

# Dokumentation: Webservices für die Arbeit mit dem GeWiss-Korpus

Stand: Februar 2015

Autoren: Daniel Jettka, David Stoppel

## Inhaltsverzeichnis

I.	Einleitung und allgemeine Hinweise .....	2
II.	Voraussetzungen für die Nutzung der Webservices .....	2
III.	Allgemeine Hinweise zur Nutzung .....	3
IV.	Liste der Webservices.....	4
1.	Auflistung von Teilkorpora (frei zugänglich) .....	4
2.	Auflistung von Kommunikationen (frei zugänglich) .....	5
3.	Auflistung von Sprechern (frei zugänglich) .....	6
4.	Ausgabe von Metadaten (frei zugänglich) .....	7
5.	Ausgabe von Transkripten (passwortgeschützt) .....	7
6.	Wiedergabe von Audiodaten (passwortgeschützt).....	9
7.	Auflistung aller Typen von Metadaten (frei zugänglich) .....	10
8.	Auflistung aller Typen von Annotationen (frei zugänglich).....	11
9.	Einfache Konkordanzsuche - Text (frei zugänglich).....	12
10.	Einfache Konkordanzsuche - Annotationen (frei zugänglich).....	13
11.	Erweiterte Konkordanzsuche – Text (passwortgeschützt) .....	15
12.	Erweiterte Konkordanzsuche – Annotationen (passwortgeschützt) .....	18
	Weiterführende Literatur.....	22
	Anhang .....	23
	Liste aller Subkorpora.....	23
	Liste aller Annotationstypen mit Codes .....	23
	Liste aller Metadatatypen mit Codes .....	23

## I. Einleitung und allgemeine Hinweise

Im Zuge des CLARIN-Kurationsprojekts CLARIN-KP-GeWiss ist eine Reihe von RESTful Java Webservices entstanden, die für grundlegende Arbeiten mit dem GeWiss-Korpus genutzt werden können. Es handelt sich dabei um Server-Anwendungen, mit denen durch standardisierte, parametrisierte Anfragen des Nutzers eine Reihe unterschiedlicher Ausgaben erhalten werden können.

Die vorhandene Funktionalität umfasst z.B. die Auflistung vorhandener Ressourcen, wie

- Teilkorpora nach Sprache, akademischem Kontext, Diskursgenre
- Kommunikationen eines Teilkorpus,
- Sprecher in einer best. Kommunikation,
- Transkripte der Kommunikationen,
- Audioaufnahmen der Kommunikationen,
- Typen von Metadaten und Annotationen,

sowie

- die Anzeige von Metadaten für Kommunikationen oder Sprecher
- die Suche nach Transkriptionen und Annotationen in Transkripten
- die Ausgabe von Konkordanzen, also die Suche nach Suchbegriffen im Korpus, inkl. zugehöriger Metadaten.

Die Daten werden dabei in unterschiedlichen Formaten wiedergegeben, wie z.B. einem XML-Format, welches die Weiterverarbeitung bzw. Auswertung durch eigene Webservices oder Anwendungen erleichtert, oder dem HTML-Format, das die Anzeige der gelieferten Ergebnisse im Webbrowser erlaubt.

Für Hintergrundinformationen sowohl zum Projekt als auch zu den im Korpus enthaltenen Kommunikationen und zugehörigen Sprechern, sowie zu den Transkripten und Annotationen sei auf das Handbuch zum GeWiss-Korpus verwiesen (Gräfe et al. 2015). Nähere Informationen zu den Metadaten finden sich außerdem unter <https://gewiss.uni-leipzig.de/index.php?id=metadata>; zu den verschiedenen Typen von Annotationen siehe auch <https://gewiss.uni-leipzig.de/index.php?id=annotations>.

## II. Voraussetzungen für die Nutzung der Webservices

Für die Nutzung der Webservices ist zu beachten, dass einige von Ihnen auf sensible Korpusdaten zugreifen bzw. diese ausliefern und daher speziellen Zugriffsbeschränkungen unterliegen, d.h. nur passwortgeschützt zugänglich sind. Registrierte Nutzer des GeWiss-Portals können auf Wunsch Zugang zu den geschützten Webservices erhalten (eine kurze Nachricht an [gewiss-korpus@uni-leipzig.de](mailto:gewiss-korpus@uni-leipzig.de) genügt). Eine Ausnahme bildet hierbei der Webservice zum Abrufen ganzer Audiodateien (PlayRecording), welcher zusätzlichen datenschutzrechtlichen Beschränkungen unterliegt und daher nach persönlicher Vereinbarung zugänglich gemacht werden kann.

Die nicht passwortgeschützten Webservices sind frei verwendbar und können auch von nicht registrierten Nutzern eingesetzt werden. Die Zugänglichkeit der einzelnen Webservices ist in den folgenden Abschnitten ersichtlich, in denen die Webservices im Detail vorgestellt werden. Sie können bspw. mit wget oder cURL (Programme zum Herunterladen von Daten aus dem WWW über das HTTP- und FTP-Protokoll von der Kommandozeile aus, die für verschiedene Plattformen verfügbar sind) aufgerufen werden, oder z.B. auch über das Browser-Plugin RESTClient. Ein Abruf direkt über die Adressleiste des Webbrowsers ist für die GET-Methode ebenfalls möglich.

### III. Allgemeine Hinweise zur Nutzung

Die Ausgabeformate der jeweiligen Webservices sind im Text angegeben. Sofern nur ein Ausgabeformat für einen Webservice zur Verfügung steht, wird dieses durch den Webservice automatisch zurückgegeben.

Durch das Hinzufügen von Parametern (es handelt sich dabei in der Regel um die Auswahl von Metadaten) können Ergebnisse spezifiziert werden. Die Liste aller jeweils verfügbaren Parameter ist für jeden Webservice gesondert aufgeführt. Die meisten Parameter sind optional, während einige zwingend gesetzt werden müssen.

Die Parameter werden als Bestandteil der URL im Query String aufgeführt; das bedeutet, nach der jeweiligen URL des Webservice wird ein Fragezeichen „?“ gesetzt, auf das die Parameter in der Form „Parametername=Parameterwert“ folgen:

*Bsp.: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSubcorpora?acad-context=UK>*

Erlaubt ein Webservice unterschiedliche Parameter, so sind diese auch miteinander kombinierbar, wenn nicht anders angegeben.

*Bsp.: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSubcorpora?acad-context=UK&lang=DEU&lang-competence=L2&disc-genre=EV>*

Die einzelnen Parameter werden, wie aus obigem Beispiel ersichtlich ist, durch „&“ abgetrennt. Wenn nicht anders angegeben, kann jeder einzelne Parametername dabei nur einmal verwendet werden.

*nicht möglich: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSubcorpora?lang=DEU&lang=ENG>*

Bei einigen optionalen Parametern ist unter den zulässigen Parameterwerten ein Default-Wert angegeben. Wird der Parameter nicht manuell gesetzt, so wird automatisch dieser Default-Wert gesetzt.

*Bsp.: ListMetadata; Parameter codes ist optional, default ist „true“. Der Aufruf <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/ListMetadata> entspricht damit <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/ListMetadata?codes=true>*

Hinweis zu den passwortgeschützten Webservices: Beim Aufruf mittels wget sind Benutzername und Passwort wie folgt in den auszuführenden Befehl zu integrieren:

Bsp.: `-http-user="USERNAME" -http-password="PASSWORD" http://....`

Beim Aufruf des Webservice im Browser öffnet sich ein Popup-Fenster, in dem man zur Eingabe der Nutzendaten aufgefordert wird.

## IV. Liste der Webservices

### 1. Auflistung von Teilkorpora (frei zugänglich)

Beschreibung:

Liefert eine Liste aller Teilkorpora des GeWiss-Korpus in Form der ihnen zugewiesenen Kürzel. Die Teilkorpora lassen sich nach Metadaten wie Sprache, akademischer Kontext usw. differenzieren. Die Kürzel finden als Parameter in weiteren Webservices Verwendung. Für eine vollständige Liste aller Teilkorpora siehe Anhang.

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSubcorpora>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSubcorpora/html>

Methode: GET

Output: text/html

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
acad-context	Akademischer Kontext	ja	UK, PL, BG
lang	Sprache	ja	DEU, ENG, POL, ITA, BUL
lang-competence	Sprachkompetenz	ja	L1, L2
disc-genre	Diskursgenre	ja	EV, PG, SV

Beispielaufruf:

`http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSubcorpora?acad-context=UK&lang=DEU&lang-competence=L2&disc-genre=EV`

Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<result-list>
  <result>
    <data name="corpus-name">DEU_L2_UK_EV</data>
  </result>
</result-list>
```

## 2. Auflistung von Kommunikationen (frei zugänglich)

### Beschreibung:

Liefert eine Auflistung aller Kommunikationen eines Teilkorpus in Form der ihnen zugewiesenen Kürzel. Die Kürzel finden als Parameter in weiteren Webservices Verwendung.

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListCommunications>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListCommunications/html>

Methode: GET

Output: text/html

### Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
cor-name	Kürzel eines Teilkorpus	nein	Bsp.: DEU_L1_EV, DEU_L2_BG_SV vollständige Liste, siehe Anhang

### Beispielaufruf:

[http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListCommunications?cor-name=DEU\\_L1\\_EV](http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListCommunications?cor-name=DEU_L1_EV)

### Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<result-list>
  <result>
    <data name="communication-name">EV_DE_004</data>
  </result>
  <result>
    <data name="communication-name">EV_DE_005</data>
  </result>

  <!-- ... gekürzt ... -->

  <result>
    <data name="communication-name">EV_DE_105</data>
  </result>
  <result>
    <data name="communication-name">EV_DE_106</data>
  </result>
</result-list>
```

### 3. Auflistung von Sprechern (frei zugänglich)

Beschreibung:

Liefert eine Liste aller an einem Kommunikationsereignis beteiligten Sprecher in Form der ihnen zugewiesenen Kürzel. Die Kürzel finden als Parameter in weiteren Webservices Verwendung.

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSpeakers>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSpeakers/html>

Methode: GET

Output: text/html

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
com-name	Kürzel einer Kommunikation	nein	Bsp.: EV_DE_004, PG_DE_047

Beispielaufruf:

[http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSpeakers?com-name=EV\\_DE\\_004](http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ListSpeakers?com-name=EV_DE_004)

Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<result-list>
  <result>
    <data name="speaker-name">LV_0295</data>
  </result>
  <result>
    <data name="speaker-name">JS_0215</data>
  </result>

  <!-- ... gekürzt ... -->
</result-list>
```

## 4. Ausgabe von Metadaten (frei zugänglich)

Beschreibung:

Liefert die zu einem Teilkorpus, einer Kommunikation oder einem einzelnen Sprecher hinterlegten Metadaten.

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewMetadata>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewMetadata/html>

Methode: GET

Output: text/html

Parameter:

Name	Bedeutung	Optional*	Zulässige Parameterwerte
cor-name	Kürzel eines Teilkorpus	ja*	Bsp.: DEU_L1_EV, DEU_L2_BG_SV vollständige Liste siehe Anhang
com-name	Kürzel einer Kommunikation	ja*	Bsp.: EV_DE_004, PG_PL_090
spk-name	Kürzel eines Sprechers	ja*	Bsp.: MH_0226, RO_1505

\* (genau) ein Parameter ist obligatorisch zu setzen

Beispielaufruf:

[http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewMetadata?com-name=EV\\_IT\\_002](http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewMetadata?com-name=EV_IT_002)

Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<Corpus Name="GeWiss" Id="ID7C223B9A-3273-F969-8822-4E006B5D93B8-X6">
  <DBNode/>
  <CorpusData>
    <Communication Id="ID56B4CC45-81ED-AD08-4318-3F064E48BF6E-ITA_L1_EV"
      Name="EV_IT_002">
      <!-- ... gekürzt ... -->
    </Communication>
  </CorpusData>
</Corpus>
```

## 5. Ausgabe von Transkripten (passwortgeschützt)

Beschreibung:

Dieser Webservice dient dem Abruf der dem Korpus zugrunde liegenden Kommunikationsereignisse in transkribierter Form.

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewTranscript>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewTranscript/html>

Methode: GET  
Output: text/html

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewTranscript/pdf>  
Methode: GET  
Output: application/pdf

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
tra-name	Kürzel einer Kommunikation	nein	Bsp.: EV_DE_004, PG_PL_090

Beispielaufruf (Benutzerdaten werden angefordert):

`http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewTranscript?tra-name=EV_DE_004`

Beispielaufruf per wget:

`wget -http-user="USERNAME" -http-password="PASSWORD" http://gewiss.uni-leipzig.de:8282/gewiss/webresources/ViewTranscript?tra-name=EV_DE_004`

Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- (c) http://www.rrz.uni-hamburg.de/exmaralda -->
<basic-transcription Id="CIDID2AB43F30-3B69-F4E6-5CCF-20345564498C">
  <head>
    <meta-information>
      <project-name>GeWiss</project-name>
      <transcription-name>EV_DE_004</transcription-name>
      <referenced-file url="../../Audio/EV_DE_004.wav" />

      <!-- ... gekürzt ... -->

    </meta-information>
  </head>
  <basic-body>
    <common-timeline>
      <tli id="T0" time="0.0" />
      <tli id="T2" time="0.43333277368063916" />
      <tli id="T4" time="5.693326576545131" />

      <!-- ... gekürzt ... -->

    </common-timeline>
    <tier id="TIE0" speaker="SPK0" category="v" type="t" display-name="MOD
[v]">
      <event start="T0" end="T2">(0.4) </event>
      <event start="T2" end="T4">ich freue mich die nächste vortragende
vorstellen zu dürfen </event>

      <!-- ... gekürzt ... -->

    </tier>
  </basic-body>
</basic-transcription>
```



## 6. Wiedergabe von Audiodaten (passwortgeschützt)

Beschreibung:

Liefert die dem Korpus zugrunde liegenden Kommunikationseignisse in abspielbarem Audioformat.

Hinweis:

Aus datenschutzrechtlichen Gründen ist eine gesonderte Registrierung zur Nutzung dieses Webservice erforderlich, die nur im Einzelfall und nach Prüfung eines berechtigten Forschungsinteresses erteilt werden kann. Bitte setzen Sie sich mit uns in Verbindung.

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/PlayRecording/wav>

Methode: GET

Output: audio/wav

URL: <http://gewiss.uni-leipzig.de:8282/gewiss/webresources/PlayRecording/mp3>

Methode: GET

Output: audio/mpeg3

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
rec-name	Name einer Aufnahme	nein	Bsp.: EV_DE_004, PG_PL_090
seg-name	Name einer Transkriptionsevents	ja	Bsp.: T169
start	Startzeit des wiedergegebenen Audioausschnitts in Sek.	ja (Angabe nur zusammen mit end)	Bsp.: 15
end	Endzeit des wiedergegebenen Audioausschnitts in Sek.	ja (Angabe nur zusammen mit start)	Bsp.: 25

Hinweis:

Die Angabe von seg-name, start und end ist z.Z. nur für die Ausgabe von audio/wav möglich. Für audio/mp3 wird aktuell immer die gesamte Audioaufnahme ausgeliefert.

Beispielaufruf (per wget):

```
wget -http-user="USERNAME" -http-password="PASSWORD" -header="Accept: audio/wav"
http://gewiss.uni-leipzig.de:8282/gewiss/webresources/PlayRecording/wav?rec-
name=EV_DE_004&seg-name=T169
```

## 7. Auflistung aller Typen von Metadaten (frei zugänglich)

Beschreibung:

Liefert eine einfache Liste aller Typen von Metadaten und deren mögliche Werte. Optional ist die Angabe der Codes für die erweiterten Suchfunktionen der Konkordanzsuche (s.u.); mit „C“ = communication, „S“ = speaker. Nicht alle Metadaten verfügen über festgelegte codierte Werte (z.B. C4 – siehe Bsp.)

Für eine vollständige Liste der Metadattentypen siehe Anhang.

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/ListMetadata>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/ListMetadata/html>

Methode: GET

Output: text/html

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
codes	Codes der Metadattentypen & -werte	ja	true – Anzeige der Codes (default) false – keine Anzeige der Codes

Beispielaufwurf:

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/ListMetadata?codes=true>

Beispielergbnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<metadata-list>
  <meta type="C" code="C1" value="1" name="Teilkorpus">Deutschland</meta>
  <meta type="C" code="C1" value="2"
    name="Teilkorpus">Großbritannien</meta>
  <meta type="C" code="C1" value="3" name="Teilkorpus">Polen</meta>
  <meta type="C" code="C1" value="4" name="Teilkorpus">Bulgarien</meta>
  <meta type="C" code="C1" value="5" name="Teilkorpus">Italien</meta>
  <meta type="C" code="C2" value="1" name="Muttersprachliche
    Kommunikation">ja</meta>
  <meta type="C" code="C2" value="2" name="Muttersprachliche
    Kommunikation">nein</meta>

  <!-- ... gekürzt ... -->

  <meta type="C" code="C4" name="Art" />
  <meta type="C" code="C5" name="Kurzbezeichnung" />

  <!-- ... gekürzt ... -->
</metadata-list>
```

## 8. Auflistung aller Typen von Annotationen (frei zugänglich)

Beschreibung:

Liefert eine einfache Liste aller Typen von Annotationen (inkl. Subtypen für D1, D2, Verweis und Zitat) und deren Codes für die Webservices der Annotationssuche (s.u.). Zu den Annotationen siehe auch <https://gewiss.uni-leipzig.de/index.php?id=annotations>.

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/ListAnnotations>

Methode: GET

Output: text/xml

Parameter: keine

Aufruf:

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/ListAnnotations>

Ergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<annotation-list>
  <anno code="0" type="Wechsel" />
  <anno code="1" type="D1">Daten</anno>
  <anno code="1" type="D1">Ende</anno>
  <anno code="1" type="D1">Fazit</anno>
  <anno code="1" type="D1">Makrostruktur</anno>
  <anno code="1" type="D1">Performanz</anno>
  <anno code="1" type="D1">Rueckbezug</anno>
  <anno code="1" type="D1">Sprechhandlung-A</anno>
  <anno code="1" type="D1">Sprechhandlung-F</anno>
  <anno code="1" type="D1">Thema</anno>
  <anno code="1" type="D1">Zeit</anno>
  <anno code="2" type="D2">Anfang</anno>
  <anno code="2" type="D2">Diskussion</anno>
  <anno code="2" type="D2">Rederecht</anno>
  <anno code="2" type="D2">Vorstellung</anno>
  <anno code="3" type="D3" />
  <anno code="4" type="Situation" />
  <anno code="5" type="Verweis">Konzept</anno>
  <anno code="5" type="Verweis">Studie</anno>
  <anno code="5" type="Verweis">Publikation</anno>
  <anno code="5" type="Verweis">unspezifisch</anno>
  <anno code="6" type="Zitat">sinn</anno>
  <anno code="6" type="Zitat">woertl</anno>
</annotation-list>
```

## 9. Einfache Konkordanzsuche - Text (frei zugänglich)

Beschreibung:

Liefert eine Liste der Kommunikationen, in denen der Suchtext enthalten ist und die Gesamtzahl der Suchergebnisse.

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/SimpleSearch/Text>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/SimpleSearch/Text/html>

Methode: GET

Output: text/html

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
search	Suchtext	nein	kleinbuchstaben,_,- sowie * als Platzhalter für beliebige Zeichenfolgen innerhalb eines Suchbegriffs: test test* -> teste, tester they_re
acad-context	Akademischer Kontext	ja	D, UK, PL, BG
lang	Sprache	ja	DEU, ENG, POL, ITA, BUL
lang-competence	Sprachkompetenz	ja	L1, L2
disc-genre	Diskursgenre	ja	EV, PG, SV

Beispielaufruf:

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/SimpleSearch/Text?search=inhaltlich&lang-competence=L1>

Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<result-list total-occurrences="7">
  <result>
    <data name="communication-name">PG_DE_132</data>
  </result>
  <result>
    <data name="communication-name">PG_DE_136</data>
  </result>

  <!-- ... gekürzt ... -->
</result-list>
```

## 10. Einfache Konkordanzsuche - Annotationen (frei zugänglich)

Beschreibung:

Liefert eine Liste der Kommunikationen, in denen eine bestimmte Annotation enthalten ist (ggf. Subtyp - siehe Abschnitt 8). Filtern nach Teilkorpora möglich.

Die Suche kann mit der Textsuche kombiniert werden, indem zusätzlich der Parameter *search* gesetzt wird. Damit werden nur Annotationen ausgegeben, die auch den Suchbegriff enthalten.

URL:

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/SimpleSearch/Anno>

Methode: GET

Output: text/xml

URL:

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/SimpleSearch/Anno/html>

Methode: GET

Output: text/html

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
anno-type	Annotationstyp	ja	0(default), 1,2,3,4,5,6 – siehe Liste aller Annotationstypen im Anhang
search	Suchtext	ja	kleinbuchstaben,_,- sowie * als Platzhalter für beliebige Zeichenfolgen innerhalb eines Suchbegriffs: test test* -> teste, tester they_re
subtype	Subtyp	ja	nur bei Annotationstypen mit definierten Subtypen; z.B. Subtyp „Daten“ für Typ D1 – siehe Liste aller Annotationstypen im Anhang
acad-context	Akademischer Kontext	ja	D, UK, PL, BG
lang	Sprache	ja	DEU*
lang-competence	Sprachkompetenz	ja	L1, L2
disc-genre	Diskursgenre	ja	EV, PG, SV

\*Annotationen liegen nur für die deutschsprachigen Teilkorpora vor. In allen deutschsprachigen Daten wurden Code-switches annotiert (anno-type 0). Ausschließlich im Teilkorpus DEU\_L1\_EV wurden die Diskurskommentierungen (anno-type 1-4) mit ihren jeweiligen Subtypen annotiert (vgl. u.a. Baur et al. 2014). In den Teilkorpora DEU\_L1\_EV und DEU\_L1\_SV wurden Verweise und Zitate annotiert (vgl. Maier et al. 2015).

Beispielaufwurf:

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/SimpleSearch/Anno?anno-type=2&subtype=Diskussion&search=also&lang-competence=L1>

### Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<result-list total-occurrences="7">
  <result>
    <data name="communication-name">EV_DE_093</data>
  </result>
  <result>
    <data name="communication-name">PG_UK_063</data>
  </result>

  <!-- ... gekürzt ... -->
</result-list>
```

## 11. Erweiterte Konkordanzsuche – Text (passwortgeschützt)

Beschreibung:

Liefert alle Vorkommen des Suchtexts im (angegebenen Teil-)Korpus und zusätzliche Informationen zum Kommunikationsereignis und den Sprechern. Erweiterte Optionen bieten sich im Hinblick auf Kontextangaben und -eigenschaften, die Anzeige von Metadaten, Sortier- und Filteroptionen.

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Text>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Text/html>

Methode: GET

Output: text/html

Parameter:

Name	Bedeutung	Optional	Zulässige Parameterwerte
search	Suchtext	nein	kleinbuchstaben,_,- sowie * als Platzhalter für beliebige Zeichenfolgen innerhalb eines Suchbegriffs: test test* -> teste, tester they_re
acad-context	Akademischer Kontext	ja	D, UK, PL, BG
lang	Sprache	ja	DEU, ENG, POL, ITA, BUL
lang-competence	Sprachkompetenz	ja	L1, L2
disc-genre	Diskursgenre	ja	EV, PG, SV
first	Erster angezeigter Treffer	ja	1-10000, Default: 1
last	Letzter angezeigter Treffer	ja	1-10000, Default: 10000
context	Angezeigter Kontext, Anzahl Buchstaben	ja	ganze Zahlen, Default: 100
show	Angezeigte Metadaten	ja*	Codes für Metadatentypen (vollst. Liste siehe Anhang): {C1-C17,S1-S8}
fltr	Filtert nach Werten für einzelne Metadatentypen	ja*	Code für Metadatentyp+ IN   EX + Wert: C1IN1,C1EX1,C1IN2,...;
sort	Sortiert die Suchergebnisse	ja*	Ca,Cd,Sa,Sd; C1a,C1d,C2a,C2d...Codes für Metadatentyp+ "a"   "d"

\*diese Parameter können jeweils mehrfach verwendet werden, um unterschiedliche Werte zu kombinieren, z.B.

„PARAMETER1=PARAMETERWERT1 & PARAMETER1=PARAMETERWERT2 & PARAMETER1=PARAMETERWERT3...“

## Suche

Suche nach dem Suchtext "kurz" mittels Parameter **search** in den Teilkorpora deutscher Sprache (**lang**=DEU) und Diskursgenre Prüfungsgespräch (**disc-genre**=PG)

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Text?search=kurz&lang=DEU&disc-genre=PG>

## Eingrenzen der Ergebniszahl

Um nur einen Teil der Suchergebnisse auszugeben, verwende man die Parameter **first** und **last** als ersten bzw. letzten auszugebenden Treffer (z.B. first=100, last=199 für die Treffer 100 bis einschl. 199).

## Kontext

Mit **context** lässt sich der Kontext, in dem der jeweilige Suchtext gefunden wurde, für jeden Beleg ausgeben. Der Parameterwert entspricht dabei der Anzahl an ihn umgebenden Buchstaben in der Transkription (z.B. context=80 für 80 Buchstaben links und rechts des gefundenen Suchtexts).

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Text?search=kurz&first=100&last=199&context=80>

## Anzeigen von Metadaten

Mit dem Parameter **show** lassen sich zusätzliche Metadaten der Kommunikation und des Sprechers für jeden Treffer ausgeben. Zum Anzeigen der Metadattentypen „Teilkorpus“ (C1) und „Geschlecht“ des/der Sprechers/in (S2) ergänzt man: show=C1&show=S2 – für Codes und Werte wiederum siehe Anhang.

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Text?search=kurz&show=C1&show=S2>

## Filtern

Der Parameter **fltr** erlaubt das Filtern der Suchergebnisse nach Metadaten und ihren Werten. Die Parameterwerte haben immer die Form *Metadattentyp+IN/EX+Metadatenwert*. Der Filterbefehl IN liefert dabei alle Treffer, die den angegebenen Wert für den angegebenen Metadattentyp beinhalten; der Filter EX liefert alle Treffer, die **nicht** dem angegebenen Wert entsprechen.

Um also bspw. nur Ergebnisse zu erhalten, für die gilt: „muttersprachliche Kommunikation“=„ja“ und „Teilkorpus“≠„Polen“, setzt man fltr=C2IN1&fltr=C1EX3.

- fltr=C2IN1 liefert alle Treffer, für die gilt: der Metadattentyp C2 („muttersprachliche Kommunikation“) hat Wert 1 („ja“) – für Codes und Werte siehe Anhang
- fltr=C1EX3 liefert alle Treffer, für die gilt: der Metadattentyp C1 („Teilkorpus“) hat nicht den Wert 3 („Polen“) – für Codes und Werte siehe Anhang

**Hinweis:** C1EX3 liefert nur Treffer, bei denen ein Wert für C1 angegeben ist (aber eben nicht „Polen“). Treffer, für die gar kein Wert angegeben ist, werden weder durch C1EX3, noch



durch C1IN3 erfasst. Anders gesagt: Gesamtzahl Treffer (ohne Filter) = Treffer C1IN3 + Treffer C1EX3 + Treffer (ohne Wert für C1)

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Text?search=kurz&show=C1&show=C2&fltr=C2IN1&fltr=C1EX3>

Anmerkungen:

Es ist möglich, wie in dem angegebenen Beispiel, die Metadatentypen anzuzeigen, nach denen man filtert (C1 und C2), aber nicht notwendig.

Bei Metadatentypen ohne festgelegte Werte (siehe Anhang) kann ein Wert frei eingegeben werden. Z.B. alle Treffer mit „Auslandsstudium“(S6)=„Polen“: fltr=S6INPolen.

### Sortieren

Zur Sortierung der Ergebnisse dient der Parameter **sort**. Die Parameterwerte haben dabei die Form *Sortiergröße+Reihenfolge*; also bspw. Ca, Sa, C1d, S3a usw.:

- Sortiergrößen sind
  - C für alphabetische Sortierung nach Kürzel des Kommunikationsereignis (EV\_DE\_004, PG\_DE\_047 etc.)
  - S für alphabetische Sortierung nach Kürzel des Sprechers (MH\_0226, RO\_1505 etc.)
  - C1,C2,..S1... für alphabetische Sortierung nach dem Wert des Treffers für den jeweiligen Metadatentyp (z.B. C2 sortiert alphabetisch nach den Werten „gemischt“/„ja“/„nein“ für den Typ „muttersprachliche Kommunikation“ – siehe Anhang für Codes und Werte)
- Reihenfolgen sind
  - a für aufsteigend (ascending)
  - d für absteigend (descending)

Damit sortiert bspw. sort=Ca aufsteigend nach Kürzel der Kommunikation, sort=Sa aufsteigend nach Kürzel des Sprechers, sort=C1d absteigend nach „Teilkorpus“ (C1) usw. Mehrere Sortierungen können dabei kombiniert werden und werden von links nach rechts der Reihenfolge nach abgearbeitet. Bspw. sortiert sort=Ca&sort=Sa&sort=C1d die Ergebnisse erst aufsteigend nach dem Kürzel der Kommunikation; Ergebnisse mit dem jeweils gleichen Kürzel der Kommunikation werden nun aufsteigend nach dem Kürzel des Sprechers sortiert; Ergebnisse mit dem gleichen Kürzel der Kommunikation sowie dem gleichen Kürzel des Sprechers werden nun absteigend nach „Teilkorpus“ sortiert.

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Text?search=kurz&sort=Ca&sort=Sa&sort=C1d>

### Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<search-result-list total-occurrences="10">
<search-result selected="true" communication="EV_DE_002" speaker="SY_0644">
<locator file="" xpath=""/>
<left-context>depeschen (.) also aus korrespondenzbriefen (.) wo (.) sehr </left-
context>
<match original-match-start="">kurz</match>
<right-context> (.) und sehr knapp (.) über eine schlacht (.) oder ein (.) </right-
context>
<data name="tier">TIE2</data>
</search-result> <meta type="C1" name="Teilkorpus">Deutschland</meta>
<meta type="S2" name="Geschlecht">weiblich</meta>
</search-result>
<meta type="S6" name="Auslandsstudium">USA</meta>

<!-- ... gekürzt ... -->

</search-result-list>
```

## 12. Erweiterte Konkordanzsuche – Annotationen (passwortgeschützt)

### Beschreibung:

Liefert alle Vorkommen einer Annotation im angegebenen (Teil-) Korpus und zusätzliche Informationen zum Kommunikationsereignis und Sprecher. Erweiterte Optionen bieten sich im Hinblick auf Kontextangaben und -eigenschaften, die Anzeige von Metadaten, Sortier- und Filteroptionen.

Die Suche kann mit der Textsuche kombiniert werden, indem zusätzlich der Parameter *search* gesetzt wird. Damit werden nur Annotationen ausgegeben, die auch den Suchbegriff enthalten.

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Anno>

Methode: GET

Output: text/xml

URL: <http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Anno/html>

Methode: GET

Output: text/html

Name	Bedeutung	Optional	Zulässige Parameterwerte
anno-type	Annotationstyp	ja	0(Default), 1, 2, 3, 4, 5, 6 – siehe Liste aller Annotationstypen im Anhang
search	Suchtext	ja	kleinbuchstaben, _ , - sowie * als Platzhalter für beliebige Zeichenfolgen innerhalb eines Suchbegriffs: test test* -> teste, tester they_re

subtype	Subtyp	ja	nur bei Annotationstypen mit definierten Subtypen; z.B. Subtyp „Daten“ für Typ D1 – siehe Liste aller Annotationstypen im Anhang
acad-context	Akademischer Kontext	ja	D, UK, PL, BG
lang	Sprache	ja	DEU*
lang-competence	Sprachkompetenz	ja	L1, L2
disc-genre	Diskursgenre	ja	EV, PG, SV
first	Erster angezeigter Treffer	ja	1-10000, Default:1
last	Letzter angezeigter Treffer	ja	1-10000, Default:10000
context	Angezeigter Kontext, Anzahl Buchstaben	ja	ganze Zahlen, Default: 100
show	Angezeigte Metadaten	ja**	Codes für Metadatentypen (vollst. Liste siehe Anhang): {C1-C17,S1-S8}
fltr	Filtert nach Werten für einzelne Metadatentypen	ja**	Code für Metadatentyp+ IN   EX + Wert: C1IN1,C1EX1,C1IN2,...;
sort	Sortiert die Suchergebnisse	ja**	Ca,Cd,Sa,Sd; C1a,C1d,C2a,C2d...Codes für Metadatentyp+ "a"   "d"

\*Annotationen liegen nur für die deutschsprachigen Teilkorpora vor. In allen deutschsprachigen Daten wurden Code-switches annotiert (anno-type 0). Ausschließlich im Teilkorpus DEU\_L1\_EV wurden die Diskurskommentierungen (anno-type 1-4) mit ihren jeweiligen Subtypen annotiert (vgl. u.a. Baur et al. 2014). In den Teilkorpora DEU\_L1\_EV und DEU\_L1\_SV wurden Verweise und Zitate annotiert (vgl. Maier et al. 2015).

\*\*diese Parameter können jeweils mehrfach verwendet werden, um unterschiedliche Werte zu kombinieren

„PARAMETER1=PARAMETERWERT1 & PARAMETER1=PARAMETERWERT2 & PARAMETER1=PARAMETERWERT3...“

### Suche

Suche nach Annotationstyp D2-Diskussion (**annotype=2**, **subtype=Diskussion**; siehe Anhang für Codes) in den Teilkorpora deutscher Sprache (**lang=DEU**) und Diskursgenre Prüfungsgespräch (**disc-genre=PG**):

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Anno?anno-type=2&subtype=Diskussion&search=also&lang=DEU&disc-genre=PG>

### Eingrenzen der Ergebniszahl

Um nur einen Teil der Suchergebnisse auszugeben, verwendet man die Parameter **first** und **last** als ersten bzw. letzten auszugebenden Treffer (z.B. first=100, last=199 für die Treffer 100 bis einschl. 199).

### Kontext

Mit **context** lässt sich der Kontext, in dem die jeweilige Annotation gefunden wurde, für jeden Treffer ausgeben. Der Wert entspricht dabei der Anzahl an Buchstaben in der Transkription (z.B. context=80 für 80 Buchstaben links und rechts der gefundenen Suchtexts).

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Anno?anno-type=0&first=80&last=199&context=80>

### Anzeige von Metadaten

Mit dem Parameter **show** lassen sich zusätzliche Metadaten der Kommunikation und des Sprechers für jeden Treffer ausgeben. Zum Anzeigen der Metadattentypen „Teilkorpus“ (C1) und „Geschlecht“ des/der Sprechers/in (S2) setze `show=C1&show=S2` – für Codes und Werte siehe Anhang.

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Anno?anno-type=0&show=C1&show=S2>

### Filtern

Der Parameter **fltr** erlaubt das Filtern der Suchergebnisse nach Metadaten und ihren Werten. Die Parameterwerte haben immer die Form *Metadattentyp+IN/EX+Metadatenwert*. Der Filter IN liefert dabei alle Treffer, die den angegebenen Wert für den angegebenen Metadattentyp haben; der Filter EX liefert alle Treffer, die **nicht** den angegebenen Wert haben.

Um also bspw. nur Ergebnisse zu erhalten, für die gilt: „muttersprachliche Kommunikation“=„ja“ und „Teilkorpus“≠„Polen“, setzt man `fltr=C2IN1&fltr=C1EX3`.

- `fltr=C2IN1` liefert alle Treffer für die gilt: der Metadattentyp C2 („muttersprachliche Kommunikation“) hat Wert 1 („ja“) – für Codes und Werte siehe Anhang
- `fltr=C1EX3` liefert alle Treffer für die gilt: der Metadattentyp C1 („Teilkorpus“) hat nicht den Wert 3 („Polen“) – für Codes und Werte siehe Anhang

**Hinweis:** C1EX3 liefert nur Treffer, bei denen ein Wert für C1 angegeben ist (aber eben nicht „Polen“). Treffer, für die gar kein Wert angegeben ist, werden weder durch C1EX3, noch durch C1IN3 erfasst.

Anders gesagt: Gesamtzahl Treffer (ohne Filter) = Treffer C1IN3 + Treffer C1EX3 + Treffer (ohne Wert für C1)

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Anno?anno-type=0&show=C1&show=C2&fltr=C2IN1&fltr=C1EX3>

### Anmerkungen:

- Es ist möglich, wie in dem angegebenen Beispiel, die Metadattentypen anzuzeigen, nach denen man filtert (C1 und C2), aber nicht notwendig.
- Bei Metadattentypen ohne festgelegte Werte (siehe Anhang) kann ein Wert frei eingegeben werden. Z.B. alle Treffer mit „Auslandsstudium“(S6)=„Polen“:  
`fltr=S6INPolen`

### Sortieren

Zur Sortierung der Ergebnisse dient der Parameter **sort**. Die Parameterwerte haben dabei die Form *Sortiergröße+Reihenfolge*; also bspw. Ca, Sa, C1d, S3a usw.:

- Sortiergrößen sind
  - C für alphabetische Sortierung nach Kürzel des Kommunikationsereignis (EV\_DE\_004, PG\_DE\_047 etc.)

- S für alphabetische Sortierung nach Kürzel des Sprechers (MH\_0226, RO\_1505 etc.)
- C1,C2,..S1... für alphabetische Sortierung nach dem Wert des Treffers für den jeweiligen Metadatentyp (z.B. C2 sortiert alphabetisch nach den Werten „gemischt“/„ja“/„nein“ für den Typ „muttersprachliche Kommunikation“ – siehe Anhang für Codes und Werte)
- Reihenfolgen sind
  - a für aufsteigend (ascending)
  - d für absteigend (descending)

Damit sortiert bspw. sort=Ca aufsteigend nach Kürzel der Kommunikation, sort=Sa aufsteigend nach Kürzel des Sprechers, sort=C1d absteigend nach „Teilkorpus“ (C1) usw. Mehrere Sortierungen können dabei kombiniert werden und werden von links nach rechts der Reihenfolge nach abgearbeitet. Bspw. sortiert sort=Ca&sort=Sa&sort=C1d die Ergebnisse erst aufsteigend nach dem Kürzel der Kommunikation; Ergebnisse mit dem jeweils gleichen Kürzel der Kommunikation werden nun aufsteigend nach dem Kürzel des Sprechers sortiert; Ergebnisse mit dem gleichen Kürzel der Kommunikation sowie dem gleichen Kürzel des Sprechers werden nun absteigend nach „Teilkorpus“ sortiert.

<http://gewiss.uni-leipzig.de:8282/CorpusQuery/webresources/Search/Anno?anno-type=0&sort=Ca&sort=Sa&sort=C1d>

#### Beispielergebnis:

```
<?xml version="1.0" encoding="UTF-8"?>
<search-result-list total-occurrences="10">
<search-result selected="true" communication="EV_DE_002" speaker="SY_0644">
<locator file="" xpath=""/>
<left-context>depeschen (.) also aus korrespondenzbriefen (.) wo (.) sehr <match
original-match-start="">kurz</match>
<right-context> (.) und sehr knapp (.) über eine schlacht (.) oder ein (.) </right-
context>
<data name="tier">TIE2</data>
</search-result> <meta type="C1" name="Teilkorpus">Deutschland</meta>
<meta type="S2" name="Geschlecht">weiblich</meta>
</search-result>
<meta type="S6" name="Auslandsstudium">USA</meta>

<!-- ... gekürzt ... -->
</search-result-list>
```

#### Hinweis:

Aus technischen Gründen wird u.U. nicht die vollständige annotierte Passage im Transkript als Treffer ausgegeben, sondern nur das erste Segment. Die Annotation erstreckt sich damit ggf. auch in den rechten Kontext hinein.

## Weiterführende Literatur

Baur, Benedikt/Gräfe, Karen/Lange, Daisy/Schmidt, Julia (2014): *Dokumentation zur Annotation der Diskurskommentierungen im GeWiss-Projekt*, abrufbar unter: [https://gewiss.uni-leipzig.de/index.php?id=annotations\\_discourse](https://gewiss.uni-leipzig.de/index.php?id=annotations_discourse)

Fandrych, Christian/Meißner, Cordula/Slavcheva, Adriana (2012): "The GeWiss Corpus: Comparing Spoken Academic German, English and Polish", in: Schmidt, Thomas/Wörner, Kai (Hg.): *Multilingual corpora and multilingual corpus analysis*. Amsterdam: Benjamins. (= Hamburg Studies in Multilingualism).

Gräfe, Karen/Lange, Daisy/Sieradz, Magda/Meißner, Cordula/Slavcheva, Adriana/Stoppel, David (2015): *Handbuch zum Korpus*, abrufbar unter: <https://gewiss.uni-leipzig.de/index.php?id=help>

Lange, Daisy/Slavcheva, Adriana/Rogozińska, Marta/Morton, Ralph (2014): "GAT 2 als Transkriptionssystem für multilinguale Sprachdaten? Zur Adaption der Notationskonventionen im Rahmen des Projekts GeWiss", in: Fandrych, Christian/Meißner, Cordula/Slavcheva, Adriana (Hg.): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron-Verlag. (= Wissenschaftskommunikation), 39-55.

Maier, Elisabeth/Sadowski, Sabrina/Schmidt, Julia (2015): *Dokumentation zur Annotation der Verweise und Zitate*, abrufbar unter: [https://gewiss.uni-leipzig.de/index.php?id=annotations\\_discourse](https://gewiss.uni-leipzig.de/index.php?id=annotations_discourse)

Meißner, Cordula/Jettka, Daniel/Fandrych, Christian (2013): CLARIN-KP-GeWiss: Das zweite Kurationsprojekt der F-AG 1 Deutsche Philologie, in: *CLARIN-D-Newsletter* 4: 3-8, abrufbar unter: <http://de.clarin.eu/images/newsletter/CLARIN-D-Newsletter-2013-4.pdf>

Selting, Margret/Auer, Peter/Barth-Weingarten, Dagmar/Bergmann, Jörg/Bergmann, Pia/Birkner, Karin/Couper-Kuhlen, Elizabeth/Deppermann, Arnulf/Gilles, Peter/Günthner, Susanne/Hartung, Martin/Kern, Friederike/Mertzluft, Christine/Meyer, Christian/Morek, Miriam/Oberzaucher, Frank/Peters, Jörg/Quasthoff, Uta/Schütte, Wilfried/Stuckenbrock, Anja/Uhmann, Susanne (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2), in: *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion [Online]*, 10, 353-402, abrufbar unter: <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf>

# Anhang

## Liste aller Subkorpora

DEU\_L1\_EV  
DEU\_L1\_SV  
DEU\_L2\_PL\_PG  
DEU\_L2\_UK\_SV  
ENG\_L1\_SV  
ENG\_L2\_SV  
POL\_PG  
DEU\_L1\_PG  
DEU\_L2\_D\_PG  
DEU\_L2\_PL\_SV  
DEU\_L2\_UK\_EV  
ENG\_L1\_EV  
ENG\_L2\_EV  
DEU\_L2\_D\_SV  
DEU\_L2\_PL\_EV  
DEU\_L2\_UK\_PG  
ENG\_L1\_PG  
ENG\_L2\_PG  
POL\_EV  
POL\_SV  
ITA\_L1\_EV  
DEU\_L2\_BG\_SV

## Liste aller Annotationstypen mit Codes

```
<anno code="0" type="Wechsel" />  
<anno code="1" type="D1">Daten</anno>  
<anno code="1" type="D1">Ende</anno>  
<anno code="1" type="D1">Fazit</anno>  
<anno code="1" type="D1">Makrostruktur</anno>  
<anno code="1" type="D1">Performanz</anno>  
<anno code="1" type="D1">Rueckbezug</anno>  
<anno code="1" type="D1">Sprechhandlung-A</anno>  
<anno code="1" type="D1">Sprechhandlung-F</anno>  
<anno code="1" type="D1">Thema</anno>  
<anno code="1" type="D1">Zeit</anno>  
<anno code="2" type="D2">Anfang</anno>  
<anno code="2" type="D2">Diskussion</anno>  
<anno code="2" type="D2">Rederecht</anno>  
<anno code="2" type="D2">Vorstellung</anno>  
<anno code="3" type="D3" />  
<anno code="4" type="Situation" />  
<anno code="5" type="Verweis">Konzept</anno>  
<anno code="5" type="Verweis">Studie</anno>  
<anno code="5" type="Verweis">Publikation</anno>  
<anno code="5" type="Verweis">unspezifisch</anno>  
<anno code="6" type="Zitat">sinn</anno>  
<anno code="6" type="Zitat">woertl</anno>
```

## Liste aller Metadatentypen mit Codes

```
<meta type="C" code="C1" value="1" name="Teilkorpus">Deutschland</meta>  
<meta type="C" code="C1" value="2" name="Teilkorpus">Großbritannien</meta>  
<meta type="C" code="C1" value="3" name="Teilkorpus">Polen</meta>  
<meta type="C" code="C1" value="4" name="Teilkorpus">Bulgarien</meta>  
<meta type="C" code="C1" value="5" name="Teilkorpus">Italien</meta>  
<meta type="C" code="C2" value="1" name="Muttersprachliche Kommunikation">ja</meta>  
<meta type="C" code="C2" value="2" name="Muttersprachliche Kommunikation">nein</meta>  
<meta type="C" code="C2" value="3" name="Muttersprachliche Kommunikation">gemischt</meta>  
<meta type="C" code="C3" value="1" name="Genre">Expertenvortrag</meta>  
<meta type="C" code="C3" value="2" name="Genre">Studentischer Vortrag</meta>  
<meta type="C" code="C3" value="3" name="Genre">Prüfungsgespräch</meta>  
<meta type="C" code="C4" name="Art" />  
<meta type="C" code="C5" name="Kurzbezeichnung" />  
<meta type="C" code="C6" name="Zusammenfassung" />  
<meta type="C" code="C7" name="Jahr" />
```

<meta type="C" code="C8" name="Land"/>  
 <meta type="C" code="C9" name="Institution"/>  
 <meta type="C" code="C10" name="Dauer"/>  
 <meta type="C" code="C11" value="1" name="Raum">Computerpool</meta>  
 <meta type="C" code="C11" value="2" name="Raum">Seminarraum</meta>  
 <meta type="C" code="C11" value="3" name="Raum">Hörsaal</meta>  
 <meta type="C" code="C11" value="4" name="Raum">Konferenzraum</meta>  
 <meta type="C" code="C11" value="5" name="Raum">Büro</meta>  
 <meta type="C" code="C12" value="1" name="Basissprache">deu</meta>  
 <meta type="C" code="C12" value="2" name="Basissprache">eng</meta>  
 <meta type="C" code="C12" value="3" name="Basissprache">pol</meta>  
 <meta type="C" code="C12" value="4" name="Basissprache">ita</meta>  
 <meta type="C" code="C13" value="1" name="Grad der Mündlichkeit">frei gesprochen</meta>  
 <meta type="C" code="C13" value="2" name="Grad der Mündlichkeit">zum Teil abgelesen</meta>  
 <meta type="C" code="C13" value="3" name="Grad der Mündlichkeit">vollständig abgelesen</meta>  
 <meta type="C" code="C13" value="4" name="Grad der Mündlichkeit">scheint vorformuliert und  
 auswendig gelernt</meta>  
 <meta type="C" code="C14" value="1" name="Wechsel in andere Sprache(n)">eng</meta>  
 <meta type="C" code="C14" value="2" name="Wechsel in andere Sprache(n)">pol</meta>  
 <meta type="C" code="C14" value="3" name="Wechsel in andere Sprache(n)">deu</meta>  
 <meta type="C" code="C14" value="4" name="Wechsel in andere Sprache(n)">bul</meta>  
 <meta type="C" code="C15" value="1" name="Verwendete Medien">Präsentation</meta>  
 <meta type="C" code="C15" value="2" name="Verwendete Medien">Handout</meta>  
 <meta type="C" code="C15" value="3" name="Verwendete Medien">Skript</meta>  
 <meta type="C" code="C15" value="4" name="Verwendete Medien">Thesenpapier</meta>  
 <meta type="C" code="C15" value="5" name="Verwendete Medien">Prüfungsfragen</meta>  
 <meta type="C" code="C15" value="6" name="Verwendete Medien">OHP-Folien</meta>  
 <meta type="C" code="C15" value="7" name="Verwendete Medien">Audiobeispiel</meta>  
 <meta type="C" code="C15" value="8" name="Verwendete Medien">Videobeispiel</meta>  
 <meta type="C" code="C15" value="9" name="Verwendete Medien">Analysematerial</meta>  
 <meta type="C" code="C15" value="10" name="Verwendete Medien">Mindmaps</meta>  
 <meta type="C" code="C15" value="11" name="Verwendete Medien">Essay</meta>  
 <meta type="C" code="C15" value="12" name="Verwendete Medien">Internet</meta>  
 <meta type="C" code="C15" value="13" name="Verwendete Medien">Plakat</meta>  
 <meta type="C" code="C15" value="14" name="Verwendete Medien">Tafel</meta>  
 <meta type="C" code="C15" value="15" name="Verwendete Medien">keine</meta>  
 <meta type="C" code="C16" name="Anzahl der Teilnehmer"/>  
 <meta type="C" code="C17" name="Beziehung der Sprecher zueinander"/>  
 <meta type="S" code="S1" name="Alter"/>  
 <meta type="S" code="S2" value="1" name="Geschlecht">weiblich</meta>  
 <meta type="S" code="S2" value="2" name="Geschlecht">männlich</meta>  
 <meta type="S" code="S3" value="1" name="Rollen">Vortragender</meta>  
 <meta type="S" code="S3" value="2" name="Rollen">Prüfling</meta>  
 <meta type="S" code="S3" value="3" name="Rollen">Prüfer</meta>  
 <meta type="S" code="S3" value="4" name="Rollen">Seminarleiter</meta>  
 <meta type="S" code="S4" value="1" name="Erstsprache">ara</meta>  
 <meta type="S" code="S4" value="2" name="Erstsprache">ces</meta>  
 <meta type="S" code="S4" value="3" name="Erstsprache">chi</meta>  
 <meta type="S" code="S4" value="4" name="Erstsprache">cmn</meta>  
 <meta type="S" code="S4" value="5" name="Erstsprache">dan</meta>  
 <meta type="S" code="S4" value="6" name="Erstsprache">deu</meta>  
 <meta type="S" code="S4" value="7" name="Erstsprache">eng</meta>  
 <meta type="S" code="S4" value="8" name="Erstsprache">fon</meta>  
 <meta type="S" code="S4" value="9" name="Erstsprache">fra</meta>  
 <meta type="S" code="S4" value="10" name="Erstsprache">idd</meta>  
 <meta type="S" code="S4" value="11" name="Erstsprache">jpn</meta>  
 <meta type="S" code="S4" value="12" name="Erstsprache">k.A.</meta>  
 <meta type="S" code="S4" value="13" name="Erstsprache">ndl</meta>  
 <meta type="S" code="S4" value="14" name="Erstsprache">pol</meta>  
 <meta type="S" code="S4" value="15" name="Erstsprache">por</meta>  
 <meta type="S" code="S4" value="16" name="Erstsprache">ron</meta>  
 <meta type="S" code="S4" value="17" name="Erstsprache">rus</meta>  
 <meta type="S" code="S4" value="18" name="Erstsprache">slv</meta>  
 <meta type="S" code="S4" value="19" name="Erstsprache">spa</meta>  
 <meta type="S" code="S4" value="20" name="Erstsprache">swe</meta>  
 <meta type="S" code="S4" value="21" name="Erstsprache">tgl</meta>  
 <meta type="S" code="S4" value="22" name="Erstsprache">tur</meta>  
 <meta type="S" code="S4" value="23" name="Erstsprache">ukr</meta>  
 <meta type="S" code="S4" value="24" name="Erstsprache">zho</meta>  
 <meta type="S" code="S5" value="1" name="L2">afr</meta>  
 <meta type="S" code="S5" value="2" name="L2">amh</meta>  
 <meta type="S" code="S5" value="3" name="L2">ara</meta>  
 <meta type="S" code="S5" value="4" name="L2">arb</meta>  
 <meta type="S" code="S5" value="5" name="L2">ben</meta>  
 <meta type="S" code="S5" value="6" name="L2">ces</meta>  
 <meta type="S" code="S5" value="7" name="L2">cmn</meta>  
 <meta type="S" code="S5" value="8" name="L2">dan</meta>  
 <meta type="S" code="S5" value="9" name="L2">deu</meta>



```
<meta type="S" code="S5" value="10" name="L2">eng</meta>
<meta type="S" code="S5" value="11" name="L2">est</meta>
<meta type="S" code="S5" value="12" name="L2">eth</meta>
<meta type="S" code="S5" value="13" name="L2">fas</meta>
<meta type="S" code="S5" value="14" name="L2">fra</meta>
<meta type="S" code="S5" value="15" name="L2">hbo</meta>
<meta type="S" code="S5" value="16" name="L2">heb</meta>
<meta type="S" code="S5" value="17" name="L2">hrv</meta>
<meta type="S" code="S5" value="18" name="L2">hun</meta>
<meta type="S" code="S5" value="19" name="L2">ind</meta>
<meta type="S" code="S5" value="20" name="L2">isl</meta>
<meta type="S" code="S5" value="21" name="L2">ita</meta>
<meta type="S" code="S5" value="22" name="L2">jpn</meta>
<meta type="S" code="S5" value="23" name="L2">k.A.</meta>
<meta type="S" code="S5" value="24" name="L2">lat</meta>
<meta type="S" code="S5" value="25" name="L2">lit</meta>
<meta type="S" code="S5" value="26" name="L2">ndl</meta>
<meta type="S" code="S5" value="27" name="L2">nld</meta>
<meta type="S" code="S5" value="28" name="L2">nor</meta>
<meta type="S" code="S5" value="29" name="L2">pol</meta>
<meta type="S" code="S5" value="30" name="L2">por</meta>
<meta type="S" code="S5" value="31" name="L2">pus</meta>
<meta type="S" code="S5" value="32" name="L2">rus</meta>
<meta type="S" code="S5" value="33" name="L2">slk</meta>
<meta type="S" code="S5" value="34" name="L2">spa</meta>
<meta type="S" code="S5" value="35" name="L2">swe</meta>
<meta type="S" code="S5" value="36" name="L2">swh</meta>
<meta type="S" code="S5" value="37" name="L2">urd</meta>
<meta type="S" code="S5" value="38" name="L2">wen</meta>
<meta type="S" code="S5" value="39" name="L2">zho</meta>
<meta type="S" code="S6" name="Schulbildung" />
<meta type="S" code="S7" name="Auslandsstudium" />
<meta type="S" code="S8" name="Auslandsaufenthalt" />
```