

Annotating Events and Entities in Dialogue

Tatiana Anikina, Ivana Kruijff-Korabayová
DFKI / Saarland Informatics Campus, Saarbrücken, Germany
tatiana.anikina@dfki.de ivana.kruijff@dfki.de

1. Introduction

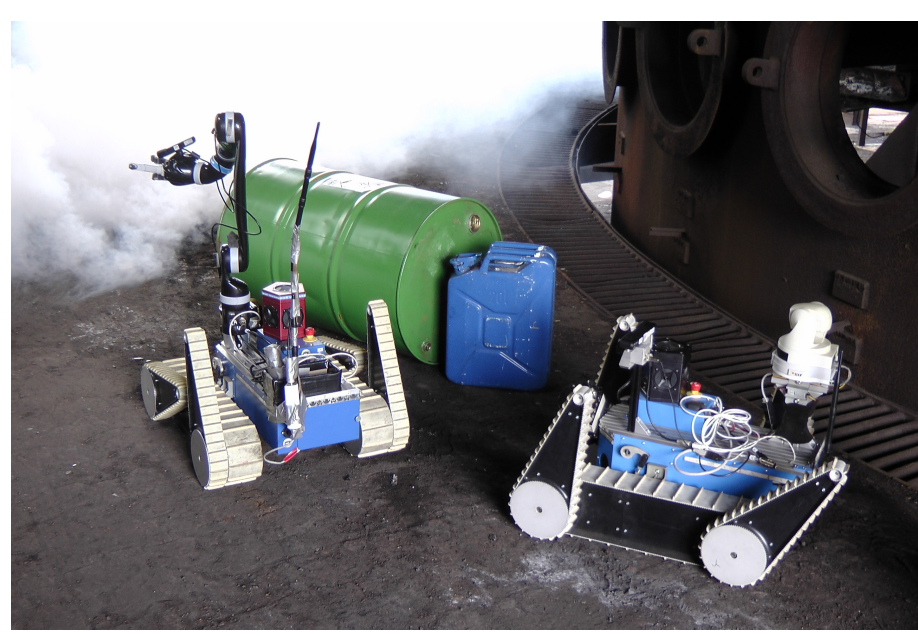
Events and entities are crucial for NLU but there are not many annotated resources available and most of them provide annotations only for one specific phenomenon (e.g., coreference or temporal relations). To create a corpus of **dialogues with semantic annotations** we developed an annotation tool EveEnti and devised a new annotation scheme that covers a wide range of annotations for events and entities and relations between them.

2. Data and Annotation Scheme

Our dataset is based on **German dialogues from the disaster response domain**. The dialogues are between the team leader and several operators who remotely operate the robots while exploring some area, locating fire, damage or victims (see Figure 1). In total, we have **2,398 transcribed dialogue turns in German and 818 turns in English**.



(a) Robot operators at work



(b) Ground robots exploring a barrel

Figure 1: Emergency response training

We annotate all events and entities in dialogue with various **semantic features and relations** (see Figures 2-3). Events are linked to each other based on temporal order, rhetorical relations or coreference and each event has a list of arguments (entities) with the corresponding thematic roles. Entities are connected via coreference, set/member, part/whole, or an unspecific bridging relations. Additionally, each turn is annotated with a communicative function.

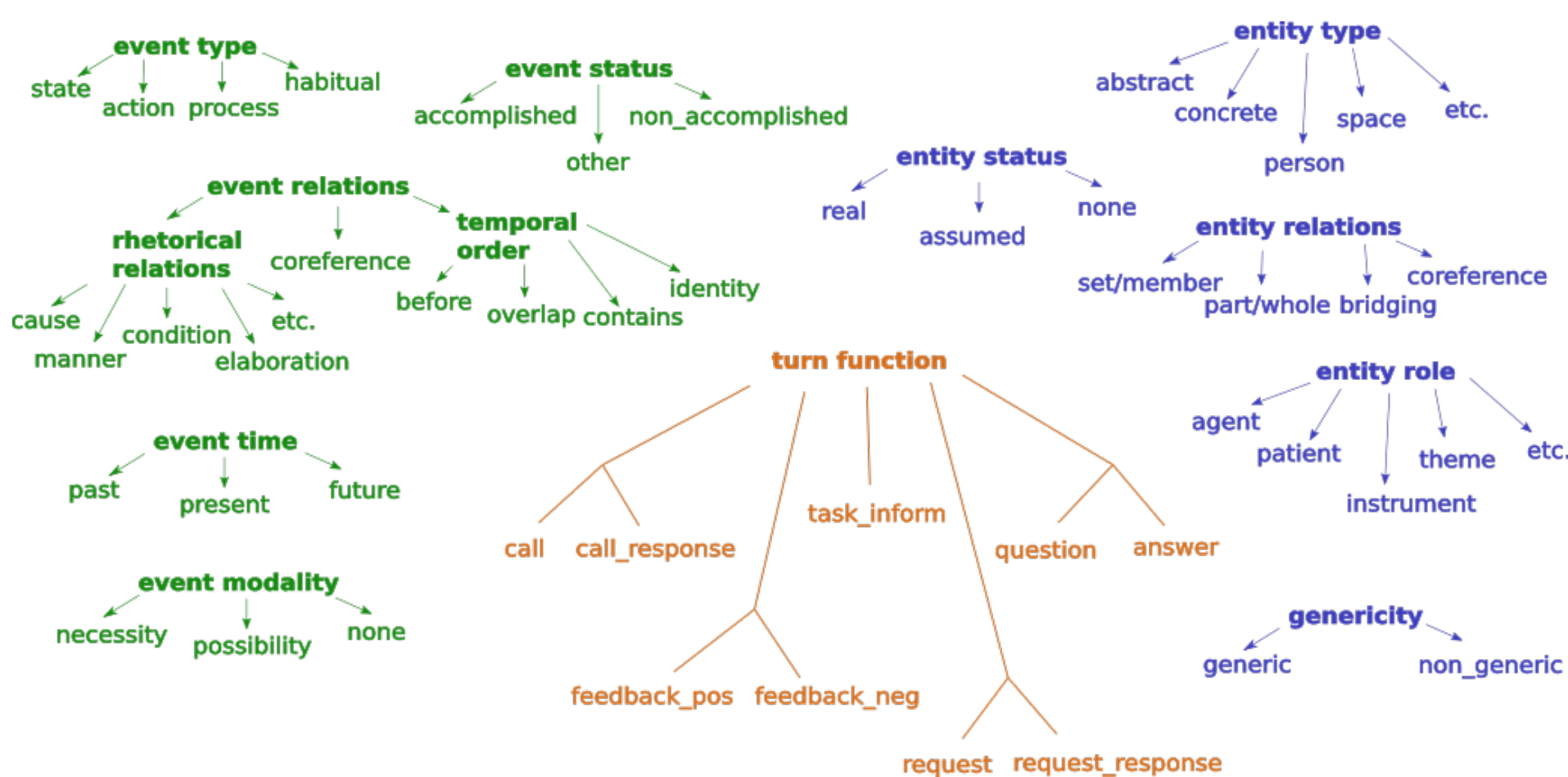


Figure 2: Taxonomy of event and entity annotations and turn functions

3. Annotation Challenges

1. Discontinuous event/entity spans: e.g., 'habe gemacht' (made) in 'ich **habe** ein paar Fotos **gemacht**' [I **made** some pictures].
2. Nested events and entities: e.g., 'hat angefangen' [started] and 'regnen' [raining] in 'es **hat angefangen zu regnen**' [it **started raining**].
3. Implicit arguments and relations: 'ich habe **dort** schon gesucht und es **gab** kein Zeichen' [I have already searched **there** and didn't **find** any sign].
4. Unclear status of events and entities (e.g., assumed or real entity): 'vorrangig weiter **Personen** suchen' [first search further for **people**].
5. Compounds that have both events and entities within the same token: e.g., 'ich setze **Personensuche** fort' [I continue **searching for people**].

6. Ambiguous relations: e.g., bridging or set/member relation between 'erstes Bildmaterial' [the first pictures] and 'das Foto' [the photo]:

A: Baldmöglich **erstes Bildmaterial** senden. [Send **the first pictures** as soon as possible.]

B: Ich habe ein Objekt ... ich schick **das Foto**. [I found an object ... I am sending **the photo**.]

(a) D3 mit Überfass ist am Standort angekommen.
(D3 arrived to the position with a barrel.)

(b) A: Die Standorte der Roboter sind im Lagebild nicht zu erkennen.
(The positions of the robots are not visible in the system.)
B: Das ist verstanden. (Understood.)

Figure 3: Event and entity annotations

Annotation Type (# of categories)	Agreement (κ)	Annotation Type (# of categories)	Agreement (κ)
Event Type (4)	0.74	Entity Role (10)	0.50
Event Time (3)	0.68	Entity Type (9)	0.53
Event Negation (2)	0.80	Entity Genericity (2)	0.72
Event Modality (3)	0.75	Entity Status (3)	0.69
Event Status (3)	0.38	Entity Relations (3)	0
Event Relations (20)	0.13	Entity Coreference (avg. f-score)	0.88
Event Coreference (avg. f-score)	0.76	Turn Function (10)	0.91

(a) Event annotation

(b) Entity annotation & Turn function

Figure 4: Inter-Annotator Agreement

4. Conclusion

We present our ongoing work on **event and entity annotations in dialogue**. The EveEnti scheme provides various kinds of semantic annotations that can be used in combination or stand-alone. We hope that these annotations will be helpful for understanding how dialogue evolves over time and how different events and entities relate to each other in discourse. We believe that such interactions are important for a variety of NLP tasks, including coreference resolution and intent recognition.