

DiÖ-DB benutzen

Benutzername & Passwort

Bevor man auf die Datenbank zugreifen kann, benötigt man einen Benutzernamen und ein Passwort – bei Bedarf bitte bei der LV-Leitung melden.

Der Benutzername besteht aus vorname.nachname, das Passwort wird per Mail zugeschickt.

Wichtig: Beim ersten Login ist das Passwort unbedingt zu ändern!

Im Reiter rechts oben auf euren Nutzer:innennamen klicken à „Passwort ändern“ auswählen. Hier könnt ihr euch auch ausloggen („Abmelden“).



Einloggen

Die Datenbank wird über den Browser (bevorzugt **Chrome**) geöffnet. Achtung: Firefox funktioniert manchmal nicht.

Bitte nur über den eigenen und keine öffentlichen PCs einloggen, immer nach der Benutzung

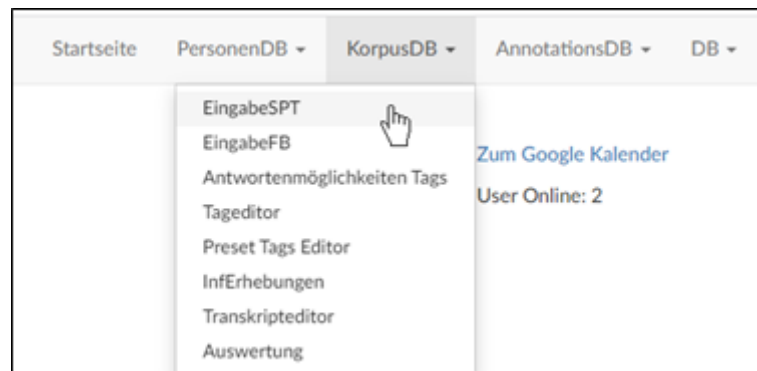
ausloggen und das Passwort bitte nicht speichern!

URL: <https://dioedb.demo.dioe.at>

Übersichtsseite

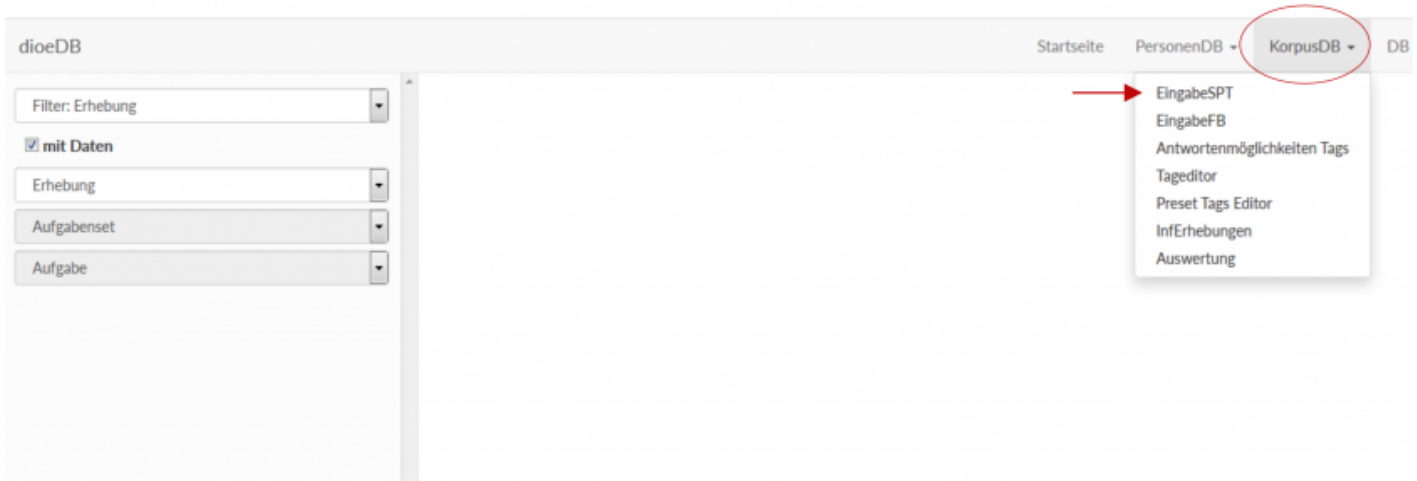
Nach dem Log-in gelangt man auf die Übersichtsseite, von der aus man die Korpus-Datenbank (Korpus DB) und die Annotations-Datenbank (AnnotationsDB) durch einen Klick aufrufen kann.

- In der **KorpusDB** passieren sämtliche Arbeiten (Anhören, Transkribieren, Annotieren) an den kontrollierten Daten (**Experimente, Leseliste, Wenker-Übersetzungen**).
- In der **AnnotationsDB** können die freien Gespräche (**Interviews, Freundesgespräch**) angehört und annotiert werden.



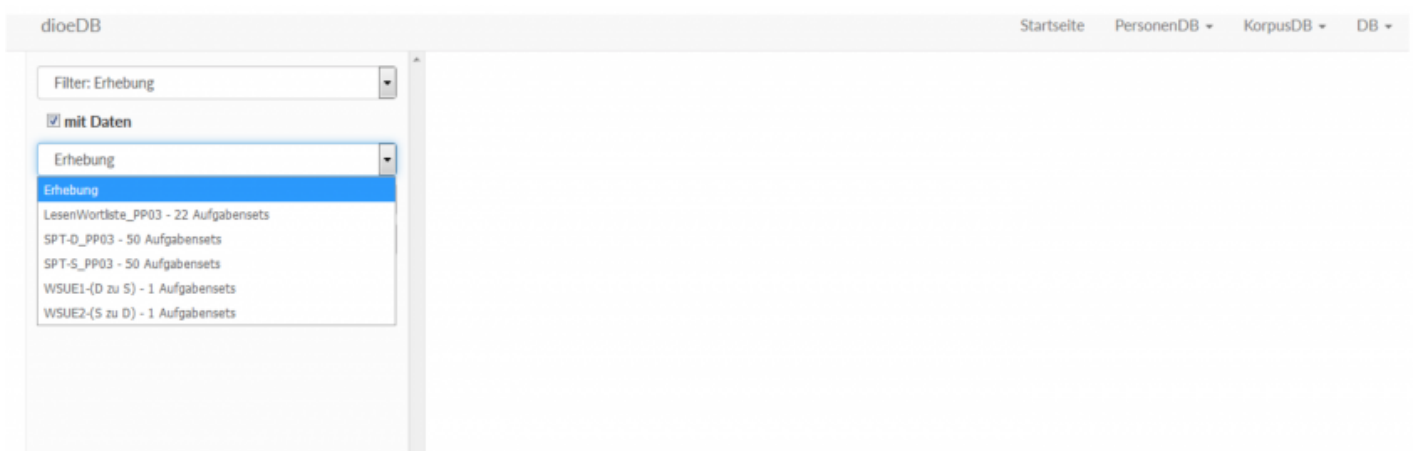
1. Korpus DB: Wenker-Übersetzungen, Sprachproduktionsexperimente, Vorlese-Wortliste

Um die KorpusDB zu öffnen, klickt man auf "**Eingabe SPT**" (siehe Bild).



Um ausgewählte Aufgaben anzuhören, muss bei „**Filter:Erhebung**“ das richtige Erhebungssetting ausgewählt werden. (Alternativ kann auch nach den einzelnen Gewährspersonen ("Filter: Informant") oder nach Phänomenen gefiltert werden.) Zur Auswahl stehen:

- **LesenWortliste_PP03**: Vorlesen der Wortliste
- **SPT-D_PP03**: Sprachproduktionsexperiment im intendierten Dialekt
- **SPT-S_PP03**: Sprachproduktionsexperiment im intendierten Standard
- **WSUE1-(D zu S)**: Übersetzung der Wenkersätze vom Dialekt in den Standard
- **WSUE2-(S zu D)**: Übersetzung der Wenkersätze vom Standard in den Dialekt



Im Fenster „**Aufgabenset**“ kann nun das Aufgabenset ausgewählt werden. Die Benennung zeigt grob an, um welches Phänomen es sich jeweils handelt, bei den SPTs z. B. KONJ = Konjunktivaufgaben; bei den Wenkersätzen z. B. "WS-D" für die Sätze im Dialekt oder "WS-S" für die Sätze im Standard; bei der Leseliste zum Beispiel "L02_ig" für Wörter auf -ig, z. B. „König“

Achtung Leseliste: Leider können Wörter, in denen mehrere Phänomene drinstecken, nicht mehreren Kategorien zugewiesen werden, daher bitte immer alle „Aufgabensets“ durchschauen (z. B. „Kleinigkeit“ für die Phänomene „-ig“ und „ei“ relevant).

dioeDB

Startseite

PersonenDB ▾

KorpusDB ▾

DB ▾

Filter: Erhebung ▾

☒ mit Daten

SPT-D_PP03 - 50 Auf; ▾

Aufgabenset ▾

DART3 (2) - 2 Aufgaben

FINI1 (2) - 2 Aufgaben

FINI2 (2) - 2 Aufgaben

FINI3 (2) - 2 Aufgaben

KONJ1 (1) - 4 Aufgaben

NEGC1 (2) - 1 Aufgaben

NEGC2 (2) - 1 Aufgaben

NEGC3 (2) - 1 Aufgaben

NEGC4 (2) - 1 Aufgaben

POSS1 (2) - 4 Aufgaben

POSS2 (2) - 4 Aufgaben

POSS3 (2) - 2 Aufgaben

POSS4 (2) - 2 Aufgaben

PROG1 (3) - 2 Aufgaben

In einem Aufgabenset sind mehrere Aufgaben zu einem Phänomen gruppiert, diese können im Fenster „**Aufgabe**“ entsprechend ausgewählt werden, z. B. KON_SPT-D_1 = erste Aufgabe mit Konjunktiv "Stell dir vor, du gewinnst morgen 10 Mio. Euro im Lotto";

z. B. 15-König - Bildbenennung = Vorlesen des Items "König", das als Bild präsentiert wurde. Im Falle der Wenkersätze findet sich hier einfach die Auflistung der einzelnen Wenkersätze, z. B. 11 - Ich schlage Dich gleich mit dem Kochlöffel um die Ohren, Du Affe!

dioeDB

Startseite

PersonenDB ▾

KorpusDB ▾

AnnotationsDB ▾

DB ▾

Filter: Erhebung ▾

☒ mit Daten

LesenWortliste_PP03 - 22 Aufgabensets ▾

L02_ig - 23 Aufgaben ▾

Aufgabe ▾

Aufgabe

1 - 20 - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

2 - 33 - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

3 - 44 - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

4 - 66 - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

5 - 99 - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 86,2% mit Tags

6 - 30.11.2033 - Vorlesen der Aufgabenstellung (L02_ig) - 3,5% Erledigt, davon 87,5% mit Tags

7 - beleidigt - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

8 - einig - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

9 - erledigt - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 86,2% mit Tags

10 - gefällt - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 86,2% mit Tags

11 - geläufig - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

12 - häufiger - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags und 2 unsicher

13 - Honig - Bildbenennung (L02_ig) - 4,1% Erledigt, davon 86,0% mit Tags

14 - Kleinigkeit - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

15 - König - Bildbenennung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

16 - königlich - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

17 - lediglich - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

18 - richtig - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,2% mit Tags

19 - Tätigkeit - Vorlesen der Aufgabenstellung (L02_ig) - 3,8% Erledigt, davon 87,1% mit Tags

Sobald alle Filter ausgewählt wurden, erscheint links eine **Liste**. Daraus kann also im letzten Schritt jene **Gewährsperson** ausgewählt werden, die bearbeitet werden soll. Die Gewährspersonen sind mit ihrer ID und einer Abkürzung des Erhebungsorts aufgelistet, z. B. "0025 - NECK" = Person mit der ID 0025 aus Neckenmarkt/Burgenland. In der Liste wird außerdem schnell sichtbar, wo bereits Antworten angelegt sind (Zahl) und wo sogar schon getaggt (grünes Symbol) wurde. Eine Antwort ist immer eine Rückmeldung einer Person zum Stimulus. Korrigiert sich die Person oder wird durch den/die Explorator*in noch einmal nachgefragt, können auch mehrere Antworten zu einem Stimulus vorliegen. Erst wenn man auf ein Item aus der Liste klickt, wird es geladen (Achtung, Ladezeit).

Als Reminder hier die Kürzel der Orte und ihre Einordnung in ein Dialektgebiet (nach Wiesinger 1983):

- *ALLE: Allentsteig/NÖ = Mittelbairisch*
- *GAW: Gaweinstal/NÖ = Mittelbairisch*
- *HÜTT: Hüttschlag/Sbg = Süd-/Mittelbairisches Übergangsgebiet*
- *NECK: Neckenmarkt/Bgld = Süd-/Mittelbairisches Übergangsgebiet*
- *NEUMY: Neumarkt/Ybbs/NÖ = Mittelbairisch*
- *OB: Oberwölz/Stmk = Südbairisch*
- *PASS: Passail/Stmk = Süd-/Mittelbairisches Übergangsgebiet*
- *RAGG: Raggal/Vbg = Höchstalemannisch*
- *STEY: Steyrling/OÖ =*
- *TARR: Tarrenz/T = Bairisch-Alemannisches Übergangsgebiet (in phonologischen Merkmalen heute v. a. Südbairisch)*
- *TAUF: Taufkirchen/OÖ = Mittelbairisch*
- *TUX: Tux/T = Südbairisch*
- *WEIS: Weißbriach/Ktn = Südbairisch*

Nach einer Ladezeit werden die angewählten Daten direkt rechts im bisher leeren/weißen Teil des Fensters angezeigt.

Achtung: Geduldig sein! Die DB muss die Inhalte immer neu laden! Nicht zu schnell zwischen den Sprecher*innen in der Liste hin- und herklicken.

dioeDB Startseite PersonenDB ▾ KorpusDB ▾ AnnotationsDB ▾ DB ▾

Filter: Erhebung ▾

☒ mit Daten

LesenWortliste_PP03 - 22 Aufgabensets ▾

L02_ig - 23 Aufgaben ▾

15 - König - Bildbenennung (L02_ig) - 3,8% Erledigt ▾

0001 - PIRI	0
0002 - PIRI	0
0003 - PIRI	0
0004 - PIRI	0
0005 - PIRI	0
0006 - PIRI	0
0024 - NECK	1
0025 - NECK	1
0026 - NECK	

PK: 6 - Ort: Neckenmarkt, Bezirk Oberpullendorf, Burgenland, 7311, Österreich

dioeDB Startseite PersonenDB ▾ KorpusDB ▾ AnnotationsDB ▾ DB ▾

Filter: Erhebung ▾

☒ mit Daten

LesenWortliste_PP03 - 22 Aufgabensets ▾

L02_ig - 23 Aufgaben ▾

15 - König - Bildbenennung (L02_ig) - 3,8% Erledigt ▾

0025 - 15 - Bildbenennung (L02_ig)

Aufgabe: ID: 456 - Aufgabenset: L02_ig - Variante: 15 - Stimulus Dialekt: Nein - Evokationsziel Dialekt: Nein - Beschreibung: Bildbenennung
Aufgabenstellung: König

1. Antwort: ☐ beeinflusst start: 00:00:00.000000 T stop: 00:00:00.000000 T
komm.: ☐ kontrolliert ☐ veröffentlichung

standard König I

transkript

IPA-Transkript kœ:nɪkʰ

IG ▾ igVar igReal P + LautNach MG + + + ★

+ Tag-Ebene

2. Antwort: ☐ beeinflusst start: 00:00:00.000000 T stop: 00:00:00.000000 T
komm.: ☐ kontrolliert ☐ veröffentlichung

Man sieht also die Antwort in der Mitte samt **Transkript**. In den verschiedenen Spuren "standard" (Standardorthografische Transkription), "transkript" (lautororientierte Transkription) und "IPA-Transkript" (phon. Transkription) kann die dementsprechende Transkription angefertigt werden. Für Wenker und SPT wird meist eine lautororientierte Umschrift verwendet, für die Leseliste auch phonetische Umschrift. **Tipp für phonetische Transkription:** Bei der Eingabe erscheint auch eine zusätzliche Tastatur zur phonetischen Transkription. Bei Bedarf hier transkribieren und nach Excel kopieren.

Oben befindet sich noch eine genauere **Beschreibung der Aufgabe**, z. B. "Bildbenennung" (manchmal muss man ins Audio hineinhören, wenn die Angabe oben nicht existiert);

Unter der Antwort ist ggf. bereits ein **Tag** (Annotation = linguistische Information) vorhanden. Mehrere Antworten werden untereinander angezeigt und können einzeln getaggt werden. Um eine Antwort mit einer (bereits bestehenden) Annotation zu versehen klicken Sie auf "+Tag Ebene" und

wählen Ihre Tags aus.

Ganz unten befindet sich der **Player**, mit der die Antwort durch Klick auf den Play-Button für die jeweilige Aufgabe angehört werden kann. Manchmal kann es vorkommen, dass Audio und Antwort nicht ganz übereinstimmen, man kann dann in den unteren Zeitbalken hineinklicken und einen größeren Ausschnitt anhören. Will man ein Wort/einen Laut mehrmals anhören, kann man in den oberen Zeitbalken hineinklicken (mehrmals) oder über das Dropdown-Menü rechts unten statt „Einfach“ „Wiederhole“ bzw. „Wiederhole Antwort“ auswählen.

Nur wenn neue Transkriptionen oder Tags hinzugefügt wurden, muss gespeichert werden (nicht fürs Anhören). Der **"Speicher"-Button** findet sich unten rechts in grün.

The screenshot shows the dioeDB interface. The top navigation bar includes 'dioeDB', 'Startseite', 'PersonenDB', 'KorpusDB', 'DB', and 'MELANIE.NELTMANN'. The left sidebar lists tasks under 'SPT-D_PP03 - 50 Aufga', 'TEST4 (3) - 1 Aufgaben', and '2 - Testitem_SPT-D_4 (1)'. The main content area displays '0032 - 2 - Testitem_SPT-D_4 (TEST4 (3))' with details: 'Aufgabe: ID: 187 - Aufgabenset: TEST4 (3) - Variante: 2 - Dialekt: Ja - Beschreibung: Testitem_SPT-D_4'. Below this are input fields for '1. Antwort', 'beeinflusst', 'start', and 'stop'. A 'komm.' field is also present. A red arrow points to the 'transkript' button. The bottom section shows a timeline with 'start: 00:00:48.864000' and 'stop: 00:01:04.850000', and a media player with a play button. A green button 'Antworten Speichern' is visible on the right.

Export: Wenn Sie selbst transkribiert/annotiert haben, können Sie die transkribierten und/oder annotierten Antworten zur Weiterbearbeitung in Excel exportieren. Klicken Sie dafür ganz oben im Drop-Down-Menü zu "KorpusDB" auf den letzten Punkt "Auswertung". Hier können Sie angeben, ob Sie nur die Antworten ("Antworten") oder die Antworten und die Annotationen ("Antworten (Tag Ebenen)") downloaden wollen. Dann ist wiederum das Setting und das Phänomen auszuwählen. Ein Klick auf "CSV" oder "XLS" (empfohlen) startet den Download.

Allgemeine Zusatzinfos zur KorpusDB:

Filtern: Die Daten können nach unterschiedlichen Kriterien **gefiltert** werden – überlegen/ausprobieren, was individuell am praktischsten ist. Wenn man z. B. nur wenige Personen analysieren möchte oder nur ein Phänomen, kann man den Filter entsprechend einstellen.

Achtung: Man muss immer alle Filter einstellen und auf ein konkretes Item aus der Liste klicken, bevor Daten angezeigt werden (Wartezeit!). Mit dem Häkchen „mit Daten“ werden nur Personen/Aufgaben angezeigt, die ausgefüllt sind, wo also Daten vorliegen. Ohne das Häkchen werden auch leere Daten angezeigt.

- Nach **„Erhebung“** filtern: entspricht quasi dem Setting. Die nächste Kategorie sind die „Aufgabensets“ (z. B. DIM = Diminutive im Falle der SPTs). Bei den Wenkersätzen gibt es immer nur ein Aufgabenset. Bei „Aufgabe“ den konkreten Wenkersatz, den konkreten SPT oder das konkrete Wort (Leseliste). Hier finden sich die Antworten aller Personen zu einer Aufgabe.
 - Nach **„Informant“** filtern: das ist der/die Sprecher*in. Personen werden in der Datenbank immer mit ihrer Informanten-ID gespeichert. Als letzte Filteroption kann hier noch die Erhebung eingestellt werden, z. B. „WSUE1-(D zu S)“ zeigt alle Wenkersätze (Dialekt zu Standard) einer Person an. Auf den konkreten Satz klicken und warten bis Daten geladen sind.
 - Nach **„Phänomen“** filtern: zeigt alle Antworten zu einem Phänomen (muss vorher definiert und angelegt worden sein). Setting wird erst danach eingestellt, dann erst die konkrete Aufgabe.
-
- Wenn Tags/Transkriptionen in der Datenbank (DB) erfolgen: Vergessen Sie nicht häufig zwischenzuspeichern – lieber einmal zu viel als zu wenig!
 - Allgemeines Troubleshooting (löst ca. 75% aller Probleme): Verwenden Sie Google Chrome! Leeren Sie den Cache! Refreshen Sie! ;)
 - Bei Fragen oder Problemen können Sie sich gern an die LV-Leitung oder unseren Technik-Experten Markus wenden (pluschkovits@univie.ac.at, telegram: @Lmpzvgnbds)!

2. Annotations DB: Interview (INT), Freundesgespräch (FG oder GFG), Vorlesetext Nordwind und Sonne (LESN)


Die Annotations DB unterteilt sich in:

1. das Annotationstool

Zugang zu den Transkripten erhalten ihr unter dem Reiter „AnnotationsDB“ im „Annotations Tool“.

Hier könnt ihr die Transkripte zum/zur jeweiligen Informant:in aufrufen.

Transkript:

0025 (9) - 4 Transkripte 

0024_0025_NECK_GFG_Vers35
15 176

0025_NECK_jungll_m_BDE_INT_Vers8
14 842

0025_NECK_jungll_m_INT_Vers2
5 14 724

0025_NECK_jungll_m_LESN_Vers6
128

Kein Transkript ausgewählt!

Für die meisten Informant:innen findet ihr ein Interview, ein Freundesgespräch, und in vielen Fällen noch einen kurzen Lesetext (Nordwind und Sonne). (Bei manche Informant:innen gibt es außerdem ein Transkript zum BDE_INT - Achtung, dabei handelt es sich nicht um das "reguläre Interview", sondern um ein Kurzinterview, das im Rahmen der **Bundesdeutschen Erhebung** von einem bundesdeutschen Sprecher geleitet wurde!) Nach Auswahl des Transkriptes wird dieses geladen.

Das Annotationstool wird verwendet, um bereits bestehende Transkripte zu annotieren. Nach der Auswahl des Transkripts in der Liste (über die SprecherInnen-ID) öffnet sich folgende Ansicht des Transkripts:

1. Blendet allgemeine Informationen zum Transkript ein und aus.
2. Blendet Ansichtsoptionen ein. Hier lassen sich die Event-Tiers und Tokenset-IDs zuschalten (standardmäßig abgewählt), die unter/über den jeweiligen Events und Tokensets angezeigt werden. Zusätzlich lassen sich die tatsächlichen Tags über Tokensets anzeigen. Mit der Auswahl "Benutzte Tagebenen anzeigen" wird nur die jeweilige Ebene über allen Tokensets angezeigt. Wählt man eine bestimmte Ebene aus, so wird der gesamte Tag auf dieser Ebene über den betreffenden Tokensets angezeigt.
3. Blendet Filteroptionen aus und ein. Mit diesen Filtern lässt sich verändern, welche SprecherInnen und Spuren angezeigt werden.

4. Blendet Suchfunktion aus und ein.
5. Blendet eine Übersicht über die Tastaturshortcuts aus und ein.
6. Zeitleiste des Audioplayers (STRG + Klick um von gewünschter Stelle an abzuspielen).
7. Steuerung des Audioplayers.
8. Speicherknopf

Annotieren in der AnotationsDB

Tagging, theoretisch – das A und O der Datenerschließung

Die explizite Annotation^[1] der Sprachdaten ist ein wichtiger Teil der Forschungsarbeit. Im SFB:DiÖ funktioniert diese Annotation^[2] über die Vergabe von Tags – diese kann man sich als linguistische Label zu einem Sprachdatum vorstellen. Tags werden beim SFB hierarchisch auf einer Tagebene vergeben. Das meint nichts anderes als das – wie bei einem Baumdiagramm – manche Tags nur auf spezifische Tags folgen können. Die individuellen Tagebenen schränken dabei ein, welche Tags überhaupt ausgewählt werden können, und bilden auch die Basis für die Auswertung, sprich die Ausgabe der Daten aus der DB. Dabei kann alles Sprachmaterial, das auf einer Tagebene getagged wurde, als .xcl Datei ausgeworfen werden. Bedenkt bei eurem Tagging bitte, dass *alles* getaggte auf einer Tagebene ausgegeben wird – sprich, wenn diese Tagebene schon benutzt wurde, dann werden alle bereits auf ihr getaggtten Daten ausgegeben. Die Hierarchie der Tags ist eine Stütze beim Annotieren. Sie hilft, dass nur wohlgeformte Annotationen entstehen. Außerdem ist die Hierarchie semantisch – wir sprechen von sog. Generationen, die in parent-child-Beziehungen zueinanderstehen. Die einzelnen Generationen tragen dabei ein Stück weit eine Bedeutung – ein Tag der Generation 0 bezeichnet gemeinhin ein linguistisches Phänomen (und kann nur von Tags der Gen. 1 gefolgt werden), Tags der Generation 1 bezeichnen Kategorien dieses Phänomens (und können nur von Tags der Gen 2 gefolgt werden), und Tags der Generation 2 bezeichnen Features der Kategorien. Kategorien ergänzen einander, während Features einander ausschließen. Betrachten wir hierzu ein Beispiel, etwa Verben – Verben als linguistische Phänomen

(Gen0) können anhand von 5 Kategorien (Gen1) bestimmt werden: Tempus, Modus, Numerus, Person, Genus Verbi. Diese Kategorien ergänzen einander. Die einzelnen Features bei einem konkreten Verb (Gen2-Tags) schließen sich jedoch aus – ein spezifisches Verb kann nicht gleichzeitig Aktiv und Passiv sein

[1] Zur Datenbank und Annotation siehe Breuer, Ludwig Maximilian & Melanie E. H. Seltmann. 2018. Sprachdaten(banken) – Aufbereitung und Visualisierung am Beispiel von SyHD und DiÖ. In Ingo Börner, Wolfgang Straub & Christian Zolles (Hgg.), Germanistik digital: digital humanities in der Sprach- und Literaturwissenschaft, 135–152. Wien: facultas.

[2] Zur Annotation im SFB spezifisch siehe Pluschkovits, Markus & Katharina Kranawetter. 2021. Annotation von Sprachdaten eines variationslinguistischen Großprojekts am Beispiel des Spezialforschungsbereichs ›Deutsch in Österreich‹. In Agnes Kim, Katharina Korecky-Kröll, Ludwig Maximilian Breuer, Jan Höll und Wolfgang Koppensteiner (Hgg.), Vom Tun nicht lassen können: Historische und rezente Perspektiven auf sprachliche Variation (in Österreich und darüber hinaus) Festgabe für Alexandra N. Lenz zum runden Geburtstag (=Wiener Linguistische Gazette 89), 167–189 (

https://wlg.univie.ac.at/fileadmin/user_upload/p_wlg/892021/FSLenz_PluschkovitsKranawetter.pdf)

Tagging, praktisch: How To Do Things With Tags

Tagverwaltung

Eine Tagebene erstellen: Zum Erstellen einer Tagebene zum Reiter „DB“ navigieren, dort „Verwaltung“ auswählen. Unter KorpusDB „Tagebenen“ anwählen, mit dem Blatt-Papier-Icon eine neue Tagebene anlegen.

Neue Tags erstellen: Im Reiter „KorpusDB“ den „Tageditor“ auswählen. Mit dem Papiericon rechts oben lässt sich ein neues Tag erstellen. Wichtig dabei ist, Generation und Tagebene, sowie etwaige Children und Parents anzugeben! Tags können auch in bereits bestehende Tagfamilien eingebaut werden – hier nur Parent-Child-Beziehungen betrachten, und bitte keine Tags Ihrer Kolleg:innen löschen!

Tag-Presets erstellen: Im Korpus-DB-Reiter zu „Preset Tags Editor“ navigieren, erneut mit dem Papier-Icon ein neues Preset erstellen. Wichtig: ordnen Sie dem Preset-Tag eine Tagebene zu! Fügen Sie die Tags zum Preset in der Reihenfolge hinzu, in der Sie angezeigt werden **sollen** – beachten Sie dabei die Generationen!

Tagging von Transkripten:

Öffnen Sie das entsprechende Transkript via dem Reiter „AnnotationsDB“. Sie haben im Transkript drei Möglichkeiten der Annotation: Sie können einzelne Token, durchgehende Spannen von Token, oder ausgewählte Token in einer Gruppe annotieren.

Um ein **einzelnes Token** zu taggen, klicken sie auf das entsprechende Token (zwei Mal). Ein Pop-Up öffnet sich. Durch den Klick auf „Antwort erstellen“ bekommen Sie die Möglichkeit, Tag-Ebene und sukzessive Tags/das Preset auszuwählen. Schließen Sie das Pop-Up, indem sie auf „Ändern“ klicken. Vergessen Sie im Abschluss nicht zu speichern!

Token

×

ID

53605

Spuren

☐ Alle Anzeigen

ortho

text

phon

event

0:11:39.099000 - 0:11:39.519000 - ID: 11745

start_timepoint

end_timepoint

ID_Inf


0044 (0044 - ID: 14)

Felder

☐ Alle Anzeigen

Antwort

Antwort erstellen



Wein

bäuerin ja, ist mir am liebsten, ja. ja. ja. nein, winemaker das habe ich nicht so gerne. ((lachen)) nein, kann ich nicht mehr. was ich kann, mache ich. ja

⏮

▶

⏭

event ▾

☒ Loop

00:11:39.099000

00:11:39.099000

00:11:39.519000

◀

Ändern

▶

Schließen

Um eine **Tokenspanne** (also mehrere Token hintereinander zu annotieren), klicken Sie auf den Token, bei dem die Spanne beginnen soll. Klicken Sie, während Sie shift links gedrückt halt, auf den Token, der am Ende der Spanne stehen soll. Alle Token dazwischen werden ausgewählt (i.e. grün unterlegt). Im linken Bereich erhalten Sie nun die Möglichkeit, diese Token in ein Tokenset umzuwandeln. Sobald das geschehen ist, sehen sie einen kleinen blauen Bogen oberhalb der ausgewählten Token. Klicken Sie diesen zwei Mal, und ein Pop-Up öffnet sich. Analog wie bei Einzelntoken können Sie hier eine Antwort erstellen und Tags vergeben.

Die Tastenkombination um eine durchgängige Folge an Token zu annotieren ist Shift + Klick. Dazu einmal auf den Beginn des zu annotierenden Bereich klicken, Shift drücken, und das letzte Token des Tokensets anklicken. Alle Token eines Sprechers, die zwischen diesen beiden Token liegen, werden zu dem Tokenset hinzugefügt (bei diesem Beispiel von *ja* bis *((lachen))*).

The screenshot shows the SFB DiÖ interface. On the left, a sidebar displays 'Auswahl Bereich: 12 Tokens' with fields for 'Informant: 0044', 'Von Token:', and 'Bis Token:'. Below these is a button 'Auswahl in Token Set umwandeln'. A red arrow points to this button. The main area shows a transcription timeline with two rows of data. The first row is for Informant 0044, showing a transcription of 'ja sage ich "ich ich (()) tue das nicht so gern" ((lachen))' with a duration of 00:12:08.553 to 00:12:09.786. The second row shows a transcription of '((O,9s)) mhm.' with a duration of 00:12:09.786 to 00:12:10.000. Below the main area, a detailed view of the first transcription shows the text 'ja sage ich "ich ich (()) tue das nicht so gern" ((lachen))' with a duration of 00:12:09.786 to 00:12:10.000. A blue bracket above the text indicates a selection of 1 token.

Um ein **diskontinuierliches Set von Token** zu erstellen, halten Sie bei der Auswahl der Token strg gedrückt. Sie können einzelne Token einer/s Sprecher:in auswählen. Wie zuvor bei der Tokenspanne lassen sich diese links zu einem Tokenset umwandeln, welches dann getaggt werden kann.

2. Anno-sent

Durch Anno-sent ist das gesamte Korpus des SFB DiÖ durchsuchbar. Dies bietet sich besonders bei Fragestellungen, die die Lexik betreffen, an. Z.b. lässt sich das Korpus durch Lemmata (wie z.B. Paradeiser versus Tomate) durchsuchen, um Einblicke zu erhalten, wie frequent die einzelnen Lexeme auftreten.

3. Anno-check

Anno-check dient primär der Überprüfung und Anpassung bereits bestehender Annotationen.

Anno-Check lädt direkt beim Öffnen *alle* „Antworten“ der Sprachproduktionsexperimente und es kommt zur Anzeige „Fehler! Kein Satz übergeben!“. Das kann hier aber getrost ignoriert werden, da noch keine Suachanfrage getätigt wurde und hier Daten "sichtbar" sind, die noch nicht transkribiert wurden. also hier ignoriert werden.

Um Anno-Check nun zu nutzen, müssen Filterkriterien festgelegt werden, i.e.: Was will ich sehen? Dabei lässt sich nach filtern nach:

Tag Ebene: meist das konkrete Phänomen, z.B. KONJ (Konjunktiv)

Tag: Teil der gesamten Annotation, z.B. Kperi (periphrastischer Konjunktiv)

Nicht Tag: darüber kann ausgeschlossen werden, welche Ergebnisse nicht auftauchen sollen, z.B. würdeK (würde Konjunktiv) (--> Suchanfrage über Tag Ebene: KONJ, Tag: Kperi, Nicht Tag: würdeK findet somit nur periphrastische Konjunktivformen mit *täte*)

Informant: Die einzelnen InformantInnen können über Ihre Sigle ausgewählt werden.

Transkript: Neben jedem einzelnen Transkript (Interviews und Freundesgespräche) können auch Teil-Korpora (Nur Transkripte bedeutet alle freien Gespräche (INT, FG); Keine Transkripte bedeutet kontrolliertere Settings (SPE, Wenker, Leseliste) geladen werden.

Aufgabensets: Um nicht die gesamten kontrollierten Daten zu laden, gibt es die Möglichkeit, dass nur ausgewählte Aufgabensets (z.B. Wenkersätze, ARTM (=Artikel vor Massennomen) der kontrollierten Erhebungen angezeigt werden.

Aufgabe: Mit "Aufgabe" lässt sich noch feiner filtern, innerhalb des Aufgabensets lässt sich weiters nur eine Aufgabe anzeigen, z.B. einzelne Wenkersätze

Tag Ebene	<input type="text" value="Alle"/>	Transkript	<input type="text" value="Alle"/>
Tag	<input type="text" value="Alle"/>	Aufgabensets	<input type="text" value="Alle"/>
Nicht Tag	<input type="text" value="Alle"/>	Aufgaben	<input type="text" value="Alle"/>
Informant	<input type="text" value="Alle"/>	<input type="checkbox"/> Anzahl anzeigen	

Nachdem die entsprechenden Filter gesetzt wurden, erscheint die Antwort-Tabelle. In der Tabelle lässt sich jede Antwort anklicken, wodurch sich ein Pop-Up-Fenster zur Bearbeitung öffnet, in dem sich die Tags in gewohnter Weise verändern lassen und die Antwort anzuhören ist.

4. Anno Auswertung

Im Reiter „AnnotationsDB“ lassen sich unter „Anno Auswertung“ die getaggten Daten ausgeben. Wähle dazu die Tagebene aus und wähle „Als XLS herunterladen“ aus. Für größere Ausgaben (+1k Ergebnisse) wähle bitte „XLS am Server generieren“ aus, und lade das Ergebnis via dem Reiter „DB“ --> „Dateien“ im Verzeichnis „Auswertung“ manuell herunter (kann einige Minuten in Anspruch nehmen bis die xls erscheint).

Export:

Kompletten Datensatz: [Als XLS herunterladen](#) - [XLS Datei am Server generieren](#)[XLS geteilt runterladen](#)

#	Transkript	tId	Informant	ild	iGebDatum	iWeiblich	iGruppe	iOrt	antId	antType	antKomme
1	None	None	0034	67	1990-03-10	True	jung (18-35) +matura	Allentsteig, Zwettl, Niederösterreich, 3804, Österreich	19	s	None
2	None	None	0034	67	1990-03-10	True	jung (18-35) +matura	Allentsteig, Zwettl, Niederösterreich, 3804, Österreich	74	s	None
3	None	None	0034	67	1990-03-10	True	jung (18-35) +matura	Allentsteig, Zwettl, Niederösterreich, 3804, Österreich	90	s	None

! Achtung: Die Datei ist leider häufig zu groß und es kommt zu einer Fehlermeldung. Sollte dies passieren, so ist die bereits erstellte (XLS Datei am Server generieren) Datei zu finden unter "DB" bei "Dateien". Als Verzeichnis ist dann unter "annotationsdb" "auswertung" zu wählen. In diesem Ordner liegen dann alle generierten Tabellen, benannt sind diese nach der Nummer der Tag Ebene (z.B. tageebene_0_25). (Tipp: ein Blick auf das Erstellungsdatum kann bei der Suche nach der richtigen Datei hilfreich sein!).

Verzeichnisse:

...	...	
...	ai04	
...	allfb02	
...	allsptd03	
...	allspts03	
...	allwort03	
...	annotationsdb	
...	auswertung	
...	bdint03	
...	bdnws03	
...	bdsptd03	
...	bdspts03	

AnnotationsDB ▾

DB ▾

Verwaltung

Diagramm

Dateien

5. Auswertung in Excel

Wichtig: Bitte bei Unsicherheiten/Problemen gerne bei der LV-Leitung nachfragen!

Revision #13

Created Mon, May 27, 2019 10:32 AM by [Ludwig Maximilian Breuer](#)Updated Thu, Jan 5, 2023 11:11 AM by [Anja Wittibschlager](#)