

CLARIN-D Showcase: Textual Emigration Analysis

André Blessing, Jens Stegmann, Jonas Kuhn
Institute for Natural Language Processing (IMS)
University of Stuttgart, Germany

SPONSORED BY THE



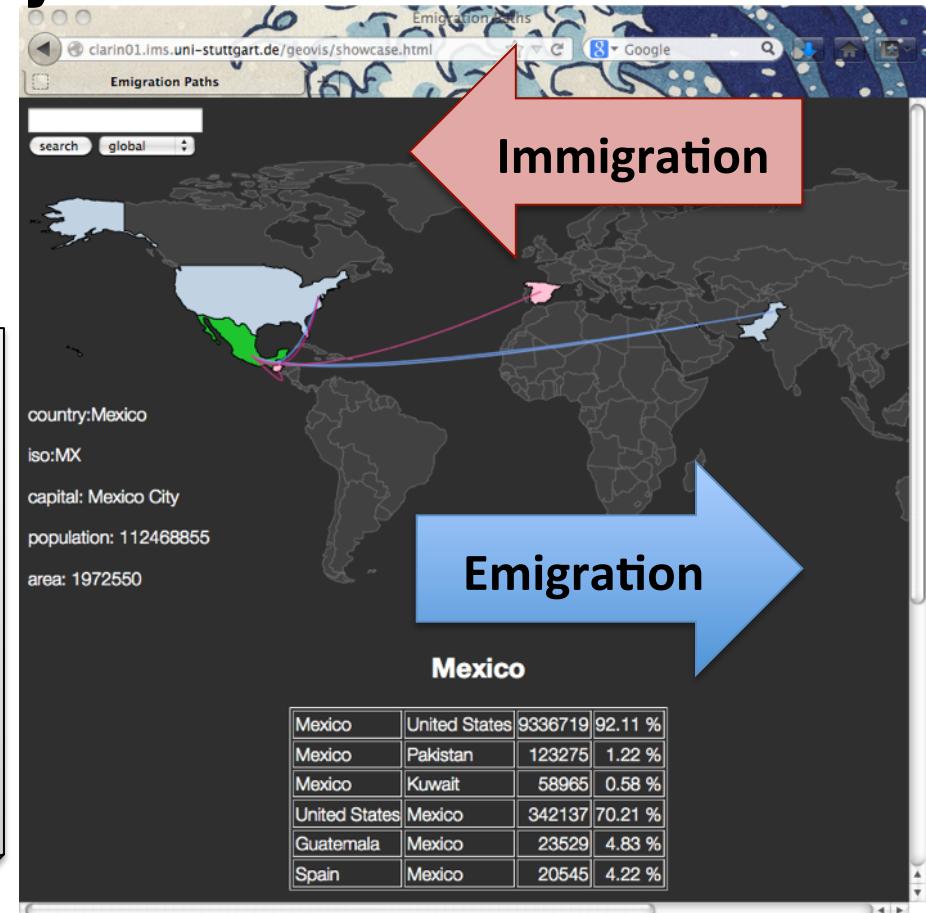
Federal Ministry
of Education
and Research

Textual Emigration Analysis

Extract and visualize
descriptions of emigration
moves from texts

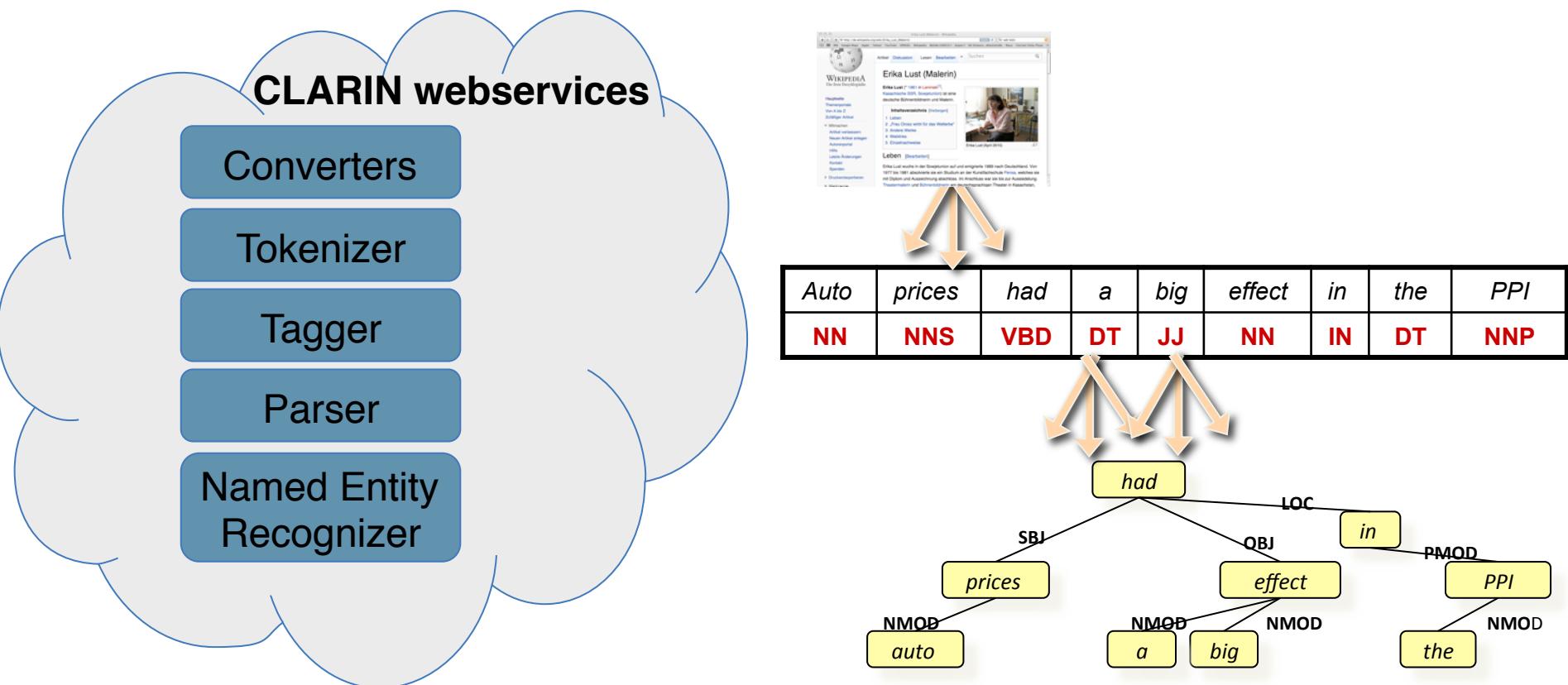
Erika Lust wuchs in Kasachstan

and
In
D
si
F
'E
'I
K
F
1906 übersiedelte
Grabmayr nach Wien.
'In 1906 Grabmayr
moved to Vienna.'



Text analysis for the eHumanities

- Exploit resource infrastructure & existing linguistic tools



Text analysis for the eHumanities

- ① Exploit resource infrastructure & existing linguistic tools

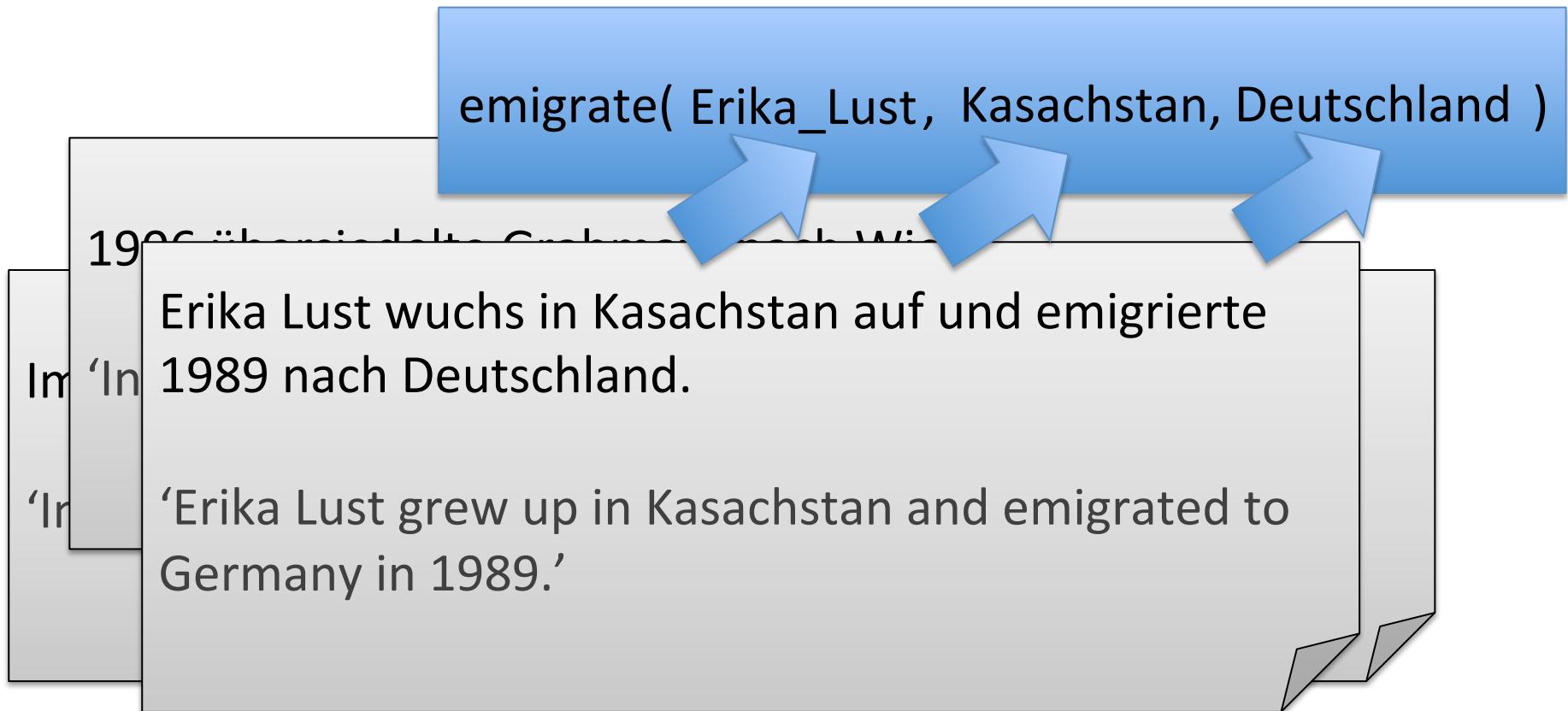
“enabling”:
facilitate innovative
research on larger-scale
data collections

Humanities:
diverse range of disciplines dealing
with aspects of text(s)
(a.o., philology, linguistics, history,
social sciences, ...)

- ② Accommodate discipline-specific concepts and relations
- ③ Aggregate information across textual sources
- ④ Link textual instances for critical reflection and correction
- ⑤ Adapt components to target language variety and domain

Sample task

Extract instances of **emigrate** relation from text



The diagram illustrates the extraction of an **emigrate** relation from a sentence. A blue callout box at the top right contains the triple: `emigrate(Erika_Lust, Kasachstan, Deutschland)`. Three blue arrows point from the words "emigrierte", "Kasachstan", and "Deutschland" in the German sentence below to the corresponding arguments in the triple. The German sentence is: "Erika Lust wuchs in Kasachstan auf und emigrierte 1989 nach Deutschland." Below it, the English translation is: "'Erika Lust grew up in Kasachstan and emigrated to Germany in 1989.'". The year "1989" is also highlighted with a blue arrow pointing to its position in the German sentence.

emigrate(Erika_Lust, Kasachstan, Deutschland)

1989

Erika Lust wuchs in Kasachstan auf und emigrierte 1989 nach Deutschland.

'Erika Lust grew up in Kasachstan and emigrated to Germany in 1989.'

Sample task

Agostino Novella – Wikipedia

http://de.wikipedia.org/wiki/Agostino_Novella

IMS Google Maps Apple Yahoo! SPIEGEL Wikipedia Beliebt (24585) Aupair Ski Schwarz...d hochstraße Maus Clarinet-Viola-Piano komoot Rout...gen planen

Artikel Diskussion Lesen Bearbeiten SUCHEN

Agostino Novella

Agostino Novella (* 28. September 1905 in Genua; † 15. September 1974 in Rom) war ein italienischer Politiker (Kommunist) und Gewerkschafter.

Inhaltsverzeichnis [Verbergen]

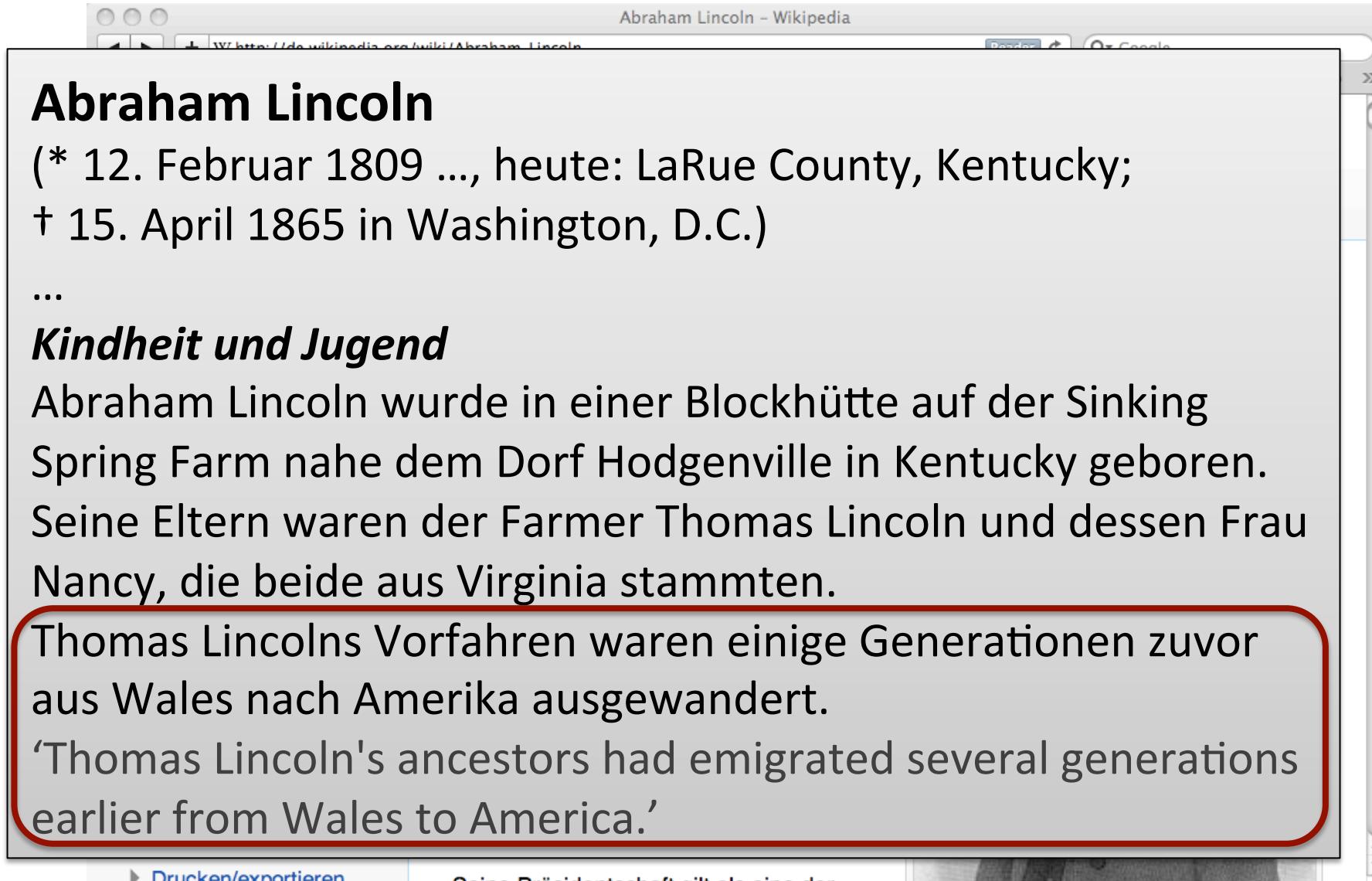
- 1 Leben
- 2 Werke
- 3 Literatur
- 4 Weblinks

Leben [Bearbeiten]

Novella, Metallarbeiter von Beruf, wurde 1923 Mitglied des Kommunistischen Jugendverbandes (FGCI; *Federazione Giovanile Comunista Italiana*) und 1924 Mitglied der Kommunistischen Partei Italiens (KPI). Von 1924 bis 1926 war er Sekretär der FGCI in der Region Ligurien und Mitglied des ZK der FGCI, sowie Stellvertreter des Sekretärs der Metallarbeitergewerkschaft FIOM (*Federazione Italiana Impiegati Operai Metallurgici*) in Genua.

1927 verurteilte ihn ein faschistisches Gericht zu vier Jahren Gefängnis 1931/32 und von 1934 bis 1938 war Novella Sekretär des ZK der FGCI. 1932 emigrierte er nach Frankreich. Im selben Jahr wurde er Mitglied des ZK der KPI. 1935 nahm er als Delegierter am VII. Kongress der Kommunistischen

A more complicated case...



Abraham Lincoln – Wikipedia

Abraham Lincoln

(* 12. Februar 1809 ..., heute: LaRue County, Kentucky;
† 15. April 1865 in Washington, D.C.)

...

Kindheit und Jugend

Abraham Lincoln wurde in einer Blockhütte auf der Sinking Spring Farm nahe dem Dorf Hodgenville in Kentucky geboren. Seine Eltern waren der Farmer Thomas Lincoln und dessen Frau Nancy, die beide aus Virginia stammten.

Thomas Lincolns Vorfahren waren einige Generationen zuvor aus Wales nach Amerika ausgewandert.
'Thomas Lincoln's ancestors had emigrated several generations earlier from Wales to America.'

Spectrum of analytical approaches

(1) General-purpose text analytics

- Meta data, document structure
- Look-up of names for geo-political entities

emigrate(Abraham_Lincoln, Kentucky, Wales)

emigrate(Abraham_Lincoln, Kentucky, Amerika)

Abraham Lincoln

(* 1809, Kentucky; † 1865 in Washington, D.C.)

Thomas Lincolns Vorfahren waren einige Generationen zuvor aus Wales nach Amerika ausgewandert.

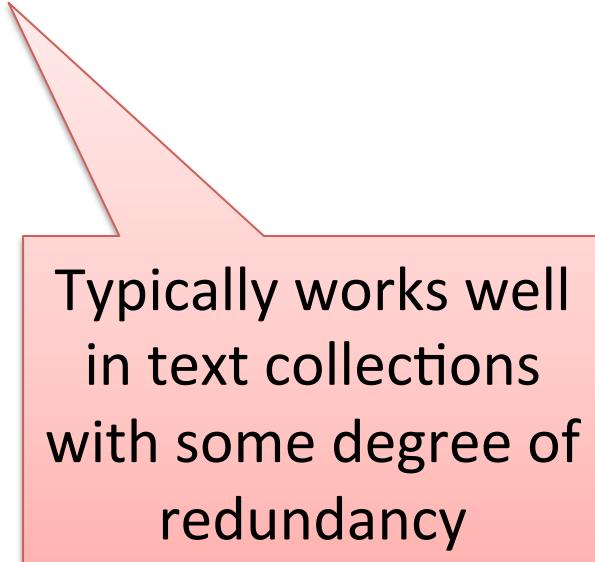
'Thomas Lincoln's ancestors had emigrated several generations earlier from Wales to America.'

Spectrum of analytical approaches

(1) General-purpose text analytics

- Meta data, document structure
- Look-up of names for geo-political entities

- ① Exploit linguistic tools
- ② Accommodate concepts
- ③ Aggregate information
- ④ Link textual instances
- ⑤ Adapt components



Typically works well
in text collections
with some degree of
redundancy

(2) Taking advantage of language analysis tools

- Tokenization, Pos-Tagging, Named Entity Recognition
- Identify semantic relation based on keywords

emigrate(Thomas_Lincoln, Wales, Amerika)

Abraham Lincoln

Thomas Lincolns Vorfahren waren einige Generationen zuvor aus Wales nach Amerika ausgewandert

'Thomas Lincoln's ancestors had emigrated several generations earlier from Wales to America.'

(2) Taking advantage of language analysis tools

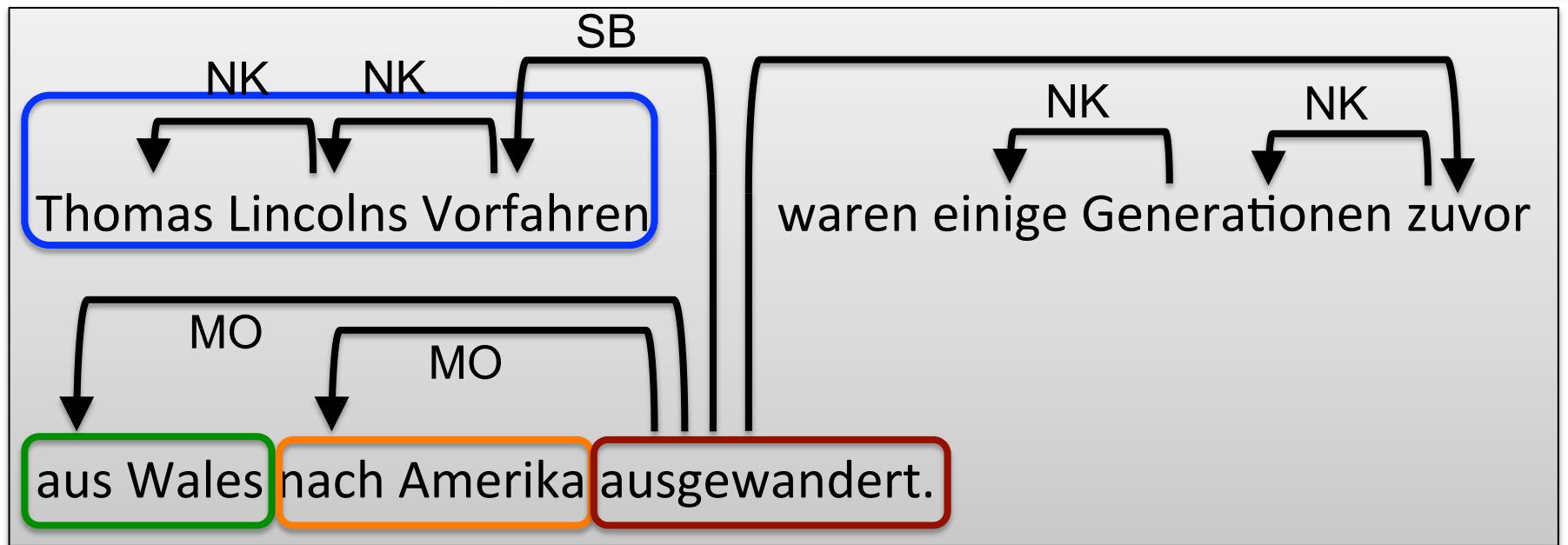
- Tokenization, Pos-Tagging, Named Entity Recognition
- Identify semantic relation based on keywords

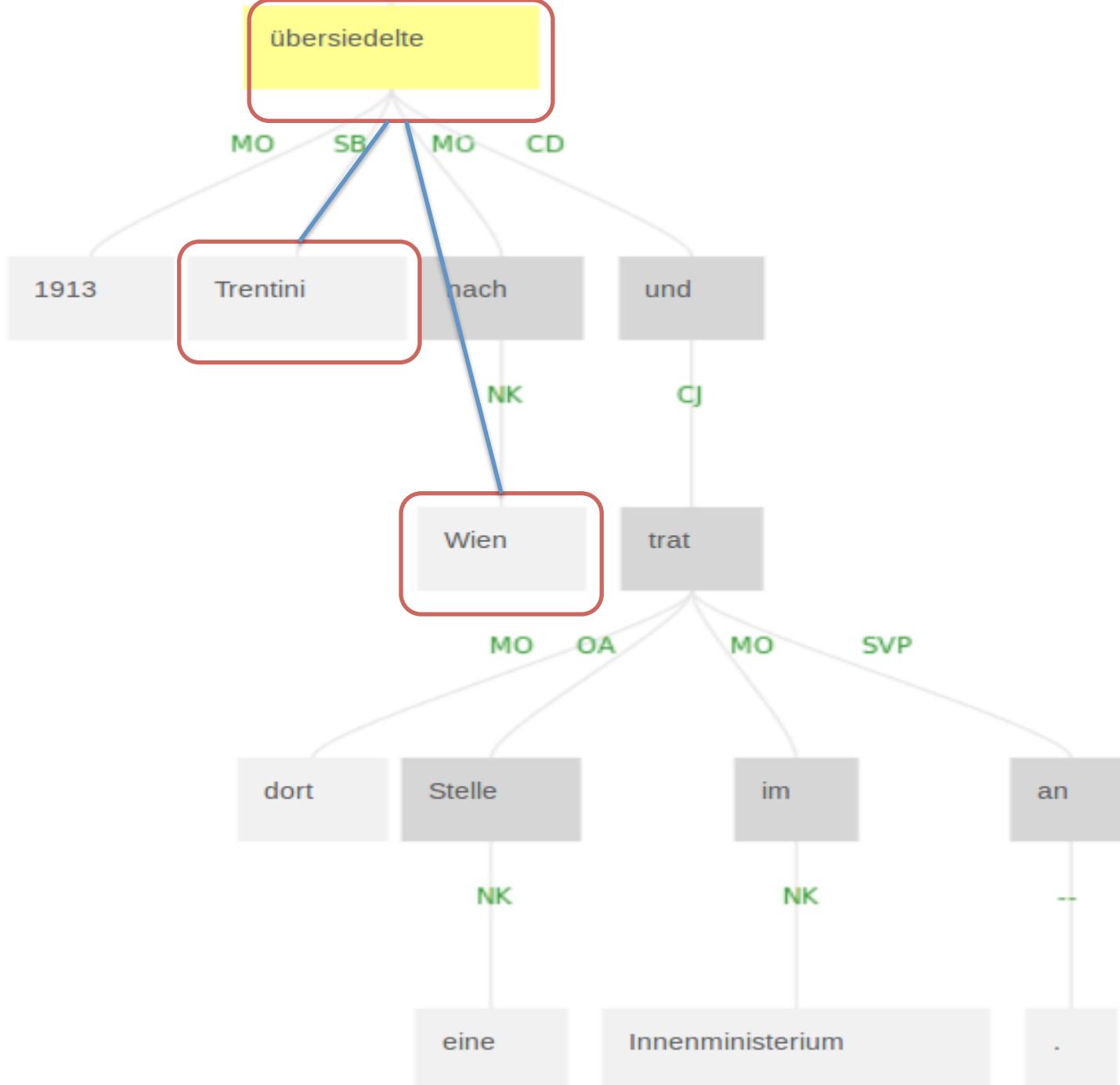
- ① Exploit linguistic tools
- ② Accommodate concepts
- ③ Aggregate information
- ④ **Link textual instances**
- ⑤ **Adapt components**

(3) Adaptable eHumanities toolkit

- Trainable relation extraction
- Adaptable/retrainable parser and additional tools

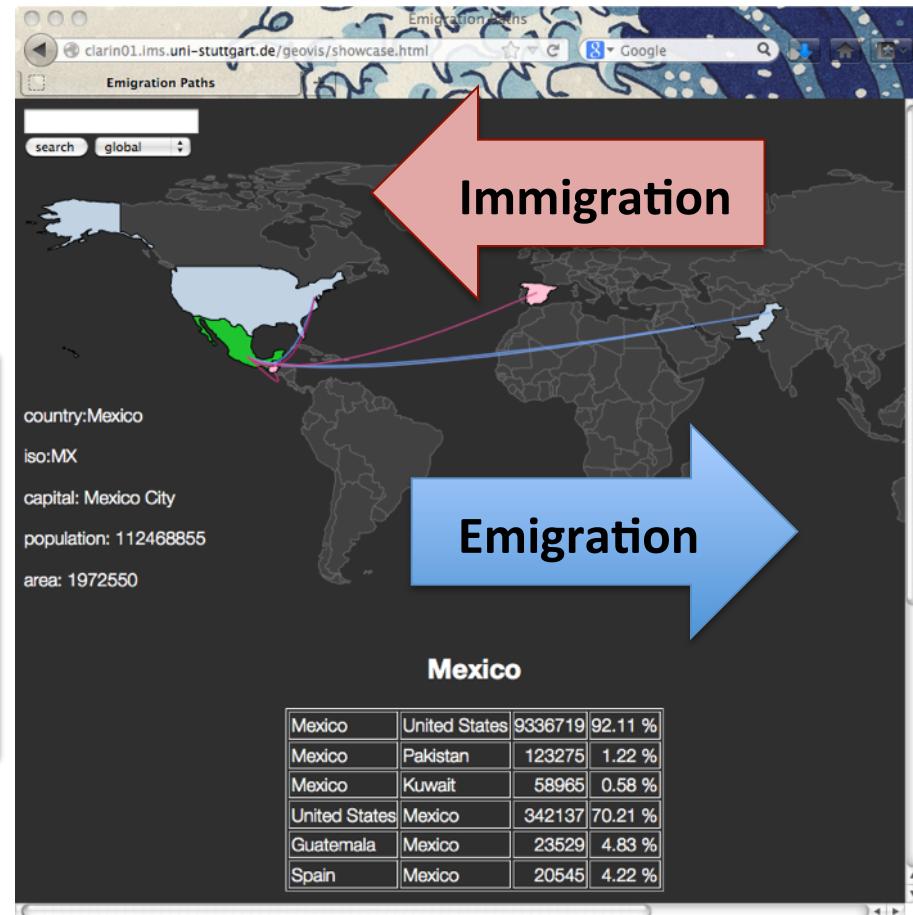
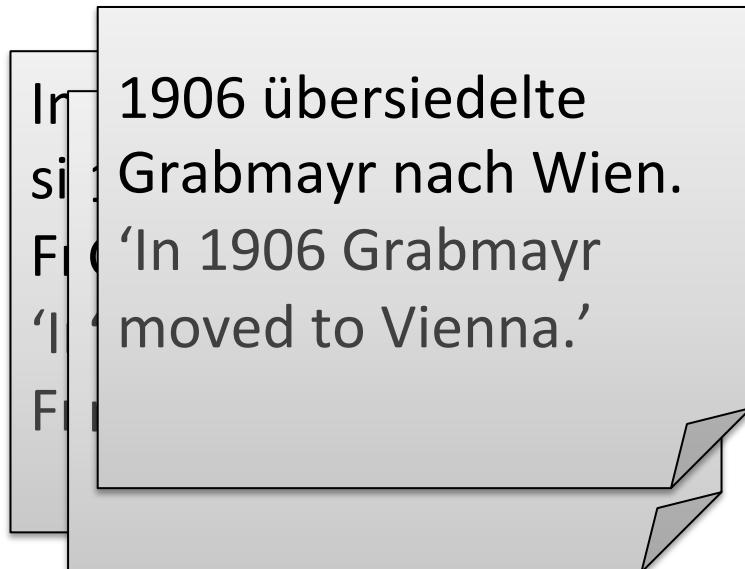
emigrate(TLs Vorfahren, Wales, Amerika)





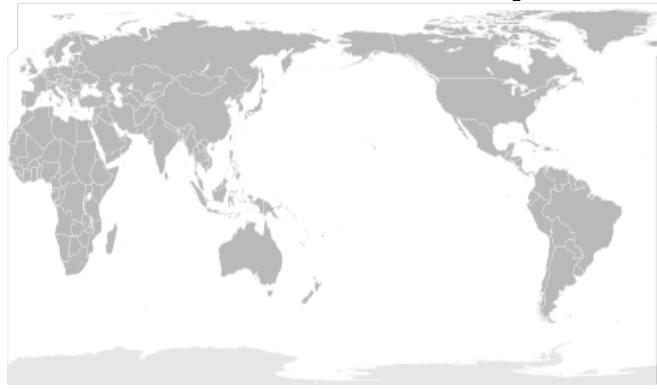
Textual Emigration Analysis

Extract and visualize
descriptions of
emigration moves
from texts



Raphaël—JavaScript Library

- Different maps can easily be integrated



World 2013



Europe 1938



Germany 1949

- NLP challenge
 - Ground named entities (toponyms) to maps
 - Simple approach
 - Use gazetteer:  GeoNames

Adaptability, user interaction

Gennadi Sosonko

Gennadi Sosonko (* 18. Mai 1943)

Sosonko emigrierte 1972 aus der UdSSR.
1978 wurde er nochmals Meister des
in Nijmegen, 1981 in Wijk aan Zee,

Sosonko emigrierte 1972 aus der UdSSR in die Niederlande.

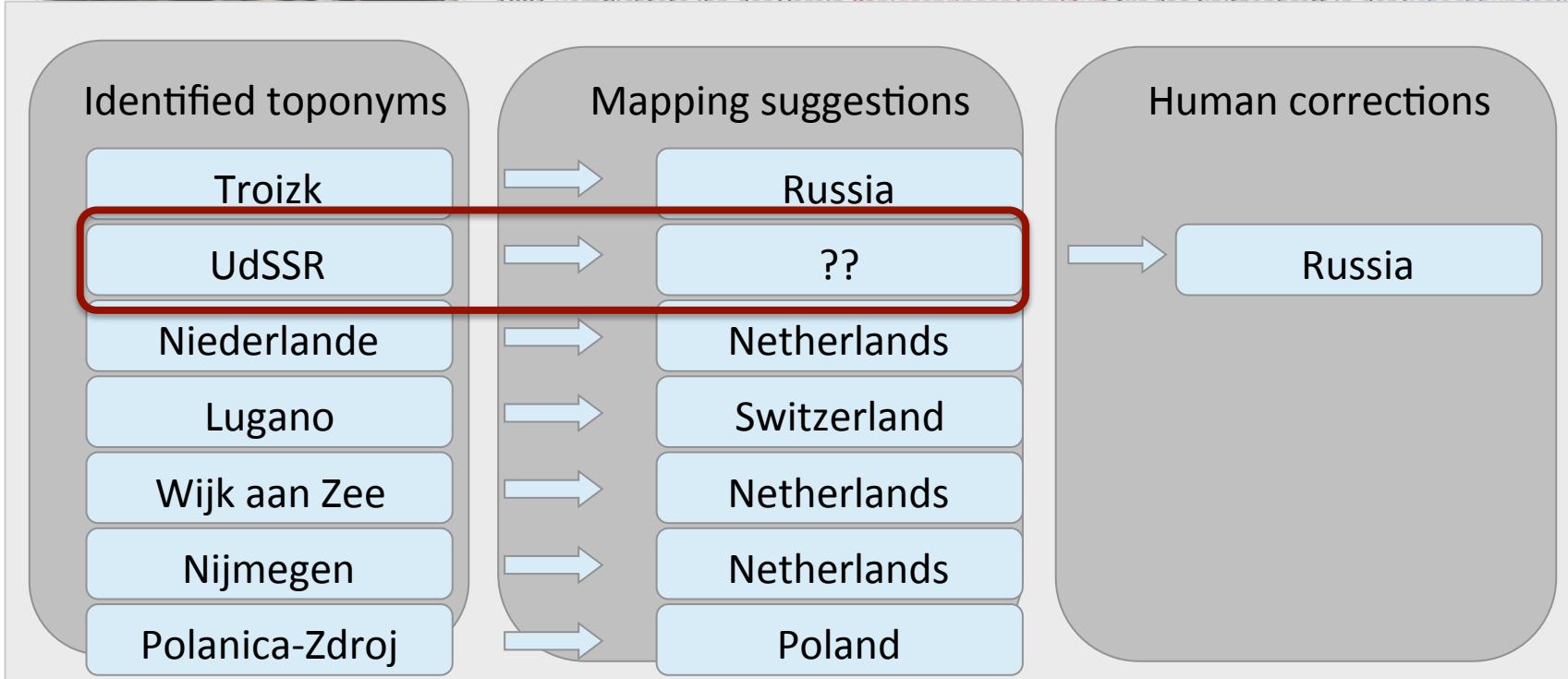
'Sosonko emigrated in 1972 from the USSR to the Netherlands.'



Sosonko spielte oft in der niederländischen Nationalmannschaft, darunter drei Schacholympiaden. Dort erzielte

60 Punkte aus 96 Partien (28 Siege, 4 Niederlagen, 64 Remis).

1981 gewann er wiederum den Titel eines Internationalen Meisters.



Outlook: Factuality

Das Ehepaar **wollte** eigentlich nach Amerika **auswandern**, aber die Geburt ihrer ... Kinder ließ sie ihre Pläne ändern.

'The couple actually wanted to **emigrate** to America, but the birth of their children made them change their plans.'

Den Zweiten Weltkrieg verbrachte er in Deutschland, weil er **nie** in die USA **auswandern wollte**.

'He spent WW II in Germany because he **never wanted to emigrate** to the USA.'

1968 im Prager Frühling marschierte die Rote Armee in die Tschechoslowakei ein, Kohoutek **entschied sich** darauf hin 1970 **zur Emigration** nach Deutschland.

'In 1968, in the Prague Spring the Red Army marched into Czechoslovakia; as a consequence, in 1970 Kohoutek **decided** in favour of **emigration** to Germany.'

(FCS)



...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

TEA application

- Interfaces to webservices
- User Interface

Parser (retrainable)

GeoGrounding (retrainable)

Relation extractor (retrainable)

CLARIN webservices

Converters

Tokenizer

Tagger

Parser

Named Entity
Recognizer

**TCF exchange
format**

- 1 Exploit linguistic tools
- 2 Accommodate concepts
- 3 Aggregate information
- 4 Link textual instances
- 5 Adapt components

Thank you!

CLARIN-D Showcase: Textual Emigration Analysis

André Blessing, Jens Stegmann, Jonas Kuhn
Institute for Natural Language Processing (IMS)
University of Stuttgart, Germany

SPONSORED BY THE



Federal Ministry
of Education
and Research