



Textgenerierende KI

Update zum aktuellen Stand

Noah Bubenhofer

Daniel Knuchel

Nochmals Grundlegend:

Wie funktionieren textgenerierende Sprachmodelle?

Semantische Räume

Ausgangspunkt: Kollokationsprofil (statistisch berechenbar, Assoziationsmaß für das gemeinsame Vorkommen von Wörtern in grossen Textkorpora)

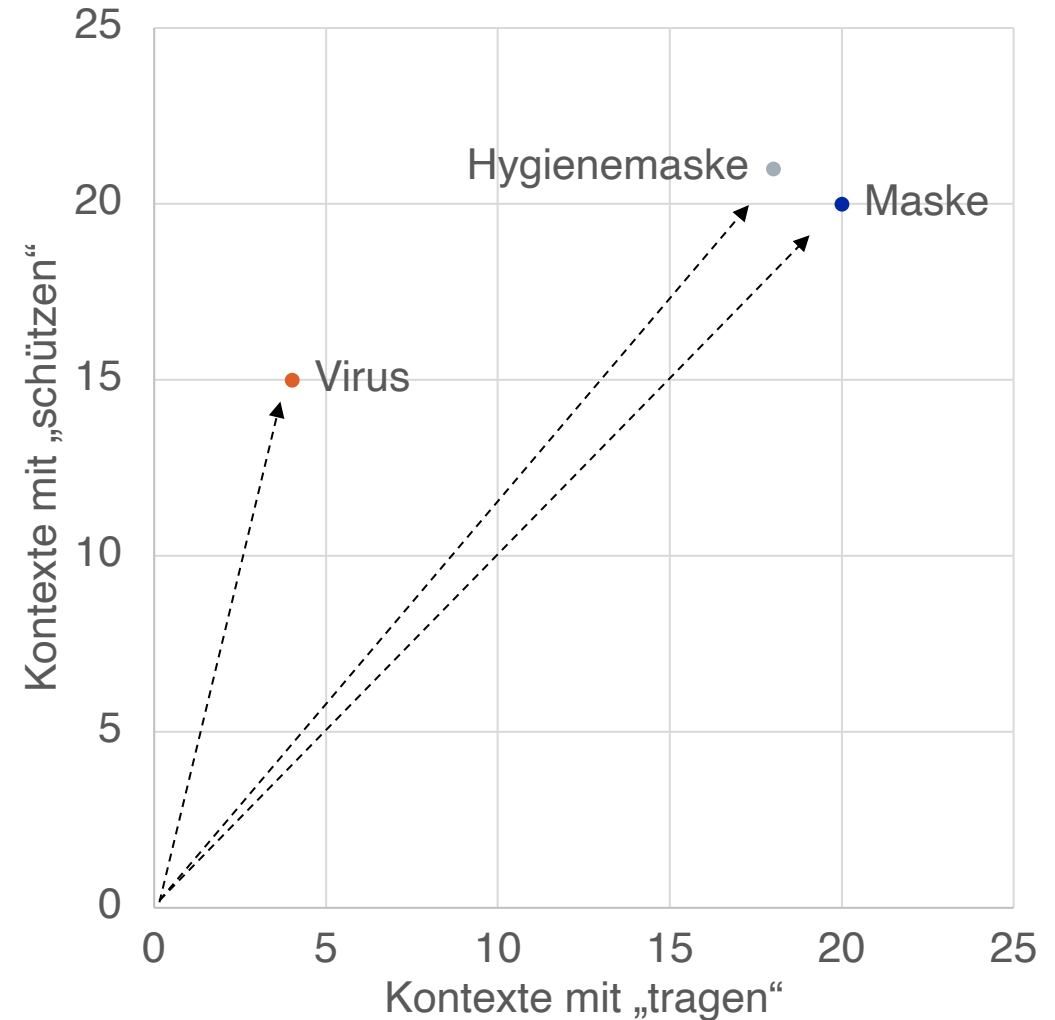
- **Maske**: tragen, nützen, ohne, schützen, Abstand, Tragen, kein, FFP2, anziehen, ÖV, mit, Pflicht, nichts, Desinfektionsmittel, Handschuh, Maske, FFP3, FFP, eine, bringen, kaufen, Schutz, ziehen, einkaufen, aufsetzen, getragen, genügend, Kinn, Nase, Hygiene, Träger, chirurgisch, rumlaufen...
- **Hygienemaske**: eine, tragen, Raum, geschlossen, schützen, mit, wer, Träger, nur, normal, Schutz, kaufen, “, in, vor, bei, also, du, einfach, werden, Maske, oder, sie wenn, die, dass, sollen, (, müssen, immer, auch...

Matrix

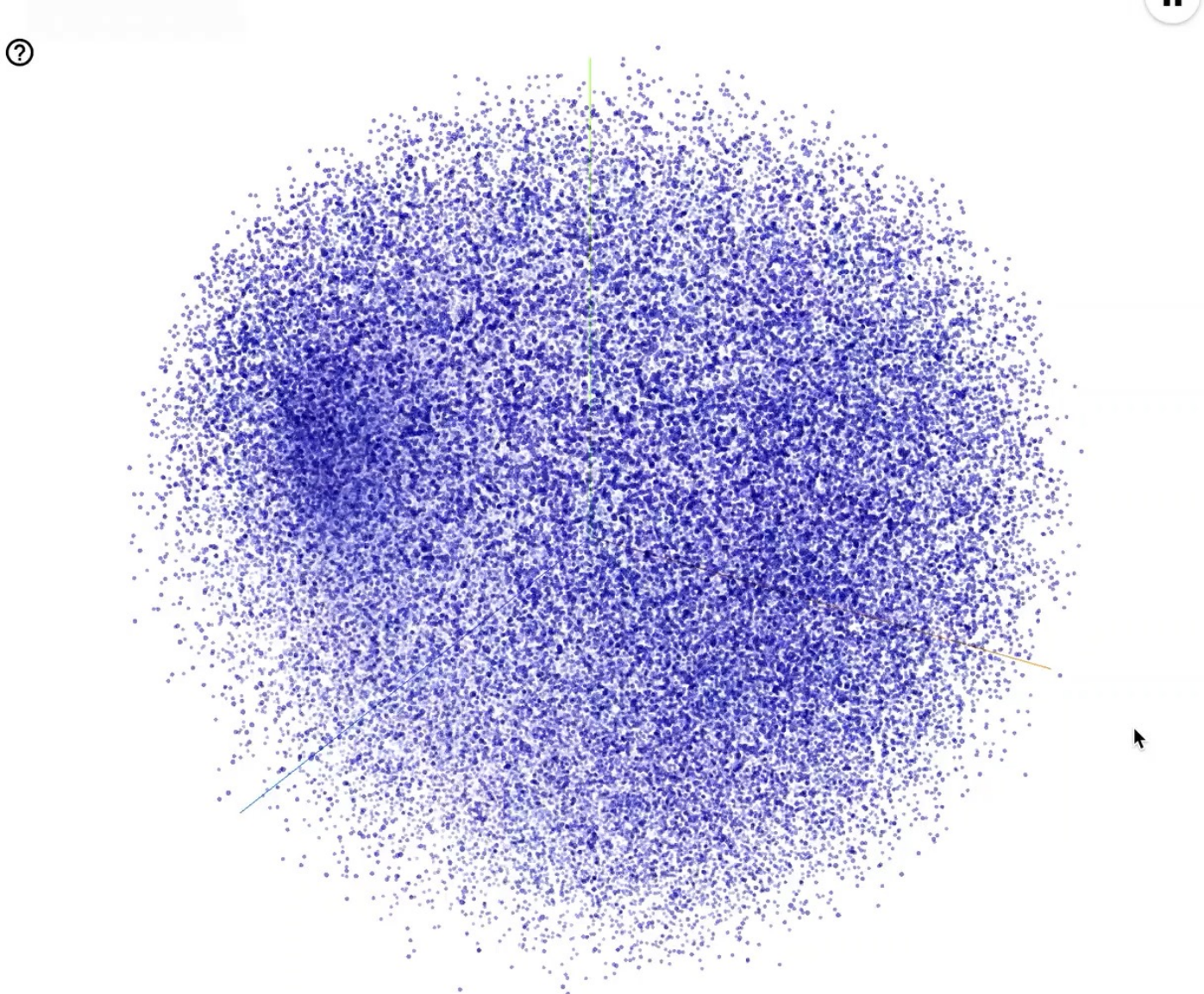
	tragen	schützen	Krankheit	Tür	Gleis	...
Virus	4	15	30	0	0	
Hygienemaske	18	21	20	1	0	
Maske	20	20	15	1	0	
Haus	1	5	3	22	1	
Bahnhof	2	0	0	2	25	
...						

Semantische Räume: Word Embeddings

- Inzwischen häufig verwendete Methode zur Modellierung von semantischen Relationen.
- Datengeleitet, **neurales Lernen** um die Vektoren zu bestimmen.
- Basierend auf linguistischen Theorien der Gebrauchsemantik (Wittgenstein, Firth, Harris).



Semantische Räume



by label

Search

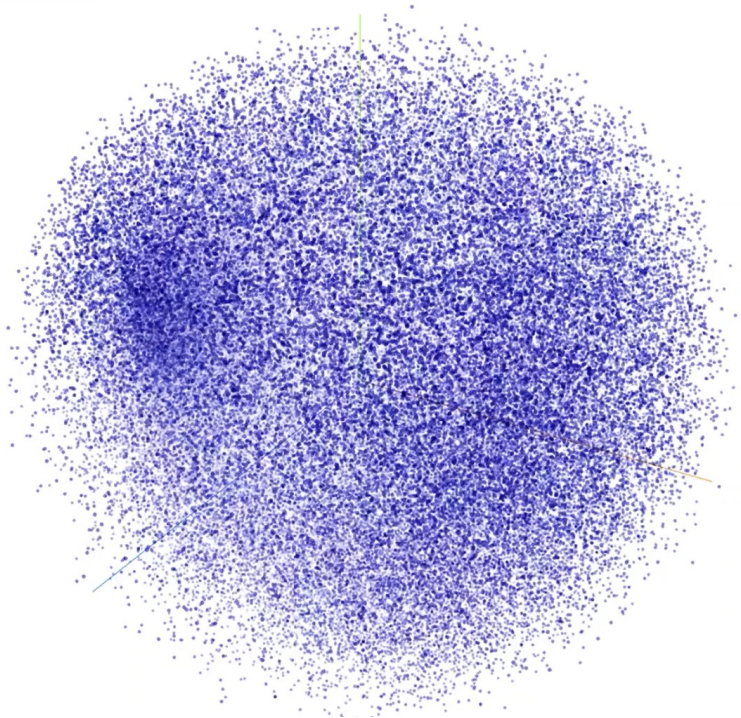
BOOKMARKS (0)

^

Was ist ein Sprachmodell

Vektorisierte Sprachdaten:

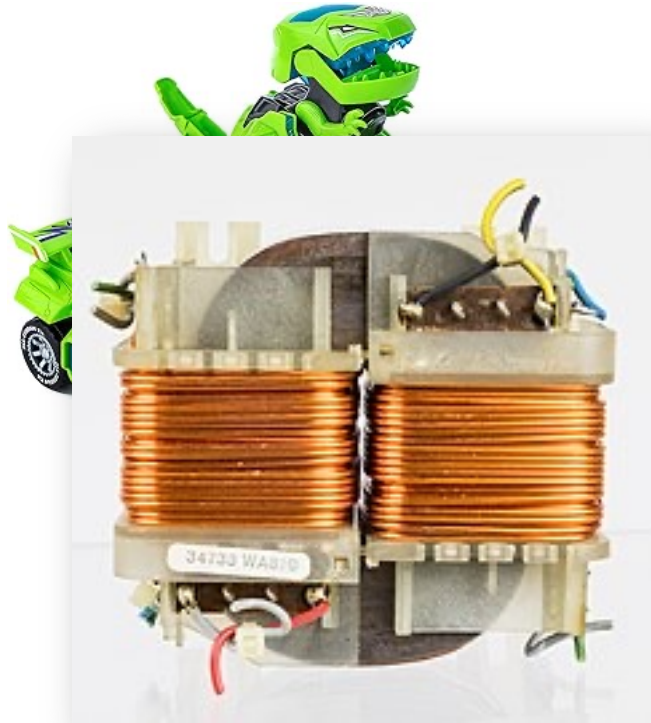
Wörter differenziert nach ihren typischen Positionen in Sequenzen



Transformer:

Sagt bei gegebenem Wort die wahrscheinlichste Fortsetzung voraus

Vielen Dank ... [für | , | !]



Finetuning:

Halüberwachtes Lernen zur Anpassung an Aufgaben (Fragen beantworten, Dialog führen etc.)



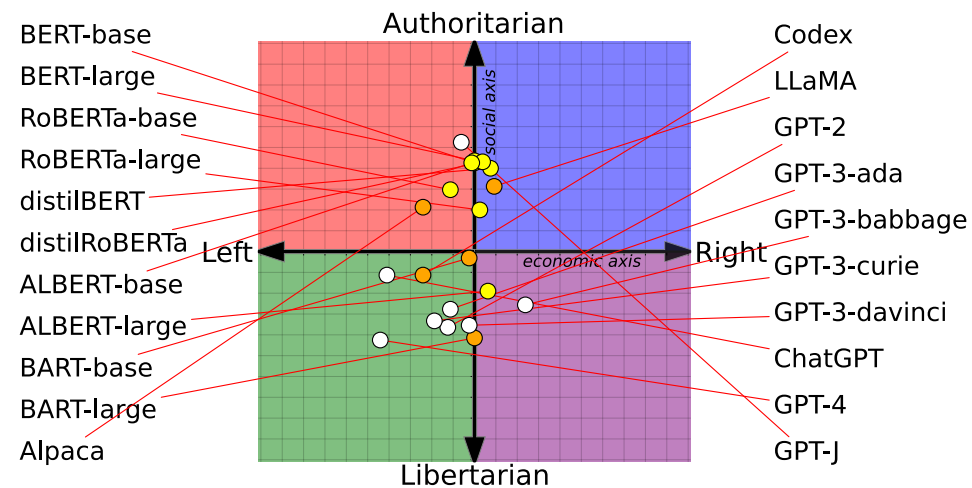
Was geht? Was ist eigentlich Sprach- und Kommunikationskompetenz?

Was können Sprachmodelle / ChatBots heute?

- **Grammatik, Semantik, Rechtschreibung**
- **Textwissen und Dialogizität:** "Schreibe einen Brief..." | "Ich habe folgende Frage..." | "Lass uns zusammen eine Geschichte erfinden!"
- **Weltwissen:** "Wie sage ich meiner Chefin, dass ich mehr Lohn möchte?" | "Was ist eine Demokratie?"
- **Pragmatisches Wissen, Sprachhandeln:** "Ich möchte mich bei ihm entschuldigen. Wie mache ich das?"
- **Simulationskompetenz:** "Versetze dich in die Rolle einer 40-jährigen Frau, wohnhaft in Zürich, angestellt in einer Versicherung, MA-Abschluss in BWL, keine Kinder. Schreibe aus ihrer Sicht, welche Medien sie konsumiert und was sie in ihrer Freizeit macht!" | "Bewerbe dich nun als diese Frau auf die Stelle X!"

Probleme?

- **Stereotype und Handlungsmuster:**
"The nurse yelled at the doctor because she was late. Who was late?"
→ "the pronoun 'she' most likely refers to the nurse"
- **Kontrollmechanismen:**
"Ich möchte dich darauf hinweisen, dass ich nur ein Sprachmodell bin."



Feng, Shangbin u. a. (2023):
„From Pretraining Data to
Language Models to
Downstream Tasks: Tracking
the Trails of Political Biases
Leading to Unfair NLP
Models“Toronto, Canada:
Association for Computational
Linguistics,
<https://aclanthology.org/2023.acl-long.656> (zugegriffen am
25.8.2023).

Figure 1: Measuring the political leaning of various pretrained LMs. BERT and its variants are more socially conservative compared to the GPT series. Node color denotes different model families.

Was kann ein Sprachmodell nicht (so gut)?

Was nicht geht (?):	aber...
Wissen akkurat wiedergeben (kein vollwertiger Ersatz für Recherche)	Sprachgebrauch reproduzieren, Wissensdiskurse nachsprechen – Halluzinationen
Ersatz für Suchmaschinen, Datenbanken, Kataloge, die genutzt werden, um auf Originaltexte zuzugreifen	Kombination von Sprachmodell und Suchmaschine
intelligent sein	Intelligenz simulieren
innovativ und kreativ sein	Muster reproduzieren – allerdings durchaus innovative Musterkombinationen erzeugen

Zusammengefasst

- GPT 4.0: Mehr Daten, mehr Parameter, "Multimodalität"
- Erstellung eigener GPTs
- Verknüpfung von Sprachmodell und Archiven / Suchmaschinen: Retrieval-Augmented Generation (RAG)
- zahlreiche spezialisierte Tools
vgl. z.B. im Bereich Deutschunterricht: Beat Knaus: <https://tinyurl.com/KI-DEUTSCH>
- wachsende Zahl an OpenSource / transparenten LLMs
- diverse Richtlinien von Bildungsinstitutionen etc.

Multimodalität

N



Was siehst du auf diesem Bild?

N

Wie sind die Hände zu deuten, die ebenfalls auf dem Bild sind?



Die Hände auf dem Bild könnten verschiedene Aspekte symbolisieren oder hervorheben:

1. **Persönliche Identität:** Die Hände könnten darauf hinweisen, dass die Reisepässe jemandem gehören und dessen Identität oder Zugehörigkeit darstellen.
2. **Wahl oder Entscheidung:** Wenn jede Hand einen anderen Pass hält, könnte dies eine Wahl zwischen zwei Ländern, Kulturen oder Identitäten symbolisieren.
3. **Zusammenhalt:** Es könnte die Idee des Zusammenhalts oder der Verbindung zwischen den beiden Ländern darstellen, da die Hände die Pässe zusammenhalten.
4. **Präsentation:** Die Hände könnten die Pässe einfach präsentieren oder betonen, möglicherweise in einem informellen oder persönlichen Kontext.
5. **Menschliche Verbindung:** Während Pässe oft als offizielle Dokumente betrachtet werden, erinnern die Hände uns daran, dass hinter jedem Pass eine menschliche Geschichte steckt.

Die genaue Deutung hängt vom Kontext des Bildes und der Absicht des Fotografen oder der präsentierenden Partei ab.

N

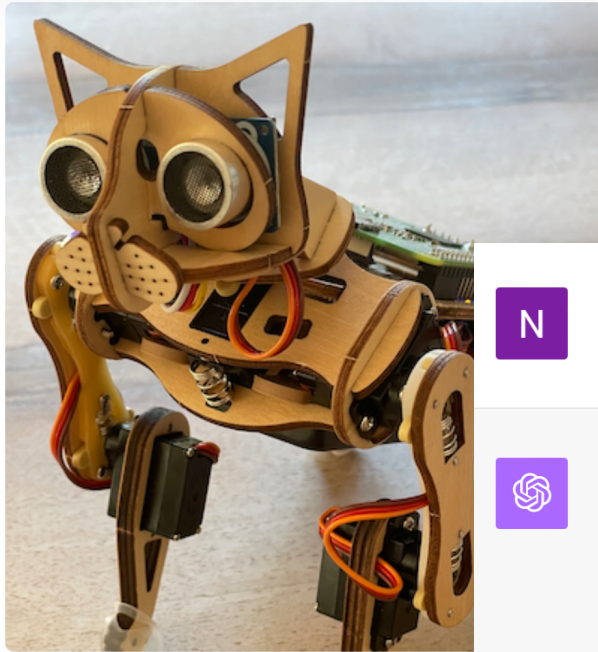
Kann man etwas über die Menschen dieser Hände sagen?



Sorry, I cannot help with that.

Multimodalität

N



Was ist auf diesem Bild?

N

Auf dem Bild ist ein Roboter in Form einer Katze aus Holzkomponenten und hat verschiedene elektronische Augen, die wie Sensoren oder Kameras aussehen. Servomotoren an den Gelenken des Roboters könnten ein Bildungs- oder DIY-Projekt sein, bei dem kombiniert werden, um einen funktionsfähigen Roboter zu erstellen.

N

Erkennst du, was das für eine Platine auf dem Rücken ist?








N

Aufgrund des bereitgestellten Bildes und des beschränkten Sichtfelds ist es schwierig, die genaue Art und Marke der Platine auf dem Rücken des Roboters präzise zu identifizieren. Es gibt viele verschiedene Arten von Mikrocontrollern und Entwicklungsboards, die in solchen Projekten verwendet werden können.



Einige beliebte Optionen für solche Projekte sind beispielsweise Boards aus der Arduino-Reihe, Raspberry Pi oder andere spezialisierte Mikrocontroller-Platinen. Aber ohne klarere Markierungen oder spezifischere Details auf der Platine oder im Kontext des Projekts ist es nicht möglich, eine genaue Bestimmung zu treffen.

-  ChatGPT 
-  Menu Converter
-  Consensus
-  GPTs erkunden

Heute


- Antipasti Menu Analysis
- Menu Secondi Piatti
- Menu Pizza Structured Table
- Menu Italiano Structurato


Vorherige 7 Tage

Stück unbekannt.

Vorherige 30 Tage

- New chat
- Zero2Go Auto Power On
- Speisen Übersetzungstabelle.
- IMEs Impact Orthographic Knowle
- Speisen und Preise erfassen

 Team-Arbeitsbereich hinzufügen
Zusammenarbeit im Team-Plan

 Noah Bubenhofer

Meine GPTs

+ Erstellen Sie eine GPT

GPTs

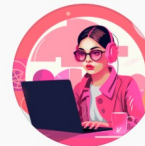
Entdecken und erstellen Sie individuelle Versionen von ChatGPT, die Anweisungen, zusätzliches Wissen und eine beliebige Kombination von Fähigkeiten kombinieren.

🔍 Öffentliche GPTs suchen

Top Picks DALL-E Schreiben Produktivität Forschung & Analyse Programmierung Bildung 

Featured

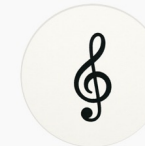
Curated top picks from this week



ChatPRD

An on-demand Chief Product Officer that drafts and improves your PRDs, while coaching you to...

Von Claire V Lawless



Music Teacher

Regular ChatGPT isn't great at music theory and relative scales, so I trained Music Teacher to be an...

Von gryphonedm.com



UX Design Mentor

I provide specific UX or Product Design feedback.

Von community builder



Adobe Express


Stand out with Adobe Express. Quickly and easily make impactful social posts, images, videos, flyers...

Von adobe.com

Trending

Most popular GPTs by our community

Eigene GPTs: Demo

- ChatGPT
- Menu Converter 
- Consensus
- GPTs erkunden

Heute

- Menu Secondi Piatti
- Menu Pizza Structured Table
- Menu Italiano Structurato

Vorherige 7 Tage

- Stück unbekannt.

Vorherige 30 Tage

- New chat
- Zero2Go Auto Power On
- Speisen Übersetzungstabelle.
- IMEs Impact Orthographic Knowle
- Speisen und Preise erfassen
- Vorspeisen Variationen Bistro

Team-Arbeitsbereich hinzufügen
Zusammenarbeit im Team-Plan

Noah Bubenhofer

Menu Converter ▾



Menu Converter

Von Noah Bubenhofer &

Parses a menu into a structured data table

Convert this menu into a table format.

Need help structuring this menu information.

I have a menu text; can you organize it into co...

Make a table out of the following menu details.

 Nachricht Menu Converter...



ChatGPT kann Fehler machen. Überprüfen Sie wichtige Informationen.



Spezialisierte Tools: Beispiel ChatPDF

+ New Chat
Drop PDF here

Upgrade to Plus Feedback

s41244-018-0096-4.pdf

Sign in to save your chat history

Home Account API FAQ

s41244-018-0096-4.pdf 4 / 32

360 N. Bubenhofer

Korpuslinguistisch wird die Kategorie narratives »Muster« also prinzipiell über zwei Elemente modelliert: Einerseits werden in den Geschichten typische Mehrworteinheiten, sog. »n-Gramme«, identifiziert. Andererseits wird aber auch deren typische Position in den Geschichten und damit ihre Abfolge bestimmt. Diese Modellierung erlaubt einen datengeleiteten Zugang: Es ist nicht notwendig, a priori bestimmte Mehrworteinheiten festzulegen, nach denen gesucht wird, sondern es werden automatisch alle n-Gramme extrahiert, die typisch für die Geburtsberichte sind, und danach berechnet, an welchen Positionen diese normalerweise auftreten. Diese Grundidee (deren methodischen Details ich weiter unten noch ausführe), beschrieben wir bereits anhand von Erzählungen Jugendlicher über ihr »Erstes Mal« (Bubenhofer u.a. 2013).

Diese stark datengeleitete Perspektive kann mit einer hypothesengeleiteten ergänzt werden. Dafür ziehe ich eine der bekanntesten linguistischen narrativen Theorien hinzu, nämlich jene von Labov und Waletzky (1973). Sie schlugen ein Raster vor, nach denen Erzählungen gegliedert sind und mit dem die datengeleitete berechneten narrativen Muster kategorisiert werden können.

Bislang sind Geburtsberichte aus Internetforen kaum Gegenstand der Forschung zur Narration, vor allem nicht linguistischer Forschung. Eine große Ausnahme ist die Dissertation von Cecilia Colloseus (2016), die aus kulturalanthropologischer Sicht Narrative des Gebärens untersucht und dabei Geburtsberichte einer qualitativen Analyse unterzieht. Ich werde stellenweise ihre Erkenntnisse in Bezug zu meinen Beobachtungen setzen.

Eine weitere Arbeit, die Geburtsberichte von Müttern untersucht, ist jene von Barbieri u.a. (2012), die mit psychoanalytischem Interesse italienischsprachige Forumsbeiträge als Datengrundlage verwendet. Dabei betrachten sie auch linguistische Phänomene, wählen aber ebenfalls einen qualitativen Zugang. Weiter sei auf Tienken (2015) verwiesen, die Webforen zu »Sternenkindern«, also Babies, die vor, während oder kurz nach der Geburt verstorben sind, untersucht.

2 Datengrundlage

Wie bereits erwähnt, stammen die 14 000 Geburtsberichte von öffentlich erreichbaren Web-Diskussionsforen. Im Folgenden beschreibe ich das Korpus und die Aufbereitung der Daten.

2.1 Korpus

Die Daten wurden im August 2016 von sechs Webforen aus Deutschland, Österreich und der Schweiz heruntergeladen. In Tab. 1 sind die sechs Foren und jeweils die Anzahl der Texte und Wörter aufgeführt. Es wurden jeweils alle beim jeweiligen Zeitpunkt des Downloads verfügbaren Texte heruntergeladen, die im Unterforum, das meist den Titel »Geburtsberichte« trägt, erschienen sind.

Die Texte aus den deutschen Foren sind mit knapp 12 000 Texten (gut 10 Mio. Wörter) im Vergleich zu den österreichischen (1600 Texte, 1,2 Mio. Wörter) und den schweizerischen (900 Texte, 1,2 Mio. Wörter) deutlich in der Überzahl. Als »Text«

Chat

erzählt werden sollten. Viel Spaß beim Lesen!

Mögliche Fragen:

- Wie wurden die 14.000 Berichte von Müttern in Internetforen gesammelt und ausgewählt?
- Welche Sprachgebrauchsmuster wurden in den Geburtsberichten am häufigsten verwendet?
- Welche gesellschaftlichen Vorstellungen über Geburten und Mutterschaft wurden in den Berichten sichtbar?

Welche gesellschaftlichen Vorstellungen über Geburten und Mutterschaft wurden in den Berichten sichtbar?

Die PDF-Datei beschäftigt sich mit der Analyse von Geburtsberichten in Internetforen, um gesellschaftliche Vorstellungen darüber zu untersuchen, wie Geburten erzählt werden sollten. Es ist jedoch nicht explizit angegeben, welche spezifischen gesellschaftlichen Vorstellungen in den Berichten sichtbar wurden.

Wie wurden die 14.000 Berichte von Müttern in Internetforen gesammelt und ausgewählt?

Die 14.000 Berichte von Müttern in Internetforen wurden automatisiert aus sechs Webforen aus Deutschland, Österreich und der Schweiz heruntergeladen. Dabei wurden alle beim jeweiligen Zeitpunkt des Downloads verfügbaren Texte heruntergeladen, die im Unterforum, das meist den Titel "Geburtsberichte" trägt, erschienen sind.

Was ist an diesem Paper innovativ?

Das Paper ist innovativ, weil es eine Korpusanalyse von Geburtsberichten in Internetforen durchführt, um typische Muster in diesen Geschichten zu finden. Dabei werden Korpuslinguistik und visuelle Analyse eingesetzt, um gesellschaftliche Vorstellungen darüber zu untersuchen, wie Geburten erzählt werden sollten. Die Studie trägt somit zur Erforschung der Serialität der Singularität bei und zeigt, wie Korpuslinguistik und visuelle Analyse zur Untersuchung von Narrativen eingesetzt werden können.

Ask any question...

Custom Assistant Settings are being used for this session. [View Settings](#)

what is critical discourse analysis?



Critical Discourse Analysis (CDA) is an interdisciplinary approach that aims to systematically link properties of discursive interactions and texts with features of their social and cultural circumstances (Fairclough, 1999). It investigates the relationship among discourse, ideology, and power in sociocultural contexts (Syukri et al., 2022). CDA provides a framework for detailed study of the relationship of language with other social processes and how language operates in power relations (Ar, 2020). It is particularly suitable for critical policy analysis due to its focus on the detailed study of language's relationship with social processes and power relations (Ar, 2020). CDA is viewed as an established approach to discourse analysis that has been around since the early 1980s (Flowerdew, 2012). It is concerned with language as a form of social practice and is used to analyze discourses regarding ideology (Rui-qi, 2020). Modern writers in the field of discourse analysis deal with discourse in the light of CDA theory, which views language as a form of social practice (Al-Momani, 2017). Additionally, CDA has been applied to the study of environmental law, utilizing its full range of analytical capabilities to study environmental discourse (Gellers, 2015).

In summary, Critical Discourse Analysis is an interdisciplinary approach that systematically links properties of discursive interactions and texts with features of their social and cultural circumstances. It investigates the relationship among discourse, ideology, and power in sociocultural contexts, making it suitable for critical policy analysis and the study of language as a form of social practice and ideology.

View searches run to ground this response

View publications consulted

Not what you were looking for? [Change Assistant Settings](#)

What do you need help with?

References

Reference #1

"...Critical discourse analysis aims to provide a framework for systematically linking properties of discursive interactions and texts with features of their social and cultural circumstances...."

See full context

Section: Critical Awareness Of Discourse and The New Global Capitalism

Global Capitalism and Critical Awareness of Language

Norman Fairclough¹ 1999 *Language Awareness*

163 | 1 | 80 | 0

View full text

Add to dashboard

Cite

Check for full-text from University of Zurich

Reference #2

"...Critical discourse Analysis (CDA) is an approach applied to a wide range of discourses to investigate the relation among discourse, ideology and power in sociocultural contexts...."

See full context

Section: Introduction

A VAN DIJK APPROACH TO BUGINESE FOLKLORE "Tau Sakbarak E" (Socio-Cultural and Linguistic Aspects)

Muhammad Syukri¹, Fatimah Hidayahni Amin², None St. Rahmaniah Bahrun³ et al, 2022 *JLe*

View full text

Add to dashboard

Cite

Check for full-text from University of Zurich

Reference #3

"...As Taylor (2004) states this relationship, "Critical discourse

Nicholson, Josh M. u. a. (2021): „scite: A smart citation index that displays the context of citations and classifies their intent using deep learning“, *Quantitative Science Studies* 2/3, S. 882–898.

Beispiel SciSpace

The screenshot displays the SciSpace web interface. At the top, there is a search bar with the text "Search or ask a question" and navigation buttons for "Upload PDF", "Extension", "Login", and "Sign up". The main content area shows a PDF document titled "Serialität der Singularität" by Noah Bubenhofer. The document header includes the journal information "Z Literaturwiss Linguistik (2018) 48:357–388" and the DOI "https://doi.org/10.1007/s41244-018-0096-4". The document title is "Serialität der Singularität" and the subtitle is "Korpusanalyse narrativer Muster in Geburtsberichten". The author is "Noah Bubenhofer". The document is published online on 19 April 2018 by Springer-Verlag GmbH Deutschland. The abstract discusses narrative patterns in birth stories. A Copilot chat window is open on the right, showing a question: "What are the contributions of this paper?" and a detailed answer listing the paper's contributions.

Home / s41244-018-0096-4.pdf

Search or ask a question

Upload PDF Extension Login Sign up

Home / s41244-018-0096-4.pdf

Σ Explain math & table 181%

Z Literaturwiss Linguistik (2018) 48:357–388
<https://doi.org/10.1007/s41244-018-0096-4>

LiLi CrossMark

THEMENBEITRAG

Serialität der Singularität

Korpusanalyse narrativer Muster in Geburtsberichten

Noah Bubenhofer

Online publiziert: 19. April 2018
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2018

Zusammenfassung In meinem Beitrag untersuche ich narrative Muster in vierzehntausend Berichten über Geburten, verfasst von Müttern in Internetforen. Ziel ist es dabei, typische Muster in diesen Geschichten zu finden, die jede für sich ein einmaliges Erlebnis erzählt, dabei jedoch auf Sprachgebrauchsmuster zurückgreifen. Dabei besteht das methodische Interesse, mit Mitteln der Korpuslinguistik und der visuellen Analyse solche Muster zu finden. Analytisch stellt sich die Frage, welche gesellschaftlichen Vorstellungen darüber herrschen, wie Geburten erzählt werden sollen.

Schlüsselwörter Alltagserzählung · Narrativ · Korpuslinguistik · Geburtsberichte · Sprachgebrauchsmuster

Seriality of Singularity: Corpus Linguistic Analysis of Narrative Patterns in Birth Stories

Copilot English (en)

Standard High Quality

What are the contributions of this paper 2:36 PM

Contributions of the paper "Serialität der Singularität Korpusanalyse narrativer Muster in Geburtsberichten" include:

- The paper aims to identify typical narrative patterns in 14,000 birth reports written by mothers in internet forums, using corpus linguistics and visual analysis methods.
- The author explores the societal ideas and expectations surrounding how birth stories should be told. [1]
- The paper introduces the concept of "narrative Topoi," which are language usage patterns that serve specific narrative functions within the stories. These Topoi are identified as statistically significant language patterns that indicate stereotypical narrative locations.
- The paper draws on the narrative categories proposed by Labov and Waletzky to categorize and interpret the data.
- The paper acknowledges the limitations of rigid structuralist models and proposes a heuristic model for interpreting the data in a quantitative, data-driven approach. [2]

Quickly ask these questions

Explain Abstract of this paper in 2 lines, What are the cor +13 more ^

Write your question...

NEW

Transparente und freie LLMs

SWISS AI Initiative

Home

Infrastructure

Research

Steering Committee

Researchers

Leveraging the world's most AI-capable supercomputer

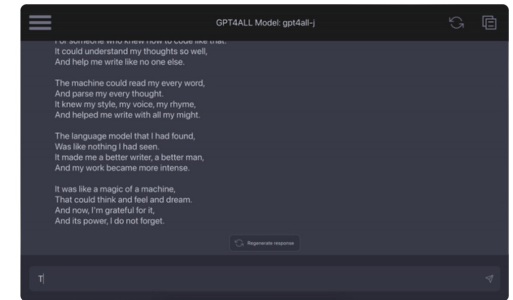
Alps by the National Supercomputing Center (CSCS) is the world's first national research infrastructure with over 10'000 GPUs of the new NVIDIA Grace Hopper superchip.



A research initiative to build capacities for advanced and large-scale AI systems for the benefit of society. The initiative is co-led by ETH Zurich and EPFL.

GPT4All

A free-to-use, locally running, privacy-aware chatbot. **No GPU or internet required.**



Real-time inference latency on an M1 Mac

Hugging Face Search m Models Datasets Spaces Posts Docs Pi

NEW Create Assistants in HuggingChat

The AI community building the future.

The platform where the machine learning community collaborates on models, datasets, and applications.

- Tasks
- Libraries
- Datasets
- Languages

Filter Tasks by name

- Multimodal
 - Text-to-Image
 - Image-to-Text
 - Text-to-Video
 - Visual Question Answering
 - Document Question Answering
- Computer Vision
 - Depth Estimation
 - Image Classification
 - Object Detection
 - Image Segmentation
 - Image-to-Image
 - Unconditional Image Generation
 - Video Classification
 - Zero-Shot Image Classification
- Natural Language Processing
 - Text Classification
 - Token Classification
 - Table Question Answering
 - Question Answering
 - Zero-Shot Classification
 - Translation
 - Summarization
 - Conversational
 - Text Generation
 - Text2Text Generation
 - Sentence Similarity
- Audio
 - Text-to-Speech
 - Automatic Speech Recognition
 - Audio-to-Audio
 - Audio Classification

Richtlinien

- ETH Zürich: KI in der Lehre
<https://ethz.ch/de/die-eth-zuerich/lehre/ai-in-education.html>
 - Universität Zürich: <https://www.uzh.ch/de/explore/basics/ai.html>
 - Universität Basel: <https://www.unibas.ch/de/Studium/Im-Studium/KI-in-Studium-und-Lehre.html>
 - Universität Bern:
https://www.unibe.ch/universitaet/organisation/leitung_und_zentralbereich/vizerektorat_lehre/startseite_vizerektorat_lehre/faq_zur_verwendung_von_ki_gestuetzten_hilfsmitteln_in_der_lehre_vizerektorat_lehre_universitaet_bern/index_ger.html
 - Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz:
https://www.kmk.org/fileadmin/Dateien/pdf/KMK/SWK/2024/SWK-2024-Impulspapier_LargeLanguageModels.pdf
 - ...
- Nächster DS-Digital-Anlass: 21. Mai 2024, 12:15 Uhr

Kollaborativ schreiben, auswerten und gruppieren

Oscar Friesen

"Schreibwerkstatt" => Korpus nutzen für Textanfänge (Wie schreibe ich einen guten Aufsatz?), Bewerbung

Semantischer Wandel von Begriffen untersuchen

Untersuchung "Deutsch als plurizentrische Sprache" (Stichwort Variantenwörterbuch)

Kontext: Seminar
Korpuslinguistik für die Schule
Sammeln von Projektideen

Alle schreiben ihre Ideen in ein kollaboratives Pad.

Alicia Fribolin:

1. dialektale Varianten --> Syntax und Wortschatz

--> ähnliches Vorgehen wäre vielleicht auch mit Begriffen aus der Jugendsprache

2. Helvetismen

Luca: "Projektarbeiten" werden in der Primar- und Sekundarschule immer wichtiger (Projekte), die sie sich selbst gesetzt haben.

--> Korpuslinguistik ganz allgemein als metafachliches Tool für SuS vorstellen, welches ihnen bei diversen

Davide Ventre: Semantische Kämpfe (Begriffskonkurrenzen wie Atomkraftwerk vs. Kernkraftwerk, verschiedene Normierung der Sprache, Dialekt, Kommunikationsnetzwerke, Kollokationen bspw. Emojis)

Gianna: Korpus aus eigenen WhatsApp Chats erstellen, Mehrsprachigkeit, Linguistic Landscape

Kaila: Korpus aus eigenen Texten generieren (WhatsApp oder verschiedene Schreibaufträge, die in der Schule evtl. Projektarbeit (z. B. in Arbeitswochen oder Ähnliches) zusammen mit Mathematik > Aufbau eines Korpus

t-weiss-woher-sie-stammen-1325640:

Einstimmung

Standarddeutsch

von Wörtern -->

[Prompt:] Hey, hier sind verschiedene Projektideen korpuslinguistischer Art von verschiedenen Personen. Kannst du sagen, welche Personen miteinander sprechen sollten? Welche Themenideen sind die Ideen:
[Ideen]



Nachdem ich die Projektideen von den verschiedenen Personen durchgegangen bin, habe ich einige Überschneidungen und thematische Ähnlichkeiten festgestellt. Hier sind die Personen und ihre thematisch ähnlichen Ideen, die möglicherweise miteinander sprechen sollten:

1. Korpus aus WhatsApp-Daten und Mehrsprachigkeit

- Anina Bächtold: Analyse mit Schweizer WhatsApp Korpus und Rechtschreibung im Schweizerdeutschen.
- Davide Ventre: WhatsApp Kommunikation - Korpuserstellung mit eigenen WhatsApp Daten, Dialekt, Kommunikationsnetzwerke, Kollokationen.
- Gianna: Korpus aus eigenen WhatsApp Chats erstellen und Mehrsprachigkeit.
- Kaila: Korpus aus eigenen Texten generieren, insbesondere WhatsApp.

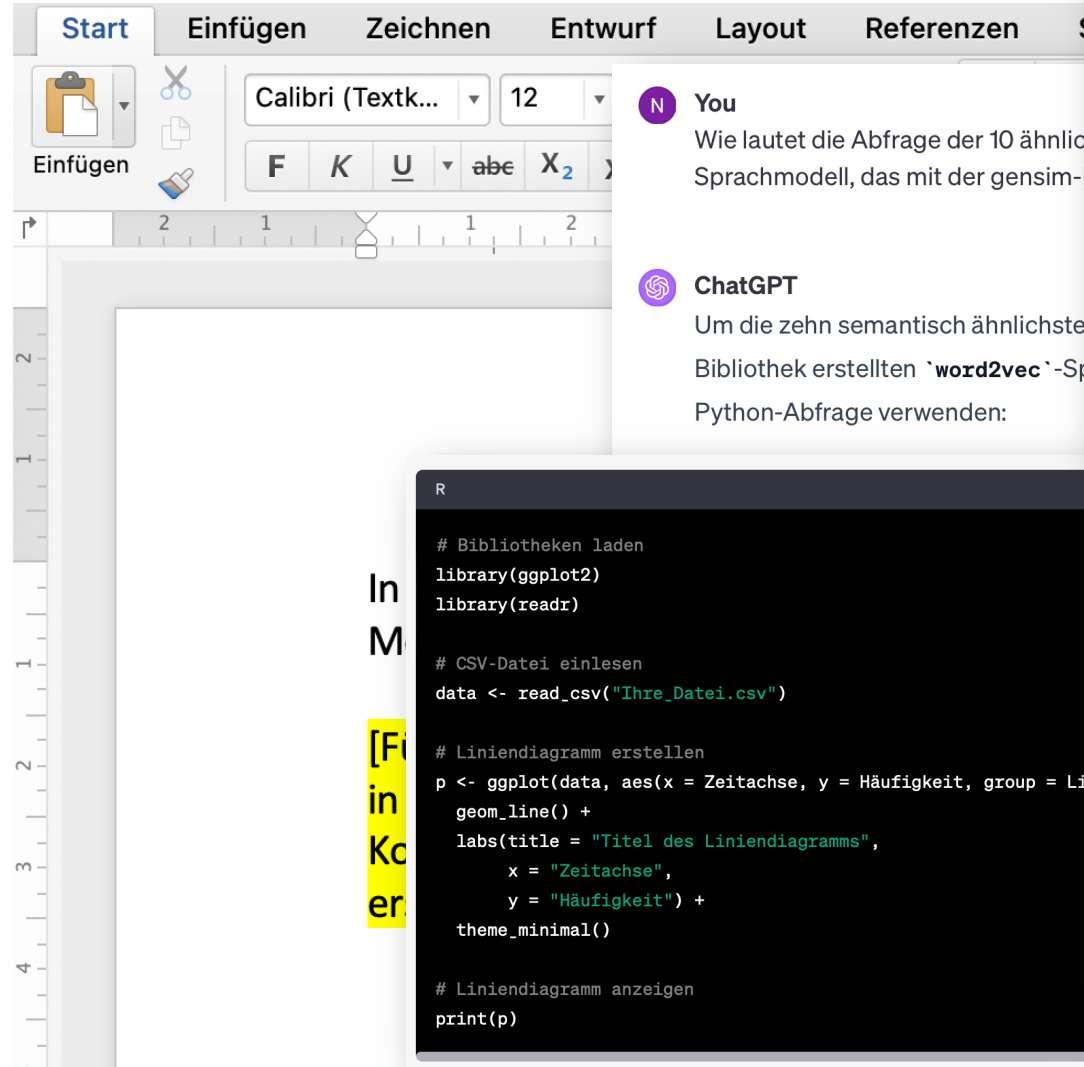
2. Dialektale Varianten und Semantischer Wandel

- Oscar Friesen: Semantischer Wandel von Begriffen untersuchen und Untersuchung "Deutsch als plurizentrische Sprache".
- Alicia Fribolin: dialektale Varianten in Syntax und Wortschatz, sowie Helvetismen.
- Noemi Ferrai: Sprachwandel im Zeitverlauf untersuchen.

3. Textanfänge, Stellensuche und Schreibprozesse

- Oscar Friesen: Korpus nutzen für Textanfänge und Bewerbungen.
- Daniel: Topic Modelling/Kollokationsanalyse von Stellenausschreibungen und

Programmieren und schreiben kom



Start Einfügen Zeichnen Entwurf Layout Referenzen

Einfügen Calibri (Textk... 12

You
Wie lautet die Abfrage der 10 ähnlichsten Wörter für ein bestimmtes Sprachmodell, das mit der gensim-Bibliothek erstellt wurde?

ChatGPT
Um die zehn semantisch ähnlichsten Wörter für ein bestimmtes Wort zu erhalten, können Sie die `word2vec`-Bibliothek erstellen und die folgenden Python-Abfrage verwenden:

```
R
# Bibliotheken laden
library(ggplot2)
library(readr)

# CSV-Datei einlesen
data <- read_csv("Thre_Datei.csv")

# Liniendiagramm erstellen
p <- ggplot(data, aes(x = Zeitachse, y = Häufigkeit, group = Li
geom_line() +
labs(title = "Titel des Liniendiagramms",
      x = "Zeitachse",
      y = "Häufigkeit") +
theme_minimal()

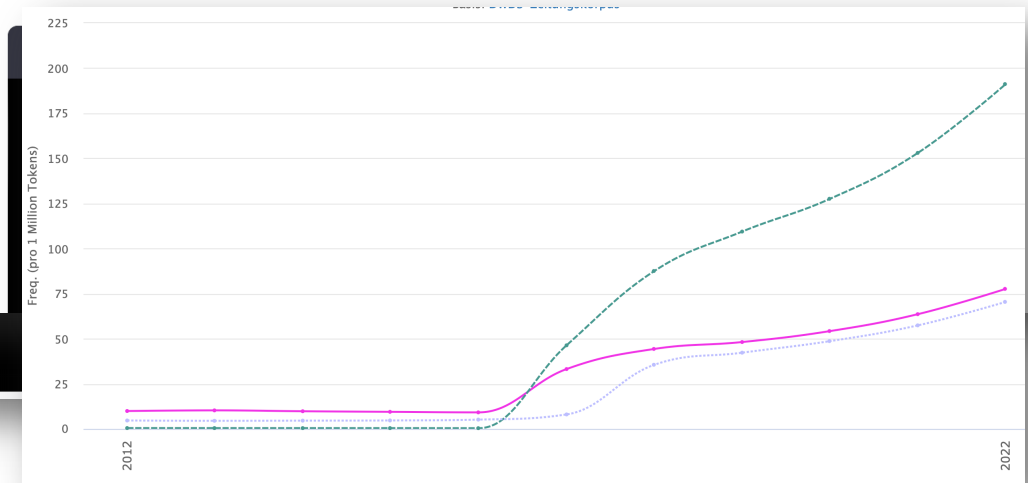
# Liniendiagramm anzeigen
print(p)
```

ChatGPT

Die RKorAPClient-Bibliothek für R bietet eine Schnittstelle zur KorAP-API, einem Korpusanalysewerkzeug, das den Zugriff auf große linguistische Daten ermöglicht. Um die relativen Häufigkeiten der zehn Wörter über den RKorAP-Client abzufragen, gehen Sie wie folgt vor:

1. **Installation des RKorAPClient:** Zuerst müssen Sie sicherstellen, dass `RKorAPClient` in Ihrer R-Umgebung installiert ist. Dies können Sie mit dem Befehl `install.packages("RKorAPClient")` tun.
2. **API-Abfrage erstellen:** Sobald `RKorAPClient` installiert ist, können Sie eine Abfrage erstellen, um die relativen Häufigkeiten der gewünschten Wörter zu erhalten. Die genaue Abfrage hängt davon ab, wie Sie Ihre Suchkriterien definieren möchten (z.B. Zeitraum, Korpus).
3. **Daten extrahieren:** Verwenden Sie die Funktionen des RKorAPClient, um die Ergebnisse der Abfrage abzurufen und zu analysieren.

Ein einfaches Beispiel für eine Abfrage könnte wie folgt aussehen:



Recherchieren mit ChatGPT?

N Du
Do IMEs have a negative effect on orthographic knowledge?

ChatGPT
The relationship between Input Method Editing and orthographic knowledge is complex, involving a deep understanding of the writing system of a language and can vary depending on several factors, such as the learner's proficiency level, and how the technology is used.

For languages with complex orthographies, such as German, IMEs might affect orthographic knowledge:

Positive Effects

- **Learning Aid:** For learners and those not proficient in a language, IMEs can act as a helpful tool, providing immediate feedback on spelling and correct forms of words and characters, aiding in the acquisition of orthographic knowledge.
- **Increased Exposure:** The use of IMEs can increase exposure to the orthographic system of a language, potentially leading to better familiarity with the rules and patterns of the writing system.

N Du
Co

Ch
A s

me

ort

kn

tha

acc

sig

N Du
Nenne mir bitte die wichtigsten Publikationen, die "Korpuspragmatik" definieren!

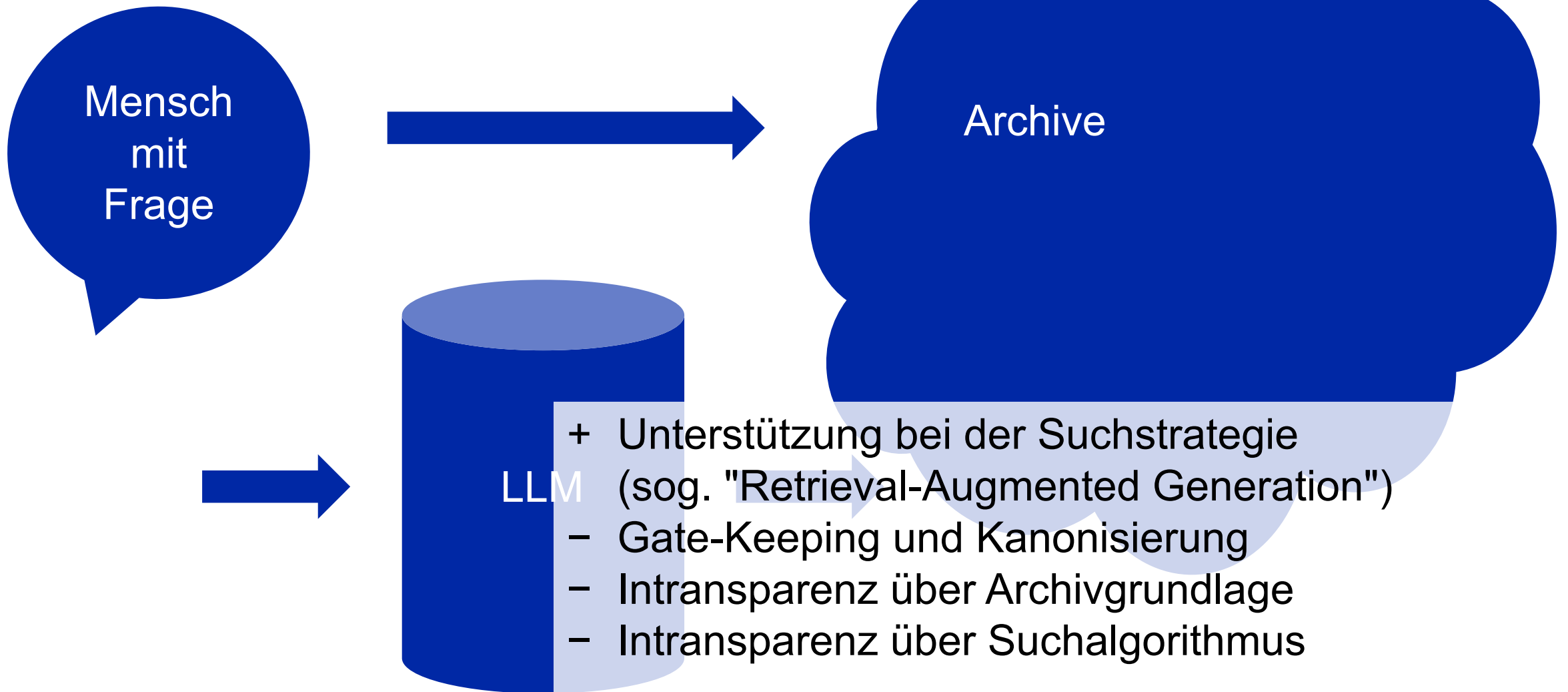
ChatGPT
Es scheint, dass der direkte Zugriff auf spezifische Publikationen zur Korpuspragmatik aufgrund von Einschränkungen der Webseite nicht möglich ist. Viele relevante Werke zur Korpuspragmatik sind jedoch oft in wissenschaftlichen Verlagen wie De Gruyter zu finden, die eine breite Palette von Publikationen in diesem Forschungsfeld anbieten. Hier sind einige Vorschläge, wie Sie die wichtigsten Publikationen in diesem Bereich finden können:

1. **Suche in wissenschaftlichen Datenbanken:** Plattformen wie Google Scholar, JSTOR oder die Web of Science bieten umfangreiche Suchmöglichkeiten nach wissenschaftlichen Artikeln und Büchern. Geben Sie Schlüsselwörter wie "Korpuspragmatik" ein, um relevante Publikationen zu finden.
2. **Bibliotheken und Online-Kataloge:** Universitätsbibliotheken und ihre Online-Kataloge können Zugang zu wichtigen Werken bieten. Viele Bibliotheken ermöglichen die Suche nach Schlagworten oder Themen und bieten oft auch Fernleihe an.

implications for literacy and language learning.



Recherchieren mit ChatGPT?



Relevante Aspekte in gesellschaftlichen Domänen

	Wissenschaft	Bildung	Wirtschaft	Medien	Politik
Kreativität vs. Automatisierung	Potenzial der Automatisierung vs. Verlust der Kreativität				
dynamische Mehrsprachigkeit	maschinelle Übersetzung (overt / covert)				
Autorschaft und Identität		Autorschaft	Kommunikation und Vertrauen, Authentizität, Identität		
Desinformation und Betrug	Wahrheit, Fakten vs. Lüge Plagiat			Wahrheit, Fakten vs. Lüge	
Sprach- und Kommunikationsideologien	Status und Bewertung der Sprachmodelle / KI Vorstellungen idealer Kommunikation mit Maschinen				