



Universiteit
Leiden

The Netherlands

Emergence of linguistic universals in neural agents via artificial language learning and communication

Lian, Y.

Citation

Lian, Y. (2025, December 12). *Emergence of linguistic universals in neural agents via artificial language learning and communication*. Retrieved from <https://hdl.handle.net/1887/4285152>

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/1887/4285152>

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

Samenvatting

Menselijke taal is voortdurend in ontwikkeling, waarbij taalkundige structuur wordt gevormd door taalgebruikers, zowel op individueel als op populatieniveau. De recente ontwikkeling van krachtige, neurale leersystemen in de AI heeft geleid tot een hernieuwde interesse in het simuleren van taalevolutie met agent-gebaseerde modellen—een krachtige methodologie om de dynamische aard van menselijke taal te simuleren.

Een recentelijk, maar rijk onderzoeksdomein, bekend als ‘emergent communication’, richt zich op interacties tussen agents die nieuwe communicatieprotocollen ontwikkelen waarmee zij succesvol informatie kunnen uitwisselen over vereenvoudigde werelden. Een belangrijk onderzoeksgebied in dit domein is het vergelijken van de ontstane taalprotocollen met abstracte structurele eigenschappen van menselijke talen om te onderzoeken of communicatieve sturing kan leiden tot menselijke taalkenmerken zoals compositionele grammatica. Echter, deze protocollen worden geïnitieerd met willekeurige symbolen, en de agents ontwikkelen hun eigen ‘gesloten codes’, wat het intrinsiek moeilijk maakt om hun talen te analyseren en te vergelijken met menselijke producties op het niveau van specifieke taalfenomenen.

Om enkele van de uitdagingen rond de standaardopzet van emergent communication aan te pakken, introduceert dit proefschrift het Neural-agent Language Learning and Communication framework (NeLLCom). Een belangrijke innovatie hierin is dat de agents eerst worden getraind in een vooraf gedefinieerde kunstmatige taal, wat aansluit bij de methoden die gebruikt worden in laboratoriumexperimenten met menselijke deelnemers. Concreet leren de NeLLCom-

agents deze kunstmatige talen eerst individueel via supervised learning, om vervolgens in paren of groepen te interacteren via een taalspel, waarbij ze dynamisch leren van deze interacties door reinforcement learning. Op deze manier biedt NeLLCom een algemene procedure voor taalverwerving en communicatie die kan worden gebruikt om veel taalfenomenen te bestuderen.

In dit proefschrift richten we ons op twee veelvoorkomende taalfenomenen: (i) de afweging tussen flexibiliteit van woordvolgorde en gebruik van naamvallen, en (ii) differentieel gebruik van naamvallen. In beide gevallen hebben we de kunstmatige taal zorgvuldig ontworpen naar voorbeelden van eerder bestudeerde systemen in experimenten met mensen. Onze resultaten tonen aan dat neurale agents beide fenomenen op een mensachtige manier kunnen repliceren via communicatie. Bovendien stelt ons nieuwe framework ons in staat de eerdere resultaten met mensen op grotere schaal uit te breiden; we hebben kunnen aantonen dat een trade-off tussen woordvolgorde en gebruik van naamvallen ook op groepsniveau ontstaat in de simulaties.

Door de wisselwerking te belichten tussen taalverwervingsprocessen en communicatieve behoeften in de vorming van menselijke taal, biedt dit proefschrift een integrerend framework dat aansluit bij moderne benaderingen in de computationele linguïstiek om taalverwerving en gebruik te simuleren. Dit framework kan worden gebruikt voor het uitvoeren van gecontroleerde experimenten, waarmee experimenteel onderzoek bij mensen over taalontwikkeling kan worden aangevuld, om zo het uiteindelijke doel te bereiken van het verklaren waarom menselijke talen zijn zoals ze zijn.