



# Falko



## Error annotations in Falko 2.x

Marc Reznicek & Cedric Krummes

**Symposium**

What's Hard in German?

Structural Difficulties, Research Approaches and Pedagogic Solutions

**Bangor University**

Monday and Tuesday, 18-19 July 2011

# Overview

---

- 1) Errors in learner texts
- 2) Target hypotheses
- 3) Annotating in Excel
- 4) Short pause
- 5) Inter-Annotator agreement
- 6) Falko target hypothesis 1
- 7) Automatic error annotation
- 8) Manual error annotation

# Errors in learner texts

## 4 steps in error annotation

- 1) identify error
- 2) classify error
- 3) describe error
- 4) explain error

# Errors in learner texts

---

## Hand-out

- 1) Please read the learner text.
- 2) For the first 3 sentences, please individually classify and count the errors.
- 3) Which error types did you find and how many of each?

# Target hypotheses

---

Typically error annotations suffer from **inconsistencies** with regard to:

- **error count** (identify error)
- **error exponent** (identify error)
- **error domain** (identify error)
- **error level** (identify error / classify error)
- **error type** (classify error / describe error)

# Target hypotheses

How to operationalise error identification:

- Error annotation necessarily is an interpretation of the learner's utterance  
→ target hypothesis

Please write down your target hypothesis of the first three learner sentences.

# Target hypotheses

---

Which linguistic levels did you take into account for the target hypothesis

exponent:

word, phrase, sentence, text

Falko: word (token)

domain:

word, phrase, sentence, text

## Operationalisation of the target hypothesis

On which restrictions could you agree to make your target hypotheses more congruent?

- linguistic level
- error exponent
- error domain



# Target hypotheses

---

Please form two groups

Please annotate the text in the Excel file  
for your target hypothesis.

We developed a simple annotation Add-In for  
Excel to help our annotators.

# Annotation in Excel

	A	B	C	D
1	tok	pos	lemma	ZH
2	Kriminalität	NN	Kriminalität	Kriminalität
3	zahlt	YVFIN	zahlen	zahlt
4	sich	PRF	er es sie	sich
5	nicht	PTKNEG	nicht	nicht
6	aus	PTKVZ	aus	aus
7	.	\$.	.	.
8	Von	APPR	von	Von
9	dieser	PDAT	dies	diesem
10	Titel	NN	Titel	Titel
11	habe	YAFIN	haben	habe
12	ich	PPER	ich	ich
13	verstanden	YVPP	verstehen	verstanden
14	,	\$,	,	,

Text is shown vertically in the "tok" column.

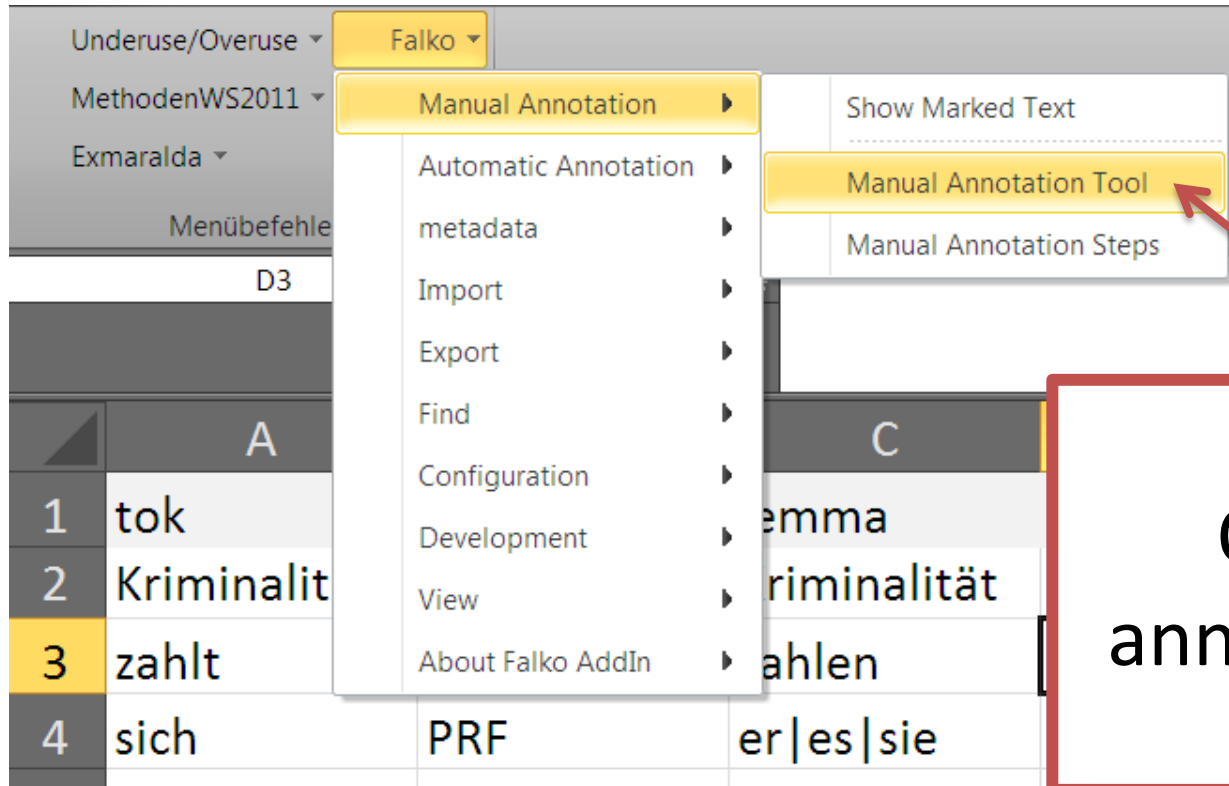
*Kriminalität zahlt sich nicht aus. Von dieser Titel habe ich verstanden, ...*

# Annotation in Excel

Target hypothesis is  
a full text.  
Starts as a copy of  
the learner text  
"tok"

	A	B	C	D
	tok	pos	lemma	ZH
	Kriminalität	NN	Kriminalität	Kriminalität
	zahlt	VVFIN	zahlen	zahlt
	sich	PRF	er es sie	sich
	nicht	PTKNEG	nicht	nicht
	aus	PTKVZ	aus	aus
	.	\$.	.	.
8	Von	APPR	von	Von
9	dieser	PDAT	dies	diesem
10	Titel	NN	Titel	Titel
11	habe	VAFIN	haben	habe
12	ich	PPER	ich	ich
13	verstanden	VVPP	verstehen	verstanden
14	,	\$.	,	,

# Annotation in Excel



The screenshot shows the Falko application interface. The 'Falko' menu is open, and the 'Manual Annotation' option is selected, which has opened a sub-menu. In this sub-menu, the 'Manual Annotation Tool' option is highlighted with a yellow background. A red arrow points from this option to a text box on the right. The background shows a spreadsheet with text in German, such as 'tok', 'Kriminalität', and 'zahlen'.

Open Falko-  
annotation Add-in



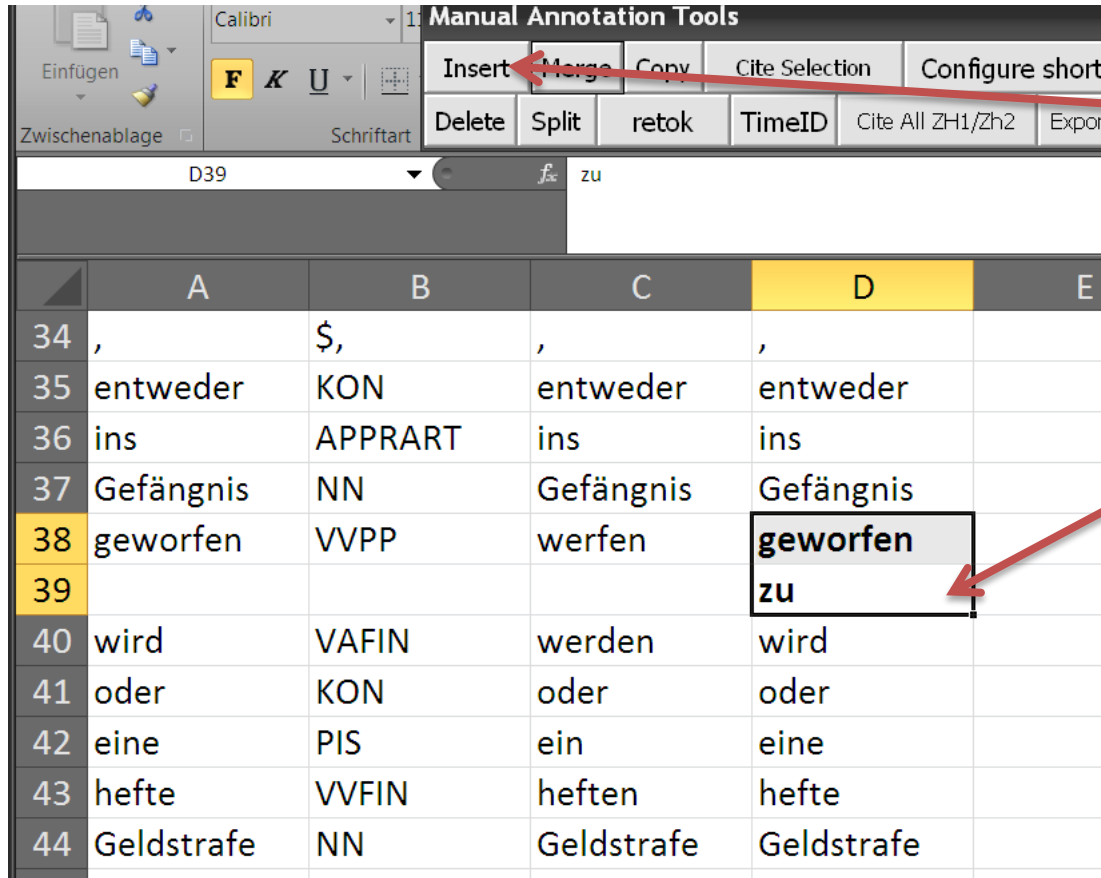
The screenshot shows the 'Manual Annotation Tools' dialog box. It has a title bar with a close button (X). The dialog contains several buttons arranged in two rows. The first row includes 'Insert', 'Merge', 'Copy', 'Cite Selection', 'Configure shortcuts', and 'Close'. The second row includes 'Delete', 'Split', 'retok', 'TimeID', 'Cite All ZH1/Zh2', and 'Export Tier'.

# Annotation in Excel

	A	B	C	D
4	tok	pos	lemma	ZH
5	nicht	PTKNEG	nicht	nicht
6	aus	PTKVZ	aus	aus
7	.	\$.	.	.
8	Von	APPR	von	Von
9	dieser	PDAT	dies	<b>diesem</b>
10	Titel	NN	Titel	Titel
11	habe	VAFIN	haben	habe
12	ich	PPER	ich	ich
13	verstanden	VVPP	verstehen	verstanden

Correct the  
text in the  
"ZH"-column

# Annotation in Excel



The screenshot shows the 'Manual Annotation Tools' ribbon in Excel. The 'Insert' button is highlighted with a red arrow. Below the ribbon, a table is displayed with columns A through E and rows 34 through 44. The cell at row 38, column D, containing the word 'geworfen', is highlighted with a black border. A red arrow points from the 'Insert' button to this cell.

	A	B	C	D	E
34	,	\$,	,	,	
35	entweder	KON	entweder	entweder	
36	ins	APPRART	ins	ins	
37	Gefängnis	NN	Gefängnis	Gefängnis	
38	geworfen	VVPP	werfen	<b>geworfen</b>	
39				<b>zu</b>	
40	wird	VAFIN	werden	wird	
41	oder	KON	oder	oder	
42	eine	PIS	ein	eine	
43	hefte	VVFIN	heften	hefte	
44	Geldstrafe	NN	Geldstrafe	Geldstrafe	

Insert a token via  
click on "Insert" or  
press  
CTRL + L

# Annotation in Excel

- word order:
- 1) Insert new token where item will move to
  - 2) Move token to new line

	A	B	C	D
13	verstanden	VVPP	verstehen	verstand
14	,	\$,	,	,
15	dass	KOUS	dass	dass
16	ein	ART	ein	ein
17	Verbrechen	NN	Verbrechen	Verbrech
18	zu	PTKZU	zu	zu
19	begehen	VVINP	begehen	begehen
20	lohnt	VVFIN	lohn	
21	sich	PRF	er es sie	sich
22	nicht	PTKNEG	nicht	nicht
23				lohnt
24	.	\$.	.	.

# Inter-Annotator-Agreement





# Falko Target Hypothesis 1

On ZH1 only **grammatical errors** are corrected

- orthography
- inflection
- argument structure
- categorical word order errors

**no correction ruled by tendencies (middle field)**

Lexical, semantic, pragmatic errors are ignored

# Falko Target Hypothesis 1

tok	ZH1
Dagegen	Dagegen
ein	
Verbrechen	
zu	
begehen	
bedeutet	bedeutet
	ein
	Verbrechen
	zu
	begehen
auch	auch
die	die
Ermorderung	Ermordung

Minimal corrections  
(instead of "den Mord")



# Falko Target Hypothesis 1

tok	ZH1
Dagegen	Dagegen
ein	
Verbrechen	
zu	
begehen	
bedeutet	bedeutet
	ein
	Verbrechen
	zu
	begehen
auch	auch
die	die
Ermorderung	Ermordung

Correction of ungrammatical syntax

# Falko Target Hypothesis 1

tok	ZH1
Man	Man
soll	soll
	etwas
	über
sein	seinen
Hintergrund	Hintergrund
wissen	wissen

Correction of ungrammatical argument structure

Corrections should conserve the verb over its arguments

# Falko Target Hypothesis 1

tok	ZH1
Man	Man
soll	soll
	etwas
	über
sein	seinen
Hintergrund	Hintergrund
wissen	kennen

Correction of ungrammatical argument structure

Corrections should conserve the verb over its arguments

# Automatic error annotation

CHANGED token

tok	ZH1	ZH1Diff
Daggegen	Dagegen	CHA
ein		MOVS
Verbrechen		MOVS
begehen		MOVS
bedeutet	bedeutet	
	ein	MOVT
	Verbrechen	MOVT
	zu	INS
	begehen	MOVT
eine		DEL
die	die	
Ermorderung	Ermordung	CHA

# Automatic error annotation

tok	ZH1	ZH1Diff
Dagegen	Dagegen	CHA
ein		MOVS
Verbrechen		MOVS
begehen		MOVS
bedeutet	bedeutet	
	ein	MOVT
	Verbrechen	MOVT
	zu	INS
	begehen	MOVT
eine		DEL
die	die	
Ermorderung	Ermordung	CHA

INSERTED token

# Automatic error annotation

tok	ZH1	ZH1Diff
Dagegen	Dagegen	CHA
ein		MOVS
Verbrechen		MOVS
begehen		MOVS
bedeutet	bedeutet	
	ein	MOVT
	Verbrechen	MOVT
	zu	INS
	begehen	MOVT
eine		DEL
die	die	
Ermorderung	Ermordung	CHA

MOVED token  
MOVS<sub>source</sub>  
MOVT<sub>target</sub>



# Automatic error annotation

tok	ZH1	ZH1Diff
Dagegen	Dagegen	CHA
ein		MOVS
Verbrechen		MOVS
begehen		MOVS
bedeutet	bedeutet	
	ein	MOVT
	Verbrechen	MOVT
	zu	INS
	begehen	MOVT
eine		DEL
die	die	
Ermorderung	Ermordung	CHA

DELETED token

# Automatic error annotation

tok	ZH1	ZH1Diff	Ortho
Daggegen	Dagegen	CHA	length
ein		MOVS	
Verbrechen		MOVS	
begehen		MOVS	
bedeutet	bedeutet		
	ein	MOVT	
	Verbrechen	MOVT	
	zu	INS	
	begehen	MOVT	
eine		DEL	
die	die		
Ermorderung	Ermordung	CHA	lex

NOW error  
classification can be  
added to the corpus  
for ...

...orthography

# Automatic error annotation

tok	ZH1	ZH1Diff	Ortho	WO
Daggegen	Dagegen	CHA	length	
ein		MOVS		V3
Verbrechen		MOVS		
begehen		MOVS		
bedeutet	bedeutet			
	ein	MOVT		
	Verbrechen	MOVT		
	zu	INS		
	begehen	MOVT		
eine		DEL		
die	die			
Ermorderung	Ermordung	CHA	lex	

...orthography

...word order

# Automatic error annotation

tok	ZH1	ZH1Diff	Ortho	WO	DEF
Dagegen	Dagegen	CHA	length		
ein		MOVS			
Verbrechen		MOVS			
begehen		MOVS			
bedeutet	bedeutet				
	ein	MOVT			
	Verbrechen	MOVT			
	zu	INS			
	begehen	MOVT		V3	
eine		DEL			
die	die				2ART
Ermorderung	Ermordung	CHA	lex		

...orthography

...word order

...definiteness

# Annotation work shop

Please split into groups of two

Try to annotate the error level of your choice in Excel.

- How many columns do you need?
- Do you need merging of cells?

# Thanks!