Postprint: Mitteilungen des deutschen Germanistenverbandes
Digitale Umbrüche. Sprache, Literatur und Deutschunterricht in Zeiten von Big Data und KI
Jahr: 2023, **Jahrgang:** 70, **Heft:** 4, **Seiten:** 327–345

Der Sprachgebrauchsautomat. Die Funktionsweise von GPT und ihre Folgen für Germanistik und Deutschdidaktik *

Hans-Georg MÜLLER
Universität Potsdam
Potsdam, Deutschland
hans-georg.mueller@uni-potsdam.de
0000-0003-3561-8233

July 21, 2025

Maurice FÜRSTENBERG
LMU München
München, Deutschland
m.fuerstenberg@lmu.de
0009-0001-1090-9299

***Abstract*— In diesem Beitrag erläutern wir – ausgehend von einer Darstellung technischer Grundlagen von künstlicher Intelligenz im Allgemeinen und GPT im Besonderen –, inwiefern sich Transformer-Netzwerke als eine Art „Sprachgebrauchsautomat" auffassen lassen. Aus diesen Kernprinzipien leiten wir sieben Thesen für die Germanistik und die Deutschdidaktik ab:**

1. **GPT ersetzt „Bedeutung" durch „Auftretenswahrscheinlichkeit".**
2. **GPT ist kein Speicher für Wissen, sondern ein Speicher für Phrasen.**
3. **GPT schreibt im wahrsten Sinne des Wortes „durchschnittlich".**
4. **GPT zementiert den Mainstream.**
5. **Der Output von GPT droht, zur sich selbst erfüllenden Prophezeiung zu werden.**
6. **GPT kann den Erwerb von Schreibkompetenz unterstützen – oder ersetzen.**
7. **Für die Philologien ist GPT in analytischer Hinsicht interessanter als in produktiver.**

Stichwörter*—*Germanistik, Sprachmodell, GPT, Funktionsweise

1 Hinführung

Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache

Der Auszug aus §43 der Philosophischen Untersuchungen Wittgensteins bildet den Kernsatz der Gebrauchstheorie der Semantik. Wittgenstein wendet sich damit – pointiert ausgedrückt – gegen die Auffassung, dass Zeichen eine festlegbare Referenz aufweisen, sondern führt die Semantik sprachlicher Zeichen auf deren Pragmatik zurück, sodass sich jene im Kern erübrigt hat.

Seit ihrer Veröffentlichung hat sich die Gebrauchstheorie der Semantik im wissenschaftlichen Diskurs als äußerst fruchtbar erwiesen – sowohl als Inspirationsquelle für neue sprachwissenschaftliche Paradigmen als auch in der kritischen Auseinandersetzung mit ihr (etwa Fodor, Katz 1963, Chomsky 1969, Grewendorf 1985, Hinzen 2004). In diese Debatte hat ChatGPT ein neues Argument eingebracht. Denn technisch betrachtet lassen sich Transformer-Netzwerke (Das „T" in GPT steht für „Transformer") als Modell des menschlichen Sprachgebrauchs begreifen: Trainiert mit einer ungeheuren Menge an Sprachdaten repräsentieren sie die typischen Auftretensmuster von Wörtern und ihre topologischen Beziehungen zueinander. Wenn menschliche Sprachnutzer:innen im Output dieser Technik Bedeutung zu erkennen glauben, dann nicht deshalb, weil GPT etwas „verstanden" hätte, sondern weil es ein Modell des Sprachgebrauchs angelegt hat, das authentisch wirkende Texte produziert, denen wir Bedeutung zuzusprechen bereit sind.

In diesem Beitrag erläutern wir, ausgehend von einer Darstellung technischer Grundlagen von künstlicher Intelligenz im Allgemeinen und GPT im Besonderen, inwiefern sich Transformer-Netzwerke als eine Art von „Sprachgebrauchsautomat" auffassen lassen, und leiten aus diesen Kernprinzipien sieben Thesen für die Germanistik und die Deutschdidaktik ab.**(327)** Zu tief in die technischen Einzelheiten muss dabei nicht vorgedrungen werden, weil sich der mit Abstand komplexeste Teil der Technik von GPT letztlich als (technisch meisterhafter) Workaround erweist, um einen vergleichsweise kruden und ungemein ressourcenhungrigen Lernalgorithmus überhaupt auf aktueller Hardware und in endlicher Zeit lauffähig zu machen. Deshalb sind die allermeisten Einzelheiten der Architektur aktueller Sprachmodelle eher ingenieurstechnisch interessant und für das prinzipielle Verständnis ihrer Arbeitsweise und damit unseren Beitrag entbehrlich.

2 Was tun und wie lernen aktuelle Sprachmodelle?

Die folgenden Ausführungen erläutern technische Aspekte von Transformer-Netzwerken wie ChatGPT und bilden damit die Grundlage der darauffolgenden Argumentation. Für den

*Zitiervorschlag: Müller, H.-G. & Fürstenberg, M. (2023). Der Sprachgebrauchsautomat. Die Funktionsweise von GPT und ihre Folgen für Germanistik und Deutschdidaktik. *Postprint: Mitteilungen des Deutschen Germanistenverbandes*(70/4), 327–345.

Nachvollzug der anschließend formulierten Thesen sind sie nicht zwingend erforderlich, wohl aber für das Verständnis, warum wir zu diesen und keinen anderen Schlüssen kommen.

2.1 KI als automatisierte Statistik

Ein Großteil dessen, was unter dem Schlagwort „künstliche Intelligenz“ kursiert, erweist sich im Kern als ein vergleichsweise simples statistisches Anpassungsverfahren, dessen Komplexität weniger aus der Raffinesse der verwendeten Methoden als aus seiner schieren Größe erwächst – und zwar sowohl der Größe des verwendeten Datensatzes als auch des Rechenaufwandes, der zur Festlegung der Modellparameter betrieben wird.

Statistische Modelle dienen bekanntlich dazu, die Regelmäßigkeiten und Muster eines vorgegebenen Datensatzes zu repräsentieren, um daraus generalisierbare Schlüsse zu ziehen. Wer beispielsweise den Schulerfolg von Kindern anhand ihrer Lebenssituation voraussagen möchte, erhebt an einer repräsentativen Versuchspopulation verschiedene Sozialvariablen und misst anschließend, in welchem Grad und mit welcher Wahrscheinlichkeit diese die Zielvariable „Schulerfolg“ beeinflussen. Künstliche neuronale Netzwerke tun im Grunde nichts anderes, denn auch sie versuchen, den wahrscheinlichsten Wert einer Zielvariable (den Output des Netzwerkes) nach Maßgabe eines repräsentativen Datensatzes (des Inputs) zu ermitteln. In Sprachmodellen besteht dieser Input stets aus *Text*, der Wort für Wort[1] präsentiert wird und zu dem das Netzwerk das jeweils nächste Wort als Output voraussagen soll.**(328)** Der wesentliche Unterschied zwischen klassischen Statistiken und künstlicher Intelligenz besteht darin, dass bei Letzterer die Einflussvariablen nicht im Voraus bekannt, sondern lediglich implizit in den Daten enthalten sein müssen. So brauchen Sprachmodelle im Voraus keine Hypothesen darüber, welche grammatischen oder semantischen Eigenschaften eines Textes das Auftreten des jeweiligen Folgewortes beeinflussen. Stattdessen passt ein Lernalgorithmus (vgl. Kap. 1.3) die Parameter des Modells während eines langwierigen „Trainings“ so an, dass das Netzwerk mit immer höherer Wahrscheinlichkeit aus dem vorgegebenen Input (dem bisherigen Text) den erwünschten Output (das richtige Folgewort) vorhersagt. Die Selbstständigkeit dieses Anpassungsprozesses ist möglicherweise das einzige, was das Etikett „Intelligenz“ verdient.

2.2 Der Aufbau künstlicher neuronaler Netzwerke

Künstliche Intelligenz modelliert in stark vereinfachter Form eine Grundfunktion natürlicher Neuronen. Diese zeichnen sich dadurch aus, dass sie elektrische Impulse als Input aufnehmen (z.B. von Sinneszellen oder von anderen Neuronen), verarbeiten und als Output an andere Neuronen oder das motorische System des Körpers (Muskulatur) weitergeben (Schema: Abbildung 1). Die Informationsverarbeitung geschieht dabei einerseits innerhalb des Neurons, das aufgrund seines Inputs entweder selbst aktiv wird oder nicht, andererseits in den Verbindungen zwischen den Neuronen (den Synapsen, in Abbildung 1 als Pfeile symbolisiert), die höchst unterschiedlich stark sein können, sodass Neuronen einander sowohl aktivieren als auch hemmen können.

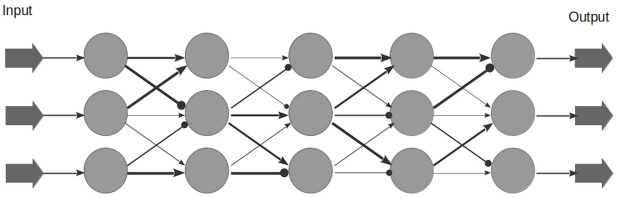

Abbildung 1: Schematisiertes neuronales Netzwerk. Die verschiedenen Strichstärken und Pfeilspitzen symbolisieren unterschiedlich starke aktivierende und hemmende Verbindungen.**(329)**

Im Gegensatz zu ihren natürlichen Vorbildern arbeiten künstliche Neuronen weitgehend dichotom, das heißt, sie reagieren auf ihren Input bis zu einem gewissen Schwellenwert gar nicht, aber leiten bei Überschreitung des Schwellenwertes selbst elektrische Signale weiter (man sagt: sie „feuern“). „Schaltern“, die bestimmte Teile des Inputs entweder weiterleiten oder unterdrücken (Alles-oder-Nichts-Prinzip). Der Schwellenwert bewirkt dadurch einen Abstraktionsprozess, weil jedes Neuron aus den potenziell unendlichen Möglichkeiten seines Inputs einen nahezu dichotomen Output erzeugt, also ein metrisches Signal in ein kategoriales überführt.

Das Alles-oder-Nichts-Prinzip führt notwendig zu Datenverlust, da ein Großteil der Inputsignale für die weitere Reizverarbeitung keine Rolle mehr spielt. Im Gegenzug kann das kategoriale Outputsignal aber so zur Repräsentation eines bestimmten regelmäßigen Ereignisses, eines Musters in den Inputdaten werden. Wenn etwa in den Trainingsdaten von GPT auf das Wort „Olaf“ häufig das Wort „Scholz“ folgt, so kann diese Regelmäßigkeit vom Netzwerk gelernt werden und schlägt sich in einem bestimmten Aktivierungsmuster der künstlichen Neuronen untereinander nieder.

Die zwischen Input und Output liegenden Neuronen dienen der Repräsentation solcher wiederkehrender Muster und jedes Neuron stellt darin einen eigenen kleinen Abstraktionsschritt dar, weil es die Signale der vorausgegangenen Neuronen aufsummiert, dem eigenen Schwellenwert unterwirft und erneut als kategoriales Signal weitergibt. Der Output des Gesamtsystems ist damit eine Abstraktion von Abstraktionen von Abstraktionen und bewirkt, dass nicht die individuellen Eigenschaften der einzelnen Inputdaten, sondern ihre generalisierbaren, wiederkehrenden Muster extrahiert werden.

Je vielschichtiger ein Netzwerk aufgebaut ist, umso komplexer können die Muster sein, die es repräsentieren kann. Aktuelle Netzwerkarchitekturen kämpfen systematisch mit dem Dilemma, möglichst viele Abstraktionsstufen erlauben zu wollen, den Rechenaufwand jedoch nicht ins Unendliche treiben zu können, da durch zusätzliche Neuronen immer mehr und mehr Parameter berechnet werden müssen.

[1] Genauer müsste man sagen: token für token, denn große Sprachmodelle werden nicht direkt mit Text trainiert, sondern mit sog. tokens, ganzzahligen Identifikationsnummern, die ein Wort, ein Wortteil, ein Satzzeichen oder ein Steuersymbol codieren. Wir werden im folgenden Text der Anschaulichkeit halber bei *Wort* bleiben.

2.3 Der Lernalgorithmus künstlicher neuronaler Netzwerke

Die Art, wie künstliche neuronale Netzwerke die Regelmäßigkeiten und Muster ihrer Trainingsdaten lernen, ist im Grunde alles andere als intelligent, sondern eher *brute force*: Das Modell verhält sich wie ein ziemlich dummer „Schüler", der die Antwort auf die Fragen seines „Lehrers" einfach rät. Nicht wesentlich klüger agiert der „Lehrer", dessen einzige didaktische Intervention darin besteht, dem „Schüler" Erfolg und Misserfolg seines Ratens zurückzumelden und ihn so zur Verhaltensänderung anzuregen. Dieses Spiel führen die beiden so lange mit den**(330)** Trainingsdaten fort, bis der „Lehrer" mit den Antworten des „Schülers" hinreichend zufrieden ist. Der Lernalgorithmus in dieser sehr einfachen Kombination aus *trial and error* einerseits und Erfolgsrückmeldung andererseits besteht darin, dass aus der Abweichung von erzeugtem und erwünschtem Output (also bei GPT aus der Diskrepanz zwischen geratenem und tatsächlichem Folgewort im Trainingstext) ein Fehlersignal errechnet und dem Netzwerk als Feedback zurückgegeben wird. Gesetzt etwa den Fall, das Netzwerk in Abbildung 1 hätte zufällig aus einem bestimmten Input X das oberste Outputneuron aktiviert, während der erwünschte Output das mittlere Neuron gewesen wäre. In diesem Fall zeigt das Fehlersignal an, dass die Aktivität des obersten Neurons offenbar verringert und die Aktivität des mittleren verstärkt werden müsse. Neuronale Netzwerke erreichen dies, indem sie alle unerwünscht hohen Verbindungsstärken um einen bestimmten Betrag (die sog. *Lernrate*) verringern und alle erwünschten um diesen Betrag erhöhen. Diese Anpassung wird mithilfe eines Rückmeldealgorithmus (*Backpropagation*, vgl. Rumelhart et al. 1986) anteilig an alle vorausgehenden Neuronen weitergeleitet, sodass der tatsächliche Output beim nächsten Trainingsdurchlauf ein wenig erwünschter ausfällt und das Fehlersignal etwas kleiner wird. Wird dieses Verfahren lang genug und mit einer geeigneten Netzwerkarchitektur fortgeführt, nähert sich er tatsächliche Output immer stärker dem erwünschten an und das Fehlersignal wird minimiert. Wird es so klein, dass tatsächlicher und erwünschter Output hinreichend übereinstimmen, repräsentieren die Parameter des Modells die Regelmäßigkeiten der Trainingsdaten, sodass das Netzwerk nun auch aus unbekanntem Input adäquaten Output generieren, also beispielsweise einen unbekannten Text angemessen fortführen kann.

Die Einfachheit des eingesetzten Lernmechanismus wird in künstlichen neuronalen Netzen teuer erkauft, denn sie macht es erforderlich, dass die Inputdaten dem Netzwerk wieder und wieder (und wieder und wieder…) präsentiert und in jedem Lernschritt die Modellparameter um den Betrag der Lernrate angepasst werden, bis die Fehlerrate nicht weiter sinkt. Das kann einen erheblichen zeitlichen und energetischen Aufwand verursachen (vgl. Rödel 2023) und bei großen Sprachmodellen hart an die Grenzen des derzeit technisch Umsetzbaren reichen. Nicht umsonst steht das „P" in GPT für „pretrained" (vortrainiert) und verweist mit dieser Bezeichnung darauf, dass ein möglichst großer Teil dieses brute-force-Lernens nur einmal (bzw. möglichst selten) erfolgen sollte,[2] bevor das Modell anschließend kopiert und mithilfe spezialisierter Trainingsdaten auf konkretere Einsatzzwecke vorbereitet wird.**(331)**

2.4 Rekurrenz und die Besonderheit von Transformer-Netzwerken

Wie erwähnt, erhalten große Sprachmodelle wie GPT Texte als Input und sollen das jeweils nächste Wort des Textes als Output voraussagen. Damit GPT dabei nicht nur den je aktuellen Input (also das unmittelbar vorangegangene Wort) berücksichtigt, sondern alle Wörter des bisherigen Textes, werden sog. rekurrente Schleifen eingesetzt.

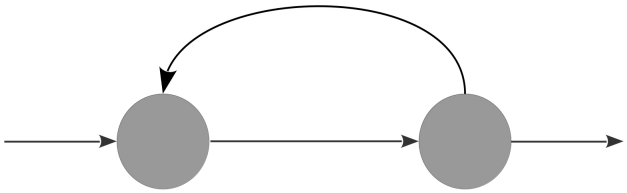

Abbildung 2: Rekurrente Verschaltung zwischen zwei Neuronen. Durch den Rückkanal wird das erste Neuron von seiner eigenen Aktivität mitbeeinflusst.

Rekurrenz entsteht bekanntlich dadurch, dass Signale nicht nur vom Sender zum Empfänger fließen, sondern vom Empfänger auch wieder an den Sender zurückgeleitet werden und damit das Folgeverhalten des Senders beeinflussen (Abbildung 2). So könnte ein mit aktuellen Zeitungstexten trainiertes Sprachmodell *ohne* rekurrente Schleife auf den Input „Olaf" stets nur mit dem Output „Scholz" reagieren, weil dieses Wort im Trainingsdatensatz mutmaßlich das häufigste Folgewort nach „Olaf" wäre und somit das typischste Muster abbildete. Durch die rekurrente Schleife hingegen wird die Verarbeitung des Inputs auch von den vorausgegangenen Wörtern beeinflusst, sodass eine differenziertere Reaktion möglich wird: Das Netzwerk reagiert dadurch kontextsensitiv und antwortet nur dann mit „Scholz", wenn der vorangegangene Input Wörter wie „Bundeskanzler" oder „SPD" enthalten hat, während es bei vorherigem Input wie „Comedian" (hoffentlich) eher der Output „Schubert" aktiviert. Mit diesem einfachen Verfahren gelingt es künstlichen neuronalen Netzwerken, auch sequenzielle Muster zu lernen und bei der Voraussage des wahrscheinlichsten Folgewortes den vorherigen Kontext mitzuberücksichtigen. Rekurrenz ist damit maßgeblich für die Herstellung von Kohärenz im Text verantwortlich (oder besser gesagt: für die *Simulation* von Kohärenz, s.u.).

Rekurrente neuronale Netzwerke (RNN) werden bereits seit längerer Zeit produktiv eingesetzt, so etwa bei Übersetzungssoftware, da durch die Berücksichtigung des Kontextes etwa zwischen Synonymen adäquater ausgewählt werden kann. Leider reicht die Kontextsensitivität klassischer RNN nur wenige Schritte (i.e. Wörter) zurück, sodass sie sich für die Repräsentation umfangreicherer**(332)** Textzusammenhänge nicht eignen. Technisch elaboriertere Netzarchitekturen wie etwa Long-Short-Term-Memory, das z.B. derzeit in Google Translate eingesetzt wird, gestalten die rekurrente Schleife etwas geschickter und steigern damit die Kontextsensitivität, kommen aber bei längeren Textpassagen ebenfalls an ihre Grenzen (im Deutschen etwa bei umfangreicheren Satzklammern).

[2] Hier liegt ein Grund dafür, dass etwa ChatGPT 3.5 zum Verfassungszeitpunkt dieses Beitrages noch auf dem Datenstand von September 2021 ist.

Die entscheidende Idee von Transformer-Netzwerken, die den Durchbruch von GPT ermöglicht haben, liegt darin, die rekurrenten Beziehungen des Modells durch eigene neuronale Netzwerke umzusetzen (sog. Attention-Module, vgl. Vaswani et al. 2017), die im Laufe des Trainings selbstständig lernen, welche vorausgegangenen Wörter des Inputs für die Vorhersage des kommenden Outputs besonderen Vorhersagewert haben und folglich besonders gewichtet werden sollten. Attention-Module werden ihrem Namen insofern gerecht, als sie die „Aufmerksamkeit" des Netzwerkes steuern und vergangenen Input verstärkt oder abgeschwächt in die jeweilige Verarbeitung einbeziehen.

In GPT kommen Attention-Module auf unterschiedlichen Ebenen der Verarbeitung zum Einsatz, was der simulierten Textkohärenz unterschiedliche Reichweiten verleiht. Dagegen bleibt unbekannt, aus welchen Gründen Attention-Module welchen Wörtern des Textes welches Gewicht verleihen, weil auch ihre Parameter dem oben dargestellten Lernalgorithmus unterliegen, also das Resultat eines kleinschrittigen Anpassungsprozesses zur Minimierung des Fehlersignals sind.

2.5 Zwischenfazit: Was kann und was tut GPT?

Ausgehend von der dargestellten Architektur und dem zugrundeliegenden Lernalgorithmus lässt sich die Arbeitsweise von Transformer-Netzwerken folgendermaßen zusammenfassen: Sie berechnen in jedem Arbeitsschritt die Wahrscheinlichkeit für das Auftreten jedes möglichen Folgewortes unter der gewichteten Bedingung des Auftretens aller vorausgehenden Wörter. Diese Wahrscheinlichkeit ergibt sich aus den komplexen Gebrauchsmustern jedes Wortes im Trainingsdatensatz, welche durch den Lernalgorithmus aus den Trainingsdaten extrahiert wurden.

Der Output von GPT repräsentiert also nichts anderes als einen extrem komplexen topologischen Durchschnitt dafür, mit welchem Wort nach Maßgabe der Trainingsdaten im aktuellen Kontext am wahrscheinlichsten zu rechnen ist. Dabei setzt ChatGPT seine Texte keineswegs immer mit dem wahrscheinlichsten Wort fort, sondern wählt mithilfe eines Zufallsgenerators aus der Gruppe der wahrscheinlichen Wörter eines aus. Dieser Umstand ist dafür verantwortlich, dass in ein und demselben Kontext gleichwohl unterschiedliche Antworten des Modells entstehen, die durch die rekurrente Weiterverarbeitung teils stark unterschiedliche Gesamttexte verursachen.**(333)**

Das eigentliche Faszinosum von GPT besteht darin, dass mit diesem prinzipiell einfachen und nur anwendungstechnisch komplexen Lernverfahren weder grammatisches Wirrwarr noch inhaltlich zusammenhangloses Kauderwelsch entsteht, sondern „Text", den menschliche Sprachbenutzer:innen als bedeutungsvoll erleben. Diese vermeintliche Bedeutung ist jedoch lediglich die gewichtete, durch ihr topologisches Umfeld bedingte statistische Auftretenswahrscheinlichkeit jedes Wortes im Trainingsdatensatz – oder, um es mit Wittgenstein zu sagen: sein Gebrauch in der Sprache[3].

GPT tut also nichts anderes, als der Sprachgemeinschaft die Regelmäßigkeiten ihres eigenen Sprachgebrauchs vorzurechnen. Dass es dabei so verblüffend differenziert antworten kann, zeigt vor allem, wie regelmäßig und wie vorhersagbar unser Sprachgebrauch eigentlich ist – so regelmäßig nämlich, dass die Extraktion topologischer Auftretensmuster hinreicht, um Bedeutung und Verständnis vorzugaukeln.

Ein wichtiger Grund dafür, warum GPT dieses Gaukelspiel so überzeugend beherrscht, liegt, wie oben erwähnt, in der enormen Größe sowohl der eingesetzten Netzwerke wie des verwendeten Trainingsdatensatzes. So repräsentiert GPT 3.5[4] die Regelmäßigkeiten seiner Trainingsdaten in 178 Milliarden Modellparametern und ist damit um ein Vielfaches größer als alle zuvor entwickelten großen Sprachmodelle zusammen. Das Training erfolgte mit einem Datensatz von 45 Terabyte Größe (zum Vergleich: Die Bibel bringt es mit 5 MB Text auf ein Neunmillionstel, die gesamte englischsprachige Wikipedia mit ihren derzeit 6,6 Mio. Artikeln auf etwa 20 GB, also weniger als ein halbes Prozent!).

Der ungeheure Datensatz, mit dem GPT 3.5 trainiert wurde, ist nicht nur ein quantitatives, sondern auch ein qualitatives Problem, denn eine solche Textmenge ist auch bei größtem Sammeleifer nur unter Verwendung von „Common Crawl" aufzubringen, also von maschinell gesammelten Internettexten, die im Trainingsmix für ChatGPT ca. 60%[5] ausmachen. Hier dürfte eine der Ursachen liegen, warum das Modell, sofern man es nicht durch Nachtraining oder gezieltes Prompting daran hindert, all die Vorurteile, Rassismen und Sexismen wiederholt, die leider nach wie vor einen Teil der Konversationen auf Websites und Internetforen ausmachen. GPT tut damit nichts anderes, als den Gebrauch in der**(334)** Sprache wiederzugeben – und der Menschheit einen recht unerfreulichen Spiegel ihrer Diskurse vorzuhalten (vgl. auch Catani 2023).

3 Sieben Thesen zu GPT

Die folgenden Ausführungen ziehen Konsequenzen aus der Arbeitsweise und den technischen Hintergründen aktueller großer Sprachmodelle. Sie sind auf (Chat-)GPT bezogen, lassen sich aber auf praktisch alle anderen derzeit verwendeten Sprachmodelle mit einer Transformer-ähnlichen Architektur übertragen. Die Aussagen sind bewusst thesenhaft zugespitzt, weil wir einen Diskurs anregen wollen, ohne im Rahmen dieses Beitrages bereits alle bedenkenswerten Argumente und Gegenargumente anführen zu können.

1. GPT ersetzt „Bedeutung" durch „Auftretenswahrscheinlichkeit"

Die Architektur und Arbeitsweise von GPT ermöglichen es, höchst differenzierte bedingte Wahrscheinlichkeiten für die topologischen Regelmäßigkeiten der im Datensatz vorkommenden Wörter zu errechnen. Das macht das Sprachmodell zu einer pragmatischen, aber nicht zu einer semantischen

[3] Sofern man den Trainingsdatensatz von GPT als repräsentativen Ausschnitt der Sprache akzeptiert, s. u.

[4] Über das zur Niederschrift dieses Beitrags aktuelle GPT 4.0 wurden von OpenAI mit Verweis auf Wettbewerbs- und Sicherheitsbedenken keine konkreten Zahlen und Fakten mehr veröffentlicht (OpenAI 2023: 2).

[5] Hinzu kommen weitere 22% Internettexte, deren inhaltliche Qualität immerhin insofern stärker sichergestellt wurde, als sie auf der amerikanischen Aggregations- und Bewertungsplattform Reddit.com mindestens dreimal positiv bewertet wurden. Der Rest der Trainingsdaten stammt aus zwei großen Buchkorpora (zusammen ca. 16%) sowie der internationalen Wikipedia (ca. 3%) (OpenAI 2020, 9).

Maschine: Das Netzwerk „weiß“ nicht, was es sagt, sondern lediglich, was nach Maßgabe des Trainingsdatensatzes im aktuellen Kontext mit hoher Wahrscheinlichkeit gesagt *würde* (eigentlich: *wurde*). Semantik kommt in den Parametern des Modells damit höchstens indirekt vor, nämlich dadurch, dass die Texte, an denen GPT trainiert wurde, von Menschen geschrieben wurden, die durch Auswahl und Anordnung von Wörtern und Stiftung grammatischer Beziehungen Bedeutung kodiert haben.

Es ist verblüffend genug, dass sich allein aus der wiederholten Errechnung bedingter Auftretenswahrscheinlichkeiten grammatisch wohlgeformte Texte ergeben. Aber dem Output von GPT Bedeutung beizumessen, entspricht dem logischen Fehlschluss, aus einem Konsequens q auf sein Antezedens p zu schließen. Dass menschliche Logik für solche ungültigen konditionalen Umkehrungen anfällig ist, ist spätestens seit den Arbeiten von Wason und Johnson-Laird (1972: 54–65) bekannt. Mit GPT erlangt der Fehlschluss jedoch insofern eine neue Brisanz, als sein Output so vertrauenerweckend menschlich wirkt.

Dass GPT die typischen Wortreihenfolgen bedeutungstragender menschlicher Texte kontextsensitiv imitieren und variieren kann, ist kein hinreichender Garant dafür, dass sein Output seinerseits wieder Bedeutung enthält. Das *kann* möglich sein (so wie es möglich ist, dass eine nasse Straße tatsächlich auf vorherigen Regen schließen lässt, vgl. ebd.) und ist umso wahrscheinlicher, je stärker sich Bedeutung in immer wieder ähnlichen sprachlichen Formen niederschlägt. Aber angesichts der aktuellen Architektur kann es niemals dazu führen, dass Sprachmodelle tatsächlich Bedeutungen lernen und abbilden – sie werden darauf schlicht nicht**(335)** trainiert. Die Bedeutung des Outputs von GPT entsteht nicht in der Maschine, sondern in ihren Anwender:innen.

2. GPT ist kein Speicher für Wissen, sondern ein Speicher für Phrasen

Indem GPT nichts als die bedingten Auftretenswahrscheinlichkeiten von Wörtern im Kontext anderer Wörter repräsentiert, produziert es in Reinform das, was Harry Frankfurt (1986) „Bullshit“ genannt hat: inhaltsleere, wahrheitsindifferente sprachliche Phrasen. Auf sie hereinzufallen und dem Sprachmodell Wissen zu unterstellen, hieße, dem phrasalen Charakter unserer Sprache auf den Leim zu gehen (vgl. Thesen 3 und 7).

Am augenfälligsten wird dieser Umstand an den teils eklatanten Defiziten in der logischen Kohärenz des Outputs von GPT. Befragt etwa nach dem sog. Flussüberquerungsrätsel antwortet das Modell mit derselben Bereitwilligkeit und demselben Lösungsweg unabhängig davon, welche Lebewesen eigentlich transportiert werden sollen: Ersetzt man im Rätsel den Wolf durch eine Kuh, schlägt GPT gleichwohl vor, zuerst die Ziege über den Fluss zu bringen, damit diese nicht von der Kuh gefressen würde. Dagegen nimmt es sorglos in Kauf, dass Kuh und Kohlkopf allein am Ufer verbleiben.

Das Beispiel ist eines von unzähligen in den Medien kursierenden Belegen für die logischen Schwierigkeiten von GPT (vgl. auch Weitz 2023a und 2023b) und weit mehr als die Kinderkrankheit einer noch jungen Technologie. Was GPT fehlt (und seiner Architektur nach auch fehlen muss), ist ein übergreifendes „mentales Modell“ der Welt, ein über die Auftretenswahrscheinlichkeiten der Wörter hinausweisendes und mindestens teilweise kontextunabhängiges System aus kognitiven Konzepten, die über spezifische Relationen wie „Teil–Ganzes“ oder „Ursache–Wirkung“ miteinander in Beziehung stehen. Deshalb kann GPT zwar das Flussüberquerungsrätsel aus den Trainingsdaten rekonstruieren und dank der kohärenzstiftenden Attention-Module auch den Wolf kontextsensitiv durch eine Kuh ersetzen; die eklatante Unsinnigkeit, die sich daraus ergibt, teilt sich dem Netzwerk hingegen nicht mit, weil es über kein mentales Konzept von Wolf und Kuh verfügt, das unabhängig vom gerade dominierenden Gebrauchskontext existiert. Grävemeyer (2023) schlussfolgert pointiert, GPT habe die „Logik gelernt, aber nicht verstanden“ (ebd. 117). „Verstehen“ ist kein Parameter derzeitiger Sprachmodelle.**(336)**

3. GPT schreibt im wahrsten Sinne des Wortes "durchschnittlich"

Die stetige Datenaggregation, die künstliche neuronale Netzwerke in jeder Verarbeitungsschicht vornehmen, ermöglicht ihnen die Extraktion und neuronale Repräsentation der jeweils stärksten Muster und Regelmäßigkeiten des Trainingsdatensatzes. Der Zufallsgenerator, der am Ende der Datenverarbeitung von GPT eines der wahrscheinlichsten, aber nicht zwingend das wahrscheinlichste Folgewort auswählt, erlaubt zwar eine gewisse Varianz des Outputs, das heißt jedoch lediglich, dass GPT *die* stärksten und nicht nur *das* stärkste sprachliche Muster unterstützt.

GPT löst sprachliche Aufgaben daher notwendigerweise so, wie der Durchschnitt derer sie gelöst hat, die den Inhalt des Trainingsdatensatzes produziert haben. Je stärker eine individuelle sprachliche Handlung von diesem Durchschnitt abweicht, umso unwahrscheinlicher ist es, dass GPT sie in seinen Parametern repräsentiert. GPT schreibt daher niemals kreativ, sondern epigonal: Es ist in der Lage, den durchschnittlichen Sprachgebrauch zu imitieren und die darin verwendeten typischen Muster zu variieren, aber es ist nicht in der Lage, neue Muster zu entwickeln. Es kann auf Befehl ein Märchen schreiben, in dem eine Maus, ein Zauberer und eine Steuerberaterin vorkommen, aber es kann nicht von selbst auf die Idee kommen, die Steuerberaterin in das Märchenmuster zu integrieren, weil seine Trainingsdaten ihm dazu keinen Anlass geben.

Die Repräsentation des sprachlichen Durchschnitts ist der möglicherweise sprachdidaktisch attraktivste Aspekt von GPT, denn damit liefert das Modell eine empirisch fundierte Antwort auf die Frage, wie „man“ ein Märchen schreibt, einen Vortrag gliedert oder eine Umfrage erstellt. Überall dort, wo diese Durchschnittslösung hinreicht, um ein wiederkehrendes sprachliches Problem zu bewältigen, liefert GPT brauchbare Vorlagen. Auch dort, wo es darum geht, sprachliche Muster und sprachbezogene Standardsituationen überhaupt erst zu erwerben (also dezidiert im sprachdidaktischen Kontext), kann das Modell seine Fähigkeiten ausspielen und Lernenden als „Sparringspartner“ dienen (vgl. These 6). Das Ungewöhnliche, Ausgefallene, Schöpferische hingegen, das menschlichen Sprachgebrauch mitunter über das Mittelmaß hinauswachsen lässt und sicherlich nicht wenig zur diachronen Dynamik der Sprache beiträgt, ist GPT aufgrund seiner induktiven Architektur fremd.

4. GPT zementiert den Mainstream

Durchschnittlich zu schreiben und durchschnittliche sprachliche Lösungen zu produzieren, ist kein Manko, sondern buchstäblich der Erwartungswert sprachlichen Handelns. Zum Problem werden sprachliche Durchschnitte allerdings, wenn sie fragwürdige Muster abbilden und/oder beginnen, auf sich selbst zurückzuwirken. Auf die klischee- und vorurteilsbekräftigenden Charakter des**(337)** Outputs von GPT war bereits hingewiesen worden; aber auch inhaltlich unverdächtige Sprachdurchschnitte können durch Selbstverstärkung auf Dauer einen systematischen Sog entwickeln und den Sprachgebrauch selbst beeinflussen.

Ein Beispiel: GPT wird bereits heute umfassend dazu verwendet, Texte zu „verbessern". Eine Suchanfrage in einer beliebigen klassischen Suchmaschine fördert hunderte Anleitungen und ähnlich viele kommerzielle Angebote zu diesem Behuf zutage. Was aber kann „Texte verbessern" angesichts der Fähigkeiten von GPT heißen? Eine Stichprobe mit einigen willkürlichen Textausschnitten bringt Erwartbares zutage: GTP löst komplexe Satzgefüge in Einzelsätze auf, wandelt Passiv- in Aktivkonstruktionen um und ersetzt seltenere durch geläufigere Formulierungen.

Erwartbar sind diese Änderungen insofern, als GPT seiner Arbeitsweise gemäß den eingegebenen Text Wort für Wort mit seinen Mustern vergleicht, und diejenigen Änderungen vornimmt, die nach Maßgabe seiner Parameter die wahrscheinlicheren Weiterführungen des Textes gewesen wären. Es misst damit den Vorlagentext am kontextsensitiven „Durchschnitt" seiner Modellparameter und nimmt Änderungen vor, die den Fortgang des Textes statistisch erwartbarer machen. Das kann in Fällen, in denen ungewöhnliche sprachliche Konstruktionen mehr aus Ungeschick denn aus sprachgestalterischer Absicht gewählt wurden, durchaus zu subjektiv erlebten Textverbesserungen führen, aber genauso gut zur sprachlichen Verflachung, sofern die gewählte Konstruktion aus einer gezielten gestalterischen Intention heraus gewählt wurde. GPT kann das eine vom anderen nicht unterscheiden, sondern lediglich den Grad der Abweichung vom Erwartungswert ermitteln. Ob diese Abweichung die Textqualität erhöht oder verringert, ist keine Frage, die sich aus den Parametern des Sprachmodells erschließen ließe.

Man muss kein Prophet sein, um vorauszusehen, dass ein zunehmender Teil zukünftigen Schriftsprachgebrauchs vor seiner Veröffentlichung einer kosmetischen Aufarbeitung durch GPT unterzogen werden wird. Für viele Texte mag das zu einer Qualitätssteigerung führen (nach welchem Maßstab auch immer), aber für die Gesamtheit der im textlichen Kosmos verwendeten sprachlichen Konstruktionen bedeutet es, dass die wahrscheinlicheren noch wahrscheinlicher und die unwahrscheinlicheren noch unwahrscheinlicher werden. GPT macht den Durchschnitt zum Standard, an dem sich jeder bedienen kann, aber dem zu folgen, den Standard selbst bekräftigt (vgl. These 5).

Eine Zementierung des statistisch Erwartbaren mag sich durchaus positiv auf die schriftsprachliche Kultur auswirken: Möglicherweise sind die sinnvollen Anwendungsfälle für hochkomplexe Satzgefüge, für idiosynkratische Formulierungen oder Passivkonstruktionen tatsächlich viel seltener als ihre tatsächlichen Anwendungen. Auch gewährleistet der sprachliche Mainstream mutmaßlich ein Maximum an Breitenwirkung und könnte sich damit dem kollektiven Textverständnis als dienlich erweisen. Unter der Hypothese hingegen, dass außergewöhnliche sprachliche Konstruktionen nicht zufällig, sondern bedeutungstragend**(338)** gewählt werden, könnte der Normierungsdruck des Sprachmodells beginnen, die textgestalterischen Potenziale der Sprache zu verengen.

5. Der Output von GPT droht, zur sich selbst erfüllenden Prophezeiung zu werden

Die mainstream-zementierende Tendenz von GPT gewinnt an exponentieller Brisanz, wenn man berücksichtigt, dass GPT gerade dabei ist, seine eigenen zukünftigen Trainingsdaten zu erzeugen. Seit der Einführung von ChatGPT berichten Technikmagazine wöchentlich von neuen Start-ups und Webservices, die GPT dazu verwenden, die Produktion unterschiedlichster Texte – vom einfachen Forumsbeitrag über den Geschäftsbrief bis zur wissenschaftlichen Abhandlung – zu erleichtern, zu unterstützen oder gleich ganz zu übernehmen (vgl. Gieselmann 2023). Bei anspruchsloseren Gebrauchstextsorten, etwa News-Tickern oder Spam-Nachrichten, ist dieses Geschäftsmodell bereits seit einigen Jahren etabliert; mit dem erheblichen Qualitätssprung seit dem Durchbruch von ChatGPT gewinnt es hingegen einen Umfang, von dem sich berufsmäßig Textschaffende zunehmend bedroht sehen (vgl. etwa Hammer 2023).

Unabhängig von den wirtschaftlichen Implikationen droht der Erfolg maschinell erzeugter Texte aber auch in schwer vorhersagbarer Weise auf sich selbst zurückzuwirken, denn der Output von GPT erzeugt durch seine Anwendung einen immer größeren Anteil des zukünftigen „Common Crawl" (s.o.) und damit des sprachlichen Materials, mit dem zukünftige Sprachmodelle trainiert werden. Dadurch wird GPT in zukünftigen Trainings seine eigenen Muster und Regelmäßigkeiten zunehmend von den Trainingsdaten bestätigt sehen und misst folglich dem selbst erzeugten Durchschnitt notwendig immer höhere Wahrscheinlichkeiten bei, während es allen Formen der sprachlichen Abweichung immer geringere Chancen gibt. Im Extremfall könnte das dazu führen, dass ein immer kleineres Repertoire sprachlicher Muster ständig den modellinternen Wahrscheinlichkeitswettbewerb gewinnt und damit den Output von GPT dominiert.

Die längerfristigen Resultate einer solchen systematischen Selbstverstärkung könnten sehr unterschiedlich ausfallen. Denkbar wäre etwa, dass sich die Sprachgemeinschaft den Normierungstendenzen der Sprachmodelle unterwirft und jedwede Abweichung vom sprachlichen Mainstream als zunehmend fremdartig oder antiquiert wahrnimmt. Unter dieser Annahme geriete auch die menschliche Textproduktion immer stärker unter Normierungsdruck, weil die Wahl außergewöhnlicher sprachlicher Konstruktionen Gefahr liefe, nicht nur stilistisches Befremden, sondern regelrechte Verständnisschwierigkeiten zu verursachen, wie sie heute etwa bei der Verwendung von Archaismen auftreten mögen. Sie werden gewissermaßen durch Nichtgebrauch aus der Sprache gedrängt.**(339)**

Ebenfalls denkbar wäre freilich, dass sich die Sprachgemeinschaft die Lust an der sprachlichen Varianz keineswegs so leicht austreiben lässt und daher nicht die subjektive sprachliche Abweichung, sondern den standardisierten Output von GPT als zunehmend befremdlich oder langweilig erlebt. Das Resul-

tat einer solchen Entwicklung könnte eine neue Wertschätzung menschlich verfasster, lektorierter und kuratierter Texte sein, während sich der Einsatz maschinell erzeugter Schriftsprache auf textgestalterisch anspruchslose Anwendungsfälle verengt, in denen die Verwendung immer gleicher sprachlicher Phrasen unproblematisch bis erwünscht ist . In diesem Szenario hielte sich auch die selbstverstärkende Wirkung der Sprachmodelle in Grenzen, weil auch zukünftige Trainingsdaten stets ein bestimmtes Maß sprachstruktureller Varianz behalten würden.

Wie sich der Sprachgebrauch und mit ihm GPT entwickeln werden, ist derzeit nicht absehbar und zweifellos von vielen Faktoren abhängig. Fest steht indes, dass GPT aufgrund seiner Funktionsweise und der Herkunft seiner Trainingsdaten eine Tendenz zur sich selbst verstärkenden Standardisierung aufweist, die auf Dauer umso stärker durchschlagen wird, je umfänglicher das Modell zur Textproduktion eingesetzt wird.

6. GPT kann den Erwerb von Schreibkompetenz unterstützen – oder ersetzen

Fast zeitgleich mit der Veröffentlichung von ChatGPT trat die Frage auf, welche Konsequenzen die neuen maschinellen Fähigkeiten für die Vermittlung und den Erwerb von Sprach- und insbesondere Textkompetenz haben (vgl. Kragl 2023, Catani 2023, Rödel 2023, Maiwald 2023). Dass die Debatte dabei zunächst vornehmlich um das Problem von Plagiat und Kompetenzvortäuschung kreiste, war angesichts der Plötzlichkeit der neuen Situation nur verständlich. Inzwischen sind die Fragen vielfältiger und die Antworten differenzierter geworden. Sie umfassen neben den Gefahren längst auch eine Diskussion der Chancen, die große Sprachmodelle für den Erwerb sprachlicher Kompetenzen eröffnen (vgl. ebd.). Worin können diese aus der Perspektive der Arbeitsweise von GPT bestehen?

Angesichts der architekturbedingten semantischen Beschränktheit von GPT (s. Thesen 1 und 2) ist es sicher keine gute Idee, dem Modell die Lösung inhaltlicher Probleme zu überlassen. GPT kann zutreffende von nicht zutreffenden sprachlichen Aussagen nur insofern unterscheiden, als Erstere (hoffentlich) eine höhere Auftretenswahrscheinlichkeit in den Trainingsdaten haben und Bedeutungen zudem eine gewisse Tendenz aufweisen, sich in einer begrenzten Anzahl wiederkehrender sprachlicher Muster auszudrücken (vgl. These 7). Das ist für**(340)** seriöse pädagogisch-didaktische Arbeit in jedem Fall zu wenig und wird sich ohne tiefgreifende Veränderungen der eingesetzten Netzwerkarchitekturen auch zukünftig nicht maßgeblich ändern lassen.

Das Feld, in dem die Fähigkeiten von GPT für den Erwerb von Schreibkompetenz nutzbar gemacht werden kann, bleibt also die sprachliche Form. Das Modell erzeugt Texte, die in hohem Maße sprachlichen Konventionen genügen – angefangen bei einer standardkonformen grammatischen und orthografischen Umsetzung über die Bedienung typischer Textsortenmerkmale bis hin zur Wahl passender sprachlicher Register. GPT verfügt damit über „Fähigkeiten“, die Sprachlernende in ihrer individuellen schriftsprachlichen Sozialisation mühsam und über viele Jahre hinweg erwerben müssen. Die sprachdidaktisch interessante Frage ist damit, auf welche Weise GPT diesen Prozess unterstützen könnte.

Dazu ist zunächst zu berücksichtigen, dass auch menschliche Schreibkompetenz zu großen Teilen induktiv erworben wird – zwar wesentlich effizienter als über den bruteforce-Ansatz des maschinellen Lernens (s.o.), aber doch insofern strukturell ähnlich, als sprachliches Lernen vor allem über die rezeptive und produktive Auseinandersetzung mit Sprache selbst erfolgt, erst sekundär über die Anwendung deduktiver Regeln (vgl. etwa Marx 2006). Mit anderen Worten: Um Schreibkompetenz zu erwerben, müssen Lernende sich vor allem schriftsprachlich mit Texten auseinandersetzen und zur Qualität ihrer Schreibprodukte ein möglichst schnelles, differenziertes und günstigstenfalls didaktisch aufbereitetes Feedback erhalten.

GPT kann in diesen Erwerbsprozess grundsätzlich auf zweierlei Weise eingebunden werden, nämlich einerseits als digitaler „Sparringpartner“, der den Lernenden Rückmeldungen zu ihren persönlichen tetxgestalterischen Lösungsansätzen zur Verfügung stellt, oder andererseits als textgenerierender „Taschenrechner“, der die Textgestaltung nicht begleitet, sondern (in Teilen) übernimmt.

Im Sparringpartner-Modell würde GPT dazu eingesetzt, um Lernenden zu ihren Schreibprodukten Rückmeldung zu geben und so einen Maßstab zur Verfügung zu stellen, aus dem sie sich ein Bild über die Vorzüge und Probleme ihrer eigenen Texte machen können. Darüber hinaus experimentieren verschiedene didaktisch orientierte Projekte bereits damit, Sprachmodelle zur direkten didaktischen Beurteilung menschlich erzeugter Texte einzusetzen.[6] Der didaktische Vorzug gegenüber klassischem Schreibunterricht liegt dabei darin, dass Schreibende nicht auf die Rückmeldungen der anleitenden Lehrkraft warten müssen,**(341)** sondern für jedes textgestalterische Problem unmittelbare digitale Hilfe anfordern können, die obendrein auf den je aktuellen Problemkontext zugeschnitten ist. Die Entwicklung von Schreibkompetenz nach dem Sparringpartner-Modell würde folglich ganz konventionell mit persönlichen schriftsprachlichen Lösungsversuchen beginnen, die der Maschine im Anschluss zur formalen Überarbeitung oder Beurteilung unterbreitet oder mit einer Musterlösung verglichen würden, um in der reflektierten, kritischen Auseinandersetzung eigene Lösungen infrage zu stellen, überzeugende Änderungsvorschläge anzunehmen und weniger überzeugende zurückzuweisen.

Wie wahrscheinlich ist es, dass GPT in nennenswertem Umfang in der geschilderten Weise eingesetzt werden wird? Mutmaßlich so wahrscheinlich wie das Szenario, den Taschenrechner im Mathematikunterricht nur zur Überprüfung der händisch berechneten Lösungen zu verwenden. Noch liegen kaum belastbare Daten zum Nutzungsverhalten Lernender im Umgang mit GPT vor, aber die Vermutung liegt nahe, dass viele das Sprachmodell nicht erst post hoc zum Vergleich mit eigenen Lösungsversuchen einsetzen werden, sondern sich umgehend eine Musterlösung ausgeben lassen, um diese im Anschluss mehr oder weniger stark zu überarbeiten. Trifft diese

[6] So etwa https://peer.edu.sot.tum.de oder https://www.fiete.ai, die mithilfe neuronaler Sprachmodelle Rückmeldungen zu Textproduktionen geben. Eine zentrale Herausforderung für automatisiertes Feedback auf inhaltlicher Ebene ist das fehlende Weltwissen bzw. Verständnis des Modells für den In- als auch den eigenen Output (s. These 2), also für den zu bewertenden Text (das kann neben dem Text selbst auch die Aufgabe und ggf. vorhandenes Material sein) als auch für die produzierte Rückmeldung. Die Komplexität dieses Urteils skizziert Rödel (2023).

Vermutung zu, wird GPT zum textlichen „Taschenrechner", der den produktiven Teil des Arbeits- und Erwerbsprozesses zugunsten des rezeptiven Teils verringert, da der zu erstellende Text nicht mehr geschrieben, sondern nur noch lesend beurteilt und allenfalls überarbeitet werden muss. Eine solche Verschiebung ist für alle, die bereits über ausgeprägte Schreibkompetenzen verfügen und für die der Prozess der Niederschrift daher lediglich die Ausführung einer mehr oder weniger lästigen sprachlichen Standardroutine darstellt, mutmaßlich unproblematisch und eine willkommene Arbeitserleichterung, da sie den vorgeschlagenen Text mit ihrem mentalen Konzept eines guten Textes abgleichen und entsprechend verändern können. Für diejenigen hingegen, die sich mit der sprachlichen Aufgabe vor allem zum Zweck des Kompetenzerwerbs auseinandersetzen, bedeutet es eine starke Veränderung – und angesichts der geringeren kognitiven Verarbeitungstiefe wohl auch Verflachung – des Lernprozesses, deren Auswirkungen aktuell schwer abschätzbar sind. Wer einen Taschenrechner in der Mathematik einsetzt, statt schriftlich oder im Kopf zu rechnen, verliert zweifellos Übung im Rechnen und gewinnt dafür Arbeitszeit, die er für andere, anspruchsvollere mathematische Aufgaben einsetzen kann. Eine ähnliche Kompetenzverschiebung ist auch zu erwarten, wenn Lernende ihre Texte nicht mehr selbst schreiben, sondern lediglich den maschinellen Schreibprozess von Sprachmodellen steuernd und kontrollierend begleiten (vgl. Rödel 2023 und Ott 2023). Allerdings ist gutes Kopfrechnen weder eine notwendige noch eine hinreichende Bedingung für den Erwerb höherer mathematischer Kompetenzen und kann daher möglicherweise tatsächlich ohne nennenswerte Verluste an eine Maschine delegiert werden (ebd.). Dagegen scheint es**(342)** uns für den sprachlichen Kompetenzerwerb alles andere als ausgemacht, dass die routinierte produktive Lösung sprachlicher Standardaufgaben ähnlich entbehrlich ist wie das Kopfrechnen in der Mathematik.

Wie sehr und wie gewinnbringend GPT den schriftsprachlichen Kompetenzerwerb beeinflussen wird, hängt stark von den didaktischen Anwendungsszenarien ab. Überall dort, wo der Output des Sprachmodells zum Vergleich und zur kritischen Reflexion des persönlichen Sprachgebrauchs herangezogen wird, kann sein Einsatz vermutlich vorbehaltlos empfohlen werden. Überall dort hingegen, wo er die Textproduktion ganz oder in Teilen ersetzt, läuft er mindestens Gefahr, den Kompetenzerwerb ebenfalls eher zu ersetzen als zu unterstützen.

7. Für die Philologien ist GPT in analytischer Hinsicht interessanter als in produktiver

Was stellen wir (als Sprachgemeinschaft, als Philologie, als Individuen) mit der neuen Maschine an, die unseren Sprachgebrauch so adäquat modellieren kann, dass sie gut und gerne den Turing-Test besteht? Was hat sich mit der Existenz von ChatGPT und Co. verändert und worin besteht das Revolutionäre, Bahnbrechende und Neue, das im ersten Rausch der Verblüffung in Foren und Feuilletons ausgerufen wurde? Und steht wirklich das „Ende vom Lernen [sic!] wie wir es kennen" (Blume 2023) vor der Tür?

Schon der Abstand einiger Monate hat den Anfangsschock relativiert und die überreizte Debatte spürbar abgekühlt. GPT ist keine starke Intelligenz, es hat kein Bewusstsein entwickelt (Rüger 2022) und seine textgestalterischen Fähigkeiten stoßen, wenn man sie mit denen natürlicher Personen vergleicht, doch eher früh als spät an ihre Grenzen.

GPT ist kein Sprach-, sondern ein Sprach*gebrauchs*automat, dessen Output semantisch ernst zu nehmen man sich tunlichst hüten sollte. Möglicherweise glücken die derzeitigen Versuche der großen Tech-Unternehmen wie Microsoft, Google und Co., GPT mit klassischer Suchmaschinentechnik zu vermählen, dem Sprachmodell damit die Verantwortung für die Inhalte zu entziehen und es nur noch für die sprachliche Aufbereitung der Ergebnisse zu verwenden. Das könnte zu einem deutlichen Zugewinn an Komfort für die Wissensgesellschaft führen, aber doch kaum zur Ausrufung eines revolutionären Neubeginns.

Als wirklich revolutionär erweisen sich GPTs Fähigkeiten unseres Erachtens eher in Bezug auf die Sprach*theorie* und damit auf Fragen, die im aktuellen Diskurs bemerkenswert selten gestellt werden. Wie rasch etwa haben wir uns daran gewöhnt, dass der Output von GPT praktisch durchgängig grammatisch wohlgeformt ist, ohne dass in seine Architektur auch nur eine einzige grammatische Regel implementiert wurde! Haben wir uns eigentlich bereits genug darüber gewundert, dass GPT damit ein System repräsentiert, um das die moderne Linguistik buchstäblich Jahrzehnte lang gerungen hat, ohne zu einem befriedigenden**(343)** Ende zu kommen – ein System nämlich, das praktisch ausschließlich grammatisch und orthografisch wohlgeformte Sätze erzeugt?

Haben wir bereits hinreichend ermessen, was es über unsere Sprache aussagt, dass sich aus der statistischen Repräsentation von Wortreihenfolgen Texte generieren lassen, die mindestens bedeutungstragend *scheinen* (und nicht nur unsinnige Wortfolgen wie Chomskys oft zitierte „farblose grüne Ideen", ebd. 1957, 15)? Übersehen wir in der Diskussion der eklatanten logischen Fehler von GPT vielleicht gerade, in wie vielen Anwendungssituationen die statistische Repräsentation von Wortreihenfolgen hinreicht, um so etwas wie „Sinn" zu erzeugen? Was bedeutet es für die Sprachtheorie, dass das topologische Blendwerk nicht schon im ersten Nebensatz auffliegt, sondern erst bei eingehender Tuchfühlung? Und dass es sogar häufig genügt, um nichttriviale sprachliche Probleme angemessen zu lösen?

Am Anfang dieses Beitrages haben wir Wittgensteins Gebrauchstheorie der Bedeutung zitiert und die These aufgestellt, dass GPT zur Angemessenheit dieser Theorie ein neues Argument eingebracht habe. Sehr bewusst haben wir uns zunächst gehütet, dieses Argument näher zu spezifizieren, ja wir haben noch nicht einmal Stellung dazu bezogen, ob die Fähigkeiten von GPT eigentlich für oder gegen Wittgensteins Auffassung sprechen.

Dezidiert wiederholen wir zum Schluss dieser Auseinandersetzung aber die Überzeugung, dass GPT in diesen Diskurs ein relevantes und bedenkenswertes Argument einbringt: Das Modell liefert eine (vorläufige und interpretationsbedürftige, aber empirisch stark fundierte) Antwort auf die Frage, wie weit man semantisch eigentlich kommt, wenn man nur die Topologie des Sprachgebrauchs statistisch repräsentiert. Ob diese Antwort als „bemerkenswert weit" oder eher als „längst nicht weit genug" zu verstehen ist, lässt sich in einem Beitrag dieses Umfangs nicht sinnvoll erörtern. Dass sie hingegen sprachtheoretisch ebenso relevant wie interessant ist, wagen wir zu unterstellen und wünschen uns und der Sprachgemeinschaft eine rege Debatte.

Literatur

[1] Blume, B. (2023). *ChatGPT. Das Ende vom Lernen wie wir es kennen.* Online verfügbar unter https://deutsches-schulportal.de/kolumnen/chatgpt-das-ende-vom-lernen-wie-wir-es-kennen/ [Zugriff: 10.08.2023].

[2] Catani, Stephanie (2023): „Mit KI schreiben – über KI schreiben. Künstliche Intelligenz als Thema im literaturwissenschaftlichen Studium". *Mitteilungen des Deutschen Germanistenverbandes* 70 (4), S. 393–405. https://doi.org/10.14220/mdge.2023.70.4.393

[3] Chomsky, N. (1957). *Syntactic Structures.* Den Haag/Paris.

[4] Chomsky, N. (1969). Some empirical assumptions in modern philosophy of language. In E. Nagel, S. Morgenbesser, P. Suppes & M. G. White (Hrsg.), *Philosophy, Science, and Method* (S. 260–285). New York: St. Martin's Press.

[5] Fodor, J. A., & Katz, J. J. (1963). The availability of what we say. *The Philosophical Review, 72*(1), 57–71.

[6] Frankfurt, H. G. (1986). *On Bullshit.* Princeton, NJ: Princeton University Press.

[7] Gieselmann, H. (2023). Wer soll das alles lesen? KI-Textgeneratoren überschwemmen das Internet. *c't, 5*, 64–68.

[8] Grewendorf, G. (1985). Sprache als Organ und Sprache als Lebensform. Zu Chomskys Wittgenstein-Kritik. In D. Birnbacher & A. Burkhardt (Hrsg.), *Sprachspiel und Methode* (S. 89–129). Berlin / New York: de Gruyter.

[9] Grävemeyer, A. (2023). KI-Sprachgeneratoren: Wie man sie von Menschen unterscheiden kann. *c't, 16*, 116–119.

[10] Hammer, P. (2023). Nachgefragt: Killt KI Arbeitsplätze in Agenturen? Online verfügbar unter https://www.wuv.de/Themen/Kreation-Design/Nachgefragt-Killt-KI-Arbeitsplaetze-in-Agenturen [Zugriff: 10.08.2023].

[11] Hinzen, W. (2004). Zum gegenwärtigen Stand der Gebrauchstheorie der Bedeutung. In A. Fuhrmann & E. J. Olsson (Hrsg.), *Pragmatisch denken* (S. 59–88). Berlin / Boston: de Gruyter.

[12] Kragl, Florian (2023): „Zwischen historischem Korpuswissen und literarästhetischer Kompetenz. Das Fach Germanistik/Deutsch im Zeitalter von Digitalität und Künstlicher Intelligenz". *Mitteilungen des Deutschen Germanistenverbandes* 70 (4), S. 346–358. https://doi.org/10.14220/mdge.2023.70.4.346

[13] Maiwald, Klaus (2023): „Digitale Textkompetenz als Bildungsaufgabe in einer Kultur der Digitalität und künstlichen Intelligenz". *Mitteilungen des Deutschen Germanistenverbandes* 70 (4), S. 359–372. https://doi.org/10.14220/mdge.2023.70.4.359

[14] Marx, E. (2006). Profitiert das kindliche Sprachsystem von anderen kognitiven Entwicklungsbereichen? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*(3), 139–145.

[15] OpenAI (2020). Language Models are Few-Shot Learners. *arXiv*: 2005.14165.

[16] OpenAI (2023). GPT-4 Technical Report. *arXiv*: 2303.08774.

[17] Ott, Christine (2023): „Bildung in der digitalen Welt: Rückwirkungen generativer künstlicher Intelligenzen auf den Deutschunterricht". *Mitteilungen des Deutschen Germanistenverbandes* 70 (4), S. 382–392. https://doi.org/10.14220/mdge.2023.70.4.382

[18] Rödel, Michael (2023): „ChatGPT und Textkompetenz: Wie sieht die Zukunft des Schreibens in der Schule aus?" *Mitteilungen des Deutschen Germanistenverbandes* 70 (4), S. 373–381. https://doi.org/10.14220/mdge.2023.70.4.373

[19] Rüger, K. (2022). Die Google-KI LaMDA: Hat sie ein Bewusstsein entwickelt? Online verfügbar unter https://www.wipub.net/die-google-ki-lamda-hat-sie-ein-bewusstsein-entwickelt/ [Zugriff: 10.08.2023].

[20] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.

[21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *arXiv*: 1706.03762.

[22] Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content.* London: B. T. Batsford.

[23] Weitz, E. (2023a). ChatGPT und die Mathematik [Video]. Online verfügbar unter https://www.youtube.com/watch?v=medmEMktMlQ [Zugriff: 10.08.2023].

[24] Weitz, E. (2023b). ChatGPT und die Logik [Video]. Online verfügbar unter https://www.youtube.com/watch?v=5cYYeuwYF_0 [Zugriff: 10.08.2023].