Projection-based Coreference Resolution Using Deep Syntax

Michal Novák, Anna Nedoluzhko and Zdeněk Žabokrtský
Charles University, Prague

CORBON 2017, Valencia, 4th April 2017

Introduction

- a system for German and Russian trained on coreferential links projected from English

Introduction

- a system for German and Russian trained on coreferential links projected from English
  - uses automatic English coreference annotation
  - uses manual translations

Introduction

- a system for German and Russian trained on coreferential links projected from English
  - uses automatic English coreference annotation
  - uses manual translations
  - operates on the level of deep syntax

Introduction

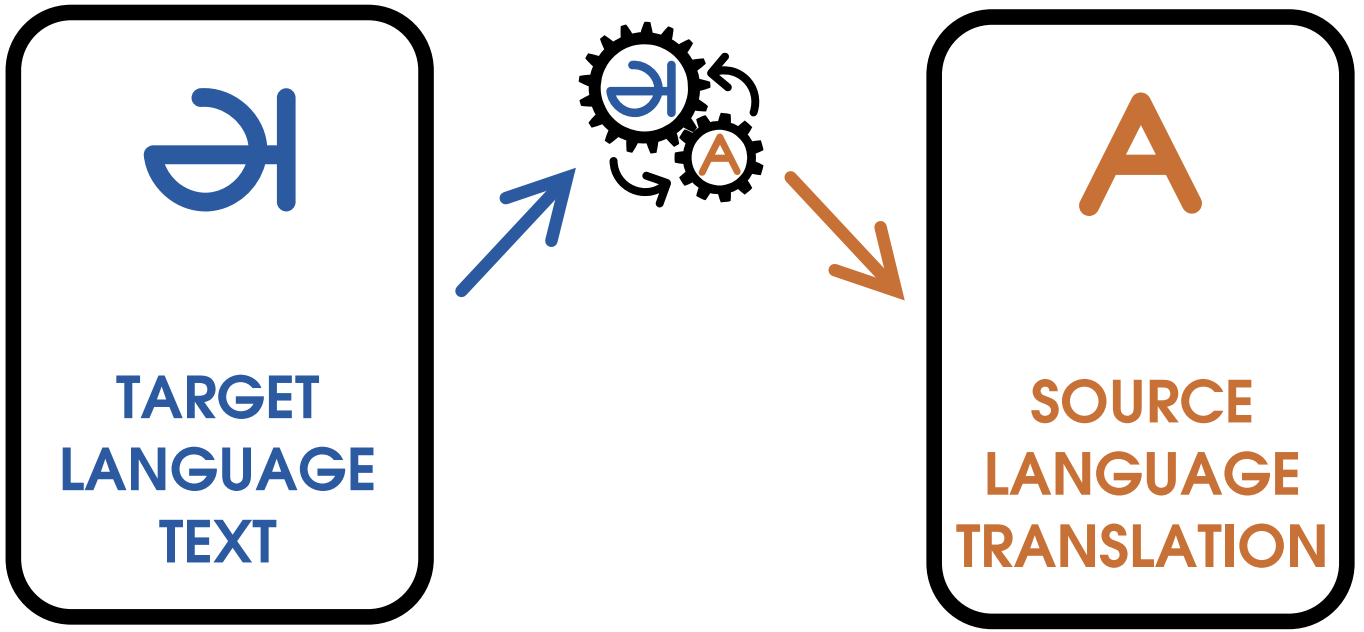
- a system for German and Russian trained on coreferential links projected from English
  - uses automatic English coreference annotation
  - uses manual translations
  - operates on the level of deep syntax
- submitted to the CORBON 2017 Shared Task
  - we won!!!

Introduction

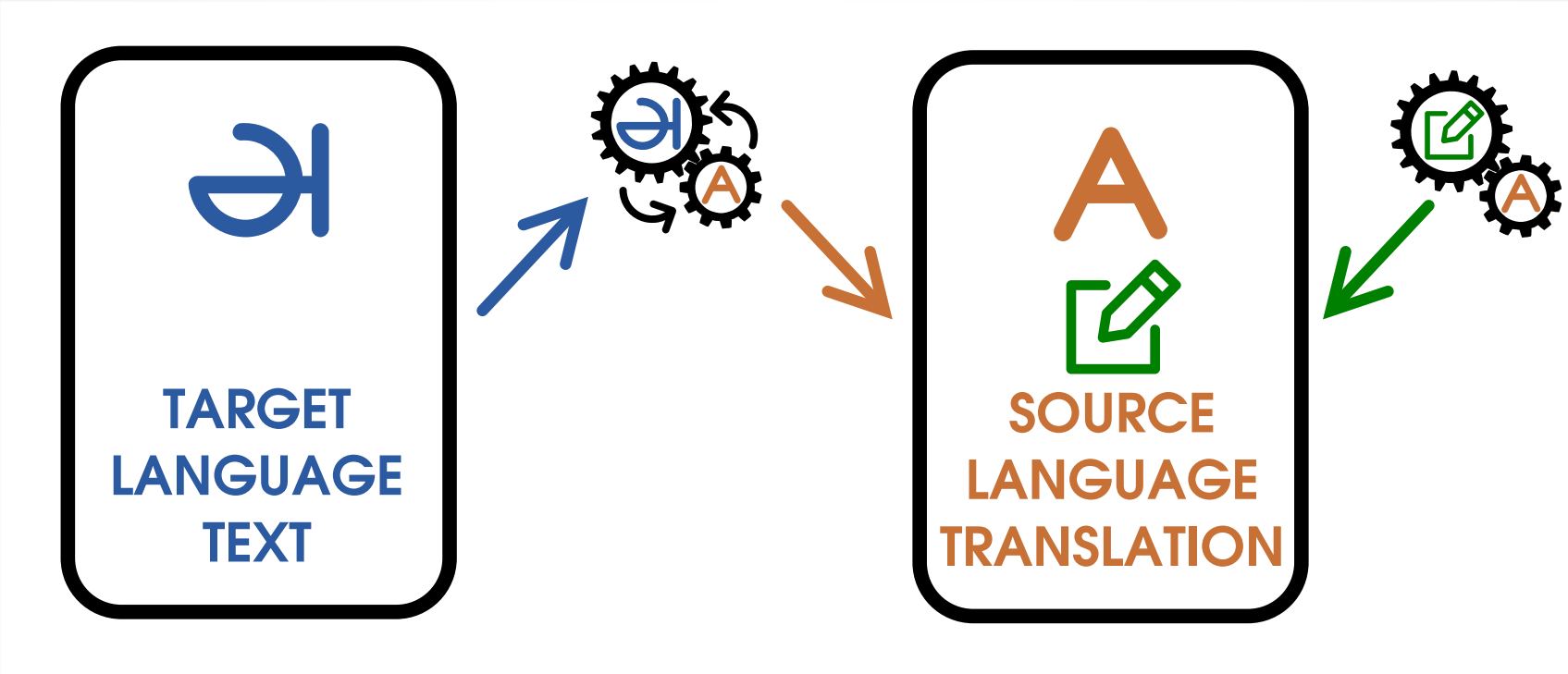
- a system for German and Russian trained on coreferential links projected from English
  - uses automatic English coreference annotation
  - uses manual translations
  - operates on the level of deep syntax
- submitted to the CORBON 2017 Shared Task
  - we won!!!
  - sadly, we were the only participant

Projection-based CR, Type I

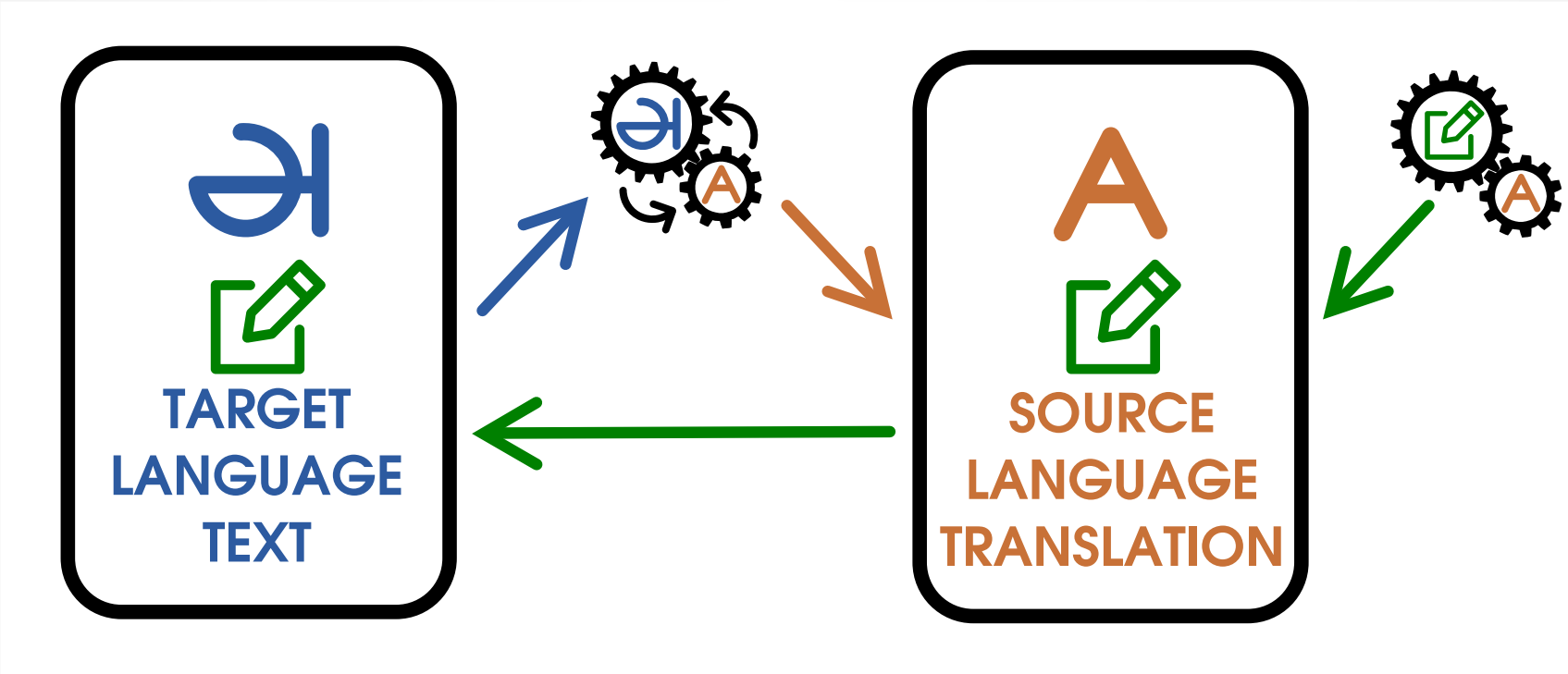
Projection-based CR, Type I



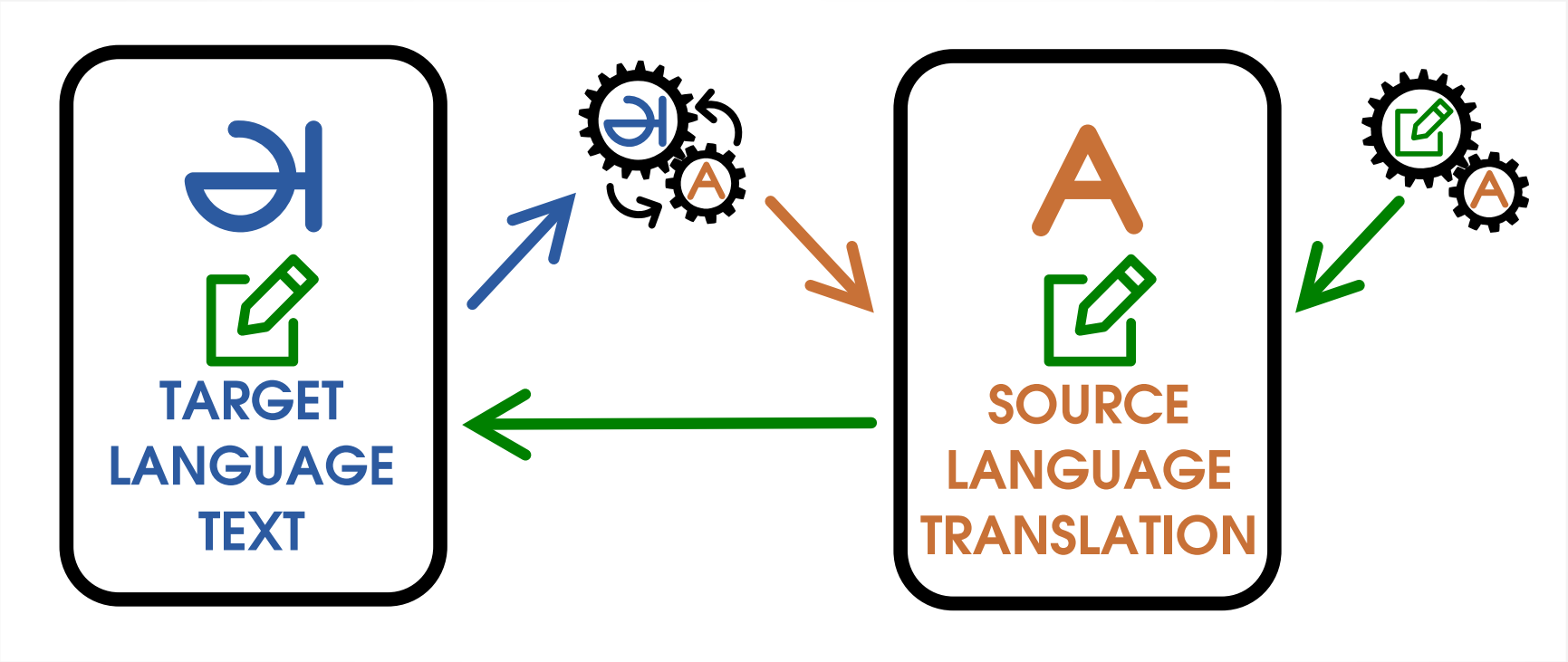
Projection-based CR, Type I



Projection-based CR, Type I

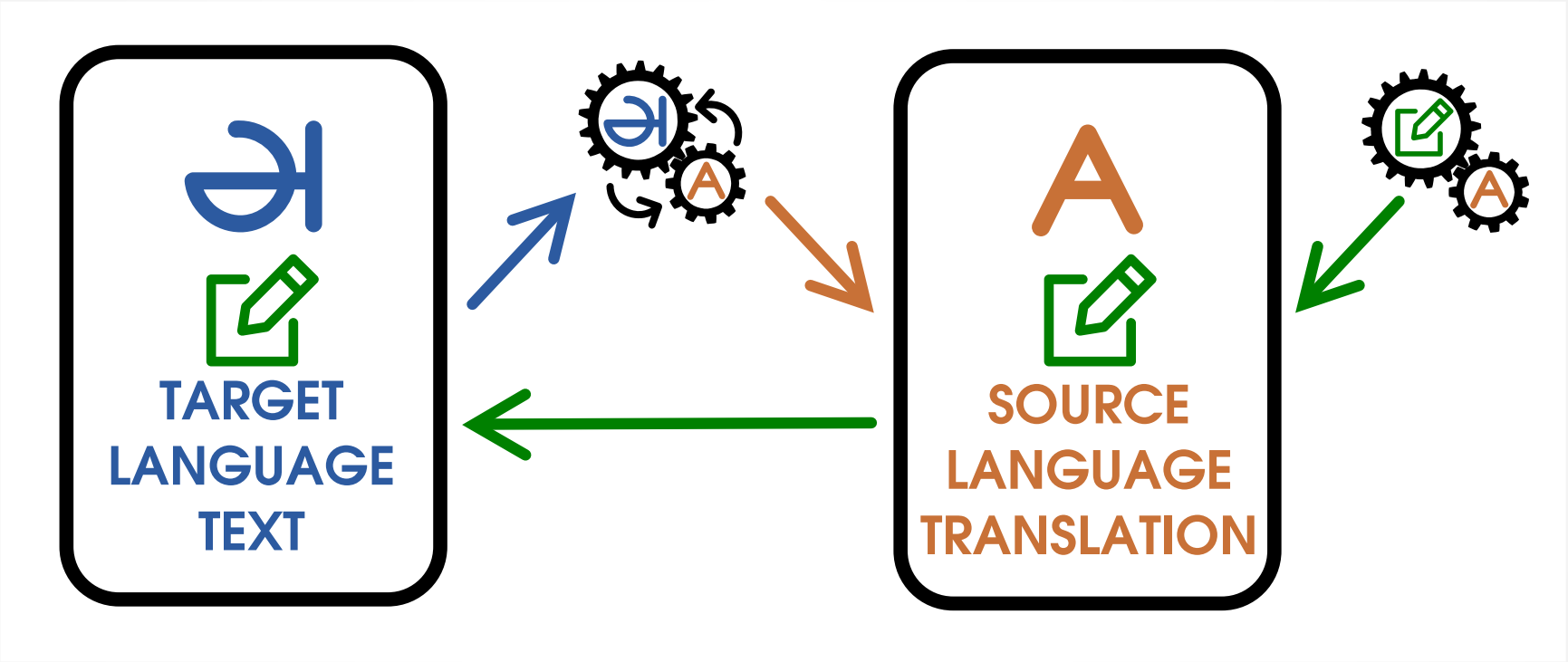


Projection-based CR, Type I



- no parallel data needed
- requires an MT service
- projection performed in test time

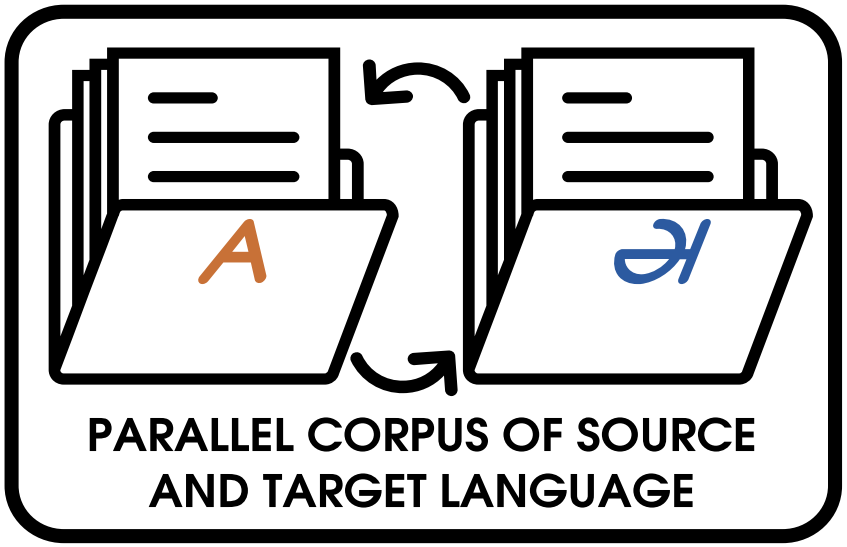
Projection-based CR, Type I



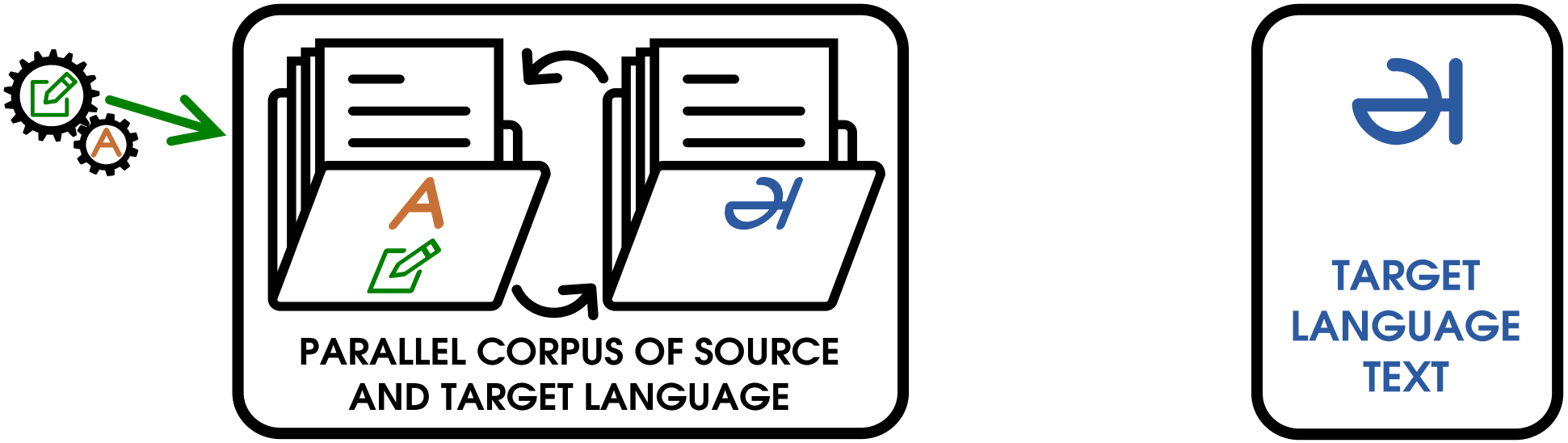
- no parallel data needed
- requires an MT service
- projection performed in test time

→ *Rahman and Ng (2012)*
→ *Ogrodniczuk (2013)*

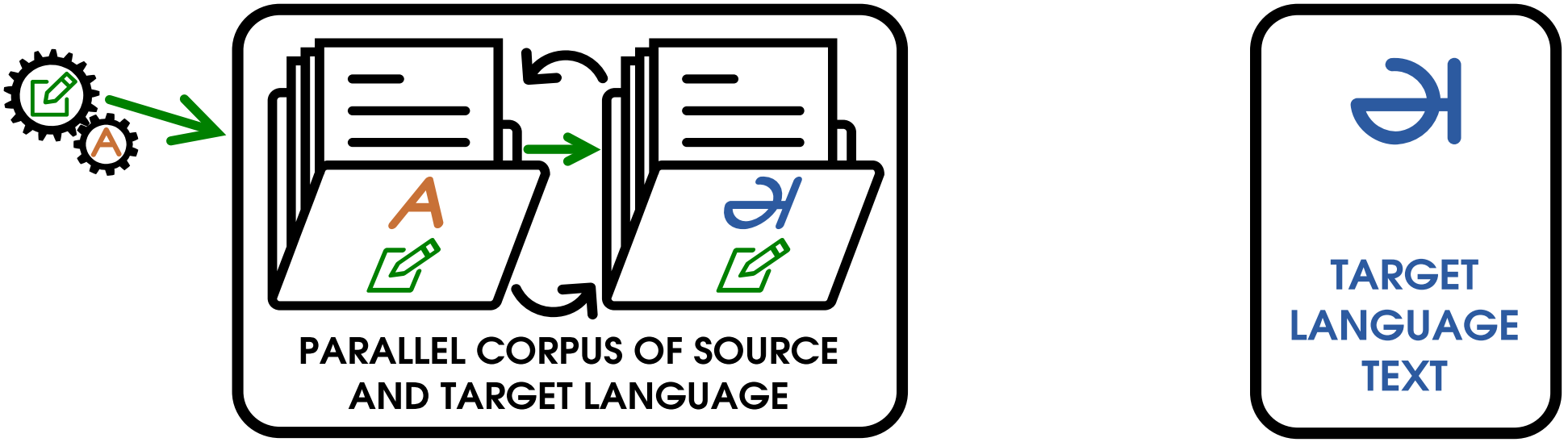
Projection-based CR, Type II



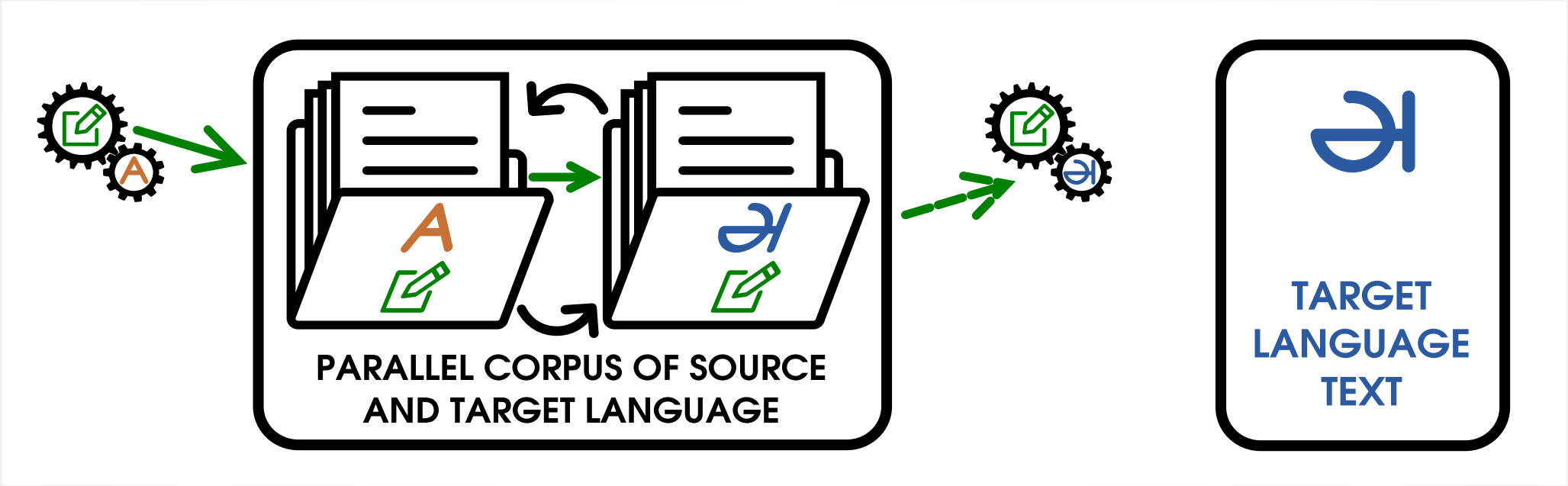
Projection-based CR, Type II



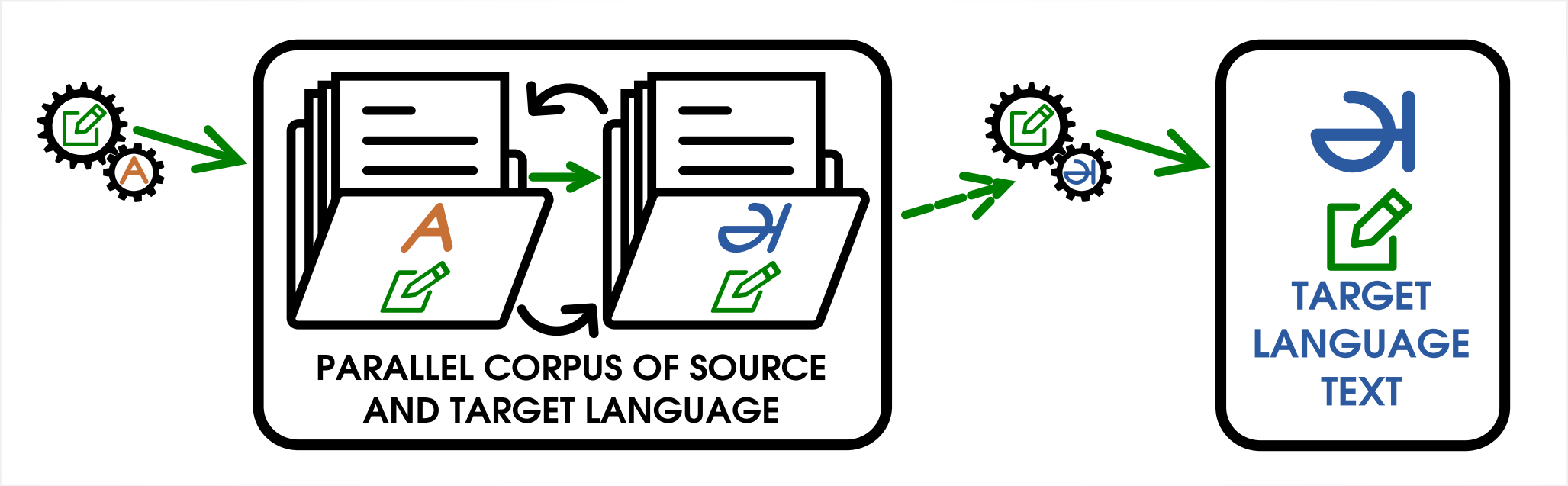
Projection-based CR, Type II



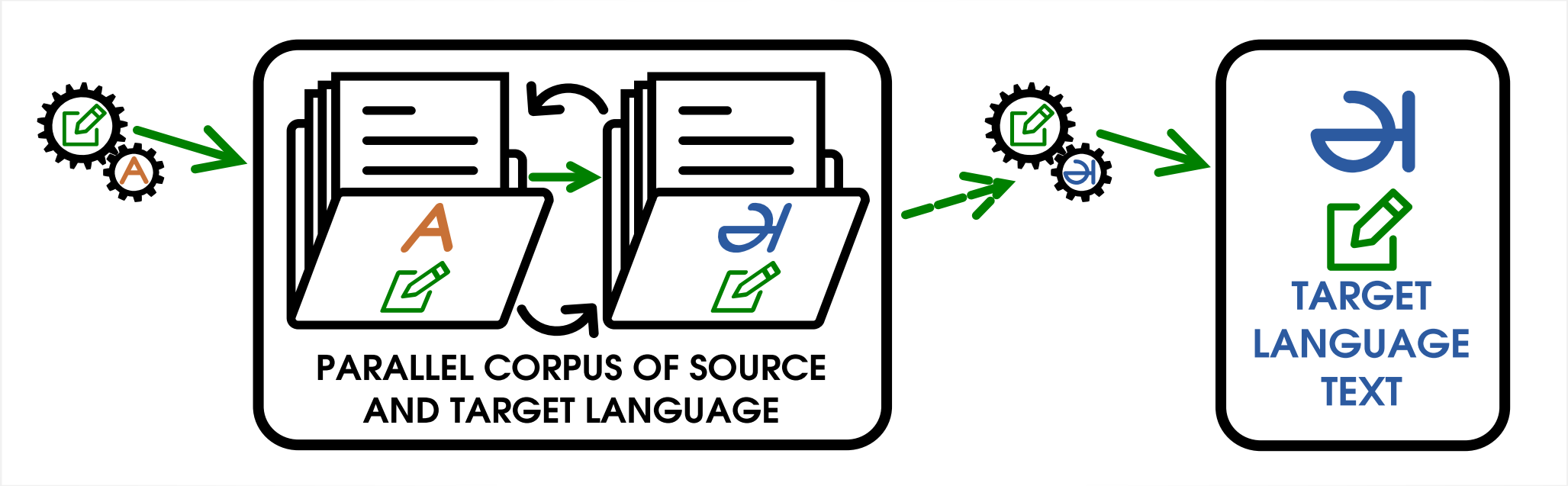
Projection-based CR, Type II



Projection-based CR, Type II

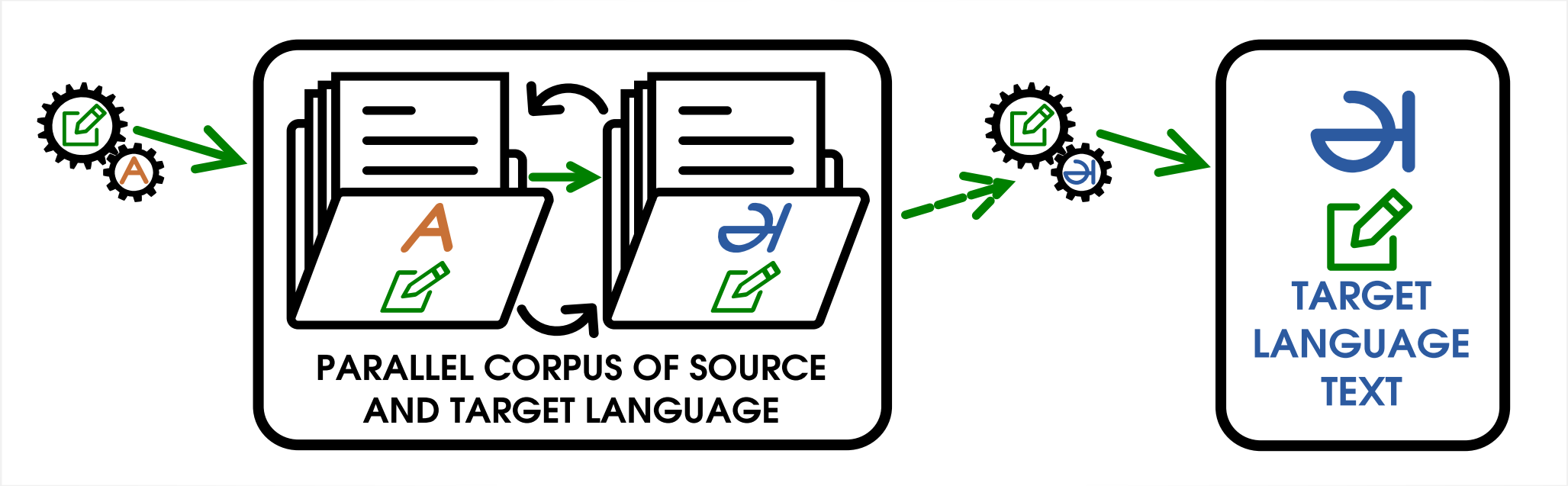


Projection-based CR, Type II



- no MT service needed
- requires a parallel corpus
- projection performed in train time

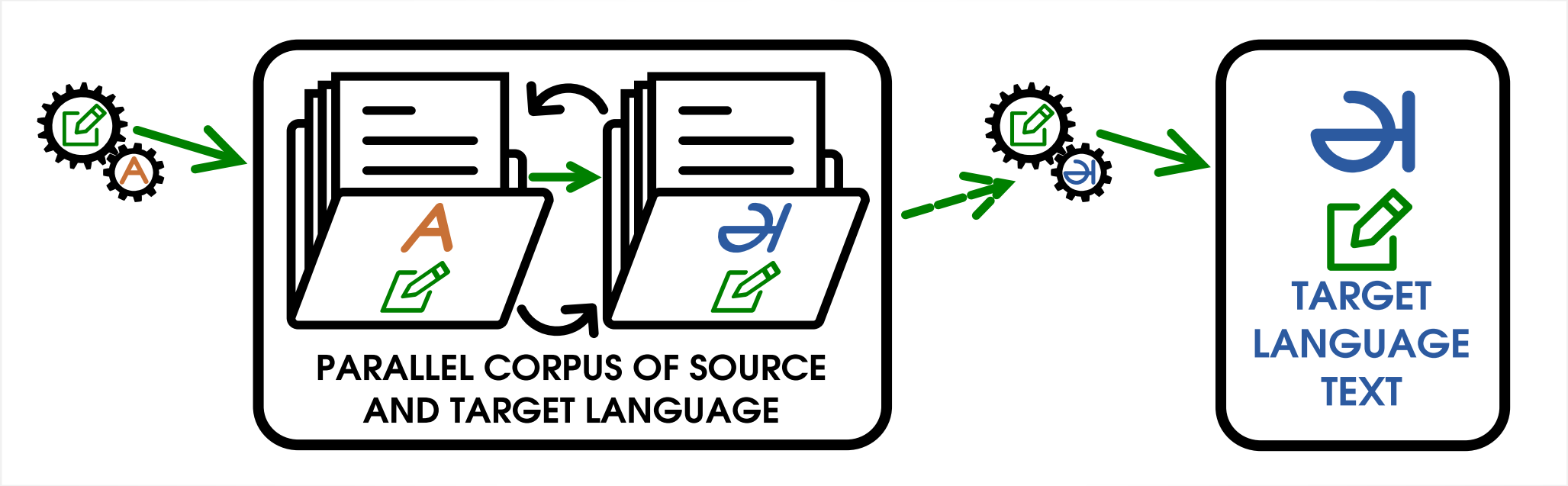
Projection-based CR, Type II



- no MT service needed
- requires a parallel corpus
- projection performed in train time

- *Postolache et al. (2006)*
- *Souza and Orasan (2013)*
- *Martins (2015)*

Projection-based CR, Type II



- no MT service needed
- requires a parallel corpus
- projection performed in train time

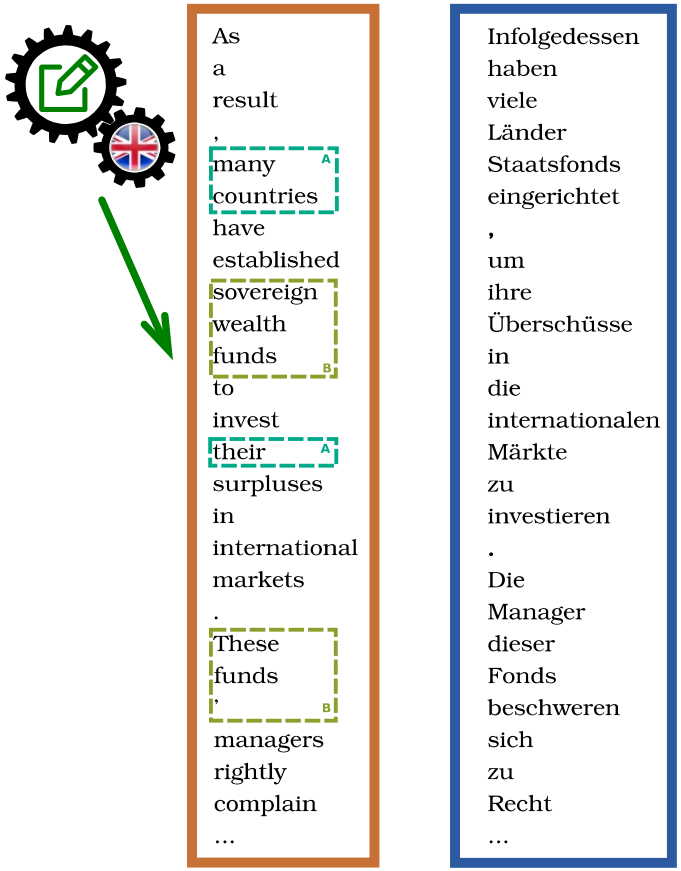
- *Postolache et al. (2006)*
- *Souza and Orasan (2013)*
- *Martins (2015)*
- ***this work***

Our projection-based CR Design

As
a
result
,
many
countries
have
established
sovereign
wealth
funds
to
invest
their
surpluses
in
international
markets
.
These
funds
'
managers
rightly
complain
...

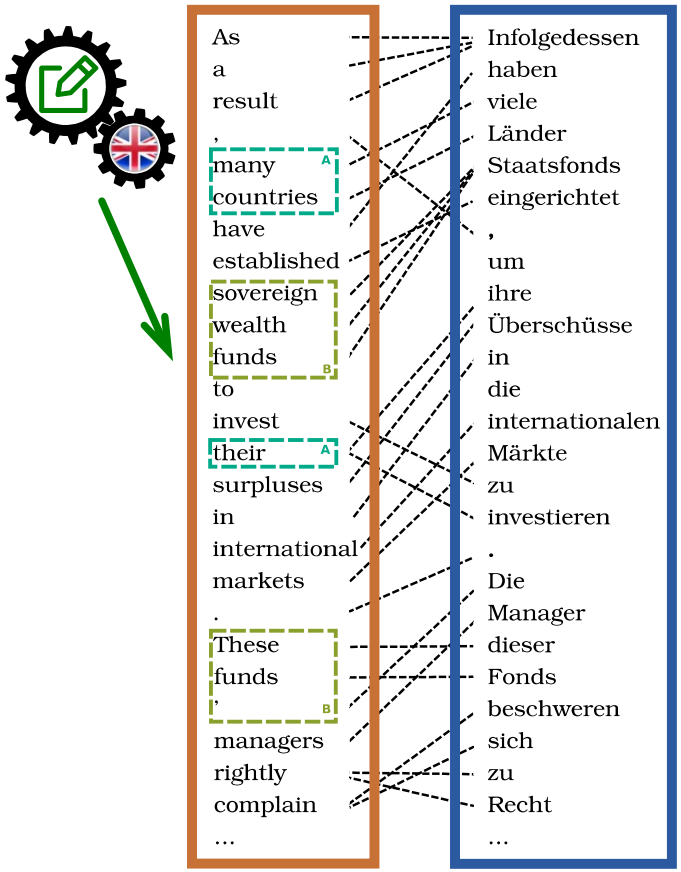
Infolgedessen
haben
viele
Länder
Staatsfonds
eingerichtet
,
um
ihre
Überschüsse
in
die
internationalen
Märkte
zu
investieren
.
Die
Manager
dieser
Fonds
beschweren
sich
zu
Recht
...

Our projection-based CR Design



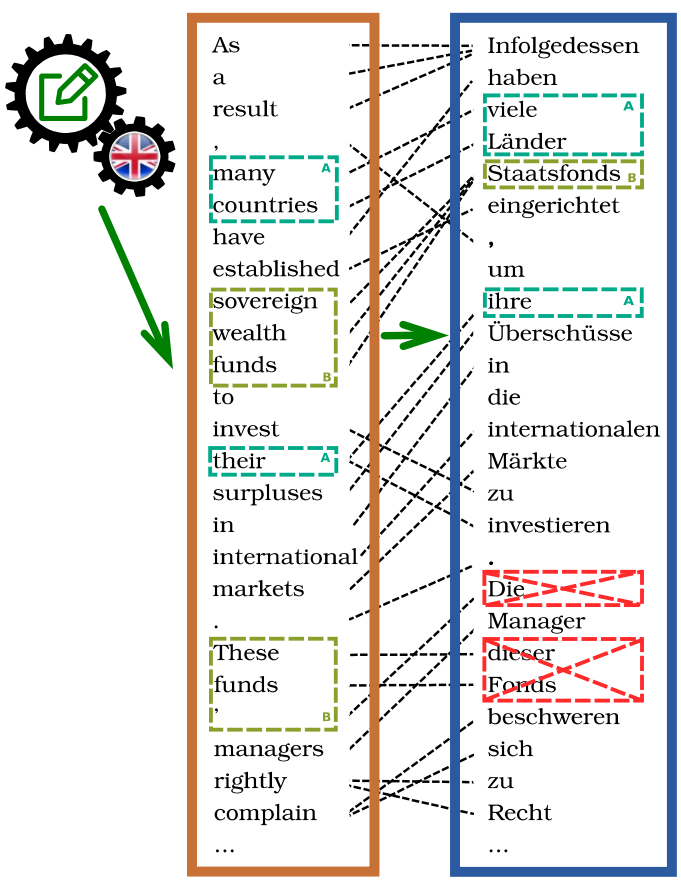
- automatic coreference resolution on the English side of a parallel corpus
  - Berkeley Entity Resolution system *(Durrett and Klein, 2014)*
  - rule-based resolution of relative pronouns by Treex framework

Our projection-based CR Design



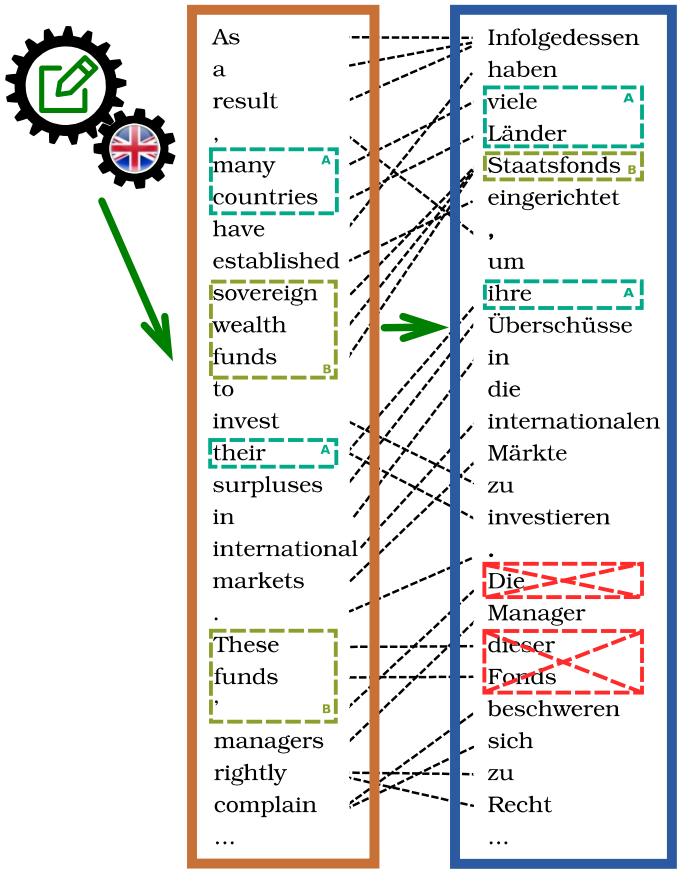
- word alignment
  - GIZA++ *(Och and Ney, 2000)*
  - run on full parallel training data with no additional data
  - both English and German/Russian texts had been lemmatized first

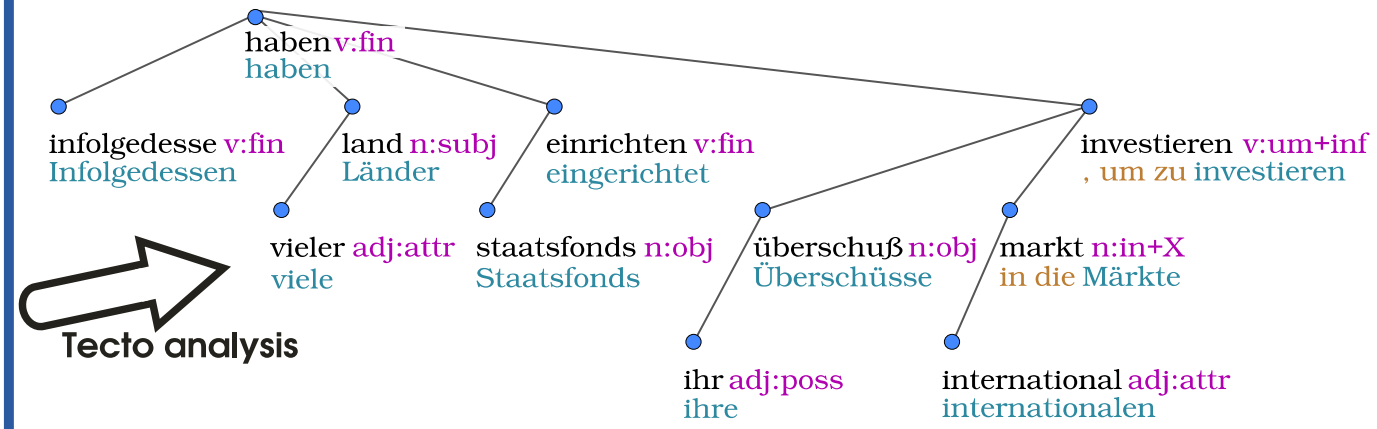
Our projection-based CR Design



- projection of mentions and entity ids
  - mention's counterpart must form a consecutive sequence of tokens

Our projection-based CR Design

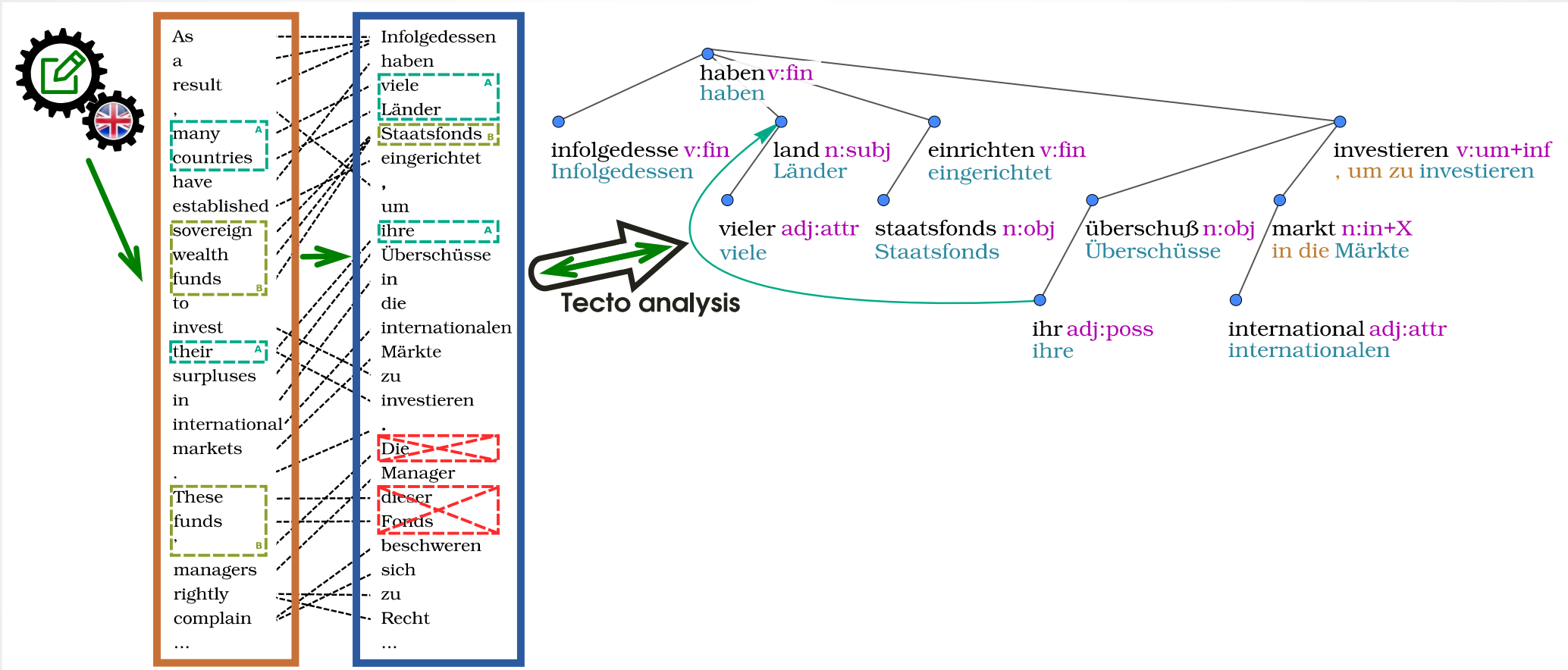




morphology and syntax	MATE tools *(Bjorkelund et al., 2010)*	TreeTagger *(Schmid, 1995)*
	HamleDT transformation *(Zeman et al., 2014)*	UDPipe *(Straka et al., 2016)*
NEs	NameTag *(Strakova et al., 2014)*	
tecto	Treex framework *(Popel and Zabokrtsky, 2010)*	

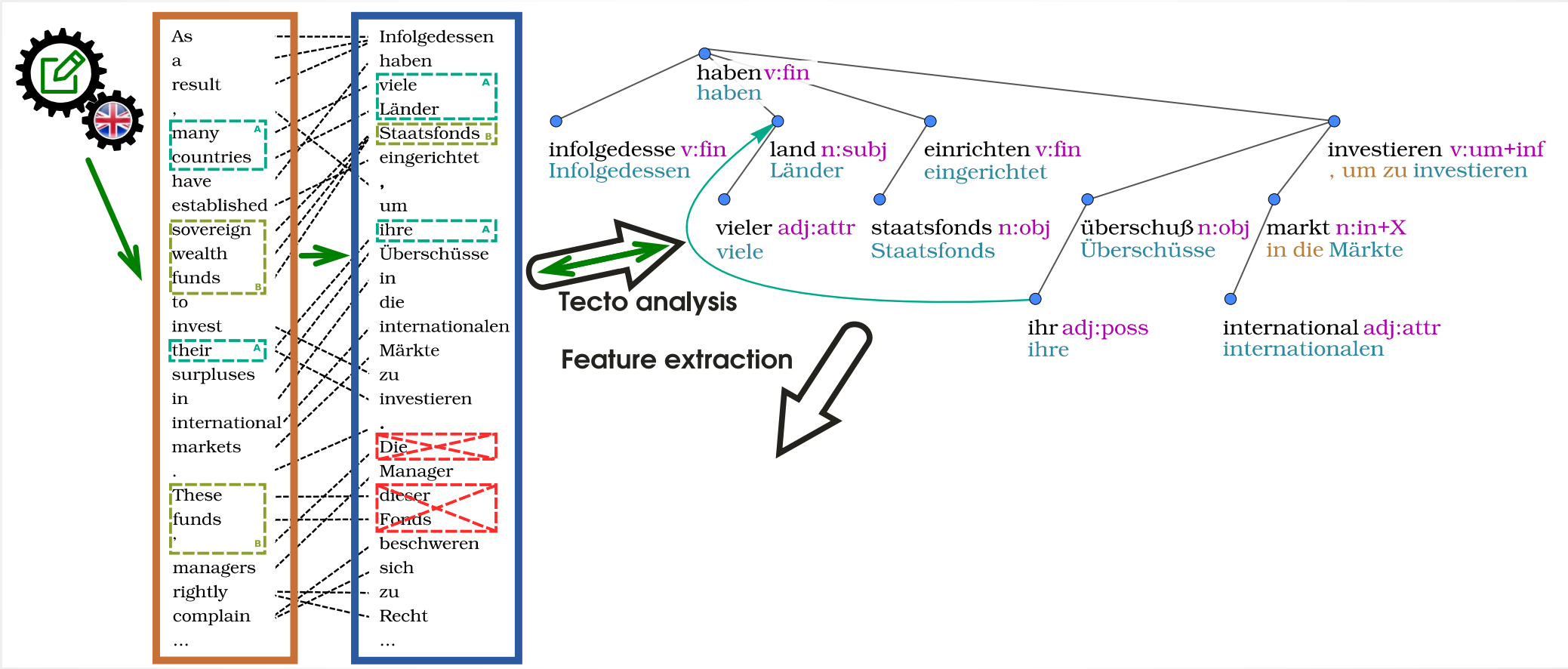
- analysis to the tectogrammatical layer
  - deep syntax dependency tree consisting only of content words

Our projection-based CR Design



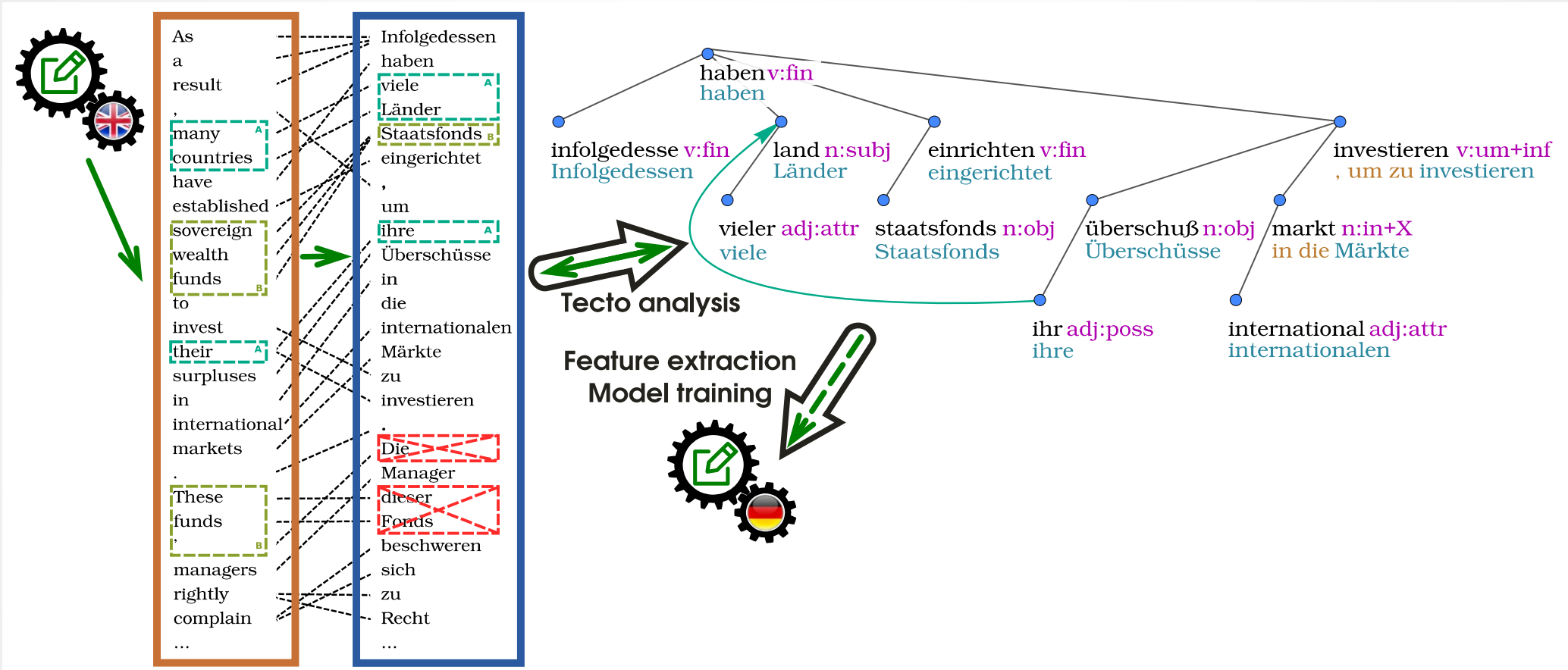
- transfer of mentions from surface and back
  - → use dependency tree to find the head
  - ← expand over the whole subtree (with some exceptions)
    - needed for surface-oriented evaluation

Our projection-based CR Design



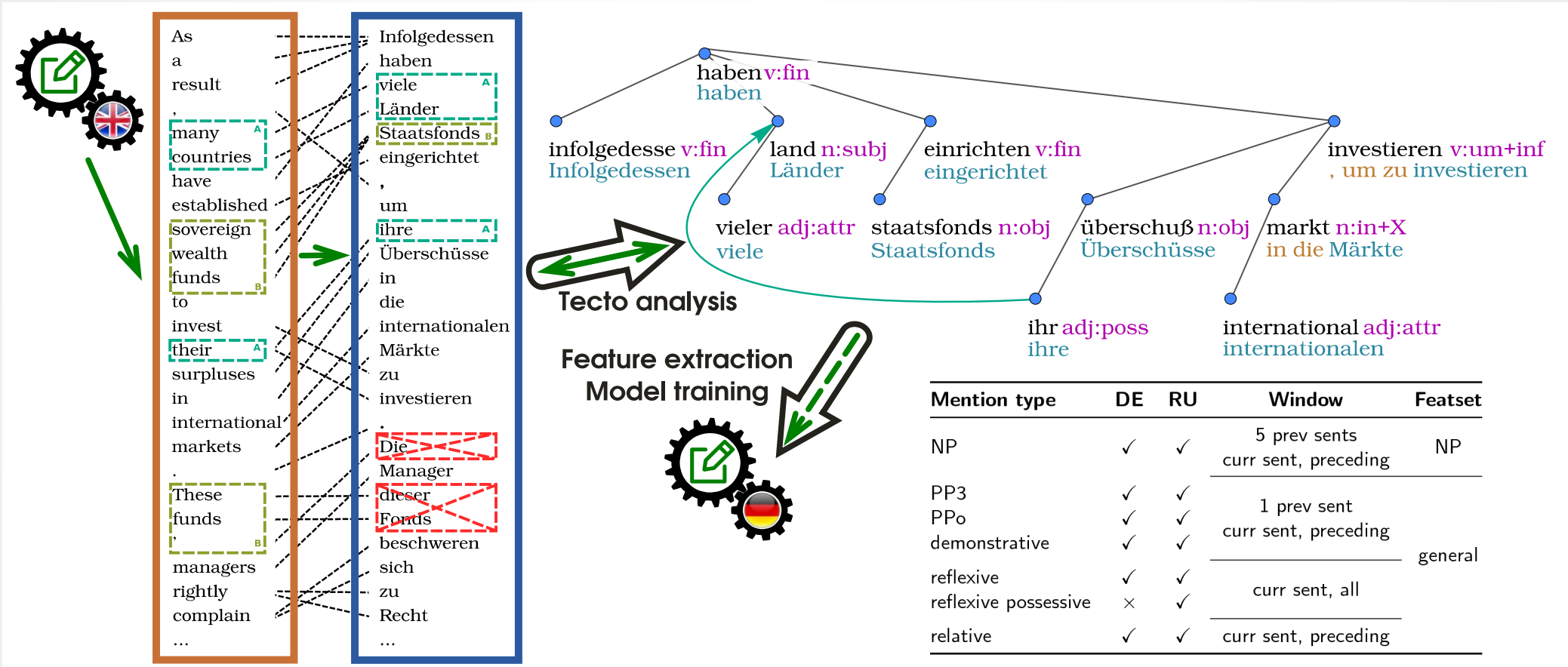
- feature sets
  - *General*: gender and number agree, other morpho, distance, named entities, depend. relations, syntactic patterns in trees
  - *NP*: General + head lemma match, Levehnstein distance, full match; similarity based on word2vec *(Mikolov et al., 2013)*

Our projection-based CR Design



- modeling
  - mention-ranking model *(Denis and Baldridge, 2007)*
  - joint anaphoricity detection and antecedent selection

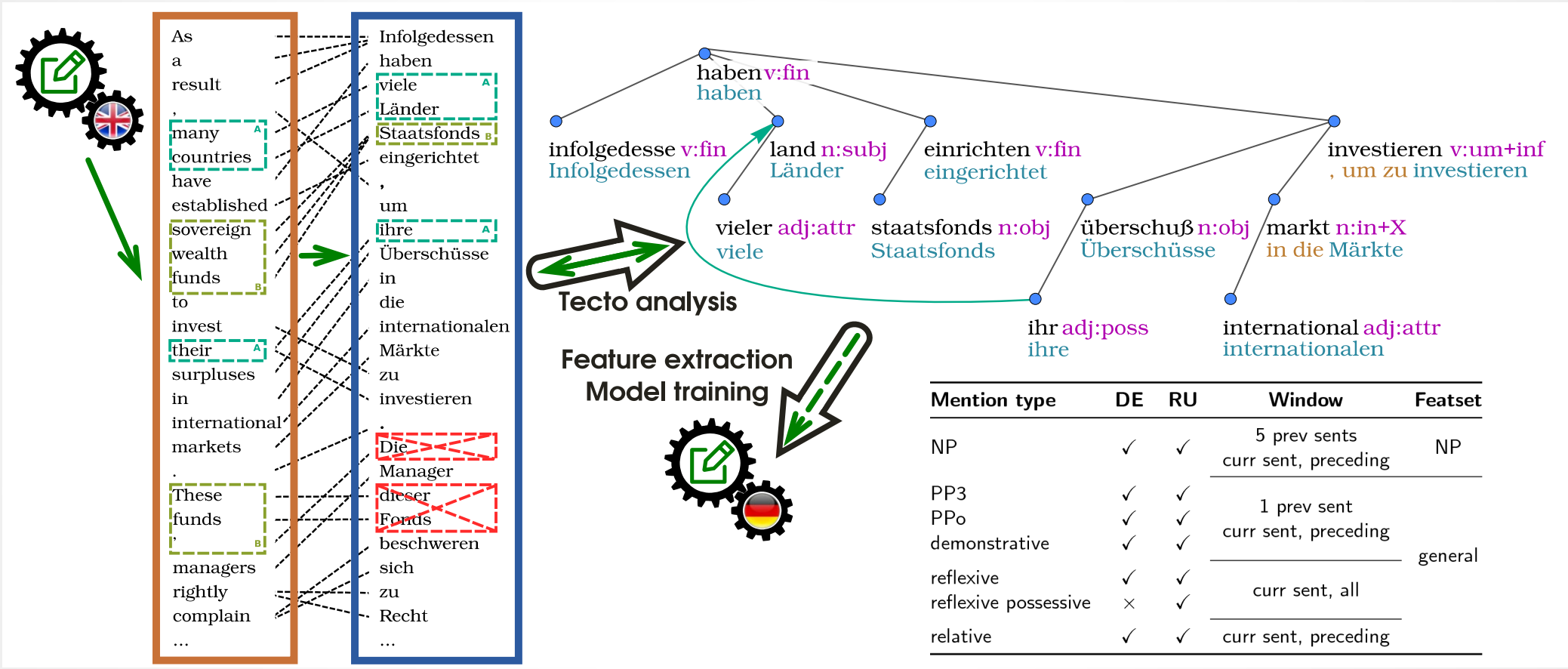
Our projection-based CR Design



Mention type	DE	RU	Window	Featset
NP	✓	✓	5 prev sents curr sent, preceding	NP
PP3	✓	✓	1 prev sent curr sent, preceding	general
PPo	✓	✓		
demonstrative	✓	✓		
reflexive	✓	✓	curr sent, all	
reflexive possessive	×	✓		
relative	✓	✓	curr sent, preceding	

- modeling
  - mention-ranking model *(Denis and Baldridge, 2007)*
  - joint anaphoricity detection and antecedent selection
  - separate models for different anaphor types

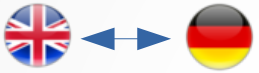
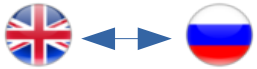
Our projection-based CR Design



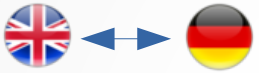
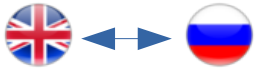
Mention type	DE	RU	Window	Featset
NP	✓	✓	5 prev sents curr sent, preceding	NP
PP3	✓	✓	1 prev sent curr sent, preceding	general
PPo	✓	✓		
demonstrative	✓	✓		
reflexive	✓	✓	curr sent, all	
reflexive possessive	×	✓		
relative	✓	✓	curr sent, preceding	

- modeling
  - mention-ranking model *(Denis and Baldridge, 2007)*
  - joint anaphoricity detection and antecedent selection
  - separate models for different anaphor types
  - logistic regression optimized with SGD in Vowpal Wabbit

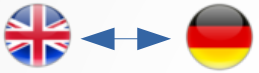
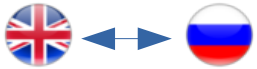
Evaluation

- Data: News-Commentary11
  - Train:  192k  155k sent. pairs
  - DevTest:  207 sent. triples
    - set aside from the Train set
    - labeled by 3 annotators using the official annotation guidelines
  - EvalTest:  404 412 sents.

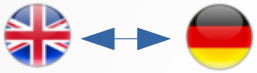
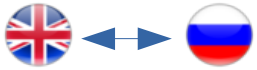
Evaluation

- Data: News-Commentary11
  - Train:  192k  155k sent. pairs
  - DevTest:  207 sent. triples
    - set aside from the Train set
    - labeled by 3 annotators using the official annotation guidelines
  - EvalTest:  404 412 sents.
- No other submissions nor baseline was available

Evaluation

- Data: News-Commentary11
  - Train:  192k  155k sent. pairs
  - DevTest:  207 sent. triples
    - set aside from the Train set
    - labeled by 3 annotators using the official annotation guidelines
  - EvalTest:  404 412 sents.
- No other submissions nor baseline was available
- Measures:
  - CoNLL score: combination of MUC, B^3 and CEAF-e F-scores

Evaluation

- Data: News-Commentary11
  - Train:  192k  155k sent. pairs
  - DevTest:  207 sent. triples
    - set aside from the Train set
    - labeled by 3 annotators using the official annotation guidelines
  - EvalTest:  404 412 sents.
- No other submissions nor baseline was available
- Measures:
  - CoNLL score: an unweighted average of MUC, B^3 and CEAF-e F-scores

	MUC	B^3	CEAF-e	**CoNLL**	MUC	B^3	CEAF-e	**CoNLL**
DevTest	24.9	18.6	23.7	**22.4**	33.7	27.5	34.2	**31.8**
EvalTest	-	-	-	**29.4**	-	-	-	**30.9**

Model ablation analysis

- measured on DevTest

Mention type		
all	**22.4**	**31.8**
- noun phrases	-4.6	-3.0
- personal, possessive 3rd person	-11.3	-10.4
- personal, possessive, other persons (PPo)	-1.0	-1.1
- demonstrative	-0.1	0
- reflexive	0	0
- reflexive possessive	-	-6.4
- relative	-1.9	-3.4

Model ablation analysis

- measured on DevTest
- models for personal and possessive pronouns in 3rd person are the most valuable

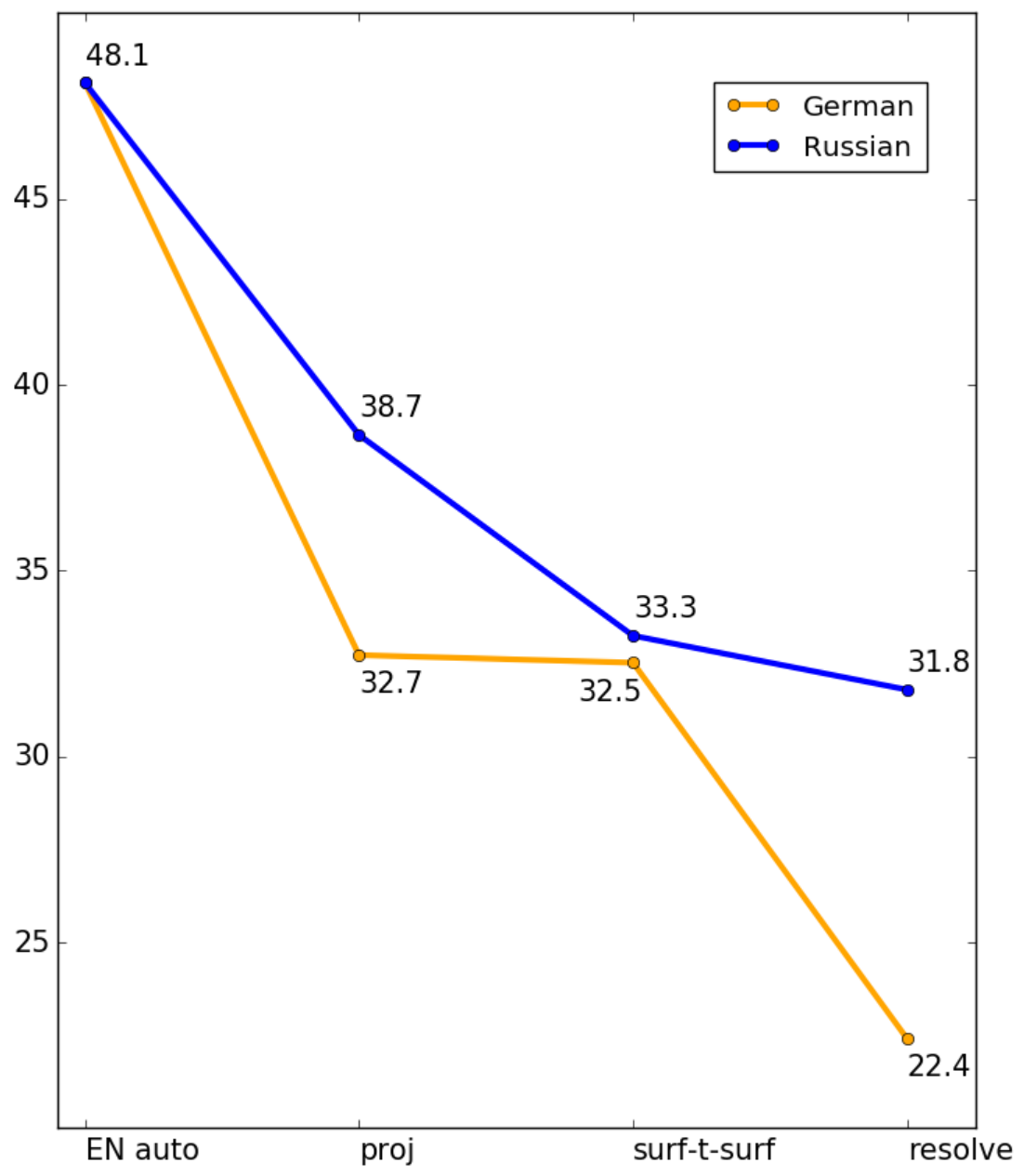
Mention type		
all	**22.4**	**31.8**
- noun phrases	-4.6	-3.0
- personal, possessive 3rd person	**-11.3**	**-10.4**
- personal, possessive, other persons (PPo)	-1.0	-1.1
- demonstrative	-0.1	0
- reflexive	0	0
- reflexive possessive	-	-6.4
- relative	-1.9	-3.4

Model ablation analysis

- measured on DevTest
- models for personal and possessive pronouns in 3rd person are the most valuable
- low impact of some models results from low frequency of anaphoric occurrences in the test (personal and possessive in other persons) or the train set (demonstratives, reflexives)

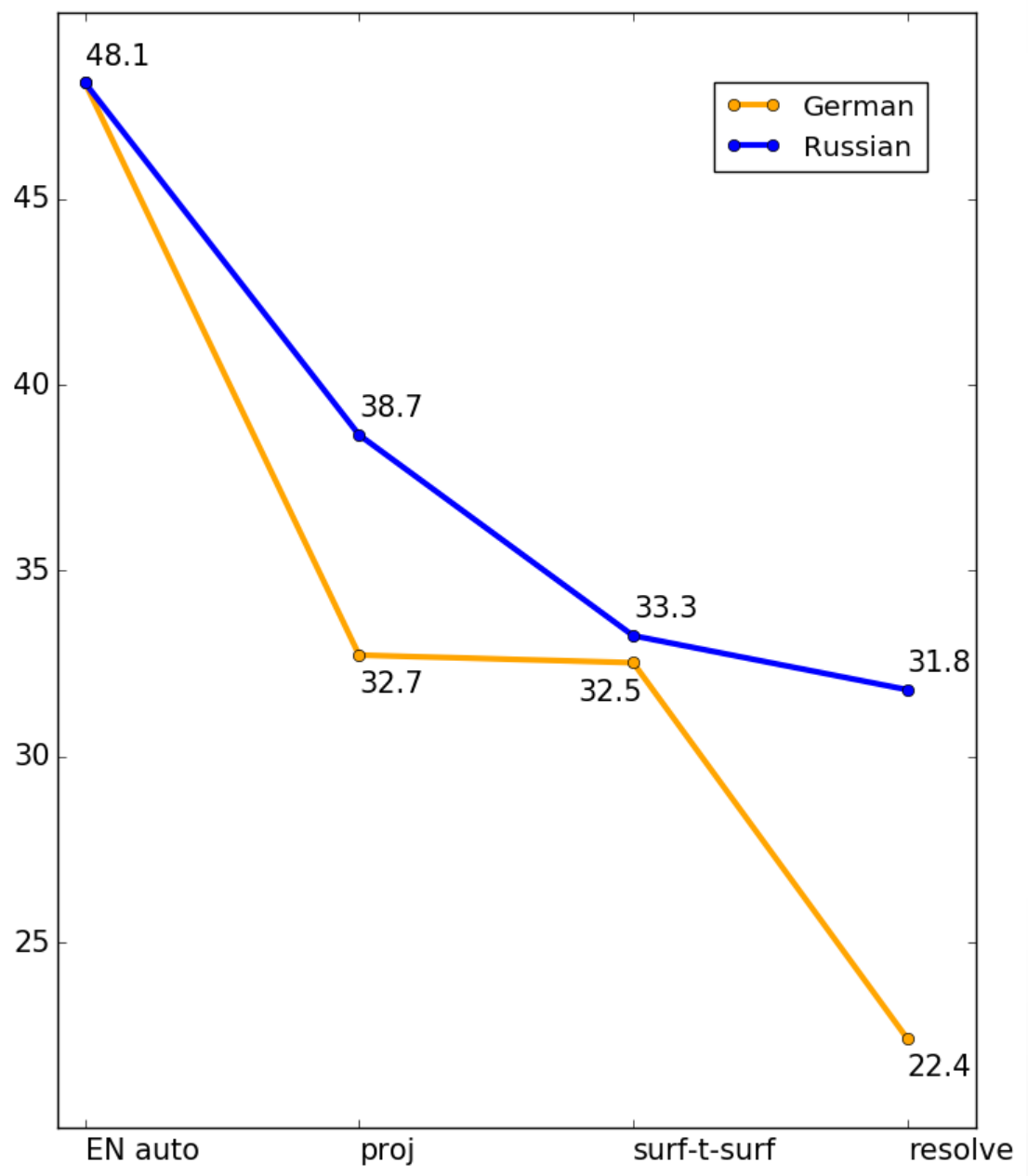
Mention type		
all	**22.4**	**31.8**
- noun phrases	-4.6	-3.0
- personal, possessive 3rd person	**-11.3**	**-10.4**
- personal, possessive, other persons (PPo)	**-1.0**	**-1.1**
- demonstrative	**-0.1**	**0**
- reflexive	**0**	**0**
- reflexive possessive	-	-6.4
- relative	-1.9	-3.4

Analysis over stages



- stages:
  - *EN auto* – English CR
  - *proj* – English mentions projected
  - *surf-t-surf* – projected mentions transferred to the t-layer and back to the surface
  - *resolve* – performance of the model trained on projected links

Analysis over stages



- stages:
  - *EN auto* – English CR
  - *proj* – English mentions projected
  - *surf-t-surf* – projected mentions transferred to the t-layer and back to the surface
  - *resolve* – performance of the model trained on projected links
- observations:
  - cross-lingual projection is the bottleneck
  - issues with coreference representation transfer most likely relate to the performance of the Russian parser
  - lower descriptive power of the German featset

Conclusion

- a projection-based coreference resolution system for German and Russian
- submitted to the CORBON 17 Shared Task

Conclusion

- a projection-based coreference resolution system for German and Russian
- submitted to the CORBON 17 Shared Task
- operates on the level of deep syntax

Conclusion

- projection-based coreference resolution system for German and Russian
- submitted to the CORBON 17 Shared Task
- operates on the level of deep syntax
- surprisingly, the English-to-Russian projection seems to be easier than the English-to-German one