The Conversational User Interface

Ronald M Kaplan

Nuance Natural Language and Artificial Intelligence Laboratory
The graphical user interface is topping out.
Conversational UI: The solution

System: Can I help you?
User: I want to go to London on May 23

System: When do you want to leave?
User: I want to arrive by noon.

System: Which airline?
User: Get me the cheapest business class seat.

System: OK, there is an SAS flight leaving at…
User: Good, I’ll take it.

Bobrow et al., 1975

Intensional
Conversation and Information

• Ordinary language to describe what you need
  “When will my package arrive?”

• Clarification/repair
  “No, tomorrow”

• Drill-down discussion in context
  “What are the 15-year rates?”

• Immediate sentiment
  “You lost my luggage!”
Conversation and Action

• E-commerce
  “Book a flight to San Diego…”
  “Mexican restaurants?” “No, Italian” “OK, table for 4 at about 7”

• TV
  Direct command: “Change to channel 5”
  Standing order: “Turn the volume down during ads”

• Thermostat
  “A little cooler in the afternoon”
  “Vacation starting Tuesday”

• Customer service
  “Change my address to xxxx.”
From then to now: Obstacles

- Typing is unnatural, speech recognition is hard
- Language is efficient: much is unsaid but understood
  - Rampant ambiguity without context and expectations
- Language is complex
  - Many overlapping patterns to encode meaning
- Conversation is a cooperative social activity
  - Speaker/hearer model each other, share conventions, plan and reason
- You need something worth talking about
  - Accessible devices and information resources
  - Detect goals, track environment, determine/execute useful actions
From then to now: Obstacles

- Typing is unnatural, speech recognition is hard

- Language is efficient: much is unsaid but understood
  - Rampant ambiguity without context and expectations

  “The chicken is ready to eat”

- Language is complex
  - Many overlapping patterns to encode meaning

- Conversation is a cooperative social activity
  - Speaker/hearer model each other, share conventions, plan and reason

- You need something worth talking about
  - Accessible devices and information resources
  - Detect goals, track environment, determine/execute useful actions
The need/opportunity

Ubiquitous computing ⇒ ubiquitous complexity

• Mass distribution of computation and confusion
  – Proliferation of hard-to-control digitized devices
    TV, thermostat, clock, car…internet of things

• Phones and wearables
  – Universal: Applification of other connected devices
  – Personal and situational: preferred and appropriate behavior
    ⇒ The illusion of simplicity

• Cloud infrastructure: shared information and processing

• Advances on key components: speech, NL, dialog, reasoning

• Public interfaces to local devices, remote services

• [God Bless Siri: The NL Summer]
Speech recognition performance

Server dictation word error rate reduction ~ 18% / year
A simple conversation

A dialog between Bob and a speech-enabled proactive Conversational Assistant (CA)

Bob> Book a table at Zingari’s after my last meeting and let Tom and Brian know to meet me there.

CA> Sorry, but there aren’t any tables open until 9pm. Would you like me to find you another Italian restaurant in the area at about 6:30pm?

Bob> Can you find a table at a restaurant with a good wine list?

CA> Barbacco has an opening. It’s in the Financial District but the travel time is about the same.

Bob> Ok. That sounds good.
A simple conversation

A dialog between Bob and a speech-enabled proactive Conversational Assistant (CA)

Bob> Book a table at Zingari’s after my last meeting and let Tom and Brian know to meet me there.

CA> Sorry, but there aren’t any tables open until 9pm. Would you like me to find you another Italian restaurant in the area at about 6:30pm?

Bob> Can you find a table at a restaurant with a good wine list?

CA> Barbacco has an opening. It’s in the Financial District but the travel time is about the same.

Bob> Ok. That sounds good.

Ambiguous: booking done now or after last meeting?

Assumption: last meeting today (check calendar) and dinner tonight (tomorrow also meets constraints)

Expectation: Bob usually spends 30 minutes on email before leaving work

Factor in travel time → 6:30 for reservation

Referent for Tom and Brian
Bob> Book a table at Zingari’s after my last meeting and let Tom and Brian know to meet me there.

CA> Sorry, but there aren’t any tables open until 9pm. Would you like me to find you another Italian restaurant in the area at about 6:30pm?

Bob> Can you find a table at a restaurant with a good wine list?

CA> Barbacco has an opening. It’s in the Financial District but the travel time is about the same.

Bob> Ok. That sounds good.

• Initial search fails
• Informative explanation, not just “I can’t”
• Relax less important constraints, propose an otherwise similar alternative: type of restaurant and table time
• Expose most salient of remaining constraints
A simple conversation

A dialog between Bob and a speech-enabled proactive Conversational Assistant (CA)

Bob> Book a table at Zingari’s after my last meeting and let Tom and Brian know to meet me there.

CA> Sorry, but there aren’t any tables open until 9pm. Would you like me to find you another Italian restaurant in the area at about 6:30pm?

Bob> Can you find a table at a restaurant with a good wine list?

CA> Barbaccio has an opening. It’s in the Financial District but the travel time is about the same.

Bob> Ok. That sounds good.
A simple conversation

A dialog between Bob and a speech-enabled proactive Conversational Assistant (CA)

Bob> Book a table at Zingari's after my last meeting and let Tom and Brian know to meet me there.

CA> Sorry, but there aren't any tables open until 9pm. Would you like me to find you another Italian restaurant in the area at about 6:30pm?

Bob> Can you find a table at a restaurant with a good wine list?

CA> Barbacco has an opening. It's in the Financial District but the travel time is about the same.

Bob> Ok. That sounds good.

Drop one of the constraints ("restaurant in the area") in preference to others ("same travel time", "Italian", "tonight")
Bob> Book a table at Zingari’s after my last meeting and let Tom and Brian know to meet me there.

CA> Sorry, but there aren’t any tables open until 9pm. Would you like me to find you another Italian restaurant in the area at about 6:30pm?

Bob> Can you find a table at a restaurant with a good wine list?

CA> Barbacco has an opening. It’s in the Financial District but the travel time is about the same.

Bob> Ok. That sounds good.

• End of Dialog. CA goes to Opentable, makes the reservations, sends emails to Tom and Brian.

• Persistence: The duties of a true assistant are not yet complete. It must monitor the plan for unexpected events such as delays.
Language and reasoning

ASR → Syntax → Semantics → Pragmatics → Dialog → Knowledge Reasoning

Managing end-to-end ambiguity through hard constraints and probabilistic reasoning

- Bridging language and logic
- Inferring intent & preferences
- Modeling collaboration
- Representing knowledge
Major technical challenges:

- Integration of independent best-of-breed components
- Global resolution of ambiguity while preserving modularity
- Deployment at scale

Managing end-to-end ambiguity through hard constraints and probabilistic reasoning

- Bridging language and logic
- Inferring intent & preferences
- Modeling collaboration
- Representing knowledge
Ambiguity is pervasive

- ASR
- Mentions & Morphology
- Parsing
- Knowledge

A Clint Eastwood movie
Actor or director

Book a table after my last meeting

- play role or song
- walks noun or verb
- untieable knot (untie)able or un(tieable)

Bourne/Bond movies with Sean Connery
Ambiguity can be explosive…

… if alternatives multiply within or across modules
Ambiguity can be explosive…

… if alternatives multiply within or across modules

But: early (most probable?) resolution may get to bad result, quickly
Ambiguity can be explosive…

... if alternatives multiply within or across modules

But: early (most probable?) resolution may get to bad result, quickly

“Watch a movie with Tom Cruise” vs. “Watch a movie with Sally Jones”
actor friend
Ambiguity can be explosive…

… if alternatives multiply within or across modules

But: early (most probable?) resolution may get to bad result, quickly

“Watch a movie with Tom Cruise”  vs.  “Watch a movie with Sally Jones”

actor  friend

Local ambiguities need global resolution
Ambiguity management: keep it going

- Pack alternatives for decision by later modules (pragmatic reasoning and domain statistics)
- Choice doesn’t depend on “meeting” structure, so never unpacked
- Bet on independence, not inconsistency: a “nearly decomposable system” (Simon, 1962)
Technical approaches: data + rules

- Data driven – learning by **observation**
  - Classification and correlation
    (on the head: lots of data)
  - Add special domain concepts by examples
  - Probabilistic preference and disambiguation
  - Hard to tune, generalize from single events…but robust (sort of)
Technical approaches: data + rules

• Data driven – learning by observation
  – Classification and correlation
    (on the head: lots of data)
  – Add special domain concepts by examples
  – Probabilistic preference and disambiguation
  – Hard to tune, generalize from single events…but robust (sort of)

• Symbolic – learning by instruction
  – Interpretation: internal structures
    (on the tail: little data)
  – Deep linguistic structures provide statistical locality
  – Less domain dependent
  – Supports meaningful explanations
  – Easy to tune, but also robust
Technical approaches: data + rules

• **Data driven – learning by observation**
  – Classification and correlation
    (on the head: lots of data)
  – Add special domain concepts by examples
  – Probabilistic preference and disambiguation
  – Hard to tune, generalize from single events…but robust (sort of)

• **Symbolic – learning by instruction**
  – Interpretation: internal structures
    (on the tail: little data)
  – Deep linguistic structures provide statistical locality
  – Less domain dependent
  – Supports meaningful explanations
  – Easy to tune, but also robust

Appropriate combination: Trade data for knowledge
Semantic analysis
Bob> “Can you find a table at a restaurant with a good wine list?”

- Syntactic structure mapped to logical representation with event tokens, individual objects, properties and relations
- Davidsonian representation (event variables) supports incremental addition of new constraints by conjunction
- Discourse Representation Structures (DRS) for ease of manipulation, with translation to (first?) order logic for more general reasoning

\[ e_1, e_2, x, y \]
Surface_request(e1,e2)
Agent(e1,Bob), Agent(e2,CA)
Find(e2), Restaurant(x),
Object(e2,x)
Food(x,Italian), Open(x)
Available(y,x), Wine(y),
Good(y)

Logical representation
\[ e_1, e_2, x, y. \]
surface_request(e1,e2) \( \land \)
agent(e1,Bob)
\( \land \)
agent(e2,MA) \( \land \)
find(e2) \( \land \)
restaurant(x) \( \land \)
object(e2,x)
\( \land \)
food(x,Italian) \( \land \)
open(x) \( \land \)
available(y,x) \( \land \)
wine(y) \( \land \)
good(y)
Pragmatics

Example: Speech acts

Bob> “Can you find a table at a restaurant with a good wine list?”

- Transform surface speech act (ability to find a table?) into a request to make a reservation

\[
\begin{align*}
e_{1}, e_{2}, x, y \\
\text{Surface\_request}(e_{1}, e_{2}) \\
\text{Agent}(e_{1}, \text{Bob}), \text{Agent}(e_{2}, \text{CA}) \\
\text{Find}(e_{2}), \text{Restaurant}(x), \text{Object}(e_{2}, x) \\
\text{Food}(x, \text{Italian}), \text{Open}(x) \\
\text{Available}(y, x), \text{Wine}(y), \text{Good}(y)
\end{align*}
\]

\[
\begin{align*}
e_{1}, e_{2}, x, y \\
\text{Request}(e_{1}, e_{2}) \\
\text{Agent}(e_{1}, \text{Bob}), \text{Agent}(e_{2}, \text{CA}) \\
\text{Reserve}(e_{2}), \text{Restaurant}(x), \text{Object}(e_{2}, x) \\
\text{Food}(x, \text{Italian}), \text{Open}(x) \\
\text{Available}(y, x), \text{Wine}(y), \text{Good}(y)
\end{align*}
\]
Conversational interaction: Plan and replan

“Book a table at Zingari’s after my last meeting …”

Task recipe library

Dynamic Intention Structures

(Ortiz & Hunsberger, 2013)
Conversational interaction: Plan and replan

“Book a table at Zingari’s after my last meeting …”

Task recipe library

Book table

Get restaurant

Get time

Reserve

Get restaurant

Get restaurant

Get Guide

Find

Reserve

Yelp

Get candidate

Compare

Opentable

Dynamic Intention Structures

Book_table(e1)
Agent(e1,CA)
Object(e1,r),
Restaurant(r)
Date(d),Time(t)

Get_rest(e2)
Agent(e2,CA)

Get_time(e4)
d=12112
t=6:30pm

From user

Agent(e5,CA)
Object(e5,r)
Source(e5,
Opentable)
Available(r,d,t)

Opentable: not available

(Ortiz & Hunsberger, 2013)
Conversational interaction: Plan and replan

“Book a table at Zingari’s after my last meeting …”

Task recipe library

Dynamic Intention Structures (Ortiz & Hunsberger, 2013)
Conversational interaction: Plan and replan

“Book a table at Zingari’s after my last meeting …”

Task recipe library

Dynamic Intention Structures (Ortiz & Hunsberger, 2013)
Conversational interaction: New recipes

Describing new tasks from known recipes

| Bob> | Can you arrange a romantic evening? |
| CA>  | I don’t know how to do that. Can you tell me how? |
| Bob> | First **make a reservation** for two at a quiet, cozy **restaurant**. Then **find a theater** near the restaurant that’s showing a **romantic comedy**, and **buy** the tickets. |
| CA>  | OK, I now know how to plan a romantic evening. When do you want to eat? |
Conversational interaction: New recipes
Describing new tasks from known recipes

Bob> Can you arrange a romantic evening?

CA> I don’t know how to do that. Can you tell me how?

Bob> First make a reservation for two at a quiet, cozy restaurant. Then find a theater near the restaurant that’s showing a romantic comedy, and buy the tickets.

CA> OK, I now know how to plan a romantic evening. When do you want to eat?

• Graceful recovery from gaps in task knowledge
• Cross-domain constraints: time to eat/travel before movie
• Recipe stored for later use
Proactive monitoring, replan on failure
Anticipate glitches, create standing orders

Believe(CA, e.time(e,1730) leave(e) agent(e,Bob)
object(e,Office))  Intend(CA, replan(e1))

“If CA comes to believe that Bob hasn’t left the office by
5:30 pm, it will form the intention to replan the book-table
action”

CA> Bob, you’re running late. Should I change the
reservation?
Bob> Yes, I’ll be ready to leave in about 30 minutes.
Standing orders and proactivity

• Specific constraints on future/hypothetical events
  “Let me know when I get close to a café—but not Peets”
  “Move $1000 to my savings when my paycheck comes in”
  – Linguistic pipeline decodes idiosyncratic intent—long tail
  – Planner creates future-situation recognizer
  – Monitor watches and initiates action (location, time, bank…)

• Data-driven approach for big-head situations
  – Infer from common interests and repeated patterns of daily life
  – Little/no linguistic analysis
  – Templatic but flexible use of general planning and monitoring
  – User model and context awareness to suppress unwanted intrusions
Extending across domains

Linguistic analysis, conventions of conversation, planning principles remain

• General vocabulary and grammatical expressions of meaning are (mostly) domain independent
  “I want…” “Can you…” “Later than that” “No, French” “Maybe Monday”

• Structured representations can be interpreted according to context

• “Upper” ontology and axioms provide stable background
  – People, places, objects, action, time, cause-effect, desire, belief, intention

• New domain: augment general framework
  – Add/specialize vocabulary and ontology
  – Define constraints and inferences
  – Provide access to domain information sources and execution interfaces

• Architecture, algorithms, background are language independent
Conversation: Natural, efficient, effective

- Universal way of interacting with
  - Ubiquitous technology: Phone, TV, thermostat...
  - Information, Institutions, and services

- (Many) core technologies now exist
  - Challenge of integration, ambiguity management

- Perfection is not required: People misunderstand too
  - Need plausible failures
  - Conversation provides for easy repair

- Confirmation is often unnatural
  - A defensive hangover from the errorful past
  - Needed for actions with consequence
The trophy would not fit in the brown suitcase because it was too big. What was too big?
Answer 0: the trophy
Answer 1: the suitcase

The Winograd Schema Challenge

• Future intelligent personal assistants will need broad coverage of commonsense knowledge and reasoning

• Nuance is sponsoring the Winograd Schema Challenge as an alternative to the Turing Test and as a way to quantify progress

• Organized and administered by www.ComonsenseReasoning.org
Thank you