Creating RRG treebanks through semi-automatic conversion of annotated corpora

Tatiana Bladier, Kilian Evang, Laura Kallmeyer, Robin Möllemann & Rainer Osswald

(RRGbank Team former member: Andreas van Cranenburgh)

RRG Conference Buffalo
19-21 August 2019
University of Buffalo, NY
Overview

Motivation behind RRGbank

Creating RRGbank
- Design of the RRG structures
- Penn Treebank to RRG conversion
- Universal Dependencies to RRG conversion

Evaluation

Conclusions and Future Work
Outline

Motivation behind RRGbank

Creating RRGbank
- Design of the RRG structures
- Penn Treebank to RRG conversion
- Universal Dependencies to RRG conversion

Evaluation

Conclusions and Future Work
RRGb ank

★ a large corpus of RRG annotated sentences;
★ starting point: English → 50 000 sentences from the Wall Street Journal;
★ future work: several languages → transformation from Universal Dependencies corpora → over 80 languages.
Why RRGbank?

★ corpus-based investigations for linguistic modeling with RRG,
★ test corpus for formalization of RRG
  → using tree grammars: Kallmeyer et al. (2013); Kallmeyer (2016); Kallmeyer and Osswald (2017),
★ test corpus for precision RRG grammars,
★ training data for supervised data-driven RRG parsing,
★ new insights into RRG for different languages.
Outline

Motivation behind RRGbank

Creating RRGbank
  Design of the RRG structures
  Penn Treebank to RRG conversion
  Universal Dependencies to RRG conversion

Evaluation

Conclusions and Future Work
RRG: Textbook Notation

Japanese were said to be heavy buyers.
Our Single-Tree Notation

SENTENCE
  CLAUSE
    CORE
      NP
      NUC
      CORE
        NUC
        NP
        CORE
          NUC
          NUC

Japanese were said to be heavy buyers

said to be heavy

operators/peripheries are marked by adding -OP/-PERI to label.

operators, peripheries and CLM are attached to constituent projection using normal edges,

crossing branches possible,
Our Single-Tree Notation

- operators, peripheries and CLM are attached to constituent projection using normal edges,
Our Single-Tree Notation

- operators, peripheries and CLM are attached to constituent projection using normal edges,
- → crossing branches possible,
Our Single-Tree Notation

- operators, peripheries and CLM are attached to constituent projection using normal edges,
- crossing branches possible,
- operators/peripheries are marked by adding -OP/-PERI to label.
Creating a Validation Treebank

- manually check and validate data,
- automatic conversion script.

RRG annotation tool: rrgbank.phil.hhu.de
Validation Treebank

RRGbank annotation tool: rrgbank.phil.hhu.de
Automatic conversion from the Penn Treebank

Tree in the PTB and the converted RRG tree
Example Conversion Rule 1/3: Adverb
Example Conversion Rule 2/3: Sentence
Example Conversion Rule 3/3: Topicalization

\[ \text{S-TPC NP-SBJ V} \]
\[ a \quad b \quad c \]
\[ \rightarrow \]
\[ \text{SENTENCE} \]
\[ \text{LDP} \]
\[ \text{CLAUSE} \]
\[ \text{CORE} \]
\[ \text{SENTENCE} \]
\[ \text{NP} \quad \text{NUC} \]
\[ a \quad b \quad c \]
Statistics on RRGbank

- 4 annotators,
- 395 gold annotated sentences → validated and adjudicated by at least two annotators,
- 1090 silver annotated sentences → validated by one annotator,
- 8500 sentences ≤ 25 tokens,
- accuracy: 95.98 (PTB2RRG) and 87.03 (UD2RRG).
Universal Dependencies to RRG: Automatic Conversion

<table>
<thead>
<tr>
<th></th>
<th>ptb2rrg</th>
<th>ud2rrg</th>
</tr>
</thead>
<tbody>
<tr>
<td>input trees</td>
<td>PTB</td>
<td>UD (converted from PTB with Stanford CoreNLP)</td>
</tr>
<tr>
<td>languages</td>
<td>1</td>
<td>83+</td>
</tr>
<tr>
<td>algorithm</td>
<td>rewrite rules</td>
<td>complete traversal</td>
</tr>
<tr>
<td>treebank-specific information</td>
<td>yes (PTB)</td>
<td>via extensions</td>
</tr>
<tr>
<td>accuracy (evalb F1)</td>
<td>95.98</td>
<td>87.03</td>
</tr>
<tr>
<td>coverage (short sent.)</td>
<td>100%</td>
<td>94.9%</td>
</tr>
<tr>
<td>converted gold sentences</td>
<td>395 (all)</td>
<td>375 (of 395)</td>
</tr>
</tbody>
</table>
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG

```
VERB  Avoiding
NOUN  failure
AUX   is
ADJ   easy
PUNCT .
```

```
ROOT
CLAUSE
CORE

CORE
NUC
V
Avoiding

NUC
AP

CORE_A
NUC_A

AUX
is
easy
```
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG
Example PTB-UD to RRG

- VERB: Avoiding
- NOUN: failure
- AUX: is
- ADJ: easy
- PUNCT: .
Outline

Motivation behind RRGbank

Creating RRGbank
- Design of the RRG structures
- Penn Treebank to RRG conversion
- Universal Dependencies to RRG conversion

Evaluation

Conclusions and Future Work
RRGbank: evaluation

Sentence: The new rate will be payable Feb. 15
Gold tree:

```
ROOT
  |---CLAUSE
  |   |---CORE
  |     |---NP
  |     |   |---CORE_N
  |     |   |   |---AP-PERI
  |     |   |   |   |---CORE_A
  |     |   |   |   |---NUC_A
  |     |   |---NUC_N
  |---DEF-OP
  |     |---A
  |     |---N
  |---TNS-OP
  |     |---A
  |     |---N
  |---QNT-OP
  |     |---15
```

Candidate tree:

```
ROOT
  |---CLAUSE
  |   |---CORE
  |     |---NP
  |     |   |---CORE_N
  |     |   |---AP-PERI
  |     |   |   |---CORE_A
  |     |   |---NUC_A
  |     |   |---NUC_N
  |---DEF-OP
  |     |---A
  |     |---N
  |---TNS-OP
  |     |---A
  |     |---N
  |---QNT-OP
  |     |---15
```

- Gold manually validated sentences \(= 395\);
- EVALB bracketing scores:
  - matching spans,
  - matching brackets,
  - matching labels.
- *zero common bracketings between the PTB and converted RRG trees \(\rightarrow\) full coverage.
**RRGbank: evaluation metrics PTB2RRG and UD2RRG**

<table>
<thead>
<tr>
<th>parameter</th>
<th>ptb2rrg</th>
<th>ud2rrg</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of converted sentences:</td>
<td>395</td>
<td>375</td>
</tr>
<tr>
<td>longest sentence:</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>gold brackets (discontinuous):</td>
<td>5096</td>
<td>5096</td>
</tr>
<tr>
<td>candidate brackets (discontinuous):</td>
<td>5076</td>
<td>4807</td>
</tr>
<tr>
<td>labeled recall:</td>
<td>95.88</td>
<td>84.54</td>
</tr>
<tr>
<td>labeled precision:</td>
<td>96.09</td>
<td>89.66</td>
</tr>
<tr>
<td><strong>labeled f-measure:</strong></td>
<td><strong>95.98</strong></td>
<td><strong>87.03</strong></td>
</tr>
<tr>
<td>exactly matched sentences:</td>
<td>66.84</td>
<td>44.30</td>
</tr>
<tr>
<td>function tags:</td>
<td>95.16</td>
<td>100.00</td>
</tr>
<tr>
<td>POS accuracy:</td>
<td>98.54</td>
<td>96.75</td>
</tr>
</tbody>
</table>
Encountered issues and problematic cases

Three types of problematic cases we encountered during conversion:

- Inconsistencies or errors in Penn Treebank → lead to inconsistencies in PTB-UD.
- Distinctions made in RRG but not in PTB → PTB-UD is even more affected.
- Analyses in PTB and PTB-UD which have no direct equivalent in RRG.
Example 1: PTB annotation inconsistencies

Erroneous annotation in PTB
Example 1: Annotation inconsistencies
Example 1: Annotation inconsistencies

- Lexical elements misanalyzed in PTB are manually corrected in the PTB input
Example 1: Annotation inconsistencies

- Lexical elements misanalyzed in PTB are manually corrected in the PTB input
- However, not all cases are as clear as "heavy" in Example 1
Example 1: Annotation inconsistencies

- Lexical elements misanalyzed in PTB are manually corrected in the PTB input.
- However, not all cases are as clear as "heavy" in Example 1.
- Some NPs in PTB are not headed by a noun, which could either be an annotation error or a possible conversion.
Example 2: Scope of negation

Negation in PTB and RRG
Example 2: Scope of negation

- PTB doesn't differentiate between internal and external negation.
- As internal negation is more common, negation is treated as a core-operator.
Example 2: Scope of negation

- PTB doesn’t differentiate between internal and external negation
Example 2: Scope of negation

- PTB doesn’t differentiate between internal and external negation
- As internal negation is more common, negation is treated as a core-operator
Example 3.1: Different junctures

We keep wondering what Mr. Gates wanted to say.

PTB structure
Example 3.2: Different junctures

We keep wondering what Mr. Gates wanted to say.

PTB-to-RRG structure
Example 3: Different junctures

• PTB doesn't differentiate between cosubordination and subordination, barring conversion on a syntactic basis alone.
• Using a lexical approach however, enables consistent conversion of some cosubordinations.
• When cooccurring with gerunds, certain verbs (like start, keep, or finish) indicate a Phase relation and therefore nuclear cosubordination Van Valin Jr (2005).
• If necessary, the traces contained in PTB can be used to further restrict the conversion context to avoid false positives.
Example 3: Different junctures

- PTB doesn’t differentiate between cosubordination and subordination, barring conversion on a syntactic basis alone.
Example 3: Different junctures

- PTB doesn’t differentiate between cosubordination and subordination, barring conversion on a syntactic basis alone.
- Using a lexical approach however, enables consistent conversion of some cosubordinations
Example 3: Different junctures

- PTB doesn’t differentiate between cosubordination and subordination, barring conversion on a syntactic basis alone.
- Using a lexical approach however, enables consistent conversion of some cosubordinations.
- When cooccurring with gerunds, certain verbs (like start, keep, or finish) indicate a Phase relation and therefore nuclear cosubordination Van Valin Jr (2005)
Example 3: Different junctures

- PTB doesn’t differentiate between cosubordination and subordination, barring conversion on a syntactic basis alone.
- Using a lexical approach however, enables consistent conversion of some cosubordinations.
- When cooccurring with gerunds, certain verbs (like start, keep, or finish) indicate a Phase relation and therefore nuclear cosubordination Van Valin Jr (2005).
- If necessary, the traces contained in PTB can be used to further restrict the conversion context to avoid false positives.
Example 4: Quantifier Phrases

QPs and nounless NPs
Example 4.1: Quantifier Phrases

```
PP-LOC-CLR

NP

in a trading range of $1.19 to $1.34
```

NP headed by a QP in PTB
Example 4.2: Quantifier Phrases

NP headed by a QP retained in RRG
Examples 4.3 - 4.6 Quantifier Phrases

A variety of different QPs
Example 4: Quantifier Phrases

QPs are inconsistent with regard to both the lexical category and internal position of their heads.

QPs can function as the only constituent within an NP leaving the NP without a lexical head.

These properties of QPs have proven highly problematic for automated conversion.

Until a consistently correct conversion can be achieved, QPs will be retained.
Example 4: Quantifier Phrases

- QPs are inconsistent with regard to both the lexical category and internal position of their heads.
Example 4: Quantifier Phrases

- QPs are inconsistent with regard to both the lexical category and internal position of their heads.
- QPs can function as the only constituent within an NP leaving the NP without a lexical head.
Example 4: Quantifier Phrases

- QPs are inconsistent with regard to both the lexical category and internal position of their heads.
- QPs can function as the only constituent within an NP leaving the NP without a lexical head.
- These properties of QPs have proven highly problematic for automated conversion.
Example 4: Quantifier Phrases

- QPs are inconsistent with regard to both the lexical category and internal position of their heads.
- QPs can function as the only constituent within an NP leaving the NP without a lexical head.
- These properties of QPs have proven highly problematic for automated conversion.
- Until a consistently correct conversion can be achieved, QPs will be retained.
Example 5.1: Complex proper NPs

Complex NP with multiple proper nouns
Example 5.2: Complex proper NPs

```
NP
  | CORE_N
  |
NP-PERI
  | CORE_N
  | NP-PERI
  |
NUC_N
  | NUC_N
  | NUC_N
  |
N-PROP
  | N-PROP
  | N-PROP
  | N-PROP
  |
Arnold
  | Celnicker
  | Assistant
  | Professor
  |
Ohio
  | State
  | University
```

Complex NP with multiple proper nouns
Outline

Motivation behind RRGbank

Creating RRGbank
- Design of the RRG structures
- Penn Treebank to RRG conversion
- Universal Dependencies to RRG conversion

Evaluation

Conclusions and Future Work
Conclusions and Future Work

- Further conversion of PTB trees and validation of RRG trees.
- Further formalization of the RRG (Kallmeyer et al., 2013).
- Extraction of fragments ("supertags") of the formalized RRG grammar.
- Statistical parsing with these fragments.
Thank you!

THANK YOU VERY MUCH FOR YOUR ATTENTION!


## Category Statistics

(10 most frequent categories / errors)

<table>
<thead>
<tr>
<th>label</th>
<th>% gold</th>
<th>recall</th>
<th>prec.</th>
<th>F1</th>
<th>recall</th>
<th>prec.</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>15.89</td>
<td>97.65</td>
<td>99.00</td>
<td>98.32</td>
<td>87.27</td>
<td>95.68</td>
<td>91.28</td>
</tr>
<tr>
<td>CORE_N</td>
<td>12.99</td>
<td>96.37</td>
<td>96.67</td>
<td>96.52</td>
<td>89.88</td>
<td>95.05</td>
<td>92.39</td>
</tr>
<tr>
<td>NUC_N</td>
<td>12.76</td>
<td>98.62</td>
<td>98.31</td>
<td>98.46</td>
<td>92.62</td>
<td>96.47</td>
<td>94.51</td>
</tr>
<tr>
<td>CORE</td>
<td>8.28</td>
<td>88.39</td>
<td>87.15</td>
<td><strong>87.76</strong></td>
<td>71.84</td>
<td>72.53</td>
<td>72.18</td>
</tr>
<tr>
<td>NUC</td>
<td>8.18</td>
<td>88.49</td>
<td>89.56</td>
<td>89.02</td>
<td>82.73</td>
<td>81.75</td>
<td>82.24</td>
</tr>
<tr>
<td>CLAUSE</td>
<td>7.46</td>
<td>99.21</td>
<td>99.21</td>
<td><strong>99.21</strong></td>
<td>85.41</td>
<td>88.71</td>
<td>87.03</td>
</tr>
<tr>
<td>SENTENCE</td>
<td>6.87</td>
<td>99.71</td>
<td>97.76</td>
<td>98.73</td>
<td>94.59</td>
<td>99.40</td>
<td>96.93</td>
</tr>
<tr>
<td>AP</td>
<td>3.71</td>
<td>91.53</td>
<td>91.53</td>
<td>91.53</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>NUC_A</td>
<td>3.69</td>
<td>98.40</td>
<td>97.37</td>
<td>97.88</td>
<td>93.09</td>
<td>99.43</td>
<td>96.15</td>
</tr>
<tr>
<td>CORE_A</td>
<td>3.69</td>
<td>91.49</td>
<td>90.53</td>
<td>91.01</td>
<td>86.70</td>
<td>92.61</td>
<td>89.56</td>
</tr>
</tbody>
</table>
## Function Tag Statistics

(most frequent tags / errors)

<table>
<thead>
<tr>
<th>func. tag</th>
<th>% gold</th>
<th>recall</th>
<th>prec.</th>
<th>F1</th>
<th>recall</th>
<th>prec.</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP</td>
<td>39.12</td>
<td>100.00</td>
<td>100.00</td>
<td><strong>100.00</strong></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>PROP</td>
<td>27.16</td>
<td>99.28</td>
<td>100.00</td>
<td>99.64</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>PERI</td>
<td>26.96</td>
<td>71.64</td>
<td>96.57</td>
<td><strong>82.25</strong></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>WH</td>
<td>4.12</td>
<td>100.00</td>
<td>95.45</td>
<td>97.67</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>DEM</td>
<td>1.96</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

PTB2RRG and UD2RRG evaluation metrics.
PTB2RRG and UD2RRG: Example 1

Japanese were said to be heavy buyers.

Erroneous annotation in PTB and PTB-UD
Negation in PTB, PTB-UD, and RRG
Different junctures
PTB2RRG and UD2RRG: Example 4

We keep wondering what Mr. Gates wanted to say.
PTB2RRG and UD2RRG: Example 5

Quantifier Phrases