Human Language Technology: Applications to Information Access

Lesson 10: Automatic Analysis of Human Interactions

December 22, 2016

EPFL Doctoral Course EE-724
Andrei Popescu-Belis, Idiap Research Institute

Text structure

- How to access information that is enclosed in the text structure as a whole?
  - as opposed to individual sentences
  - "text" includes here monologues and dialogues

- Example: movie reviews (sent. analysis data)
  - not only a list of positive and negative feelings
  - some sentences matter more than others
    - e.g. a long positive analysis of a minor point followed by a short negative statement on a major issue

Theories of text structure

- Tree-based structure
  - monologues: Rhetorical Structure Theory (Mann & Thompson 1988)
    - sentences have *functions* (e.g. elaboration, condition, contrast, temporal) and are organized into *discourse trees* (e.g., Marcu 2000)
  - dialogues: e.g., Grosz and Sidner 1986

- While sentences have tree-like underlying representations (parse trees), it is not clear whether texts have such things

- Local theories of structure: especially important for dialogues
  - *functions* of sentences: dialogue acts | decisions | argumentation
  - *local relations* between sentences: adjacency pairs

Plan

- Layers of dialogue analysis
- Example of the AMI Meeting Corpus
- Dialogue acts
- Argument modeling

Layers of dialogue analysis

- Spoken/synchronous or written/asynchronous dialogues
- Why is a conversation different from a monologue?
  - monologues focus on “informing” and have explicit structure
  - in dialogues, utterances can “do” more things and structure is less explicitly marked

- Layers
  - sentence-level: words, POS, syntax/semantics
  - segmentation into functional units → utterances
  - detection of the functions of utterances
  - construction of an argumentative structure
  - segmentation into topical episodes

Example: the AMI Meeting Corpus

- Recordings of 177 technical meetings: 100 hrs
  - multimodal recordings: audio, video, slides, pens
  - manually transcribed and annotated in layers
    - following precise guidelines (annotation manuals)
    - some annotations do not cover entire data (too expensive)
    - human annotations can serve as training/test data for automatic annotation systems ➔ definition of tasks
  - available from http://corpus.amiproject.org
- Other conversational corpora for HLT
  - Switchboard, ICSI MR, Fisher, etc.

Instrumented meeting room for recordings at Idiap

Instrumented meeting room for recordings at the U. of Edinburgh

Annotations of the AMI Corpus: layers and	number of occurrences
words	1,207,769
named entities	14,230
speech segments	69,258
topics	1,879
dialogue acts	117,043
adjacency pairs	26,825
abstractive summaries	2,578
extractive summaries	19,216
abstractive/extractive links	22,101
participant summaries	3,409
visual focus of attention	31,271
hand gesture	1,453
head gesture	36,257
argument structures	6,920
argumentation relations	4,759

Functions of utterances in dialogue

- Several dimensions with multiple values

1. Speech acts: assert, suggest, promise, question, …
2. Turn-taking management: backchannel, floor grabber, …
3. Adjacency pairs: question/answer, offer/refuse, …
4. Overall organization: opening, change topic, …
5. Politeness management: face saving, face threatening, …

- An utterance may fulfil a function in several dimensions, but often has a dominant one: dialogue act

DIALOGUE ACT SEGMENTATION AND LABELING (TAG SETS)

Automatic dialogue act segmentation over the AMI Corpus: sample study

- Goal: assign label 'boundary' or 'not boundary' to every word

- Features commonly used
  - *time-related features* (duration of words and pauses, etc.)
  - *lexical information* (current, previous, and next word and their POS)
  - *prosodic features* (pitch, energy, talking speed)
  - *segment-related features* generated online (e.g. #words in prev. seg.)

- Experiments by op den Akker and Schulz (2008)
  - best performing feature subset: pause information, mean duration of word, specific current words (*mm-hmm, but, yeah, so, okay, and) and previous words (okay),* POS, minimum/mean energy, speech flow change, length of segment

- Scores
  - baseline algorithm: a word is a segment boundary if preceded by a pause
    - low recall: 0.55 | high precision: 0.97 | F-measure: 0.70
  - more features, better classifier: F-measure of up to 0.76

A well-known DA tagset: DAMSL

- *Dialog Act Markup in Several Layers*

- Four dimensions > each with sub-dimensions > tags
  - Information Level: 4 non-exclusive tags
  - Communicative Status: 3 non-exclusive tags
  - Forward-Looking Function
    - 7 sub-dimensions with 3-4 non-exclusive tags each
      - e.g. statement, info-request, committing-speaker-future-action
  - Backward-Looking Function
    - 3 sub-dimensions
      - agreement: 6 non-exclusive tags
      - understanding: 6 non-exclusive tags
      - answer: yes/no
- About 4 million possible labels!
  - search space is very large for automatic DA annotation

SWBD-DAMSL tagset

- Derived from occurring combinations of DAMSL tags (= labels)
  - observed on ca. 200,000 tagged utterances of Switchboard data
  - 220 unique tags

  → clustered into 42 final tags

- Examples of most frequent tags
  - Statement 36%
  - Continuer 19%
  - Opinion 13%
  - Agree/Accept 5%
  - Abandoned/Turn-Exit 5%
  - Appreciation 2%
  - Yes-No-Question 2%
  - Non-verbal 2%
  - Yes answers 1%
  - Conventional-closing 1%
  - Uninterpretable 1%
  - Wh-Question 1%
  - No answers 1%
  - Response Acknowledgment 1%
  - Hedge 1%
  - Declarative Question 1%
  - Other 1%
  - Backchannel-Question 1%

- Smaller search space for automatic DA annotation

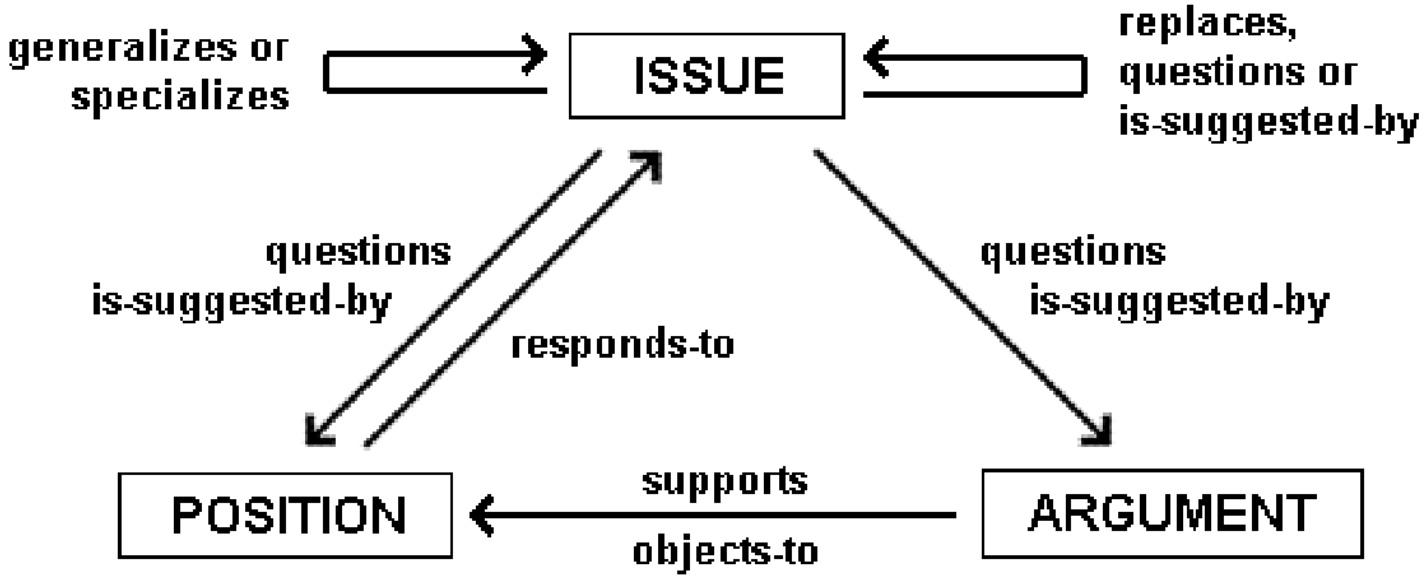
Automatic DA tagging

- Classic technique (Stolcke et al. 2000)
  - two sources of information
    - lexical items (e.g. word n-grams)
    - dialogue model (sequence of DAs)
  - machine learning
    - HMM for the dialogue model
    - classifier for each utterance using lexical items
    - decoding to find the most likely sequence of DAs

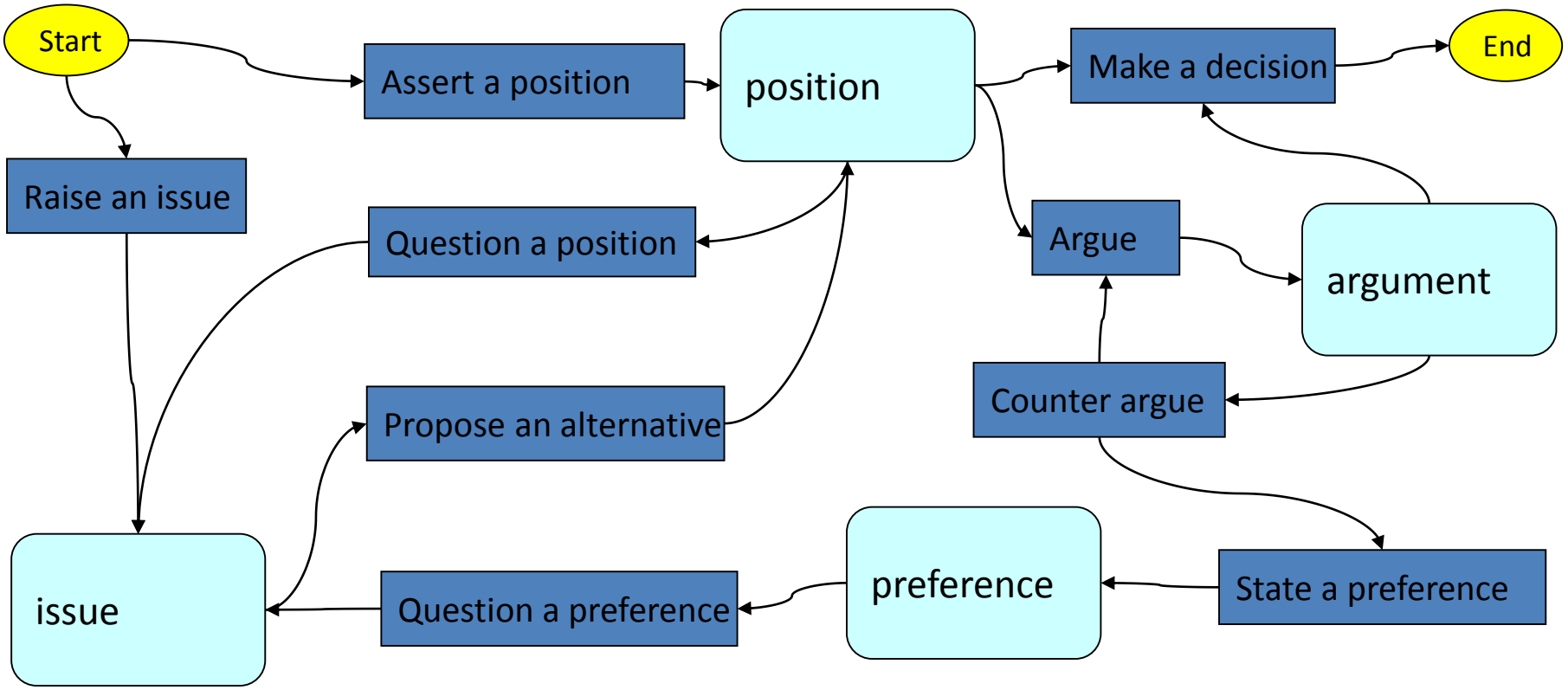
ARGUMENTATIVE STRUCTURE

IBIS model (Kunz and Rittel 1970)

- Based on Toulmin's model of argumentation
- Discussion graphs represent a multi-agent decision making processes
- Ontology: agents, objects, relations, states
- Argument types: issues, positions, arguments, criteria



A discussion model in the Zeno system (Gordon and Karakapilidis 1997)



An argumentation tree in Zeno

Husband: Honey, we've been thinking about buying a new car. Do you have something particular in mind? (I1)

Wife: Well, yes. I think we should buy a Volvo station wagon. (P2)

Husband: But, that's such a family car! (P5) Let's buy a nice fast sport car. A Porsche would be great. (P1)

Wife: Isn't a Porsche pretty expensive? (P4) And besides, I think we should buy a safer car. Volvos are built like tanks. (P6)

Husband: What makes you think Volvos are so safe? (I2)

Wife: Don't you watch TV? Haven't you seen the advertisements? (P7)

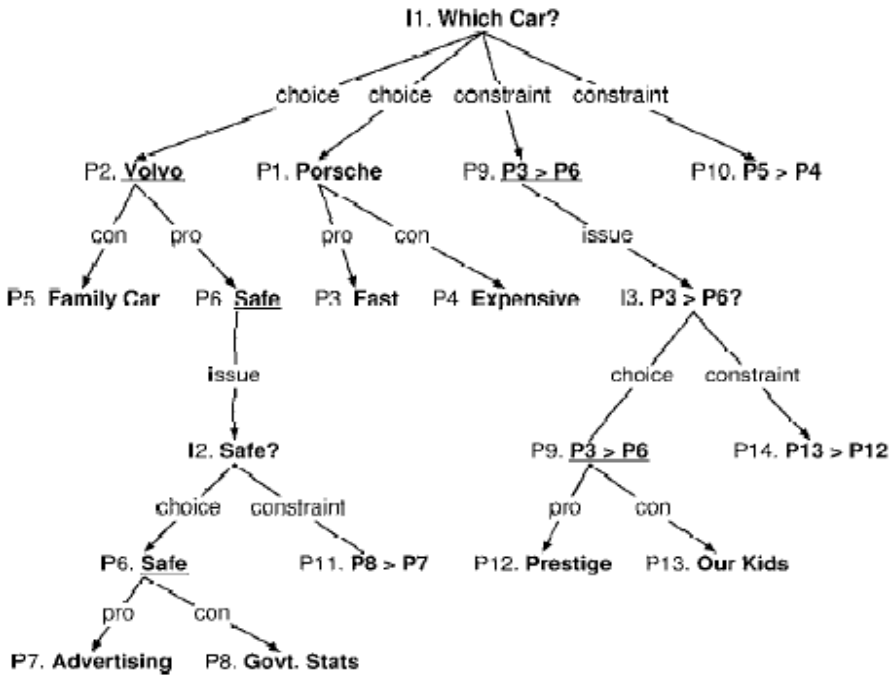
Husband: Oh, come on Honey. I read a report in "Auto Sports Today" the other day which cited some government accident statistics. Do you know what? Volvos were said to be involved in more fatal accidents than almost any other brand. (P8) And besides, having a fast car is more important to me than having a safe car. (P9)

Wife: Why?

Husband: Look, I've been wanting a fast sport car ever since I finished law school. An attorney in this town has to have a dynamic image (P12)

Wife: Yes, dear. But what about Betty and Susan. We have to think of the safety of our kids first! (P13, P14)

Husband: I guess you're right about that. But where does that leave us? I still think I rather pay a few thousand dollars more for a Porsche than drive such a boring family car. (P10)



Use for HLT

- Relationship between
  - automatic analysis of human(-human) dialogue
  - design of human-computer dialogue systems

- Understanding HH dialogue is closely related to the input side of a spoken dialogue system
  - and helps indirectly the output side