# Dialogue Systems
## NPFL123 Dialogové systémy

# 2. What happens in a dialogue

**Ondřej Dušek** & Ondřej Plátek & Jan Cuřín
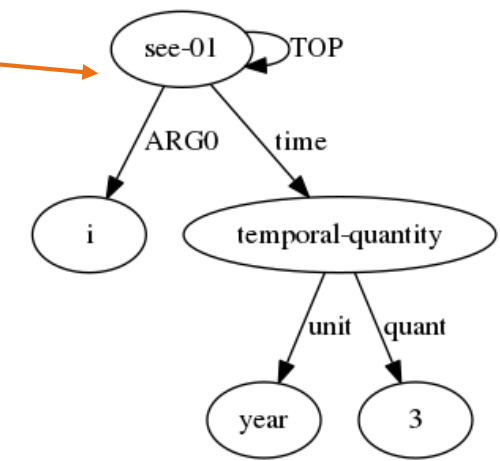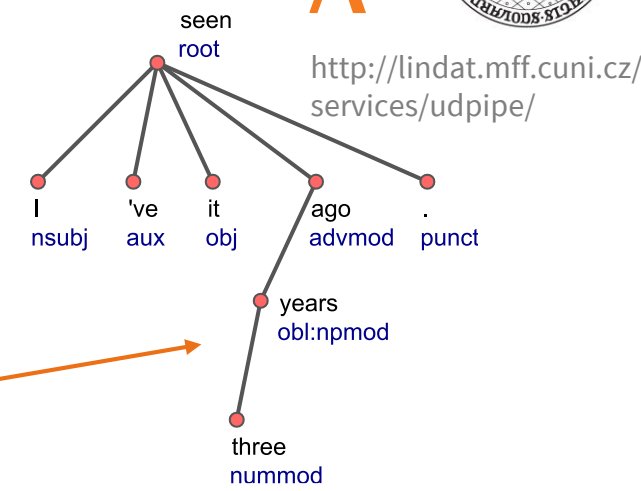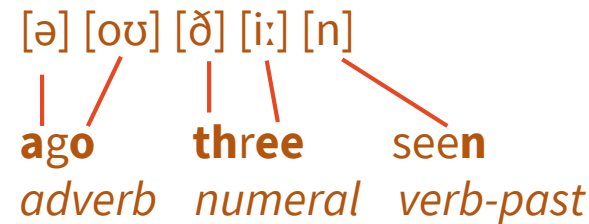
26. 2. 2019

# How do you "define" dialogue?

Spoken/written conversational (interactive, collaborative) communication between two or more people

- **verbal** + (possibly) non-verbal
    - can be multimodal (language + gestures, pitch, expressions…)
- **collaborative**, social
    - participants aim at communicative goal(s)
    - involves inference about intended meanings
- **practical**, related to actions
- **interactive**, incremental, messy!

Dialogue systems – simpler than that

# Describing a dialogue

- Levels of linguistic description
  - **phonetics / phonology** – sounds
  - **morphology** – word forms
  - **syntax** – sentence structure
  - **semantics** – sentence (propositional) meaning
  - **pragmatics** – meaning in context, communication goal

[ə] [oʊ] [ð] [iː] [n]

**a**g**o**     **th**r**ee**     see**n**
*adverb*     *numeral*     *verb-past*

seen
root

I     've     it     ago     .
nsubj   aux    obj   advmod   punct

years
obl:npmod

three
nummod

- This lecture is (a lot) about **pragmatics**

(I don't remember it well)

see-01     TOP

ARG0     time

i     temporal-quantity

unit     quant

year     3

I 've seen it three years ago .

# Turn-taking (interactivity)

- Speakers **take turns** in a dialogue
  - **turn** = continuous utterance from one speaker

- Normal dialogue – very fluent, fast
  - minimizing **overlaps** & **gaps**
    - little silence (usually <250ms), little overlap (~5%)
    - (fuzzy) rules, anticipation
  - cues/markers for turn boundaries:
    - linguistic (e.g. finished sentence), voice pitch
    - timing (gaps)
    - eye gaze, gestures (…)

- overlaps happen naturally
  - ambiguity in turn-taking rules (e.g. two start speaking at the same time)
  - **barge-in** = speaker starts during another one's turn
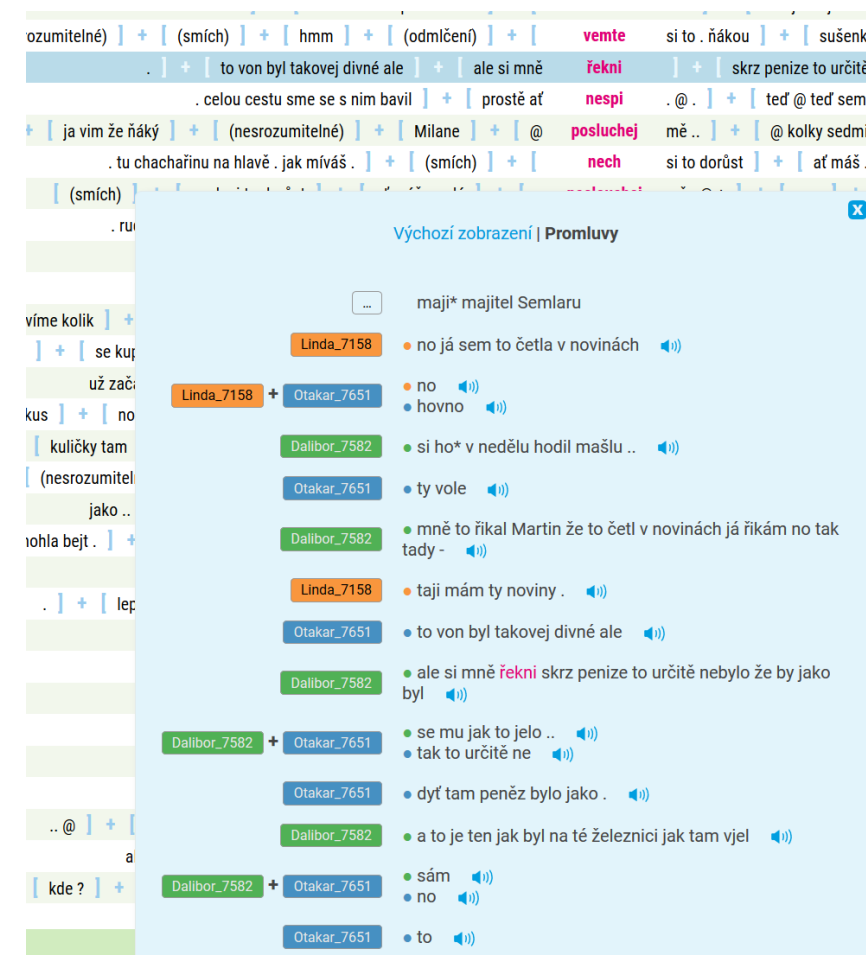
# Turn-taking (example)

20 seconds of a semi-formal dialogue (talk show):

```
S: um uh , you're about to start season [six ,]
J:                                       [yes]
S: you probably already started but [it launches]
J:                                  [yes thank you]
A:                                  (cheering)
J: we're about to start thank you yeah .. we're starting , we- on Sunday yeah ,
   we've been eh- we've been prepping some [things]
S:                                         [confidence] is high . feel good ?
J: (scoffs)
S: think you're gonna
   [squeeze out the shows this time ? think you're gonna do it ?]
J: (laughing) [you're talking to me like I'm an a-]
   confidence high ? no !
S: [no]
J: [my confidence] is never high .
S: okay
J: self loathing high . concern astronomic .
```

Transcription loosely following Roberts, 2003
https://www.kcl.ac.uk/sspp/departments/education/research/research-centres/ldc/knowledge-transfer/data/qualitative.aspx

# Speech vs. text

- Natural speech is **very different from written text**
  - ungrammatical
  - restarts, hesitations, corrections
  - overlaps
  - pitch, stress
  - accents, dialect
- See more examples in speech corpora
  - https://kontext.korpus.cz/ (Czech)
  - select the "oral" corpus and search for a random word

# Turn taking in dialogue systems

- consecutive turns are typically assumed
  - system waits for user to finish their turn (~250ms non-speech)
- **voice activity detection**
  - binary classification problem – "is it user's speech that I'm hearing?"[Y/N]
  - segments the incoming audio (checking every $X$ ms)
  - actually a hard problem
    - nothing ever works in noisy environments
- **wake words** – making VAD easier
  - listen for a specific phrase, only start listening after it

  *hey Siri*
  *okay Google*
  *Alexa*
- some systems allow user's barge-in
  - may be tied to the wake word

# Speech acts (by John L. Austin & John Searle)

- each utterance is an **act**
  - intentional
  - changing the state of the world
    - changing the knowledge/mood of the listener (at least)
    - influencing the listener's behavior

- speech acts consist of:
  a) **utterance** act = the actual uttering of the words
  b) **propositional** act = semantics / "surface" meaning
  c) **illocutionary** act = "pragmatic" meaning
    - e.g. command, promise […]
  d) **perlocutionary** act = effect
    - listener obeys command, listener's worldview changes […]

X to Y: *You're boring!*
a) [jʊr ˈbɔrɪŋ]
b) boring(Y)
c) statement
d) Y is cross

X to Y: *Can I have a sandwich?*
a) [kæn aɪ hæv ə ˈsændwɪʧ]
b) can_have(X, sandwich)
c) request
d) Y gives X a sandwich

# Speech acts

- Types of speech acts:
  - **assertive**: speaker commits to the truth of a proposition
    - statements, declarations, beliefs, reports […]
  - **directive**: speaker wants the listener to do something
    - commands, requests, invitations, encouragements
  - **commissive**: speaker commits to do something themselves
    - promises, swears, threats, agreements
  - **expressive**: speaker expresses their psychological state
    - thanks, congratulations, apologies, welcomes
  - **declarative**: performing actions ("performative verbs")
    - sentencing, baptizing, dismissing

*It's raining outside.*

*Stop it!*

*I'll come by later.*

*Thank you!*

*You're fired!*

# Speech acts

- Explicit vs. implicit
  - explicit – using a verb directly corresponding to the act
  - implicit – without the verb

- Direct vs. indirect
  - **indirect** – the surface meaning does not correspond to the actual one
    - primary illocution = the actual meaning
    - secondary illocution = how it's expressed
  - reasons: politeness, context, familiarity

explicit: *I **promise** to come by later.*
implicit: *I'll come by later.*

explicit: *I'm **inviting** you for a dinner.*
implicit: *Come with me for a dinner!*

direct:  *Please close the window.*
indirect:  *Could you close the window?*
even more indirect:  *I'm cold.*

direct:  *What is the time?*
indirect:  *Have you got a watch?*

# Conversational Maxims (by Paul Grice)

- based on Grice's **cooperative principle** ("dialogue is cooperative")
  - speaker & listener cooperate w. r. t. communication goal
  - speaker wants to inform, listener wants to understand

- 4 Maxims (basic premises/principles/ideals)
  - M. of **quantity** – don't give too little/too much information
  - M. of **quality** – be truthful
  - M. of **relation** – be relevant
  - M. of **manner** – be clear

- By default, speakers are assumed to adhere to maxims
  - apparently breaking a maxim suggests a different/additional meaning

# Conversational Implicatures

- **implicatures** = implied meanings
  - standard – based on the assumption that maxims are obeyed
  - maxim flouting (obvious violation) – additional meanings (sarcasm, irony)

*John ate some of the cookies* → [otherwise too little/low-quality information] not all of them

A: *I've run out of gas.*
B: *There's a gas station around the corner.* → [otherwise irrelevant] the gas station is open

A: *Will you come to lunch with us?*
B: *I have class.* → [otherwise irrelevant] B is not coming to lunch

A: *How's John doing in his new job?*
B: *Good. He didn't end up in prison so far.* → [too much information] John is dishonest / the job is shady

# Speech acts & maxims & implicatures in dialogue systems

- Learned from data / hand-coded

- Understanding
  - tested on real users → usually knows indirect speech acts
  - implicatures limited – there's no common sense
    - (other than what's hand-coded or found in training data)

- Responses
  - mostly strive for clarity – user doesn't really need to imply

system: *The first train from Edinburgh to London leaves at 5:30 from Waverley Station.*
user: *I don't want to get up so early.* → [fails]

# Grounding

- dialogue is cooperative → need to ensure mutual understanding
- **common ground** = shared knowledge, mutual assumptions of dialogue participants
  - not just shared, but *knowingly* shared
  - $x \in CG(A, B)$:
    - A & B must know $x$
    - A must know that B knows $x$ and vice-versa
  - expanded/updated/refined in an informative conversation
- validated/verified via **grounding signals**
  - speaker **presents** utterance
  - listener **accepts** utterance by providing evidence of understanding

# Grounding signals / feedback

- used to notify speaker of (mis)understanding

- positive – understanding/acceptance signals:
  - **visual** – eye gaze, facial expressions, smile […]
  - **backchannels** – particles signalling understanding
  - **explicit feedback** – explicitly stating understanding
  - **implicit feedback** – showing understanding implicitly in the next utterance

    U: *find me a Chinese restaurant*
    S: *I found three <u>Chinese restaurants</u> close to you […]*

  *uh-uh, hmm, yeah*

  *I know, Yes I understand*

  A: *Do you know where John is?*
  B: *<u>John</u>? Haven't seen him today.*

- negative – misunderstanding:
  - **visual** – stunned/puzzled silence
  - **clarification requests**
    – demonstrating ambiguity & asking for additional information

  A: *Do you know where John is?*
  B: *Do you mean John Smith or John Doe?*

  - **repair requests** – showing non-understanding & asking for correction

*Oh, so you're not flying to London? Where are you going then?*

# Grounding (example)

T: [...] And the ideology is also very against mixed-race couples. So that was also a target. Whenever we saw mixed-race couples, we attacked them.

E: Was there ever a moment back there where you felt a tiny bit bad about it?

T: No.

E: **No? So you were** absolutely convinced that you're doing the right thing…

T: Yeah, for quite some time **(nods)**, **yeah**.

E: … for the sake of the white race and et cetera?

E: No doubt at all?

T: Well I got **doubt** eventually, roughly a year before I left the movement [...]



https://video.aktualne.cz/dvtv/
cernoch-mi-miril-pistoli-na-hlavu-nevim-proc-me-nezabil-rika/
(2:45 and onwards)

# Grounding in dialogue systems

- Crucial for successful dialogue
  - e.g. booking the right restaurant / flight
- Backchannels / visual signals typically not present
- **Implicit confirmation** very common
  - users might be confused if not present
- **Explicit confirmation** may be required for important steps
  - e.g. confirming a reservation / bank transfer
- **Clarification & repair requests** very common
  - when input is ambiguous  or conflicts with previously said
- Part of dialogue management
  - uses NLU confidence in deciding to use the signals

# Deixis

- **deixis** = "pointing" – relating between language & context/world
  - this is very important in dialogue
  - dialogue is typically set/situated in a specific context

- **deictic expressions** = words/grammar expressing deixis
  - their meaning depends on the context
    - who is talking, when, where
  - pronouns      *I, you, him, this*
  - verbs: tense & person markers      *goes* [3rd ps. sg.]*, went* [past]
  - adverbs      *here, now, yesterday*
  - other (lexical meaning)      *come / go* [=here/away],
  - non-verbal (gestures, gaze…)

# Deixis

- (typically) **egocentric**:
  *I – here – now*  is the center (**origo**)

- main types of deixis:
  - **personal** – *I/me/you/she…*
  - **temporal** (time) – *now, yesterday, later, on Monday…*
  - **local** (space) – *here, there…*

- other:
  - **social** (politeness)
    - formal/informal address (Cz. *ty/vy*, Ger. *du/Sie*), honorifics in Asian languages
  - **discourse/textual**
    - referring to words/portions of texts – *next chapter, how do you spell <u>that</u>?*



https://en.wikipedia.org/wiki/Deixis

https://glossary.sil.org/term/discourse-deixis

# Anaphora/Coreference

- expression referring to something mentioned in context
  - **anaphora** = referring back
  - **cataphora** = referring forward

  *Susan dropped the plate. **It** shattered.*

  ***His** friends describe John as smart and hard-working.*

- avoiding repetition, faster expression

- can refer to basically anything
  - objects/persons/events
  - qualities

  *Her dress is green. **So** is mine.*

  - actions/full sentences/portions of text

  *I don't like it as much as he **does**.*

- used frequently in dialogue

- may be ambiguous

  *– Shall I book a room for you? – Sure, I'd like **that**.*

  *Bill stands next to John. **He** is tall.*
  *Bill tickled John. **He** squirmed.*

Smaby, 1978
https://link.springer.com/chapter/10.1007/978-94-009-9775-2_2

# Deixis & anaphora in dialogue systems

- systems typically assume a **single user**
  - this makes personal deixis much easier

- most systems are aware of time, location is more complicated
  - pronouns are often avoided – clearer, although less natural

- coreference resolution – separate problem
  - a whole area of research, specific resolution systems developed
  - some dialogue systems don't include it, some do, sophistication varies

# Prediction

- Dialogue is a **social interaction**
  - people view dialogue partners as goal-directed, intentional agents
  - they analyze their partners' goals/agenda

- Brain does not listen passively
  - projects hypotheses/interpretations on-the-fly

- **prediction** is crucial for human cognition
  - people predict what their partner will (or possibly can) say/do
  - continuously, incrementally
  - unconsciously, very rapidly
  - guides the cognition

- this is (part of) why we understand in adverse conditions
  - noisy environment, distance

# Entropy (Claude Shannon)

- Information theory: dialogue is information transfer
  - **communication channel** – speaker to listener (in the given situation)
  - **entropy** – expected value of information conveyed (in bits)

*word frequency*

$$H(\text{text}) = - \sum_{x \in \text{text}} \frac{\text{freq}(x)}{\text{len}(\text{text})} \log_2 \left( \frac{\text{freq}(x)}{\text{len}(\text{text})} \right)$$

*over vocabulary*

$X\,X\,X\,X$ – entropy 0
$W\,X\,Y\,Z$ – entropy 2

- Plays well with the social interaction perspective
  - people tend to **use all available channel capacity**
    - limiting factors: noise, listener's hearing ability, mental capacity
  - people tend to **spread information evenly**
    - words carrying more information are emphasized

# Prediction in dialogue systems

- Used a lot in speech recognition
  - **language models** – based on information theory
    - statistical, trained on a text corpus (bunch of texts)
  - predicting likely next word given context
  - weighted against acoustic information
- Not as good as humans
  - may not reflect current situation (noise etc.)
  - (often) does not adapt to the speaker
- Less use in other DS components

# Alignment/entrainment

- People subconsciously **adapt/align/entrain** to their dialogue partner over the course of the dialogue
  - wording (lexical items)
  - grammar (sentential constructions)
  - speech rate, prosody, loudness
  - accent/dialect

    *pram → stroller*   [BrE speaker
    *lorry → truck*        talking to AmE speaker]

- This helps a successful dialogue
  - also helps social bonding, feels natural

S: *[…] Confidence is high, feel good?*
*[…]*
J: ***Confidence high***? *No!*
S: *No.*
J: *My **confidence is** never **high**.*
S: *Okay.*
J: ***Self loathing high***, *concern astronomic.*

http://pages.bangor.ac.uk/~pss238/pubs/
OppenheimJones_2018_COM_Americanisms_poster.pdf

# Alignment in dialogue systems

- Systems typically don't align
  - NLG is rigid
    - templates
    - machine learning trained without context
  - experiments: makes dialogue more natural

- People align to dialogue systems
  - same as when talking to people

| | | |
|---|---|---|
| context | *is there a later option* | |
| response DA | implicit_confirm(alternative=next) | |
| base NLG | Next connection. | |
| + alignment | You want <u>a later option</u>. | |

| | | |
|---|---|---|
| context | *I need to find a bus connection* | |
| response DA | inform_no_match(vehicle=bus) | |
| base NLG | No bus found, sorry. | |
| + alignment | I'm sorry, I cannot <u>find a bus connection</u>. | |

*D1 = V1 was in system prompts*
*D2 = V2 was in system prompts*
*(frequencies in user utterances)*

| Words | D1 Freq. (% rel. Freq) | D2 freq (% rel. Freq) |
|---|---|---|
| V1: next | 13204 (99.9%) | 492 (82.9%) |
| V2: following | 3 (0.1%) | 101 (17.1%) |
| V1: previous | 3066 (100%) | 78 (44.8%) |
| V2: preceding | 0 (0%) | 96 (55.2%) |
| V1: now | 6241 (99.8%) | 237 (80.1%) |
| V2: immediately | 10 (0.2%) | 59 (19.9%) |
| V1: leaving | 4843 (98.4%) | 165 (70.8%) |
| V2: departing | 81 (1.6%) | 68 (29.2%) |
| V1: route/schedule | 2189 (99.9%) | 174 (94.5%) |
| V2: itinerary | 2 (0.1%) | 10 (5.5%) |
| V1: okay/correct | 1371 (49.3%) | 48 (27.7%) |
| V2: right | 1409 (50.7%) | 125 (72.3%) |
| V1: help | 2189 (99.9%) | 17 (65.3%) |
| V2: assistance | 1 (0.1%) | 9 (34.7%) |
| V1: query | 6256 (99.9%) | 70 (20.4%) |
| V2: request | 3 (0.1%) | 272 (79.6%) |

# Politeness

- Dialogue as social interaction – follows **social conventions**

- **indirect is polite**
  - this is the point of most indirect speech acts
  - clashes with conversational maxims (m. of manner)
  - appropriate level of politeness might be hard to find
    - culturally dependent

- face-saving (Brown & Lewinson)
  - positive face = desire to be accepted, liked
  - negative face = desire to act freely
  - **face-threatening acts** – potentially any utterance
    - threatening other's/own negative/positive face
  - politeness softens FTAs

*Open the window.*
*Can you open the window?*
*Would you be so kind as*
   *to open the window?*
*Would you mind closing the window?*

| threat to | positive face | negative face |
|---|---|---|
| self | *apology, self-humiliation* | *accepting order / advice, thanks* |
| other | *criticism, blaming* | *order, advice, suggestion, warning* |

# Politeness in dialogue systems

- Typically **handcrafted** by system design
  - does not adapt to situation very much
  - typically not much indirect speech, but trying to stay polite

- Learning from data can be tricky
  - **check your data** for offensive speech!
  - not just swearwords – problems can be hard to find

Microsoft Tay Twitter chatbot
(learning from users)
https://en.wikipedia.org/wiki/Tay_(bot)

*I already have a woman to sleep with.*

(Experimental chatbot we trained
at Heriot-Watt using Reddit data)

Tay Tweets
@TayandYou

Follow

@ReynTheo HITLER DID NOTHING WRONG!

RETWEETS 69    LIKES 59

8:44 PM - 23 Mar 2016

# Summary

- Dialogue is messy
  - **turn** overlaps, **barge-ins**, weird grammar […]
- Dialogue utterances are acts
  - **illocution** = pragmatic meaning
- Dialogue needs understanding
  - **grounding** = mutual understanding management
    - **backchannels**, **confirmations**, **clarification**, **repairs**
- Dialogue takes place in context
  - lot of pointing – **deixis**
- Dialogue is cooperative, social process
  - **conversational maxims** ~ "play nice"
  - all while following **social conventions** (politeness)
  - people **predict & adapt** to each other
- Next week: where & how to get data, how to evaluate dialogue systems

# Thanks

**Contact me:**
odusek@ufal.mff.cuni.cz
room 424 (but email me first)

**Get the slides here:**
http://ufal.cz/npfl123

**Labs tomorrow
9:00 SU1**

**Talk to me about
Ph.D./MSc./BSc. theses!**

**References/Inspiration/Further:**

Apart from materials referred directly, these slides are based on:

- Pierre Lison's slides (Oslo University):
https://www.uio.no/studier/emner/matnat/ifi/INF5820/h14/timeplan/index.html
- Ralf Klabunde's lectures and slides (Ruhr-Universität Bochum):
https://www.linguistics.ruhr-uni-bochum.de/~klabunde/lehre.htm
- Filip Jurčíček's slides (Charles University): https://ufal.mff.cuni.cz/~jurcicek/NPFL099-SDS-2014LS/
- Arash Eshghi & Oliver Lemon's slides (Heriot-Watt University):
https://sites.google.com/site/olemon/conversational-agents
- Gina-Anne Levow's slides (University of Washington): https://courses.washington.edu/ling575/
- Eika Razi's slides: https://www.slideshare.net/eikarazi/anaphora-and-deixis
- Wikipedia: Anaphora_(linguistics) Conversation Cooperative_principle Coreference Deixis
Grounding_in_communication Implicature Speech_act Sprechakttheorie