

## Sprache – Ton – Bild Codierung gesprochener Sprache

### 1. Jahrestagung des Zentrums Sprachwissenschaften, Bild- und Tondokumentation (SBT)

#### Kurzfassungen | Abstracts

**Reinhard Bachmaier | Ulrike Kramer**  
WBÖ&DBÖ | DINAMLEX, ÖAW

Teuthonista versus IPA – Vergleich zweier  
Transkriptionssysteme

Das Wörterbuch der bairischen Mundarten in Österreich (WBÖ) verwendet für seine phonetischen Transkriptionen das System Teuthonista, das eine lange Tradition im Bereich der hochdeutschen Dialektologie aufweist. Ausgehend von der täglichen Beschäftigung mit der Lautschrift Teuthonista und konfrontiert mit der starken Präsenz der IPA in vielen linguistischen Teilbereichen drängte es sich auf, einen Vergleich zwischen beiden Systemen anzustellen.

Berücksichtigt werden die unterschiedlichen Entstehungskontexte und vor allem die Anwendungsbereiche und Leistungsspektren. Abseits vorhandener Stereotypen geht es darum, die in der Literatur bereits angerissene Diskussion erneut aufzugreifen, Differenzen wie Gemeinsamkeiten beider Systeme zu verdeutlichen und – last but not least – eine Lücke zu schließen, die in der Literatur zur Teuthonista augenscheinlich ist: eine systematisierte Übersicht über das Lautinventar der Teuthonista zu geben, das als Pendant zum bekannten IPA-Chart verstanden werden kann.

Diese Übersicht entspricht, der Praxis folgend, den Konventionen des WBÖ und erhebt somit nicht den Anspruch, eine alleingültige Norm darzustellen. Sie will die Wichtigkeit einer möglichen Vereinheitlichung der Teuthonista, wie sie für die IPA bereits erfolgt ist, aufzeigen. Über die Webpage des Instituts Dinamlex der Österreichischen Akademie der Wissenschaften wird sie der interessierten Öffentlichkeit in Form von PDF-Dateien zugänglich gemacht und unter anderem auch über die Seite des Instituts für Germanistik der Universität Wien erreichbar sein.

---

**Hubert Bergmann |  
Barbara Piringer | Eveline Wandl-Vogt**  
WBÖ&DBÖ | DINAMLEX, ÖAW

Dialekt – Notation und Interpretation im Umfeld des  
*Wörterbuchs der bairischen Mundarten in Österreich*

Das *Wörterbuch der bairischen Mundarten in Österreich (WBÖ)* ist ein Dialektwörterbuch, das auf der Verschriftlichung gesprochener Sprache beruht. Notation und Interpretation gehen dabei Hand in Hand. In den Sammlungen der Kartothek, des so genannten Hauptkatalogs, begegnen uns individuelle und normierte Notationen. Das Transkriptionssystem des WBÖ wiederum ist ein weitgehend standardisiertes. Betreffs der phonetischen Umschrift wird ein System basierend auf Teuthonista verwendet (vgl. Beitrag von Reinhard Bachmaier | Ulrike Kramer). In der *Datenbank der bairischen Mundarten in Österreich (DBÖ)* werden die Notationssysteme des Hauptkatalogs originalgetreu übernommen; durch eine Standardisierung (Stichwort) wird wie im WBÖ der Zugriff gesichert. Auch der Lemmaansatz mit seiner etymologisch-historischen Aussagekraft kann als standardisiertes Notationssystem angesehen werden. Eine Form von Codierung stellen schließlich auch Abbildungen dar.

Das Belegmaterial der WBÖ-Belegzettelkartothek setzt sich aus drei wesentlichen Teilen zusammen: aus über Fragebogen indirekt gesammeltem Material, aus Belegen, die auf Kundfahrten direkt erhoben wurden sowie aus Exzerpten. Während das von geschulten Fachkräften aus dem Umkreis der Wörterbuchkanzlei direkt erhobene Material hinsichtlich der verwendeten Notation relativ einheitlich zu nennen ist, treffen wir bei den beiden anderen erwähnten Segmenten auf eine große Vielfalt unterschiedlichster Transkriptionsweisen. So stammen etwa einerseits die Exzerpte zu einem nicht unerheblichen Teil aus Fachliteratur, die sich unterschiedlicher Transkriptionen bediente, während die Sammlerinnen und Sammler vor Ort beim Ausfüllen der Fragebogen nur

bedingt der 1913 gedruckten *Belehrung für die Sammler des Bayrisch-österreichischen Wortschatzes* folgten. Eine nicht unwesentliche Herausforderung für die Autorinnen und Autoren des seit 1963 erscheinenden *WBÖ* bildet deshalb die Interpretation der Belegografie und ihre Transkription in die *WBÖ*-Lautschrift. Quellenkundliche Nachforschungen zu den einzelnen Sammlerinnen und Sammlern, wie sie an der Arbeitsstelle schon seit einiger Zeit betrieben werden, können hier interessante Aufschlüsse liefern. Neue Möglichkeiten einer raschen und übersichtlichen Vergleichbarkeit von Belegen hinsichtlich individuellen Notationsverhaltens bietet die digitale Verfügbarkeit des Materials durch dessen Eingabe in die *DBÖ*.

Zu den wichtigsten Kriterien bei der Umsetzung der Belege aus dem Hauptkatalog gehört es dabei, die Belege möglichst originalgetreu abzubilden. Wo dies aus technischen Gründen nicht möglich ist, müssen normierte Ersatzdarstellungen verwendet werden. Zu diesem Zweck wurde institutsintern eine auf dem Programm TUSTEP basierende Codierung entwickelt.

Bei der Eingabe stößt man vor allem auf zwei Problemkreise: einerseits mangelnde Lesbarkeit der Belege, andererseits Umgang mit der Codierung unerwarteter individueller Notierungen. Die Lesbarkeit hängt nicht allein von der Deutlichkeit der Schrift ab, sondern auch davon, welche verwendet wurde. Die handschriftlichen Belege aus dem Hauptkatalog sind zum größten Teil in Kurrentschrift verfasst. Auch verwenden einige Sammlerinnen und Sammler bei Bedeutungsangaben Kurzschrift, meist die heute nicht mehr gebräuchliche nach Gabelsberger.

Neue Medien und digitale Publikationsformen bedingen neue Fragestellungen und Herausforderungen. Der im Rahmen des Projekts *Datenbank der bairischen Mundarten in Österreich electronically mapped (dbo@ema)* entwickelte und kosten- und registrierungsfrei zur Verfügung gestellte Font *dinamlex* ist eine um rund 300 Sonderzeichen erweiterte Schrift und erleichtert die Darstellung spezifischer Lautphänomene und (historischer) Verschriftlichungen. Die Präsentation wird durch einen Ausblick auf den Webauftritt unter dem Blickwinkel der Sonderzeichenrecherche abgeschlossen.

---

**Verena Blaschitz**  
BEST | I DINAMLEX, ÖAW

Erhebung, Analyse und Auswertung von kindlichen Sprachdaten im FWF-Projekt „Bildungserfolg bei Sprachtod?“

Eines der Teilziele des FWF-Projekts „B.E.S.T.“ ist die Untersuchung und in weiterer Folge die Einschätzung von Aspekten bildungserfolgsrelevanter Sprachkompetenz von ca. 200 neun- bis zehnjährigen Kindern mit Migrationshintergrund in Wien.

Im Beitrag werden die Aufgabenstellungen sowie erste Analyse- und Auswertungsschritte der Sprachdaten präsentiert: Die schriftliche Aufgabe ist eine in beiden Sprachen zu absolvierende Handlungsablaufsbeschreibung eines Films, die mündliche Aufgabe kann als „Erzählen auf Anforderung – in zwei Sprachen“ paraphrasiert werden.

Durch eine explorative, heuristische Zugangsweise wurden linguistische Analyse- und Auswertungsmöglichkeiten entwickelt, die Aussagen über wesentliche Aspekte bildungssprachlich relevanter kindlicher Sprachkompetenz erlauben soll. Wenngleich damit kein Testen und Prüfen im herkömmlichen Sinn möglich ist, ist es dazu gedacht, Sprachdaten linguistisch plausibel argumentiert einzustufen.

---

Mit der samtenen Revolution in Tschechien öffneten sich die Grenzen und der Zugang zu den ehemals deutschsprachigen Gebieten wurde zunehmend leichter. Seit Beginn der 90er Jahre lenkten Dialektologen vermehrt ihre Aufmerksamkeit auf eine der spannendsten Mundartregionen im deutschsprachigen Raum. Nach einigen Aufnahmen, die von den angrenzenden Sprachatlasunternehmen aus Bayern und Österreich mit ihren Fragebüchern durchgeführt wurden und nachdem man sah, dass noch ausreichend geeignete Mundartsprecher in diesen Gebieten leben, entschied man sich, den gesamten ehemals deutschsprachigen Raum zu erheben. Es wurde unter widrigen Umständen ein internationales Forschungsprojekt ins Leben gerufen, an dem tschechische, österreichische und deutsche Wissenschaftler mitarbeiten.

Nach langer Diskussion zu Beginn des Projektes entschied man sich, die Aufnahmen in Teuthonista und nicht in IPA zu transkribieren.

Die internationale Dimension des Projekts hat auch Auswirkungen auf die Aus- und Weiterbildung der Exploratoren und der laufenden Koordination ihrer Arbeit. Nur durch intensive Schulung und fortlaufenden Austausch über die konkrete Arbeit im Feld, kann und konnte ein zuverlässiger Standard in der Verschriftlichung der Mundartbelege über Länder- und Mundartgrenzen hinweg realisiert werden. Ein weiteres Mittel zur Verifizierung der Verschriftlichung ist der Tonbandmitschnitt. Standardmäßig wurden von allen Befragungen Tonaufnahmen angefertigt, die eine zusätzliche Kontrolle der einzelnen Mitschriften ermöglichen.

Die erhobenen Daten werden in Adetes verschriftlicht, einem Zeichensatz, der sich an den am Sprachatlas von Mittelfranken entwickelten anschließt. Dieses Sprachmaterial wird so aufbereitet, dass eine computergestützte Kartierung möglich wird. Alle Fragebücher werden eingescannt und sollen so der wissenschaftlichen Öffentlichkeit zugänglich gemacht werden.

In der germanistischen Varietäten- und Areallinguistik sind mit IPA und Teuthonista zwei unterschiedliche Lautschriftsysteme im Einsatz, deren Vor- und Nachteile ausführlich diskutiert wurden und hinreichend bekannt sind. Für den Vergleich corpusbasierter Publikationen und Forschungsprojekte (Wörterbücher, Sprachatlanten usw.) stellt die Verwendung unterschiedlicher Kodierungssysteme an sich aber bereits ein signifikantes Hindernis dar.

Bei der weitgehend auf den oberdeutschen Sprachraum beschränkten Verwendung von Teuthonista kommt erschwerend hinzu, dass es trotz mancher Ansätze nie zu einer wirklichen Standardisierung gekommen ist. Dieser Mangel bedingt nicht nur Mehrdeutigkeit und Redundanz zahlreicher phonetischer Symbole, sondern hat in verschiedenen Forschungstraditionen letztlich auch zur Ausbildung unterschiedlicher Subtypen dieser Lautschrift geführt.

Für das Projekt „Sprachatlas von Oberösterreich“, das mit mehreren divergierenden Teuthonista-Traditionen zu kämpfen hat, lag daher der Einsatz einer lautschriftunabhängigen Kodierung nahe. Dabei werden nicht die phonetischen Symbole eines bestimmten Transkriptionssystems kodiert, sondern die phonetischen Realisate an sich. Grundlage der Kodierung sind dabei klassische phonetische Parameter wie Artikulationsart und -ort, Qualität und Quantität, Stimmhaftigkeit, Rundung, Nasalierung, aber auch einige suprasegmentale Eigenschaften.

Die Kodierung der Belege erfolgt mittels EDV in einer formalen Sprache, deren syntaktische Struktur in einem offenen Standard (EBNF) definiert ist und zugleich erweiterbar ist. Neben der maschinellen semantischen Verwertung erlaubt diese Kodierung auch die Umwandlung von Belegen (Konvertierung in anderen Kode, Darstellung in IPA, Teuthonista usw.), statistische Funktionen (Berechnung von Indizes) sowie die komplexe Suche nach phonologischen und morphologischen Strukturen.

Der Einsatz einer standardisierten Abfragesprache („Phonological Query Language“) ermöglicht darüber hinaus die kombinierte Suche und das Vergleichen von Belegen auf Basis linguistischer Kriterien und zusätzlicher Metadaten (Gewährspersonen, Erhebungszeitraum und -ort, Fragebuchstellen etc.).

---

Dieser Vortrag präsentiert Kodierungs- und Analysemöglichkeiten von gesprochener Sprache innerhalb des CHILDES-Systems. CHILDES steht für Child Language Data Exchange System und ist die für die Kindersprache entwickelte Komponente des TalkBank-Systems [<http://talkbank.org/>]. CHILDES wurde 1984 von den amerikanischen Psycholinguisten Brian MacWhinney und Catherine Snow gegründet und zählt mittlerweile zu den führenden Methodologien in der internationalen Erstspracherwerbsforschung. CHILDES besteht aus 3 Teilen: 1) einem standardisierten computerbasierten Format zur Transkription und Kodierung kindersprachlicher Daten (CHAT), 2) einem Computerprogrammpaket zur Analyse des kodierten Sprachmaterials (CLAN) und 3) einer allgemein zugänglichen, sich laufend erweiternden Kindersprachdatenbank (CHILDES DATABASE).

Nach einem kurzen Überblick über die derzeit in Bearbeitung befindlichen Kindersprachdaten der Kommission zeigen wir anhand von Beispielen grundlegende Schritte der Verschriftung, Kodierung und Analyse mit CHILDES. Hauptaugenmerk wird auf eine semi-automatische, lexikon-basierte morphologische Kodierungsmöglichkeit gelegt, welche mit Hilfe der Programmkomponente MOR erfolgt. Abschließend präsentieren wir einige Analyseprogramme von CHILDES (z.B. *FREQ* zur Berechnung von Type- und Tokenfrequenzen), mit denen MOR-kodierte Sprachdaten quantitativ ausgewertet werden können.

Für die Erforschung der Dialekte wird primär ein sensibles und trotzdem sehr gut handhabbares Instrumentarium benötigt, um die feinen Nuancen der regionalen Lautungen notieren zu können. Traditionell wird hierzu im oberdeutschen Sprachraum die Verschriftung durch das System Teuthonista verwendet, welche zum Großteil auf dem lateinischen Alphabet fußt und darüber hinaus ermöglicht, unterschiedliche lautliche Abstufungen mit Hilfe von Sonderzeichen (Diakritika) wiederzugeben. Somit geschieht praktisch bereits bei der Erhebung im Feld eine Kodierung entsprechender akustischer Äußerungen in eine optische Variante, abgebildet mit den graphischen Elementen einer Lautschrift.

Mein Vortrag soll die sekundären Weiterverarbeitungsmethoden (Kartierung in Atlasbänden, akustischer ‚Sprechender Sprachatlas‘) des gewonnenen Dialektmaterials beschreiben und zwar exemplarisch am ‚Sprachatlas von Niederbayern‘.

Doch nicht nur auf abgeschlossene, sondern auch auf ganz aktuelle Projekte wie die Ortsnamenforschung an der Universität Passau und die dabei nicht zu unterschätzende Rolle der Teuthonista soll näher eingegangen werden.

---