

# Annotation historischer Korpora

05/08/2024 04:02:50

[FAQ Article Print](#)

<b>Category:</b>	FAQ Clarin::Ressourcen	<b>Votes:</b>	0
<b>State:</b>	public (all)	<b>Result:</b>	0.00 %
<b>Language:</b>	de	<b>Last update:</b>	11:29:31 - 07/25/2016 (Europe/Berlin)

## Keywords

Historisch , Annotation , Korpuslinguistik

## Symptom (public)

Auf welche historischen Grammatiken bzw. Wörterbücher/Lexika wird als Ausgangspunkt bei der Erstellung von Annotationen historischer Texte zurückgegriffen?

Konkrete Suchanfrage: Grammatikalisierungsstand des Umlauts am Ende der mhd. Periode; Inwieweit kann auf verschiedene, in den Handschriften unterschiedlich realisierte grafische Kennzeichnungen des Umlauts zugegriffen werden? (Problem: nach Lachmann'scher Normalorthografie geglättete Texte als Basis mhd. Grammatiken).

## Problem (public)

## Solution (public)

Die Annotation historischer und diachroner Primärdaten, insbesondere zum Zweck der Analyse von Grammatikalisierungsphänomenen, stellt hohe Anforderungen an die zu verwendenden Annotationsschemata.

Standardisierte und weitgehend adaptierte Tagsets wie das Stuttgart-Tübingen Tagset ([1]STTS) basieren auf einer synchronen und gegenwartsbezogenen Beschreibung der Sprache. Eine Annotation mithilfe dieser Schemata würde also Hypothesen hinsichtlich des Grammatikalisierungsgrades vorwegnehmen. Es empfiehlt sich daher, allgemeine und weitgehend zeitstabile Kategorien bei der Annotation der Daten zu verwenden. Eine mögliche Lösung ist beispielsweise, nur übergeordnete Kategorien des etablierten Tagsets zu verwenden, sich also auf einem allgemeineren Level an ihnen zu orientieren und so die Möglichkeit zu einer vergleichenden Analyse historischer und gegenwartssprachlicher Ressourcen zu schaffen.

Eine Suche nach verschiedenen Realisierungen von Umlauten wiederum, ist mithilfe einer Volltextsuche unter Verwendung [2]Regulärer Ausdrücke ohne weiteres möglich, und erfordert daher keine Annotation. Bei stark normalisierten Transkriptionen, bei denen Informationen über verschiedene Realisierungen fehlen, ist eine solche Analyse hingegen nur möglich, wenn diese mit deren Facsimiles verknüpft werden, z. B., mithilfe des Text-Bild-Link-Editors, der Teil der Aufbereitungssoftware [3]TextGrid Laboratory ist.

[1] <http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html>

[2] [http://de.wikipedia.org/wiki/Regul%C3%A4rer\\_Ausdruck](http://de.wikipedia.org/wiki/Regul%C3%A4rer_Ausdruck)

[3] <http://www.textgrid.de>