

Falko (Parent
Corpus)



Summaries
(L1)

Summaries
(L2)

Essays (L1)

Essays (L2)

e.g. Danish
L1

e.g.
Hungarian L1

WHiG

Y2

Y4



Which modern languages do you know? (including your mother tongue[s]) for ancient languages, see other table

Please indicate the name of each language . The number sequence indicates the level of your command of the language: 1 = best command 2 = second-best ... 8 = least command		How old were you when you started using this language? If it is from birth, please put "0".	Do you consider this language your mother tongue ? Tick if yes. Leave unticked if no.	Did you receive instruction in this language? Tick if yes. Leave unticked if no.	If applicable, how many years:months of instruction did you receive in that language? (2:3 = 2 years and 3 months)	If applicable, where did you receive instruction? s = school; u = university; l = language school	Did you spend some time abroad in a country where they speak that language? Tick if yes. Leave unticked if no.	If applicable, how many years:months did you spend in each foreign country? (1:2 = 1 year and 2 months)	If applicable, where did you spend your time abroad using that language?
*	e.g. English	0	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	13 :	<input checked="" type="checkbox"/> s <input type="checkbox"/> u <input type="checkbox"/> l	<input type="checkbox"/>	:	
*	e.g. German	15	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5 : 5	<input checked="" type="checkbox"/> s <input checked="" type="checkbox"/> u <input type="checkbox"/> l	<input checked="" type="checkbox"/>	: 6	Berlin
1			<input type="checkbox"/>	<input type="checkbox"/>	:	<input type="checkbox"/> s <input type="checkbox"/> u <input type="checkbox"/> l	<input type="checkbox"/>	:	
2			<input type="checkbox"/>	<input type="checkbox"/>	:	<input type="checkbox"/> s <input type="checkbox"/> u <input type="checkbox"/> l	<input type="checkbox"/>	:	

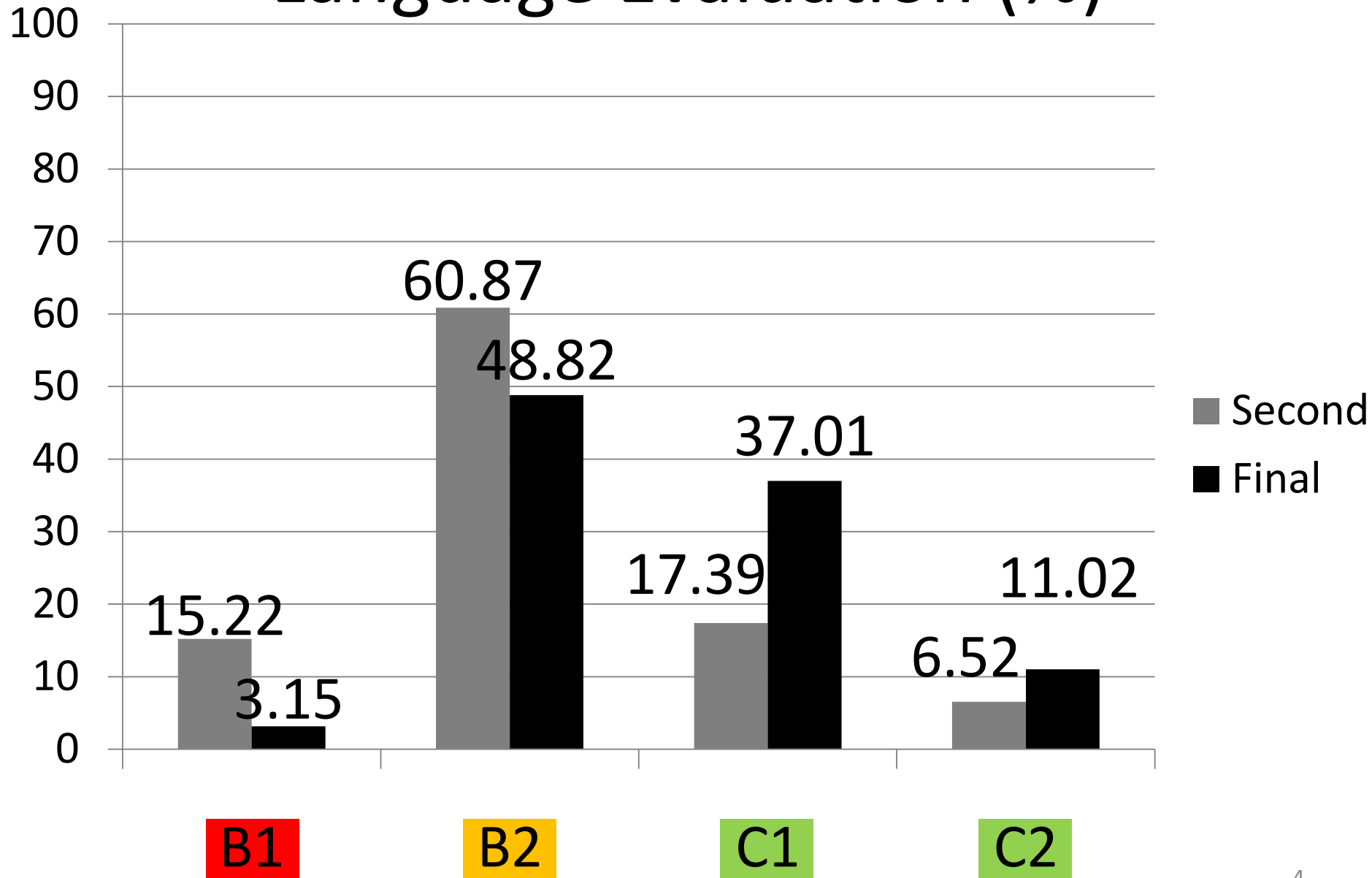
Essay

500 words in 90 minutes

Notepad, no language tools

Choice of 3-4 topics

Language Evaluation (%)



Word Search "Student" in Annis

ANNIS² Tutorial logged in as "demo"

Search Form

AnnisQL: "Student"

Query Builder: Show >>

Result: 76

More Corpora

<input type="checkbox"/>	Name	Texts	Tokens	
<input type="checkbox"/>	FalkoEssayL1V2_0	95	70608	i
<input checked="" type="checkbox"/>	FalkoEssayL2V2_0	248	131599	i
<input type="checkbox"/>	FalkoSummaryL1V1_2	57	21211	i
<input type="checkbox"/>	FalkoSummaryL2V1_2	107	40865	i
<input type="checkbox"/>	FalkoSummaryLV1_0	12	11114	i
<input type="checkbox"/>	HUKonstruktionsgram...	22	19340	i

Search Result - "Student" (5, 5)

Page 1 of 8 Token Annotations Show Citation URL

[i](#) sind sehr hilfreich . Ein **Student** könnte mit Leidenschaft ein Fach sein sehr hilfreich . eine Student können mit Leidenschaft ein Fach
 VAFIN ADV ADJD \$. ART NN VMFIN APPR NN ART NN
 ZHverb (grid)
 falko (grid)

Select Displayed Annotation Levels ▾

lemma	sein	sehr	hilfreich	.	eine	Student	können	mit	Leidenschaft	ein
word	sind	sehr	hilfreich	.	Ein	Student	könnte	mit	Leidenschaft	ein
pos	VAFIN	ADV	ADJD	\$.	ART	NN	VMFIN	APPR	NN	ART
tok	sind	sehr	hilfreich	.	Ein	Student	könnte	mit	Leidenschaft	ein

ZH2 (grid)
 ZH1 (grid)
 text (grid)
 Volltext

[i](#) praxioniert , aber wenn der **Student** gut studieren will , und , aber wenn d Student gut studieren wollen , und
 ADJD \$, KON KOUS ART NN ADJD VWINF VMFIN \$, KON
 text (grid)
 ZHverb (grid)
 falko (grid)
 ZH1 (grid)
 ZH2 (grid)
 Volltext

[i](#) Universität , sondern auch der **Student** (auch) versuchen
 Universität , sondern auch d Student (auch) versuchen
 NN \$, KON ADV ART NN \$(ADV \$(VMFIN
 text (grid)

Target Hypotheses in Annis

Show >>

690

	Texts	Tokens	
2_0	95	70608	i
2_0	248	131599	i
1V1_2	57	21211	i
2V1_2	107	40865	i
LV1_0	12	11114	i
gram...	22	19340	i

Search Result - ZH2Diff="INS"& ZH2lemma="d"& #1=_#2 (5, 5)

Page 1 of 69 Token Annotations Show Citation URL

falko (grid)

Select Displayed Annotation Levels

lemma	machen	.	zum	Schluß	sein	Feminismus	wegen	d	Ungerechtigkeit	in
word	machen	.	Zum	Schluss	ist	Feminismus	wegen	der	Ungerechtigkeiten	in
pos	VVINF	\$.	APPRART	NN	VAFIN	NN	APPR	ART	NN	APPR
tok	machen	.	Zum	Schluss	ist	Feminismus	wegen	der	Ungerechtigkeiten	in

text (grid)

ZH2 (grid)

Select Displayed Annotation Levels

ZH2pos	VVINF	\$.	ADJD	VAFIN	ART	NN	APPR	ART	NN	APPR
ZH2	machen	.	Letztendlich	ist	der	Feminismus	aufgrund	der	Ungerechtigkeiten	in
ZH2lemma	machen	.	letztendlich	sein	d	Feminismus	aufgrund	d	Ungerechtigkeit	in
ZH2Diff			MERGE		INS		CHA			
tok	machen	.	Zum	Schluss	ist	Feminismus	wegen	der	Ungerechtigkeiten	in

ZH1 (grid)

ZHverb (grid)

Volltext

[i](#) der Gesellschaft und besonders in Werbungen , im Fernsehen oder
 d Gesellschaft und besonders in Werbung , im Fernsehen oder
 ART NN KON ADV APPR NN \$, APPRART NN KON

falko (grid)

text (grid)

Formulaic Language: Definitions

- **variety of constructions** : ranging from completely fixed strings of words (e.g. *alles in allem*) to more syntactically, morphologically and lexically flexible structures (e.g. *ich bin der Meinung/ Ansicht/ Überzeugung; es stellt/ erhebt sich die Frage nach.../ob/warum/wie ...*)

Definition I (Psycholinguistics/
Research on language loss):

multi-word sequences that appear to be “prefabricated: that is, **stored and retrieved whole from memory** at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray 2001: 9).

Definition II (Corpus Linguistics):

word combinations (lexical bundles) that occur **frequently** in a language or a language variety and have customary pragmatic and/or discourse functions (Chen & Baker 2010, Biber et al 2004).

- **ubiquitous linguistic phenomenon**:
 - Erman & Warren (2000): 60% of spoken English and more than 50% of written English
 - Altenberg (1998): 80% of natural language

Formulaic Language and Second Language Acquisition

- **distinctive marker of fluency** and central to successful foreign language learning (Hyland 2008, Wray 2001, Howarth 1996).
- **A serious stumbling block.**

- L2 learners **rely heavily on creativity** and produce grammatically well-formed expressions not used by native speakers (Pawley & Syder, 1983).
- L2 learners tend to **underuse native-like expressions** but **overuse** those **word pairs which have direct L1 equivalents** (Granger 1998).
- L2 learners tend to have a **smaller repertoire of collocations** and **restricted collocations** are the most problematic (Nesselhauf 2003).
- non native writers tend to **be stylistically more verbose** and shows **less control of cautious language** (hedging devices) (Chen & Baker 2010).

The most frequent 4-word sequences

FALKO-L1 (11 in total)	WHiG (21 in total)
IN DEN LETZTEN JAHREN (13)	SIND/(ICH) BIN/ICH DER MEINUNG DASS (DIE) + noun/ES (57)
KANN MAN SAGEN DASS (SICH) (13)	AUF DER ANDEREN/ANDERE SEITE (31)
IN DEN MEISTEN FÄLLEN (10)	(MAN) KÖNNTE/KANN MAN SAGEN DASS (29)
AUF DER ANDEREN SEITE (8)	IST ES / ES IST KLAR DASS (19)
AUF DER GANZEN WELT (7)	MEINER MEINUNG NACH IST (ES)/SIND (19)
ICH BIN DER MEINUNG (7)	(ZUM)SCHLUSS GLAUBE ICH (DASS) (12)
IN DER HEUTIGEN GESELLSCHAFT (7)	ZUM BEISPIEL WENN MAN (10)
MEINER MEINUNG NACH IST (6)	IM GROßEN UND GANZEN (8)
STELLT SICH DIE FRAGE (6)	IN DEN LETZTEN JAHREN (8)
IM GROßEN UND GANZEN (5)	IN DER HEUTIGEN GESELLSCHAFT (8)
IN DER HEUTIGEN ZEIT (5)	NICHT ZU LEUGNEN DASS (7)

3-word sequences (Top 11)

FALKO-L1 (83 in total)	WHiG (134 in total)
IN DER (HEUTIGEN)/UNSERER GESELLSCHAFT (81)	ES IST / IST ES To the left: ABER... (21), UND... (19), DESHALB... (10) To the right: ...NICHT (45) ...EIN/EINE (30)...AUCH (18) ...WICHTIG (15)
MEINER MEINUNG NACH (IST) (48)	GIBT ES / ES GIBT To the left: UND... (16), ABER... (15), HEUTZUTAGE... (9) To the right:VIEL/VIELE (56), ...AUCH (40), ...EIN/EINE (35)
GIBT ES / ES GIBT To the left: NATÜRLICH... (6), DENNOCH... (5), DOCH... (5) To the right: ...AUCH (14) ...VIELE (11) ...NOCH (5)	WENN MAN To the left: AUCH... (12) DASS... (12) UND... (12) To the right: EIN/EINE... (52) ETWAS... (13) SICH... (11) DIE... (10)
(ICH BIN) DER MEINUNG/ÜBERZEUGUNG/ANSICHT DASS (39)	AN/IN DER/DIE UNIVERSITÄT/UNI (106)
KANN MAN SAGEN DASS/*DAS (SICH) (35)	MEINER/*MEINE MEINUNG NACH (46)
IN DEN LETZTEN JAHREN (34)	(ABER) (AUF/AN) DER ANDEREN/ANDERE SEITE (39)
WENN MAN TO THE LEFT: UND... (9) TO THE RIGHT: ...SICH (10) ... (AN+HALTEN / MIT+BEFASSEN / VORSTELLEN) ...DIE (5)	(IN/IM) BEZUG AUF (DIE) (38)
IN DEN MEISTEN/VIELEN (FÄLLEN) (15)	ICH/BIN/VIELE LEUTE/ BIN/ICH/SIND DER MEINUNG DASS (ES) (35)
(ES STELLT) SICH (NUN) DIE FRAGE (30)	ICH/GLAUBE GLAUBE/ICH DASS (31)
AUF DER (GANZEN) WELT (25)	KANN/KÖNNTE/KONNTE/MUSS/WILL MAN/NIEMAND SAGEN DASS (28)
(WIE) ZUM BEISPIEL (DIE) (20)	IN DIESEM AUFSATZ (WERDE ICH) (22)

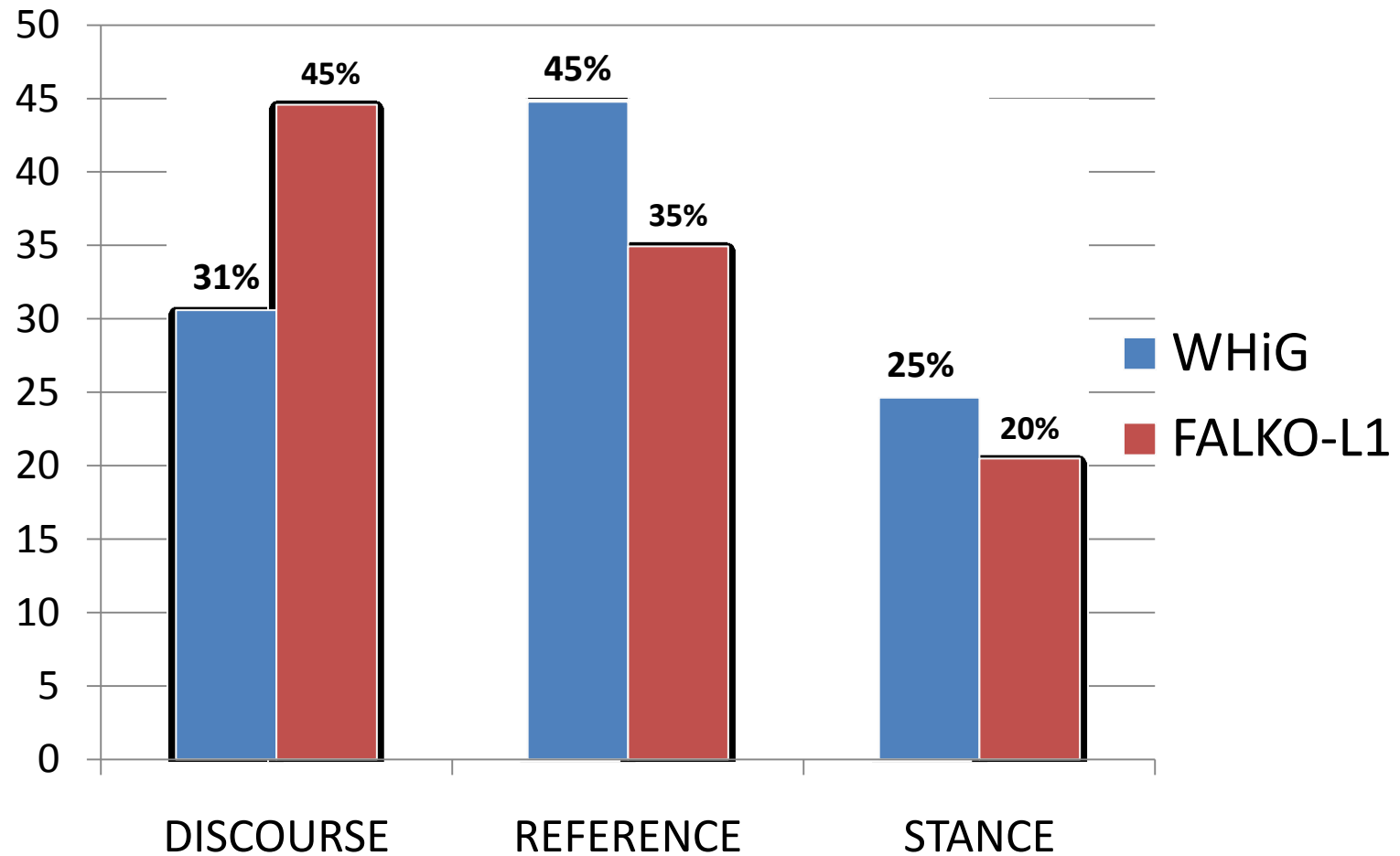
Variation in the ES·GIBT construction

WHiG	und (16) aber (15) heutzutage (9) deshalb (7) Deutschland (6) auch (5) jedoch (5) jetzt (5) Seite (5) Warum (6)	es gibt (203) gibt es (191)	viel/viele (56) auch (40) ein/eine (35) noch (19) immer (14) die (12) aber (11) so (7) jedoch (5) keine (5) natürlich (5)
Falko-L1	natürlich (6) dennoch (5) doch (5) und (5)	gibt es (94) es gibt (88)	auch (14) viele (11) noch (5)

Variation in the ES·IST construction

WHiG	aber (21) und (19) deshalb (10) also (9) deswegen (9) vielleicht (9) heutzutage (8) jedoch (7) (meiner meinung nach) nach (6) heute (5) leider (5)	ist es (263) es ist (201)	nicht (45) ein/eine (30) auch (18) wichtig (15) klar (14) sehr (14) aber (13) oft (13) ganz (12) möglich (9) schwer (7)
Falko-L1	so (7)	ist es (129) es ist (75)	nicht (15) auch (9) also (6) ein (6)

Functions of 3-word sequences



the most frequent 3-word sequences & functions

	DISCOURSE BUNDLES
WHiG	WIE/ WENN ZUM BEISPIEL EIN/MIT/WENN AUF/*AN DER ANDEREN/ANDERE SEITE ABER/UND ES GIBT GIBT ES AUCH/ABER/NOCH AUCH/UND/ABER WENN MAN /SIE/ES/ETWAS ES KOMMT DARAUF ... AN (IN/IM) BEZUG AUF IN DIESEM AUFSATZ (WERDE ICH) ZUM SCHLUSS + KANN/KÖNNTE/MÖCHTE/MUSS/SOLLTE
FALKO-L1	UND ES GIBT/ NATÜRLICH GIBT ES (ABER) AUCH/VIELE/NOCH (ES STELLT) SICH (NUN) DIE FRAGE (WIE) ZUM BEISPIEL (DIE) DIE FRAGE NACH (DEM) (AUF) DER ANDEREN SEITE ALLES IN ALLEM AN DIESER STELLE

the most frequent 3-word sequences & functions

	REFERENCE BUNDLES
WHiG	AN/IN DER/DIE UNIVERSITÄT/UNI IN DER (HEUTIGEN)/UNSERER/*UNSERE GESELLSCHAFT VIELE LEUTE DIE IN DER ARBEITSWELT (MAN) NACH DEM STUDIUM/ DER UNI FÜR VIELE/MANCHE LEUTE IN DER VERGANGENHEIT
FALKO-L1	IN DER (HEUTIGEN)/UNSERER GESELLSCHAFT IN DEN LETZTEN JAHREN IN DEN MEISTEN/VIELEN (FÄLLEN/FÄCHERN) AUF DER (GANZEN) WELT IN DER SCHULE AUCH HEUTE NOCH DIE MEISTEN MENSCHEN IN DEN # (+ YEAR)

the most frequent 3-word sequences & functions

STANCE BUNDLES

WHiG

IST/ES ES/IST SCHWER/WICHTIG/ KLAR
MEINER/*MEINE MEINUNG NACH
ICH/BIN/VIELE LEUTE/ BIN/ICH/SIND DER MEINUNG DASS
ICH/GLAUBE GLAUBE/ICH DASS (47)
KANN (27)/KÖNNTE (29)/MUSS (6)/MAN SAGEN DASS
(UND) MAN/MUSS (NICHT) MUSS/MAN (AUCH)
ES/IST IST/ES WICHTIG/MÖGLICH/NICHT/RICHTIG/EINFACH ZU
SAGEN DASS

FALCO-L1

MEINER MEINUNG NACH (IST)
ICH BIN DER MEINUNG/ÜBERZEUGUNG/ANSICHT DASS
KANN (13)/KÖNNTE (5) MAN SAGEN DASS/ *DAS (SICH)
ICH/DENKE DENKE/ICH DASS (33)
AUF JEDEN FALL
IN DER LAGE
AUF KEINEN FALL

Collocations: “Beispiel”

	Falko-L1	WHiG	
zum Beispiel	0.95	2.12	overuse
beispielsweise	0.47	0.14	underuse
z.B.	0.41	0.31	underuse
nehmen wir beispielsweise	1	0	underuse
nehmen wir {als, das, ein zum} Beispiel	0.06	0.03	underuse
als {zum Beispiel, z.B.} + noun phrase	0	0.04	overuse

Collocations: adjectives + Beispiel

Corpus	adjectives (in common)
Falko-L1	ander , angeführt, bekanntest, best, extrem , genannt, gut , heutig-, hyperbolisch, markant, offensichtlich, prägnant, schlimmer, schön, typisch, weiter, zahlreich, zitiert
WHiG	aktuell, ander , extrem , gut , modern, perfekt, stark, verschieden, wichtig

Conclusion for WHiG

- more formulaic sequences but:
- simple syntactic structures (*es ist .../ es gibt.../ wenn man...*) transferred from English - '**Englisch im deutschen Gewande**'
- less lexical variation (collocations)

WHiG participants do not take morphosyntactic risks.

How can we help them?

- more references, fewer discourse-structuring devices
- more devices structuring macro discourse (*in diesem Aufsatz.../ Zum Schluss*)
very few devices structuring micro discourse (*an dieser Stelle.../ stellt sich die Frage*)
- more cautious language

WHiG participants imagine their essays as a finished product. German native speakers imagine theirs as a sequence of events and an ongoing process (cf. Fandrych & Graefen 2002). **Different paradigms?**



Programme

Symposium **What's Hard in German?**
Structural Difficulties, Research
Approaches and Pedagogic Solutions

Bangor University Monday and Tuesday, 18-19 July 2011

Day One – Monday, 18 July 2011

Time	Session description
0900	Meet & greet, tea & coffee
0930	Astrid Ensslin Welcome and introduction
0945	Anke Lüdeling & Astrid Ensslin Falko and WHiG: Introducing our learner corpora
1045	Tea, coffee, and biscuits
1115	Cedric Krummes & Marc Reznicek Hands-on session: Annotation of the learner data in the Falko corpus
1300	Lunch
1400	Hagen Hirschmann & Anke Lüdeling Hands-on session: Linguistic searches in annotated Falko learner data
1600	Tea, coffee, and biscuits
1630	Astrid Ensslin (moderator) Feedback from tutors and discussion: Student needs and applications
1730	Downtime for delegates
1900	Dinner

To register, please email
c.krummes@bangor.ac.uk

Day Two – Tuesday, 19 July 2011

Time	Session description
0900	Detmar Meurers TBA
1015	Tea, coffee, and biscuits
1030	Sylvia Jaworska & Astrid Ensslin (moderators) Developing German L&T materials from WHiG and Falko
1200	Lunch