Reference in Team Communication for Robot-Assisted Disaster Response: An Initial Analysis

Natalia Skachkova and Ivana Kruijff-Korbayová
DFKI / Saarland Informatics Campus / 66123 Saarbrücken Germany
`natalia.skachkova@dfki.de;ivana.kruijff@dfki.de`

Abstract

We analyze reference phenomena in a corpus of robot-assisted disaster response team communication. The annotation scheme we designed for this purpose distinguishes different types of entities, roles, reference units and relations. We focus particularly on mission-relevant objects, locations and actors and also annotate a rich set of reference links, including co-reference and various other kinds of relations. We explain the categories used in our annotation, present their distribution in the corpus and discuss challenging cases.

1 Introduction

We present the findings of an initial analysis of contextual reference phenomena in team communication in robot-assisted disaster response. Disaster response teams operate in high risk situations and must make critical decisions quickly, despite partial and uncertain information. For better safety and operational capability, first responders increasingly deploy mobile robots for remote reconnaissance of an incident site. The work in this paper contributes to our ongoing effort to develop methods for interpreting the verbal communication in a response team, in order to extract *run-time mission knowledge* from it. Mission knowledge encompasses the mission goals, which tasks have been assigned to whom, the state of their execution, the relevant points of interest (POIs) and the possibly changing information about them, etc. We work on using mission knowledge extracted from the verbal team communication and integrated with information from other sources, such as the sensors carried by the robots, to provide situation awareness and teamwork assistance both during and after a mission, as described in (Willms et al., 2019).

As part of extracting mission knowledge from the verbal team communication, it is important to identify *mission-relevant objects, locations, actors, tasks, events etc.* that are being referred to and the links between them. This is the goal of reference resolution. In order to get a better understanding of how this task can be performed, we analysed the corpus of robot-assisted disaster response team communication from the TRADR project (TRADR project website, 2020), (Kruijff-Korbayová et al., 2015). The corpus contains the communication in teams of first responders who are using ground and airborne robots to explore an area, searching for victims and hazards, to carry out measurements and gather samples in the aftermath of an industrial incident, such as a fire or explosion. In the first phase of reference phenomena analysis, the results of which we present in this paper, we focused on the references to and links between mission-relevant objects, locations and actors. We aimed to gain insight regarding the kinds of reference cases in the data, their distribution and the challenges for reference resolution. We annotated the data in order to systematically capture the various cases and be able to access them later for deeper analysis. Our aim was not to create an ultimate annotated resource and/or a novel annotation scheme. We designed the annotation scheme specifically for our analysis, and it was evolving as the annotation progressed. The present paper reports our findings and indicates what our reference resolution system would need to deal with. In Section 2 we overview existing approaches to co-reference and anaphora annotation in text and dialogue. Section 3 gives more details about our data. Section 4 describes the cases of reference to

and links between mission-relevant objects, locations and actors that we identified and how we captured them in our annotation scheme, accompanied by typical examples as well as illustrations of some tricky cases and challenges. In Section 5 we summarize and indicate our future steps.

2 Approaches to co-reference and anaphora annotation

The study of co-reference and anaphoric relations has long tradition in linguistics, and there exist numerous approaches to annotation. They are rather difficult to systematise, due to proliferating terminologies and combinations of heterogeneous phenomena, such as different types of references, referents and relations. Since a proper discussion of the similarities and differences would exceed the space we have available here, we present the relevant previous work in a close to chronological order.

The first annotation scheme for anaphoric relations appeared in 1992 (Fligelstone, 1992). In the late 1990s this field of research got a push, when the 6th and 7th Message Understanding Conferences (MUC-6, MUC-7) took place. MUC-7 Coreference Task Definitions (Hirschman and Chinchor, 1998) defined *co-reference* as a symmetric identity relation between two noun phrases (NPs) if both of them refer to the same entity. Among the researchers who worked on co-reference annotation schemes around this time are McEnery et al. (1997), Ge (1998), Rocha (1999). There also appeared some works investigating not only relations between entities introduced by NPs, but also event co-reference and temporal relations between events, e.g. by Bagga and Baldwin (1999), or Setzer and Gaizauskas (2000).

At the same time, interest in the annotation of a wider range of semantic relations emerged. Among these relations is *anaphoric reference*. While co-reference is an equivalence relation, anaphoric reference is not - the interpretation of an anaphoric expression always depends on its antecedent. Deemter and Kibble (2000) discussed the differences between anaphora and co-reference in detail. They stressed that they can coincide, but are not interchangeable, and pointed out that co-reference is not to be mixed with bound anaphora, where an anaphor relates to a generic antecedent, which does not actually refer to any specific entity. They also showed the difference between co-reference and an intensional relation between an entity and a predicative expression that refers to a whole set of entities.

One of the more fine-grained approaches to co-reference annotation was presented by Hasler et al. (2006). Aiming at creating corpora for event processing, they investigated NP and event co-reference and created a co-reference annotation scheme. They introduced relations between NPs (identity, synonymy, generalisation, specialisation and other) and co-reference types (NP, copula, apposition, bracketed text, speech pronoun and other).

Among more recent papers on co-reference annotation are the works by Cohen et al. (2017) about identity and appositive relations in biomedical journal articles, Dakle et al. (2020) on co-reference in emails, Wright-Bettner et al. (2019) on cross-document co-reference.

While all the above mentioned works mostly deal with text data, Poesio et al. (1999) developed the MATE 'meta-scheme' for anaphora annotation in dialogue. This generic scheme consists of a core scheme for annotating identity relations (co-reference) between the entities introduced by NPs, and three extensions for annotating references to the visual situation, bridging and anaphoric relations involving an extended range of anaphoric expressions and antecedents. The bridging extension was later realized in the GNOME annotation project (Poesio, 2004).

Poesio et al. (1999) also formulated the difficulties that any designer of an annotation scheme for anaphora faces, namely, that almost every word or phrase in a coherent text can potentially be linked to something that was introduced earlier (cf. also the concept of cohesion (Halliday and Hasan, 1976)).

Speaking of anaphora, sometimes researches try to concentrate only on the identity relation between entities, e.g. (Poesio, 2004) or Aktaş et al. (2018). But often anaphora is understood in a broader sense. Nissim et al. (2004) and Elango (2005) discussed in various types of anaphors and antecedents. Zinsmeister and Dipper (2010) researched the annotation of abstract (discourse-deictic) anaphora. Anaphoric relations between different types of events were also studied by Caselli and Prodanof (2010). Poesio et al. (2008), annotating the ARRAU corpus, introduced an anaphoric relation between single anaphoric expression and plural antecedents, as well as references to events, actions and plans.

Kruijff-Korbayová and Kruijff (2004) developed a discourse-level annotation scheme that covered a

broad range of discourse reference properties, e.g. semantic sort, delimitation, quantification, familiarity status, and anaphoric links (co-reference and bridging).

There also emerged some classifications of anaphora types and other relations. So, Tetreault et al. (2004) annotated the Monroe corpus, consisting of task-oriented dialogues from an emergency rescue domain. They focused on co-referential pronouns and NPs (identity relation), but also presented a classification of relations for non-co-referential pronouns with the following relation types: indexicals, action, demonstrative, functional, set, hard and dummy. Botley (2006) distinguished three types of abstract anaphora: label anaphora (encapsulates stretches of text), which has several sub-types, situation anaphora (for linking events, processes, states, facts, propositions) and text deixis. Another classification by Eckart de Castilho et al. (2016) included the following anaphora types: individual anaphors, reference to abstract objects, vague anaphors, inferrable-evoked pronouns and unmarked anaphors.

All these classifications were developed with a certain corpus and task in mind, and none can be considered universal, standard and pervasive. Although none of the existing classifications was entirely suitable to us, given our domain and the goal of our analysis, they were fundamentally helpful in devising our own annotation scheme.

So, following Hasler et al. (2006), we try to differentiate between an explicit identity, when expressions are linked via a copula, and an implicit one. Defining the bridging relation, we relied on works of Poesio (2004) and Kruijff-Korbayová and Kruijff (2004). Our intensional relation can be traced back to both bound anaphora and intensional predicates presented by Deemter and Kibble (2000), and the notion of vague anaphor is similar to that defined by Eckart de Castilho et al. (2016).

We started the annotation effort with the aim to keep the annotation scheme quite simple, distinguishing between main types of mission-relevant entities, locations and actors, and focusing on the reference relation types. As we proceeded with the annotation, we found it necessary to extend the scheme with certain more fine-grained distinctions to capture sometimes quite special cases.

3   Data: The TRADR Team Communication Corpus

The TRADR corpus consists of dialogues that represent human-human team communication in robot-assisted disaster response. The dialogues were recorded during exercises on different industrial sites performed as part of the TRADR project (TRADR project website, 2020), (Kruijff-Korbayová et al., 2015). The exercises simulated situations after a industrial accident, such as fire, explosion, etc. and involved teams of firefighters using ground and airborne robots for reconnaissance.

There are 15 files with dialogues in the corpus, each corresponding to a mission or sometimes a part of a mission. Nine files contain dialogues in German, and six - in English. The German dialogues were recorded in 2015 and 2016, the English data is from 2017. The firefighters who took part in the 2017 experiment were Dutch, and so non-native English speakers. In total the joint corpus contains about 2,9k dialogue turns (see Table 1).

Recording	Mission	Duration	Turns
TJex2015			**363**
	Day 1	48:21 min	186
	Day 2	33:21 min	177
TEval 2015			**1,279**
	Day 1	58:23 min	359
	Day 2	65:04 min	356
	Day 3	57:15 min	272
	Day 4	53:22 min	292
TEval 2016			**422**
	Day 1	n.a.	312
	Day 2	n.a.	110
TEval 2017			**811**
	Day 1	64:02 min	239
	Day 2	149:20 min	400
	Day 3	56:36 min	172

Table 1: TRADR corpus composition (based on (Anikina and Kruijff-Korbayová, 2019))

The TRADR experiments involve teams of first responders exploring complex dynamic environments using robots, namely unmanned ground vehicles (UGVs) and unmanned aerial vehicles (UAVs). The robots are used for reconnaissance, mainly to look for points of interest (POIs), including victims and hazard sources, such as smoke, fire, or contamination; and check if the site is safe enough for human first responders to enter. The robots are equipped with gas detectors, a standard camera and an infrared one. Pictures taken by the robot cameras can be shared among the team members. Some UGVs have a mechanical arm for picking up, turning, pushing or moving objects.

The team consists of operators (UGV-1, UGV-2, UAV) who control the robots, a team leader (TL) and

in some missions also a mission commander (MC). A MC is in charge of the whole mission and gives tasks to teams. The TL distributes the tasks between the operators, coordinates their actions and reports to the MC (if present). The operators use robots to perform the tasks assigned to them and report to the TL about the results or possible difficulties.

The team members use a shared situation awareness interface, consisting of a digital map on which POIs are marked and robots' positions are displayed; a repository of shared photos made with the robot camera; and in 2017 also a task list which the TL can manually edit.

The team communication in the TRADR scenarios has the following characteristics:

- Participants follow the radio communication protocol (albeit somewhat loosely), i.e. they use special phrases to start/finish a conversation, check the connection quality, accept/reject requests, etc.
- Information flows through a rather complex communication pipeline with several participants. This sometimes leads to repeating information or requests.
- TL switches between the operators, so the flow of information is usually split into several interlaced threads. This sometimes leads to confusion and misunderstandings.
- The participants sometimes refer to objects on the display, i.e., the shared digital map, photos or task list.
- The fact that the participants perceive the environment via a medium (here the robot's camera(s)) is reflected in language usage. Often, when the TL assigns tasks and gives commands, they speak to an operator, but mean a robot. Similarly, an operator may refer to an icon on the digital map as a real object or location, and vice versa. We call this *double reality representation*.
- Like any spontaneous speech TRADR dialogues are characterized by repetitions, elliptical constructions, fillers/hesitation markers, such as 'erm', 'uh', etc. and other disfluencies, incomplete and/or ungrammatical utterances.

4 Annotation and Analysis of Reference in the TRADR Corpus

The analysis we present aims to gain initial insight in the kinds and distribution of references to entities and relations between them in the TRADR corpus, as a preparatory step before developing reference resolution modules as part of the team communication interpretation in our system (Willms et al., 2019).

In this section we explain the categories that we distinguished in the analysis, show their distribution in the TRADR corpus and discuss the challenges we encountered during annotation.

We used the WebAnno tool (Eckart de Castilho et al., 2016) to perform the annotation. Originally, only one annotator, the first author, performed the annotation, under the guidance of the second author. All spurious cases were discussed by both authors and the annotation was updated based on the decision. We adjusted and extended our annotation scheme in the process. We did not involve multiple independent annotators, because our aim was mainly to get an overview of the reference resolution issues. To test the reliability of the resulting annotation scheme, another person annotated a small subset of the corpus, consisting of one dialogue, which contained 57 utterances. We measured inter-annotator agreement using Cohen's kappa following (Carletta, 1996). We obtained a kappa score of 0.704 for the *Entities* layer, 1.0 for *Comments*, 0.895 for *Roles*, 0.573 for *Reference Units* and 0.845 for *Reference Links*. This shows good agreement, except for reference units.

4.1 Annotation Scheme

Our annotation scheme has four separate layers: *Entities, Roles, Reference units* and *Comments*. We use separate layers, so that each layer can have its own set of markable expressions and a separate corresponding tag set. We keep the tag sets flat for practical reasons.

At the *Entities* layer we annotate mission-relevant objects (POIs), locations, mission participants (actors) and other mentioned entities. At the *Roles* layer we annotate the role of each mention of a mission participant, such as MC, TL, OP-UGV1, OP-UAV, etc. The purpose of the *Reference Units* layer is to annotate reference links as well as the syntactic category of the expressions that constitute the source and target of the link. Finally, at the *Comments* layer we annotate several special cases: expletive pronouns, deictic pronouns referring to displayed objects, incorrect transcriptions, uncertain and vague cases. The annotation of entities, roles and reference units and links is discussed in more detail below.

A given expression may be marked simultaneously at different layers. For example, the expression 'UGV one' may be marked at the *Entities* layer as an actor, at the *Roles* layer as OP-UGV1 and at the *Reference Unit* layer as an NP.

Table 2 shows the full list of tags for each layer and their distribution in the TRADR corpus. Of the 7067 entities in total 51.2% are actors, 17.48% are various kinds of objects, 11.07% locations, 0.92% displayed POIs and 19.34% do not fit into one of these classes and are labeled as other. As for roles, 31.35% refer to the TL and 1.2% to the MC, 53.61% to the robot operators, 5.15% to the robots and 1.41% are other roles. We marked a total of 4385 reference units, of which 80.21% are nominal expressions and 19.79% other markables.

Table 3 shows the distribution of reference links. We annotated in total 2502 relation instances. The largest group is basic anaphora, which makes up almost 55% of all relation instances. Bridging constitutes 12.35%, implicit identity and base identity are also among the common relations with 9.6% and 7.5%, respectively. Other relations occur much less often.

4.2 Entities

We marked expressions referring to mission-relevant entities and assigned them a tag characterizing their semantic type. As we were particularly interested in mission-relevant objects (POIs), locations and actors, we distinguished these explicitly, and the rest received the tag "other". We considered NPs and NP-like expressions as primary markables. For locations we included also other types of expressions, esp. prepositional phrases and adverbials.

From the viewpoint of reference resolution we identified the need to make the following distinctions: (a) *object/location* that the participants know exists (object) or where it is (location), (b) *potential object/location*: the participants are not sure it exists (object) or it is a hypothetical place, (c) *undefined object/location*: the participants know it exists but not what it is (object) or it is an unknown place (location). Example 4.1 illustrates the three cases in (a), (b) and (c) respectively.[1]

Layer	Tag set and distribution
Entities Total: 7067	actor (3618 / 51.2%), object (1082 / 15.31%), potential object (128 / 1.81%), undefined object 25 / 0.35%), location (741 / 10.49%), potential location (9 / 0.13%), undefined location (32 / 0.45%), POI (65 / 0.92%), other (1367 / 19.34%)
Roles Total: 3764	MC (45 / 1.2%), TL (1180 / 31.35%), TEAM (136 / 3.61%), OP-UGV1 (986 / 26.2%), OP-UGV2 (649 / 17.24%), OP-UAV (312 / 8.29%), OP-PL (71 / 1.89%), UGV1 (59 / 1.57%), UGV2 (65 / 1.73%), UAV (38 / 1.01%), OP+ROBOT (138 / 3.67%), ROBOT-SG (13 / 0.35%), ROBOTS-PL (19 / 0.5%), OTHER-SG (45 / 1.2%), OTHER-PL (8 / 0.21%)
Reference units Total:4385	np (2904 / 66.23%), pro (581 / 13.25%), num (20 / 0.46%), name (12 / 0.27%), adv (582 / 13.27%), vp (40 / 0.91%), pp (18 / 0.41%), clause (156 / 3.56%), discourse (72 / 1.64%)
Comments Total: 303	EXPLETIVE (121 / 39.93%), INCORRECT (68 / 22.44%), DEICTIC (36 / 11.88%), UNCERTAIN (53 / 17.49%), VAGUE (25 / 8.25%)

Table 2: Reference annotation layers, tag sets and their distribution

Reference links Total: 2502	basic anaphora (1375 / 54.96%), bridging (309 / 12.35%), discourse anaphora (37 / 1.48%), propositional anaphora (103 / 4.12%), identity (183 / 7.31%), potential identity (32 / 1.28%), implicit identity (240 / 9.59%), implicit potential identity (44 / 1.76%), asking for identity (11 / 0.44%), intensional reference (82 / 3.28%), negative reference (38 / 1.52%), continuation (43 / 1.72%), metonymy (5 / 0.2%)

Table 3: Reference link types and their distribution

Example 4.1 *Object/location: (a) real, (b) potential, (c) undefined*
(a) UAV: Ich habe jetzt bei [der zweiten Person] auch eventuell Rauch gefunden. (I also found possible smoke near [the second person].)
(b) TL: Yes, only the outside, looking for [smoke] or [victims]. Over.
(c) TL: Can you see a... [what is a... that smoke from]? Over.

---

[1] Square brackets enclose the markable(s) in focus in each example. We do not indicate all markables for the sake of legibility. For German examples we provide an English translation that we make as near-literal as possible.

The analysis of inter-annotated agreement showed that it is especially difficult to distinguish between displayed POIs and objects, and to decide whether we have a mission-relevant POI (*object* or *location*) or an irrelevant one (*other*). Furthermore, we detected the following aspects that need further consideration in the future. First, in most cases locations are not clearly delimited, e.g., *'the north-west corner of the plant'*, which is a challenge for example for rendering them in the digital map. Second, we currently do not distinguish between absolute locations, e.g., *'the north-west corner of the plant'* and locations relative to the current position of the robot. Third, the same or similar expression may be interpreted as a reference to an object in one context, and location in another. For example, the two staircases in Example 4.2(a) are objects, but a staircase can be a location in another case. Fourth, the expressions referring to locations (and objects) can be nested, cf. Example 4.2(b). The nested structure can be quite complex.

Example 4.2 *Entities – challenges: (a) object vs. location, (b) nesting*
(a) UGV-1: Ja, ich hab gerade ein Foto geschossen, hier sind [zwei Treppen] auf der linken Seite, allerdings kein Angriffstrupp. (Yes, I have just taken a photo, here are [two staircases] on the left but there's no attack squad.)
(b) UGV-1: eine Person gesichtet, ist [[auf der Ebene, auf der auch die Leckage ist], [hinter dem ersten Hochofen]]. (One person sighted, is [[at the same floor as the leakage] [behind the first furnace]].)

A special type of entity is a *displayed POI*. This is an icon on the digital map that represents a physical POI, like in Example 4.3(a). We use the displayed POI tag when it is clear that an expression refers to an icon on the map. In many cases it is difficult or even impossible for the annotator to distinguish between a physical POI and its displayed POI, they are used interchangeably. Example 4.3(b) illustrates this.

Example 4.3 *(a) displayed POI, (b) physical and displayed POI mix-up*
(a) TL: Can you make [a POI from the victim the photo you sent]? Over.
(b) TL: Yeah. Tango goes to [fire thirty nine] and Romeo goes to [victim thirty eight].

4.3 Roles

Role resolution is needed as a basis for tracking who is assigned which task. We mark every expression that refers to a mission participant (or a group thereof), and label it with a tag reflecting their role. Besides full NPs, including names and personal pronouns, which are also annotated as actors at the *Entity* layer, the markables at the *Roles* layer include reflexive and possessive pronouns, in order to capture really all references to mission participants. The primary roles are MC, TL, the robot operators and the robots themselves. The examples above and below provide various illustrations. In addition we had to introduce a tag for the entire team (TEAM) and tags for subgroups (OP+ROBOT, ROBOTS-PL, OTHER-PL).

Role annotation has the following challenges. First, the resolution of roles for pronouns and personal names is context dependent, and sometimes the annotator cannot figure it out. Second, an operator and their robot often act and are perceived as a single unit, which makes it difficult to distinguish between them in the annotation. Third, reference is sometimes made to some (sub)group, of the participants, and it is not clear who is meant.

4.4 Reference Units and Relations

Reference units are those expressions whose referents are, or have the potential to be, linked to the referents of other expressions by a reference relation. We annotate the syntactic type of the markable expression. Because we originally focused on entities, nominal expressions are the primary markables: we distinguish between noun phrases (*np*), pronouns (*pro*) and numerals (*num*). As we proceeded with the annotation we added other markable types: adverbial (*adv*), verb phrase (*vp*), prepositional phrase (*pp*), *clause* and *discourse*. The examples discussed further below provide various illustrations.

The inter-annotator agreement is the lowest for this layer. Disagreement concerned mostly the following pairs of labels: *adv* vs. *pp*, *np* vs. *clause*, *discourse* vs. *clause*.

Reference units are connected via *Reference links*. Our annotation scheme includes the traditional link types that we apply according to their usual definitions: *basic anaphora* as in (Deemter and Kibble, 2000), *bridging* (associative relations), *discourse anaphora* (reference to a multi-sentence descriptive passage in the dialogue) and *propositional anaphora* (reference to a statement, proposition or fact not

longer than a clause). In addition, we introduce several new link types in order to capture the kinds of relationships we observe in our data. The latter we describe below.

The *identity* link captures the cases where it is more or less explicitly asserted that two expressions have identical referents or describe the same phenomenon. Identity is typically expressed by a copula construction, but other forms are also found, like in Example 4.4(a). When the identity relation needs to be inferred, we label it *implicit identity* (see Example 4.4(b) and (c)). Sometimes, the speakers may not even know that they refer to the same entity. This may happen throughout a dialogue, and it is important that reference resolution recognizes it, to keep track of mission-relevant objects, locations and tasks. The difficulty here is that the related reference units may be far apart.

Sometimes it is difficult to differentiate between *identity* and *implicit identity*, for example, when someone is explaining something giving additional details or paraphrasing, as in Example 4.4(d).

Example 4.4 *identity vs. implicit identity link*
(a) Op1: Eeh. Team leader for operator one. I sent you a picture with [the BIO hazards]. [Codes eight and three].
(b) TL: Ok. I have [a task] for you. Emm. I create an area and then you have to [explore it].
(c) Op2: I will give [my status] for about thirty seconds. Over. ... Op2: [I am at the position of the north-west corner of the plant.]
(d) TL: Das gleiche Bild, was du jetzt gemacht hast, [da in die Mitte] reinzoomen, [da wo die Rauchen-twicklung ist]. (The same picture you've taken now but zooming [in the middle] [there where the smoke development is].)

For cases where it is not clear (to the dialogue participants or the annotator) whether identity holds between some reference units, we introduce the link types *potential identity* and *implicit potential identity*.

Example 4.5 *(a) potential vs. (b) implicit potential identity link*
(a) UAV: Es könnten [Personen] sein [das was hell leuchtet]. (It could be [people], [that what is brightly glowing].)
(b) UGV1: I see a victim. It's looks like he's sitting on [a chair]. Is that the same victim you see? UGV2: Negative. It's an... erm... maybe. M- my victim is also sitting on [an chair]. ... UGV2: UGV one, I think, I'm seeing your victim. Is also sitting on a blue chair. –probably the same chairs

These relations pose a challenge for reference resolution, because the hypothetical identity may turn out to be untrue later in the dialogue (not because of an annotator's mistake, but because of belief revision due to additional information available to the participants).

The link *asking for identity* is applied in cases when a participant poses a question concerning the identity of two or more entities. In this case the speaker may suggest a possible candidate for identity, or the speaker does not have any candidates in mind and wants to have an answer (see Example 4.6).

Example 4.6 *Asking for identity link*
UAV: Also - die Person müsste hinter [dem großen Hochofen] auf dem freien Weg sein. TL: Ist das [welcher Hochofen] ist es, der rechte oder der linke von uns aus gesehen? (So, the person must be behind [the big furnace] on the free path. [Which furnace] is it, the right one or the left one seen from our direction?) welcher Hochofen → dem großen Hochofen

Example 4.5(b) above demonstrates a combination of basic anaphora, asking for identity, implicit identity and potential implicit identity regarding the reference to the victim(s).

Because we differentiate between actual and hypothetical (or potential) entities, we also need a separate reference link type for the latter, in order to be consistent. The *intensional reference* link, illustrated in Example 4.7, serves this purpose. We use this link type also for references to generic objects.

Example 4.7 *Intensional reference link*
TL: Kannst du mir mal [ein Foto von deiner Position] machen? UGV-2: Ja ich mach dir mal [ein Snapshot]. (Can you make a photo from your position for me? - Yes, I'll make you a snapshot.)
–a photo does not exist yet

To correctly keep track of things during a mission, we also need information about entities referred to in the scope of a negation operator. We introduce the link *negative reference* illustrated in Example 4.8.

Example 4.8 *Negative reference link*
TL: Habt ihr neuen Status zum [Standort der Person oder der Chemikalien]? Ich krieg [keinen Standort]. (Do you have a new status of [the location of the person or the chemicals]? I don't get [any location].)

Challenges   Our annotation scheme includes various data-, domain- and task-specific types of reference units and relations. Such an extensive approach gives rise to additional challenges.

First of all, since we only use one type of *bridging* link, the annotated cases encompass a range of different association types. This includes the typical ones, such as set membership, part-whole, entity-attribute, as well as some rather special ones, for instance, contextual association as in Example 4.9(a) between a physical mission entity (POI) and a corresponding displayed POI, or (b) between a picture and what is depicted, as well as (c) between an adverbial pronoun (in German) and an entity. We knowingly overloaded the bridging link for the sake of the initial analysis, and intend to split it in the future.

Example 4.9 *Bridging relation: specific associations*
(a) Op1: Emm. [The fire] is on the north site. North, north-east site. Took on the corner. TL: Can eeh you put [the point of interest]?   the point of interest → the fire
(b) UGV-2: Ich hab dir gerade [ein Bild] geschickt. [Da] steht ein Stuhl auf dem Stuhl liegt ein Paket und vor dem Stuhl steht ein Paket. (I've just sent you [a picture]. [There] is a chair, on the chair lies a package, and in front of the chair lies another package.)   – da (there) → ein Bild (a picture)
(c) UGV-1: [Grüne Fass] mit Flasche [drauf]. ([Green barrel] with a bottle [on it].)
– drauf (on it) → grüne Fass (green barrel)

German adverbial pronouns sometimes refer to propositions or larger pieces of dialogue. We annotate these cases as *propositional* or *discourse anaphora*. An exception is the case when an entity referred to by an adverbial pronoun is introduced after it within the same sentence, as in Example 4.10. For now we use the identity link here, but it actually does not fully capture this kind of relation.

Example 4.10 *Adverbial pronouns: identity relation*
TL: Bitte [darauf] achten, [eine Bezeichnung auf dem Kanister zu erkennen]. (Please take care to recognize a label on the canister.)

Moreover, there are some other associative relations that we currently do not annotate. Here we have cases when an entity in plural form has several singular antecedents (Example 4.11(a)), cases with negative noun phrases as antecedents (Example 4.11(b)), and cases where a noun phrase to be resolved contains words 'more', 'another' and so on (Example 4.11(c)).

Example 4.11 *Unlabeled associative relations*
(a) UAV: [Ein Lagebild von oben] komplette Lage und [ein Lagebild zwischen den beiden Türen], verstanden. ... UAV: Ja, [beide Bilder] in Infrarot ebenfalls. ([A picture of the whole situation from above] and [a picture of the situation between the two doors], roger. ... Yes, [both pictures] also in infrared.)
(b) UGV2: UGV two. [No alert]. Over. TL: All right. I've got [a couple of them] from location you are now. Over.
(c) TL: Only I have [a picture from mmm that is him]. I don't, I don't have [more information].

Next, there are cases of *implicit identity* that we do not currently annotate, such as when we have two different antecedents and a singular entity that refers to them, like in Example 4.12.

Example 4.12 *Implicit identity: several antecedents for a singular entity*
TL: Operator one for team leader. I create an area. Can you [explore that area], please? ... TL: It's operator two for team leader. I create an area. Can you [explore that area], please? TL: Operator one and operator two. Here is team leader. Can you accept [your task]? Over.
– your task → explore that area (operator one), your task → explore that area (operator two)

Communication problems, such as in Example 4.13, also evoke annotation difficulties. Currently we do not annotate the cases, when the participants mishear, misspeak or misjudge a situation or what is on the screen, although such cases are also relevant for keeping track of entities or events accurately.

Example 4.13 *Communication problems*

Op2: Yeah. [What's area] do you mean? TL: [Area on the west site]. Over. Op2: [The area on the left site]. Ok. TL: [The area on the west site]. Over. Op2: [On the west site].

One more issue that requires further consideration is how the *double representation of reality* is annotated. Generally, we follow the convention that we treat physical objects and the corresponding symbols on the screen as the same entities, because they are used interchangeably by the mission participants. However, this is not always possible, as it can happen that the difference between a physical object and its symbolic representation is brought under discussion, like in Example 4.14. In such cases we differentiate them, but this sometimes leads to complications, for example with co-reference chains.

Example 4.14 *Real object vs. its symbolic representation*

UGV2: I'm searching for [$_{object}$ the victim] in area, where it's... where- where the picture I can see in the plot. Over. TL: You can see [$_{poi}$ a victim] in the plot. [$_{object}$ The real victim] will be more to the right side of the- [$_{poi}$ it]. Over.

Finally, our annotation of *interrogative pronouns* and noun phrases that include them is currently not quite consistent. In Example 4.15(a), we connect 'welchem Bild' with 'einem Wärmebild' and 'einem richtigen Foto' via two potential identity links, as these phrases are the only possible candidates. But we do not link 'what' and 'the whole area' in Example 4.15(b).

Example 4.15 *Interrogative pronouns*

(a) TL: Auf [welchem Bild] jetzt auf [einem Wärmebild] oder auf [einem richtigen Foto]? (On [which picture], on [a heat image] or on [a usual photo]?)

(b) TL: [What] did you explore? What- UGV-1: I did explore [the whole area].

5   Conclusions and Outlook

We presented a preliminary analysis of reference phenomena in a corpus of team communication for robot-assisted disaster response, done in preparation for developing reference resolution modules for a system that interprets such team communication to extract run-time mission knowledge and use it for various forms of teamwork assistance. Our annotation scheme has separate layers for mission-relevant entities, roles of actors, reference units with links between them, as well as comments for special cases.

Among the mission-relevant entities we focused on objects, locations and actors. Although these constitute the majority of objects referred to, there is a large number of other entities that remain to be classified in more detail. For the sake of reference resolution we found it important to distinguish between objects and locations that are known to exist, and those that are potential/hypothetical or undefined. This is however very difficult to do during annotation. Moreover, information about mission entities evolves during the dialogue, and this creates challenges for co-referential links. Another interesting challenge is the double reality representation, which means that mission objects in the physical reality and those displayed in a digital map are mostly referred to interchangeably, but sometimes need to be distinguished. In this regard we plan to review the literature on visual co-reference resolution as a next step.

Our analysis shows that a content representation using slots and fillers, as is commonly done in dialogue systems, clearly does not suffice for this domain, a proper discourse representation is required.

As for referential links, basic anaphora together with identity relationships dominate. Bridging is next, and then there are various different cases which are not that frequent but quite tricky for reference resolution. Analyzing different kinds of bridging in more detail and properly describing the other kinds of links remains a topic for our future work. The annotated corpus is currently being used for testing existing co-reference resolution models, including the AllenNLP model (Lee et al., 2017), NeuralCoref by HuggingFace (Wolf et al., 2020) and the CoreNLP framework (Manning et al., 2014). The results of these experiments will help determine our future steps for reference resolution.

Acknowledgements

References

Berfin Aktaş, Tatjana Scheffler, and Manfred Stede. 2018. Anaphora resolution for Twitter conversations: An exploratory study. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 1–10.

Tatiana Anikina and Ivana Kruijff-Korbayová. 2019. Dialogue act classification in team communication for robot assisted disaster response. In *Proceedings of SIGDIAL 2019*.

Amit Bagga and Breck Baldwin. 1999. Cross-document event coreference: Annotations, experiments, and observations. In *Coreference and Its Applications*.

Simon Philip Botley. 2006. Indirect anaphora: Testing the limits of corpus-based linguistics. *International Journal of Corpus Linguistics*, 11(1):73–112.

Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Comput. Linguist.*, 22(2):249–254, June.

Tommaso Caselli and Irina Prodanof. 2010. Annotating event anaphora: A case study. In *LREC*.

K. Bretonnel Cohen, Arrick Lanfranchi, Miji Joo-young Choi, Michael Bada, William A. Baumgartner, Natalya Panteleyeva, Karin Verspoor, Martha Palmer, and Lawrence E Hunter. 2017. Coreference annotation and resolution in the Colorado Richly Annotated Full Text (CRAFT) corpus of biomedical journal articles. *BMC bioinformatics*, 18(1):1–14.

Parag Pravin Dakle, Takshak Desai, and Dan Moldovan. 2020. A study on entity resolution for email conversations. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 65–73.

Kees van Deemter and Rodger Kibble. 2000. On coreferring: Coreference in MUC and related annotation schemes. *Computational linguistics*, 26(4):629–637.

Richard Eckart de Castilho, Éva Mújdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann. 2016. A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 76–84, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Pradheep Elango. 2005. Coreference resolution: A survey. *University of Wisconsin, Madison, WI*.

Steve Fligelstone. 1992. Developing a scheme for annotating text to show anaphoric relations. *New Directions in English Language Corpora: Methodology, Results, Software Developments*, pages 153–170.

Niyu Ge. 1998. Annotating the Penn treebank with coreference information. *Technical report*.

Michael A.K. Halliday and R. Hasan. 1976. *Cohesion in English*. Longman, London, U.K.

Laura Hasler, Constantin Orasan, and Karin Naumann. 2006. NPs for events: Experiments in coreference annotation. In *LREC*, pages 1167–1172. Citeseer.

Lynette Hirschman and Nancy Chinchor. 1998. Appendix F: MUC-7 coreference task definition (version 3.0). In *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998*.

Ivana Kruijff-Korbayová and Geert-Jan M Kruijff. 2004. Discourse-level annotation for investigating information structure. In *Proceedings of the Workshop on Discourse Annotation*, pages 41–48.

---

[2]rettungsrobotik.de

Ivana Kruijff-Korbayová, Francis Colas, Mario Gianni, Fiora Pirri, Joachim de Greeff, Koen Hindriks, Mark Neerincx, Petter Ögren, Tomáš Svoboda, and Rainer Worst. 2015. TRADR project: Long-term human-robot teaming for robot assisted disaster response. *KI - Künstliche Intelligenz*, 29(2):193–201, Jun.

Kenton Lee, Luheng He, M. Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *EMNLP*.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Tony McEnery, Izumi Tanaka, and Simon Botley. 1997. Corpus annotation and reference resolution. In *Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*.

Malvina Nissim, Shipra Dingare, Jean Carletta, and Mark Steedman. 2004. An annotation scheme for information status in dialogue. In *LREC*. Citeseer.

Massimo Poesio, Florence Bruneseaux, and Laurent Romary. 1999. The MATE meta-scheme for coreference in dialogues in multiple languages. In *Towards Standards and Tools for Discourse Tagging*.

Massimo Poesio, Ron Artstein, et al. 2008. Anaphoric annotation in the ARRAU corpus. In *LREC*.

Massimo Poesio. 2004. The MATE/GNOME proposals for anaphoric annotation, revisited. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004*, pages 154–162.

TRADR project website. 2020. Long-term human-robot teaming for robot-assisted disaster response (TRADR). http://www.tradr-project.eu/. Accessed: 2020-04-30.

Marco Rocha. 1999. Coreference resolution in dialogues in English and Portuguese. In *Coreference and Its Applications*.

Andrea Setzer and Robert J Gaizauskas. 2000. Annotating events and temporal information in newswire texts. In *LREC*, volume 2000, pages 1287–1294. Citeseer.

Joel Tetreault, Mary Swift, Preethum Prithviraj, Myroslava O Dzikovska, and James Allen. 2004. Discourse annotation in the Monroe corpus. In *Proceedings of the Workshop on Discourse Annotation*, pages 103–109.

Christian Willms, Constantin Houy, Jana-Rebecca Rehse, Peter Fettke, and Ivana Kruijff-Korbayová. 2019. Team communication processing and process analytics for supporting robot-assisted emergency response. In *International Conference on Safety, Security, and Rescue Robotics (SSRR)*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Huggingface's transformers: State-of-the-art natural language processing.

Kristin Wright-Bettner, Martha Palmer, Guergana Savova, Piet de Groen, and Timothy Miller. 2019. Cross-document coreference: An approach to capturing coreference without context. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 1–10.

Heike Zinsmeister and Stefanie Dipper. 2010. Towards a standard for annotating abstract anaphora. In *LREC*, pages 54–59.