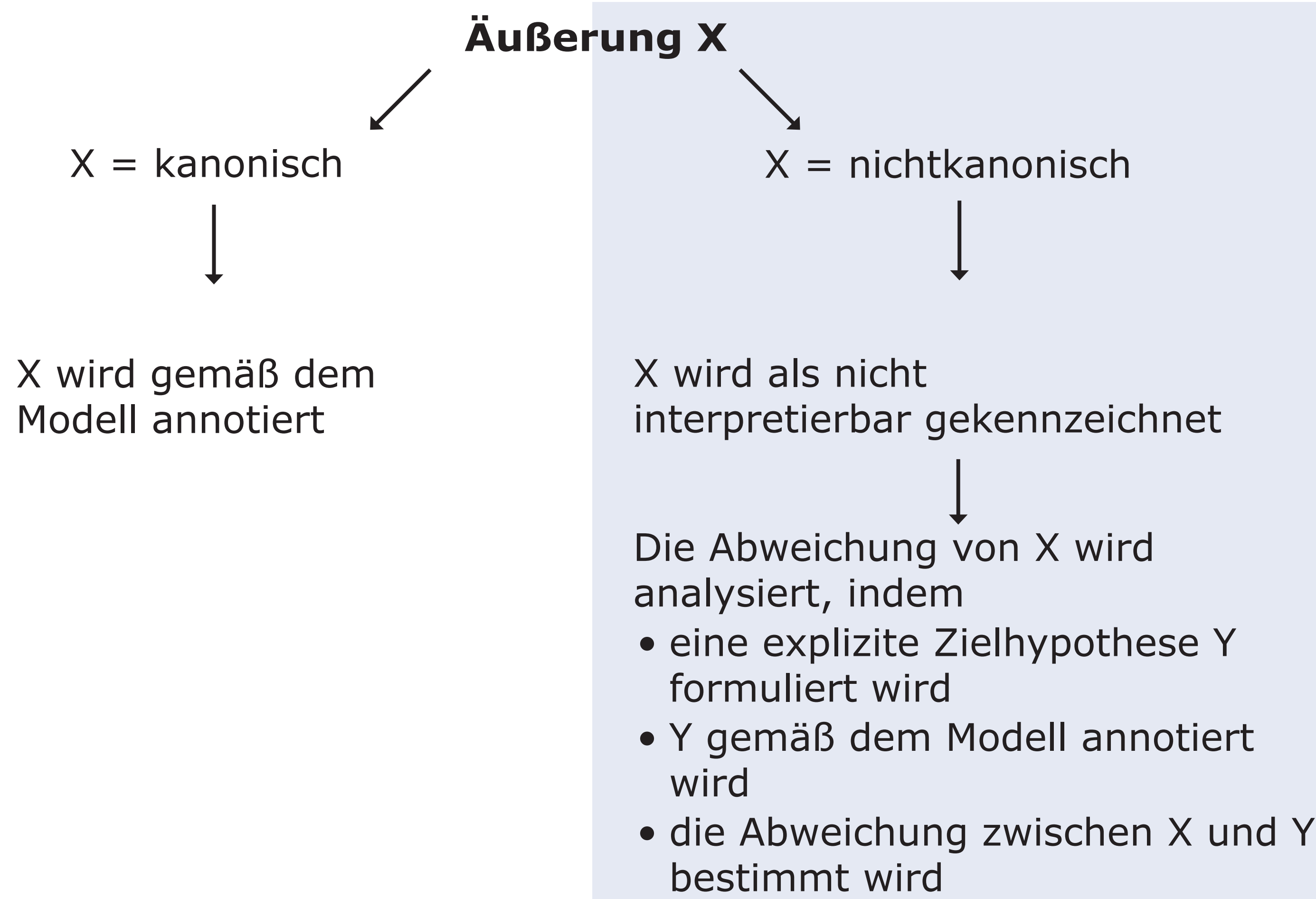


Annotation von kanonischen und nichtkanonischen Äußerungen in Korpora



Dieses Poster stellt eine Methode zur integrierten Annotation von kanonischen und nichtkanonischen Äußerungen in demselben Korpus vor. Existierende (syntaktische) Annotationsschemata bauen meist auf einer Theorie auf, die für eine standardisierte (schriftsprachliche) Varietät entwickelt wurde. In vielen Korpora gibt es jedoch Äußerungen, die durch die gegebenen Schemata nicht befriedigend annotiert werden können. Für viele linguistische Fragestellungen ist es allerdings interessant, sowohl kanonische als auch nichtkanonische Äußerungen in einem Korpus zu finden (bzw. systematisch unterscheiden zu können) und gerade unterschiedliche Arten der Abweichung von einem gegebenen Schema suchen zu können. (Hirschmann, Doolittle & Lüdeling 2007)

Annotationsschema



„Kanonisch“ bezeichnet alle Äußerungen in einem Korpus, die gemäß einem dem Annotationsschema zugrunde liegenden Modell annotierbar sind.

„Nichtkanonisch“ bezeichnet alle Äußerungen in einem Korpus, die gemäß einem dem Annotationsschema zugrunde liegenden Modell **nicht** annotierbar sind.

Kanonizität bzw. Nichtkanonizität entscheidet sich demnach an der Frage, ob ein gegebenes Schema eine bestimmte Äußerung beschreiben kann oder nicht. Nichtkanonizität muss interpretiert werden als Fehler in der Äußerung (z.B. in Lernerkorpora) oder als mangelnde Abdeckung des Modells (z.B. bei gesprochener Sprache).

Zielhypothese

Für jede Interpretation/Annotation nichtkanonischer Äußerungen ist die Formulierung einer Zielhypothese unabdingbar. Sie ist eine kanonische, annotierbare Entsprechung der nichtkanonischen Äußerung.

Ursprünglich findet die Auseinandersetzung mit Zielhypothesen im Kontext der Fehlerbeschreibung statt: „Eine notwendige Voraussetzung für jede Fehlerannotation ist die Annahme einer im Kontext ‚korrekten‘ Form, die wir in Anlehnung an Ellis (1994) Zielhypothese nennen.“ (Lüdeling et al. 2008) In unserem Modell wird für jede nichtkanonische Äußerung eine Zielhypothese angenommen. Die Angabe einer Zielhypothese erlaubt eine genaue Beschreibung der Abweichung.

Für viele nichtkanonische Äußerungen sind mehrere Zielhypothesen denkbar, die dann zu unterschiedlichen Analysen führen (Lüdeling 2008). Durch die Angabe einer expliziten Zielhypothese ist die Datengrundlage nachvollziehbar.

Fallbeispiel aus dem Lernerkorpus Falko (<http://korpling.german.hu-berlin.de/falko/index.jsp>)

Lerneräußerung: Er tatsächlich war sehr wohlhabend gewesen. Seine Vorfahren waren die Kaiser im Deutschland.
 Wortart (POS): PPER.ADJ VAFIN.ADV.ADJ VAPP.\$ POSPART.NN VAFIN.ART.NN APPRART.NE \$
 Lemma: er tatsächlich sein sehr wohlhabend gewesen sein Vorfahr|Vorfahren sein d Kaiser im Deutschland.

Seine Vorfahren waren die Kaiser im Deutschland Er tatsächlich war sehr wohlhabend gewesen

Annotation gemäß dem Stellungsfeldermodell

[Äußerung]	Seine	Vorfahren	waren	die	Kaiser	im	Deutschland
[Felder]	VF		LSK				MF

Bei kanonischen Äußerungen entspricht die Zielhypothese der Äußerung

[Äußerung]	Er	tatsächlich	war	sehr	wohlhabend	gewesen
[Felder]	f_ = mit Feldermodell nicht annotierbar					

Formulierung einer expliziten Zielhypothese (Zh)

[Äußerung]	Er		tatsächlich	war	sehr	wohlhabend	gewesen
[Felder]	f_ = mit Feldermodell nicht annotierbar						
[Zh]	Er	war	tatsächlich		sehr	wohlhabend	gewesen

Annotation der Zielhypothese gemäß dem Feldermodell

[Äußerung]	Er		tatsächlich	war	sehr	wohlhabend	gewesen
[Felder]	f_ = mit Feldermodell nicht annotierbar						
[Zh]	Er	war	tatsächlich		sehr	wohlhabend	gewesen
[Felder]	VF	LSK			MF		RSK

Beschreibung und Interpretation der Abweichung zwischen Äußerung und Zielhypothese

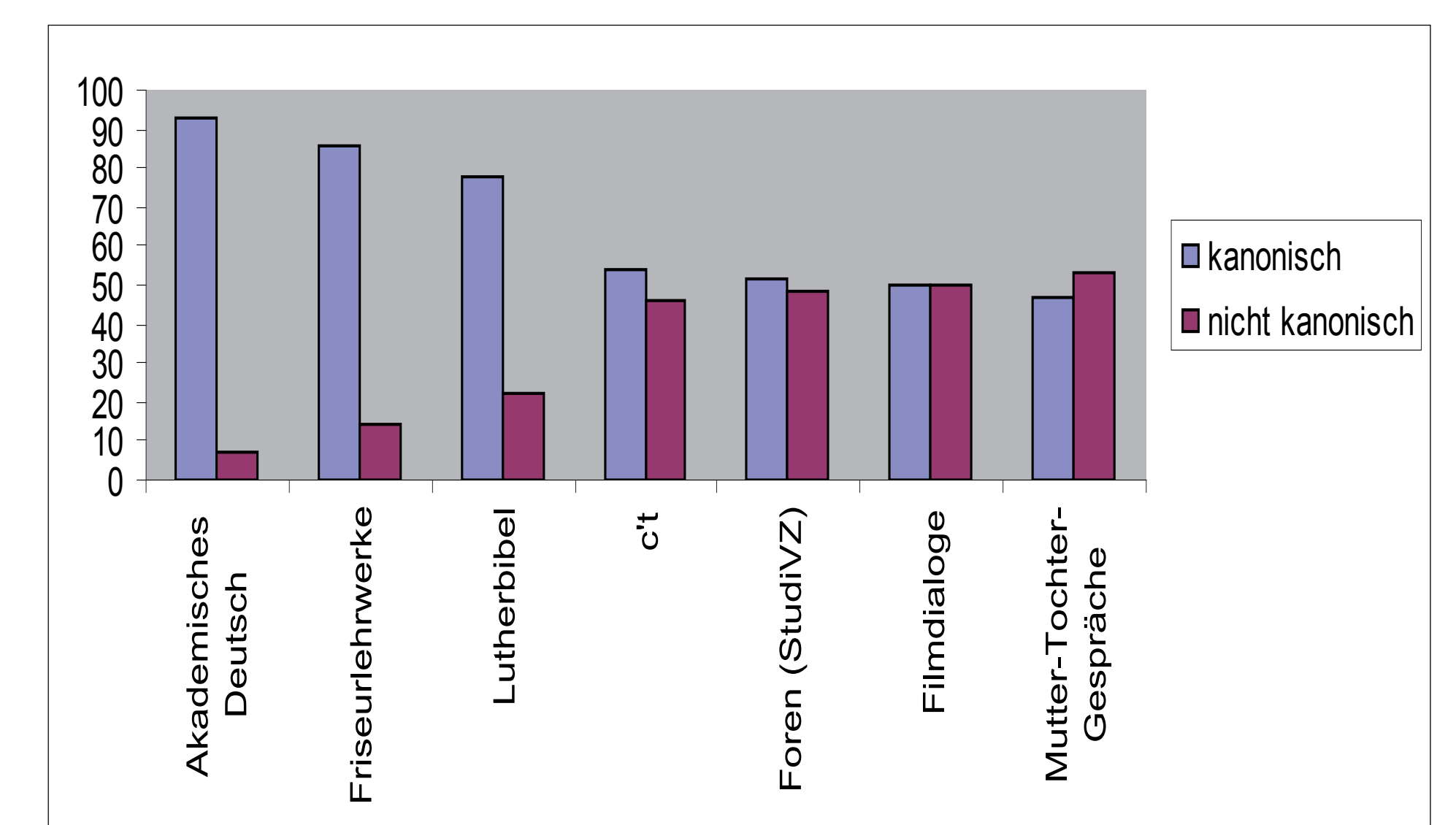
[Äußerung]	Er		tatsächlich	war	sehr	wohlhabend	gewesen
[Felder]	f_ = mit Feldermodell nicht annotierbar						
[Zh]	Er	war	tatsächlich		sehr	wohlhabend	gewesen
[Felder]	VF	LSK			MF		RSK
Beschreibung Abweichung	token gelöscht			token eingefügt			
Interpretation Abweichung	Verbstellungsfehler						

Stellungsfeldermodell

Die Methode ist unabhängig von einem bestimmten syntaktischen Modell. Für unsere Experimente verwenden wir hier das topologische Stellungsfeldermodell nach Drach (1937) und Höhle (1986).

Verwendete Tags:
 VF - Vorfeld
 LSK - linke Satzklammer
 MF - Mittelfeld
 RSK - rechte Satzklammer
 NF - Nachfeld

Relevanz des Ansatzes



Experiment aus einem Hauptseminar, je ca. 500 ‚Sätze‘ aus 7 Varietäten. Kanonizität bezogen auf das Stellungsfeldermodell

Anwendung auf weitere Korpusbeispiele

Gesprochene Sprache

[Äußerung]	aber	unsre			in	ruh	lassen	okay
[Zielhypothese]	aber	unsre	sollst	du	in	ruh	lassen	okay
Beschreibung Abweichung			token eingefügt	token eingefügt				
[Felderannotation der Zielhypothese]	(kon-junction)	VF	LSK	MF			RSK	NF

Annotation einer (topologisch) nicht kanonischen Äußerung aus einem Korpus des gesprochenen Deutschen.

Sonderforschungsbereich 245: „Sprechen und Sprachverstehen im sozialen Kontext“ in Heidelberg und Mannheim zwischen 1988 und 1992 (<http://www.ids-mannheim.de/kskd/agd/korpora/ekkorpus.html>)

Computer-mediated Communication CMC

[Äußerung]	Ok	erst			Level	10	
[Zielhypothese]	Ok	erst	muss	ich	Level	10	machen
Beschreibung Abweichung			token eingefügt	token eingefügt			
[Felderannotation der Zielhypothese]	(diskurs-marker)	VF	LSK	MF			RSK

Annotation einer (topologisch) nicht kanonischen Äußerung aus einem CMC Korpus

Beispiel von <http://www.worldofgothic.de>

Literatur

Drach, E. (1937) Grundgedanken der deutschen Satzlehre. Frankfurt a. M.: Diesterweg.
 Ellis, R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
 Hirschmann, H., S. Doolittle & A. Lüdeling (2007) Syntactic annotation of non-canonical linguistic structures. In: *Proceedings of Corpus Linguistics 2007*, Birmingham. [Online unter <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/mitarbeiter-innen/anke/pdf/HirschmannDoolittleLuedelingCL2007.pdf>]
 Höhle, T. (1986) Der Begriff „Mittelfeld“. Anmerkungen über Theorie der topologischen Felder, in M. Reis et al. (ed.) *Actes du VII. Kongresses der Internationalen Vereinigung für germanistische Sprach- und Literaturwissenschaft*, pp. 329-340. Tübingen: Niemeyer.
 Lüdeling, Anke; S. Doolittle, H. Hirschmann, K. Schmidt & M. Walter (erscheint 2008) *Das Lernerkorpus Falko*. In: *Deutsch als Fremdsprache 2/2008*
 Lüdeling, Anke (2008) Mehrdeutigkeiten und Kategorisierung. Probleme bei der Annotation von Lernerkorpora. In: M. Walter & P. Grommes (Hrsg.) *Fortgeschrittene Lernervarietäten*. Tübingen: Niemeyer, 119-140.
 Lüdeling, A., M. Walter, E. Kroymann and P. Adolphs (2005): Multi-level error annotation in learner corpora, in: *Proceedings of Corpus Linguistics 2005*. Birmingham. [Online unter <http://www.corpus.bham.ac.uk/PCLC/>]
 Schwitalla, J. (2006) *Gesprochenes Deutsch. Eine Einführung*. Berlin: Erich Schmitt Verlag.