Clarification Requirements

What clarification requests tell us about dialogue system design

NIST ATP: Bosch RTC, Volkswagen, CSLI, SRI

Clarification Requests

- Questions about an antecedent (sub-)utterance
- Can concern meaning or form
- A characteristic dialogue phenomenon
- (Purver et al., 2003; Rodriguez & Schlangen, 2004; Rieser, 2005)
  - Quite common (3-6% of turns)
  - Lots of different types

Ben: No, ever, everything we say she laughs at.	Laura: Can I have some toast please?
Frances: **Who Emma?**	Jan: **Some?**
Ben: Oh yeah.	Laura: Toast

Restrictions on representations

- (Ginzburg & Cooper, 2004):
  - Must represent clarifiable information for each possible antecedent
  - *fractal heterogeneity*
  - *utterance reference & accessibility*
  - *clarificational potential*
- Representation of clarifiable elements via abstraction
  - Simultaneously abstracted set of parameters
  - C-PARAMS in a HPSG grammar
- Explicit record of sub-constituents, form and content
- Association of sub-constituents with their abstracted parameters
- Phrase types associated with suitable contents
- Amalgamation vs. inheritance

Restrictions on representations

Ben: No, ever, everything we say she laughs at.	Laura: Can I have some toast please?
Frances: **Who Emma?**	Jan: **Some?**
Ben: Oh yeah.	Laura: Toast

- Frances' question needs to be able to ask about the semantic reference of "she"
- Need to know how that fits into the overall proposition
- Jan's question needs to be able to ask about the utterance "toast"
- Need to know that "toast" came after "some"

Restrictions on semantics

- (Purver & Ginzburg, 2004): *reprise content hypothesis*
  - CRs query the semantic content of their antecedents
  - So knowing what a CR can/cannot ask about tells us something about that antecedent's contents
- Stricter than standard compositionality
  - Phrases held to account, not just sentences
- Suggests that generalised quantifiers might not be ideally suited
  - NP CRs really seem to ask about *individuals*
  - V and N CRs seem to ask about *predicates*

Restrictions on semantics

Monica: You pikey! Typical! Andy: **Pikey?** Nick: Pikey! Andy: What's pikey? What does pikey mean? Monica: I dunno. Crusty.	Terry: Richard hit the ball on the car. Nick: **What car?** Terry: The car that was going past. Nick: **What ball?** Terry: James [last name]'s football.

- Andy's question asks about the semantic content of "pikey"
  - A property or predicate of individuals
- It doesn't ask about the individual reference of "you pikey"
- Nick's questions ask about the semantic content of "the car", "the ball"
  - Individuals
  - Not properties-of-properties

Restrictions on context models

- We need a record of:
  - Utterances and words
  - Content associated (parameter assignments)
  - Inter-utterance dependency (QUD)
  - What hasn't been grounded (PENDING)
- Even the first takes us away from a purely finite-state-based model
- Examining possible clarification sequences can tell us about possible protocols:
  - Nested clarification pairs: a stack-based model?
  - Crossing clarification pairs: a set-based model?
  - Clarifications-of-clarifications: processing CRs as normal utterances

CLARIE (Purver, 2006)

- Dialogue system specifically designed to model human CR capabilities
  - Building on GoDiS, IBiS (Larsson (et al.), 2000, 2002)
- Reflects all the requirements explicitly
  - Compatible semantic representation
  - Explicit representation of clarificational potential
  - Fractally heterogeneous
  - PENDING stack, empirically grounded protocols
- Can engage in clarification dialogue in either direction
  - Learn unknown words, check contradictory information, …
  - Answer user queries
- But it's not a "serious" dialogue system
  - Text-based
  - Small domain
  - Small lexicon
  - HPSG grammar with minimal coverage

CHAT (Weng et al., 2006)

- Interactive in-car device control
  - Music player
  - Phone/addressbook
  - Point-of-interest database query
  - Navigation
- Information-state-update approach
  - CSLI Dialogue Manager as used in WITAS, SCoT
  - (Lemon & Gruenstein, 2004; Pon-Barry et al., 2006)
- Tree-based context representation
  - Dialogue moves as nodes
  - Update effects determined via node properties, structural relations
- Tested on real users with pretty good success rates
- Quite advanced system CR behaviour (see Stanley's talk)

CHAT clarification dialogues

- Use confidence scores at various interpretation levels:
  - Hypothesize most likely (pragmatic) interpretation
  - (including DMT attachment point)
  - Ask confirmation question
  - Positive answers lead to full attachment
  - Negative answers remove attachment, report
- Questions targetted at problematic levels

  "I couldn't hear you"

  "Which song do you want, the one by X or the one by Y?"
- Incorporated into dialogue (allow further information in the answer)

  "Are you looking for a cheap Chinese restaurant?"

  "Yes, a casual one"

  "No, an expensive one"

CHAT = CH(e)AT ?

- However, it doesn't fulfill all the "requirements" ...
- Semantics:
  - Uses an intermediate LF representation
  - Represents clarifiable information, albeit implicitly
- Utterance representation:
  - Not fractally heterogeneous
- Context model:
  - Does have a "pending" equivalent
  - Not restricted to stack-based processing
- How do we get away with this?
  - Utterance-based clarification/confirmation (not phrase-level)
  - Explicit questions (rather than reprise behaviour)

Why should we care?

- Spoken dialogue system designers might say we don't have to …
  - Systems certainly must be able to confirm, but you can get away with something much less
  - Utterance-level confirmation requires much less complexity
- With system error-handling, we always have speech recognition in the loop
  - Perhaps we can't be confident enough to be very subtle anyway
- We might assume we know all about the domain
  - Semantic CRs irrelevant – just acoustic, pragmatic
- We don't necessarily want to ask CRs like a human anyway
  - If we can't get prosody/timing quite right, might be safer to be explicit (non-reprise)

What if users request clarification?

- With user CRs, we have no way of controlling them
  - Could ask about any word/phrase
  - Could ask about any information level
  - Could use any syntactic form (incl. reprises/NSUs)
- Standard (implicit) assumption is that this won't happen
  - Users are more intelligent than systems
  - Users know more than systems about the domain
- But in some domains this no longer holds:
  - Tutoring systems
    - Users not understanding terms or explanations
    - Mitigated by system strategies (e.g. Socratic question-asking)
  - Interactive navigation systems (in-car systems)
    - Noise
    - User attention divided, cognitive load
    - Unfamiliarity with the domain

User CRs really do happen

- Wizard-of-Oz experiments in the navigation domain
  - User driving a simulated car in a simulated city
  - 1100 user turns
- About 7% of user turns are CRs of one kind or another
- About ¾ of these are requests for repetition
  - Often requests for a particular word or phrase, rather than the whole utterance
  - Often verbatim repetition is the wrong thing to do
- The rest are queries about reference or content
- About ¾ of them are explicitly marked (“please repeat X”, “what was the Y”)
- Some are more difficult to spot/interpret (reprise questions)
- Haven’t seen any about underlying intentions, syntax etc, or corrections
  - Probably not surprising

What can happen (1)

TTS: I calculated a route to Brian's house. It is 4.13 miles long and should take around 9.0 minutes .
SR: what is my next turn
TTS: Go East on Miranda Ave
SR: what is the next turn
TTS: Make sharp right on Arastradero Rd
SR: what road
TTS: I'm sorry, I do not understand what you mean by: what road
TTS: Please try rephrasing that.
SR: what road do i turn on
TTS: I'm sorry, I do not understand what you mean by: what road do i turn on
TTS: Please try rephrasing that.
SR: what road do i turn on and where is alma street from there
TTS: Did you want to plan a route from there?
SR: what was the last instruction you gave me
TTS: Sorry , there is no instruction to repeat
SR: what was the last route instruction you gave me
TTS: Sorry , there is no instruction to repeat
SR: i want to go to brian's
TTS: Ok , I will plan a route to Brian . Would you like anything else?

Simplest case: repetition

- Most obvious example: simple repetition

  Wiz: there is no traffic at wall street
  Subj: repeat that
  Wiz: there is no traffic at wall street

- Slightly more subtle: repetition of a particular move type

  Wiz: go straight for three blocks turn left at madison avenue
  …
  Subj: repeat the last directions
  Wiz: turn left at madison avenue

- Repetition of a phrase or word

  Wiz: turn right at sage road
  Subj: was that left or right
  Wiz: turn right at sage road

Pseudo-repetition

- Even requests for repetition are not always straightforward
- Intervening material means we have to look things up in context
- Intervening time means that previous information may no longer be correct

Wiz: did you want directions to dave's house
Subj: that is affirmative
Wiz: go straight for one mile turn left at columbia road
Subj: please repeat
Wiz: go straight for one mile turn left at columbia road
Subj: please provide location to nearest steakhouse relative to my current position
Wiz: carver's steakhouse is two point one miles away
Subj: please repeat directions to dave's current location
Wiz: go straight for three blocks turn left at columbia road

- New information, but pseudo-repetition form

Reprise questions

- We also get reprised fragments, so far all with wh-substitution:

  Wiz: go straight for four blocks turn left at wall street
  Subj: turn left where
  Wiz: turn left at wall street

  TTS: Make sharp right on Arastradero Rd
  SR: what road

- Seem to be ambiguous in general
  - Can be asking for verbatim repetition of queried element
  - Can be asking for clarification of reference
- Would like to know which (although could answer for both)
- Either way, need to establish which element is being queried
  - Don't want to repeat whole utterance

Non-matching CRs

- Reference questions may involve non-identical terms:

  Subj: how long

  Wiz: dave's house is sixteen minutes away

  Subj: was that one six or six zero minutes

  Wiz: six minutes away

- Even apparent requests for repetition:

  Wiz: after left at elm street turn right at lois lane

  Subj: was that right on lois lane or left on lois lane

  Wiz: turn right at lois lane

- Of course, this may be a result of ASR errors
- Antecedent identification becomes vital

Incorrect CR hypotheses

- Importantly, the user's hypothesis may be wrong:

  Subj: how long
  Wiz: dave's house is six minutes away
  Subj: was that one six or six zero minutes
  Wiz: six minutes away

  Wiz: go straight for three blocks turn right at wall street
  Subj: please repeat left where

  Wiz: go straight for three blocks turn right at wall street
  Subj: left where

- Antecedent identification becomes vital

What do we need to do?

- Need to be able to recognize the particular CR type
  - Some CR types are easier to recognize than others
- Need to be able to identify the antecedent
  - Association of words/phrases with their semantic content
- Need to be able to find the required information in context
  - Utterance history
  - Move type history
  - Semantic representation
  - Real-world (dynamic) context

Small steps …

- The current system can handle various repetition requests
- Repeat last utterance
  - Temporal utterance record
  - Slight complication, as we need to avoid e.g. error messages
- Repeat last navigation instruction
  - Requires move history
  - Really requires semantic check (re-generate rather than repeat)
- Actually takes care of a lot of user CRs in this domain

Medium steps …

- Queries about fragment reference and/or repetition
  - Minimally requires representation of constituency
  - Requires association of phrases & contents
- Need to spot reprise fragment CRs
  - Repeated fragments
    - Unless interpretable as other relevant move
    - Must repeat semantically potent element
  - WH-substituted fragments
    - Similar approach
- But this misses reformulations
  - Could perhaps treat with domain-specific lists
  - A general approach is more difficult (hard to know what counts as an alternative in context)
- Also misses possible ASR errors

Further steps

- Interpreting paraphrased or incorrect-hypothesis CRs
  - Phrase co-reference goes some way
  - Intended co-reference is rather more difficult to spot (“one six or six zero”)
- Disambiguation
  - Determine what’s being asked about
  - Determine how to answer it (repetition/reference)
  - Determine whether this is a CR in the first place
    - Dialogue systems usually try very hard to interpret things
    - CRs often interpretable as other commands/queries
- Incrementality
  - Most spoken systems now allow barge-in
  - We know that human-human CRs often occur mid-utterance
  - What might this mean for us?

Questions we need to ask

- What are CRs likely to ask about
  - Possible/likely phrase & word types
  - Requirements for lexical & semantic representation
- How are CRs likely to be phrased
  - Can surface form tell us what's going on?
- When are CRs likely to appear
  - Position relative to antecedent turn
  - Turn-by-turn: antecedent detection
  - Phrase-by-phrase: incremental processing
- How should CRs be answered?

CR Antecedents

- Corpus data can tell us what lexical and phrasal types are likely to be antecedents
- (Purver, Ginzburg & Healey, 2003)
  - Conversational English dialogue (BNC)
- (Rodriguez & Schlangen, 2004)
  - Task-oriented German dialogue (Bielefeld)
- (Rieser & Moore, 2005)
  - Task-oriented English dialogue (Communicator)

CR Antecedents: BNC results

Whole utterances		44%	
Nominal phrases		41%	
	det-N		*30%*
	pronoun		*23%*
	proper		*21%*
	CN		*27%*
Modifiers		7%	
Verbs & verb phrases		3%	
Function words		3%	

CR Antecedents: task-oriented

- No direct antecedent data, but can infer some

Intention recognition (whole utterance)	22%
Acoustic problems	12%
NP reference	24%
Deictic reference	27%
Action reference	0
Syntactic problems	0

CR Antecedents

- Most phrase-level CRs ask about nominals
- Very few ask about function words
- Almost all function word CRs were determiners
  - Numbers & quantifiers, rather than articles
- Very few ask about verbs or VPs (or actions)
- 94% nominals, modifiers & determiners
  - Can we get away with expecting just these?

Just a frequency effect?

- If this is actually just a frequency effect, that would be a dangerous assumption
- Not the case for the content/function distinction:

	CRs	General
Content	92.4%	69.2%
Function	7.6%	30.8%

- Not the case for the verb/noun distinction:

	CRs	General
Noun	93.9%	39.0%
Verb	6.1%	61.0%

Explaining the distinction (1)

- So why do we see these distinctions?
- With the content/function case, perhaps this is clear
  - Content words carry the semantic information
  - That's why they're called content words
- In that case, might see a variance effect
  - High variance of word counts across documents = high information content (high context-dependence)
  - (Kilgarriff, 1997; Francis & Kucera, 1982)
- Indeed, across the BNC (and other corpora), content words have a much higher average count variance
- Another possibility might be how likely words are to be rare (and therefore possibly not mutually known)
  - Ratio of average rarity matches the ratio of CR frequencies very well

Explaining the distinction (2)

- But we can't explain the verb/noun distinction so easily
- Verbs are no less common than nouns
- Verbs are no less contentful than nouns
  - Comparing average variances shows the opposite, in fact
- Verbs are more likely to be rare …
  - but not enough (about 3 times more)
- We see more verb types than noun types
  - But not enough (about 3 times more)
- Verb fragments are no less easy to interpret as CRs
  - Chat tool experiments (Healey et al., 2003)
- Less fine-grained semantics?
- Incremental processing effects?

What do we need?

- Semantic representation of nominal phrases
  - Usually present in any ISU system to some extent
    - Database entries, slot/value pairs …
  - Intermediate representation (LF) or database reference
  - Probably would prefer a non-GQ representation
  - Destination, waypoints, POIs …
- Semantic representation of interesting determiners
  - Present in intermediate representations
  - Only implicitly present in database reference
  - Cardinality of results sets (number of restaurants)
- Association of phrases with their semantic content
  - This is by no means standard, so must be added

CR Disambiguation

- Form-content correlations from BNC study
  - Reprise sentences tend to have y/n (clausal) readings
  - Reprise fragments similarly
- Suspect these may not generalize to the domain, though. But:
  - Domain data suggests often lexically specified
  - Domain data suggests strong bias to repetition
  - (not just of whole utterance, though)
- Seems likely that intonation will help
  - Pitch contours (Srinivasan & Massaro, 2003; Grice et al., 1995)
    - (see David's talk)
    - How well does this translate to HCI, though?
  - Available from standard speech recognizers?

When do CRs occur?

- All corpus studies show a strong preference for immediate clarification
  - 85% within 1 turn in the BNC
  - Most long-distance examples were unrepresentative
    - Multi-party dialogue
    - Repeated clarification sequences
- Stronger effect in task-oriented dialogue
  - 93-95% within 1 turn
  - All the NAV data is the immediately preceding turn
- Long-distance CRs tend to be explicit forms
  - Non-reprise e.g. “repeat the last directions”
- A default strategy of checking the immediately preceding turn unless incompatible seems OK

CRs and incrementality

- CHAT (as many systems) allows barge-in
- Do people ask CRs mid-utterance?
- Do we know anything about when to expect them?
  - Help us decide whether a turn is a CR or not
  - Help us identify the correct antecedent
  - Might also tell us something about how sentence processing works …
- Can't use existing test data (no user CR capability)
- Can't use existing WOZ data (not annotated for barge-in)
- BNC is annotated for speaker overlap
- Can use existing CR corpus to investigate possible patterns

Incrementality

- We know human-human CRs occur mid-sentence:

  A: They X-rayed me, and took a urine sample, took a blood sample.
  A: Er, the doctor
  B: Chorlton?
  A: Chorlton, mhm, he examined me, erm, he, he said now they were on about a slight [shadow] on my heart. Mhm, he couldn't find it.

- No clear examples in WOZ data, but we can imagine:

  Sys: take the next exit left, and then …
  Usr: which exit? This one?

- Incidentally, not many CRs occur mid-phrase
  - Because NPs need to be completed before resolution attempted?
- Do we really need incremental processing for this?
  - Fortunately, probably not
  - Need a representation of what's said/meant, but we already did

Avoid Mind-reading?

- Clearly, 0% of CRs ask about something that hasn't been said yet
- (Actually, this isn't quite true:
  - In human-human conversation, we see "fillers" (c. 4% of CRs)
  - Suggested completions after clear pauses/problems
  - But with systems delivering fully-formed utterances, shouldn't be possible)
- We need to know what we've said so far
  - Possible antecedents restricted to completed portions
- This may not be at all trivial
  - Most NLG systems are pipelined
  - The dialogue manager forms a complete move and passes it on
  - We need a TTS module which knows what it has said (bookmarking)
  - We need a representation from which we can determine what has therefore been expressed

How should we answer CRs?

- Again, we can look at the BNC:
  - Sluices: fragment answers
  - Conventional repetitions: full utterances
  - Reprise sentences: y/n
  - Reprise fragments: depends on the intention
    - Clausal “checks”: y/n
    - Constituent: fragments
- With human-computer dialogue, contrast:
  - Longer answers may be clearer
  - Brevity may be important in general (especially with in-car navigation)

Re-formulation

- Some apparent “repetitions” may need reformulation
  - Ensure information is still correct
- Really need to recognize the user’s intention in asking the CR
  - Asking about the words used: repeat
  - Asking about reference: reformulate
  - The difference could be really important:

    “Turn left here”

    “Where?”

    “Here.” vs. “The first exit”, “After Safeway” etc.
- This requires some departures from standard NLG
  - The best NP realization now depends on the context
  - Alignment with previous turns may be *undesirable*

Over-answering

- Hockey et al. (1997) looked at check-questions in the MapTask corpus
  - Only 40% of y/n answers were bare
  - More likely for negative answers
  - (distinct from general y/n answers: 64% bare)
- Perhaps a good strategy is to over-answer: y/n + fragment
- Particularly important for negative answers:

  "Turn left into Arastradero Road."
  "The first exit?"
  "No." vs. "No, the second exit."
- Especially so with possible ASR errors:

  "Turn left into Arastradero Road."
  "The first exit?" *ASR: THE THIRD EXIT*
  "No." vs. "No, the first exit."

Conclusions (roughly in order)

- Users do request clarification
  - Not necessarily in the same way they would from other people
- Need to represent what can be clarified
  - Need a record of what has been said
  - Need an intermediate semantic representation
  - Need an underlying model of the world
- Need to associate words/phrases with contents
- Need non-standard NLG for answering
- Need TTS bookmarking for barge-in
- Need more investigation for disambiguation
  - May include prosodic features

What can happen (2)

- SR: turn left where
- TTS: I'm sorry, I do not understand what you mean by: turn left where

- SR: turn right at what
- TTS: I'm sorry, I do not understand what you mean by: turn right at what

- SR: uh what's my next turn
- TTS: Go East on Miranda Ave
- SR: was that right on alma street
- TTS: I'm sorry, I do not understand what you mean by: was that right on alma street

- SR: where do i turn again
- TTS: I'm sorry, I do not understand what you mean by: where do i turn again