Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright © 2026. All rights reserved. Draft of January 6, 2026.

CHAPTER

25

Conversation and its Structure

Conversation is an intricate and complex joint activity, and conversations have structure. This is true of all conversations, whether they are conversations between people or conversations between people and language models. Understanding the structure of human conversations is an important social science and linguistic task. The concepts we introduce in studying human conversation can also be a useful tool for analyzing human-LLM conversations.

[This draft is the initial stub of a chapter that will introduce different kinds of conversational structure and how to annotate them computationally.]

25.1 Properties of Human Conversation

What are the conversational phenomena that take place when humans converse with each other? Are conversations between humans and machines different? Consider what goes on in the conversation between a human travel agent and a human client excerpted in Fig. 25.1.

C_1: . . . I need to travel in May.
A_2: And, what day in May did you want to travel?
C_3: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A_4: And you're flying into what city?
C_5: Seattle.
A_6: And what time would you like to leave Pittsburgh?
C_7: Uh hmm I don't think there's many options for non-stop.
A_8: Right. There's three non-stops today.
C_9: What are they?
A_{10}: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
C_{11}: OK I'll take the 5ish flight on the night before on the 11th.
A_{12}: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_{13}: OK.
A_{14}: And you said returning on May 15th?
C_{15}: Uh, yeah, at the end of the day.
A_{16}: OK. There's #two non-stops . . . #
C_{17}: #Act. . . actually #, what day of the week is the 15th?
A_{18}: It's a Friday.
C_{19}: Uh hmm. I would consider staying there an extra day til Sunday.
A_{20}: OK. . . OK. On Sunday I have . . .

Figure 25.1 Part of a phone conversation between a human travel agent (A) and human client (C). The passages framed by # in A_{16} and C_{17} indicate overlaps in speech.

25.1.1 Turns

turn

A dialogue is a sequence of **turns** (C_1, A_2, C_3, and so on), each a single contribution from one speaker to the dialogue (as if in a game: I take a turn, then you take a turn, then me, and so on). There are 20 turns in Fig. 25.1. A turn can consist of a sentence (like C_1), although it might be as short as a single word (C_{13}) or as long as multiple sentences (A_{10}).

Turn structure has important implications for spoken dialogue. A human has to know when to stop talking; the client interrupts (in A_{16} and C_{17}), so a system that was performing this role must know to stop talking (and that the user might be making a correction).

The same issues come up for LLMs; a system also has to know when to start talking. For example, most of the time in conversation, speakers start their turns almost immediately after the other speaker finishes, without a long pause, because people are can usually predict when the other person is about to finish talking.

endpointing

Spoken language models must also detect whether a user is done speaking, so they can process the utterance and respond. This task—called **endpointing** or **endpoint detection**— can be quite challenging because of noise and because people often pause in the middle of turns.

25.1.2 Speech Acts

speech acts

A key insight into conversation—due originally to the philosopher Wittgenstein (1953) but worked out more fully by Austin (1962)—is that each utterance in a dialogue is a kind of **action** being performed by the speaker. These actions are commonly called **speech acts** or **dialogue acts**: here's one taxonomy consisting of 4 major classes (Bach and Harnish, 1979):

Constatives:	committing the speaker to something's being the case (*answering*, *claiming*, *confirming*, *denying*, *disagreeing*, *stating*)
Directives:	attempts by the speaker to get the addressee to do something (*advising*, *asking*, *forbidding*, *inviting*, *ordering*, *requesting*)
Commissives:	committing the speaker to some future course of action (*promising, planning, vowing, betting, opposing*)
Acknowledgments:	express the speaker's attitude regarding the hearer with respect to some social action (*apologizing*, *greeting*, *thanking*, *accepting an acknowledgment*)

A user asking a person or a dialogue system to do something ('Turn up the music') is issuing a DIRECTIVE. Asking a question that requires an answer is also a way of issuing a DIRECTIVE: in a sense when the system says (A_2) "what day in May did you want to travel?" it's as if the system is (very politely) commanding the user to answer. By contrast, a user stating a constraint (like C_1 'I need to travel in May') is issuing a CONSTATIVE. A user thanking the system is issuing an ACKNOWLEDGMENT. The speech act expresses an important component of the intention of the speaker (or writer) in saying what they said.

25.1.3 Grounding

common ground
grounding

A dialogue is not just a series of independent speech acts, but rather a collective act performed by the speaker and the hearer. Like all collective acts, it's important for the participants to establish what they both agree on, called the **common ground** (Stalnaker, 1978). Speakers do this by **grounding** each other's utterances. Grounding means acknowledging that the hearer has understood the speaker (Clark, 1996). (People need grounding for non-linguistic actions as well; the reason an elevator button lights up when it's pressed is to acknowledge that the elevator has indeed been called, essentially grounding your action of pushing the button (Norman, 1988).)

Grounding is also important when the hearer needs to indicate that the speaker has *not* succeeded in performing an action. If the hearer has problems in understanding, she must indicate these problems to the speaker, again so that mutual understanding can eventually be achieved.

contribution

How is closure achieved? Clark and Schaefer (1989) introduce the idea that each joint linguistic act or **contribution** has two phases, called **presentation** and **acceptance**. In the first phase, a speaker presents the hearer with an utterance, performing a sort of speech act. In the acceptance phase, the hearer has to ground the utterance, indicating to the speaker whether understanding was achieved.

What methods can the hearer B use to ground the speaker A's utterance? Clark and Schaefer (1989) discuss a continuum of methods ordered from weakest to strongest:

Continued attention:	B shows she is continuing to attend and therefore remains satisfied with A's presentation.
Next contribution:	B starts in on the next relevant contribution.
Acknowledgment:	B nods or says a continuer like *uh-huh*, *yeah*, or the like, or an **assessment** like *that's great*.
Demonstration:	B demonstrates all or part of what she has understood A to mean, for example, by **reformulating** (paraphrasing) A's utterance or by **collaborative completion** of A's utterance.
Display:	B displays verbatim all or part of A's presentation.

Examples of these kind of grounding occur in the travel agent conversation. We can ground by explicitly saying "OK", as the agent does in A_8 or A_{10}. Or we can ground by repeating what the other person says; in utterance A_2 the agent repeats "in May", demonstrating her understanding to the client. Or notice that when the client answers a question, the agent begins the next question with "And". The "And" implies that the new question is 'in addition' to the old question, again indicating to the client that the agent has successfully understood the answer to the last question.

This particular fragment doesn't have an example of an *acknowledgment*, but there's an example in another fragment:

C: He wants to fly from Boston to Baltimore
A: **Uh huh**

continuer
backchannel

The word *uh-huh* here is a **continuer**, also often called an **acknowledgment token** or a **backchannel**. A continuer is a (short) optional utterance that acknowledges the content of the utterance of the other and that doesn't require an acknowledgment by the other (Yngve, 1970; Jefferson, 1984; Schegloff, 1982; Ward and Tsukahara, 2000).

25.1.4 Subdialogues and Dialogue Structure

conversation analysis
adjacency pair

Conversations have structure. Consider, for example, the local structure between speech acts discussed in the field of **conversation analysis** (Sacks et al., 1974). QUESTIONS set up an expectation for an ANSWER. PROPOSALS are followed by ACCEPTANCE (or REJECTION). COMPLIMENTS ("Nice jacket!") often give rise to DOWNPLAYERS ("Oh, this old thing?"). These pairs, called **adjacency pairs**, are composed of a **first pair part** and a **second pair part** (Schegloff, 1968), and these expectations can help systems decide what actions to take.

side sequence
subdialogue

However, dialogue acts aren't always followed immediately by their second pair part. The two parts can be separated by a **side sequence** (Jefferson 1972) or **subdialogue**. For example utterances C_{17} to A_{20} constitute a **correction subdialogue** (Litman 1985, Litman and Allen 1987, Chu-Carroll and Carberry 1998):

C_{17}: #Act. . . actually#, what day of the week is the 15th?
A_{18}: It's a Friday.
C_{19}: Uh hmm. I would consider staying there an extra day til Sunday.
A_{20}: OK. . . OK. On Sunday I have . . .

The question in C_{17} interrupts the prior discourse, in which the agent was looking for a May 15 return flight. The agent must answer the question and also realize that ''I would consider staying...til Sunday" means that the client would probably like to change their plan, and now go back to finding return flights, but for the 17th.

Another side sequence is the **clarification question**, which can form a subdialogue between a REQUEST and a RESPONSE. This is especially common in dialogue systems where speech recognition errors causes the system to have to ask for clarifications or repetitions like the following:

User: What do you have going to UNKNOWN_WORD on the 5th?
System: Let's see, going where on the 5th?
User: Going to Hong Kong.
System: OK, here are some flights...

presequence

In addition to side-sequences, questions often have **presequences**, like the following example where a user starts with a question about the system's capabilities ("Can you make train reservations") before making a request.

User: Can you make train reservations?
System: Yes I can.
User: Great, I'd like to reserve a seat on the 4pm train to New York.

25.1.5 Initiative

initiative

Sometimes a conversation is completely controlled by one participant. For example a reporter interviewing a chef might ask questions, and the chef responds. We say that the reporter in this case has the conversational **initiative** (Carbonell, 1970; Nickerson, 1976). In normal human-human dialogue, however, it's more common for initiative to shift back and forth between the participants, as they sometimes answer questions, sometimes ask them, sometimes take the conversations in new directions, sometimes not. You may ask me a question, and then I respond asking you to clarify something you said, which leads the conversation in all sorts of ways. We call such interactions **mixed initiative** (Carbonell, 1970).

Full mixed initiative, while the norm for human-human conversations, can be difficult for dialogue systems. The most primitive dialogue systems tend to use **system-initiative**, where the system asks a question and the user can't do anything until they answer it, or *user-initiative* like simple search engines, where the user specifies a query and the system passively responds. Even modern large language model-based dialogue systems, which come much closer to using full mixed initiative, often don't have completely natural initiative switching. Getting this right is an important goal for modern systems.

25.1.6 Inference and Implicature

Inference is also important in dialogue understanding. Consider the client's response C_2, repeated here:

A_2: And, what day in May did you want to travel?
C_3: OK uh I need to be there for a meeting that's from the 12th to the 15th.

Notice that the client does not in fact answer the agent's question. The client merely mentions a meeting at a certain time. What is it that licenses the agent to infer that the client is mentioning this meeting so as to inform the agent of the travel dates?

The speaker seems to expect the hearer to draw certain inferences; in other words, the speaker is communicating more information than seems to be present in the uttered words. This kind of example was pointed out by Grice (1975, 1978) as part of his theory of **conversational implicature**. **Implicature** means a particular class of licensed inferences. Grice proposed that what enables hearers to draw these inferences is that conversation is guided by a set of **maxims**, general heuristics that play a guiding role in the interpretation of conversational utterances. One such maxim is the maxim of **relevance** which says that speakers attempt to be relevant, they don't just utter random speech acts. When the client mentions a meeting on the 12th, the agent reasons 'There must be some relevance for mentioning this meeting. What could it be?'. The agent knows that one precondition for having a meeting (at least before Web conferencing) is being at the place where the meeting is held, and therefore that maybe the meeting is a reason for the travel, and if so, then since people like to arrive the day before a meeting, the agent should infer that the flight should be on the 11th.

implicature

relevance

These subtle characteristics of human conversations (**turns**, **speech acts**, **grounding**, **dialogue structure**, **initiative**, and **implicature**) are among the reasons it is difficult to build dialogue systems that can carry on natural conversations with humans. Many of these challenges are active areas of dialogue systems research.

25.2 Dialog Acts and Corpora

The ideas of speech acts and grounding are combined in a single kind of action called a **dialogue act**, a tag which represents the interactive function of the sentence being tagged.

dialogue act

Dialog acts can be used to analyze human-human conversation or human-LLM conversation. Both the nature of the participants and the type of dialogue (task-based or not task-based) influence the development of dialogue act tagsets.

Figure 25.2 shows a domain-specific tagset for the task of two people scheduling meetings. It has tags specific to the domain of scheduling, such as SUGGEST, used for the proposal of a particular date to meet, and ACCEPT and REJECT, used for acceptance or rejection of a proposal for a date, but also tags that have a more general function, like CLARIFY, used to request a user to clarify an ambiguous proposal.

Figure 25.3 shows a tagset for a restaurant recommendation system, and Fig. 25.4 shows these tags labeling a sample dialogue from the HIS system (Young et al., 2010). This example also shows the content of each dialogue acts, which are the slot fillers being communicated.

There are a number of more general and domain-independent dialogue act tagsets. In the DAMSL (Dialogue Act Markup in Several Layers) architecture inspired by

Tag	Example
THANK	*Thanks*
GREET	*Hello Dan*
INTRODUCE	*It's me again*
BYE	*Alright bye*
REQUEST-COMMENT	*How does that look?*
SUGGEST	*from thirteenth through seventeenth June*
REJECT	*No Friday I'm booked all day*
ACCEPT	*Saturday sounds fine*
REQUEST-SUGGEST	*What is a good day of the week for you?*
INIT	*I wanted to make an appointment with you*
GIVE_REASON	*Because I have meetings all afternoon*
FEEDBACK	*Okay*
DELIBERATE	*Let me check my calendar here*
CONFIRM	*Okay, that would be wonderful*
CLARIFY	*Okay, do you mean Tuesday the 23rd?*
DIGRESS	*[we could meet for lunch] and eat lots of ice cream*
MOTIVATE	*We should go to visit our subsidiary in Munich*
GARBAGE	*Oops, I-*

Figure 25.2 The 18 high-level dialogue acts for a meeting scheduling task, from the Verbmobil-1 system (Jekat et al., 1995).

Tag	Sys	User	Description
HELLO($a = x, b = y, \ldots$)	✓	✓	Open a dialogue and give info $a = x, b = y, \ldots$
INFORM($a = x, b = y, \ldots$)	✓	✓	Give info $a = x, b = y, \ldots$
REQUEST($a, b = x, \ldots$)	✓	✓	Request value for a given $b = x, \ldots$
REQALTS($a = x, \ldots$)	χ	✓	Request alternative with $a = x, \ldots$
CONFIRM($a = x, b = y, \ldots$)	✓	✓	Explicitly confirm $a = x, b = y, \ldots$
CONFREQ($a = x, \ldots, d$)	✓	χ	Implicitly confirm $a = x, \ldots$ and request value of d
SELECT($a = x, a = y$)	✓	χ	Implicitly confirm $a = x, \ldots$ and request value of d
AFFIRM($a = x, b = y, \ldots$)	✓	✓	Affirm and give further info $a = x, b = y, \ldots$
NEGATE($a = x$)	χ	✓	Negate and give corrected value $a = x$
DENY($a = x$)	χ	✓	Deny that $a = x$
BYE()	✓	✓	Close a dialogue

Figure 25.3 Dialogue acts used by the HIS restaurant recommendation system of Young et al. (2010). The Sys and User columns indicate which acts are valid as system outputs and user inputs, respectively.

the work of Clark and Schaefer (1989), Allwood et al. (1992), and (Allwood, 1995), each utterance is tagged for two types of functions, **forward-looking functions** like speech act functions, and **backward-looking** functions, like grounding and answering, which "look back" to the interlocutor's previous utterance (Allen and Core, 1997; Walker et al., 1996; Carletta et al., 1997; Core et al., 1999).

	Utterance	**Dialogue act**
U:	Hi, I am looking for somewhere to eat.	`hello(task = find,type=restaurant)`
S:	You are looking for a restaurant. What type of food do you like?	`confreq(type = restaurant, food)`
U:	I'd like an Italian somewhere near the museum.	`inform(food = Italian, near=museum)`
S:	Roma is a nice Italian restaurant near the museum.	`inform(name = "Roma", type = restaurant, food = Italian, near = museum)`
U:	Is it reasonably priced?	`confirm(pricerange = moderate)`
S:	Yes, Roma is in the moderate price range.	`affirm(name = "Roma", pricerange = moderate)`
U:	What is the phone number?	`request(phone)`
S:	The number of Roma is 385456.	`inform(name = "Roma", phone = "385456")`
U:	Ok, thank you goodbye.	bye()

Figure 25.4 A sample dialogue from the HIS System of Young et al. (2010) using the dialogue acts in Fig. 25.3.

Forward Looking Function	
STATEMENT	a claim made by the speaker
INFO-REQUEST	a question by the speaker
CHECK	a question for confirming information
INFLUENCE-ON-ADDRESSEE	(=Bach's directives)
OPEN-OPTION	a weak suggestion or listing of options
ACTION-DIRECTIVE	an actual command
INFLUENCE-ON-SPEAKER	(=Austin's commissives)
OFFER	speaker offers to do something, (subject to confirmation)
COMMIT	speaker is committed to doing something
CONVENTIONAL	other
OPENING	greetings
CLOSING	farewells
THANKING	thanking and responding to thanks

The backward looking function of DAMSL focuses on the relationship of an utterance to previous utterances by the other speaker. These include accepting and rejecting proposals (since DAMSL is focused on task-oriented dialogue), and grounding and repair acts:

Backward Looking Function	
AGREEMENT	speaker's response to previous proposal
ACCEPT	accepting the proposal
ACCEPT-PART	accepting some part of the proposal
MAYBE	neither accepting nor rejecting the proposal
REJECT-PART	rejecting some part of the proposal
REJECT	rejecting the proposal
HOLD	putting off response, usually via subdialogue
ANSWER	answering a question
UNDERSTANDING	whether speaker understood previous
SIGNAL-NON-UNDER.	speaker didn't understand
SIGNAL-UNDER.	speaker did understand
ACK	demonstrated via continuer or assessment
REPEAT-REPHRASE	demonstrated via repetition or reformulation
COMPLETION	demonstrated via collaborative completion

Fig. 25.5 shows a labeling of parts of our sample conversation using versions of

the DAMSL Forward and Backward tags.

[assert]	C_1:	. . . I need to travel in May.
[info-req,ack]	A_2:	And, what day in May did you want to travel?
[assert, answer]	C_3:	OK uh I need to be there for a meeting that's from the 12th to the 15th.
[info-req,ack]	A_4:	And you're flying into what city?
[assert,answer]	C_5:	Seattle.
[info-req,ack]	A_6:	And what time would you like to leave Pittsburgh?
[check,hold]	C_7:	Uh hmm I don't think there's many options for non-stop.
[accept,ack]	A_7:	Right.
[assert]		There's three non-stops today.
[info-req]	C_8:	What are they?
[assert, open-option]	A_9:	The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
[accept,ack]	C_{10}:	OK I'll take the 5ish flight on the night before on the 11th.
[check,ack]	A_{11}:	On the 11th?
[assert,ack]		OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
[ack]	C_{12}:	OK.

Figure 25.5 A potential DAMSL labeling of the beginning of the conversational fragment in Fig. 25.1.

Allen, J. and M. Core. 1997. Draft of DAMSL: Dialog act markup in several layers. Unpublished manuscript.

Allwood, J. 1995. An activity-based approach to pragmatics. *Gothenburg Papers in Theoretical Linguistics*, 76:1–38.

Allwood, J., J. Nivre, and E. Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26.

Austin, J. L. 1962. *How to Do Things with Words*. Harvard University Press.

Bach, K. and R. Harnish. 1979. *Linguistic communication and speech acts*. MIT Press.

Carbonell, J. R. 1970. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4):190–202.

Carletta, J., N. Dahlbäck, N. Reithinger, and M. A. Walker. 1997. Standards for dialogue coding in natural language processing. Technical Report 167, Dagstuhl Seminars. Report from Dagstuhl seminar number 9706.

Chu-Carroll, J. and S. Carberry. 1998. Collaborative response generation in planning dialogues. *Computational Linguistics*, 24(3):355–400.

Clark, H. H. 1996. *Using Language*. Cambridge University Press.

Clark, H. H. and E. F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

Core, M., M. Ishizaki, J. D. Moore, C. Nakatani, N. Reithinger, D. R. Traum, and S. Tutiya. 1999. The Report of the 3rd workshop of the Discourse Resource Initiative. Technical Report No.3 CC-TR-99-1, Chiba Corpus Project, Chiba, Japan.

Grice, H. P. 1975. Logic and conversation. In P. Cole and J. L. Morgan, eds, *Speech Acts: Syntax and Semantics Volume 3*, 41–58. Academic Press.

Grice, H. P. 1978. Further notes on logic and conversation. In P. Cole, ed., *Pragmatics: Syntax and Semantics Volume 9*, 113–127. Academic Press.

Jefferson, G. 1972. Side sequences. In D. Sudnow, ed., *Studies in social interaction*, 294–333. Free Press, New York.

Jefferson, G. 1984. Notes on a systematic deployment of the acknowledgement tokens 'yeah' and 'mm hm'. *Papers in Linguistics*, 17(2):197–216.

Jekat, S., A. Klein, E. Maier, I. Maleck, M. Mast, and J. Quantz. 1995. Dialogue acts in verbmobil. Verbmobil–Report–65–95.

Litman, D. J. 1985. *Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues*. Ph.D. thesis, University of Rochester, Rochester, NY.

Litman, D. J. and J. Allen. 1987. A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11:163–200.

Nickerson, R. S. 1976. On conversational interaction with computers. *Proceedings of the ACM/SIGGRAPH workshop on User-oriented design of interactive graphics systems*.

Norman, D. A. 1988. *The Design of Everyday Things*. Basic Books.

Sacks, H., E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Schegloff, E. A. 1968. Sequencing in conversational openings. *American Anthropologist*, 70:1075–1095.

Schegloff, E. A. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen, ed., *Analyzing Discourse: Text and Talk*, 71–93. Georgetown University Press, Washington, D.C.

Stalnaker, R. C. 1978. Assertion. In P. Cole, ed., *Pragmatics: Syntax and Semantics Volume 9*, 315–332. Academic Press.

Walker, M. A., E. Maier, J. Allen, J. Carletta, S. Condon, G. Flammia, J. Hirschberg, S. Isard, M. Ishizaki, L. Levin, S. Luperfoy, D. R. Traum, and S. Whittaker. 1996. Penn multiparty standard coding scheme: Draft annotation manual. `www.cis.upenn.edu/~ircs/discourse-tagging/newcoding.html`.

Ward, N. and W. Tsukahara. 2000. Prosodic features which cue back-channel feedback in English and Japanese. *Journal of Pragmatics*, 32:1177–1207.

Wittgenstein, L. 1953. *Philosophical Investigations. (Translated by Anscombe, G.E.M.)*. Blackwell.

Yngve, V. H. 1970. On getting a word in edgewise. *CLS-70*. University of Chicago.

Young, S. J., M. Gašić, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. 2010. The Hidden Information State model: A practical framework for POMDP-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174.