Spatial descriptions in discourse: choosing a perspective

Simon Dobnik
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg, Sweden

Aims
Create a corpus of dialogue interactions including projective spatial descriptions of perceptual scenes. Examine how speakers integrate each other’s choice of the reference frame in the dialogue common ground and align it. Build and evaluate a conversational agent using these strategies.

Reference frame/perspective choice

Where is the table?

- **Intrinsic**: The table is behind the chair.
- **Relative**: The table is to the left/right of the chair (from my point of view/from there).
- **Extrinsic**: The table is north of/above the chair.

Reference frames in conversation

Strategies that speakers may follow:
1. Always use the relative RF to yourself.
2. Always align to the RF used by your partner in the previous turn.
3. If the reference objects provides orientation, choose its intrinsic RF.
4. Align to the RF used by your partner after they have used it for n turns.
5. For each turn select the RF randomly.
6. Keep a randomly chosen RF for n turns, then change.

Online experimental environment

http://goo.gl/8KLja

Study #1: Participants at the opposite side of the room are instructed to describe the room to each other in a free dialogue. 7 conversations.

- Participants focused less on spatial descriptions but expressed opinions, etc.
- They aligned to a role where one was asking questions and the other was describing.
- (Only) a few examples of the intended interaction.

Results

- Participants align their perspective but only if their confederate uses a particular perspective consistently over several turns (priming).
- The majority of descriptions involve perspective relative to the speaker or the hearer: “the first chair on my right.”
- < 10% of descriptions contain intrinsic perspective to reference objects: “the chair opposite the chair I chose previously”.
- Intrinsic reference frame is likely to be aligned in the following confederate’s utterance.
- There may be anaphoric descriptions: “the chair that is to the left of the one you chose previously”, “the same one” and “the chair on the left that wasn’t selected in the preceding two steps”.
- Agents must keep track of all previously chosen objects and reference frames.
- In the study #1 the participants mostly omit the explicit definition of perspective while in the study #2 this is mostly always defined (“from your side”) even if they are aligned.
- Consequently, the accuracy (measured by successful clicks) is high: > 95%.
- Speakers not only align perspective but also their choice of lexical items and the way the scene is described syntactically.

Further work

1. Collect more conversations online and tag them.
2. Estimate frequencies of perspective use, distribution of a particular perspective use over a conversation, the mean number of consistent perspective use before the alignment occurs, etc.
3. Excluding anaphoric uses (minority cases) we have for each object a set of best referring expressions with perspective information.
4. Integrate the descriptions in a dialogue manager and repeat the experiment where one partner is a computer. Allow humans to rate the performance of the system.

Example of annotation

```
<p utterance="description" at="1345032975" speaker="1" object="4"> izbral sem si <lo>srednji stol</lo>
<relation perspective="relative" origin="s1" explicit="1" correct="1">na moji desnili strani</relation> <lo>
<p utterance="interpretation" at="1345033044" speaker="2" object="4" perspective="relative" origin="s1" correct="1">Sogovornik 2 je izbral stol.</p>
```

Human annotation involves identifying:
1. Reference objects (<ref>), located objects (<lo>) and relations (<relation>) in descriptions.
2. The attributes perspective, origin, explicit and correct in relations.

Study #2: Only one kind of objects are used. Rotating between pairs of turns: (i) one person describes the location of a pre-selected chair; (ii) the other confirms that they understood the description by clicking on that chair. 11 conversations in Slovenian and 2 conversations in English each containing 20 turns per participant.

Speaker A: “I chose the middle chair on my north of/above the chair.”

Speaker B: Clicks on one of the chairs.

Automatic logging to XML files with partial annotation (linguistic and perceptual data).