CS772: Deep Learning for Natural Language Processing (DL-NLP)

LM, CAI

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

Week 12 of 27th March, 2023

Re-cap

Recap: Conversational AI

Attempts at Automation

- InstructGPT:
  - ***Command/Request/Order → Response***
- ChatGPT:
  - Carry out a ***conversation***
  - Respect context (state), personalization, quality and quantity and respond
    - Input: *I have been promoted*
    - Appropriate response: *I am delighted/congratulations/great* ..
    - Inappropriate: *why did they promote you?*

Gricean Maxims: Cooperative Principle in Converstaion (Wikipedia)

- **Quantity, Quality, Relation, and Manner**
- Paul Grice, philosopher of language
- "*Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged*".
- Captures the LINK between utterances

Maxim of Quantity (length and depth)

- *Be* ***informative,*** and submaxims are:
  - Make your contribution as informative as is required (for the current purposes of the exchange).
  - Do not make your contribution more informative than is required.
- Grice’s analogy: "If you are assisting me to mend a car, I expect your contribution to be neither more nor less than is required. If, for example, at a particular stage I need four screws, I expect you to hand me four, rather than two or six."

Maxim of Quality (truth)

- Be ***Truthful***
- Submaxims:
  - Do not say what you believe is false.
  - Do not say that for which you lack adequate evidence
- Grice’s analogy: "I expect your contributions to be genuine and not spurious. If I need sugar as an ingredient in the cake you are assisting me to make, I do not expect you to hand me salt; if I need a spoon, I do not expect a trick spoon made of rubber."

Maxim of Relation (relevance)

- Information is ***relevant*** to the current exchange; therefore omitting any irrelevant information
- Grice’s analogy for this maxim: "I expect a partner’s contribution to be appropriate to the immediate needs at each stage of the transaction. If I am mixing ingredients for a cake, I do not expect to be handed a good book, or even an oven cloth (though this might be an appropriate contribution at a later stage)."

Maxim of Manner (clarity)

- Be ***perspicuous***
- Submaxims:
  - Avoid obscurity of expression — i.e., avoid language that is difficult to understand.
  - Avoid ambiguity — i.e., avoid language that can be interpreted in multiple ways.
  - Be brief — i.e., avoid unnecessary prolixity.
  - Be orderly — i.e., provide information in an order that makes sense, and makes it easy for the recipient to process it.

Examples of Gricean Maxims (1/3)

- ***Quantity:***
  - Question: *Do you know where the IITB library is*
  - *Answer: Yes I do, it is near the maths building*
  - *(violation): Yes, I do*
- ***Quality:***
  - Question: *Do you know where the IITB library is*
  - *Answer: Yes I do, it is near the maths building*
  - *(violation, if the answerer purposefully misleads): Yes, it is near the main gate*

Examples of Gricean Maxims (2/3)

- ***Relation:***
  - Question: *Do you know where the IITB library is*
  - *Answer: Yes I do, it is near the maths building*
  - *(violation): Mumbai is a big city*
  - *(violation of a different kind and degree): IITB campus is big! (so I do not know)*
  - *(flouting, but ok): I saw Ram enter the library from the maths buidling*
- ***Manner:***
  - Question: *Do you know where the IITB library is*
  - *Answer: Yes I do, it is near the maths building*

Examples of Gricean Maxims (3/3)

- ***Manner:***
  - Question: *Do you know where the IITB library is*
  - *Answer: Yes I do, it is near the maths building*
  - *(violation): Up yonder in the citadel of learning where polynominals are the bread and operators are the butter and where Hardy and Ramanujam permeate the atmosphere, thither will thee find the storehouse of what bibliophiles love* (maybe used for humorous effect)

What does ChatGP “think”? (1/2)

If for the question, "where is the library?" someone answers "Up yonder in the citadel of learning where polynominals are the bread and operators are the butter and where Hardy and Ramanujam permeate the atmosphere, thither will thee find the storehouse of what bibliophiles love", what would you say?

What does ChatGP “think”? (2/2)

Based on the answer given, it seems that the person is using figurative and poetic language to describe the location of the library. They are painting a vivid picture of the library being located in a place of high learning where mathematics and academic excellence are highly valued.

However, their answer may not be helpful if the person asking the question simply needs a straightforward answer to their question. In such a case, it might be more helpful to ask the person to provide a more direct and concrete answer to the question of where the library is located.

AI chatbots compared: Bard vs. Bing vs. ChatGPT

https://www.theverge.com/2023/3/24/23653377/ai-chatbots-comparison-bard-bing-chatgpt-gpt-4

Comparison: Chatbots

Google’s Bard (https://bard.google.com/),

Microsoft’s Bing (https://www.theverge.com/2023/3/24/23653377/ai-chatbots-comparison-bard-bing-chatgpt-gpt-4),

OpenAI’s ChatGPT (https://chat.openai.com/chat#)

Comparison: what is compared

Range of questions spanning
common requests

from holiday tips to
gaming advice to
mortgage calculations.

Summary from the blog: "*ChatGPT is the most verbally dextrous, Bing is best for getting information from the web, and Bard is... doing its best*"

Underlying systems

- OpenAI's ChatGPT: uses GPT-4
- Bing: GPT4; but Bing has other abilities
  - Can generate images
  - Can access the web
  - Offers sources for its responses
- However, "OpenAI announced it's launching plug-ins for ChatGPT that will allow the chatbot to also access real-time data from the internet."

“Give me a recipe for chocolate cake”

- Cake recipes offer room for creativity
- “Shift around the ratio of flour to water to oil to butter to sugar to eggs, and you’ll get a slightly different version of your cake: maybe drier, or moister, or fluffier”
- Chatbots combine different recipes to achieve a desired effect

Recipe for chocolate cake: ChatGPT, the best

- “chose a chocolate cake recipe from one site, a buttercream recipe from another, shared the link for one of the two, and reproduced both of their ingredients correctly. It even added some helpful instructions, like suggesting the use of parchment paper and offering some (slightly rough) tips on how to assemble the cake’s layers, neither of which were found in the original sources. This is a recipe bot I can trust!”

Recipe for chocolate cake: Bing

- “Bing gets in the ballpark but misses in some strange ways. It cites a specific recipe but then changes some of the quantities for important ingredients like flour, although only by a small margin. For the buttercream, it fully halves the instructed amount of sugar to include. Having made buttercream recently, I think this is probably a good edit! But it’s not what the author called for.”

Recipe for chocolate cake: Bard

- “Bard, … makes some changes that meaningfully affect flavor: it swaps buttermilk for milk and coffee for water. Later on, it fails to include milk or heavy cream in its buttercream recipe, so the frosting is going to end up far too thick. The buttercream recipe also seems to have come from an entirely different source than the one it cited. If you follow ChatGPT or Bing, I think you’d end up with a decent cake. But right now, it’s a bad idea to ask Bard for a hand in the kitchen”

“How do I install RAM into my PC?”

“The instructions should definitely guide people to their motherboard manual to ensure RAM is being installed optimally. ChatGPT does pick up on a key part of the RAM install process — checking your system BIOS afterward — but it doesn’t go through another all-important BIOS step. If you’ve picked up some Intel XMP-compatible RAM, you’ll typically need to enable this in the BIOS settings afterward, and likewise for AMD’s equivalent. Otherwise, you’re not running your RAM at the most optimized timings to get the best performance. Overall, the advice is solid but still very basic. It’s better than some PC building guides, but I’d like to have seen the BIOS changes or dual-channel

“Write me a poem about a worm” (1/2)

- Anapestic tetrameter: an arcane meter
  - 1.“Twas the night before Christmas, when all through the house/Not a creature was stirring, not even a mouse;
    2. The stockings were hung by the chimney with care,/In hopes that St. Nicholas soon would be there;
    3. The children were nestled all snug in their beds;/While visions of sugar-plums danced in their heads;”

Question Answering

- "… answer questions about passages taken from fiction (mostly Iain M. Banks books, as those were the nearest ebooks I had to hand). Again, ChatGPT/GPT-4 was the best, able to parse all sorts of nuances in the text and make human-like inferences about what was being described, with Bard making very general and unspecific comments (though often identifying the source text too, which is a nice bonus). Clearly, ChatGPT is the superior system if you want verbal reasoning."

Basic Math

- “asked each chatbot to determine monthly repayments and total repayment for a mortgage of $125,000 repaid over 25 years at 3.9 percent interest. None offered the answer supplied by several online mortgage calculators, and Bard and Bing gave different results when queried multiples times.
- GPT-4 was at least consistent, but failed the task because
- It insisted on explaining its methodology (good!) and then was so long-winded it ran out of space to answer (bad!).”
- 20% increase of 2,230=2676; chatGPT and Bing got it right, but not BARD

Basic Math (cntd.)

- "Bing, for example, booted me to a mortgage calculator site when I asked about mortgages, and
- ChatGPT's forthcoming plugins include a Wolfram Alpha option which should be fantastic for all sorts of complicated sums. But in the meantime, don't trust a language model to do a math model's work. Just grab a calculator."

"What's the average salary for a plumber in NYC? (And cite your sources)" (1/2)

- Bing's cited sources include Zippia, CareerExplorer, and Glassdoor

- ChatGPT best: gave a ballpark figure, explained that there were caveats, and told about what sources one could check for more detailed numbers

- Bing: gives specific numbers, cites its sources, and even gives links. But fudges the final two numbers. Both are close to their actual total, but for some reason, the bot just decided to change them up a bit.

“What’s the average salary for a plumber in NYC? (And cite your sources)” (2/2)

- Bard: lot of hallucination. “Was the median wage for plumbers in the US $52,590 in May 2020? Nope, that was in May 2017.
- Did a 2021 survey from the National Association of Plumbers and Pipefitters determine the average NYC salary was $76,810? Probably not because, as far as I can tell, that organization doesn’t exist. Did the New York State Department of Labor find the exact same number in its own survey? I can’t find it if the agency did. My guess: Bard took that number from CareerExplorer and then made up two different sources to attribute it to.”

“Design a training plan to run a marathon” (1/2)

- “In the race to make a marathon training plan, ChatGPT is the winner by many miles.
- Bing barely bothered to make a recommendation, instead linking out to a *Runner’s World* article. This isn’t necessarily an irresponsible decision — I suspect that *Runner’s World* is an expert on marathon training plans! — but if I had just wanted a chatbot to tell me what to do, I would have been disappointed.

“Design a training plan to run a marathon” (2/2)

- Bard’s plan was just confusing. It promised to lay out a three-month training plan but only listed specific training schedules for three weeks, despite saying later that the full plan “gradually increases your mileage over the course of three months.” The given schedules and some general tips provided near the end of its plan seemed good, but Bard didn’t quite go the distance.”

“When in Rome? Holiday tips”

- “all three chat bots seem fine. They’re very broad, choosing whole neighborhoods or areas, but the initial question prompt was also fairly broad. Rome is a unique place because you can cover a lot of touristy things in the heart of the city on foot, but it’s busy as all hell and you constantly get hounded by annoying grifters and scam artists at the touristy hotbeds. Many of these suggestions from Bing, Bard, and ChatGPT are fine for getting away from those busiest areas.”

"Testing reasoning: let's play find the diamond" (1/2)

- "Read the following story:
  'I wake up and get dressed, putting on my favorite tuxedo and slipping my lucky diamond into the inside breast pocket, tucked inside a small envelope. As I walk to my job at the paperclip bending factory where I'm gainfully employed I accidentally tumble into an open manhole cover, and emerge, dripping and slimy with human effluence. Much irritated by this distraction, I traipse home to get changed, emptying all my tuxedo pockets onto my dresser, before putting on a new suit and taking my tux to a dry cleaners.'
  Now answer the following question: where is the narrator's diamond?"

“Testing reasoning: let’s play find the diamond” (2/2)

“Bard and Bing sometimes got the answer right, and ChatGPT occasionally got it wrong (and all models switched their answer when asked to try again). Do these results prove or disprove that these systems have some sort of reasoning capability? This is a question that people with decades of experience in computer science, cognition, and linguistics are currently tearing chunks out of each other trying to answer, so I won’t venture an opinion on *that*. But just in terms of comparing the systems, ChatGPT/GPT-4 is again the most accomplished.”

I gave this test (1/2)

Read this story and examine everything is right about this story: "once a ship was going from England to France. After a few days of journey, it was caught in a violent storm. The captain tried a lot to steer the ship clear from the eye of the storm, but the ship sped to a cliff and dashed against it. All those on board died and only the captain survived. With a binocular, he could spot a spot of jungle very far. With great difficulty, he took his articles near the jungle and after going there fell unconscious. He regained consciousness after 3 days. After becoming conscious, the captain felt very hungry. He fortunately found some potatoes amongst his articles. (cntd.)

I gave this test (2/2)

He managed to get a fire ignited by striking two pieces of stone. There were some dry pieces of wood which he lighted and boiled the potatoes on the fire and appeased his hunger. When he had slept for some time, he woke up and found some shining objects on the sand where the fire was burnt. He kept these objects with him. After several days, he spied a ship far away with the binocular. He managed to attract attention of the ship with his torn shirt waving it vigorously. The ship rescued him from the island. The captain took those shining objects with him to England. Later they found interesting properties of these shining objects. These objects were transparent, but could be used to make containers. The material was given the name "glass". This is how glass

ChatGPT response (1/2)

There are several issues with this story, indicating that it is not entirely accurate.

Firstly, glass was not invented by a shipwrecked captain as described in the story. The history of glassmaking goes back thousands of years and involves the contributions of many different cultures.

Secondly, the story implies that glass is made from potatoes, which is completely false. Glass is made from a mixture of silica, soda ash, and limestone, among other materials.

ChatGPT response (2/2)

Thirdly, the story suggests that the captain discovered glass by chance, simply by finding some shiny objects on the sand. However, the process of making glass is complex and requires specialized knowledge and equipment.

Finally, the story is written in a somewhat simplistic and unconvincing manner, with several inconsistencies and unrealistic details, such as the captain managing to survive on a deserted island with only a few potatoes and no other resources for several days.

In summary, while the story may be entertaining, it is not a reliable source of information about the invention of glass.

Regenerate response (1/2)

There are a few issues with the story:

- The story describes the invention of glass, but glass was actually first produced by the ancient Mesopotamians around 3500 BCE, long before the time period in which the story is set.
- The story suggests that the captain single-handedly invented glass by discovering it on a deserted island. However, the invention of glass was a complex and gradual process that involved experimentation and refinement by many people over many centuries. (Cntd.)

Regenerate response (2/2)

Cntd.

- The story does not provide any specific details about the captain's journey or the location of the island, which makes it difficult to verify the accuracy of the events described.
- In conclusion, while the story may be an interesting work of fiction, it is not an accurate portrayal of the invention of glass.

3 stages of LLM based CAI

- Generative Pretraining (GP)
- Supervised Fine Tuning (SFT)
- Reinforcement Learning based on Human Feedback (RLHF)

Enter Pragmatics

Modeling

***P(e)*: "language" model**

$$e^* = \arg\max_e P(e \mid f)$$
$$= \arg\max_e [P(e)P(f \mid e)]$$

- Dialogue Act Classification (DAC): *f* → *Dialogue Sequence, e* → *Dialogue turn labels*
- Dialogue Intent: *f* → *dialogue sequence, e* → *dialogue turns with Intent like 'question', 'elaboration', 'affirmation', 'command/request' etc.*

Example of sequence of dialogue and dialogue acts

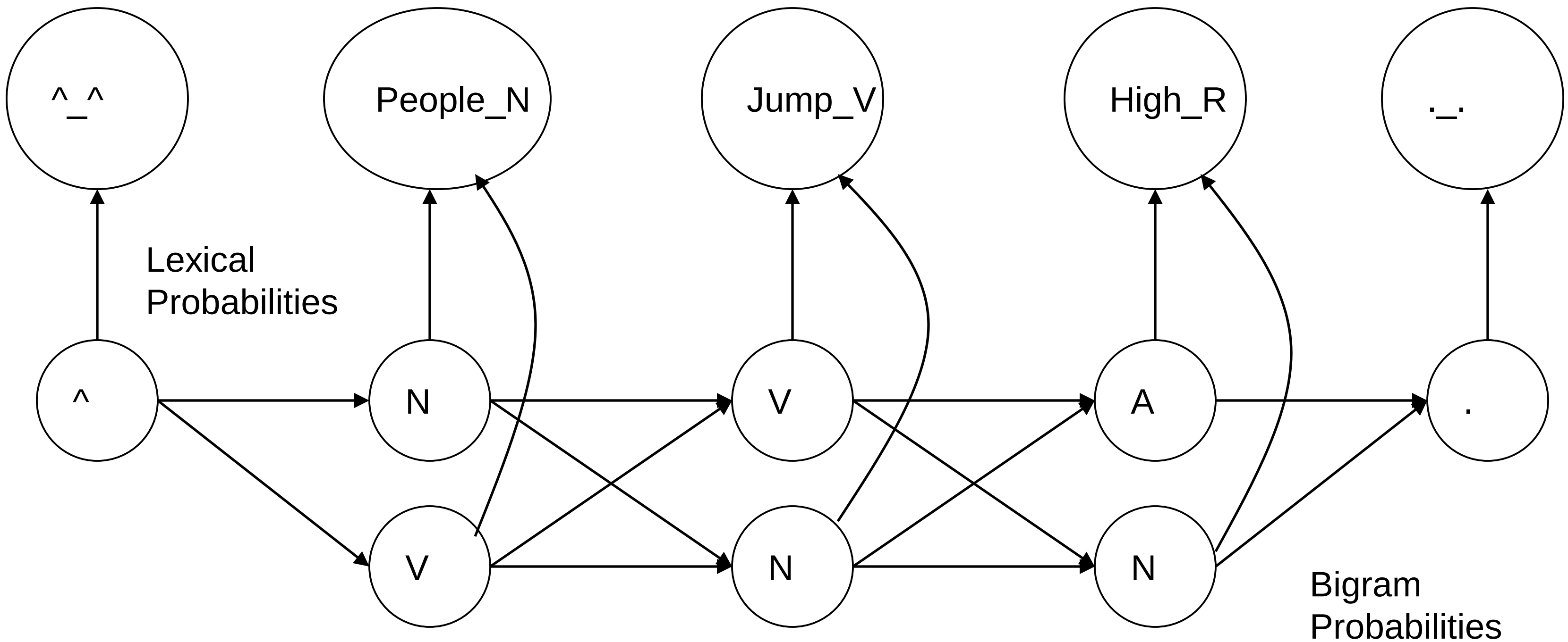
Fragment of a labeled conversation (from the Switchboard corpus).

Speaker	Dialogue Act	Utterance
A	YES-NO-QUESTION	So do you go to college right now?
A	ABANDONED	Are yo-,
B	YES-ANSWER	*Yeah,*
B	STATEMENT	*it's my last year [laughter].*
A	DECLARATIVE-QUESTION	You're a, so you're a senior now.
B	YES-ANSWER	*Yeah,*
B	STATEMENT	*I'm working on my projects trying to graduate [laughter].*
A	APPRECIATION	Oh, good for you.
B	BACKCHANNEL	*Yeah.*
A	APPRECIATION	That's great,
A	YES-NO-QUESTION	um, is, is N C University is that, uh, State,
B	STATEMENT	*N C State.*
A	SIGNAL-NON-UNDERSTANDING	What did you say?
B	STATEMENT	*N C State.*

Digression: What if there are many “abandoned”s

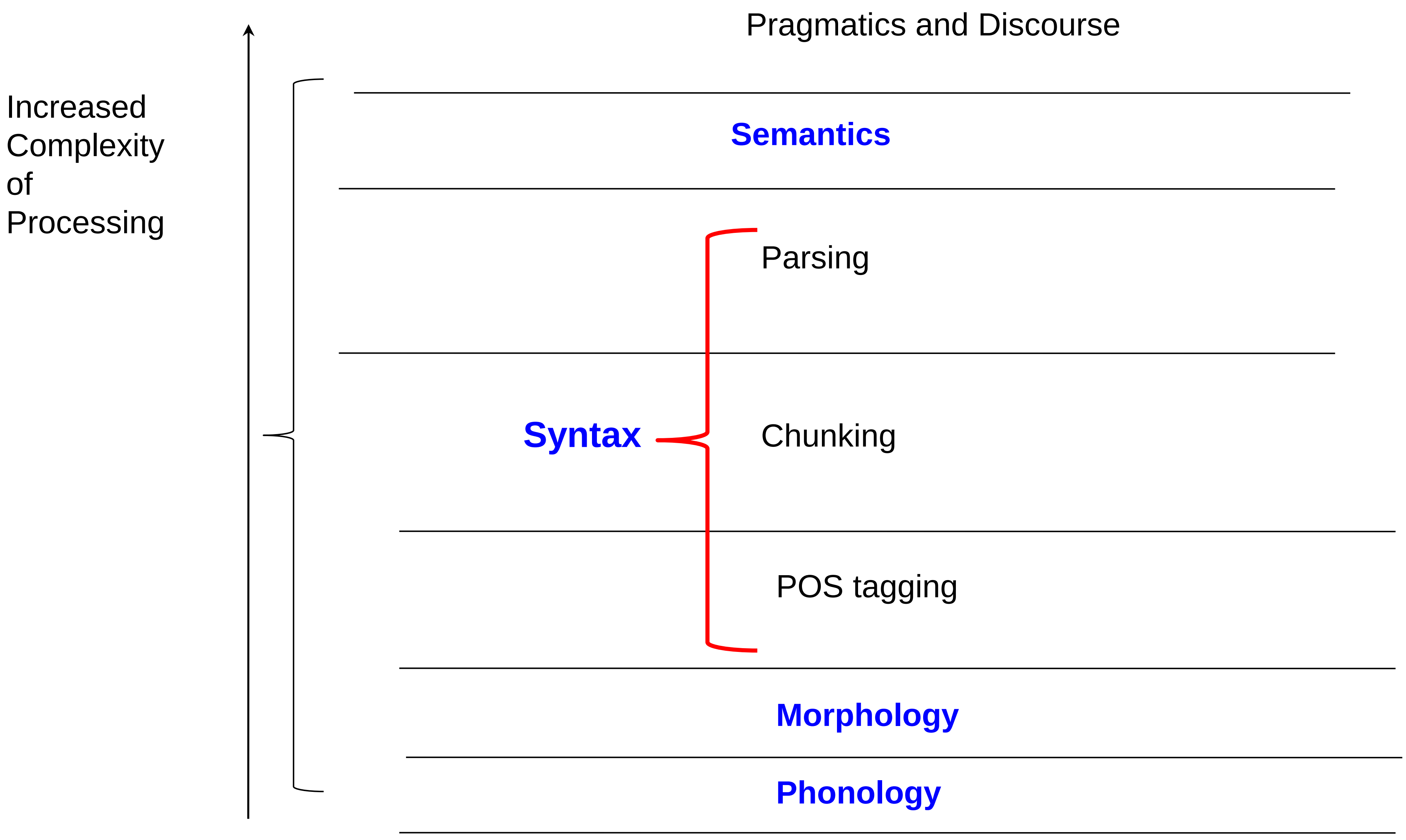
- Raise a flag
- Linguistic Limitation
- Or, Mental Health Problem
- Mental health doctors would like help of chatbots that can give preliminary help to mental health patients by engaging in a dialogue
- Or, they can do a preliminary screening based on disfluenices, abndonded statements and so on

HMM based POS tagging: Generative Model



This model is called Generative model.
Here words are observed from tags as states.
This is similar to HMM.

NLP (and linguistics) Layers

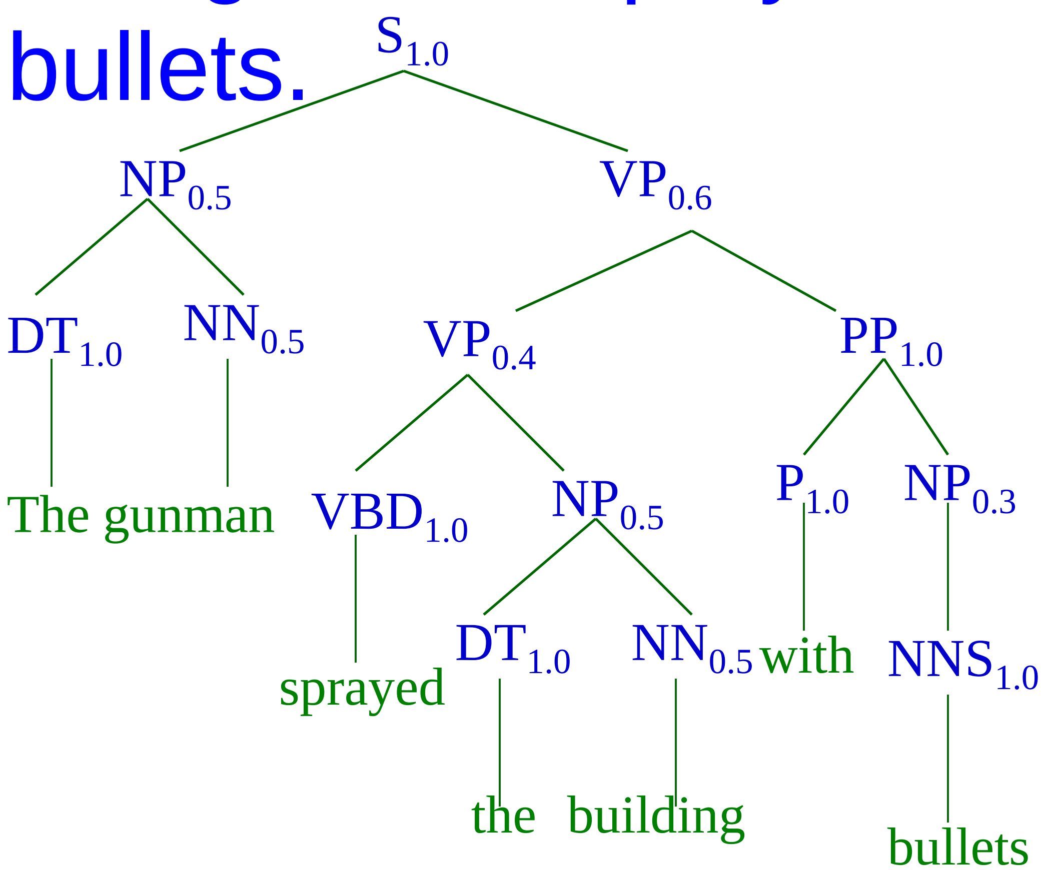


Which meaning of "*the gunman sprayed the building with bullets*" is more likely?: Semantics-Pragmatics interface

- The gunman has the bullets
- The building has the bullets
- The former is more likely
- Corroborated by data

Parse t_1`

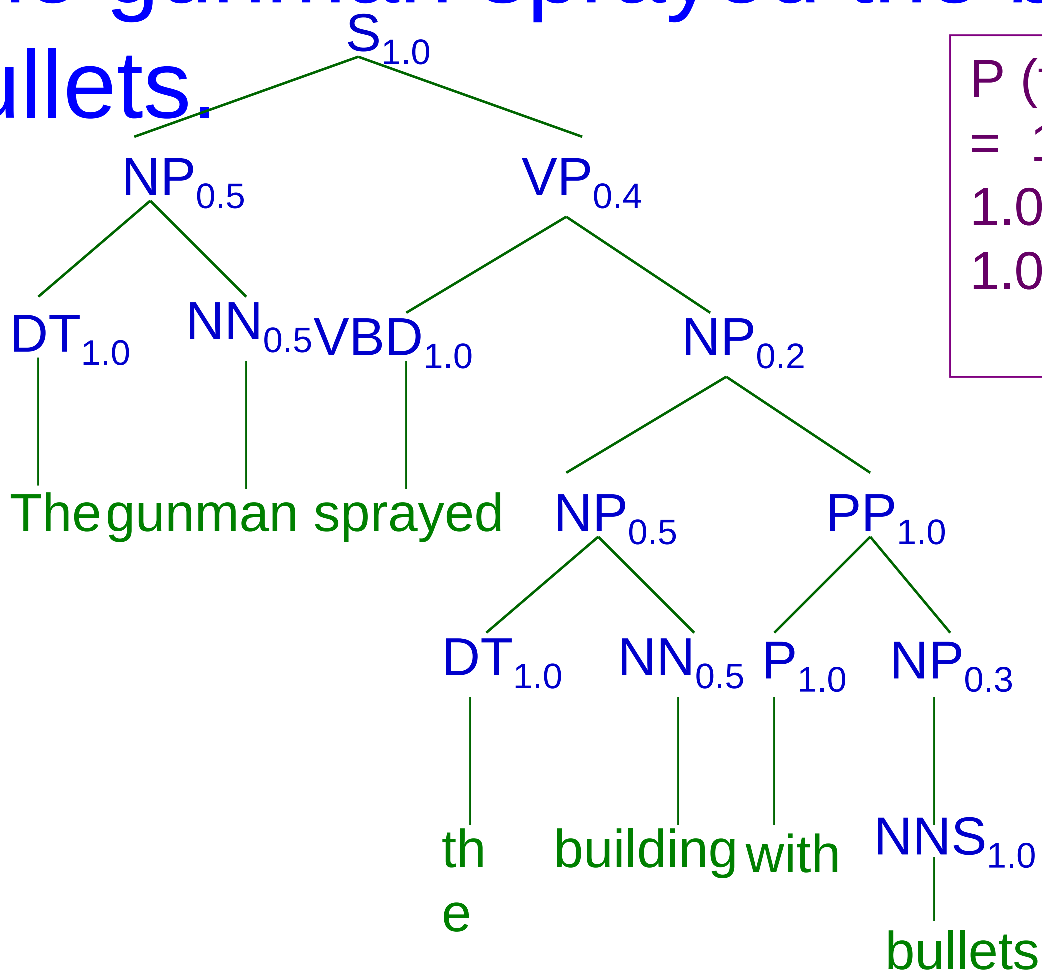
- The gunman sprayed the building with bullets.



$P(t_1) = 1.0 * 0.5 * 1.0 * 0.5 * 0.6 * 0.4 * 1.0 * 0.5 * 1.0 * 0.5 * 1.0 * 1.0 * 0.3 * 1.0 = 0.00225$

Parse t_2

- The gunman sprayed the building with bullets.

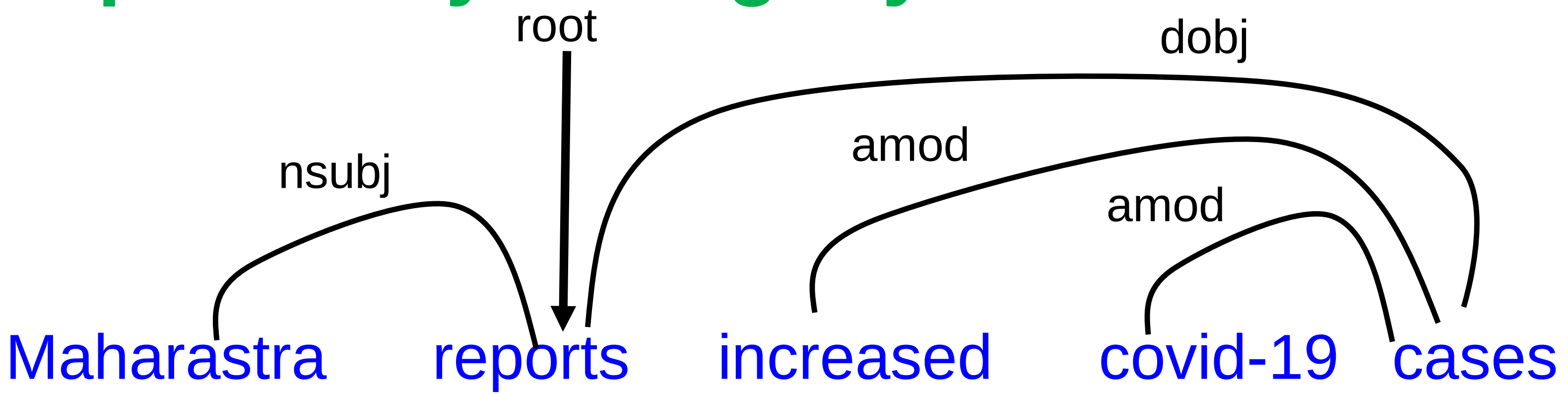


$P(t_2)$
$= 1.0 * 0.5 * 1.0 * 0.5 * 0.4 * 1.0 * 0.2 * 0.5 * 1.0 * 0.5 * 1.0 * 1.0 * 0.3 * 1.0$
$= 0.0015$

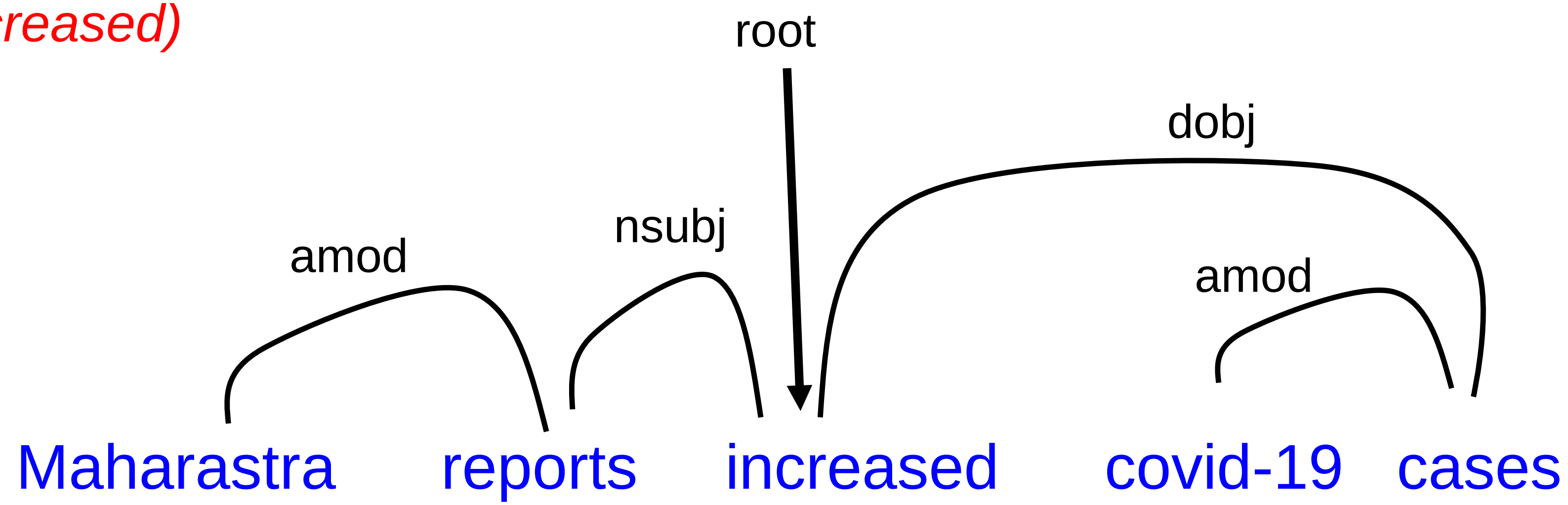
More examples of pragmatics constraining semantics

- **Command** Center to Track Best Buses (ToI 30Jan21)
  - Improbable meaning- *we should command the centre, i.e., Delhi to monitor the buses in Mumbai (BEST buses)*
- Elderly with young **face** increased covid 19 risk (ToI Oct 20)
  - Improbable: *old people who look young have increased covid 19 risk*

Dependency Ambiguity



(it is reported by Maharastra Govt. that covid-19 cases have increased)



(it is the Maharastra reports that have increased covid-19 cases!!!)

Meaning of "Pragmatics"

- Meaning in a Context
- To be contrasted with **Lexical Semantics**- word meanings
- **Sentential semantics**- truth value of a sentence and entailment (*a la* Montague)
  - *Today, the sky is blue* |= *Today, there is no rain*
- Pragmatics is extra-sentential
- Arises due to the limitations of lexical and formal semantics

Example of Pragmatics

- Dialogue or Conversation Setting

  - Speaker A: *shall we go for a walk?*
  - Speaker B: *It is raining outside*

- Implication: answer is NO

More examples

- (Person entering a room where there is an AC running): *Isn't it cold here*?
- Implication- stop or reduce the AC.

Another dimension of pragmatics

- Pragmatics is language is use

- Priest conducting a wedding ceremony in church: *I pronounce you man and wife*

- Leads to solemnisation of marriage

Elements of Pragmatics (1/2)

- Deixis (literally, ‘pointing with words’: temporal- *now, then;* spatial- *here, there*; personal- *I*, *you, he, they*; definite-indefinite- *this, that, those)*

- Presupposition: (*untie the shoe* → presupposes *the shoe was tied before*)

Elements of Pragmatics (1/2)

- Speech Acts: (*I pronounce you man and wife)-* **locutionary, illocutionary, and perlocutionary**
- Implicatures: (*A: shall we go for a walk? B: It is raining outside*)
- Politeness: (*close the door → please close the door → can you close the door → would you mind closing the door)*
- Information Structure: ordering of information (??*The table is under the flower pot*- odd; smaller object first mention) credit: Handke

Let us remember: Sanskrit Tradition

- **Shabdshakti (power inherent in word)**
  - Abhidha, Lakshana, Vyanjana

- **Meaning of Hall:**
  - *The hall is packed* (avidha)
  - *The hall burst into laughing* (lakshana)
  - *The Hall is full* (unsaid: and so we cannot enter) (vyanjana)

Abhidha, Lakshana, Vyanjana giving rise to

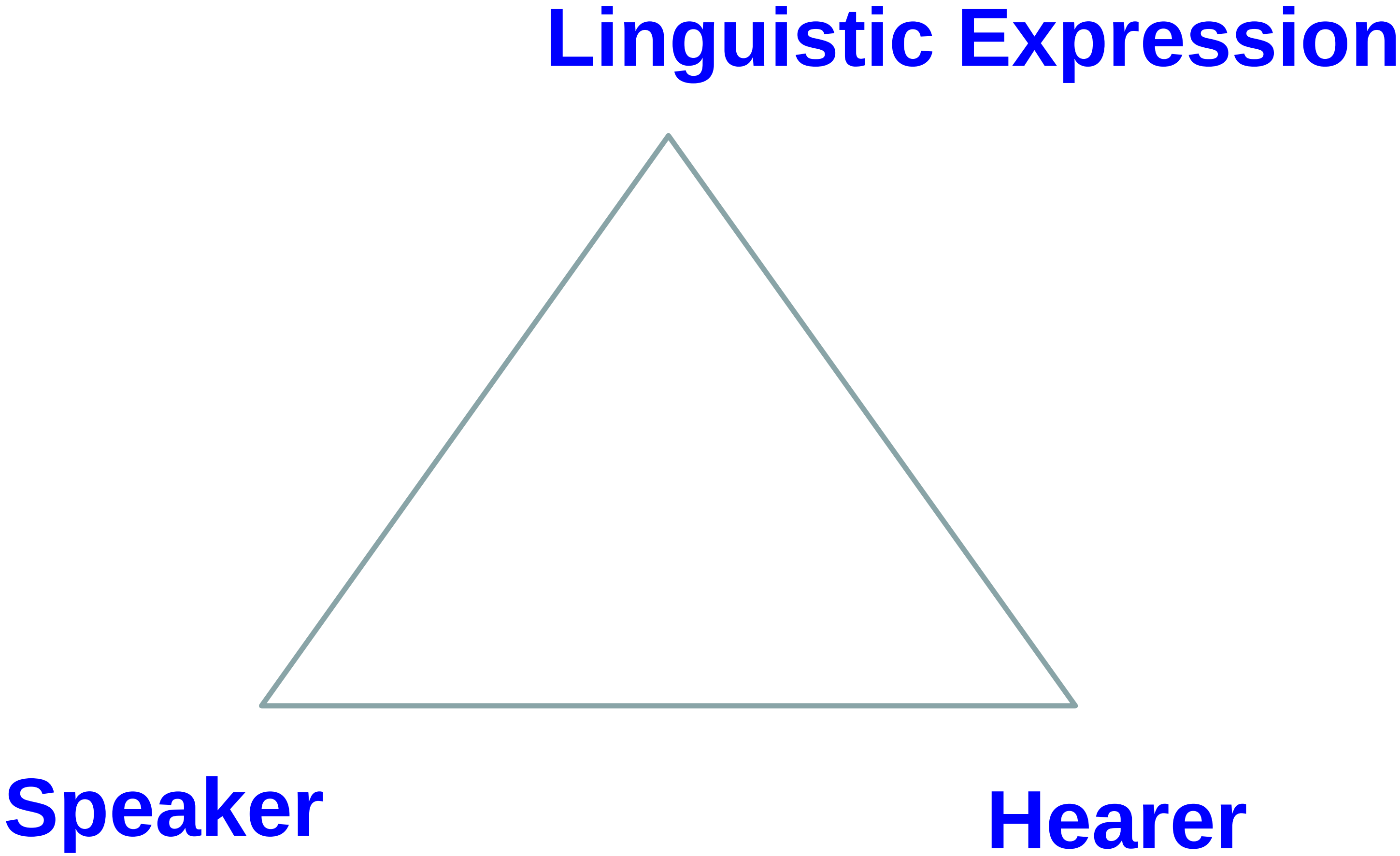
- Vachyartha, Lakshyartha, Vyangaartha

- **“Gangaa”:**
  - *vaachyaartha:* The river Gangaa (due to abhidhaa)
  - *lakshyaartha: gangaayaaM ghoshaH:* the house on river gangaa, meaning “on the bank of” (due to lakshanaa)
  - *Vyangaartha:* the house will have nice view, breeze etc. (unsaid) (due to vyanjana)

Crucial to Pragmatics

- Sentence vs. Utterance
- Semantics + Intent ➔ Pragmatics

The Trinity of Pragmatics



Communicative Aspects of language: nobody's baby? (Akmajian 2010) (1/2)

"Linguistics, focusing on structural properties of language, has tended to view communicative phenomena as outside its official domain. Likewise, it seems possible to pursue philosophical concerns about meaning, truth, and reference without investigating the details of communication. … cntd.

Communicative Aspects of language: nobody’s baby? (Akmajian 2010) (2/2)

(from prev slide)…Traditional psychology of language has focused on the processing of sentences, but without much concern for the specifics of communicative phenomena. Finally, some sociologists and anthropologists concern themselves with conversations, but have bypassed (or assumed an answer to) the question of the nature of communication itself.”

Syntax and semantics not enough

- "communicative process does not end with processing structural properties and decoding meaning."
- Syntactic tree → uncovers the structure
- Model theoretic semantics → uncovers lexical semantics and compositional meaning

Problems beyond reach of plain syntax and semantics (1/2)

- Ambiguity: “Flying planes can be dangerous” → what is dangerous? Act of flying or the planes?- airport zoning meeting vs. Pilot Insurance Board
- Reference: “The weather here is good”: which weather? Where?
- Intention: “mei tumhe bataataa hu”: promise (I will tell you)? Threat (I will teach you a lesson)?

Problems beyond reach of plain syntax and semantics (2/2)

- Non-literality: Sarcasm, Metaphor: “I love being ignored”
- Indirection: “My car has a flat tire” to a car mechanic is not just stating a fact, but wants and action
- Non-communicative acts: “I pronounce you man and wife”: the act of legalizing the marriage is not exactly in the message which has a normative, formal standing

Conversational Presumptions

- Relevance: The speaker's remarks are relevant to the conversation.
- Sincerity: The speaker is being sincere.
- Truthfulness: The speaker is attempting to say something true.
- Quantity: The speaker contributes the appropriate amount of information.
- Quality: The speaker has adequate evidence for what she says.

Diexis

Credit:
https://doi.org/10.1093/acrefore/9780199384655.013.213

Deictic Expressions

- Universal across languages
- “Used to individuate objects in the immediate context in which they are uttered, by pointing at them so as to direct attention to them.”
- Results in the Speaker (Spr) and Addressee (Adr) attending to the same referential object.
  - A: Oh, <u>there’s that</u> guy again (pointing)
  - B:Oh yeah, <u>now</u> I see him (fixing gaze on the guy)

Endophoric and Exophoric deixis

- Endophoric- refers to an object of discourse
- E.g., Anaphoric usage
  - “So you went to Boston, did you like it there?”

- Exophoric- Deictic (token) denotes an object in the extralinguistic context
  - “here, have a sip” (extending beverage to addressee)

Other Categorizations (Wikipedia)

- Personal: Grammatical person referred to, “do you know him?”
- Spatial: the place referred to, “do you enjoy living here?”
- Temporal: The time referred to, “he has gone now”
- Discourse: “This is a great story”; “that was a great account” (different from anaphora which refers to an ENTITY in the discourse, “I know the man, he live sin Delhi.”)
- Social: “thou, you” (En), “tu, tum, aap” (Hi), (honorifc) “aap ki shikshaa aallahabaad me hui” (“he” with respect)

Classifiers in Bengali: *ti, ta, te, to*

- Introduces definitiveness: shared understanding between the speaker and the addressee
  - ছেলেটি ভালো (Chēlēṭi bhālō): the boy is good
  - দুটো আম (duṭō ām): two mangoes
  - চারটে বেড়াল (Cāraṭē bēṛāla): four cats
  - An aside: East Asian languages, including Chinese, Korean, Japanese, and Vietnamese have classifiers. Classifiers are absent or marginal in European languages. In English, the work "piece": *three pieces of paper*

Speech Act

Definition

- "**speech act** is something expressed by an individual that not only presents information but performs an action as well" (Wikipedia)

- Purpose of language is not only to pass on information, but also to achieve an end

- Speech act is Speech+ Act
  - "I hereby resign from this job"

Kinds of Speech Act

- Locutionary
- Illocutionary
- Perlocutionary
- Performative Speech acts

Locutionary Speech Act

- The meaning that is on the surface of the utterance
  - *It is raining* → Stating the fact that it is raining

Recall…

- Vachyartha, Lakshyartha, Vyangaartha

- **“Gangaa”:**
  - *vaachyaartha:* The river Gangaa (due to abhidhaa)
  - *lakshyaartha: gangaayaaM ghoshaH:* the house on river gangaa, meaning “on the bank of” (due to lakshanaa)
  - *Vyangaartha:* the house will have nice view, breeze etc. (unsaid) (due to vyanjana)

Illocutionary Speech Acts

- "By saying something, we do something"- J. L. Austin 1962 (The classic book- "How to do things with words", Harvard University Press)

- Example:
  - A to B on a dining table, pointing to a jug: *Is that water,* meaning a request: *pl pass me the water*

Perlocutionary Speech Acts

- Perlocutionary acts always have a 'perlocutionary effect' which is the effect a speech act has on a listener

- Example:
  - A to B: *I am hungry*
  - (B goes to the fridge) *here have this sandwitch*

Performative Speech Acts

- Action that the sentence describes is performed by the utterance of the sentence itself
- Has self-reference!
- Examples
  - *I nominate you the chairman* (as opposed to *you are the chairman of the)*
  - *I pronounce you man and wife* (as opposed to *you now become man and wife)*
  - *I promise to pay you back* (as opposed to *I will pay you back)*

Subtle Differences between illocutionary, perlocutionary and performative (1/2)

- Illocutionary: express the intent (speaker centric)

- Perlocutionary: effect on the addressee (listener centric)

- Performative: self reference

Difference cntd.

- Example: *I promise you to pay back*
- Illocutionary: Intent to stick to the utterance
- Perlocutionary: The addressee accepts/rejects
- Performative: the utterance itself is the promise!

Implicatures

Examples all around

- Sign on a room in Amsterdam Airport:
  - BABY CHANGING ROOM (*what is changed in the room!*)

- A to B: *shall we go for a walk?*
- B*: it is raining outside* (implies ‘no’)

Implicatures and Abductive Reasoning

- Both are defeasible (can be ‘cancelled’)
- Digression
  - Deductive reasoning: *all men are mortal, Shakespeare is a man* → *Shakespeare is mortal* (indefeasible, provided axioms are indeed true)
  - Inductive Reasoning: *crows in Delhi are black, crows in Mumbai are black, …* → *all crows are black* (defeasible)
  - Abductive Reasoning: *if rain no cricket, no cricket* → *rain* (defeasible)

Implicatures are defeasible

- A: *shall we go for a walk?*
- B: *it is raining* (so no?)
- B: *I will take my rain coat* (yes☺)

Another categorization of speech act (1/2) *(J & M, 2006)*

- **Assertives:** committing the speaker to something's being the case (*suggesting*, *putting forward*, *swearing*, *boasting*, *concluding*).

- **Directives:** attempts by the speaker to get the addressee to do something (*asking*, *ordering*, *requesting*, *inviting*, *advising*, *begging*).

Another categorization of speech act (2/2)

- **Commissives:** committing the speaker to some future course of action (*promising, planning, vowing, betting, opposing*).
- **Expressives:** expressing the psychological state of the speaker about a state of affairs *thanking*, *apologizing*, *welcoming*, *deploring*.
- **Declarations:** bringing about a different state of the world via the utterance (including many of the performative examples above; *I resign*, *You're fired*.)

Think about these two pieces of dialogues in the context of pragmatics

1. (ToI, Oct 28, 2010)

Son: mother, I broke the neighbour's lamp shade.

Mother: then we have to give them a new one.

Son: no need, aunty said the lamp shade is irreplaceable.

2. (ToI, 3/11/10)

Ram: I got a Jaguar car for my wife.

Shyam: Wow, that's a great exchange!

Computational Perspective: Conversational AI

A dialogue

S:	How may I help you?
U:	*can you tell me how much it is to tokyo?*
S:	You want to know the cost of a call?
U:	*yes that's right*
S:	Please hold on for rate information.
S:	How may I help you?
U:	*what's the area code for chicago?*
S:	You want an area code.
U:	*(silence)*
S:	The area code for Chicago is 312.

Figure 22.2 The call routing domain: two conversations between a user (U) and the AT&T How May I Help You (HMIHY) system (S) (Gorin et al., 1997).

Modeling

$$e^* = \arg\max_e P(e \mid f)$$

$$= \arg\max_e [P(e)P(f \mid e)]$$

***P(e)*: “language” model**

- Dialogue Act Classification (DAC): *f→ Dialogue Sequence, e→Dialogue turn labels*
- Dialogue Intent: *f→ dialogue sequence, e→ dialogue turns with Intent like ‘question’, ‘elaboration’, ‘affirmation’, ‘command/request’ etc.*

Dialogue Based Computation

Zihao He, Leili Tavabi, Kristina Lerman, and Mohammad Soleymani. 2021. Speaker Turn Modeling for Dialogue Act Classification. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 2150–2157, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Tulika Saha, Aditya Patra, Sriparna Saha and Pushpak Bhattacharyya, *Towards Emotion-aided Multi-modal Dialogue Act Classification*, Association of Computational Linguistics Conference (**ACL 2020**), Seattle USA, 5-10 July, 2020.

Dialogue Act Classification and Sentiment

Tulika Saha, Aditya Patra, Sriparna Saha and Pushpak Bhattacharyya, *Towards Emotion-aided Multi-modal Dialogue Act Classification*, Association of Computational Linguistics Conference (**ACL 2020**), Seattle USA, 5-10 July, 2020.

Contributions (1/2)

- Investigated the role of emotion and multi-modality in determining DAs of an utterance
- Created of a novel dataset, EMOTyDA, containing emotion-rich videos of dialogues collected from various open-source datasets manually annotated with DAs
- Given an attention based (self, inter-modal, inter-task) multi-modal, multi-task framework for joint optimization of DAs and emotions

Contributions (2/2)

- Multi-modality and multi-tasking boosted the performance of DA identification compared to its unimodal and single task DAC variants

- Plan in future to incorporate conversation history, speaker information, fine-grained modality encoding to predict DA with more accuracy and precision

Dialogue Act Classification (DAC) and Multimodality

- DAC → Intent
- Each turn primarily *a question, a statement, or a request for action*
- Prior work: Jurafsky et al. (1997), Stolcke et al (2000), Verbree et al (2006), Kalchbrenner and Blunsom (2013), Liu et al. (2017), Ortega et al (2019), Saha et al (2019) etc.

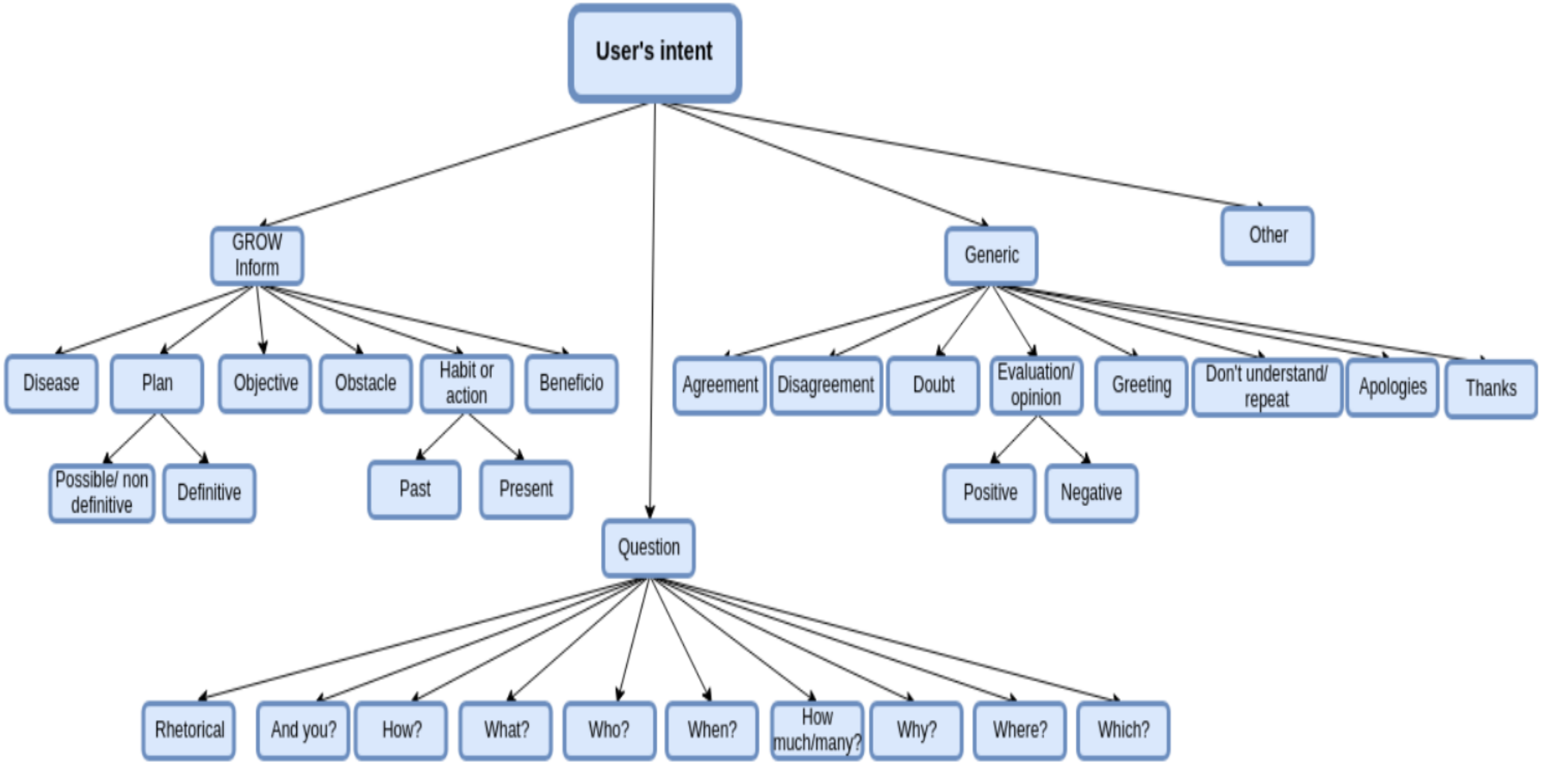
Emotion and Dialogue

- Non-verbal features
  - change of tone, facial expressions
  - provide beneficial cues to identify DAs
  - Emotion aided multi-modal DAC
  - “ha ji ha” in Hindi can denote agreement (statement) or disagreement *(sarcasm)*
- Contributions
  - Emotion-aware DA dataset (EMOTyDA)
  - Multi-modal, multi-task DNN for DAs and emotions identification
  - Showed: Multi-modality and multi-tasking DAC better than uni-modal and single task DAC

Prior Work

- Dating back to late 1990's (Reithinger and Klesen, 1997), (Stolcke et al., 1998) and early 2000's (Stolcke et al., 2000), (Grau et al., 2004)
- DA specification and taxonomy got established
- ***DAC = sequence labelling***
- DL Based:
  - Khanpour et al. (2016)- stacked LSTM
  - Kumar et al (2018)- hierarchical bi-LSTM and CRF
  - Raheja and Tetreault (2019)- contextual self-attention framework fused with hierarchical recurrent units
  - Yu et al (2019)- CNN

Dialogue Taxonomy



Example of dialogue and dialogue acts

Fragment of a labeled conversation (from the Switchboard corpus).

Speaker	Dialogue Act	Utterance
A	Yes-No-Question	So do you go to college right now?
A	Abandoned	Are yo-,
B	Yes-Answer	*Yeah,*
B	Statement	*it's my last year [laughter].*
A	Declarative-Question	You're a, so you're a senior now.
B	Yes-Answer	*Yeah,*
B	Statement	*I'm working on my projects trying to graduate [laughter].*
A	Appreciation	Oh, good for you.
B	Backchannel	*Yeah.*
A	Appreciation	That's great,
A	Yes-No-Question	um, is, is N C University is that, uh, State,
B	Statement	*N C State.*
A	Signal-Non-Understanding	What did you say?
B	Statement	*N C State.*

Dataset: EMOTyDA

- Short videos of dialogue conversations manually annotated with its DA along with its pre-annotated emotions
- Studied existing emotion recognition data sources
  - Youtube (Morency et al 2011), MOUD (P´erez-Rosas et al., 2013), IEMOCAP (Busso et al., 2008), ICT-MMMO (W¨ollmer et al., 2013), CMU-MOSI (Zadeh et al., 2016), CMU-MOSEI (Zadeh et al., 2018) and MELD (Poria et al., 2019)
- Zeroed down on IEMOCAP and MELD datasets
- Manually annotated for DAs

Data Annotation (1/2)

- SWBD-DAMSL tag-set consisting of 42 DAs (Jurafsky, 1997) for task-independent dyadic conversation such as SWBD corpus used
- Out of the 42 DAs of the SWBD-DAMSL tag-set, 12 most commonly occurring tags selected
- 12 frequently occurring chosen tags are
  - Greeting (g), Apology (ap), Command (c),
  - Question (q), Answer (ans), Agreement (ag),
  - Disagreement (dag), Statement-Opinion (o), Statement-Non-Opinion (s),
  - Acknowledge (a), Backchannel (b) and Others (oth).

Data Annotation (2/2)

- A subset of 1039 dialogues from MELD amounting to 9989 utterances and the entire IEMOCAP dataset of 302 dialogues amounting to 9376 utterances
- Three annotators graduate in English were assigned to annotate the utterances
- The inter-annotator score with more than 80% was considered as reliable agreement
- Mapped the *joy* tag of MELD to *happy* tag of the IEMOCAP

Particulars of EMOTyDA

1341 dyadic and multi-party conversations resulting in a total of 19,365 utterances or annotated videos with the corresponding DA and emotion tags considering the dialogue history.

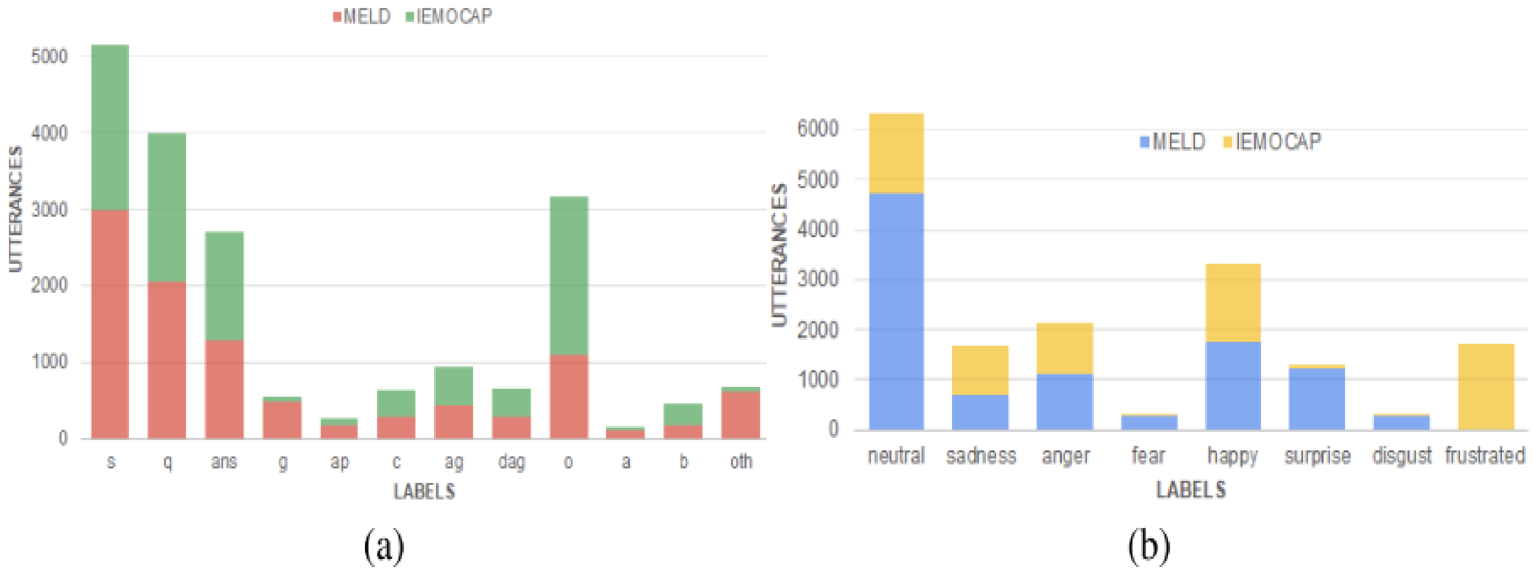


Figure 1: Statistics across the datasets : (a) Distribution of DA labels, (b) Distribution of emotion labels

Case for modality

Figure 2: (a) Incongruent modalities in DAC, (b) Importance of emotion in DAC.

Technique: Feature Extraction (1/2)

- **Text**: transcripts of each video; concatenation of pretrained GloVe (Pennington et al., 2014)
- **Audio**: OpenSMILE (Eyben et al., 2010), an open source software used
  - 12 Mel-frequency coefficients, glottal source parameters (Drugman et al., 2011), maxima dispersion quotients (Kane and Gobl, 2013),
  - several low-level descriptors (LLD) such as voice intensity, MFCC, voiced/unvoiced segmented features (Drugman and Alwan, 2011), pitch and their statistics (for example, root quadratic mean, mean etc.), voice quality (for example, jitter and shimmer), etc.

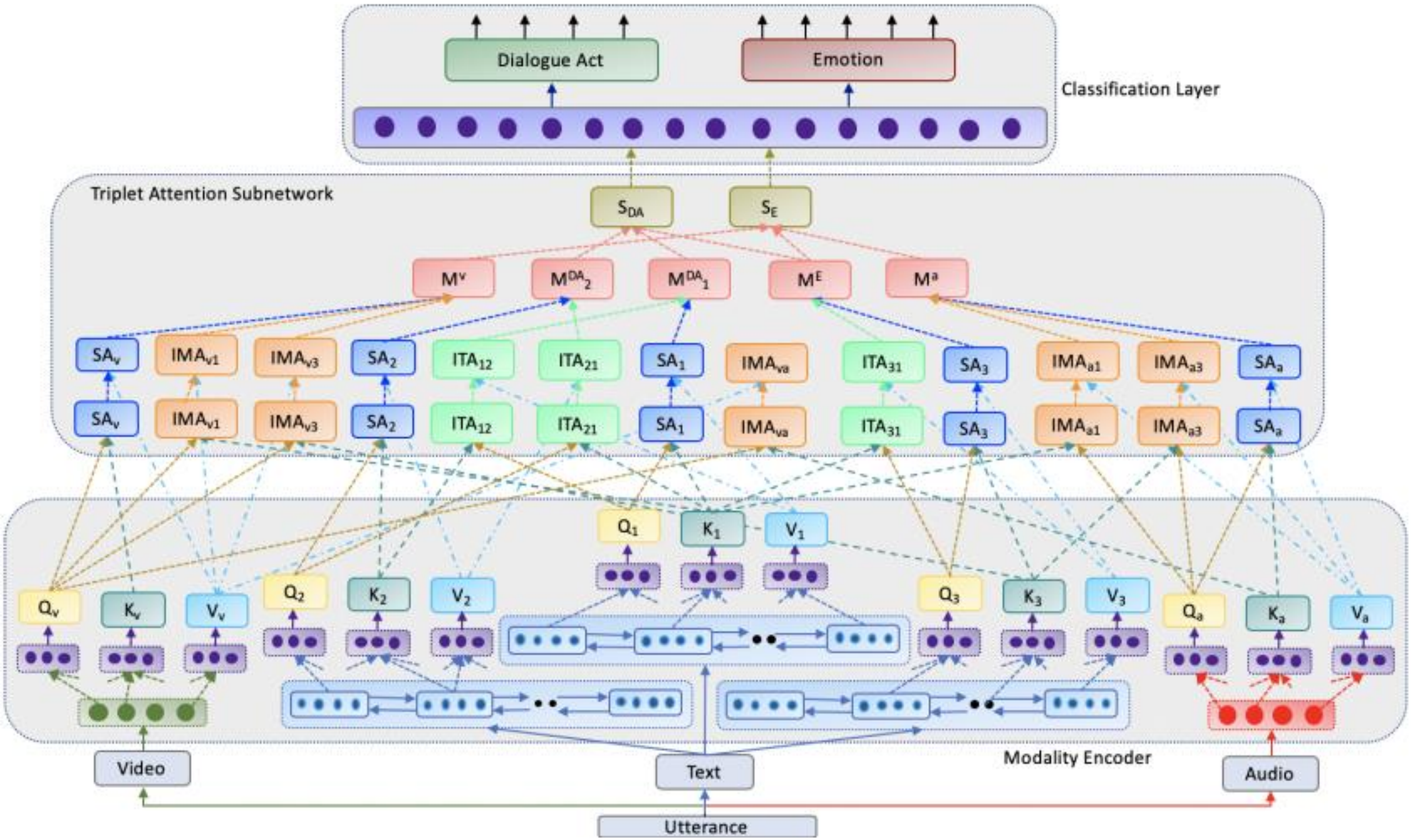
Technique: Feature Extraction (2/2)

- **Audio (cntd):**
  - Extracted features concatenated together to form a dq = 256 dimensional representation for each window. The final audio representation of
  - each utterance (A) is obtained by concatenating dq for every window
- **Video:**
  - ImageNet (Deng et al., 2009) pretrained ResNet-152 (He et al., 2016) used
  - Visual representation of each utterance (F) is obtained by concatenating the obtained df = 4096 dimensional feature vector for every frame

Network Architecture

- Three main components:
  - (i) Modality Enocoders (ME) which typically takes as input the uni-modal features and outputs the modality encodings,
  - (ii) Triplet Attention Subnetwork (TAS) that encompasses self, inter-modal and inter-task attention and
  - (iii) Classification layer that encompasses outputs of both the tasks (DAC and ER) to be learned jointly conditioned on the output of the TAS

Network Diagram



SA, IMA, ITA represents self, inter-modal and inter-task attentions respectively.

Results

	Dataset											
	EMOTyDA:dyadic				EMOTyDA:multiparty				EMOTyDA			
	DA		DA + ER		DA		DA + ER		DA		DA + ER	
Modality	**Acc.**	**F1-score**	**Acc.**	**F1-score**	**Acc.**	**F1-score**	**Acc.**	**F1-score**	**Acc.**	**F1-score**	**Acc.**	**F1-score**
Text (T)	63.75	60.67	65.23	62.35	46.20	39.23	48.90	41.10	53.56	49.17	53.02	50.22
Audio (A)	32.06	24.95	35.42	38.92	25.76	19.45	26.58	21.01	27.13	23.09	28.65	24.87
Video (V)	35.94	29.71	36.88	30.34	27.23	20.26	28.12	21.03	30.16	26.85	32.09	27.73
T + A	65.43	60.67	**66.98**	**62.08**	47.17	40.30	**49.42**	**41.69**	54.12	50.00	**56.62**	**51.99**
A + V	38.59	34.98	40.07	36.00	27.91	22.76	28.95	23.89	32.09	28.86	33.76	29.13
T + V	67.12	64.14	**70.55**	**68.12**	49.80	41.90	**51.00**	**44.52**	57.31	53.20	**60.88**	**57.96**
T + A + V	66.35	62.30	**69.45**	**67.00**	49.02	41.00	**50.65**	**44.00**	56.77	52.09	**59.86**	**56.05**
T + V (emotional cue)	65.26	60.20	-	-	46.88	39.70	-	-	54.31	50.02	-	-

Table 1: Results of the various models. Higher the values of accuracy and F1-score, better the performance of the corresponding model. All the reported results are statistically significant

Dyadic- IEMOCAP; multiparty- MELD;
EMOyDA- combined IEMOCAP and MELD

Case Study

Utterance	True Label	MT(T+V)	ST (T+V)
She is not Larry's girl	dag	dag	s
I know, it was amazing! I mean, we totally nailed it, it was beautiful.	ag	ag	o
Then why is she still single?,New York is full of men.,Why hasn't she married? Probably a hundred people told her she's foolish, but she's waited.	o	s	q
God, I,feel so guilty about Ross.	ap	ap	s

Table 2: Sample utterances with its predicted labels for the best performing multi-task (MT) (T+V) model and its single task (ST) DAC variants; These examples show that ER as an auxiliary task helps DAC for better performance in MT.

Greeting (g), Apology (ap), Command (c),
Question (q), Answer (ans), Agreement (ag),
Disagreement (dag), Statement-Opinion (o), Statement-Non-Opinion (s),
Acknowledge (a), Backchannel (b) and Others (oth).

Summary

- Investigated the role of emotion and multi-modality in determining DAs of an utterance
- Created of a novel dataset, EMOTyDA, containing emotion-rich videos of dialogues collected from various open-source datasets manually annotated with DAs
- Given an attention based (self, inter-modal, inter-task) multi-modal, multi-task framework for joint optimization of DAs and emotions
- Multi-modality and multi-tasking boosted the performance of DA identification compared to its unimodal and single task DAC variants