

Towards a Gold Standard for RRG Annotation

(and Automatic Conversion from UD to RRG)

Kilian Evang

University of Düsseldorf

2019-05-15

TreeGraSP Meeting #3

Outline

Introduction

Example Disagreements

UD2RRG

Outline

Introduction

Example Disagreements

UD2RRG

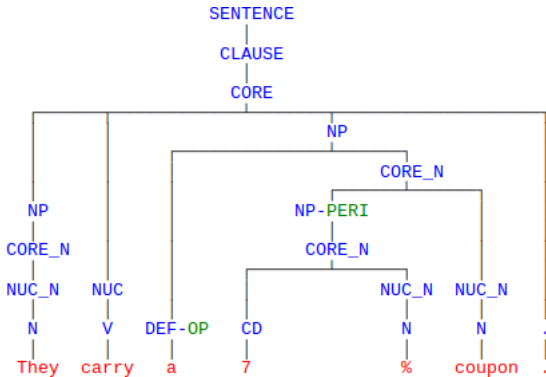
WP 1: Formalizing and Implementing RRG

- gain new insights about the mathematical properties of grammar
- provide tools for RRG implementation that will be useful to the community
- insights on grammar architecture can be transferred from RRG to TAG
- overall goal: integrated TAG/RRG grammar development framework
 - need evaluation (and training) data

RRGbank

- <https://rrgbank.phil.hhu.de/>
- ongoing effort to create a corpus annotated with RRG structures
- iterative conversion and correction
 - automatic conversion from (for now) Penn Treebank
 - correction by annotators
 - improvement of converter
- multilingual UD converter underway
- challenge: need annotation guideline! Not everything specified in RRG literature

Example Tree



Towards a First (Small) Gold Standard Dataset

- our two most active annotators currently have 285 sentences in common
- of these, 189 have 10 tokens or less (starting point)
- of these, 76 have no disagreements, 113 have some (ignoring attachment of punctuation)
- overall, 0.9 disagreements per sentence (the way I counted)
- Let's try to resolve them!

Classification of Disagreements

class	count
extra or missing nodes	8
auxiliary or operator?	8
POS tag	28
phrasal category	10
internal structure of pronoun NPs	7
flat vs. hierarchical phrase structure	1
analysis of topicalization	1
which layer?	68
periphery or not?	13
juncture-nexus type	22
what kind of nucleus?	8
total	177

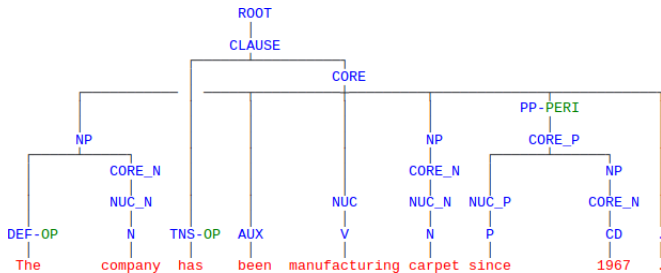
Outline

Introduction

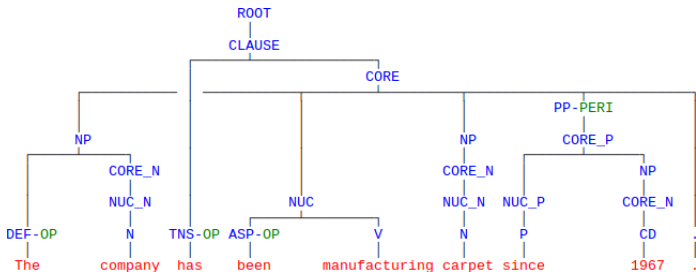
Example Disagreements

UD2RRG

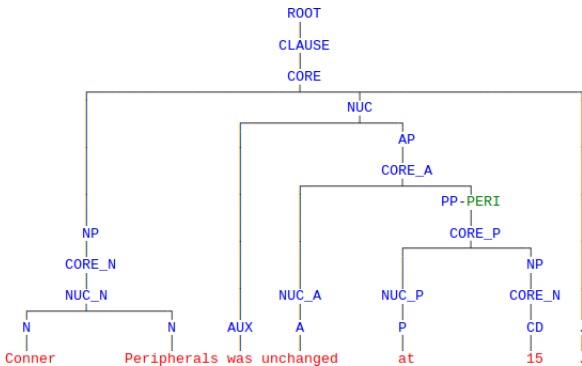
AUX or ASP-OP?



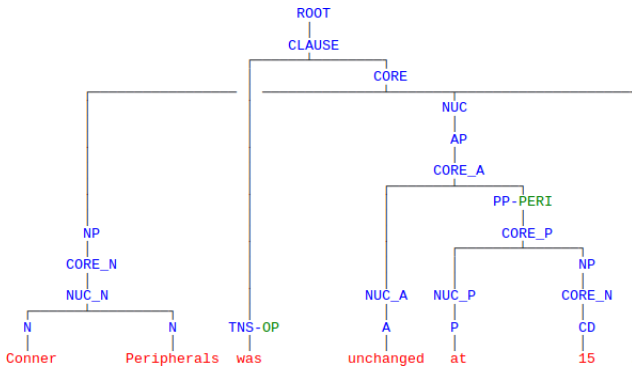
AUX or ASP-OP?



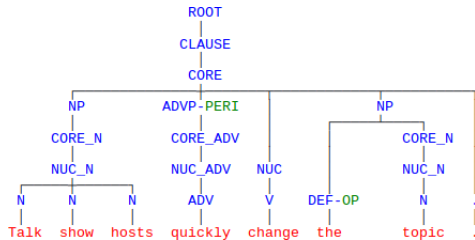
AUX or TNS-OP?



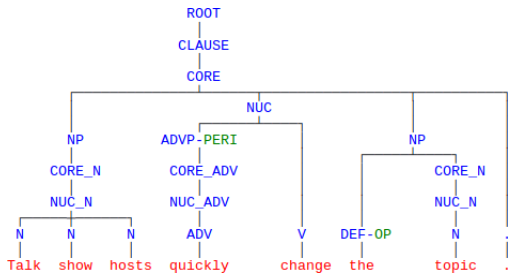
AUX or TNS-OP?



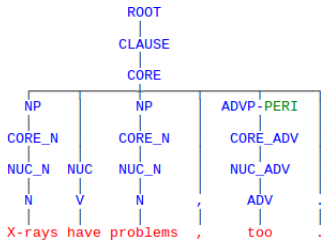
ADVP-PERI → CORE or NUC? (1)



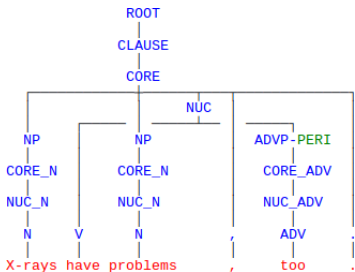
ADVP-PERI → CORE or NUC? (1)



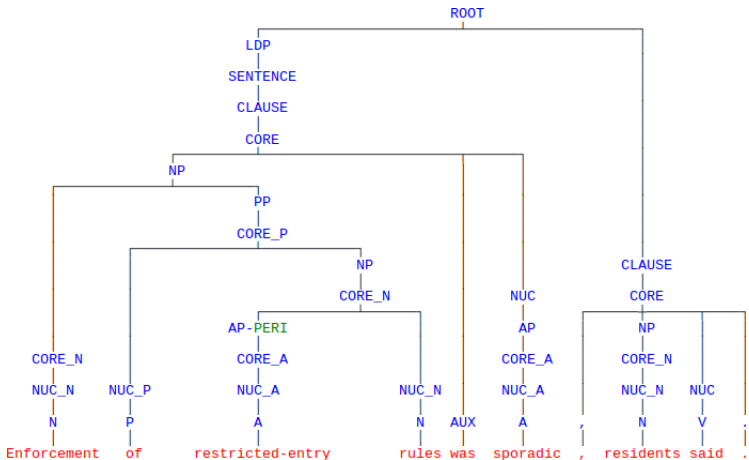
ADVP-PERI → CORE or NUC? (2)



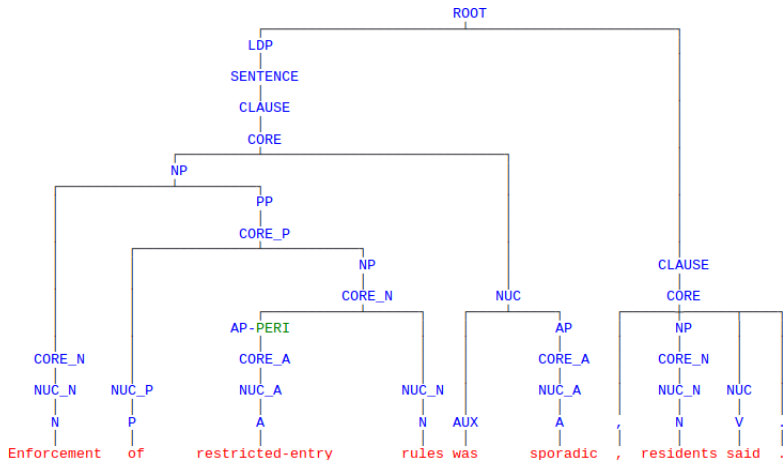
ADVP-PERI → CORE or NUC? (2)



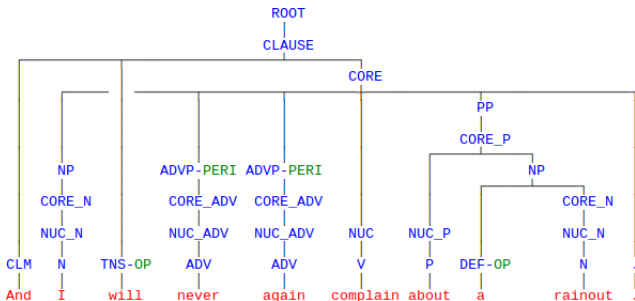
AUX → CORE or NUC?



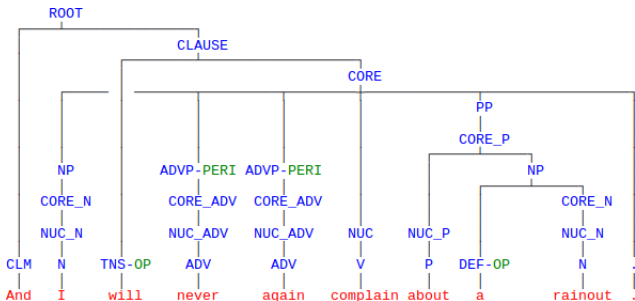
AUX → CORE or NUC?



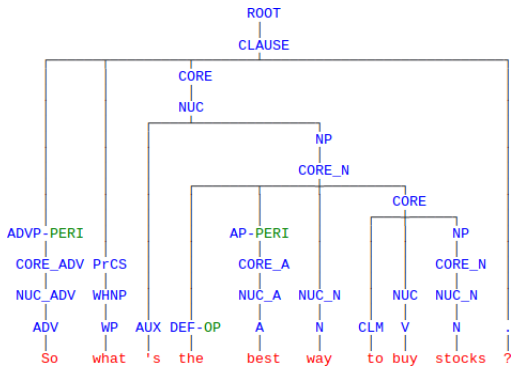
CLM → CLAUSE or SENTENCE?



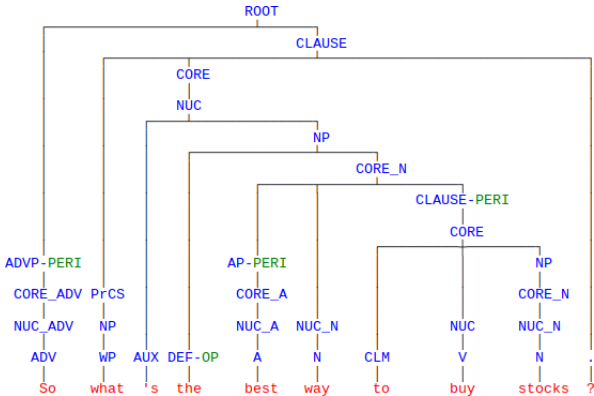
CLM → CLAUSE or SENTENCE?



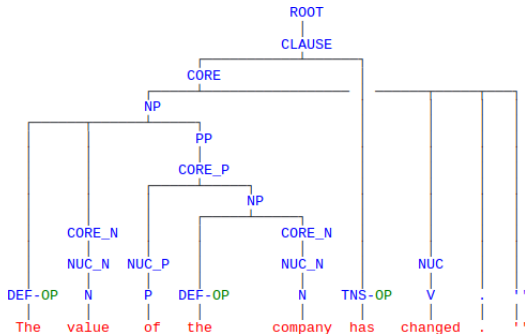
DEF-OP → CORE_N or NP?



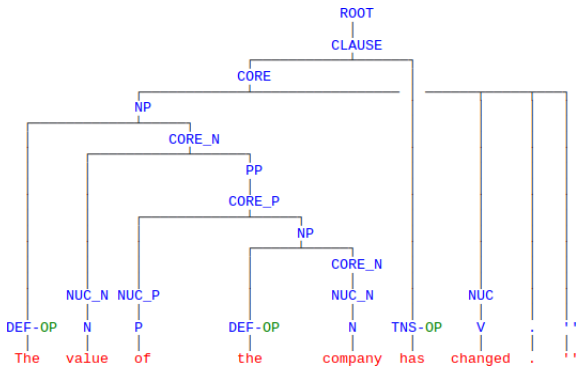
DEF-OP → CORE_N or NP?



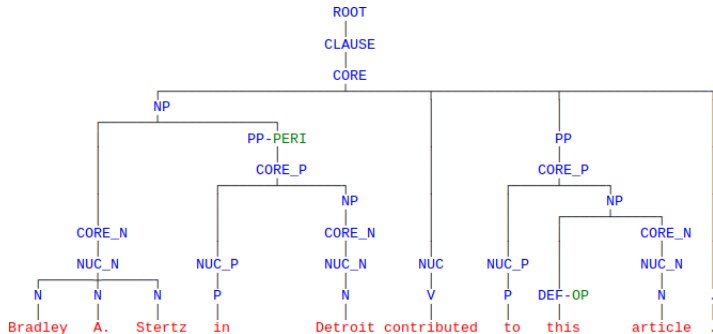
PP → NP or CORE_N?



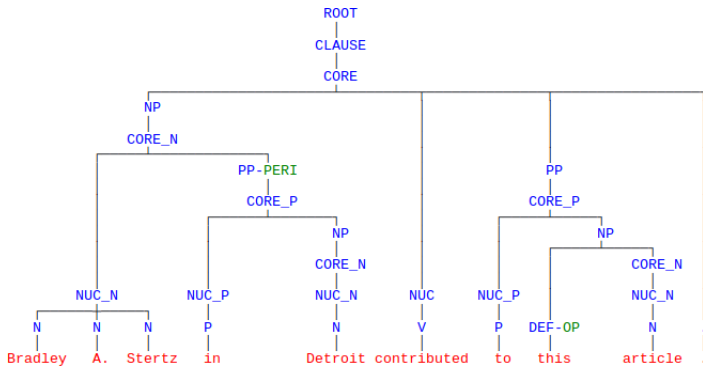
PP → NP or CORE_N?



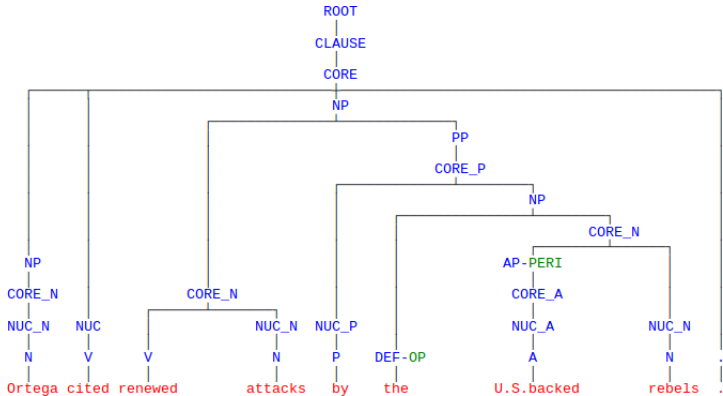
PP-PERI → NP or CORE_N?



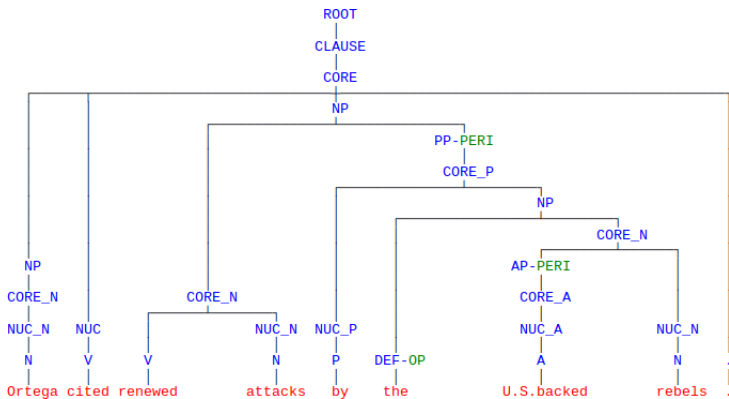
PP-PERI → NP or CORE_N?



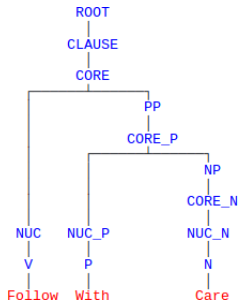
PP or PP-PERI? (1)



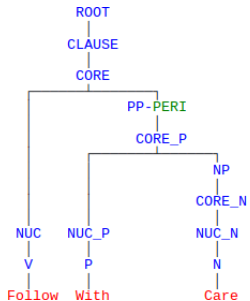
PP or PP-PERI? (1)



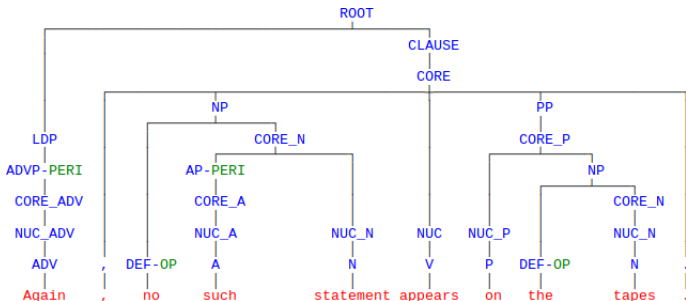
PP or PP-PERI? (2)



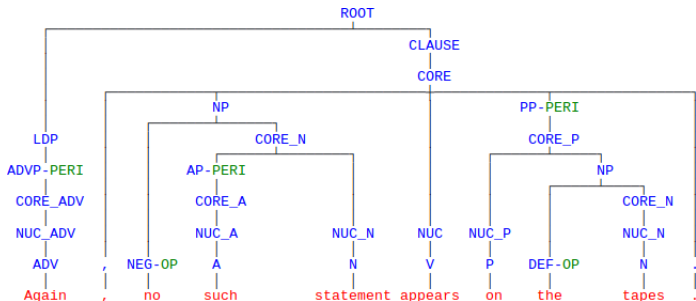
PP or PP-PERI? (2)



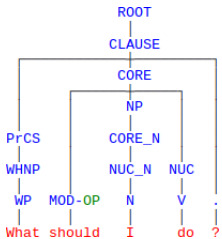
DEF-OP or NEG-OP?



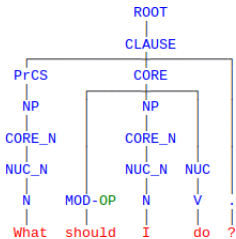
DEF-OP or NEG-OP?



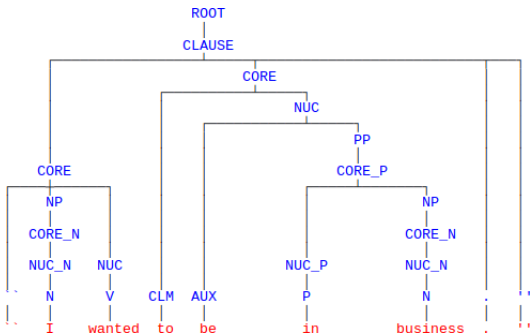
wh-pronouns



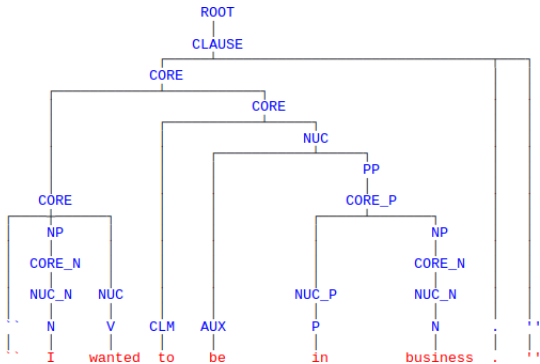
wh-pronouns



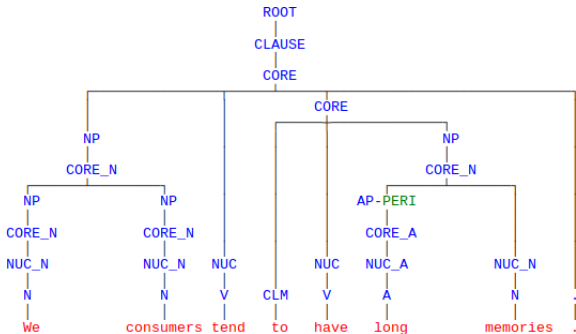
Core Coordination or Core Cosubordination?



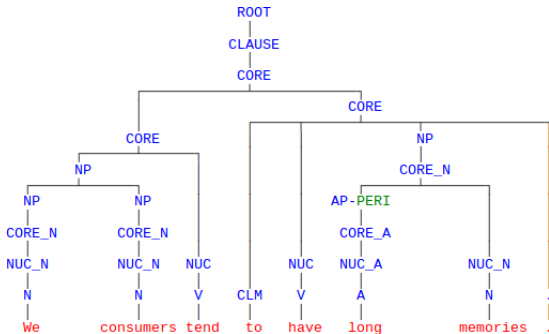
Core Coordination or Core Cosubordination?



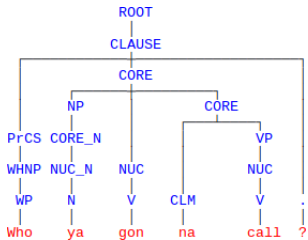
Core Subordination or Core Cosubordination?



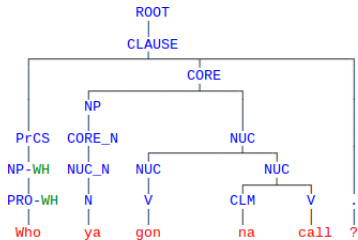
Core Subordination or Core Cosubordination?



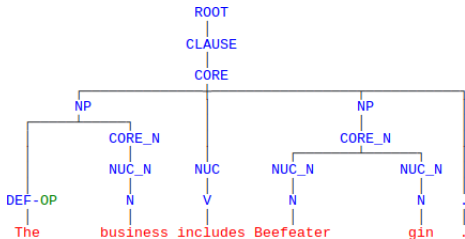
Core Subordination or Nuclear Cosubordination?



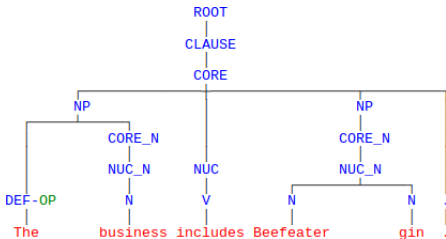
Core Subordination or Nuclear Cosubordination?



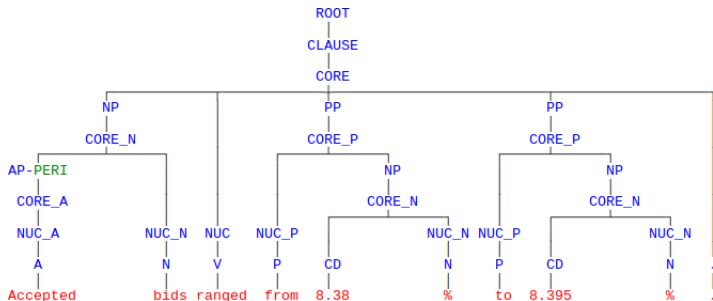
Compounds: Separate Nuclei or Single Nucleus?



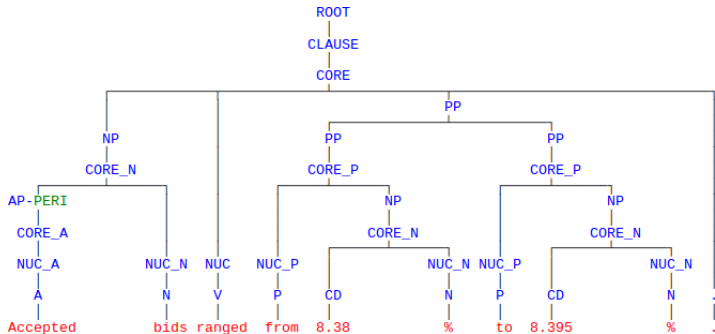
Compounds: Separate Nuclei or Single Nucleus?



PP Coordination or PP Cosubordination?



PP Coordination or PP Cosubordination?



Outline

Introduction

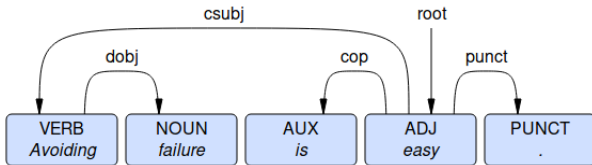
Example Disagreements

UD2RRG

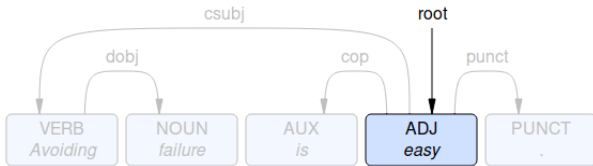
Automatic Conversion to RRG

	ptb2rrg.py	ud2rrg.py
input trees	PTB	UD
languages	1	83+
algorithm	rewrite rules	complete traversal
treebank-specific information	yes (PTB)	via extensions
coverage (short sent.)	100%	91.7%+
accuracy	?	?

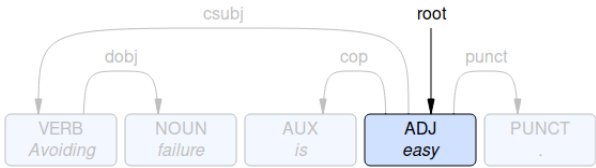
Example



Example

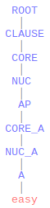
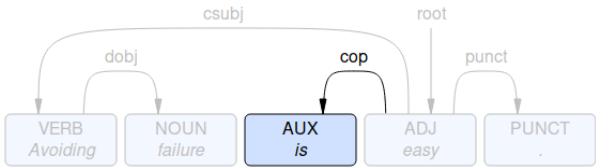


Example

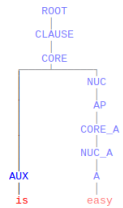
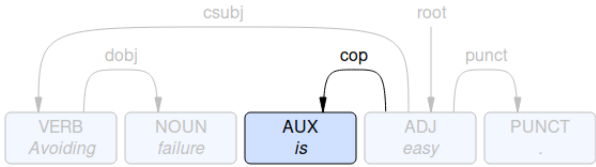


ROOT
|
CLAUSE
|
CORE
|
NUC
|
AP
|
CORE_A
|
NUC_A
|
A
|
easy

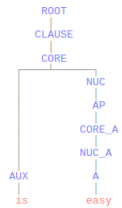
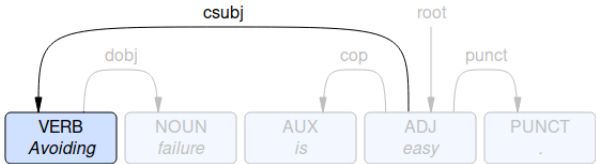
Example



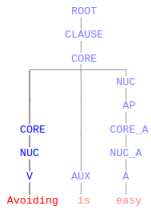
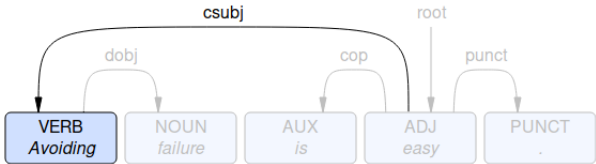
Example



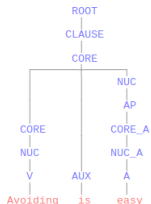
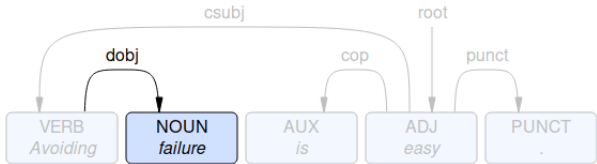
Example



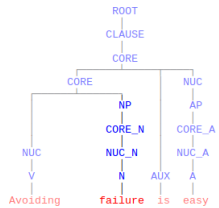
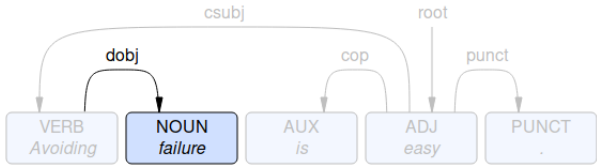
Example



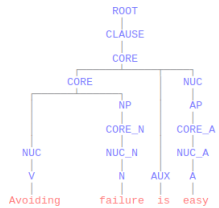
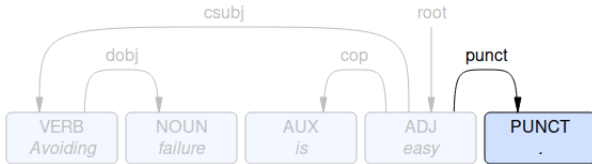
Example



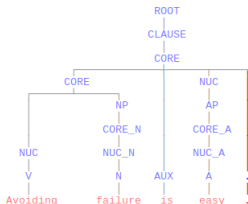
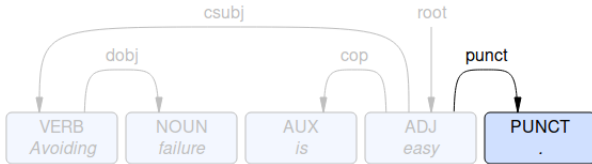
Example



Example



Example



Example

