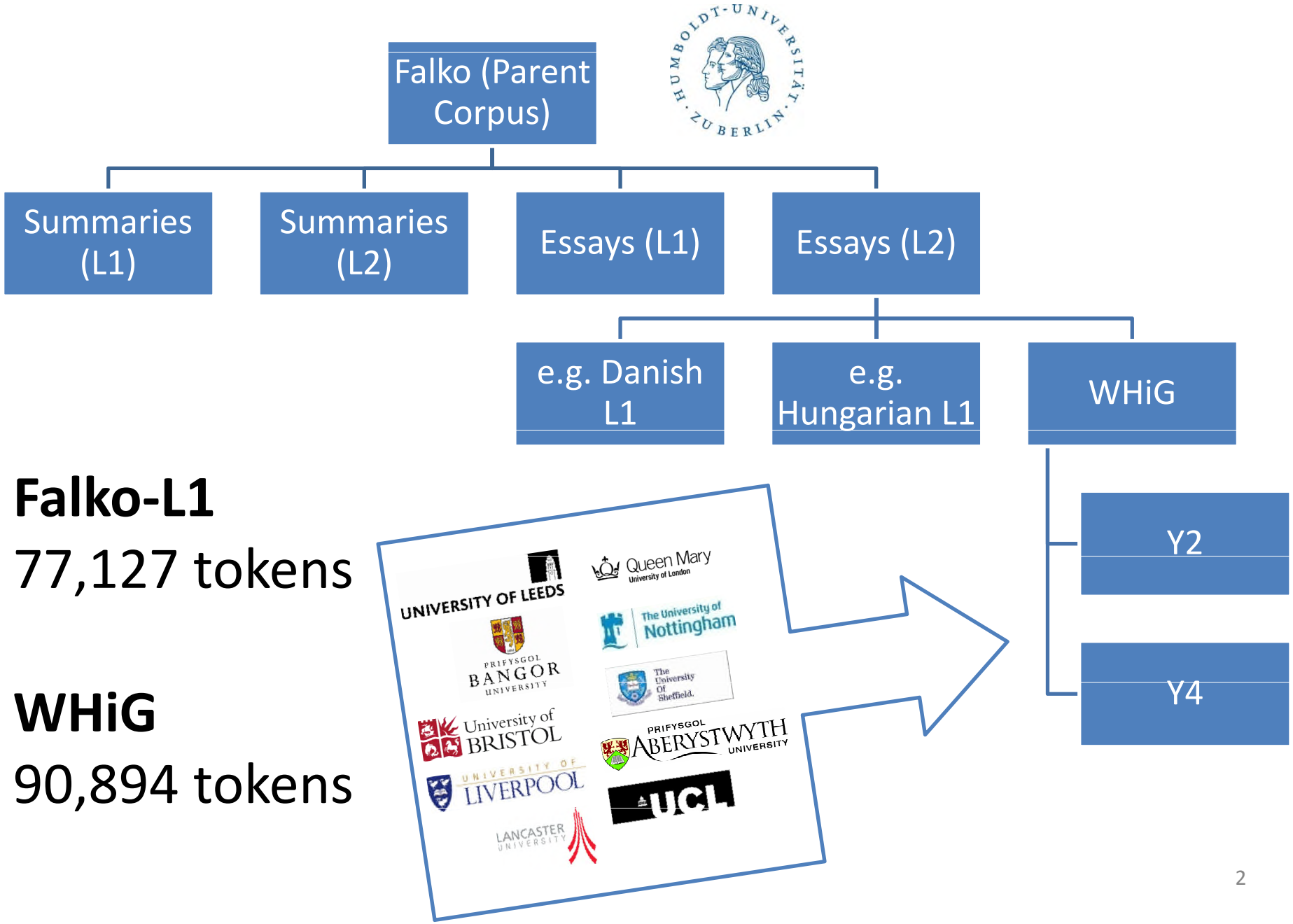
Formulaic Sequences in German Essays: Comparing L1 with L2 data

Cedric Krummes, Sylvia Jaworska, and Astrid Ensslin

WHiG? *funded by* Arts & Humanities Research Council *and* DFG



Falko-L1

77,127 tokens

WHiG

90,894 tokens

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

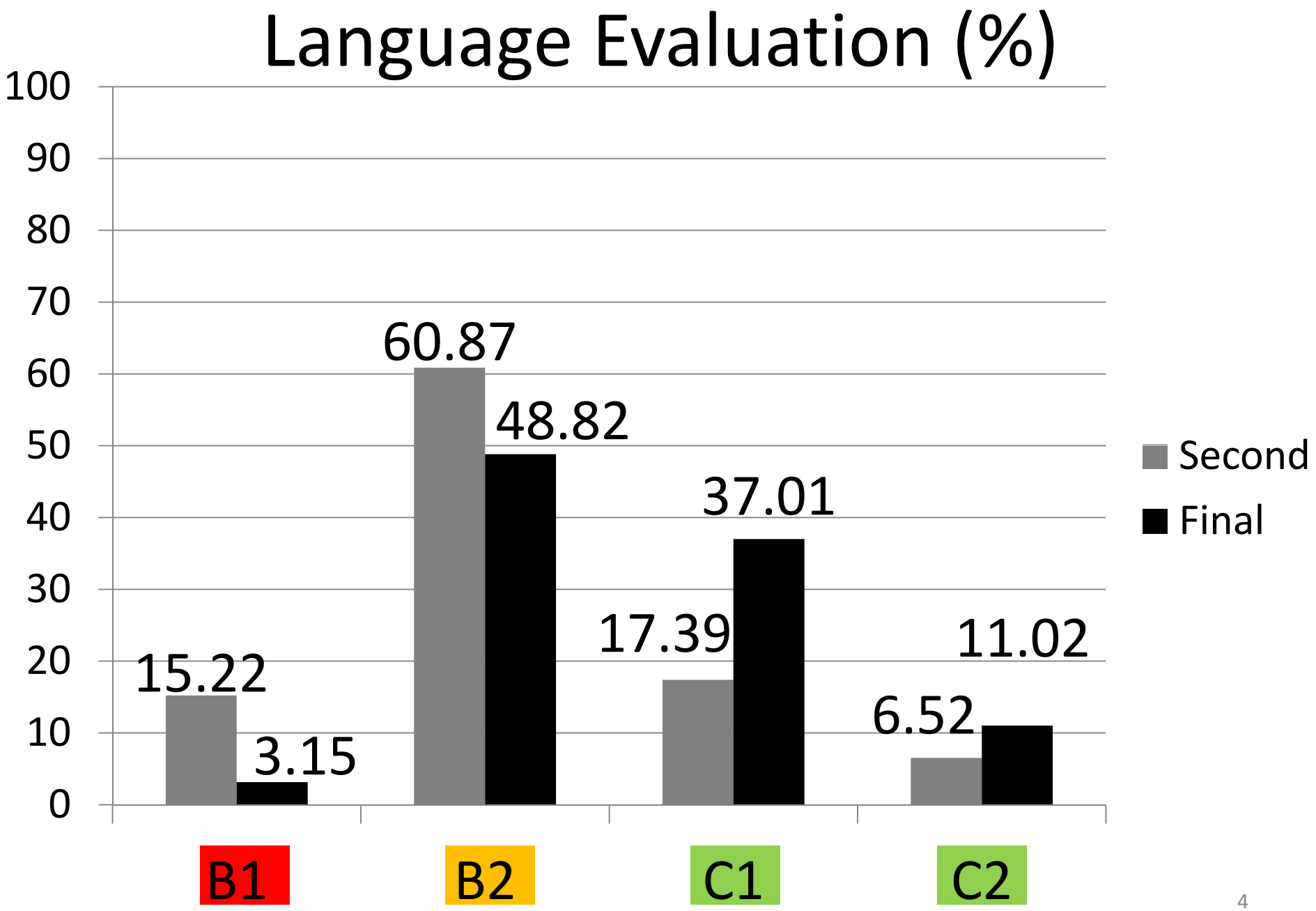
Please indicate the name of each **language.** The number sequence indicates the level of your command of the language: 1 = best command 2 = second-best … 8 = least command		**How old** were you when you started using this language? If it is from birth, please put "0".	Do you consider this language your **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	☑	☑	13 :	☑ s ☐ u ☐ l	☐	:	
*	e.g. German	15	☐	☑	5 : 5	☑ s ☑ u ☐ l	☑	: 6	Berlin
1			☐	☐	:	☐ s ☐ u ☐ l	☐	:	
2			☐	☐	:	☐ s ☐ u ☐ l	☐	:	

Essay

500 words in 90 minutes

Notepad, no language tools

Choice of 3-4 topics


Language Evaluation (%)
100
90
80
70
60
50
40
30
20
10
0
15.22
3.15
60.87
48.82
17.39
37.01
6.52
11.02
B1
B2
C1
C2
Second
Final

Defining Formulaic Languge

- variety of constructions: from fixed expressions to flexible structures
- Psycholinguistics: “multi-word sequences […] stored and retrieved whole from memory” (Wray 2001: 19)
- ★ Corpus Linguistics: frequent word combinations (lexical bundles) (Chen & Baker 2010, Biber et al 2004, Juknevičienė 2009)
- 60% of spoken English and >50% of written English (Erman & Warren 2000)
- 80% of natural language (Altenberg 1998)

Formulaic Language & L2 Learners

- ★ distinctive marker of fluency (Hyland 2008, Wray 2001, Howarth 1996)
- serious stumbling block
- relying heavily on linguistic creativity (Pawley & Syder 1983)
- underuse native-like expressions but overuse those word pairs which have direct L1 equivalents (Granger 1998)
- smaller repertoire of collocations and restricted collocations (Nesselhauf 2003)
- be stylistically more verbose & less control of cautious language (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)

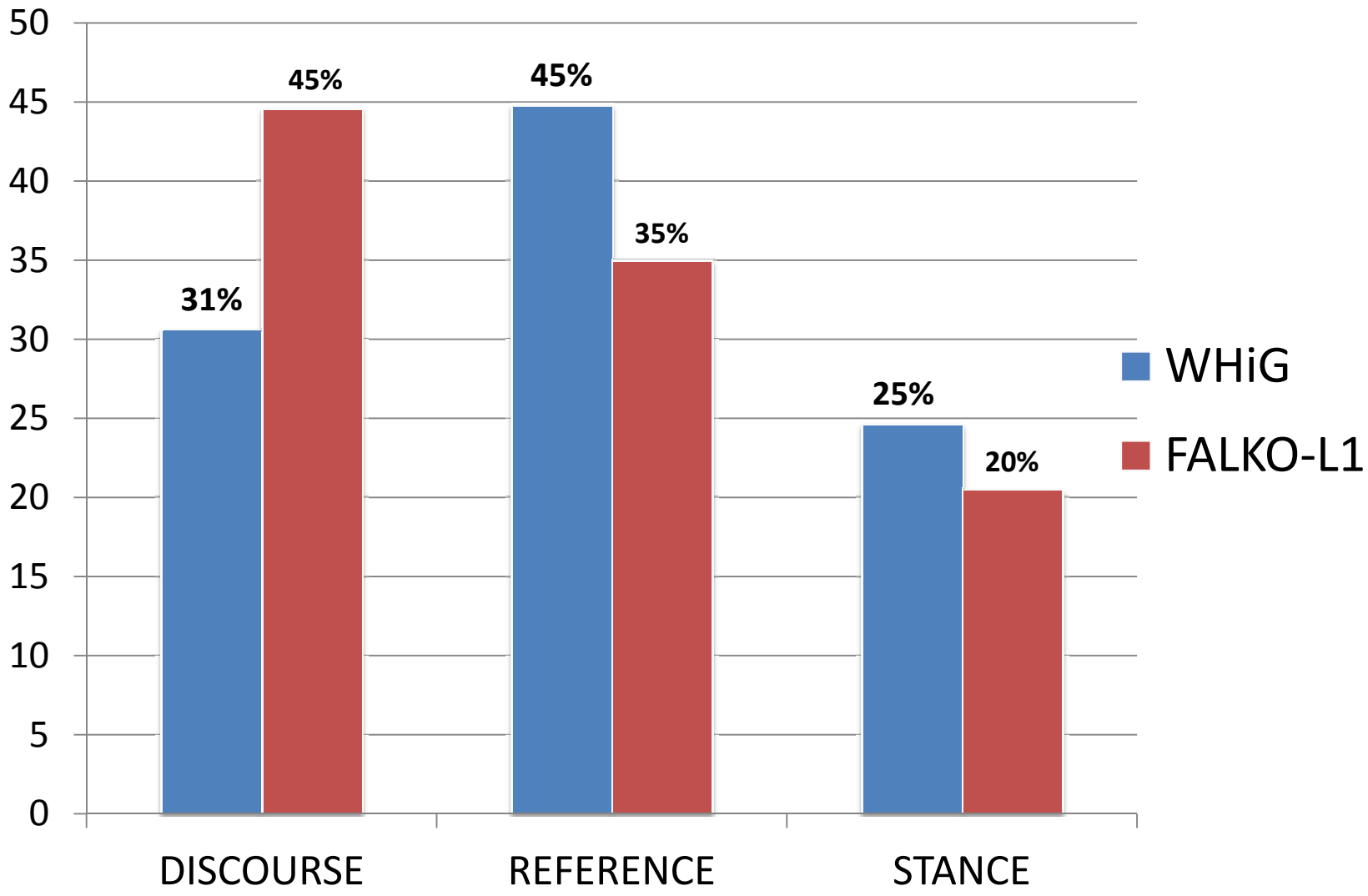
Variation in the ES·GIBT construction

WHiG	und (16) aber (15) heutzutage (9) deshalb (7) Deutschland (6) Warum (6) auch (5) jedoch (5) jetzt (5) Seite (5)	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 (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
FALKO-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

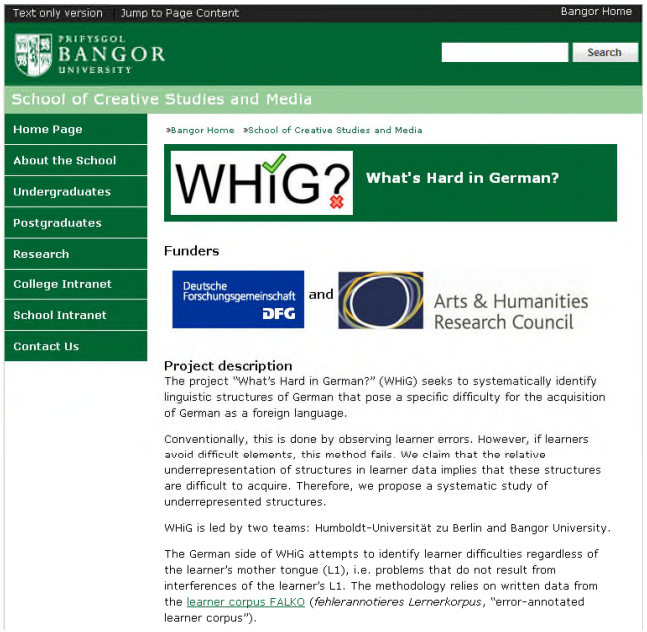
- more formulaic sequences but:
- simple syntactic structures (*es ist …/ es gibt…/ wenn man…*) transferred from English – **"English wine in German bottles"**
- Or: **non-transfer L1-independent** preference for topic-complement constructions
- less lexical variation (collocations)

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?**

Search Engine: whig bangor

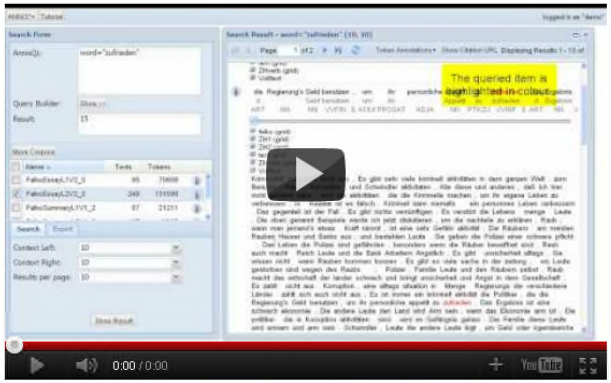


Conference papers and posters

- Krummes, Cedric, Marc Reznicek, Chan Jia Wei, Hagen Hirschmann, Thomas Krause, Amir Zeldes, Astrid Ensslin & Anke Lüdeling (2010) '"What's Hard in German?": Touching the Void of Over- and Underuse', paper given at the Forum for Germanic Language Studies Conference, Gregynog/Aberystwyth, January 8-9, 2010.
- Reznicek, Marc, Cedric Krummes, Hagen Hirschmann, Anke Lüdeling, Astrid Ensslin, Jia Wei Chan, Amir Zeldes, Thomas Krause and Florian Zipser (2010) '"Dass wenn man etwas will, muss man dafür arbeiten"- Zielhypothesen im Lernerkorpus Falko', poster given at 31. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Berlin, February 25, 2010.
- Astrid Ensslin, Cedric Krummes, Marc Reznicek, Anke Lüdeling (2010) 'How to Spot Error and Compare Learners of German: the Falko and WHiG Projects, paper given at the 73rd conference of the Association for German Studies in Great Britain and Ireland, Reading, March 29-31.
- Sylvia Jaworska, Cedric Krummes & Astrid Ensslin (2011). 'Auf der einen Hand und auf die andere Seite: Corpus Evidence of Formulaic Sequences in Essays by Advanced Learners of German', paper given at the 74th conference of the Association for German Studies in Great Britain and Ireland, London, April 13-15. (Slides and Handout)
- Cedric Krummes, Sylvia Jaworska & Astrid Ensslin (2011). 'The Use of Discourse-Structuring Sequences by Advanced Learners of German: Corpus-Driven Approaches', poster given at Corpus Linguistics 2011, Birmingham University, July 20-22 2011.

Invited Lectures

- Krummes, Cedric (2010) 'WHiG: Finding out what's hard in German at the University of Leeds', departmental research seminar, University of Leeds, March 11, 2010.
- Krummes, Cedric (2011) 'Mining *Meinung*: WHiG's Corpus Evidence of German Learner Data', departmental research seminar, Lancaster University, February 03, 2011.
- Krummes, Cedric & Ensslin, Astrid (2011) 'From Learner Essays to Textbook Examples: What's Hard in German?', IntelliText Workshop, University of Leeds, March 11, 2011.
- Ensslin, Astrid, Cedric Krummes and Sylvia Jaworska (2011) 'What's Hard in

Overview of Annis

Overview of Annis results