

# Investigating the Morphological Complexity of German Named Entities

## The Case of the GermEval NER Challenge

Bettina Klimek<sup>2</sup>   Markus Ackermann<sup>1</sup>   Amit Kirschenbaum<sup>2</sup>  
Sebastian Hellmann<sup>1,2</sup>

<sup>1</sup>Institut für Informatik  
Universität Leipzig

<sup>2</sup>Institut für Angewandte Informatik Leipzig

International Conference of the German Society for Computational  
Linguistics and Language Technology

# English and German NER - A Performance Gap

English (ConLL 2003)	
system	$F_1$
Passos et al. 2014	90.90
Huang et al. 2015	90.10
Florian et al. 2003	88.76

German (NoSta-D 2014)	
system	$F_1$
Agerri and Rigau 2016	76.43
Hänig et al. 2014	76.38
Reimers et al. 2014	75.09

# German NER - Morphological Challenges

- frequent and extensive compounding

*'Bibelforscherfrage' (Bible researchers' question)*

# German NER - Morphological Challenges

- frequent and extensive compounding
- morpho-phonologically conditioned inner modifications (e.g. *außereuropäisch* (*Europa*) - 'non-European', Häfler - Hafel)

'*außereuropäisch*' - '*Europa*' (*non-European* - *Europe*)  
'Häfler' - '*Hafel*' (citizen of German city Hafel)

# German NER - Morphological Challenges

- frequent and extensive compounding
- morpho-phonologically conditioned inner modifications (e.g. *außereuropäisch* (*Europa*) - 'non-European', Häfler - Hafel)
- NEs in different word-classes after derivation (e.g. *lutherischen*, an adjective, derived from the proper noun *Martin Luther*)

'lutherischen' - *Martin Luther* (*Lutheran*)

# Research Questions

Do morphological alternations of proper nouns constitute another difficulty layer which needs to be addressed to shrink the performance gap between German and English NER?

# Research Questions

Do morphological alternations of proper nouns constitute another difficulty layer which needs to be addressed to shrink the performance gap between German and English NER?

- How to formalize morphological complexity in NER context to allow quantitative comparisons?
- Is there a measurable relation between NER errors and occurrence of morphological alterations?
- Which morphological alterations pose harder challenges to German NER system?

# The GermEval NER Challenge

Benikova et al. 2014b

- based on the NoSta-D dataset (Benikova et al. 2014a)
- follow up on the ConLL 2003 NER Challenge for English and German (Tjong Kim Sang and De Meulder 2003)
- extensions of challenge task:
  - nested named entities (*'Real Madrid'*, *'Deutschlandfunk'*) as own annotation layer
  - additional NE classification to mark compounding (*part*) and derivation (*deriv*): *'Berlin'* - LOC, *'Berlin-Liebhaber'* - LOC*part*, *'Berliner'* - LOC*deriv*
- 11 submitted systems
- best achieved official F1-Measure: 76.38 % (Hänig et al. 2014)



# The NoSta-D Corpus

Benikova et al. 2014a

- primary text: sentence wise samples from Wikipedia articles and online newspapers, from Leipzig Corpora Collection (Richter et al. 2006)

	<b>sentences</b>	<b>tokens</b>	<b>NEs</b>
<b>Train</b>	24,000	$\approx$ 452800	$\approx$ 31500
<b>Dev/Tune</b>	2,200	$\approx$ 41600	$\approx$ 2880
<b>Test</b>	5,100	$\approx$ 96400	$\approx$ 6690
<b>Total</b>	31,300	$\approx$ 591000	$\approx$ 41100

- our investigation focused on (subsets of) the Test split

# The NoSta-D Corpus

Benikova et al. 2014a

1951 bis 1953 wurde der nördliche Teil als Jugendburg **OTH**  
des Kolpingwerkes gebaut.

---

Beschreibung Die Kanadalilie erreicht eine **LOCpart**  
Wuchshöhe von 60 bis 180 cm und wird bis zu  
25 cm breit.

---

Um 1800 wurde im ehemaligen Hartung'schen Amt- **PERderiv**  
shaus eine Färberei eingerichtet.

---

1911 wurde er Mitglied der sozialistischen Partei, **ORG**  
aus der er aber ein Jahr später wieder austrat.

# GermEval System Predictions

- automatic NE annotations of the Test split from the competing systems
- now publicly available for further research after consent of all participants
- comparative analysis of true positives and false negatives<sup>1</sup> across the systems
- construction of subsets of Test split:

## GermEval Predictions Subsets

- True Pos. Intersection (TPi)** recognised and classified correctly by all systems ➔ NE phrases posing low or no challenges
- False Neg. Intersection (FNi)** not recognised and classified correctly by any system ➔ NE phrases with 'unsolved' challenges
- False Neg. Best System (FN ExB)** not recognised and classified correctly by best system

	NEs	Sentences
<b>TPi</b>	1008	791
<b>FNi</b>	692	530
<b>FN ExB</b>	1690	1093

# Operationalising Morphological Complexity

Source Named Entity (SNE) as found in the Corpus

Target Named Entity (TNE) the referenced entity in lexical canonical form (LCF)

<u>Luxemburger</u>	<i>sehen</i>	<i>überwiegend</i>	<u>deutsches</u>	<i>Fernsehen.</i>
<u>Luxembourgers</u>	<i>watch</i>	<i>predominantly</i>	<u>German</u>	<i>television.</i>

# Operationalising Morphological Complexity

Source Named Entity (SNE) as found in the Corpus

Target Named Entity (TNE) the referenced entity in lexical canonical form (LCF)

<u>Luxemburger</u>	<i>sehen</i>	<i>überwiegend</i>	<u>deutsches</u>	<i>Fernsehen.</i>
<u>Luxembourgers</u>	<i>watch</i>	<i>predominantly</i>	<u>German</u>	<i>television.</i>

*Luxemburg*  
*Luxembourg*

*Deutschland*  
*Germany*

# Operationalising Morphological Complexity

Source Named Entity (SNE) as found in the Corpus

Target Named Entity (TNE) the referenced entity in lexical canonical form (LCF)

<u>Luxemburger</u>	<i>sehen</i>	<i>überwiegend</i>	<u>deutsches</u>	<i>Fernsehen.</i>
<u>Luxembourgers</u>	<i>watch</i>	<i>predominantly</i>	<u>German</u>	<i>television.</i>
▼?			▼?	
<i>Luxemburg</i>			<i>Deutschland</i>	
<i>Luxembourg</i>			<i>Germany</i>	

# Operationalising Morphological Complexity

$\mathcal{C}_k$  compounding alterations were applied

$\mathcal{D}_l$  derivations applied

$c$  a derivation applied resulted in a word-class change between SNE and TNE

$m$  an inner modification of the SNE stem compared to its LCF

$f$  SNE is inflected



# Operationalising Morphological Complexity

$C_k$  compounding alterations were applied

$D_l$  derivations applied

$c$  a derivation applied resulted in a word-class change between SNE and TNE

$m$  an inner modification of the SNE stem compared to its LCF

$f$  SNE is inflected

complexity  $> C_0D_0$ : morphologically relevant

complexity with  $C + D \geq 1$ : morphologically complex

# Operationalising Morphological Complexity

SNE: Transatlantikflüge      transatlantic flights

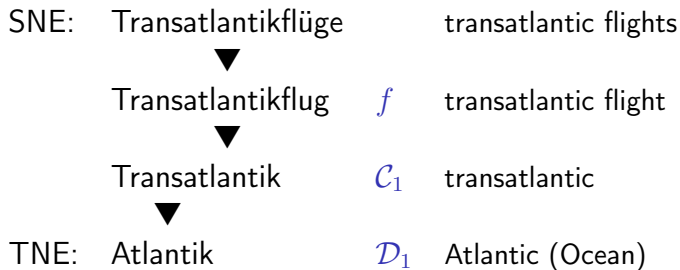
# Operationalising Morphological Complexity

SNE: Transatlantikflüge      transatlantic flights  
    ▼  
Transatlantikflug      *f*      transatlantic flight

# Operationalising Morphological Complexity

SNE:	Transatlantikflüge		transatlantic flights
	▼		
	Transatlantikflug	$f$	transatlantic flight
	▼		
	Transatlantik	$C_1$	transatlantic

# Operationalising Morphological Complexity



# Operationalising Morphological Complexity

SNE:	Transatlantikflüge		transatlantic flights
	▼		
	Transatlantikflug	$f$	transatlantic flight
	▼		
	Transatlantik	$C_1$	transatlantic
	▼		
TNE:	Atlantik	$D_1$	Atlantic (Ocean)

**overall complexity:**  $C_1 D_1 f$

# Morphological Complexities

Mozart

Mozart

$\mathcal{C}_0\mathcal{D}_0$

1352 (84.3 %)

# Morphological Complexities

Mozart	Mozart	$\mathcal{C}_0\mathcal{D}_0$	1352	(84.3 %)
Mozarts	Mozart	$\mathcal{C}_0\mathcal{D}_0f$	74	(4.6%)



# Morphological Complexities

Mozart	Mozart	$C_0D_0$	1352	(84.3 %)
Mozarts	Mozart	$C_0D_0f$	74	(4.6%)
Reichtagssabgeordneten	Reichstag	$C_1D_0mf$	3	

# Morphological Complexities

Mozart	Mozart	$C_0D_0$	1352	(84.3 %)
Mozarts	Mozart	$C_0D_0f$	74	(4.6%)
Reichtagssabgeordneten	Reichstag	$C_1D_0mf$	3	
Transatlantikflüge	Atlantik	$C_1D_1f$	2	

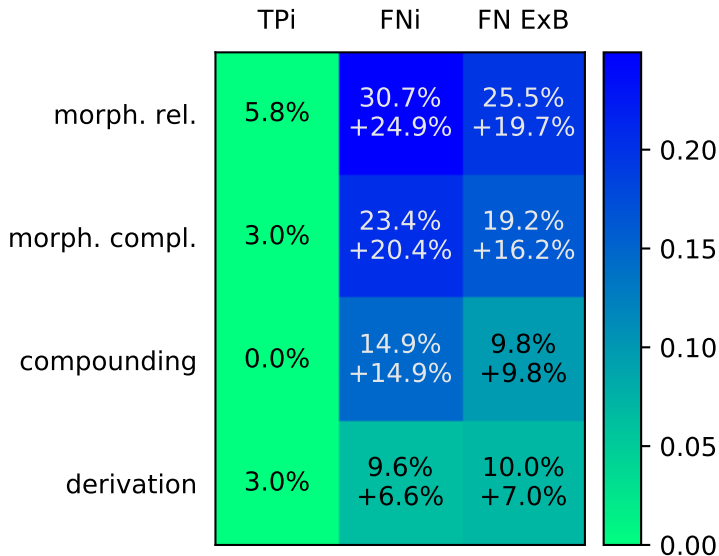
# Morphological Complexities

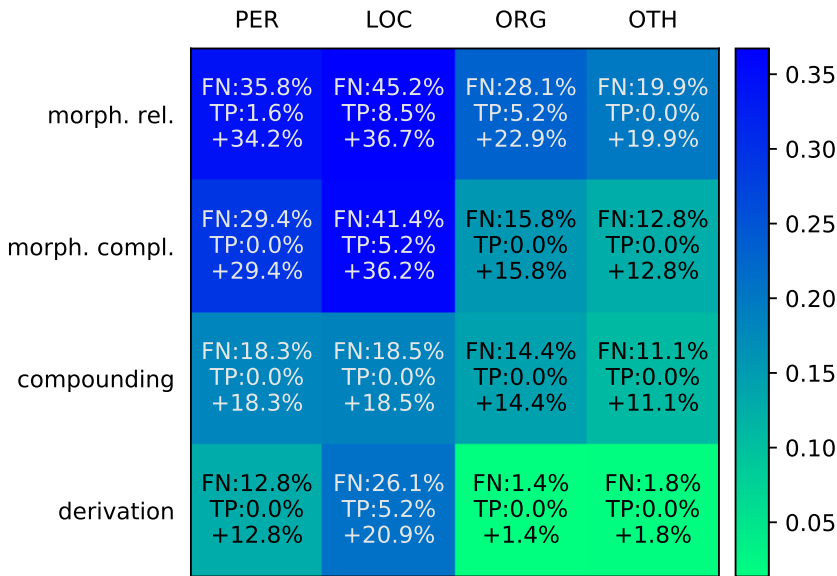
Mozart	Mozart	$\mathcal{C}_0\mathcal{D}_0$	1352	(84.3 %)
Mozarts	Mozart	$\mathcal{C}_0\mathcal{D}_0f$	74	(4.6%)
Reichtagsabgeordneten	Reichstag	$\mathcal{C}_1\mathcal{D}_0mf$	3	
Transatlantikflüge	Atlantik	$\mathcal{C}_1\mathcal{D}_1f$	2	
gesamtschweizerischen	Schweiz	$\mathcal{C}_1\mathcal{D}_2cf$	1	

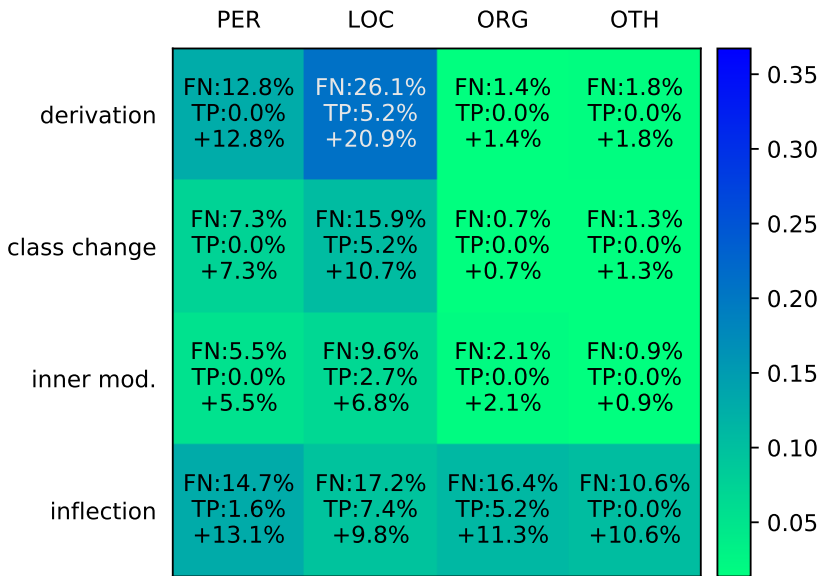
- 27 distinct complexity levels
- evidence for about 71% of combinatorial space until  $\mathcal{C}_1\mathcal{D}_1cmf$

compl.	TPI	FNi	FN ExB	example SNE	example TNE
$C_0D_0$	910 (94.20%)	442 (69.28%)	1149 (74.47%)	Mozart	Mozart
$C_0D_0f$	27 (2.80%)	47 (7.37%)	98 (6.35%)	Mozarts	Mozart
$C_1D_0$	0 (0.00%)	62 (9.72%)	101 (6.55%)	Mozart-Konzert	Mozart
$C_1D_0f$	0 (0.00%)	15 (2.35%)	24 (1.56%)	Mozart-Konzerten	Mozart
$C_1D_0m$	0 (0.00%)	3 (0.47%)	5 (0.32%)	Pieterskirche	Pieter
$C_1D_0mf$	0 (0.00%)	3 (0.47%)	4 (0.26%)	Reichstagsabgeordneten	Reichstag
$C_0D_1$	0 (0.00%)	9 (1.41%)	20 (1.30%)	Donaldismus	Donald
$C_0D_1f$	0 (0.00%)	1 (0.16%)	4 (0.26%)	Donaldismusses	Donald
$C_0D_1m$	0 (0.00%)	7 (1.10%)	10 (0.65%)	Nestorianismus	Nestorius
$C_0D_1mf$	0 (0.00%)	1 (0.16%)	2 (0.13%)	Spartiaten	Sparta
$C_0D_1c$	5 (0.52%)	16 (2.51%)	61 (3.95%)	japanisch	Japan
$C_0D_1cf$	9 (0.93%)	8 (1.25%)	14 (0.91%)	japanischen	Japan
$C_0D_1cm$	1 (0.10%)	1 (0.16%)	6 (0.39%)	europäisch	Europa
$C_0D_1cmf$	10 (1.04%)	8 (1.25%)	19 (1.23%)	europäischen	Europa
$C_2D_0$	0 (0.00%)	3 (0.47%)	5 (0.32%)	Bibelforscherfrage	Bibel
$C_2D_0mf$	0 (0.00%)	1 (0.16%)	1 (0.06%)	Erderkundungssatelliten	Erde
$C_1D_1$	0 (0.00%)	1 (0.16%)	2 (0.13%)	Benediktinerstift	Benedikt
$C_1D_1f$	0 (0.00%)	2 (0.31%)	2 (0.13%)	Transatlantikflüge	Atlantik
$C_1D_1m$	0 (0.00%)	1 (0.16%)	2 (0.13%)	Römerstrasse	Rom
$C_0D_2$	0 (0.00%)	1 (0.16%)	2 (0.13%)	Geismarerin	Geismar
$C_0D_2f$	0 (0.00%)	1 (0.16%)	2 (0.13%)	Hüttenbergerinnen	Hüttenberg
$C_0D_2m$	0 (0.00%)	0 (0.00%)	1 (0.06%)	Rheinländerin	Rheinland
$C_0D_2cf$	0 (0.00%)	1 (0.16%)	1 (0.06%)	austropolnischen	Polen
$C_0D_2cmf$	4 (0.41%)	0 (0.00%)	3 (0.19%)	transatlantischen	Atlantik
$C_3D_0$	0 (0.00%)	1 (0.16%)	1 (0.06%)	25-US-Dollar-Marke	US
$C_1D_2cf$	0 (0.00%)	2 (0.31%)	2 (0.13%)	gesamtschweizerischen	Schweiz
$C_1D_2cmf$	0 (0.00%)	1 (0.16%)	2 (0.13%)	Skialpinisten	Alpen
total	966	638	1543		

# Prevalence of Morph. Complexity







## A Morpho-Gap?

- false negative NEs with corresponding true positive NE in TPi (same TNE)
- 19 cases in FNi, 38 additional cases for FN ExB

False Neg.	TPi NE	shared TNE
austropolnischen	polnischen	Polen
gesamtschweizerischen	Schweizer	Schweiz
Japan-Aufenthaltes	Japans	Japan

- 3.4 % of false negatives of best system of GermEval are such cases



## Reference Annotation Deficiencies

issue	prevalence	ann. spec.
Wrong NE Type	62 (20.8 %)	✗
No NE	18 (6.0 %)	✗
Invalid Reference	7 (2.4 %)	✗
Not Derived	94 (31.5 %)	✓
TNE Unclear	66 (22.2 %)	✓
Wrong Spelling	51 (17.1 %)	✓

### Wrong NE Type:

SNE = *'barocker'* (ORG-deriv) with TNE = Barock, "Baroque" is an epoch (OTH-deriv)

## Reference Annotation Deficiencies

issue	prevalence	ann. spec.
Wrong NE Type	62 (20.8 %)	✗
No NE	18 (6.0 %)	✗
Invalid Reference	7 (2.4 %)	✗
Not Derived	94 (31.5 %)	✓
TNE Unclear	66 (22.2 %)	✓
Wrong Spelling	51 (17.1 %)	✓

### No NE:

'*Junta*' (SNE) is a common noun, there is no TNE

# Reference Annotation Deficiencies

issue	prevalence	ann. spec.
Wrong NE Type	62 (20.8 %)	✗
No NE	18 (6.0 %)	✗
Invalid Reference	7 (2.4 %)	✗
Not Derived	94 (31.5 %)	✓
TNE Unclear	66 (22.2 %)	✓
Wrong Spelling	51 (17.1 %)	✓

## Invalid Reference:

*'Was ist theoretische Biologie ?'* is an HTML link label, which is not related to any NE

## Reference Annotation Deficiencies

issue	prevalence	ann. spec.
Wrong NE Type	62 (20.8 %)	✗
No NE	18 (6.0 %)	✗
Invalid Reference	7 (2.4 %)	✗
Not Derived	94 (31.5 %)	✓
TNE Unclear	66 (22.2 %)	✓
Wrong Spelling	51 (17.1 %)	✓

### Not Derived:

'*Kirgisische*' (LOC-deriv) is derived from "*Kirgise*", but LOC TNE is Kirgistan

## Reference Annotation Deficiencies

issue	prevalence	ann. spec.
Wrong NE Type	62 (20.8 %)	✗
No NE	18 (6.0 %)	✗
Invalid Reference	7 (2.4 %)	✗
Not Derived	94 (31.5 %)	✓
TNE Unclear	66 (22.2 %)	✓
Wrong Spelling	51 (17.1 %)	✓

### TNE Unclear:

*'Köln/Weimar/Wien'* - TNE is unclear, unknown to which of the three named entities is referred to

## Reference Annotation Deficiencies

issue	prevalence	ann. spec.
Wrong NE Type	62 (20.8 %)	✗
No NE	18 (6.0 %)	✗
Invalid Reference	7 (2.4 %)	✗
Not Derived	94 (31.5 %)	✓
TNE Unclear	66 (22.2 %)	✓
Wrong Spelling	51 (17.1 %)	✓

### Wrong Spelling:

'Freiburg/31:52' with TNE Freiburg

# Richer Morph. Features from Labeled Morphological Segmentation

Labeled Morphological Segmentation (Cotterell et al. 2015) for  
'*gesamtschweizerisch*' ('pertaining to the whole of Switzerland'):

<b>gesamt</b>	<b>schweiz</b>	<b>er</b>	<b>isch</b>
AFFIX		AFFIX	AFFIX
DERIV	STEM	DERIV	DERIV

- stem-based matching of SNEs against gazetteers (match '*schweiz*' substring instead of '*gesamtschweizerisch*')
- mine distinctive sequential patterns:

~ (AFFIX DERIV) ('isch' AFFIX DERIV) \$

~ ('gesamt' AFFIX DERIV) ('isch' AFFIX DERIV) \$

# Conclusion

- error analysis for GermEval focused on morphological complexity
- linguistically motivated, principled approach to operationalize morphological complexity
- false negatives were considerably more likely to be complex
- varying variants of complexity across NE types

## Further Directions:

- test how well additional morph. features mitigate remaining morph. challenges
- extend annotations (inter-annotator agreement; top 4 systems)



# Thank you for your attention.

These research activities were funded by grants from the H2020 EU projects ALIGNED (GA-644055) and FREME (GA-644771) and the Smart Data Web BMWi project (GA-01MD15010B).