



Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik – Korpuslinguistik

**Das Falko-Handbuch
Korpusaufbau und Annotationen
Version 1.0.1**

Reznicek, Marc; Walter, Maik; Schmid, Karin; Lüdeling, Anke;
Hirschmann, Hagen; Krummes, Cedric; Andreas, Thorsten

Stand vom:
18.Oktober 2010

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

Übersicht

1. Falko-Korpus	4
1.1. Format der Metadaten:	6
2. Falko Zusammenfassungskorpus(Summary-Korpus)	7
2.1 Lernertexte (FalkoSummaryL2):	8
2.2 Muttersprachlertexte (FalkoSummaryL1):	9
2.3 Vorlagentexte (FalkoSummaryVL):	9
2.4. FalkoSummaryL2 1.1	9
2.4.1. Annotationen in FalkoSummaryL2	16
Zielhypothesen in Summary L2	17
Annotation topologischer Felder und syntaktischer Beschreibung	17
2.5. FalkoSummaryL1 1.1	17
2.6. FalkoSummaryVL 1.0	18
3. Falko-Aufsatzkorpus (Essay-Korpus)	19
3.1. FalkoEssayL2 2.0	20
3.1.1 Übersicht über Sprache und Geschlecht der Lerner für die einzelnen Erhebungen ...	21
3.1.2 Übersicht über die Orte und Textgrößen bezüglich der einzelnen Erhebungen	24
3.1.3 Verteilung der C-Test-Ergebnisse in FalkoEssayL2 2.0	25
3.2. FalkoEssayL11.2	26
3.2.1 Übersicht über Sprache und Geschlecht der Lerner für die einzelnen Erhebungen ...	26
3.2.2 Übersicht über die Orte und Textgrößen bezüglich der einzelnen Erhebungen	27
4. Richtlinien für die Annotation im Falko-Essay-Korpus 2.0	28
4.2 Textgrenzen: [TXTstructure]	29
4.3 Korrigierte Tokenebene: [ctok]	29
4.4 Makrostrukturebene: [macro]	30
5. Zielhypothesen	32
5.1 Technische Vorgaben für die Erstellung der Zielhypothesen	33
5.2 Minimale Zielhypothese (Satzebene, Orthografie, Morphosyntax, satzinterne	
Koreferenz, temporale und modale Kompatibilität) [ZH1]	36
Orthografie	36
Morphologie	37
Syntax	38
Referenz	44
Lexik	45
Semantik	45
5.2.1 Auf der Ebene der minimalen Zielhypothese [ZH1] NICHT annotierte Fehler	46
Syntax	46
Semantik	48
Lexik	48
Pragmatik	49
5.3 Maximale Zielhypothese [ZH2]	49



(Textebene, Semantik, Pragmatik, satzübergreifende Koreferenz, informationsstrukturelle Gliederung, Stil)	49
Syntax	50
Semantik.....	51
Morphologie	54
Lexik.....	54
Pragmatik.....	55
Stil.....	57
5.4 Zielhypothese für die Annotation der komplexen Verben [ZHverb]	58
Abweichungen der Zielhypothesen von der ctok-Ebene	58
6. Komplexe Verben	60
6.1. Annotationsebenen für die komplexen Verben	61
6.1.1 verbkategorie.....	61
6.1.2 verblemma.....	62
6.1.3 verbfehlertyp	62
6.1.4 verbform	63
7. Literatur:.....	64
8. Kontakt.....	65

1. Falko-Korpus

Das Falko-Gesamtkorpus V2.0 setzt sich aus fünf Subkorpora zusammen. Sie unterscheiden sich in zwei Faktoren: Schreibaufgabe und Muttersprache. Für die Zusammenfassungen liegen außerdem die Originalvorlagen vor.

	Lernerkorpus	muttersprachliches Kontrollkorpus	Vorlagenkorpus	Σ
Zusammenfassungen	FalkoSummaryL2 V1.2 (40.865 Tokens ¹)	FalkoSummaryL1 V1.2 (21.184 Tokens)	FalkoSummaryVL (11.114 Tokens)	73.163
Aufsätze	FalkoEssayL2 V2.0 (122.778 Tokens)	FalkoEssayL1 V1.2 (68.491 Tokens)		191.269
Σ	163.643	89.675	11.114	264.432

Für alle Lerner wurden neben den Texten auch umfangreiche Metadaten zu Alter, Geschlecht, akademischem Hintergrund, sprachlicher Biografie und Erhebungssituation erfasst und so aufbereitet, dass sie für die Generierung individueller Subkorpora dienen können.

Alle Texte sind unter Prüfungsbedingungen entstanden. Die Kontrollkorpora wurden unter den gleichen Bedingungen und mit den gleichen Anforderungen erhoben wie die Lernerkorpora.

Für alle Daten wurden die Wortarten und Lemmata mit dem *Treetagger* (Schmid 1994) automatisch annotiert. Die händischen Annotationen wurden in *EXMARaLDA* (Schmidt 2004,2005) und *Microsoft Excel* vorgenommen. Mithilfe des Treetaggers wurden dann auch für die Zielhypothesen Wortarten und Lemmata automatisch hinzugefügt. Im nächsten Schritt wurden diese Daten dann mit *SaltNPepper 1.0* (Zipser 2009) nach *RelANNIS* konvertiert, das als genuines relationales Datenbankformat für das Suchwerkzeug *ANNIS2* (Zeldes et al. 2009) dient. Die Korpora sind frei zugänglich und können auf der Internetseite des Instituts für deutsche Sprache und Linguistik der Humboldt-Universität nach einer Registrierung online durchsucht werden.

<http://korpling.german.hu-berlin.de/falko-suche/search.html>

Die Subkorpora beinhalten folgende Annotationsebenen:

Subkorpus	Ebenen	Tag
FalkoSummaryVL	Lernertext :	word
	Wortart (automatisch):	pos
	Lemma (automatisch):	lemma
FalkoSummaryL1	Lernertext:	word
	Wortart (automatisch):	pos
	Lemma (automatisch):	lemma
FalkoSummaryL2	Lernertext:	word
	Wortart (automatisch):	pos
	Lemma (automatisch):	lemma

¹ Tokenanzahlen beziehen sich auf die Ebene der Lernertexte ohne die durch die Zielhypothesen verursachten Leertokens.

	Zielhypothese: ²	target hypothesis
	korrigierte Wortart:	cpos
	Kommentar des Transkribenten:	transcripator comment
	Topologische Felder:	matrix-satz matrix-satz_felder konstituenten-satz_1 konstituenten-satz_1_felder konstituenten-satz_1_felder_2 matrix-satz_2 konstituenten-satz_2 konstituenten-satz_2_felder konstituenten-satz_2_felder_2 konstituenten-satz_3 konstituenten-satz_3_felder konstituenten-satz_3_felder_2
	Syntaktische Beschreibung:	[syntax_description_1] [syntax_classification_1] [syntax_classification_pos_1] [syntax_hypothesis_1] [syntax_description_2] [syntax_classification_2] [syntax_classification_pos_2] [syntax_hypothesis_2]
FalkoEssayL1	Lernertext:	word
	Wortart (automatisch):	pos
	Lemma (automatisch):	lemma
FalkoEssayL2	Basistext:	tok
	Lernertext:	word
	Wortart (automatisch):	pos
	Lemma (automatisch):	lemma
	Korrigierte Tokenisierung	ctok
	Textstruktur:	macro
	Minimale Zielhypothese:	ZH1
	Abweichung ZH1:	ZH1Diff
	Maximale Zielhypothese:	ZH2
	Abweichung ZH2:	ZH2Diff
	Zielhypothese komplexe Verben:	ZHverb
	Abweichung ZHverb:	ZHverbDiff
	Komplexe Verben:	verbkategorie verblemma verbfehler verbform

² Die Zielhypothesen der Zusammenfassungen folgen nicht den Richtlinien dieses Manuals und sind beschrieben im Handbuch der Annotation der Stellungsfelder bei Falko (2006).

1.1. Format der Metadaten:

Für alle Texte in Falko wurden Metadaten erhoben. Obwohl das volle Spektrum an Kategorien erst in zukünftigen Erhebungen aufgenommen werden, wurden auch die bereits vorhandenen Daten an das neue Format angepasst. Die Kategorien, die in der vorliegenden Version 2.0 bereits vorhanden sind, sind mit dem Index „V2.0“ gekennzeichnet.

		Kategorie	Beispiel	Tag	Kommentar
Allgemeine Informationen zum Lerner		Dateiname ^{V2.0}	HUB_001_2005_10.txt	Subkorpus + laufende Nummer	dreistelliges Kürzel der erhebenden Institution (wird vor der Erhebung mit dem Falkoteam besprochen), dreistellige laufende Nummer, 2005 und 10 stehen für den Erhebungszeitraum im Oktober 2005 – alles durch Unterstriche getrennt
		Name ^{V2.0}	Mustermann	name	werden gelöscht und durch MD5-Hashwerte ersetzt
		Vorname ^{V2.0}	Max	first-name	
		Geburtsjahr ^{V2.0}	1980	birth-year	in vierstelliger Form: JJJJ
		Geschlecht ^{V2.0}	m	sex	männlich = m, weiblich = f
		Bildungsgrad	M.A.	degree	höchster erreichter akademischer Abschluss
		Thema ^{V2.0}	Studium	topic	eines der vier vorgegebenen Themen (Entlohnung, Studium, Feminismus, Kriminalität)
		Fach ^{V2.0}	Deutsch	major-subject	Studienfach
		Transkribent ^{V2.0}	RM	transcripitor	Kürzel des Transkribenten, der die Daten digitalisiert
		Erhebungsdatum ^{V2.0}	25.05.2009	transcription-date	Datum, an dem die Erhebung stattfand.
		C-Test ^{V2.0}	85	cctest	Ergebnis aus dem C-Test
		Textsorte ^{V2.0}	Essay	production-modality	Textsorte
Sprache		Sprachbezeichnung ^{V2.0}	eng	11_x 12_x	Kürzel aus der Sprachliste 11 = Muttersprache, 12 = Fremd-/Zweitsprache
		gesprochen seit	0	11_x_since 12_x_since	ab welchem Alter gesprochen/gelernt, 0 = ab Geburt
	Unterricht	Monate ^{V2.0}	6	11_x_duration 12_x_duration	Dauer des Unterrichts der Sprache in MOonaten

	Institution	Schule	Ja	l1_x_school l2_x_school	Sprache in einer Schule gelernt?
		Uni	Ja	l1_x_university l2_x_university	Sprache an einer Universität gelernt?
		Sprachschule	Ja	l1_x_langschool l2_x_langschool	Sprache an einer Sprachschule gelernt?
	Auslands-ufenthalt	Monate	26	l1_x_awayMonths l2_x_awayMonths	Aufenthalt in einem Land der Zielsprache in Monaten
		Ort	Frankfurt	l1_x_awayPlace l2_x_awayPlace	Ort des Aufenthalts
	Indexfelder zur Suche mit regulären Ausdrücken ^{V2.0}		dan:N/A:0:Ja:Ja:Nein : N/A:N/A	l1index	Abfolge aller Werte für die Muttersprachen <i>Sprache1:Wert1</i> _{Sprache1} : <i>Wert2</i> _{Sprache1} : <i>Sprache2:Wert1</i> _{Sprache2} ...
			deu:48:12:Nein:Ja:Nein:12:Berlin	l2index	Abfolge aller Werte für die Fremdsprachen <i>Sprache1:Wert1</i> _{Sprache1} : <i>Wert2</i> _{Sprache1} : <i>Sprache2:Wert1</i> _{Sprache2} ...
			l1:dan,l1_duration:N/A,l1_since:N/A,l1_school:N/A,l1_university:N/A,l1_langschool:N/A,l1_awayMonths:N/A,l1_awayPlace:N/A;l2:eng,l2_duration:N/A,l2_since:N/A,l2_school:N/A,l2_university:N/A,l2_langschool:N/A,l2_awayMonths:N/A,l2_awayPlace:N/A	reg	Abfolge aller Sprachen und ihrer Werte als Variablen-Wert-Paare <i>Variable₁: Wert_{Sprache1}</i> , <i>Variable₂: Wert_{Sprache1}</i> ; <i>Variable₁: Wert_{Sprache2}</i> , <i>Variable₂: Wert_{Sprache2}</i> ; ...

1.2. Falko Ländercodes nach ISO 639-3

Code	Sprache
afr	Afrikaans
cat	Katalanisch
ces	Tschechisch
cma	Maa
dan	Dänisch
ell	Neugriechisch
eng	Englisch
fin	Finnisch

fra	Französisch
hbs	Serbokroatisch
hin	Hindi
hun	Ungarisch
iii	Sichuan-Yi
ita	Italienisch
jpn	Japanisch
kik	Kikuyu
kor	Koreanisch
kua	Oshivambo
lub	Kiluba
luy	Luhya
nde	Nord-Ndebele
nld	Niederländisch
nor	Norwegisch
pol	Polnisch
ron	Rumänisch
rus	Russisch
slk	Slowakisch
sme	Nord-Sami
spa	Spanisch
sqi	Albanisch
swe	Schwedisch
tat	Tatarische
tur	Türkisch
ukr	Ukrainisch
uzb	Usbekisch
vie	Vietnamesische
zho	Chinesisch

2. Falko Zusammenfassungskorpus(Summary-Korpus)

Das Falko-Summary-Korpus besteht aus drei Subkorpora.

2.1 Lernertexte (FalkoSummaryL2):

Dieser Teil enthält Textzusammenfassungen, die von fortgeschrittenen Lernern des Deutschen erstellt wurden. Die Texte sind Zusammenfassungen von linguistischen und literaturwissenschaftlichen Fachtexten, die als Teil der obligatorischen Sprachstandsbestimmung für ausländische Studierende verfasst wurden. Die Daten wurden an der Freien Universität Berlin erhoben. Ausländische Studierende, die in einem germanistischen Hauptfach eingeschrieben sind, müssen nach dem Grundstudium eine Sprachprüfung absolvieren, in der sie nachweisen, dass sie einen germanistischen Fachtext verstehen und sich fachsprachlich ausdrücken können. Diese Sprachstandsbestimmung ist eine Voraussetzung für die Zulassung zur Zwischenprüfung. Die Prüfung wird durch das Studienggebiet Deutsch als Fremdsprache des Instituts für Deutsche und Niederländische Philologie verantwortet. Die Textvorlagen wurden von Maik Walter (Linguistik) und Almut Hille

(Literaturwissenschaft) ausgewählt. Neben dem schriftlichen Teil absolvieren die Studierenden einen mündlichen Teil. Die Verfasser der Texte haben die DSH-Prüfung erfolgreich absolviert und werden deshalb als fortgeschrittene Lerner (auf dem Niveau C1 - C2 des Europäischen Referenzrahmens) eingestuft. Der Prüfungskontext ist unten beschrieben. Die Texte wurden von Julia Kassubek, Katja Jansen und Karin Schmidt digitalisiert und mehrfach von verschiedenen Mitarbeitern und Studierenden korrigiert.

2.2 Muttersprachlertexte (FalkoSummaryL1):

Dieser Teil enthält Textzusammenfassungen, die von deutschen Muttersprachlern (Studierenden der Freien Universität Berlin und der Humboldt-Universität zu Berlin) erstellt wurden. Die Texte sind Zusammenfassungen derselben linguistischen und literaturwissenschaftlichen Fachtexte, die auch von den Lernern bearbeitet wurden. Die Rahmenbedingungen für die Erhebungen waren vergleichbar (90 Minuten, keine Hilfsmittel), allerdings wurden die Texte nicht als Prüfungsleistung erhoben. Auch hier wurden mithilfe eines Fragebogens Metadaten über die Verfasser erhoben.

2.3 Vorlagentexte (FalkoSummaryVL):

Dieser Teil enthält die linguistischen und literaturwissenschaftlichen Fachtexte, die als Vorlage für die Textzusammenfassungen in den anderen Subkorpora verwendet wurden.

Im Folgenden sind die einzelnen Erhebungszeiträume für FalkoSummaryL2 und die Zusammensetzung aller Subkorpora dokumentiert. In der Dokumentation sind die Vorlagentexte abkürzend benannt – die genauen Angaben zu jedem Vorlagentext finden sich in der Dokumentation der FalkoSummaryVL.

2.4. FalkoSummaryL2 1.1

Die folgenden 6 Datenerhebungen wurden als Grundlage für das Korpus FalkoSummaryL2 1.1 verwendet: Datum der Erhebung	Anzahl der Lernertexte		
	männlich	weiblich	Σ
09.02.2004	5	19	24
01.07.2004	7	13	20
20.01.2005	1	14	15
27.06.2005	3	20	23
06.02.2005	0	16	16
02.02.2006	3	6	9
Σ	19	88	107

Tokenanzahl

Lerner	98	Texte	107	Tokens	40923	Ø Text	382,46
--------	----	-------	-----	--------	-------	--------	--------

Insgesamt haben 98 Lerner 197 Texte verfasst, von 9 Lernern sind daher zwei verschiedene Texte im Subkorpus enthalten. Die folgenden Texte wurden zusammengefasst:

Vorlagentext	N
Berlinromane	5
Entscheidungen	6
Epochen	5
Hermeneutik	18
Pragmatik	11
Realismus	9
Schlaf	9
Semantik	11
Syntax	4
Textgrenzen	12
Valenz	14
Volksmärchen	3
Σ	107

Datenerhebung vom 09.02.2004

Die Aufgabe bestand darin, einen literaturwissenschaftlichen (N=18) bzw. linguistischen (N=6) Fachtext zusammenzufassen.

Angaben zu den Ausgangstexten:

- (a) Witte, Bernd (1993): Das Gericht, das Gesetz, die Schrift. Über die Grenzen der Hermeneutik am Beispiel von Kafkas Türhüter - Legende. In: Bogdal, Klaus-Michael (Hg.): Neue Literaturtheorien in der Praxis. Textanalysen von Kafkas "Vor dem Gesetz". Opladen: Westdeutscher Verlag, S. 94-97. Als Datei hermeneutik.rtf Teil des Korpus.
- (b) Miller, George A. (1993): Unterscheidungen treffen. In: ders.: Wörter. Streifzüge durch die Psycholinguistik. Spektrum. Heidelberg, Berlin, New York: Akademischer Verlag, S. 223. Als Datei entscheidungen.rtf Teil des Korpus.

Aufgabenstellung

- a) Beantworten Sie bitte folgende Fragen anhand des Textes.
1. Was ist Hermeneutik?
 2. Warum ist Franz Kafkas Legende "Vor dem Gesetz" für eine hermeneutische Analyse geeignet?
 3. Was ist das "Paradoxe" in Kafkas Text?
- b)
1. Fassen Sie den folgenden Text mit eigenen Worten zusammen.
 2. Geben Sie ein Beispiel für eine nicht informationsübermittelnde Kommunikation (mit nicht ernsthaften Menschen).

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis

- keine Hilfsmittel
- handschriftlich verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

Datum der Erhebung	Anzahl der Teilnehmer		L1	L2
	männlich	weiblich		
09.02.2004	5	19	Polnisch (11) Portugiesisch (2) Russisch (2) Georgisch (2) Koreanisch (2) Französisch (1) Bulgarisch (1) Weißrussisch (1) Englisch (1) Persisch (1)	Deutsch (24) Englisch (20) Französisch (1) Russisch (11) Ukrainisch (1) Spanisch (3) Niederländisch (4) Japanisch (1) Chinesisch (1) Italienisch (1)
Σ	24			

Datenerhebung vom 01.07.2004

Die Aufgabe bestand darin, einen literaturwissenschaftlichen (N=9) bzw. linguistischen (N=11) Fachtext zusammenzufassen.

Angaben zu den Ausgangstexten:

- (a) Sprengel, Peter (1998): III. Stile und Richtungen. 1. Realismus. In: ders.: Geschichte der deutschsprachigen Literatur 1870-1900. Von der Reichsgründung bis zur Jahrhundertwende. München: Verlag C.H. Beck, S. 99-101. Als Datei realismus.rtf Teil des Korpus.
- (b) Meibauer, Jörg (1999): Pragmatische Erwerbsprinzipien. In: ders.: Pragmatik. Eine Einführung. Tübingen: Stauffenburg, S. 170-172. Als Datei pragmatik.rtf Teil des Korpus.

Aufgabenstellung:

3. Fassen Sie bitte den folgenden Text zusammen.
4. Fassen Sie den Text mit eigenen Worten zusammen.

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- handschriftlich verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

Datum der Erhebung	Anzahl der Teilnehmer		L1	L2
	männlich	weiblich		
01.07.2004	7	13	Polnisch (5)	Deutsch (20)



Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik – Korpuslinguistik

Spezifikationen des Falko Korpus 2.0 – Version 1.0

		Chinesisch (3)	Englisch (14)
		Russisch (2)	Russisch (4)
		Japanisch (2)	Spanisch (3) Franzö-
		Georgisch (1)	sisch (2) Chinesisch (1)
		Persisch (1)	Italienisch (1)
		Slowenisch (1)	
		Arabisch (1)	
		Ungarisch (1)	
		Türkisch (1)	
		Deutsch (1)	
		Litauisch (1)	
		Thai (1)	
Σ	20		

Die Aufgabe bestand darin, einen literaturwissenschaftlichen (N=3) bzw. linguistischen (N=12) Fachtext zusammenzufassen.

Angaben zu den Ausgangstexten:

- (a) Klotz, Volker (2002): Kunstmärchen: Name und Sachverhalt. In: ders.: Das europäische Kunstmärchen. Fünfundzwanzig Kapitel seiner Geschichte von der Renaissance bis zur Moderne. 3. überarbeitete und erweiterte Auflage. München: Wilhelm Fink Verlag, S. 7-8. Als Datei volksmaerchen.rtf Teil des Korpus.
- (b) Linke, Angelika / Nussbaumer, Markus / Portmann, Paul R. (21994): Textgrenzen. In: dies.: Studienbuch Linguistik. Tübingen: Max Niemeyer Verlag, S. 255-256. Als Datei textgrenzen.rtf Teil des Korpus.

Aufgabenstellung:

1. Fassen Sie bitte den folgenden Text zusammen.
2. Fassen Sie den Text mit eigenen Worten zusammen.

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- handschriftlich verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

Datum der Erhebung	Anzahl der Teilnehmer		L1	L2
	männlich	weiblich		
20.01.2005	1	14	Polnisch (4)	Deutsch (15)



Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik – Korpuslinguistik
Spezifikationen des Falko Korpus 2.0 – Version 1.0

		Russisch (4) Bulgarisch (2) Ukrainisch (1) Serbo-Kroatisch (1) Japanisch (1) Armenisch (1) Englisch (1) Chinesisch (1)	Englisch (13) Russisch (3) Franzö- sisch (2) Spanisch (2) Italienisch (2) Rumä- nisch (1) Latein (1) Bosnisch (1)
Σ	15		

Datenerhebung vom 27.06.2005

Die Aufgabe bestand darin, einen linguistischen (N=9) bzw. literaturwissenschaftlichen (N=14) Fachtext zusammenzufassen.

Angaben zu den Ausgangstexten:

- (a) Eisenberg, Peter (2004): 3.2.2 Valenz und Bedeutung. Grundpositionen. In: ders.: Grundriss der deutschen Grammatik. Band 2: Der Satz. 2., überarbeitete und aktualisierte Auflage. Stuttgart, Weimar: Metzler, S. 71-72. Als Datei valenz.rtf Teil des Korpus.
- (b) Alt, Peter-André (2002): Der Schlaf der Vernunft. Literatur und Traum in der Kulturgeschichte der Neuzeit. München: Beck, S. 10-12. Als Datei schlaf.rtf Teil des Korpus.

Aufgabenstellung:

- Fassen Sie bitte den folgenden Text zusammen.
- Fassen Sie den Text mit eigenen Worten zusammen.

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- handschriftlich verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

Datum der Erhebung	Anzahl der Teilnehmer		L1	L2
	männlich	weiblich		
27.06.2005	3	20	Polnisch (10)	Deutsch (23)



Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik – Korpuslinguistik
 Spezifikationen des Falko Korpus 2.0 – Version 1.0

		Russisch (10) Weißrussisch (3) Ukrai- nisch (3) Portugie- sisch(1) Mongolisch (1)	Englisch (20) Russisch (9) Franzö- sisch (5) Spanisch (4) Italienisch (1) Rumä- nisch (1) Latein (1) Polnisch (1) Niederländisch (1) Ja- panisch (1)
Σ	23		

Datenerhebung vom 02.02.2006

Die Aufgabe bestand darin, einen literaturwissenschaftlichen (N=5) bzw. linguistischen (N=11) Fachtext zusammenzufassen.

Angaben zu den Ausgangstexten:

(a) Rosenberg, Rainer (2001): Epochen. In: Brackert, Helmut/ Stückrath, Jörn (Hg.): Literaturwissenschaft. Ein Grundkurs. Reinbek: Rowohlt Taschenbuch Verlag, S. 269-272.

Als Datei epochen.rtf Teil des Korpus.

(b) Wunderlich, Dieter (1991): Welche Verfahren gibt es zur Bedeutungsanalyse? In: ders.: Arbeitsbuch Semantik. 2., ergänzte Auflage. Frankfurt am Main: Hain, S. 124-126.

Als Datei semantik.rtf Teil des Korpus.

Aufgabenstellung:

5. Fassen Sie bitte den folgenden Text zusammen.
6. Fassen Sie den Text mit eigenen Worten zusammen.

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- handschriftlich verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

Datum der Erhebung	Anzahl der Teilnehmer		L1	L2
	männlich	weiblich		
6	0	16	Polnisch (5) Russisch (4) Mongolisch (1) Bulgarisch (1) Kroatisch (1) Italienisch (1) Japanisch (1)	Arabisch (1) Englisch (16) Deutsch (16) Schwedisch (2) Französisch (2) Spanisch (3) Italienisch (3)



Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik – Korpuslinguistik

Spezifikationen des Falko Korpus 2.0 – Version 1.0

		Koreanisch (1) Litauisch (1)	Portugiesisch (1) Russisch (3) Litauisch (1)
Σ	16		

Datenerhebung vom 06.02.2007

Die Aufgabe bestand darin, einen linguistischen (N=4) bzw. literaturwissenschaftlichen (N=5) Fachtext zusammenzufassen.

Angaben zu den Ausgangstexten:

(a) Eroms, Hans-Werner (2000): Syntax der deutschen Sprache. Berlin, New York: Walter de Gruyter, S. 47-48.

Als Datei syntax.rtf Teil des Korpus.

(b) Siebenpfeiffer, Hania (2001): Topographien des Seelischen. Berlinromane der neunziger Jahre. In: Harder, Matthias (Hg.): Bestandsaufnahmen. Deutschsprachige Literatur der neunziger Jahre aus interkultureller Sicht. Würzburg: Königshausen & Neumann, S. 85-87.

Als Datei berlinromane.rtf Teil des Korpus.

Aufgabenstellung:

7. Fassen Sie bitte den folgenden Text zusammen.
8. Fassen Sie den Text mit eigenen Worten zusammen.

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- handschriftlich verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

Datum der Erhebung	Anzahl der Teilnehmer		L1	L2
	männlich	weiblich		
06.02.2007	3	6	Polnisch (4)	Deutsch (9)

		Russisch (3) Englisch (2) Baschkirisch (1)	Englisch (6) Französisch (2) Latein (2) Altgriechisch (1) Arabisch(1) Hebräisch (1) Italienisch (1) Japanisch (1) Niederländisch(1) Spanisch (1) Türkisch (1) Tschechisch (1)
Σ	9		

2.4.1. Annotationen in FalkoSummaryL2

Anotationsebene	Kürzel
Lernertext:	word
Wortart (automatisch):	pos
Lemma (automatisch):	lemma
einfache Zielhypothese:	target hypothesis
korrigierte Wortart:	cpos
Kommentar des Transkribenten:	transcriptor comment
Topologische Felder:	matrix-satz matrix-satz_felder konstituenten-satz_1 konstituenten-satz_1_felder konstituenten-satz_1_felder_2 matrix-satz_2 konstituenten-satz_2 konstituenten-satz_2_felder konstituenten-satz_2_felder_2 konstituenten-satz_3 konstituenten-satz_3_felder konstituenten-satz_3_felder_2
Syntaktische Beschreibung:	[syntax_description_1] [syntax_classification_1] [syntax_classification_pos_1] [syntax_hypothesis_1] [syntax_description_2] [syntax_classification_2] [syntax_classification_pos_2] [syntax_hypothesis_2]

Zielhypothesen in Summary L2

Die Zielhypothesen im Summary-Korpus entsprechen den Entwürfen in Lüdeling al. 2005. Sie weichen somit von den hier entwickelten Richtlinien ab.

Annotation topologischer Felder und syntaktischer Beschreibung

Im Rahmen der Magisterarbeit von Seanna Doolittle (Doolittle 2009) wurden kanonische und unkanonische Sätze annotiert und für erstere die folgenden Felderannotationen vergeben

Vorfeld	Linke Satzklammer	Mittelfeld	Rechte Satzklammer	Nachfeld
VF	LSK	MF	RSK	NF

Weiterhin wurde auf der Grundlage dieser Annotationen eine syntaktische Beschreibung vorgenommen. Weitere Details zu diesen Annotationen finden Sie im **Fehler! Verweisquelle konnte nicht gefunden werden...**

2.5. FalkoSummaryL1 1.1

Vier Datenerhebungen wurden als Grundlage für das Korpus FalkoSummaryL1 1.1 verwendet. Die Zusammenfassungen wurden an der Freien Universität Berlin und an der Humboldt-Universität zu Berlin von Studierenden eines germanistischen Faches im Hauptstudium verfasst. Ein Teil der Studierenden (N=39) absolvierte den Zusatzstudiengang Deutsch als Fremdsprache an der Freien Universität Berlin. Alle Texte wurden unter den identischen Bedingungen erhoben, das betrifft insbesondere die Aufgabenstellung und die kontrollierte Datenerhebung.

Datum (Ort) der Erhebung	Anzahl der Texte			Σ
	männlich	weiblich	N/A	
17.02.2005 (FU Berlin)	2 (2)	5 (5)	11(11)	18
22.05.2007 (FU Berlin)	0	10 (10)	0	10
15.07./20.07./01.08.2007 (FU Berlin)	0	11 (8)	0	11
03.05./07.06./13.06./09.07./20.07.2007 (HU Berlin)	8 (8)	10 (10)	0	18
Σ	10	36	11	57

Tokenanzahl

Texte	36	Lerner	33	Tokens	21184	Ø/Text	370,62
-------	----	--------	----	--------	-------	--------	--------

Unterschied zur Version 1.0 – proportional zum L1-Subkorpus kompiliert

Zu jedem Vorlagentext wurden (aus ökonomischen Gründen) die halbe Anzahl der Texte (der L2) in der L1 Deutsch erhoben. Bei einer ungeraden Zahl wurde aufgerundet. Nach der ersten Erhebung wurden ebenfalls die Metadaten (L1, L2, L3,..., Dauer des Erwerbs, Alter, Geschlecht) erhoben. In 11 Fällen liegen keine Metadaten vor.

Aufgabenstellung:

- identisch mit den L2-Erhebungen (s.o.)

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- handschriftlich verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

Vorlagentext	N
Berlinromane	5
Entscheidungen	6
Epochen	5
Hermeneutik	18
Pragmatik	11
Realismus	9
Schlaf	9
Semantik	11
Syntax	4
Textgrenzen	12
Valenz	14
Volksmärchen	3
Σ	57

2.6. FalkoSummaryVL 1.0

Die folgenden Texte bilden die Textbasis für das Subkorpus FalkoSummaryVL1.0:

Signle	Quelle
Hermeneutik	Witte, Bernd (1993): Das Gericht, das Gesetz, die Schrift. Über die Grenzen der Hermeneutik am Beispiel von Kafkas Türhüter - Legende. In: Bogdal, Klaus-Michael (Hg.): Neue Literaturtheorien in der Praxis. Textanalysen von Kafkas "Vor dem Gesetz". Opladen: Westdeutscher Verlag, S. 94-97.
Entscheidungen	Miller, George A. (1993): Unterscheidungen treffen. In: ders.: Wörter. Streifzüge durch die Psycholinguistik. Heidelberg, Berlin, New York: Spektrum. Akademischer Verlag, S. 223.
Pragmatik	Meibauer, Jörg (1999): Pragmatische Erwerbsprinzipien. In: ders.: Pragmatik. Eine Einführung. Tübingen: Stauffenburg, S. 170-172.
Realismus	Sprengel, Peter (1998): III. Stile und Richtungen. 1. Realismus. In: ders.: Geschichte der deutschsprachigen Literatur 1870-1900. Von der Reichs-

	gründung bis zur Jahrhundertwende. München: Verlag C.H. Beck, S. 99-101.
Volksmärchen	Klotz, Volker (2002): Kunstmärchen: Name und Sachverhalt. In: ders.: Das europäische Kunstmärchen. Fünfundzwanzig Kapitel seiner Geschichte von der Renaissance bis zur Moderne. 3. überarbeitete und erweiterte Auflage. München: Wilhelm Fink Verlag, S.7-8.
Textgrenzen	Linke, Angelika / Nussbaumer, Markus / Portmann, Paul R. (21994): Textgrenzen. In: dies.: Studienbuch Linguistik. Tübingen: Max Niemeyer Verlag, S. 255/-256.
Schlaf	Alt, Peter-André (2002): Der Schlaf der Vernunft. Literatur und Traum in der Kulturgeschichte der Neuzeit. München: Beck, S. 10-12.
Valenz	(a) Eisenberg, Peter (2004): 3.2.2 Valenz und Bedeutung. Grundpositionen. In: ders.: Grundriss der deutschen Grammatik. Band 2: Der Satz. 2., überarbeitete und aktualisierte Auflage. Stuttgart, Weimar: Metzler, S. 71-72.
Semantik	Wunderlich, Dieter (1991): Welche Verfahren gibt es zur Bedeutungsanalyse? In: ders.: Arbeitsbuch Semantik. 2., ergänzte Auflage. Frankfurt am Main: Hain, S. 124-126.
Epochen	Rosenberg, Rainer (2001): Epochen. In: Brackert, Helmut/ Stückrath, Jörn (Hg.): Literaturwissenschaft. Ein Grundkurs. Reinbek: Rowohlt Taschenbuch Verlag, S. 269-272.
Syntax	Eroms, Hans-Werner (2000): Syntax der deutschen Sprache. Berlin, New York: Walter de Gruyter, S. 47-48.
Berlinromane	Siebenpfeiffer, Hania (2001): Topographien des Seelischen. Berlinromane der neunziger Jahre. In: Harder, Matthias (Hg.): Bestandsaufnahmen. Deutschsprachige Literatur der neunziger Jahre aus interkultureller Sicht. Würzburg: Königshausen & Neumann, S. 85-87.

Tokenanzahl

Texte	12	Tokens	11114	Ø/Text	926,17
-------	----	--------	-------	--------	--------

3. Falko-Aufsatzkorpus (Essay-Korpus)

Das Falko-Essay-Korpus besteht aus zwei Subkorpora.

Lernertexte (FalkoEssayL2):

Dieser Teil enthält Aufsätze, die von fortgeschrittenen Lernern des Deutschen erstellt wurden. Die Texte sind argumentative Aufsätze zu einem von vier vorgegeben Themen, die aus der Gesamtmenge der im *International Corpus of Learner English* (ICLE) (Granger 1993, 2003) verwendeten Aufsatzthemen ausgewählt wurden.

Die Lernertexte stammen von Nicht-Muttersprachlern, die teilweise an Feriensprachkursen an der Freien Universität Berlin und der Humboldt-Universität zu Berlin und teilweise an ausländischen Universitäten und Goethe-Instituten erhoben wurden. Alle Lerner mussten einen Fragebogen für

die Erfassung der Lernerdaten ausfüllen und in einem C-Test mindestens 60 von 100 Punkten erreichen, der vom Sprachenzentrum der Humboldt-Universität zu Berlin entwickelt wurde und dort ebenfalls eingesetzt wird.

Ergebnis im C-Test	Einstufungsniveau in Maßen des GER
60-79	B2
80-89	C1
90-100	C2

Die Texte wurden unter Aufsicht direkt in einem Texteditor geschrieben, der keine Rechtschreibkorrektur beinhaltet. Jeglicher Zugriff auf weitere Hilfsmittel bzw. das Internet wurde vorher ausgeschlossen.

Muttersprachlertexte (FalkoEssayL1):

Dieser Teil enthält Aufsätze, die von deutschen Muttersprachlern in den Abschlussklassen dreier Gymnasien in Berlin, Eichwalde und Potsdam, sowie in einem Kurs im Studiengang „Deutsch als Fremdsprache“ an der Freien Universität Berlin erhoben wurden. Die Texte sind Aufsätze zu denselben Themen, die auch von den Lernern bearbeitet wurden. Die Rahmenbedingungen für die Erhebungen waren vergleichbar (90 Minuten, keine Hilfsmittel). Auch hier wurden mithilfe eines Fragebogens Metadaten über die Verfasser erhoben.

Im Folgenden sind die einzelnen Erhebungszeiträume für Falko-Essay und die Zusammensetzung aller Subkorpora dokumentiert.

3.1. FalkoEssayL2 2.0

Die Aufgabe bestand darin, zu einem der folgenden vier Themen einen argumentativen Aufsatz zu schreiben:

- Der Feminismus hat den Frauen mehr geschadet als genutzt.
- Kriminalität zahlt sich nicht aus.
- Die meisten Universitätsabschlüsse bereiten die Studenten nicht auf die wirkliche Welt vor. Sie sind deswegen von geringem Wert
- Die finanzielle Entlohnung eines Menschen sollte dem Beitrag entsprechen, den er/sie für die Gesellschaft geleistet hat.

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- in einem Texteditor verfasste Klausuren unter Aufsicht

• Zeit: 90 Minuten

3.1.1 Übersicht über Sprache und Geschlecht der Lerner für die einzelnen Erhebungen

(Mehrfachangaben für Sprachen wurden auch mehrfach gezählt)

Erhebungsdatum	Anzahl der Teilnehmer		L1	L2			
	männlich	weiblich					
02.05.2006	3	5	Türkisch (8)	Deutsch (8)			
09.05.2006				Englisch (8)			
10.05.2006				Französisch (1)			
28.06.2006	5	13	Suaheli (9) Kikuyu (5) Luo (5) Luhya (2) Meru (2) Embu (1) Gusii (1) Kalenjin (1) Nandi (1) KiTaita (1)	Deutsch (18)			
11.07.2006				Englisch (18)			
12.07.2006				Suaheli (9)			
17.07.2006				Französisch (2)			
18.07.2006				Chinesisch (1)			
10.08.2006				Italienisch (1)			
12.09.2006				Kikuyu (1)			
29.09.2006				Luhya (1)			
				Meru (1)			
				Giryama (1)			
	Spanisch (1)						
	Schwedisch (1)						
27.07.2006	7	17	Englisch(8) Französisch(5) Neugriechisch (3) Schwedisch(3) Italienisch(2) Chinesisch(1) Dänisch(1) Hebräisch(1) Niederländisch(1) Norwegisch(1) Polnisch(1) Rumänisch(1) Russisch(1) Tschechisch(1) Ukrainisch(1)	Deutsch(24)			
				Englisch(16)			
				Französisch(10)			
				Latein(10)			
				Russisch(6)			
				Spanisch(6)			
				Niederländisch(2)			
				Altenglisch(1)			
				Indonesisch(1)			
				Italienisch(1)			
				Japanisch(1)			
17.08.2006							
19.09.2006				6	7	Englisch (5) Norwegisch (2) Griechisch (1) Finnisch (1) Französisch (1) Niederländisch (1)	Deutsch (13)
							Englisch (12)
							Französisch (8)
							Italienisch (3)
							Spanisch (3)
	Griechisch (1)						
26.09.2006			Polnisch (1) Spanisch (1)	Latein (1) Niederländisch (2)			
			<i>(Fast alle Teilneh-</i>	<i>(Fast alle Teilneh-</i>			

			<i>mer haben 2 Texte geschrieben)</i>	<i>mer hatten 2 Texte geschrieben)</i>
29.09.2006	4	27	Dänisch(31) Schwedisch(2) Norwegisch(1)	Englisch(31) Deutsch(31) Französisch(17) Latein(8) Spanisch(4) Russisch(3) Schwedisch(3) Norwegisch(2) Italienisch(1)
01.10.2007				
04.10.2006	6	9	Usbekisch (11) Russisch (4) Tadschikisch (3)	Deutsch (15) Englisch (14) Russisch (11) Usbekisch (3) Tadschikisch (2) Persisch (1) Französisch (1) Koreanisch (1)
24.10.2006	5	10	Englisch (3) Japanisch (3) Chamorro (2) Meru (2) Angika (1) Embu (1) Maithili (1) Hindi (1) Koreanisch (1) Norwegisch (1) Polnisch (1) Russisch (1) Ukrainisch (1)	Deutsch (15) Englisch (11) Französisch (3) Japanisch (1) Russisch (1) Jiddisch (1)
20.11.2007				
07.12.2006	1	9	Afrikaans (8) Englisch (2)	Deutsch (10) Englisch (7) Xhosa (3) Afrikaans (2) Französisch (2) Chinesisch (1)
05.03.2007	1	1	Englisch (2)	Deutsch (2) Französisch (2)
18.05.2007				Tschechisch (1)
25.06.2007				Latein (1) Norwegisch (1)

			Chinesisch (1)
09.08.2007			Russisch (13) Englisch (12) Französisch (6) Dänisch (5) Spanisch (5) Ukrainisch (4) Niederländisch (3) Polnisch (3) Rumänisch (3) Italienisch (2) Türkisch (2) Tschechisch (2) Ungarisch (2) Vietnamesisch (2) Albanisch (1) Finnisch (1) Hindi (1) Irish (1) Katalanisch (1) Neugriechisch (1) Slovakisch (1)
20.11.2007			Englisch (55) Deutsch (50) Französisch (28) Spanisch (18) Latein (17) Russisch (7) Italienisch (6) Niederländisch (4) Altgriechisch (2) Katalanisch (2) Schwedisch (2) Tschechisch (2) Chinesisch (3) Walisisch (2) Baskisch (1) Japanisch (1)
25.07.2008	12	38	
06.08.2008			
Σ	50	136	Dänisch(37) Deutsch (186) Englisch (32) Englisch (172) Russisch (19) Französisch (74) Französisch (12) Spanisch (32) Usbekisch (11) Latein (29) Türkisch (10) Russisch (28) Suaheli (9) Italienisch (12) Afrikaans (8) Suaheli (9) Polnisch (6) Niederländisch (8) Spanisch (6) Chinesisch (6) Ukrainisch (6) Schwedisch (6) Luo (5) Japanisch (3) Kikuyu (5) Norwegisch (5) Niederländisch (5) Tschechisch (3) Norwegisch (5) Usbekisch (3) Schwedisch(5) Xhosa (3) Neugriechisch (4) Afrikaans (2) Italienisch (4) Altgriechisch (2) Rumänisch (4) Katalanisch (2) Japanisch (3) Tadschikisch (2) Tadschikisch (3) Tschechisch (2) Tschechisch (3) Walisisch (2) Chamorro (2) Altenglisch(1)

			Embu (2)	Baskisch (1)
			Finnisch (2)	Giryama (1)
			Hindi (2)	Griechisch (1)
			Luhya (2)	Indonesisch(1)
			Ungarisch (2)	Jiddisch (1)
			Vietnamesisch (2)	Kikuyu (1)
			Albanisch (1)	Koreanisch (1)
			Angika (1)	Luhya (1)
			Chinesisch(1)	Meru (1)
			Griechisch (1)	Persisch (1)
			Gusii (1)	
			Hebräisch(1)	
			Irish (1)	
			Kalenjin (1)	
			Katalanisch (1)	
			KiTaita (1)	
			Koreanisch (1)	
			Maithili (1)	
			Meru (4)	
			Nandi (1)	
			Slovakisch (1)	
Σ		186		

3.1.2 Übersicht über die Orte und Textgrößen bezüglich der einzelnen Erhebungen

Erhebungsdatum	Ort	Texte	Token	Token/Text
02.05.2006		2		
09.05.2006	Cukurova University, Türkei (TRK)	2	2772	346,50
10.05.2006		4		
28.06.2006		1		
11.07.2006	Goethe-Institut Nairobi, Kenia (KNE)	1	5016	278,67
12.07.2006		2		
17.07.2006		6		
18.07.2006		1		
10.08.2006		2		
12.09.2006		1		
29.09.2006		4		
27.07.2006		HU Berlin (Ferienkurs), Deutschland(FK)		
17.08.2006	19			
19.09.2006	Humboldt-Universität, Deutschland (HU)	13	12926	538,58
26.09.2006		11		
29.09.2006	Copenhagen Business School, Dänemark (CBS)	16	16463	531,06
01.10.2007		15		
04.10.2006	National University of Uzbekistan, Usbekistan (USB)	15	6044	402,93



Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik – Korpuslinguistik
Spezifikationen des Falko Korpus 2.0 – Version 1.0

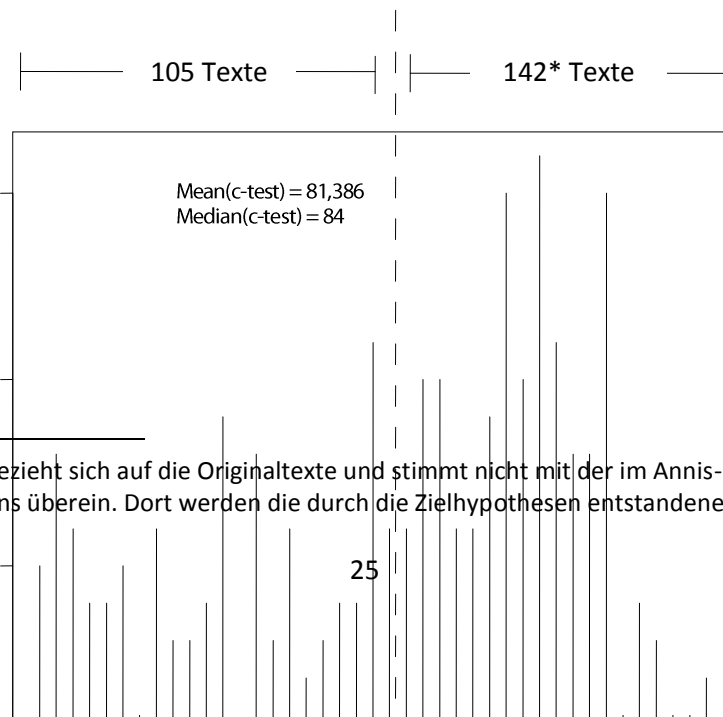
24.10.2006	Freie Universität, Deutschland (FU)	9	10166	442,00
20.11.2007		14		
07.12.2006	Stellenbosch University, Südafrika (SA)	10	6802	680,20
05.03.2007	Auckland University, New Zealand (NZ)	1	2148	537
18.05.2007		1		
25.06.2007		2		
09.08.2007	Humboldt-Universität Ferienkurs, Deutschland (FKB)	12	36559	494,04
20.11.2007		37		
25.07.2008		1		
06.08.2008		24		
Σ		248	122778³	495,08

Tokenanzahl

Texte	248	Lerner	186	Tokens	122.778	Ø/Text	495,08
-------	-----	--------	-----	--------	---------	--------	--------

3.1.3 Verteilung der C-Test-Ergebnisse in FalkoEssayL2 2.0

Insgesamt überwiegt die Zahl der Texte von sehr fortgeschrittenen Lernern (c-test > 80).



³ Diese Tokenanzahl bezieht sich auf die Originaltexte und stimmt nicht mit der im Annis-Interface angezeigten Zahl von 1315 überein. Dort werden die durch die Zielhypothesen entstandenen Leertokens mitgezählt.

³ Diese Tokenanzahl bezieht sich auf die Originaltexte und stimmt nicht mit der im Annis-Interface angezeigten Zahl von 1315 überein. Dort werden die durch die Zielhypothesen entstandenen Leertokens mitgezählt.

* Daten für einen Lerner fehlen

3.2. FalkoEssayL11.2

Auch für die Muttersprachler bestand die Aufgabe darin, zu einem der folgenden vier Themen einen argumentativen Aufsatz zu schreiben:

- Der Feminismus hat den Frauen mehr geschadet als genutzt.
- Kriminalität zahlt sich nicht aus.
- Die meisten Universitätsabschlüsse bereiten die Studenten nicht auf die wirkliche Welt vor. Sie sind deswegen von geringem Wert
- Die finanzielle Entlohnung eines Menschen sollte dem Beitrag entsprechen, der er/sie für die Gesellschaft geleistet hat.

Prüfungskontext

- keine Vorbereitungszeit
- keine Textkenntnis
- keine Hilfsmittel
- in einem Texteditor verfasste Klausuren unter Aufsicht
- Zeit: 90 Minuten

3.2.1 Übersicht über Sprache und Geschlecht der Lerner für die einzelnen Erhebungen

(Mehrfachangaben für Sprachen wurden auch mehrfach gezählt)

Datum der Erhebung	Anzahl der Teilnehmer		L1	L2
	männlich	weiblich		
25.10.06	5	2	Deutsch(7)	Englisch(7) Französisch(5) Latein(3) Spanisch(2) Schwedisch(1) Altgriechisch(1) Russisch(1)



Humboldt-Universität zu Berlin

Institut für deutsche Sprache und Linguistik – Korpuslinguistik

Spezifikationen des Falko Korpus 2.0 – Version 1.0

15.06.2007	8	31	Deutsch(39)	Englisch(39) Französisch(34) Latein(19) Russisch(3) Spanisch(2) Altgriechisch (2) Chinesisch(1)
11.09.2007	5	10	Deutsch(15)	Englisch(15) Latein(12) Französisch(10) Russisch(2) Spanisch(1)
13.09.07	2	11	Deutsch(13)	Englisch(13) Französisch(12) Latein(8) Spanisch(2) Russisch(1)
09.10.2007	7	8	Deutsch(15) Thailändisch(1)	Englisch(15) Französisch(15) Latein(7) Spanisch(1) Japanisch(1)
23.10.07	2	4	Deutsch(6)	Englisch(6) Französisch(6) Spanisch(4) Latein(3) Russisch(1) Jiddisch(1)
Σ	29	66	Deutsch(95) Thailändisch(1)	Englisch(95) Französisch(85) Latein(54) Spanisch(12) Russisch(8) Altgriechisch (3) Jiddisch(1) Japanisch(1) Chinesisch(1)
Σ	95			

3.2.2 Übersicht über die Orte und Textgrößen bezüglich der einzelnen Erhebungen

Erhebungsdatum	Ort	Texte	Token	Token/Text
25.10.06	Freie Universität Berlin, Deutschland (FUD)	7	4670	778,33

15.06.2007	Evangelisches Gymnasium Hermannswerder, Deutschland (DHW)	39	34502	884,67
11.09.2007	Humboldt-Gymnasium Eichwalde, Deutschland (DEW),	15	8828	588,53
13.09.07	Humboldt-Gymnasium Eichwalde, Deutschland (DEW),	13	6399	492,23
09.10.2007	Carl-Siemens-Schule Berlin, Deutschland (DCS)	15	9302	620,13
23.10.07	Freie Universität Berlin, Deutschland (FUD)	6	4790	684,29
Σ		95	68491	720,96

Tokenanzahl

Texte	95	Muttersprachler	95	Tokens	68491	Ø/Text	720,96
-------	----	-----------------	----	--------	-------	--------	--------

4. Richtlinien für die Annotation im Falko-Essay-Korpus 2.0

Das Essaykorpus ist neben dem Summary- und dem Georgetown-Essay-Subkorpus ein Teil des fehlerannotierten Lernerkorpus des Deutschen (Falko). Mit der Version 2.0 wird Falko zum ersten Mal in einem Format vorliegen, das mithilfe des Suchwerkzeugs ANNIS2 (Zeldes et al. 2009) durchsucht werden kann, das speziell für Abfragen tiefannotierter Korpora wie Falko entwickelt wurde.

Diese Version wird vom Augenblick der Publikation an nicht mehr verändert, um die auf ihr basierenden Ergebnisse replizierbar zu halten. Die Daten wurden zwar alle von mindestens zwei Annotatoren kontrolliert, allerdings ist damit zu rechnen, dass auch weiterhin vereinzelt Dinge übersehen wurden.

Wir möchten daher alle Nutzer dazu auffordern, uns Fehler und Verbesserungsvorschläge immer sofort zu melden, damit wir zeitnah verbesserte Versionen herausgeben können.

Das Falko-Essay-Korpus enthält bisher folgende Annotationsebenen:

Annotationsebene	Tag
Basistext reiner Lernertext mit nicht-adjazenten Tokens	[tok]
Lernertext Lernertext, mit adjazenten Tokens	[word]
Wortartenebene für den Lernertext sowie alle drei Zielhypothesen	[pos] ,[ZH1pos],[ZH2pos],[ZHverbpos] (automatisch: Treetagger (Schmidt 1994))
Lemmaebene für den Lernertext sowie alle drei Zielhypothesen	[lemma],[ZH1lemma],[ZH2lemma], [ZHverblemma] (automatisch: Treetagger (Schmidt 1994))
Textgrenzen	[TXTstructure]
Korrigierte Tokenebene: [ctok]	[ctok]
Makrotextuelle Ebene	[macro]

Minimale Zielhypothese Satzebene, Orthografie, Morphosyntax, satzinterne Koreferenz, temporale und modale Kompatibilität	[ZH1]
Maximale Zielhypothese Semantik, Pragmatik, Informationsstruktur, Le- xik, Kollokation	[ZH2]
Zielhypothese für die Annotation komplexer Verben	[ZHverb]
Abweichungsebene für jede der Zielhypothesen	[DiffZH1, DiffZH2 & DiffZHverb]
Komplexe Verben	[verbkategorie, verblemma, verbfehler, verb- form]

4.2 Textgrenzen: [TXTstructure]

Um in ANNIS2 die Anzahl von Texten suchen zu können, in der eine bestimmte Struktur auf-
taucht, wurde auf dieser Ebene jedes erste Token mit dem Tag „start“ und jedes letzte mit dem
Tag „end“ versehen.

Erstes Token im Text

Das erste Token erhält das Tag „start“.

tok	Der	Feminismus	hat	den
TXTstructure	start			

Letztes Token im Text

Das letzte Token erhält das Tag „end“.

tok	müssen	berücksichtigt	werden	.
TXTstructure				end

4.3 Korrigierte Tokenebene: [ctok]

Die Korrigierte Tokenebene stellt die Grundlage für die Berechnung jeglicher Abweichungen
der Zielhypothesen dar. Hier werden lediglich Fehler in der automatischen Tokenisierung ver-
bessert. Die Korrigierte Tokenebene besteht aus einem kompletten Text, der die Änderungen
enthält. Sie erlaubt darüber hinaus die Simulation von Subtokens, also einem einzelnen Token
auf der ctok-Ebene, das mehreren Tokens auf einer Zielhypothesenebene entspricht.

Tokens zusammenfassen

Durch die Tokenisierung fälschlich getrennte Zeichen werden auf der ctok-Ebene in einer Span-
ne zusammengeführt.

tok	das	20	.	Jhd
-----	-----	----	---	-----

ctok	das	20.	Jhd
------	-----	-----	-----

pos-Konstanz

Werden Tokens getrennt, wird darauf geachtet, dass gleiche Wortarten untereinander stehen.

tok	ob	er/		sie	kommt
pos	KOUS	<unknown>		PPER	VVFIN
ctok	ob	er	/	sie	kommt
	KOUS	PPER	&(PPER	VVFIN

Basis für Subtokenisierung auf einer Zielhypothese

Werden in den Zielhypothesen Tokens aus dem Lernertext getrennt, so wird die ctok-Ebene zur Spanne über alle „Subtokens“ der Zielhypothese.

ctok	oder	auf	der	Jungfrau
ZH2	oder	auf	eine	junge Frau

4.4 Makrostrukturebene: [macro]

Auf der Makrostrukturebene werden Textabschnitte gekennzeichnet, die nicht zum eigentlichen Textverlauf beitragen, um sie in einer späteren Suche gesondert behandeln zu können. Hierzu zählt auch die Markierung fremdsprachlichen Materials.

start

Beginnt ein Essay mit der Beschreibung der Umstände, dem Namen oder Datum, so wird dies mit „start“ gekennzeichnet.

ctok	Im	Zuge	des	DaF-Seminars	soll	ff.	These	diskutiert	werden	:
macro	start									

(fu081d_2007_10)

title

Viele Essays beginnen mit der Wiederholung der Fragestellung. Diese wird als „title“ gekennzeichnet. Wird der Titel wortwörtlich später im Text zitiert, so wird er auch dort als „title“ gekennzeichnet.

ctok	Kriminalität	zahlt	sich	nicht	aus	.	Die	Frage	
macro	title								

subtitle

Tauchen in den Texten Zwischenüberschriften auf, so werden sie als „subtitle“ markiert, damit sie beispielsweise für syntaktische Auswertungen ignoriert werden können.

ctok	nicht	ändern	wollen	.	Der	Feminismus	Feminismus	hat	schon
------	-------	--------	--------	---	-----	------------	------------	-----	-------

macro					subtitle			
-------	--	--	--	--	----------	--	--	--

structure

In Texten auftretende Strukturierungszeichen wie Listenzeichen (* - • → etc.) werden als „structure“ gekennzeichnet, um gezielt nach ihnen suchen zu können.

ctok	geht	.	-	Alle	Leute	wissen	viel	von
macro			structure					

citation

Zitate im Lenertext werden als solche mit „citation“ markiert.

ctok	"	Does	feminsit	mean	ugly	,	loud	woman	or	does
macro	citation									

ctok	it	mean	someone	who	stands	up	for	their	own	rights
macro	citation									

ctok	.	I	believe	it	is	is	the	latter	.	"	Margaret	Atwood
macro	citation											

(fk010_2006_07)

comment

Kommentare, die sich auf die Schreibsituation oder die Erhebungssituation beziehen, werden als „comment“ gekennzeichnet. Diese können bei einer Analyse des Textaufbaus ignoriert werden.

ctok	hat	.	Ach	so	!	Also	,
macro			comment				

alternative

Bietet der Lerner in Klammern alternative syntaktische Varianten an, so werden diese mit dem Label (alternative) gekennzeichnet, um für eine spätere syntaktische Verarbeitung besser ausgeschlossen werden zu können.

ctok	jeder	sollte	(=	jedem	ist	erlaubt)	so	viel	herausohlen	als	möglich	ist
macro			alternative											

fm:[Sprache]

Fremdsprachliches Material wird mit dem Tag „fm“ markiert. Ist die Sprache bekannt, wird das passende Kürzel nach einem Doppelpunkt angehängt. Dieses Tag wird nur außerhalb konkurrierender Tags auf dieser Ebene vergeben.

ctok	learning	by	doing
macro	fm:eng		

end

Metatextuelle Zusätze am Ende des Textes, die sich keinen Einfluss mehr auf den Textverlauf haben, werden mit „end“ gekennzeichnet, um sie später gesondert behandeln zu können.

ctok	Vorname	Name	Studierende	und	Sachbearbeiterin
macro	end				

5. Zielhypothesen

Gegenstand:

Eine wichtige Komponente eines Lernerkorpus ist die Fehlerannotation. Um Fehler annotieren zu können, muss man sie identifizieren. Laut Corder(1986:37) müssen dafür die Lerneräußerungen mit den „korrekt“ rekonstruierten Äußerungen verglichen werden.

We identify errors by comparing original utterances with what I shall call reconstructed utterances, that is, correct utterances having the meaning intended by the learner.

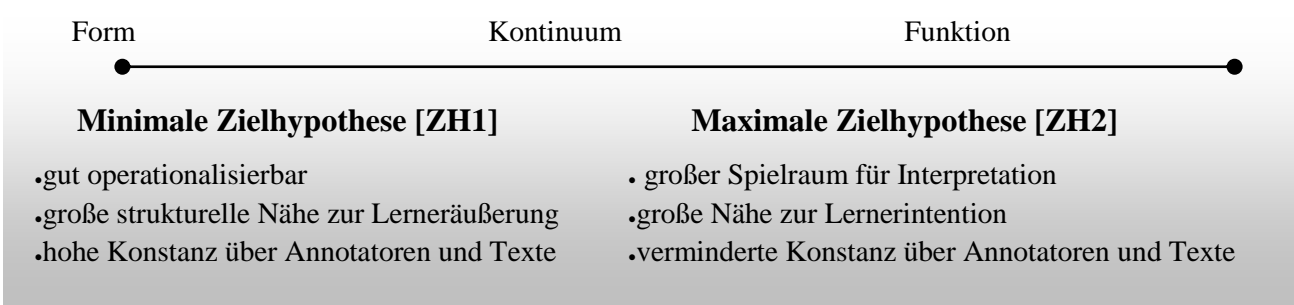
Was man unter der rekonstruierten Äußerung zu verstehen hat, wird erkennbar, wenn man sich Lençons(1991:182) Definition für Fehler ansieht. Unter einem Lernerfehler versteht er „*a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts.*“ Dabei führt der Begriff der „rekonstruierten Äußerung“ leicht in die Irre, denn eine direkte Verbindung zwischen Lerneräußerung und muttersprachlicher Rekonstruktion existiert nicht. Jede Lerneräußerung erlaubt eine Vielzahl gleichwertiger Äußerungen in der Muttersprache. Lüdeling(2008) zeigt in einer Studie mit verschiedenen professionellen Deutschlehrern, dass die Abweichung zwischen „Fehlerannotatoren“ sehr groß ist. Aus diesem Grund spricht sie von einer „Zielhypothese“, welche eine Interpretation der Lerneräußerung durch einen geschulten Annotator darstellt. Drei Konsequenzen erwachsen daraus. Zum Ersten muss eine Zielhypothese explizit sein, damit darauf basierende Fehlerannotationen transparent werden. Zweitens muss für Benutzer des Korpus die Möglichkeit bestehen, eigene, konkurrierende Zielhypothesen zu entwickeln und einzubinden. Drittens wird deutlich, dass die Erarbeitung einer Zielhypothese einer Operationalisierung mit mehr oder weniger engen Grenzen der Interpretation bedarf. Diese Richtlinien liegen hiermit vor.

Mehrere Zielhypothesen

Während der Erarbeitung der Operationalisierung der Falko-Zielhypothese sind zwei widerstrebende Tendenzen deutlich geworden. Je stärker man die Interpretation der Lerneräußerung lenkt und beschränkt, desto weniger Spielraum bleibt dem Annotatoren, die eigentliche Intention, des Lerners sinngemäß zu rekonstruieren. Je mehr von dieser Intention er jedoch versucht in der Zielhypothese widerzuspiegeln, desto weniger Einheitlichkeit verspricht das Resultat. Aus diesem Grund wurde für das Falko-Essay-Korpus entschieden, beide Strategien zu verfolgen, indem zwei verschiedene Zielhypothesen für den gleichen Text erstellt wurden. Eine minimale Zielhypothese versucht, die Anzahl der Abweichungen von der Lerneräußerung zu minimieren und nimmt dafür in Kauf, sich vor allem auf Fehler niedriger sprachlicher Ebenen (Orthografie, Morphologie, Syntax) zu beschränken. Die maximale Zielhypothese versucht eine große Bandbreite von Fehlern auch auf höheren Ebenen (Semantik, Lexik, Pragmatik, Stilistik) einzufangen auf Kosten der Operationalisierbarkeit und Konstanz über Annotatoren und Texte. Auf diese Weise kann eine geschickt ausgenutzte Kombination beider Zielhy-

pothesen bereits ohne weitere explizite Fehlerannotation ein großes Spektrum von Abweichungen der Lernertexte erfassen.

Zusammengefasst handelt es sich bei den Zielhypothesen NICHT um explizite nach Fehlerkategorien unterteilte Fehlerannotationen, sondern um eine implizite Annotation von Abweichungen der Lernertexte von einer postulierten Standardvariante, vor deren Hintergrund eine Fehlerannotation möglich wird und die auf dieser aufbauen kann.



- Die Richtlinien für die Annotation der Zielhypothesen sind in vier Teile unterteilt:
- Allgemeine Information über das technische Vorgehen bei der Annotation der Daten mit MS Excel oder EXMaRALDA⁴.
 - Kriterien für die Annotation der minimalen Zielhypothese [ZH1]
 - Kriterien für die Annotation der maximalen Zielhypothese [ZH2]
 - Richtlinien für die Annotation der Differenzmarkierungen zwischen den Zielhypothesen und der ctok-Ebene

5.1 Technische Vorgaben für die Erstellung der Zielhypothesen

Geringe Korrekturen
 Was nicht geändert werden muss, um eine Version zu erzeugen, die den oben beschriebenen Kriterien entspricht, wird nicht verändert. Je weniger Tokens verändert werden müssen, desto besser.

ctok	Frauen	konnten		solchen		gesellschaftlichen	Zustand	verändern
ZH1	Frauen	konnten	einen	solchen		gesellschaftlichen	Zustand	verändern
*ZH1	Frauen	konnten		solch	einen	gesellschaftlichen	Zustand	verändern

(fk024_2006_07)

Bewegen statt Tauschen
 Zielposition für eine bewegte Konstituente in der Zielhypothese ist immer ein leeres Token. Konstituenten werden nicht getauscht.

ctok	Man	hat		ihr	es	geglaubt	.
------	-----	-----	--	-----	----	----------	---

⁴ siehe Schmidt 2001

ZH1	Man	hat	es	ihr		geglaubt	.
*ZH1	Man	hat		es	ihr	geglaubt	.

Bewegung nach links

Spricht keine der vorangegangenen Regeln dagegen, werden Bewegungen nach links denen nach rechts vorgezogen.

ctok	Man	hat		ihr	es		geglaubt	.
ZH1	Man	hat	es	ihr			geglaubt	.
*ZH1	Man	hat			es	ihr	geglaubt	.

Kurze Bewegungen

Weicht die Position eines Tokens in der Zielhypothese von ihrem Pendant auf der ctok-Ebene ab, so soll der Abstand zwischen beiden Positionen minimal bleiben.

ctok		Die	Frauen	haben		die	Macht	auch	.
ZH1		Die	Frauen	haben	auch	die	Macht		.
*ZH1	Auch	die	Frauen	haben		die	Macht		.

Bewegung leichter Konstituenten

Tokens werden so in einem Satz verschoben, dass möglichst wenige Konstituenten und Tokens davon betroffen sind.

ctok	Es	hat					Geld	seiner	Firma	in	Polen		gebracht	.
ZH1	Es	hat						seiner	Firma	in	Polen	Geld	gebracht	.
*ZH1	Es	hat	seiner	Firma	in	Polen	Geld						gebracht	.

Erst löschen, dann einfügen

Wird ein Token an einer Stelle im Satz eingefügt, an der eine Löschung vorgenommen werden muss, so wird zuerst gelöscht und dann rechts der Löschung eingefügt.

ctok	sondern		sind		sie			auch	.
ZH1	sondern	sie	sind			auch	groß		.
*ZH1	sondern	sie	sind	auch			groß		.

Interpunktionszeichen werden für Regel 6a nicht berücksichtigt.

Wird ein Token an einer Stelle im Satz eingefügt, an der ein Interpunktionszeichen gelöscht wird, so kann das Token links oder rechts eingefügt werden, um eine Zielstruktur zu erhalten, die der Ausgangsstruktur ähnelt.

ctok	Heute	,		jeder	Mensch	hat	seine	Aufgabe	.
ZH1	Heute		hat	jeder	Mensch		seine	Aufgabe	.

ctok	nach	der	Meinung	der	Feministinnen		,	als	altmodisch
ZH1		der	Meinung	der	Feministinnen	nach		als	altmodisch

Token aufsplitten mit zwei unterschiedlichen pos-tags der Zieltoken

Wird ein Token aufgesplittet, geschieht dies so, dass die part-of-speech-Annotation in der ZH1 und die des Ursprungstextes untereinanderstehen, die ctok-Ebene wird dabei über den neuen Tokens verbunden.

ctok	um	neue	Kontakte		anzuknüpfen	
pos	APPR	ADJA	NN		VVIZU	
ctok	um	neue	Kontakte	anzuknüpfen		
ZH1	um	neue	Kontakte	zu	knüpfen	
ZH1pos	APPR	ADJA	NN	PTKZU	VV _{INF}	
*ZH1	um	neue	Kontakte		zu	knüpfen
*ZH1pos1	APPR	ADJA	NN		PTKZU	VV _{INF}

Token verschmelzen

Werden zwei Token miteinander verschmolzen, werden auch die Zellen zu einer Spanne verschmolzen.

ctok	an	dem	Wohlstand
ZH1	am		Wohlstand
*ZH1		am	
*ZH1	am		Wohlstand

Doppelte Sequenzen löschen

Fälschlicherweise verdoppelte Sequenzen werden rechts gelöscht.

ctok	es	nicht	es	nicht
ZH1	es	nicht		
*ZH1			es	nicht

Verschmelzung von zwei nichtadjazenten Tokens

Verschmelzen zwei nichtadjazente Tokens, so wird der eine gelöscht und auf ZH1Diff mit dem Tag „MOVS“ markiert, während der andere die verschmolzene Form darstellt und das Tag „MOVT“ erhält.

ctok	nicht	mehr	einen	Job		.
ZH1			keinen	Job	mehr	.
*ZH1	keinen			Job	mehr	.
ZH1Diff	MOVS		MOVT			

Gleiche Funktion mit anderen Mitteln

Werden in der Zielhypothese bestimmte Satzfunktionen durch andere Phrasentypen erfüllt als im Lernertext, so werden beide Phrasen (wenn möglich) als koindizierte¹ Spannen konstruiert.

ctok	ob	Forschung	frei	oder	mit	einem	gegebenen	Ziel	sein	solle	.
ZH2	ob	Forschung	frei	oder	zielgerichtet				sein	sollte	.
ZH2Diff					MERGE					CHA	

ctok: Zu dieser Frage von Praxis versus Teori hört zusätzlich auch eine Diskussion zwischen Gruppen von Forscher und Gruppen von Politiker und Unternehmer über die Thema ob Forschung frei oder mit einem gegebenen Ziel sein solle.

ZH2: Zu dieser Kontroverse zwischen Praxis und Theorie gehört zusätzlich auch eine Diskussion zwischen Gruppen von Forschern und Gruppen von Politikern und Unternehmern über das Thema, ob Forschung frei oder zielgerichtet sein sollte. (fk021_2006_08.)

ctok	über	sich	selbst	und	ihre	Erwachsenwerdenprobleme				schreiben
ZH2	über	sich	selbst	und	ihre	Probleme	mit	dem	Erwachsenwerden	schreiben
ZH2Diff						SPLIT				

ctok: Plötzlich können sie, über sich selbst und ihre **Erwachsenwerdenprobleme** schreiben - und es ist interessant für die Gesellschaft.

ZH2: Plötzlich können sie über sich selbst und ihre **Probleme mit dem Erwachsenwerden** schreiben und es interessiert die Gesellschaft. (fk022_2006_07.)

5.2 Minimale Zielhypothese (Satzebene, Orthografie, Morphosyntax, satzinterne Koreferenz, temporale und modale Kompatibilität) [ZH1]

Das Ziel der minimalen Zielhypothese ist es, durch minimale Änderungen einen Satz zu konstruieren, der den Regeln der deutschen **Morphosyntax** und **Orthografie** gehorcht. Außerdem wird darauf geachtet, dass die **satzinterne Koreferenz** und temporale sowie modale Kompatibilität gewahrt wird. Abweichungen werden gegenüber der ctok-Ebene vorgenommen. Die Zielhypothese besteht aus einem kompletten Text, der die Änderungen enthält.

Orthografie

Betonung durch Großbuchstaben

In Großbuchstaben geschriebene Silben oder Wörter werden der deutschen Orthografie angepasst.

ctok	Vater	UND	Mutter	arbeiten
ZH1	Vater	und	Mutter	arbeiten

Kreative Orthografie

Kreativer Sprachgebrauch, der zu einer nicht-standardsprachlichen Orthografie führt, wird korrigiert.

ctok: Vor fünfzig Jahre hat diese ganze Frauenbewegung vielleicht noch Sinn gemacht aber sein wir doch mal ehrlich, wann trifft **frau** den heutzutage noch auf einen richtigen Macho?

ZH1: Vor fünfzig Jahren hat diese ganze Frauenbewegung vielleicht noch Sinn gemacht, aber seien wir doch mal ehrlich, wann trifft **eine Frau** denn heutzutage noch auf einen richtigen Macho? (dhw019_2007_06)

Morphologie

Überflüssige Affixe löschen

Trägt ein Wort inkompatible Affixe, so werden die äußeren gelöscht und die inneren beibehalten, wenn kein inhaltlicher Grund dagegen spricht.

ctok	hinzubeifügen
ZH1	beifügen
*ZH1	hinzufügen

Pluralitanta & Singularitanta

Im Singular gebrauchte Plurariatanta und im Plural gebrauchte Singularitanta werden korrigiert.

ctok: Kriminalitäten wo etwas schief geht, und man deshalb immer auf der flucht sein wird, und deshalb nicht das Leben genießen kann obwohl man vielleicht einen Haufen Geld geklaut hat, würde ich nicht als schön bezeichnen.

ZH1: Kriminalität, bei der etwas schief geht und durch die man deshalb immer auf der Flucht sein wird und deshalb das Leben nicht genießen kann, obwohl man vielleicht einen Haufen Geld geklaut hat, würde ich nicht als schön bezeichnen. (cbs002_2007_10)

mündliche Verschmelzungen aus Präposition und Artikel

Im mündlichen Sprachgebrauch übliche Verschmelzungen von Präposition und Artikel werden nicht korrigiert

ctok: Heutzutage ist es selbstverständlich, sich mit dem sogenannten " Ellbogen-Prinzip" durchs Leben zu schlagen.

ZH1: Heutzutage ist es selbstverständlich, sich mit dem sogenannten " Ellbogen-Prinzip" durchs Leben zu schlagen.(dhw_010_2007_06)

Syntax

Abweichung von minimaler Bewegung

Führt eine minimale Bewegung zu einer Struktur, die einerseits sehr stark von der Lernaltersyntax abweicht, die intendierte Zielstruktur aber andererseits nicht gut abbildet, so darf die minimale Bewegung verletzt werden.

ctok	In	Neu	Seeland	auch	haben		wir		eine	Prime	Minister
ZH1	In	Neuseeland			haben		wir	auch	eine	Premministerin	
*ZH1	In	Neuseeland			haben	auch	wir		eine	Premministerin	

ctok: In Neu Seeland auch haben wir eine Frau für Prime Minister.

ZH1: In Neuseeland haben wir auch eine Frau als Premierministerin. (nz002_2007_05)

Konnektoren und Wortstellung

Findet sich in einem Nebensatz eine falsche Kombination aus Konnektorvorgabe und Wortstellung, so wird die Wortstellung auch dann angepasst, wenn eine Veränderung des Konnektors weniger Abweichung insgesamt bedeuten würde.

ctok	Ich	denke	,	dass	unsere	Zivilisation	besteht	aus	Gruppen		.
ZH1	Ich	denke	,	dass	unsere	Zivilisation		aus	Gruppen	besteht	.
*ZH1	Ich	denke	,		unsere	Zivilisation	besteht	aus			.

ctok: Und ich kann es nicht verstehen, dass viele Leute glauben an diese Wesen(Terroristen), die ich als Menschen nicht nennen kann.

ZH1: Und ich kann es nicht verstehen, dass viele Leute an diese Wesen (Terroristen) glauben, die ich nicht Menschen nennen kann. (usb015_2006_10)

Konkurrierendes „dass“ und Infinitiv mit „zu“

Tauchen in einem Satz der „dass“-Subordinierer und ein Infinitiv mit „zu“ auf, wird die Variante mit der kleineren Abweichung gewählt.

ctok: Aber... Einige streben danach, dass noch reicher zu sein, noch gewaltiger;

ZH1: Aber: Einige streben danach, noch reicher zu sein und noch gewaltiger;

(usb015_2006_10)

Verbstellung subordinierten Nebensätzen

In subordinierten Nebensätzen wird Verbzweitstellung korrigiert. Stellungsvarianten mit eingebettetem V2-Satz nach „dass“, wie sie etwa Freywald (2008,2009) auch bei Muttersprachlern findet, werden hier nicht berücksichtigt.

ctok: Die reiche Studenten möchten nicht studieren" - hatte ich gesagt, aber ist das falsche Meinung, weil viele reichen Studenten **erwerben** in diesem Moment in großem Ergebnisse als arme Studenten.

ZH1: Die reichen Studenten möchten nicht studieren", hatte ich gesagt, aber das ist eine falsche Meinung, weil viele reiche Studenten in diesem Moment größere Ergebnisse als arme Studenten **erwerben**.(usb007_2006_10)

Linke Satzklammer beibehalten

In Hauptsätzen wird das finite Verb als linke Satzklammer behandelt. Diese wird nicht bewegt.

ctok	Heute		Männer	gehen		nicht	oft	einkaufen	.
ZH1	Heute			gehen	Männer	nicht	oft	einkaufen	.
*ZH1	Heute	gehen	Männer			nicht	oft	einkaufen	.

Linke Satzklammer erzeugen

Wird aus einem Originalsatz mit besetzter linker (LSK) und rechter Satzklammer (RSK) eine Zielhypothese erstellt, in der lediglich die LSK besetzt ist, wird der Inhalt der ehemaligen LSK gelöscht und das Token der ehemaligen RSK rechts von der Löschung eingefügt.

ctok: So wie es keine einzige Bewegung, um das adequate Verhältnis des Männerns zu Frauen beschreiben **gibt**, so sollte es keinen einzigen Feminismus sein.

ZH1: So wie es keine einzige Bewegung **gibt**, um das adäquate Verhältnis der Männern zu Frauen beschreiben, so sollte es kein einziger Feminismus sein.
(fk010_2006_07)

Rechte Satzklammer

Die rechte Satzklammer hat keinen besonderen Status in Bezug auf die Annotation, sodass das finite Verb in diesem Fall bewegt werden darf.

ctok	was	ist	darüber	gemeint		.
ZH1:	was		damit	gemeint	ist	.
ZH1Diff		MOVS	CHA		MOVT	

ctok: Wenn wir über" Feminismus" oder" die Interessen der Frauen" reden, sollen wir erklären, was ist darüber gemeint.

ZH1: Wenn wir über" Feminismus" oder" die Interessen der Frauen" reden, sollen wir erklären, was damit gemeint ist. (fkb031_2008_07)

Verbale Kongruenzhierarchie

Bei fehlender Kongruenz zwischen Verb, Kopula und Komplementen unterschiedliche Varianten einer Zielhypothese zu, so wird gemäß der folgenden Hierarchie versucht, Veränderungen zu vermeiden und die gegebene Information zu erhalten:

höchste Priorität



Finites Verb

Akkusativ-Objekt

niedrigste Priorität Dativobjekt
 Präpositionalobjekt
 Subjekt

ctok	dass	unsere	Zivilisation	sich	aus	Gruppen	besteht	.
ZH1	dass	unsere	Zivilisation		aus	Gruppen	besteht	.
*ZH1	dass	unsere	Zivilisation	sich	aus	Gruppen	zusammensetzt	.

ctok: In diesem Fall kann ich nur über mein Land sagen, aber vermutlich sieht diese Situation auch in anderen Ländern ähnlich.

ZH1: In diesem Fall kann ich nur **etwas** über mein Land sagen, aber vermutlich sieht diese Situation auch in anderen Ländern ähnlich. (fkb028_20)

ctok: Man sollte lieber die Ausbildung und das ganze Wissen, was man" im Rucksack" mit sich hat wenn man fertig ist, als ein Werkzeug betrachten.

ZH1: Man sollte lieber die Ausbildung und das ganze Wissen, das man" im Rucksack" mit hat, wenn man fertig ist, als ein Werkzeug betrachten. (cbs003_2006_09)

Valenz
 Valenzen werden gemäß der Darstellung in E-VALBU (Schneider 2008) angenommen, hierbei wird nur die standardsprachliche Variante akzeptiert.

ctok: Wenn wir über "Feminismus" oder "die Interessen der Frauen" reden, sollen wir erklären, was ist **darüber gemeint**.

ZH1: Wenn wir über "Feminismus" oder "die Interessen der Frauen" reden, sollen wir erklären, was **damit gemeint** ist. (fkb031_2008_07)

ctok: Die Universitäten sollten praxisorientiert sein, damit die Studenten, wenn sie auf dem Arbeitsmarkt kommen, **auf die Herausforderungen gut gewappnet** sind.

ZH1: Die Universitäten sollten praxisorientiert sein, damit die Studenten, wenn sie auf den Arbeitsmarkt kommen, **für die Herausforderungen gut gewappnet** sind. (cbs006_2007_10.)

Kasusreaktion durch Präpositionen
 Kasusreaktion durch Präpositionen wird gemäß der Darstellung in Grammis (Breindl 2000) behandelt, hierbei wird nur die standardsprachliche Variante akzeptiert

ctok: Während **den letzten Jahrhundert** haben die Feministen eine wichtige Rolle gespielt, sie haben geholfen Frauen eine Stimme und eine Chance zu geben.

ZH1: Während **des letzten Jahrhunderts** haben die Feministen eine wichtige Rolle gespielt, sie haben geholfen den Frauen eine Stimme und eine Chance zu geben. (fk008_2006_07)

Korrelate
 Für die Entscheidung, welche Korrelate obligatorisch, fakultativ oder unzulässig sind, orientie-

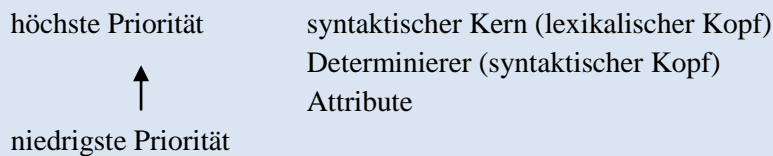
ren wir uns an den Listen in Engel (1988: 255ff.) .

ctok: Man kann es ja behaupten, dass ohne ihre Bemühungen, hätten die Frauen noch heute kein Wahlrecht.

ZH1: Man kann es ja behaupten, dass ohne ihre Bemühungen, die Frauen noch heute kein Wahlrecht hätten. (fkb037_2008_07)

Nominale Kongruenzhierarchie

Bei fehlender Kongruenz zwischen Determinierern, Attributen und Nominalkern wird vom Kern ausgegangen. Danach wird zugunsten des Determinierers entschieden.



ctok	keine	Arbeit	bekommen	,	die	ihrer	Qualifikationen	entspricht	.
ZH1	keine	Arbeit	bekommen	,	die	ihren	Qualifikationen	entspricht	.
*ZH1	keine	Arbeit	bekommen	,	die	ihrer	Qualifikation	entspricht	.

ctok: Das Problem entsteht auch, weil man zwischen **öffentliche** und **private** Unternehmen entscheiden muss.

ZH1: Das Problem entsteht auch, weil man **sich** zwischen **öffentlichen** und **privaten** Unternehmen entscheiden muss. (fkb010_2008_07)

Wortstellung im Mittelfeld

Abfolgen von Konstituenten im Mittelfeld werden nur dann korrigiert, wenn dadurch ungrammatische Strukturen entstehen. Wortstellungstendenzen werden ignoriert (vgl. Eisenberg 2006:405ff.).

ctok	Er	hat		meiner	Firma	es	gesagt	.
ZH1	Er	hat	es	meiner	Firma		gesagt	.
*ZH1	Er	hat		meiner	Firma	es	gesagt	.

ctok: Aber vielleicht ist das Problem auch, dass man **nicht** auf eine Universität alles lernen kann.

ZH1: Aber vielleicht ist das Problem auch, dass man auf einer Universität **nicht** alles lernen kann. (cbs003_2007_10)

ctok: Einerseits muss ich sagen, dass jeder Student muss unbedingt gute theoretische Grundkenntnisse haben, weil ohne diesen er keine Chance um ein(egal ob Mathematisch- oder soziales) Problem zu Lösen hat.

ZH1: Einerseits muss ich sagen, dass jeder Student unbedingt gute theoretische Grundkenntnisse haben muss, weil er ohne diese keine Chance hat, ein(egal ob mathematisches oder soziales) Problem zu lösen. (fkb011_2007_09)

ABER:

ctok: Also für mich **am wichtigsten an der Universität** ist nicht der Abschluss, sondern die Erfahrung. (fkb010_2007_09)

ZH1: Also, für mich ist **am wichtigsten an der Universität** nicht der Abschluss, sondern die Erfahrung. (fkb010_2007_09)

Koordinierte Strukturen nicht teilen

Führt eine kurze Bewegung zur Teilung koordinierter Strukturen, wird eine längere Bewegung bevorzugt.

wor d	ob- wohl	ich	erkläre	meine	Mei- nung		un d		von	meinen	Freunden	
ZH1	ob- wohl	ich		meine	Mei- nung		un d	die	von	meinen	Freunden	erkläre
*ZH 1	ob- wohl	ich		meine	Mei- nung	erkläre	un d	die	von	meinen	Freunden	

Verschmelzung von Präposition und Artikel

Die Verschmelzung von Artikel und Präposition wird da korrigiert, wo eine alternative Lesart mit definitivem Artikel ausgeschlossen ist.⁵

ctok	diese	weiblichen	Kulturen	bleiben	oft	in	dem	Hintergrund
ZH1	diese	weiblichen	Kulturen	bleiben	oft	im		Hintergrund

Semantisch konfligierendes Tempus & Modus

In Fällen, in denen Tempus oder Modus semantisch oder pragmatisch durch eine Äußerung gefordert wird, wird nicht korrigiert.

ctok	Erst	verliert	er	das	Geld	und	dann	floh	er
ZH1	Erst	verliert	er	das	Geld	und	dann	floh	er
*ZH1	Erst	verlor	er	das	Geld	und	dann	floh	er

ctok: So **verhält** es sich jedenfalls früher in Dänemark

ZH1: So **verhält** es sich jedenfalls früher in Dänemark (cbs005_2007_10)

Morphosyntaktisch konfligierendes Tempus & Modus

In Fällen, in denen Tempus oder Modus morphosyntaktisch durch eine Äußerung gefordert wird, wird korrigiert. Das ist immer dann der Fall wenn die zu verbessernde Äußerung durch eine weitere Äußerung modifiziert wird, die selbst **ein** finites Verb beinhaltet.

⁵ vgl. Eisenberg 2006, S.198ff.

ctok: Weiter haben die Studenten dabei auch die Möglichkeit später zum Forscher zu studieren was sonst schwierig wäre, wenn die Studenten nicht schon beim Studium wissenschaftlich fundiert gewesen **sein**.

ZH1: Weiterhin haben die Studenten dabei auch die Möglichkeit, später auf den Beruf des Forschers hin zu studieren, was sonst schwierig wäre, wenn die Studenten nicht schon beim Studium wissenschaftlich fundiert gewesen **wären**.
(cbs007_2007_10)

ctok: Hier soll auch daran gedacht werden, dass die Studenten und Studentinnen vielmehr praxisorientierter denken würden, wenn sie ein Studienjob **hatten**, die mit ihrem Studium zu tun **hatte**.

ZH1: Hier soll auch daran gedacht werden, dass die Studenten und Studentinnen viel praxisorientierter denken würden, wenn sie einen Studienjob **hätten**, der mit ihrem Studium zu tun **hätte**. (cbs016_2006_09)

ctok: Ich finde, es wäre schön, **dass** die Universitäten praxisorientiert **wurden**,

ZH1: Ich finde, es wäre schön, **wenn** die Universitäten praxisorientiert **würden**,
(fkb061_2008_08)

ctok: Als diese Gesellschaft (noch?) nicht existiert, **kann** das Antwort nur auf der Ebene der Theorie gegeben wird.

ZH1: Als diese Gesellschaft(noch?) nicht existierte, **konnte** die Antwort nur auf der Ebene der Theorie gegeben werden.

ctok: Und jedes Mal denk ich mir," ich **wünsche**, es wäre nicht so, ich **wünsche**, ich **konnte** diese Frage beantworten.

ZH1: Und jedes Mal denke ich mir:" ich **wünschte**, es wäre nicht so, ich **wünschte**, ich **könnte** diese Frage beantworten. (fk004_2006_08)

Artikel

Fehlende Artikel werden nur eingefügt, wenn eine falsche Verwendung zu Ungrammatikalität führt. Kann für die Korrektur ein definitiver und ein indefinitiver Artikel gleichermaßen verwendet werden, fällt die Entscheidung zugunsten des indefiniten Artikels.

ctok: Nächstes Jahr werde ich mein Studium enden und ich neide meinen Freundinnen aus Gymnasium, die sich für Jura- oder Medizinstudium entschieden.

ZH1: Nächstes Jahr wird mein Studium enden und ich beneide meine Freundinnen aus dem Gymnasium, die sich für ein Jura- oder Medizinstudium entschieden.
(fk015_2006_08)

ABER:

ctok: Seit dem Anfang des zwanzigsten Jahrhunderts ist Feminismus einer der umstrittensten Themen gewesen, am wenigstens in demokratischen Länder.

ZH1: Seit dem Anfang des zwanzigsten Jahrhunderts ist Feminismus eines der umstrittensten Themen gewesen, wenigstens in demokratischen Ländern.
(fkb037_2008_07)

ctok: In eigenen Teilen der Welt ist heute Feminismus zu radikal geworden
(fkb016_2008_07)

ZH1: In eigenen Teilen der Welt ist heute Feminismus zu radikal geworden

Bezugsnommen und Ergänzungssatz

Bei falscher Kombination aus einem Bezugsnomen und einem Ergänzungssatz, wird letzterer in der ZH1 angepasst.

ctok: Ich denke, dass die Qualität und Zahl von praxisorientierten Studiengänge und Unterrichten vor allem von dem Land, der Anzahl von Universitäten, der Leitungskader, Staats- und Einrichtungstradition oder der Tatsache, **ob** es eine private oder staatliche Universität ist, abhängig ist.

ZH1: Ich denke, dass die Qualität und Zahl von praxisorientierten Studiengängen und Unterrichten vor allem vom Land, der Anzahl von Universitäten, der Leitungskader, der Staats- und Einrichtungstradition oder der Tatsache, **dass** es eine private oder staatliche Universität ist, abhängig ist. (fkb011_2007_09)

Hanging topic-Ausdrücke

Isoliertstehende Ausdrücke mit „hanging topic“ werden mit einem Komma verbunden.

ctok: Und was die politische Welt betrifft? Die erste berühmte Öffentliche Frauen waren Feministinnen

ZH1: Und was die politische Welt betrifft, die ersten berühmten öffentlichen Frauen waren Feministinnen (fk012_2006_07)

Referenz

Referenz innerhalb eines Matrixsatzes

Wird innerhalb eines Satzes durch Anaphern und Kataphern falsch verwiesen, wird dies korrigiert. Falsche Referenz über die Satzgrenze hinaus, wird nicht korrigiert.

ctok	Wenn	man	isst	,	muss	er	auch	trinken
ZH1	Wenn	man	isst	,	muss	man	auch	trinken
*ZH1	Wenn	man	isst	,	muss	er	auch	trinken

ctok: Wenn man aber jetzt keine Gruppenarbeit mehr haben und wenn man als Student kein Praxis erreichen, könnte man sagen, dass **sie** dann die Anforderungen der Wirtschaft erfüllen.

ZH1: Wenn man aber jetzt keine Gruppenarbeit mehr hat und wenn man als Student keine Praxis erreicht, könnte man sagen, dass **man** dann die Anforderungen der Wirtschaft erfüllt. (cbs005_2007_10)

Referenzverdoppelung innerhalb eines Matrixsatzes

Fehlt in einem komplexen Satz ein Referent als Subjekt oder Objekt und taucht es im zweiten Teilsatz auf, so wird die volle NP an erster Stelle eingefügt und an der zweiten Stelle pronominalisiert.

ctok: Mehrere Lehrer anspornen nicht, wenn **die Studenten** arbeiten.

ZH1: Mehrere Lehrer spornen **die Studenten** nicht an, wenn **sie** arbeiten.
(fkb044_2008_08)

Referenz über den Matrixsatz hinaus

Die Referenz wird nur dann über den Matrixsatz hinaus angeglichen, wenn innerhalb eines Matrixsatzes zwei mögliche Versionen denkbar sind.

ctok: Es kann passieren dass die Leute verändern sich und auch seine Meinungen im Knast.

Aber bekommen sie noch eine andere Chance vom Leben und vom Bekannten?

Wird **ihm** jemand helfen, oder werden **sie** kein Chance haben sondern nur eins - noch einmal ausprobieren ein besserer und kluger Kriminalist zu werden?

ZH1: Es kann passieren, dass die Leute im Knast sich und auch ihre Meinungen verändern.

Aber bekommen sie noch eine andere Chance vom Leben und vom Bekannten?

Wird **ihnen** jemand helfen oder werden **sie** keine Chance haben, sondern nur eins - noch einmal ausprobieren ein besserer und kluger Kriminalist zu werden?

(fkb005_2007_09)

Lexik

Fremdsprachliches Material

Fremdsprachliches Material außerhalb von Klammern wird wenn möglich übersetzt.

ctok: Die Stunden an den dänischen Unis, sollten nicht zu lange Vorträge von das gelesene beinhalten, und mehr Übungen in den Stunden einbauen lassen in Form von z. B Cases.

ZH1: Die Stunden an den dänischen Unis sollten nicht zu lange Vorträge des Gelesenen beinhalten und mehr Übungen in die Stunden einbauen lassen z. B. in Form von Fällen. (cbs006_2007_10)

Semantik

Semantik

Die Semantik wird bei der Korrektur nur dann in Betracht gezogen, wenn eine alternative Korrektur deutlich kompliziertere Veränderungen mit sich brächte, und die außerdem der Intention des Lerners widerspräche. In diesem Fall dürfen auch finite Verben ersetzt werden.

ctok: Man sagt, dass die Arbeitsbedingungen fuer Arbeitseinsteiger können nicht verbessert werden, es sei denn der Staat **passt** neue Arbeitsmarkt regelnde Gesetze.

ZH1: Man sagt, dass die Arbeitsbedingungen für Arbeitseinsteiger nicht verbessert werden können, es sei denn, der Staat **verabschiedet** neue arbeitsmarktregelnde Gesetze. (fk016_2006_08)

ctok: Es passiert aber oft im Berufswelt, dass junge Leuten mit einem oder mehr Grad das Berufswelt hineingehen und sich nicht in den Konkurrenz für Jobs **gelingen**.

ZH1: Es passiert aber oft in der Berufswelt, dass junge Leute mit einem oder mehreren Graden in die Berufswelt hineingehen und sich nicht gegen die Konkurrenz um Jobs **durchsetzen**. (sa002_2006_09)

ctok: Der Feminismus in solchen Situation zu nützen schadet die gute Konzepte der Feminismus sehr und ist ein Beispiel, darin die Konzepte des Feminismus dumm ausgesehen werden, nur denn sie werden in die falsche Situation **benutzt**.

ZH1: Den Feminismus in einer solchen Situation zu benutzen, schadet den guten Konzepten des Feminismus sehr und ist ein Beispiel dafür, dass die Konzepte des Feminismus dumm aussehen werden, nur weil sie in der falschen Situation **benutzt werden**. (fkb003_2008_07)

ctok: In Großbritannien, zum Beispiel, hat die Blair-Regierung das Ziel **genommen**, die Anzahl der Hochschulstudenten bis 2010 zu verdoppeln.

ZH1: In Großbritannien, zum Beispiel, hat die Blair-Regierung das Ziel **anvisiert**, die Anzahl der Hochschulstudenten bis 2010 zu verdoppeln. (hu001_2006_10)

ctok: Sie **wachsen** sowohl Kind als auch arbeiten.

ZH1: Sie **erziehen** Kinder und arbeiten auch. (trk007_2006_05)

ctok: Der Beitrag, den ein Mensch für die Gesellschaft leistet, sollte in Acht **genommen werden**

ZH1: Der Beitrag, den ein Mensch für die Gesellschaft leistet, sollte zur Kenntnis **genommen werden** (fkb014_2008_07)

5.2.1 Auf der Ebene der minimalen Zielhypothese [ZH1] NICHT annotierte Fehler

Syntax

Verberstdeklarativsätze ohne Subjekt

Verberstdeklarativsätze ohne Subjekt, in denen Topikdrop stattgefunden haben kann, werden nicht korrigiert, indem ein Subjekt eingesetzt wird.

ctok: Sie stossen auf Situationen wo sie nuicht genau wissen was sie tun sollten. **Fühlen** sich etwas orientierungslos.

ZH1: Sie stoßen auf Situationen, in denen sie nicht genau wissen, was sie tun sollten.

Fühlen sich etwas orientierungslos. (cbs006_2007_10)

Verblose Sätze

Verblose Sätze, die nicht als Paraphrase an den vorhergehenden oder den nachfolgenden Satz angehängt werden können, werden nicht korrigiert.

ctok: Ins Besondere die sprachliche Gewinne (natürlich), aber nicht nur die!

ZH1: Insbesondere die sprachlichen Gewinne (natürlich), aber nicht nur die!
(cbs006_2006_09)

ctok: Falls ich mich, während des Studiums, am besten für eine Stelle und ein Berufstätigkeit, vorbereiten wollte, sollte ich mich viel mehr mit Bauhandlungen, Vermassung, Bauökonomie und Präsentationen von Projekten beschäftigen. **Überhaupt** nicht mit den Raumtheorien.

ZH1: Falls ich mich während des Studiums am besten auf eine Stelle und eine Berufstätigkeit vorbereiten wollte, sollte ich mich viel mehr mit Bauhandlungen, Vermassung, Bauökonomie und Präsentationen von Projekten beschäftigen, **überhaupt nicht mit den Raumtheorien.** (fk022_2006_08)

Uneingeleitete Nebensätze

Uneingeleitete Nebensätze werden nicht korrigiert.

ctok: Ich meine, die staatlichen Arbeiter leisten der Gesellschaft auf keinen Fall einen geringeren Beitrag.

ZH1: Ich meine, die staatlichen Arbeiter leisten der Gesellschaft auf keinen Fall einen geringeren Beitrag. (fkb028_2008_07)

Isoliert stehende Nebensätze

Isoliert stehende Nebensätze werden nicht korrigiert.

ctok: Deswegen sollen sie ein bisschen besser vom Staat unterstützt werden.

ZH1: Deswegen sollen sie ein bisschen besser vom Staat unterstützt werden.
(fkb028_2008_07)

Uneingeleitete Nebensätze: Zweiteilige Konnektoren

Zweiteilige Konnektoren wie „zwar... , aber“, „je...desto“, „nicht nur..., sondern auch“ müssen in einem Matrixsatz verbunden werden.

ctok: Zwar ist es wahr, dass noch viele Frauen zu Hause bleiben und sich mit wenig mehr als Kinder und Haushalt beschäftigen. Es ist aber die Entscheidung die sie selbst getroffen haben - im Prinzip könnten sie auch etwas Anderes mit ihrem Leben tun.

ZH1: Zwar ist es wahr, dass noch viele Frauen zu Hause bleiben und sich mit wenig mehr als Kindern und Haushalt beschäftigen, es ist aber die Entscheidung, die sie

selbst getroffen haben - im Prinzip könnten sie auch etwas anderes mit ihrem Leben tun. (fkb037_2008_07)

Semantik

Semantik wird ignoriert

Bis auf Ausnahmen gemäß Regel 21 wird die Semantik bei der Erstellung der ZH1 nicht berücksichtigt.

ctok: In den meisten Fällen aber sieht man, dass eine Frau ungefähr alles machen muss.
ZH1: In den meisten Fällen aber sieht man, dass eine Frau ungefähr alles machen muss. (fk006_2006_07)

ctok: Beide Naturwissenschaften und Geisteswissenschaften sind, wenn man richtig darüber überlegt, akademische Fächer ohne eine gewisse Orientierung.
ZH1: Beide Naturwissenschaften und Geisteswissenschaften sind, wenn man richtig überlegt, akademische Fächer ohne eine gewisse Orientierung. (fk014_2006_08)

Lexik

Neologismen und falsche Bedeutungen

Morphosyntaktisch korrekte Neologismen und semantisch falsch verwendete Ausdrücke werden **nicht** korrigiert.

ctok: Ich finde diesen Zitat sehr schön, weil er erlaubt ein Raum für eine kreative Bildung des **Frauentums**, die nicht von einer patriarchalischen Gesellschaft diktiert ist und die nicht von einen einseitigen Sicht an Feminismus diktiert ist.
ZH1: Ich finde dieses Zitat sehr schön, weil es einen Raum für eine kreative Bildung des **Frauentums** erlaubt, die nicht von einer patriarchalischen Gesellschaft diktiert ist und die nicht aus einer einseitigen Sicht vom Feminismus diktiert ist. (fk010_2006_07)

ctok: Journalisten haben photographien von Frauen gemacht während sie Gefängnisse brutalischerweise schlugen.
ZH1: Journalisten haben Fotografien von Frauen gemacht, während sie Gefängnisse brutalerweise schlugen. (fk010_2006_07)

ctok: Sie meinten, dass Männer die mit der kultur beschäftigt sind sind Gewaltiger und Unrechter als Frauen (weil Kultur kann auch zum Krieg führen), die mit der "schöne Natur" verbunden sind.
ZH1: Sie meinten, dass Männer, die mit der Kultur beschäftigt sind, gewaltiger und ungerechter sind als Frauen (weil Kultur auch zum Krieg führen kann), die mit der "schönen Natur" verbunden sind. (fk010_2006_07)

Relativpronomen „was“

Relativpronomen „was“, das sich nicht auf ein sächliches Demonstrativ- oder Indefinitpronomen, ein sächliches unbestimmtes Zahladjektiv oder ein sächliches substantiviertes Adjektiv im Superlativ bezieht, wird korrigiert.

ctok: Man sollte lieber die Ausbildung und das ganze Wissen, was man" im Rucksack" mit sich hat wenn man fertig ist, als ein Werkzeug betrachten.

ZH1: Man sollte lieber die Ausbildung und das ganze Wissen, das man" im Rucksack" mit hat, wenn man fertig ist, als ein Werkzeug betrachten. (cbs003_20)

Nichtstandardsprachliche Relativpronomen

Nichtstandardsprachliche Relativpronomen werden nicht korrigiert.

ctok: In einer Gesellschaft wie die dänische, **wo** alle in unserem Sozialstaat teilnehmen, wird es übel aufgenommen, wenn einige Leute versucht das System zu entkommen.

ZH1: In einer Gesellschaft wie der dänischen, **wo** alle an unserem Sozialstaat teilnehmen, wird es übel aufgenommen, wenn einige Leute versuchen dem System zu entkommen. (cbs008_2006_09)

ctok: das Studiums sollte Studenten für eine Welt bereit machen, **wo** Kapitalismus und Wirtschaft unseres Leben reguliert.

ZH1: Das Studiums sollte Studenten für eine Welt bereit machen, **wo** Kapitalismus und Wirtschaft unser Leben regulieren. (fk004_2006_08)

Lexematisch geblockte morphologisch zulässige Varianten werden nicht korrigiert

Fremdsprachliches Material außerhalb von Klammern wird wenn möglich übersetzt.

ctok: Oder man probiert aus **Leichtsinnigkeit** illegale Drogen aus.

ZH1: Oder man probiert aus **Leichtsinnigkeit** illegale Drogen aus. (dew09_2007_09_ta)

Pragmatik

Sprechereinstellungen

Fehlende Mittel zum korrekten Ausdruck von Sprechereinstellungen werden nicht verbessert.

ctok: Wir sollen, also, die Qualität und Quantität unterscheiden.

ZH1: Wir sollen also die Qualität und Quantität unterscheiden. (usb001_2006_10)

5.3 Maximale Zielhypothese [ZH2]

(Textebene, Semantik, Pragmatik, satzübergreifende Koreferenz, informationsstrukturelle Gliederung, Stil)

Das Ziel der maximalen Zielhypothese ist es, ein möglichst weites Spektrum dessen zu erfassen, was in der Fehlerliteratur unter dem Begriff „acceptability“ gefasst wird. Die maximale Zielhypothese ist derart, dass sie dem Annotatoren muttersprachlich vorkommt. Hierzu gehört, dass sowohl der semantische als auch der pragmatische Gehalt sowie die informationsstrukturelle Gliederung des Satzes im Kontext des gesamten Textes korrekt bzw. angemessen ist. Nicht verbessert wird eine aus Sicht der Annotatoren falsche oder widersinnige Interpretation oder Einstellung bezüglich des besprochenen Themas im Aufsatz. Dass hier teilweise Überschneidungen auftauchen können, ist uns bewusst.

Die Vielzahl möglicher, notwendiger und unvermeidlich subjektiver Abweichungen vom Lernertext, die in die maximale Zielhypothese mit einbezogen werden, macht eine vollständige Liste von Annotationsregeln für diese unmöglich. Aus diesem Grund sollen hier lediglich eine grobe Leitlinie und einige exemplarische Entscheidungen für die Annotation aufgeführt werden. Bei der Arbeit mit dieser Zielhypothese muss man sich der Subjektivität und nur teilweise Reproduzierbarkeit durch andere Annotatoren bewusst sein.

Syntax

Verberstdeklarativsätze ohne Subjekt

Verberstdeklarativsätze ohne werden korrigiert, indem ein Subjekt eingesetzt wird.

ctok: Sie stossen auf Situationen wo sie nicht genau wissen was sie tun sollten. **Fühlen** sich etwas orientierungslos.

ZH2: Sie stoßen auf Situationen, in denen sie nicht genau wissen, was sie tun sollten. **Sie** fühlen sich etwas orientierungslos. (cbs006_2007_10)

Uneingeleitete Nebensätze

Uneingeleitete Nebensätze werden nicht korrigiert, wenn sie nicht selbst mit einem eingebetteten Nebensatz beginnen, wenn Verbletzstellung vorliegt oder wenn das Verb typischerweise mit eingeleiteten Nebensatz vorkommt.

ctok: Ich meine, es lohnt sich nicht zu denken, dass Terror der einzige Ausgang ist.

ZH1: Ich meine, es lohnt sich nicht zu denken, dass Terror der einzige Ausgang ist.

ZH2: Ich meine, es lohnt sich nicht zu denken, dass Terror die einzig mögliche Folge ist. (usb015_2006_10)

ctok: Mein Fazit lautet deshalb, wenn man im Leben Schwein hat, kann sich auch Kriminalität lohnen,

ZH2: Mein Fazit lautet deshalb, **dass** sich, wenn man im Leben Schwein hat, auch Kriminalität lohnen kann, (cbs002_2007_10)

ctok: Deshalb finde ich nicht, dass man sagen kann, die Universitätsabschlüsse von einem geringeren Wert sind.

ZH2: Deshalb finde ich nicht, dass man sagen kann, **dass** die Universitätsabschlüsse von einem geringeren Wert sind (cbs003_2007_10)

ctok: Man soll auch anerkennen, ein Universitätsabschluss **bereitet** die Jugend auf die Welt vor, weil die meisten einflussreichen Menschen der Welt schön einen Abschluss haben.

ZH2: Man sollte auch anerkennen, **dass** ein Universitätsabschluss die Jugend auf die Welt vorbereitet, weil die meisten einflussreichen Menschen der Welt schon einen Abschluss haben. (fu120_2006_10a)

ctok: Beim Abschluss erkennt der Student, er **kann** schwierige Ziele erreichen und sein Selbstwert nimmt zu.

ZH2: Beim Abschluss erkennt der Student, **dass** er schwierige Ziele erreichen **kann** und sein Selbstwert nimmt zu. (fkb055_2008_08)

Besetzung des Nachfeldes

Umfangreiche Präpositionalphrasen, Vergleichssätzen, Relativsätze und Appositionen werden ins Nachfeld ausgelagert.

ctok: Oder Vorteile besser zu sagen. Wir können, sogar müssen, das selbe wie die Männer schaffen. (fkb034_2008_07)

ZH2: Oder, um eher die Vorteile zu nennen: Wir können und müssen sogar, dasselbe schaffen wie die Männer. (fkb034_2008_07)

„dass“-Sätze und Infinitivsätze mit „zu“

Ein Objektsatz, in dem das Subjekt einen Referenten des Hauptsatzes wiederaufnimmt, wird durch Infinitivsatz mit „zu“ konstruiert.

ctok: Sie suchen nach der Lösung und anklagen die Universitäten, dass die keine gute Spezialisten bereiten können.

ZH2: Sie suchen nach der Lösung und klagen die Universitäten an, keine guten Spezialisten auszubilden. (fkb044_2008_08)

Semantik

Verb ersetzen

Anders als in der ZH1 hat das Verb in der ZH2 keinen besonderen Status. In Funktionsverbgefügen bildet der nominale Teil den Kern, um den korrigiert werden soll.

ctok: Vielleicht auch manche Frauen würden gerne diese Meinung vertreten, vor allem die, die anspruchsvolle Arbeit ausüben müssen um Geld zu gewinnen.

ZH2: Vielleicht würden auch manche Frauen gerne diese Meinung vertreten, vor allem die, die anspruchsvolle Arbeit ausüben müssen, um Geld zu verdienen. (fkb030_2008_07)

ctok: **Oder Vorteile besser zu sagen**. Wir können, sogar müssen, das selbe wie die Männer schaffen. (fkb034_2008_07)

ZH2: **Oder, um eher die Vorteile zu nennen**: Wir können und müssen sogar, dasselbe schaffen wie die Männer. (fkb034_2008_07)

Definitheit

Der fehlerhafte Ausdruck einer kontextabhängigen Definitheit von Referenten, wird korrigiert.

ctok: Das ist wahr, dass es nicht so für die Frauen in allen Ländern ist, aber das ist keine Folge **von** Feminismus.

ZH2: Es ist wahr, dass es nicht für die Frauen in allen Ländern so ist, aber das ist keine Folge **des** Feminismus. (fkb031_2008_07)

ctok: Ich finde, dass **eine** finanzielle Entlohnung eines Menschen dem Beitrag entsprechen soll, den er / sie für das Unternehmen leistet.

ZH2: Ich finde, dass **die** finanzielle Entlohnung eines Menschen dem Beitrag entsprechen sollte, den er / sie für ein Unternehmen leistet. (fkb010_2008_07)

Tempus & Modus

Tempus und Modus werden dem Kontext angepasst.

ctok: Wie die Berliner haben von Barrack Obama gerade gestern gehört, unsere Sozietät **glaube**, und muss glauben bleiben, der Idee der Ebenheit und der Gleichkeit aller Personen, Mann oder Frau.

ZH2: Wie die Berliner von Barrack Obama gerade gestern gehört **haben**, glaubt unsere Gesellschaft und muss weiter glauben an die Idee der Gleichheit aller Personen, ob Mann oder Frau.

(fkb031_2008_07)

ctok: Wegen des Obenstehendes ist es also nicht in Ordnung zu äußern, dass die Universitätsabschlüsse von geringem Wert **sind**.

ZH2: Deshalb ist es also nicht in Ordnung zu sagen, dass die Universitätsabschlüsse von geringem Wert **seien**. (cbs007_2006_09.)

Ausdruck von Modus und Aspekt

Für das deutsche ungebräuchliche Formen des Ausdrucks von Modus und Aspekt werden durch gebräuchliche ersetzt.

ctok: Wie die Berliner haben von Barrack Obama gerade gestern gehört, unsere Sozietät glaube, und muss glauben **bleiben**, der Idee der Ebenheit und der Gleichkeit aller Personen, Mann oder Frau.

ZH2: Wie die Berliner von Barrack Obama gerade gestern gehört haben, glaubt unsere Gesellschaft und muss **weiter** glauben an die Idee der Gleichheit aller Personen, ob Mann oder Frau.

(fkb031_2008_07)

Modalverben

Eine dem deutschen untypische Verwendung von Modalverben, wird angepasst.

ctok: Obwohl die Lehren so wichtig ist, **können** die Lehrer nicht gut verdienen.

ZH2: Obwohl die Lehrer so wichtig sind, **verdienen** sie nicht gut. (trk008_2006_05)

Konnektoren

Drücken die verwendeten Konnektoren eine im Kontext unstimmmige Relation aus, werden sie ersetzt.

ctok: Weil für meisten Menschen das Geld als die beste Entlohnung gilt, wollen die andere andere Entlohnungen, schätzen die andere Werte, glauben an die unsiegbare Macht des Gelds nicht.

ZH2: Während für die meisten Menschen das Geld als die beste Entlohnung gilt, wollen andere eine andere Entlohnung, schätzen andere Werte und glauben nicht an die unbesiegbare Macht des Gelds. (fu127_2006_10c)

ctok: müssen uns auch fragen, ob die These des oberliegenden Titels heutzutage wirklich erreichbar sei, **weil** in diesem politischen Klima, wer hat die Macht, solche Entscheidungen über Löhne und entsprechende gesellschaftliche Beiträge eigentlich zu treffen?

ZH2: Wir müssen uns auch fragen, ob die These des obigen Titels heutzutage wirklich erreichbar ist, **denn** wer hat in diesem politischen Klima eigentlich die Macht, solche Entscheidungen über Löhne und entsprechende gesellschaftliche Beiträge zu treffen? (hu006_2006_09)

Fehlende Objekte

Semantisch fehlende Objekte werden ergänzt.

ctok: Nach seine Universitätsabschluss kennt er wahrscheinlich jeder Wirkung von eine besondere Maschine aber er hat noch nie früher die Gelegenheit physisch mit dieser Maschine um zu gehen und sich genauer zu beschauen.

ZH2: Nach seinem Universitätsabschluss kennt er wahrscheinlich jede Wirkung **einer** besonderen Maschine, aber er hatte früher noch nie die Gelegenheit, physisch mit dieser Maschine umzugehen und sie sich genauer zu beschauen. (sa001_2006_09)

Numerus

Ein unplausibler Numerus wird korrigiert.

ctok: . Unsere Krankenschwester und andere Mitarbeiter des öffentlichen Unternehmens haben ihre Meinungen hierzu geäußert.

ZH2: . Unsere Krankenschwestern und andere Mitarbeiter in öffentlichen Unternehmen haben ihre Meinung hierzu geäußert. (fkb010_2008_07.)

Konfligierende Referenz

Widersprechen sich satzinterne Referenzen, so werden sie korrigiert.

ctok: Sie laufen mit Messern und Pistolen rum, weil man Angst haben muss, dass man im nächsten Moment von einem anderen Gangmitglied der einer anderen Gang angehört getötet werden kann.

ZH2: Sie laufen mit Messern und Pistolen herum, weil man Angst haben muss, dass man im nächsten Moment von einem Mitglied einer **anderen** Gang getötet werden kann. (dew10_2007_09)

Morphologie

Nichtstandardsprachliche Reduktion von Partikelverben

Eine nichtstandardsprachliche Reduktion von Partikeln in Partikelverben werden korrigiert.

ctok: Und ich bin der Meinung, dass es dann schwerer ist da wieder rauszukommen.

ZH2: Und ich bin der Meinung, dass es dann schwerer ist, da wieder herauszukommen.
(dew18_2007_09)

Lexik

Feste Wendungen

Gibt es für eine unübliche Formulierung eine feste Wendung, dann wird diese bevorzugt.

ctok: In der dänischen Gesellschaft **werden** diese Ausbildungen sehr nachgefragt, weil es zur Zeit ein Mangel an diesen Fachkräften besteht.

ZH2: In der dänischen Gesellschaft **sind** diese Ausbildungen sehr gefragt, weil zur Zeit ein Mangel an diesen Fachkräften besteht. (fkb049_2008_08)

ctok: Europa war im 20. Jh zu viel **in den Kriegen tätig**.

ZH2: Europa war im 20. Jh. zu viel **mit Kriegen beschäftigt**. (fkb051_2008_08)

ctok: **Wegen des Obenstehendes** ist es also nicht in Ordnung zu äußern, dass die Universitätsabschlüsse von geringem Wert sind.

ZH2: **Deshalb** ist es also nicht in Ordnung zu sagen, dass die Universitätsabschlüsse von geringem Wert seien. (cbs007_2006_09.)

Konstruktionen: Modalpartikel

Fehlt in bestimmte Konstruktionen eine sonst übliche Modalpartikel, wird diese ergänzt.

ctok: Sei die im Titel benannte These provozierend, und auf ersten Blick völlig unbegründet scheinen, auf jeden Fall ist sie eine Diskussion wert.

ZH2: Sei die im Titel benannte These auch provozierend und mag sie auf den ersten Blick völlig unbegründet scheinen, auf jeden Fall ist sie eine Diskussion wert.
(fkb058_2008_08)

ctok: Das heißt, dass sie versuchen ihren Studenten so viel wie möglich die universelle Werte der Menschheit beizubringen. Das kann nur dann passieren, wenn sie sich mehr mit dem Geist des Menschen beschäftigen.

ZH2: Das heißt, dass sie versucht ihren Studenten so viele universelle menschliche Werte beizubringen wie möglich. Das kann aber nur dann passieren, wenn sie sich mehr mit dem menschlichen Geist beschäftigen. (fk006_2006_08.)

Übliche Bezeichnungen

Steht im Lernertext anstatt einer üblichen Bezeichnung für einen bestimmten Referenten ein anderer Begriff, bei dem nicht erkennbar ist, dass der Lerner bewusst auf die üblichere Variante verzichtet hat, wird letztere durch erstere ersetzt.

ctok: Die Errungenschaften der deutschen Philosophen, der deutschen Historiker des 19. Jhs bewertet man sich sehr hoch in der **Weltwissenschaft**.

ZH2: Die Errungenschaften der deutschen Philosophen und der deutschen Historiker des 19. Jhs. bewertet man in der **globalen Wissenschaftsgemeinde** sehr hoch. (fkb051_2008_08)

ctok: Die wenige Schicksallose, denen gelang es ihre **Fatum** zu verändern und finanziell zu wachsen erreichten das wieder durch Macht, Schönheit, **Schlaueheit**, aber nur in einzelne Außnahmefälle durch Arbeit oder durch seinen Kopf, die könnten manchmal die Annerkennung des Menschen in der Gesellschaft zu gewinnen, aber wenig Geld zu bringen.

ZH2: Die wenigen Schicksalslosen, denen es gelänge, ihr **Schicksal** zu verändern und finanziell zu wachsen, erreichten das wieder entweder durch Macht, Schönheit und **Intelligenz**, aber nur in einzelnen Ausnahmefällen durch Arbeit, die manchmal die Anerkennung der Menschen in der Gesellschaft gewinnen könnte, aber wenig Geld brächte oder durch ihren Kopf. (fu127_2006_10c)

ctok: Es gibt gewaltsam Kriminalität wie Mord, Kriminalität gegen **humanität**

ZH2: Es gibt gewaltsame Kriminalität wie Mord, Kriminalität gegen die **Menschlichkeit** (kne20_2006_07)

Pragmatik

Referenzen anpassen

Referenzen werden dem satzübergreifenden Kontext angepasst.

ctok: Sie hatten keine Lust mehr sich nur um eigene Kinder zu kümmern, Essen zu kochen, Wäsche zu waschen, kurz zusammengefasst: zu Hause wie in einem Käfig zu sitzen.

ZH2: Sie hatten keine Lust mehr, sich nur um die eigenen Kinder zu kümmern, Essen zu kochen, Wäsche zu waschen, kurz gesagt, zu Hause wie in einem Käfig zu sitzen. (fkb030_2008_07)

ctok: Die wenige Schicksallose, denen gelang es ihre Fatum zu verändern und finanziell zu wachsen erreichten das wieder durch Macht, Schönheit, Schlaueheit, aber nur in einzelne Außnahmefälle durch Arbeit oder durch **seinen Kopf**, die könnten

manchmal die Anerkennung des Menschen in der Gesellschaft zu gewinnen, aber wenig Geld zu bringen.

ZH2: Die wenigen Schicksalslosen, denen es gelänge, ihr Schicksal zu verändern und finanziell zu wachsen, erreichten das wieder entweder durch Macht, Schönheit und Intelligenz, aber nur in einzelnen Ausnahmefällen durch Arbeit, die manchmal die Anerkennung der Menschen in der Gesellschaft gewinnen könnte, aber wenig Geld brächte oder durch **ihren** Kopf. (fu127_2006_10c)

Informationsstruktur: Informationsstatus

Eine Nominalphrase, die auf einen im gleichen Satz bereits erwähnte Referenten verweist, wird pronominalisiert.

ctok: Die Unis sollten auch besser den Studenten erklären, wofür sie etwas lernen. Das, was **man** lernt sollte **der Student** zu einer bestimmten Anwendungssituation verbinden können.

ZH2: Die Unis sollten den Studenten auch besser erklären, wozu sie etwas lernen. Das, was **sie** lernen, sollten **sie** mit einer bestimmten Anwendungssituation verbinden können. (cbs006_2007_10)

ctok: Die meisten Frauen möchten, damit sich die Männer mit ihnen manchmal wie Gentlemen umgehen, **ihnen** Hilfe leisten.

ZH2: Die meisten Frauen möchten den Männern Hilfe leisten, damit **sie mit ihnen** manchmal wie Gentlemen umgehen. (fk019_2006_07)

ctok: Der Mann spielte eine zentrale Rolle in der Gesellschaft. **Der Mann** war nicht der Frau untertan

ZH2: Der Mann spielte eine zentrale Rolle in der Gesellschaft. **Er** war der Frau nicht untertan (fk024_2006_07)

Informationsstruktur: Fokus

Lerneräußerungen mit einer Fokusgliederung, die dem Kontext nicht entspricht wird korrigiert.

ctok: **Die besonders schwere** Lage war in Deutschland.

ZH2: Besonders schwer war die Lage in Deutschland.
(fkb051_2008_08)

ctok: Es gibt aber Häuser, wo die Mutter zu Hause nur faulenz.

Es ist so beispielweise in der Familie meines Freundes

ZH2: Es gibt aber Haushalte, in denen die Mutter zu Hause nur faulenz.

So ist es beispielweise in der Familie meines Freundes (fk015_2006_07)

Informationsstruktur: Fokuspartikel

Die Stellung der Fokuspartikel wird an den intendierten Skopus im Satz angepasst.

ctok: Wir müssen uns auch fragen, ob die These des oberliegenden Titels heutzutage wirklich erreichbar sei, weil in diesem politischen Klima, wer hat die Macht, solche Entscheidungen über Löhne und entsprechende gesellschaftliche Beiträge **eigentlich** zu treffen?

ZH2: Wir müssen uns auch fragen, ob die These des obigen Titels heutzutage wirklich erreichbar ist, denn wer hat in diesem politischen Klima **eigentlich** die Macht, solche Entscheidungen über Löhne und entsprechende gesellschaftliche Beiträge zu treffen? (hu006_2006_09)

ctok: Dies Erkenntnis ist auch **wesentlich** in finanziellen Berufen, und in Berufen, in den man muss Argumenten machen und schützen, zum Beispiel politische Berufen.

ZH2: Diese Erkenntnis ist auch in finanziellen Berufen **wesentlich** und in Berufen, in denen man Argumente bringen und unterstützen muss, zum Beispiel in der Politik. (fk013_2006_08.)

Informationsstruktur: Topik

Topikausdrücke stehen vor Fokusaussagen, wenn letztere nicht im Kontext als topikalisiert angesehen werden können.

ctok: Z. B. Leibniz hatte keine Möglichkeit an der Leipziger Universität zu promovieren, denn **die grosse Korruption** herrschte **dort**, die Professoren hielten die schlechten Vorlesungen, nicht immer kamen zu den Vorlesungen u. s. w.

ZH2: Z. B. hatte Leibniz keine Möglichkeit, an der Leipziger Universität zu promovieren, denn **dort** herrschte **große Korruption**, die Professoren hielten schlechte Vorlesungen, sie kamen nicht immer zu den Vorlesungen usw. (fkb051_2008_08)

Stil

Direkte Rede

Direkte Rede wird nur korrigiert, wenn sie fehlerhaft ist.

ctok: Wozu nutzt es, wenn einer beim Bewerbungsgespräch sagt: "Ja, ich habe die Theorien von Marx, Weber und Durkheim im Griff und ich bin ein netter und pünktlicher Mensch."?

ZH2: Wozu nutzt es, wenn einer beim Bewerbungsgespräch sagt: "Ja, ich habe die Theorien von Marx, Weber und Durkheim im Griff und ich bin ein netter und pünktlicher Mensch."? (cbs011_2006_09)

ctok: Die Frauen sagen: "**Wir** konnten früher so viele Dinge nicht machen und jetzt wollen **wir** gerne alles nachholen!"

ZH2: Die Frauen sagen, dass **sie** früher so viele Dinge nicht machen konnten und jetzt alles nachholen wollen. (fkb002_2007_09)

ctok: Was soll die Lösung sein? Sicherlich nicht die Satz "Universitätsabschlüsse sind von geringem Wert".

ZH2: Was soll die Lösung sein? Sicherlich nicht der Satz: "Universitätsabschlüsse sind von geringem Wert." (fk004_2006_08.)

Einleitungen im Zeitungsstil werden nicht korrigiert

Uneingeleitete Nebensätze werden nicht korrigiert, wenn sie nicht selbst mit einem eingebetteten Nebensatz beginnen, oder Verbletzstellung vorliegt.

ctok: Das Unterschied: das Praktikum muss völlig anhand des Studenten ausgehen und wird nur bis zum 50 Prozent als Teil des Studiums anerkannt.

ZH2: Der Unterschied: Das Praktikum muss völlig von dem Studenten ausgehen und wird nur bis zu 50 Prozent als Teil des Studiums anerkannt. (cbs005_2006_09)

Relativpronomen

Nichtstandardsprachliche Relativpronomen werden ersetzt

ctok: Die Schule, wo man streng kontrolliert wird, ist schon in der Vergangenheit geblieben.

ZH2: Die Schule, in der man streng kontrolliert wurde, ist schon Vergangenheit. (fkb044_2008_08)

Umgangssprachliche Lexik

Sehr umgangssprachliche Lexik wird durch durch im Kontext angemessenere ersetzt.

ctok: Wegen der methodischen / theoretischen Vorgangsweise **kriegt** eine wissenschaftliche Arbeit auch eine besondere Form.

ZH2: Wegen der methodischen bzw. theoretischen Vorgehensweise **bekommt** eine wissenschaftliche Arbeit auch eine besondere Form. (cbs013_2006_09)

5.4 Zielhypothese für die Annotation der komplexen Verben [ZHverb]

Diese Zielhypothese stimmt fast komplett mit der maximalen Zielhypothese überein und unterscheidet sich nur an einigen Stellen, an denen die Annotatorin der komplexen Verben eine Zielstruktur gewählt hat, die entweder der minimalen Zielhypothese oder einer dritten Variante entspricht.

Abweichungen der Zielhypothesen von der ctok-Ebene

Für jede Zielhypothese wird automatisch die Abweichung gegenüber der ctok-Ebene berechnet und in einer DiffZH1, DiffZH2, DiffZHverb etc. eingetragen.

Diese Abweichungen erlauben eine gezielte Suche nach fehlerhaften Strukturen in den Lernerdaten.

Fehlendes Token (INS)

Für ein Token, das in der Zielhypothese gefüllt ist, in der ctok-Ebene dagegen nicht, wird die-

ses Token mit „INS“ gekennzeichnet.

ctok	Kriminalität	zahlt	sich	nicht		.
ZH1	Kriminalität	zahlt	sich	nicht	aus	.
ZH1Diff					INS	

(cbs002_2007_10)

Überflüssiges Token (DEL)

Für ein Token, das in der ctok-Ebene gefüllt ist, in der Zielhypothese dagegen nicht, wird dieses Token mit „DEL“ gekennzeichnet.

ctok	Sich	ständig	in	der	kriminellen	Welt	aufzuhalten	lohnt	sich	auch	nicht	aus	,
ZH1	Sich	ständig	in	der	kriminellen	Welt	aufzuhalten	lohnt	sich	auch	nicht		,
ZH1Diff												DEL	

(cbs002_2007_10)

Verändertes Token (CHA)

Ein Token, das in der Zielhypothese einen anderen Wert hat als in der ctok-Ebene, das aber beide male nicht leer ist, wird mit „CHA“ markiert.

ctok	die	Unterricht	an	der	Uni	geht	auch	so	.
ZH1	der	Unterricht	an	der	Uni	geht	auch	so	.
ZH1Diff	CHA								

(usb007_2006_10)

Geteiltes Token (SPLIT)

Ein Token der ctok-Ebene, das zwei Tokens in der Zielhypothese entspricht, wird mit „SPLIT“ markiert.

ctok	,	die	er	im	Universität	verbracht	hat	
ZH1	,	die	er	in	der	Universität	verbracht	hat
ZH1Diff				SPLIT				

(usb008_2006_10)

Token mit anderer Position (MOVS / MOVT)

Taucht in einem Satz das gleiche Token an unterschiedlichen Positionen in der ctok-Ebene und der Zielhypothese auf, so werden sie koindiziert⁶ und als Start (MOVS) und Ziel (MOVT) einer „Bewegung“ markiert.

ctok	Und	meiner	Meinung	nach	,	die	Frau	muss			selbst	entscheiden
ZH1	Und	meiner	Meinung	nach				muss	die	Frau	selbst	entscheiden
ZH1Diff					DEL	MOVS	MOVS		MOVT	MOVT		

(usb006_2006_10)

⁶ Dieser Index kann in der aktuellen Version noch nicht dargestellt werden, soll aber in Zukunft die Grundlage für die Realisation als Parallelkorpus dienen.

ctok	sich	wohl	draußen	von	der	akademischen	Umgebung		zu	fühlen	,
ZH1	sich		außerhalb	von	der	akademischen	Umgebung	wohlfühlen			,
ZH1Diff		MOVS	CHA					MOVT			

(fkb057_2008_08)

Token zusammenführen (MERGE)

Sollen in einer Zielhypothese mehrere Token des Ausgangstextes verbunden werden, so wird in der ZHDiff-Ebene mit dem Tag „MERGE“ versehen.

ctok	Man	hat	so	zu	sagen	die	Welt	vor	sich	.
ZH1	Man	hat	sozusagen		die	Welt	vor	sich	.	
ZH1Diff			MERGE							

6. Komplexe Verben

Annotatorin: Anke Lüdeling

Gegenstand:

Annotiert werden Präfixverben und Partikelverben. Andere komplexe Verben (z.B. neoklassische) werden nicht betrachtet. Wir wollen möglichst viele Phänomene bzgl. komplexer Verben erfassen, daher sind Zweifelsfälle mit erfasst.

Was zählt dazu?

alle finiten und infiniten zusammengeschiedenen Präfix- oder Partikelverben, auch solche, die auch auseinandergeschrieben werden könnten, zB auch:

deren Mitglieder sich irgendwo **festgekettet** haben (cbs001_2007_10)⁷

Da es keine klare Definition von Partikelverben und Präfixverben gibt, haben wir hier vieles eingeschlossen, also auch komplexe Verben mit Infinitiv (*kennenlernen*) oder Nomen (*klavierspielen*) als Erstglied. Bei Präfixverben haben wir auch ‚halbverdunkelte‘ Formen wie *empfinden* aufgeführt. Semantische Transparenz war kein Kriterium.

syntaktisch getrennte Partikelverben.

morphologisch getrennte Partikelverben

Partizipien in Zustandspassivkonstruktionen:

ctok: ..., dass es nicht selten für den Erfolg der Kommunikation **entscheidend** ist, ob ...

ZHverb: ..., dass es nicht selten für den Erfolg der Kommunikation **entscheidend** ist, ob ... (cbs002_2006_09)

⁷ Zu den Fehlern bei komplexen Verben siehe weiter unten. Alle anderen Fehler in den Beispielen sind auf anderen Ebenen zu behandeln und werden hier nicht betrachtet. Die betrachtete Form ist jeweils fett markiert.

aber nicht Partizipien in pränominaler Adjektivposition

Wir haben auch solche Adj/Adv/Part+V-Kombinationen aufgenommen, die getrennt geschrieben wurden, aber zusammen geschrieben werden können (*schiefgehen, ernstnehmen* etc.). Wir haben hier die zusammengeschriebene Form in die Zielhypothese genommen und einen Orthographiefehler annotiert (orth, siehe unten). .

ctok: Kriminalitäten wo etwas schief geht,

ZHverb: Kriminalität, bei der etwas schiefgeht (cbs002_2007_10)

6.1. Annotationsebenen für die komplexen Verben

Die komplexen Verben wurden immer am Verb der Lerneräußerung annotiert. Bei korrekten Verben (keine Änderung in der ZH-Verb) wurden die Ebenen **verbkategorie**, **verblemma** und **verbform** annotiert. Bei Fehlern zusätzlich die Ebene **verbfehlertyp**.

6.1.1 verbkategorie

<i>Annotationswert</i>	<i>Beschreibung</i>
vpart	bei Partikelverben, immer am Verb
ppart	bei getrennt stehenden Partikeln
vpräf	bei Präfixverben
ppräf	bei getrennt stehenden Präfixen (das sind dann Fehler) <i>während die Arbeitenden den Steuer be zahlen müssen</i> (fkb008_2008_07) (ZH: bezahlen)
vpartx	an Stellen, bei denen ein Partikelverb hätte stehen müssen (ZHVerb), aber keines stand <i>Das Leben auf die Universität führt zu dem zunehmenden Selbstbewusstsein und bereitet man in diese Weise auf die wirkliche Welt.</i> (fk008_2006_08) ZH-Verb: <i>Das Leben an der Universität führt zu einem zunehmenden Selbstbewusstsein und bereitet einen in dieser Weise auf die wirkliche Welt vor.</i>
ppartx	an Stellen, an denen eine Partikel hätte stehen müssen (ZHVerb), aber keine stand
vpräfx	an Stellen, an denen ein Präfixverb hätte stehen müssen (ZHVerb), aber keines stand <i>Diese Frage, nämlich, ob die finanzielle Entlohnung eines Menschend dem Beitrag entsprechen sollte, den er für die Gesellschaft geleistet hat, ist nicht einfach zu antworten</i> (fk013_2006_07) (ZHVerb: beantworten)

--	--

Es gibt (einige wenige) Lerneräußerungen, in denen klar ein Verb fehlt, das dann in ZH-Verb eingesetzt wurde. Solche Verben, auch wenn sie komplex sind, werden gar nicht annotiert, da hier der Lerner keinen Wortbildungsfehler gemacht hat.

6.1.2 verblemma

Das Verblemma wird beim finiten Verb notiert. Üblicherweise steht hier das Verb aus der Lerneräußerung, auch wenn es ein ‚unmögliches‘ Verb ist.

In den x-Fällen wird hier das Lemma aus der Zielhypothese eingesetzt.

6.1.3 verbfehlertyp

Die Fehlertypen beziehen sich auf die ZHverb

<i>Annotationswert</i>	<i>Beschreibung</i>
sem	Verblemma existiert, ist aber falsch verwendet (hier muss in der ZHverb ein anderes Verb angegeben sein) <i>Infolgedessen kann man nicht vernichten, dass der in der XXten Jahrhundert entwickelte Feminismus [...] eine entscheidende Rolle gehabt hat.</i> (fkb009_2008_07) (ZH: <i>verneinen</i>)
orth	orthographischer Fehler <i>Vielleicht können die dann anschliesend auch versuchen zu erklären was [...]</i> (hu007_2006_10) (ZH: <i>erklären</i>) Dazu zählen auch Getrennt- und Zusammenschreibungsfehler (dann in der ZH anders). <i>Der Arbeiter, der nicht weiterausgebildet wird, wird bald entkündigt.</i> (fkb008_2008_07) (ZH: <i>weiter ausgebildet</i>) Es gibt ‚Partikelverben‘ wie z.B. <i>ernstnehmen</i> , die getrennt oder zusammen geschrieben werden können. Hier wird eine einheitliche (zusammengeschriebene) Form in der ZH angenommen. Daher wird hier bei Getrenntschreibung ein Fehler angemerkt, der nach Orthographieprinzipien keiner ist.
lex	Verblemma ist nicht gebräuchlich (hier muss in der ZH ein anderes Verb angegeben sein) <i>Der Arbeiter, der nicht weiterausgebildet wird, wird bald entkündigt.</i> (fkb008_2008_07) (ZH: <i>gekündigt</i>)
part	Partikel zu viel <i>Sich ständig in der kriminellen Welt aufzuhalten lohnt sich auch nicht aus.</i> (cbs002_2007_10)

	(ZH-Verb: ... <i>lohnt sich auch nicht</i> .)	
as	Argumentstrukturfehler. Hier wird Argumentstruktur sehr weit gefasst – interessant ist, ob die Lerner die korrekte Verwendung des betreffenden Verbs kennen. <i>In Dies verhält sich auch mit der Praxisorientierung und Praxiserfahrung</i> (cbs003_2006_09) fehlt z.B. ein <i>so</i> . Hierunter werden auch falsche Präpositionen (bei subkategorisierten Ps) gefasst.	
flex	Flexionsform falsch ⁸ <i>[...] haben sie trotzdem viel gelernt und erfahrt</i> (fkb010_2007_09) (ZH: <i>erfahren</i>) Hier sind auch falsche Flexionsformen an Auxiliaren markiert (<i>sein</i> statt <i>haben</i> , <i>sein</i> statt <i>werden</i> etc.) <i>[...] oder dass noch eine Frau Vergewaltigt ist</i> (sa008_2006_09) (ZH <i>vergewaltigt wurde</i>)	
ws	Wortstellung (betrachtet wird hier nur die Verbstellung) falsch (vor allem V2-Fehler). Wortstellung innerhalb des Mittelfelds (die mit dem Verb nichts zu tun hat) wird nicht betrachtet.	
Besondere Flexionsprobleme bei komplexen Verben	ge	<i>ge</i> steht falsch <i>Aber auf jeden Fall wird es nicht rechtgefertigt</i> (usb015_2006_10) oder <i>ge</i> fehlt <i>Die Rollen können gelentglich austauscht werden</i> (fkb040_2008_08)
	zu	<i>zu</i> steht falsch <i>Als ich an dem Hochschulsommerkurs zu teilnehmen begann</i> (usb012_2006_10)
Besonderes Wortstellungsprobleme bei komplexen Verben	sep	Trennung falsch <i>und nun anbieten die meisten davon eine praxisorientierte Ausbildung auf dem Niveau BA</i> (fkb007_2007_09)

6.1.4 verbform

Hier wird die Flexionsform angegeben. Bei Fehlern wird hier die korrigierte Form aus der ZHVerb angegeben.

Annotationswert	Beschreibung
fin	finit
finsep	finit und syntaktisch getrennt
inf	Infinitiv ohne <i>zu</i>

⁸ Manchmal ist es schwierig zwischen flex und orth zu unterscheiden, wie z. B. in *und man **wachst** einfach auf* (fkb010_2007_09, von *aufwachsen*). Wir haben in solchen Fällen flex markiert.

infzu	Infinitiv mit <i>zu</i>
p2	Partizip II
p1	Partizip I
nn	In einigen (ganz wenigen) Fällen steht ein Verb, wo eigentlich ein Nomen stehen müsste (ZH). In solchen Fällen wurde als Fehler flex angegeben und bei Verbform nn.

7. Literatur:

(2006): Handbuch der Annotation der Stellungsfelder bei Falko.

ANNIS2: <http://www.sfb632.uni-potsdam.de/d1/annis/>, 26.5.2010.

Breindl, Eva et al. (2000): GRAMMIS - ein Projekt stellt sich vor. In: Sprachreport. Informationen und Meinungen zur deutschen Sprache 2000 (1), S. 19–24.

Corder, Stephen Pit (1986): The role of interpretation in the study. In: Corder, Stephen P. (Hrsg.): *Error analysis and interlanguage*. 4. impr. Oxford: Oxford University Press, S. 35–44.

Das elektronische Valenzwörterbuch deutscher Verben (E-VALBU): <http://hypermedia2.ids-mannheim.de/evalbu/index.html>, 26.5.2010.

Doolittle, Seanna (2009): Entwicklung und Evaluierung eines auf dem Stellungsfeldermodell basierenden syntaktischen Annotationsverfahrens für Lernerkorpora innerhalb einer Mehrebenen-Architektur mit Schwerpunkt auf schriftlichen Texten fortgeschrittener Deutschlerner. Magisterarbeit HU-Berlin.

Eisenberg, Peter (2006): Der Satz. 3., durchges. Aufl. Stuttgart, Weimar: Metzler (= Grundriss der deutschen Grammatik; Bd. 2).

Engel, Ulrich (1996): Deutsche Grammatik. 3., korr. Aufl. Heidelberg: Groos.

Freywald, Ulrike (2008): Zur Syntax und Funktion von dass-Sätzen mit Verbzweitstellung. In: Deutsche Sprache 36), S. 246–285.

Freywald, Ulrike (2009): Kontexte für nicht-kanonische Verbzweitstellung. V2 nach dass und Verwandtes. In: Ehrich, Veronika et al. (Hrsg.): *Koordination und Subordination im Deutschen*. Hamburg: Buske (= Linguistische Berichte; Sonderheft 16), S. 113–134.

- Granger, Sylviane (1993):** The International Corpus of Learner English. In: Aarts, Jan/Haan, P. de/Oostdijk, Nelleke (Hrsg.): *English Language Corpora. Design, Analysis and Exploitation*. Amsterdam: Rodopi, S. 57–69.
- Granger, Sylviane et al. (2009):** The International Corpus of Learner English. Version 2. Handbook and CD-ROM. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Lennon, Paul (1991):** Error. Some Problems of Definition, Identification, and Distinction. In: *Applied Linguistics* 12 (2), S. 180–196.
- Lüdeling, Anke; Walter, Maik; Kroymann, Emil; Adolphs, Peter (2005):** Multi-level error annotation in learner corpora. In: *Proceedings of Corpus Linguistics 2005*. Birmingham.
- Lüdeling, Anke (2008):** Mehrdeutigkeiten und Kategorisierung. Probleme bei der Annotation von Lernerkorpora. In: Walter, Maik/Grommes, Patrick (Hrsg.): *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung*. Deutsche Gesellschaft für Sprachwissenschaft. Tübingen: Niemeyer (= Linguistische Arbeiten; 520), S. 119–140.
- Schmid, Helmut (1994):** Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of the International Conference on New Methods in Language Processing.
- Schmidt, Thomas (2001):** The transcription system EXMARaLDA: An application of the annotation graph formalism as the Basis of a Database of Multilingual Spoken Discourse. In: *Proceedings of the IRCS Workshop On Linguistic Databases, 11-13 December 2001*. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania, S. 219–227. URL: http://www.exmaralda.org/files/IRCS_Paper.pdf.
- Schneider, Roman (2008):** E-VALBU. Advanced SQL/XML processing of dictionary data using an object-relational XML database. In: *Sprache und Datenverarbeitung, International Journal for Language Data Processing* 32 (1), S. 33–44.
- Zeldes, Amir et al. (2009):** ANNIS. A Search Tool for Multi-Layer Annotated Corpora. In: *Proceedings of Corpus Linguistics 2009, Liverpool, July 20-23, 2009*.
- Zipser, Florian (2009):** Entwicklung eines Konverterframeworks für linguistisch annotierte Daten auf Basis eines gemeinsamen (Meta-)modells. Diplomarbeit. Institut für Informatik. Berlin.

8. Kontakt

Marc Reznicek

Wissenschaftlicher Mitarbeiter

Institut für deutsche Sprache und Linguistik

Korpuslinguistik und Morphologie

Dorotheenstraße 24

Raum 3.310

Tel: +49 (30) 2093-9720

Marc.Reznicek@staff.hu-berlin.de