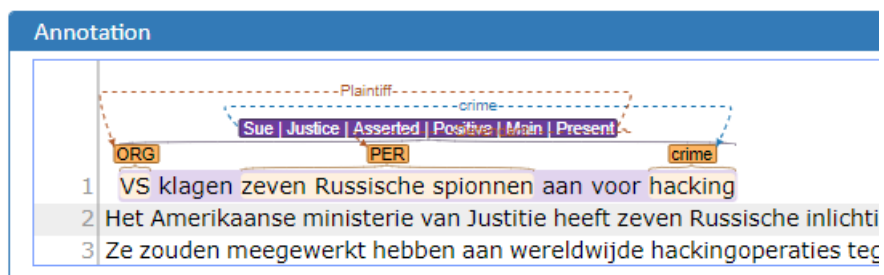
Comment

Event subtype

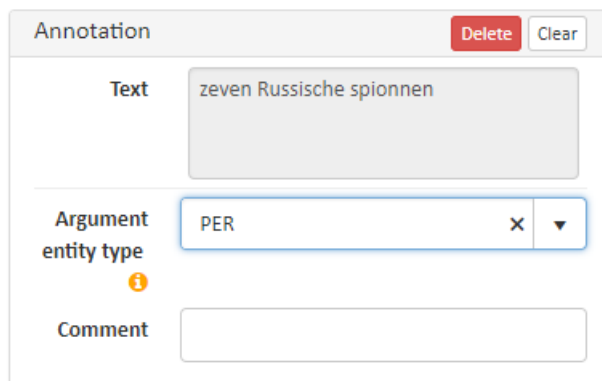
Event type Sue ▼

Justice ▼

To fill in an argument role, click it to activate it, then select the text span for that role. This will create an annotation on the *Events - arguments* layer and an arc from the event to that argument. An event annotated with its arguments can become visually very busy. Remember it is possible to turn off the rendering of specific layers through the settings menu.



Afterwards, you will need to select the argument annotation separately to fill in its *Argument entity type* feature. **If the argument corresponds to an entity annotated in the *Entities* layer, make sure their string spans overlap exactly.**



It is possible for an event to be non-contiguous. In that case, you can use the slot feature called **same clause**. Annotate the first part of the event clause as an event, then annotate the subsequent parts of that event clause using the same clause slot feature. The reason we do this is because it is not possible to directly annotate non-contiguous spans in WebAnno.

As with entities, you can indicate coreference across event by drawing arcs between them. This will create a coreference relation annotation.

Chapter 4

Annotating IPTC Media Topics

Since 2010, the International Press Telecommunications Council (IPTC) maintains a taxonomy for the categorization of news text. This taxonomy takes the form of a tree with seventeen top-level topics such as *politics*, *society* or *sport*. Each topic branches into subtopics until very specific topics are reached, such as *adult education*, *impeachment* or *missing person*. Newswriters may use this open standard to categorize their articles.

In EventDNA , we annotate articles with their relevant IPTC Media Topics. We allow an arbitrary number of topics for each article.

The taxonomy can be visualized at <http://show.newscodes.org/index.html?newscodes=medtop&lang=en-GB&startTo=Show>.

4.1 Choosing the correct topics

This section describes the procedure the annotator should follow to choose relevant media topics.

1. Using the visualization found at the above address (reached by following the links on the IPTC Media Topics website), go over the topics at the top level of the taxonomy. If it applies to the article, write it down. If it is not, ignore it.

2. For each topic you have written down, expand it and go over its subtopics. If it applies to the article, write it down, otherwise ignore it.
3. Repeat this process on each level of the taxonomy for each relevant topic until you have exhausted all possible topics.
4. To build a final list of topics, it is only necessary to retain the most specific topics, discarding the intermediate levels. This is because choosing a specific topic implies that you have also chosen the more general topics above it.

At the end of this process, you have a list of topics which apply to the articles.

Consider as an example the following article snippet.

Onverhoopt staakt-het-vuren in hongerend Jemen

Zweedse vredesbesprekingen over Jemen leveren meer op dan verwacht

Na een jarenlange impasse pakken hernieuwde vredesgesprekken voor Jemen onverwacht positief uit. In de belegerde havenstad Hodeida wordt een staakt-het-vuren van kracht. Dat betekent een "verbetering voor de leefomstandigheden van miljoenen Jemenieten", zegt VN-secretaris-generaal Antonio Guterres.

Going over the top level topics, we find the topics *politics* and *conflict, war and peace* quite relevant. Under *politics, international relations* is the only topic that fits the bill; we are thinking about the relations between Yemen and the UN. Under this topic we find *diplomacy* and *international organization* (the UN). We judge that other topics are not quite relevant. *Foreign aid* (the transfer of resources from one country to another) is not discussed in the snippet. Neither are *refugees and internally displaced people*, even though we imagine they might exist in this situation. **We only aim to annotate topics that are discussed in the article, and not those that we imagine are part of the situation described in the article.** *Summit meetings* does not apply, since we read in the description of the topic that it *includes meetings of leaders, foreign and finance ministers from the Group of Eight major nations and Group of Seven industrialized nations*. This does not cover the Yemen peace talks.

We have picked all relevant topics under *politics*. We repeat the process for *conflict, war and peace*. We follow a line through *armed conflict* to *war*, ending at *civil war*. We also pick *peace process* under the top level, and find the relevant topic *peace talks*. We then discard the intermediate topics we found and arrive at this final list: *international organization, civil war* and *peace talks*.

It is normal for the annotator to encounter tags during this process that they are not certain apply to the article or not. For this purpose, we allow annotating the certain tags and the uncertain tags separately.

4.2 Annotating media topics in WebAnno

WebAnno does not allow making annotations over an entire text, and does not allow for an arbitrary number of annotations of a certain feature. For this reason, we use a workaround to annotate IPTC Media Topics. The process of annotation is as follows:

1. While viewing the document to be annotated, select the IPTC Media Topic layer.
2. Create an annotation on the first token of the document.
3. The features *Certain topics* and *Uncertain topics* can be filled in with typed text. The *Certain topics* feature is required, *Uncertain topics* is not. You should write the topics you have found in these articles as a list, using the colon (":") as the delimiter. *It is important to write the topic names exactly as they appear in the tree view provided. It is important to use the colon as delimiter, since the normal comma is already used in some topic names.*

Annotation

(IPTMedTop)

- 1 Onverhoopt staakt-het-vuren in hongerend Jemen
- 2 Zweedse vredesbesprekingen over Jemen leveren meer op dan verwacht
- 3 Na een jarenlange impasse pakken hernieuwde vredesgesprekken voor Jemen onverwacht positief uit .
- 4 In de belegerde havenstad Hodeida wordt een staakt-het-vuren van kracht .
- 5 Dat betekent een verbetering voor de leefomstandigheden van miljoenen Jemenieten , zegt VN-secretaris-generaal Antonio Guterres .

Layer IPT

Annotation

Text On

Certain topics

Uncertain topics

Annotation

international organization; civil war; peace talks

- 1 Onverhoopt staakt-het-
- 2 Zweedse vredesbesprekingen over Jemen lev
- 3 Na een jarenlange impasse pakken hernieuwc
- 4 In de belegerde havenstad Hodeida wordt eer
- 5 Dat betekent een verbetering voor de leefoms

Annotation Delete Clear

Text Onverhoopt

Certain topics international organization; civil war; peace t

Uncertain topics

Bibliography

- [1] Linguistic Data Consortium. *DEFT Rich ERE Annotation Guidelines: Events (V3.0)*. 2015. URL: http://www.nist.gov/tac/2016/KBP/guidelines/summary_rich_ere_v4.2.pdf.

Appendix A

EventDNA event type table

| | | | |
|-------------------|-----------------------|---------------------|--------------------------------------|
| Business | <i>StartOrg</i> | Agent | PER, ORG |
| | | Org | ORG |
| | <i>EndOrg</i> | Org | ORG |
| | <i>MergeOrg</i> | Org | ORG |
| Conflict | <i>Attack</i> | Attacker | PER, ORG |
| | | Target | PER, ORG, vehicle, weapon, commodity |
| | | Instrument | Weapon, vehicle, commodity |
| | <i>Demonstrate</i> | DemonstrateEntity | PER, ORG |
| | <i>Conflict</i> | Entity | PER, ORG |
| Contact | <i>Meet</i> | Entity | PER, ORG |
| | <i>Correspondence</i> | Entity | PER, ORG |
| | <i>Contact</i> | Entity | PER, ORG |
| | <i>Broadcast</i> | Sender | PER, ORG |
| | | Receiver | PER, ORG |
| Journalism | <i>Publication</i> | Publisher | PER, ORG |
| | | Author | PER, ORG |
| | | Content | content |
| | <i>Investigation</i> | Investigator | PER, ORG |
| | | InvestigationTarget | PER, ORG, target |
| Justice | <i>ArrestJail</i> | Agent | PER, ORG |
| | | Person | PER |
| | | Crime | crime |
| | <i>ReleaseParole</i> | Agent | PER, ORG |
| | | Person | PER |
| | | crime | crime |
| | <i>TrialHearing</i> | Prosecutor | PER, ORG |
| | | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |
| | | crime | crime |
| | <i>Sentence</i> | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |
| | | sentence | sentence |
| | | crime | crime |
| | <i>Fine</i> | Adjudicator | PER, ORG |
| | | Entity | PER, ORG |
| | | money | money |
| | | crime | crime |
| | <i>ChargeIndict</i> | Prosecutor | PER, ORG |
| | | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |
| | | crime | crime |
| | <i>Sue</i> | Plaintiff | PER, ORG |
| | | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |

| | | | |
|--------------------|--------------------------|---------------------|---------------------------------|
| | | crime | crime |
| | <i>Extradite</i> | Agent | PER, ORG |
| | | Person | PER |
| | | Crime | crime |
| | | Origin | LOC |
| | | Destination | LOC |
| | <i>Acquit</i> | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |
| | | Crime | crime |
| | <i>Convict</i> | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |
| | | crime | crime |
| | <i>Appeal</i> | Prosecutor | PER, ORG |
| | | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |
| | | Crime | crime |
| | <i>Execute</i> | Agent | PER, ORG |
| | | Person | PER |
| | | Crime | crime |
| | <i>Pardon</i> | Adjudicator | PER, ORG |
| | | Defendant | PER, ORG |
| | | Crime | crime |
| | <i>Investigation</i> | Investigator | PER, ORG |
| | | InvestigationTarget | PER, ORG, target |
| | <i>Justice</i> | Entity | PER, ORG |
| | | Crime | crime |
| Life | <i>Injure</i> | Agent | PER, ORG |
| | | Victim | PER |
| | | Instrument | Weapon, vehicle, commodity |
| | <i>Die</i> | Agent | PER, ORG |
| | | Victim | PER |
| | | Instrument | Weapon, vehicle, commodity |
| | <i>BeBorn</i> | Person | PER |
| | <i>Marry</i> | Person | PER |
| | <i>Divorce</i> | Person | PER |
| Manufacture | <i>Artifact</i> | Agent | PER, ORG |
| | | Artifact | vehicle, weapon, commodity, LOC |
| | | Instrument | Weapon, vehicle, commodity |
| Movement | <i>TransportPerson</i> | Agent | PER, ORG |
| | | Person | PER |
| | | TransportInstrument | vehicle, weapon |
| | | Origin | LOC |
| | | Destination | LOC |
| | <i>TransportArtifact</i> | Agent | PER, ORG |

| | | | |
|--------------------|--------------------------|---------------------|--------------------------------------|
| | | Artifact | weapon, vehicle, commodity, LOC |
| | | TransportInstrument | vehicle, weapon |
| | | Origin | LOC |
| | | Destination | LOC |
| Personnel | <i>StartPosition</i> | PersonnelEntity | ORG |
| | | Person | PER |
| | | Position | title |
| | <i>EndPosition</i> | PersonnelEntity | ORG |
| | | Person | PER |
| | | Position | title |
| | <i>Elect</i> | Agent | PER, ORG |
| | | Person | PER |
| | | Position | title |
| Politics | <i>Vote</i> | Voter | PER, ORG |
| | | Topic | topic |
| Transaction | <i>TransferOwnership</i> | Giver | PER, ORG |
| | | Recipient | PER, ORG |
| | | Beneficiary | PER, ORG |
| | | Thing | vehicle, weapon, LOC, ORG, commodity |
| | <i>TransferMoney</i> | Giver | PER, ORG |
| | | Recipient | PER, ORG |
| | | Beneficiary | PER, ORG |
| | | Money | money |
| | <i>Transaction</i> | Giver | PER, ORG |
| | | Recipient | PER, ORG |
| | | Beneficiary | PER, ORG |
| Unknown | <i>Unknown</i> | Participant | PER, ORG, specifier |