

GRAMMATIK Textgenerierende *large language models* (LLMs) rütteln an einigen lieb gewonnenen Vorstellungen, die wir von Sprache haben. Bemerkenswerterweise tun sie das unter dem Etikett <Sprachmodell>. Diese Bezeichnung suggeriert, die statistischen Modelle würden <die Sprache> repräsentieren – aber eine solche Sprache gibt es nicht. Und dass textgenerierende LLMs so gut funktionieren, zeigt das deutlich.

Führen wir uns eine einfache, aber gängige Idee davon vor Augen, was Sprache ist. Wie wir auch an diesem Text sehen können, leben sprachliche Äußerungen davon, dass sie einen begrenzten Zeichenvorrat, Wörter, in eine syntaktische Ordnung bringen, welche die Zeichen in Beziehung setzt, woraus eine unbegrenzte Zahl von Sätzen entstehen kann. «Irgendjemand tut irgendetwas mit irgendjemandem» ist eine dieser syntaktischen Strukturen – also ein Prädikat (irgendetwas tun), das ein Subjekt (irgendjemand) mit einem Objekt (irgendjemandem) in Verbindung bringt. Aus dieser grundlegenden Struktur ergeben sich Regeln über morphologische Anpassungen, die als Kasus, Numerus, Genus usw. beschrieben werden. Wenn ich eine Sprache lernen will, dann muss ich einen möglichst großen Teil des Zeichenvorrats kennen (das Lexikon) sowie die Regeln, nach denen die syntaktischen Ordnungen immer wieder neu erzeugt werden können (die Grammatik). <Sprache> ist, so die Vorstellung, das Zusammenspiel von Lexikon und Grammatik.

Textgenerierende LLMs wie ChatGPT und andere Chatbots erzeugen nun problemlos Sätze und ganze Texte, die lexikalisch wie grammatisch einwandfrei sind und die aktuell geltenden Regeln der Rechtschreibung einhalten. Sie tun dies aber ohne jegliches Wissen einer Grammatik und auch ohne ein eigentliches Lexikon. Und sie können noch mehr: Die erzeugten Texte sind semantisch kohärent und pragmatisch plausibel. Sie simulieren ganz verschiedene Sprachhandlungen (soweit das ohne Körper und mittels eines Webinterfaces möglich ist): loben, beleidigen, widersprechen,

kritisieren – und spielen so ein kommunikatives Gegenüber. Wie geht das ohne Grammatik und ohne Lexikon?

Es gibt in der Linguistik einen alten Streit darüber, wie Sprache am besten modelliert werden kann. Sehr vereinfacht gesagt herrscht auf der einen Seite die Überzeugung vor, Sprache sei ein System, das sich über die beiden Komponenten Grammatik und Lexikon angemessen beschreiben lasse (wobei wiederum darüber gestritten wird, welche Grammatik die beste Modellierung darstellt). Aufgabe der Linguistik sei es, herauszufinden, welche <Kompetenz>, also welches Wissen von Grammatik ein*e Sprecher*in besitzen müsse, um sprachliche Äußerungen generieren zu können.

Auf der anderen Seite, im Kontextualismus, Distributionalismus und Poststrukturalismus, wird zumindest angezweifelt, dass es Sprache überhaupt als System gibt, und vielmehr darauf hingewiesen, dass wir letztlich nur Sprachgebrauch, <Performanz>, sehen und darin etwa typische Muster identifizieren können, die immer wieder (in mitunter angepassten Formen) eingesetzt werden. Wittgensteins bekannte Aussage, die «Bedeutung eines Wortes ist sein Gebrauch in der Sprache» (Wittgenstein 2003, § 43), steht für diese Überzeugung. Um ein Wort zu verstehen, schlagen wir nicht gleichsam in unserem Kopf in einem großen Lexikon nach, sondern wir kennen typische Kontexte, in denen das Wort so oder so eingesetzt werden kann, wodurch ein kommunikativer Zweck erreicht wird. Diese Kontexte stehen für zweierlei: typische syntaktische Strukturen und typische Verwendungskontexte, also die pragmatischen, sprachhandelnden Aspekte.

Die Computerlinguistik indessen ist längst von Kompetenz-orientierten Modellierungen der Sprache zu Performanz-orientierten Modellen übergegangen. Maschinelle Übersetzung wurde in ihren Anfängen mit regelbasierten Ansätzen angegangen, d.h. mit programmierten Grammatiken und Lexika. Heute ist klar, dass Systeme,

G



Abb. 1 Roboterkatze CatGPT (Foto: Noah Bubenhofer)

G die auf sprachliche Performanz rekurren und Sprachgebrauch statistisch <erlernen>, ungleich besser funktionieren. Und dies zeigt sich besonders eindrücklich an textgenerierenden LLMs. Solche Sprachmodelle werden zwar schon seit längerer Zeit erstellt. Aber die neueste Generation der LLMs führt vor Augen, welchen Effekt der gigantische Anstieg von Textdaten hat, mit denen die Modelle trainiert werden. Die Simulation von immer mehr Facetten des Sprachgebrauchs wird möglich, sodass Chatbots ganz verschiedene Formen von Sprachhandeln beherrschen (Dialoge führen, Wissen wiedergeben, Texte analysieren und generieren).

Die Idee der Grammatik ist damit eigentlich tot. Zumindest hat sie sich als deutlich weniger plausible Modellierung des Sprachgebrauchs erwiesen. (Ganz nutzlos sind Grammatiken freilich nicht, bleiben sie doch praktische Hilfsgerüste, um z. B. eine Fremdsprache zu lernen. Sie reichen jedoch bei Weitem nicht aus, um kommunikative Routinen in ihrer Komplexität verstehen und reproduzieren zu können.) Jetzt

könnte man einwenden, dass diese Erkenntnis aus sprachtheoretischer und linguistischer Sicht zwar ganz interessant ist, darüber hinaus aber wenig Relevanz hat. Doch scheint mir, dass wir bei den Versuchen der Modellierung von Performanz erst ganz am Anfang stehen und dass die weitere Entwicklung größere Auswirkungen auf Theorien der Kommunikation und der Medialität haben wird.

Das lässt sich anhand meiner Roboterkatze <CatGPT> erahnen (vgl. Abb. 1). Sie basiert auf einem Bausatz der Firma Petoï und besitzt elf Gelenke, d. h. Servos, sowie einen Mikrocomputer auf Arduino-Basis, der die Servos koordiniert, damit Bewegungen wie gehen, sitzen, Pfote heben usw. möglich sind. Ich erweiterte die Katze um einen weiteren Mikrocomputer (Raspberry Pi) mit Mikrofon und Spracherkennung. Nun ist es möglich, mit der Katze zu sprechen. Der menschliche Sprachinput wird zu einem Prompt (menschliche Anweisung an den ChatBot) für ChatGPT kombiniert:

[Prompt:] Du bist eine intelligente Katze. Antworte auf die folgende Frage mit Bewegungen. Du kannst folgende Bewegungen: kbalance (stehen), kbuttUp (Hintern hoch), krest (pausieren) [...]. Wenn du keine Bewegung weißt, dann sage einfach ksit. Gib die Befehle komma-separiert zurück.

[Sprachinput:] «Hallo Katze, was machst du, wenn du fröhlich bist?»

ChatGPT erzeugt aufgrund des Prompts und auf Basis des Sprachmodells eine verbale Antwort auf den Sprachinput, ist durch die Anweisung jedoch gezwungen, diese in Bewegung auszudrücken, um die Katze entsprechend zu steuern. Eine Antwort von ChatGPT auf den oben genannten Prompt waren die Bewegungen ksit (sitzen), kstr (sich strecken), kwh (Kopf links/rechts bewegen), kvf (auf der Stelle treten), kwh (den Kopf links/rechts bewegen), kcmh (herkommen), kp (Push-ups) und kfv (ein von ChatGPT neu erfundener Code, der deshalb auch nicht in Bewegung umgesetzt werden kann!).

Auf die Frage «Was machst du, wenn du traurig bist?» folgten die Bewegungen krest (pausieren), ksit (sitzen), kzero (Körperstellung neutralisieren). Hier wurden also merklich weniger Bewegungen erzeugt, was eine plausible körperliche Umsetzung oder Performanz von Traurigkeit darstellt.

Die linguistischen Teildisziplinen, die sich mehr für Performanz als für Kompetenz interessieren, beschäftigen sich seit längerer Zeit mit Aspekten der Sprachlichkeit abseits von Lexik und Grammatik: gesprochene Sprache, Texte und Textfunktionen, sprachliches Handeln, sprachlich vermittelte und geprägte gesellschaftliche Diskurse, Kultur und Sprache, Gesellschaft und Sprache, Sprache in Interaktion und als körperliche Praktiken usw. Der Fall der Roboterkatze zeigt exemplarisch, dass Sprache und Körperlichkeit zusammengedacht werden müssen. Das Beispiel zeigt aber auch, dass selbst in einem Sprachmodell, das mit Texten trainiert wurde, Spuren von körperlichen Praktiken repräsentiert sind! Sprachgebrauch gedacht als

Sprachhandeln in bestimmten Kontexten von Raum, Körper und Zeit, das diese Kontexte zugleich repräsentiert und hervorbringt, ist der Kern dessen, was wir als Sprachlichkeit bezeichnen. Eine Grammatik will und kann diese Art von Sprache nicht modellieren.

Was wird in den kommenden Jahren passieren? Wenn die Trainingsdaten nicht mehr nur aus Texten bestehen, sondern aus multimodalen Kontexten von Sprachgebrauch, könnten immer umfassendere Modelle entstehen. Meine Katze würde davon profitieren: Ihre körperlichen Antworten auf verbalen und nonverbalen Input würden differenzierter – sofern ihr <Körper> solche Bewegungen ermöglicht. Aber auch die kommerziellen Hersteller der Modelle werden profitieren. Es werden mächtige Instrumente entstehen, die weit über *natural language processing* und ein <Sprachmodell> hinausgehen, indem sie andere Zeichensysteme und komplexe Interaktionen integrieren.

Es wird entscheidend sein, eine *AI literacy* zu vermitteln, die auch ein Verständnis dafür gibt, was Sprache ist und wie sie modelliert werden kann. Wenn in den Köpfen der Benutzer*innen von Künstlichen Intelligenzen die Vorstellung von Sprache als System, Sprache als Lexikon und Grammatik vorherrscht, werden die Möglichkeiten von LLMs falsch eingeschätzt. Sie werden womöglich als <Intelligenz> überschätzt, als LLM jedoch massiv unterschätzt. Mit dem Bedeutungsverlust von Grammatik in der maschinellen Verarbeitung und Modellierung von Sprache geht also ein Verlust an Übersichtlichkeit einher: Sprachgebrauch, Kommunikation und Interaktion sind weit mehr als bloße Grammatik.

NOAH BUBENHOFER

Lit.: Wittgenstein, Ludwig (2003): *Philosophische Untersuchungen*, Frankfurt/M.

G