

Deutsch ist nicht gleich Deutsch: Einblicke in das L1-Lernerkorpus KoKo

Andrea Abel

Institut für Fachkommunikation und Mehrsprachigkeit

Europäische Akademie Bozen (EURAC)

Workshop „Plurizentrik im Deutschunterricht“, Wien, 10.10.2014

Übersicht

1. Projekt „KoKo“
2. L1-Lernerkorpus „KoKo“
3. Beispielanalysen: Rektionsfehler
4. Ausblick

→ Einblicke in das Projekt und die Erarbeitung des Lernerkorpus im Hinblick auf die Ausarbeitung eines länderübergreifenden Projekts zum Thema: „Schulisches Schreiben im deutschen Sprachraum. Plurizentrische und pluriareale Aspekte“ (Arbeitstitel)



1. Projekt „KoKo“

Projekt: „KoKo: Bildungssprache im Vergleich: korpusunterstützte Analyse der Sprachkompetenz bei Lernenden im deutschen Sprachraum“

Leitung:
(Andrea Abel)



Partner:



Im Rahmen der Initiative „Korpus Südtirol“ (Leitung: **EURAC** research)

<http://www.korpus-suedtirol.it>





1. Projekt „KoKo“

Projektziele:

1. Aussagen über **Sprachkompetenzen** von SchülerInnen

- Schreibkompetenz: Bildungssprache, Deutsch als Erst-/Unterrichtssprache
- Jugendliche mit höherem Bildungsniveau (Ende Sekundarstufe II)

2. Analyse von **Kontextvariablen**

- Diatopische, sprachbiographische und sozio-ökonomische Faktoren

3. Aufbau eines digitalen **Lernerkorpus** und Entwicklung von Instrumenten

- wissenschaftlich fundierte Basis für Erfassung und Analyse schriftlicher Sprachproduktion und nachhaltige Dokumentation
- Entwicklung computerlinguistischer Werkzeuge zur vergleichenden Analyse (z.B. Anstein 2013)



1. Projekt „KoKo“

Design:

- **TeilnehmerInnen:** ca. 1500 OberschülerInnen 1 Jahr vor der Matura bzw. dem Abitur:
 - Südtirol ca. 520
 - Nordtirol ca. 460
 - Thüringen ca. 520
- **Stratifizierte Zufallsstichprobe** (nach Schulklassen) für jede Region mit folgenden Schichtungsmerkmalen:
 - Stadtgröße (Schule): Groß-, Mittel- u. Kleinstadt
 - Schultyp: allgemeinbildend vs. berufs-/fachspezifisch
- **Zeitpunkt der Datenerhebung:** Stichtag 25.05.2011



1. Projekt „KoKo“

Methode:

- **Erörterung** zu 1 Thema (Zitat von H.M. Enzensberger) in der Schule
- **Schüler- und Lehrerfragebogen (Metadatenerhebung)**
 - Kontrollvariablen (Geschlecht, Alter, Schultyp ...)
 - Sprachliche Biographie der TeilnehmerInnen
- **Datenaufbereitung und Korpuserstellung** (manuelle Transkription, automatische Annotation)
- qualitative, **manuelle Analyse** (linguistische Annotation nach Analyseraster)
- quantitative, **halb-automatische Analyse** (korpuslinguistische Tools)
- Verknüpfung von **Textanalyse und Metadaten**



1. Projekt „KoKo“

Forschungsleitende Fragen für die Analyse der Texte:

1. Welche Merkmale weisen Schülertexte (im bildungssprachlichen Kontext) auf?
2. Welche Arten von Merkmalen können beschrieben werden?
3. Welche Faktoren (Schultyp, Region, Geschlecht, sozioökonomischer Hintergrund, Sprachbiographie etc.) bestimmen die Verteilung der Merkmale?

(vgl. Abel/ Glaznieks 2014)



1. Projekt „KoKo“

Wie oft kommt es vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und dem Lauten Orgeln, der in der Ecke ~~es~~ ^{kontinuierlich} spielenden Altherren, lauscht die sich über die Jugend von heute beschwerten.

Ich finde die Jugend von heute ist nicht schlechter als die Jugend von ~~vor~~ ^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld ~~verändert~~ ^{verändert} wir sind globaler, Informierter aber wir haben auch Zeit für Sonstigen Musik mein ~~großes~~ ^{großes} noch von ~~6~~ ⁶ Uhr morgens bis 9 Uhr Abends auf dem Bauernhof halbe so haben wir mehr Freizeit, mehr Freiheit

Beispiel – Ausschnitt L1-
Lernertext



1. Projekt „KoKo“

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt ~~und~~ dem lauten Organen, der in der Ecke ~~es~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch zeit für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr Abends auf dem Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und ~~z~~ dem lauten Organen, der in der Ecke ~~s~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch zeit für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr Abends auf den Bauernhof helfen so haben wir mehr Freizeit, meh Freiheit

(...)



1. Projekt „KoKo“

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und dem lauten Organen, der in der Ecke ~~er~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von ~~f~~ ^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch zeit für Sonstiges. Musste mein ~~er~~ Großvater noch von ~~er~~ 6 Uhr morgens bis 9 Uhr Abends auf dem Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und ~~er~~ dem lauten Organen, der in der Ecke ~~er~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen. **<ABSATZ>**

<LEERZEILE>

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von ~~f~~ ^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch zeit für Sonstiges. Musste mein ~~er~~ Großvater noch von ~~er~~ 6 Uhr morgens bis 9 Uhr Abends auf den Bauernhof helfen so haben wir mehr Freizeit, meh Freiheit

(...)



1. Projekt „KoKo“

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und dem lauten Organen, der in der Ecke ~~es~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch zeit für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr Abends auf den Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)
Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und ~~z~~ dem lauten Organen, der in der Ecke ~~es~~ **Kartenspielenden** Altherren, lauscht, die sich über die Jugend von heute **beschwehen**.

Ich finde, die Jugend von **Heute** ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren, im **gegenteil**, nur hat sich das Umfeld verändert, wir sind **Globaler, Informierter**, aber wir haben auch **zeit** für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr **Abends** auf den Bauernhof helfen, so haben wir mehr Freizeit, **meh** Freiheit
(...)



1. Projekt „KoKo“

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und ~~dem~~ lauten Organen, der in der Ecke ~~es~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von heute ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind globaler, Informierter aber wir haben auch Zeit für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr Abends auf dem Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und ~~dem~~ lauten Organen, der in der Ecke ~~es~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch Zeit für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr Abends auf ~~den~~ Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)



1. Projekt „KoKo“

Wie oft kommt es ~~f~~ vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und dem lauten Organen, der in der Ecke ~~es~~ Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von heute ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind globaler, Informierter aber wir haben auch Zeit für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr Abends auf dem Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)

Wie oft kommt es ~~f~~ vor, dass man in einer **Bar** sitzt, seinen Apfelsaft trinkt und ~~f~~ dem lauten Organen, der in der Ecke ~~es~~ Kartenspielenden **Altherren**, lauscht die sich über die Jugend von heute beschwehen.

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von ~~f~~^{Vor} 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch Zeit für Sonstiges. Musste mein ~~o~~ Großvater noch von ~~z~~^z 6 Uhr morgens bis 9 Uhr Abends auf dem Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)

2. L1-Lernerkorpus „KoKo“

KoKo -

Korpuszusammensetzung:

Subkorpus (nach Region)	Gesamtanzahl	L1 Deutsch
Nordtirol (NT):	233.098 Tokens (457 Texte)	206.439 Tokens (404 Texte)
Südtirol (ST):	222.209 Tokens (520 Texte)	192.891 Tokens (451 Texte)
Thüringen (TH):	353.674 Tokens (521 Texte)	317.075 Tokens (464 Texte)
ohne Angabe	2.349 Tokens (5 Texte)	---
gesamt	811.330 Tokens (1503 Texte)	716.405 Tokens (1319 Texte)

Derzeit verfügbare Metadaten:

- Informationen zu soziolinguistisch relevanten Variablen (Region, L1, Schultyp, Geschlecht, Schulnote Deutsch - Fragebogenerhebung)
- Autor-ID, Transkribierer, Projekt, Sprache

(Version „KoKo_2“, Stand: Dezember 2012)

2. L1-Lernerkorpus „KoKo“

Tools und Formate zur Datenaufbereitung und –abfrage:

- **Transkription:** XMLmind (<http://www.xmlmind.com/>)
- **automatische Annotation:**
Tokenisierung, Sentence Splitting, POS-Tagging, Lemmatisierung: IMS Tree Tagger (Schmid 1994)
- **manuelle Annotation:** XMLmind und MMAX2 (Müller/Strube 2006)
- **Konvertierungsframework:** SaltNPepper (Zipser/Romary et al. 2010)
- **Versionskontrolle:** Subversion (<http://subversion.tigris.org/>)
- **Such- und Visualisierungswerkzeug:** ANNIS (Zeldes et al. 2009)

2. L1-Lernerkorpus „KoKo“

Transkription und manuelle Annotation:

- **Digitalisierung** der handgeschriebenen Texte
 - Annotation der **Texteigenschaften** (Ziel → „Fotografie“, u.a. Selbstkorrekturen, Tilgungen, Emoticons ...)
 - Annotation **nicht standardkonformer Schreibungen** & Ergänzung von **Zielformen** als Vorbereitung der korpuslinguistischen Aufbereitung

(Überblick: Abel et al., in Druck)

2. L1-Lernerkorpus „KoKo“

Transkription und manuelle Annotation (XMLMind)

The screenshot shows the XMLMind interface with the following content:

Transcriber: Vera
Author ID: 1244
Exam type: Koko

Exam language: Deutsch
Exercise 1

„Die Jugend ist sowieso keine beneidenswerte Phase“, dieses Zitat ~~ist~~^{wurde} in der Zeitung [!] die Die[!] Zeit vom 4. Mai 2001 von Essayist Hans Magnus Enzensberger geschrieben. Er beschreibt darin [!] das dass[!] die Jugend ein nicht zu [!] erstrebenswertes ~~erstrebenswertes~~[!] Alter, ist, da in dieser Zeitspanne des Lebens man weder Souveränität hat, noch sich der Kultur anpasst. Auch beschreibt er [!] Jugendliche[!] als: „labil, unsicher, schwankend und als solche Menschen die jede Dummheit mitmachen.“

Als junger Mensch kann man dieses^{diesem} Zitat nicht teilen zustimmen da die Jugend auch die Phase des Lebens sein soll um sich in einer [!] Gesellschaft Gesellschaft[!] zu [!] integrieren integrieren[!]. Somit ist es zwar richtig [!] das dass[!] man [??] als Jugendlicher unsicher ist aber diese Unsicherheit [!] benötigt benötigt[!] man um sich den richtigen Weg für Leben [!] herauszu zu suchen herauszusuchen[!]. Auch ist es [!] war wahr[!] [!] das dass[!] man als Teenager jede Dummheit [!] mit macht mitmacht[!], ob aber wenn man diese Erfahrung nie machen würde. könnte man es nie als Dummheit ansehen. Ein

The XML structure on the right shows the following elements:

- `<transcriber>Vera </transcriber>`
- `<author_id>1244 </author_id>`
- `<exam_type>Koko </exam_type>`
- `<text_language>Deutsch </text_lang </text_language>`
- `<body>`
 - `<exercise number="1" xml:space="preserve">`
 - „Die Jugend ist sowieso keine beneidenswerte Phase“, dieses Zitat
 - `<correction>`
 - `<deletion>ist </deletion>`
 - `<insertion>wurde </insertion>`
 - `</correction>`
 - in der Zeitung
 - `<error eType="28 Eigennarr">`
 - `<originalForm>di </originalForm>`
 - `<targetForm>Die </targetForm>`

2. L1-Lernerkorpus „KoKo“

Manuelle Annotation:

- **Annotation von sprachlichen Merkmalen (stand-off)** – Kombination:
 - Linguistische Annotation (Orthographie, Grammatik, Lexik, (Textebene) sowie Subkategorien)
 - „Target modification classification“ (Tilgung, Einfügungen ...) (vgl. Díaz-Negrillo/Domínguez 2006)
- Fehler → **Zielhypothesenannotation** (vgl. Lüdeling et al. 2005) → Zielform(en)

(Überblick: Abel et al., in Druck)

2. L1-Lernerkorpus „KoKo“

Manuelle Annotation (MMAx2)

MMAX2 1.13.003 C:\Users\AAbel\Documents\Koko\Mmax2\Silvia\ID1063.mmax

File Settings Display Tools Plugins Info Show ML Panel

Transcriber: Vera Author_id: 1063 Exam Type: Koko Language: Deutsch

Erörterung

Darüber ob die Jugend eine beneidenswerte Phase des Lebens ist, lässt sich streiten. Der deutsche Schriftsteller und Essayist Hans Enzensberger sagte dazu in einem Interview der Wochenzeitung beneidenswerte Phase, denn der junge Mensch sei labil{**\$**} unsicher, schwankte jede Dummheit mit. Er zählt verschiedene Beispiele wie die Klamottensucht und sein, wenn man diese Zeit überstanden habe, behauptet der Schriftsteller. Mir stellt sich die Frage, ob ich, wo ich gerade mitten in {diese} Phase stecke diese Zeit # [!]**\$** zurück denke[!] und froh bin, dass ich die hinter mir habe Ich glaube{**\$**} ich werde kaum an diese Phase [!]**\$** zurück denken[!] und der Meir Phasen meines Lebens gewesen sei. Eher werde ich an meine Jugendzeit [!]**\$** zu viel Wunderbares und Aufregendes ich dort erlebt habe. Schon aus Sprichwörtern wie „sich nochmal jung fühlen“{**\$**} oder „Man ist nur e viele Menschen die Zeit des [!]**\$** jung seins[!] gerne zweimal mitmachen würden. {Als Jugendlicher hat man {weniger} Verantwortung {wie} als ein Erwachsene, und Strom zu bezahlen und um Kinder zu ernähren und [!]**\$** groß zu ziehen[!]**\$**} Faktoren sind für viele Menschen eine Last. Als Jugendlicher hat man {dieser} oder arbeitet, wobei das Geld nicht für Miete, Strom, Kinder usw. [!]**\$** drauf geht[uneingeschränkter, denn sie {können} das Geld {zum} Feiern, Shoppen oder a {ausgeben}. {Wir} {Jugendlichen} haben den Luxus{**\$**} im „Hotel Mama“ zu lebe und geputzt. So manch einer regt sich dann schon mal auf, wenn er # einmal c Erwachsener muss man diese kleinen Hausarbeiten jeden Tag machen. [!]**\$** Tag Jugendliche sind können sich ihre Wochen und Tage viel abwechslungsreicher andere Party gehen, neue Leute kennen lernen und hier und da mal einen Blö

[d i e s e] [dirty]

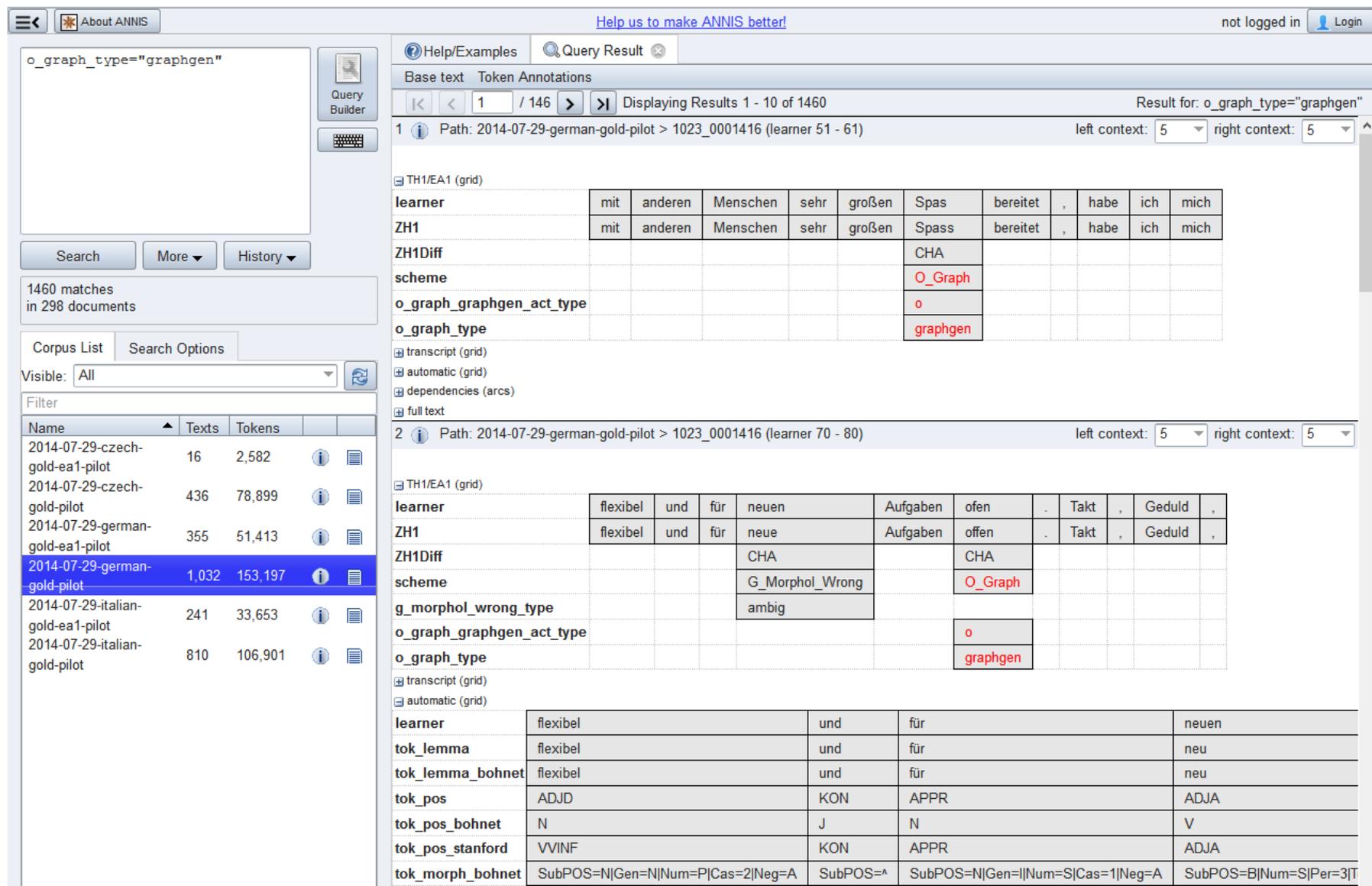
One-click annotation Panel Settings

nbol	theme	title	unreadable	anakoluth	cohesionMarker	correspondence
<>	unit_type	determiner				
<>	determiner_type	demonstrative				
<>	error_type	case				
	target	dative				
	target_word	dieser				
	acceptability_in_written_language	no				
	acceptability_in_spoken_language	no				
	inherited_error	no				
	occurance_in_quotation	no				

Suppress check Warn on extra attributes

2. L1-Lernerkorpus „KoKo“

Suche und
Visualisierung
über ANNIS
(Beispiel: Projekt
MERLIN – L2-
Lernerkorpus)



The screenshot displays the ANNIS (Annotation Network Based Interface) interface. The search query is `o_graph_type="graphgen"`. The results are filtered to show 1460 matches across 298 documents. The interface is divided into several sections:

- Search Results Table:** A table listing corpus documents with columns for Name, Texts, and Tokens. The document `2014-07-29-german-gold-pilot` is highlighted, showing 1,032 texts and 153,197 tokens.
- Query Result View:** Shows the base text and token annotations for two specific results.

learner	mit	anderen	Menschen	sehr	großen	Spas	bereitet	,	habe	ich	mich
ZH1	mit	anderen	Menschen	sehr	großen	Spass	bereitet	,	habe	ich	mich
ZH1Diff											
scheme											
o_graph_graphgen_act_type											
o_graph_type											

learner	flexibel	und	für	neuen	Aufgaben	ofen	.	Takt	,	Geduld	.
ZH1	flexibel	und	für	neue	Aufgaben	offen	.	Takt	,	Geduld	.
ZH1Diff											
scheme											
g_morphol_wrong_type											
o_graph_graphgen_act_type											
o_graph_type											
- Annotations:** Below the text, various annotations are shown, including `o_graph_type` with values like `graphgen` and `o`.
- Navigation:** The interface includes navigation buttons for previous/next results and context windows (left/right context: 5).

3. Beispielanalysen: Rektionsfehler

Rektion:

„Lexemspezifische Eigenschaft von Verben, Adjektiven, Präpositionen und Substantiven, die morphologische Kategorie (insb. den Kasus) abhängiger Elemente zu bestimmen“ (Bußmann 1990: 639)

Fehler:

Grammatische Abweichungen von der geschriebenen Standardsprache nach DUDEN Bd.4 bzw. grammis (i.e. Systemfehler & Normfehler (Eisenberg 2007) + Angabe der Akzeptanz für geschriebene und gesprochene Varietäten)

3. Beispielanalysen: Rektionsfehler

Fokus Beispielanalysen:

a) Präpositionalphrase: Fehlerhafte Kasusmarkierung an der Nominalphrase

b) Verbreaktion: Fehlerhafte Kasusmarkierung an den Komplementen

(Fokus: Dativ und Akkusativ)

Fehler werden pro Fehlerträger annotiert!

3. Beispielanalysen: Rektionsfehler

Grundlage Beispielanalysen – KoKo-Teilkorpus:

Subkorpus (nach Region)	Gesamtanzahl	L1 Deutsch
Nordtirol (NT):	233.098 Tokens (457 Texte)	206.439 Tokens (404 Texte)
Südtirol (ST):	222.209 Tokens (520 Texte)	192.891 Tokens (451 Texte)
Thüringen (TH):	353.674 Tokens (521 Texte)	317.075 Tokens (464 Texte)
ohne Angabe	2.349 Tokens (5 Texte)	---
gesamt	811.330 Tokens (1503 Texte)	716.405 Tokens (1319 Texte)



Subkorpus (nach Region)	L1 Deutsch
Nordtirol (NT):	116.077 Tokens (199 Texte)
Südtirol (ST):	96.612 Tokens (199 Texte)
Thüringen (TH):	160.100 Tokens (199 Texte)
ohne Angabe	---
gesamt	372.789 Tokens (597 Texte)

3. Beispielanalysen: Rektionsfehler

a) Präpositionalphrase:

- Zielphrase im Dativ:

Mir stellt sich die Frage, ob ich, wo ich gerade mitten *{in}* *{diese}* Phase stecke, das später auch so sehen werde. (ST ID1063)

Weiters finde ich, dass die Zeit der Jugend eindeutig Vorteile hat *{gegenüber}* *{die}* eines Erwachsenen. (NT ID2708)

Die Jugend von heute hat *[!]fiel[!]* zu leisten und muss *{mit}* gewaltigem Druck, sowohl *{privaten}* wie auch *{[!]gesellschaftlichen[!]}*, *[!]klar kommen[!]* (TH ID2821)

3. Beispielanalysen: Rektionsfehler

a) Präpositionalphrase:

- Zielphrase im Akkusativ:

Man kann Phasen des Lebens nicht {in} {einem} Topf stecken und verallgemeinern. (ST ID2259)

Die meisten von {ihnen} haben nur Dummheiten im Kopf und hören oft auch nicht {auf} {der} Mutter oder {dem} Vater. (ST ID2030)

3. Beispielanalysen: Rektionsfehler

b) Verbreaktion:

- **Dativ:**

Sei es{\$} um {andere} zu {*imponieren*} oder sich selbst etwas zu beweisen. (NT ID1184)



Es {*gefällt*} {einen}, was sich die größeren Geschwister anziehen, was sie sagen, welche Musik sie hören, einfach alles! (ST ID1215)



Ich {*stimme*} {den} Autor auch {*zu*}, dass ein junger Mensch labil und unsicher ist, denn er hat keine Erfahrung vom Leben und muss erst ~~alles~~^alles^ ausprobieren. (TH ID1391)



3. Beispielanalysen: Rektionsfehler

b) Verbreaktion:

- **Akkusativ:**

Vieles{ $\$$ } [!]dass[!] man erlebt, {wird} {einem} das Leben lang {begleiten} und auch die Erfahrungen{ $\$$ } die man macht{ $\$$ } will man nicht missen. (NT ID1632)

Die meisten Erfahrungen werden in der Jugendzeit gemacht, die {einem} Menschen nach und nach reifer {werden lassen}. (ST ID1880)

Hat er wohlmöglich schlechte Erfahrungen gemacht die {ihm} bis heute [!]immernoch[!] {prägten}? (TH ID2223)

3. Beispielanalysen: Rektionsfehler

Beispielanalysen & Ergebnisse: diatopische Verteilung

Korrespondenz		absolut			prozentuell			log likelihood		
		NT	ST	TH	NT	ST	TH	NT vs. ST	NT vs. TH	ST vs. TH
Rektion	PP	96	119	136	0.083	0.123	0.085	8,50**	0.04	8,64**
	Dativ	66	74	100	0.057	0.077	0.062	3.10	0.35	1.75
	Akkusativ	22	28	22	0.019	0.029	0.014	2.24	1.13	6,91**
	VP	54	45	69	0.047	0.047	0.043	0.00	0.18	0.16
	Dativ	24	20	36	0.021	0.021	0.022	0.00	0.10	0.09
	Akkusativ	22	22	18	0.019	0.023	0.011	0.37	2.72	4,95*

log likelihood - * signifikanter Unterschied

95th percentile; 5% level; $p < 0.05$; critical value = 3.84

99th percentile; 1% level; $p < 0.01$; critical value = 6.63

99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83

99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13

(Festlegung Schwellen nach der Cochran rule, vgl. Rayson et al. 2004)

3. Beispielanalysen: Rektionsfehler

Beispielanalysen & Ergebnisse: Verteilung nach Schultyp

Korrespondenz		absolut		prozentuell		log likelihood
		AHS	BHS	AHS	BHS	AHS vs. BHS
Rektion	PP	180	171	0.083	0.110	7,30**
	Dativ	117	123	0.054	0.079	9.12**
	Akkusativ	35	37	0.016	0.024	2.81
VP		76	92	0.035	0.059	11,82***
	Dativ	37	43	0.017	0.028	4,80*
	Akkusativ	24	38	0.011	0.025	9,75**

4. Ausblick

Für einen länderübergreifenden Projektantrag:

- erstmals Verbindung Schreibkompetenz- und L1-Lernerkorporaforschung in diesem Kontext
- Aufbau auf bestehender Korpusinfrastruktur und etabliertem Workflow zur Lernerkorpusaufbereitung
- Anpassung bestehender Annotationsschemata
- Umfassende, korpusbasierte Auswertung von Lernerdaten (kontrastive Lenersprachenanalyse: Über-/Mindergebrauch sprachlicher Einheiten, Fehleranalyse/Analyse ausgewählter sprachlicher Merkmale) & Verknüpfung mit Metadaten
- Interoperabilität und langfristige Speicherung für Nutzung in unterschiedlichen Kontexten (Synergien ICLTT, ACDH, ISCM)

Vielen Dank für Ihre Aufmerksamkeit!

Andrea Abel
(andrea.abel@eurac.edu)



Literaturauswahl

- Abel, A. & A. Glaznieks (2014): Wo Sprachkompetenzforschung auf Varietätenlinguistik trifft: Empirische Befunde aus dem Varietäten-Lernerkorpus „KoKo“. In: A. Lenz & M. Glauninger [Hgg.]: *Variation und Varietäten des Deutschen in Österreich – Theoretische und empirische Perspektiven*. Frankfurt am Main: Peter Lang.
- Ágel, V. & M. Hennig (2010): Einleitung. In: V. Ágel & M. Hennig [Hgg.]: *Nähe und Distanz im Kontextvariationslinguistischer Forschung*. Berlin / New York: de Gruyter, 1-22.
- Augst, G., Disselhoff, K., Henrich, A., Pohl, T., and Völzing, P. (2007). *Text-Sorten-Kompetenz*. Peter Lang.
- Becker-Mrotzek, Michael / Böttcher, Ingrid 2006. *Schreibkompetenz entwickeln und beurteilen*. Berlin: Cornelsen Scriptor Verlag.
- Ammon, U. (1995): *Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten*. Berlin / New York: de Gruyter.
- Berruto, G. (1993a): Le varietà del repertorio. In: A. A. Sobrero [Hg.]: *Introduzione all'italiano contemporaneo. Le variazioni e gli usi*. Roma-Bari: Gius. Laterza & Figli Spa, 3-36.
- Berruto, G. (1993b): Varietà diamesiche, diastratiche, diafasiche. In: A. A. Sobrero [Hg.]: *Introduzione all'italiano contemporaneo. Le variazioni e gli usi*. Roma-Bari: Gius. Laterza & Figli Spa, 37-92.
- Berruto, G. (2004): Sprachvarietäten – Sprache (Gesamtsprache, historische Sprache. In: U. Ammon et al. [Hgg.]: *Sociolinguistics. An International Handbook of the Science of Language and Society. Handbücher zur Sprach- und Kommunikationswissenschaft 3.1*. Berlin / New York: de Gruyter, 2., vollständig neu bearbeitete und erweiterte Auflage, 188-195.
- Christ, O. (1994): A Modular and Flexible Architecture for an Integrated Corpus Query System. In: *Proceedings of COMPLEX 1994*. Ungarn, 23-32.
- Coseriu, E. (1988): *Sprachkompetenz. Grundzüge der Theorie des Sprechens*. Tübingen: Francke.
- Davies, Alan (2003). *The Native Speaker: Myth and reality*. Clevedon: Multilingual Matters.
- Dipper, St., Lüdeling, A. & M. Reznicek (2013): NoSta-D: A Corpus of German Non-Standard Varieties. In: M. Zampieri & S. Diwersy [Hgg.]: *Non-Standard Data Sources in Corpus-Based Research*. Shaker Verlag, 69-76.

Literaturauswahl

Dittmar, N. (1997): *Grundlagen der Soziolinguistik: ein Arbeitsbuch mit Aufgaben*. Tübingen: Narr.

Dürscheid, C. (2003): Medienkommunikation im Kontinuum von Mündlichkeit und Schriftlichkeit. Theoretische und empirische Befunde. In: *Zeitschrift für Angewandte Linguistik* 38, 37-56.

Ehlich, Konrad (2007a): Deixis und Anapher. In: Konrad Ehlich: *Sprache und sprachliches Handeln. Bd. 2: Prozeduren des sprachlichen Handelns*. Berlin, New York: de Gruyter, 5-24

Ehlich, Konrad (2007b): Anadeixis und Anapher. In: Konrad Ehlich: *Sprache und sprachliches Handeln. Bd. 2: Prozeduren des sprachlichen Handelns*. Berlin, New York: de Gruyter, 25-44

Granger, S. (in Druck) Contrastive Interlanguage Analysis: A Reappraisal. In: *International Journal of Learner Corpus Research*, Volume 1, issue 1.

Hoffmann, Ludger (2000): Anapher im Text. In: G. Antos/K. Brinker et al. (Hg.): *Text- und Gesprächslinguistik*, Bd.1. HSK 16.1. Berlin/New York de Gruyter, 295-305.

Koch, P. & W. Oesterreicher (2007): Schriftlichkeit und kommunikative Distanz. In: *Zeitschrift für germanistische Linguistik* 35, 346-375.

Kretschmer, E. (2013): Über die Kategorisierung der Sprachvarietäten und die des Mediolektivs im Besonderen. In: *Revista de Estudos Alemães* 4, 152-175.

Löffler, H. (2010): *Germanistische Soziolinguistik* [1985]. Berlin: Schmidt, 4., neu bearbeitete Auflage.

Lüdeling, A. (2012): A corpus-linguistics perspective on language documentation, data, and the challenge of small corpora. In: F. Seifart et al. [Hgg.]: *Potentials of Language Documentation: Methods, Analyses, and Utilization. Language Documentation & Conservation. Special Publication No. 3* at the University of Hawai'i Press, 32-38.

Mioni, A. M. (1983): Italiano tendenziale: osservazioni su alcuni aspetti della standardizzazione. In: P. Benincà et al. [Hgg.]: *Scritti linguistici in onore di Giovan Battista Pellegrini*. Vol. 1, 495-517.

Mukherjee, J. (2005). The native speaker is alive and kicking – Linguistic and language-pedagogical perspectives. In: *Anglistik* 16/2, 7-23.

Literaturauswahl

Rayson, P., D. Berridge & B. Francis (2004): Extending the Cochran rule for the comparison of word frequencies between corpora. In: G. Purnelle, C. Fairon & A. Dister [Hgg.]: *Proceedings of the 7th International Conference on Statistical Analysis of Textual Data (JADT)*. Louvain-la-Neuve: Presses universitaires de Louvain, 926–936.

Rickheit, G. / Schade, U. (2000): Kohärenz und Kohäsion. In: K. Brinker / Antos, G, Heineman, W. & Sager, S.F. (Hrsgg.): *Text- und Gesprächslinguistik*. (1. Halbband). Berlin - New York: de Gruyter, 275-284.

Schmid, H. (1994): Probabilistic Part-of-Speech-Tagging Using Decision Trees. In: *Proceedings of the International Conference on New Methods in Language Processing*.

Storrer, A. (2013): Sprachstil und Sprachvariation in sozialen Netzwerken. In: B. Frank-Job, A. Mehler & T. Sutter [Hgg.]: *Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen an Beispielen des WWW*. Wiesbaden: VS Verlag für Sozialwissenschaften, 331-366.

Walther, M. & P. Grommes (2008): Die Entdeckung des fortgeschrittenen Lerners in der der Varietätenlinguistik. In: M. Walther & P. Grommes [Hgg.]: *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer, 3-28.

2. L1-Lernerkorpus „KoKo“

Manuelle Annotation (MMAx2)

MMAx2 1.13.003 C:\Users\AAbel\Documents\Koko\Mmax2\Silvia\ID1063.mmax

File Settings Display Tools Plugins Info Show ML Panel

Transcriber: Vera Author_id: 1063 Exam Type: Koko Language: Deutsch

Erörterung

Darüber ob die Jugend eine beneidenswerte Phase des Lebens ist, lässt sich streiten. Der deutsche Schriftsteller und Essayist Hans Enzensberger sagte dazu in einem Interview der Wochenzeitung *Frankfurter Allgemeine Zeitung*, dass die Jugend eine beneidenswerte Phase, denn der junge Mensch sei labil{**\$**} unsicher, schwankend und jede Dummheit mit. Er zählt verschiedene Beispiele wie die Klamottensucht und das Verstecken von Geld, um zu sein, wenn man diese Zeit überstanden habe, behauptet der Schriftsteller. Mir stellt sich die Frage, ob ich, wo ich gerade mitten **{in}** diese Phase stecke diese Zeit **{}** zurück denke[!] und froh bin, dass ich die hinter mir habe. Ich glaube{**\$**} ich werde kaum an diese Phase [!]zurück denken[!] und der Meinung sein, dass diese Phasen meines Lebens gewesen sei. Eher werde ich an meine Jugendzeit [!]zu viel Wunderbares und Aufregendes ich dort erlebt habe. Schon aus Sprichwörtern wie „sich nochmal jung fühlen“{**\$**} oder „Man ist nur ein Kind, wenn man viele Menschen die Zeit des [!]jung seins[!] gerne zweimal mitmachen würden. {Als Jugendlicher hat man {weniger} Verantwortung {wie} als ein Erwachsene, und Strom zu bezahlen und um Kinder zu ernähren und [!]groß zu ziehen[!]}{**\$**} Faktoren sind für viele Menschen eine Last. Als Jugendlicher hat man {dieser} oder arbeitet, wobei das Geld nicht für Miete, Strom, Kinder usw. [!]drauf geht[!] uneingeschränkter, denn sie {können} das Geld {zum} Feiern, Shoppen oder zu {ausgeben}. {Wir} {Jugendlichen} haben den Luxus{**\$**} im „Hotel Mama“ zu leben und geputzt. So manch einer regt sich dann schon mal auf, wenn er **{}** einmal ein Erwachsener muss man diese kleinen Hausarbeiten jeden Tag machen. [!]Tag für Tag. Jugendliche **sind** können sich ihre Wochen und Tage viel abwechslungsreicher und anders gestalten, andere Party gehen, neue Leute kennen lernen und hier und da mal einen Blö

[i n]

One-click annotation Panel Settings

readable anacoluth cohesionMarker correspondence **correspondenceReferent**

unit type preposition

marker correspondence:76

Suppress check Warn on extra attributes

Wortgrenzen

Wie oft kommt es vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und dem lauten Organen, der in der Ecke ~~es~~ Kartenspielenden Altherren, lauscht von heute beschwehen.

Satzgrenzen

Wortarten

Ich finde die Jugend von heute schlechter als die Jugend von vor 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind globaler, Informierter aber wir haben auch Zeit für Sonstiges. Musste mein Großvater noch von 6 Uhr morgens bis 9 Uhr Abends auf dem Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

Lemmata

(...)
Wie oft kommt es vor, dass man in einer Bar sitzt, seinen Apfelsaft trinkt und dem lauten Organen, der in der Ecke Kartenspielenden Altherren, lauscht die sich über die Jugend von heute beschwehen.

VFIN

finden

Ich finde die Jugend von Heute ist nicht schlechter als die Jugend von vor 60 Jahren im Gegenteil, nur hat sich das Umfeld verändert, wir sind Globaler, Informierter aber wir haben auch Zeit für Sonstiges. Musste mein Großvater noch von 6 Uhr morgens bis 9 Uhr Abends auf dem Bauernhof helfen so haben wir mehr Freizeit, mehr Freiheit

(...)



1. Projekt „KoKo“

Grundannahme:

H_0 : Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein Lernerphänomen.

Mögliche Alternativannahmen:

H_1 : Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein diatopisches Phänomen.

...

H_n : Das sprachlich auffällige Merkmal x im Lernerkorpus „KoKo“ ist ein Y-Phänomen.

(vgl. Abel/ Glaznieks 2014)