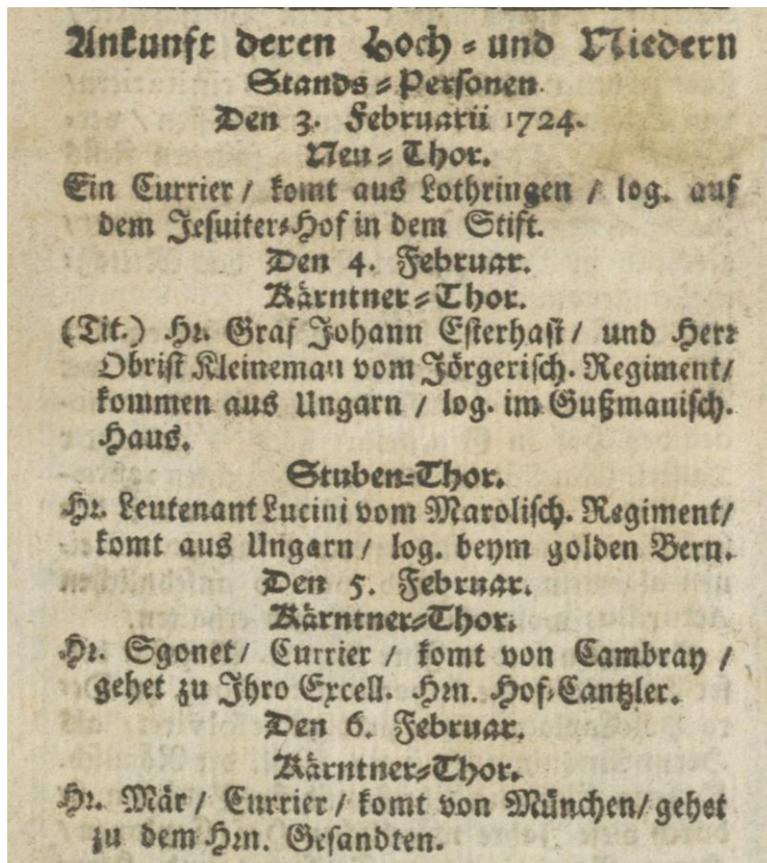


Informationsextraktion aus frühneuzeitlichen Ankunftslisten – das Projekt „Visiting Vienna“ als Fallstudie zur Named Entity Recognition mit GPT-3.5

Nina C. Rastinger

Workshop „Generative KI, LLMs und GPT bei
digitalen Editionen“ (DHd2024, 27.02.2024)

Ankunftslisten des *Wien[n]erischen Diariums*

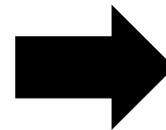


Wienerisches Diarium (09.02.1724: 8)

- 1703-1725: zweimal wöchentlich in der Zeitung *Wien[n]erisches Diarium* (seit 1780: *Wiener Zeitung*) publiziert
- Erhalt über kaiserliches Privileg (Mader-Kratky, Resch & Scheutz 2019: 99)
- Dokumentation der Ankunft hochrangiger Personen (z.B. Adelige, Individuen mit militärischer Funktion, Geistliche, Kuriere)

Den 5. Dito.
Kärntner-Thor. Herz Graf Jörger/kome
aus Spannen/ log. im wilden Mann.

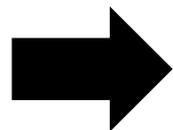
WD 07.01.1711:7



- Hohe Dichte an Named Entities (NEs)
- Möglichst automatische Aufbereitung und Auswertung
- Semi-strukturierte Texte > strukturierte Daten

Datenbasis

- Layoutanalyse und Handwritten Text Recognition (HTR) mit Transkribus (<https://readcoop.eu/de/transkribus>)
- HTR-Modell: „German Fraktur 18th Century – WrDiarium_M9“ (Resch & Kampkaspar 2020, <https://readcoop.eu/de/modelle/german-fraktur-18th-century>, CER: 0,8 %)
- + manuelle Nachkorrektur
- Präprozessierung: Entfernen von Zeilenumbrüchen; Vereinheitlichung von doppelten zu einfachen Bindestrichen



1.447 verlässliche Volltexte (176.452 Tokens) als Ausgangsbasis

The screenshot shows the Transkribus interface with a handwritten document. The text is in German Fraktur script. The document contains two entries, one for January 30, 1706, and one for January 31, 1706. The text is segmented into lines, and a legend at the bottom explains the segmentation. A green circle with the number 6 is overlaid on the text.

Legend:

- 5-1 Ankunfft aller Hoch= und nidrigen Stands=Personen.
- 6-1 Den 30. Januari 1706.
- 6-2 Cärntner=Thor. Herr Häckel / vom
- 6-3 **Kriegs=Commissariat, kombt auß Jta=**
- 6-4 lien / log. beym wilden Mann.
- 6-5 Den 31. dito.
- 6-6 Cärntner=Thor. Herr Graff von Leng=
- 6-7 heim / kombt von Grätz / log. auf der
- 6-8 Mehl=Gruben.

Transkribus

NER mit GPT-3.5 & Promptify

- **Promptify** = Open-Source-Python-Bibliothek für Prompt Engineering
- <https://github.com/prompts-lab/Promptify>
- Zugriff auf **GPT-Modelle über OpenAI-API** (hier: **gpt-3.5-turbo**)
- Workflow für NER – u.a. folgende Parameter setzbar:

Domäne	Labels	Beispiele	Beschreibung	Format
Domäne des Textmaterials, z.B. Medizin	Annotationskategorien, z.B. ORT	Exemplarische Input-Output-Paare	Details zum Textmaterial, z.B. Zeitungen	Output-Format, z.B. JSON



NER mit GPT-3.5 & Promptify

Domäne	Labels	Beispiele	Beschreibung	Format
Domäne des Textmaterials, z.B. Medizin	Annotationskategorien, z.B. ORT	Exemplarische Input-Output-Paare	Details zum Textmaterial, z.B. Zeitungen	Output-Format, z.B. JSON



NER mit GPT-3.5 & Promptify

Domäne	Labels	Beispiele	Beschreibung	Format
Domäne des Textmaterials, z.B. Medizin	Annotationskategorien, z.B. ORT	Exemplarische Input-Output-Paare	Details zum Textmaterial, z.B. Zeitungen	Output-Format, z.B. JSON

```
labels = ["Person", "Stadtort", "Datum",  
"Herkunftsort", "Unterkunftsort", "Zielort"]
```

Deutschsprachige, auf Textstruktur basierende Labels



NER mit GPT-3.5 & Promptify



Domäne	Labels	Beispiele	Beschreibung	Format
Domäne des Textmaterials, z.B. Medizin	Annotationskategorien, z.B. ORT	Exemplarische Input-Output-Paare	Details zum Textmaterial, z.B. Zeitungen	Output-Format, z.B. JSON

```
one_shot_text = "Ankunft aller Hoch- und nidrigen Stands Persohnen. Den 25. May. Rothen-Thurn. Herr Graff Rechtern Almelo / Holländis. Gesandter  
one_shot = [[one_shot_text, [{"E": "25. May", "T": "Datum", "S": 56, "En": 63},  
{"E": "Rothen-Thurn", "T": "Stadtort", "S": 65, "En": 77},  
{"E": "Herr Graff Rechtern Almelo / Holländis. Gesandter", "T": "Person", "S": 79,  
{"E": "Regensburg", "T": "Herkunftsort", "S": 141, "En": 151},  
{"E": "Sintzendorffis. Garten", "T": "Unterkunftsort", "S": 164, "En": 186},  
{"E": "26. dito", "T": "Datum", "S": 192, "En": 200},  
{"E": "Herr Obrist Baron Hauben", "T": "Person", "S": 202, "En": 226},
```

One/few shot: 1 Ankunftsliste mit 9 Items als Input → 37 Entitäten mit Label, Start- und Endpunkt als Output

NER mit GPT-3.5 & Promptify

Domäne	Labels	Beispiele	Beschreibung	Format
Domäne des Textmaterials, z.B. Medizin	Annotationskategorien, z.B. ORT	Exemplarische Input-Output-Paare	Details zum Textmaterial, z.B. Zeitungen	Output-Format, z.B. JSON

Your output format is only `{{ output_format|default("[[{'T': type of entity from predefined entity types, 'E': entity in the input text, 'S' : Start Position, 'En' : End Position}],...]") }}` form, no other form.



NER mit GPT-3.5 & Promptify

You are a highly intelligent and accurate Named-entity recognition(NER) system. You take Passage as input and your task is to recognize and extract specific types of named entities in that given passage and classify into a set of following predefined entity types: {{ labels }}

Your output format is only {{ output_format|default("[{{'T': type of entity from predefined entity types, 'E': entity in the input text, 'S' : Start Position, 'En' : End Position}},...]") }} form, no other form.

Examples:

Input: {{ sentence }}

Output: [{{ label }}]

Auszug aus jinja-template



NER mit GPT-3.5 & Promptify

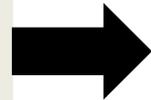
```
model = OpenAI("api-key")
prompter = Prompter('ner_adapted.jinja')
pipe = Pipeline(prompter, model)
result = pipe.fit(text,
                  domain = None,
                  examples = one_shot,
                  labels = ["Person", "Stadttor", "Datum", "Herkunftsort", "Unterkunftsort", "Zielort"])
```



Jeweils eine Ankunftsliste übergeben und Output gemeinsam mit bereits vorhandenen Metadaten abgespeichert

NER mit GPT-3.5 & Promptify

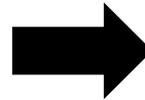
Ankunft derer Hoh- und Niedern Stands-Personen. Den 24. Septemb. 1712. Roten-Thurn. Herr Graf Scherschoffski / von Prag / komtlog. beyn 3. Hacken. Kärntner-Thor. Herr Sarasini, Cavalier, komt von Trient / log. im goldnen Hirschen. Den 25. Dito. Herr Wüttich / Currier / komt von der Reichs-Armee / geht ins Post-Haus. Stuben-Thor. Herr Graf Carl Zitschi / komt von Presburg / log. im Erdödischen Hauß. Den 27. Dito. Schotten-Thor. Herr Hof-Kammer-Rath Mora / komt von Salzburg / log. in der Kärntnerstrassen / im Zinngiesserischen Hauß.



```
{'E': '24. Septemb. 1712', 'T': 'Datum', 'S': 40, 'En': 57}  
{'E': 'Roten-Thurn', 'T': 'Stadttor', 'S': 59, 'En': 71}  
{'E': 'Herr Graf Scherschoffski', 'T': 'Person', 'S': 73, 'En': 97}  
{'E': 'Prag', 'T': 'Herkunftsort', 'S': 110, 'En': 114}  
{'E': 'beyn 3. Hacken', 'T': 'Unterkunftsort', 'S': 127, 'En': 140}  
{'E': 'Kärntner-Thor', 'T': 'Stadttor', 'S': 142, 'En': 155}  
{'E': 'Herr Sarasini', 'T': 'Person', 'S': 157, 'En': 170}  
{'E': 'Trient', 'T': 'Herkunftsort', 'S': 183, 'En': 189}  
{'E': 'goldnen Hirschen', 'T': 'Unterkunftsort', 'S': 202, 'En': 217}  
{'E': '25. Dito', 'T': 'Datum', 'S': 223, 'En': 231}  
{'E': 'Herr Wüttich', 'T': 'Person', 'S': 233, 'En': 245}  
{'E': 'Reichs-Armee', 'T': 'Herkunftsort', 'S': 258, 'En': 270}  
{'E': 'Post-Haus', 'T': 'Zielort', 'S': 283, 'En': 292}  
{'E': 'Stuben-Thor', 'T': 'Stadttor', 'S': 294, 'En': 305}  
{'E': 'Herr Graf Carl Zitschi', 'T': 'Person', 'S': 307, 'En': 328}  
{'E': 'Presburg', 'T': 'Herkunftsort', 'S': 341, 'En': 349}  
{'E': 'Erdödischen Hauß', 'T': 'Unterkunftsort', 'S': 362, 'En': 377}  
{'E': '27. Dito', 'T': 'Datum', 'S': 383, 'En': 391}  
{'E': 'Schotten-Thor', 'T': 'Stadttor', 'S': 393, 'En': 406}  
{'E': 'Herr Hof-Kammer-Rath Mora', 'T': 'Person', 'S': 408, 'En': 432}  
{'E': 'Salzburg', 'T': 'Herkunftsort', 'S': 445, 'En': 452}  
{'E': 'Kärntnerstrassen', 'T': 'Unterkunftsort', 'S': 465, 'En': 481}  
{'E': 'Zinngiesserischen Hauß', 'T': 'Unterkunftsort', 'S': 484, 'En': 506}
```

NER mit GPT-3.5 & Promptify

```
{'E': '24. Septemb. 1712', 'T': 'Datum', 'S': 40, 'En': 57}  
{'E': 'Roten-Thurn', 'T': 'Stadtort', 'S': 59, 'En': 71}  
{'E': 'Herr Graf Scherschoffski', 'T': 'Person', 'S': 73, 'En': 97}  
{'E': 'Prag', 'T': 'Herkunftsart', 'S': 110, 'En': 114}  
{'E': 'beyn 3. Hacken', 'T': 'Unterkunftsart', 'S': 127, 'En': 140}  
{'E': 'Kärntner-Thor', 'T': 'Stadtort', 'S': 142, 'En': 155}  
{'E': 'Herr Sarasini', 'T': 'Person', 'S': 157, 'En': 170}  
{'E': 'Trient', 'T': 'Herkunftsart', 'S': 183, 'En': 189}  
{'E': 'goldnen Hirschen', 'T': 'Unterkunftsart', 'S': 202, 'En': 217}  
{'E': '25. Dito', 'T': 'Datum', 'S': 223, 'En': 231}  
{'E': 'Herr Wüttich', 'T': 'Person', 'S': 233, 'En': 245}  
{'E': 'Reichs-Armee', 'T': 'Herkunftsart', 'S': 258, 'En': 270}  
{'E': 'Post-Haus', 'T': 'Zielort', 'S': 283, 'En': 292}  
{'E': 'Stuben-Thor', 'T': 'Stadtort', 'S': 294, 'En': 305}  
{'E': 'Herr Graf Carl Zitschi', 'T': 'Person', 'S': 307, 'En': 328}  
{'E': 'Presburg', 'T': 'Herkunftsart', 'S': 341, 'En': 349}  
{'E': 'Erdödischen Hauß', 'T': 'Unterkunftsart', 'S': 362, 'En': 377}  
{'E': '27. Dito', 'T': 'Datum', 'S': 383, 'En': 391}  
{'E': 'Schotten-Thor', 'T': 'Stadtort', 'S': 393, 'En': 406}  
{'E': 'Herr Hof-Kammer-Rath Mora', 'T': 'Person', 'S': 408, 'En': 432}  
{'E': 'Salzburg', 'T': 'Herkunftsart', 'S': 445, 'En': 452}  
{'E': 'Kärntnerstrassen', 'T': 'Unterkunftsart', 'S': 465, 'En': 481}  
{'E': 'Zinngiesserischen Hauß', 'T': 'Unterkunftsart', 'S': 484, 'En': 506}
```

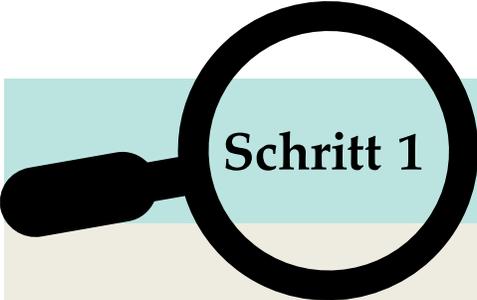


Ankunft derer Hoh- und Niedern Stands-Personen. Den 24.
Septemb. 1712. Roten-Thurn. Herr Graf Scherschoffski / von Prag /
komtlog. beyn 3. Hacken. Kärntner-Thor. Herr Sarasini, Cavalier,
komt von Trient / log. im goldnen Hirschen. Den 25. Dito. Herr
Wüttich / Currier / komt von der Reichs-Armee / geht ins Post-
Haus. Stuben-Thor. Herr Graf Carl Zitschi / komt von Presburg /
log. im Erdödischen Hauß. Den 27. Dito. Schotten-Thor. Herr Hof-
Kammer-Rath Mora / komt von Salzburg / log. in der
Kärntnerstrassen / im Zinngiesserischen Hauß.



Verortung der Entitäten

- Meist Ausgabe inkorrektur Entitätenstart- und endpunkte
- Aber: Abfolge der Entitäten in (fast) allen Fällen korrekt
- Eruiereu tatsächlicher Positionen über zweiteiligen Suchprozess:



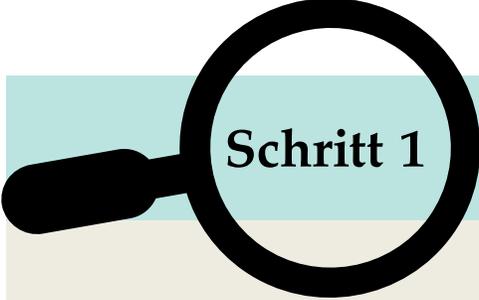
Schritt 1

Lokalisierung von
gefundenen Entitäten,
die nur **einmal** im
Inputtext auftreten

Ankunft deren hoh- und niederen Stands-Personen. Den 10. Augusti / 1720. Schotten-Thor. Herr Graf St. Julian / komt von Walsee / log. im Hoisischen Hauß. Karntner-Thor. Herr von Dütigen / Hauptmann vom Althannischen Regiment / komt aus Hungarn / log. im weisen Lambel. Den 11. Dito. Herr von Ränfdelshoffen / Probst zu Strasburg / in Kärnten / komt von dannen / log. im wilden Mann. Herr Baron Poztazky / und Herr Hauptmann Alberti / vom Lothringis. Regiment / kommen aus Hungarn / log. in 3. Haasen. [...] Den 13. Dito. Karntner-Thor. Herr Wolf Ehrenreich / Graf von Auersperg / komt von St. Pölten / log. im Matschacker Hof. Herr Vogel / Ritmaister vom Croyschen Regiment / komt aus Hungarn / log. bey der Heil. Dreyfaltigkeit.

Verortung der Entitäten

- Meist Ausgabe inkorrektter Entitätenstart- und endpunkte
- Aber: Abfolge der Entitäten in (fast) allen Fällen korrekt
- Eruieren tatsächlicher Positionen über zweiteiligen Suchprozess:



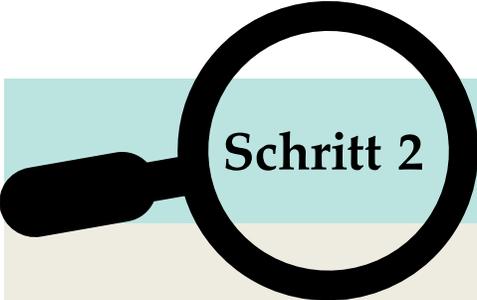
Schritt 1

Lokalisierung von
gefundenen Entitäten,
die nur **einmal** im
Inputtext auftreten

Ankunft deren hoh- und niederen Stands-Personen. Den 10. Augusti / 1720. Schotten-Thor. Herr Graf St. Julian / komt von Walsee / log. im Hoisischen Hauß. Karntner-Thor. Herr von Dütigen / Hauptmann vom Althannischen Regiment / komt aus Hungarn / log. im weisen Lambel. Den 11. Dito. Herr von Ränfdelshoffen / Probst zu Strasburg / in Kärnten / komt von dannen / log. im wilden Mann. Herr Baron Poztazky / und Herr Hauptmann Alberti / vom Lothringis. Regiment / kommen aus Hungarn / log. in 3. Haasen. [...] Den 13. Dito. Karntner-Thor. Herr Wolf Ehrenreich / Graf von Auersperg / komt von St. Pölten / log. im Matschacker Hof. Herr Vogel / Ritmaister vom Croyschen Regiment / komt aus Hungarn / log. bey der Heil. Dreyfaltigkeit.

Verortung der Entitäten

- Meist Ausgabe inkorrektor Entitätenstart- und endpunkte
- Aber: Abfolge der Entitäten in (fast) allen Fällen korrekt
- Eruieren tatsächlicher Positionen über zweiteiligen Suchprozess:



Schritt 2

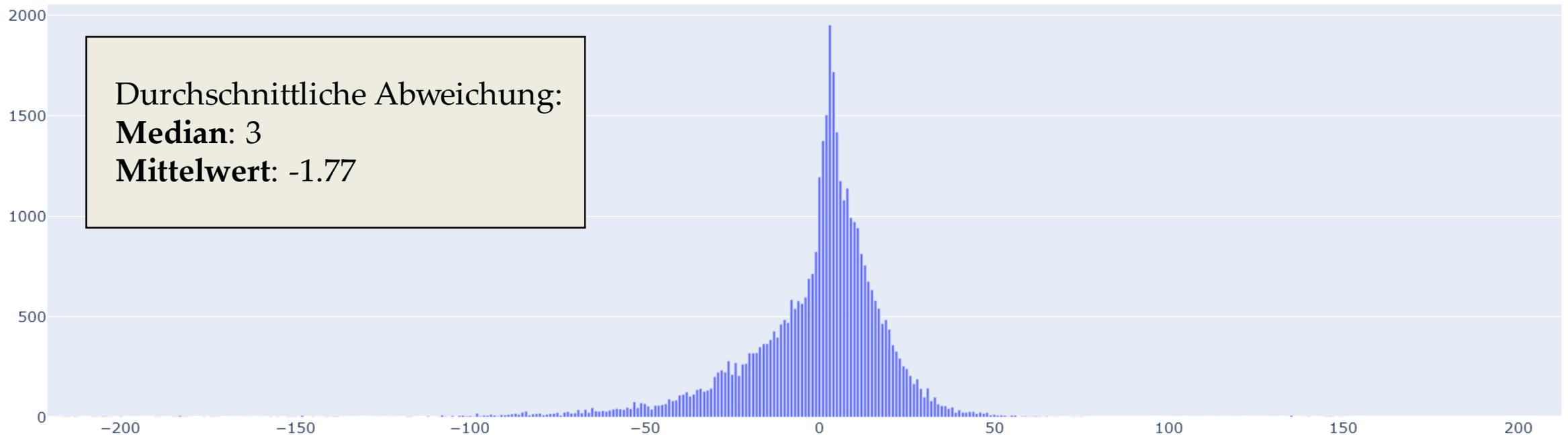
Mehrmals → Suchfenster
anhand von davor und
danach bereits
identifizierten Entitäten

Ankunft deren hoh- und niederen Stands-Personen. Den 10. Augusti / 1720. Schotten-Thor. Herr Graf St. Julian / komt von Walsee / log. im Hoisischen Hauß. Karntner-Thor. Herr von Dütigen / Hauptmann vom Althannischen Regiment / komt aus Hungarn / log. im weisen Lambel. Den 11. Dito. Herr von Ränfdelshoffen / Probst zu Strasburg / in Kärnten / komt von dannen / log. im wilden Mann. Herr Baron Poztazky / und Herr Hauptmann Alberti / vom Lothringis. Regiment / kommen aus Hungarn / log. in 3. Haasen. [...] Den 13. Dito. Karntner-Thor. Herr Wolf Ehrenreich / Graf von Auersperg / komt von St. Pölten / log. im Matschacker Hof. Herr Vogel / Ritmaister vom Croyschen Regiment / komt aus Hungarn / log. bey der Heil. Dreyfaltigkeit.



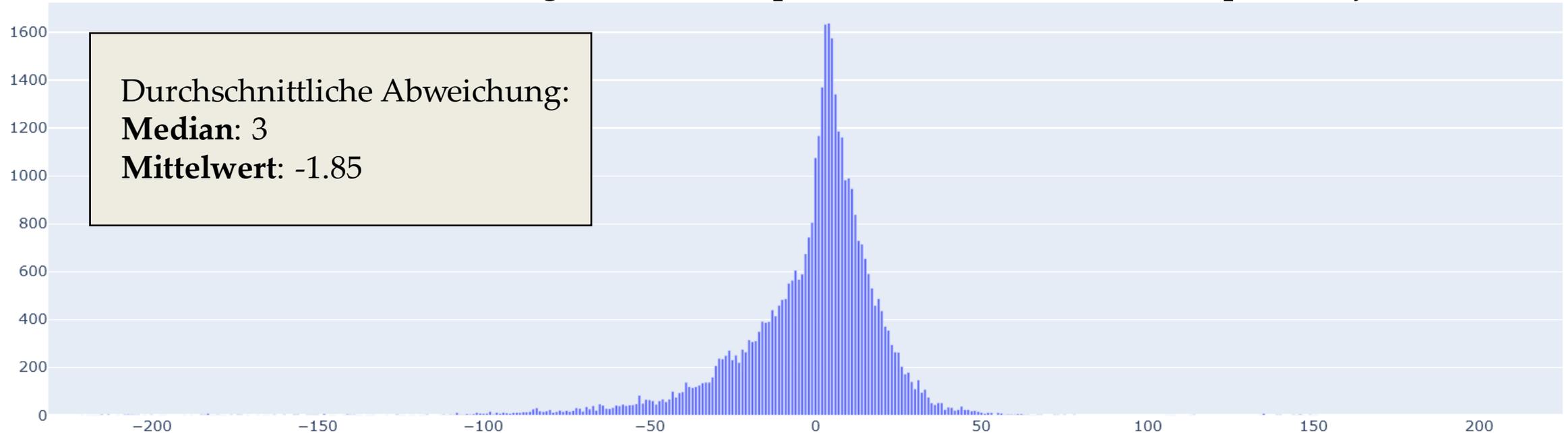
Positionsgenauigkeit

Differenz zwischen von GPT-3.5 geschätzter Startposition und tatsächlicher Startposition je Entität



Positionsgenauigkeit

Differenz zwischen von GPT-3.5 geschätzter Endposition und tatsächlicher Endposition je Entität



Output

39.512 Entitäten
aus 1.447 Listen



Evaluation

Ankunfft Hoch- und Niedriger Stands-Persohnen. Stuben-Thor stadttor den 4. April Datum .
Gestert spath kombt Herr Graf Lobkowitz Person von Prag Herkunftsort / logirt im
Nostitzischen Hauß auff der Hohen-Brucken Unterkunftsor . Rothen-Thurn stadttor . Herr
Wandt / Fourrier von Rehmischen Person / kombt gestern spath von Thiernkruth Herkunftsort
logirt in der guldenen Ganß Unterkunftsor Herr Prangheim / Hof-Cammer-Rath Person /
kombt von Breßlau Herkunftsort / logirt beym schwarzen Thor Unterkunftsor . Stuben-
Thor stadttor den 5. dito Datum . Herr Peter Gradoni Person / kombt von Görtz
Herkunftsor / logirt in der Cärnter-Strassen im weissen Schwan Unterkunftsor . Cärnter-
Thor stadttor . Herr Zeiler / Kayserl. Courier Person / kombt von der Rabau vom Herrn
General Heister Herkunftsort / logirt im Hof-Sailerischen Hauß Unterkunftsor . Herr Baron
von Schönau Person / kombt von Venedig Herkunftsort / logirt im guldenen Hirschen
Unterkunftsor . Stuben-Thor stadttor den 6. dito Datum . Herr Graf Frantz Adam von
Lamberg Person / kombt von Stockerau Herkunftsort / logirt im Orellischen Hauß

Vorgehen bei Evaluation

- Visualisierung über **displaCy**
- Manuelle Durchsicht aller bearbeiteten Ankunftslisten
- Evaluation auf **Zeichen-Level** und über Python-Framework **segeval** (IOB-Format)

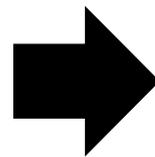
→ **Ziel:** Erkenntnisse zu Ergebnis- und Fehlerqualität von GPT-3.5

Halluzinationen



Bild generiert mit DALL-E

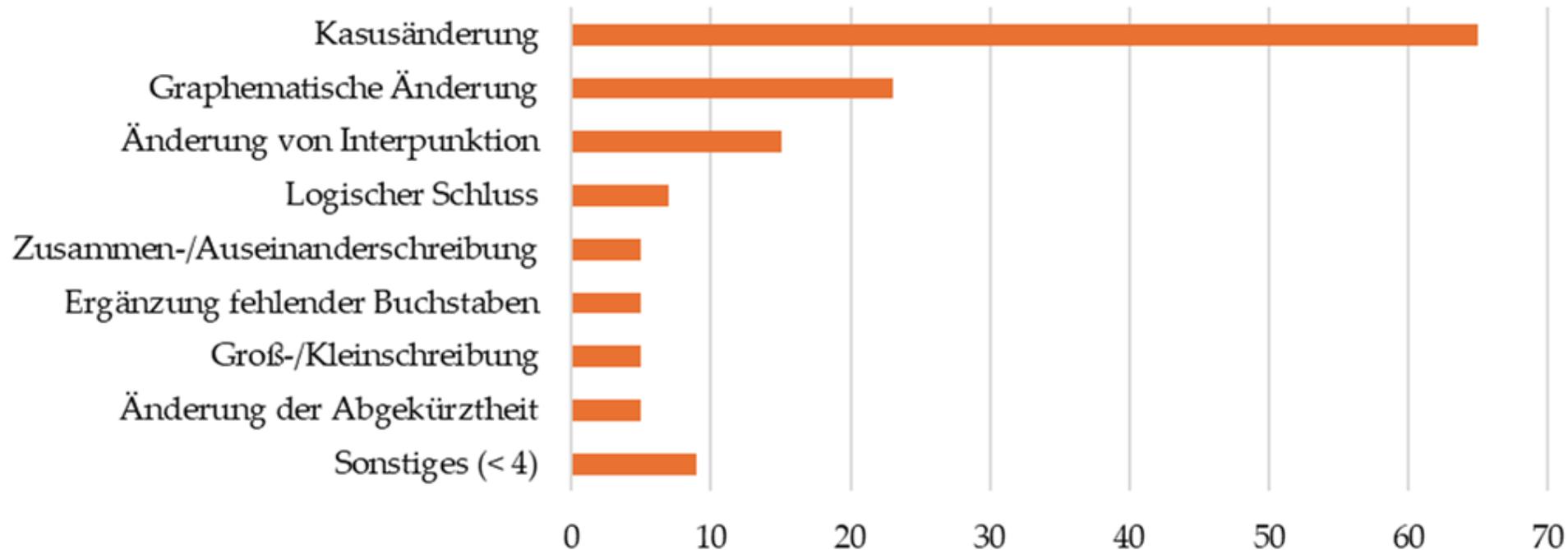
- Halluzinationen auf zwei Arten:
 - (1) Ausgabe inkorrektter Start- und Endpunkte
 - (2) Ausgabe von **Entitäten, die nicht (identisch) im Inputtext vorkommen**
- **139 Halluzinationen** des zweiten Typs bei 39.512 von GPT-3.5 gefundenen Entitäten → **geringe Halluzinationsquote von 0,35 %**



**Frage nach Systematik von Halluzinationen:
Muster in Fehlern erkennbar?**

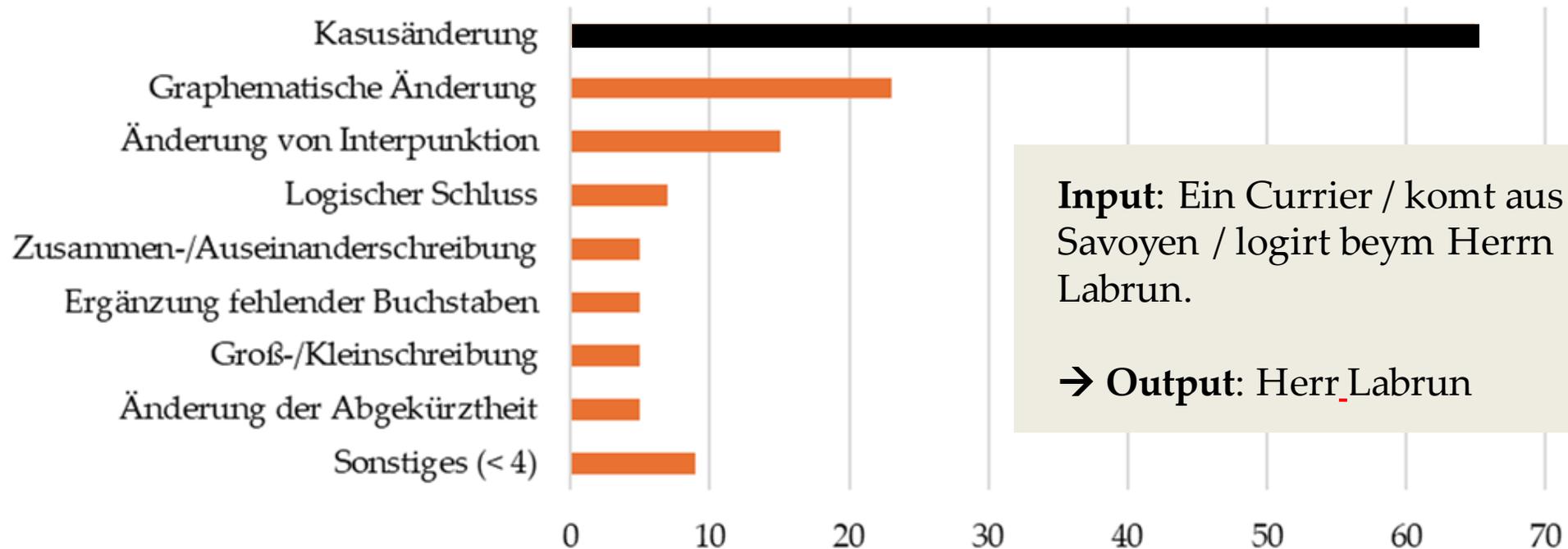
Halluzinationen

Arten von Halluzinationen



Halluzinationen

Arten von Halluzinationen

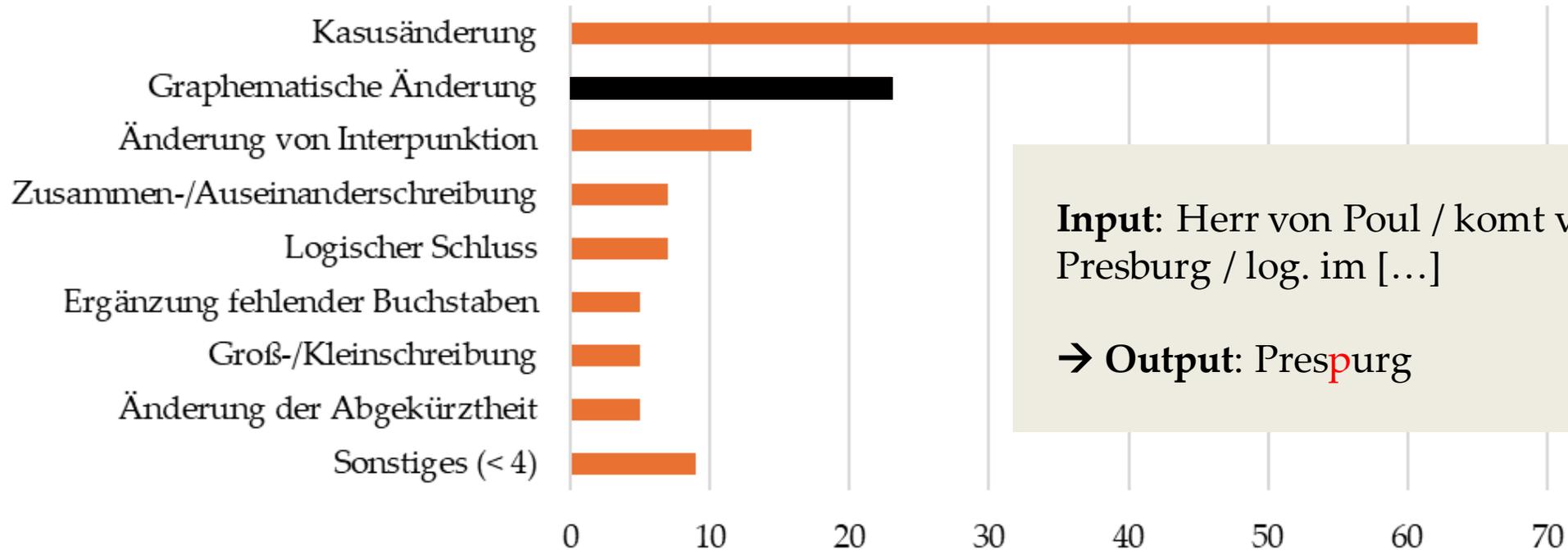


Input: Ein Currier / komt aus Savoyen / logirt beym Herrn Labrun.

→ **Output:** Herr Labrun

Halluzinationen

Arten von Halluzinationen

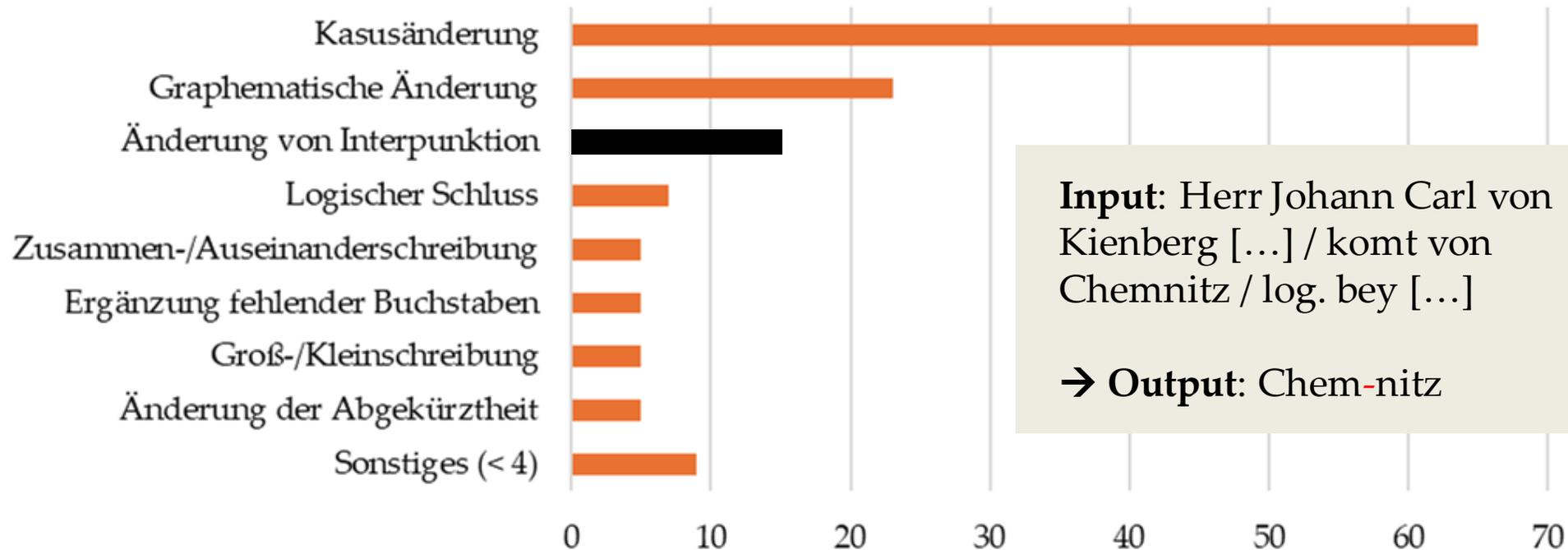


Input: Herr von Poul / komt von
Presburg / log. im [...]

→ **Output:** Presburg

Halluzinationen

Arten von Halluzinationen

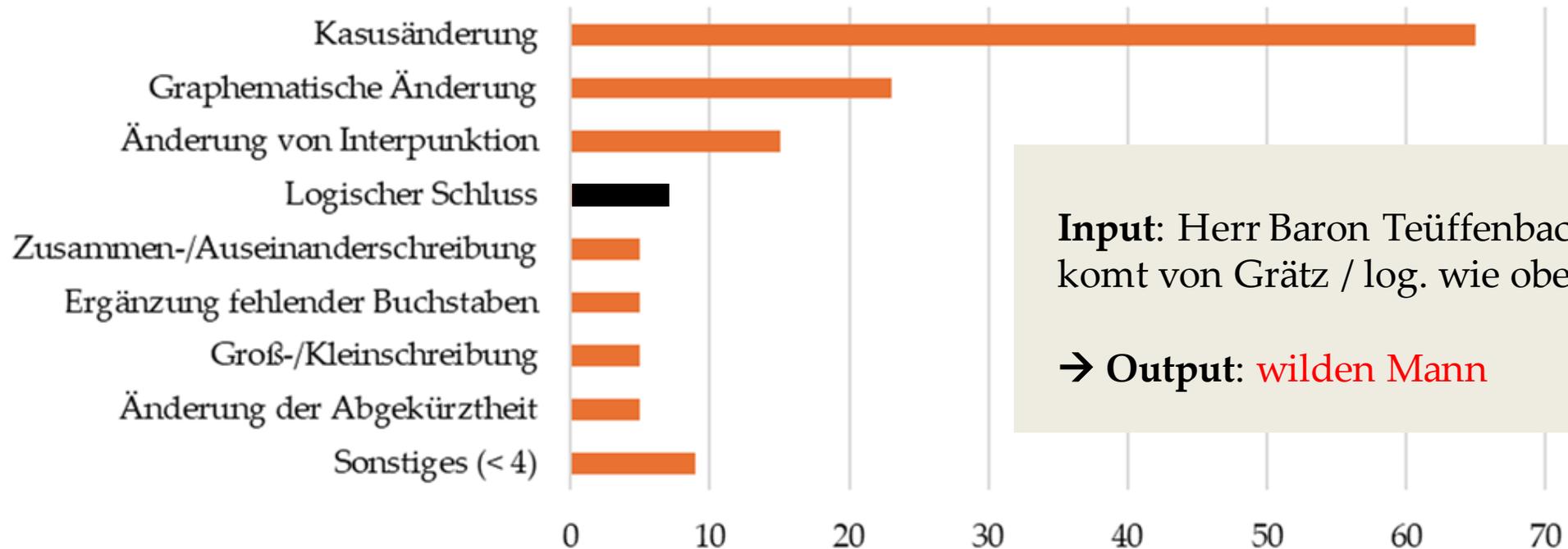


Input: Herr Johann Carl von
Kienberg [...] / komt von
Chemnitz / log. bey [...]

→ **Output:** Chem-nitz

Halluzinationen

Arten von Halluzinationen

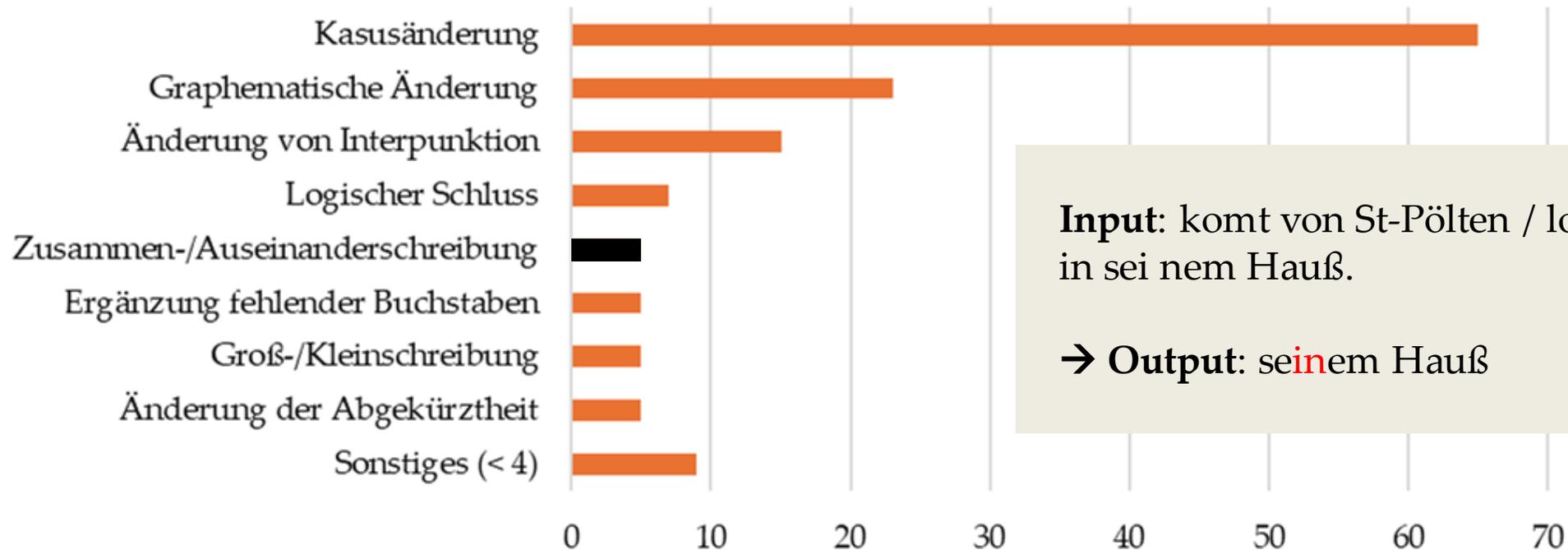


Input: Herr Baron Teuffenbach /
kommt von Grätz / log. wie oben.

→ **Output:** wilden Mann

Halluzinationen

Arten von Halluzinationen

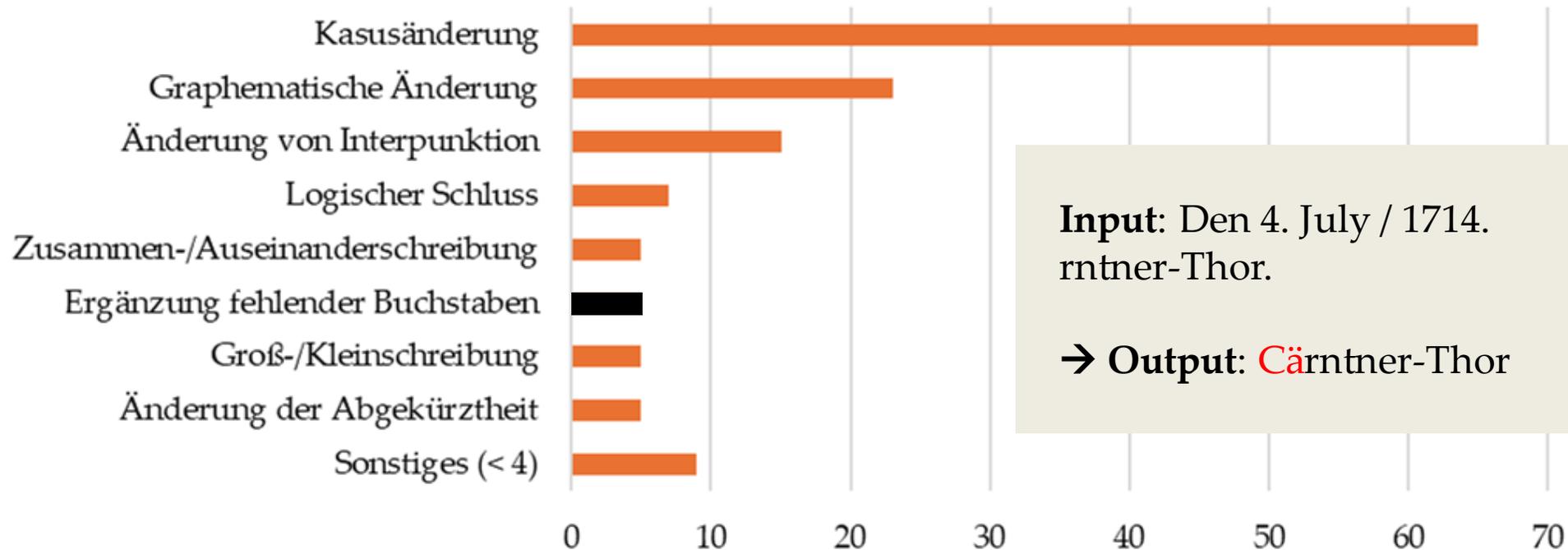


Input: komt von St-Pölten / log.
in sei nem Hauß.

→ **Output:** seinem Hauß

Halluzinationen

Arten von Halluzinationen

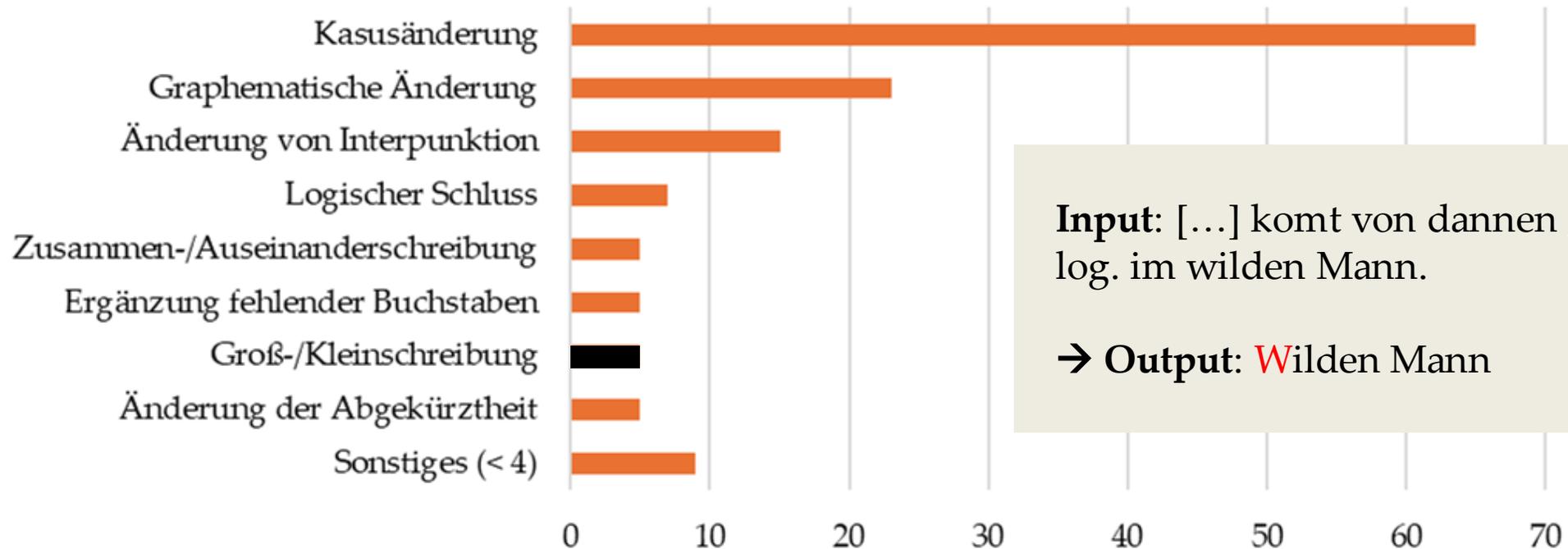


Input: Den 4. July / 1714.
rntner-Thor.

→ **Output:** Cärintner-Thor

Halluzinationen

Arten von Halluzinationen

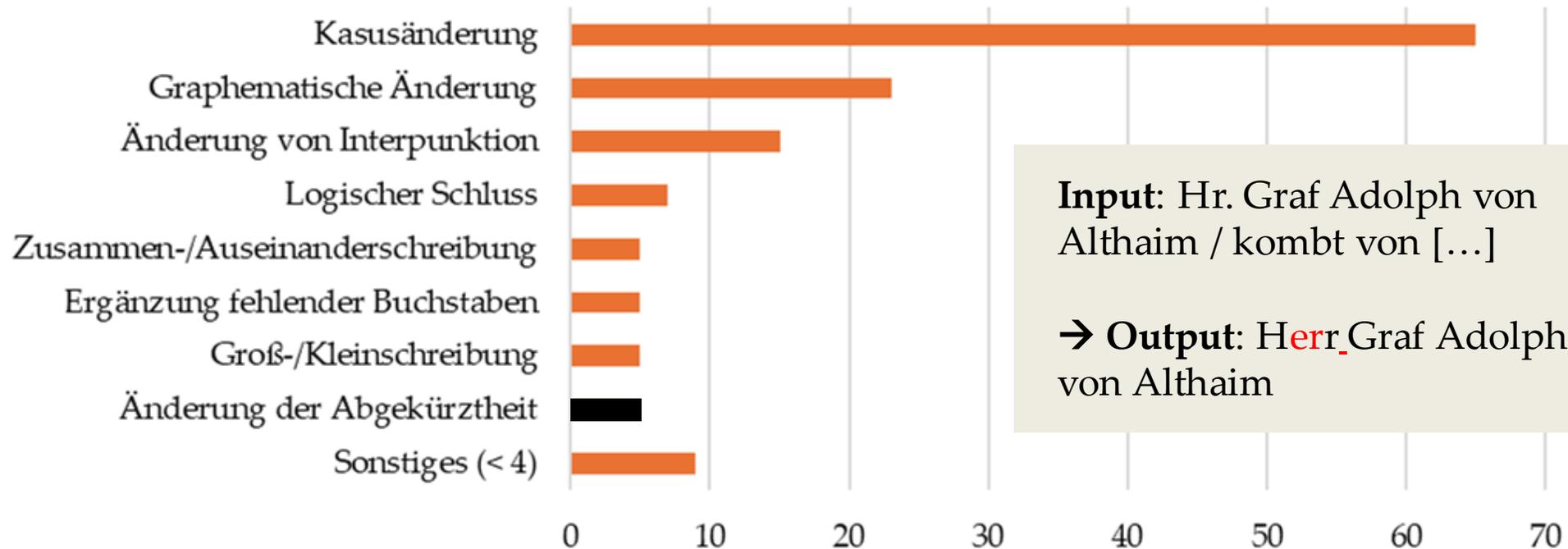


Input: [...] komt von dannen /
log. im wilden Mann.

→ **Output:** Wilden Mann

Halluzinationen

Arten von Halluzinationen

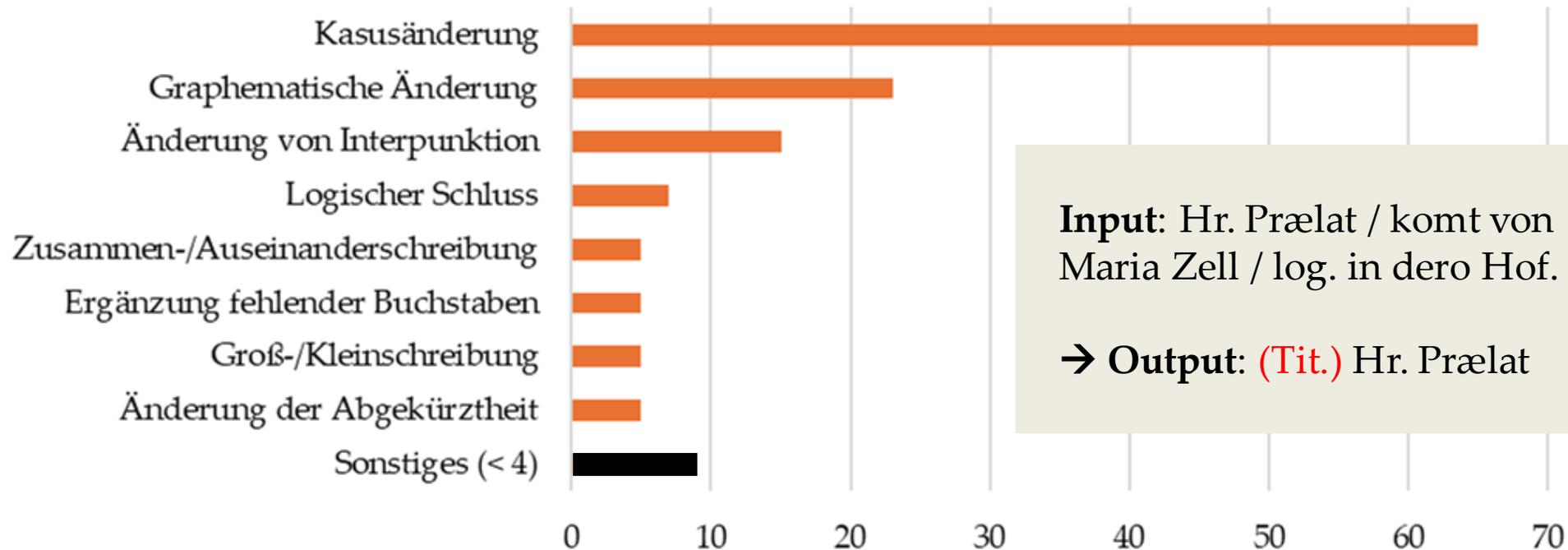


Input: Hr. Graf Adolph von
Althaim / kombt von [...]

→ **Output:** Herr Graf Adolph
von Althaim

Halluzinationen

Arten von Halluzinationen



Input: Hr. Prælat / komt von
Maria Zell / log. in dero Hof.

→ **Output:** (Tit.) Hr. Prælat

Halluzinationen

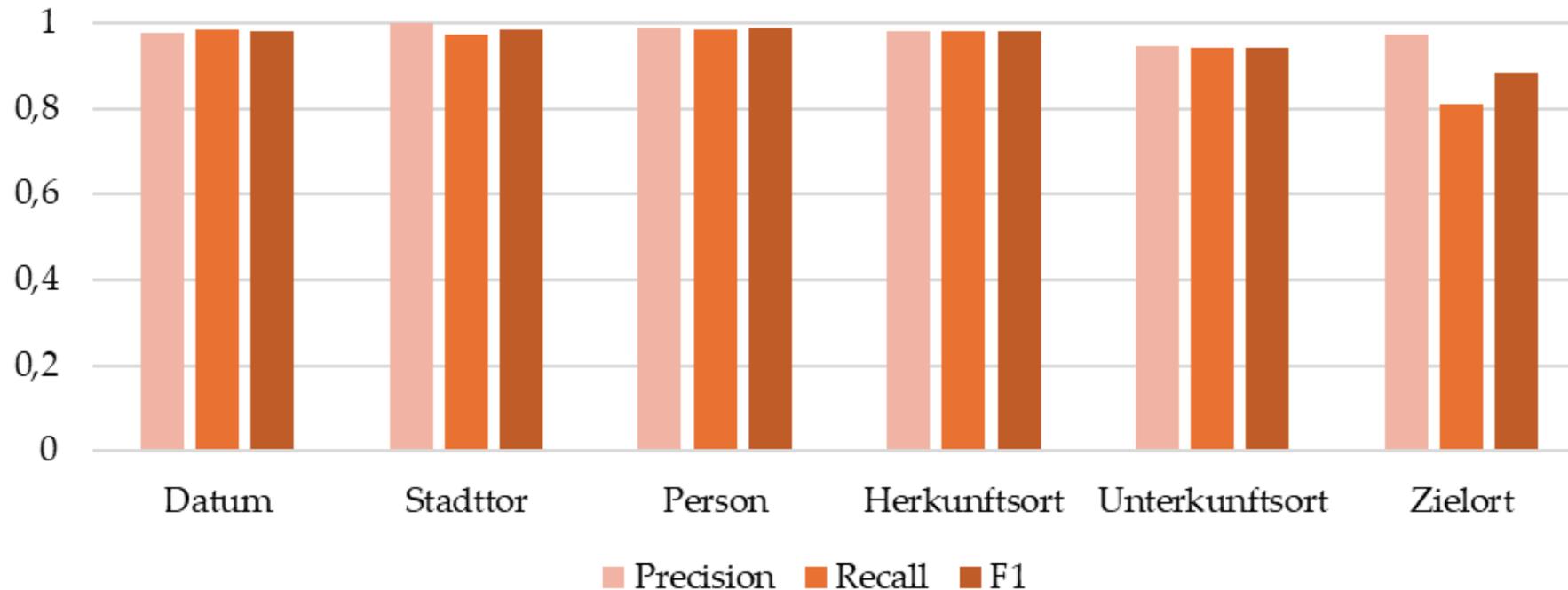


Bild generiert mit DALL-E

- Einzelne Arten von Halluzinationen abhängig nach Erkenntnisinteresse möglicherweise auch nutzbar (z.B. parallele Post-Korrektur bei OCR-Volltexten)
- **Herausforderung:** Einberechnung von Halluzinationen bei Evaluationsmetriken (z.B. F1-Score)
- Hier: **möglichst strikte Evaluation** = Halluzinationen als zur Gänze nicht gefundene Entitäten

Evaluation

Evaluationsergebnisse je Entitätenkategorie



Insgesamt

Precision: 0.98

Recall: 0.97

F1-Score: 0.97

Evaluation

- **Tendenziell selten grobe Fehler:**
 - z.B. nur 318 von 39.788 tatsächlichen Entitäten (= 0,79 %) gar nicht erkannt
- **In meisten Fällen kleine(re) Fehler – etwa:**
 - Artikel/Präposition mit extrahiert – z.B. *zu Gumpendorf* > *Gumpendorf*
 - Interpunktion fehlt – z.B. *30. Aug* > *30. Aug.*
 - Mehrere Herkunftsorte statt einem erkannt – z.B. *Gut, Steuermarck* > *Gut aus Steuermarck*

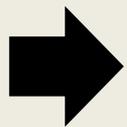
Ankunfft Hoch- und Niedriger Stands-Persohnen. Stuben-Thor stadttor den 4. April Datum .
Gestert spath kombt Herr Graf Lobkowitz Person von Prag Herkunftsort / logirt im
Nostitzischen Hauß auff der Hohen-Brucken Unterkunftsart . Rothen-Thurn stadttor . Herr
Wandt / Fourrier von Rehmischen Person / kombt gestern spath von Thiernkruth Herkunftsort /
logirt in der guldenen Ganß Unterkunftsart Herr Prangheim / Hof-Cammer-Rath Person / kombt
von Breßlau Herkunftsort / logirt beym schwarzen Thor Unterkunftsart . Stuben-Thor
stadttor den 5. dito Datum . Herr Peter Gradoni Person / kombt von Görtz Herkunftsort /
logirt in der Cärnter-Strassen im weissen Schwan Unterkunftsart . Cärnter-Thor stadttor .
Herr Zeiler / Kayserl. Courier Person / kombt von der Rabau vom Herrn General Heister
Herkunftsart / logirt im Hof-Sailerischen Hauß Unterkunftsart . Herr Baron von Schönau
Person / kombt von Venedig Herkunftsort / logirt im guldenen Hirschen Unterkunftsart .
Stuben-Thor stadttor den 6. dito Datum . Herr Graf Frantz Adam von Lamberg Person /

(Zwischen-)Fazit & Ausblick

Potenziale



- **Hohe Ergebnisqualität** für frühneuzeitliche, semi-strukturierte Zeitungstexte
- **Starke Anpassbarkeit** an Textmaterial und Erkenntnisinteresse (Labels, One-/Few-Shot-Ansatz, ...)

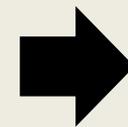


Vergleichbare Nutzbarkeit für unstrukturierte(re) Texte?

Limitationen



- **Inputlimit** → Aufteilen von Text
- **Halluzinationen** → geringer Anteil
- **Positionsbestimmung** notwendig
- **Zugänglichkeit** (API-Kosten)



Vergleichbare Ergebnisse kostenfreier Open-Source-Modelle?

Referenzen

- Mader-Kratky, A. / Resch, C. / Scheutz, M. (2019): „Das Wien[n]erische Diarium im 18. Jahrhundert. Neue Sichtweisen auf ein Periodikum im Zeitalter der Digitalisierung“. *Wiener Geschichtsblätter* 74(2), 93–113.
- Gius, E., Meister, J. C., Meister, M., Petris, M., Bruck, C., Jacke, J., Schumacher, M., Gerstorfer, D., Flüh, M., & Horstmann, J. (Hg.). *CATMA (6.5.0)*. <https://doi.org/10.5281/zenodo.6419805>
- OpenAI (Hg.). *OpenAI API*. <https://openai.com/blog/openai-api>
- Promptslab (Hg.). *Promptify*. <https://github.com/promptslab/Promptify>
- READ-COOP (Hg.). *Transkribus*. <https://readcoop.eu/de/transkribus>
- Resch, C. / Kampkaspar, D. (Hg.). *German Fraktur 18th Century – WrDiarium_M9*. <https://readcoop.eu/de/modelle/german-fraktur-18th-century/>
- *SpaCy*. <https://spacy.io>
- Nakayama, H. (2018): *seqeval. A Python framework for sequence labeling evaluation*. <https://github.com/chakki-works/seqeval>