Computational Semantics

with Pictures

Week 3 - The Tasks & Language Data

Project Module, Summer 2019
David Schlangen
david.schlangen@uni-potsdam.de

organisational matters

- you will need access to our file & compute server

- you can get an account from Siegfrid Wrobel & assistants (2.25 or 2.21)

Recap

language & the world (via vision)

A couple of women standing on a tennis court shaking hands.

Two female tennis players shaking hands across the net.

Denotation

Implication

Situations

A woman in sandals stands next to clocks.

A woman's feet standing next to four clock faces on a floor.

Types of relations

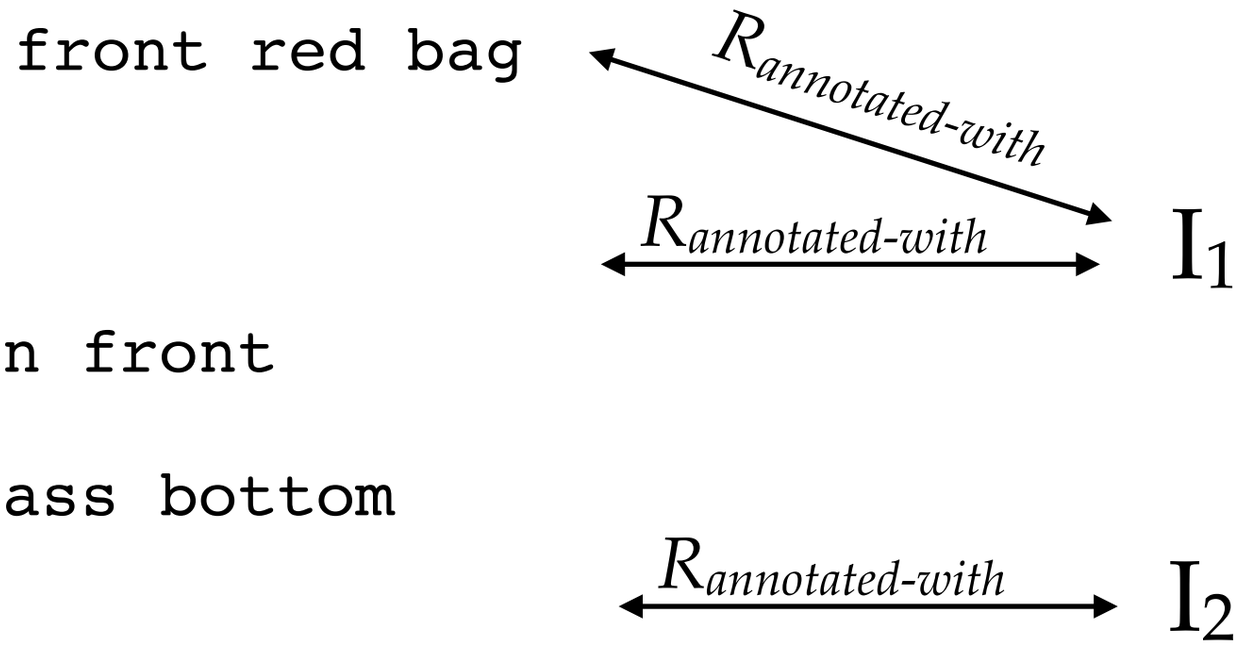

Image Corpora

- SAIAPR: holiday snaps. Objects in images segmented and classified.

- Flickr30k: misc pictures, focus on events. Some objects segmented and classified (WordNet).

- COCO: common objects in context. 80 types of objects, in various kinds of context. These objects segmented and classified.

- Visual Genome, ADE20k, CUB Birds

today

- The Image Data (cont.)
- The Language Data
  - Tasks, Environment, and Games
  - Example Tasks (which might give you ideas for your projects)
- Homework: Small projects (analysis, more datasets, etc.)

Our Data

- The Image Corpora provide… the images.
- Often, the corpus distribution also includes object annnotation / segmentation.
- Then there is the task specific language data that relates to the images. This is sometimes provided together with the original distribution, and sometimes (often) added by other people and distributed separately.
- We will make a distinction between the Image Data, and the Language Data, with exemplifies a Language Task.

Task, Data, Competence

- A Language Task is a *mapping* from a *state* to an *action*, where either state or action (or both) involve language.

- An LT is typically defined intensionally through a verbal description ("provide a description that captures the essence of this image").

- In reality, the more common specification is extensionally through a *data set*.

- Data sets should be *verified*, and *validated*. (Are there error free? Do they actually represent the intensional definition well?)

- Tasks should be *motivated*. If useful for an application, great. If not, need to argue that they represent *language competence* which is separable from other competences.

Tasks

Definition 1 *A* Language Task *is a tuple* $(S, A, \mathcal{L}, D_T)$*, where:*

- S *is a (possibly infinite) set of* states,
- A *is a (possibly infinite) set of* actions,
- *with either the states in* S *or the actions in* A *(or both) having as part* natural language expressions*, and*
- $\mathcal{L} : S \rightarrow A$ *is a function that maps a state* $s \in S$ *to an action* $a \in A$*, where*
- *the mapping* $\mathcal{L}$ *conforms to* task description D_T.

And Motivations

Motivation The original paper by Antol et al. (2015) that introduced the task and the first large-scale dataset provides a veritable smorgasbord of motivations, which are worth citing in full and relating to the discussion from above:[7]

- *"[There is] a belief that multi-discipline tasks like image captioning are a step towards solving AI."*

  The goal is "solving AI", and this goal can be approached in steps.

- *"What makes for a compelling AI-complete task? We believe that in order to spawn the next generation of AI algorithms, an ideal task should (i) require multi-modal knowledge beyond a single sub-domain (such as CV) and (ii) have a well-defined quantitative evaluation metric to track progress."*

  The task of visual question answering is "AI-complete"—which is typically used to mean that it comprises all aspects of general intelligence, and solving it is equivalent to solving general AI—as it requires multi-modal knowledge.[8]

- *"Open-ended questions require a potentially vast set of AI capabilities to answer – fine-grained recognition (e.g., What kind of cheese is on the pizza?), object detection (e.g., How many bikes are there?), activity recognition (e.g., Is this man crying?), knowledge base reasoning (e.g., Is this a vegetarian pizza?), and commonsense reasoning (e.g., Does this person have 20/20 vision?, Is this person expecting company?)."*

  In particular, it involves these capabilities, which then appear to be necessary (or, if the term "AI-complete" is to be taken seriously, sufficient) for solving AI.

Our Data

- We will look at existing datasets (for tasks), and at datasets (and tasks) that we can derive. (Like the entailment / implication example that we've seen last week.)

Homework

- In a group of 2-3 people, pick one of the tasks from *Intermediate Projects I* (or come up with something similar).
    - Analysis tasks ask you to investigate the data. Make linguistically interesting observations. Validate the data. Etc.
    - Processing tasks ask you to bring more datasets into our joint format, or to automatically clean the dataset.
- Discuss general idea via piazza, by Monday April 29 (so that I know what is being done, and not everyone is doing the same thing).
- Send me documentation (ideally, notebook) by Tuesday May 7.
- (No class next week.)

The rest of the course

- see website, schedule