Grounding

LING 575: Spoken Dialog Systems
May 12th, 2016

What is Grounding?

Spoken Dialog is special way of communication

It is the result of a joint collaboration

Achieving a common ground of mutually believed facts of what is being talked about, that serves as a basis for furthers acts of communication

System: Did you want to review your profile?
User: No

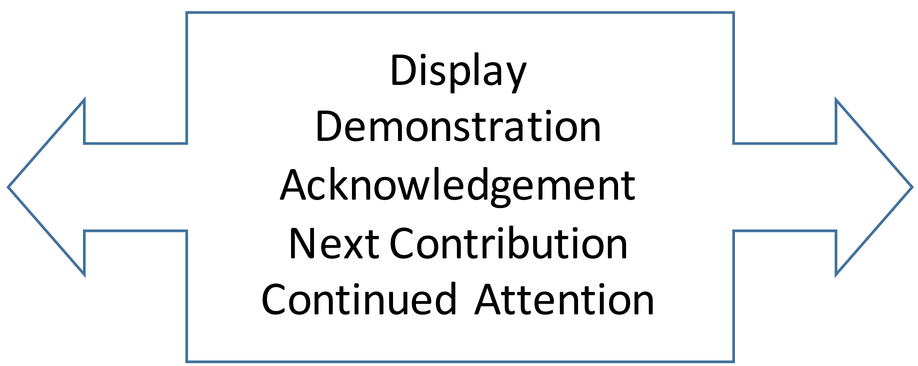
System: Okay, what is next? OR System: What is next?

Grounding 101

Contributional Model (Clark & Schaefer)

Dialog is a collaborative process

Presentation

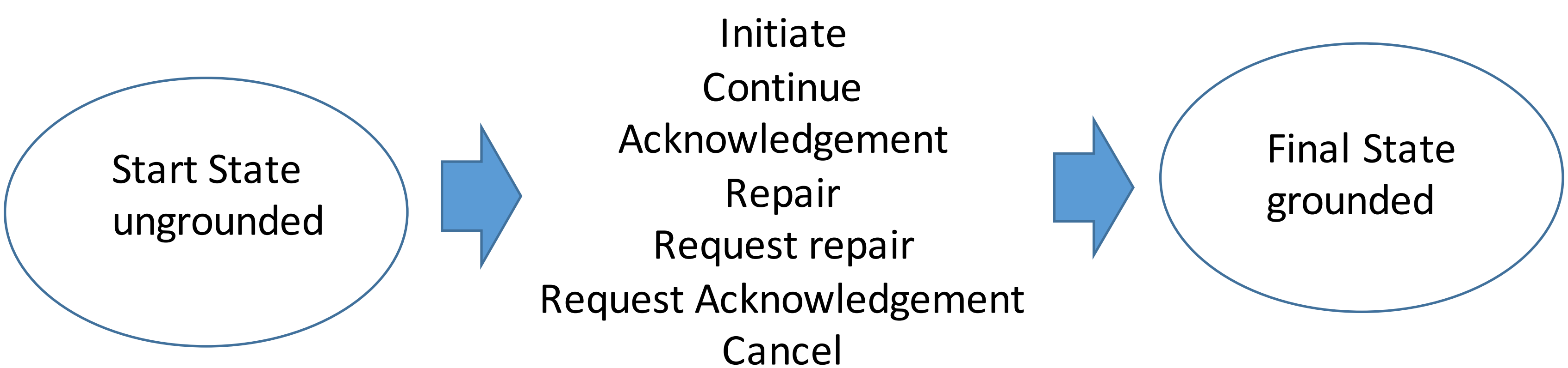


Acceptance

Grounding 102

Grounding Act Model (Traum)

Utterances are identified with a grounding act (discourse units) that work towards achievement of common ground



Grounding 201

Decision Models under Uncertainty

What kind of evidence to choose?

When to ground?

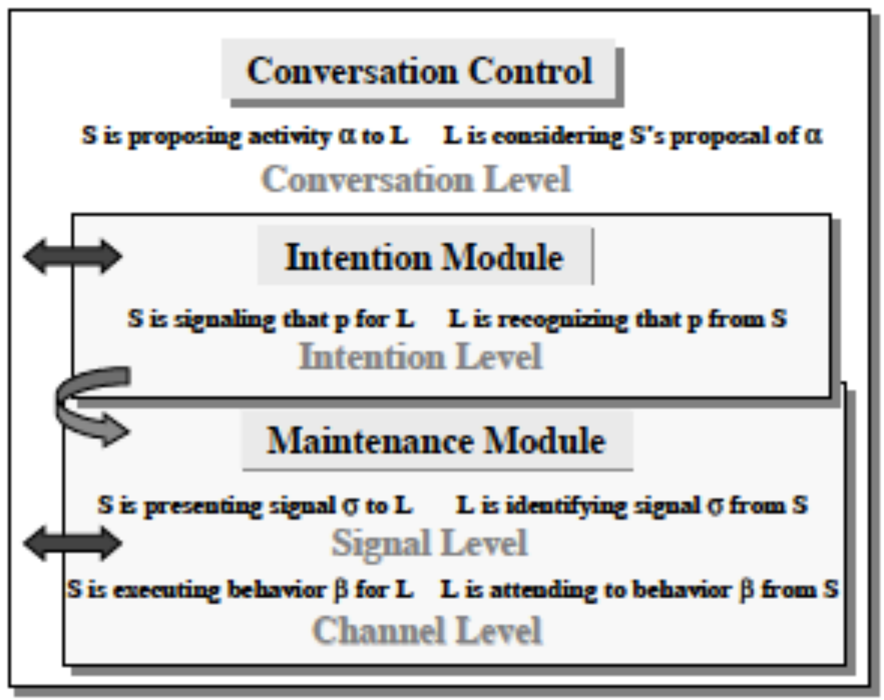
Utility Problem by minimizing costs when performing a action :
Accept, Display, clarify, reject

$$GA = \arg min_a \; (P(a, correct) * \quad Costs\,(a, correct) \\ + P(a, incorrect) * Costs\,(\mathrm{a, incorrect})\;)$$

Grounding 202

Quartet Model : Conversation Model under Uncertainty

→ Exploit Uncertainties in order to disambiguate



Grounding 203

Degrees of Grounding (Traum)

Given a new utterance → Keeping track of state

Common Ground Unit (CGU)

Types of evidence:	**Degrees of groundness**
submit, repeat back,	unknown, misunderstood
resubmit, acknowledge,	unacknowledged,accessible
request repair, move-on, use,	agreed-signal, agreed-content
lack of response	assumed

Thanks!

¡Gracias !

Danke schön !

ありがとうございます！

Questions

1. What annotation scheme or other empirical data was used to reach some of these conclusions? And do they suffer from low kappa values?

2. The idea of ambiguity influencing the ability to determine the nature of acceptance of a particular utterance in response to an initial utterance seems well-suited for a probabilisitic model. There was some hinting at that but no detailed description. Has that been done and how successful has it been for grounding?

3. As an extension of question #2, what features have been used? I'd expect that phrase-level or discourse-level units could be predictive.

(nperk)

Questions

In the primary paper, for Clark and Schaefer's model, the author mentioned that the graded evidence of understanding has several problems, like for example how to differentiate between "little or no evidence needed" from "evidence not needed" ?. I received that point well.

However, down in the paper, in the Grounding Acts model ,he mentioned that one of it's deficiencies is that the binary "grounded or ungrounded" distinction in the grounding acts model is clearly an oversimplification.

It seems to me that both extremes have problems, does this mean that we need to seek a middle approach ?

(eslam)

Questions

In the primary paper for grounding, Traum discusses two theories of grounding. The goal of both of these theories is to be able to understand when a given piece of information enters the shared context between the interlocutors. However, he spends little time discussing what this shared context actually looks like. What are your thoughts on, for example, the need to ground information that is already in shared context, or what information is already shared at the beginning of a dialogue?
(erroday)

Questions

Based on primary paper
How many utterances were used? The authors mentioned 16 participants. Would you know how engaged these participants were(i.e average length of the whole conversation in terms of utterances) ?
(lopez380)

One of the discussion questions by Traum asks whether models of this type should explicitly be used in HCI systems, rather than just incorporating grounding feedback. Since this was in 1999, now 17 years later, are we doing that?
(mcsummer)

Miscommunications, Repairs, and Disfluencies

Laurie Dermer – George Cooper

5/12/2016

S: *What time do you want to leave?*
U: Eight o'clock a.m.
S: *Do you want to leave around 10:00 p.m.?*
U: Eight o'clock
S: *Do you want to leave around 10:00 p.m.?*
U: No
S: *What time do you want to leave?*
U: Eight o'clock a.m.

Source papers and topics

Topic group #1: Detecting corrections

- Three papers, including the primary paper, were primarily on detecting corrections:
  - Litman et al. 2006: "Characterizing and Predicting Corrections in Spoken Dialogue Systems"
  - Levow 2004 "Identifying Local Corrections in Human-Computer Dialogue"
  - Levitan & Elson 2014 "Detecting Retries of Voice Search Queries"

Topic group #2: Detecting disfluencies

- Two papers were on detecting disfluencies:
  - Zayatset al. 2014: "Multi-Domain Disfluency and Repair Detection"
  - Schriberg 2001: "To 'errrr' is human: ecology and acoustics of speech disfluencies."

Topic group #3: Handling Corrections

- Four papers discussed methods for handling corrections:
  - Liu et al., 2014: "Detecting Inappropriate Clarification Requests in Spoken Dialogue Systems"
  - Stoyanchev et al, 2013: "Modelling Human Clarification Strategies"
  - Jiang et al., 2013: "How do users respond to voice input errors?: lexical and phonetic query reformulation in voice search."
  - Bohus & Rudnicky, 2005: "A principled approach for rejection threshold optimization in spoken dialog systems."

Some general background

Miscommunications and Repairs

- **Disfluencies** happen all the time in speech.
  - "One study observed disfluencies once in every 20 words, affecting up to 1/3 of utterances." (Zayats et al. 2014)
- We use **repair techniques** to “correct” disfluencies for listeners.
- **Miscommunication** is also an everyday part of speech, and in natural language use we have techniques (prosody, hyper-articulation, repetition) for correcting miscommunications when they occur.

Types of miscommunications

- **Speech disfluencies** include most kinds of disrupted speech
  - Disfluencies include filled pauses ("uh"), repetitions ("I want – I want to go to..."), (self-)repairs, and false starts.
- **Miscommunications** are generally when a system misinterprets a user's utterance.
  - A user might respond by rejecting ("no!", "go back") or correcting ("I meant the *sixth* of December", "No, *Toronto*") the system's utterance.

Implications for NLP

- Humans account for repairs fairly naturally. Computers do not.
- Filled pauses are trivial to detect.
- Disfluencies with a repair are harder to detect, but detecting them (and fixing the transcription or accounting for them) aids NLP tasks.
- Detecting corrections during a system's use can boost system quality, and detecting them after the fact can help with error analysis.

Detecting corrections

How do we do it?

Also, when do they happen?

How do they happen?

What types of corrections do people make?

- **Omissions** (of part of the utterance), **paraphrases**, and simple **repetition** of the utterance are common tactics.
- Omissions were more common after a misrecognized utterance
- Repetitions were more likely after a rejected turn.
    - *Speaking of which…*

Table 2
Distribution of correction types.

Type	ADD		ADD/OMIT		OMIT		PAR		REP		N
Total	51	8%	14	2%	215	32%	127	19%	265	39%	672
Post-Misrec	39	7%	13	3%	203	40%	90	18%	173	32%	518
Post-Rej	8	6%	0	0%	9	7%	36	28%	75	59%	128
Non-Immed	4	15%	1	4%	3	12%	1	4%	17	65%	26

System Design Matters

- Part of why repetitions were more likely after a rejected turn in that paper (Litman et. Al.) was that the system prompted the user to "repeat the utterance."
- Levow (2004) pointed out lack of feedback by systems leading users to be less local in corrections.
- **It's important to craft prompts that favor the type of correction most easily recognized by the system, and/or most useful to the system.**

Systems

- The authors of the papers typically built classifiers (boosters, logistic regression) and used features that varied depending on their exact task.
- Some features:
  - Prosody, pitch, intensity
  - Silence within an utterance (hyperarticulation)
  - Confidence score
  - LM score
  - Interaction (or lack thereof) by the user
  - Preceding pause
- All systems had very good error reduction on the task they were handling (~50%)

Some major findings from the papers

- Litman et al. (2006) noted that hyperarticulation can lead to misinterpretation of an utterance by the system, and other prosodic differences can also lead to problems.
- Generally, speech recognizers were more likely to misinterpret something that was hyperarticulated.
- Even when a person can't distinguish hyperarticulation, an unrecognized utterance often has features of hyperarticulation.

Some major findings from the papers

- Levow (2004) – used prosodic cues to detect the location of a local correction. Remember these phrases from an earlier slide?

  ("I meant the *sixth* of December", "No, *Toronto*")
- This paper was about detecting local corrections – in other words, corrections of just one part of an utterance.
- People often do not use specific syntactic structures or cue phrases for local corrections, but use prosodic cues instead.

Some major findings from the papers

- Levitan & Elson (2014) used logistic regression on Google voice data to detect retries.
- Their three features included
  - Similarity between the queries (based on edit distance),
  - Correctness (based on confidence, user behavior, retry interval, and hyperarticulation features), and
  - Recognizability (low LM score, # of alternate pronunciations, length of query).

Modelling Human Clarification Strategies

Svetlana Stoyanchev, Alex Liu, Julia Hirschberg

Clarification questions

- Non-targeted: e.g. "what did you say"
- Targeted:
  - Example:
    - Speaker A: "Can I have some toast please"
    - Speaker B: "Some what?"
  - Repeat the understood part of the question
  - Also serve as a form of grounding

Current approaches to clarification questions

- Most Spoken Dialogue systems set an arbitrary threshold
- Stoyanchev et al. built classifiers for whether to ask a clarification question, and if so whether it should be targeted or non-targeted

Data

- Utterances were drawn from interactions with IraqComm, a speech-to-speech translation system
- Misrecognized words were replaced with XXX
- Annotators on Mechanical Turk marked whether to ask a clarification question or not, and if so which kind

Inter-annotator agreement

Clarify-or-not classifier	Targeted/non-targeted classifier
39%	25%

Classifier description

- Two binary classifiers
- Built using WEKA machine learning framework
- Feature classes:
    - Missing word position
    - POS
    - Dependency parse information
    - Semantic roles

Results

	Clarify-or-not classifier	Targeted/non-targeted classifier
Accuracy	72.8%	74.6%
Baseline	59.1%	71.8%

Disfluencies

The parts of a disfluency

- Reparandum: The words that are corrected or repeated
- Editing phase:
  - Filler words
  - Serves to stall for time or signal disfluency
- Repair: The correction for, or repetition of, the reparandum

"We had the cat, uh, the dog first"

reparandum

repair

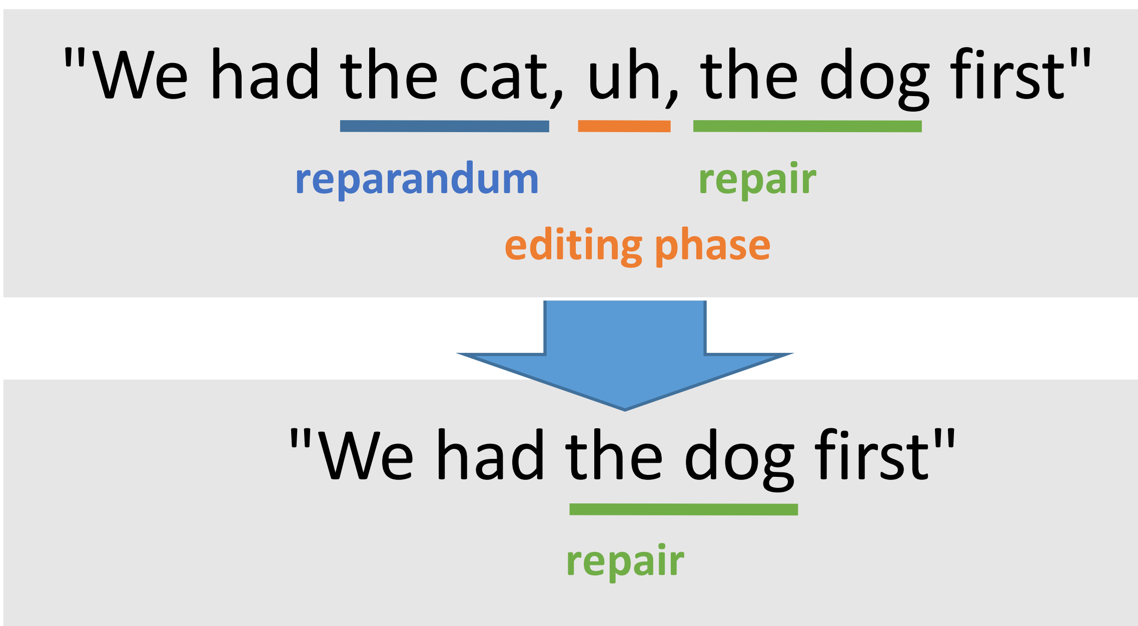
editing phase

Problems for spoken dialogue systems

- ASR: Truncated words
  - Partial words unlikely to be in vocabulary
  - Including partial words in the vocabulary would cause them to be used too often
- NLU
  - They would be difficult to incorporate into hand-built grammars
  - They present statistical noise for machine-learning based systems

Removing disfluencies

- Disfluencies can be corrected by removing the reparandum and editing phase



Automatic disfluency detection

- Often treated as a sequence labeling problem, similar to NER
- Uses labeling schemes similar to BIO
- Corpora include switchboard
- Features include word and POS n-grams, syllable length

Questions/Discussion

- How has new system design (aka neural networks) affected robustness vs. things like hyperarticulation, false starts?
- What are the most common/most useful strategies used by spoken dialogue systems to repair errors once they've been detected?
- Hyperarticulations are less likely to be recognized, and hyperarticulated corrections are less likely to be recognized – does this lead to a cycle of corrections? - yes!

Questions/Discussion

- Do (human-human and human-computer) error correction strategies vary by (age, gender, region, native language)?

  Yes! Individuals vary in their repair techniques.

  "Some people are "repeaters" and others are "deleters" -- in other words, they tend to favor one strategy over the other." (Zayats et al. 2014) (see next slide for more)

- If so, are those variations significant enough to effect the results of this system, and suggest using targeted subsystems?

- "Note, however, that there were overall differences in the corrections produced by native and non-native speakers, normalized by value of first turn in task: mean f0 was higher for native speakers than for non-native speakers (t stat = −2.72, df = 602, p = .0067), tempo was faster (t stat = −3.18, df = 670, p = .0015), and duration was shorter (t stat = 2.20, df = 670, p = .028). These differences do not occur in non-correction utterances.
- Gender of the speaker was also annotated in the corpus for the primary paper – they didn't say much about it though.

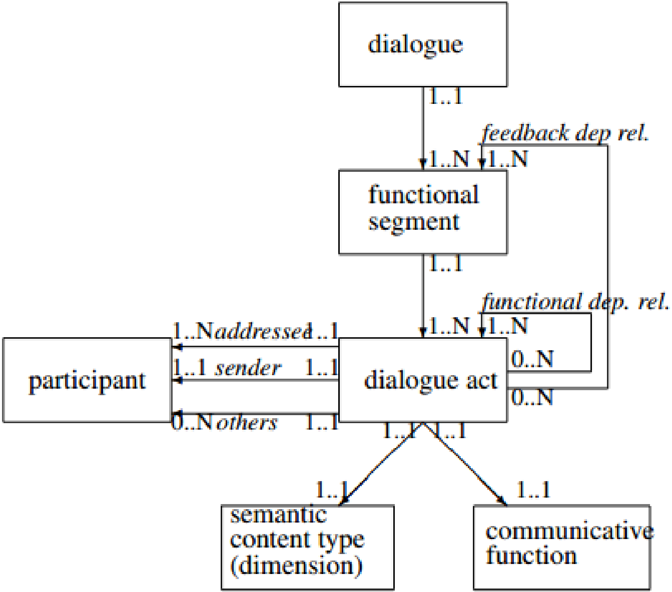
Dialog Act Taxonomies

May 12, 2016

Basic concepts and metamodel

1. Functional segmentation
   - Communicative functions can be assigned more accurately to smaller units, which we call functional segments
   - at least 2 participants
     1.1 an agent whose communicative behaviour is interpreted, the “speaker”, or “sender”
     1.2 a participant to whom he is speaking and whose information state he wants to influence, called the “addressee”
2. Dependency relations

Metamodel



dialogue
1..1
feedback dep rel.
1..N
1..N
functional segment
1..1
functional dep. rel.
1..N
1..N
1..N addressee 1..1
1..1 sender 1..1
0..N others 1..1
participant
dialogue act
0..N
0..N
1..1
1..1
1..1
1..1
semantic content type (dimension)
communicative function

Communicative functions

1. Approaches to communicative function definition
    - communicative functions use one or both of the following definitions:
        1.1 in terms of the intended effects on addressees
        1.2 in terms of properties of the signals that are used
2. Communicative function recognition
    - depends on addressees understanding the communicative functions of the speaker's utterances
    - use of 1 hierarchies of communicative functions, and 2 function qualifiers, which make a base communicative function more specific

Dimensions

dialogue utterances can have multiple communicative functions multidimensional schema addresses this ‘dimension’ refers to various types of semantic content – the types of communicative activity concerned with these types of information

Core Concepts: Dimensions

First four of these criteria apply to the identification of dimensions more generally; the fourth criterion applies to the choice of a coherent set of dimensions, and the final fifth criterion applies specifically to 'core' dimensions.

1. Each dimension has a clear empirical basis,
2. Each dimension is theoretically justified,
3. Each dimension is recognizable with acceptable precision by human analysts, in particular by annotators, as well as by dialogue understanding and dialogue annotation systems.
4. Each dimension in a multidimensional system can be addressed by dialogue acts independent from addressing other dimensions (the dimensions are independent or orthogonal).
5. Each core dimension is present in many existing dialogue act annotation schemes.

Nine dimensions that qualify as core dimensions.

- Task (or Activity):
- Auto-and Allo-Feedback, eliciting information about the processing of previous utterances by speaker (auto) or addressee (allo);
- Turn Management
- Time Management
- Discourse Structuring: dealing with topic management and structuring the dialogue
- Own and Partner Communication Management: actions by the speaker for editing his current contribution, or for editing contribution of another
- Social obligations Management: introducing oneself, apologizing, and thanking, and responses to these acts, such as accepting an apology

Communicative Functions

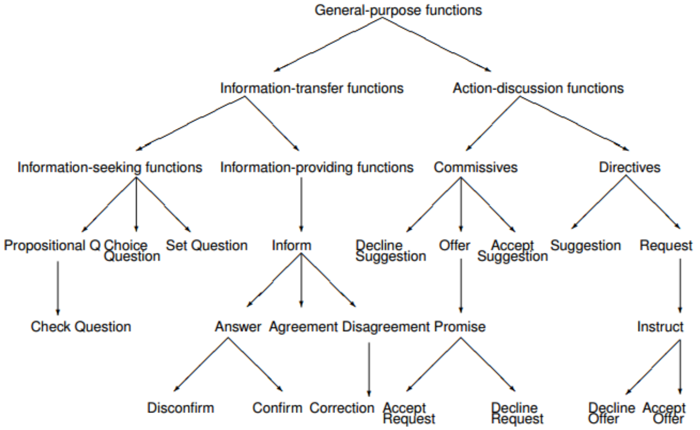
populate a multidimensional schema can be based on similar criteria as the choice of core dimensions The following six criteria have been identified:

1. Empirical validity: for every communicative function there exist linguistic or nonverbal means which can be used by speakers to indicate that their behaviour has that function.
2. Theoretical validity: every communicative function has a precise definition which distinguishes it semantically from other functions.
3. The set of communicative functions applicable in a certain dimension provides a good coverage of the phenomena in that dimension.
4. Each communicative function can be recognized with acceptable precision by humans and by machines.
5. Each core communicative function occurs in many existing annotation schemas.
6. Any two communicative functions that can be used in a given dimension are either mutually exclusive, i.e. if one of them applies to a given functional segment then the other one does not, or one function is a specialization of the other.

Dimension-specific and general-purpose functions

- general-purpose functions:
  - 4 information-seeking functions,
  - 7 information-providing functions,
  - 6 commissive functions,
  - 5 directive functions;
- dimension-specific functions:
  - 2 auto-feedback functions;
  - 3 allo-feedback functions;
  - 2 time management functions;
  - 6 turn management functions;
  - 3 discourse structuring functions;
  - 2 own communication management functions;
  - 2 partner communication management functions;
  - 10 social obligation management functions.

Taxonomy of general-purpose functions



Function Qualifiers

Qualifier attributes, values, and function categories

qualifier attribute	qualifier values	CF category
modality	uncertain, certain	info-providing functions
mode	angry, happy, surprised, ...	info-providing functions; feedback functions
conditionality	conditional, unconditional	action-discussion functions
partiality	partial, complete	responsive functions; feedback functions

DiAML: Dialogue Act Markup Language

(11)

P1:	*Do you know what time the next train to Utrecht leaves?*
P2:	*The next train to Utrecht leaves I think at 8:32.*
AuFB	The next train to Utrecht [*positiveAutoFeedback*]
TA	The next train to Utrecht leaves at 8:32. [*answer, uncertain*]

(12)

```
<diaml
xmlns:"http://www.iso.org/diaml/">
<dialogueAct xml:id="da1" sender="#p1"
addressee="#p2" target="#fs1"
communicativeFunction="request"
dimension="task"
conditionality="conditional"/>
<dialogueAct xml:id="da2" sender="#p2"
addressee="#p1" target="#fs2"
communicativeFunction="overallPositive"
dimension="autoFeedback"
feedbackDependenceTo="fs1"/>
<dialogueAct xml:id="da3" sender="#p2"
addressee="#p1" target="#fs2"
communicativeFunction="answer"
dimension="task"
functionalDependenceTo="da1"/>
</diaml>
```

Dialogue Structure Coding Scheme

Dialogue structure coding scheme based on utterance function, game structure, and higher-level transaction structure

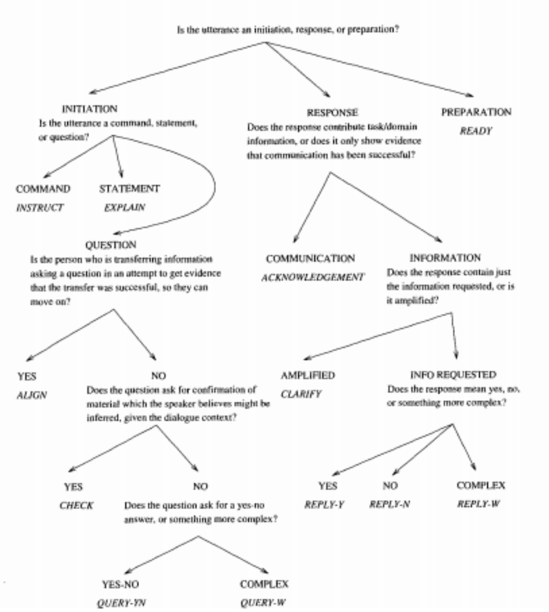
Structure

Dialogues are divided into transactions Transactions are conversational games Game analysis differentiates between:

- initiations, which set up a discourse expectation about what will follow
- responses, which fulfill those expectations.

Games are themselves made up of conversational moves

Conversational move categories



Initial Moves

- INSTRUCT commands the partner to carry out an action. Where actions are observable, the expected response could be performance of the action.
- EXPLAIN states information that has not been directly elicited by the partner
- ALIGN move checks the partner's attention, agreement, or readiness for the next move
- QUERY-YN asks the partner any question that takes a yes or no answer and does not count as a CHECK or an ALIGN
- QUERY-W is any query not covered by the other categories. Made of are wh-questions and otherwise unclassifiable queries

Response moves

- ACKNOWLEDGE move is a verbal response that minimally shows that the speaker has heard the move to which it responds, and often also demonstrates that the move was understood and accepted.
- REPLY-Y is any reply to any query with a yes-no surface form that means "yes", however that is expressed.
- REPLY-N Move. Similar to REPLY-Y, a reply to a query with a yes-no surface form, that means "no" is a REPLY-N.
- REPLY-W is any reply to any type of query that doesn't simply mean "yes" or "no."
- CLARIFY move is a reply to some kind of question in which the speaker tells the partner something over and above what was strictly asked.

The READY Move

- Moves that occur after the close of a dialogue game and prepare the conversation for a new game to be initiated.
- Speakers often use utterances such as "OK" and "right" to serve this purpose.
- whether READY moves should form a distinct move class or discourse markers attached to the subsequent moves, but the
- It is sometimes appropriate to consider READY moves as distinct, complete moves in order to emphasize the comparison with ACKNOWLEDGE moves

Transaction Coding Scheme

Gives the subdialogue structure of complete task-oriented dialogues each transaction being built up of several dialogue games The coding system has two components:

1. how route givers divide conveying the route into subtasks and what parts of the dialogue serve each of the subtasks, 2. what actions the route follower takes and when.

The basic route giver coding identifies the start and end of each segment and the subdialogue that conveys that route segment Transaction types:

- NORMAL
- REVIEW
- OVERVIEW
- IRRELEVANT

The ICSI Meeting Recorder Dialog Act (MRDA) Corpus

corpus of over 180,000 handannotated dialog act tags and accompanying adjacency pair annotations for roughly 72 hours of speech from 75 naturally-occurring meetings Annotation involved three types of information:

- marking of DA segment boundaries
- marking of DAs themselves
- marking of correspondences between DAs (adjacency pairs).

Segmentation methods were developed based on separating out speech regions having different discourse functions and paying attention to pauses and intonational grouping

MRDA tags to SWBD-DAMSL tags

TAG TITLE	SWBD-DAMSL	MRDA
Indecipherable	%	%
Abandoned	%-	%--
Interruption		%-
Nonspeech	x	x
Self-Talk	t1	t1
3rd-Party Talk	t3	t3
Task-Management	t	t
Communication-Management	c	
Statement	*sd*	s
Subjective Statement	*sv*	s
Wh- Question	qw	qw
Y/N Question	qy	qy
Open-Ended Question	qo	qo
Or Question	qr	qr
Or Clause After Y/N Question	qrr	qrr
Rhetorical Question	qh	qh
Declarative- Question	d	d
Tag Question	g	g
Open-Option	oo	
Command	*ad*	*co*
Suggestion	*co*	*cs*
Commit (self-inclusive)	cc	cc

TAG TITLE	SWBD-DAMSL	MRDA
Conventional-Opening	fp	
Conventional-Closing	fc	
Topic Change		**tc**
Explicit-Performative	fx	
Exclamation	fe	fe
Other-Forward-Function	fo	
Thanks	ft	ft
Welcome	fw	fw
Apology	fa	fa
Floor-Holder		**fh**
Floor-Grabber		**fg**
Accept, Yes Answers	*ny, aa*	*aa*
Partial Accept	aap	aap
Partial Reject	arp	arp
Maybe	am	am
Reject, No Answers	*nn, ar*	*ar*
Hold	h	h
Collaborative-Completion	2	2
Backchannel	b	b
Acknowledgment	bk	bk
Mimic	m	m
Repeat		**r**

TAG TITLE	SWBD-DAMSL	MRDA
Reformulation	bf	bs
Appreciation	ba	ba
Sympathy	by	by
Downplayer	bd	bd
Misspeak Correction	bc	bc
Rhetorical-Question Backchannel	bh	bh
Signal Non understanding	br	br
Understanding Check		**bu**
Defending/Explanation		**df**
Misspeak Self-Correction		**bsc**
"Follow Me"		**f**
Expansion/Supporting addition	*e*	*e*
Narrative-affirmative answers	na	na
Narrative-negative answers	ng	ng
No knowledge answers	no	no
Dispreferred answers	nd	nd
Quoted Material	q	
Humorous Material		**j**
Continued from previous line	+	
Hedge	h	
Nonlabeled		**z**

Reliability

Table 1: Results for strict segmentation agreement metric

Reference Labeler	Comparison Labeler	Agree	Total	Agree %
1	2	3004	4915	61.1
1	3	2797	3820	73.2
2	1	3004	4908	61.2
2	3	5253	7906	66.4
3	1	2797	3808	73.5
3	2	5253	7889	66.6
Overall		22108	33246	66.5

Table 2: Kappa values for DAs using different class mappings. Map 1: Disruptions vs. backchannels vs. fillers vs. statements vs. questions vs. unlabelable; does not break at the "|". Map 2: Same as Map 1 but breaks at the "|". Map 3: Same as Map 2

Questions

The paper talks very little about the ISO standard itself, just giving a brief example on the last page, and they neglect to give an example that has multiple function dimensions, a major point in their paper. So how would you represent multiple function dimensions? Their example <dialogueAct> tags seem to have a communicativeFunction="" attribute, but I believe that XML does not allow multiple attributes with the same name in one tag.

Example transcript

```
3. P1:
---------------------------------------------
we are going to go due south | NONVOC_noise ... | # |
fs3.1 Task: inform
fs3.2 TimeM: stalling
      TurnM: turnKeep
---------------------------------------------
straight south | ... and NONVOC_noise ... | then we're going to g-.
fs3.3 OCM: Self-Correction
fs3.4 TimeM: Stalling
      TurnM: TurnKeep
fs3.5 Task: Instruct
---------------------------------------------
      g- ..
fs3.6 OCM: Retraction
fs3.7 turn turn straight back round and head north... past an old mill ... on the right ... hand side
fs3.8: Task: Instruct
```

Example xml


```
<dialogueAct xml:id="da7" target="#fs3.2" sender="#p1" addressee="#p2"
communicativeFunction="turnKeep" dimension="turnManagement"/>
<dialogueAct xml:id="da8" target="#fs3.2" sender="#p1" addressee="#p2"
communicativeFunction="stalling" dimension="timeManagement"/>
<dialogueAct xml:id="da9" target="#fs3.3" sender="#p1" addressee="#p2"
communicativeFunction="selfCorrection" dimension="ownCommManagement"/>
<dialogueAct xml:id="da10" target="#fs3.4" sender="#p1" addressee="#p2"
communicativeFunction="stalling" dimension="timeManagement"/>
<dialogueAct xml:id="da11" target="#fs3.4" sender="#p1" addressee="#p2"
communicativeFunction="turnKeep" dimension="turnManagement"/>
<dialogueAct xml:id="da12" target="#fs3.5" sender="#p1" addressee="#p2"
communicativeFunction="instruct" dimension="task"/>
<dialogueAct xml:id="da13" target="#fs3.6" sender="#p1" addressee="#p2"
communicativeFunction="retraction" dimension="ownCommManagement"/>
<dialogueAct xml:id="da14" target="#fs3.7" sender="#p1" addressee="#p2"
communicativeFunction="selfCorrection" dimension="ownCommunicationManagement"/>
<dialogueAct xml:id="da15" target="#fs3.8" sender="#p1" addressee="#p2"
communicativeFunction="instruct" dimension="task"/>
```

Questions

Clarify the use of dimensions for annotating data. The dimensions are meant to cluster communicative functions into mutually exclusive clusters, but then the authors go on to say that some communicative functions are dimension specific (turn accept/turn release are only in turn management) while other are general purpose (check question). What then makes using dimensions more powerful than some other alternative method?

Questions

There are some strange relations there:

- why is accept/decline request under offer/promise?
- Why is decline/accept offer under request/instruct?

Questions

There are hierarchies of communicative functions, so that human annotators can use more fine-tuned labels and machine annotators can use more surface-level labels for dialog acts. This distinction is made because humans possess more capable context-reading skills that allow them to make more fine-tuned distinctions that computers wouldn't catch when labeling communicative functions. Couldn't other cues such as prosody, lexical content, and other more quantifiable aspects than context be combined and used by machines to provide fairly accurate classifications, even when it came to the more complex communicative functions? The justification that computers are completely limited simply because they do not possess human-level context awareness seemed to completely omit the possibility of labeling based upon these other aspects of speech.

Questions

One criterion for communicative functions is that "Each communicative function can be recognized with acceptable precision by humans and by machines." Should it say "can theoretically be recognized"

Questions

What is the distinction between ‘side-participants’, ‘bystanders’, and ‘overhearers’?