Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Interaktive systemer

Pierre Lison
plison@nr.no

IN2110: Språkteknologiske metoder

27. april 2022

Oblig 2b

- 2 deler:
  - **MT**: Oversettelseseksperimenter med filmtekstinger fra Ringenes Herre (tysk → engelsk)
  - **Interaktive systemer**: utvikling av en veldig enkel, datadrevet praterobot (trent på filmtekstinger)

- Oppgaven skal publiseres i de neste dagene
- Innleveringsfrist: **16. mai**

Plan for i dag

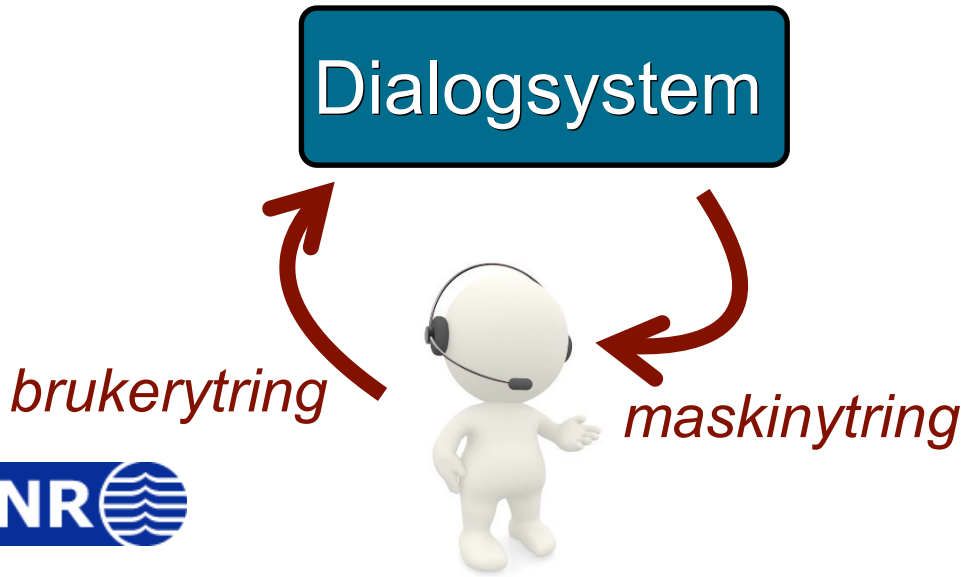
- Introduksjon
- Hva er en samtale?
- Prateroboter
  - Regelbaserte systemer
  - IR og seq2seq modeller
  - Intentgjenkjenning
- Dialogstyring
- Etiske betrakninger

Plan for i dag

- **Introduksjon**
- Hva er en samtale?
- Prateroboter
  - Regelbaserte systemer
  - IR og seq2seq modeller
  - Intentgjenkjenning
- Dialogstyring
- Etiske betrakninger

Dialogsystemer?

= dataprogrammer ("bots") utviklet til å samhandle med mennesker ved hjelp av (talte eller tekstbaserte) naturlige språk

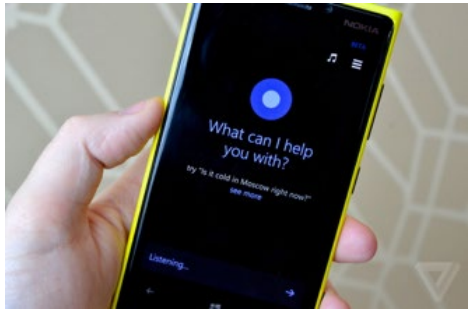


Hvorfor?

- *Brukervennlig*: du trenger bare å kunne snake/skrive!
- Kan utrykke *komplekse forespørsler*
- For talebaserte systemer: *berøringsfri* grensesnitt (viktig for f.eks. bilkjøring)

Anvendelser

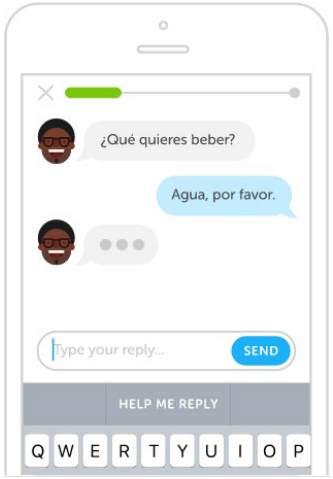
Virtuelle assistenter (Siri, Cortana, osv.)



Smarthusprodukter

Stemmebasert navigering og styring i bilen

Lærings platformer

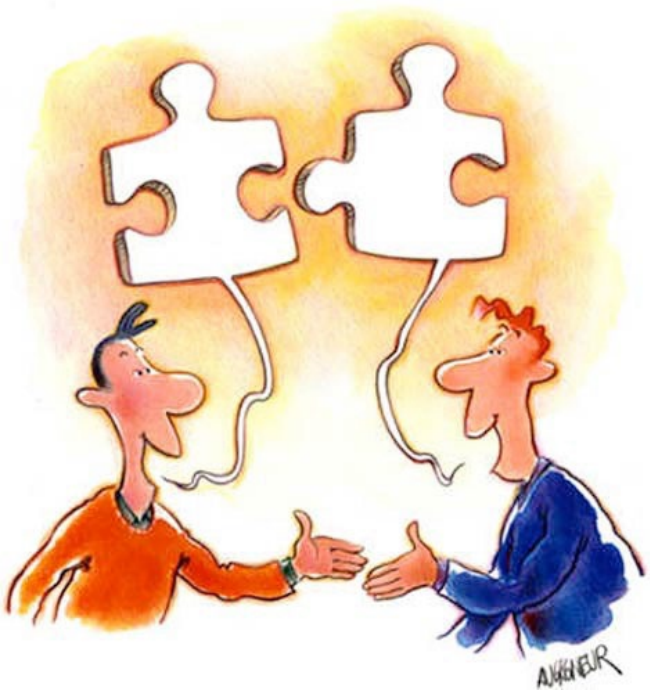


Prateroboter

Tjenesteroboter

Plan for i dag

- Introduksjon
- **Hva er en samtale?**
- Prateroboter
  - Regelbaserte systemer
  - IR og seq2seq modeller
  - Intentgjenkjenning
- Dialogstyring
- Etiske betrakninger



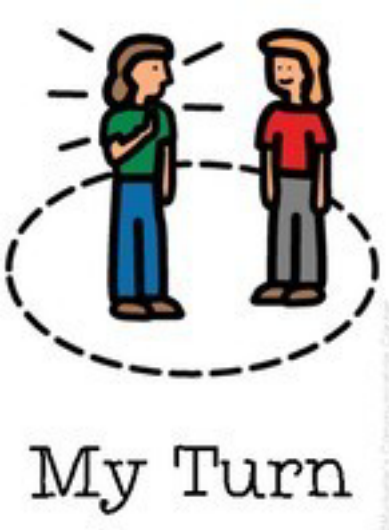
Turtaking (turn-taking)

- Deltakerne i dialog tar *turer*:
  - En tur = et kontinuerlig bidrag fra en samtalepartner
- Turtaking er en beskrivelse av hvordan ordet fordeles mellom samtalepartnere
- Overraskende flytende i dagligdagssamtaler
  - Forsøk på å minimere både tomrom (ingen som snakker) og overlapp (flere som snakker over hverandre)
- Intervallet mellom turer er rundt 250 ms

[Duncan (1972): «Some Signals and Rules for Taking Speaking Turns in Conversations», in *Journal of Personality and Social Psychology*]

Turtaking

- Hvordan er turer tildelt (tatt/frigitt)?
- Ulike markører:
  - Komplett syntaktisk/semantisk enhet?
  - Dialogstruktur (hilsen → hilsen, spørsmål → svar osv.)
  - Intonasjon (fallende intonasjon signaliserer ofte at turen er ferdig)
  - Ikke-verbale signaler som blikk, gester
  - Markører for stillhet og nøling som "eh" (utfylte pauser ≠ fylte pauser)
  - Sosiale konvensjoner



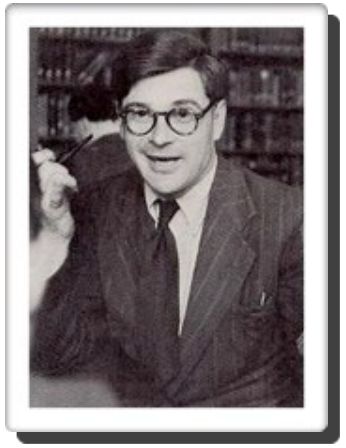
Eksempel av turtaking

Person 1:	han vil bo i skogen ?
Person 2:	# altså hvis jeg hadde kommet og sagt " skal vi flytte i skogen ? " så hadde han sagt ja
Person 1:	mm
Person 2:	men jeg vil ikke bo i skogen
Person 1:	nei det skjønner jeg
Person 2:	så vi må jo finne et sted som er mellomting og det jeg vil ikke bo utpå landet # i hvilken som helst (uforståelig) ...
Person 1:	* men det kommer jo an på hvor i skogen da

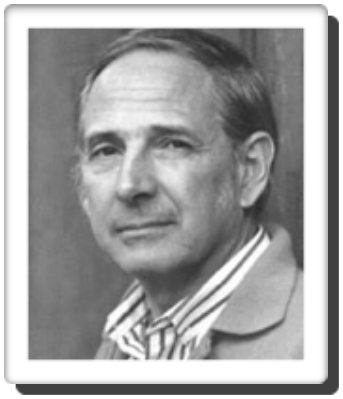
[«*Norske talespråkskorpus - Oslo delen*» (NoTa), samlet og annotert av Tekstlaboratoriet]

Språkhandlinger

- Hver ytring er en **handling** utført av taleren
- Taleren har et spesifikt mål (som kanskje "bare" er å etablere eller opprettholde et sosial kontakt med samtalepartnere)
- Ytringer leder til bestemte *virkninger* på samtalepartnere (eller på verden forøvrig)
- «Språk som handling» -perspektiv



J.L. Austin (1911-1960)
språkfilosof



J. Searle (1932, -)
språkfilosof

[J. L. Austin (1955), *How to do things with words*.]

Språkhandlinger

- Morens reaksjonen har et bestemt *formål,* nemlig å formidle hennes forbauselse/sinne og stoppe Calvin
- Spørsmålet hennes leder til bestemte virkninger:
  - En psykologisk reaksjon fra Calvin (overraskelse)
  - + en virkning på selve verdenen (Calvin stopper)

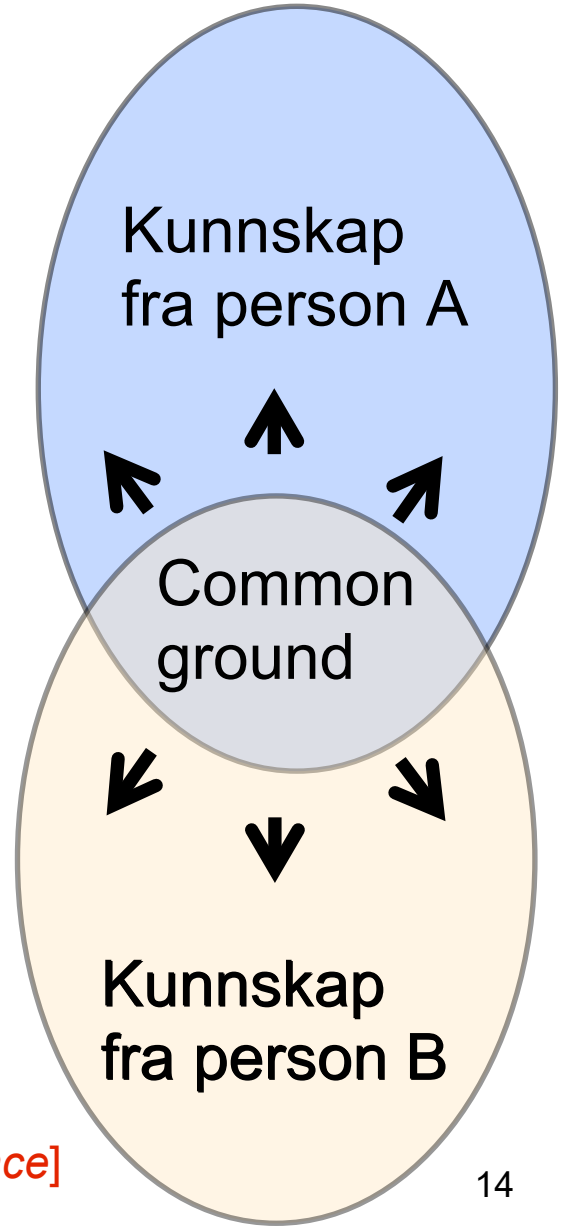
Searles taksonomi

- **Assertives** : committing the speaker to the truth of a proposition. E.g.: *«Eksamen vil finne sted 3. juni»*
- **Directives**: attempts by the speaker to get the addressee to do something. E.g. : *«Kan du rydde opp soverommet ditt?»*
- **Commissives**: committing the speaker to some future course of action. E.g.: *«Jeg lover deg å rydde opp rommet mitt»*.
- **Expressives**: expressing the psychological state of the speaker. E.g.: *«Takk for at du ryddet opp rommet ditt».*
- **Declaratives**: bringing about a different state of the world by the utterance. E.g.: *«Du er sparket.».*

For norske termer: https://ndla.no/subject:1:b9e86c43-93b8-49e9-81af-09dbc7d79401/topic:2:193544/topic:2:82776/resource:1:82785

Grounding

- Dialog er et samarbeid mellom deltakerne i samtalen
  - Trenger å sikre gjensidig forståelse – "vi er på samme linje"
- Gradvis utvidelse og berikelse av vår "common ground"
- "common ground" = kunnskap som er felles for samtaledeltakerne

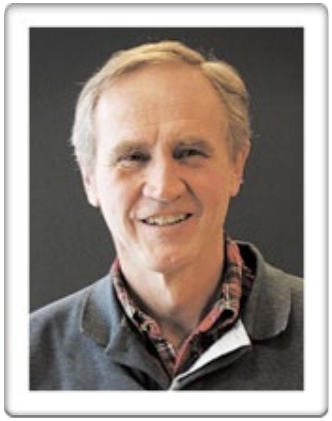


[H. H. Clark and E. F. Schaefer (1989), «Contributing to discourse», in *Cognitive Science*]

Grounding

- Grounding betegner prosessen av å gradvis utvide vår "common ground" i løpet av samtalen
  - Bred register av signaler og strategier
- Flere nivåer
  - Kontakt (oppmerksomhet overfor partneren)
  - Oppfatning av ytringen
  - Forståelse av ytringen
  - Holdninger (enighet, empati osv.)

Herbert H. Clark
psykolingvist



Jens Allwood
lingvist

[Jens Allwood (1992), «On discourse cohesion», in *Gothenburg papers in Theoretical Linguistics*.]

Grounding actions

- Backchannels: «*uh-uh*», «*mm*», «*ja*»
- Eksplisitte tilbakemeldinger: «*ja det skjønner jeg*»
- Implisitte tilbakemeldinger: A: «*Jeg ønsker å fly til Roma*» → B: «*På onsdag er det to flyvninger til Roma: ... »*
- Klarifiseringsstrategier: «*Mente du til Roma eller til Goa?*", "*Kan du bekrefte at …* "
- Fiksingsstrategier: «*OK, dere ønsker altså ikke å fly til Goa. Hvor ønsker dere å fly da?*"

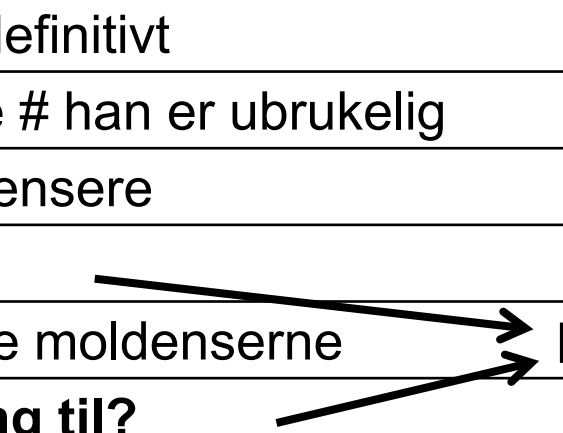
Eksempel av grounding

Person 1:	vi vasker den hver dag vi # vi har mopp
Person 2:	**mm** ## **ja** det er fort og faren til M27 legger nytt teppe han # det er gjort på to timer ## så det er fort gjort
Person 1:	**ja** ## da er ikke noe sak
Person 2:	vi har skifta teppe tre ganger allerede han gjør det gratis
Person 1:	**hæ ?**
Person 2:	vi har skifta teppe tre ganger og # han han ...
Person 1:	* **jeg skjønner ikke** hvorfor dere har teppe
Person 2:	jeg syns det var rart jeg **òg** # men e # (sibilant)

[«*Norske talespråkskorpus - Oslo delen*» (NoTa), samlet og annotert av Tekstlaboratoriet]

Eksempel av grounding (2)

Person 1:	e # **nei** det er ikke mange
Person 2:	**ja * nei**
Person 1:	men heldigvis så var ikke Petter Rudi tatt ut denne gangen da
Person 2:	**ja** # jeg skjønner ikke hva han skal på landslaget å gjøre
Person 1:	*** nei han har ingen ting på landslaget**
Person 2:	**nei** # definitivt
Person 1:	å gjøre # han er ubrukelig
Person 2:	* moldensere
Person 1:	**hm?**
Person 2:	ja disse moldenserne
Person 1:	**en gang til?**
Person 2:	disse moldenserne
Person 1:	* **å ja** (fremre klikkelyd) # unnskyld # jeg hørte ikke hva du sa

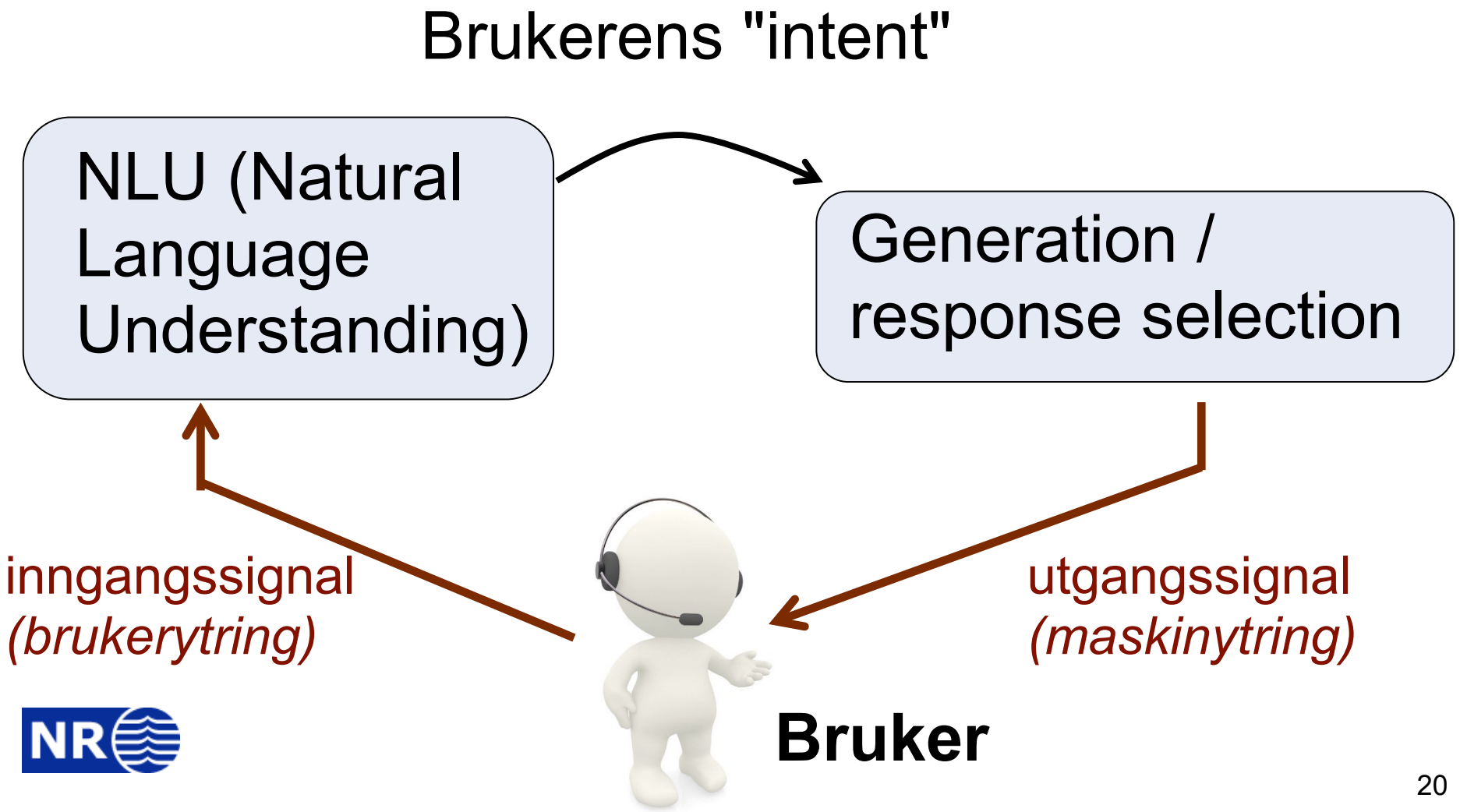


[«*Norske talespråkskorpus - Oslo delen*» (NoTa), samlet og annotert av Tekstlaboratoriet]

Plan for i dag

- Introduksjon
- Hva er en samtale?
- **Prateroboter**
  - **Regelbaserte systemer**
  - **IR og seq2seq modeller**
  - **Intentgjenkjenning**
- Dialogstyring
- Etiske betrakninger

Grunnarkitektur



Regelbaserte systemer

- Pattern-action rule:

```
(0 YOU 0 ME)      [pattern]
→
(WHAT MAKES YOU THINK I 3 YOU) [transform]
```

- For eksempel:

```
You hate me
WHAT MAKES YOU THINK I HATE YOU
```

[eksempel fra D. Jurafsky]

IR modeller

- Alternativ: bruke en datadrevet tilnærming og svare brukeren basert på et *dialogkorpus*
- Prosedyren:
  - Gitt en brukerinngang **q**, finn ytringen **t** i dialogkorpuset som ligner mest på **q**
  - Så returner som svar enten (a) selve **t** eller (b) ytringen som følger **t** i korpuset

IR modeller

$$r = response\left(\underset{t \in C}{\text{argmax}} \frac{q^T t}{||q||t||}\right)$$

Hvordan kan vi bestemme hvilken ytring **t** fra korpuset «ligner mest» på inputytringen **q**?

- Cosine similarity mellom vektorer
- Vektorene kan være TF-IDF vektorer (hvor hver ytring tilsvarer et eget "dokument")
- Eller embeddings beregnet fra et nevralt nettverk

Eksempel

Korpuset:

1. hei ! →
2. hei ! har du det bra ? →
3. ja , hva med deg ? →
4. bare bra →
5. har du spist ? →
6. ja →

TF vektorer:

bare	bra	deg	det	du	ja	har	hei	hva	med	spist	,	!	?
							1					1	
	1		1	1		1	1					1	1
		1			1			1	1		1		1
1	1												
				1		1				1			
					1								

Eksempel

$\log(6) \approx 0.78$

$\log\left(\frac{6}{2}\right) \approx 0.48$

Korpuset:

1. hei !
2. hei ! har du det bra ?
3. ja , hva med deg ?
4. bare bra
5. har du spist ?
6. ja

TF-IDF vektorer:

	bare	bra	deg	det	du	ja	har	hei	hva	med	spist	,	!	?
1. hei !								.48					.48	
2. hei ! har du det bra ?		.48		.78	.48		.48	.48					.48	.48
3. ja , hva med deg ?			.78			.48			.78	.78		.78		.48
4. bare bra	.78	.48												
5. har du spist ?					.48		.48				.78			
6. ja						.48								

Ny brukerinput *q*: "går det bra med deg?"

TF-IDF vektor:

	.48	.78	.78						.78			.48

Eksempel

	bare	bra	deg	det	du	ja	har	hei	hva	med	spist	,	!	?
1.								.48					.48	
2.		.48		.78	.48		.48	.48					.48	.48
3.			.78			.48			.78	.78		.78		.48
4.	.78	.48												
5.					.48		.48				.78			
6.						.48								

$q^T t$	$\frac{q^T t}{\|q\|\|t\|}$
0	0
1.07	0.50
1.45	0.56
0.23	0.17
0	0
0	0

	.48	.78	.78						.78				.48

Eksempel

$$\frac{q^T t}{\|q\|\|t\|}$$

Korpuset:

1. hei ! → 0
2. hei ! har du det bra ? → 0.50
3. ja , hva med deg ? → 0.56
4. bare bra → 0.17
5. har du spist ? → 0
6. ja → 0

→ Det betyr at ytringen som ligner mest på *q* er ytringen 3: "ja, hva med deg?"

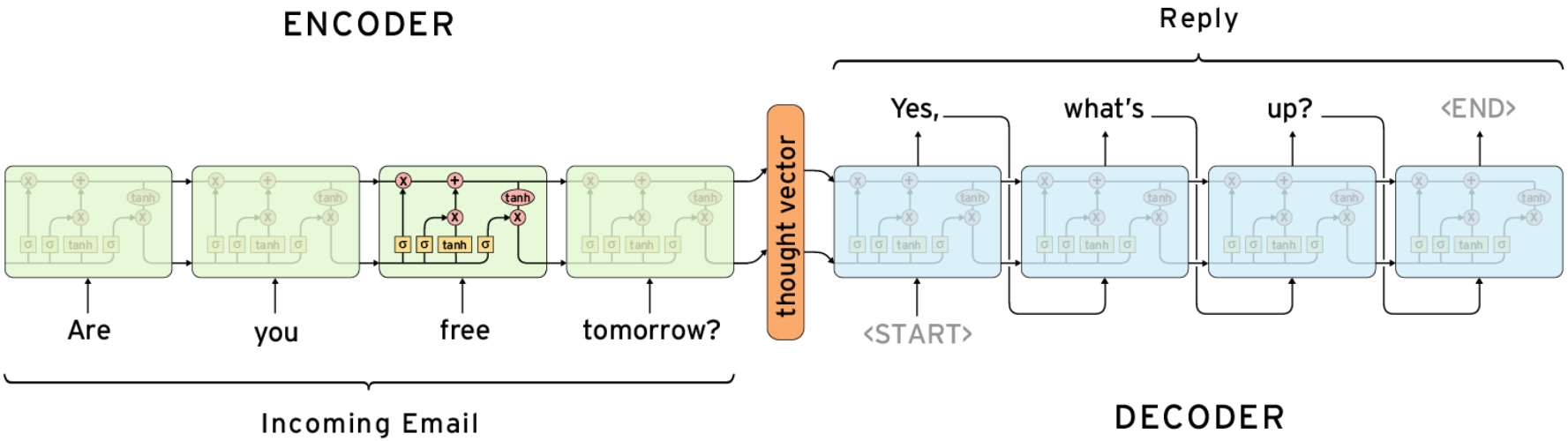
... og at systemet kan velge som respons ytringen som kommer rett etter 3

Ny brukerinput *q*: "går det bra med deg?"

→ **Respons fra systemet**: "bare bra"

Seq2seq modeller

- IR modeller kan kun velge ytringer fra korpuset
- Seq2seq modeller kan generere *nye* ytringer, ord for ord
  - Veldig lik modellene brukt for maskinoversetelse
  - Skritt 1: konvertere brukerytringen til en vektor (*encoding*)
  - Skritt 2: generere svaret ord for ord (*decoding*)

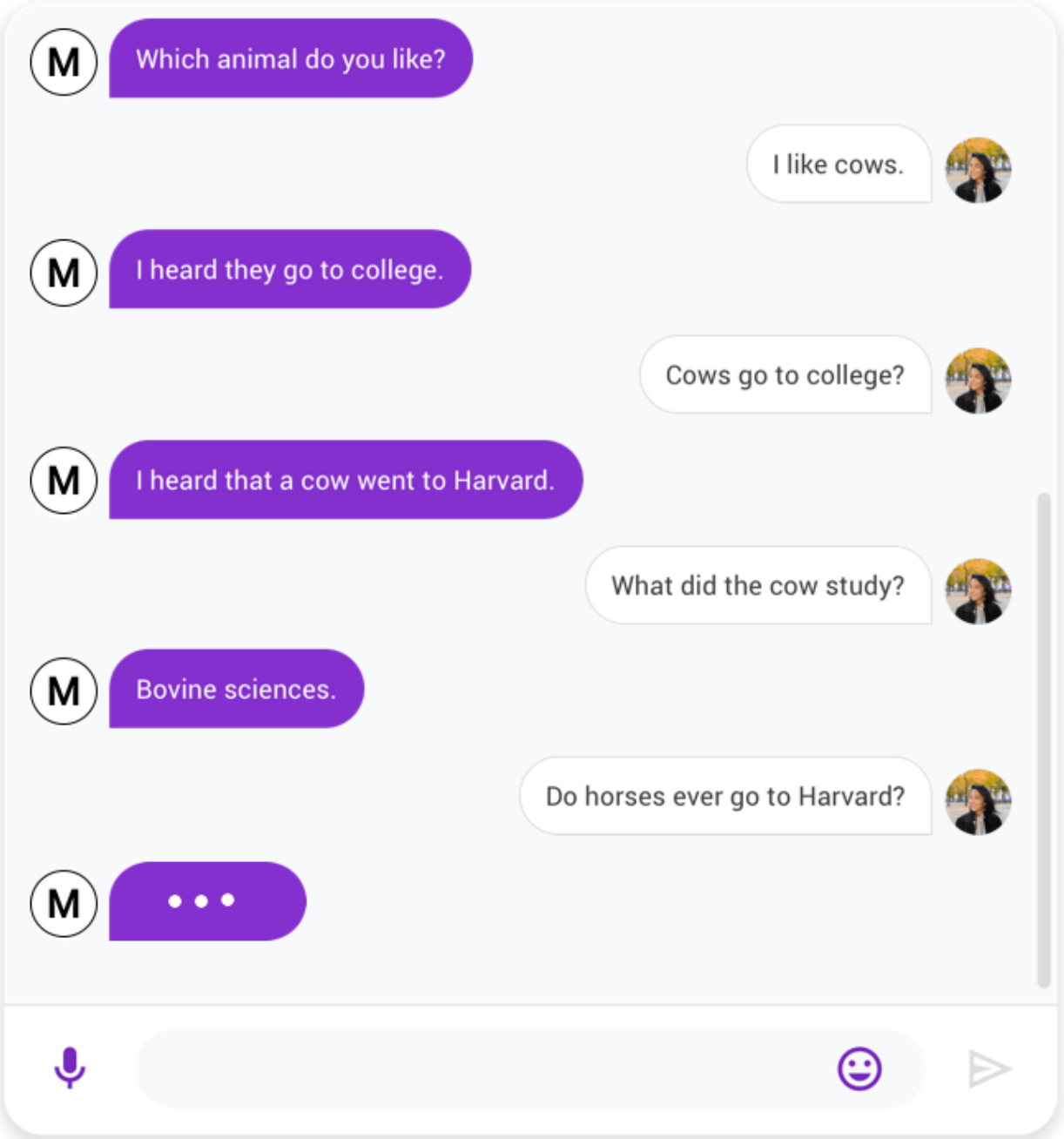


[Bildet fra Deep Learning for Chatbots: Part 1]

Eksempel fra "Meena" (Google)

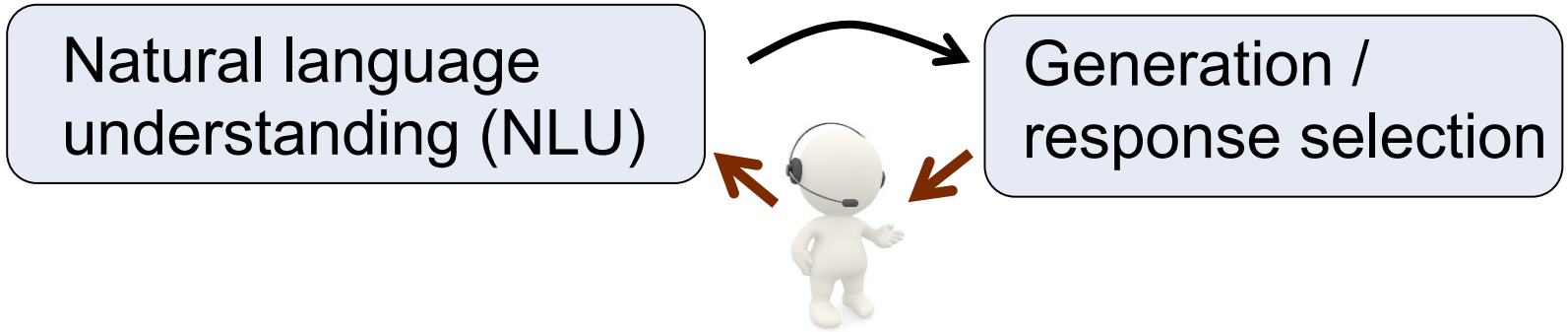
Seq2seq modeller: ←

+ Kan produsere helt «nye» svar
- Vanskelig å «styre» (hva kommer systemet til å svare, og hvorfor?)



https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html

Hybride arkitekturer



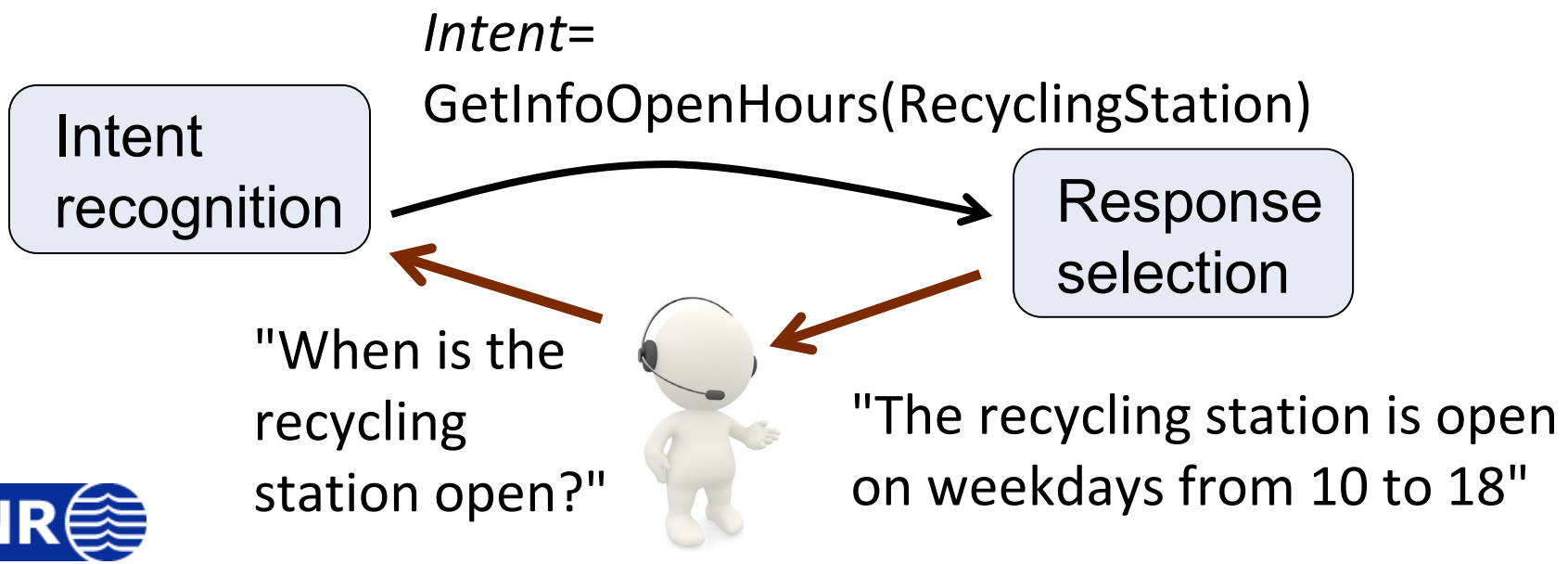
Mange prateroboter utviklet av/for virksomheter tar i bruk en hybrid arkitektur hvor:

- NLU-komponenten er trent fra markerte data (klassifiseringsproblem: koble brukerytringen til dets *intent*)
- Svarene er forhåndsskrevet av utvikleren

Intentgjenkjenning

Definert fra en liste forhåndsdefinerte kategorier (muligens med argumenter, ofte kalt slots)

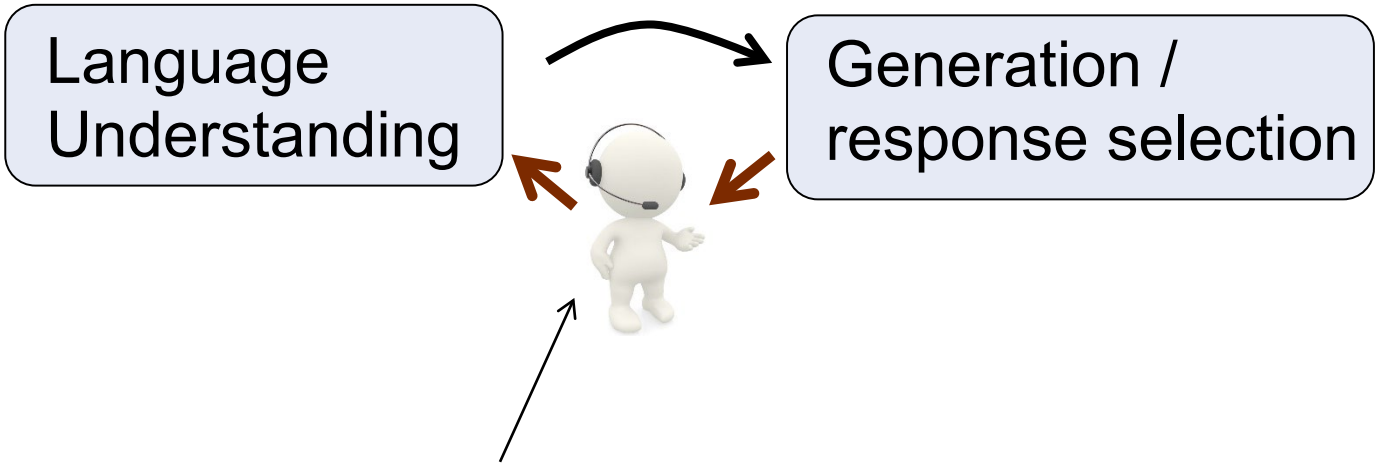
- *Input*: selve brukerytringen (+ kontekst)
- *Output*: intent (= hva brukeren forsøker å oppnå)



Plan for i dag

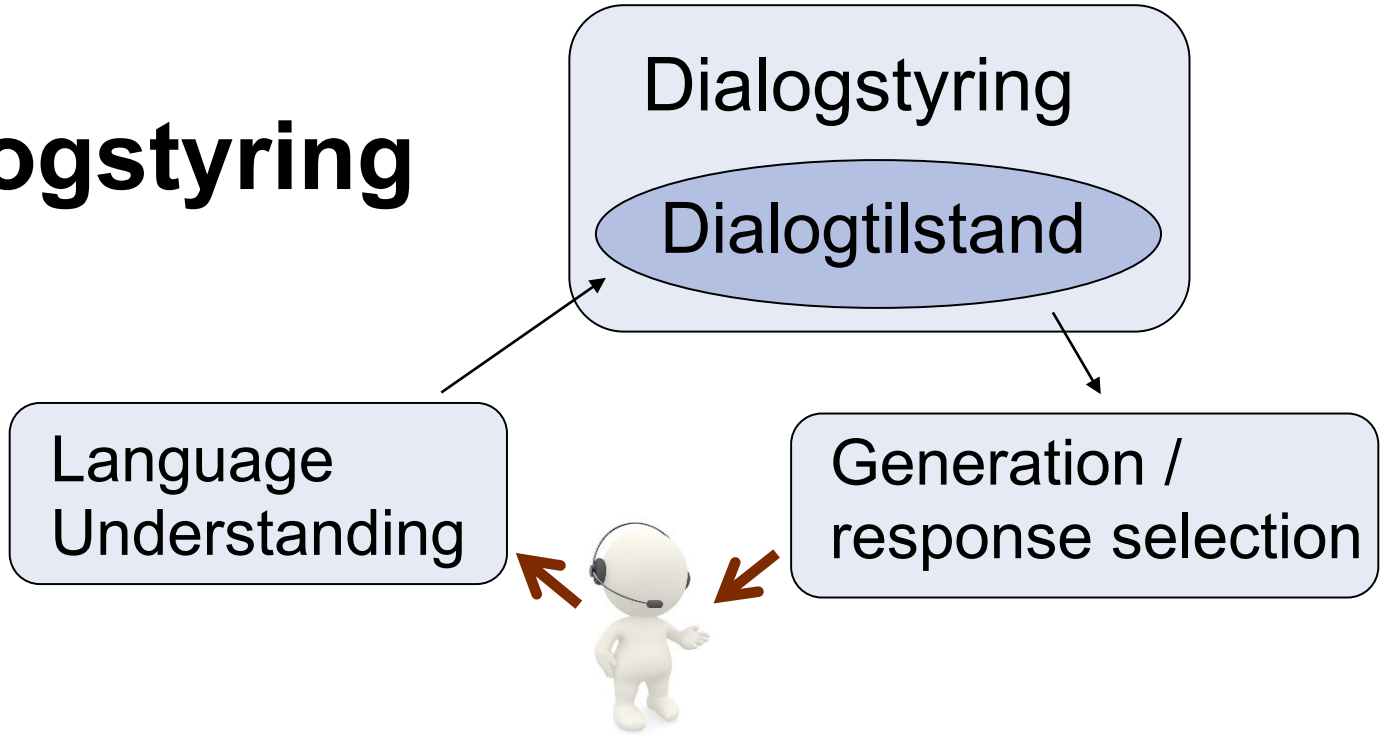
- Introduksjon
- Hva er dialog?
- Prateroboter
  - Regelbaserte systemer
  - IR og seq2seq modeller
  - Intentgjenkjenning
- **Dialogstyring**
- Etiske betrakninger

Dialogstyring



- **Begrensning av prateroboter**: ofte ingen/lite "minne" av samtalehistorikk (utover den gjeldende ytringen)
- Vanskelig å skalere til mer kompliserte interaksjoner, hvor samtalen kan utfolde seg på mange "turer"

Dialogstyring



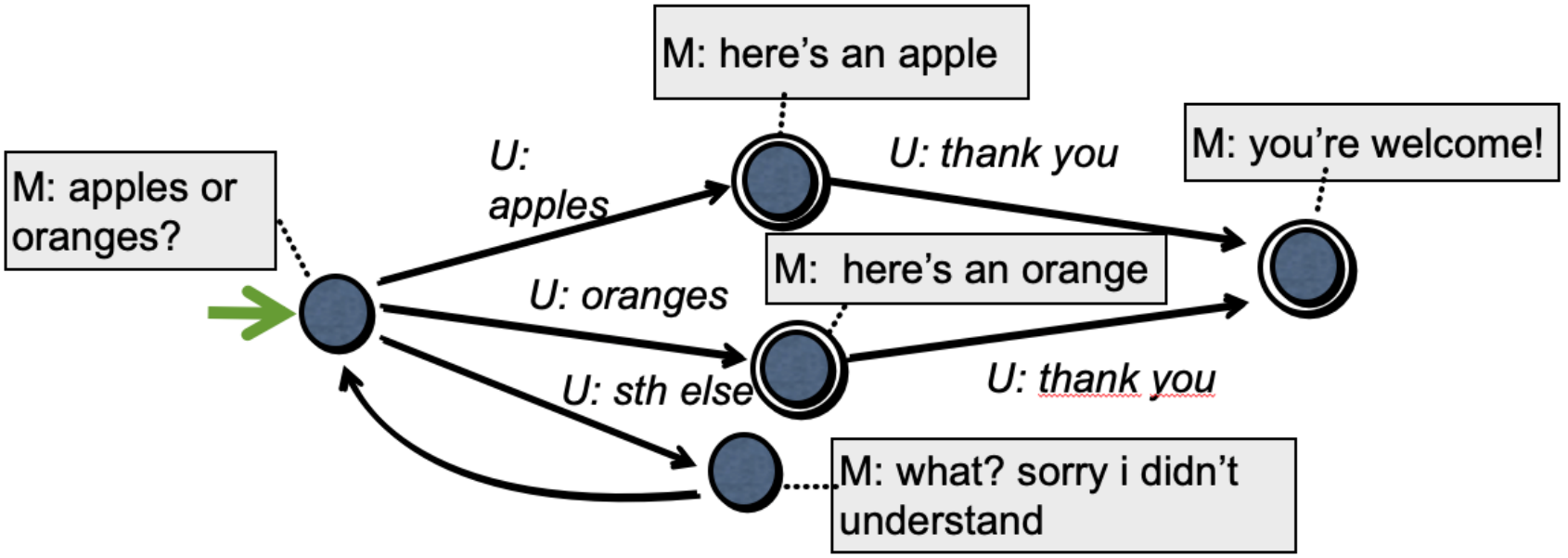
= Tar ansvar for *flyten* av samtalen over tid

- Styre turtakingen
- Sikre gjensidige forståelse ("grounding")
- Oppdatere representasjonen av *dialogtilstanden* jevnlig
- Finne ut hva systemet bør si eller gjøre på et vist tidspunkt

Endelige tilstandsmaskiner

Enkleste dialogstyringsmetode: *endelig tilstandsautomata*

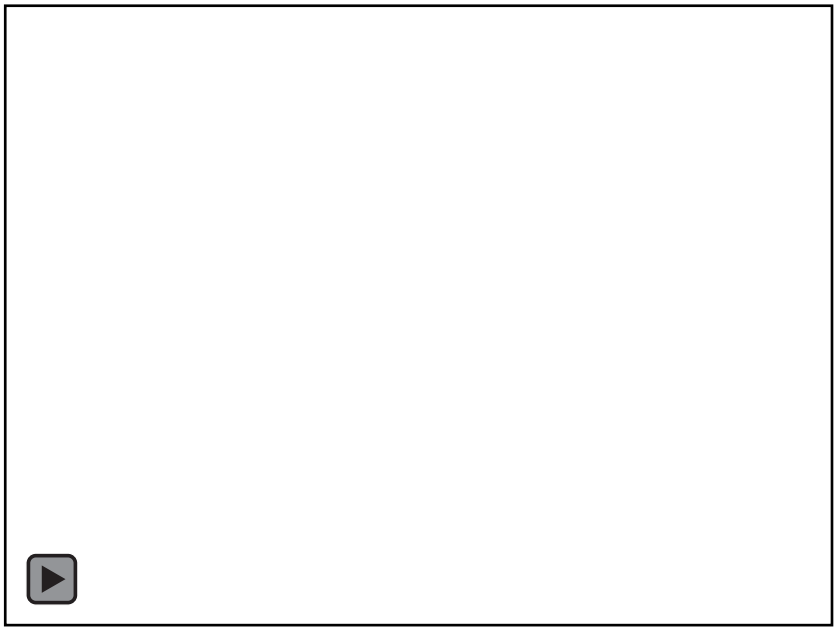
- Hver node representerer en spesifikk *dialogtilstand*, og er assosiert med en systemrespons
- Hver kant representer en mulig svar fra brukeren, og brukes til å bevege systemet fra tilstand til tilstand



Interaksjonsstyl

Begrensninger ved endelige tilstandsmaskiner og lignende metoder:

- Stiv, repetitiv (og ofte irriterende) adferd
- Ikke egnet til å handtere støy og usikkerhet
- Ingen bruker- eller domenetilpasning

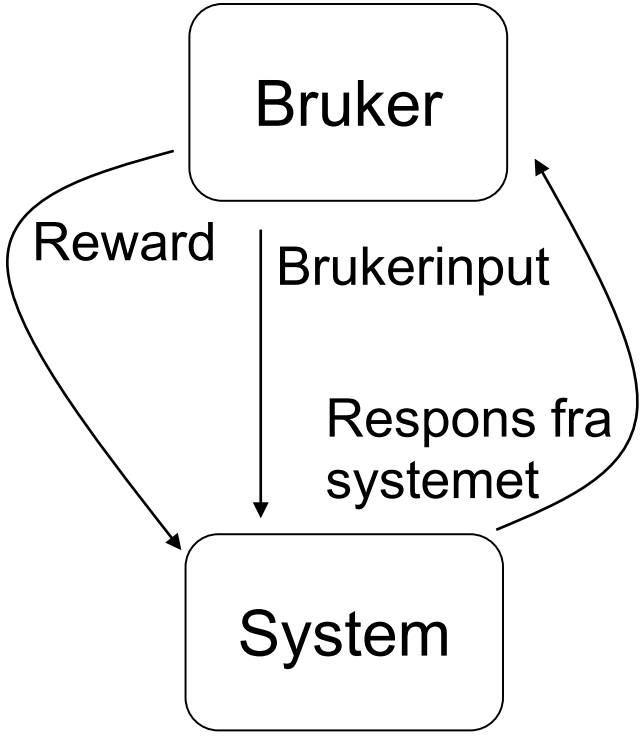


"Saturday night live" sketch comedy, 2005

- Begrenset til et mål ... men reelle interaksjoner er *avveininger* mellom ulike (konkurrerende) hensyn

Datadrevet dialogsstyring

- **Løsning**: automatisk optimere dialogstyring ut fra erfaringer samlet av systemet
  - Ofte basert på *reinforcement learning*
  - "Erfaringer" = samtaler med enten ekte eller simulerte brukere
  - Samtaler er belønnet med positive og negative tilbakemeldinger (*rewards*)
- Dessverre ikke tid til å beskrive slike modeller i dag!



Plan for i dag

- Introduksjon
- Hva er en samtale?
- Prateroboter
  - Regelbaserte systemer
  - IR og seq2seq modeller
  - Intentgjenkjenning
- Dialogstyring
- **Etiske betrakninger**

Etiske betrakninger

- Maskinlærings kan gjenskape (og noen ganger forsterke) skjevheter som oppstod i treningsdataene
  - Hatytringer, fordommer, krenkende språk, kjønnsskjevhet, osv.
- Hvordan beskytte brukerens personvern?
- Forsterker virtuelle assistenter kjønnsstereotyper?
- Hvordan unngå at prateroboter brukes til å bedra folk?

«Nei Siri, det kan du ikke si!»,
https://blogg.forskning.no/akademiet-for-yngre-forskere