Køpsala

Transition-Based Graph Parsing via Efficient Training and Effective Encoding

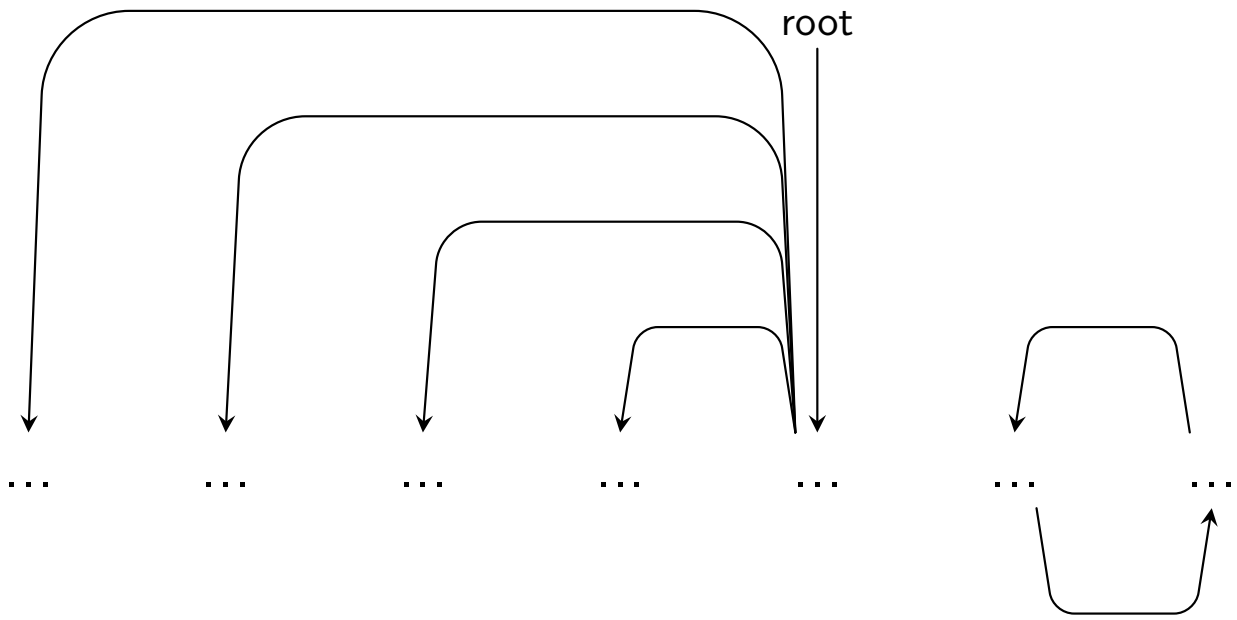
Daniel Hershcovich[*♢] **Miryam de Lhoneux**[*♢]
Artur Kulmizev[♡] **Elham Pejhan**[♢]
Joakim Nivre[♡]

[♢]University of Copenhagen
[♡]Uppsala University

IWPT 2020 Shared Task

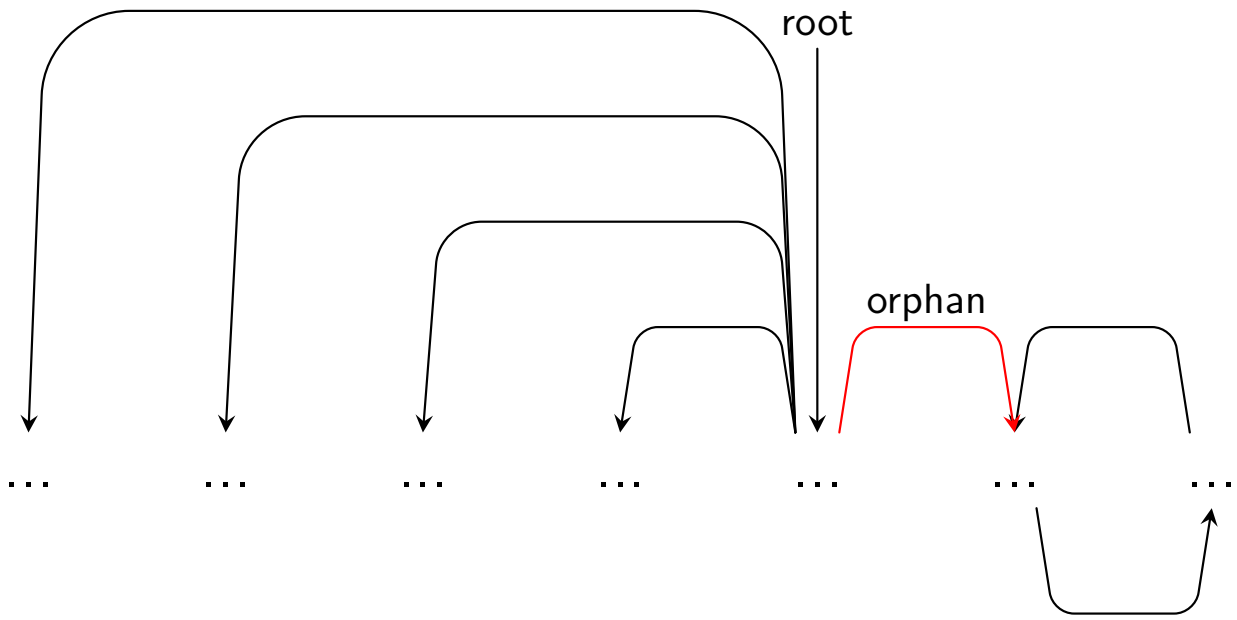
Parsing Tragedy

Unconnected graphs last-minute fix...

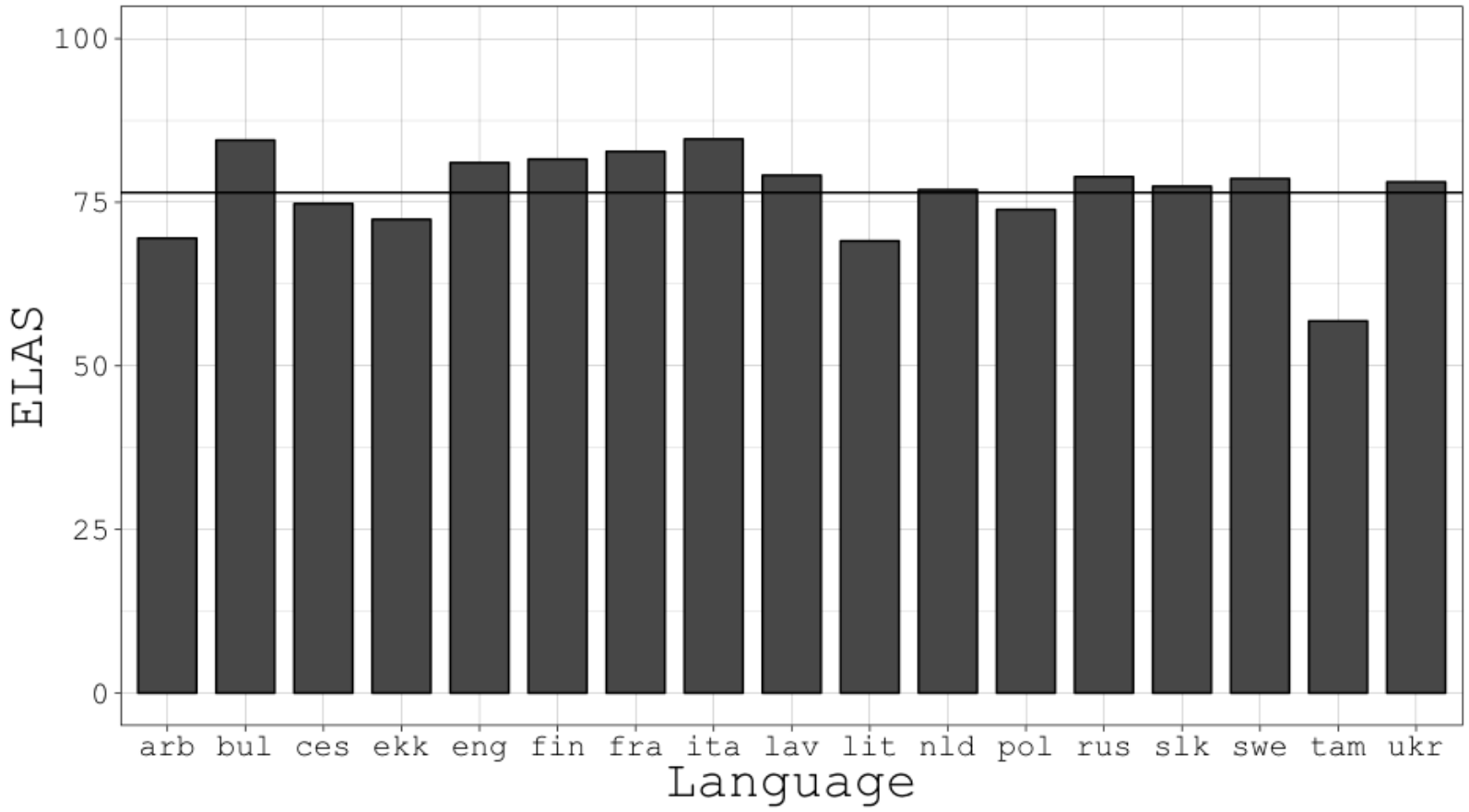


Parsing Tragedy

Unconnected graphs last-minute fix...

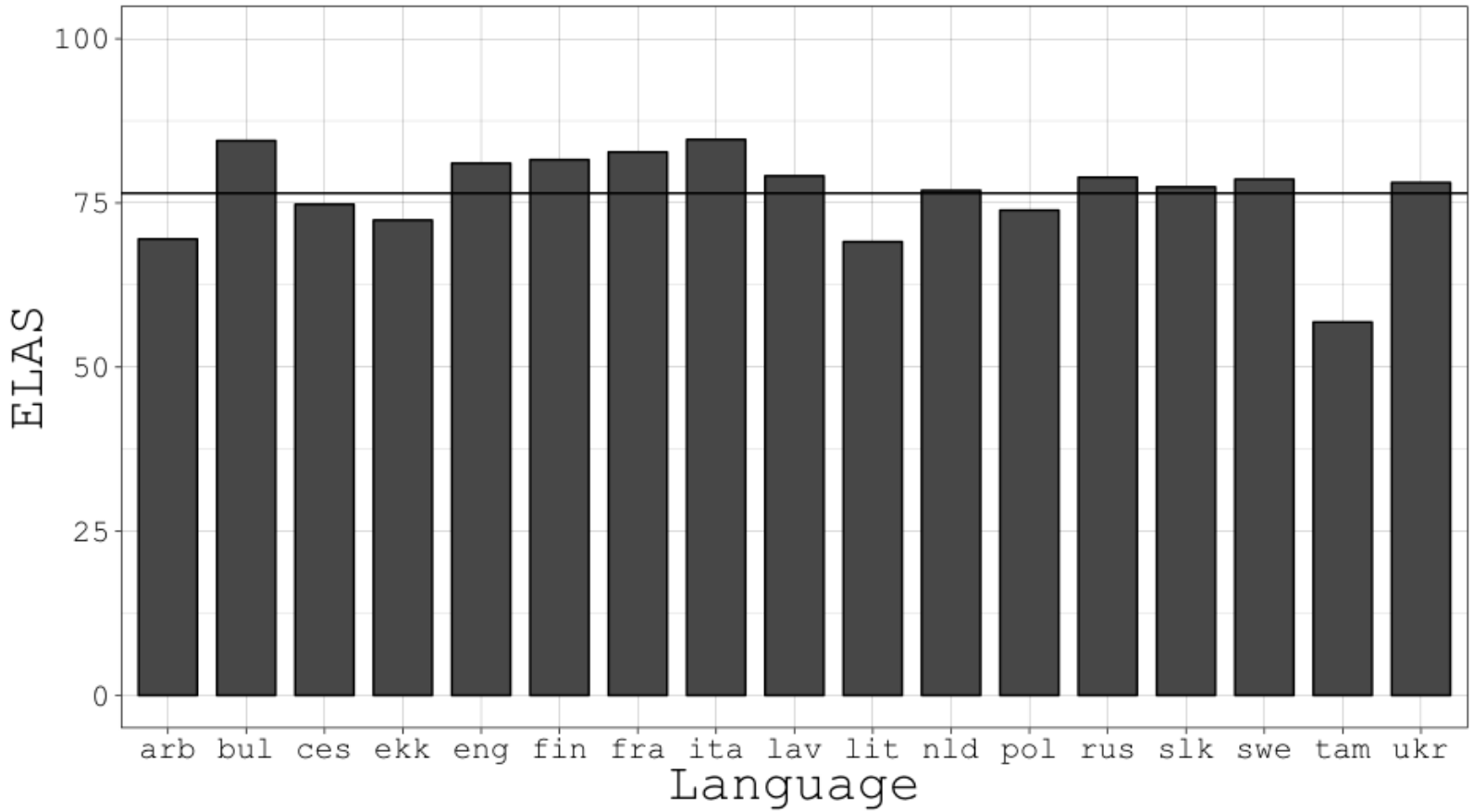


(Unofficial) Fixed Test Results



Unofficial 6th place, tied with FASTPARSE

(Unofficial) Fixed Test Results



Unofficial 6th place, tied with FASTPARSE

A Transition-based Parser

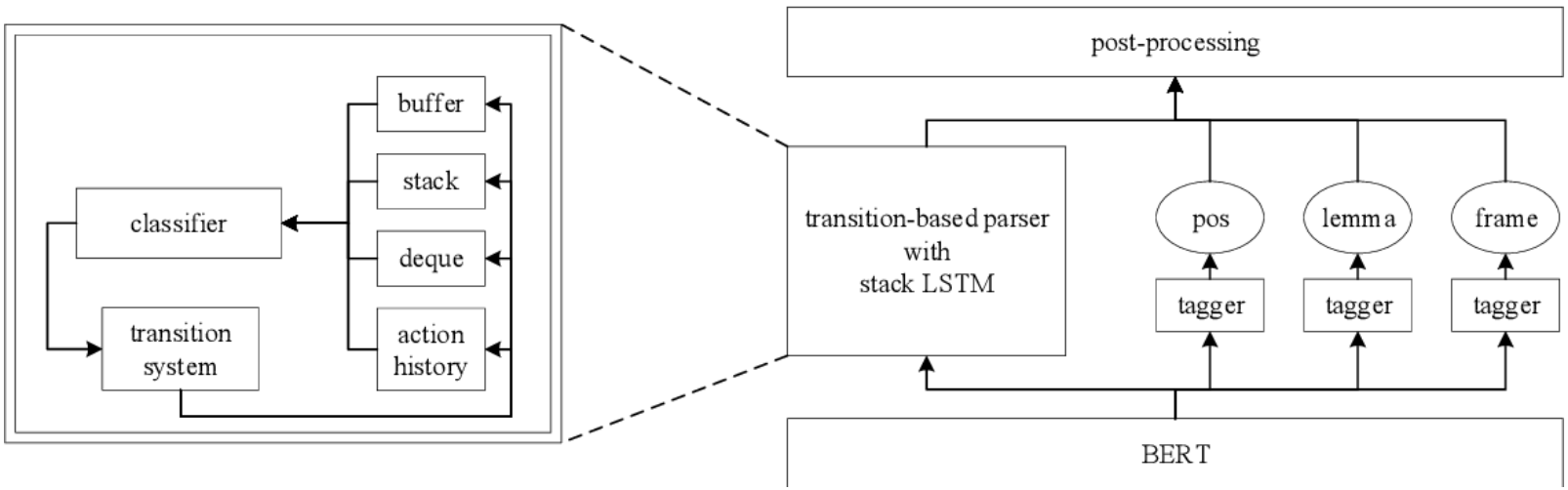
HIT-SCIR at MRP 2019:
A Unified Pipeline for Meaning Representation Parsing via Efficient Training and Effective Encoding

Wanxiang Che, Longxu Dou, Yang Xu, Yuxuan Wang, Yijia Liu, Ting Liu
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{car,lxdou,yxu,yxwang,yjliu,tliu}@ir.hit.edu.cn

A Transition-based Parser

HIT-SCIR at MRP 2019:
A Unified Pipeline for Meaning Representation Parsing via Efficient Training and Effective Encoding

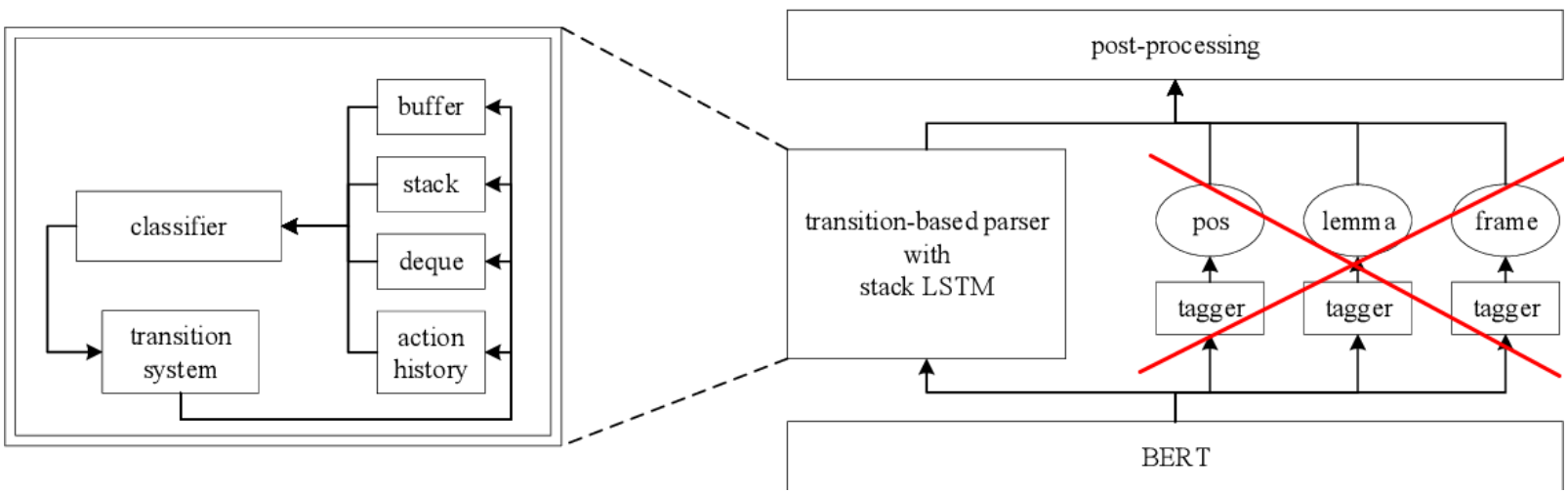
Wanxiang Che, Longxu Dou, Yang Xu, Yuxuan Wang, Yijia Liu, Ting Liu
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{car,lxdou,yxu,yxwang,yjliu,tliu}@ir.hit.edu.cn



A Transition-based Parser

HIT-SCIR at MRP 2019:
A Unified Pipeline for Meaning Representation Parsing via Efficient Training and Effective Encoding

Wanxiang Che, Longxu Dou, Yang Xu, Yuxuan Wang, Yijia Liu, Ting Liu
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{car,lxdou,yxu,yxwang,yjliu,tliu}@ir.hit.edu.cn



A Transition-based Parser

HIT-SCIR at MRP 2019:
A Unified Pipeline for Meaning Representation Parsing via Efficient Training and Effective Encoding

Wanxiang Che, Longxu Dou, Yang Xu, Yuxuan Wang, Yijia Liu, Ting Liu
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{car,lxdou,yxu,yxwang,yjliu,tliu}@ir.hit.edu.cn

DM & PSD	UCCA	EDS	AMR
Shift	Shift	Shift	Shift
Reduce	Reduce	Reduce	Reduce
Left-Edge	Left-Edge	Left-Edge	Left-Edge
Right-Edge	Right-Edge	Right-Edge	Right-Edge
Pass	Left-Remote	Drop	Drop
Finish	Right-Remote	Node-Start	Pass
	Node	Node-End	Merge
	Swap	Pass	Confirm
	Finish	Finish	Entity
			New
			Finish

Køpsala

Transition-Based Graph Parsing via Efficient Training and Effective Encoding

Daniel Hershcovich[*♢] **Miryam de Lhoneux**[*♢]
Artur Kulmizev[♡] **Elham Pejhan**[♢]
Joakim Nivre[♡]

[♢]University of Copenhagen
[♡]Uppsala University

IWPT 2020 Shared Task

Transition system for Enhanced UD

- Left-Edge:label
- Right-Edge:label
- Shift
- Swap
- Node
- Reduce-0
- Reduce-1
- Finish

Transitions

Transitions

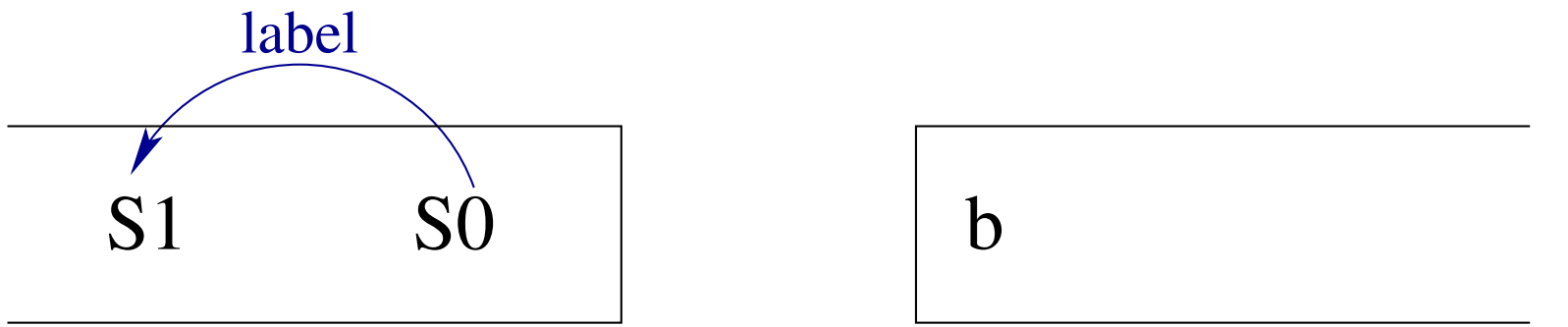
LEFT-EDGE:LABEL

stack

buffer

Transitions

LEFT-EDGE:LABEL



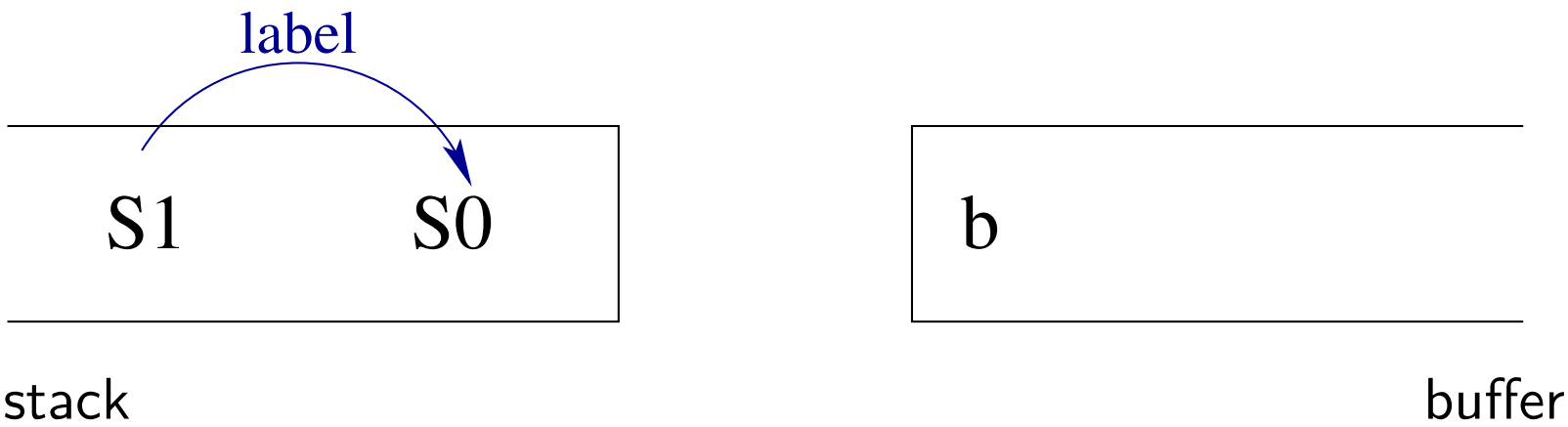
Transitions

RIGHT-EDGE:LABEL

stack buffer

Transitions

RIGHT-EDGE:LABEL



Transitions

SHIFT

stack buffer

Transitions

SHIFT

Transitions

SWAP

stack buffer

Transitions

SWAP

stack buffer

Transitions

Node

stack

buffer

Transitions

Node

Transitions

REDUCE-0

stack

buffer

Transitions

REDUCE-0

stack buffer

Transitions

Reduce-1

stack

buffer

Transitions

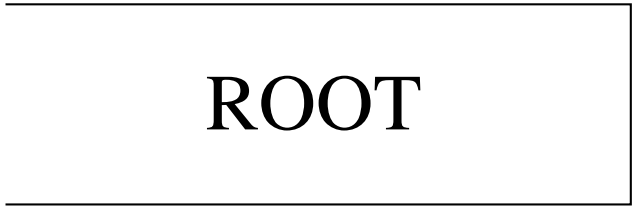
Reduce-1

stack

buffer

Transitions

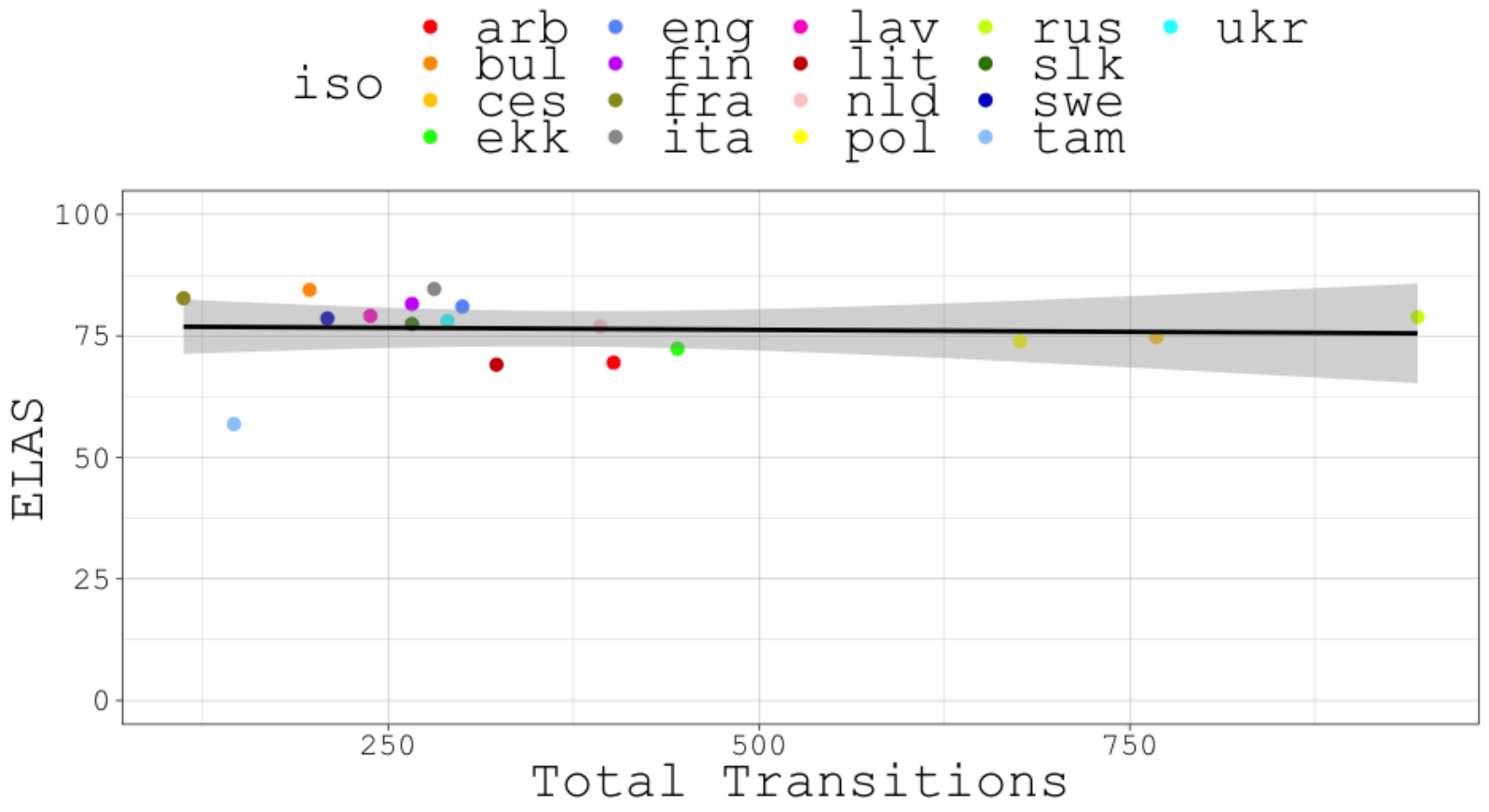
Finish



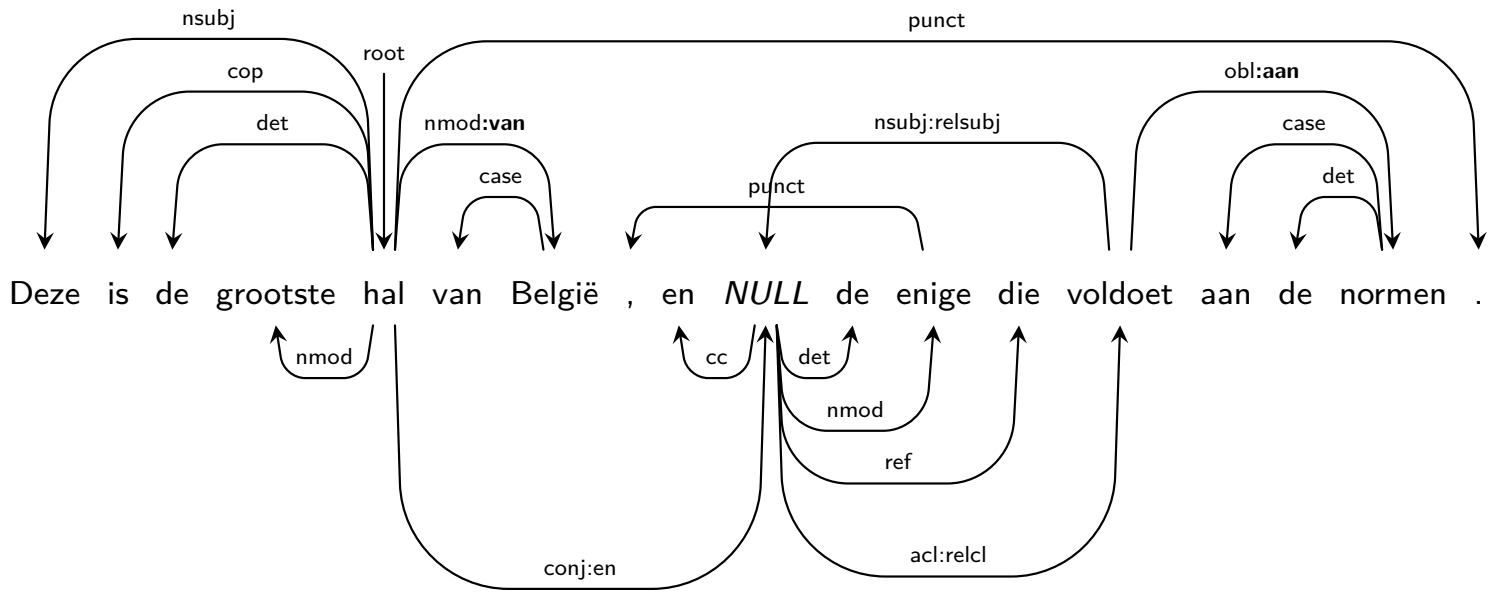
stack

buffer

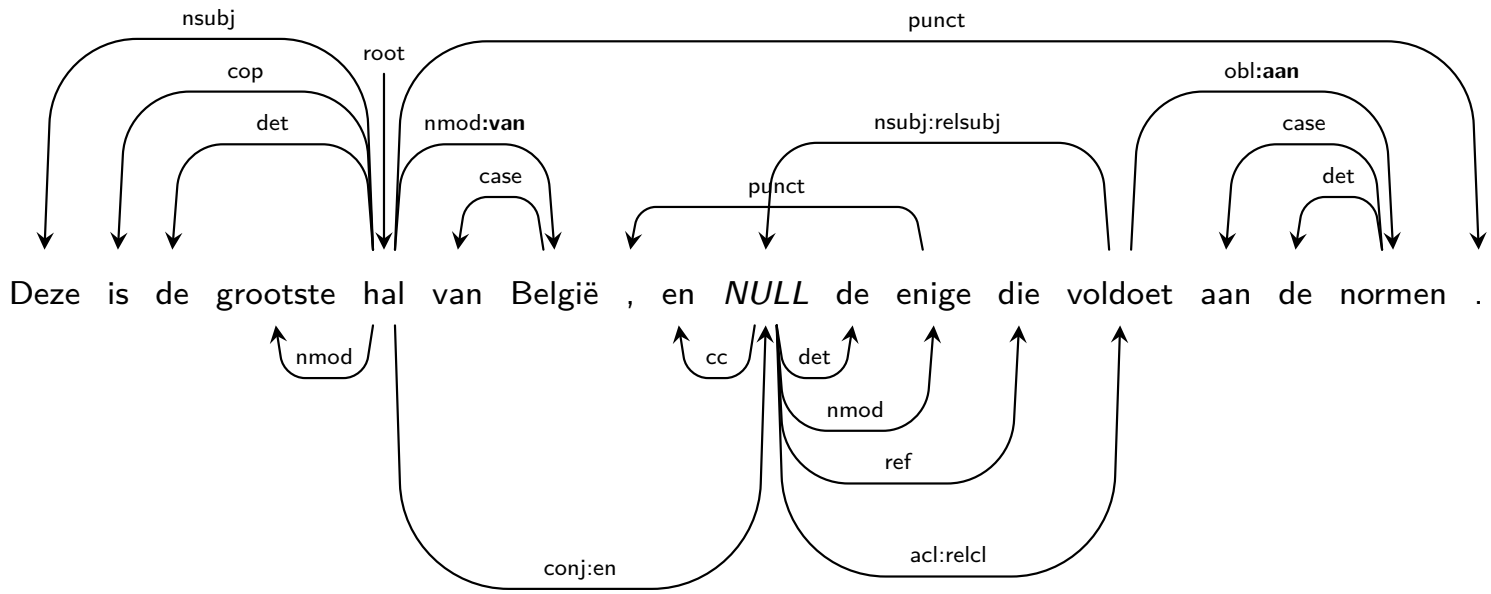
Transition Set Size



Transition sequence

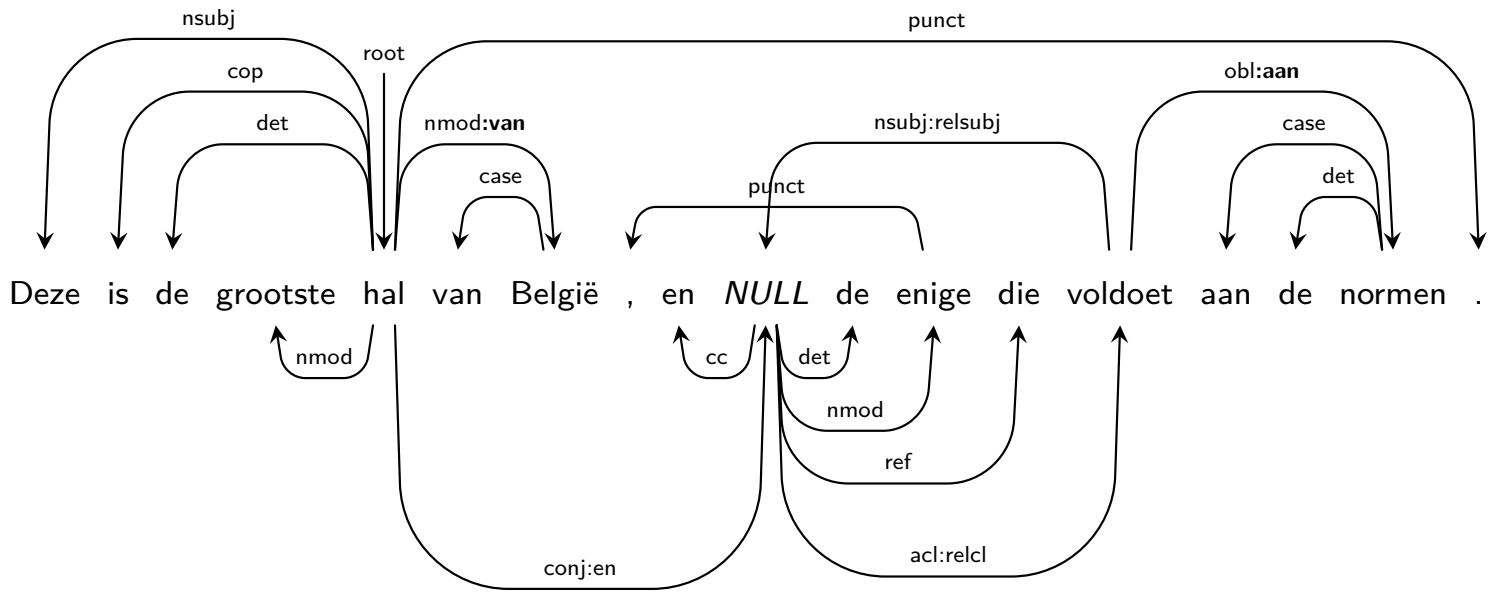


Transition sequence



[ROOT] [Deze is de grootste hal van België , en de enige die voldoet aan de normen .]

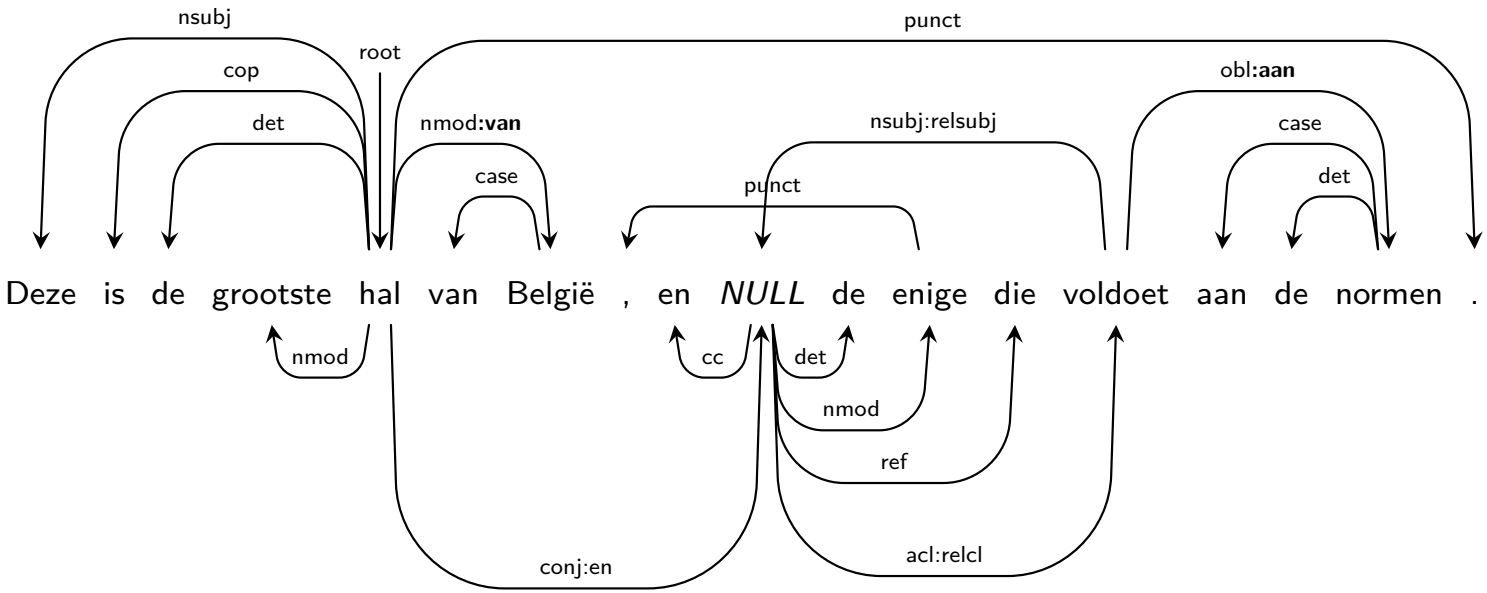
Transition sequence



[ROOT] [Deze is de grootste hal van België , en de (...)]

Transition sequence

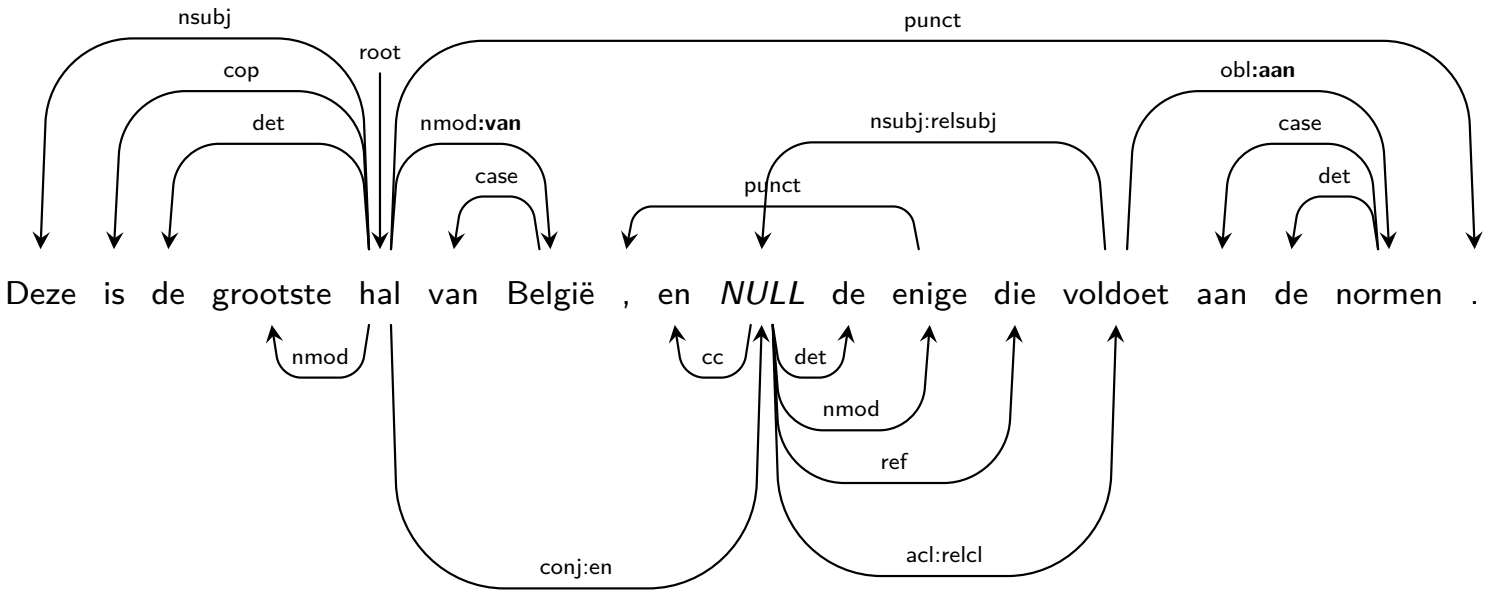
Shift



[ROOT Deze] [is de grootste hal van België , en de (...)]

Transition sequence

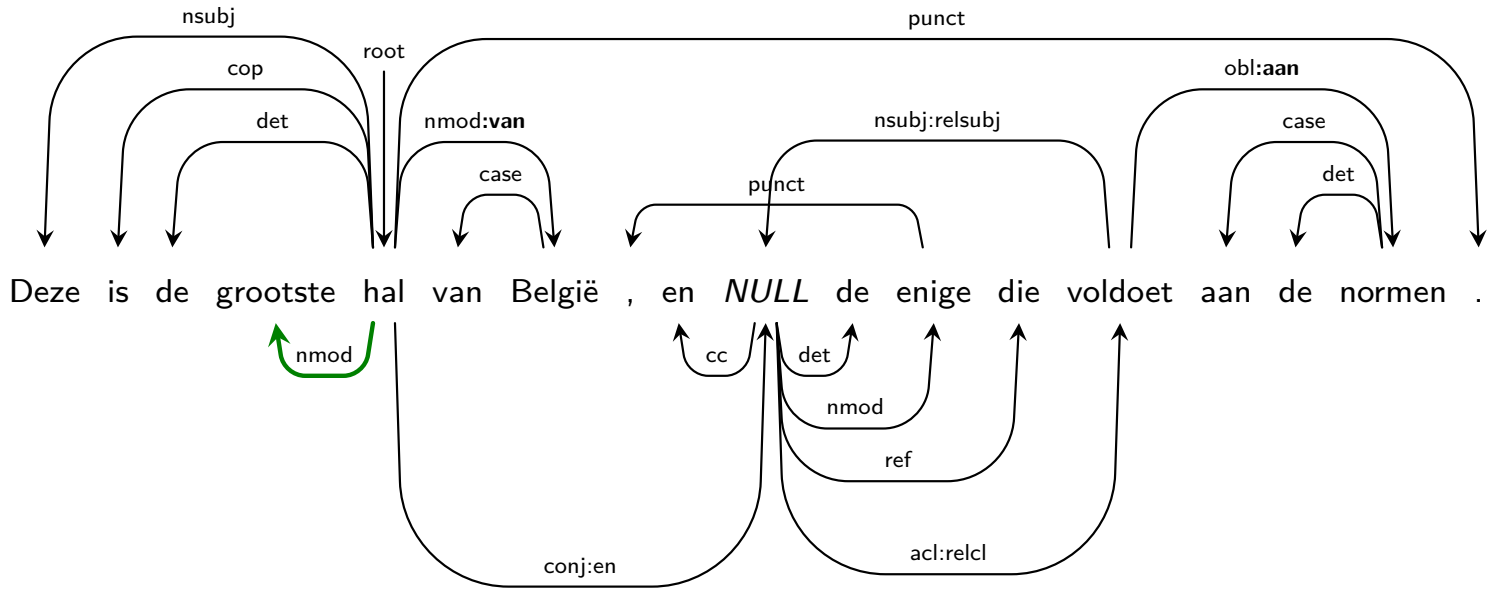
SHIFT x4



[ROOT Deze is de grootste hal] [van België , en de (...)]

Transition sequence

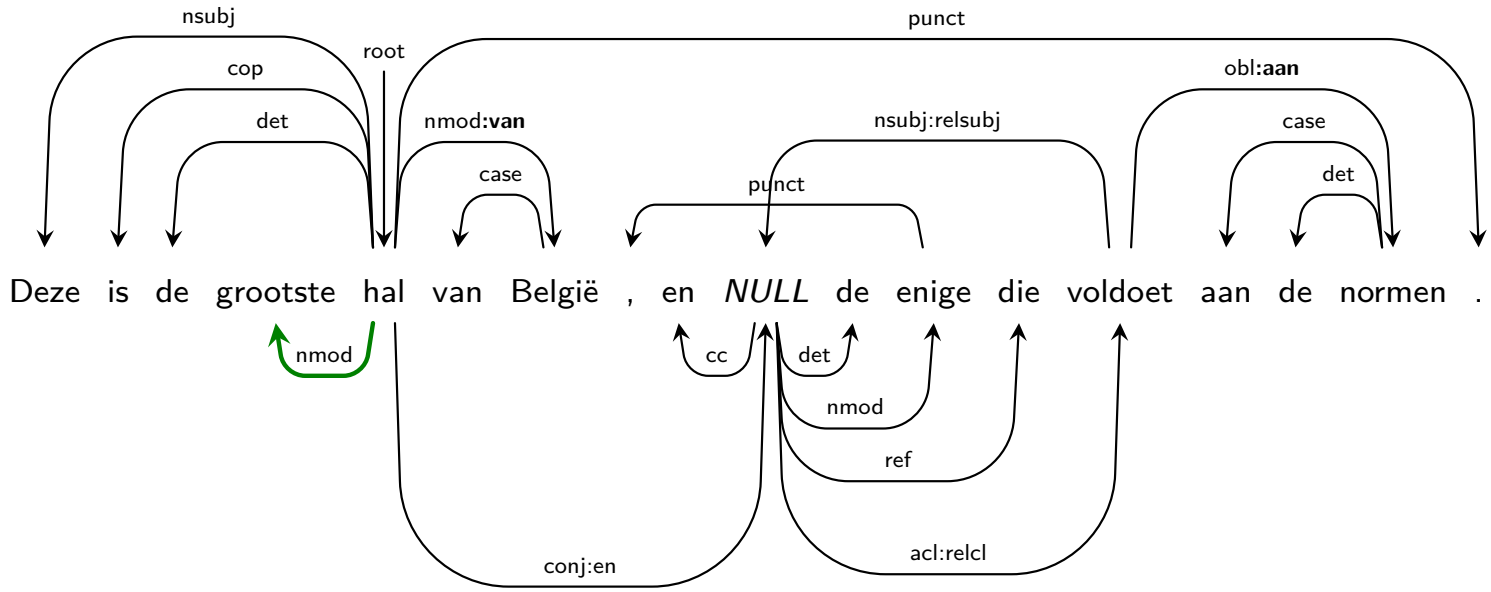
Left-Edge:nmod



[ROOT Deze is de grootste hal] [van België , en de (...)]

Transition sequence

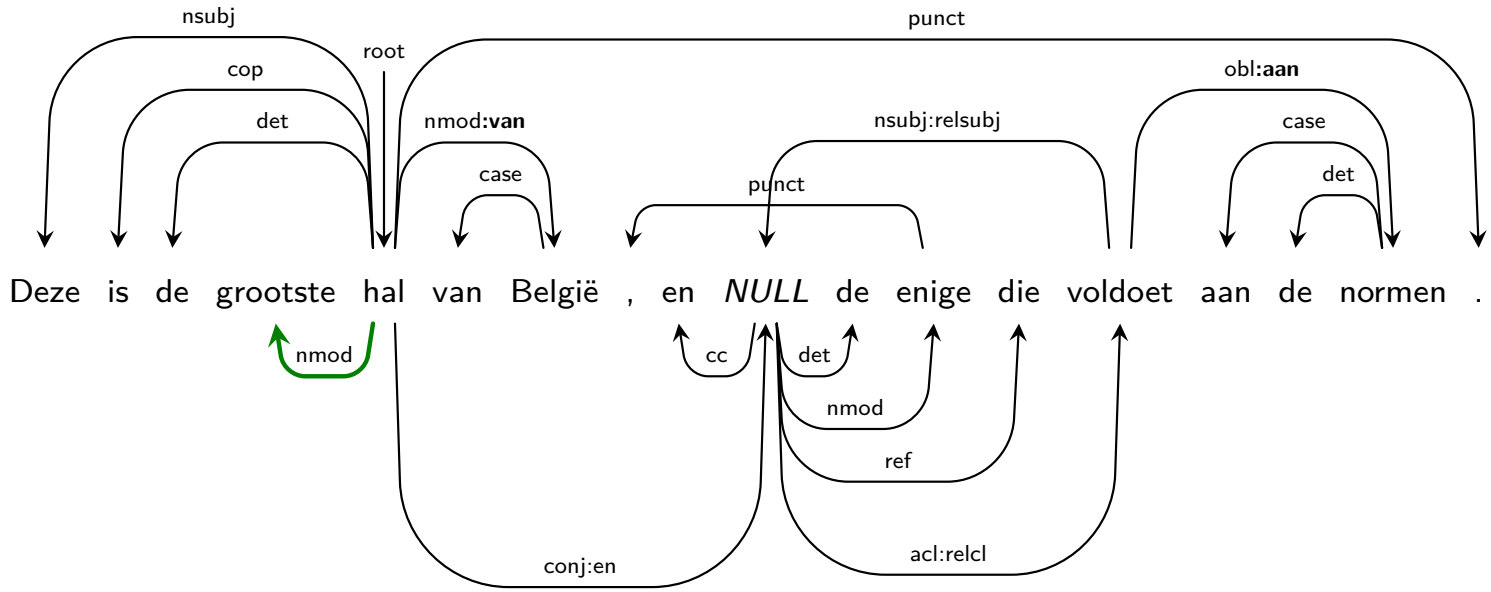
Node



[ROOT Deze is de grootste hal] [NULL van België , en (...)]

Transition sequence

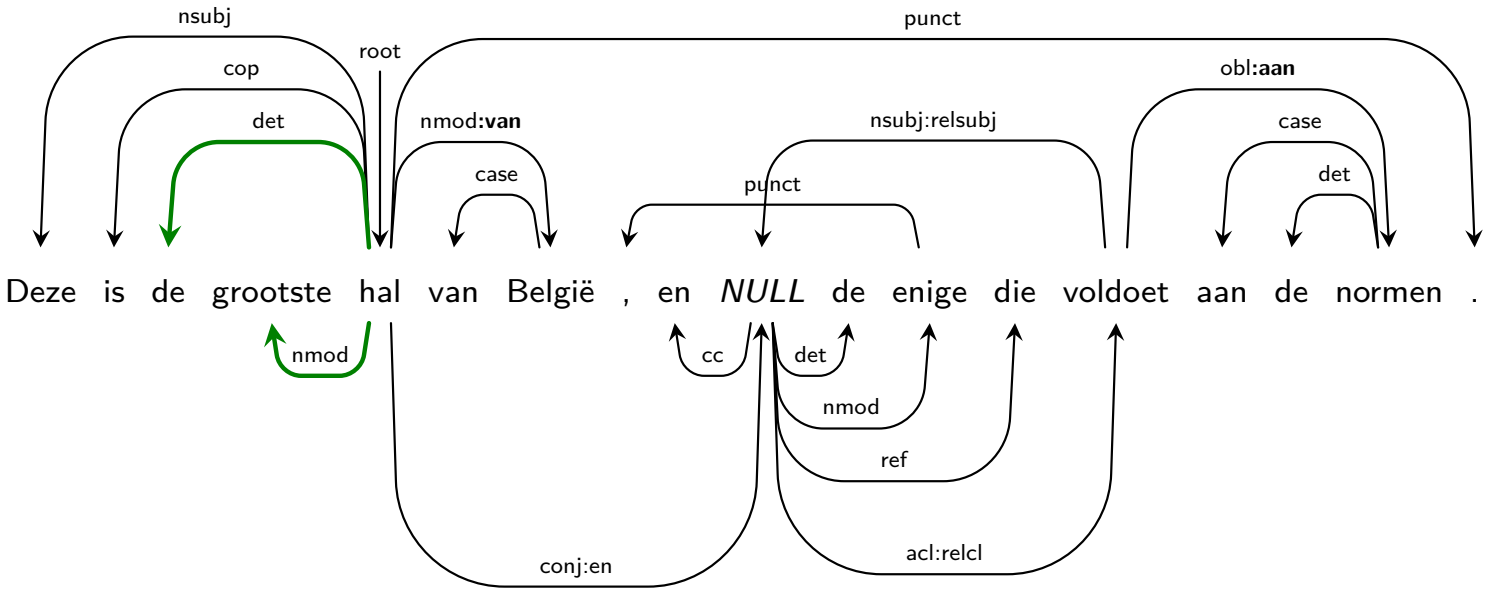
Reduce-1



[ROOT Deze is de hal] [NULL van België , en de enige (...)]

Transition sequence

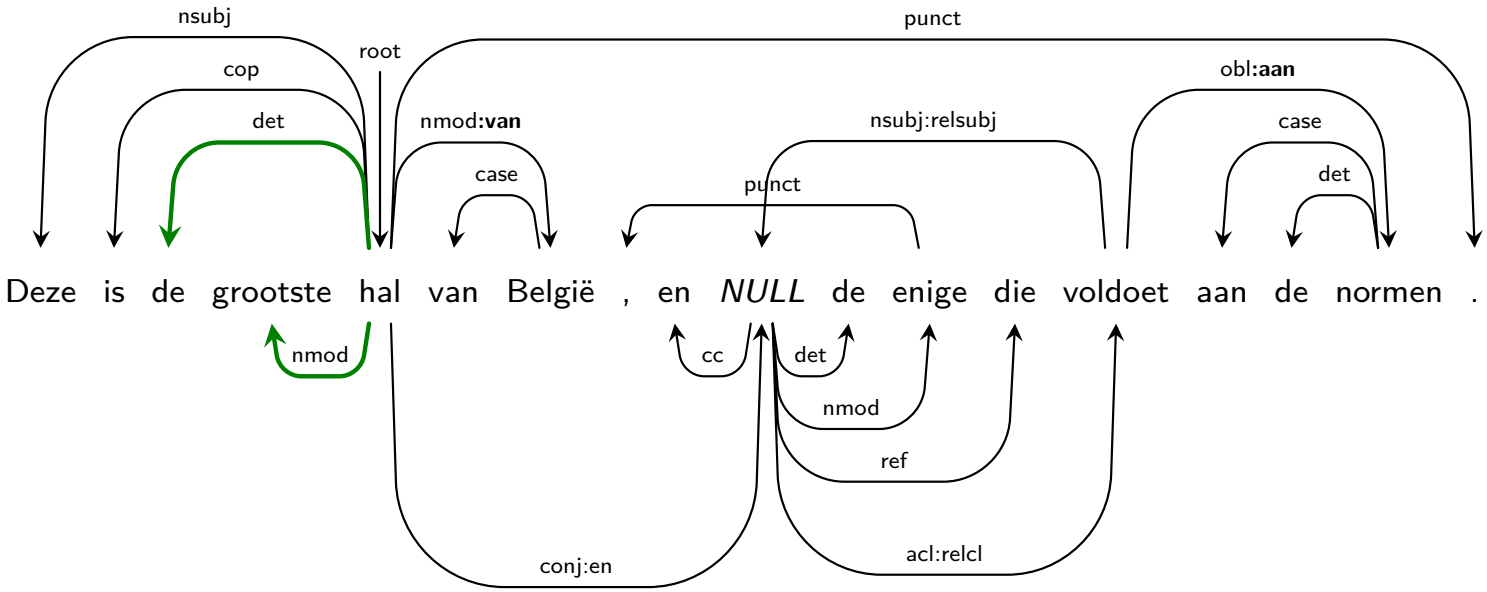
Left-Edge:det



[ROOT Deze is de hal] [NULL van België , en de enige (...)]

Transition sequence

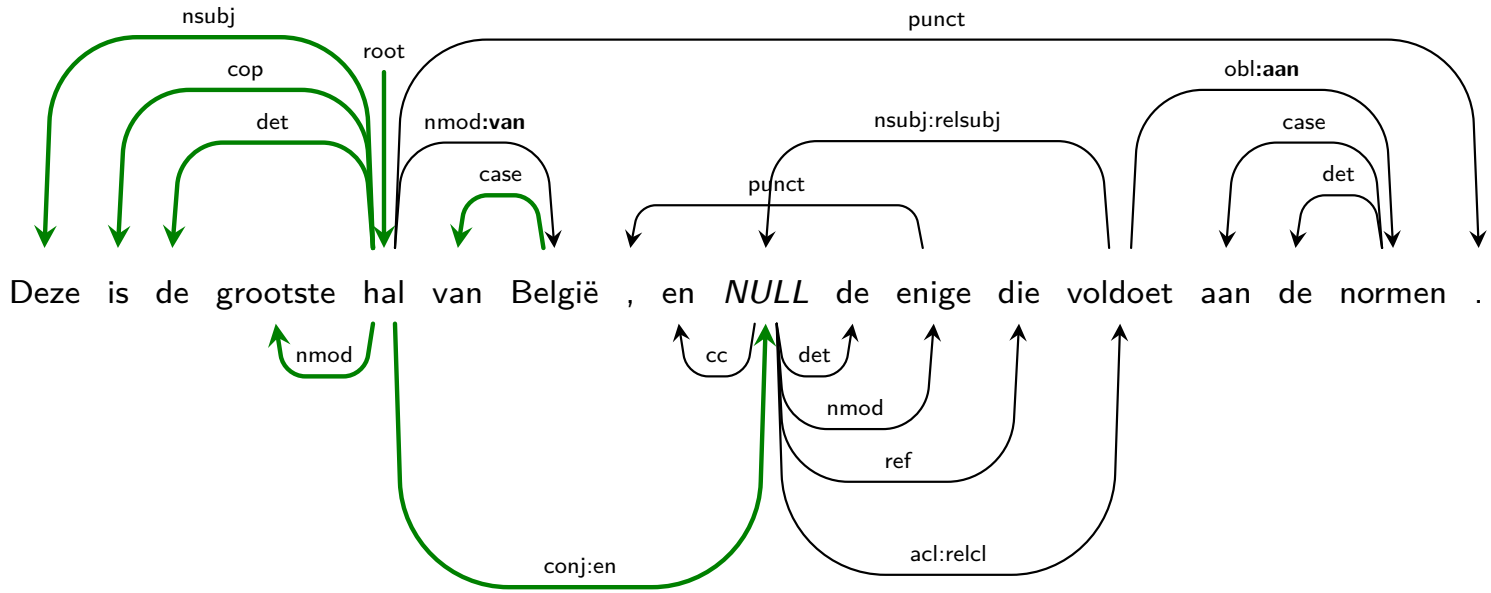
Reduce-1



[ROOT Deze is hal] [NULL van België , en de enige die (...)]

Transition sequence

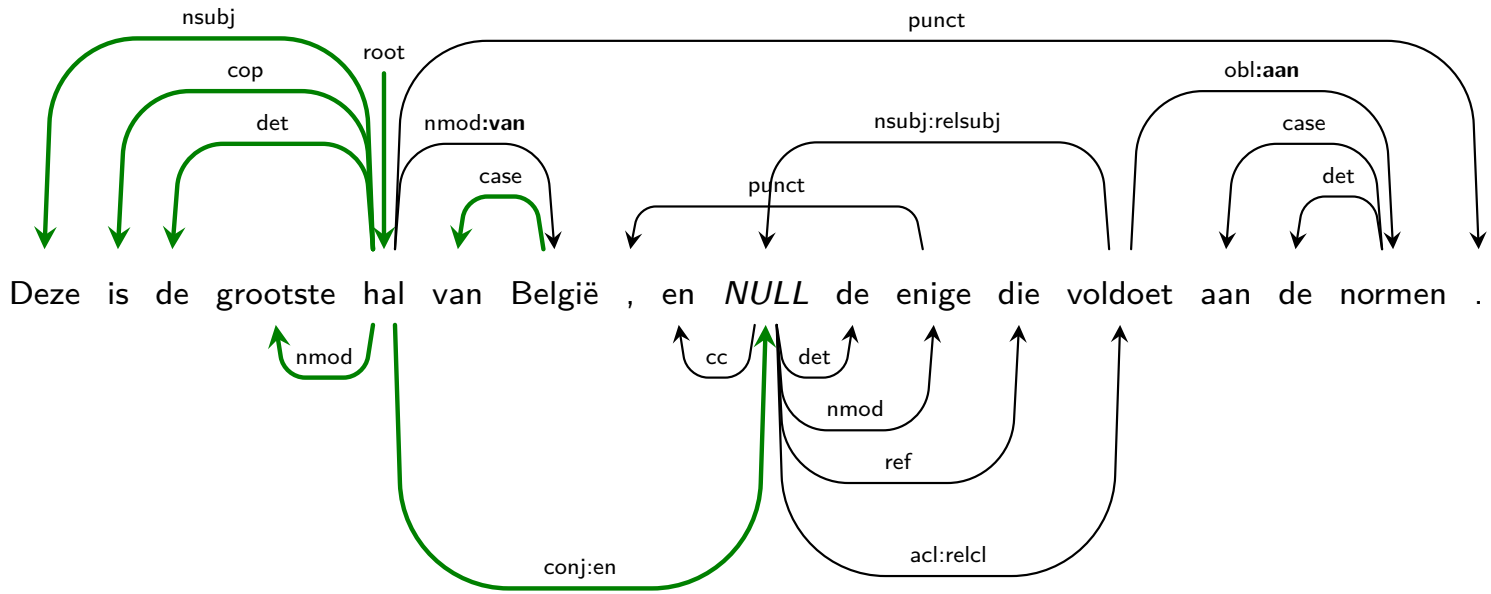
...



[ROOT hal NULL België] [, en de enige die voldoet aan (...)]

Transition sequence

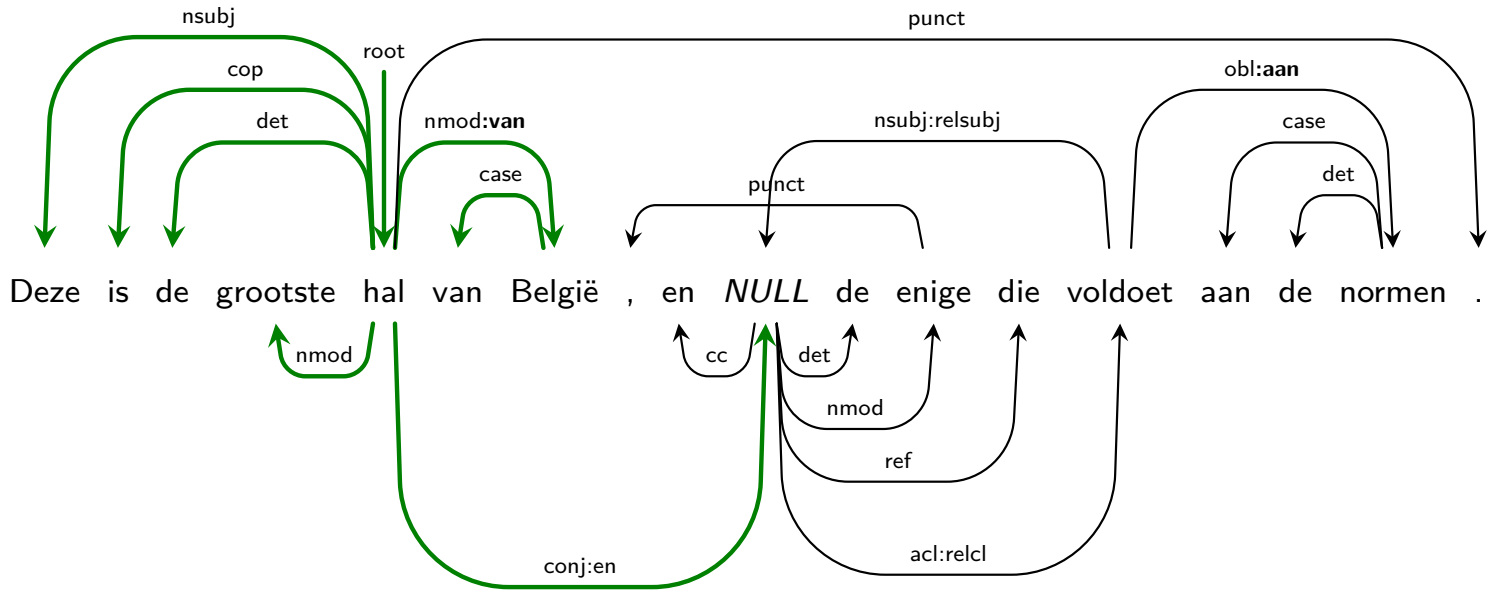
Swap



[ROOT hal België] [NULL , en de enige die voldoet aan (...)]

Transition sequence

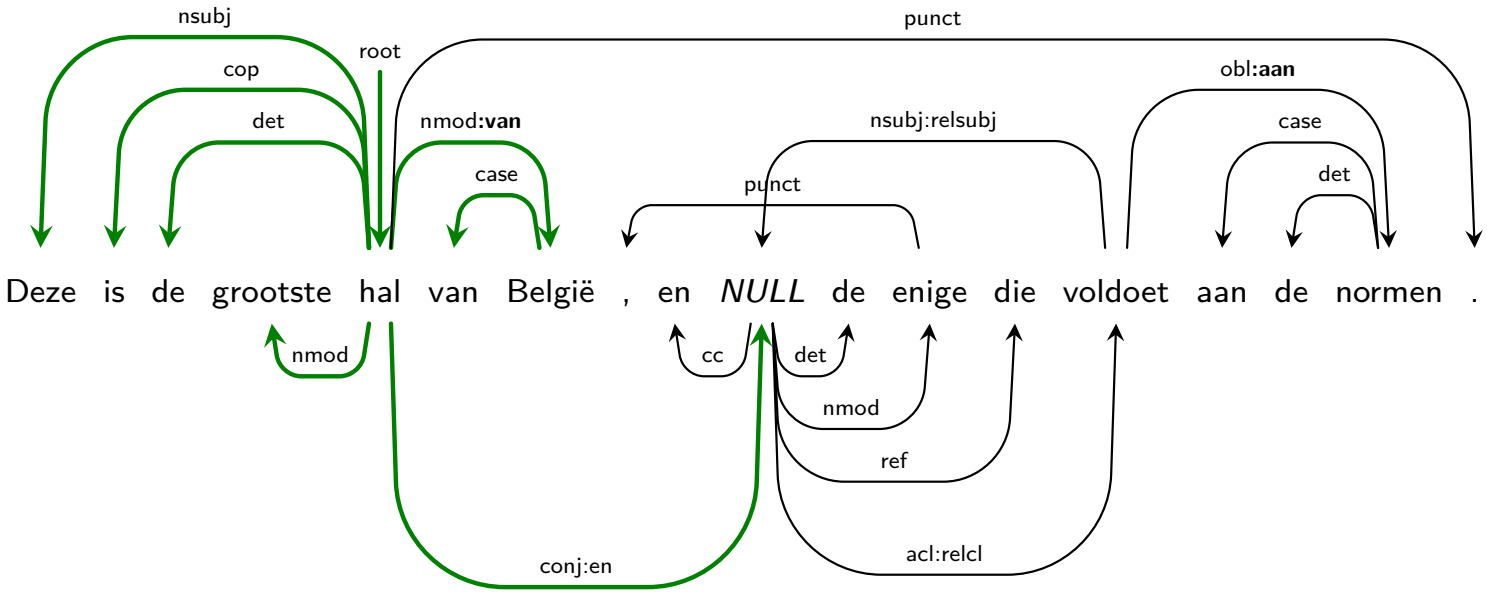
RIGHT-EDGE:NMOD:VAN



[ROOT hal België] [NULL , en de enige die voldoet aan (...)]

Transition sequence

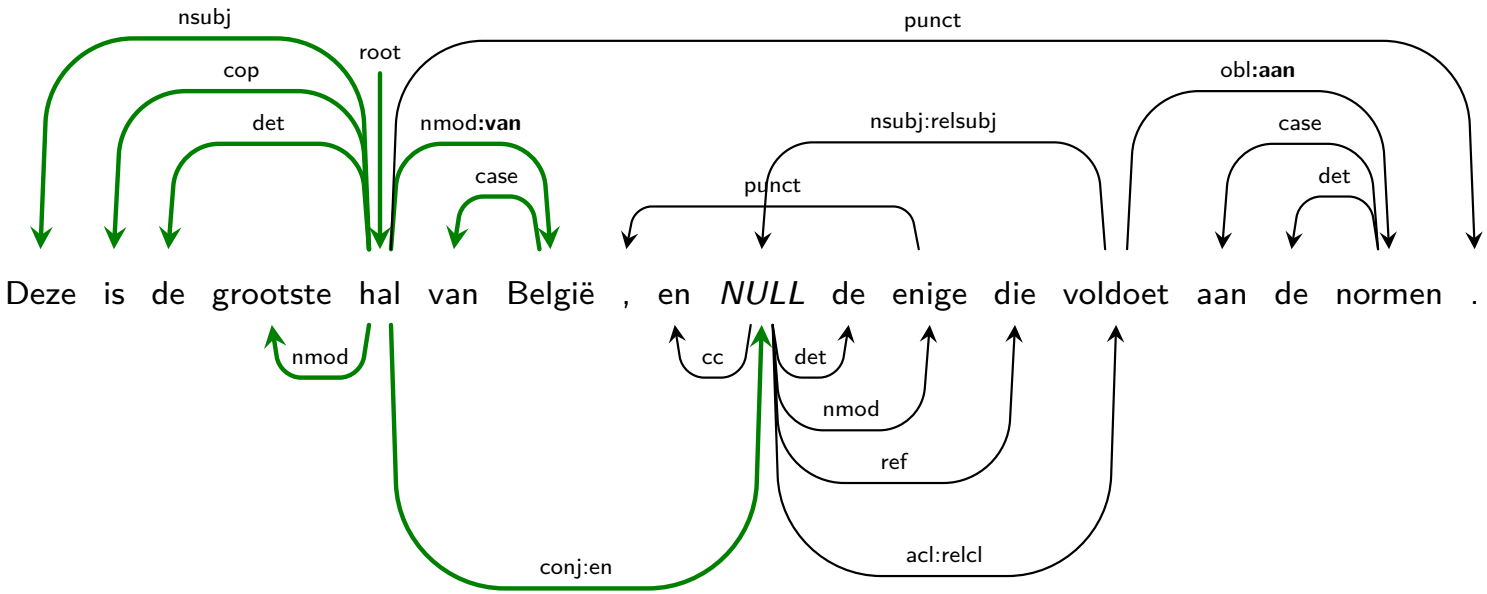
Reduce-0



[ROOT hal] [NULL , en de enige die voldoet aan de normen]

Transition sequence

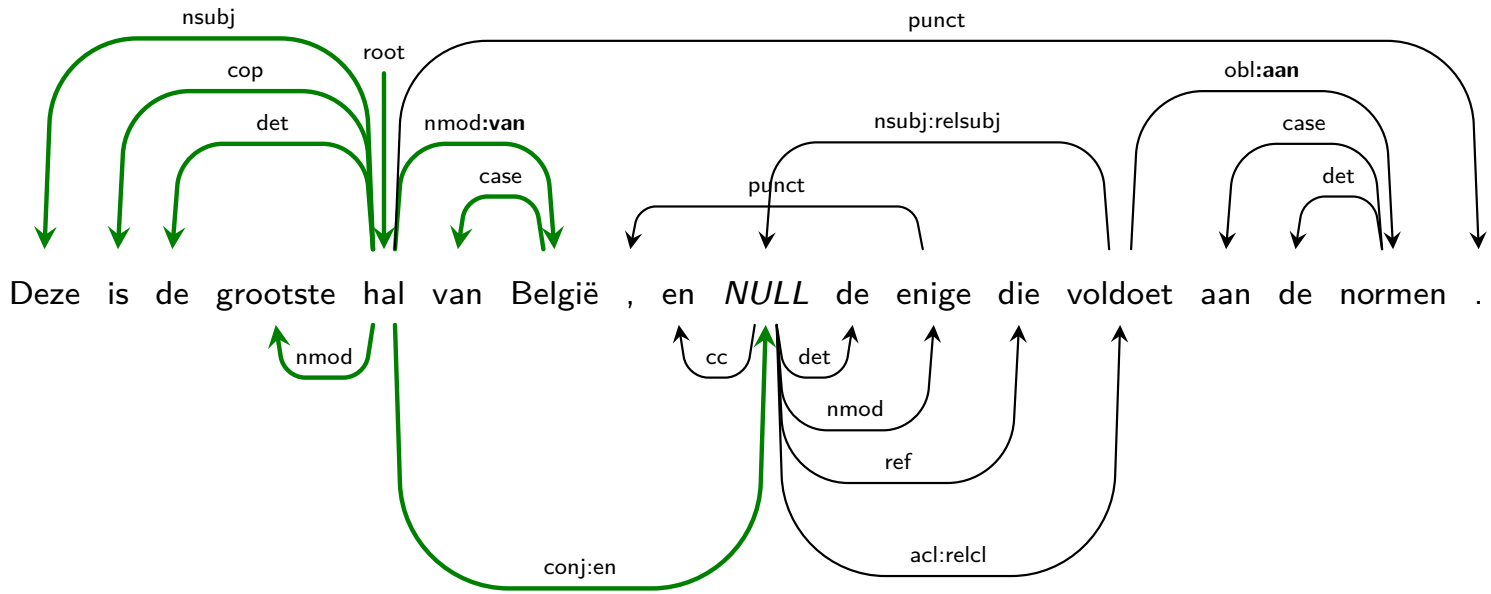
Shift x3



[ROOT hal NULL , en] [de enige die voldoet aan de normen]

Transition sequence

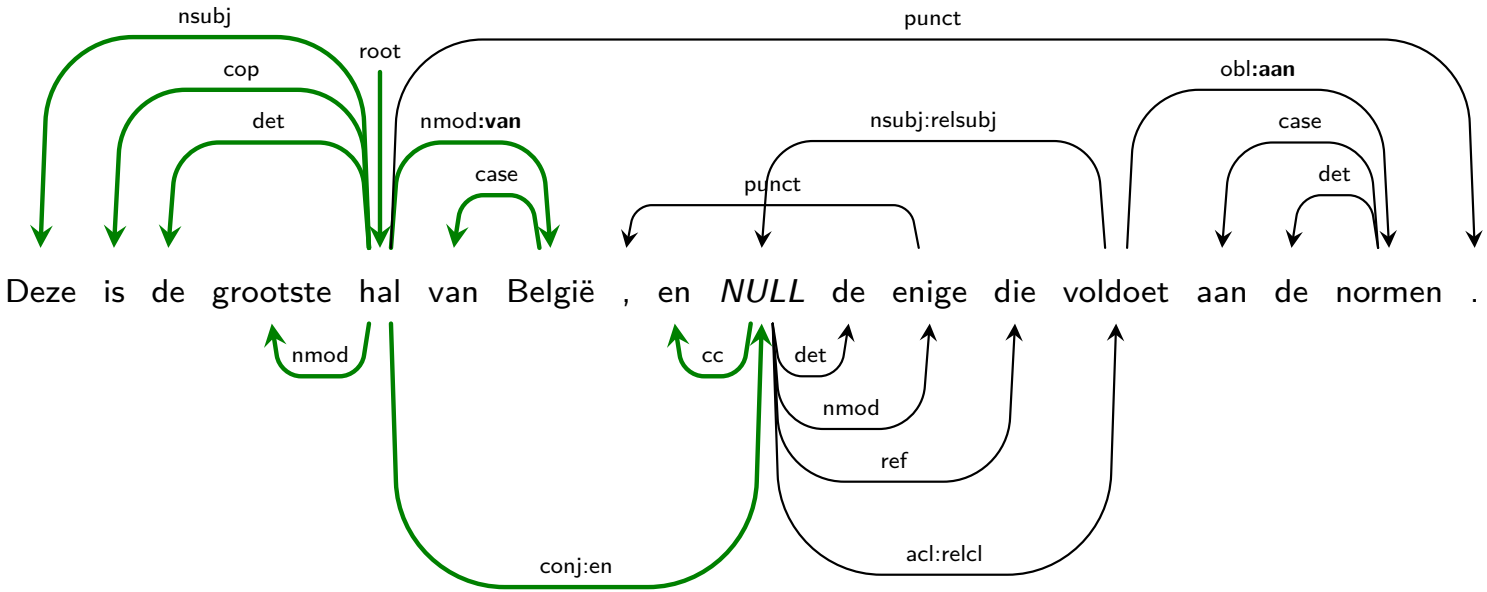
Swap



[ROOT hal NULL en] [, de enige die voldoet aan de normen]

Transition sequence

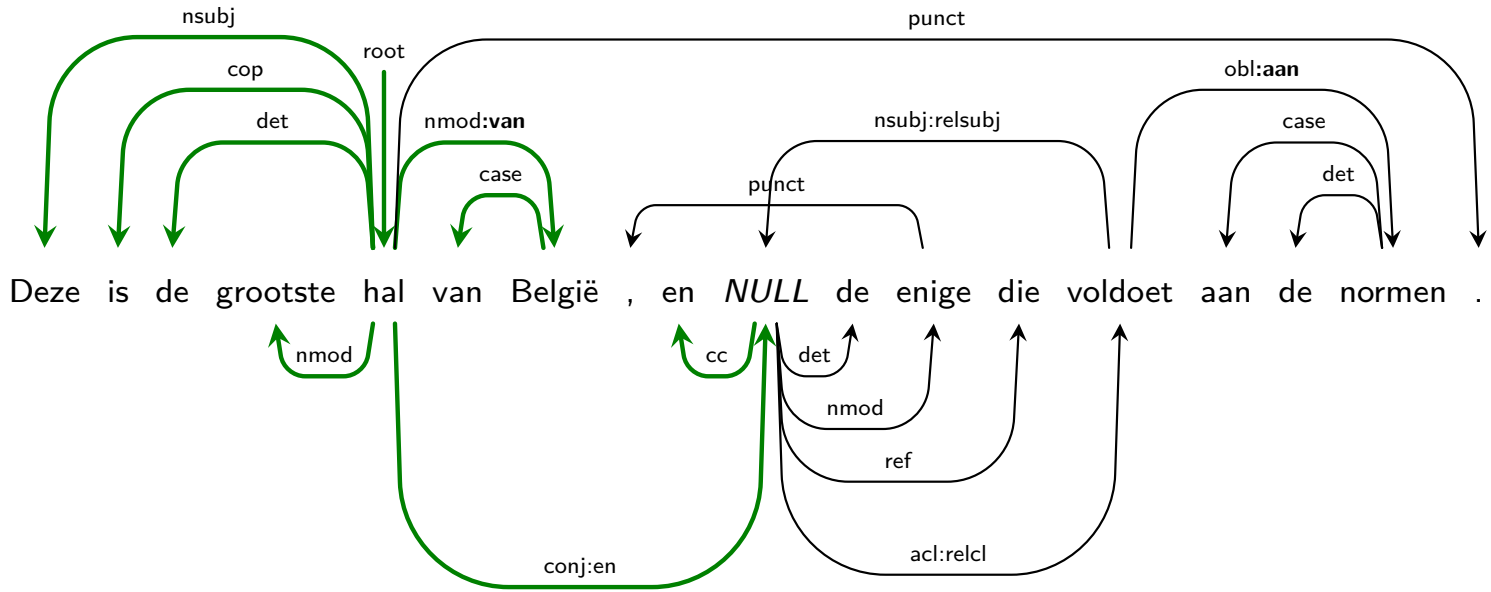
RIGHT-EDGE:CC



[ROOT hal NULL en] [, de enige die voldoet aan de normen]

Transition sequence

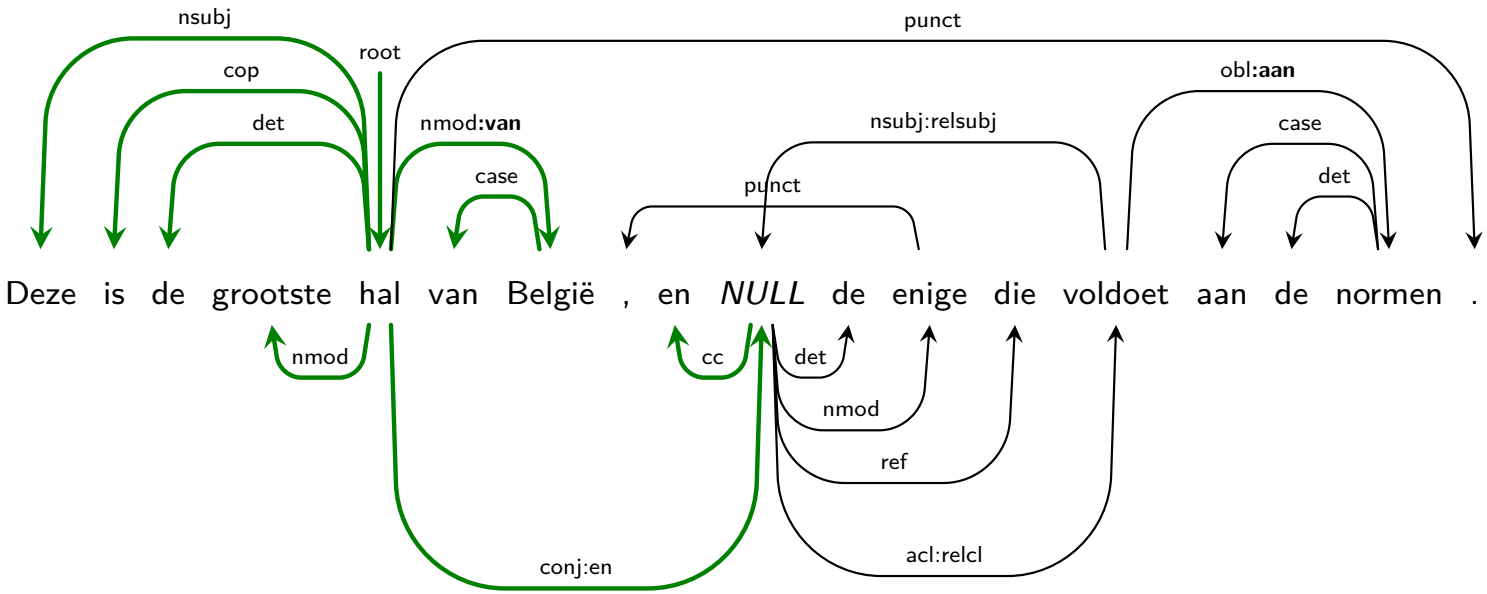
Reduce-0



[ROOT hal NULL] [, de enige die voldoet aan de normen]

Transition sequence

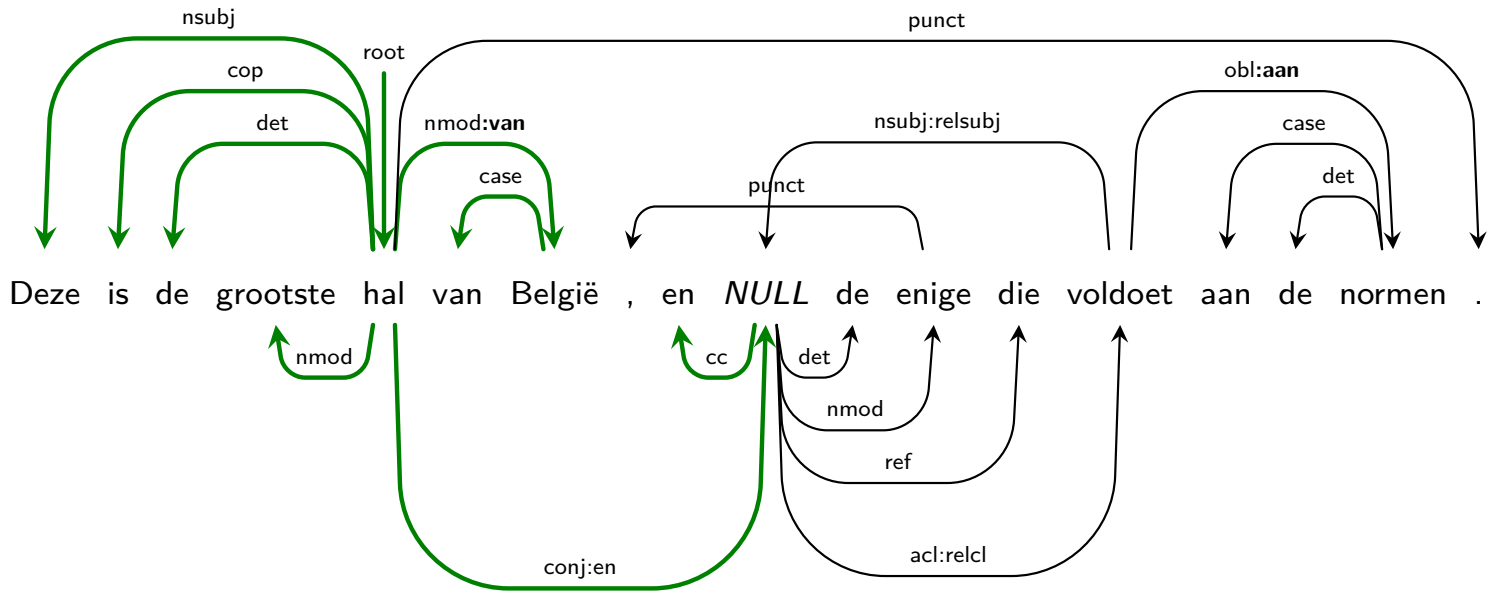
Shift x2



[ROOT hal NULL , de] [enige die voldoet aan de normen]

Transition sequence

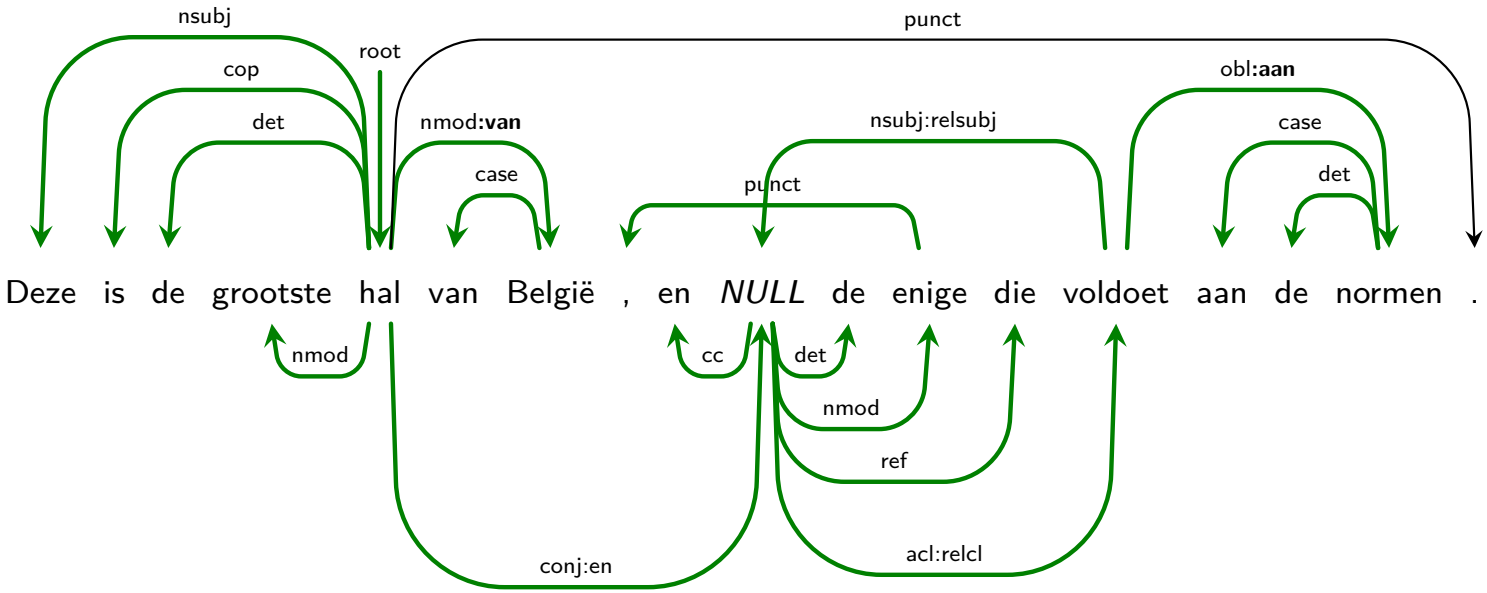
Swap



[ROOT hal NULL de] [, enige die voldoet aan de normen]

Transition sequence

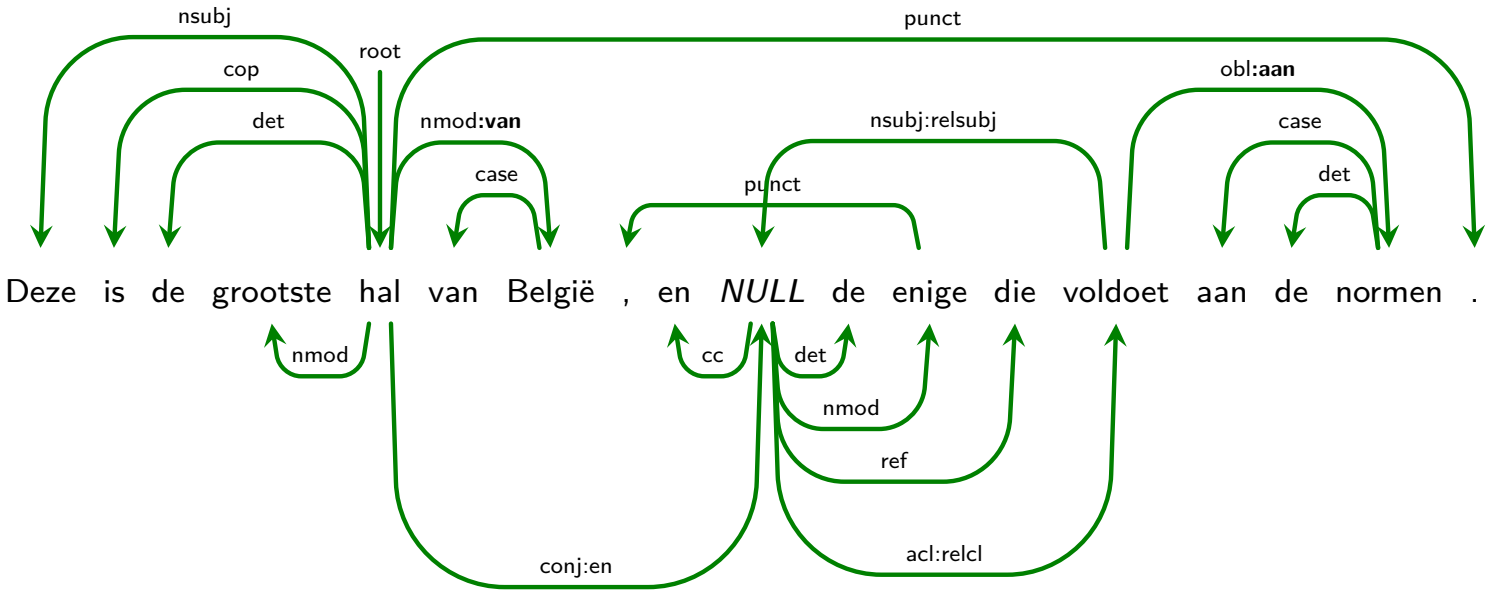
...



[ROOT hal .] [ ]

Transition sequence

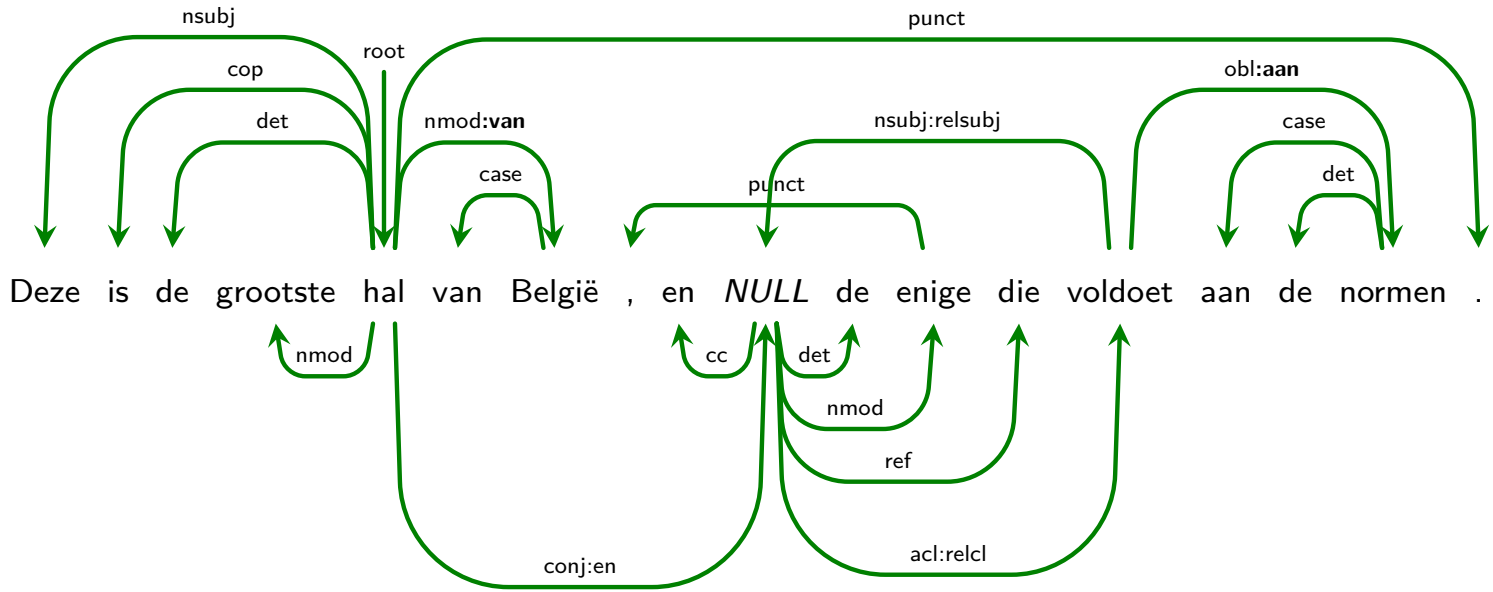
Right-Edge:punct



[ROOT hal .] [ ]

Transition sequence

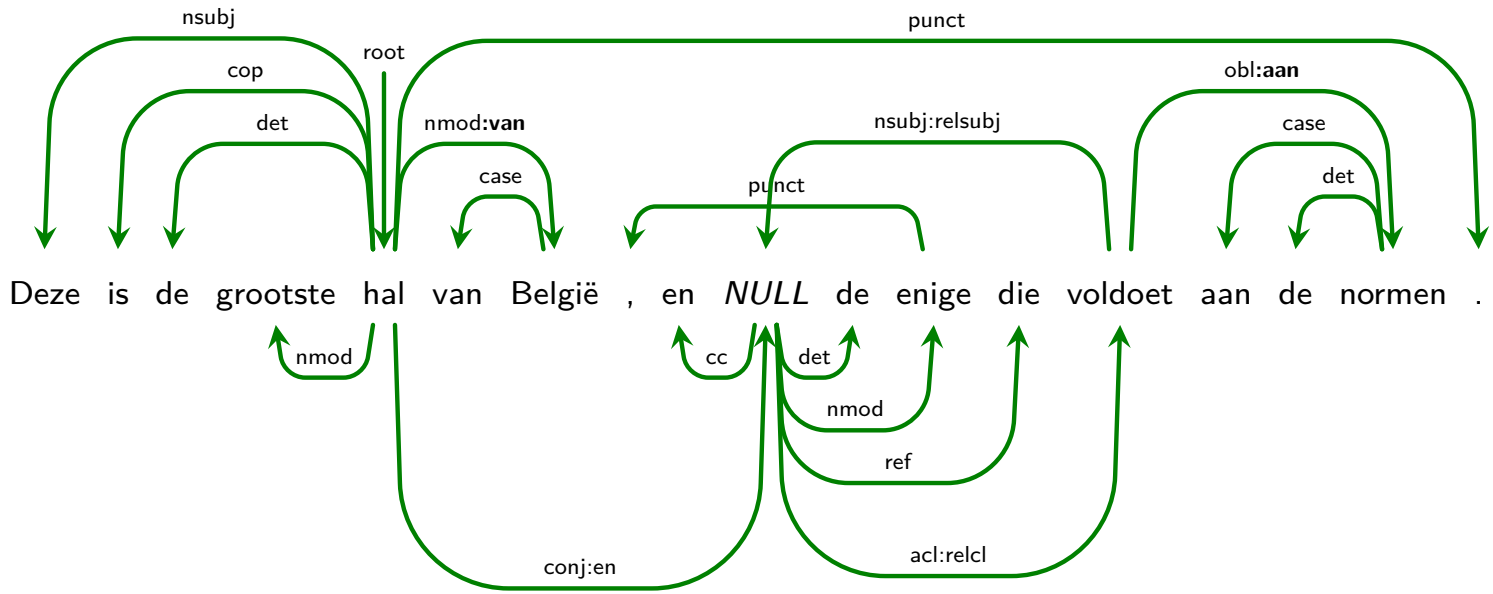
Reduce-0 x2



[ROOT] [ ]

Transition sequence

Finish



[ROOT] [ ]

Preprocessing

- Trained on gold inputs
- Tried two pipelines for dev and test:
  - STANZA
  - UDPIPE
- Including:
  - sentence segmentation
  - tokenization
  - lemmatization
  - PoS tagging
  - morphological tagging
  - basic UD parsing
- For languages with more than one treebank:
  - Using language's largest treebank

Preprocessing

- Trained on gold inputs
- Tried two pipelines for dev and test:
  - STANZA
  - UDPIPE
- Including:
  - sentence segmentation
  - tokenization
  - lemmatization
  - PoS tagging
  - morphological tagging
  - basic UD parsing
- For languages with more than one treebank:
  - Using language's largest treebank

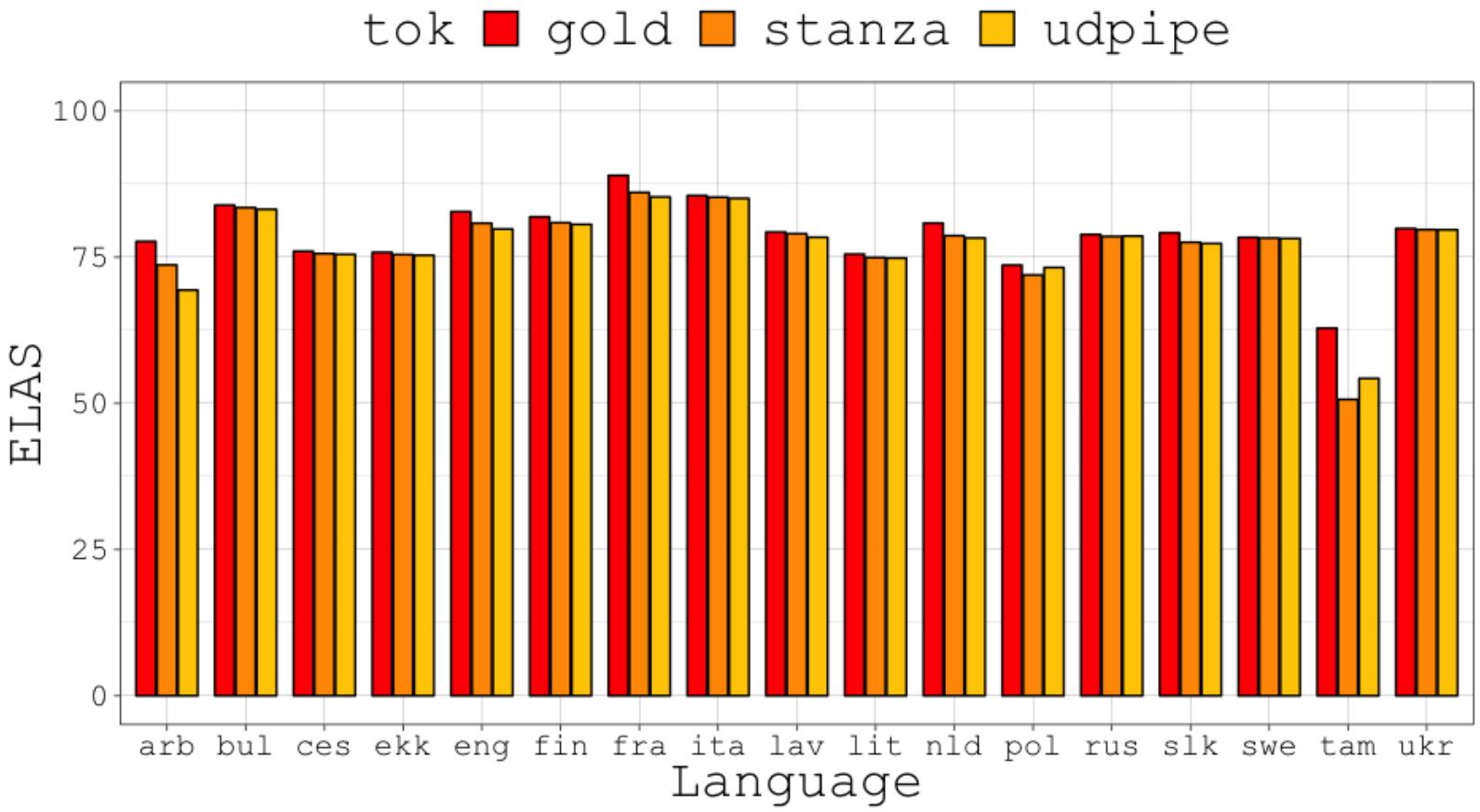
Preprocessing

- Trained on gold inputs
- Tried two pipelines for dev and test:
  - STANZA
  - UDPIPE
- Including:
  - sentence segmentation
  - tokenization
  - lemmatization
  - PoS tagging
  - morphological tagging
  - basic UD parsing
- For languages with more than one treebank:
  - Using language's largest treebank

Preprocessing

- Trained on gold inputs
- Tried two pipelines for dev and test:
  - STANZA
  - UDPIPE
- Including:
  - sentence segmentation
  - tokenization
  - lemmatization
  - PoS tagging
  - morphological tagging
  - basic UD parsing
- For languages with more than one treebank:
  - Using language's largest treebank

The Effect of Preprocessing



Køpsala

Transition-Based Graph Parsing via Efficient Training and Effective Encoding

Daniel Hershcovich[*♢] **Miryam de Lhoneux**[*♢]
Artur Kulmizev[♡] **Elham Pejhan**[♢]
Joakim Nivre[♡]

[♢]University of Copenhagen
[♡]Uppsala University

IWPT 2020 Shared Task