

Annotation Guidelines for Entity and Event Coreference

Marta Recasens*

November 10, 2011

Contents

1	Previous Work	1
1.1	OntoNotes (Pradhan et al., 2007)	1
1.1.1	Overview	1
1.1.2	Commented examples	2
1.2	NP4E (Hasler et al., 2006)	3
1.3	ECB - EventCorefBank (Bejan and Harabagiu, 2008)	4
1.3.1	Overview	4
1.3.2	Commented examples	4
2	Reannotating the ECB Corpus	5
2.1	Mentions	6
2.2	Coreferent events	7
2.3	Entity coreference	8

1 Previous Work

1.1 OntoNotes (Pradhan et al., 2007)

1.1.1 Overview

The event annotation in OntoNotes is restricted to those events that have been referred to by NPs. There might be verb–verb connections only if there is an NP linking them. Thus the annotation guidelines do not explicitly mention event coreference. All they say is:

*Stanford University, Stanford, CA 94305. E-mail: recasens@stanford.edu

Verbs are added as single-word spans **if** (and only if) they can be coreferenced with an existing NP. This includes morphologically related nominalizations (1) and NPs that refer to the same event but are lexically distinct (2).

- (1) Sales of passenger cars [grew]₁ 22%. [The strong growth]₁ followed year-to-year increases.
- (2) Japan’s domestic sales of cars, trucks and buses in October [rose]₁ 18% from a year earlier to 500,004 units, a record for the month, the Japan Automobile Dealers’ Association said. [The strong growth]₁ followed year-to-year increases of 21% in August and 12% in September.

Only the single-word head of the verb phrase is included in the span, even in cases where the entire verb phrase is the logical coreferent. Singleton events are not annotated.

1.1.2 Commented examples

Most of the event coreference relations annotated in OntoNotes involve a discourse-deictic anaphor (Webber, 1988).

- (3) You know we’ve been here a long time and I think most people would [meet] them halfway, but [that] doesn’t seem to be what they want.
- (4) Iowa’s governor is now looking to follow Postville’s example, [relaxing] immigration quotas. [It]’s a goal that seems possible...
- (5) This was an all-white, all-christian community that all the sudden was taken over – not taken over, that’s a very bad choice of words, but [invaded] by, perhaps different groups. [It] began when a Hasidic Jewish family bought one of the town’s two meat-packing plants 13 years ago.

*Note the rewording: “not taken over but . . .”, so the event is not annotated.

- (6) . . . the gigantic international auction house Sotheby’s pleaded guilty to [price-fixing] with Christie’s – its only real competition in an industry that does \$4 billion in business every year. [This] is a case of the rich and the powerful stifling competition. . . . Arch rivals for nearly three centuries, the two auction houses agreed to [fix] prices on what they charged the buyers and sellers of high-priced art. They have already agreed to pay more than \$500 million to settle the civil suits spawned by [their crimes].

*Notice *their crimes* as an event noun.

- (7) Slobodan Milosevic finally [admitted]₁ he [is]₂ no longer the President of Yugoslavia, and he said he had just received [that information]₂. [It]₁ followed a day of intense behind-the-scenes negotiation to decide his future and the future of the country. ... In a surprise live speech tonight, Slobodan Milosevic [conceded]₁ his loss and congratulated his opponent.

*The verb *to be* is annotated.

*Note that *that information* is not referring to the event itself, but to the verbal report of the event.

- (8) Senior administration officials are now saying privately that if Milosevic never [comes] to trial, the U.S. might not like [it], but could live with [it].

*Negation (*never*) is not annotated as verbs are annotated as single words.

- (9) The chief prosecutor at the International War Crimes Tribunal has [demanded] the new Yugoslav President hand Slobodan Milosevic over to face trial for war crimes ... I believe we can not ignore [the actions of the war crimes tribunal].
- (10) Israelis have so often [felt] insulated from this kind of violence. The last 13 days have changed [all of that].

1.2 NP4E (Hasler et al., 2006)

An **event** is defined as a thing that happens or takes places; a single specific occurrence, either instantaneous or ongoing, that is **unique** and can be anchored at a **point in time**. This does not include generic events. In (11), *Zaire's counter-offensive* is considered as an event, but not *flying* because it describes a repeated action over a longer period of time.

- (11) ... the planes were flying from Kisangani, the headquarters for [Zaire's counter-offensive launched last month] ...

The annotation is limited: only event categories related to the domain of terrorism/security (i.e., attack, defend, injure, die, contact) are annotated.

Further annotations:

- Polarity: whether an event does or does not occur.

- (12) a. Positive: ... a vast stretch of territory seized by Tutsi-led rebels who [took up arms] last October.
- b. Negative: The MRTA have refused to [meet] with government negotiators.

- Modality: the speaker’s/writer’s certainty that an event is real or actually happens.
- (13) a. Yes: Police say up to one hundred people might have [lost their lives] last October.
- b. No: ... a vast stretch of territory seized by Tutsi-led rebels who [took up arms] last October.
- Tense: past / present / future / unspecified.

1.3 ECB - EventCorefBank (Bejan and Harabagiu, 2008)

1.3.1 Overview

Unlike OntoNotes and NP4E, the focus of ECB is on cross-document coreference. The corpus consists of related documents from the Google News archive that describe the same seminal event.

# Topics	43
# Documents	482
# Within-topic events	339
# Cross-document events	208
# Within-document events	1,302

Apart from coreference, they also consider the following relations: subevent, reason, purpose, enablement, precedence, and related. Event coreference is annotated if two requirements are fulfilled:

- Both events are expressed by the same predicate or their synonyms or hyponyms.
- Whenever specified, both predicates have the same arguments.

1.3.2 Commented examples

- (14) [taking] over the director’s chair for ... [direct] the sequel to Twilight ... [take] on the sequel to Twilight ... negotiating to [direct] the cinematic sequel to Twilight ... is in negotiations to [helm] “New Moon”, the sequel to Summit Entertainment’s vampire breakout Twilight ... has been [hired] to ... hired to [direct] the second film in ... is expected to [direct] New Moon ... is close to [signing] on for the blockbuster’s sequel ... [getting] behind the Twilight steering wheel

*We should not have V+V and both coreferential as in *has been [hired] to [direct]*. The being hired event is different from the directing event.

*The signing event is different from the directing event, but very similar to the being hired event.

- (15) [rumored] to be the front runner earlier ... despite having [directed] the vampire romance

*The rumored event should not be coreferent with the directing event.

- (16) [confirmed] that Chris Weitz ... [claiming] Chris Weitz is getting behind the Twilight steering wheel
- (17) [directed] The Golden Compass
- (18) is [negotiating] to ... is in [negotiations] to helm
- (19) “Chris Weitz, who directed “The Golden Compass,” is negotiating to direct the cinematic sequel to Twilight,” industry sources [say] ... Days after Catherine Hardwicke was [ousted]

*Non-coreferent events.

*One event mention is missing: Chris Weitz will direct the sequel to Twilight, New Moon, [replacing] Catherine Hardwicke. This replacing event should be annotated as coreferent with *ousted*.

- (20) Summit Entertainment [announced] today that filmmaker
- (21) a [battle] is going on between a studio saying there’s no news to report

2 Reannotating the ECB Corpus

We will both complete and revise the current annotation of ECB. With respect to completing the corpus, we will annotate every entity and event in the sentences that are (partially) annotated. With respect to revising the corpus, we should modify the extension of some of the annotated spans (see Section 2.1) and make the event annotation more consistent (see Section 2.2). Some of these changes imply modifying the annotation of entity coreference as well (see Section 2.3). For example, we will reannotate the following example (only the relevant annotations are shown)

he has <MENTION CHAIN=“1”> made </MENTION> <Entity COREFID=“1”> an eight-year offer </Entity>

as

he has made <MENTION CHAIN=“1”> an eight-year offer </MENTION>

2.1 Mentions

We annotate all event mentions, regardless of whether they are referred to again later in the text or they are singletons. Event mentions include two kinds of syntactic units, verbs and NPs. See the details below.

1. Verbs

- In general, annotate a single word: the semantic head of the VP (can be finite or non-finite). Auxiliaries should not be annotated.

(22) has been [hired]

- When a phrasal verb is involved, annotate both the verb and the preposition or adverb. Note that there might be a discontinuity (24-a), in which case it should be annotated as in (24-b), ie., the second part of the mention is marked with an asterisk.

(23) [taking over] the director's chair

(24) a. beloved pin-up icon Bettie Page [passed] yesterday
[away] in Los Angeles
b. beloved pin-up icon Bettie Page <MENTION
CHAIN="2"> passed </MENTION> yesterday
<MENTION CHAIN="2*"> away </MENTION>
in Los Angeles

- Light verbs (i.e., *give*, *take*, *have*, *make*, *do*) followed by an event noun should not be annotated. Annotate the noun (or verb) that carries the semantic content.

(25) the Angels have made [an offer] to the free agent first
baseman

(26) They're not gonna [leave]

- Idioms should be annotated with all their components. Like (24-b), if there is a discontinuity, the second part of the mention is annotated with an asterisk (27-b).

(27) a. Chris Weitz is [getting behind the] Twilight [steering
wheel]
b. <MENTION CHAIN="23"> getting behind the
</MENTION> Twilight <MENTION
CHAIN="23*"> wheels </MENTION>

2. NPs

- Annotate the full NP including all modifiers.

(28) Wladimir Klitschko easily set aside [the challenge of substitute Hasim Rahman], stopping the former heavyweight champion . . .

- Nominal event mentions include: verbal nominalizations (29-a), nouns that have the same verbal form (29-b), report nouns (29-c), event-related nouns (29-d), and pronouns (29-e).

(29) a. Employees at were coping Saturday with [the death of their boss, who was gunned down at their office Christmas party].
b. [a battle] is going on between a studio saying there's no news to report
c. With [last night's news from Los Angeles Angels]
d. with [a lopsided victory over the American Hasim Rahman]
e. Iowa's governor is now looking to follow Postville's example, [relaxing] immigration quotas. [It]'s a goal that seems possible...

Any mention other than a verb or an NP should not be annotated with event coreference. Therefore, in the following example, the preposition *according* needs to be removed from the current annotation of the ECB corpus.

(30) Legendary pinup queen Bettie Page died of pneumonia at the age of 85 in a Los Angeles, California, hospital Thursday, a week after suffering a heart attack, according to her agent.

2.2 Coreferent events

Annotate the mentions that refer to the same event. Hints to know if it is the 'same' event:

- Both events happened in the same place at the same time.
- Both predicates have the same arguments: subject, object, cause, goal, etc.

We annotate coreference, not narrative chains or subevents, so refrain from annotating a coreference relation between verbs like *direct* and *hire* in (31), as you direct a movie but hire somebody.

- (31) [direct] the sequel to Twilight ... is in negotiations to [helm] “New Moon”, the sequel to Summit Entertainment’s vampire breakout Twilight ... has been hired to

Coreference between generic (or repeated) events should also be annotated:

- (32) Israelis have so often [felt] insulated from this kind of violence. The last 13 days have changed [all of that].

Report nouns can refer to either the communicative event (33-a) or the communication object itself (33-b) :

- (33) a. Slobodan Milosevic finally [admitted] he is no longer the President of Yugoslavia. [His announcement] was met with a standing ovation.
 b. Slobodan Milosevic finally admitted he [is] no longer the President of Yugoslavia, and he said he had just received [that information].

2.3 Entity coreference

Changes with respect to the previous version of the ECB corpus that we have used so far:

1. Singleton entities are included.
2. The whole NP is annotated. So a case like (34-a) becomes (34-b).

(34) a. [Jack Ketsoyan], Tara Reid’s representative
 b. [Jack Ketsoyan, Tara Reid’s representative]
3. Following OntoNotes, we will NOT annotate appositive phrases and nominal predicates as coreferent. So the highlighted phrases in (35-a) and (35-b) should be unannotated.

(35) a. It was [her decision].
 b. The organization behind the Academy Awards named Australian actor Hugh Jackman as [the host of February’s annual show].
4. Only premodifiers that are proper nouns (36-a) or possessive phrases (36-b) should be annotated.

(36) a. [[U.S.] TV show “Scrubs”]
 b. [[the American Pie star’s] rep]

References

- Bejan, C. A. and Harabagiu, S. (2008). A linguistic resource for discovering event structures and resolving event coreference. In *Proceedings of LREC 2008*, pages 2881–2887.
- Hasler, L., Orasan, C., and Naumann, K. (2006). NPs for Events: Experiments in coreference annotation. In *Proceedings of LREC 2006*, pages 1167–1172.
- Pradhan, S. S., Ramshaw, L., Weischedel, R., MacBride, J., and Micciulla, L. (2007). Unrestricted coreference: Identifying entities and events in OntoNotes. In *Proceedings of ICSC 2007*, pages 446–453.
- Webber, B. L. (1988). Discourse deixis: reference to discourse segments. In *Proceedings of ACL 1988*, pages 113–122, Buffalo, New York.