

## MIKO: Übersicht Spurstruktur

Im Folgenden wird ein Überblick über die einzelnen Transkriptions- und Annotationsspuren der Vorlesungstranskripte gegeben. Das Kürzel *X* wird in allen Spuren durch die Sprechersigle der/des jeweiligen Dozierenden ersetzt (Metadatenvariable *speaker\_acronym*).

### Transkriptionsspuren<sup>1</sup>

Kürzel	Beschreibung
X[v]	Transkription der verbalen Äußerungen und Pausen der/des Dozierenden
Assist[v]	Transkription verbaler Äußerungen und Pausen des Assistenten sowie ausgewählter relevanter außersprachlicher Handlungen
NN[v]	Anonymisierte Transkription verbaler Äußerungen oder außersprachlicher Handlungen aus dem Auditorium

*Tabelle 1: Transkriptionsspuren*

### Manuelle Annotationsspuren

Kürzel	Beschreibung
X[akz]	Annotation besonderer Betonungen
X[word]	Annotationen auf Wortebene (Versprecher, Pausen und Verzögerungssignale innerhalb von Wörtern, Vorschläge bei schwer verständlichen Wörtern oder Passagen)
X[sprech]	Annotation besonderer Sprechweisen
X[meta]	Annotation sprachlicher Erläuterungen
X[nv]	Annotation nonverbaler Handlungen
X[nv2]	Annotation parallel stattfindender nonverbaler Handlungen
X[bild]	Annotation von Text-Bild-Bezügen, d. h. zeitliche Bezüge zwischen Verbalem und auf Medien sichtbaren Begriffen
X[komm]	Anmerkungen zum Transkript, freie Beschreibung von nicht festgelegten Phänomenen

*Tabelle 2: Manuelle Annotationsspuren*

Die tokenisierten Versionen der Vorlesungstranskripte, die im Langzeitarchiv zur Verfügung stehen, enthalten darüber hinaus automatische Transkriptions- und Annotationsspuren.

---

<sup>1</sup> Es wurden je Vorlesungstranskript nur die Spuren verwendet, in denen mindestens ein Phänomen transkribiert bzw. annotiert werden konnte. Daher enthält nicht jedes Transkript alle Spuren.

### Automatische Transkriptions- und Annotationsspuren

Kürzel	Beschreibung
X[tok]	Tokenisierung der Verbalspur der/des Dozierenden $X[v]$ ohne Satzzeichentokenisierung
X[pos]	POS-Tagging
X[p-pos]	Angabe der Wahrscheinlichkeit für die Korrektheit des POS-Taggings
X[lemma]	Lemmatisierung
Assist[tok]	Tokenisierung der Verbalspur des Assistenten $Assist[v]$ ohne Satzzeichentokenisierung
Assist[pos]	POS-Tagging
Assist[p-pos]	Angabe der Wahrscheinlichkeit für die Korrektheit des POS-Taggings
Assist[lemma]	Lemmatisierung
NN[tok]	Tokenisierung der Auditoriumsspur $NN[v]$ ohne Satzzeichentokenisierung

*Tabelle 3: Automatische Transkriptions- und Annotationsspuren*

Da die Spur  $NN[v]$  des Auditoriums keine eigentlichen Verbaläußerungen enthält, stehen die Spuren *pos*, *p-pos* und *lemma* zu dieser Spur nicht zur Verfügung.