

SPROGMUSEET

Redaktør: Ole Stig Andersen



Fra Wittgenstein til Ontology Engineering 1: Analytisk Sprogfilosofi og Moderne Kunstig Intelligens

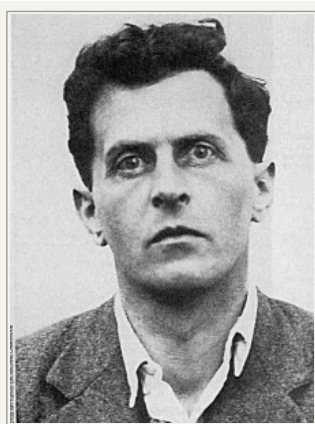
 Af Erik David Johnson 24. april 2012 • I kategorien [Sprogteknologi](#) •

Hvor langt er vi nået?

Hvis man kort gør status på den sproglige side af kunstig intelligens i dag, er der mange avancerede løsninger som imponerer ved deres tekniske fremskridt og praktiske anvendelighed. Computere kan selv oversætte tekst fra et sprog til et andet, tæve os i skak og Jeopardy og i visse tilfælde endda kommunikere på vores præmisser indenfor begrænsede emner, såsom at svare på spørgsmål om rejsetider og lignende over telefonen.

Alligevel hersker der blandt forskere i kunstig intelligens og sprogteknologi forholdsvis bred enighed om at målet om at udvikle kunstig intelligens der kan simulere en form for bevidsthed gennem kommunikation via menneskelige sprog, også kaldet 'natursprog', er stødt på en naturlig grænse. Eksempler på gnidningsfri kommunikation mellem mennesker og computere via natursprog lader simpelthen vente på sig. Eksempelvis burde en søgning på Google fungere som en afslappet samtale – ikke en programmeringsopgave i semantisk snedighed og forståelse af den bagvedliggende algoritmes underliggende funktionalitet.

Det er som om denne type forskning har nået et sidste plateau, hvorfra man ikke kan komme videre. Spørgsmålet er så om der kan findes en helt ny vej – en ny tilgang til den sprogteknologiske side af kunstig intelligens, og om den i så tilfælde vil kunne udvise potentiale for nye og mere overbevisende resultater end det vi kender til i dag. Vores rejse mod dette mål starter ved den anerkendte, østrigske sprogfilosof [Ludwig Wittgenstein](#).



Ludwig Wittgenstein (1889-1951)

Wittgensteins tidlige sprogforståelse

Kendetegnende for Wittgenstein er måden hvorpå hans liv kan opdeles på i to komplet divergerende perioder og hovedværker. Hver repræsenterer de et syn på det menneskelige sprog som på mange måder er komplet uforenelige med hinanden.

I sin tidlige periode var Wittgenstein dybt optaget af at afdække sprogets underliggende *logiske form*. Man kan sige at han forsøgte at finde 'matematikken i sproget' ved udelukkende at forholde sig til sprogets deskriptive egenskaber, og dette indenfor de strengeste logisk formaliserede rammer. En sætnings mening var udelukkende defineret ved dens sandhedsbetingelser; dvs. en forståelse af hvorledes verden ville se ud hvis den pågældende sætning var sand.

Dette deskriptive og logisk-matematiske syn på det menneskelige sprog, udmøntede sig i Wittgensteins første hovedværk "[Tractatus Logico-Philosophicus](#)". Værket er notorisk utilgængeligt for den gængse læser, og opstiller i hierarkisk punktform et regelsæt der beskriver hvorledes sproget bør anvendes, og i særdeleshed hvorledes det *ikke* bør anvendes.

Sagt i al korthed, viser Wittgenstein os hvorledes filosofiske problemer blot er et resultat af en forkert anvendelse af sproget, hvor der fejlagtigt fremsættes spørgsmål som ikke kan bringes på en *logisk form*. Konklusionen på værket lyder derfor også at "*Det, hvorom man ikke kan tale, om det må man tie*". Ifølge Wittgenstein er "*det hvorom man ikke kan tale...*" så de sproglige instanser som ikke i sig selv kan bringes på en logisk form, som for eksempel metafysiske eller eksistentielle spørgsmål som 'hvad er meningen med livet?'.

Hovedkritikken af værket er koncentreret omkring den *logiske atomisme* som bliver afgørende for mekanikken i Wittgensteins tidlige sprogforståelse. Wittgenstein hævder i Tractatus at sammenhængen mellem sproget og den verden som det beskriver, er ganske entydig via relationen mellem virkelighedens og sprogets 'mindste bestanddele'. Problemet er bare at Wittgenstein er totalt ude af stand til at præcisere hvad disse mindste bestanddele skulle være. Herved kan man til dels sige at selve fundamentet for Tractatus falder sammen.

Status på den moderne sprogforståelse

Hvis vi et øjeblik vender tilbage til det dominerende syn på sproget indenfor nutidens forskning, så er det interessant at se hvorledes der er mange fællestræk med Tractatus. For eksempel er der i dag et overhængende fokus på den beskrivende del af sproget der blot sætter navn på de ting og begreber vi kender til. Som vi senere skal se, er dette en mangelfuld sprogforståelse, der ikke formår at indkapsle samtlige relevante aspekter af et natursprogs dynamik og mangfoldighed.

Den mest afgørende, fælles fejlslutning må dog være antagelsen om at nøglen til en givtig systematisering

Seneste sprognyheder

4/5	Sprogforskerne fandt en skat i skoven politiken.dk
27/4	Lille indsats styrker små børns sprog www.dr.dk
26/4	Lad os komme det danske 'jantekomma' til livs politiken.dk
23/4	Ud med sproget - Berlingske Mener www.b.dk
20/4	Unikt runefund i centrum af Odense videnskab.dk

4/1	John Holm, Pioneer in Linguistics, Dies at 72 www.nytimes.com
10/8	Young women, give up the vocal fry and reclaim your strong female voice www.theguardian.com
20/5	Bill Funding Native Language Programs Passes mtpr.org
17/5	Sounds Of The Pilbara II: Songs In Language finishes recording - WAM - West Australian Music wam.org.au
13/5	Seven US Senators Introduce Bill to Promote Preservation of American Indian Languages - Native News Online nativenewsonline.net

FLERE NYHEDER >>>

Verdens sprog på Sproguseet.dk på et større kort

Mere i kategorien 'Sprogteknologi'

Fra Wittgenstein til Ontology Engineering 3: Kognitive mønstre og neurale netværk

Fra Wittgenstein til Ontology Engineering 2: Kropslighedstesen og den kognitive semantik

Nye kommentarer

Arturo til Hebraisk: Et genoplivet sprog eller et nyt sprog?

Yunus til Two Turkish Loanwords in Swedish

Sonstige til Den Danske Ordbog på nettet

Mads Haupt til Er det forkert at splitte sammensatte ord?

Monica Scheuer til Jødiske efternavne

jane til Jødiske efternavne

Inglês til Sprogene i Mozambique

Birgit Eggert til Hvad er der blevet af Maren?

Grethe Movsing til Hvad er der blevet af Maren?

Henrik Klindt-Jensen til Ded borrihjolmska måled

Artikler om

af sproget skulle ligge i en form for logisk atomisme der fokuserer på sprogets mindste bestanddele. På samme måde som i Tractatus, er nutidens sprogteknologiske løsninger nødt til at finde en systematik i sproget, så det kan håndteres af en computer. Den mest udbredte strategi hertil, som man ser anvendt i dag, kan med rette anses for en type af logisk atomisme idet den næsten udelukkende arbejder med sprogets mindste bestanddele; dvs. enkelte bogstaver og ord.

Der er helt konkret tale om *statistisk funderede løsninger*, som for eksempel kan oversætte en sætning ved at regne på sandsynligheden af en række ord- eller bogstavkombinationer mellem to forskellige sprog. De fleste har dog oplevet at selv de bedste oversættelsesprogrammer fungerer ganske utilfredsstillende. Den logiske atomisme som er indeholdt i den førnævnte moderne sprogforståelse, består altså helt konkret i et fejlbehæftet fokus på kombinatorikken af elementer fra de allernederste *niveauer af sproganalyse*. Spørgsmålet er hvad alternativet til denne logiske atomisme så er. Svaret finder vi i den sene Wittgensteins forkastelse af sine egne tidligere teorier i hans andet centrale hovedværk: "[Filosofiske Undersøgelser](#)".



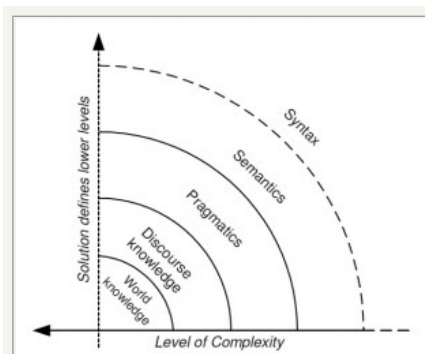
I dag er der blandt natursprogssystemer en udbredt tendens til at anvende statistisk funderede løsninger som fokuserer på de nederste niveauer af sproganalyse, såsom morfologi som beskæftiger sig med de enkelte ords struktur, samt grammatik og syntaks som beskæftiger sig med hvorledes disse ord kan arrangeres i forhold til hinanden efter gængse regler.

Wittgensteins lektie

I "Filosofiske Undersøgelser" forkaster Wittgenstein sin tidligere tro på logikken og ideen om logisk atomisme. I stedet for at anse sproget som en strengt logisk affære hvis hovedformål udelukkende er at *beskrive* verden, skifter han fokus til hverdagssproget og interpersonel kommunikation indenfor en social og kulturel kontekst. Meningen med sproget er ikke længere sandhedsbetingelser, men dets egentlige, kontekstuelle *brug*.

Denne nye sprogforståelse udmønter sig i det Wittgenstein kalder "sprogspil". Al sprogbrug er baseret på sprogspil, som er endnu mere grundlæggende end sproget selv. Wittgenstein eksemplificerer dette ved at gøre os opmærksomme på, at selv når børn lærer sproget via ostensiv definition (det at pege på en ting og sige dets navn), så indebærer dette et sprogspil hvis regler skal mestres før selve tillæringen af et sprog kan begynde. Sprogspil er desuden unikke for hver sproginstans. Enhver samtale har for eksempel et helt unikt sæt regler som aldrig kommer igen. Det kan således undre at vi i det hele taget er i stand til at kommunikere, til trods for denne ekstremt kontekstualitet som vores sprogbrug er underlagt.

Årsagen hertil er, ifølge Wittgenstein, at vi alle har en medfødt kapacitet for sprogspil, og at vi alle deler det han kalder en "fælles livsform". Wittgenstein selv beskæftiger sig ikke meget med hvad denne medfødte kapacitet for sprogspil kommer af, men indenfor de sidste par år er der blandt andet gjort opdagelser indenfor neurobiologien der understøtter synspunktet. I den næste artikel i serien ser vi blandt andet nærmere på opdagelsen af de såkaldte "spejlneuroner", som peger på at vi som mennesker har en medfødt kapacitet for samarbejde og empati, samt at disse egenskaber lader til at være stærkt forbundne med vores sprog. Ydermere kan vores ensartede kroppe, fysiske miljø og kulturelle omstændigheder også betegnes som delelementer af det som udgør vores fælles livsform.



Wittgensteins lektie fortæller os, at vi bør undlade at tage udgangspunkt i sprogets mindste bestanddele som man gør i dag, og i stedet starte ved det øverste niveau af sproganalyse, 'world knowledge', som så successivt kan definere hvorledes vi bedst håndterer de nedre niveauer af sproganalyse.

Wittgensteins lektie er en vigtig pointe. Hvis forskere indenfor de sprogteknologiske aspekter af kunstig intelligens skal tage ved lære af hans voldsomme revidering af sin egen sprogforståelse, så må vi først og fremmest forkaste alle afarter af den logiske atomisme som stadig nyder stor indflydelse i dag. I stedet for at fokusere på de nederste niveauer af sproganalyse må svaret ligge i at starte ved det *øverste* niveau og arbejde sig nedad – ikke omvendt.

Grunden til at tendensen til at starte med de nederste niveauer af sproganalyse i den grad har vundet fodfæste, er at dette hurtigt har kunnet give konkrete resultater via relativt simple, statistiske metoder. Det betyder dog ikke nødvendigvis at de indeholder potentiale for at kunne videreudvikles til at inkludere de højere niveauer. Faktisk har de sidste års forskning vist det modsatte. Fokus bør altså rettes mod det øverste sproganalyseniveau "World Knowledge", som er den viden sprogbrugere

må have for at kunne kommunikere med hinanden. Samtidigt skal vores indgangsvinkel hertil være, at denne viden skal organiseres ud fra en forståelse af vores fælles livsform og måden hvorpå den konkret udmønter sig i sproget.

I den anden del af denne lille serie på tre artikler, vil vi se nærmere på hvordan dette rent praktisk kan lade sig gøre, og gå dybere ind i nogle af de mest centrale konsekvenser for fremtidens forskning indenfor den sprogteknologiske side af moderne kunstig intelligens.

Erik David Johnson, cand.it
www.erk.dk

Læs også:

1. [Fra Wittgenstein til Ontology Engineering 2: Kropslighedstesen og den kognitive semantik](#) George Lakoff

aktuelle sprog [Alfabeter](#) [Anmeldelser](#) [arabisk](#)

Biblen [bogstaver](#) [børn](#) [Danmark](#) **Dansk** [Dialekter](#)

[engelsk](#) [esperanto](#) [Formidling](#) [fransk](#) [identitet](#)

[konsonanter](#) [Medier](#) [modersmål](#) [Musik](#) [Navne](#) [norsk](#) **Ord**

[ordbøger](#) [ordforråd](#) [oversættelse](#) [Plansprog](#) [religion](#)

[romanske sprog](#) [russisk](#) [Sjov](#) [skriftsprog](#) [sprogdød](#) [Sproggeografi](#)

[sprogkort](#) **Sprogpolitik** [sprogteknologi](#)

[svensk](#) [truede sprog](#) [tv](#) [tyrkisk](#) [tysk](#) [Udtale](#)

[Underholdning](#) [video](#) [vokaler](#)

Arkiv	Resources
januar 2015	Ethnologue: Languages of the World
december 2014	Forvo – All the Words in the World. Pronounced.
november 2014	LL-Map: Language and Location
maj 2014	Minority Rights Group
marts 2014	Omniglot. Writing Systems and Languages of the World
februar 2014	UNESCO Atlas of the World's Languages in Danger
oktober 2013	World Atlas of Linguistic Structures (WALS)
august 2013	
marts 2013	
januar 2013	
december 2012	
november 2012	
oktober 2012	
september 2012	Resurser
juli 2012	Bogstavlyd
juni 2012	Dansk sprognævn
maj 2012	Den danske ordbog
april 2012	Dialekt.dk
marts 2012	dk.kultur.sprog
februar 2012	Korpus.dk
januar 2012	Nye ord i dansk på nettet (NOID)
december 2011	Ordbog over det danske sprog
november 2011	Ordned. Dansk sprog i ordbøger og korpus
oktober 2011	Sproget.dk
september 2011	Svenska Akademien
august 2011	Ø (Schwa).dk
juli 2011	
juni 2011	
maj 2011	
april 2011	
marts 2011	
februar 2011	
januar 2011	
december 2010	
november 2010	
oktober 2010	
september 2010	
juni 2010	
maj 2010	
april 2010	
marts 2010	
februar 2010	
januar 2010	
december 2009	
november 2009	
oktober 2009	
september 2009	
august 2009	
juli 2009	
juni 2009	

er en fremtrædende figur indenfor kognitiv semantik, og betragtes af mange som faderen til kropslighedstesen. Udover sine analyser af politisk billedsprog, er han især også kendt for sit...

2. [Fra Wittgenstein til Ontology Engineering 3: Kognitive mønstre og neurale netværk](#) I denne tredje og sidste artikel i en serie af tre ser vi nærmere på de menneskelige, kognitive mønstre som den anden artikel i serien viste, var ganske essentielle. Fokus...

maj 2009

april 2009

marts 2009

Tagget med: [AI](#), [kunstig intelligens](#), [metafor](#), [ontologi](#), [spejlneuron](#), [Wittgenstein](#)

Skriv en kommentar

Navn (kræves)

E-mail (kræves)

Hjemmeside

Send mig en e-mail når der kommer flere kommentarer.