

Catching the Common Cause:

Extraction and Annotation of Causal Relations and their Participants

Ines Rehbein & Josef Ruppenhofer

3. April 2017

LAW XI

New resource for causality in German

- Building a resource for describing causality in German
 - following Dunietz et al. (2015)...
 - ...but adding FN flavor to PDTB style analysis of arguments

(1) Dieser verrückte Mächtegernpolitiker beschert uns durch
 This crazy pseudopolitician bestows us through
 seine Kriegsgelilitheit noch mehr Pack, Gesockse,
 his lusting of the war even more vermin, riff-raff,
 Frauenbelästiger und Schmarotzer ...
 molesters of women and parasites ...

“Through his lusting for war^{Cause}, this crazy pseudopolitician^{Actor}
bestows upon us^{Affected} even more vermin, riff-raff, molesters of
 women and parasites^{Effect}”

Annotation scheme (Dunietz et al. 2015)

- causality types
 1. CONSEQUENCE
 2. MOTIVATION
 3. PURPOSE
 4. INFERENCE

- arguments
 1. CAUSE
 2. EFFECT
 3. ACTOR_{new}
 4. AFFECTED_{new}

- degrees of causality
 1. facilitate
 2. inhibit

Annotation scheme (Dunietz et al. 2015)

- causality types
 1. CONSEQUENCE
 2. MOTIVATION
 3. PURPOSE
 4. INFERENCE

- arguments
 1. CAUSE
 2. EFFECT
 3. ACTOR_{new}
 4. AFFECTED_{new}

- degrees of causality
 1. facilitate
 2. inhibit

smoking_{Cause} causes cancer_{Effect}
 CONSEQUENCE, facilitate

he_{Actor} causes me_{Affected}
 to stand on the heights_{Effect}
 CONSEQUENCE, facilitate

A resource for describing causality in German

- Lexicon
 - **Task 1:** detect causal triggers to be included in the lexicon
- Corpus
 - **Task 2:** extract instances for that trigger to be included in the corpus → training data for system development

A resource for describing causality in German

- Lexicon
 - **Task 1:** detect causal triggers to be included in the lexicon
- Corpus
 - **Task 2:** extract instances for that trigger to be included in the corpus → training data for system development

This work

- Identification of transitive causal verbs:

<NOUN1> *causes* <NOUN2>

Related work

- Girju (2003)
 - identified instances of noun-verb-noun causal relations in WordNet glosses N1 *starvation* causes *bonyness* N2
 - uses extracted noun pairs to search a large corpus for causal verbs that link one of the noun pairs from the list

Related work

- Girju (2003)
 - identified instances of noun-verb-noun causal relations in WordNet glosses *N1 starvation causes bonyness N2*
 - uses extracted noun pairs to search a large corpus for causal verbs that link one of the noun pairs from the list
- Hidey & McKeown (2016)
 - use monolingual comparable corpora to find alternative lexicalisations for causal DRs
- Versley (2010)
 - bootstrapping approach for a connective dictionary
 - distribution-based heuristics on word-aligned German-English text

Related work

- Girju (2003)
 - identified instances of noun-verb-noun causal relations in WordNet glosses *N1 starvation causes bonyness N2*
 - uses extracted noun pairs to search a large corpus for causal verbs that link one of the noun pairs from the list
- Hidey & McKeown (2016)
 - use monolingual comparable corpora to find alternative lexicalisations for causal DRs
- Versley (2010)
 - bootstrapping approach for a connective dictionary
 - distribution-based heuristics on word-aligned German-English text
- Our approach:
 - knowledge-lean, based on parallel multi-lingual text (EN-GE)
 - focussing on causal events and their participants

Extraction of causal triggers from parallel text

- English-German part of Europarl (Koehn 2005)
 - > 1,9 mio parallel sentences
 - Preprocessing:
 - word-aligned (Berkeley Aligner, Denero & Klein 2007)
 - dependency-parsed (Chen & Manning 2014; Lei et al. 2014)

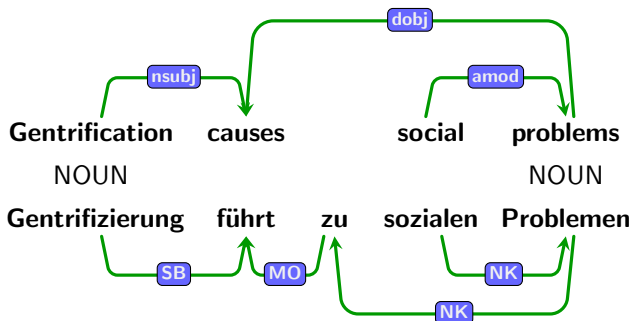
Extraction of causal triggers from parallel text

- English-German part of Europarl (Koehn 2005)
 - > 1,9 mio parallel sentences
 - Preprocessing:
 - word-aligned (Berkeley Aligner, Denero & Klein 2007)
 - dependency-parsed (Chen & Manning 2014; Lei et al. 2014)

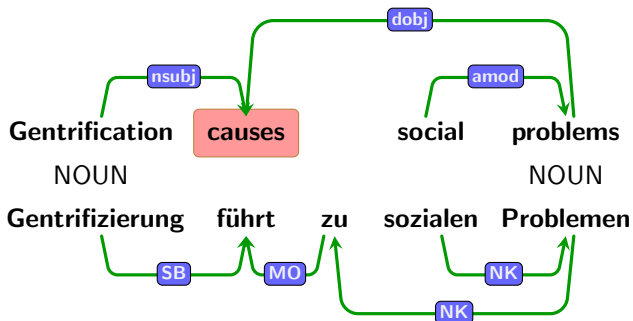
2 Steps

1. **Noun pair extraction** from parallel text
2. Extraction of **causal German triggers**

Step 1: Noun pair extraction

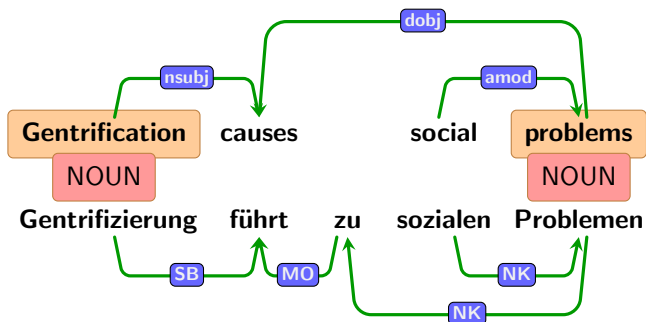


Step 1: Noun pair extraction



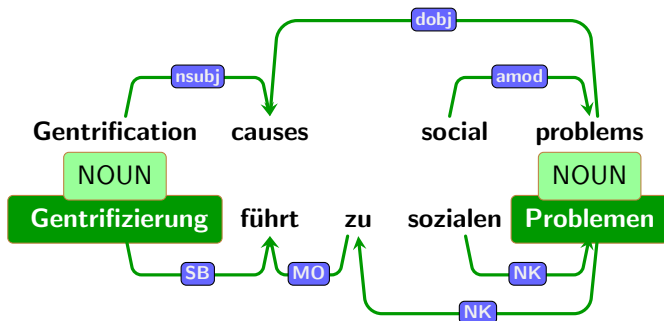
step 1-1: select English sentences that include *cause*

Step 1: Noun pair extraction



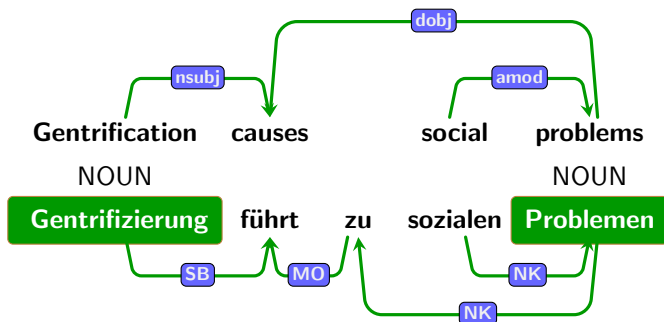
step 1-2: *nsubj*, *dobj* realised as nouns

Step 1: Noun pair extraction



step 1-3: *nsubj*, *dobj* aligned to nouns in German

Step 1: Noun pair extraction



step 1-4: extract noun pair <Gentrifizierung, Problem>

Extraction of causal triggers from parallel text

Step 1

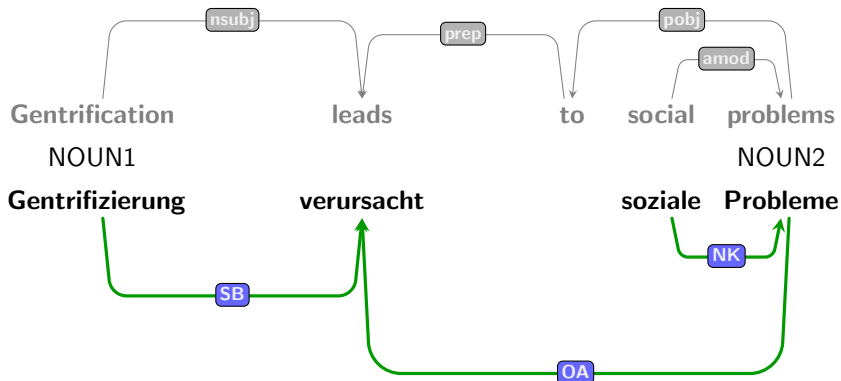
- Noun pair extraction from parallel text
- Input: word-aligned, dependency-parsed English-German data
- Output: list of German noun pairs ✓

Step 2

- Use noun pairs to identify potentially causal triggers in monolingual German text

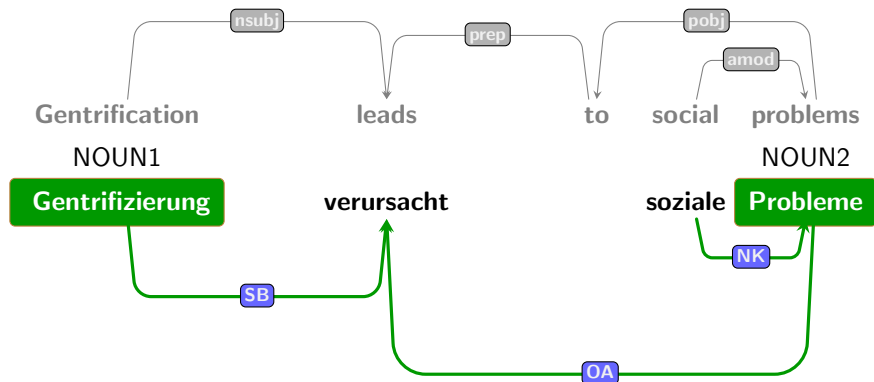
Step 2: Extraction of German triggers

Input: noun pair list from step 1



Step 2: Extraction of German triggers

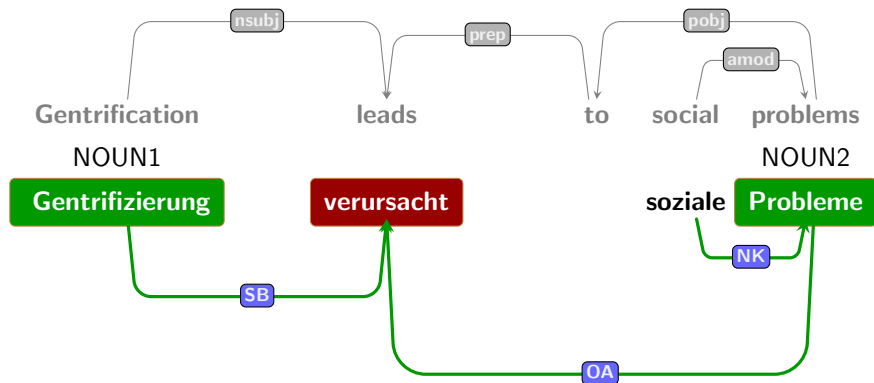
Input: noun pair list from step 1



step 2-1: select German sentences that include such a noun pair

Step 2: Extraction of German triggers

Input: noun pair list from step 1



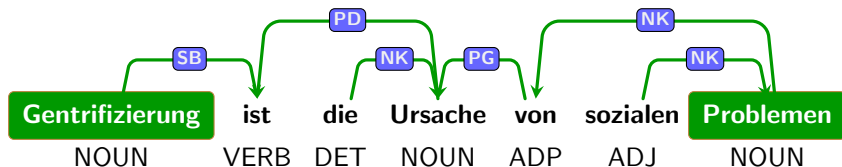
step 2-2: select the verb that links the two nouns

Extraction from parallel text: settings

- Settings
 1. **strict**: restrict noun pairs to sentences where aligned German nouns are also subj and dobj

Extraction from parallel text: settings

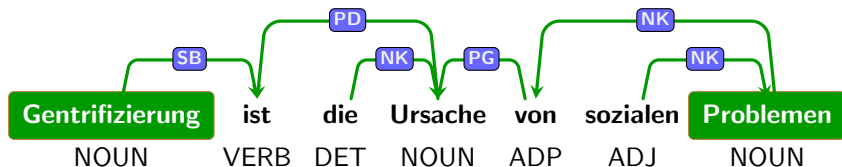
- Settings
 - strict**: restrict noun pairs to sentences where aligned German nouns are also subj and dobj
 - loose**: ignore grammatical function of German nouns, extract all nouns that are linked to the same verb (max. distance 3)



Extraction from parallel text: settings

- Settings

- strict:** restrict noun pairs to sentences where aligned German nouns are also subj and dobj
- loose:** ignore grammatical function of German nouns, extract all nouns that are linked to the same verb (max. distance 3)



- boost:** generalise over seen noun pairs using word2vec embeddings (Reimers et al. 2014)

boost: generalise over seen noun pairs

- For each noun pair,
 - compute cosine similarity to each noun in the embeddings
 - add 10 nouns most similar to noun 1
 - add 10 nouns most similar to noun 2
(to avoid noise, use similarity threshold of 0.75)
- ⇒ create new noun pairs

Unsicherheit	uncertainty	cos
Verunsicherung	uncertainty	0.87
Unsicherheiten	insecurities	0.80
Unzufriedenheit	dissatisfaction	0.78
Frustration	frustration	0.78
Nervosität	nervousness	0.75
Ungewissheit	incertitude	0.74
Unruhe	concern	0.74
Ratlosigkeit	perplexity	0.74
Überforderung	excessive demands	0.73

Extraction from parallel text: results

- Step 1:

setting	# types
noun pairs	343
+ word2vec	585

Extraction from parallel text: results

- Step 1:

setting	# types
noun pairs	343
+ word2vec	585

- Step 2:

causal triggers		# types
setting 1	(strict)	22
+ setting 2	(loose)	79
+ setting 3	(boost)	100

- strict:** mostly direct translations of *cause*, $\approx 75\%$ causal
- loose:** more variety, also some support verb constructions
- boost:** detects a high number of verbal triggers, at low cost

Annotation study

- (2) Die bevorstehende **Wiedereröffnung** des Tunnels
The imminent reopening of the tunnel
hat allerdings viele **Kontroversen** entfacht
has indeed many controversies new ignited.
“The imminent reopening of the tunnel has, however,
revived a number of controversies.”

Annotation study

- (3) Die bevorstehende **Wiedereröffnung** des Tunnels
The imminent reopening of the tunnel
hat allerdings viele **Kontroversen** entfacht
has indeed many controversies new ignited.
“The imminent reopening of the tunnel has, however,
revived a number of controversies.”
1. Does *entfachen* (ignite) have a *causal* meaning
(in this particular context)?

Annotation study

- (4) Die bevorstehende **Wiedereröffnung** des Tunnels
The imminent reopening of the tunnel
hat allerdings viele **Kontroversen** entfacht
has indeed many controversies new ignited.
“The imminent reopening of the tunnel has, however,
revived a number of controversies.”

1. Does *entfachen* (ignite) have a *causal* meaning (in this particular context)?
2. If *causal*:
 - argument of NOUN1: Wiedereröffnung (reopening)?
 - argument of NOUN2: Kontroversen (controversies)?

Annotation study – IAA

	no.	% agr.	κ
causal	427	94.4	0.78
NOUN1	352	94.9	0.74
NOUN2	352	99.1	0.95

Table : Annotation of causal transitive verbs: number of instances and IAA (percentage agreement and Fleiss' κ) for a subset of the data (427 sentences, 352 instances annotated as causal by both annotators)

Error analysis

- Disagreements mostly systematic, easy to resolve
 - *causal vs. non-causal*
 - (5) zum Ausdruck bringen
to the expression bring
“to express something”

Error analysis

- Disagreements mostly systematic, easy to resolve
 - *causal vs. non-causal*
 - (7) zum Ausdruck bringen
to the expression bring
“to express something”
 - *Cause vs. Actor*
 - Organisations: commission, European Union, member state
 - Animals, ghosts, ...
 - (8) das Gespenst des Kommunismus
the spectre of communism

Sum-up

Done so far

- Lexicon: 100 causal triggers (mostly verbs)
- Corpus: 1337 annotated instances (720 causal, 617 non-causal)

Sum-up

Done so far

- Lexicon: 100 causal triggers (mostly verbs)
- Corpus: 1337 annotated instances (720 causal, 617 non-causal)

Work in progress

- Build the lexicon:
 - Identify more causal triggers
(connectives, nouns, prepositions, alternative lexicalisations ...)
 - Add another language (triangulation)
 - Learn to filter out noise

Sum-up

Done so far

- Lexicon: 100 causal triggers (mostly verbs)
- Corpus: 1337 annotated instances (720 causal, 617 non-causal)

Work in progress

- Build the lexicon:
 - Identify more causal triggers
(connectives, nouns, prepositions, alternative lexicalisations ...)
 - Add another language (triangulation)
 - Learn to filter out noise

Future work

- Annotate more data (crowdsourcing)
- Use data to develop a causal tagger

Thanks for listening!

If this talk *CAUSE* has left you *AFFECTED*
puzzled *EFFECT*,
there is time for questions 😊

Referenzen

- Altenberg and Tapper (1998): The use of adverbial connectors in advanced Swedish learners' written English. *Learner English on Computer*.
- Becher (2011): Explicitation and implicitation in translation. A corpus-based study of English- German and German-English translations of business texts. Ph.D. thesis, Universität Hamburg.
- Blum-Kulka (1986): Shifts of cohesion and coherence in translation. In Juliane House and Shoshana Blum-Kulka, editors, *Interlingual and intercultural communication*, pages 17–35. Gunter Narr, Tübingen.
- Field and Yip (1992): A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal*, 23 (1), 15–28.
- Granger and Tyson (1996): Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17-27.
- Hoek & Zufferey (2015): The role of expectedness in the implicitation and explicitation of discourse relations. *Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT)* (pp. 41–46). Association for Computational Linguistics, Lisbon, Portugal.
- Lapshinova-Koltunski (2015): *Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT)*, Lisbon, Portugal.
- Narita et al. 2004): The Use of Linking Adverbials in the English Essay Writing of Japanese EFL Learners. In *Proceedings of ASIALEX 2003* (pp. 440–444).
- Sanders (2005): Coherence, causality and cognitive complexity in discourse. In M. Aurnague & M. Bras (Eds.), *Proceedings of the First International Symposium on the Exploration and Modelling of Meaning* (pp. 31–46). Toulouse, France: Universite de Toulouse-le-Mirail.
- Teich (2003): *Cross-Linguistic Variation in System und Text. A Methodology for the Investigation of Translations and Comparable Texts*. Mouton de Gruyter, Berlin.
- Vinay and Darbelnet (1958): *Stylistique Comparee du Francais et de l'Anglais. Methode de Traduction*. Didier, Paris.