Underuse of Syntactic Categories in Falko
-
A Case Study on Modification

Hagen Hirschmann
Anke Lüdeling
Ines Rehbein
Marc Reznicek
Amir Zeldes

LEARNER CORPUS RESEARCH 2011
LOUVAIN-LA-NEUVE

research questions & approach

- how can syntactic analyses of L2 learner data help in understanding interlanguage/acquisition processes?
- what is the relationship between lexical elements and syntactic classes?

➢ phenomenon: modification

➢ data: dependency-parsed corpus of advanced L2 learners of German

➢ CIA study (underuse statistics)

Falko

- freely available annotated learner corpus of German as a foreign language
- advanced learners (tutored acquisition)
- written language / controlled, unaided writing
- several text types (sub-corpora);
  here essays (ca. 130000 tokens)
- comparable native speaker corpora (ca. 70000 tokens)
- meta-data for each learner
  (bibliographic data, linguistic history, c-test score)
- Lüdeling et al. (2008), Reznicek et al. (2010),
  http://www.linguistik.hu-berlin.de/institut/professuren/-korpuslinguistik/forschung/falko/standardseite

annotations in Falko

- standoff format (token annotation, span annotation, graphs, pointers etc.), annotation layers can be freely added (Lüdeling et al. 2005)
- learner utterance
  - pos & lemma (automatic, manual correction) (TreeTagger, Schmid 1994)

  ➢ target hypotheses (manual, as many as necessary)
  - pos & lemma
  - error annotation (automatic)
  - parses (dependencies; automatic, manual correction)
  - manual error annotation of some phenomena
  - …

annotation of learner data: conceptual issues

- annotation of learner data is highly problematic
  - data is not systematic according to L1 grammar (especially if there are different L1s)
  - difficult for automatic tools (taggers, parsers)
  - for error analysis and contrastive interlanguage analysis: data has to be interpreted

- Corder (1981), Izumi/Uchimoto/Isahara (2005), Tenfjord/Hagen/Johansen (2004), Diaz-Negrillo et al. (2010) etc.

conceptual problems: pos

- word forms in L2 data sometimes correspond to different pos (Diaz-Negrillo et al. 2010)

Most	important	of	all	was	the	conscious	that
RBS	JJ	IN	DT	VBD	DT	JJ	IN/that
most	important	of	all	be	the	conscious	that

(ICLE)

- every assignment of a pos is an interpretation (*conscious*/NN?JJ → *consciousness*/NN)

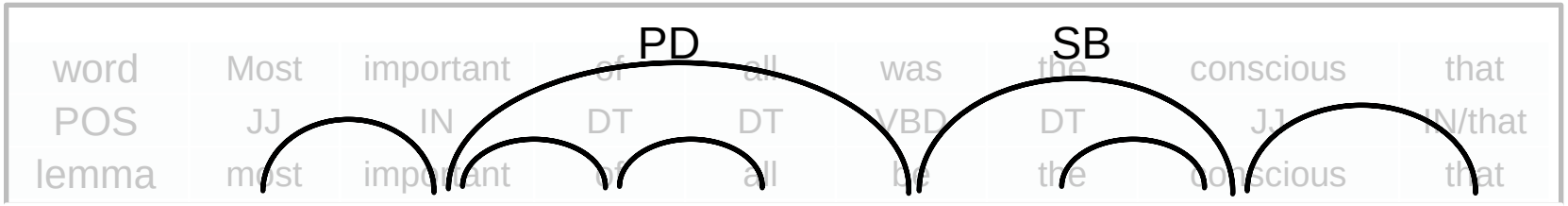
conceptual problems: syntax

Most	important	of	all	was	the	conscious	that
RBS	JJ	IN	DT	VBD	DT	JJ	IN/that
most	important	of	all	be	the	conscious	that

- no possible/useful parse of this structure
- utterance must be transformed into a canonical structure (Hirschmann et al. 2007)

➢ target hypothesis

parsing approach: target hypotheses



- note: conflicting th may be formulated:

TH	Most	important	of	all	was	the	conscious	thought	that
TH_Diff								**INS**	
TH_POS	JJ	IN	DT	DT	VBD	DT	JJ	NN	IN/that

annotation of learner data: target hypothesis in Falko

- th1: sentence-based, very close to original text, mainly ‚genuine' grammatical errors
- th2: text-based, also stylistic errors
- the differences between a target hypothesis and the original data is automatically annotated with edit tags (change, insert, replace etc.)

- (Lüdeling 2011, Reznicek et al. submitted)

target hypotheses …

- are just as necessary for L1 data, btw

research question

- we want to find structural features/problems in German L2 interlanguage
- structural problems are those problems that
  - occur independent of the learners' L1
  - and are therefore attributed to the structure of the target grammar

underuse

- L2 distributions are compared to L1 distributions
- overuse, underuse are defined as (statistically significant) differences between the varieties
- a category can be underused in L2 because
  - the learners do not know it
  - the learners do know it but (unconsciously) avoid it

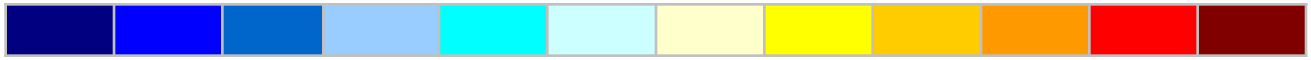
underuse

- L2 distributions are compared to L1 distributions
- overuse, underuse are defined as (statistically significant) differences between the varieties
- a category can be underused in L2 because
  - the learners do not know it
  - the learners do know it but (unconsciously) avoid it → a diagnostics for detecting structural acquisition problems

visualization of overuse and underuse

- underuse: cold colours
- overuse: warm colours
- intensity of colour signals strength of overuse/underuse

Underuse **Overuse**



- Excel add in by Amir Zeldes available at http://korpling.german.hu-berlin.de/~amir/uoaddin.htm

visualization of overuse and underuse: lexical categories

lemma	tot_norm	de	da	en	fr	pl	ru
in	0.013188	0.012261	0.014041	0.014247	0.015272	0.012135	0.009534
es	0.010897	0.011945	0.010900	0.011379	0.013347	0.008163	0.012385
sie	0.010618	0.008193	0.010643	0.008835	0.010909	0.006067	0.005613
man	0.010164	0.007900	0.012438	0.008742	0.009754	0.006950	0.007306
dass	0.009522	0.007404	0.012823	0.008789	0.009625	0.008880	0.009890
von	0.007982	0.007122	0.007309	0.006846	0.007315	0.010259	0.007930
auch	0.007028	0.008362	0.008527	0.005828	0.005775	0.005461	0.004455
für	0.006683	0.007201	0.006091	0.007216	0.006802	0.005736	0.004188
sind	0.006465	0.004271	0.008976	0.007308	0.006930	0.004964	0.005346
sich	**0.006309**	**0.011697**	**0.006283**	**0.006291**	**0.006930**	**0.007170**	**0.005435**
ich	0.006262	0.003877	0.013272	0.005366	0.003465	0.001434	0.001426
aber	0.006048	0.003347	0.007309	0.006245	0.007315	0.003365	0.003831

sich (reflexive pronoun) is underused in all L1 groups

visualization of overuse and underuse: bigrams of pos-categories

bigram	tot_norm	de	da	en	fr	pl	ru
$.-PPER	0.042384	0.005297	0.009748	0.007963	0.006166	0.005801	0.007409
VVFIN-$,	0.042131	0.006457	0.00776	0.006343	0.006937	0.006243	0.008391
PPOSAT-NN	0.041739	0.008058	0.007247	0.007269	0.007066	0.006298	0.005802
ADV-ADV	**0.041604**	**0.012858**	**0.010518**	**0.006111**	**0.006166**	**0.003094**	**0.002856**
ADV-APPR	0.039742	0.009117	0.008016	0.005324	0.007837	0.004807	0.004642
PDAT-NN	0.03956	0.005409	0.004233	0.005509	0.007837	0.007735	0.008837
ADV-ART	0.037125	0.007629	0.006349	0.006898	0.005653	0.006133	0.004463

adverb chains are underused in all L1 groups

modification

- corpus-based studies of adverbs in GFL
  - typically based on lexical items and (rarely) word classes (form-based)
  - typically for one language pair (Möllering 2004, Vyatkina 2007 etc.)
- ADV underuse points to a more general phenomenon: modification

modification

- are the effects form-based or function-based?
  - are all adverbs underused?
  - are certain adverbs (forms) underused?
  - are certain adverbs (forms) underused in certain functions?
  - are certain adverbial functions underused?
  - is modification generally underused?
    (or do learners make up for the underuse of adverbs by other means of modification?)

modification

- are the effects form-based or function-based?
  - are all adverbs underused?
    no; *auch*, *noch* etc. overused
  - are certain adverbs (forms) underused?
    yes
  - are certain adverbial functions underused?
  - are certain adverbs (forms) underused in certain functions?
  - is modification generally underused?
    (or do learners make up for the underuse of adverbs by other means of modification?)

underuse of adverbs: function

- pos tag ADV is not fine-grained enough

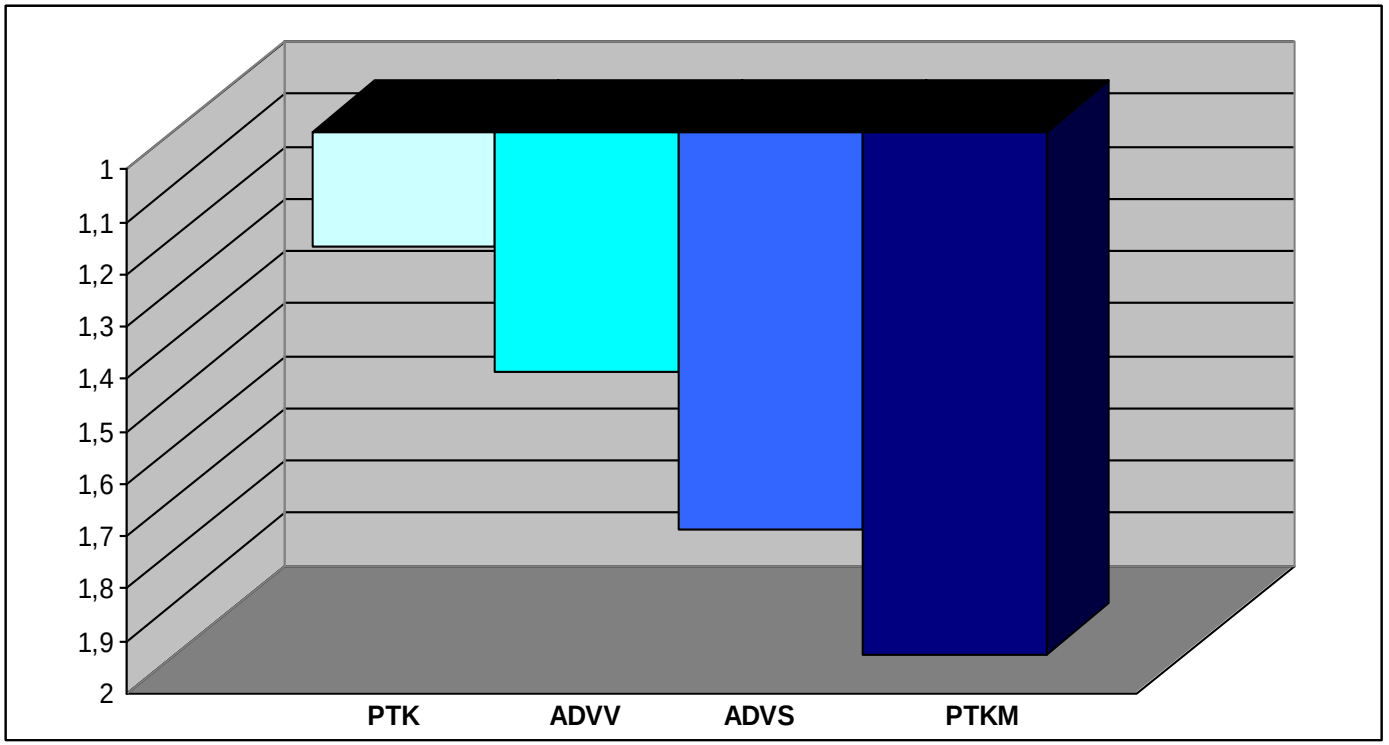
➢ better classification, different functions

  – classes show different distributions

  – only some of these classes are underused by the learners

- Hirschmann (2011, in preparation)

strength of underuse of different syntactic ADV classes



PTK: particles (***sehr*** *gut* - ***very*** *good*)

ADVV: modal adverbs (***Bald*** *schneit es* – ***Soon*** *it will snow*)

ADVS: sentence adverbs (***Bestimmt*** *schneit es bald* – ***Certainly,*** *it will snow soon*)

PTKM: modal particles (*Es schneit* ***wohl*** *gerade* – *It is* ***?apparently?*** *snowing now*)

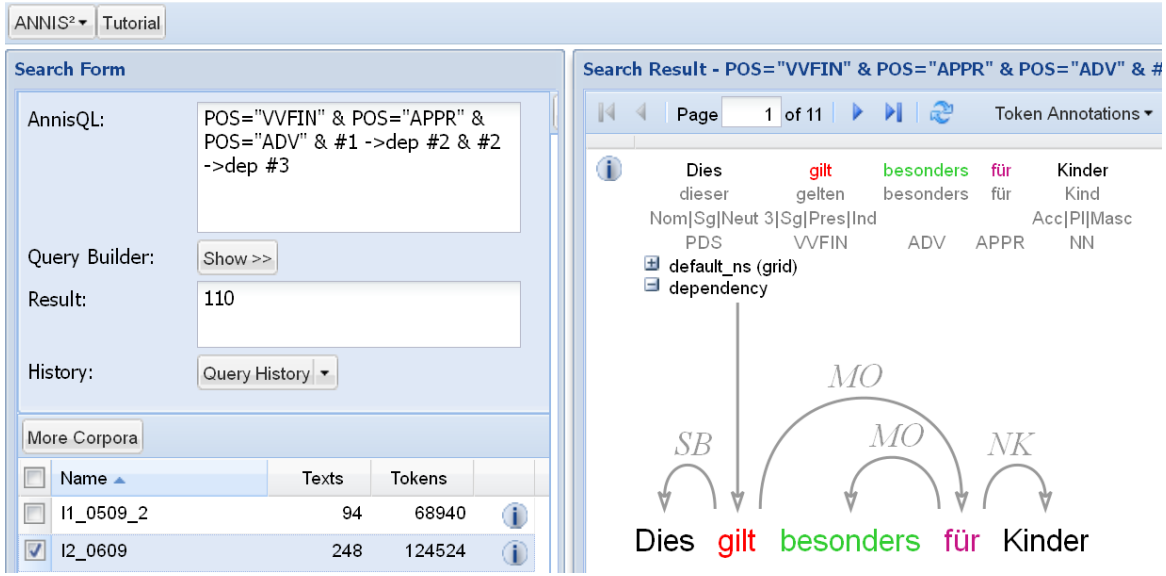
underuse of adverbs: function

- underuse differences between different adverbial functions
- but classification still word based
- compensation strategies?

➢ necessity to code syntactic functions independent of filler category

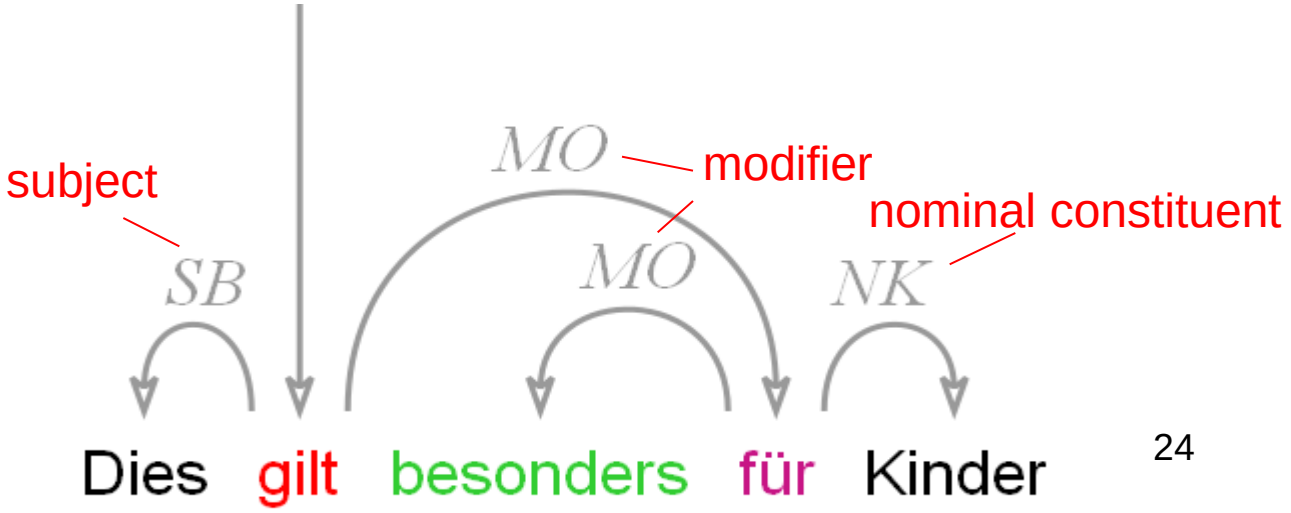
Falko – syntactic annotation

- target hypothesis1 of Falko L1 and L2 corpora
- manually corrected pos tags
- semi-automatic sentence segmentation
- dependency parser by Bernd Bohnet (2010; Syntactic Analyser)
- training data: TiGer dependency bank (derived from ~50000 trees of the TiGer treebank)
- result: very accurate dependency parses with syntactic functions



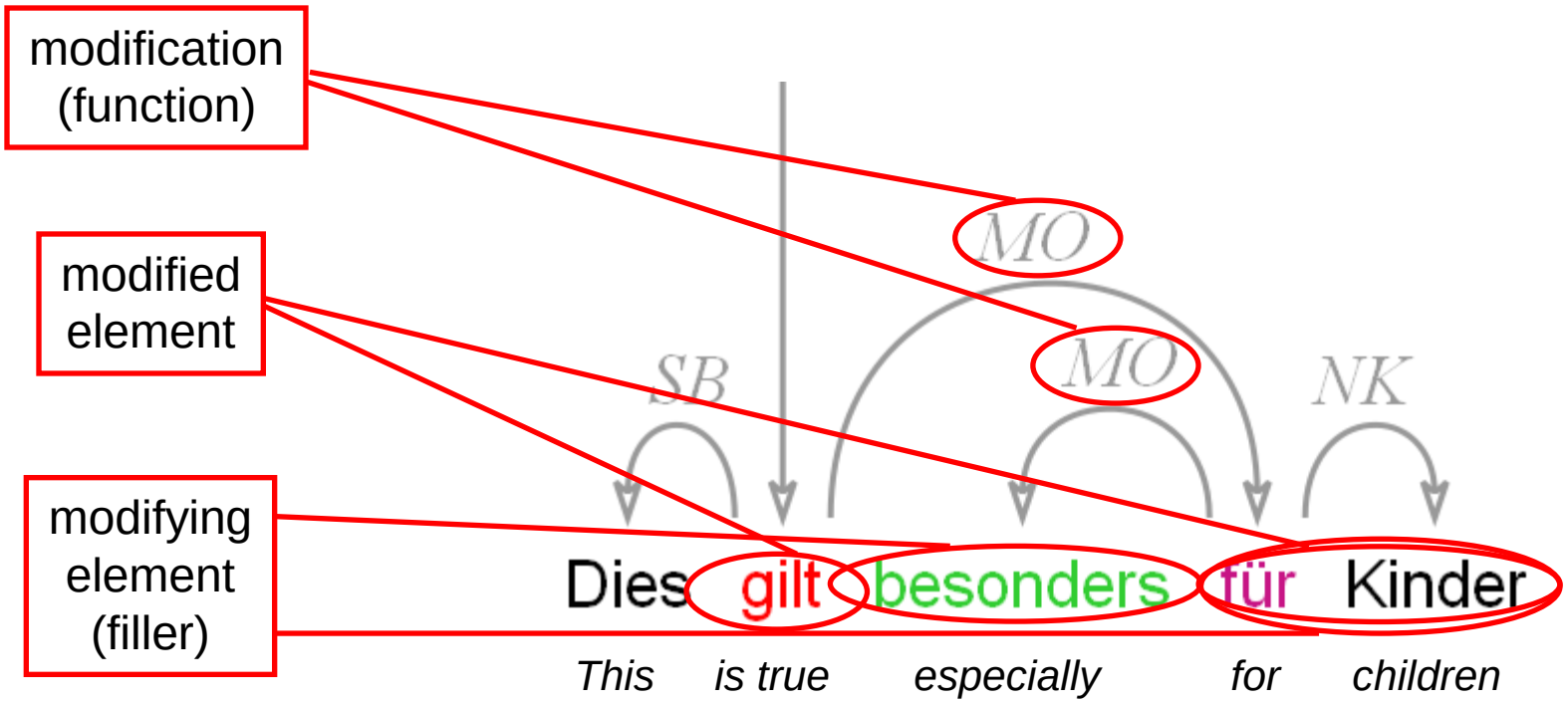
syntax schema (very briefly)

- every word is connected with its dependent(s)
- arrows point to hierachically lower dependent
- each arrow (dependency) has a function label

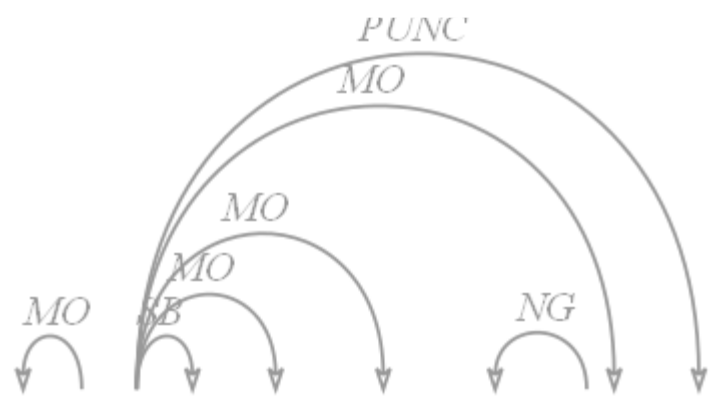


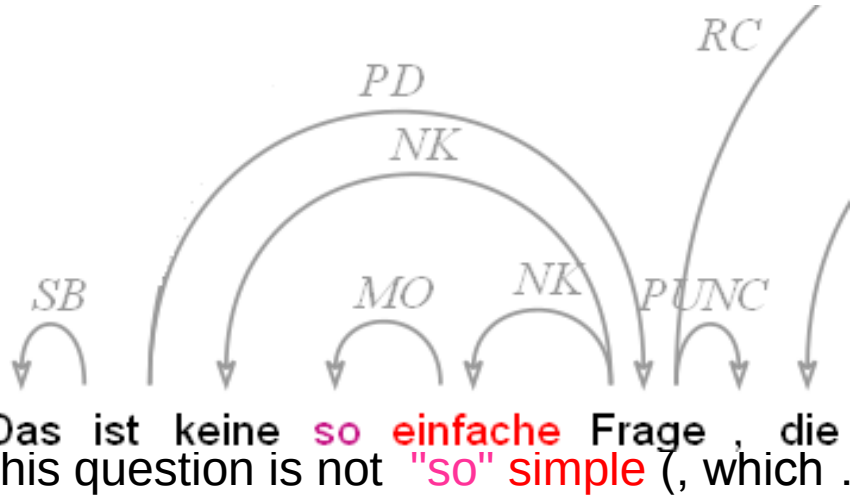
searching for modification in Falko

- different aspects of the problem
  - is the syntactic function ‚modification' underused?
  - what is the target of the modification?
  - what are the categories used for modification?



polyfunctional lexemes: *so*





modification

- are the effects form-based or function-based?
  - are all adverbs underused?
    no; *auch*, *noch* etc. overused
  - are certain adverbs (forms) underused?
    yes
  - are certain adverbs (forms) underused in certain functions?
    yes
  - are certain adverbial functions underused?
  - is modification generally underused?
    (or do learners make up for the underuse of adverbs by other means of modification?)

overuse / underuse of syntactic functions

label	de	da	en	fr	ru	usb
NK	0,264067	0,278546	0,284881	0,303271	0,29552	0,295136
HD	0,156192	0,155622	0,157178	0,154275	0,15809	0,156483
MO	0,141968	0,12789	0,113704	0,110112	0,112513	0,108707
SB	0,07398	0,078506	0,077099	0,075093	0,078852	0,085512
CJ	0,059604	0,053397	0,056411	0,050632	0,059274	0,072183
AC	0,057051	0,059317	0,057215	0,054796	0,054012	0,04916
OC	0,050335	0,053039	0,050008	0,049888	0,047125	0,040679
OA	0,044213	0,042352	0,044097	0,043643	0,046119	0,046218
CD	0,026549	0,024632	0,025639	0,022156	0,024917	0,030466
CP	0,017653	0,021732	0,020325	0,018141	0,017256	0,014887
PD	0,014435	0,014462	0,015943	0,015019	0,016947	0,018002
NG	0,011065	0,011561	0,010914	0,00974	0,00975	0,011252
MNR	0,010995	0,013707	0,013429	0,013383	0,010679	0,009521
RC	0,010051	0,008979	0,009385	0,011375	0,006268	0,005366

overuse / underuse of syntactic functions – significant results

label	de	da	en	fr	ru	usb
NK	0,264067	0,278546	0,284881	0,303271	0,29552	0,295136
HD	0,156192					
MO	0,141968	0,12789	0,113704	0,110112	0,112513	0,108707
SB	0,07398	0,078506			0,078852	0,085512
CJ	0,059604	0,053397	0,056411	0,050632		0,072183
AC	0,057051					0,04916
OC	0,050335					0,040679
OA	0,044213					
CD	0,026549			0,022156		
CP	0,017653	0,021732	0,020325			
PD	0,014435		0,015943		0,016947	0,018002
NG	0,011065					
MNR	0,010995	0,013707	0,013429	0,013383		
RC	0,010051				0,006268	0,005366

MO (modification) is significantly underused independent of L1

modified element

func	L2 (norm)	L1 (norm)	
V	117,635562	139,407446	In my opinion this statement holds .
ADJ	11,8629809	14,5772595	the often very theoretical approach
PREP	4,24891865	6,05986598	especially in Denmark where …
PROADV	0,08497837	0,15264146	…and exactly for this reason …
NEG	1,22368857	2,57964068	Perhaps not when
ADV	2,85527333	5,08296063	Only then do they develop…

frequencies normalized per 1000 edges

modified element – results

- all categories are frequently modified in both L1 and L2
- but *all* syntactic relations possible for modification are underused
- modifiers of adverbs show the strongest underuse

modifiers

func	L2 (norm)	L1 (norm)	
V	14,6162802	12,8218827	If she makes her career, …
PROADV	7,41011413	6,73148841	Some have success [with this] …
COMPARE	0,26343296	0,27475463	One can, as mentioned above …
PREP	44,8600831	48,5857769	To make money on a criminal basis
ADJ	12,7722495	17,5842962	… criminality increases steadily …
ADV	61,8302642	87,7230473	which still exists …

frequencies normalized per 1000 edges

modifier – results

- categories of different complexity (lexemes to sentences) are used for modification; modification is frequent in L2 and L1
- some categories are underused by the learners, two categories are slightly overused
- adverbs and (adverbially used) adjectives show the strongest underuse

modification

- are the effects form-based or function-based?
  - are all adverbs underused?
    no; *auch*, *noch* etc. overused
  - are certain adverbs (forms) underused?
    yes
  - are certain adverbial functions underused?
    yes
  - are certain adverbs (forms) underused in certain functions?
    yes
  - is modification generally underused?
    (or do learners make up for the underuse of adverbs by other means of modification?)
    yes

summary: modification in Falko

- modification is a difficult category for learners of GFL
  - previous evidence: form-based
  - previous hypotheses: ‚transfer', polyfunctionality
- additional syntactic evidence shows the syntactic function ‚modification' is underused, independent of form & independent of L1 of the learners

methodological conclusions

- in annotation separation of form and function necessary
- parsing of learner data necessary to find syntactic functions
- explicit target hypotheses: making interpretation visible and learner language parsable
- multi-layer architectures

Thank you!

Merci!

Danke!

Falko:
http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko

contact: anke.luedeling@rz.hu-berlin.de

analysis of syntactic annotation: modifiers

- certain syntactic classes of adverbs are underused
- adverbs are syntactically analyzed as modifiers
- research question:
  - is adverb underuse due to lexical properties of certain adverbs?
    do learners compensate for this underuse with other means of modifications (e.g. PPs)?
  - or do learners simply underuse modifiers (of any kind) (adverb underuse would then be a result of the general underuse of modifiers)?

summary

- categorization
- additionally, there is a purely syntactic effect: MO is structurally underused by the learners
- why is MO difficult?
  - semantics: now we would have to look at different semantic classes of modification (temporal, local, ….) – further research …
  - word order (topology): placement problems in the German middle field – further research
  - categorial effect: does the complexity of categories play a role?
  - ...

MO – summary

- the lexical ADV underuse is still visible
- additionally, there is a purely syntactic effect: MO is structurally underused by the learners
- why is MO difficult?
  - semantics: now we would have to look at different semantic classes of modification (temporal, local, ….) – further research …
  - word order (topology): placement problems in the German middle field – further research
  - categorial effect: does the syntactic complexity of categories play a role?
  - ...
- in order to abstract away from semantic and word order effects we look at the vorfeld

summary

- reaerch question: how does syntactic annotation of L2 learner data and interpretations of it help in understanding interlanguage/acquisition processes?
- interlanguage ← learner corpus
- underuse as a diagnostic for structural difficulties
- Falko
  - design: advanced learners of German, written, essays, metadata, control group
  - annotation: target hypotheses, automatic edit errors, pos, lemma, more error annotation, syntactic annotation (Berkeley parser) of target hypotheses
  - architecture: multi-layer, standoff, searchable with Annis2

summary – adverbs and modification

- from lexical studies we know that learners underuse adverbs
- modification is also generally underused

➢ combination of factors

➢ syntactic annotation helps us in finding acquisition patterns that combine lexical, categorial, topological and functional properties

interlanguage & data

- further assumption: interlanguage can be researched through the analysis of (naturally occurring) learner data
- one type of data: learner corpora
- analysis
  - error analysis
    → analysis of learner data wrt a 'correct' form
  - contrastive interlanguage analysis (CIA)
    → analysis of the learner data wrt to another corpus

background: interlanguage

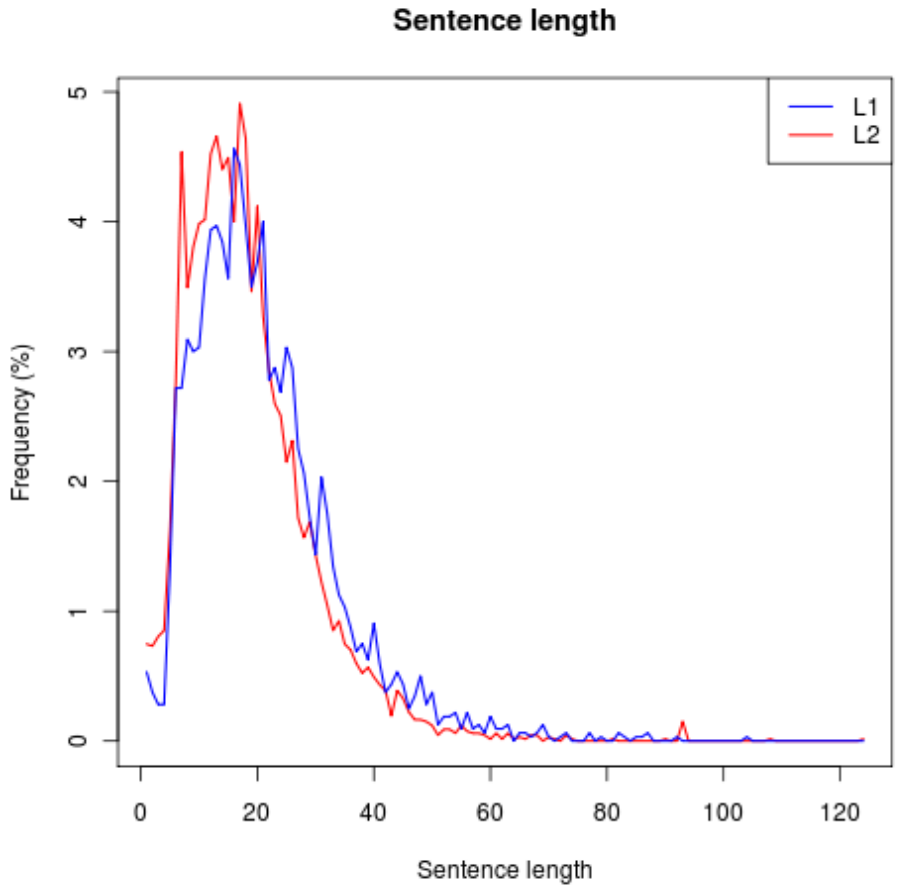
- assumption: learners of a second/foreign language have a systematic internal grammar (interlanguage), different from the internal grammar of L1 speakers of the target language
- interlanguage is influenced by
  - the learners' L1 (transfer, interference)
  - the structure of the L2
  - general learning principles
  - mode of acquisition / teaching method / learning strategies

- Selinker (1972), Nickel (1998) and many others

interlanguage & data

- further assumption: interlanguage can be researched through the analysis of (naturally occurring) learner data
- one type of data: learner corpora
- analysis
  - error analysis
    → analysis of learner data wrt a 'correct' form
  - contrastive interlanguage analysis (CIA)
    → analysis of the learner data wrt to another corpus

comparison of sentence length



data used in the study XXX

- Falko subcorpus
- the largest L1 groups (da: Danish, en: English, fr: French, pl: Polish, rz: Russian)
- 58210 tokens of too small L1s groups (pre-hoc control)

German		L2	
de	88736	da	15593
		en	21600
		fr	7786
		pl	18100
		ru	11203
	88736		74282
total 163018			

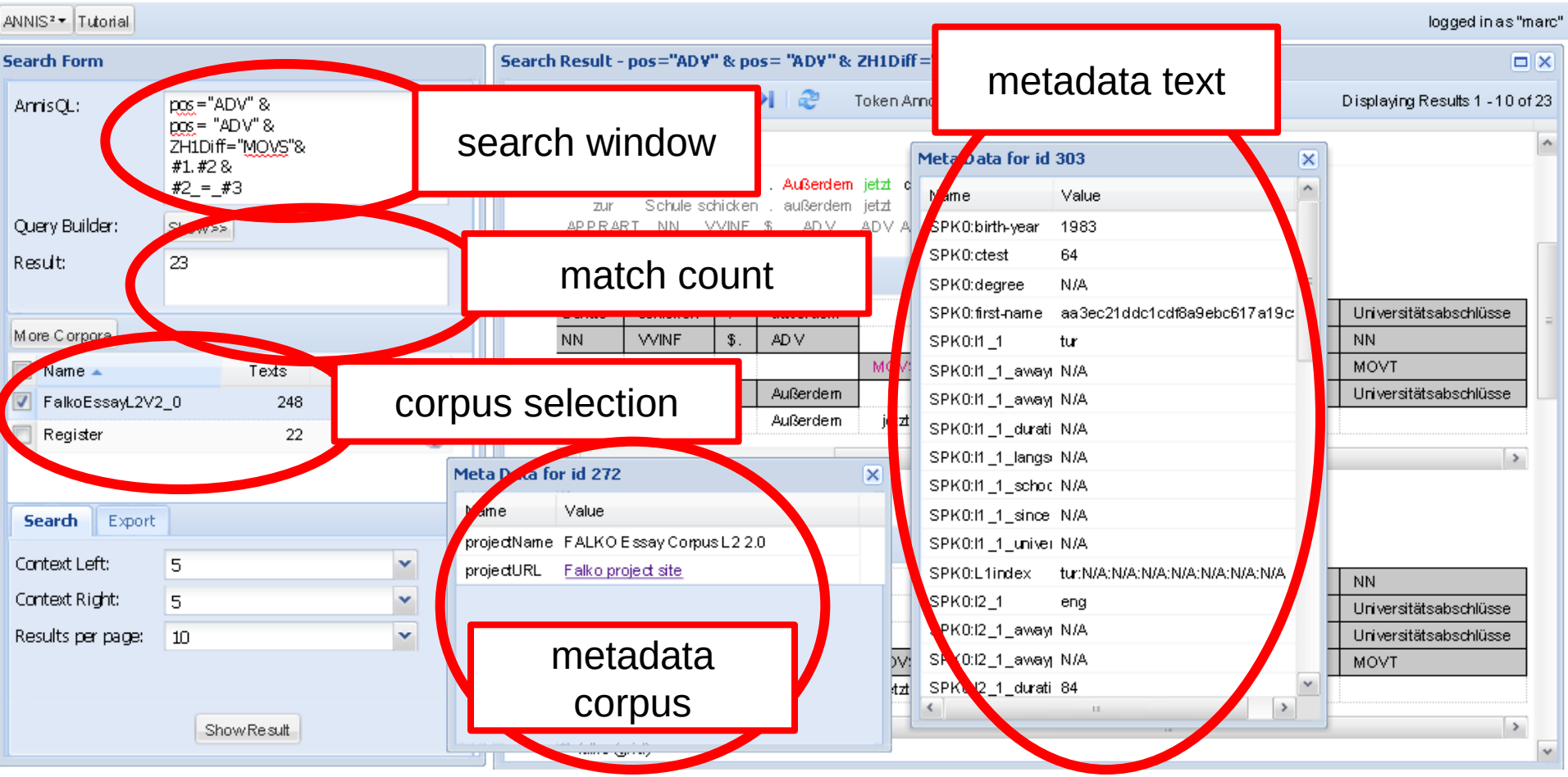
grammatical function in the vorfeld: subject

func_cat/pos	L1 (norm)	L2 (norm)
SB	185,4354354	222,0760826
SB_PPER	84,45945946	120,4977742
SB_PDS	27,4024024	29,54269527
SB_PIS	29,27927928	28,73330635
SB_NN	11,07357357	11,33144476
SB_NE	6,569069069	4,552812626
SB_PWS	3,753753754	3,439902873
SB_NP	142,454955	169,3646297

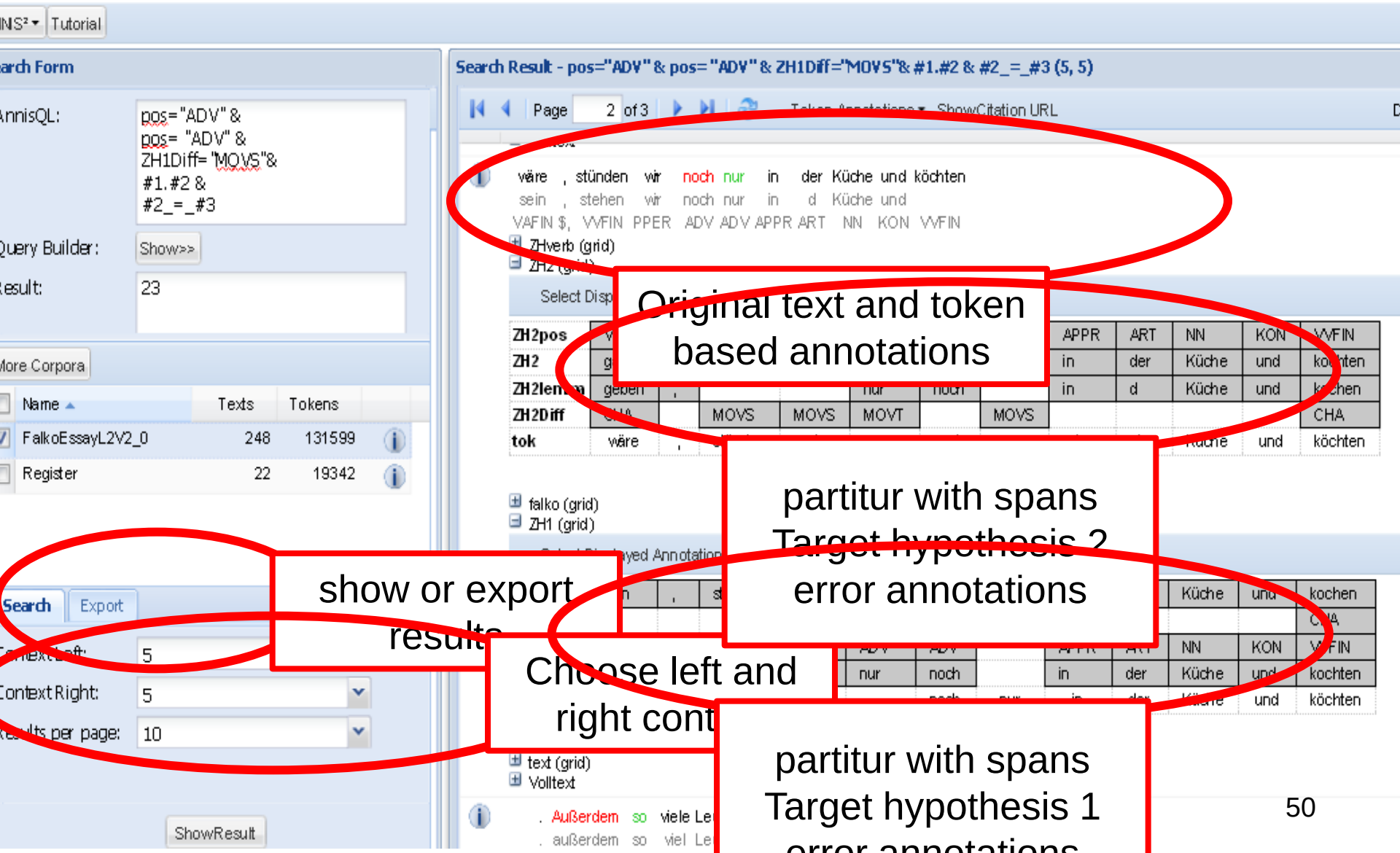
comparison of different subjects in L2 and L1

frequencies normalized per 1000 main clauses

aside: Annis


ANNIS²
Tutorial
logged in as "marc"
Search Form
AnnisQL:
pos="ADV" &
pos= "ADV" &
ZH1Diff="MOVS"&
#1.#2 &
#2_=_#3
search window
Query Builder:
Result:
23
match count
More Corpora
Name
Texts
FalkoEssayL2V2_0
248
Register
22
corpus selection
Search
Export
Context Left:
5
Context Right:
5
Results per page:
10
Show Result
Search Result - pos="ADV" & pos= "ADV" & ZH1Diff=
Displaying Results 1 - 10 of 23
Meta Data for id 272
Name
Value
projectName
FALKO Essay Corpus L2 2.0
projectURL
Falko project site
metadata corpus
Meta Data for id 303
Name
Value
SPK0:birth-year
1983
SPK0:ctest
64
SPK0:degree
N/A
SPK0:first-name
aa3ec21ddc1cdf8a9ebc617a19c
SPK0:l1_1
tur
SPK0:l1_1_away
N/A
SPK0:l1_1_away
N/A
SPK0:l1_1_durati
N/A
SPK0:l1_1_langs
N/A
SPK0:l1_1_schoc
N/A
SPK0:l1_1_since
N/A
SPK0:l1_1_univer
N/A
SPK0:L1index
tur:N/A:N/A:N/A:N/A:N/A:N/A:N/A
SPK0:l2_1
eng
SPK0:l2_1_away
N/A
SPK0:l2_1_away
N/A
SPK0:l2_1_durati
84
metadata text

aside: Annis



Original text and token based annotations

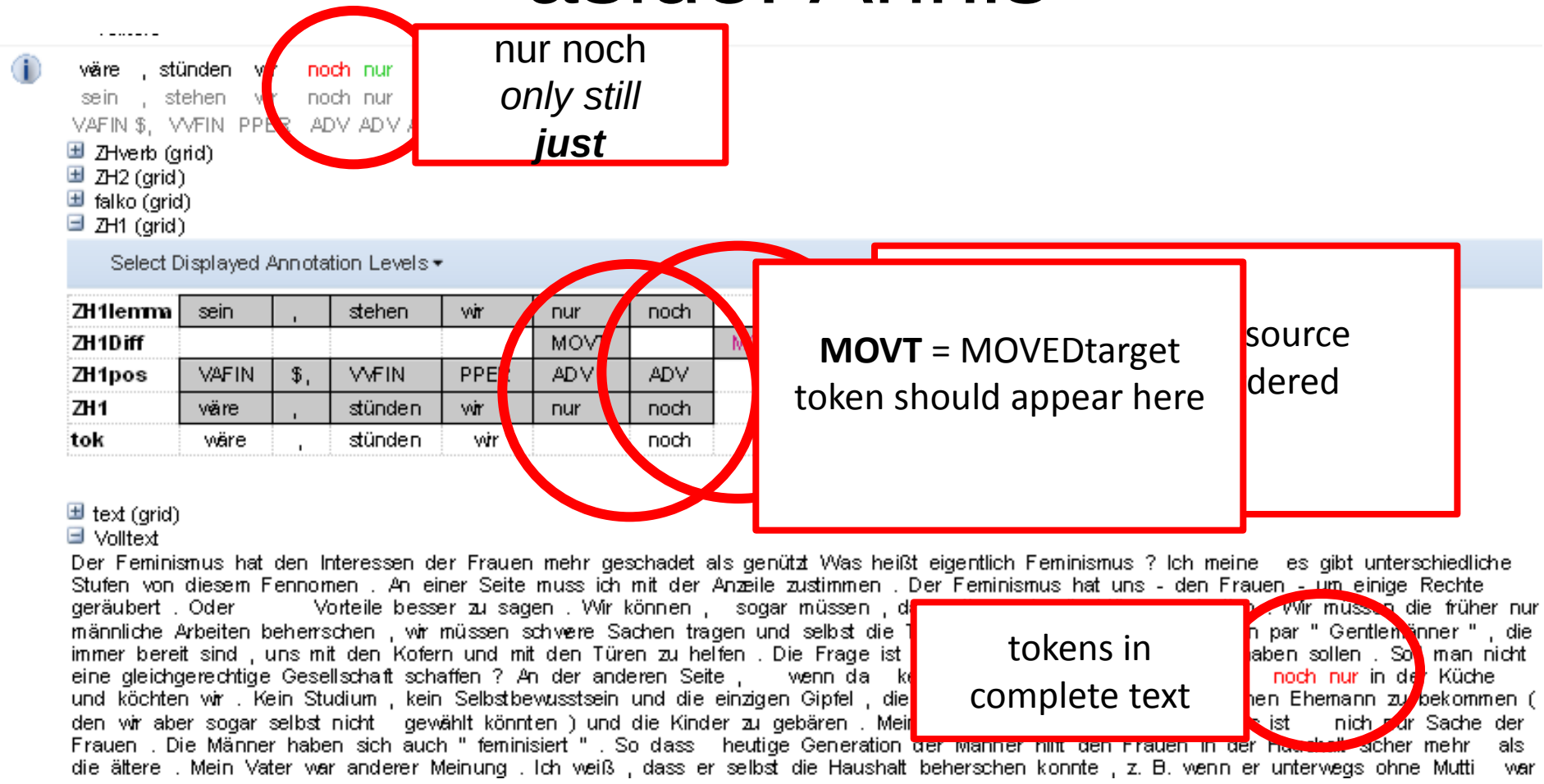
partitur with spans Target hypothesis 2 error annotations

show or export results

Choose left and right cont

partitur with spans Target hypothesis 1 error annotations

aside: Annis



ZH1lemma	sein	,	stehen	wir	nur	noch	
ZH1Diff					MOVT		M
ZH1pos	VAFIN	$,	VVFIN	PPE	ADV	ADV	
ZH1	wäre	,	stünden	wir	nur	noch	
tok	wäre	,	stünden	wir		noch	

Der Feminismus hat den Interessen der Frauen mehr geschadet als genützt Was heißt eigentlich Feminismus ? Ich meine es gibt unterschiedliche Stufen von diesem Fennomen . An einer Seite muss ich mit der Anzeile zustimmen . Der Feminismus hat uns - den Frauen - um einige Rechte geräubert . Oder Vorteile besser zu sagen . Wir können , sogar müssen , d . Wir müssen die früher nur männliche Arbeiten beherrschen , wir müssen schvere Sachen tragen und selbst die T n par " Gentlemänner " , die immer bereit sind , uns mit den Kofern und mit den Türen zu helfen . Die Frage ist aben sollen . So man nicht eine gleichgerechtige Gesellschaft schaffen ? An der anderen Seite , wenn da k noch nur in der Küche und köchten wir . Kein Studium , kein Selbstbewusstsein und die einzigen Gipfel , die nen Ehemann zu bekommen (den wir aber sogar selbst nicht gewählt könnten) und die Kinder zu gebären . Mei ist nich Sache der Frauen . Die Männer haben sich auch " feminisiert " . So dass heutige Generation der Männer hilft den Frauen in der H sicher mehr als die ältere . Mein Vater war anderer Meinung . Ich weiß , dass er selbst die Haushalt beherschen konnte , z. B. wenn er unterwegs ohne Mutti war

http://korpling.german.hu-berlin.de/falko-suche

Parser Evaluation on L1/L2

Evaluation of contituent structure with GF labels (evalb)

	Precision	Recall	F-Score	Tagging acc.
L1	63.93	64.27	64.10	91.93
L2	68.16	69.44	68.79	92.85
Tiger*	69.23	70.41	69.81	

Tiger* Berkeley results on the Tiger Treebank (Petrov & Klein, 2008)

vorfeld

- it is often assumed that in German only one constituent is allowed before the finite verb
  (V2-constraint, vorfeld-constraint)
- the vorfeld is often studied in learner language
  (indication of advancedness, information structure)
- in Falko: there is no significant difference in the vorfeld *complexity* between L1 and L2 –
  but do learners and native speakers use the same elements in the vorfeld?

➢ combination of topological information, functional information and categorial information

- Haberzettl (1998), Walter, Doolittle & Schmidt (2007)

elements in the vorfeld (independent of function)

cat/pos	L1 (norm)	L2 (norm)
NP	152,965465	179,481991
NN	10,3228228	11,3314448
PP	63,2507508	90,0445164
PPER	96,4714715	130,210441
ADV	89,9024024	75,9813841
AVP	15,2027027	10,0161878
PDS	28,3408408	27,9239174
PIS	28,5285285	27,7215702
AP	3,19069069	3,43990287

comparison of vorfeld-elements in L2 and L1

frequencies normalized per 1000 main clauses

modifiers in the vorfeld

MO	207,7702703	223,6948604
MO_PP	63,43843844	90,95507892
MO_AVP	15,2027027	10,01618778
MO_ADV	91,02852853	76,79077297
MO_PROAV	21,02102102	29,0368272
MO_AP	2,064564565	1,214083367
MO_ADJD	6,193693694	6,475111291
MO_PWAV	6,193693694	5,261027924

comparison of different modifier categories in L2 and L1

frequencies normalized per 1000 main clauses

summary: modification in the vorfeld

- learners generally use modification in the vorfeld as often as the native speakers
- learners have a different distribution of elements in the vorfeld
  - they overuse PPs (although they slightly underuse modifying PPs generally)
  - they also significantly overuse simple personal pronouns
  - the same categories (adverbs, adverbial phrases) that are underused everywhere as modifiers are also underused in the vorfeld
  - learners 'compensate' this by overusing prepositional phrases and pronominal adverbs
- syntactic complexity does not seem to be the relevant category

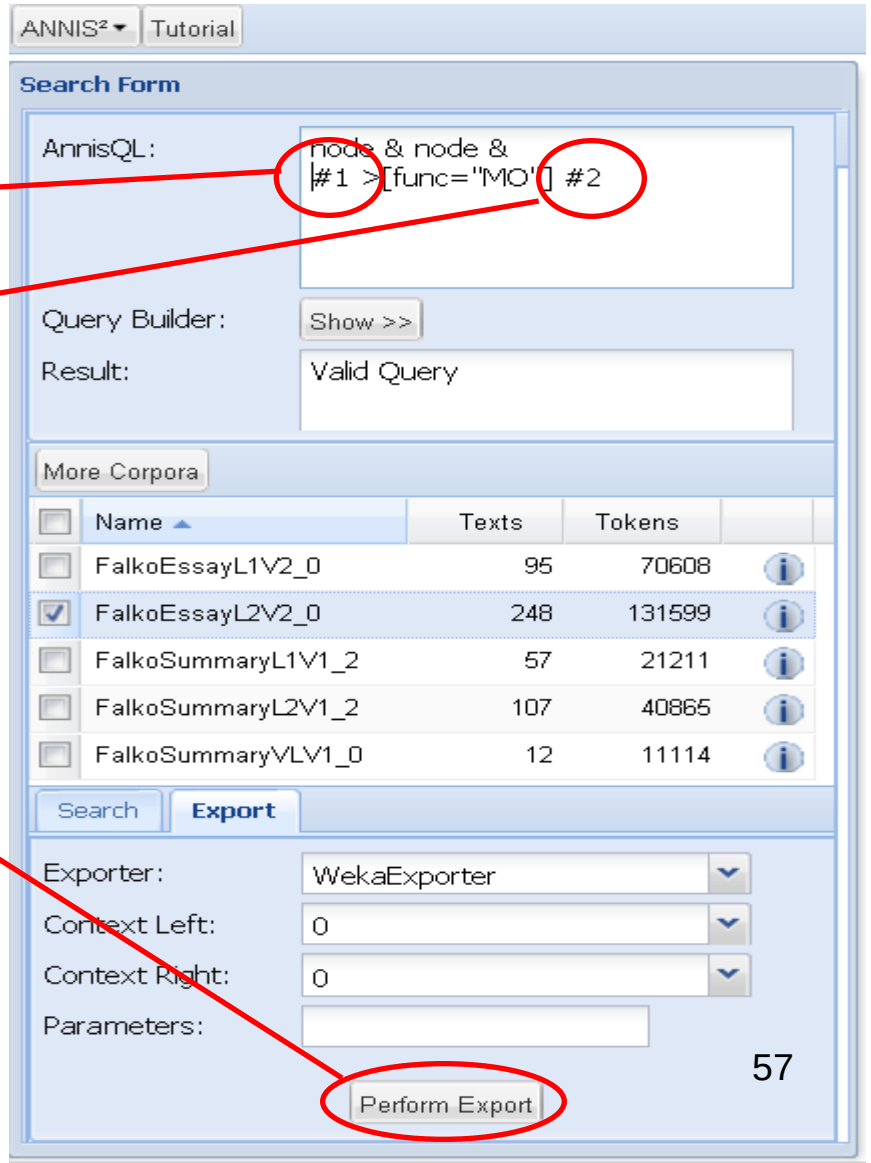
➢ back to lexical and semantic factors ...
➢ further studies: other topological areas in the sentence

aside: annis search&statistics

represents modified element

represents modifier

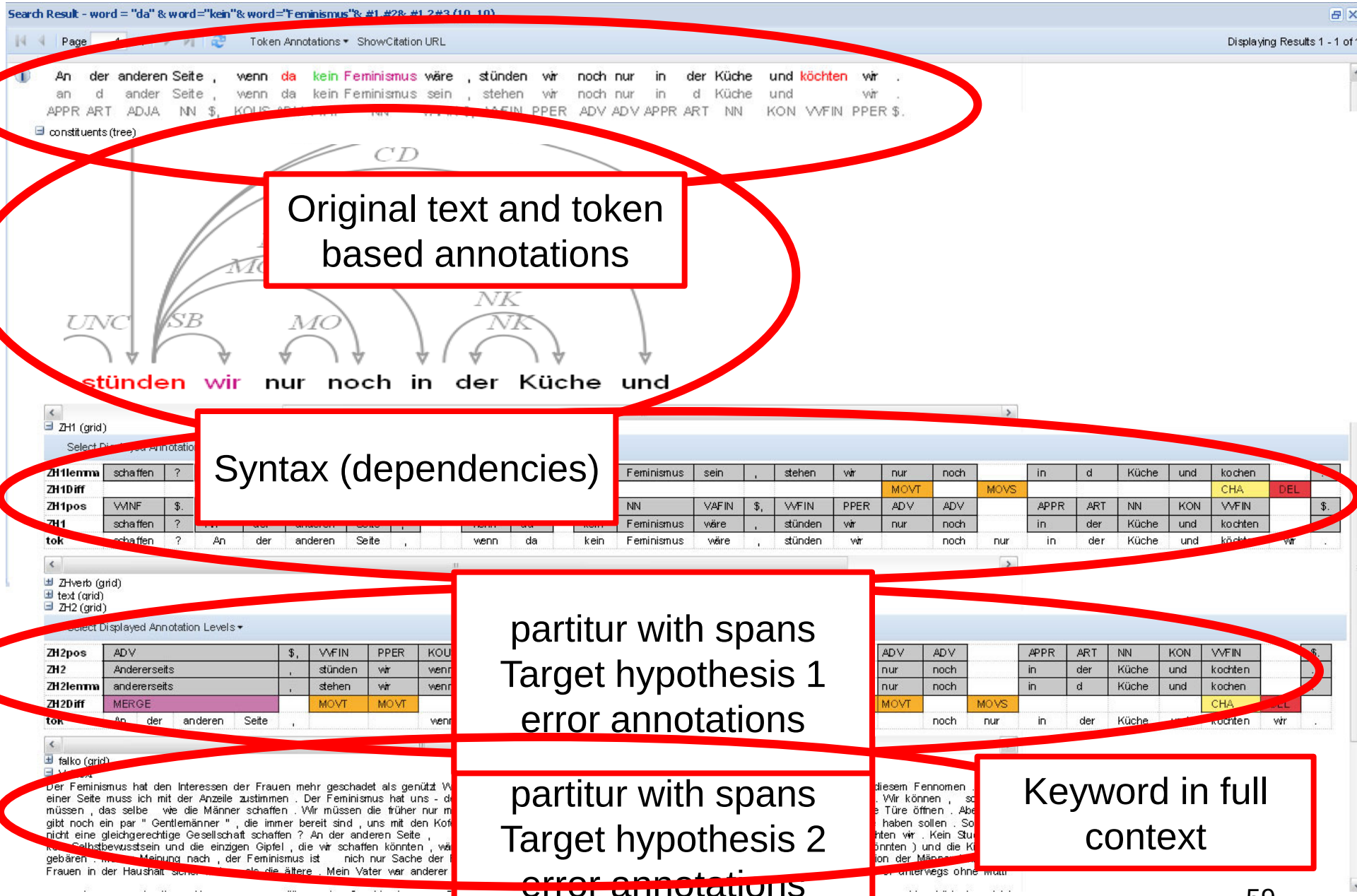
gets frequencies for #1 and #2



aside: Annis

- we search Falko in our freely available search tool Annis2
- multi-layer standoff model (token annotation span annotation, graphs, pointing relations)
- search across all annotation layers

- Chiarcos et al. (2008), Zeldes et al. (2009), Zipser & Romary (2010), http://www.sfb632.uni-potsdam.de/d1/annis/


Original text and token based annotations
Syntax (dependencies)
partitur with spans Target hypothesis 1 error annotations
partitur with spans Target hypothesis 2 error annotations
Keyword in full context

syntactic annotation of learner corpora for acquisition research

- many studies of syntactic phenomena in learner corpora, usually on the basis of surface structures (manually, pos tags, lexical cues etc.)
  for German see e.g. Diehl et al. (2000), Ahrenholtz (2008), Doolittle (2008), Breckle & Zinsmeister (submitted)
- several (very few) parsed learner corpora, often not publically available
  Dickinson & Ragheb (2009), Rosén & de Smedt (to appear)

syntactic annotation of learner corpora for CALL

- parsing learner data would help in generating intelligent answers to learner errors in call systems – a lot of reserach in this area – usually not helpful for our research question
  - often very restricted domains (question answering, fill in the blanks exercises etc.)
  - sometimes errors are explicitly introduced into 'native' data
- another goal: making parsers robust against data errors – again not directly helpful for our research question
- still: interesting results wrt to parsing techniques / evaluation techniques etc.

- Menzel & Schröder (1999), Vandeventer Faltin (2003), Ule & Simov (2004), Dickinson & Meurers (2005), Metcalf & Boyd (2006), Dickinson & Lee (2009), Amaral/Meurers/Ziai (to appear) etc.

learner corpora for GFL

- many learner corpora for English,
  more and more learner corpora for other languages
- Granger/Hung/Petch-Tyson (2002), Cobb (2003), Tono (2003), Myles/Mitchell (2004), Nesselhauf (2004), Tenfjord/Meurer/Hofland (2004), Granger (2008), Lüdeling/Walter (2009) etc.

- For German very few freely available learner corpora
  - LeaP (spoken)
  - AleSKO (in construction)
  - Ursula Weinberger (" ")
  - Falko

annotation of learner data: format

- many learner corpora are not annotated
- some are annotated with error tags, usually tabular formats or tree formats (XML), typically not standoff, typically not amendable by the user
- some (few) are annotated on other levels (pos, lemma etc.)

annotation of learner data: target hypothesis

- consider: *An der anderen Seite, wenn da kein Feminismus wäre, stünden wir noch nur in der Küche und köchten wir.* (fkb034_2008_07)
  ~ "On the other hand, if there were no feminism, we would still only stand in the kitchen and cook."

annotation of learner data: target hypothesis

- consider: *An der anderen Seite, wenn da kein Feminismus wäre, stünden wir noch nur in der Küche und köchten wir.* (fkb034_2008_07)
  ~ "On the other hand, if there were no feminism, we would still only stand in the kitchen and cook."

annotation of learner data: target hypothesis

- all error tags depend on an (at least implicit) correct version of a learner utterance
  → target hypothesis
- Falko: explicit target hypotheses
- often there are several ways of correcting an utterance
  th1: *Auf der anderen Seite, wenn da kein Feminismus wäre, stünden wir nur noch in der Küche und kochten.*
  th2: *Andererseits stünden wir, wenn es keinen Feminismus gäbe, nur noch in der Küche und kochten.*

learner utterance	target hypothesis 1	errors	target hypothesis 2	errors
An	Auf	CHA		
der	der			
anderen	anderen			
Seite	Seite		Andererseits	MERGE
,	,		,	
			stünden	MOVT
			wir	MOVT
			,	INS
wenn	wenn		wenn	
da	da			DEL
			es	INS
kein	kein		keinen	CHA
Feminismus	Feminismus		Feminismus	
wäre	wäre		gäbe	CHA
,	,		,	
stünden	stünden			MOVS
wir	wir			MOVS
	nur	MOVT	nur	MOVT
noch	noch		noch	
nur		MOVS		MOVS
in	in		in	
der	der		der	
Küche	Küche		Küche	
und	und		und	

how can we detect acquisition problems?

- structures that are unique for the L2 or different from the learners' L1s (transfer)
- structures that are judged to be difficult by the learners
- structures that contain many errors
- underused structures

how can we detect acquisition problems?

- structures that are unique for the L2 or different from the learners' L1s (transfer)
  - ➢ grammatical analysis
  - ➢ proved to be extremely problematic; no straightforward transfer
- structures that are judged to be difficult by the learners
  - ➢ intuition of the learners (unsystematic, dependent on teaching)
  - ➢ experiments
- structures that contain many errors
  - ➢ intuition of the teachers (unsystematic)
  - ➢ corpus analysis, error analysis (Corder 1991, Diehl/Albrecht/Zoch 1991, Granger 2008, Lüdeling 2008 etc.)
- underused structures
  - ➢ corpus analysis, Contrastive Interlanguage Analysis (Corder 1991, Ringbom 1998, Cobb 2003, Nesselhauf 2003 etc.)

how can we detect acquisition problems?

- structures that are unique for the L2 or different from the learners' L1s (transfer)
  - ➢ grammatical analysis
  - ➢ proved to be extremely problematic; no straightforward transfer
- structures that are judged to be difficult by the learners
  - ➢ intuition of the learners (unsystematic, dependent on teaching)
  - ➢ experiments
- structures that contain many errors
  - ➢ intuition of the teachers (unsystematic)
  - ➢ corpus analysis, error analysis (Corder 1991, Diehl/Albrecht/Zoch 1991, Granger 2008, Lüdeling 2008 etc.)
- underused structures
  - ➢ corpus analysis, Contrastive Interlanguage Analysis (Corder 1991, Ringbom 1998, Cobb 2003, Nesselhauf 2003 etc.)

parser evaluation

- evaluation of constituent structure (evalb)

>40	Precision	Recall	F-Score	Tagging acc.
L1	73.61	74.00	73.80	91.93
L2	77.59	79.04	78.31	92.85
Negra*	80.01	80.01	80.01	

- L2 easier to parse than L1
- possible reasons: sentence length / L1 syntactic structure might be more complex

➢ we can use parser output to compare L1 and L2

- *Berkeley results on the Negra Treebank (Petrov & Klein, 2007)

exemplary study about modifiers

- further probes into the adverb underuse
  - underuse statistics of syntactic categories: types of modification
  - underuse statistics & a combined search over positions (fields), categories and functions (vorfeld)

underuse of adverb chains

- the syntactic adverb classes were (manually) annotated (in essence this is a more fine-grained pos categorization)
- many studies about adverbs in learner language – analysis purely lexical
- the different distributions suggest that syntax might be relevant for understanding learner language
- however, the syntactic information codable at token level is too limited: we need hierarchical relations (dependencies, constituents)
- Möllering (2004), Vyatkina (2007) etc.