

Zur Nomenmodifikation in argumentativen Texten von L1- und L2-Sprecher*innen

Julia Lukassek

SFB 1412 „Register Knowledge“
Projekt C04 „Register knowledge in advanced learner language“

10.02.2021



- 1 Projektvorstellung
- 2 Studie zu Modifikation in der Nominaldomäne
- 3 Morphologische Annotation von Nomina
- 4 Weiteres Vorgehen



- 1 Projektvorstellung
- 2 Studie zu Modifikation in der Nominaldomäne
- 3 Morphologische Annotation von Nomina
- 4 Weiteres Vorgehen



Rudabe Akbari (SHK)
Hagen Hirschmann (assoziiert)
Anke Lüdeling (Projektleitung)
Julia Lukassek (Postdoc)
Anna Shadrova (assoziiert)
Shujun Wan (Stipendiatin)



- Erwerb von Registerkompetenz bei Lerner*innen des Deutschen als Fremdsprache (DaF)
- Analysen in schrift- und gesprochensprachlichen Korpora von Lerner*innen und Muttersprachler*innen im Vergleich (Falko, BeMaTaC)
- variationistische Perspektive kombiniert mit multi-dimensionalen und Regressionsmodellen



Hypothese 1

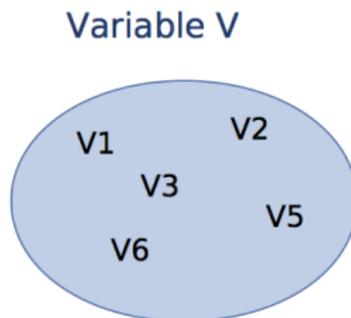
Selbst fortgeschrittene Lerner*innen einer Fremdsprache unterscheiden nicht in gleicher Weise zwischen Registern wie Muttersprachler*innen. Wir erwarten, dass wir andere Varianten derselben Variable in Lerner Sprache im Vergleich zur Muttersprache im gleichen Register finden.



Hypothesen

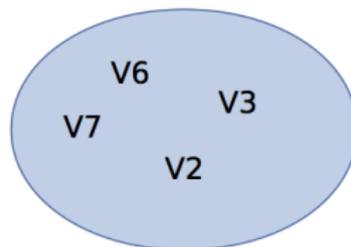
L1

argumentative
essays



L2

argumentative
essays



Hypthese 2

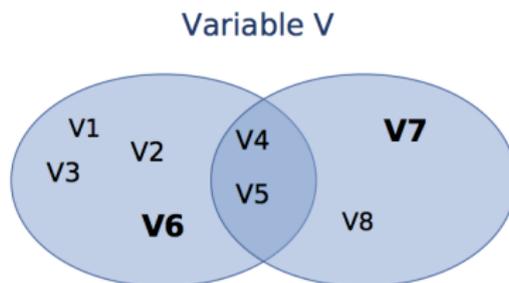
Selbst fortgeschrittene Lerner*innen einer Fremdsprache unterscheiden nicht so stark zwischen Registern wie es Muttersprachler*innen tun. Lerner*sprache ist homogener über Register hinweg als Muttersprache. Wir erwarten, dass wir bei Lerner*innen weniger Varianten für eine Variable finden als bei Muttersprachler*innen.



Hypothesen

L1

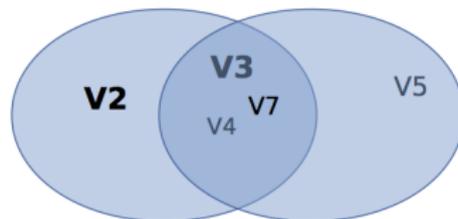
Argumentative essay



Map description

L2

Argumentative essay



Map description



Wir testen die Hypothesen an zwei verschiedenen linguistischen Phänomenen:

- **Modifikation in der Nominaldomäne**

- (1) wildes Tier; Tier aus der Wildnis; Tier der Wildnis; Wildtier;
Tier, das in der Wildnis lebt



Wir testen die Hypothesen an zwei verschiedenen linguistischen Phänomenen:

- **Modifikation in der Nominaldomäne**

- (1) wildes Tier; Tier aus der Wildnis; Tier der Wildnis; Wildtier;
Tier, das in der Wildnis lebt

- **Modifikation in der Satzdomäne**

- (2) möglicherweise; es ist möglich, dass ...; [morphologische Modusmarker]



Modifikation ist ein repräsentatives Phänomen für L1-L2-Unterschiede, weil Modifikation von Lerner*innen generell weniger verwendet wird als von Muttersprachler*innen.

(Hirschmann 2015, Hirschmann et al. 2013)

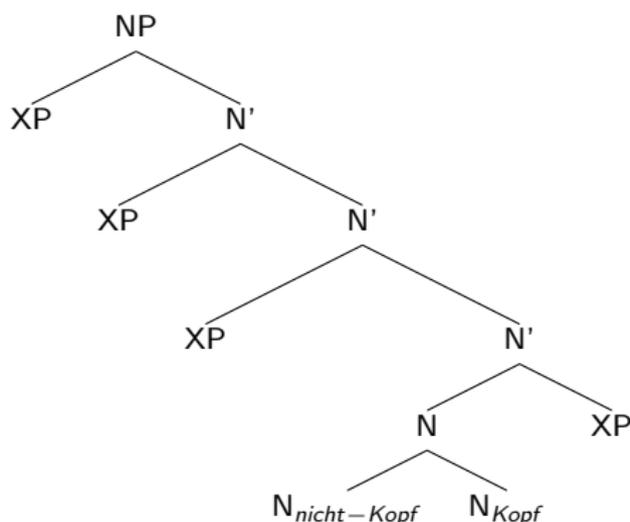


- 1 Projektvorstellung
- 2 Studie zu Modifikation in der Nominaldomäne
- 3 Morphologische Annotation von Nomina
- 4 Weiteres Vorgehen



Definition

Unter Modifikation von Nomina verstehen wir alle strukturellen Konfigurationen, in denen ein nominaler Kopf bzw. eine seiner Projektionsebenen um eine Konstituente erweitert wird.



Eckdaten zu **Falko** (= Fehlerannotiertes Lernerkorpus)

- tief annotiertes Korpus geschriebener deutscher Texte
- hier relevant: Zielhypothese 1 (ZH1) und davon abhängige POS- und Dependenz-Annotationen
- 381447 Token
- Subkorpora: L1 vs. L2, Essays vs. Zusammenfassungen
- durchsuchbar in ANNIS



Eckdaten zu **Falko** (= Fehlerannotiertes Lernerkorpus)

- tief annotiertes Korpus geschriebener deutscher Texte
- hier relevant: Zielhypothese 1 (ZH1) und davon abhängige POS- und Dependenz-Annotationen
- 381447 Token
- Subkorpora: L1 vs. L2, Essays vs. Zusammenfassungen
- durchsuchbar in ANNIS

Eckdaten zu **Falko Essay**

- Essays zu vier Erörterungsthemen (Studium, Kriminalität, Feminismus, Entlohnung)
- L2: 122791 Token
- L1: 68480 Token



Hypothese 1 Nomenmodifikation

Für die Variable **Nomenmodifikation** erwarten wir, dass Lerner*innen andere Varianten verwenden als Muttersprachler*innen. In Falko-Essay L2 werden wir also andere Varianten von Nomenmodifikation finden als in Falko-Essay L1.



- 1 Explorative Suche: Alle von N abhängigen Token



- 1 Explorative Suche: Alle von N abhängigen Token
- 2 Sichtung und Bereinigung um:
 - ▶ Determinierer
 - ▶ adverbiale Modifikatoren
 - ▶ Partikel, deren Target eine andere Konstituente ist
 - ▶ Negation mit Skopus über andere Konstituenten
 - ▶ zweites oder folgendes Glied einer Aufzählung
 - ▶ Koordinationen



- 1 Explorative Suche: Alle von N abhängigen Token
 - 2 Sichtung und Bereinigung um:
 - ▶ Determinierer
 - ▶ adverbiale Modifikatoren
 - ▶ Partikel, deren Target eine andere Konstituente ist
 - ▶ Negation mit Skopus über andere Konstituenten
 - ▶ zweites oder folgendes Glied einer Aufzählung
 - ▶ Koordinationen
- ⇒ Liste der zu erwartenden Modifikationsstrukturen



- 1 Explorative Suche: Alle von N abhängigen Token
 - 2 Sichtung und Bereinigung um:
 - ▶ Determinierer
 - ▶ adverbiale Modifikatoren
 - ▶ Partikel, deren Target eine andere Konstituente ist
 - ▶ Negation mit Skopus über andere Konstituenten
 - ▶ zweites oder folgendes Glied einer Aufzählung
 - ▶ Koordinationen
- ⇒ Liste der zu erwartenden Modifikationsstrukturen
- 3 Entwicklung von spezifischen Suchanfragen für einzelne Modifikationsstrukturtypen



Zu erwartende Modifikationsstrukturen:

- Adjektivattribute
- Genitivattribute
- Präpositionalattribute
- Adverbattribute
- Fokuspartikeln
- Adjunktorattribute
- Attributsätze
- Appositionen
- Prädikative



Suchstrategie: Datenerhebung

Ergebnis der Suche: **Modifikatordatenbank**

H	K	N
Nomen_ZH1	abh_Token_ZH1	ganzer_Beleg
Grade	gewissen	4. " Meiner Meinung nach stimmt dieser Aussage bis zu einem gewissen Grade - besonders in <u>Dänemark</u> wo es z. B. kein
Pflicht	für	6. - besonders in <u>Dänemark</u> wo es z. B. kein <u>Pflicht</u> für die Studenten an der Kopenhagener <u>Wirtschaftsuniversität</u>
Studenten	an	8. <u>Dänemark</u> wo es z. B. kein <u>Pflicht</u> für die Studenten an der Kopenhagener <u>Wirtschaftsuniversität</u> ein Praktikum
Grade	hohem	13. . Aber ob die Universitätsabschlüsse praxisorientiert sind , hängt in hohem Grade auch davon ab , womit sich die Studenten s
Zugang	theoretische	17. ein Job als Forscher <u>wünscht</u> , kann der oft sehr theoretischen Zugang der Universitätsabschlüsse von <u>grossen</u> Nutzen sein . U
Zugang	Universitätsabschlü	18. Job als Forscher <u>wünscht</u> , kann der oft sehr theoretischen Zugang der Universitätsabschlüsse von <u>grossen</u> Nutzen sein . Um e
Universitätsabschlüsse	von	20. Forscher <u>wünscht</u> , kann der oft sehr theoretischen Zugang der Universitätsabschlüsse von <u>grossen</u> Nutzen sein . Um ein wisse
Nutzen	großem	21. , kann der oft sehr theoretischen Zugang der Universitätsabschlüsse von <u>grossen</u> Nutzen sein . Um ein wissenschaftliches Ge
Gebiet	wissenschaftliches	23. Zugang der Universitätsabschlüsse von <u>grossen</u> Nutzen sein . Um ein wissenschaftliches Gebiet in der Tiefe zu behandeln , br
Zugang	systematisierten	26. Gebiet in der Tiefe zu behandeln. braucht man eine systematisierte und theoretische Zugänge als Werkzeuge. In Bezug auf die

Manuelle Ergänzung:

- Modifikatorphrase
- genaue syntaktische Funktion des Modifikators



- **Modikatorphrasen** können wir mit den vorhandenen Mitteln in Falko auffinden.



- **Modikatorphrasen** können wir mit den vorhandenen Mitteln in Falko auffinden.
- **Modifikation innerhalb der Nomina** (= Nicht-Köpfe von Komposita) sind bisher nicht automatisiert auffindbar.
 - (3) Arbeit, Arbeiten, Arbeitsaufgaben, Gruppenarbeit, Arbeiter, Verarbeitung



- **Modikatorphrasen** können wir mit den vorhandenen Mitteln in Falko auffinden.
- **Modifikation innerhalb der Nomina** (= Nicht-Köpfe von Komposita) sind bisher nicht automatisiert auffindbar.
 - (3) Arbeit, Arbeiten, Arbeitsaufgaben, Gruppenarbeit, Arbeiter, Verarbeitung

⇒ Annotation der morphologischen Eigenschaften von Nomina



- 1 Projektvorstellung
- 2 Studie zu Modifikation in der Nominaldomäne
- 3 Morphologische Annotation von Nomina**
- 4 Weiteres Vorgehen



Umfassendes Klassifikationssystem für die morphologischen Eigenschaften von Nomina

- **Primäres Ziel:** Auffindbarkeit von Komposita; Sichtbarmachung einer der Varianten der Variable *Nomenmodifikation*



Umfassendes Klassifikationssystem für die morphologischen Eigenschaften von Nomina

- **Primäres Ziel:** Auffindbarkeit von Komposita; Sichtbarmachung einer der Varianten der Variable *Nomenmodifikation*
- **Sekundäres Ziel:** Dokumentation von wortbildungsmorphologischer Komplexität in der Nominaldomäne



Umfassendes Klassifikationssystem für die morphologischen Eigenschaften von Nomina

- **Primäres Ziel:** Auffindbarkeit von Komposita; Sichtbarmachung einer der Varianten der Variable *Nomenmodifikation*
- **Sekundäres Ziel:** Dokumentation von wortbildungsmorphologischer Komplexität in der Nominaldomäne

⇒ Annotation kann für andere Fragestellungen benutzt werden



Umfassendes Klassifikationssystem für die morphologischen Eigenschaften von Nomina

- **Primäres Ziel:** Auffindbarkeit von Komposita; Sichtbarmachung einer der Varianten der Variable *Nomenmodifikation*
- **Sekundäres Ziel:** Dokumentation von wortbildungsmorphologischer Komplexität in der Nominaldomäne

⇒ Annotation kann für andere Fragestellungen benutzt werden

⇒ Annotationsrichtlinie kann für andere Korpora benutzt werden.



Maxime: Die Annotationsrichtlinie soll so gestaltet sein, dass auf ihrer Basis zwei oder mehr informierte Annotator*innen zu einem nachvollziehbaren Konsens für jedes zu annotierende Nomen kommen können.



Maxime: Die Annotationsrichtlinie soll so gestaltet sein, dass auf ihrer Basis zwei oder mehr informierte Annotator*innen zu einem nachvollziehbaren Konsens für jedes zu annotierende Nomen kommen können.

- Erarbeitung in mehreren Runden mit Testung und Abgleich der Ergebnisse
- Testläufe mit projektinternen und -externen Annotator*innen



- 5 Texte aus dem L2-Subkorpus
- 3 externe Annotatorinnen
- 287 Nomen zur Annotation
- beachtliche Übereinstimmung (substantial agreement)



- 5 Texte aus dem L2-Subkorpus
- 3 externe Annotatorinnen
- 287 Nomen zur Annotation
- beachtliche Übereinstimmung (substantial agreement)

Inter-annotator Agreement

Raters	=	3
Subjects	=	287
Fleiss' Kappa	=	0.776
z	=	46.6
p-value	=	0



Annotationsebenen

B	C	D	E	F
ZH1:ZH1	ZH1:ZH1gpos	Morph-Anno	Morph-Anno-fremd	Herkunft
2	CARD			
"	\$(
Die	ART			
meisten	PIAT			
Universitätsabschlüsse	NN	kdet		neokl
sind	VAFIN			
nicht	PTKNEG			
praxisorientiert	ADJD			
und	KON			
bereiten	VVFIN			
die	ART			
Studenten	NN	nnat	com	neokl
nicht	PTKNEG			
auf	APPR			
die	ART			
wirkliche	ADJA			
Welt	NN	sim		nat
vor	PTKVZ			
.	\$.			
Sie	PPER			
sind	VAFIN			
deswegen	PAV			
von	APPR			
geringem	ADJA			
Wert	NN	kon		nat
.	\$.			
"	\$(
In	APPR			
dem	ART			
Folgenden	NN	trans		nat
wird	VAFIN			

- manuelle Annotation aller ZH1-Token, die das POS-Tag NN haben, im Kontext



Annotationsebenen

B	C	D	E	F
ZH1:ZH1	ZH1:ZH1gpos	Morph-Anno	Morph-Anno-fremd	Herkunft
2	CARD			
"	\$(
Die	ART			
meisten	PIAT			
Universitätsabschlüsse	NN	kdet		neokl
sind	VAFIN			
nicht	PTKNEG			
praxisorientiert	ADID			
und	KON			
bereiten	VVFIN			
die	ART			
Studenten	NN	nnat	com	neokl
nicht	PTKNEG			
auf	APPR			
die	ART			
wirkliche	ADJA			
Welt	NN	sim		nat
vor	PTKVZ			
.	\$.			
Sie	PPER			
sind	VAFIN			
deswegen	PAV			
von	APPR			
geringem	ADJA			
Wert	NN	kon		nat
.	\$.			
"	\$(
In	APPR			
dem	ART			
Folgenden	NN	trans		nat
wird	VAFIN			

- manuelle Annotation aller ZH1-Token, die das POS-Tag NN haben, im Kontext

- Annotation des hierarchisch höchsten Wortbildungsprozesses

[[[Ein[reis_v]v]e_N][[ver[bot_v]v]N]N]



Annotationsebenen

B	C	D	E	F
ZH1:ZH1	ZH1:ZH1gpos	Morph-Anno	Morph-Anno-fremd	Herkunft
2	CARD			
"	\$(
Die	ART			
meisten	PIAT			
Universitätsabschlüsse	NN	kdet		neokl
sind	VAFIN			
nicht	PTKNEG			
praxisorientiert	ADJD			
und	KON			
bereiten	VVFIN			
die	ART			
Studenten	NN	nnat	com	neokl
nicht	PTKNEG			
auf	APPR			
die	ART			
wirkliche	ADJA			
Welt	NN	sim		nat
vor	PTKVZ			
.	\$.			
Sie	PPER			
sind	VAFIN			
deswegen	PAV			
von	APPR			
geringem	ADJA			
Wert	NN	kon		nat
.	\$.			
"	\$(
In	APPR			
dem	ART			
Folgenden	NN	trans		nat
wird	VAFIN			

- manuelle Annotation aller ZH1-Token, die das POS-Tag NN haben, im Kontext

- Annotation des hierarchisch höchsten Wortbildungsprozesses

$[[[Ein[reis_V]_V]e_N][[ver[bot_V]_V]N]_N]_N$

- drei Annotationsebenen:
 - ▶ Morph-Anno
 - ▶ Morph-Anno-fremd
 - ▶ Herkunft

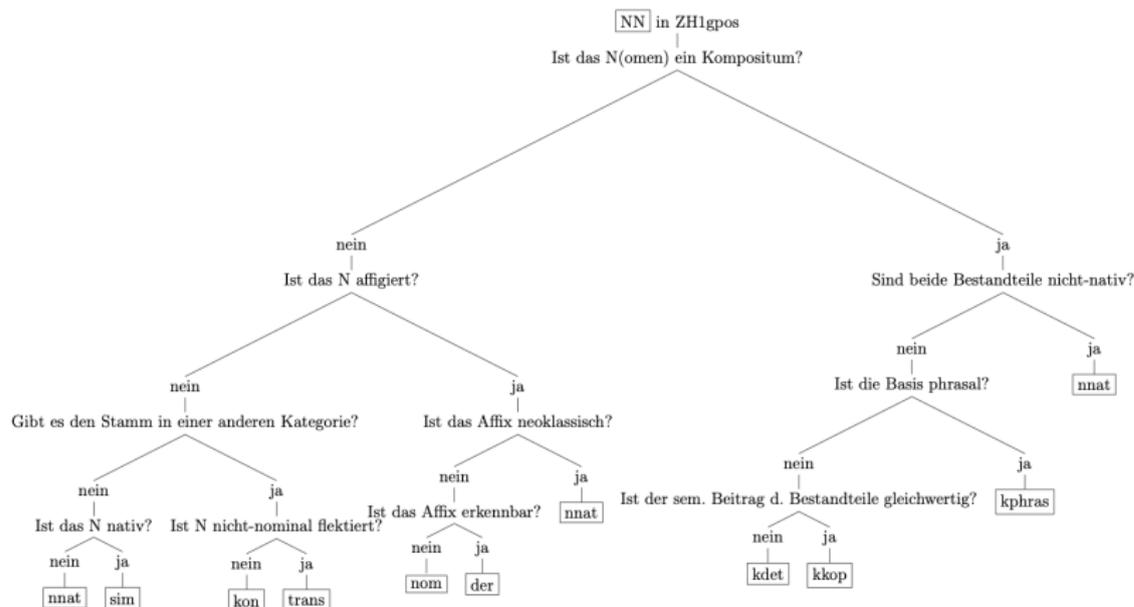


Annotationsebenen: Morph-Anno

Determinativkomposita	kdet	<i>Arbeitszeit</i>
Kopulativkomposita	kkop	<i>Spielertrainer</i>
Phrasenkomposita	kphras	<i>Trimm-Dich-Pfad</i>
Derivat	der	<i>Verarbeitung</i>
Konversion	kon	<i>Arbeit</i>
Transposition	trans	<i>Arbeiten</i>
Andere Nominalisierung	nom	<i>Fahrt</i>
Simplex	sim	<i>Tisch</i>
Nicht-native Nomen	nnat	<i>Universität</i>



Annotationsebenen: Morph-Anno



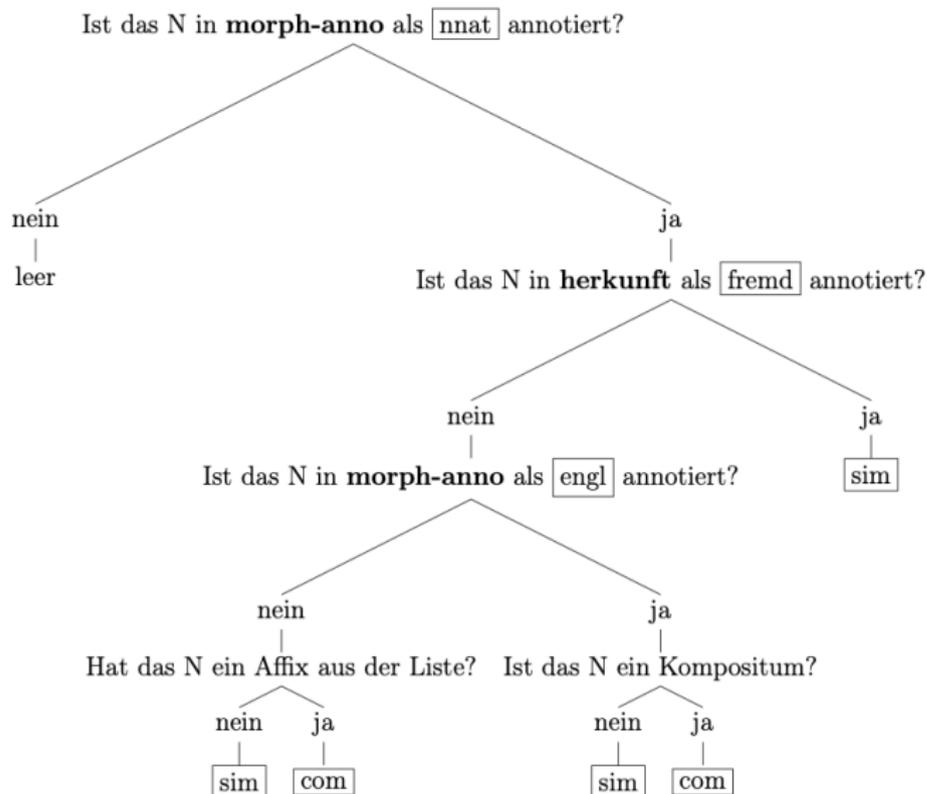
Wenn auf der Ebene Morph-Anno das Tag nnat steht, werden die morpholog. Eigenschaften auf der Ebene Morph-Anno-fremd spezifiziert.

morphologisch einfach sim
morphologisch komplex com

Job, Profil, Niveau
Kontext, Universität, Teamwork



Annotationsebenen: Morph-Anno-fremd



Annotationsebenen: Herkunft

Alle Nomen mit Beteiligung eines nicht-nativen Bestandteils werden mit einem der drei folgenden Tags annotiert:

lat. oder grie. Ursprung

neokl

Automatismus, Universität

engl. Ursprung

engl

Job, Jobbeschreibung

sonstiger fremdspr. Ursprung

fremd

Algebra, Kriminalitätsniveau

Alle Nomen ohne Beteiligung eines nicht-nativen Bestandteils haben das Tag **nat**.



Annotationsebenen: Herkunft

Ist das N in **morph-anno** als **nnat** annotiert?

nein

Hat das N einen nicht-nativen Wortbestandteil?

nein

nat

ja

Ist das N lateinischen oder griechischen Ursprungs?

nein

Ist die Aussprache am Englischen angelehnt?

nein

fremd

ja

engl

ja

neokl

ja

Ist das N lateinischen oder griechischen Ursprungs?

nein

Ist die Aussprache am Englischen angelehnt?

nein

fremd

ja

engl

ja

neokl



Beispiele aus der Annotation

ZH1	ZH1gpos	Morph-Anno	Morph-Anno-fremd	Herkunft
Schwierigkeiten	NN	der		nat
Universitätsabschlüsse	NN	kdet		neokl
Welt	NN	sim		nat
Studium	NN	nnat	sim	neokl
Studenten	NN	nnat	com	neokl
Studienjob	NN	kdet		engl



Nomenanntation ermöglicht:

- die Auffindbarkeit von wortinternen Modifikatoren
- die Nutzung für andere morphologisch inspirierte Fragestellungen



Nomenanntation ermöglicht:

- die Auffindbarkeit von wortinternen Modifikatoren
- die Nutzung für andere morphologisch inspirierte Fragestellungen

Die Annotation wirft neue Fragen auf:

- Welche Art von Komplexität wird dokumentiert?
- Welche Aussagen über sprachliches Wissen von Lerner*innen kann man auf Basis dessen machen?



- 1 Projektvorstellung
- 2 Studie zu Modifikation in der Nominaldomäne
- 3 Morphologische Annotation von Nomina
- 4 Weiteres Vorgehen



- 1 Abschluss der Nomen-Annotation
- 2 Datenerhebung: gruppenweise pro Modifikatortyp
- 3 Annotation: Kopf und (kompletter) Modifikator
- 4 Auswertung: Vergleich zwischen L1- und L2-Sprecher*innen
- 5 Iteration des gesamten Verfahrens für BeMaTaC



Vielen Dank!



- 1 BeMaTac (2013) Ein tief annotiertes multimodales Map-Task-Korpus gesprochener Lerner- und Muttersprache; verfügbar unter <https://hu.berlin/59537>
- 2 Falko (2010) Ein fehlerannotiertes Lernerkorpus des Deutschen als Fremdsprache; verfügbar unter <https://hu.berlin/27559>
- 3 Hirschmann, H. (2015) *Modifikatoren im Deutschen. Ihre Klassifizierung und varietätenspezifische Verwendung*. Studien zur Deutschen Grammatik 89. Stauffenburg.
- 4 Hirschmann, H., A. Lüdeling, I. Rehbein, M. Reznicek & A. Zeldes (2013) *Underuse of syntactic categories in Falko. A case study on modification*. In: 20 Years of Learner Corpus Research. Looking Back, Moving Ahead. Ed. by S. Granger, G. Gilquin & F. Meunier. Presses Universitaires de Louvain, 223-234.

