

14. Meyer K. Wie sagt man in der Schweiz? Wörterbuch der schweizerischen Besonderheiten. Duden-Taschenbücher. – Mannheim / Zürich: Dudenverlag, 1989. – 380 S.

15. Panizzolo P. Die schweizerische Variante des Hochseutschen // Deutsche Dialektographie. – Bd. 108. – Marburg, Lahn: Elwert, 1982. – 92 S.

\*\*\*\*\*

**Bernd Pompino-Marschall**  
**Humboldt-Universität zu Berlin, Deutschland**

**DEUTSCHE DIALEKTAUFNAHMEN AUS DEM LAUTARCHIV DER  
HUMBOLDT-UNIVERSITÄT ZU BERLIN**  
**Ein linguistisches Datenbankprojekt**

*Представлений у статті пілотний проект має на меті перевести у цифровий формат і зберегти в базі даних звукового архіву Берлінського університету ім. Гумбольдта аудіозаписи німецьких діалектів, зібраних у 20-х роках минулого століття.*

**Ключові слова:** записи діалектів, база даних аудіозаписів, фонетика.

*Целью пилотного проекта, представленного в статье, является перевод в цифровой формат и перенос в базу данных звукового архива Берлинского университета имени Гумбольдта аудиозаписей немецких диалектов, собранных в 20-х годах прошлого века.*

**Ключевые слова:** записи диалектов, база данных аудиозаписей, фонетика.

*The paper describes a small pilot project to systematically transfer the audio material of German dialects collected in the 1920s as part of the 'Lautarchiv' of the Humboldt-Universität zu Berlin into a linguistically annotated audio database. It was first presented at the 'Long night of science' on June 2<sup>nd</sup> 2012.*

**Key words:** historical dialect recordings, annotated audio database, phonetics.

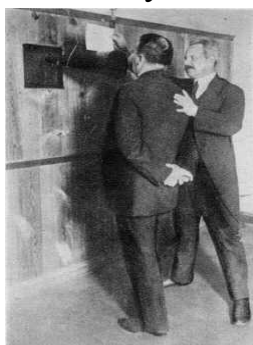
## **1. Einleitung**

Unser Projekt bildet einen Auftakt zur systematischen Erschließung der Sprachaufnahmen des Lautarchivs der Humboldt-Universität zu Berlin in Form einer linguistisch annotierten Datenbank zu Forschungs- und Lehrzwecken.

Die digitalisierten Sprachaufnahmen werden per Hand mittels PRAAT in Satz- und Wortäußerungen segmentiert, orthographisch und in phonetischer Umschrift annotiert und auf der Grundlage der orthographischen Wortannotation anschließend automatisch mit der phonologischen Repräsentation sowie der standardsprachlichen Aussprache versehen und nach Wortarten klassifiziert und zusammen mit den Metadaten (Dialekt, Sprecher, Aufnahmezeiten etc.) in eine Datenbank eingespeist. Hiermit ist die Grundlage für eine Datenabfrage nach linguistischen Kriterien und auch der gezielten Suche nach einzelnen Lautvarianten und das wort- und satzweise Abhören der jeweiligen Fundstellen sowie deren Extraktion zur weiteren Analyse geschaffen.

## **2. Zur Geschichte des Lautarchivs**

Das Lautarchiv der Humboldt-Universität zu Berlin wurde am 1. April 1920 als Lautabteilung an der Preußischen Staatsbibliothek gegründet. Initiator und erster Direktor war der Gymnasiallehrer und Sprachwissenschaftler Wilhelm Doegen.



Wilhelm Doegen  
(1877–1967)

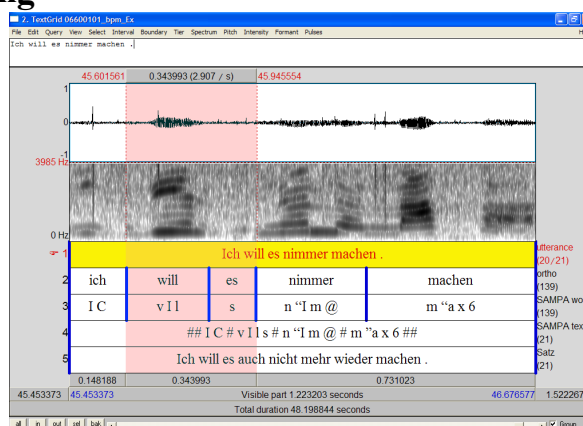
hier als Aufnahmeleiter mit einem Sprecher vor dem Schalltrichter

Ab 1922 begann die Lautabteilung mit Aufnahmen zur umfassenden Dokumentation deutscher Mundarten. Grundlage vieler dieser Aufnahmen bildeten die sogenannten Wenkersätze, die in der deutschen Mundartforschung seit der zweiten Hälfte des 19. Jahrhunderts zur systematischen Datenerhebung verwendet wurden. Es handelt sich dabei um insgesamt 40, von Georg Wenker zusammengestellte, hochsprachliche Sätze, die von Sprechern lokaler Dialekte in ihre jeweiligen Mundarten zu “übersetzen” waren und die so ausgewählt waren, dass die typischen lautlichen und grammatischen Merkmale der einzelnen Dialekte dabei hervortreten mussten.

Hier die für unsere Datenbankdemonstration verwendeten ersten zehn dieser Sätze:

1. Im Winter fliegen die trockenen Blätter in der Luft herum.
2. Es hört gleich auf zu schneien, dann wird das Wetter wieder besser.
3. Tu Kohlen in den Ofen, damit die Milch bald zu kochen anfängt.
4. Der gute alte Mann ist mit dem Pferd(e) auf dem Eis eingebrochen und in das kalte Wasser gefallen.
5. Er ist vor vier oder sechs Wochen gestorben.
6. Das Feuer war zu heiß, die Kuchen sind ja unten ganz schwarz gebrannt.
7. Er isst die Eier immer ohne Salz und Pfeffer.
8. Die Füße tun mir (so sehr) weh, ich glaube, ich habe sie (mir) durchgelaufen.
9. Ich bin selber bei der Frau gewesen und habe es ihr gesagt, und sie sagte, sie wolle es auch ihrer Tochter sagen.
10. Ich will es auch nicht mehr wieder tun/machen.

### 3. Datenaufbereitung



Obiges Bild zeigt den Ausschnitt des 10. Wenkersatzes einer schlesischen Sprecherin mit der Wortannotation in Orthographie und phonetischer Umschrift (SAMPA; unten wiederholt in der Zuordnung der Worte zur Transkription in IPA, wie sie sich aus der Textebene ergibt).

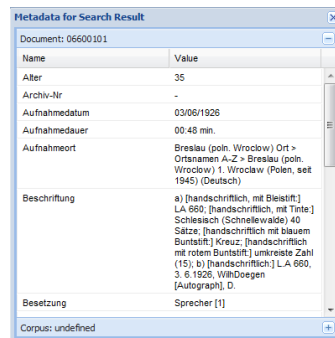
10. Ich will es nimmer machen .  
 [ ɪç vɪls nɪmə maxŋ ]

### 4. Implementation im Datenbanksystem ANNIS

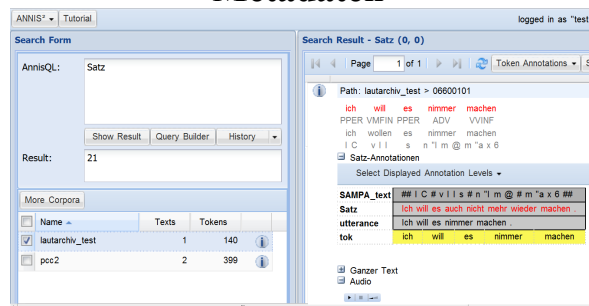
Die annotierten Daten werden schließlich zusammen mit den zu den Einzelaufnahmen vorliegenden Metadaten aus der Sammlungsdatenbank der Universität (SAM) in das Datenbanksystem ANNIS überführt. Dabei werden die orthographischen Wortdaten automatisch nach Wortarten klassifiziert (sog. POS-tagging) und lemmatisiert, sowie über ein Lexikon jeweils mit ihrer standardsprachlichen phonologischen Form und ihrer kanonischen Aussprache annotiert.

orthogr.	phonolog.	kanonisch
ich	ɪç	ɪç
will	vɪl	vɪl
es	ɛs	ɛs
nimmer	nɪmər	nɪmə
machen	maxŋ	maxŋ

## Lexikon



## Metadaten



## Datenbankdarstellung

Hiermit ist durch Verknüpfung der unterschiedlichen Annotationsebenen die Möglichkeit für Datenbank-Abfragen gegeben – z.B. nach:

- Einzelworten
- Wortarten
- Einzellauten in bestimmter Umgebung
- Vokalen und deren dialektaler Realisierung etc.

Ebenso besteht damit die Möglichkeit des selektiven Abhörens bzw. der Extraktion von Äußerungsabschnitten zur weitergehenden phonetischen Analyse.

## LITERATUR

1. Bayer, Kirsten/Mahrenholz, Jürgen-K.: "Stimmen der Völker" – Das Berliner Lautarchiv. In: Theater der Natur und Kunst (Katalog), hgg. v. Horst Bredekamp, Jochen Brüning und Cornelia Weber. Berlin: Henschel, 2000. – S. 117–128.
2. Mehnert, Dieter: Historische Schallaufnahmen – Das Lautarchiv an der Humboldt-Universität zu Berlin. In: Elektronische Sprachsignalverarbeitung – Siebente Konferenz Berlin. Studententexte zur Sprachkommunikation, Heft 13, 1996. – S. 28–45.
3. Simon, Artur: Das Berliner Phonogramm-Archiv 1900-2000. Berlin: VWB, 2000.
4. Ziegler, Susanne: Die akustischen Sammlungen – Historische Tondokumente im Phonogramm-Archiv und im Lautarchiv. In: Theater der Natur und Kunst (Essays), hgg. v. Horst Bredekamp, Jochen Brüning und Cornelia Weber. Berlin: Henschel, 2000. – S. 197-206.
5. Webpage des Lautarchivs <http://publicus.culture.hu-berlin.de/lautarchiv/>
6. ANNIS-Webpage: <https://korpling.german.hu-berlin.de/annis3/>

\*\*\*\*\*