

AMR und LLMs

Lea Hirlimann – Profilierungsmodul 1

Inhalt



Abstract Meaning Representation (AMR)



Linguistisches Verständnis großer Sprachmodelle



Experiment: Wie gut parst ChatGPT in Bezug auf AMR?



Untersuchung der Limitierung von LLMs in Bezug auf AMR Annotierung von Ettinger et al. (2023)



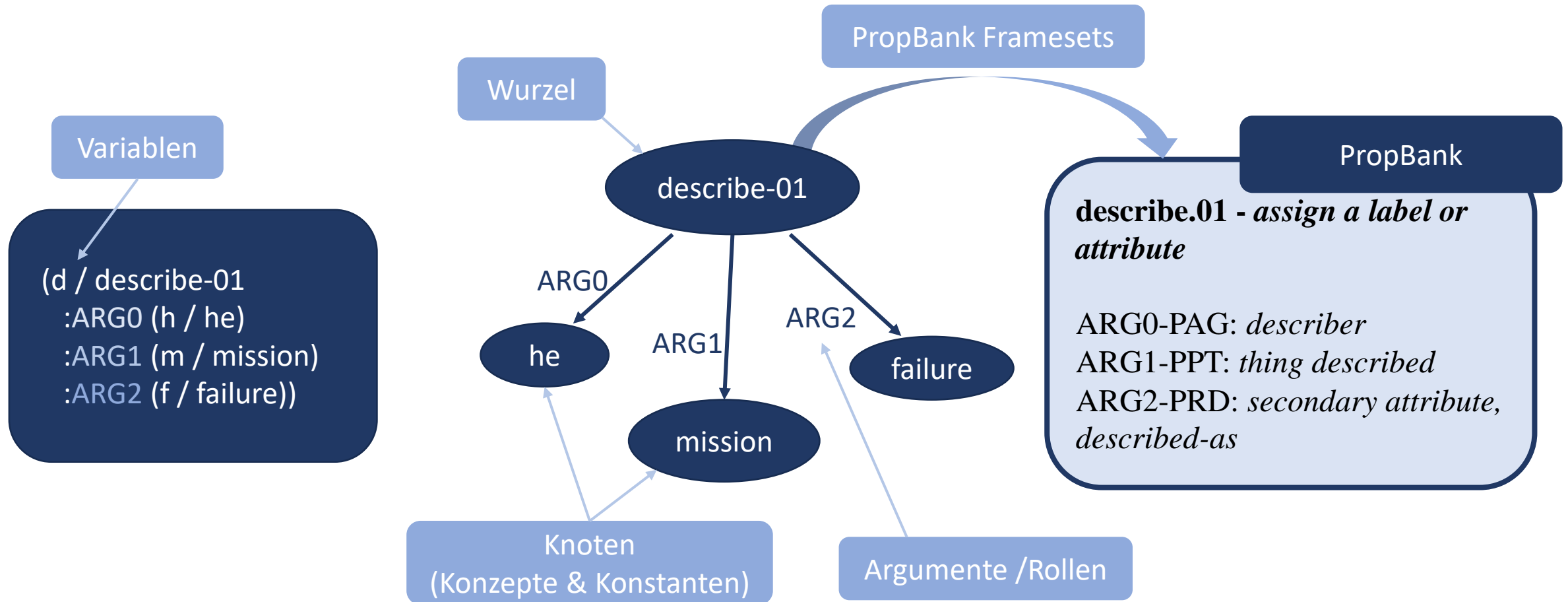
Experiment: Verstehen Sprachmodelle die Graphen besser als sie sie generieren?



Ausblick

AMR

Abstract Meaning Representation



Linguistisches Wissen in Sprachmodellen

- Da große Sprachmodelle sehr natürlichen Text generieren können, ist es interessant herauszufinden, wie gut sie sprachliche Merkmale erkennen können
- Sprachmodelle wurden bereits auf morphologisches, syntaktisches und auch semantisches Wissen getestet
- Forschende erhoffen sich mit dem grundlegenden Wissen der Sprachmodelle und einigen zusätzlichen Methoden sehr gute Ergebnisse zu erzielen und Sprache effizient zu parsen

Fragen wir ChatGPT

ChatGPT soll Abstract Meaning Representations für Sätze des kleinen Prinzen erstellen.

„It has horns.“

Chat GPT: (h / have-03
:ARG0 (i / it)
:ARG1 (h2 / horn))

✓ Korrekt laut der Annotationen

? Welche semantischen Rollen hat **have-03**?

Chat GPT:

PropBank

1.ARG0 (Possessor): The entity that possesses or owns something.

2.ARG1 (Possession): The entity or thing that is possessed or owned.

3.ARG2 (Source): The source or origin from which the possession originates.

4.ARG3 (Beneficiary): The entity that benefits from the possession.

„He thought, perhaps, that I was like himself.”

Chat GPT: (t / think-01
:ARG0 (h / he)
:ARG1 (I / like-01
:ARG1 (i / I)
:ARG2 (h2 / he))
:mod (p / perhaps))

- ✘ Nicht ganz korrekt laut der Annotationen
(t / think-01
:ARG0 (h / he)
:ARG1 (r / resemble-01
:ARG1 (i / i)
:ARG2 h)
:mod (p / perhaps))

h kann direkt verwendet werden für den ‚he‘ Knoten.

- ? Was ist der Unterschied von **resemble-01** und **like-01**?

like.01 - *have affection towards, be fond of, enjoy (habitually)*
ARG0-PAG: *liker*
ARG1-PPT: *object of affection*

PropBank

resemble.01 - *seem like*
ARG1-PPT: *copy*
ARG2-DIR: *original*

“You see yourself,” he said, “that this is not a sheep.”

Chat GPT:

```
( s / say-01
  :ARG0 ( h / he )
  :ARG1 ( s2 / see-01
    :ARG0 ( y / you )
    :ARG1 ( y2 / you )
    :ARG2 ( t / thing
      :op1 ( s3 / self )))
  :ARG2 ( n / not
    :ARG1 ( s4 / sheep )))
```

- ✘ Nicht korrekt laut der Annotationen
(s / say-01
:ARG0 (h / he)
:ARG1 (s2 / see-01
:ARG0 (y / you)
:ARG1 (s4 / sheep
:domain (t2 / this)
:polarity -)))

- **Chat GPT** nutzt weder domain noch polarity
- **op x** kann z.B. für Konjunktionen und Namen eingesetzt werden

PropBank

see.01 – view

ARG0-PAG: *viewer*

ARG1-PPT: *thing viewed*

ARG2-PRD: *attribute of arg1, further description*

say.01 - say

ARG0-PAG: *Sayer*

ARG1-PPT: *Utterance*

ARG2-GOL: *Hearer*

ARG3-ADV: *Attributive*

Ergebnisse

- Chat GPT erkennt Basis Strukturen der Annotationen
- Das Sprachmodel kennt einige Prädikate, setzt diese aber nicht zwingend richtig ein
 - es kommt zu Verwechslungen wie z.B. like-01 statt resemble-01
 - Operatoren wie quant, mod, polarity werden wenig eingesetzt
- Ein Beispiel zu nennen hilft, allerdings behart es auch auf einige Sachen

“You Are An Expert Linguistic Annotator”: Limits of LLMs as Analyzers of Abstract Meaning Representation

Allyson Ettinger, Jena D. Hwang, Valentina Pyatkin, Chandra Bhagavatula,
Yejin Choi

Experiments on GPT-3, ChatGPT, GPT-4

Die Autoren nutzen 20 Sätze mit ihren Goldstandard Annotationen von den Datensätzen Little Prince und AMR 3.0 und 10 manuell annotierte Sätze von Websites aus 2023.

Zero-Shot Parsing

- Das Model soll als linguistischer Annotator AMR Graphen im korrekten Format generieren

Few-Shot Parsing

- Das Modell erhält 5 ähnliche Beispiele von richtigen AMR-Graphen und soll basierend darauf für einen gegebenen Satz einen eigenen Graphen generieren

Metalinguistische Antworten in natürlicher Sprache

- Statt dem klassischen AMR Format soll das Model in natürlicher Sprache Informationen über die wichtigsten Events, Argumente, Prädikate und Modifikatoren zusammenfassen

Prompts

Zero-Shot Parsing

(System: You are an expert linguistic annotator.)
Sentence: <input sentence>
Provide an AMR (Abstract Meaning Representation) parse for this sentence.

Few-Shot Parsing

(System: You are an expert linguistic annotator.)
Sentence: <input sentence>
Provide an AMR (Abstract Meaning Representation) parse for this sentence.
I will first show some examples.
<5 sentences & parses>

Metalinguistische Antworten in natürlicher Sprache

(System: You are an expert linguistic annotator.)
Sentence: <input sentence>
Identify the **primary event** of this sentence, and the **predicate** corresponding to that event. If there are multiple equally primary events connected by a conjunction like “and”, identify the conjunction, and then identify each of the primary events and their corresponding predicates.

For each primary event, identify the **arguments** of the event predicate, and identify the **modifiers** of those arguments. Then for each primary event, identify any **additional modifiers** of that event.

Ettinger et al. (2023) p. 9

Evaluation

Level 1

Kriterium	Bedeutung
Basic Form	Graph hat das richtige Format
Top Node	Wurde der oberste Konzept Knoten erkannt?
Main Relation	Wurde das oberste Event erkannt?
Overall Acceptable	Wurde der Graph von menschlichen Annotatoren als akzeptabel eingestuft?

→ Zero- & Few-Shot Tests

Level 2

Kriterium	Bedeutung
Event Args	Alle Argumente sind vorhanden
Event Mods	Alle Modifikatoren des Events sind erkannt worden
Arg Mods	Die Modifikatoren der Argumente sind erkannt worden
Extra Mods	Das Model hat überschüssige Modifikatoren hinzugefügt

→ Few-Shot Tests und Metalinguistische Antworten

Level 1

Das Format wird auch ohne Beispiele in vielen Graphen richtig verwendet

Zero-Shot Parsing

Deutliche Probleme Knoten und Prädikate richtig zu benennen...

	Model	Basic Form	Main Rel	Top Node	Accept
0-shot	GPT-3	0.7	0.4	0.3	0.0
	ChatGPT	0.7	0.4	0.2	0.0
	GPT-4	1.0	0.4	0.4	0.0
5-shot	GPT-3	1.0	0.7	0.5	0.0
	ChatGPT	1.0	0.8	0.5	0.0
	GPT-4	1.0	0.7	0.6	0.1

Die Konzept Knoten und Events werden deutlich häufiger richtig benannt

Few-Shot Parsing

Aber noch keine akzeptablen Graphen...

Level 2

Weniger als die Hälfte der Graphen beinhaltet die richtigen Argumente und Modifikatoren

Few-Shot Parsing

Es werden häufig zusätzliche Elemente erfunden...

	Model	Event Args	Event Mods	Arg Mods	Extra Mods (↓)
AMR (5-shot)	GPT-3	0.4	0.4	0.1	0.4
	ChatGPT	0.4	0.5	0.2	0.5
	GPT-4	0.5	0.5	0.4	0.4
Meta- linguistic	GPT-3	0.5	0.3	0.2	0.4
	ChatGPT	0.6	0.4	0.2	0.3
	GPT-4	0.6	0.4	0.4	0.3

Metalinguistische Antworten in natürlicher Sprache

- Ergebnisse ähneln den der Few-Shot Tests
- Das Format der Parse ist also vermutlich nicht, was die Sprachmodelle limitiert

Ergebnisse

- Die groben Strukturen von AMR können gut von den Sprachmodellen generiert und verstanden werden
- Ettinger et al. (2023) vermuten jedoch grundlegende Limitationen der Sprachmodelle, die die vielen Fehler und willkürlichen Knoten erklären
- Die Autoren geben jedoch zu, dass ihre Sample Größe zu klein war um klare Aussagen über diese zu treffen

Fragen wir ChatGPT

Kann ChatGPT Sätze aus den AMR-Graphen formulieren?

(m / make-01
:ARG0 (y / you)
:ARG1 (a / another)
:ARG3 (i / i)
:mode imperative)

Chat GPT: „Make another me.“

✘ Nicht korrekt.

→ „Make me another.“

? Welche semantischen Rollen hat **make-01**?

Chat GPT:

- 1.ARG0 (Agent):** The entity that performs the action of making or creating.
- 2.ARG1 (Result):** The entity or thing that is being made or created.
- 3.ARG2 (Purpose):** The goal or purpose for which something is made.
- 4.ARG3 (Material):** The substance or material used in the making or creation process.

make.01 - create

ARG0-PAG: *creator*

ARG1-PRD: *creation*

ARG2-VSP: *created-from, thing changed*

ARG3-GOL: *benefactive*

PropBank

(a / answer-01
:ARG0 (h / he)
:ARG1 (i / i)
:ARG2 (m2 / matter-01
:ARG1 (t / that)
:polarity -))

Chat GPT:

„He answered that it doesn't matter.“



Ziemlich verständlich.

→ He answered me: „That doesn't matter.“

Chat GPT:

In this sentence:

- "He" is the one answering (ARG0).
- "I" is the one being answered to (ARG1).
- The answer involves the matter specified by "that" (ARG2).
- The polarity is negative ("-"), indicating that it doesn't matter.

```
(z1 / and
:op1 (z2 / issue-02
:polarity
:time (z3 / board-01
:ARG0 (z4 / we)
:ARG1 (z5 / train)))
:op2 (z6 / and
:op1 (z7 / sit-down-02
:ARG1 z4)
:op2 (z8 / put-01
:ARG0 z4
:ARG1 (z9 / luggage)
:ARG2 (z10 / next-to
:op1 z4))))
```

Chat GPT:

„We boarded the train without any issue and sat down, placing the luggage next to us.



Korrekt.

→ „There was no issue as we boarded the train, and we sat down and put the luggage next to us.”

Ausblick

- Bisher können Sprachmodelle nur einige wenige grundlegende Bestandteile der AMR richtig generieren
- Autoren wie Ettinger et al. sind sich sicher, dass diese Ergebnisse mit einigen Hilfsmittel effektiv gesteigert werden können
- Es sollte davor aber definitiv nochmal eine tiefergehende Studie auf mehr Beispielen durchgeführt werden
- Außerdem könnten die Ergebnisse auch mit anderen Parsing Formaten verglichen werden, um übergreifende Probleme aufzudecken

Quellen

- Ettinger, A., Hwang, J., Pyatkin, V., Bhagavatula, C., & Choi, Y. (2023, December). “You Are An Expert Linguistic Annotator”: Limits of LLMs as Analyzers of Abstract Meaning Representation. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 8250-8263)
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, Nathan Schneider (2014): Abstract Meaning Representation (AMR) 1.2 Specification. URL: <https://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf>
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, Nathan Schneider (2013): Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse* (pp. 178-186)
- Martha Palmer, Dan Gildea, Paul Kingsbury, The Proposition Bank: A Corpus Annotated with Semantic Roles *Computational Linguistics Journal*, 31:1, 2005.