Natural Language and Spatial Reasoning

Stefanie Tellex
What is he doing?
Where should the robot go?

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a whiteboard. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
What semantic structures can enable a system to understand and use spatial language in realistic situations?
Methodology

- Pick a subset of language.
- Collect a corpus focusing on that subset.
- Make a program that understands language in the corpus.
- See how well the program works.
- Analyze why the program works.
Methodology

- Pick a subset of language.
- Collect a corpus focusing on that subset.
- Make a program that understands language in the corpus.
- See how well the program works.
- Analyze why the program works.
- Do this in more than one domain.
Outline

- Problem Statement
- Spatial Prepositions
- Spatial Language Video Retrieval
- Direction Understanding
- Next Steps
- Contributions
Spatial Language is Pervasive

“Where is the bus stop?”

“Take a right at the elevators.”

“Through the years you've never let me down.”

“This book jumps across many different topics.”
He's going across the kitchen.
Grounding

He's going across the kitchen.
He's going across the kitchen.
Grounding

He's going across the kitchen.
Grounding

He's going across the kitchen.
Spatial Prepositions
Spatial Prepositions in English

about, above, across, after, against, along, alongside, amid(st), among(st), around, at, atop, behind, below, beneath, beside, between, betwixt, beyond, by, down, from, in, inside, into, near, nearby, off, on, onto, opposite, out, outside, over, past, through, throughout, to, toward, under, underneath, up, upon, via, with, within, without, far from, in back of, in between, in front of, in line with, on top of, to the left of, to the right of, to the side of

(from Landau and Jackendoff, 1993)
the, of, and, to, a, in, that, is, was, he, for, it, with, as, his, on, be, at, by, i, this, had, not, are, but, from, or, have, an, they, which, one, were, you, all, her, she, there, would, their, we, him, been, has, when, who, will, no, more, if, out, its, so, up, said, what, about, than, into, them, can, only, other, time, new, some, could, these, two, may, first, then, do, any, like, my, now, over, such, our, man, me, even, most, made, after, also, well, did, many, before, must, years, back, through, much, where, your, way, down, should, because, long, each, just, state, people, those, too, how, Mr., little, good, world, make, very, year, still, see, own, work, men, day, get, here, old, between, both, life, being, under, three, never, know, same, last, another, while, us, off, might, great, states, go, come, since, against, right, came, take, used, himself, few, house, American, use, place, during, high, without, again, home, around, small, however, found, mrs, part, school, thought, went, say, general, once, upon, every, left, war, don't, does, got, united, number, hand, course, water, until, always, away, public, something, fact, less, through, far, put, head, think, called, set, almost, enough, end, took, government, night, yet, system, better, four, nothing, told, eyes, city, going, president, why, days, present, point, didn't, look, find, asked, second, group, later, next, room, social, business, knew, program, give, half, side, face, toward, white, five, let, young, form, given, per, order, large, several, national, important, possible, rather, big, among, case, often, early, john, things, looked, ever, become, best, need, within, felt, along, children, saw, church, light, power, least, family, development, interest, others, open, thing, seemed, want, area, god, members, mind, help, country, service, turned, door, done, law, although, whole, line, problem, sense, certain, different, kind, began, thus, means, matter, perhaps, name, times, york, itself, action, human, above, week, company, free, example, hands, local, show, history, whether, act, either, gave, death, feet, today, across, body, past, quite, taken, anything, field, having, seen, word, car, experience, I'm, money, really, class, words, already, college, information, tell, making, sure, themselves, together, full, air, shall, held, known, period, keep, political, real, miss, probably, century, question, seems, behind, cannot, major, office, brought, special, whose, boy, cost, federal, economic, self, south, problems, heard, six, study, ago, became, moment, run, available, job, street, result, short, west, age, change, position, board, individual, reason,
With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a whiteboard. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
With your back to the windows, walk the elevators. Continue to walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a whiteboard. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Modeling Spatial Prepositions

- Functions that take geometric arguments.
- Classifiers for “to,” “across,” “towards,” “through,” “around,” etc.
- Library of features.
- Train and test on labeled examples.
Across Video
peakDistanceToAxes = max(dist(figure, axes))
endPointsInGroundBoundingBox
displacementFromGround = d2 - d1
Two Different Domains

- Video Retrieval
- Direction Understanding
The person is going...
Corpus

to the counter.
along the east side of the island.
from the refrigerator.
to the cabinet.
across the kitchen.
Training Classifiers for Spatial Prepositions

Binary classifiers.

Positive examples of “across” are negative examples of “to.”

Some exceptions (e.g., “to” and “towards.”)
Corpus

Positive example of “to the counter”
Negative example of “across the counter”
endPointsInGroundBoundingBox
“to the sink”

Full model: 0.99
Bounding boxes: 0.76
“to the sink”

Full model: 0.10
Bounding boxes: 0.63
Results

- Modeled six spatial prepositions.
- Evaluated on a corpus of natural language descriptions of video clips.
- Analyzed which features work.
Limitations

- Phrases, not sentences or paragraphs.
- All in one room.
- Only six spatial prepositions.
Direction Understanding

Joint work with Thomas Kollar
Where should the robot go?

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a whiteboard. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Structure of Natural Language Directions

- Sequential
With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Sequential

With your back to the windows, walk straight through the door near the elevators. **Continue to walk straight**, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Sequential

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Sequential

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Sequential

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. **Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant")**.
Sequential

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Structure of Natural Language Directions

- Sequential.
- Directive verbs.
Directive Verbs

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Structure of Natural Language Directions

- Sequential.
- Directive verbs.
- Landmarks.
With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Structure of Natural Language
Directions

- Sequential.
- Directive verbs.
- Landmarks.
- Spatial relations.
Spatial Relations

With your back to the windows, walk straight through the door near the elevators. Continue to walk straight, going through one door until you come to an intersection just past a white board. Turn left, turn right, and enter the second door on your right (sign says "Administrative Assistant").
Structure of Natural Language Directions

- Sequential.
- Directive verbs.
- Landmarks.
- Spatial relations.
Spatial Description Clause

Go through the door.

- Figure: <you>
- Verb: “Go”
- Spatial relation: “through”
- Landmark: “the door.”
Spatial Description Clause

Continue to walk straight, going through one door until you come to an intersection just past a white board.

V: Continue to walk, SR: straight
Spatial Description Clause

Continue to walk straight, going through one door until you come to an intersection just past a white board.

V: going, SR: through, L: one door
Spatial Description Clause

Continue to walk straight, going through one door until you come to an intersection just past a white board.

V: going, SR: until, L: you come to an intersection just past a white board
Spatial Description Clause

Continue to walk straight, going through one door until you come to an intersection just past a white board.

F: you, V: come, SR: to, L: an intersection just past a white board
Spatial Description Clause

Continue to walk straight, going through one door until you come to an intersection just past a white board.

F: an intersection, SR: just past, L: a white board
Spatial Description Clause

- S
- SDC
- SR
- L

- V
  - Continue to walk
    - going
      - through
        - one door
          - until
            - SDC
              - SR
                - L

- V
  - SDC
    - SR
      - L

- F
  - you
    - come
      - to
        - SDC
          - SR
            - L

- F
  - an intersection
    - just past
      - a whiteboard
Model

$$\text{argmax}_{\text{path}} p(\text{path}, \text{directions}, \text{observations})$$
Model

\[ p(path, directions, observations) = \]
\[ p(directions | path, observations) \cdot p(path, observations) \]
(definition of conditional probability)

\[ p(directions | path, observations) \cdot p(path) \cdot p(observations) \]
(path and observations are independent)

\[ p(sdc_1 \ldots sdc_M | vp_1 \ldots vp_{M+1}, o_1 \ldots o_K) \cdot p(vp_1 \ldots vp_{M+1}) \cdot p(o_1 \ldots o_K) \]
(topological map of viewpoints, and a sequence of SDCs)
Topological Map
Model

\[ p(path, directions, observations) = \]
\[ p(directions|path, observations) p(path, observations) \]
(definition of conditional probability)

\[ p(directions|path, observations) p(path) p(observations) \]
(path and observations are independent)

\[ p(sdc_1 ... sdc_M|vp_1 ... vp_{M+1}, o_1 ... o_K) p(vp_1 ... vp_{M+1}) p(o_1 ... o_K) \]
(topological map of viewpoints, and a sequence of SDCs)

\[ \prod_{i=1}^{M} p(sdc_i|vp_i, vp_{i+1}, o_1 ... o_K) \prod_{i=1}^{M} p(vp_{i+1}|vp_1 ... vp_i) p(o_1 ... o_K) \]
(Paths are Markov)
Model

\[
p(path, \text{directions}, \text{observations}) = p(directions|path, \text{observations}) p(path, \text{observations})
\]

(definition of conditional probability)

\[
p(directions|path, \text{observations}) p(path) p(\text{observations})
\]

(path and observations are independent)

\[
p(sdc_1 \ldots sdc_M|vp_1 \ldots vp_{M+1}, o_1 \ldots o_K) p(vp_1 \ldots vp_{M+1}) p(o_1 \ldots o_K)
\]

(topological map of viewpoints, and a sequence of SDCs)

\[
\prod_{i=1}^{M} p(sdc_i|vp_i, vp_{i+1}, o_1 \ldots o_K) \prod_{i=1}^{M} p(vp_{i+1}|vp_1 \ldots vp_i) p(o_1 \ldots o_K)
\]

(Paths are Markov)
Probability of Directions Given a Possible Movement

\[ p(sdc_i | \text{vp}_i, \text{vp}_{i+1}, o_1 \ldots o_K) \]

\[ = p(\text{figure}_i, \text{verb}_i, \text{spatial}_i, \text{landmark}_i | \text{vp}_i, \text{vp}_{i+1}, o_1 \ldots o_K) \]

(The structure of a spatial description clause.)
Spatial Description Clause

Go through the door.

• Figure: <you>
• Verb: “Go”
• Spatial relation: “through”
• Landmark: “the door.”
Probability of Directions Given a Possible Movement

\[ p(sdc_i|vp_i, vp_{i+1}, o_1...o_K) \]
\[ = p(fig i, verb_i, spatial\_relation_i, landmark_i|vp_i, vp_{i+1}, o_1...o_K) \]
(The structure of a spatial description clause.)
\[ = p(fig i|vp_i, vp_{i+1}, o_1...o_K) \times p(verb_i|vp_i, vp_{i+1}) \times \]
\[ p(spatial\_relation_i|landmark_i, v_i, v_{i+1}, o_1...o_K) \times \]
\[ p(landmark_i|v_i, v_{i+1}, o_1...o_K) \]
(Factor according to certain independence assumptions.)
Probability of Directions Given a Possible Movement

\[ p(sdc_i|vp_i, vp_{i+1}, o_1...o_K) \]

\[ = p(figure_i, verb_i, spatial\_relation_i, landmark_i|vp_i, vp_{i+1}, o_1...o_K) \]

(The structure of a spatial description clause.)

\[ = p(figure_i|vp_i, vp_{i+1}, o_1...o_K) p(verb_i|vp_i, vp_{i+1}) \times \]
\[ p(spatial\_relation_i|landmark_i, v_i, v_{i+1}, o_1...o_K) \times \]
\[ p(landmark_i|v_i, v_{i+1}, o_1...o_K) \]

(Factor according to certain independence assumptions.)
Observations
Contextual relationships

Tags include: desktop, monitor, computer, keyboard, and mouse
Probability of Directions Given a Possible Movement

\[
p(sdc_i|vp_i, vp_{i+1}, o_1...o_K) \\
= p(figure_i, verb_i, spatial\_relation_i, landmark_i|vp_i, vp_{i+1}, o_1...o_K) \\
(\text{The structure of a spatial description clause.}) \\
= p(figure_i|vp_i, vp_{i+1}, o_1...o_K) \times p(verb_i|vp_i, vp_{i+1}) \times \\
p(spatial\_relation_i|landmark_i, v_i, v_{i+1}, o_1...o_K) \times p(landmark_i|v_i, v_{i+1}, o_1...o_K) \\
(\text{Factor according to certain independence assumptions.})
\]
Probability of Directions Given a Possible Movement

\[ p(sdc_i|vp_i, vp_{i+1}, o_1...o_K) \]

\[ = p(figure_i, verb_i, spatial\_relation_i, landmark_i|vp_i, vp_{i+1}, o_1...o_K) \]

(The structure of a spatial description clause.)

\[ = p(figure_i|vp_i, vp_{i+1}, o_1...o_K) \times p(verb_i|vp_i, vp_{i+1}) \times \]

\[ p(spatial\_relation_i|landmark_i, v_i, v_{i+1}, o_1...o_K) \times \]

\[ p(landmark_i|v_i, v_{i+1}, o_1...o_K) \]

(Factor according to certain independence assumptions.)
Learning Distributions for Spatial Relations

Draw a path going down the hallway.
Learning Distributions for Spatial Relations

Draw a path going down the hallway.
Performance of Classifiers
High Scoring Examples of “past”
Low Scoring Examples of “past”
Corpus of Natural Language
Directions

Instructions:
Write, in whatever way is natural to you, directions from the start label to the end label, with the goal of communicating this route to another person who does not have a map.
Corpus of Natural Language Directions

- Study 1 (Stata 3)
  - 10 people, 16 routes
- Study 2 (Stata 8)
  - 15 people, 10 routes
- Study 3 (Stata 1)
  - 15 people, 10 routes
- Study 4 (outdoors)
  - ongoing
# Human Performance

<table>
<thead>
<tr>
<th></th>
<th>Stata 3</th>
<th>Stata 8</th>
<th>Stata 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>All directions</td>
<td>63%</td>
<td>85%</td>
<td>86%</td>
</tr>
<tr>
<td>Best Direction Giver</td>
<td>87%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Worst Direction Giver</td>
<td>43%</td>
<td>30%</td>
<td>20%</td>
</tr>
</tbody>
</table>
Inference

• Global.
  – Viterbi-style algorithm.
  – Last-SDC only.
  – Landmarks only.

• Local.

• Random.
Comparison to Baselines

Performance Over All Subjects

proportion correct

distance from destination (m)

Random
Comparison to Baselines

Performance Over All Subjects

- **Proportion Correct**
- **Distance from Destination (m)**

- **Random**
- **Last SDC only**
Comparison to Baselines

Performance Over All Subjects

- Random
- Last SDC only
- Landmarks only
Comparison to Baselines

Performance Over All Subjects

- Random
- Last SDC only
- Landmarks only
- Local inference
Comparison to Baselines

Performance Over All Subjects

proportion correct

Random
Last SDC only
Landmarks only
Local inference
Global inference

distance from destination (m)
With and Without Spatial Relations

Performance Over All Subjects

- Local inference, +sr
- Global inference, +sr
With and Without Spatial Relations

Performance Over All Subjects

- Local inference, +sr
- Global inference, +sr
- Global inference, -sr

Proportion correct vs. distance from destination (m)
With and Without Spatial Relations

Performance Over All Subjects

- Local inference, +sr
- Global inference, +sr
- Global inference, -sr
- Local inference, -sr

proportion correct

distance from destination (m)
With and Without Spatial Relations

Performance Over One Subject

- Local inference, +sr
- Local inference, -sr
- Global inference, +sr
- Global inference, -sr
Improving Direction Understanding

Results

- Polygon landmarks.
- Add more spatial relations.
- Use more complex environments.
- Changing the model.
  - Alignment.
  - Conditional Random Fields.
- Exploration.
Next Steps

• Video retrieval.
  – Collect a corpus of paragraph-length descriptions of a person's activity.

• Direction generation.

• Compare spatial prepositions across domains.

• Create a lexicon of spatial prepositions.
Schedule

- December 15, 2010 – JAIR, direction understanding
- January 22, 2010 – SIGIR, video retrieval
- March 1, 2010 - TiCS on generation.
- April 1, 2010 – JAIR, lexicon of spatial prepositions.
- April 14, 2010 - thesis outline to Deb.
- July, 2010 – thesis defense
Related Work

- Talmy (2005)
- Landau and Jackendoff (1993)
- MacMahon et al. (2005)
- Regier (1992)
- Regier and Carlson (2001)
- Levit and Roy (2007)
- Skubic et al. 2002
- Bauer et al. 2009
- Look et al. (2005)
- This Work
Related Work – Video Retrieval

• Katz et al. (2004) - “Show me all the cars leaving the garage.”

• Ivanov and Wren (2006) – Query by example trajectory

• Fleischman et al. (2006) - “Show me people making coffee.”

• Naphade et al. (2006) - Large-Scale Concept Ontology for Multimedia
Contributions

- Create a library of features for grounding spatial prepositions.
- Analyze which features perform best for specific prepositions.
- Compare semantics of spatial prepositions in two different domains.
- Model higher level structures.
Acknowledgements

Tom Kollar
Deb, Yuri, Boris, Cynthia,
Cogmac
ML support staff
Gremio, Dave, Carie, Kai-yuh, Lin
Ania & Seth & Moby
Grandma Patches and Grandma Nicholas
Mom, Dad, Staci & Seth, Shannon & Jon, Scott
Contributions

- Create a library of features for grounding spatial prepositions.
- Analyze which features perform best for specific prepositions.
- Compare semantics of spatial prepositions in two different domains.
- Model higher level structures.
Collaboration

- My thesis
  - Library of features for spatial prepositions.
  - Analyze which features perform the best.
  - Comparison of meanings of spatial relations in different domains.

- Tom's thesis
  - Grounding landmarks with flickr.
  - Planning to find objects.

- Both
  - Spatial description clauses.
  - MRF model of spatial language.
Directions With Only Landmarks

the elevators.

you

door

one door

the windows, an intersection a whiteboard.

the second door (sign

"Administrative Assistant").
Why is this a Media Lab thesis?

- Towards a new (or old) form of human-machine interaction.
- Multi-modal, multi-domain interface.
- Defining a new problem, with a new corpus.
Extracting Spatial Description
Clauses

• CRF chunker tags each word.
• Trained from annotated data.
• Features (tri-grams)
  o Words
  o Part-of-speech tags
  o Label sequence
• Greedy algorithm groups tagged words together.
Continue to walk straight, going through one door until you come to an intersection just past a white board.
Continue to walk straight, going through one door until you come to an intersection just past a white board.

V: Continue to walk straight
Extracting Spatial Description Clauses

Continue to walk straight, going through one door until you come to an intersection just past a white board.

V: going, SR: through, L: one door
Continue to walk straight, going through one door until you come to an intersection just past a white board.

SR: until, L: you come to an intersection
Extracting Spatial Description Clauses

Continue to walk straight, going through one door until you come to an intersection just past a white board.

SR: just past, L: a white board
Extracting Spatial Description Clauses

![Bar chart showing fraction of matched annotations for different categories: F, V, SR, L, SR and L, All. The chart indicates varying degrees of matched annotations across these categories.](image-url)
Computational Linguistics

• Maron (1961)
  – First bag-of-words paper, using Naïve Bayes.

• Cleverdon (1967)
  – First corpus-based information retrieval evaluation.
Related Work

- **Cognitive Semantics**
  - Landau and Jackendoff (1993)
  - Talmy (2005)

- **Cognitive Science**
  - Regier (1992)
  - Regier and Carlson (2001) – attention vector sum for “above”
  - Carlson and Covey (2005) - “How far is near? Inferring distance from spatial descriptions.”
Related Work

- Turing (1950)
- Winograd (1971)
  - Shrdlu
- Harnad (1990)
  - Grounding