

# Automatische Auflösung metonymisch gebrauchter Toponyme

[Gritta et al., 2017]

Leo Born

Hauptseminar: Figurative Language Resolution  
Prof. Dr. Katja Markert  
Institut für Computerlinguistik  
Ruprecht-Karls-Universität Heidelberg

03.07.2019

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion

# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion

# Was ist eine Metonymie?



Abbildung: *Apocalypse '42* [Schreckengost, 1942]

# Was ist eine Metonymie?

“Ersatz einer Benennung [...] durch eine verwandte Bezeichnung, die mit dem Gemeinten *im Unterschied zur Metapher* durch einen sachlichen (z.B. räumlichen, zeitlichen, kausalen) Zusammenhang bzw. durch semantische Kontiguität verknüpft ist”

[Bußmann, 2008]

# Arten von Metonymien

- Autor-Werk: *Pynchon lesen*
- Material-Produkt: *Seide tragen*
- Gefäß-Inhalt: *eine Flasche trinken*
- Ort-Bewohner: *Island trauert*
- Zeit-Zeitgenossen: *das Mittelalter dachte anders*
- Teil-Ganzes (aka *pars pro toto*): *das macht 10 EUR pro Nase*
- Subjektschub: *Dieses Referat behandelt das Papier...*

# Ursprünglich metonymische Alltagswörter

- abgeleitet von Personennamen:
  - *Algorithmus*
  - *Guillotine*
  - *röntgen*
  - *lynchen*
- abgeleitet von Toponymen:
  - *Cognac*
  - *Pils*
- abgeleitet von Firmen- bzw. Markennamen:
  - *Tempo*
  - *Tesa*
  - *Uhu*

# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion



# Motivation

- Question Answering und Opinion Mining
  - *The Sony is great.*
  - *Charlie Hebdo cost 12 lives.*
- Anaphernresolution
  - *China has agreed to let a UN investigator conduct an independent probe into [ . . . ] But it was unclear whether Beijing would meet past UN demands [ . . . ].*
  - Ask **Seat 19** whether **he** wants to swap.

# Motivation

- Geoparsing und Geographic Information Retrieval
  - *It is warm in **Berlin**.*
  - ***Berlin** announced tax cuts.*
- NER
  - ***Seoul**<sub>GPE</sub><sup>1</sup> agreed to peace talks.*
- Maschinelle Übersetzung
  - *He bought a **BMW**.* [ungültig im Chin.]
  - ***Washington** declared martial law.* [ungültig im Jap.]

---

<sup>1</sup>Laut Dokumentation von spaCy werden "Countries, cities, states" als **GPE** annotiert.

# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution**
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion

# Automatische Metonymieresolution

## Vancouver Welcomes You! Minimalist Location Metonymy Resolution [Gritta et al., 2017]

- Erkennung (nicht-)metonymischen Gebrauchs von Toponymen
- Erstellung eines neuen, balancierten Datensatzes
- LSTM-basierte NN-Architektur
- *predicate window* vom Kopf des Zielwortes als Kontext
- Kompetitive Ergebnisse ohne *feature engineering* oder externe Ressourcen

# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion

# SemEval-07 [Markert and Nissim, 2007]

- Enthält Annotation von Organisationen und Orten aus dem BNC
  - Relevant sind Toponyme: 925 Instanzen in Trainings-, 908 in Testdaten
  - Länderliste aus dem “The World Factbook” der CIA<sup>2</sup>
- Drei Klassen:
  - literal (80%)
  - metonymic (18%)
  - mixed (2%)
- “This seems to be the approximate natural distribution of the classes for location metonymy” [Gritta et al., 2017]

---

<sup>2</sup><https://www.cia.gov/library/publications/the-world-factbook/>

# SemEval-07 [Markert and Nissim, 2007]

- Shared Task on Metonymy Resolution
- Aufgabe: Klassifikation (Orte und Organisationen getrennt)
  - *coarse*: literal vs. non-literal
  - *medium*: literal vs. metonymic vs. mixed
  - *fine*: literal vs. mixed vs. metonymic patterns (place-for-people, place-for-product etc.)
- Fünf teilnehmende Systeme
  - Grundfeatures: POS-Tags, syntaktische Rollen etc.
  - Starke Nutzung externer Ressourcen wie WordNet, FrameNet, BNC
  - Beste *accuracy*:<sup>3</sup> 85.%

---

<sup>3</sup>Für Evaluationssetting *coarse*.

# ReLocaR

- Gleiche Klassen wie SemEval-07
  - literal (49%)
  - metonymic (49%)
  - mixed (2%)→ Balancierte Klassenverteilung
- Daten von Wikipedia Random Article API
  - Manuell kompilierte Länder- und Hauptstadtliste  $L$
  - Wenn  $l \in L$  in einem Satz → füge Satz Datensatz hinzu
- Training: 1.026 Instanzen
- Test: 1.000 Instanzen
  - Hohes IAA, aber ohne Angabe zu Kappa



# ReLocaR: Annotationsrichtlinien

- Ort ist **literal**, wenn...
  - Ortsverweis unbelebter Natur ist, d.h. auf ein konkretes Territorium verweist
    - *The Eiffel Tower is located in Paris.*
- Ort ist **metonymisch**, wenn...
  - Ortsverweis Belebtheit ausdrückt
    - *Washington's swift reaction surprised them.*
  - er stellvertretend für die dort befindlichen Personen steht
  - er stellvertretend für ein Produkt steht
  - er stellvertretend für eine Sportmannschaft steht
  - er stellvertretend für eine damit assoziierte politische Entität steht
  - er stellvertretend für ein dort stattgefundenes Ereignis steht

# ReLocaR: Unterschiede zu SemEval-07

- Keine feinere Unterscheidung metonymischer Instanzen
- Klassenverteilung ist balanciert
- Politische Entitäten werden als metonymisch eingestuft
  - SemEval:<sup>4</sup> *Britain's current account deficit* → literal
  - ReLocaR: *Britain's current account deficit* → metonymic
  - **Aber:** wenn Ort für assoziierte Personen/Organisation inkl. Regierungen steht, gilt Ort auch bei SemEval als metonymisch

---

<sup>4</sup>“Literal readings [...] comprise [...] political entity interpretations”

# CoNLL-2003

- Teilmenge der CoNLL-2003 NER Shared Task annotiert
  - 4.089 wörtliche, 2.126 metonymische Instanzen
- Nutzung als zusätzliche Trainingsdaten

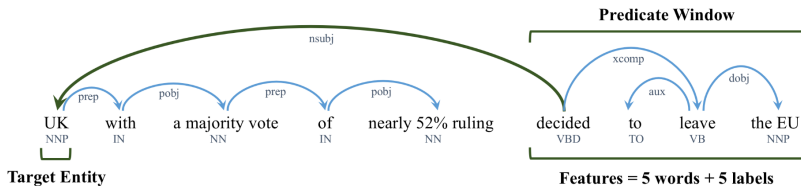
# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion

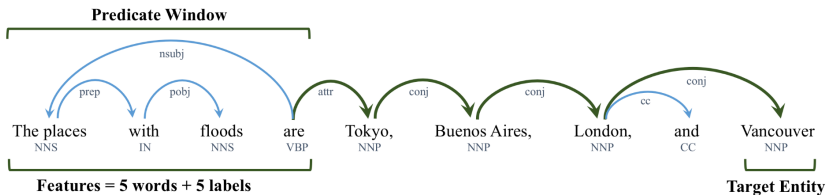
# Kontextreduktion per Predicate Window

- Featureextraktion anhand des **Predicate Window** (PreWin)
  - Erweiterung der Kontextreduktionsmethode von [Nissim and Markert, 2003]
- Input sind (bis zu) 5 Kontexttokens und -labels:
  - Parsing mit spaCy
  - Beginnend beim Kopf des Zielworts, extrahiere bis zu 5 Tokens und Labels
  - Richtung: vom Zielwort weg
  - Konjunktionen werden übersprungen
  - **Wichtig:** Das Zielwort wird nicht inkludiert

# Kontextreduktion per Predicate Window



# Kontextreduktion per Predicate Window



# Kontextreduktion per Predicate Window: Ausnahmen

- Dependenzlabels: *case, amod, compound*
- Possessivpartikel
  - ***Vancouver**'s security is the best in the world.*
- Adjektivmodifikation
  - *The pub is located in southern **Zambia**.*
- Komposita
  - *The gig takes place in downtown **Manhattan**.*

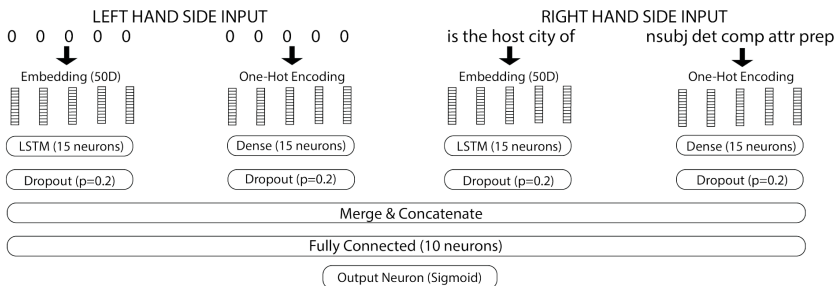


# Architektur

- LSTM-Architektur
  - Input: Kontext + Dependenzlabels
  - Output: **literal** | **non-literal**
- Zwei LSTMs, für linken und rechten Kontext
- One-hot layer für Dependenzlabels
- Nutzung vortrainierter, 50-dimensionaler GloVe embeddings [Pennington et al., 2014]
  - Vokabular: ersten 100.000 Wörter

# Architektur

- *Vancouver is the host city of the ACL 2017*



# Architektur: Warum zwei LSTMs?

- Verarbeitungsrichtung der LSTMs ist unterschiedlich
    - Linkes LSTM: links-rechts
    - Rechtes LSTM: rechts-links
- Jeweils höhere Gewichtung dem letzten (d.i. das dem Zielwort nächste) Wort beimessen
- Für jeden Satz bleibt eine Seite immer leer

# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion

# Baselines

- *Immediate*
  - 5 bis 10 Wörter zur rechten und linken als Kontext
- *Paragraph*
  - 50 Wörter zur rechten und linken als Kontext
- *Ensemble*
  - Mehrheitsvotum dreier Modelle, die auf unterschiedlichen Daten und/oder mit unterschiedlichen Parametern trainiert wurden
  - Code undurchsichtig, was die Unterschiede genau sind...
- [Nastase and Strube, 2013]
  - SVM und Wissensbasis (Wikipedia als Konzeptnetzwerk) mit Kontextinformation
  - SOTA mit 86.2%

# Experimente

- Training auf entweder SemEval-07-, CoNLL- oder ReLocaR-Daten
- Test auf SemEval-07 und ReLocaR
- Durchschnitt von zehn Testläufen
- Evaluation auf **literal** vs. **non-literal**
  - *mixed* und *metonymic* werden als *non-literal* zusammengefasst
  - accuracy und (pro Klasse) f-score

# Ergebnisse

Method	Training (Size)	Acc (STD)
PreWin	SemEval (925)	<b>62.4</b> (2.30)
Immediate 5	SemEval (925)	60.6 (2.34)
Immediate 10	SemEval (925)	59.2 (2.26)
Paragraph	SemEval (925)	58.0 (2.49)
PreWin	CoNLL (6,215)	<b>82.8</b> (0.46)
Immediate 5	CoNLL (6,215)	78.2 (0.61)
Immediate 10	CoNLL (6,215)	79.1 (0.76)
Paragraph	CoNLL (6,215)	79.5 (1.50)
PreWin	ReLocaR (1,026)	<b>83.6</b> (0.71)
Immediate 5	ReLocaR (1,026)	81.4 (1.34)
Immediate 10	ReLocaR (1,026)	81.3 (1.44)
Paragraph	ReLocaR (1,026)	80.0 (2.25)
Ensemble	ReLocaR/CoNLL	<b>84.8</b> (0.34)

Table 1: Results for ReLocaR data. Figures are averaged over 10 runs. STD - Standard deviation.

# Ergebnisse

Method	Training (Size)	Acc (STD)
PreWin	SemEval (925)	<b>83.1</b> (0.64)
Immediate 5	SemEval (925)	81.3 (1.11)
Immediate 10	SemEval (925)	81.9 (0.89)
Paragraph	SemEval (925)	81.3 (0.88)
PreWin	CoNLL (6,215)	<b>79.5</b> (0.34)
Immediate 5	CoNLL (6,215)	77.8 (1.47)
Immediate 10	CoNLL (6,215)	77.8 (1.22)
Paragraph	CoNLL (6,215)	77.2 (2.10)
PreWin	ReLocaR (1,026)	<b>69.0</b> (3.13)
Immediate 5	ReLocaR (1,026)	63.6 (5.42)
Immediate 10	ReLocaR (1,026)	64.2 (4.12)
Paragraph	ReLocaR (1,026)	64.4 (7.76)
Nastase et al.	SemEval (925)	<b>86.2</b> (N/A)
Ensemble	SemEval/CoNLL	84.6 (0.43)

Table 2: Results for SemEval data. Figures are averaged over 10 runs. STD - standard deviation.



# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion**
- 5 Diskussion

# Konklusion

- Kompetitive Ergebnisse trotz simpler Architektur
- Bereitstellung eines neues, balancierten Datensatzes
- Nicht nur Länder, sondern auch Hauptstädte
  
- Keine feinere Unterscheidung metonymischer Kategorien
- Immer noch lediglich Fokus auf Orten
- Evaluation des Ensemble-Modells

# Gliederung

- 1 Was ist eine Metonymie?
- 2 Warum uns Metonymien in der CL interessieren (sollten)
- 3 Automatische Metonymieresolution
  - Daten
  - Verfahren
  - Experimente
- 4 Konklusion
- 5 Diskussion

# Diskussion

- Was für Vor/Nachteile hätte es gehabt, das Zielwort in den Input mit aufzunehmen? Warum wurde es nicht gemacht?
- Wie gut können die Daten sein, wenn sie kontextlos annotiert werden?

# Bibliographie I



Bußmann, H. (2008).  
*Lexikon der Sprachwissenschaft*.  
Kröner, Stuttgart, 4. edition.



Gritta, M., Pilehvar, M. T., Limsopatham, N., and Collier, N. (2017).  
Vancouver Welcomes You! Minimalist Location Metonymy Resolution.  
In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 1248–1259, Vancouver, Canada.



Markert, K. and Nissim, M. (2007).  
SemEval-2007 Task 08: Metonymy Resolution at SemEval-2007.  
In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pages 36–41.



Nastase, V. and Strube, M. (2013).  
Transforming Wikipedia into a large scale multilingual concept network.  
*Artificial Intelligence*, 194:62–85.

# Bibliographie II



Nissim, M. and Markert, K. (2003).

Syntactic features and word similarity for supervised metonymy resolution.  
In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, volume 1, pages 56–63.



Pennington, J., Socher, R., and Manning, C. D. (2014).

GloVe: Global Vectors for Word Representation.

In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543, Doha, Qatar.



Schreckengost, V. (1942).

Apocalypse '42.