Spotting, Collecting and Documenting NPIs

Frank Richter
Jan-Philipp Soehn
Beata Trawiński

SFB 441
Universität Tübingen

Workshop on Negation and Polarity
University of Tübingen, Germany
March 08–10, 2007
Goals of the project A5

- Extract NPIs from partially annotated corpora
  → Improve the database for German NPIs
- Theoretical interpretation in the HPSG/LRS framework
  → Formulation of a theory of collocational restrictions
- Sample grammar implementation using the TRALE/CLLRS system
Outline

1. Introduction
2. A little background information
3. Extraction of NPI candidates
4. Modification to capture multi-word NPIs
5. Refining even more
6. Towards a collection of NPIs
Spotting, Collecting and Documenting NPIs
Part I: Spotting NPIs in partially annotated corpora – without batting an eyelid!

Frank Richter
Jan-Philipp Soehn
Beata Trawiński

SFB 441
Universität Tübingen

Workshop on Negation and Polarity
University of Tübingen, Germany
March 08–10, 2007
Spotting NPIs

- Spotting NPIs in partially annotated corpora
- Goal: Get a list of **NPI candidates**, that can be processed by linguists
- Not: Get a validated and exhaustive list of German NPIs
- Idea: Negative Polarity is a collocational phenomenon (van der Wouden 1997).
Outline

1. Introduction
2. A little background information
3. Extraction of NPI candidates
4. Modification to capture multi-word NPIs
5. Refining even more
6. Towards a collection of NPIs
A little background information on NPIs

- great variety in form and function
- not bound to a particular syntactic or semantic category
- single-word vs. multi-word NPIs
  (sonderlich (‘particularly’) vs. wahrhaben wollen (‘accept sth. to be true’))
- non-polysemous vs. polysemous
  (wahrhaben wollen vs. einen Schimmer haben (‘have the slightest idea’))
- Documentation of German NPIs relatively poor:
  Kürschner(1983) gives a list of 344 German items (NPIs?) whereas e.g. Hoeksema(2005) describes about 700 Dutch NPIs.
- Subclasses regarding strength of negativity of their licensors (Zwarts 1997)
NPI classification according to Zwarts (1997)

Negative contexts

<table>
<thead>
<tr>
<th>classical</th>
<th>regular</th>
<th>minimal</th>
<th>questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>nicht, ohne dass</td>
<td>nichts</td>
<td>höchstens</td>
<td>?</td>
</tr>
<tr>
<td>nicht glauben, dass</td>
<td>niemand</td>
<td>kaum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>kein</td>
<td>wenig</td>
<td></td>
</tr>
<tr>
<td>anti-morphic</td>
<td>anti-additive</td>
<td>downward monotonous</td>
<td>non-veridical</td>
</tr>
</tbody>
</table>

NPI

<table>
<thead>
<tr>
<th>Negation</th>
<th>classical</th>
<th>regular</th>
<th>minimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>superstrong</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>strong</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>weak</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. A little background information
3. Extraction of NPI candidates
4. Modification to capture multi-word NPIs
5. Refining even more
6. Towards a collection of NPIs
The Extraction algorithm

- **Corpus**
- **Sentence list**
- **Lemma list**

- **Reduce & add**
- **Count**

- **NPI candidates**
- **Statistics**
Our corpus: TüPP-D/Z

- *Tübingen Partially Parsed Corpus of Written German* (TüPP)
- based on the daily newspaper *die tageszeitung* (*taz*)
- Lemmatized, POS-Tags, Chunks, clause boundaries, ...
- ca. 200 Million words in total
- we use a sample of ca. 5.8 Million sentences from the years 1990 to 1998
The sentence list

- The sentence list contains ...
  - lemma form of words,
  - subclause boundaries

- NPI licensers are substituted by markers (e.g. DEINT)

- Example:

  **CLstart1** können etwas wirklich gut sein allein aus das Grund,
  **CLstart2** weil es immer so sein **CLende2 DEINT CLende1**

  „Kann etwas wirklich gut sein allein aus dem Grund, weil es immer so ist?“
  ‘Can something really be good just because it has always been like that?’
The List of Lemmas

Counts for each lemma:
- overall frequency
- occurrence with a marker (in same subclause)

<table>
<thead>
<tr>
<th>...</th>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studium</td>
<td>2781</td>
<td>363</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

We disregard lemmas with an overall frequency \( \leq 30 \)!
Number of lemmas: 1 058 462 → 64 867
Quantitative Evaluation

- We calculate for each lemma its **context ratio (CR):**

\[
CR = \frac{\text{frequency of occurrence with a marker}}{\text{overall frequency}}
\]

<table>
<thead>
<tr>
<th>...</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studium</td>
<td>2781</td>
<td>363</td>
<td>.13</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>64952</td>
<td>0.0</td>
<td>1.0</td>
<td>0.1269</td>
<td>0.10377</td>
</tr>
</tbody>
</table>
**Quantitative Evaluation**

- We calculate for each lemma its **context ratio (CR)**:
  \[ CR = \frac{\text{frequency of occurrence with a marker}}{\text{overall frequency}} \]

- Then we sort the lemmas by their CR value:

<table>
<thead>
<tr>
<th></th>
<th>unversucht</th>
<th>89</th>
<th>89</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
CR not very different from Mutual Information

Alternative possibilities of ranking: $\chi^2$, Log-likelihood

Problems: bias towards high frequencies, have to be scaled by overall frequency

Standard score: $z = \frac{x - \text{mean}}{\text{standard deviation}}$

The z-score reveals how many units of the standard deviation a case is above or below the mean.

Very good for validation, shows us the “unusual cases”
Finally: The candidate list

<table>
<thead>
<tr>
<th>#</th>
<th>lemma</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>unversucht</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>unterschätzender</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>umhin</td>
<td>0.99</td>
</tr>
<tr>
<td>4</td>
<td>nachstehen</td>
<td>0.98</td>
</tr>
<tr>
<td>5</td>
<td>lumpen</td>
<td>0.98</td>
</tr>
<tr>
<td>6</td>
<td>verhehlen</td>
<td>0.97</td>
</tr>
<tr>
<td>7</td>
<td>geheuer</td>
<td>0.96</td>
</tr>
<tr>
<td>8</td>
<td>beirren</td>
<td>0.96</td>
</tr>
<tr>
<td>9</td>
<td>Genaueres</td>
<td>0.95</td>
</tr>
<tr>
<td>10</td>
<td>wegdenken</td>
<td>0.95</td>
</tr>
<tr>
<td>11</td>
<td>unähnlich</td>
<td>0.94</td>
</tr>
<tr>
<td>12</td>
<td>allzuviel</td>
<td>0.92</td>
</tr>
<tr>
<td>13</td>
<td>sonderlich</td>
<td>0.90</td>
</tr>
<tr>
<td>14</td>
<td>hinwegtäuschen</td>
<td>0.89</td>
</tr>
<tr>
<td>15</td>
<td>dagewesen</td>
<td>0.89</td>
</tr>
<tr>
<td>16</td>
<td>abneigen</td>
<td>0.89</td>
</tr>
<tr>
<td>17</td>
<td>behagen</td>
<td>0.85</td>
</tr>
<tr>
<td>18</td>
<td>verdenken</td>
<td>0.85</td>
</tr>
<tr>
<td>19</td>
<td>missen</td>
<td>0.84</td>
</tr>
<tr>
<td>20</td>
<td>fruchten</td>
<td>0.83</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Finally: The candidate list

The candidate list looks very promising, however ...

- The candidates are single-worded, whereas many NPIs are multi-worded.
- Parts of multi-word NPIs are lower ranked and have to be ‘disambiguated’
  (nicht) alle Tassen im Schrank haben (‘to have lost one’s marbles’) ➔ Tasse at position 6934
Outline

1. Introduction
2. A little background information
3. Extraction of NPI candidates
4. **Modification to capture multi-word NPIs**
5. Refining even more
6. Towards a collection of NPIs
Modification to capture multi-word NPIs

lemma list → list of lemma chains

for each lemma chain

evaluation test

for 1 lemma chain and each of its collocates

new lemma list

collocation test

if negative – write lemma chain without collocate

lemma list

list of collocates
Modification to capture multi-word NPIs

candidate list with multi-word NPIs

reichen hinten vorne
Zweifel lassen daran er daß
schlecht staunen Blitz solange löschen
aufgehen Rezept obwohl
bekannt zunächst über
geheuer ganz
ausstehen können
zuletzt verdanken
umhin VVIZU kommen zu
ganz geheuer
unterschätzender zu
Sorge Sie brauchen machen
Stellungnahme Redaktionsschluß bis
noch dementieren wollen bestätigen
Sache jedermanns
unversucht lassen
entbehren gewiß
ejedermanns Sache
allzu lang Zeit vor
recht glauben so
übrig ander bleiben als VVIZU

darüber hinweg täuschen der können daß
angehen es können daß
ändern daran auch
gar wissen ich
kommen umhin VVIZU
so gar schlimm
Wahl ander bleiben
einwenden VVIZU gegen
verkneifen können Sie
genug gehen weit
verstehen Aufregung
vermögen sagen VVIZU zu
erkennen woran Frau
müde betonen
erwähnen Wort mit
halt vor machen auch
verhehlen
...

...without batting an eyelid!
Modification to capture multi-word NPIs

candidate list with multi-word NPIs

- 'Disambiguating' of polysemous candidates:
  Tasse $\rightarrow$ Tasse Schrank (0.8, #432)
- 112 elements in the list ($\leq$ #1000) are also listed in Kürschners NPI collection.
- Downers (ca. 80):
  notwendigerweise geben die auf Meinung wieder Seite erscheinend (1.00)
  $\rightarrow$ „Die auf dieser Seite erscheinenden Leserbriefe geben nicht notwendigerweise die Meinung der taz wieder."
  ‘The readers’ letters on this page do not necessarily reflect the opinion of the editor.’
Outline

1. Introduction
2. A little background information
3. Extraction of NPI candidates
4. Modification to capture multi-word NPIs
5. Refining even more
6. Towards a collection of NPIs
More frequencies

- Frequencies counted in a more fine-grained manner

<table>
<thead>
<tr>
<th>Negative contexts</th>
<th>classical</th>
<th>regular</th>
<th>minimal</th>
<th>questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>sonderlich (443)</td>
<td>357</td>
<td>39</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Ahnung haben (548)</td>
<td>49</td>
<td>387</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>jemals (939)</td>
<td>150</td>
<td>101</td>
<td>53</td>
<td>213</td>
</tr>
<tr>
<td>Tasse Schrank (22)</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>brauchen fürchten zu (64)</td>
<td>44</td>
<td>12</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- Frequencies helpful in verifying classifications
- Statistical methods may be employed on these counts
Outline

1. Introduction
2. A little background information
3. Extraction of NPI candidates
4. Modification to capture multi-word NPIs
5. Refining even more
6. Towards a collection of NPIs
Extraction of NPIs – Open Questions

- Candidate Lists – where is the cut-off?
- Either: much noise (items often used in negative contexts due to text type) → no “balanced” occurrence
- Or: few data (polysemous NPIs such as gar) → “standard examples” often not among the “Top 10”
- Not all licensers can be annotated → “unlicensed” NPIs are found
- Intuition of the linguist has to play a role!
Validating NPIs

Different approach:
Validation of already available collections of NPIs
→ no new data!
Database may even shrink...
Towards a collection of NPIs

The happy medium:
Validation of available NPI collections AND
Extraction of new NPIs

Our Goal:
CoDII-NPI.de — Collection of Distributionally Idiosyncratic Items
German Negative Polarity Items

Status: corpus evidence for 75 out of 200 items.

CoDII-NPI.ro finished (G. Iordachioaia)

http://www.sfb441.uni-tuebingen.de/a5/codii/index.xhtml
Acknowledgements

Thanks to **Timm Lichte** who did most of the practical work on extracting NPIs.


### Some more lemma chains...

<table>
<thead>
<tr>
<th>#</th>
<th>lemma chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>that be the <strong>end</strong></td>
</tr>
<tr>
<td>156</td>
<td><strong>thank</strong> audience</td>
</tr>
<tr>
<td>167</td>
<td>hand over <strong>Beata</strong></td>
</tr>
</tbody>
</table>

...without batting an eyelid!
Spotting, Collecting and Documenting NPIs
Part II: Collecting and Documenting NPIs

Frank Richter
Jan-Philipp Soehn
Beata Trawiński

SFB 441
Universität Tübingen

Workshop on Negation and Polarity
University of Tübingen, Germany
March 08–10, 2007
Overview

- The Collection of Distributionally Idiosyncratic Items = CoDII (Idea and Design)

- Collections:
  - The Collection of German BWs
  - The Collection of English BWs
  - The Collection of Romanian NPIs
  - The Collection of German NPIs

- Modeling and Visualizing CoDII-Entries

- Outlook
Towards a collection of NPIs

CoDII: The Essential Idea

- The essential idea of CoDII is to provide an empirical basis for linguistic investigations of lexical items showing distributional idiosyncrasies.
- This includes
  - listing appropriate items,
  - providing existing linguistic documentation,
  - specifying possibilities for extraction of data related to these items.
CoDII: Conceptual Design and Data Structure

The conceptual design and the data structure of CoDII have been conceived in such a way that

- subcollections of various types of distributionally idiosyncratic items can be modeled (anaphora, negative and positive polarity items, bound words, etc.),

- collections of distributionally idiosyncratic items from various languages can be compiled.
CoDII: Currently Available Collections

Currently, four collections of distributionally idiosyncratic items are available in CoDII:

- Bound Words of German (CoDII-BW.de),
- Bound Words of English (CoDII-BW.en),
- NPIs of Romanian (CoDII-NPI.ro),
- NPIs of German (CoDII-NPI.de).

All collections can be entered at
http://www.sfb441.uni-tuebingen.de/a5/codii
CoDII-BW.de and CoDII-BW.en are being developed

- in Project A5 *Distributional Idiosyncrasies* of the Collaborative Research Center 441 *Linguistic Data Structures: On the Relation between Data and Theory in Linguistics* at the University of Tübingen and

- at the Department of English Studies at the University of Göttingen (Manfred Sailer).
CoDII-BW.de and CoDII-BW.en contain lexical items with extremely restricted distribution ("bound words = cranberry words = unique lexemes"), e.g.:

- to make headway, by dint of, to play footsie with somebody,

Towards a collection of NPIs

CoDII-BW.de and CoDII-BW.en: Current Status

- Currently, CoDII-BW.de includes about 450 items and CoDII-BW.en includes about 100 items.
- Both collections have been presented at LREC’06: Sailer and Trawiński (2006).
- CoDII-BW.de and CoDII-BW.en are still work in progress.
- In consideration of the focus of the present phase of our project (NPIs), our current concern is the collection and documentation of NPIs.
Towards a collection of NPIs

NPIs in CoDII

At present, there are two collections of NPIs available in CoDII:

- CoDII-NPI.ro and
- CoDII-NPI.de.
CoDII-NPI.ro currently includes about 60 items.

The items of CoDII-NPI.ro correspond to the English, German and Dutch NPIs given in the linguistic literature (since there is no specific collection of Romanian NPIs available in the literature).

The following people have contributed to the design and the compilation of CoDII-NPI.ro: Gianina Iordăchioaia, Jan-Philipp Soehn and Beata Trawiński.
CoDII-NPI.de currently includes above 30 items.

At present, adverbial, prepositional and nominal NPIs are collected and documented.

The sources used for acquiring the NPIs for CoDII-NPI.de were:
- the collections of NPIs in Welte (1978) and Kürschner (1983) and
- the list of the NPI candidates automatically extracted from the Tübingen Partially Parsed Corpus of Written German (TüPP) based on the German newspaper die tageszeitung (taz).

The following people have contributed to the design and the compilation of CoDII-NPI.de: Johannes Fiegenbaum, Jan-Philipp Soehn, Beata Trawiński and Bela Usabaev.
Modeling NPIs in CoDII

Each NPI is characterized in CoDII by four information blocks:

- General Information,
- Syntactic Information,
- Licensing Contexts,
- Class.
Towards a collection of NPIs

General Information

The General Information block identifies NPIs by providing

- a particular NPI,
- the English gloss of the (German / Romanian) NPI,
- the English translation of the (German / Romanian) NPI,
- the possible expressions in which the NPI occurs,
- the set of possible paraphrases of these expressions.
Towards a collection of NPIs

Syntactic Information

The block Syntactic Information provides information on

- the syntactic category of an NPI,
- the syntactic structure of the expression in which the NPI occurs,
- possible syntactic variations such as passivization, pronominalization, modification, occurrence in raising constructions, etc.

For syntactic description of NPIs and expressions in which they occur,

- the *Stuttgart-Tübingen Tagset* (STTS) (for German NPIs), and
- the tagset from the Multilingual Text Tools and Corpora for Central and Eastern European Languages (MULTEXT-East) (for Romanian NPIs)

have been used.
Licensing Contexts

The block Licensing Contexts provides information on the environment licensing a given NPI. The following LCs are specified in CoDII:

- Clausemate Negation (CMN),
- Non-Clausemate Negation (nCMN),
- N-Word (NW),
- *kein* ‘kein-negation’,
- *without*,
- Restrictor of Universal Quantifier (UNIV),
- Downward-Entailing (DENT),
- *only*,
- Negative Verb (NV),
- Question (QUE),
- Conditional (IF),
- Comparative (COMP),
- Superlative (SUP),
- Imperative (IMP),
- Non-Affirmative Verb (NAV).

Also, exceptional cases can be specified.
Examples

For each licensing context, appropriate examples are provided from the following sources:

- **for the Romanian NPIs**
  - Romanian electronic corpus developed by Rada Mihalcea,
  - Romanian electronic corpus developed at the Romanian Academy Center for Artificial Intelligence (RACAI),
  - Internet via Google,
  - personal examples of Gianina Iordăchioaia.

- **for the German NPIs**
  - corpora of the Institute of German Language in Mannheim ([http://www.ids-mannheim.de/cosmas2/](http://www.ids-mannheim.de/cosmas2/)),
  - Internet via Google.
Class

The block Class specifies the type of polarity (negative) and the class associated with a given NPI according to the following sources:

- Welte (1978),
- Kürschner (1983),
- A5 (project-internal classification).

The following classes of NPIs are used in CoDII:

- OPEN (assigned to NPIs listed in Welte (1978) and Kürschner (1983)),
- superstrong, strong and weak (assigned to NPIs classified project-internally according to the contexts in which these NPIs are licensed; these classes have been adopted from Zwarts (1997) and are defined as in van der Wouden (1997)).
Technical Realization

CoDII-NPI.ro and CoDII-NPI.de have been internally encoded in XML. The DTD has been specified in such a way that:

- The element `codii` is the document root and its instance is identified by attributes `type` (for specifying collection type) and `xml:lang` (for specifying language the data come from).
- The content model of the element `codii` consists of two elements: `dii-list`, whose content is a list of NPIs, and `dii-examples`, whose content is a list of examples.
Towards a collection of NPIs

Encoding NPIs

The content model of the element `dii-list` consists of a list of the `dii-entry` elements, whose content model consists of a set of elements which

- identify NPIs (`dii`),
- describe documentation on each NPI (`dii-classification`),
- present syntactic properties of NPIs (`dii-syntax`),
- specify licensing contexts of NPIs (`licensers`).
A Fragment of the CoDII-XML-Encoding of the NPI *beileibe* ‘by no means’

```xml
<dii-entry id="beileibe">
  <dii>
    <ol>beileibe</ol>
    <en>by no means, at all</en>
  </dii>
  <dii-classification>
    <dii-class category="pi" subcategory="npi" type="A5" class="weak" original-class="no">
      <bibliography bib-item="A5" />
    </dii-class>
    <dii-class category="pi" subcategory="npi" type="Kuerschner83" class="OPEN" original-class="no">
      <bibliography bib-item="Kuerschner:83" />
    </dii-class>
    <dii-class category="pi" subcategory="npi" type="Welte78" class="OPEN" original-class="no">
      <bibliography bib-item="Welte:78" />
    </dii-class>
  </dii-classification>
  <dii-syntax hits="beileibe-01 beileibe-02 beileibe-03 beileibe-04 beileibe-05 beileibe-06 beileibe-07" cat="ADV">
    <dii-expression-syntax>ADV</dii-expression-syntax>
  </dii-syntax>
  <licensors>
    <cmn given="yes" hits="beileibe-01" />
    <ncmn given="no" />
    <kein given="yes" hits="beileibe-02" />
    <nw given="yes" hits="beileibe-03" />
    <dent given="yes" hits="beileibe-04" />
    <nv given="no" />
    <que given="no" />
    <imp given="no" />
    <if given="no" />
    <without given="no" />
    <only given="yes" hits="beileibe-05" />
    <univ given="no" />
    <comp given="yes" hits="beileibe-06" />
    <sup given="yes" hits="beileibe-07" />
    <exc given="no" />
  </licensors>
</dii-entry>
```
The content model of the elements `dii-examples` consists of a list of the example elements.

The example elements are linked to the appropriate NPIs by dint of the attributes `dii` and `id`. 
Towards a collection of NPIs

The CoDII-XML-description of examples for *beileibe* ‘by no means’

```xml
<example dii="beileibe" id="beileibe-01">
  <source corpus="cosmasII">Mannheimer Morgen, 15.04.1991, Politik</source>
  <ol>An Bem"hungen, als Nachfolger des ermordeten Rohwedder wieder einen Sozialdemokraten an die Spitze der Treuhandanstalt zu stellen, hat es ja beileibe nicht gefehlt.
  </ol>
</example>

<example dii="beileibe" id="beileibe-02">
  <source corpus="cosmasII">Mannheimer Morgen, 25.01.1991, Lokales</source>
  <ol>Breites soziales Engagement. Er ist einer der ganz prominenten Sozialdemokraten der Stadt, leistet aber seine Arbeit im stillen und ist beileibe kein Freund großer Auftritte.
  </ol>
</example>

<example dii="beileibe" id="beileibe-03">
  <source corpus="google">http://www.f5.parsimony.net/forum5557/messages/28281.htm (9.5.06)
  </source>
  <ol>Ich bin beileibe niemand, der die Frau und Mutter ausschließlich am Herd und bei den Kindern sehen will, aber das Gegenteil muß ihnen auch institutionell ermöglicht werden.
  </ol>
</example>
```
Visualization

- CoDII-NPI.ro and CoDII-NPI.de are available via the Internet in the form of a set of XHTML files generated by an XSLT script.

- The URL:

  http://www.sfb441.uni-tuebingen.de/a5/codii
Towards a collection of NPIs

Browser Display for the NPI *beileibe* ‘by no means’
The General Characteristic of CoDII

- CoDII is a research platform.
- The architecture of the data structure of CoDII is linguistically motivated.
- CoDII does not provide its own corpus but refers to existing corpora and gives sample queries.
- The target audience of CoDII are primarily linguists.
- Information gathered in CoDII allow for a systematic study of distributionally idiosyncratic items.
- CoDII is becoming a multilingual resource.
Outlook

- The data structure design of CoDII makes it possible to add further classifications, corpora and search tools, as well as further collections of distributionally idiosyncratic items.
- An extension to more languages is equally possible.
- It is also planned to extend the collection to other types of distributionally idiosyncratic items. In particular, we intend to include a documentation of the use of positive polarity items of German, i.e., items which require a positive context (in cooperation with the project *Positive Polarity Items*, associated with SFB 441).
- On the technical side, CoDII will be converted into a database to allow for a dynamic and more flexible access to the data.
Bibliography

_Studien zur Negation im Deutschen._
Tübingen: Gunter Narr.


_Negative Contexts. Collocation, polarity and multiple negation._

_Negationslinguistik. Ansätze zur Beschreibung und Erklärung von Aspekten der Negation im Englischen._
München: Wilhelm Fink Verlag.

Three Types of Polarity.