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 Möchtegernpolitiker besichert uns durch seine Kriegsgeilheit noch mehr Pack, Gesockse, seine Lust auf den Krieg noch mehr vermin, riff-raff, Frauenbelästiger und Schmarotzer... molesters of women and parasites...

“Through his lusting for war, this crazy pseudopolitician bestows upon us even more vermin, riff-raff, molesters of women and parasites.”
Annotation scheme (Dunietz et al. 2015)

- causality types
  1. Consequence
  2. Motivation
  3. Purpose
  4. Inference

- arguments
  1. Cause
  2. Effect
  3. Actor\textit{new}
  4. Affected\textit{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
  4. Affected

- 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>`
Motivation
Extraction method
Annotation study
Conclusions

Related work

- Girju (2003)
  - identified instances of noun-verb-noun causal relations in WordNet glosses: `starvation causes bonyness`
  - 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
  - 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
    \[ N1 \textit{starvation} \textit{causes} \textit{bonyness} \ N2 \]
  • 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 monilingual 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)
  • focusing 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

Gentrification causes social problems

Gentrifizierung führt zu sozialen Problemen
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

**Motivation**

**Extraction method**

**Annotation study**

**Conclusions**

---

step 1-4: extract noun pair  &lt;Gentrifizierung, Problem&gt;
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

- Gentrification
  - *NOUN1*
  - Gentrifizierung
  - *SB*
  - *OA*

- leads
  - *VERB*
  - verursacht

- to
  - *prep*
  - *pobj*
  - sozial
  - *amod*
  - *NK*

- social problems
  - *NOUN2*
  - Probleme
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

  Gentrifizierung ist die Ursache von sozialen Problemen

  NOUN VERB DET NOUN ADP ADJ NOUN SB PD PG NK NK

  3. **boost**: generalise over seen noun pairs using word2vec embeddings (Reimers et al. 2014)
Extraction from parallel text: settings

- **Settings**
  1. **strict**: restrict noun pairs to sentences where aligned German nouns are also subj and dobj
  2. **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**
  1. **strict**: restrict noun pairs to sentences where aligned German nouns are also subj and dobj

  2. **loose**: ignore grammatical function of German nouns, extract all nouns that are linked to the same verb (max. distance 3)

  Gentrifizierung ist die Ursache von sozialen Problemen

  - NOUN  - VERB  - DET  - NOUN  - ADP  - ADJ  - NOUN

  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

<table>
<thead>
<tr>
<th>Unsicherheit</th>
<th>uncertainty</th>
<th>cos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verunsicherung</td>
<td>uncertainty</td>
<td>0.87</td>
</tr>
<tr>
<td>Unsicherheiten</td>
<td>insecurities</td>
<td>0.80</td>
</tr>
<tr>
<td>Unzufriedenheit</td>
<td>dissatisfaction</td>
<td>0.78</td>
</tr>
<tr>
<td>Frustration</td>
<td>frustration</td>
<td>0.78</td>
</tr>
<tr>
<td>Nervosität</td>
<td>nervousness</td>
<td>0.75</td>
</tr>
<tr>
<td>Ungewissheit</td>
<td>incertitude</td>
<td>0.74</td>
</tr>
<tr>
<td>Unruhe</td>
<td>concern</td>
<td>0.74</td>
</tr>
<tr>
<td>Ratlosigkeit</td>
<td>perplexity</td>
<td>0.74</td>
</tr>
<tr>
<td>Überforderung</td>
<td>excessive demands</td>
<td>0.73</td>
</tr>
</tbody>
</table>
Extraction from parallel text: results

- Step 1:

<table>
<thead>
<tr>
<th>setting</th>
<th># types</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun pairs</td>
<td>343</td>
</tr>
<tr>
<td>+ word2vec</td>
<td>585</td>
</tr>
</tbody>
</table>
Extraction from parallel text: results

- **Step 1:**

<table>
<thead>
<tr>
<th>setting</th>
<th># types</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun pairs</td>
<td>343</td>
</tr>
<tr>
<td>+ word2vec</td>
<td>585</td>
</tr>
</tbody>
</table>

- **Step 2:**

<table>
<thead>
<tr>
<th>causal triggers</th>
<th># types</th>
</tr>
</thead>
<tbody>
<tr>
<td>setting 1 (strict)</td>
<td>22</td>
</tr>
<tr>
<td>+ setting 2 (loose)</td>
<td>79</td>
</tr>
<tr>
<td>+ setting 3 (boost)</td>
<td>100</td>
</tr>
</tbody>
</table>

  - **strict:** mostly direct translations of *cause*, ≈75% causal
  - **loose:** more variety, also some support verb constructions
  - **boost:** detects a high number of verbal triggers, at low cost
(2) Die bevorstehende \textit{Wiedereröffnung} des Tunnels hat allerdings viele Kontroversen \textit{entfacht}.

“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 hat allerdings viele Kontroversen entfacht

“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

<table>
<thead>
<tr>
<th></th>
<th>no.</th>
<th>% agr.</th>
<th>(\kappa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>causal</td>
<td>427</td>
<td>94.4</td>
<td>0.78</td>
</tr>
<tr>
<td>NOUN1</td>
<td>352</td>
<td>94.9</td>
<td>0.74</td>
</tr>
<tr>
<td>NOUN2</td>
<td>352</td>
<td>99.1</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table: Annotation of causal transitive verbs: number of instances and IAA (percentage agreement and Fleiss’ \(\kappa\)) 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


