

Evaluation der Lernforschriftte anhand des Korpusaufbaus und der korpusbasierten Erforschung der Lernalersprache



Prof. Dr. Yuan Li, Zhejiang Universität, China
Forschungsbereiche: Didaktik Deutsch als Fremdsprache,
Korpuslinguistik, Spracherwerbsforschung
E-mail: liyuan1972@zju.edu.cn

Chinese
Deutschler-
Korpus
(CDLK)





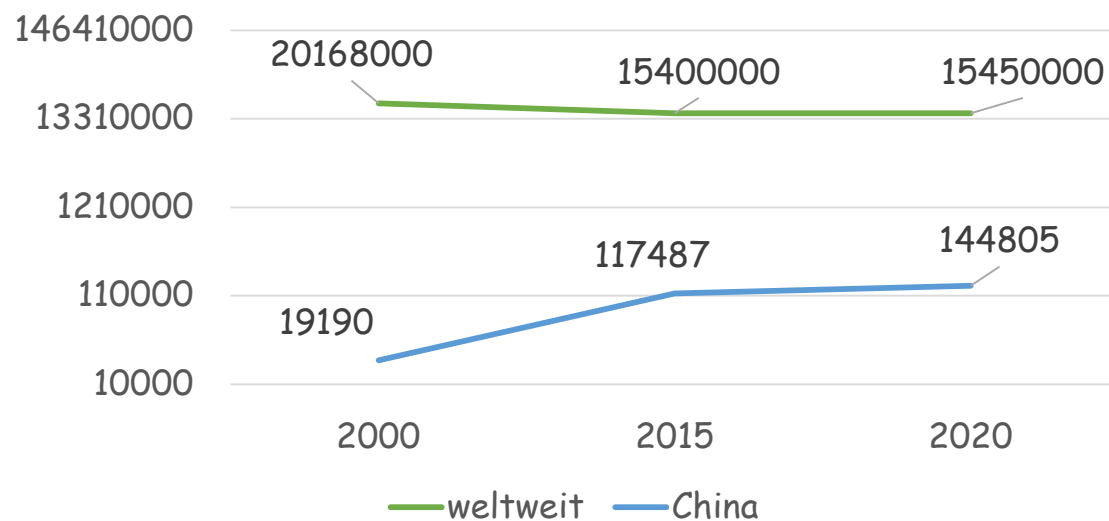
Warum wird CDLK
aufgebaut?

Was ist CDLK?

Wie kann man mithilfe CDLK
Untersuchungen durchführen?

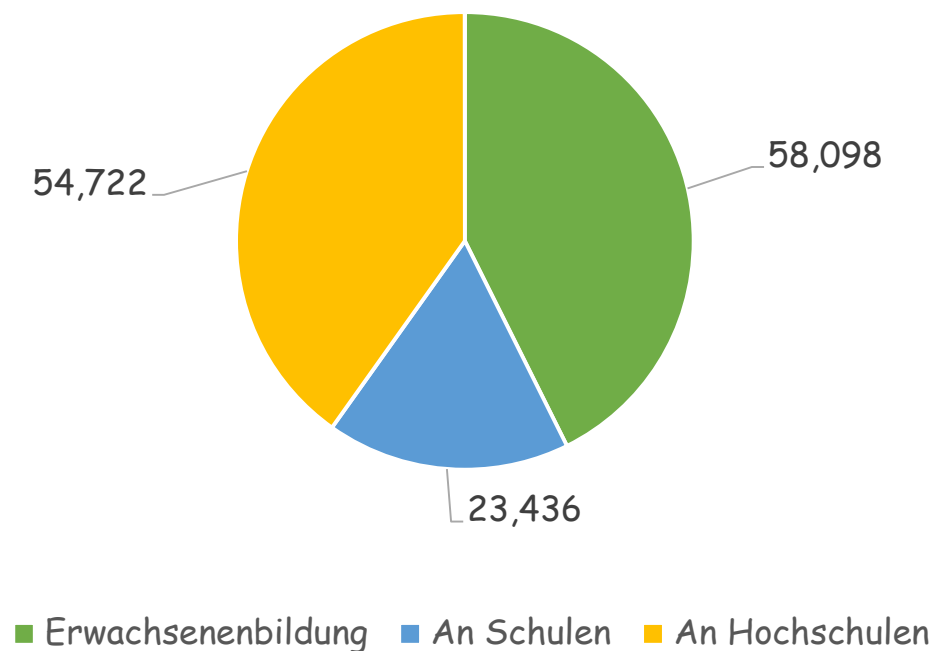
Warum?

Zahl der Deutschlernenden weltweit und in China



(Quelle: Auswärtiges Amt 2000, 2015, 2020)

Verteilung der Deutschlernenden in China* (2020)



Quelle: Auswärtiges Amt, 2020
(*exkl. Honkong, Taiwan)

Warum?

Aber nur **wenige** Forschungen dazu ...

Und:

-
- **meist** einen **qualitativen** und **induktiven** Ansatz verfolgen
- sich **selten** auf umfangreiche **Korpora** stützen

• chine

Deutsch

(Quelle: Datenerhebungen der ständigen ... auf 2000, 2005, des Netzwerk Deutsch 2010 und des Auswärtigen Amts 2019, 2020, eigene Darstellung)

Warum?

Korpora mit chinesischen Deutschlernern

| | Falko | AleSKo | Kobalt | BFSU (KICG) |
|---------------------------------------|--|--------|--------|--------------------------------------|
| Zugang | öffentlich zugänglich | | | nicht öffentlich zugänglich |
| Tokenanzahl | etwa 10.000 | | | etwa 700.000 |
| Forschungsgegenstand | fortgeschrittene Deutschlerner auf dem Niveau B2 | | | chinesische Germanistik-studierenden |
| Textanzahl chinesischer Deutschlerner | 43 | 39 | 20 | / |

Warum?

□ b

□ r

□ b



schülerInnen

Wie DFG Projekt)

Was?

schriftliche Texte im Unterricht von chinesischen Deutschlernern

- Aufgaben auf Chinesisch gegeben
- möglichst viel schreiben
- innerhalb 30 Minuten
- ohne Hilfsmittel

Was?

6 Themen, 4 Textsorten:

| Zielgruppe | Thema 1 | Thema 2 | Thema 3 | Thema 4 | Thema 5 |
|-------------------|---------------|---|-----------------------|--------------------------------------|-------------------------------|
| Schüler- Innen | | Meine Hobbys | | | |
| | | deskriptiv | | | |
| Student- Innen | Meine Familie | Arbeiten oder Weiterstudie ren nach dem Abschluss | Bild- beschreibung | Handy im Unterricht oder nicht | Ein besonderes Erlebnis |
| | deskriptiv | argumentativ | bild- beschreibend | argumentativ | erzählend |

Was?

das **erste** Chinesisches
Deutschlerner-Korpus,
das ...

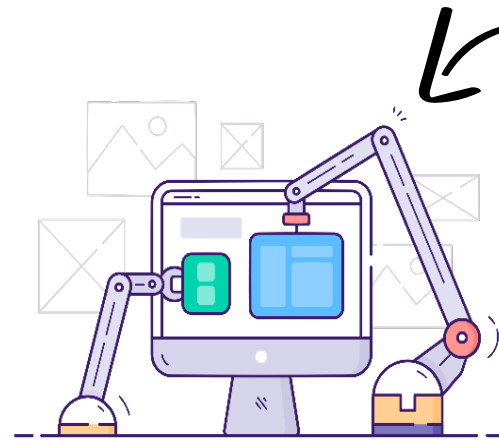


“Mehr-
ebenen-
annota-
tion”

“umfang-
reich”

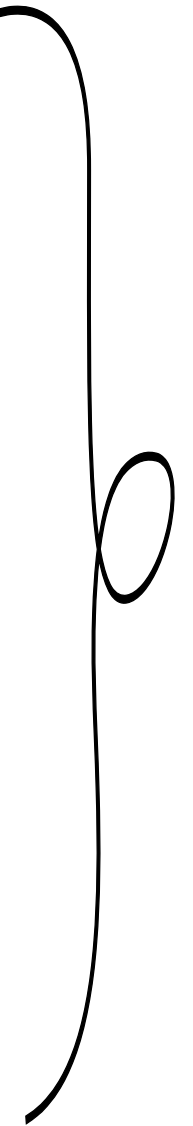


“multidi-
mensionale
Meta-
daten”



“umfangreich”

geografische
Verbreitung



| Semester | Unterstufen der Schulen | | Oberstufen der Schulen | | Universitäten | | Gesamt | |
|---------------|-------------------------|-----------|------------------------|-----------|---------------|-----------|-----------|-----------|
| | gesammelt | annotiert | gesammelt | annotiert | gesammelt | annotiert | gesammelt | annotiert |
| 1 | 276 | 0 | 262 | 51 | 302 | 109 | 840 | 160 |
| 2 | 257 | 151 | 339 | 67 | 448 | 117 | 1044 | 335 |
| 3 | 188 | 76 | 315 | 84 | 249 | 84 | 752 | 244 |
| 4 | 289 | 118 | 320 | 128 | 390 | 137 | 999 | 383 |
| 5 | 238 | 92 | 147 | 55 | 115 | 34 | 500 | 181 |
| 6 | 247 | 38 | 7 | 1 | 223 | 79 | 477 | 118 |
| 7 | 136 | 65 | | | 93 | 21 | 229 | 86 |
| 8 | 230 | 0 | | | 30 | 16 | 260 | 16 |
| 9 | 170 | 62 | | | | | 170 | 62 |
| 10 | 207 | 57 | | | | | 207 | 57 |
| 11 | 107 | 58 | | | | | 107 | 58 |
| 12 | 102 | 0 | | | | | 102 | 0 |
| 13 | 28 | 0 | | | | | 28 | 0 |
| Gesamt | 2475 | 717 | 1390 | 386 | 1850 | 597 | 5715 | 1700 |

Stand: 23.01.2022

„umfangreich“

Querschnitt:

- unterschiedliche Niveaus
- unterschiedliche Schulen/Unis

Längsschnitt:

- 06.2019 bis jetzt (jedes Semester)
- ausgewählte Gruppen einer Schule und Uni

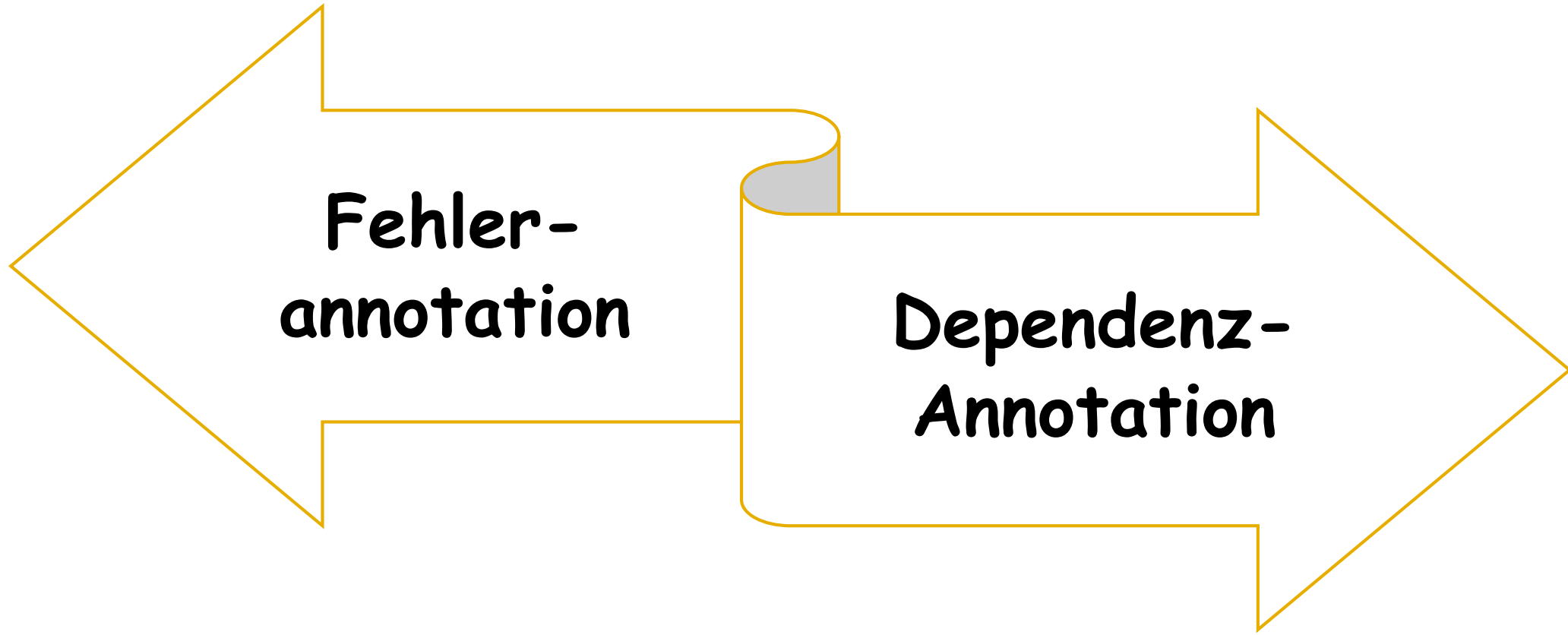
geografische
Verbreitung

Schulen &
Universitäten

umfassende
Lernphasen

Querschnitt &
Längsschnitt

"Mehrebenen- -annotation"



“Mehrebenen-annotation”



Fehler-annotation

Theoretische Grundlage:

in Anlehnung an die Fehlertaxonomie von Dr. Andreas Nolda und Dr. Hagen Hirschmann

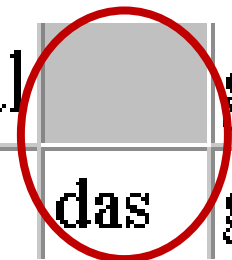
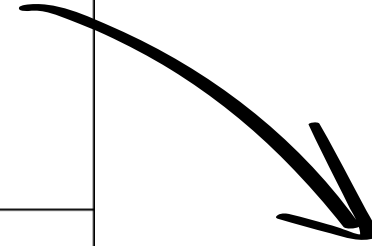
| | |
|----------------------|---------------------------|
| Normalisierungsebene | orthographische Fehler |
| | morphosyntaktische Fehler |
| | syntaktische Fehler |
| | lexikalische Fehler |
| | semantische Fehler |
| höhere Ebene | Fehler bei der Logik |
| | Fehler bei dem Stil |
| | Transfer-Fehler |

"Mehrebenen-annotation"



Fehler-annotation

| Klasse | Unterkategorie | Typ | Abkürzung |
|--------|----------------|------------------------------------|-----------|
| Syn | Valenz | Verbvalenz | ValV |
| Syn | Valenz | Reflexivpronomen in der Verbvalenz | |
| Syn | Valenz | Subst | |
| Syn | Valenz | Adjektivvalenz | ValADJ |
| Syn | Valenz | Adpositionsvalenz | ValAP |
| Syn | Valenz | Wechselpräpositionsvalenz | ValWP |



Nach Hause habe ich manchmal gemacht .
Zu Hause habe ich manchmal das gemacht .

“Mehrebenen-annotation”



Fehler-annotation

Originaläußerung

automatisch
annotiert, manuell
überprüft

Zielhypothese

automatisch
annotiert, manuell
überprüft

Normalisierungsebene (manuelle
Annotation mit
computergestützter
Hilfe)

höhere Ebene
(manuelle
Annotation mit
computergestützter
Hilfe)

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|-----------------|------|------------|----------|------|-----|------|-------|-----|-----|---|-------|------|--------|--------|--------|------------|------------|-----|
| X [word] | Ein | besonderes | Erlebnis | Wenn | ich | fünf | Jahre | alt | war | | g | ich | mit | meiner | Mutter | nach | Australien | . |
| X [S] | Ü | | | s1 | | | | | | | | | | | | | | |
| X [pos] | ART | ADJA | NP | | | | | | | | VVFIN | PPER | APPR | PPOSAT | NN | APPR | NE | \$. |
| X [lemma] | eine | bes | | | | | | | | | | ich | mit | mein | Mutter | nach | Australien | . |
| X [ZH] | | | | | | | | | | | | mit | meiner | Mutter | nach | Australien | . | |
| X [ZHDiff] | | | | | | | | | | | | | | | | | | |
| X [ZHS] | | | | | | | | | | | | | | | | | | |
| X [ZHpos] | | | | | | | | | | | | | R | PPOSAT | NN | APPR | NE | \$. |
| X [ZHlemma] | | | | | | | | | | | | | mein | Mutter | nach | Australien | . | |
| X [FehlerOrth] | | | | | | | | | | | | | | | | | | |
| X [FehlerMorph] | | | | | | | | | | | | | | | | | | |
| X [FehlerSyn] | | | | | | | | | | | | | | | | | | |
| X [FehlerLex] | | | | | | | | | | | | | | | | | | |
| X [FehlerSem] | | | | | | | | | | | | | | | | | | |
| X [Stil] | | | | | | | | | | | | | | | | | | |
| X [Logik] | | | | | | | | | | | | | | | | | | |
| X [Transfer] | | | | | | | | | | | | | | | | | | |

EXMARaLDA

"Mehrebenen-annotation"



Dependenz-Annotation

```
Python 3.8.5 Shell
File Edit Shell Debug Options Window Help
Python 3.8.5 (tags/v3.8.5:580fbb0, Jul 20 2020, 15:57:54) [MSC v.1924 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> from txt_to_conllu import *
>>> to_conllu(r'D:\Test\word', r'D:\Test2\Ergebnisse')
```

Ergebnisse0210112006YX001 - 记事本

| # | sent_id | word | lemma | POS | NER | chunk | is_start | is_end | SpaceAfter |
|-------------|---------|---------|---------|-------|--------|-------|----------|--------|--------------|
| 1 | 1 | Meine | Meine | DET | PPOSAT | - | 2 | nk | - |
| 2 | 2 | Familie | Familie | NOUN | NN | - | 0 | root | - |
| 3 | 3 | | | SPACE | _SP | - | 2 | - | SpaceAfter=N |
| # sent_id 2 | | | | | | | | | |
| 1 | 1 | Meine | Meine | DET | PPOSAT | - | 2 | nk | - |
| 2 | 2 | Familie | Familie | NOUN | NN | - | 3 | sb | - |
| 3 | 3 | wohnt | wohnen | VERB | VVFIN | - | 0 | root | - |
| 4 | 4 | in | in | ADP | APPR | - | 3 | mo | - |
| 5 | 5 | Huaning | Huaning | PROPN | NE | - | 4 | nk | SpaceAfter=N |
| 6 | 6 | . | . | PUNCT | \$. | - | 3 | punct | - |
| # sent_id 3 | | | | | | | | | |
| 1 | 1 | Huaning | Huaning | PROPN | NE | - | 2 | sb | - |
| 2 | 2 | ist | sein | AUX | VAFIN | - | 0 | root | - |
| 3 | 3 | eine | einen | DET | ART | - | 5 | nk | - |
| 4 | 4 | schöne | schöne | ADJ | ADJA | - | 5 | nk | - |
| 5 | 5 | Stadt | Stadt | NOUN | NN | - | 2 | pd | SpaceAfter=N |
| 6 | 6 | . | . | PUNCT | \$. | - | 2 | punct | - |
| # sent_id 4 | | | | | | | | | |
| 1 | 1 | Meine | Meine | DET | PPOSAT | - | 2 | nk | - |
| 2 | 2 | Familie | Familie | NOUN | NN | - | 3 | sb | - |
| 3 | 3 | gibt | geben | VERB | VVFIN | - | 0 | root | SpaceAfter=N |
| 4 | 4 | , | , | X | FM | - | 3 | oa | SpaceAfter=N |

Dependenzrelation
automatisch mit SpaCy
annotieren



manuell
überprüfen

"multidimensionale Metadaten"

Thema 1

| | | | | | | | | | | |
|--|---|--|------------|----------------------|--------|--|--------------|--|------------|--|
| Bitte füllen Sie das Formular auf Chinesisch aus (möglichst vollständig) | Name | | Schule/Uni | | Klasse | | Geburtsdatum | | Geschlecht | |
| | Deutschlerndauer (Semester) | | | bestandene Prüfungen | | | | | | |
| | Ich habe auch andere(n) Fremdsprache(n) gelernt (Bitte detaillierte Informationen angeben, z.B. Lerndauer, bestandene Prüfungen usw.) | | | | | | | | | |

Bitte schreiben Sie einen Aufsatz zum Thema "**Meine Familie**" auf Deutsch. Verwenden Sie so **viele und vielfältige** Ausdrücke wie möglich.

Für LehrerInnen

| | | | | | | | | |
|--------------------|---------------------|---------------------|---------------|--|---------------------------|-----------------------|----------------------|---|
| Schule/Uni | | Name | | Erwartete Abschlussziele (Hochschulaufnahmeprüfung auf Englisch/Deutsch, Auslandsstudium usw.) | | | | |
| PaSCH Schule? | | E-Mail | | | | | | Lernmaterialien (Bei mehreren Lehrmaterialien geben Sie bitte den Anteil der Nutzung verschiedener Lehrmaterialien an) |
| DSD Schule? | | | | | | | | |
| Schul-/Studienjahr | Deutschkurs? | | Lerneranzahl | | | Lernstunden | | |
| | als 1. Fremdsprache | als 2. Fremdsprache | Klassenanzahl | Lerneranzahl pro Klasse | Lerneranzahl pro Jahrgang | Lernstunden pro Woche | Lernstunden pro Jahr | |
| 1. | | | | | | | | |
| 2. | | | | | | | | |
| 3. | | | | | | | | |
| ... | | | | | | | | |
| Angabedatum | | | | | | | | |

“multidimensionale Metadaten”



是?
für Sie?
方式是否有所改变? (有哪些变化?)
lernt hat] Hat sich Ihre Art, Deutsch zu

段?
lernen? Welche Stufe werden Sie lernen?
划? 为什么? 德语学习对您的规划有

den?

当前

oder
und

老师、

(z.B.
Anzahl
teilen

1. 您的性别是? [单选

A. 男 (männlich)

B. 女 (weiblich)

2. 您的出生年份是? [单

1960

1961

1962

3. 可以描述一节
提高学生积极
Können Sie un
Lehrmethoden/
Hören, Spreche

4. 您对目前的德
核标准、教材
Sind Sie mit
(Qualität/Fähigkeit, die Lehrer haben sollten; Bewertungsstandards für Lehrer;
Lehrbücher und Hilfsmaterialien, Klassengröße und Anzahl der Lehrer, Anzahl der
Unterrichtsstunden, Digitalisierung beim Unterricht usw.)

5. 在推动德语学科发展过程中您有更多的需求吗?
Haben Sie weiteren Bedarf bei der Förderung der Entwicklung des deutschen Fachs?

| | Komplexität | Korrektheit | Flüssigkeit | Andere |
|--------|--|---|--|--|
| Lexik | Der Gebrauch von Konnektoren in den argumentativen Texten chinesischer DaF-Studierender (längsschnittlich) | | | Eigenschaften und Wandel der Wortlänge im Deutschen (querschnittlich + längsschnittlich) |
| | | Zur diachronen Entwicklung deutscher Präpositionen (längsschnittlich) | | |
| | | Entwicklung der lexikalischen Fehler chinesischer Deutschlerner (längsschnittlich) | | |
| | Eine dependenzbasierte Studie zum Verbgebrauch in der Schriftsprache chinesischer Deutschlerner (längsschnittlich) | | | |
| | Eine korpusbasierte Untersuchung zum Gebrauch der lexikalischen Chunks chinesischer Deutschlerner (längsschnittlich) | | | |
| | Studie zur Korrelation zwischen dem Alter beim Lernbeginn und der Lerngeschwindigkeit chinesischer Deutschlerner (querschnittlich) | Transfer bei dem Schriftdeutsch chinesischer Schüler*innen (längsschnittlich) | Studie zur Korrelation zwischen dem Alter beim Lernbeginn und der Korrektheit sowie Flüssigkeit chinesischer Deutschlerner (querschnittlich) | |
| Syntax | Die Entwicklung der syntaktischen Komplexität des Schriftdeutschen chinesischer SchülerInnen: Eine dependenzbasierte Studie (längsschnittlich) | Eine Studie zu Wortstellungsfehlern von Verben im Texten chinesischer Deutschschüler (längsschnittlich) | | Das Nachfeld im Deutschen (querschnittlich + längsschnittlich) |
| | Zur syntaktischen Komplexität des Schriftdeutschen der chinesischen Deutschlerner: Eine korpusbasierte Profilanalyse (querschnittlich) | | | Zum Gebrauch der Satzklammern bei chinesischen Deutschlernenden (längsschnittlich) |
| | Eine Untersuchung von Indikatoren der syntaktischen Komplexität in der geschriebenen Sprache bei Deutschlernenden | | | |
| Text | | Zur thematischen Progression des Schriftdeutschen chinesischer Deutschlerner (längsschnittlich) | | |

Vielen Dank für
Ihre
Aufmerksamkeit!

Prof. Dr. Yuan LI: liyuan1972@zju.edu.cn

