

# Why learner texts are easy to tag

A comparative evaluation of  
part-of-speech tagging of Kobalt

Marc Reznicek and Heike Zinsmeister

Workshop: Modeling non-standardized writing

DGfS Jahrestagung, Potsdam

March 13, 2013

# Parts-of-speech in learner texts

Und wenn unsere Eltern in ihrer Freizeit im Park ...

spaziereng**a**ngen

**TRUNC**

spaziereng**i**ngen

**VVFIN**

... und Volleyball spielten, sitzen wir ständig vor dem Computer ...

...und per Internet mit irrealen Freunden

**verkehren**

**VVIN**

... und

**verkehren**

**VVFIN**

per Internet mit irrealen Freunden.

*'And if our parents in their free-time in-the park strolled and volley ball played, sit we constantly in-front-of the computer and via internet with unreal friends chat.'*

# Parts-of-speech in learner texts

Und wenn unsere Eltern in ihrer Freizeit im Park ...

spaziereng<sup>a</sup>ngen

TRUNC

spaziereng<sup>i</sup>ngen

VVFIN

... und Volleyball spielten, sitzen wir ständig vor dem Computer ...

...und per Internet mit irrealen Freunden

verkehren

VVINF

... und verkehren per Internet mit irrealen Freunden.

VVFIN

How does tagging of **non-native speaker argument essays** differ from those of native speakers?

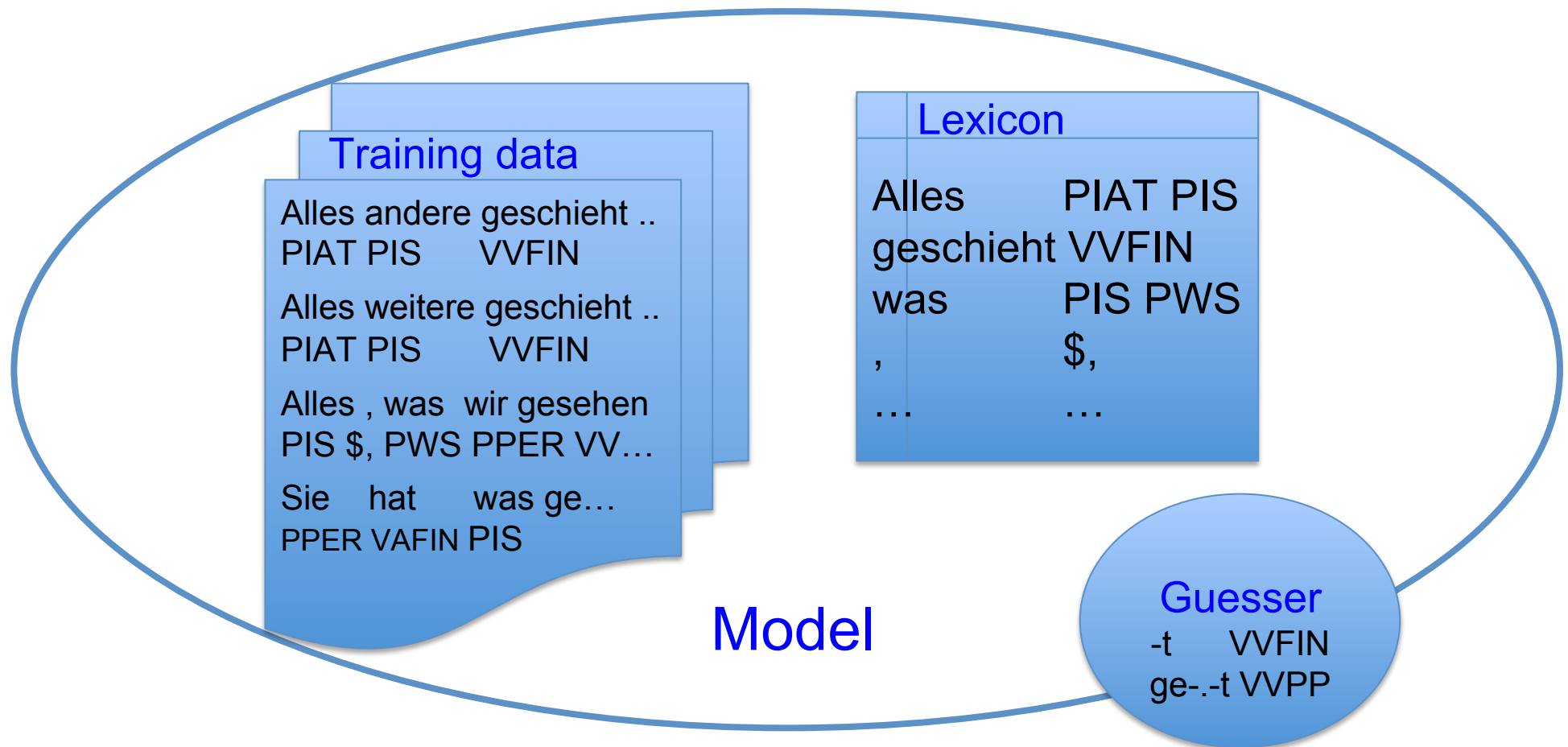
- authors
- text type

→ **two-fold non-standardized variety**

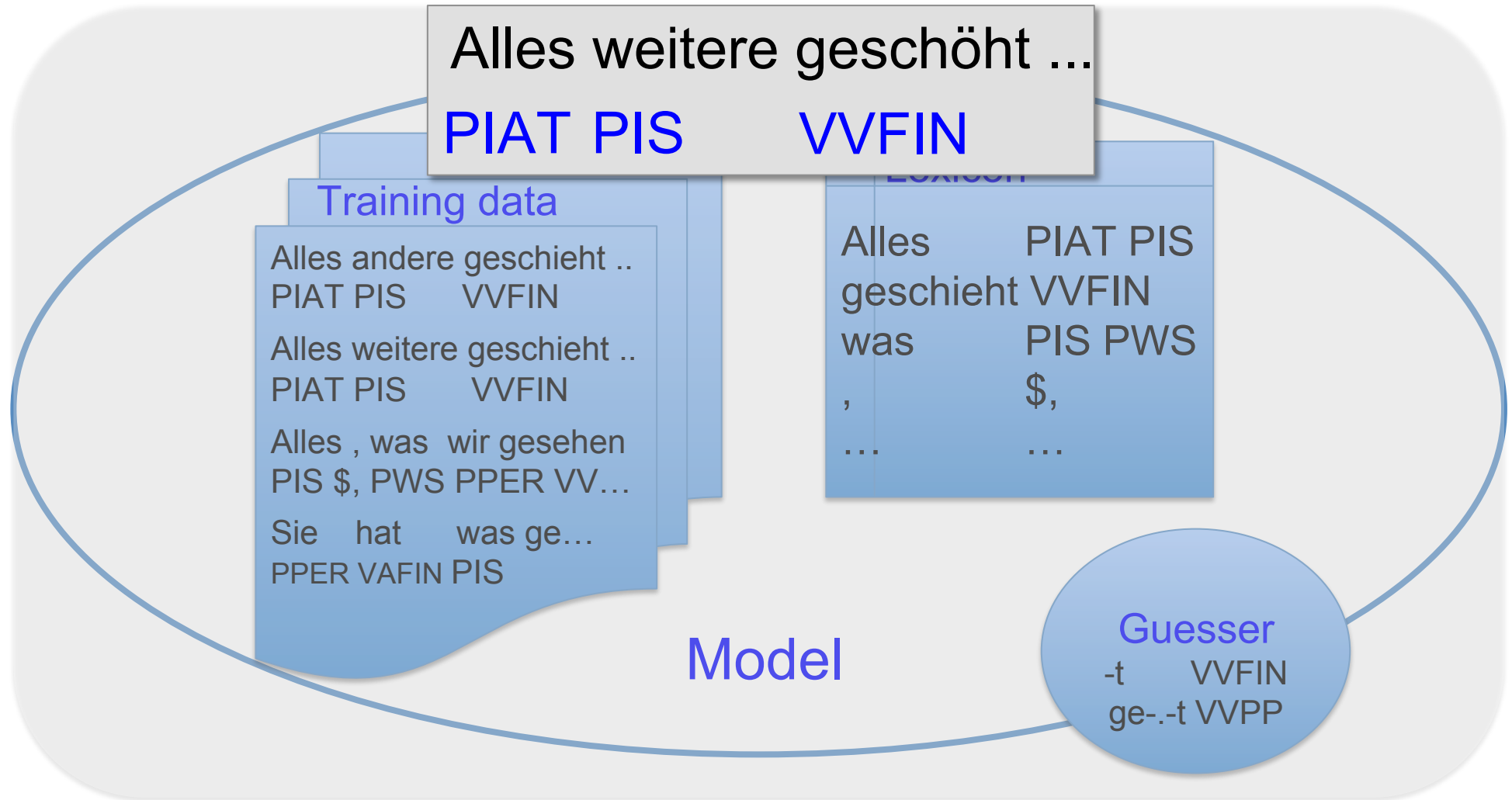
# Program

- Background
- Research question & hypotheses
- Experiment
- Conclusion & future work

# Statistical POS-Tagging



# Statistical POS-Tagging



# Tagging learner language

## Parts-of-speech multi-level cues

- POS of a word is determined by
  - its syntactical distribution
  - its morphological marking
  - its lexical stem (Díaz-Negrillo et al. 2010)

## Learner language systematically deviates from native language

*Jeden Tag viele **Kriminal/NN** Aktivitäten **passiert/VVPP** in der **Heutzutager/NN** Gesellschaft.* (TreeTagger@FalkoEssayL2v2.3)

*Every day many criminal activities happen in todays society.*

→ Different level cues for POS can contradict each other.

# Target hypothesis (TH)

## STTS guidelines

"Wenn der Sinn erkennbar ist, wird die WF verbessert, und es wird so getaggt, wie die richtige Wortform ausgesehen hätte."

If the sense is accessible, the word form is corrected and tagged like the correct word form. (Schiller et al. 1999:10)

**minimal target hypotheses (TH1)** corrects only morpho-syntax and orthography (Lüdeling et al. 2005, Reznicek et al. to appear)

*Jeden Tag viel **kriminelle/ADJA** Aktivität **passiert/VVFIN** in der **heutigen/ADJA** Gesellschaft.*

## quality:

- POS-Tags for TH1 (rfTagger) 98.9% (Rehbein et al. 2012)



# Target hypothesis (TH)

## edit tags for deviations

- TH1 = full parallel text to the original learner text
- Differences between the text (and its annotations) are marked with **edit tags**.

### Tag

INS
DEL
CHA
MERGE
SPLIT
MOVS
MOVT

# Target hypothesis (TH)

## edit tags for deviations

Tag
INS
DEL
CHA
MERGE
SPLIT
MOVS
MOVT

tok	pos	TH1	TH1pos	TH1Diff	TH1posDiff
Jeden	PIAT	Jeden	PIAT		
Tag	NN	Tag	NN		
viele	ADJA			MOVS	MOVS
Kriminal	NN			MOVS	MOVS
Aktivitäten	NN			MOVS	MOVS
passiert	VVPP	passiert	VVFIN		CHA
		viel	ADV	MOVT	MOVT
		kriminelle	ADJA	MOVT	MOVT
		Aktivität	NN	MOVT	MOVT
in	APPR	in	APPR		
der	ART	der	ART		
Heutzutage	NN	heutigen	ADJA	CHA	CHA
Gesellschaft	NN	Gesellschaft	NN		

# Program

- Background
- **Research question & hypotheses**
- Experiment
- Conclusion & future work

# Research question

## two options to achieve better tags

1) improve input data (explicit TH)

- time-consuming, offline, manual annotation

**not possible for unsupervised data processing like ICALL**

2) improve taggers (implicit TH)

- fast, on-the-fly annotation

## this study

How close do standard tools get to the desired output?

# Null hypotheses

Off-the-shelf statistically trained POS taggers **perform worse**

- on **unknown words** than on known words
  - on **mis-ordered words** than on words in target language order
- on **essays written by L2 learners** of German than on essays written by native speakers.

# Program

- Background
- Research question & hypotheses
- **Experiment**
- Conclusion & future work

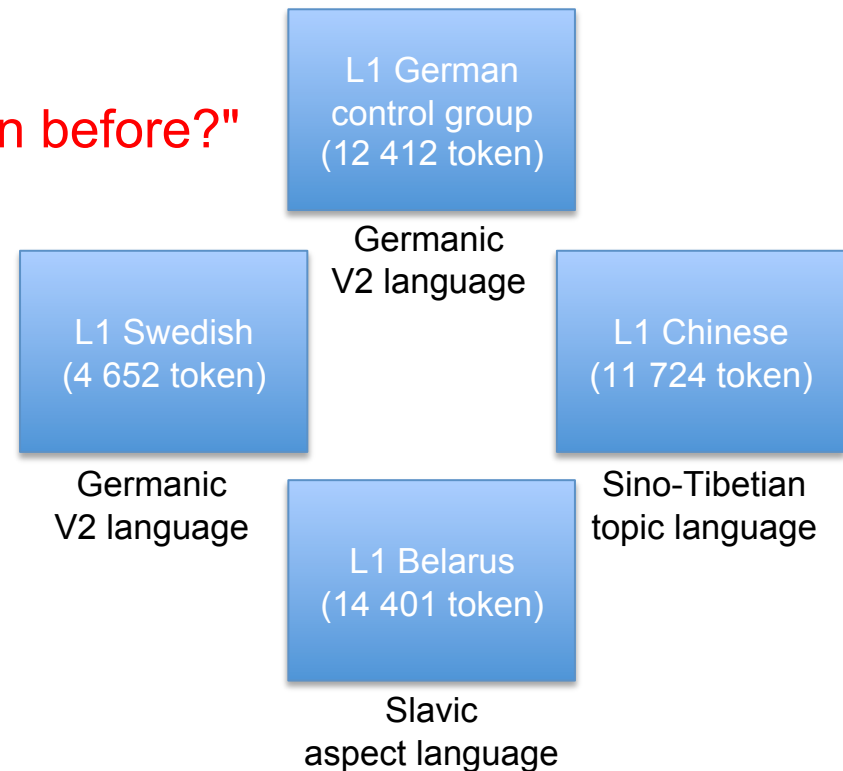
# Kobalt corpus

## sampling

- advanced learners (OnDaF: ~B2)
- argumentative essays:  
    “Is the youth better off today than before?”
- 20 texts / L1
- 90 min (~500 words)

## annotation

- target hypotheses
- parts-of-speech (STTS), lemmas
- grammatical functions
- topological fields
- edit tags for deviations



(version 1.2 03/2013)

# Experiment: data

**gold standard** (tok & TH1 → gold)

- 4 texts: 1 per L1 (BEL, CMN, SWE, DEU)
- TH1 tagged and corrected (2 consolidated annotations)

→ **How well does the TreeTagger reproduce gold POS tags?**

- on the TH1
- on the learner text

**test corpus** (tok → TH1)

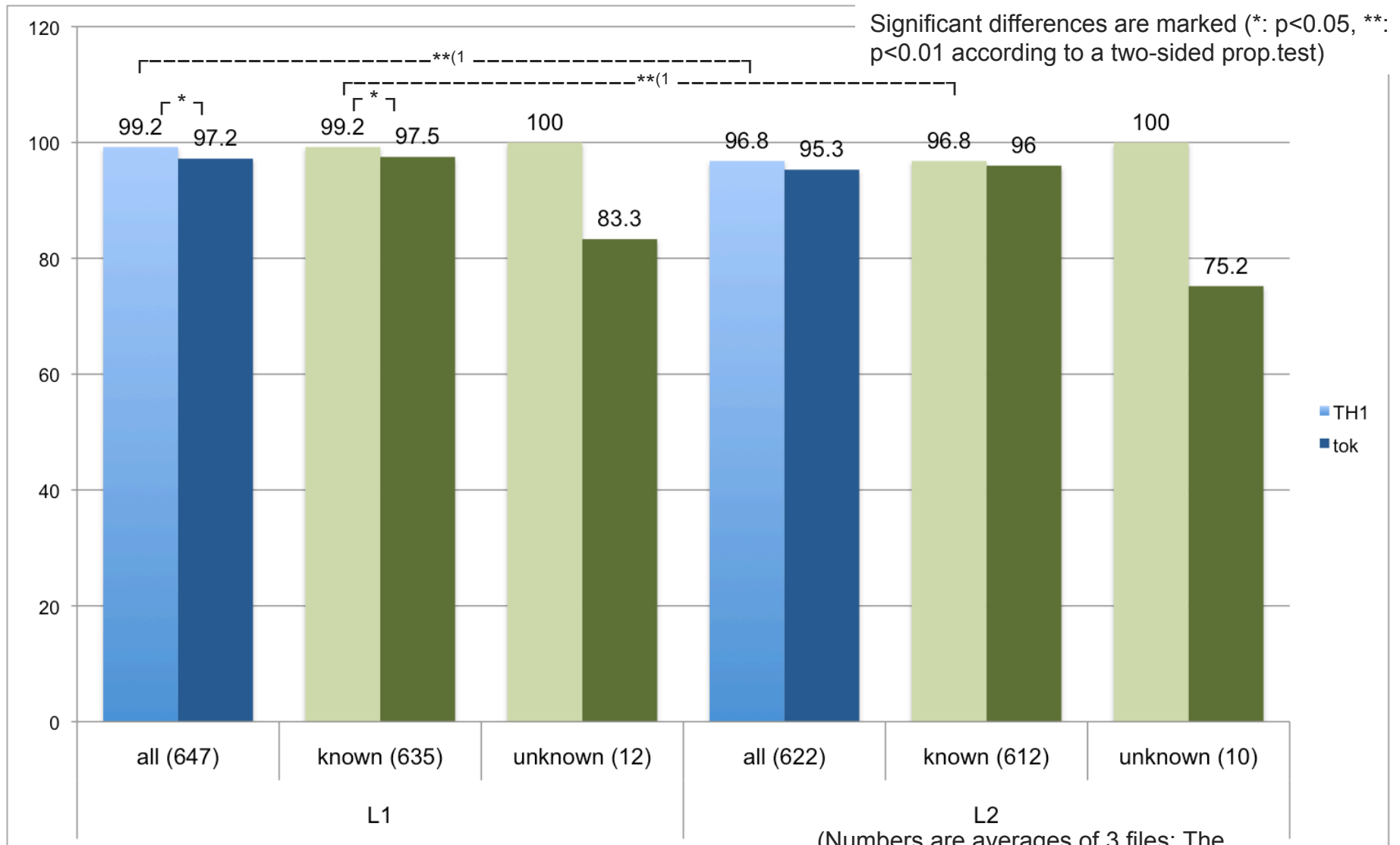
- 69 texts: 20 Chinese, 20 Belarus, 9 Swedish, 20 German

→ **How well does the TreeTagger reproduce TH1 POS tags?**

- on the learner text (tok)



# Results: tagging accuracy



(Numbers are averages of 3 files; The averaged differences are not significant)<sup>a</sup>

# Results: tok-TH1 match

Und wenn unsere Eltern in ihrer Freizeit im Park ...

tok:

spazierengangen

TRUNC

TH1:

spazierengingen

VVFIN

... und Volleyball spielten, sitzen wir ständig vor dem Computer ...

tok:

...und per Internet mit irrealen Freunden

verkehren

VVINF

TH1:

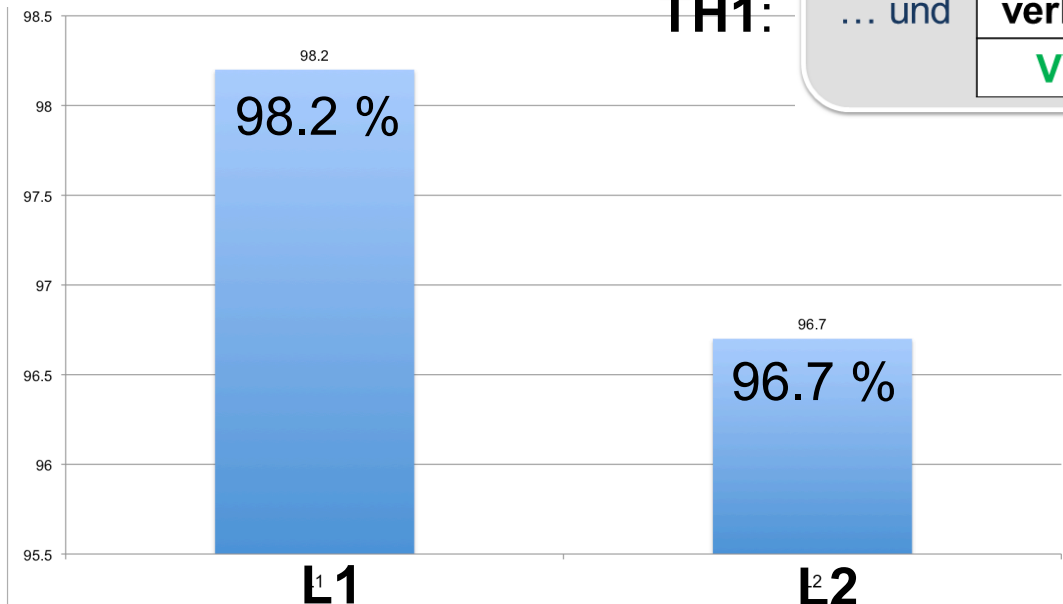
... und

verkehren

per Internet mit irrealen Freunden.

VVFIN

Match between TH1-pos and tok-pos

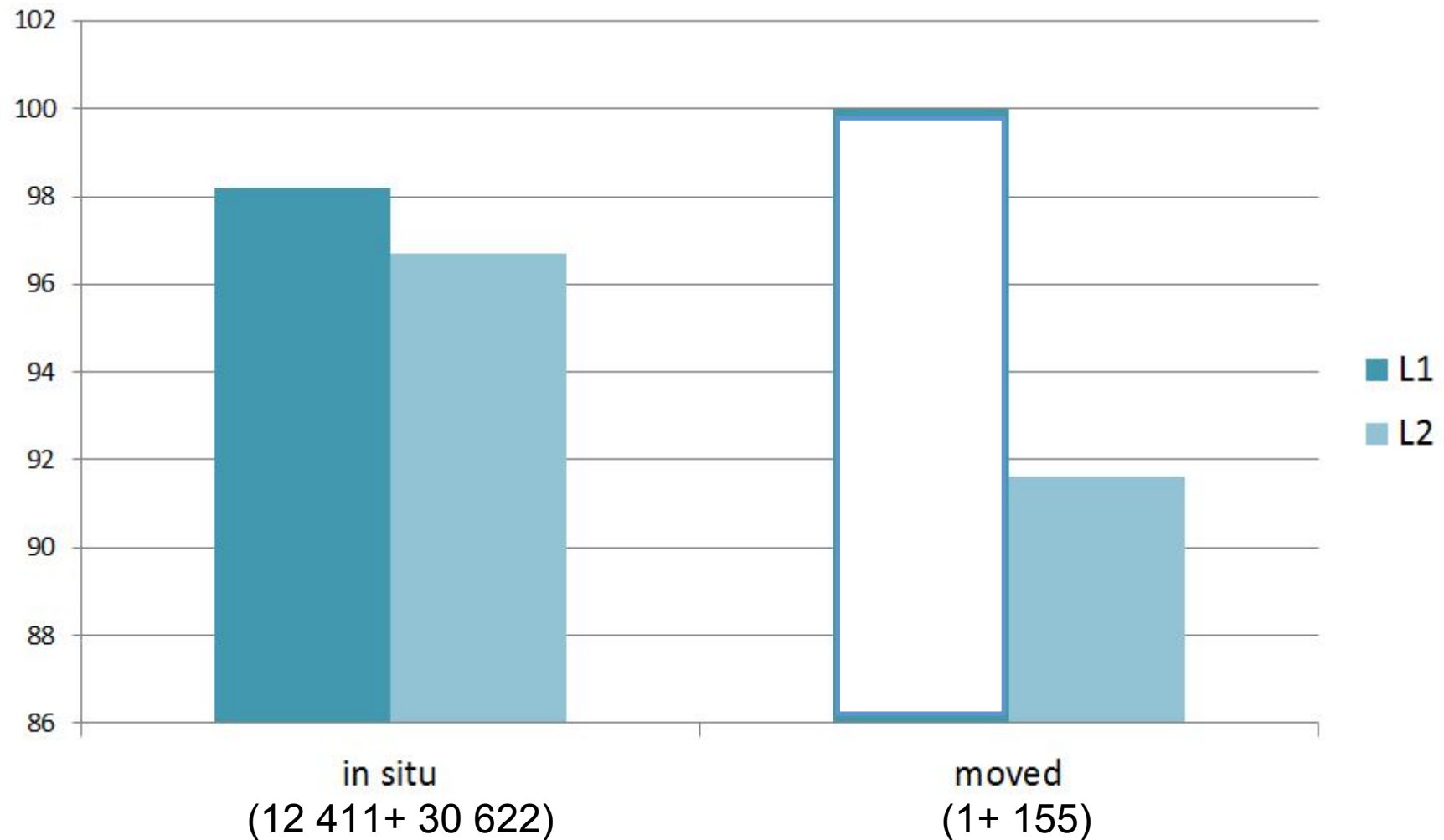


# Results: confusion matrix

		TH1Pos											
		NA	ADJD	ADV	APPR	KOKOM	KOUI	KOUS	NN	PRELS	VVFIN	VVIN	VVIZU
TokPos	\$,	23											
	ADJA				3				13				
	ADJD			10					5				
	ADV								5				
	APPR						5						
	ART									3			
	KOKOM							3					
	KOUS					7							
	NE										3		
	NN		7	4									
	PIAT									3			
	PIS									3			
	VVFIN			3					8			6	
	VVIN										3		4

confusions with freq  $\geq 3$

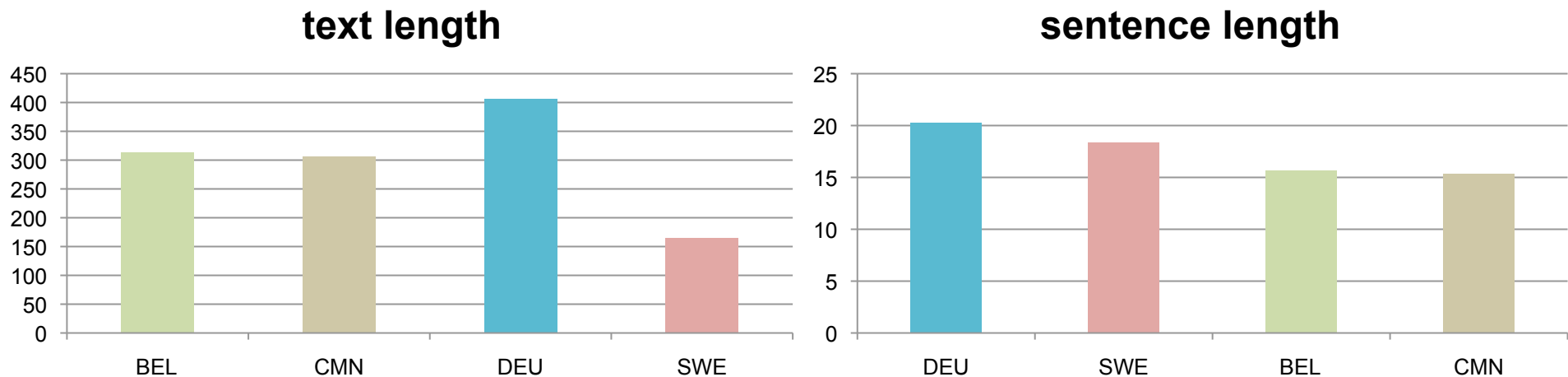
# Results: accuracy – moved words



# Experiment: analysis of other factors

## ■ sentence length

- longer sentences ~ more complex sentence structure
- shorter sentences ~ higher information density ??



# Program

- Background
- Research questions & hypotheses
- Experiment
- Conclusion & future work

# Conclusion

TreeTagger performance drops ...

- for unknown words → TRUE
- on mis-ordered words → TRUE
- for learner language vs. native language → TRUE

BUT:

Performance stays close to newspaper standard results.

# Conclusion





# Future work

## Ensemble-Tagging

- known improvement on majority-vote

(Van Halteren et al. 2001)

- TreeTagger (Schmid 95)
- RFTagger (Laws & Schmid 2009)
- Stanford Tagger (Toutanova & Manning 2000)

## Classifier-Training

- tagger training on tag-combinations

# Thanks to

## Our colleagues in the Kobalt-DaF network:

- Ute Bohnacker (Uppsala/Sweden)
- Margit Breckle (Vasa/Finland)
- Eva Breindl (Erlangen)
- Sigrid Dentler (Gothenburg/Sweden)
- Hagen Hirschmann (Berlin)
- Anke Lüdeling (Berlin)
- Detmar Meurers (Tübingen)
- Julia Ricart Brede (Heidelberg)
- Christina Rosén (Växjö/Sweden)
- Dirk Skiba (Jena)
- Maik Walter (Berlin)



# References

- Díaz-Negrillo, Ana; Meurers, Walt Detmar; Valera, Salvador; Wunsch, Holger (2010):** Towards Interlanguage POS Annotation for Effective Learner Corpora in SLA and FLT. In: *Language Forum*.
- Giesbrecht, Eugenie & Evert, Stefan (2009):** Part-of-speech Tagging - a Solved Task? An Evaluation of POS Taggers for the Web as Corpus. Alegria, I.; Leturia, I. & Sharoff, S. (ed.) Proceedings of the 5th Web as Corpus Workshop (WAC5)
- Lüdeling, Anke; Walter, Maik; Kroymann, Emil; Adolphs, Peter (2005):** Multi-level Error Annotation in Learner Corpora. In: Proceedings of Corpus Linguistics 2005. Birmingham.
- Rehbein, Ines; Hirschmann, Hagen; Lüdeling, Anke; Reznicek, Marc (2012):** Better Tags Give Better Trees or do they? In: *LiLT* 7 (10).
- Reznicek, Marc; Lüdeling, Anke; Hirschmann, Hagen (to appear):** Competing Target Hypotheses in the Falko Corpus. A Flexible Multi-Layer Corpus Architecture. In: Ana Dí-az-Negrillo (ed.): Automatic Treatment and Analysis of Learner Corpus Data: John Benjamins.
- Schiller, Anne; Teufel, Simone; Stöckert, Christine; Thielen, Christine (1999):** Guidelines für das Tagging deutscher Textkorpora mit STTS. Technical Report. University of Stuttgart; University of Tübingen.
- Schmid, Helmut (1995):** Improvements in Part-of-Speech Tagging with an Application to German. In: Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland.
- Schmid, Helmut; Laws, Florian (2008):** Estimation of Conditional Probabilities with Decision Trees and an Application to Fine-grained POS Tagging. In: Donia Scott (ed.): 22nd International Conference on Computational Linguistics. Coling 2008. International Conference on Computational Linguistics. Manchester, United Kingdom, 18 - 22 August 2008. COLING. Stroudsburg, Pa: Association for Computational Linguistics (ACL), S. 777–784.
- Toutanova, Kristina; Manning, Christopher D. (ed.) (2000):** Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora: Association for Computational Linguistics (ACL).
- Van Halteren, Hans, Walter Daelemans, and Jakub Zavrel (2001):** "Improving Accuracy in Word Class Tagging through the Combination of Machine Learning Systems." *Computational Linguistics* 27.2.199-229.APA