

# A Neural Verb Lexicon Model with Source-Side Syntactic Context for String-to-Tree MT

Maria Nădejde<sup>1</sup>

Alexandra Birch<sup>1</sup> and Philipp Koehn<sup>2</sup>

<sup>1</sup>University of Edinburgh

<sup>2</sup>Johns Hopkins University

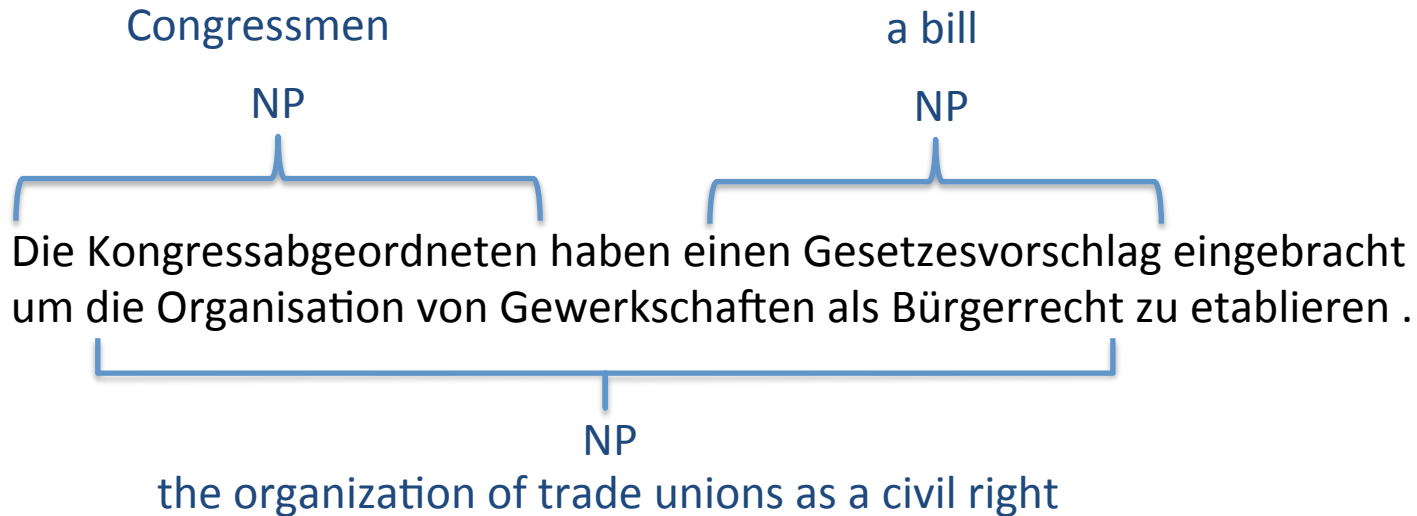
# Outline

- Verb translation with string-to-tree MT
- Related Work
- Neural Verb Lexicon Model
  - Source-side Syntactic Context
- Machine Translation Evaluation
  - Verb Translation Quality
- Examples

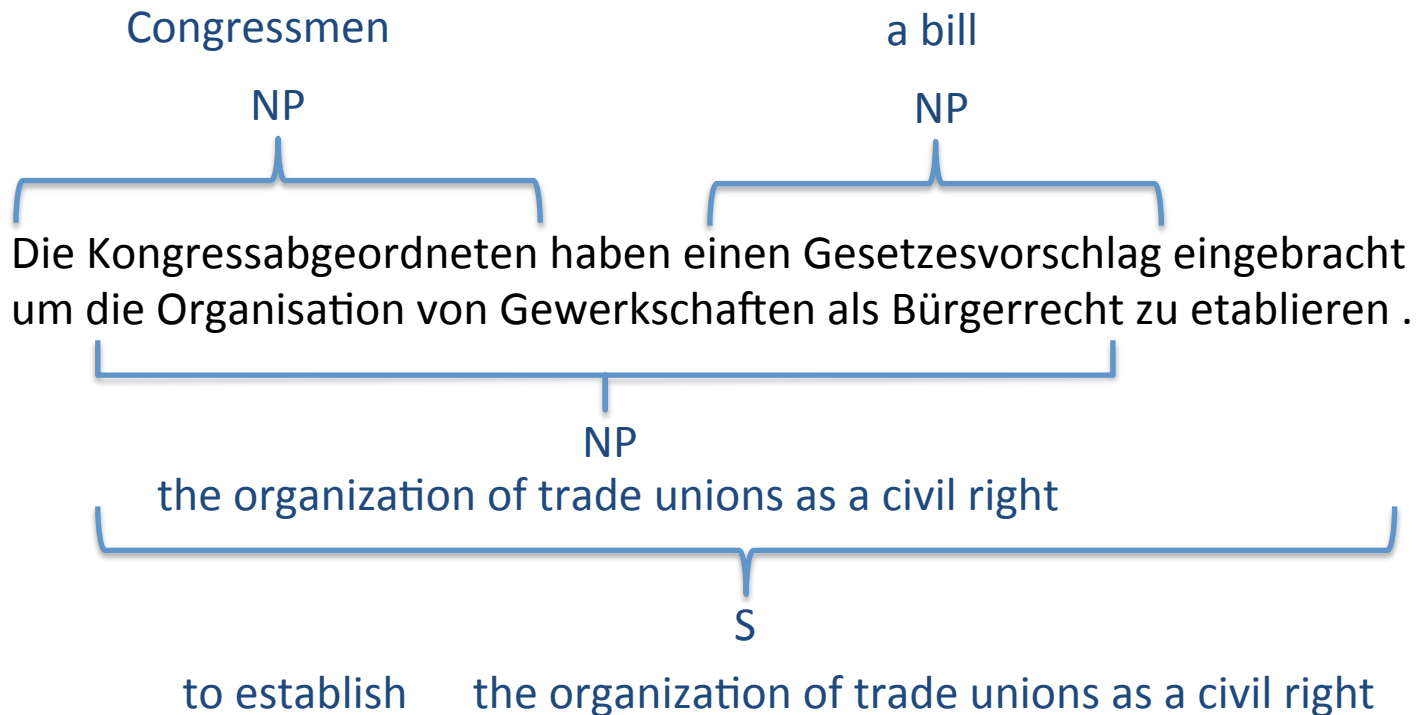
*Congressmen have proposed legislation to protect union organizing as a civil right .*

Die Kongressabgeordneten haben einen Gesetzesvorschlag eingebracht um die Organisation von Gewerkschaften als Bürgerrecht zu etablieren .

*Congressmen have proposed legislation to protect union organizing as a civil right .*

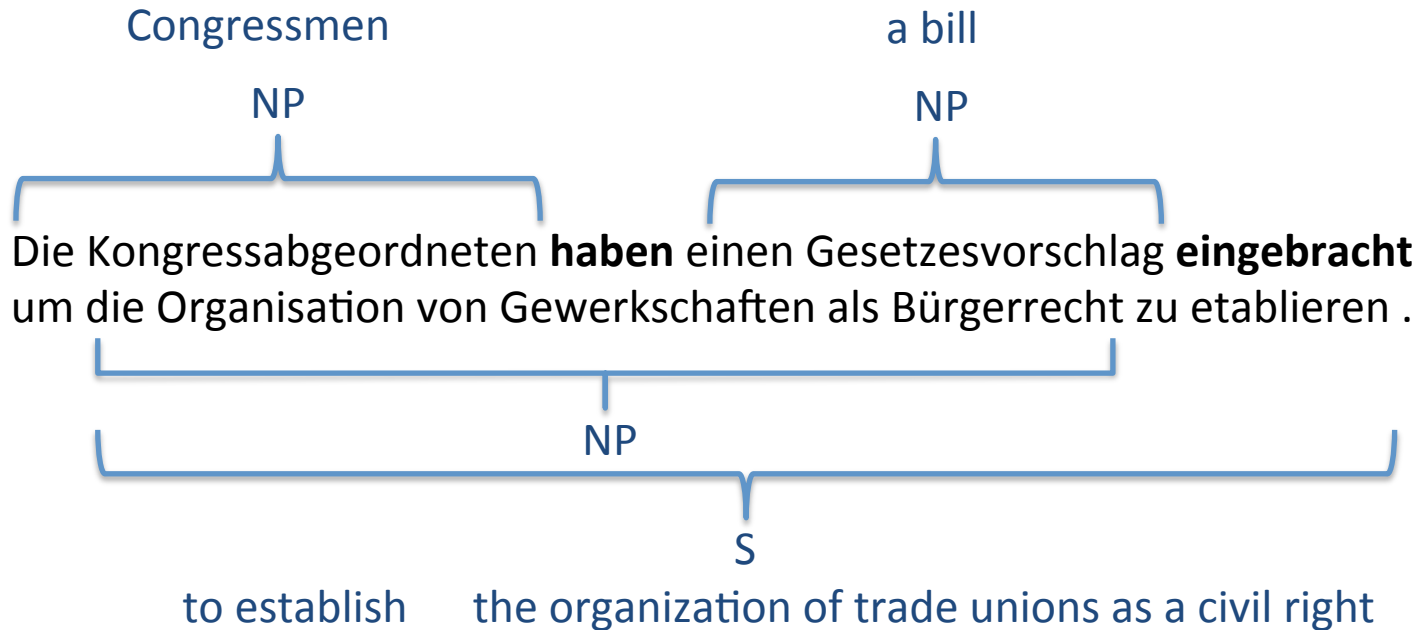


*Congressmen have proposed legislation to protect union organizing as a civil right .*

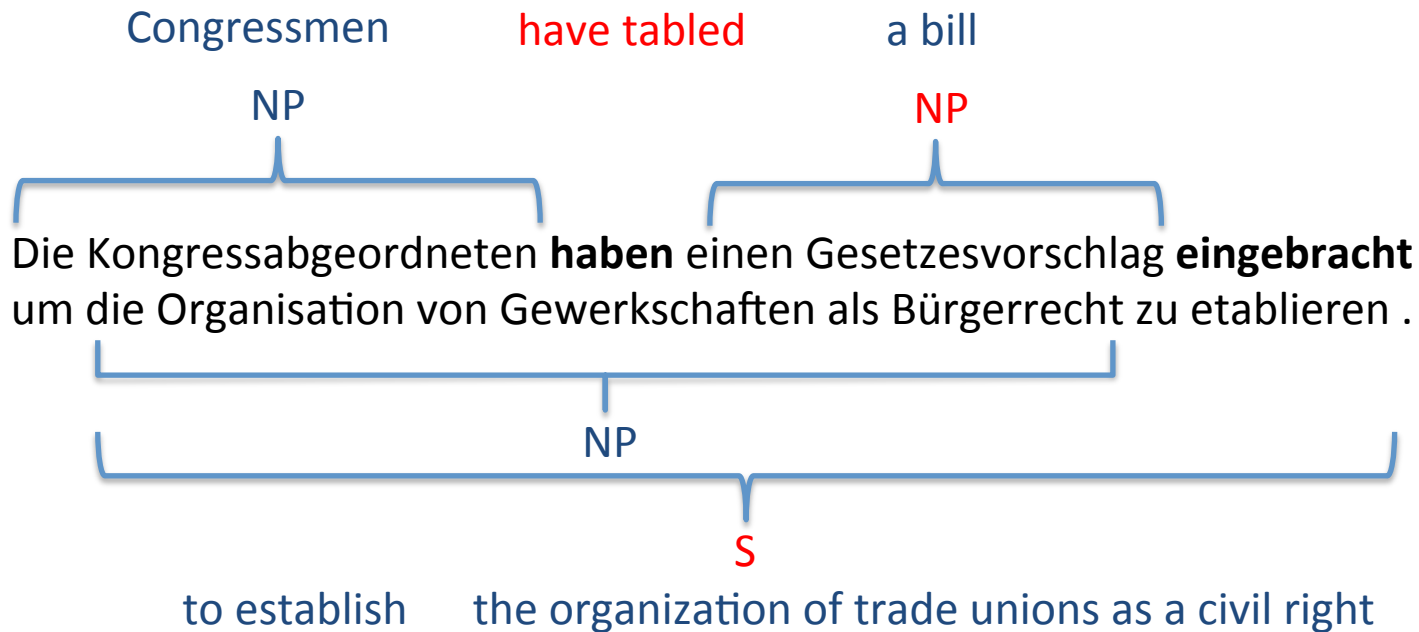


$S \rightarrow \langle \text{NP zu etablieren, to establish NP} \rangle$

*Congressmen have **proposed** legislation to protect union organizing as a civil right .*

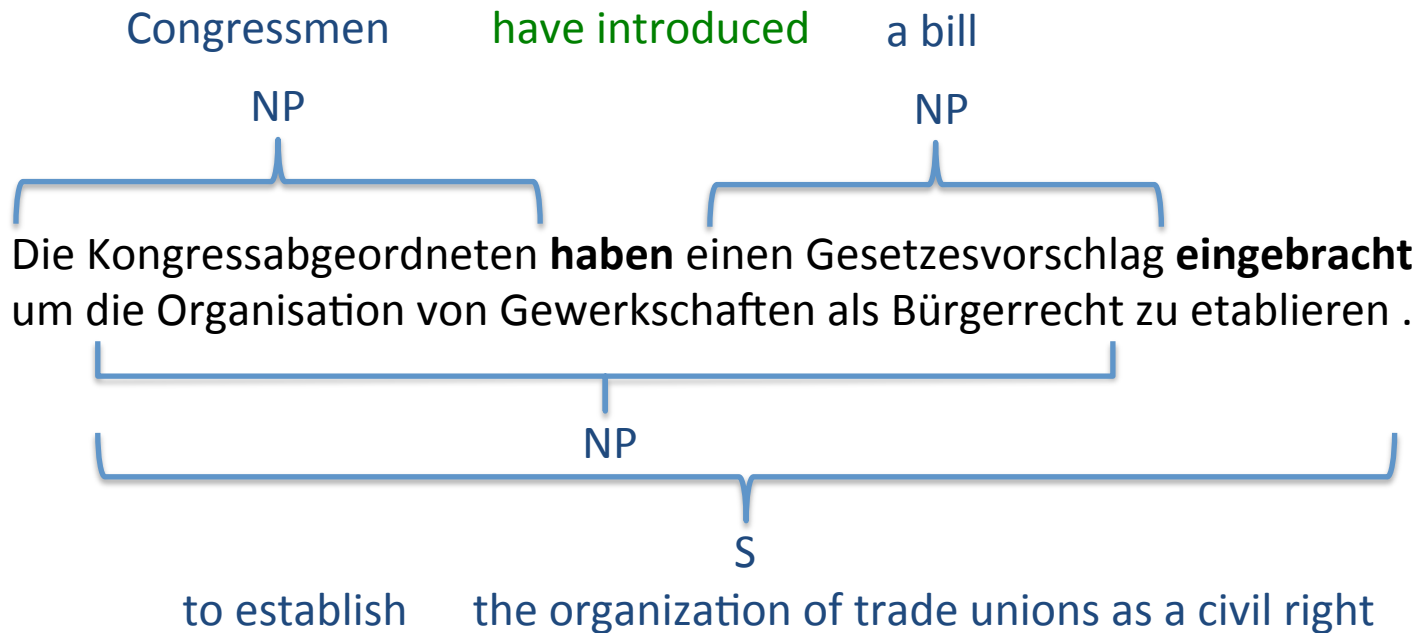


*Congressmen have **proposed** legislation to protect union organizing as a civil right .*



VP → < haben NP eingebracht um S, **have tabled NP S** >

*Congressmen have **proposed** legislation to protect union organizing as a civil right .*



VLM (**eingebracht**, Kongress. , Gesetzesvorschlag , etablieren ) → **introduced**



# Verb Translation Recall

- Low verb recall for string-to-tree system
  - Better verb translations available in n-best list

source	token	lemma
1-best	45.54	53.14
1000-best	72.87	79.24
Rule table	91.85	

Verb recall computed over 11,161 verbs from newstest2013-2015.  
WMT15 string-to-tree system for DE-EN

# Related Work

- Discriminative Word Lexicon (DWL)
  - Source-side features (Mauser et al., 2009)
    - Target-side features (Niehues et al., 2013, Tamchyna et al., 2016)
    - Structural features (Niehues et al., 2013, Hermann et al., 2015)
  - One classifier per target word/phrase

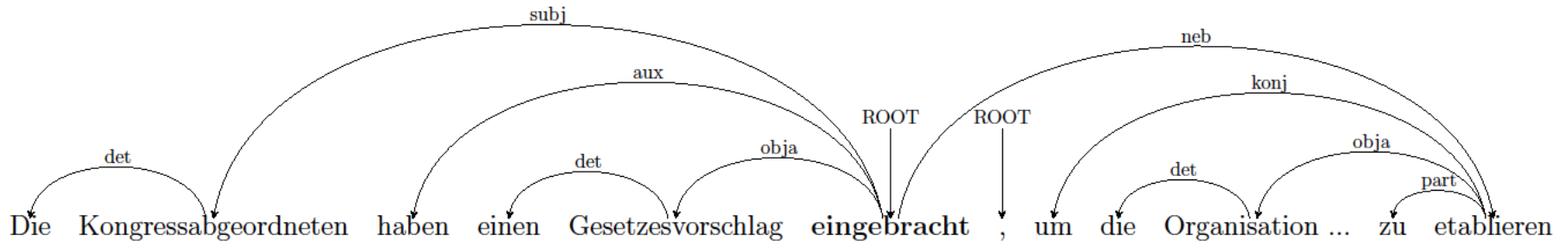
# Related Work

- **Global Neural Lexicon** (Ha et al, 2014)
  - Shared parameters across target words
    - Implicit feature combinations
  - Global context
  - Does not scale to large vocabulary

# This Work

- Neural Verb Lexicon
  - Shared parameters
  - Global context via dependency relations
  - Verb specific

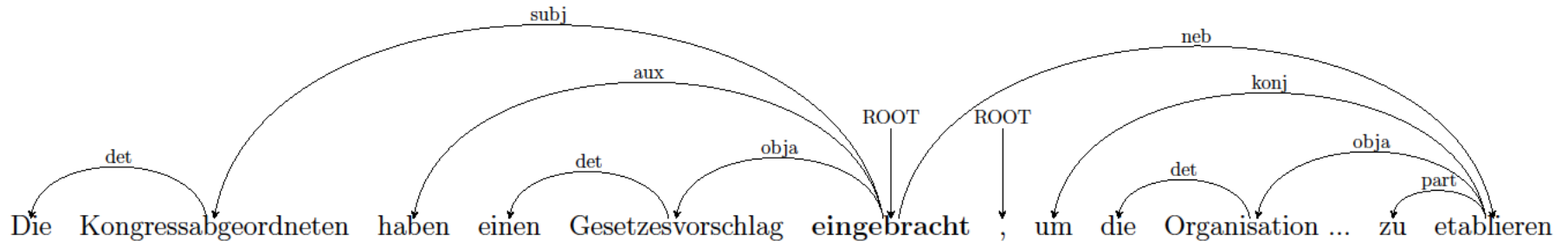
# Source-side Window Context



Ref: Congressmen have **proposed** legislation to protect the organization [...]

Window context: **eingebracht** einen Gesetzesvorschlag, um

# Source-side Syntactic Context



Ref: Congressmen have **proposed** legislation to protect the organization [...]

Window context: **eingebracht** einen Gesetzesvorschlag , um

Syntactic context:

	<i>source verb</i>	<i>parent</i>	<i>dependents</i>			<i>subcat</i>
word	<b>eingebracht</b>	haben	Kongress abgeordneten	Gesetzesvorschlag	etablieren	Subj_obja_neb
lemma	einbrechen	haben	Kongress abgeordnete	Gesetzesvorschlag	etablieren	

# Neural Verb Lexicon Model

- Feed forward neural network (NPLM Vaswani et al. 2013)
  - Predicts target verb
- Source-side context:
  - Window
  - Syntactic
    - Parent, 3 dependents, prep modifier, particle
- Factors
  - Word, lemma, sub-categorization token
  - Max input size: 16

# Experimental Setup

- Neural network architecture
  - One hidden layer
  - 200 dimension for input and hidden layers
  - Rectifier activation function (ReLU)
  - Noise-contrastive estimation (NCE)
  - Best of 25 epochs
- Vocabulary size:
  - 500,000 words
  - 80,000 subcat tokens



# Experimental Setup

- Data
  - Train: Parallel data from WMT15
  - Test: newstest2013-2015

	Train	Tune	Test
Sentences	4,472,694	2,000	8,172
Verb tokens	5,945,637	2,419	11,161

WMT15 data for German - English.

# Generic Lexicon Model

- Generic model has lower accuracy for verbs
  - Input: 5-word window
  - Trained on all words

test samples	perplexity	acc@1	acc@5
All words	23.62	50.62	70.51
Verbs only	+2.58	-6.95	-2.63

Results computed over all words and over 11,161 verbs from newstest2013-2015

# Verb Lexicon Model

- Window context
  - Better than generic lexicon model
  - Wider context not helpful

model	factors	size	ppl	acc@1	acc@5
VLM	Word	5	27.81	50.57	76.27
	Word	7	+0.17	+0.00	-0.72
	Word, lemma	5	-0.61	-0.03	-0.37
Generic	Word	5	-1.61	-6.9	-8.39

Results computed over 11,161 verbs from newstest2013-2015

# Verb Lexicon Model

- Syntactic context
  - All factors improve model accuracy

context	factors	size	ppl	acc@1	acc@5
Window	Word	5	27.81	50.57	76.27
Syntactic	Word	7	-1.32	+0.64	-0.01
	Word, lemma	7	-2.82	+0.89	+0.85
	Word, lemma, subcat	7	-2.65	+0.97	+0.56
	<b>Word, lemma, subcat, particle</b>	<b>7</b>	<b>-2.97</b>	<b>+1.42</b>	<b>+1.27</b>

Results computed over 11,161 verbs from newstest2013-2015

# Re-ranking with VLM

- String-to-tree WMT 2015 (DE-EN)
  - 1000-best list
- New feature scores:
  - Verb lexicon model score
    - Aggregated over target verbs output by decoder
  - Count of source verbs translated
- Re-tune all features with MERT (3 runs)
  - BLEU objective

# Machine Translation Evaluation

- Small decrease in evaluation scores with VLM

system	factors	BLEU	METEOR
Baseline	-	26.10	29.95
+ VLM (window)	Word	-0.39 <sub>±0.26</sub>	-0.13 <sub>±0.14</sub>
+ VLM (syntactic)	Word, lemma, subcat, particle	-0.37 <sub>±0.19</sub>	-0.14 <sub>±0.06</sub>

Results for re-ranking 1000-best list for newstest2015.  
Results averaged over 3 tuning runs

# Machine Translation Evaluation

- Verb translation accuracy
  - Both precision and recall improve with VLM

system	factors	precision	recall	F1
Baseline	-	56.91	47.86	51.99
+ VLM (window)	Word	+1.95 <sub>±0.66</sub>	+7.45 <sub>±0.42</sub>	+5.04 <sub>±0.32</sub>
+ VLM (syntactic)	Word, lemma, subcat, particle	+2.70 <sub>±0.89</sub>	+7.36 <sub>±0.40</sub>	+5.34 <sub>±0.56</sub>

Token level precision, recall and F1 scores for re-ranking 1000-best list for newstest2015. Results averaged over 3 tuning runs

# Example

- Particle verb

Source	Die Ankläger <b>legten</b> am Freitag dem Büro des Staatsanwaltes von Mallorca Beweise für Erpressungen durch Polizisten und Angestellte der Stadt Calvia <b>vor</b> .
Reference	The claimants <b>presented</b> proof of extortion by policemen and Calvia Town Hall civil servants at Mallorca's public prosecutor's office on Friday .
Baseline	The prosecutor <b>went to</b> the office of the prosecutor of Mallorca Calvi evidence of extortion by police officers and employees of the city on Friday .
Translation rule	VP → < <b>legten</b> VP, <b>went</b> VP > PP → < NP vor, <b>to</b> NP >



# Example

- Particle verb

Source	Die Ankläger <b>legten</b> am Freitag dem Büro des Staatsanwaltes von Mallorca Beweise für Erpressungen durch Polizisten und Angestellte der Stadt Calvia <b>vor</b> .
Reference	The claimants <b>presented</b> proof of extortion by policemen and Calvia Town Hall civil servants at Mallorca's public prosecutor's office on Friday .
Baseline	The prosecutor <b>went to</b> the office of the prosecutor of Mallorca Calvia evidence of extortion by police officers and employees of the city on Friday .
Verb lexicon	The prosecutor <b>presented</b> evidence of extortion by police officers and employees of the city on Friday the office of the prosecutor of Mallorca Calvia before .
Syntactic context	<b>Legten</b> , Ankläger, Büro , Staatsanwaltes , am Freitag, <b>vor</b> Subj_pp_objd_obja_pp_pp_avz

# Example

- Re-ordering problem

Source	Webster wird darüber hinaus <b>vorgeworfen</b> , am 4. Mai 2014 eine zweite Frau im Golf View Hotel in Naim im schottischen Hochland angegriffen zu haben .
Reference	Webster is then <b>charged</b> with attacking a second woman at the Golf View Hotel in Nairn in the Highlands on May 4 , 2014 .
Baseline	Webster is also <b>alleged</b> to have attacked a second woman in Naim's Golf View Hotel in the Scottish Highlands on 4 May 2014 .
Translation rule	VBN → ⟨ <b>vorgeworfen</b> , <b>alleged</b> ⟩ VP → ⟨ VBN , VP zu haben , VBN to have VP ⟩

# Example

- Re-ordering problem

Source	Webster wird darüber hinaus <b>vorgeworfen</b> , am 4. Mai 2014 eine zweite Frau im Golf View Hotel in Naim im schottischen Hochland angegriffen zu haben .
Reference	Webster is then <b>charged</b> with attacking a second woman at the Golf View Hotel in Nairn in the Highlands on May 4 , 2014 .
Baseline	Webster is also <b>alleged</b> to have attacked a second woman in Naim's Golf View Hotel in the Scottish Highlands on 4 May 2014 .
Verb lexicon	Webster is also <b>accused</b> of being a second wife in the Golf View Hotel on 4 May 2014 in Naim attacked in the Scottish Highlands .
Syntactic context	<b>vorgeworfen</b> , wird, Webster, haben, darüber hinaus objd_pp_subjc

# Example

- Re-ordering problem

Source	Webster wird darüber hinaus <b>vorgeworfen</b> , am 4. Mai 2014 eine zweite Frau im Golf View Hotel in Naim im schottischen Hochland <b>angegriffen zu haben</b> .
Reference	Webster is then <b>charged</b> with attacking a second woman at the Golf View Hotel in Nairn in the Highlands on May 4 , 2014 .
Baseline	Webster is also <b>alleged to have attacked</b> a second woman in Naim's Golf View Hotel in the Scottish Highlands on 4 May 2014 .
Verb lexicon	Webster is also <b>accused of being</b> a second wife in the Golf View Hotel on 4 May 2014 in Naim <b>attacked</b> in the Scottish Highlands .
Syntactic context	<b>vorgeworfen</b> , wird, Webster, haben, darüber hinaus objd_pp_subjc

# Conclusions

- We proposed a neural verb lexicon model
  - Wide source-side context
  - Syntactic context better than window (+1.4 acc)
- In re-ranking model improves verb translation
  - +2.7 precision and +7.4 recall
  - Small decrease in BLEU score ( $< 0.5$ )

# Thank you!

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreements 644402 (HimL) and 645452 (QT21)







# Breakdown of Verb Categories

- Less noisy data w. selecting only target verbs

	count	percentage
Source verbs	23,492	100.0
-- main verbs	17,210	73.2
-- particle verbs	1,589	6.7
-- <b>target verbs</b>	<b>11,161</b>	<b>47.5</b>
-- misaligned verbs	2,850	12.1

Counts for source-side verbs in newstest2013-2015.

# Automatic Metrics

- Small decrease in evaluation scores with VLM

system	factors	BLEU		METEOR
		dev	test	test
Baseline	-	26.18	26.10	29.95
+ VLM (window)	Word	-0.13 <sub>±0.08</sub>	-0.39 <sub>±0.26</sub>	-0.13 <sub>±0.14</sub>
+ VLM (syntactic)	Word, lemma, subcat, particle	-0.13 <sub>±0.06</sub>	-0.37 <sub>±0.19</sub>	-0.14 <sub>±0.06</sub>

Results for re-ranking 1000-best list for newstest2015.  
Results averaged over 3 tuning runs

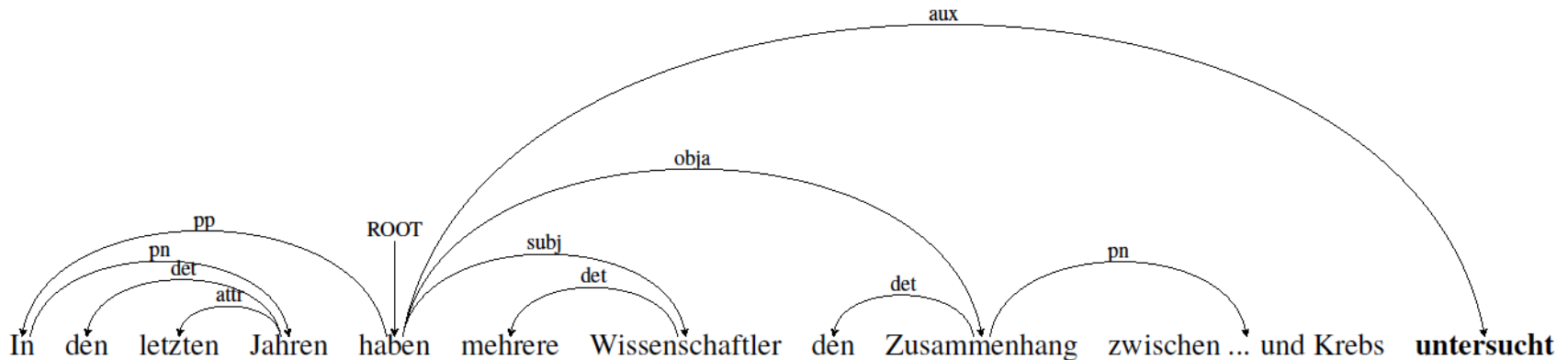
# Breakdown of Verb Categories

	count	percentage
Source verbs	23,492	100.0
-- auxiliary verbs	4,689	20.0
-- misaligned verbs	934	3.9
-- main verbs	17,210	73.2
-- particle verbs	1,589	6.7
-- target verbs	11,161	47.5
-- misaligned verbs	2,850	12.1
-- modals + other	1,593	6.8
-- lexical rules	4,905	20.8

# Related Work

- Discriminative Rule Selection
  - Select target side of scfg rule given source-side
    - Local features of the rule itself (Liu et al. 2008, Braune et al al, 2015)
    - Wider source features (Braune et al al, 2016)
  - Scaling problems
    - Very large number of training examples
    - Cross-products of sparse features

# Source-side Syntactic Context



Ref: In recent years, a number of scientists have **studied** the links [between ... and cancer]

Window context: {**untersucht**, und, Krebs, </s>, </s>}

Syntactic context:

	<i>source verb</i>	<i>parent</i>	<i>dependents</i>		<i>pp modifier</i>		<i>subcat</i>
word	<b>untersucht</b>	haben	Wissenschaftler	Zusammenhang	In	Jahren	pp_subj_objs
lemma	untersuchen	haben	Wissenschaftler	Zusammenhang	In	Jahr	

# Generic Lexicon Model

- FFNN, [NPLM]
- Input: 5-word window from source sentence
- Output: 1 target word

	perplexity	acc@1	acc@5	acc@15
All words	23.62	50.62	70.51	78.47
Verbs only	26.20	43.67	67.88	78.69