Dialogue Act Annotation

Ivana Kruijff-Korbayová
korbay@coli.uni-sb.de
http://www.coli.uni-sb.de/~korbay/Lehre/DM-beschreibung.html

Sommersemester 2002

Outline

- Introduction
  - What is the goal of DA annotation?
  - Why is DA annotation useful?
- Examples of dialogue corpora and DA annotation schemes
  - The TRAINS Corpus
  - The Map Task Corpus
  - The Verbmobil Corpus
  - The Switchboard Corpus
- Emerging DA annotation standard: DAMSL
- Measuring DA annotation reliability
- Use of DAs in applications

Introduction

The goal of DA annotation:

- Create data for corpus-based approaches

The usefulness of DA annotation:

- Empirical grounding and verification of dialogue theories
- Empirically based development of automated systems
- Training and testing of automatic DA recognition

Dialogue Corpora

- (Some) Characteristics
  - human-human vs. human-computer
  - contact (face-to-face) vs. distant (e.g., telephone)
  - general conversation vs. task-oriented
  - controlled vs. spontaneous
  - keyboard input vs. speech
  - etc.
- DA annotation schemes
  - corpus-specific
  - derived from a standard

TRAINS Corpus

Dialogue annotation in terms of conversation acts (Traum&Hinkelman 1992):

- Turn-taking CAs, e.g., take-turn, keep-turn, release-turn, assign-turn
- Grounding CAs, e.g., initiate, continue, acknowledge, repair, request-repair, request-acknowledgment, cancel
- Core speech acts, e.g., inform, wh-question, y/n-question, accept, request, reject, suggest, evaluate, request-permission, offer, promise
- Argumentation CAs, e.g., elaborate, summarize, clarify, Q&A, convince, find-plan

HCRC Map Task Corpus

Dialogue structure annotation at three levels:

- Transactions: subdialogues that accomplish one major step in task (e.g., dealing with one route segment) — normal, review, overview, irrelevant
- Conversational games (dialogue games, interactions, exchanges):
  - initiations: set up discourse expectations
  - responses: fulfill these expectations
  - a game consists of a set of utterances consisting of an initiation and all utterances up until the purpose of the game is fulfilled or abandoned
- Conversational moves: different kinds of initiations and responses classified according to their purposes

TRAINS Corpus

Characteristics

- simulated human-computer
- spontaneous speech
- task-oriented: collaborative transportation-problem solving
  - manager specifies a task
  - system helps to find a plan for accomplishing the task
  - manager and system share a map
  - system has additional information about task execution

HCRC Map Task Corpus

Characteristics:

- human-human (familiar and unfamiliar)
- with and without eye-contact
- spontaneous speech
- task-oriented:
  - route-giver has a map with landmarks and a route
  - route-follower has a map with (different) landmarks
  - route-giver instructs route-follower to replicate the route

The Verbmobil-2 Corpus

Dialogue annotation at two levels:

- Dialogue Phases (for negotiation):
  - Hello: greeting and introduction
  - Opening: introduction of topic to negotiate
  - Negotiation
  - Closing: negotiation finished, (sometimes) agreed-upon topic summarized
  - Good_bye: parting greeting
- Dialogue Acts:
  - organized into a hierarchy
  - each DA characterized by name, dialogue phase in which the DA may occur, related propositional content if any, definition, example(s)

The Switchboard Corpus

DA annotation using a modified version of the DAMSL:

- Initial multidimensional tag set: 50 basic tags plus diacritics for orthogonal information; approx. 220 unique combinations used by coders
- Reduced tag set for statistical dialogue modeling: 42 utterance types (DAs)
- Further reduced tag set with 6 categories of DAs was used in a study of DAs and prosodic features

The Verbmobil-2 Corpus

Characteristics

- human-human
- task-oriented:
  - appointment scheduling negotiation
  - travel-information seeking and giving
- spontaneous speech

The Switchboard Corpus

Characteristics:

- human-human
- distant: telephone
- general conversation
- spontaneous speech

DAMSL: Forward-looking functions

- Statement:
  - Assert vs. Reassert vs. Other
- Addressee's future action:
  - "Open option"
  - Directive (info-request, action-directive)
- Speaker's commitment to future intention to act:
  - "Offer" vs. Commit
- Other specific functions:
  - Explicit performative
  - Exclamations
  - Conventional Openings
  - Conventional Closings
- None

DA Annotation Reliability

Two methods used predominantly:

- Confusion matrices to show how frequently are DAs confused with one another
- κ-statistic to measure replicability as pairwise agreement among coders in a classification task (Krippendorf 1980, Carletta et al. 1996)
  - stability (inter-test variance): same coder's judgments over time
  - reproducibility (inter-coder variance): different coders' judgments
  - accuracy: against some "golden" standard

Dialogue Act Markup in Several Layers (DAMSL)

- DAMSL emerges as a DA annotation standard, endorsed by the Discourse Resource Initiative (DRI)
- described in (Core and Allen 1997) and (Carletta et al. 1997)
- domain (task) independent, abstract DA categories
- we'll look at two dimensions:
  - Forward-looking communicative functions (initiations)
  - Backward-looking communicative functions (responses)

DAMSL: Backward-looking functions

They are characterized along 4 dimensions:

- Understanding: speaker's ability to recover the semantic content
  - signal understanding, signal non-understanding, completion, re-realize
- Agreement: speaker's attitude toward an action, plan, object, etc.
  - closing acts: accept, accept part, non-accept, reject part, reject;
  - non-closing acts: hold
- Informational relations: relationships between the content of the current utterance and the utterance it responds to (e.g., consequence)
- Answer

DA Annotation Replicability: Kappa Statistic

(Carletta et al. 1996)

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

where

$P(A)$ is the proportion of times the coders agree
$P(E)$ is the proportion of times the coders are expected to agree by chance

Loose interpretation of the agreement scale:

- $\kappa = 0$: no agreement
- $\kappa = 1$: total agreement
- $\kappa > 0.8$: good replicability
- $0.67 < \kappa < 0.8$: tentative conclusions can be drawn

DA Annotation Replicability: Kappa Statistic

Reproducibility results for Map Task corpus coding (Carletta et al. 1996):

- Move coding:
  - Move segmentation: $\kappa = 0.92$ using word boundaries as units ($N = 4079$ units, $k = 4$ coders)
  - Move classification: $\kappa = 0.83$ on move segments where the boundaries were agreed ($N = 563$, $k = 4$); largest confusions between check vs. query-yn, instruct vs. clarify, acknowledge vs. ready vs. reply-yn
- Game coding:
  - Game beginning: $\kappa = 0.7$ ($N = 203$, $k = 4$)
  - Game type: $\kappa = 0.86$ ($N = 154$, $k = 4$)
- Transaction coding: $\kappa = 0.59$ ($N = 657$, $k = 5$, $k = 4$) for whether conversational move boundaries are transaction boundaries

Use of DAs in Applications

Generally, a conversational agent needs to know whether it was asked a question or ordered to do something, whether a previous utterance is being corrected or an alternative is being proposed, etc.

(Can you think of examples?)

Use of DAs in Verbmobil

- Transfer: DAs help to identify the best translation.
  - *vielleicht* usually translates as *possibly*, but when used as a politeness marker in requests or suggestions it is left out
  - *wiederholen* translates as *repeat* in opening or negotiation dialogue phase, and as *recapitulate* in closing phase
  - *Termin aus/abmachen/vereinbaren* translates as *make an appointment* in opening phase, and as *agree on a time* in negotiation or closing phase
- Shallow processing: Selection of templates used to generate target language expressions.
- Summary generation: Determination of the dialogue segments containing the core of the dialogue.

Robust Methods of Automatic DA Recognition

(cf. Stolcke et al. (2000) for an overview)

- use of n-grams to model the probabilities of DA sequences, or to predict upcoming DAs online
- use of n-grams as miniature models for DAs to improve speech recognition
- use of other trainable language models to estimate DA likelihoods (stochastic CFGs, FSAs)
- use of decision trees on word patterns and neural nets to model DAs
- dialogue HMMs combining word and dialogue n-grams
- prosodic DA modeling: extraction of various prosodic parameters (cf. overview in Shriberg et al. 1998)
- nonprobabilistic DA classification using transformation-based learning

Use of DAs in Speech Recognition

- Task oriented dialogue (Map Task corpus): Taylor et al. (1998) combine DA likelihoods from prosodic models with those from 1-best recognition output to condition the recognizer language model. Word error rate reduction 1%.

- Conversational speech (Switchboard corpus): Stolcke et al. (2000): probabilistic integration of speech recognition with dialogue modeling by statistical dialogue grammars modeling sequencing of DAs, and DA likelihood models expressing local lexical and prosodic cues for DAs. Word error rate reduction 0.3%.

Summary

- Goals of DA annotation
- Examples of dialogue corpora and DA annotation schemes
  - The TRAINS Corpus
  - The Map Task Corpus
  - The Verbmobil Corpus
  - The Switchboard Corpus
  - Others: see *Linguistic Data Consortium* or *Discourse Resource Initiative*
- DAMSL annotation standard
- Measuring DA annotation reliability
- Use of DAs in applications