

# ‘What’s Hard in German?’: Touching the Void of Over- and Underuse

*funded by*  *and*   
Arts & Humanities  
Research Council

**Cedric Krummes<sup>a</sup>**, Marc Reznicek<sup>b</sup>, Chan Jia Wei<sup>b</sup>,  
Hagen Hirschmann<sup>b</sup>, Thomas Krause<sup>b</sup>, Amir Zeldes<sup>b</sup>,  
**Astrid Ensslin<sup>a</sup>**, Anke Lüdeling<sup>b</sup>

<sup>a</sup>Bangor University, <sup>b</sup>Humboldt-Universität zu Berlin

Contact: [c.krummes@bangor.ac.uk](mailto:c.krummes@bangor.ac.uk), [a.ensslin@bangor.ac.uk](mailto:a.ensslin@bangor.ac.uk)

# Outline

1. Context: Falko and WHiG
2. Data
  - Collection
  - Processing
  - Analysis

# Falko Design » Key Data

- Learners
  - 132,187 tokens
- L1 speakers (control group)
  - 88,730 tokens
- Written texts
  - handwritten
  - computer-typed
- Tasks
  - summaries
  - essays
- Design
  - cross-sectional data
  - longitudinal data

Lüdeling et al. (2008)

Falko

# Falko Design » Key Data

- L1
  - 49 different languages indicated as L1
  - self-evaluation
  - most common L1: Danish, English, French, Polish, Russian
- Advanced Learners
  - external criterium
    - Deutsche Sprachprüfung für den Hochschulzugang (Summaries)
    - 2 years of German Studies (Georgetown)
  - internal criterium
    - C-test (Essays)

# Falko Design » Annotations

- Automatic Annotations
  - using TreeTagger (Schmidt 2001)
  - POS
  - lemmas
- Manual Annotations
  - corrected POS
  - topological fields (Lüdeling 2007, Doolittle 2009)
  - normalisation
  - target hypotheses

# What's Hard in German? (WHiG)

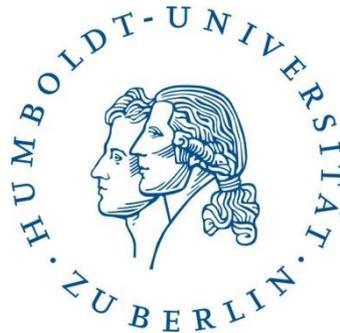
- Finding structural difficulties produced by British learners of German as a foreign language
- 2 project universities
  - Humboldt Universität zu Berlin
  - Bangor University
- 2 funding bodies
  - Deutsche Forschungsgemeinschaft
  - AHRC

WHiG? 

# Research in Falko

- **Contrastive Interlanguage Analysis (CIA)**
  - systematic deviations of interlanguages from the L1 variety of the control group
  - **Overuse & Underuse** (Cobb 2003, Tono 2003, Granger 2008, Walter & Grommes 2008, Mukherjee 2008)
- **Error Analysis**
  - deviation of interlanguages of the norm of the target language
  - error analysis (see Corder 1981)

# Our partners



# Data Collection

- Metadata questionnaire
- C-test
- Essay (in-class)

## C-test (example not used during data collection)

In Österreich müssen alle Kinder ab sechs oder sieben Jahren in die Schule gehen, und zwar vier Jahre lang. Danach können sie vier Jahre in die Hauptschule gehen. Anschließend müssen sie 1-5 Jahre lang eine berufsbildende berufsorientierte Schule besuchen. In dieser Zeit können sie eine Ausbildung oder eine Lehre für einen praktischen Beruf machen. Oder sie gehen nach der Grundschule in eine höhere Schule, das Gymnasium. Da können sie nach 8 Jahren die Matura (das Abitur) machen.

C-test score	CEFR
0 – 29	A1
30 – 39	A2
40 – 59	B1
60 – 79	B2
80 – 89	C1
90 – 100	C2

# Essay (in-class)

- 90 minutes
- Notepad, no language tools
- Choice of 4 topics (general, familiar to students)

Die finanzielle Entlohnung eines Menschen ist offensichtlich eine Frage der Politik und kommt natürlich darauf an, in welcher Gesellschaft und zu welcher Zeit man wohnt. Im sogenannten "Westen" liegt die Entlohnung im Herzen der herrschenden kapitalistischen Übereinstimmung. Was wir hier erörtern werden ist eigentlich die Frage der Rolle und Wichtigkeit einer Entlohnung, und natürlich auch das dazu entsprechende Thema der menschlichen Natur.

Heute spielt den gesellschaftlichen Beitrag eines Menschen kaum eine Rolle bei der Rechnung eines Lohns. Das Gehalt ist normalerweise höher, je lang man arbeitet und besonders an welchem Niveau. Zum Beispiel, ein Angestellter vom Dresdner Bank in Frankfurt wird ohne Frage viel mehr verdienen als ein Geschäftsmann in einer kleinen Stadt in Oberbayern. Der Angestellte der Bank wird vielleicht ein bisschen länger arbeiten und muss bei der Arbeit viel mehr finanzielle Verantwortung tragen, aber im Grunde sind beide sehr fleißig. Unsere Gesellschaft gibt dem einen ein riesiges Gehalt und bewertet den ländlichen Geschäftsmann als niedriger, unwichtiger. Diese Lage ist normal, nicht aussergewöhnlich aber doch unrecht, da die Konsequenzen der Arbeit des Bankangestellten wirken nur Folgen auf die Größe des Geldes irgendeines reichen Mannes, und die Arbeit für den ländlichen Geschäftsmann wirkt Folgen auf das Leben aller Mitglieder seiner Familie.

Wenn wir dieser Unrecht im Zusammenhang der ganzen Gesellschaft beantrachten, kann man sehen, dass in den meisten Fällen, wird man heute mehr bezahlt, desto weniger man für die Gesellschaft leistet. LehrerInnen bilden die wichtigste Gruppe einer Gesellschaft, denn sie die Zukunft davon ist nur durch effektive Ausbildung möglich. Ärzte und Krankenschwestern, bzw. Krankenpflegern, verdienen normalerweise viel weniger als ihre Genossen in der Geschäftswelt, trotz eines vergleichbaren Ausbildungsniveau. Ob sie mehr verdienen sollten ist ja eine Frage der politischen Weltanschauung. Der Moral ist tief im Menschen gewurzelt. Es wird immer einige geben, die sich mit der Verbesserung unserer Gesellschaft beschäftigen wollen werden. Doch ohne die entsprechende Entlohnung wird es sicher weniger Leute geben, die sich geschäftlich aktiv sein würden, wenn sie mehr verdienen könnten als Ärzte oder Lehrern. Was wir normalerweise vergessen ist, dass gesellschaftliche Leistung ist nicht nur eine Sache der Ethik, besonders im christlichen Sinne, sondern auch eine kapitalistische Sache. Man könnte sogar sagen, dass die Förderung einer Verbrauchergesellschaft auch eine gesellschaftliche Leistung ist, besonders wenn die meisten übereinstimmen, dass es das beste System ist.

Man muss für etwas arbeiten, ob Geld, ob Begeisterung, ob religiöse Ziele. Die finanzielle Entlohnung ist eigentlich was wir benutzen, eine schwere Arbeit erträglicher zu machen und ohne die Produkte, die eine Verbrauchergesellschaft herstellt, könnten wir keinen Vorsprung uns leisten. Wir würden nur in die Vergangenheit zurückkehren.

Es nützt eigentlich nicht, vom Kommunismus oder von der marxistischen Ideologie zu sprechen. Bis jetzt hat keine Gesellschaft die nötige Bedingungen dafür geleistet. Eine Gesellschaft in der die finanzielle Entlohnung eines Menschen dem gesellschaftlichen Beitrag entspricht, würde gewiss eine Meritokratie der Ethik sein, denn man würde eine Entlohnung erhalten, für die Arbeit zu welcher man im ethischen Sinne verpflichtet sein sollte. Die menschliche Natur ist meines Erachtens sicher anders - nicht unbedingt negativ, aber immerhin selbstorientiert. Die finanzielle Entlohnung eines Menschen sollte folglich dem Regeln der Ökonomie folgen. Wir haben ein System, das gut funktioniert. Wieso es ändern, wenn man davon keine Verbesserung leisten kann?

Raw_File_Tagged_And_Lemmatized.txt - Editor				
Datei Bearbeiten Format Ansicht ?				
Die	ART	d		
meisten	PIAT	meist		
Univerisitätsabschlüsse	NN	<unknown>		
sind	VAFIN	sein		
nicht	PTKNEG	nicht		
praxisorientiert		ADJD	praxisorientiert	
und	KON	und		
bereiten		VVFIN	bereiten	
die	ART	d		
Studenten		NN	Student	
nicht	PTKNEG	nicht		
auf	APPR	auf		
die	ART	d		
wirkliche		ADJA	wirklich	
Welt	NN	Welt		
vor	PTKVZ	vor		
.	\$.	.		
Sie	PPER	Sie sie sie		
sind	VAFIN	sein		
deswegen		PROAV	deswegen	
von	APPR	von		
geringem		ADJA	gering	
Wert	NN	Wert		
.	\$.	.		
Offensichtlich		ADJD	offensichtlich	
erhalten		VVFIN	erhalten	
immer	ADV	immer		
mehr	PIAT	mehr		
Studenten		NN	Student	
heute	ADV	heute		
ein	ART	ein		
Universitätsabschluß		NN	Universitätsabschluß	
.	\$.	.		
Deswegen		PROAV	deswegen	
ist	VAFIN	sein		
normalerweise		ADV	normalerweise	
die	ART	d		
Frage	NN	Frage		



ADJD

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
[word]	Die	finanzielle	Entlohnung	eines	Menschen	ist	offensichtlich	eine	Frage	der	Politik	und	kommt	natürlich	darauf	an	,	in	welcher	Ges
[pos]	ART	ADJA	NN	ART	NN	VAFIN	ADJD	ART	NN	ART	NN	KON	VVFIN	ADV	PROAV	PTKVZ	\$,	APPR	PRELS	NN
[lemma]	d	finanziell	Entlohnung	ein	Mensch	sein	offensichtlich	ein	Frage	d	Politik	und	kommen	natürlich	darauf	an	,	in	welch	Gese

Keyboard

HIAT + Germanic languages supplement

◌̊	◌̋	◌̌	◌̍	◌̎	◌̏	◌̐	◌̑
◌̒	◌̓	◌̔	◌̕	◌̖	◌̗	◌̘	◌̙
◌̚	((_s))	◌̛	Ä	Å	Æ		
Ð	Ö	Ø	Ü	Þ	ä	å	æ
ð	ö	ø	ü	þ	β		

Search in events

Search string

Search area [word] [pos] [lemma]

Case sensitive search Search 0 items found.

Result

Go to

Next

Save as...

Close

Done.



Madonna-Studien

	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273
[word]	"	Fächer	der	neunzigen	Jahren	.	Obwohl				Madonna-Studien	für	sich	keine	Stelle	bekommen	wird	,
[pos]	\$(	NN	ART	ADJA	NN	\$.	KOUS				NN	APPR	PRF	PIAT	NN	VVPP	VAFIN	\$,
[lemma]	↓>	"	Fach Fächer	d	<unknown>	Jahr	.	obwohl			<unknown>	für	er es sie Sie	kein	Stelle	bekommen	werden	,
[norm]	"	Fächer	der	neunzigen	Jahren	.	Obwohl				Madonna-Studien	für	sich	keine	Stelle	bekommen	wird	,
[macro]																		
[ZH1]	"	Fächer	der	neunziger	Jahren	.	Obwohl			eine	Madonna-Studie	für	sich	keine	Stelle	bekommen	wird	,
[ZH2]	"	Fächer	der	neunziger	Jahren	.	Obwohl	man	mit		Madonna-Studien			keine	Stelle	bekommen	wird	,

Done.

### Search Form

AnnisQL:

Match Count:

More Corpora

<input type="checkbox"/>	Name	Texts	Token	
<input checked="" type="checkbox"/>	Falko_Essay_L2	225	112674	<a href="#">i</a>
<input type="checkbox"/>	pcc-3	3	573	<a href="#">i</a>

[Simple Search](#)
[Query Builder](#)
[Statistics](#)

Context Left:

Context Right:

Results per page:

### Search Result - "denn" (10, 10)

Page 1 of 5 | Token Annotations | Show Citation URL | Displaying Results 1 - 20 of 94

Erzig hat der Feminismus während der Oktoberrevolution überlebt, **denn** die Frauen haben schon die Vorkriegszeit und unter gleichem Na...

[exmaralda](#)  
[Paula](#)  
[Paula Text](#)

[i](#) nur auf der Ebene der Theorie gegeben wird . Was **denn** gemacht werden sollte , sollte ein Schritt weiter , als

[exmaralda](#)

Select Displayed Annotation Levels

lemma	nur	auf	d	Ebene	d	Theorie	geben	werden	.	was	denn	machen	werden	soll
pos	ADV	APPR	ART	NN	ART	NN	VVPP	VAFIN	\$.	PWS	ADV	VVPP	VAINF	VMFI
tok	nur	auf	der	Ebene	der	Theorie	gegeben	wird	.	Was	<b>denn</b>	gemacht	werden	sollt

[Paula](#)  
[Paula Text](#)

[i](#) viele Stellen ( zum Beispiel in den Geschäften , geschweige **denn** in Bauwerk ) leer stehen . Die Studenten lernen Daten

[exmaralda](#)  
[Paula](#)  
[Paula Text](#)

[i](#) , werden wir das Problem wieder treffen , es sei **denn** , das Arbeitsbildungssystem wird grundlich erneuert . Zum Schluss kommt

[exmaralda](#)  
[Paula](#)  
[Paula Text](#)

[i](#) befürworten kann , ist es notwendig zu wissen , was **denn** die wesentlichen Unterschiede zwischen Männern und Frauen sind . Dort

[exmaralda](#)  
[Paula](#)  
[Paula Text](#)

[i](#) . Und die Frage lautet : Wie soll man sich **denn** eigentlich verhalten , wenn man versteht , dass man wegen

[exmaralda](#)  
[Paula](#)  
[Paula Text](#)

[i](#) zu meist die Frauen provoziert und ängstlich macht . Ist **denn** Geld mehr Wert als ein Leben ? Es ist sehr

[exmaralda](#)  
[Paula](#)  
[Paula Text](#)

	A	B	F	H	K	Z	AA
1	word	de	da	en	fr	pl	ru
2	,	0.055799	0.062897	0.064648	0.070923	0.050833	0.080987
61	dann	0.002754	0.002885	0.001459	0.001736		0.002165
62	einer	0.002666	0.001603	0.002465	0.002671	0.001412	0.003465
63	schon	0.002503	0.001154	0.000704	0.000801	0.001977	0.003032
64	da	0.002488	0.001154	0.000503	0.000801	0.000282	
65	Frau	0.002458	0.000769	0.001157	0.00187	0.001694	0.005197
66	muss	0.002384	0.002372	0.002515	0.001469	0.003389	0.007796
67	Geld	0.002369	0.000769	0.002515	0.000534		0.002599
68	viele	0.002281	0.003847	0.002214	0.00187	0.005083	0.002599
69	zum	0.002147	0.001282	0.002666	0.002938	0.000565	0.000866
70	des	0.002132	0.001731	0.003471	0.004408	0.001694	0.001299
71	bei	0.002103	0.001539	0.000755	0.000134	0.000282	0.000433
72	können	0.002088	0.003526	0.002616	0.003206	0.001977	0.001732
73	denn	0.002073	0.001154	0.001006	0.001336	0.00113	
74	Leben	0.002058	0.001411	0.001811	0.002003	0.000847	0.003465
75	sollte	0.00197	0.001026	0.000956	0.001469	0.000847	
76	einem	0.001925	0.001731	0.001459	0.000935	0.001694	0.000866

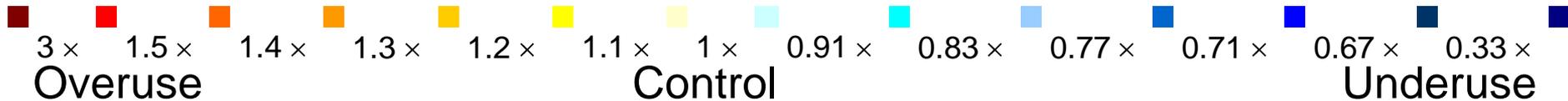


Overuse

Control

Underuse

	A	B	C	D	E	F	G
1	c_pos_bi	de	da	en	fr	pl	ru
2	VVINF-\$.	0,007144789	0,008080549	0,010046296	0,010149024	0,00679558	0,008033562
3	PTKZU-	0,007223675	0,007759892	0,011481481	0,010149024	0,006629834	0,006962421
4	VVINF-\$,	0,007032095	0,008721862	0,009675926	0,012589928	0,00519337	0,006962421
5	\$.-PRELS	0,007764605	0,007054448	0,006759259	0,012204522	0,008839779	0,007230206
6	NN-ADV	0,010638298	0,008850125	0,007037037	0,00668037	0,00718232	0,00642685
7	APPRART-	0,007989993	0,006477265	0,006898148	0,006423433	0,008176796	0,008836919
8	ADV-ADJD	0,008159034	0,008785994	0,008796296	0,007965057	0,005303867	0,004730876
9	\$.-PPER	0,00529661	0,009747964	0,007962963	0,006166495	0,005801105	0,00740873
10	VVFIN-\$,	0,006457357	0,007759892	0,006342593	0,006937307	0,006243094	0,00839061
11	PPOSAT-NN	0,008057609	0,007246842	0,007268519	0,007065776	0,006298343	0,005802017
12	ADV-ADV	0,012858366	0,01051754	0,006111111	0,006166495	0,003093923	0,002856378
13	ADV-APPR	0,009116931	0,008016418	0,005324074	0,007836588	0,00480663	0,004641614
14	PDAT-NN	0,005409304	0,004232669	0,005509259	0,007836588	0,007734807	0,008836919
15	ADV-ART	0,007629373	0,006349003	0,006898148	0,005652621	0,006132597	0,00446309
16	VAFIN-\$.	0,005262802	0,007246842	0,005972222	0,005395683	0,006353591	0,006694635
17	VVFIN-\$.	0,005781194	0,004681588	0,004907407	0,005138746	0,008839779	0,007497992



merci

Wini Davies & the committee, for accepting  
our paper  
the AHRC & the DFG, for their funding  
our partner universities, for their cooperation  
Marc Reznicek, for drafting slides 3 to 8  
the Sprachenzentrum at Humboldt-  
Universität, for our C-test