



Universiteit
Leiden

The Netherlands

Collaborative meaning-making: the emergence of novel languages in humans, machines, and human-machine interactions

Kouwenhoven, T.

Citation

Kouwenhoven, T. (2025, October 30). *Collaborative meaning-making: the emergence of novel languages in humans, machines, and human-machine interactions*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/4281976>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4281976>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

Taal is een unieke menselijke eigenschap waarmee we dagelijks samenwerken en ideeën uitwisselen. Hoewel het nu een gegeven is dat we elkaars taal begrijpen, is de manier waarop dit gedeeld begrip is ontstaan tot op heden nog een raadsel. Een prominent gedachtegoed in taalevolutie stelt dat herhaaldelijke interacties een ijkmoment bieden waarin signalen en betekenissen aan elkaar worden gelinkt. Culturele processen kunnen deze betekenisvolle signalen vervolgens verspreiden op populatieniveau. Samen zorgen ze er voor dat een taal zich aanpast aan de cognitieve voorkeuren van gebruikers, zoals een voorkeur voor simpele en gestructureerde signalen. Langdurige druk van zulke voorkeuren resulteert in een gedeelde taal die gestructureerd, expressief en te leren is.

Een nieuw soort taalgebruiker krijgt een steeds prominenter rol in onze samenleving. Kunstmatig intelligente systemen, zoals Large Language Models, kunnen inmiddels worden beschouwd als volwaardige taalgebruikers. Echter verschilt de manier waarop ze beslissingen nemen fundamenteel van hoe mensen dat doen. Daarom is het essentieel om deze grotendeels onbekende vormen van cognitie te onderzoeken. Dit proefschrift doet dat vanuit het perspectief van taalevolutie. Het gebruikt methoden uit dat veld, die niet alleen toepasbaar zijn op mensen, maar ook geschikt zijn om cognitieve patronen en voorkeuren bij niet-menselijke systemen te ontrafelen. De bevindingen die hieruit voortvloeien dragen bij aan het begrip van taalevolutie én informeren hoe niet-menselijke cognitie talige en visuele informatie verwerkt. Hiermee beantwoordt dit proefschrift de vraag hoe menselijke en niet-menselijke cognitie elkaar kunnen complementeren in onderzoek naar de evolutie van taal.

Allereerst wordt onderzocht hoe rudimentaire signalen ontstaan, en welke rol neurale netwerken daarbij kunnen spelen. Experimenten met mensen bevestigen dat gedeelde rudimentaire signalen ontstaan uit herhaaldelijke interacties en dat het kan helpen als gesprekspartners verschillen in hun behoefte aan structuur. Dit laatste nuanceert bestaand gedachtegoed dat gedeelde voorkeuren voordelig zijn. Door het gedrag van mensen in deze taak na te bootsen met computermodellen zien we dat bidirectionele mechanismen dit gedrag het beste kunnen verklaren, wat wil zeggen dat een focus op de omliggende context (zowel voor als na) van een communicatief signaal essentieel is, net als bij het verwerken van taal.

Vervolgens wordt onderzocht of multimodale modellen menselijke cross-modale associaties

vertonen en bestudeert dit proefschrift, met behulp van zelflerende modellen, de evolutie van gestructureerde taal waarin signalen opgebouwd zijn uit rudimentaire elementen. Waar mensen consistent een rondvormig object ‘bouba’ zouden noemen in plaats van ‘kiki’, vertonen vier multimodale modellen dit patroon nauwelijks. Dit bevestigt dat het voor moderne multimodale modellen lastig is om visuele en tekstuele informatie aan elkaar te linken. Zelfs wanneer dit soort multimodale modellen een communicatiespel spelen waarbij een eigen taal ontwikkelt wordt blijkt dat er andere visuele kenmerken worden gebruikt dan de kenmerken die mensen gebruiken. Hierdoor zijn ze niet in staat om met de geleerde taal twee plaatjes met dezelfde soort objecten in een andere opstelling van elkaar te onderscheiden—zelfs niet wanneer de modellen expliciet gedwongen worden om de juiste kenmerken te leren.

Afsluitend onderzoekt dit proefschrift of moderne taalmodellen kunnen fungeren als proefpersonen in taalevolutie-experimenten en bevraagt het in een collaboratief experiment of mens en machine samen een taal kunnen leren. Wanneer taalmodellen herhaaldelijk met elkaar interacteren in een communicatiespel ontstaat er een taal met compositionele eigenschappen waarin lettergrepen in nieuwe combinaties worden gebruikt om verschillende objecten te beschrijven. Dit heeft opvallende overeenkomsten met menselijke taalevolutie, ondanks dat de onderliggende mechanismen fundamenteel van elkaar verschillen. Ook uit simulaties met culturele overdracht—waarin opeenvolgende generaties taal leren en gebruiken—blijkt dat de taal zich aanpast naar gebruikersvoorkeuren van de taalmodellen op een manier die vergelijkbaar is met mensen. Toch zijn er ook verschillen: taalmodellen produceren vaker één signaal met meerdere betekenissen (homoniemen), en lange signalen, terwijl mensen de voorkeur geven aan korte, expressieve uitingen. De laatste bijdrage van dit proefschrift betreft een experiment waarin participanten samenwerken met een Llama3-70B taalmodel en een taal ontwikkelen. Ondanks de fundamenteel verschillende mechanismen voor taalverwerving en gebruik, ontstaan er expressieve en betekenisvolle talen. Deze talen bevatten compositionele eigenschappen en vertonen meer menselijke eigenschappen dan talen ontwikkeld zonder menselijke inbreng. Hiermee ondersteunen de resultaten het idee dat taal zich aanpast aan haar gebruikers en suggereren ze dat interacties, waarin communicatief succes belangrijk is, een rol kunnen spelen in natuurlijke taalverwerving door machines.

Met deze interdisciplinaire benadering van taalevolutie draagt dit proefschrift bij aan een onderzoekslijn die vanuit een menselijk oogpunt zowel menselijke als niet-menselijke cognitie onderzoekt. Het laat zien dat beide vormen van cognitie elkaar kunnen complementeren omdat ze van elkaar verschillen. Deze bevinding biedt veelbelovende mogelijkheden voor de ontwikkeling van talen die zijn aangepast aan de cognitieve capaciteiten van zowel mensen als machines waardoor we mogelijk op een natuurlijkere manier kunnen interacteren.