

Übersicht über die in DISKO vorhandenen Transkriptions- und Annotationsspuren

Name der Spur	Beschreibung	DISKO _L2	DISKO _L1	DISKO _DSH	DISKO _WebTestDaF
tok	tokenisierte Transkription (die möglichst wenige Interpretationen beinhalten soll)	ja	ja	ja	ja
tok_lemma_auto	automatisch generierte Lemma-Annotation zu <i>tok</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	ja	ja	ja	ja
tok_pos_auto	automatisch generierte POS-Annotation zu <i>tok</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	ja	ja	ja	ja
tok_s_auto	automatisch generierte Satzannotation zu <i>tok</i>	ja	ja	ja	ja
tok_lemma_auto_no_unknown	automatisch generierte Lemma-Annotation zu <i>tok</i> ; bei Wörtern, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, erscheint die Form aus <i>tok</i> als Lemma-Tag	ja	ja	ja	ja
tok_lemma_74doc	manuell korrigierte Lemma-Annotation zu <i>tok</i> (nur für einen Teil der Transkripte vorhanden)	74/510	nein	nein	nein
tok_pos_74doc	manuell korrigierte POS-Annotation zu <i>tok</i> (nur für einen Teil der Transkripte vorhanden)	74/510	nein	nein	nein
tok_s_74doc	manuell korrigierte Satzannotation zu <i>tok</i> (nur für einen Teil der Transkripte vorhanden)	74/510	nein	nein	nein

Tabelle 1: Übersicht der DISKO-Annotationsspuren zu tok

Name der Spur	Beschreibung	DISKO _L2	DISKO _L1	DISKO _DSH	DISKO _WebTestDaF
norm	teilautomatisch nach Rechtschreibung normierte Transkription (als <i><unknown></i> getaggte Wörter mit Häufigkeit >3)	ja	ja	nein	ja
norm_tok_Diff	Abweichungen zwischen <i>tok</i> und <i>norm</i>	ja	ja	nein	ja
norm_lemma_auto	automatisch generierte Lemma-Annotation zu <i>norm</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	ja	ja	nein	ja
norm_pos_auto	automatisch generierte POS-Annotation zu <i>norm</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	ja	ja	nein	ja
norm_s_auto	automatisch generierte Satzannotation zu <i>norm</i>	ja	ja	nein	ja
norm_lemma_auto _no_unknown	automatisch generierte Lemma-Annotation zu <i>norm</i> ; bei Wörtern, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, erscheint die Form aus <i>norm</i> als Lemma-Tag	ja	ja	nein	ja
norm_lemma	nach bestimmten Phänomenen teilautomatisch korrigierte Lemma-Annotation zu <i>norm</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	ja	ja	nein	ja
norm_pos	nach bestimmten Phänomenen teilautomatisch korrigierte POS-Annotation zu <i>norm</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	ja	ja	nein	ja
norm_s	nach bestimmten Phänomenen teilautomatisch korrigierte Satzannotation zu <i>norm</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	ja	ja	nein	ja

Tabelle 2: Übersicht der DISKO-Annotationsspuren zu *norm*

Name der Spur	Beschreibung	DISKO _L2	DISKO _L1	DISKO _DSH	DISKO _WebTestDaF
TH1_92doc	minimale Zielhypothese (Normalisierungsebene: Orthografie, Morphosyntax)	92/510	nein	nein	nein
TH1_tok_Diff_92doc	Abweichungen zwischen <i>TH1_92doc</i> und <i>tok</i>	92/510	nein	nein	nein
TH1_lemma_auto_92doc	automatisch generierte Lemma-Annotation zu <i>TH1_92doc</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	92/510	nein	nein	nein
TH1_pos_auto_92doc	automatisch generierte POS-Annotation zu <i>TH1_92doc</i> ; Wörter, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, werden mit <i><unknown></i> annotiert	92/510	nein	nein	nein
TH1_s_auto_92doc	automatisch generierte Satzannotation zu <i>TH1_92doc</i>	92/510	nein	nein	nein
TH1_lemma_auto_no_unknown_92doc	automatisch generierte Lemma-Annotation zu <i>TH1_92doc</i> ; bei Wörtern, die nicht in der Lemma-Datenbank des TreeTaggers hinterlegt sind, erscheint die Form aus <i>TH1</i> als Lemma-Tag	92/510	nein	nein	nein

Tabelle 3: Übersicht der DISKO-Annotationsspuren zu TH1

Name der Spur	Beschreibung	DISKO _L2	DISKO _L1	DISKO _DSH	DISKO _WebTestDaF
borrowed_203doc	Annotation zu aus der Aufgabenstellung übernommenen Textteilen	203/510	nein	nein	nein
task_119doc	Annotation markiert, welche Textabschnitte welchen Teil der Aufgabenstellung behandeln	119/510	nein	nein	nein

Tabelle 4: Übersicht der DISKO-Annotationsspuren bezogen auf die Aufgabenstellung

Name der Spur	Beschreibung	DISKO _L2	DISKO _L1	DISKO _DSH	DISKO _WebTestDaF
handwriting	Annotationen zu Auffälligkeiten des handschriftlichen Texts auf Wortebene, Silbentrennung und Wörter, die keine eindeutige Lesart ermöglichen	ja	ja	ja	nein
macro	gestalterische bzw. strukturierende Informationen wie Zwischenüberschriften, Listenzeichen, Absätze und Hervorhebungen; allgemeine Kommentare, die den gesamten Text betreffen, sind in den Metadaten unter der Variable <i>text_annotatorComment</i> zu finden.	ja	ja	ja	nein

Tabelle 5: Übersicht der DISKO-Annotationsspuren zu Merkmalen der handschriftlichen Texte

Name der Spur	Beschreibung	DISKO _L2	DISKO _L1	DISKO _DSH	DISKO _WebTestDaF
text ¹	Tag, das den Export der Volltexte vereinfacht	ja	ja	ja	ja

Tabelle 6: DISKO-Annotationsspur text

¹ Die Spur *text* beinhaltet keine Annotation im eigentlichen Sinne: Ein alle Token des Transkripts umfassendes Tag (*text*) markiert den gesamten Text, um den Export vollständiger Texte aus ANNIS zu vereinfachen. Dieser ist in Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.** genauer erläutert.