

**Hauke Bartels / Kamil Thorquindt-Stumpf**

## **Ein neues Ton- und Textarchiv des muttersprachlich-dialektalen Niedersorbischen**

### **1. Die Bedeutung des muttersprachlich-dialektalen Niedersorbischen für die Sprachdokumentation**

#### **1.1. Zur Notwendigkeit eines neuen Korpus**

Es ist bekannt, dass die jüngsten noch aktiven Sprecher des Niedersorbischen in seiner primären Form, den Dialekten, mittlerweile über 70 Jahre alt sind (JODLBAUER ET AL. 2001). Das Sprachgebiet ist gegenüber dem häufig beschriebenen Stand von Mitte des 20. Jahrhunderts<sup>1</sup> weiter erheblich geschrumpft, und die unmittelbar nach dem Zweiten Weltkrieg südlich von Cottbus zumindest von der älteren Generation noch gesprochenen Dialekte gelten mittlerweile als ausgestorben. Das niedersorbische Dialektkontinuum ist de facto nicht mehr existent, das Siedlungsgebiet nicht mehr kompakt. Damit geht einher, dass die traditionellen kommunikativen Netze weitgehend zerrissen sind und daher die Anwendbarkeit der Sprache im Alltag kaum mehr gegeben ist. Das Niedersorbische in dieser Form ist in der Tat eine der am stärksten bedrohten Sprachen Europas.

In dieser Situation ist eine gründliche Dokumentation nicht nur der niedersorbischen Schriftsprache, sondern auch des heute noch gesprochenen dialektalen Niedersorbischen notwendig, da sich beide Varietäten in vielerlei Hinsicht unterscheiden und eine Dokumentation und Beschreibung der niedersorbischen Sprache ohne Einbeziehung der Dialekte unvollständig bleiben müsste. Solange sich die Gelegenheit noch bietet, sollen daher möglichst umfassende Informationen zum gegenwärtigen dialektalen Sprachgebrauch gesammelt und verfügbar gemacht werden. Damit wird nicht nur die Grundlage für zukünftige Forschungen deutlich erweitert, sondern es kann auch die materielle Basis der laufenden Bemühungen um eine Revitalisierung des Niedersorbischen insgesamt ausgebaut werden, indem im Projekt gesammelte Daten zum Beispiel Informationen zu bislang unterrepräsentierten Bereichen des Wortschatzes oder zum Gebrauch des verbalen Aspekts und der Aktionsarten liefern können.

In Anbetracht der oben grob skizzierten Situation des Niedersorbischen und der aktuellen bzw. in naher Zukunft zu erwartenden Konsequenzen für sprachwissenschaftliche Forschung kommt einer gut aufbereiteten Datenbasis für zukünftige Untersuchungen zum Niedersorbischen eine entscheidende Bedeutung zu. Insbesondere ein umfangreiches Textkorpus wird unverzichtbar sein. Dabei war stets klar, dass die langjährige und den Umständen geschuldete bisherige Beschränkung auf gedrucktes Schrifttum eine Unzulänglichkeit darstellte, da dadurch ein wesentlicher Teil der niedersorbischen Sprache unberücksichtigt blieb. Dieser Mangel soll nun, soweit im Rahmen des hier beschriebenen Vorhabens möglich, behoben werden.

---

<sup>1</sup> Vgl. zum Beispiel FASSKE 1998: 18–22 und FASSKE 2003; dies freilich aus Mangel an verlässlichen aktuellen Daten.

## 1.2. Bisherige Maßnahmen zur Dokumentation des Niedersorbischen

Das hier vorgestellte Forschungsvorhaben ist eingebettet in eine Reihe weiterer (bereits abgeschlossener, laufender oder geplanter) Maßnahmen der Cottbuser Zweigstelle des Sorbischen Instituts zur Erstellung einer solchen breiten Datenbasis und damit zur langfristige möglichst umfassenden Dokumentation des Niedersorbischen. Die entsprechenden Projekte werden im Folgenden nur kurz umrissen; ein ausführlicher Überblick findet sich in BARTELS 2012b.

- (a) **Niedersorbisch-deutsches Wörterbuch** – Bereits 1999 erschien nach mehrjähriger Arbeit an der Cottbuser Zweigstelle das „Niedersorbisch-deutsche Wörterbuch“ von Manfred Starosta, das primär als Hilfsmittel zum Textverstehen konzipiert wurde.
- (b) **Digitales Textkorpus der niedersorbischen Schriftsprache** – Das bald nach Gründung der niedersorbischen Zweigstelle 1992 in Angriff genommene Textkorpus konnte in den letzten Jahren dank umfangreicher Drittmittel<sup>2</sup> für Digitalisierungsmaßnahmen stark ausgebaut werden, sodass mittlerweile eine Größe von ca. 25 Mio. laufenden Wortformen (tokens) erreicht ist. Das Korpus repräsentiert die niedersorbische Schriftsprache von den frühesten Zeugnissen (1. Hälfte des 16. Jahrhunderts) bis in die Gegenwart. Schon jetzt ist ein wesentlicher Teil des niedersorbischen Schrifttums erfasst. Ziel ist jedoch ein Vollkorpus, das sämtliche noch erreichbaren Texte in digitaler Form enthält. Teile des Korpus sind im Internet verfügbar.<sup>3</sup>
- (c) **Deutsch-niedersorbisches Wörterbuch (DNW)** – Das seit 2001 entstehende, derzeit bereits weit fortgeschrittene Deutsch-niedersorbische Wörterbuch (DNW) ist makro- wie mikrostrukturell das bislang umfangreichste niedersorbische Wörterbuch (mehr als 80.000 Stichwörter mit ca. 50.000 Beispielsätzen). Es ist als aktives Wörterbuch konzipiert und hat das Ziel, den speziellen Anforderungen, die sich aus der schwierigen soziolinguistischen Situation des Niedersorbischen ergeben, durch eine möglichst genaue und besondere Schwierigkeiten berücksichtigende Beschreibung des Wortschatzes gerecht zu werden (BARTELS 2010).
- (d) **Digitalisierung niedersorbisch-deutscher Wörterbücher** – In diesem 2011 abgeschlossenen Projekt wurden die wichtigsten niedersorbisch-deutschen Wörterbücher (ZWAHR 1847; MUCKE 1911–15/1926, 1928; ŠWJELA 1961, STAROSTA 1999) nicht nur digitalisiert, sondern zugleich im XML-Format neu modelliert. Auf dieser Grundlage erstellte Internet-Versionen wurden Anfang 2012 bereitgestellt.<sup>4</sup>
- (e) **Weiterbearbeitung der Datenbasis aus den digitalisierten Wörterbüchern** – Die in (d) entstandene Datenbasis wird derzeit in zwei Anschlussprojekten weiterbearbeitet und ausgewertet. Zum einen wird der Datenbestand so erweitert, dass möglichst viele Informationen auch über deutsche Suchbegriffe zugänglich werden. Zum anderen enthalten die digitalisierten Wörterbücher und das ohnehin digitale DNW eine große Menge von idiomatischen Ausdrücken und erläuternden Informationen.

<sup>2</sup> Die Digitalisierung niedersorbischen Schrifttums wurde in den Jahren 2006 bis 2011 zunächst von der Stiftung für das sorbische Volk und der Vattenfall Europe Mining AG gemeinsam (2006–2007), in der Folge dann im Rahmen einer Vereinbarung zwischen Vattenfall und der Domowina – Bund Lausitzer Sorben e. V. von Vattenfall allein finanziell gefördert (2008–2011). Auch die unter (d) und (e) genannten Projekte wurden im Rahmen der genannten Vereinbarung von Vattenfall finanziert.

<sup>3</sup> Das Textkorpus sowie alle übrigen im Internet bereitgestellten Ressourcen sind über die Adresse [www.niedersorbisch.de](http://www.niedersorbisch.de) zugänglich.

<sup>4</sup> Das Projekt wurde in groben Zügen in BARTELS 2012a und SZCZEPAŃSKI 2012 beschrieben.

Diese derzeit teilweise nur schwer auffindbaren Wendungen werden systematisch erfasst und so aufbereitet, dass ein mehrfacher und sicherer Zugriff ermöglicht wird. Zugleich entsteht auf diese Weise die Grundlage für eine spätere wissenschaftliche lexikografische Bearbeitung und damit für ein niedersorbisches phraseologisches Wörterbuch.

- (f) **SorbLex Niedersorbisch** – Alle o. g. Vorhaben sind auch als Vorarbeiten zu einem umfassenden historisch-dokumentierenden Wortschatz-Informationssystem der niedersorbischen Sprache zu verstehen. Ziel von *SorbLex* ist eine annähernd vollständige lexikografische Beschreibung der heute noch erreichbaren niedersorbischen Lexik vom Beginn der schriftlichen Überlieferung im 16. Jahrhundert bis zur Gegenwart. Das geplante Vorhaben wird an anderer Stelle näher beschrieben (BARTELS 2013).

Ergänzend ist hier das GENIE-Projekt (GESprochenes NIEDersorbisch) der Universität Saarbrücken zu nennen, in dem ältere Tonaufnahmen aus den Beständen des Sorbischen Kulturarchivs im Sorbischen Institut, eine Auswahl von niedersorbischen „O-Tönen“ aus dem Archiv des Sorbischen Rundfunks sowie Neuaufnahmen aus den Jahren 2005–2006 zu einem reinen Audio-Korpus zusammengestellt wurden (es liegen keine Transkriptionen usw. vor).<sup>5</sup> Ursprünglich war geplant, mit Blick auf die im hier vorgestellten Projekt verfolgten Ziele einen Teil der in GENIE bereitgestellten Daten sowie weitere Aufnahmen aus dem Rundfunk-Archiv einzubeziehen, zu transkribieren und damit gemeinsam über die Verschriftung zugänglich zu machen. Dies lässt sich aber zurzeit nicht realisieren und muss daher Folgeprojekten vorbehalten bleiben.

## 2. Das Forschungsprojekt „Text and audio corpus of native Lower Sorbian“

### 2.1. Projektziel und Gegenstand der Dokumentation

Ziel des Vorhabens ist der Aufbau eines bisher fehlenden und für Sprachdokumentation und Forschung dringend notwendigen, auf Transkriptionen von Audioaufnahmen basierenden Textkorpus des muttersprachlich-dialektalen Niedersorbischen. Das auf insgesamt etwa drei bis vier Jahre angelegte Projekt läuft seit Oktober 2011 und wird im Förderprogramm „Dokumentation bedrohter Sprachen“ (DoBeS)<sup>6</sup> von der Volkswagen-Stiftung finanziert. Der Begriff *T e x t* korpus, wie er im englischsprachigen Titel für die

---

<sup>5</sup> Die Ergebnisse des Vorhabens sind im Internet zugänglich: <http://genie.coli.uni-saarland.de/> (letzter Zugriff: 6. 2. 2013).

<sup>6</sup> Im Jahr 2000 startete die VolkswagenStiftung ihre Förderinitiative „Dokumentation bedrohter Sprachen“ (DoBeS), deren Ziel es ist, „möglichst viele in ihrer Existenz bedrohte Sprachen so weit aufzuzeichnen, dass spätere Generationen anhand des dokumentierten Materials noch die ganze Sprache beschreiben können“. (DOBES-BROSCHÜRE 2004, 3) Das Programm, in dessen Rahmen bisher 60 Forscherteams etwa 80 vom Aussterben bedrohte oder bereits ausgestorbene Sprachen dokumentiert haben (KLEIN 2012, 11 f; vgl. auch HAIG ET AL. 2011), ist mittlerweile ausgelaufen; es werden nur noch bereits bewilligte Projekte zu Ende geführt (vgl. <http://www.volkswagenstiftung.de/foerderung/beendet/bedrohtesprachen.html>; letzter Zugriff: 5. 2. 2013). Niedersorbisch ist neben Minderico (Portugal) und einer Gruppe von Sami-Sprachen in Russland eine der wenigen europäischen Sprachen, die in das Förderprogramm aufgenommen wurden.

einjährige Pilotphase<sup>7</sup> verwendet wurde, bezeichnete dabei den für zukünftige Untersuchungen zum Niedersorbischen wichtigen Zugang über die Textform (vgl. 2.5.) zwar zutreffend, betonte diesen Aspekt aber zu stark, sodass für das Anschlussvorhaben der Titel modifiziert wurde (Text and audio corpus of native Lower Sorbian). Durch das Projekt entsteht ein Text- und Tonarchiv aus Sprachaufnahmen mitsamt den zugehörigen Transkriptionen und Übersetzungen.

Gegenstand der Dokumentation ist das heute noch von älteren Sprechern in mündlicher Kommunikation verwendete dialektale Niedersorbische. Es handelt sich damit um die Sprache der heute in der Regel über 80 Jahre<sup>8</sup> alten muttersprachlichen Dialektsprecher, die in der ersten Hälfte des 20. Jahrhunderts als Kinder das Niedersorbische noch als Erstsprache erworben und verwendet haben (informell auch „Volksprache“, niedersorbisch „ludowa rěc“ genannt).<sup>9</sup> Die Kennzeichnung „muttersprachlich“ bzw. „native“ (im englischen Projekttitel) bezeichnet eben jenen Unterschied zwischen der Projekt-Zielgruppe und der heute im öffentlichen Leben dominierenden Mehrheit von Sprechern des Niedersorbischen, deren Sprachkompetenz nicht mehr einem ungesteuerten Erstspracherwerb entstammt.

Das Vorhaben versteht sich nicht als eine Fortsetzung oder Ergänzung des Sorbischen Sprachatlas (SSA). Ziel des Projekts ist nicht die Beschreibung sorbischer Dialekte. Der Begriff *dialektal* ist zunächst nur insofern von Bedeutung, als wir davon ausgehen, dass die für das Projekt definierte Informanten-Zielgruppe (vgl. 2.3.) das Niedersorbische als Erstsprache auf der jeweils vor Ort gegebenen dialektalen Basis erlernt hat.<sup>10</sup>

Das im Sorbischen Sprachatlas beschriebene Sprachgebiet ist trotz seines begrenzten Umfangs bekanntlich dialektal relativ stark differenziert. Es werden für das Niedersorbische sechs Dialekte unterschieden: Neben dem (a) Cottbuser Dialekt (mit mehreren Varianten) sind dies (b) der Vetschauer Dialekt, (c) der nordwestliche Dialekt mit Burg, Schmogrow und Fehrow, (d) der nordöstliche Dialekt um Peitz mit Drachhausen bis Jänschwalde, (e) die Mundart von Horno sowie (f) der Spremberger Dialekt (FASSE 2003: 49). Allerdings basiert der SSA auf in den 1960er-Jahren erhobenen Daten von Informanten, die das 60. Lebensjahr in der Regel überschritten hatten (SSA, Bd. 1, S. 9), sodass die dortige Beschreibung de facto die Sprachkompetenz einer um oder in den Jahrzehnten vor 1900 geborenen Sprechergeneration zum Gegenstand hat. Heute gelten mehrere der o. g. Dialekte als nahezu<sup>11</sup> ausgestorben, sodass für die jetzt laufende

<sup>7</sup> <http://www.volkswagenstiftung.de/foerderung/beendet/bedrohtesprachen/bewilligungen-2011.html> (letzter Zugriff: 24. 1. 2013).

<sup>8</sup> Nur wenige Informanten sind jünger, mindestens jedoch um die 70 Jahre alt.

<sup>9</sup> Diese Definition schließt nicht aus, dass auch Informanten in das Projekt einbezogen werden, die ein solches Niedersorbisch sprechen, obwohl sie es erst später als Zweitsprache, zum Beispiel im Grundschulalter, in der Sprachgemeinschaft ungesteuert erworben haben. Daneben gibt es natürlich auch Erwerbssituationen, bei denen man von zwei Erstsprachen (Deutsch/Niedersorbisch bzw. Niedersorbisch/Deutsch) ausgehen kann (vgl. WOLFF 2003: 835).

<sup>10</sup> Die Schriftsprache wird von diesen Sprechern in der Regel nur passiv rezipiert, nur eine kleine Minderheit beherrscht die Schriftsprache auch aktiv.

<sup>11</sup> Es liegen hierzu keine zuverlässigen Daten vor. Für die Vetschauer Mundart zum Beispiel schreibt FASSE (1964: 8), gestützt auf Daten spätestens aus den Jahren 1960/61: „Der Rückgang des Anteils der sorbischsprechenden Bevölkerung wirkt sich u.a. darin aus, daß sich nur noch die älteste Generation im Verkehr miteinander des Sorbischen bedient, die jüngere Generation versteht noch z.Teil [sic] das Sorbische. Die Kinder sind weder aktiv noch passiv des Sorbischen mächtig. Nur in Naundorf und Fleißdorf beherrscht auch die mittlere Gene-

Datenerhebung Informanten vorwiegend aus dem Cottbuser, dem nordwestlichen, vor allem aber aus dem nordöstlichen Dialektgebiet zu erwarten sind (vgl. Abb. 1).

Die Frage, inwieweit die im Rahmen des Projekts aufgenommenen Sprecher als Repräsentanten eines dieser Dialekte gelten können, ist offen. Es liegen keine Untersuchungen zur heutigen sprachlichen Qualität in Dörfern vor, die nach dem SSA einem bestimmten Dialektgebiet zuzuordnen sind. Es stellt sich daher die Frage, inwieweit in Dörfern, in denen traditionelle kommunikative Netze teilweise bereits seit Jahrzehnten zerrissen sind, der vom jeweiligen (ortsansässigen) Interviewpartner realisierte Idiolekt noch als für den lokalen Dialekt bzw. die örtliche Mundart repräsentativ zu werten ist. Dies werden ggf. Untersuchungen zeigen können, welche die im Projekt gesammelten Sprachdaten auswerten. Die bisherigen Informanten stammen aus folgenden Geburtsorten: Bärenbrück/Barbuk, Burg-Dorf/Wobsedne Bórkowy, Dissen/Dešno, Döbbrick/Depsk (2 Informanten), Drachhausen/Hochoza (4), Drehnow/Drjenow (3), Drewitz/Drjejece, Fehrow/Prjawoz, Heinersbrück/Móst, Jänschwalde/Janšojce (6), Maiberg/Majberk, Milkersdorf/Górnej, Saspow/Zaspy, Sielow/Žyłow (3), Skadow/Škódow, Striesow/Strjažow, Turnow (2); für die Übergangsmundarten des Schleifer Dialekts<sup>12</sup>: Rohne/Rowne (2), Schleife/Slěpe.

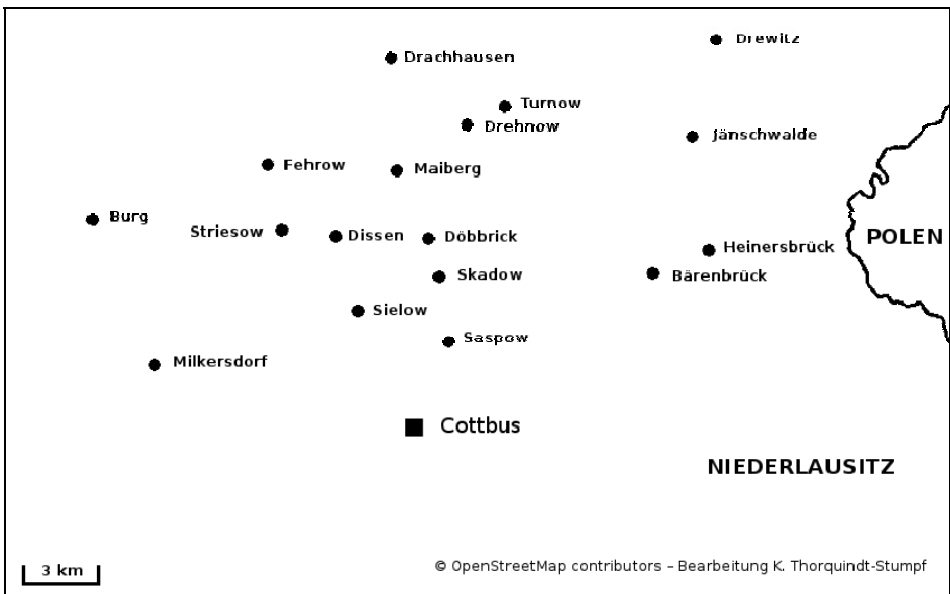


Abb. 1: Geburtsorte bisheriger Gesprächspartner in der Niederlausitz

Bei der Erfassung der Metadaten zu den Interviews (Name, Alter, Geschlecht, Hofzugehörigkeit<sup>13</sup>, Familiensprache, Ausbildung und Beruf, Aufnahmezeiten usw.) werden

ration (30 Jahre und älter) die im Ort gesprochene Mundart.“ Demnach wäre es für die beiden genannten Dörfer theoretisch noch möglich, Sprecher zu finden.

<sup>12</sup> Für die Dörfer des Schleifer Dialektgebiets werden die dialektalen Ortsnamen angegeben, die teilweise von den nach dem Sächsischen Sorbengesetz gültigen obersorbisierten Formen abweichen.

<sup>13</sup> Ein bis heute in den niedersorbischen Dorfgemeinschaften verwendeter Zuname, vergeben „nach den bewirtschafteten Höfen“ (JAHN 2010: 69).

für eine spätere dialektologische Bewertung, aber auch zur Klärung möglicher sprachlicher Besonderheiten wichtige Informationen wie Geburtsort, Orte längerer Aufenthalts usw. erhoben. Diese können eine sprachliche Analyse der Daten ergänzen und so ggf. nachträglich eine dialektale Zuordnung ermöglichen. Auf Grundlage der aktuellen Sprachsituation und der damit verbundenen allgemein geringen Verfügbarkeit muttersprachlicher Informanten (vgl. 2.3., mehr dazu in BARTELS 2012b) sowie der für die Auswahl geeigneter Interviewpartner notwendigen Kriterien wird zunächst versucht, so viele Gesprächspartner wie möglich zu gewinnen. Dabei soll der räumliche Einzugsbereich so weit gefasst werden, dass sich eine größtmögliche dialektale Varianz ergibt. Sofern die Verfügbarkeit von Informanten es zulässt, soll in einer späteren Phase des Projekts bei der Auswahl der Gesprächspartner verstärkt darauf geachtet werden, dass alle der noch lebendigen Ortsmundarten bzw. die verschiedenen niedersorbischen Dialekte repräsentiert sind.

Mit obiger Gegenstandsbestimmung wird zugleich deutlich, dass es im Projekt nicht um das erst sekundär und überwiegend an Schulen erworbene Niedersorbisch geht, das heute (neben Deutsch) u. a. in sorbischen Institutionen, auf öffentlichen Veranstaltungen und in den Medien als Kommunikationsmittel dient. Diese (heute auch mündlich realisierte) Schriftsprache hat in der Regel – ausgenommen sind hier diejenigen Dialektsprecher, die sekundär auch Kenntnisse der niedersorbischen Schriftsprache erworben haben – den Charakter einer mehr oder weniger gut beherrschten Zweit- oder Drittsprache, meist mit Deutsch, seltener mit Obersorbisch als Erstsprache. Im Niedersorbischen existiert – anders als für das Obersorbische angenommen (SCHOLZE 2008) – keine regionale Umgangssprache, die als dritte Sprachform neben den Dialekten und der niedersorbischen Schriftsprache (mit ihrer mündlichen Variante) anzusiedeln wäre.

Während der Pilotphase wurde angesichts neuer Informationen über die dort sehr kritische Sprachsituation entschieden, neben dem eigentlichen niedersorbischen Sprachgebiet auch das Schleifer Kirchspiel in das Dokumentationsprojekt einzubeziehen, das im Norden an das südlich von Cottbus gelegene ehemalige niedersorbische Dialektgebiet, im Süden an das obersorbische Sprachgebiet grenzt. Die in den zugehörigen Dörfern Groß Düben/Džewin, Halbendorf/Brjezowka, Mulkwitz/Mulkoce, Mühlrose/Miłoraz, Rohne/Rowne, Schleife/Slěpe und Trebendorf/Trjebin gesprochenen Mundarten bilden den Schleifer Übergangsdialekt, der dem Niedersorbischen nahesteht (SSA Bd. 10, 396 ff.). Nach unseren Informationen stehen dort mittlerweile nur noch wenige Muttersprachler der ältesten Generation für Aufnahmen zur Verfügung, sodass eine spätere Dokumentation des dialektalen Sorbischen dieses Gebiets nicht mehr möglich sein wird.<sup>14</sup> Es scheint daher geboten, das noch Mögliche im Rahmen unseres Do-BeS-Projekts zu tun. Bisher liegen Aufnahmen mit drei Informanten aus diesem Gebiet vor.<sup>15</sup>

<sup>14</sup> Schon BRIJNEN (2004: 11) wies vor fast zehn Jahren darauf hin, dass der Schleifer Dialekt „nur noch in einigen Dörfern gesprochen [wird], und zwar nur noch von alten Leuten“. Es heißt weiter: „Wer noch richtig Schleifersorbisch [gemeint ist das dialektale Sorbische der Gegend im Unterschied etwa zum in der Schule gelernten Obersorbischen] spricht, ist unter den Leuten bekannt. Es sind hauptsächlich Frauen, die meisten zwischen 75 und 95 Jahre alt.“ (S. 15)

<sup>15</sup> Es soll außerdem geprüft werden, ob in den übrigen dem Niedersorbischen nahestehenden Übergangsdialekten, d. h. dem Muskauer Dialekt, dem sog. mittleren Grenzdialekt sowie ggf. der Mundart von Großkoschen, noch einzelne Sprecher zu finden sind. Zumindest für die beiden letztgenannten scheint dies nicht ausgeschlossen zu sein.

## 2.2. Datenbasis

Das Ton- und Textarchiv des muttersprachlich-dialektalen Niedersorbischen wird aus Audioaufnahmen (in einigen Fällen auch Videoaufnahmen), den zugehörigen Transkriptionen sowie deren Übersetzungen ins Deutsche bestehen. Für einen repräsentativen Teil der Aufnahmen werden darüber hinaus eine Übersetzung ins Englische sowie eine phonetische Transkription angefertigt. Während der Pilotphase (Oktober 2011 bis Februar 2013) lag der Arbeitsschwerpunkt wegen der sich zunehmend verschlechternden Verfügbarkeit von Informanten auf der Erstellung von Neuaufnahmen. Derzeit (Stand: Februar 2013) liegen insgesamt schon 54 Stunden Audio- und sieben Stunden Videoaufnahmen mit 34 Informanten vor. Die Aufnahmen werden überwiegend von Exploratoren durchgeführt, die über gute Kontakte zur Zielgruppe verfügen und damit in der Lage sind, mögliche Sprechbarrieren zu überwinden.<sup>16</sup> Außerdem standen anfangs die Erprobung von Arbeitsabläufen bei Transkription und technischer Weiterbearbeitung der Aufnahmen, später die möglichst effektive Herstellung von Transkriptionen und Übersetzungen auf dem Programm. In der nun folgenden Projektlaufzeit bis etwa Mitte 2015 wird die zeitlich sehr aufwendige Erstellung der Transkriptionen im Mittelpunkt stehen. Es sollen aber auch weiter Tonaufnahmen gemacht werden, um die Datenbasis zu erweitern.

Neben den im Projekt dominierenden Einzelinterviews werden auch Mehr-Personen-Interviews geführt, etwa in den Konstellationen bzw. Situationen Ehepaar, Verwandte, Nachbarn, Seniorenheim. Die bisherigen Aufnahmen mit mehreren Informanten weisen eine relativ hohe sprachliche Qualität auf, die zum Beispiel in einer hohen Anzahl unterschiedlicher verwendeter Lexeme (Type-Token-Relation) zum Ausdruck kommt.

Bei der Herstellung neuer Sprachaufnahmen wird neben normalen Interviews von Projektmitarbeitern mit Informanten in sog. „Kind-Eltern-Gesprächen“<sup>17</sup> eine neue Methode zur Datengewinnung erprobt. In der Niederlausitz ist häufig eine Konstellation anzutreffen, in der ein (nahezu) voll kompetenter Sprecher der ältesten Generation, der nicht mehr oder nur noch selten in der Öffentlichkeit auftritt und/oder in der Öffentlichkeit bzw. mit Fremden kein Niedersorbisch (mehr) spricht, Kinder hat, die häufig als Semi-Sprecher (GRINEWALD/BERT 2011) einzustufen sind. Die Kinder selbst stehen noch in der Öffentlichkeit und haben Interesse daran, die Sprache ihrer Eltern dokumentiert zu sehen, kommen aber selbst wegen ihrer begrenzten oder nicht der Zielgruppe entsprechenden Sprachkompetenz als Informanten nur sehr eingeschränkt in Frage. Aufgrund der Nahbeziehung zum voll kompetenten Sprecher und ihrer noch vorhandenen kommunikativen Fähigkeiten in Niedersorbisch sind sie aber in der Lage, ein Interview mit ihren Eltern zu führen. Diese „indirekte“ Interviewtechnik kann möglicherweise zusätzliche, ansonsten nicht zu gewinnende Datenquellen erschließen.

## 2.3. Auswahl der Sprecher

Die durch die derzeitige Sprachsituation begründete Erwartung, dass der Sprechertyp „dialektal geprägter Muttersprachler“ bald nicht mehr zur Verfügung steht, führte zur

---

<sup>16</sup> Als Exploratoren sind bisher tätig: Maria Elikowska-Winkler, Gregor Wiczorek, Juliana Kaufmüller (Region Schleife), Marcus Koinzer. Vgl. dazu auch 2.4.1.

<sup>17</sup> Der Begriff „Kind-Eltern-Gespräch“ steht für den erwarteten typischen Fall; ggf. können auch andere nahestehende Personen (jüngere Geschwister, Enkelkinder usw.) die Funktion des „Kindes“ übernehmen.

Bestimmung des Gegenstands unseres Dokumentationsvorhabens (vgl. 2.1.). Die weitgehende Beschränkung auf diesen Sprechertyp ergibt sich allein aus der Dringlichkeit, der eine Dokumentation der Sprache dieser Gruppe unterliegt. Es ist durchaus sinnvoll und wünschenswert, auch neuere Formen des gesprochenen Niedersorbischen, d. h. Sprachformen einer jüngeren Generation von Sprechern aufzuzeichnen und zu beschreiben und damit eine weitere Ergänzung der Datenbasis vorzunehmen. Dies gilt umso mehr, als dadurch wieder andere Erscheinungen des Sprachwandels usw. ins Blickfeld kämen. Ein solches Vorhaben ist jedoch weniger dringlich und kann daher noch zurückgestellt werden.

So ist das erste Kriterium bei der Auswahl von Interviewpartnern die Zugehörigkeit zum genannten Sprechertyp (muttersprachlich-dialektale Sprachkompetenz). Dabei spielt keine Rolle, ob die Sprache des Informanten etwa mehr oder weniger Einflüsse des Deutschen zeigt, ob es häufig zu Code-Switching kommt oder ob, was bisher seltener der Fall war, sogar einige Passagen des Gesprächs auf Deutsch erfolgen. Entscheidend war allein, dass es möglich sein musste, ein Interview im Wesentlichen, d. h. zum überwiegenden Teil auf Niedersorbisch zu führen. Konkrete Grade der Beeinflussung durch das Deutsche, von Sprachmischung usw. zeigen sich ohnehin erst im Gesprächsverlauf. Hinsichtlich der Einbeziehung von Tonaufnahmen in das Korpus bzw. Archiv wird diesbezüglich keine Auswahl erfolgen.

Aus praktischen Gründen spielen bei der Suche nach Informanten allerdings zwei weitere Kriterien eine Rolle. Zum einen geht es um eine allgemeine Sprechbereitschaft. Es ist bekannt, dass manche Personen mit guter niedersorbischer Sprachkompetenz im Sinne der Projekt-Zielgruppe den aktiven Gebrauch der niedersorbischen Sprache entweder ganz aufgegeben oder auf engste persönliche Kontakte beschränkt haben. Solche Personen fallen de facto als Informanten aus, sofern es den für das Projekt tätigen Exploratoren nicht gelingt, diese Barrieren zu überwinden. Als weiteres Kriterium ist schließlich die Interviewtauglichkeit zu nennen, d. h. die Frage, inwieweit die Informanten nicht nur bereit waren, ein persönliches Gespräch auf Niedersorbisch zu führen, sondern dies auch noch „im Beisein“ eines Aufnahmeapparats bzw. eines Mikrofons zu tun.

Im obigen Sinne geeignete Informanten, in deren Interviews gute Anknüpfungspunkte für weitere interessante Gespräche erkennbar sind, können auch mehrmals besucht werden. In den Zweit- und Drittgesprächen werden Themen wieder aufgenommen, die während einer ersten Aufnahme aufgefallen sind, aber nur ansatzweise besprochen wurden. Dies ergibt sich oft aus interessanten biografischen Begebenheiten.

## 2.4. Methodik

### 2.4.1. Gesprächsgestaltung

Für Interviews, in denen nicht nur gezielt Informationen abgefragt werden, sondern eher ein Gespräch mit möglichst hohen Anteilen des Informanten angestrebt wird (biografisch-narrative Gesprächsführung), ist es in vielen Fällen nicht nur nötig, zu Beginn Eingangsimpulse zu setzen, sondern auch im Gesprächsverlauf ggf. weitere Impulse für eine Vertiefung oder thematische Richtungsänderung zu geben. Als Hilfestellung für die Exploratoren wurden zwei Fragenkataloge ausgearbeitet, die zur Orientierung und Vorbereitung auf die Gesprächsführung dienen, nicht jedoch zum „Abarbeiten“ gedacht sind.

Beim ersten Fragenkatalog handelt es sich um eine laufend ergänzte, thematisch geordnete Sammlung in den Interviews gestellter (ausformulierter) Fragen in niedersorbi-

scher Sprache, die vor allem für solche Interviewer zur Vorbereitung dienen kann, die nicht regelmäßig Niedersorbisch sprechen und daher unter Umständen Schwierigkeiten hätten, ein fließendes Gespräch auf Niedersorbisch zu führen (zum Beispiel in Kind-Eltern-Gesprächen). Hierzu zählen Standardfragen zu Name, Familie, Hochzeit, Werdegang, beruflichen Erfahrungen, Ruhestand oder Krankheiten sowie Fragen zu anderen Themen, die sich im Hinblick auf die Zielgruppe in der Regel eignen, so etwa zu Sprachgebrauch (Sprache, Schule, Kirche, Bibel), Dorfleben, Haus- und Gartenarbeiten, Landwirtschaft (Ackerbau und Tierzucht, Imkerei), Natur (Wald, Spreewald, Pilze), Bräuchen und Sitten (regionale und sorbische Feste, Feiertage, Aberglauben, Legenden), Hobbys (Handarbeit, Küche, Bücher, Geschichten), Flurnamen.

Der zweite Fragenkatalog umfasst thematisch geordnete Schlagwörter, die die inhaltliche Vorbereitung und Moderation der Interviews für erfahrene Exploratoren mit guten Niedersorbisch-Kenntnissen erleichtern können. Als Beispiel dient der Themenbereich Imkerei (pcołkarstwo): kak zachopiś, kak wuknuś, wjele ludow, kóše, matka, gelée royale, woploženje, kaženje, w zymje, w nalěšu, mjod, wósk, propolis, štapaś, parazy, pesticidy, chórosći, orientacija a komunikacija pcołkow.

#### **2.4.2. Weiterbearbeitung der Aufnahmen**

Die in der Regel von den Exploratoren angefertigten Rohaufnahmen werden katalogisiert und mit einer Registernummer versehen. Diese kodiert Explorator, Aufnahmenummer und -datum. Weitergehende Informationen werden in Datenblättern erfasst, die Namen der Informanten werden ansonsten nur in kodierter Form verwendet. Die Aufnahme-Sets (Audio- bzw. Videodatei und Datenblatt) werden mehrfach gesichert: lokal, auf dem Server der Zweigstelle Cottbus, im Internet (Online-Speicher) sowie, in einer weiter bearbeiteten Fassung, im DoBeS-Archiv (vgl. 2.5.). Der Online-Speicher dient zugleich dem Datenaustausch zwischen Projektmitarbeitern.

Die mehrstufige Nachbearbeitung einzelner Aufnahmen durch verschiedene Mitarbeiter wird zentral verwaltet:

1. Einlesen des Sprachmaterials in ELAN, Einbindung der Annotationszeilen, Anpassung der Codes nach konkreter Sprecherkonstellation (vgl. 2.4.3.)
2. Segmentierung der Audiospur zur Vorbereitung der Transkription: Die Audioaufnahme wird in kurze Segmente (ca. fünf Sekunden) gegliedert, wobei Sprechpausen sowie die Verständlichkeit des Segments berücksichtigt werden (vgl. Abb. 3 mit einzelnen Segmenten und der jeweiligen Übersetzung).
3. Erstellung der orthografischen Transkription
4. Anfertigung der Übersetzung ins Deutsche
5. Ergänzung der Topics, Durchführung von Tokenisierung und Anonymisierung (vgl. 2.4.4.)

Für einen Teil der Aufnahmen erfolgen weitere Schritte:

6. Erstellung der phonetischen Transkription im IPA
7. Anfertigung der Übersetzung ins Englische

Für alle Bearbeitungsschritte sind Kontrollen vorgesehen. Die Daten (auch Zwischenergebnisse) werden, um Metadaten (vgl. 2.4.5.) ergänzt, laufend ins DoBeS-Archiv übertragen.

### 2.4.3. Transkription

Im Transkriptionsverfahren wird, wie in DoBeS-Projekten üblich, die Annotationssoftware ELAN eingesetzt (Abb. 2).<sup>18</sup>

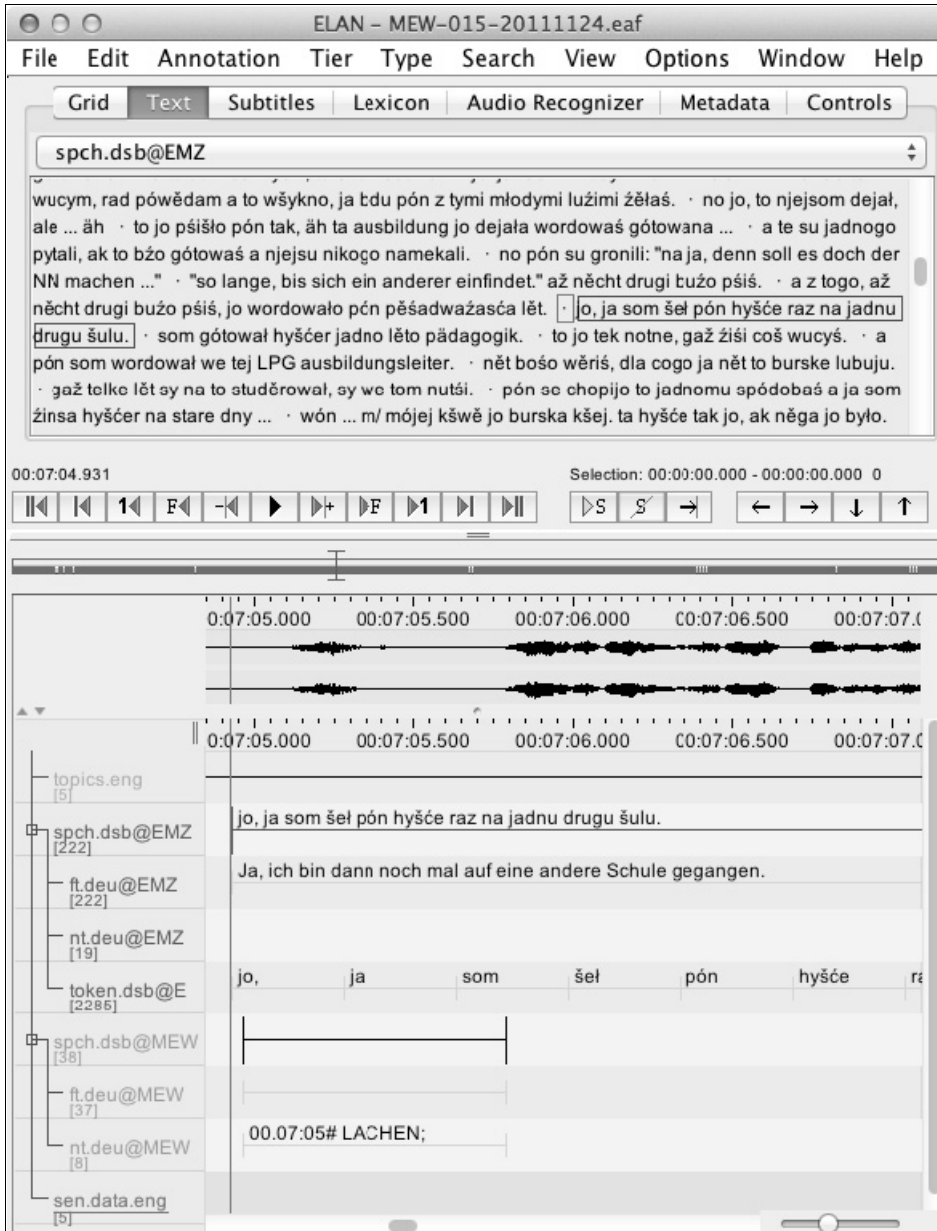


Abb. 2: Screenshot ELAN (s/w)

<sup>18</sup> <http://tla.mpi.nl/tools/tla-tools/elan/> (letzter Zugriff: 23. 1. 2013).

Für die Annotationen werden jedem Sprecher (in der Regel Informant und Explorator) jeweils mindestens drei Zeilen für folgende Informationen zugeordnet:

1. (in der Regel niedersorbischer) Sprechtext in orthografischer Transkription  
spch.dsb@XYZ<sup>19</sup> – speech (spch): Sprecherzeile
2. deutsche Übersetzung der Transkription  
ft.deu@XYZ – free translation (ft): freie Übersetzung ins Deutsche
3. Bemerkungen  
nt.deu@XYZ – notes (nt): Bemerkungen  
Diese Zeile wird u. a. für die Notation (nicht transkribierter, vgl. 2.4.3.1.6.)  
phonetischer Besonderheiten genutzt. Darüber hinaus werden zum Beispiel notiert:  
flexionsmorphologische und Wortbildungsfehler; nicht allgemein bekannte Realia  
und ortsspezifische Besonderheiten; Husten, Lachen und Störungen der Gesprächs-  
situation, die den Verlauf des Interviews beeinflussen.

Für den Informanten ist eine zusätzliche Zeile vorgesehen, in der die Transkription tokenisiert wird (token). Allen Sprechern gleichermaßen zugeordnet sind ergänzende Zeilen für „Topics“ (topics: Themen der Aufnahme auf Englisch) und für sensible Daten (sen.data.eng – sensitive data: eine „Arbeitszeile“ für spätere Anonymisierung der Aufnahme, der Transkription und der Übersetzung; vgl. 2.4.4.).

Für einen repräsentativen Teil des Datenbestandes werden außerdem eine phonetische Transkription im Internationalen Phonetischen Alphabet (IPA) sowie eine Übersetzung ins Englische angefertigt. Entsprechend sind folgende Zeilen zu ergänzen:

4. englische Übersetzung der Transkription (jeweils für Informant und Explorator)  
ft.eng@XYZ – free translation (ft): freie Übersetzung ins Englische
5. phonetische Transkription im IPA  
ipa@XYZ

#### 2.4.3.1. Verfahren der orthografischen Transkription

Die folgenden Ausführungen beziehen sich ausschließlich auf das orthografische Transkriptionsverfahren, das für den gesamten Datenbestand gilt. Die für einen Teil der Aufnahmen ebenfalls zu erstellende Transkription im Internationalen Phonetischen Alphabet (IPA) wird hier nicht thematisiert.

Die Erfahrungen verschiedener korpuslinguistischer Projekte bestätigen die Notwendigkeit einer möglichst exakten Festlegung der Transkriptionskonventionen schon vor Beginn dieses Arbeitsschritts; es wird generell empfohlen, eine Testphase einzulegen (CARSTENSEN ET AL. 2010: 194). Im hier vorgestellten Vorhaben wird weitgehend nach den Regeln der geltenden Standardorthografie transkribiert. Die vor Projektbeginn festgelegten Konventionen wurden weitgehend aus dem Dissertationsvorhaben von Kamil Stumpf übernommen, für das bereits ein kleineres, ebenfalls aus Transkriptionen von Audioaufnahmen bestehendes Korpus des dialektalen Niedersorbischen aufgebaut wurde.<sup>20</sup> Die dort erprobten Transkriptionsregeln wurden in der Pilotphase an einigen Stellen präzisiert. Bei der Ausarbeitung der Konventionen wurden ältere Transkriptio-

<sup>19</sup> XYZ steht für den jeweiligen Informanten- bzw. Mitarbeiter-Code; dsb (dolnoserbski), deu (deutsch) und eng (English) geben die jeweilige Sprache an (nach ISO 639-3).

<sup>20</sup> Für das Korpus wurde ein Teil der im GENIE-Projekt (vgl. Kap. 1.2.) bereitgestellten Audio-dateien transkribiert.

nen sorbischer Dialekttexte (FASSKE/MICHALK 1965, 1969) berücksichtigt. Das Transkriptionsverfahren selbst wie auch die punktuellen Abweichungen vom orthografischen Standard werden im Folgenden erläutert.

Die orthografische Transkription hält sich relativ streng an die geltenden Regeln des Sprachstandards. Auf die Notation phonetischer Merkmale konkreter Dialekte wird bewusst verzichtet; allerdings werden bestimmte (im Unterschied zur Schriftsprache) auffällige Merkmale des dialektalen Niedersorbischen insgesamt in der Transkription beibehalten, um den besonderen Charakter dieser Varietät in der Verschriftung zu bewahren. Außerdem werden zusätzlich sowohl Wortabbrüche, Verzögerungsphänomene (engl. *hesitation phenomena*) usw. mit einbezogen. Die für das Projekt geltenden Transkriptionskonventionen wurden nach drei Hauptquellen ausgearbeitet:

1. Orthografischer Standard des Niedersorbischen (STAROSTA 1982, 1999)
2. Regeln für die Transkription der Aufnahmen für das Korpus des gesprochenen Tschechischen (*Pravidla nahrávání a přepisu sond pro mluvený korpus češtiny – verze 2008*)<sup>21</sup>
3. Transkriptionskonventionen für das Korpus OS – Deutsche Mundarten: ehemalige deutsche Ostgebiete (Institut für Deutsche Sprache, Mannheim)<sup>22</sup>

Die Verwendung der orthografischen Transkription ermöglicht den Aufbau eines Textkorpus, das mit dem schriftsprachlichen Korpus vor allem mit Blick auf die Wortsuche „kompatibel“ ist und somit eine einheitliche Nutzung beider Korpora zulässt. Im Hinblick auf das Ziel, dennoch einige typische Eigenschaften des dialektalen Niedersorbischen in die Transkription zu übernehmen, spielen Aufwand und die unter den gegebenen Arbeitsbedingungen vorhandenen technischen und zeitlichen Möglichkeiten der Transkribenten eine wichtige Rolle. Insofern ist die Auswahl der in die Transkription übernommenen dialektalen Eigenschaften stark von pragmatischen Gründen geprägt: Nur solche „typisch dialektalen“ Merkmale, die regelmäßig ohne Hilfe von Geräten und ohne weitergehende Analysen eindeutig erkennbar sind, werden in die Transkription übernommen.

#### 2.4.3.1.1. Groß- und Kleinschreibung und Interpunktion

Die Groß- und Kleinschreibung sowie die Interpunktion orientieren sich weitgehend an den geltenden orthografischen Regeln (STAROSTA 1982 und 1999). Großgeschrieben werden grundsätzlich nur Eigennamen und, soweit den Regeln entsprechend, von ihnen abgeleitete Adjektive (STAROSTA 1982: §§ 181–185). Abkürzungen vom Typ „LPG“ oder „GmbH“ werden nach dem deutschen Vorbild übernommen. Kleingeschrieben wird in der Transkription nach dem Punkt, da dieser zwar, durch den Transkribenten eingefügt, eine deutliche Zäsur in der Rede markiert, nicht jedoch dieselbe Funktion zur Satzgrenzenmarkierung einnimmt wie in der geschriebenen Sprache.<sup>23</sup> Für Lehnwörter (Germanismen) gilt Kleinschreibung, auch im Falle des Code-Switchings. Die Schreib-

<sup>21</sup> [http://ucnk.ff.cuni.cz/oral/Prep\\_pravidla\\_2008.doc](http://ucnk.ff.cuni.cz/oral/Prep_pravidla_2008.doc) (letzter Zugriff: 16. 1. 2013); wir danken Martina Waclawičová für ihre freundliche Unterstützung.

<sup>22</sup> [http://dsav-oeff.ids-mannheim.de/DSAv/KORPORA/OS/OS\\_KONV.HTM](http://dsav-oeff.ids-mannheim.de/DSAv/KORPORA/OS/OS_KONV.HTM) (letzter Zugriff: 6. 1. 2013).

<sup>23</sup> Außerdem erleichtert diese Vorgehensweise spätere computerlinguistische Bearbeitungsschritte.

weise der Personennamen kann nicht immer zuverlässig überprüft werden, vor allem bei weniger geläufigen Namen von Personen, die nicht mehr am Leben sind. (Zur Anonymisierung der Personennamen vgl. Kap. 2.4.4.)

(1) Beispiel:

spch.dsb@XYZ      no, a pón tak. *ja* ga som äh žěłała LPG swinjecej groži, tak ak {Panicojc Mina}. *a* pón jo była hyšći {Lejnikojc Frida}.

#### 2.4.3.1.2. Direkte Rede

Direkte Rede tritt in Form erzählter Gespräche zwischen dem Informanten und einer dritten Person auf. Bis auf die konsequente Kleinschreibung nach dem Punkt als Zäsur-Marker werden alle orthografischen und Interpunktionsregeln eingehalten.

(2) Beispiel:

spch.dsb@XYZ      a ten Mato jo mjenjał: „*wy* *verfluchske bengele*, *co wy sćo mě* *nagotowali!*“

#### 2.4.3.1.3. Präpositionen

In den Dialekten werden Dativ und Präpositiv häufig präpositionslos verwendet, wo die Schriftsprache konsequent die Präpositionen *k* ‘zu’ und *w* ‘in’ fordert. In die Transkription werden nur tatsächlich realisierte Präpositionen aufgenommen.

(3) Beispiel:

spch.dsb@XYZ      a te ga su pśecej *měšće* byli. jo, a *w* *Janšojcach* tek.

#### 2.4.3.1.4. Lehnwörter

Die Transkription von Lehnwörtern, in den meisten Fällen von Germanismen, musste gesondert geregelt werden. In STAROSTA 1999 erfasste Lehnwörter werden in der dort notierten integrierten (sorabisierten) Form in die Transkription übernommen; bei dort nicht registrierten Formen wird auch Muckes „Słownik dolnosersbskeje rěcy a jeje narěcow“ (MUCKE 1911–15/1926, 1928) zur Orientierung herangezogen und dortige sorabisierte Formen werden ggf. für die Transkription an die geltende Orthografie angepasst.

Anders war bei bisher lexikografisch nicht erfassten Lehnwörtern zu verfahren. Einerseits handelt es sich um direkt übernommene deutsche Lexeme, andererseits um hybride Komposita und Derivate. Hier musste entschieden werden, wann nach der deutschen und wann nach der niedersorbischen Orthografie transkribiert wird. Um unnötige Varianten bei der Transkription zu vermeiden und um nicht die konkrete (mehr oder weniger deutsche bzw. niedersorbische) Aussprache zur Entscheidungsgrundlage machen zu müssen, was in der Praxis zu zahlreichen Zweifelsfällen führen würde, wird wie folgt verfahren:<sup>24</sup>

<sup>24</sup> Auf eine Einbeziehung des digitalen Textkorpus (vgl. 1.2. (b)) – einige der bisher nicht gebuchten Lehnwörter sind dort in unterschiedlicher Häufigkeit und in der Regel in unterschiedlichen Schreibungen vertreten – muss hier wegen des zu hohen zeitlichen Aufwands im laufenden Transkriptionsprozess verzichtet werden. Es wird stattdessen eine Liste der in der

- bei völlig fehlender<sup>25</sup> oder ausschließlich flexionsmorphologischer Integration eines deutschen Stammes wird dieser nach der deutschen Orthografie transkribiert: *ein-stellowaś, einziehowany, einbüßował, molkereija*
- bei Lehnwörtern mit niedersorbischen Präfixen wird das Wort vollständig nach der niedersorbischen Orthografie transkribiert: *wobštelowas, wuštafěrowas, wobštref-chowany*
- hybride Komposita werden getrennt nach den einzelnen Gliedern analysiert und transkribiert: *sommerkuchnja, haushaltungšula*.

#### 2.4.3.1.5. Nicht gebuchte dialektale lexikalische Formen

Bereits lexikografisch gebuchte indigene dialektale Lexik (*hyšćer, hyšće(n), telik* usw. nach STAROSTA 1999) wird, ähnlich wie bei den Lehnwörtern, entsprechend transkribiert. Bisher nicht gebuchte dialektale Formen wie *namekaś* (statt *namakaś*) oder *dyjko* (statt *dhujko*) werden in Anlehnung an die orthografischen Regeln transkribiert und zusätzlich in eine interne Liste aufgenommen, um ein einheitliches Vorgehen zu gewährleisten.

#### 2.4.3.1.6. Phonetische Phänomene in der orthografischen Transkription

Phonetisch vom Standard abweichende Realisierungen werden in die Transkription übernommen, wenn sie (a) für das dialektale Niedersorbische als sehr typisch betrachtet werden können und (b) vom Transkribenten ohne technische Unterstützung zuverlässig erkennbar sind, sodass eine konsistente Transkription gewährleistet werden kann.<sup>26</sup> Für nicht durchgängig gut erkennbare oder weniger typische Phänomene steht dem Transkribenten ggf. die Annotationszeile für Bemerkungen zur Verfügung.

##### 2.4.3.1.6.1. Die phonetische Realisierung der schriftsprachlichen Grapheme <o, ó>

Die Transkription von Lauten, die schriftsprachlich mit den Graphemen <o, ó> wiedergegeben werden, entspricht nicht den geltenden orthografischen Regeln. Es ist im Transkriptionsprozess problemlos möglich, die phonetische Realisierung entweder als o-Laut, d. h. als [ɔ] oder [o], oder aber als nicht-o-Laut, d. h. in der Regel als [i], [ɛ], [ʊ] akustisch zu unterscheiden, während die konkrete Realisierung bei den nicht-o-Lauten ohne technische Hilfe nicht immer eindeutig entscheidbar ist. Daher wird in der Transkription nicht die konkrete phonetische Realisierung des Lauts als [ɔ, o, i, ɛ, ʊ, ...] erfasst, sondern lediglich die Tatsache, ob ein o-Laut realisiert wird oder nicht. Dieses für die Dialekte typische Phänomen wird auf solche Weise zumindest teilweise erfasst und seine Untersuchung auf Basis der Transkripte erleichtert. Für einen o-Laut wird das Graphem <o> verwendet, für einen nicht-o-Laut das Graphem <ó>:

---

Transkription verwendeten Schreibungen geführt, die eine spätere Überarbeitung bzw. Systematisierung ermöglicht (vgl. etwa *emt* 'eben' in Bsp. (9)).

<sup>25</sup> Hier ist die Grenze zum Code-Switching schwer zu ziehen.

<sup>26</sup> Um die als typisch eingestuft (und gut erkennbaren) Eigenschaften in den Transkriptionen zu erhalten, wird eine (durch die Formvariation bedingte) erschwerte Suche im Korpus in Kauf genommen. Auch das schriftsprachliche Textkorpus zeichnet sich durch eine solche Formvariation aus, daher ist ohnehin eine durchgängige Lemmatisierung aller Korpora geplant.

<o ó> nach geltender Orthografie

→ <o>, wenn = [o ɔ]

→ <ó>, wenn ≠ [o ɔ], sondern zum Beispiel [i ε ʊ]

Diese Transkriptionsregel betrifft somit das aus der niedersorbischen Phonetik und Orthografie bekannte ó-Phänomen, ist aber nicht mit diesem identisch. Sie erfasst auch Ausdrücke vom Typ *sobota*, *sromota*, in denen das jeweils erste <o> als [ɔ], das jeweils zweite jedoch konstant als [ε] realisiert wird. Bei dessen Schreibung wird nach der geltenden Orthografie kein <ó> gesetzt, da es sich nach gängiger Lehrmeinung phonetisch um ein anderes Phänomen handelt (Realisierung von urslav. \*o nach Labial in nichtakzentuierter Silbe; vgl. SSA, Bd. 9, Karte 107) und demzufolge von den geltenden orthografischen Regeln für die Verwendung des <ó> nicht betroffen ist. Die beiden Grapheme <o, ó> haben somit in der Transkription eine andere Funktion als in der derzeitigen niedersorbischen Orthografie.

#### 2.4.3.1.6.2. „Stumme“ <w, ɸ>

Das sogenannte „stumme <w>“ in den Präfixen und Präpositionen *wu(-)*, *wo(-)* oder *wob(-)* wird nach der orthografischen Norm transkribiert. Die Schreibung mit <w> ersetzt seit der in der Niederlausitz umstrittenen Rechtschreibreform von 1952 die zuvor übliche Schreibung mit <h>. Dieses <h> stand – ebenso wie im Prinzip das heutige <w> – für einen „weichen“ Vokaleinsatz ohne den im Deutschen üblichen stimmlosen glottalen Plosiv (auch: Glottisschlag oder Knacklaut) und wurde phonetisch entweder nur mittelbar, d. h. durch das Fehlen des Plosivs oder auch als „leichter Hauchlaut“ realisiert, wobei der Übergang zwischen diesen Realisierungen fließend ist. Die konkrete phonetische Realisierung ist in vielen Fällen so undeutlich, dass im laufenden Transkriptionsprozess ohne Hilfsmittel nicht entschieden werden kann, ob ein gesonderter Laut realisiert wird und ggf. welcher Qualität dieser ist. Diese Regel betrifft in ähnlicher Form das Graphem <ɸ> in Wörtern wie *luka* oder *Lužyca*, wo sich die Realisierung zwischen weichem Vokaleinsatz und der in diesem Kontext normalen Aussprache als [ɸ] oder [w] bewegt. Die Kommentar-Zeile kann zur Kennzeichnung einzelner Beispiele genutzt werden.

(4) Beispiel:

spch.dsb@XYZ	tam ga jo była pjerwjej luka
nt.deu@XYZ	00:12:05# [ˈuka];

#### 2.4.3.1.6.3. Bilabial vs. lateral realisiertes <ɸ>

In der Transkription wird zwischen der lateralen [ɸ] und bilabialen [w] Realisierung des Phonems /ɸ/ (graphemisch <ɸ>) nicht unterschieden. Um eine genaue Unterscheidung einzuhalten, wäre es ebenso notwendig, die Opposition zum Phonem /l/ (graphemisch <ɺ>) und seine Varianten [lj] und [l] zu kennzeichnen. Es wird somit nach der geltenden Orthografie transkribiert, obwohl es sich um ein auffälliges Merkmal zur Unterscheidung verschiedener niedersorbischer Dialekte handelt.

#### 2.4.3.1.6.4. Reduktion

Die formale Reduktion der konjugierten Formen der Verben *móc* und *byś* (im Futur) wird in der Transkription wiedergegeben. Ohne Apostroph erscheinen folglich im Text-

korpus Formen wie *boš*, *bžoš*, *bdu*; *moš* oder *móš*. Diese Formen sind für das dialektale Niedersorbische sehr typisch und teilweise deutlich frequenter als die regulären Formen.<sup>27</sup> Bei *bžo* vs. *bužo* ist das Verhältnis etwa 1,6 : 1 ( $n = 558$ ), bei *moš*/*móš* vs. *móžoš*/*možoš* sogar 2,5 : 1 ( $n = 84$ ). Beim Futur von *byš* überwiegt dagegen die reguläre Form: 1 : 2,7 für *bdu* vs. *budu* ( $n = 238$ ).<sup>28</sup>

#### 2.4.3.1.6.5. Assimilation der Zischlaute und Fernassimilation

Die assimilierte Aussprache der ursprünglich nichtpalatalisierten Zischlaute wird nicht beachtet: *kakošć*, *šešćadwašća* statt zum Beispiel *kakošć*, *šešćadwašća*. Gerade bei den Zahlwörtern führt Assimilation zu einer großen Formvariation. Alle Zahlwörter werden jedoch in Übereinstimmung mit der Grundregel nach dem orthografischen Standard wiedergegeben. Hier gilt das oben genannte „pragmatische“ Prinzip: durch einfaches Zuhören ist es oft nicht möglich, sich in der Transkription für eine konkrete phonetische Realisation wie [ˈdwazast̪ea], [ˈdwazast̪ea], [ˈdwajast̪ea], [ˈdwaːst̪ea], [ˈdwast̪ea], [ˈdwaest̪ea] usw. zu entscheiden.

Die Fernassimilation wird beim Transkribieren dagegen berücksichtigt. Die Anzahl der entsprechenden auftretenden Formen ist relativ gering. Bisher wurden registriert *šlušaš*, *šlužabny*, *šlužba*, *šlužyš*, *šlyšaš*, *šotša*, *šuš*, *šušyš*, *žajtšo*, *žewš* / *žewšym* / *žowš* / *žowšy(m)*. Zwar wird die spätere Korpussuche durch die Kennzeichnung der Fernassimilation erschwert, es handelt sich aber um eine typische und in der Regel bei der Transkription zuverlässig erkennbare Spezifik, die in den Texten erhalten werden soll. Die Erfassung aller solcher Formen kann für die Korpussuche genutzt werden und den o. g. Nachteil kompensieren.

#### 2.4.3.1.7. Nicht-sprachliche Laute und unterbrochene Aussagen

##### 2.4.3.1.7.1. Verzögerungslaute, Rezeptionssignale, nicht-sprachliche Laute

Für Verzögerungsphänomene bzw. Rezeptionssignale wurden die Konventionen des Archivs für Gesprochenes Deutsch<sup>29</sup> am Institut für Deutsche Sprache in Mannheim (IDS) weitgehend übernommen. Anders als in den IDS-Regeln werden ein- und mehrgipflige Rezeptionssignale einheitlich als *hmm* notiert. Übereinstimmend erfolgt dagegen die Notation der Verzögerungslaute nur in zwei Varianten: *äh* für kurzes Zögern und *ähm* für längere gefüllte Pausen.

(5) Beispiel:

spch.dsb@MKR	ně, ale z twójeje jsy, to ja měnim.
spch.dsb@XYZ	äh z mójeje jsy, ně. we mójej jsy <i>ähm</i> stej bytej hyšći dvě drugej.

Nicht-sprachliches, wie Husten oder Lachen, wird in einer separaten Annotationszeile „notes“ notiert.

<sup>27</sup> Die Zahlen wurden auf Grundlage eines kleineren Korpus des dialektalen Niedersorbischen ermittelt, das von Kamil Stumpf im Rahmen seiner Dissertation zusammengestellt wurde (vgl. 2.4.3.1.). Für das DoBeS-Korpus liegen noch keine quantitativen Auswertungen vor, da es sich noch im Aufbau befindet.

<sup>28</sup> In den neuen Daten des DoBeS-Korpus findet sich darüber hinaus relativ häufig die Form *bud*.

<sup>29</sup> <http://agd.ids-mannheim.de/download/dida-trl.pdf> (letzter Zugriff: 17. 1. 2013).

(6) Beispiel:

spch.dsb@XYZ a te wšy ... te wšy su je kusali!  
nt.deu@XYZ 00:12:05# *LACHEN*;

2.4.3.1.7.2. Nicht abgeschlossene Aussagen und Wortabbrüche  
Wortabbrüche werden mit einem Schrägstrich ohne Leerzeichen notiert.

(7) Beispiel:

spch.dsb@XYZ a pón som hyšći kšěł *wulic/* äh groniš.

Pausen und nicht abgeschlossene bzw. wieder aufgenommene Aussagen werden durch Auslassungspunkte gekennzeichnet.

(8) Beispiel:

spch.dsb@XYZ a pón som pšišła do šule, a ako som ... ähm, ako som pón  
pšišła do ...  
spch.dsb@MKR žo sćo chójžili do šule?  
spch.dsb@XYZ do Janšojc ... jo, a pón Janšojcach.

2.4.3.1.8. Undeutliche Stellen und Bemerkungen

Trotz mehrfacher Kontrolle ist es nicht immer möglich, undeutliche Stellen zu entschlüsseln. Auf eine unzuverlässige Interpretation wird verzichtet. Stattdessen werden solche Passagen im Korpus in der Transkriptionszeile mit eckigen Klammern gekennzeichnet.

(9) Beispiel:

spch.dsb@XYZ no jo, a pón su se emt [...] na tom amše.

Vorschläge für eine mögliche Interpretation werden ggf. in einer gesonderten Annotationszeile „notes“ festgehalten.

(10) Beispiel:

spch.dsb@XYZ a co ga ty how tak [...]  
nt.deu@XYZ 00:52:23# *tšawkaš/tšawkaš?*

In den „notes“ werden ebenfalls Bemerkungen zu Realien (geografische, historische, kulturelle Besonderheiten) und zu anderen „Auffälligkeiten“ notiert.

(11) Beispiel:

spch.dsb@XYZ ta jo wuknuła tkalc, jo, we Chošebuzu TKC.  
nt.deu@XYZ 00:52:23# *TKC: Textil- und Konfektionsbetrieb Cottbus*

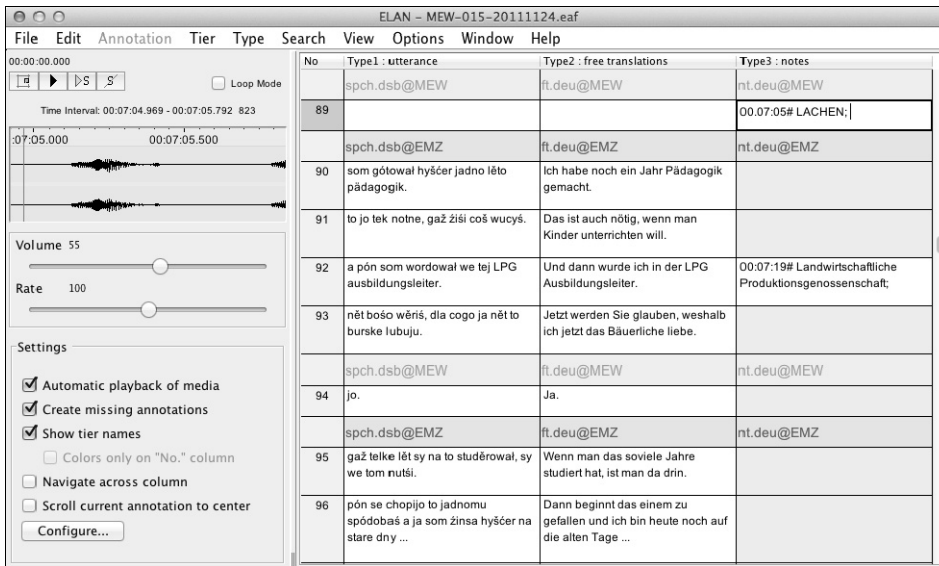


Abb. 3: Screenshot ELAN im Transkriptions-Modus (s/w)

#### 2.4.4. Anonymisierung

Personennamen und andere sensible Inhalte werden vor der Veröffentlichung aus Gründen des Datenschutzes und den DoBeS-Richtlinien<sup>30</sup> entsprechend anonymisiert. Während des Transkriptionsverfahrens werden entsprechende Stellen markiert und in einem späteren Bearbeitungsschritt im Text durch Platzhalter ersetzt: NN für Personennamen und SC für sensible Inhalte. In der Audiodatei werden die entsprechenden Sequenzen mit einem Pfeifton überblendet. Sowohl originale als auch anonymisierte Annotationen und Dateien werden archiviert. Noch nicht abschließend geklärt ist die diesbezügliche Behandlung von Videos.

(12) Beispiel [entspricht Bsp. (1); hier bereits anonymisiert]:

spch.dsb@XYZ           no, a pón tak. ja ga som äh žěłała LPG swinjecej groži,  
tak ak NN. a pón jo była hyšci NN.

#### 2.4.5. Datenformate und Technik

Bei Audioaufnahmen wird als Datenformat WAVE (48 KHz, 16 bit, linear PCM, Stereo), bei Videoaufnahmen MPEG-2 verwendet. Annotationen erfolgen im XML-Format (EAF) mit Metadaten im IMDI- und ELAN-Format.<sup>31</sup>

Neben Computern bzw. Notebooks zur Datenverarbeitung werden Aufnahmegeräte des Typs Olympus LS-11 eingesetzt. Für die Transkription finden außerdem Fußpedale Verwendung, u. a. um das Segmentierungsverfahren (vgl. 2.4.2.) zu beschleunigen.

<sup>30</sup> [http://www.mpi.nl/DOBES/ethical\\_legal\\_aspects/](http://www.mpi.nl/DOBES/ethical_legal_aspects/) (letzter Zugriff: 17. 1. 2013).

<sup>31</sup> Derzeit erfolgt die Umstellung auf ein neues Format CMDI = component-meta-data-infra-structure (DRUDE ET AL. 2012: 3266).

Als Software zur Transkription dient ELAN (vgl. 2.4.3.), zur Rechteverwaltung im DoBeS-Archiv das ebenfalls vom MPI bereitgestellte Access Management System (AMS), zur Strukturierung der Archivdaten der sog. Archive Builder.

## 2.5. Archivierung, Veröffentlichung und Zugänglichkeit

Die Ergebnisse des Projekts werden auf zweierlei Weise veröffentlicht. Zum einen werden die Sprachaufnahmen mitsamt den Transkriptionen und Übersetzungen in das DoBeS-Archiv am Max-Planck-Institut Nijmegen (Nimwegen) integriert, das der langfristigen und sicheren Aufbewahrung der Daten aus diversen Sprachdokumentationsprojekten und damit des Wissens über stark bedrohte oder bereits gar nicht mehr verwendete Sprachen dient [DoBES-BROSCHÜRE 2004, 10–15]. Es enthält in erster Linie Audio- und Videoaufnahmen, Fotos und diverse ergänzende Texte wie Transkriptionen, Übersetzungen, Annotationen, linguistische oder ethnografische Erläuterungen. Die Zugangsmöglichkeiten zu den Daten des DoBeS-Archivs unterliegen gewissen Beschränkungen, die mit der jeweiligen urheberrechtlichen Situation und verschiedenen Kulturen des Umgangs mit den Sprachdaten zusammenhängen. Für das DoBeS-Archiv wurde eine abgestufte Zugänglichkeit eingeführt, von freiem Zugang ohne oder mit Registrierung über einen genehmigungspflichtigen Zugang bis zu völliger Sperrung für die Öffentlichkeit. Obwohl generell eine möglichst weitgehende Verfügbarkeit der Daten des DoBeS-Archivs angestrebt wird, sind derzeit nur etwa 25 % der Bestände frei zugänglich, ggf. mit vorheriger Registrierung (DRUDE ET AL. 2012, 3266; KLEIN 2012, 13). Für die (anonymisierten, vgl. 2.4.4.) Daten dieses Projekts ist generell ein freier Zugang mit Registrierung vorgesehen. Das DoBeS-Archiv und mit ihm auch die Daten zum Niedersorbischen wird demnächst überführt in das Ende 2011 neu gegründete „The Language Archive“ (TLA), das von mehreren Institutionen gemeinsam getragen wird, aber ebenfalls am MPI in Nijmegen angesiedelt ist.<sup>32</sup> Es enthält schon heute Daten aus etwa 200 Sprachen, darunter auch die aus den DoBeS-Projekten (KLEIN 2012, DRUDE ET AL. 2012). Die Integration in das DoBeS-Archiv stellt die eigentliche Archivierung und Veröffentlichung des im Projekt erarbeiteten Materials dar und ermöglicht einen Zugriff primär über die Audiodateien.

Nicht primär über die Audiodateien, sondern zunächst in Form eines aus den Transkriptionen bestehenden Textkorpus werden die Ergebnisse des Projekts in das Informationsportal zum Niedersorbischen ([www.niedersorbisch.de](http://www.niedersorbisch.de)) eingebunden. Das Textkorpus des heutigen dialektalen Niedersorbischen wird somit das dort ebenfalls teilweise bereits zugängliche schriftsprachliche Korpus ergänzen und mit denselben Methoden nutzbar sein. Die Audiodateien sind hier über eine Verknüpfung mit den Transkriptionen erreichbar. Die Übersetzungen der niedersorbischen Transkriptionen ins Deutsche werden eine entsprechende Suche auch über deutsche Ausdrücke erlauben.

---

<sup>32</sup> <http://www.mpi.nl/tla> (letzter Zugriff: 24. 1. 2013); die tragenden Institutionen sind neben dem MPI in Nijmegen die Berlin-Brandenburgische Akademie der Wissenschaften ([www.bbaw.de](http://www.bbaw.de)) und die Koninklijke Nederlandse Akademie van Wetenschappen ([www.knaw.nl](http://www.knaw.nl)). Siehe auch <http://www.bbaw.de/veranstaltungen/2011/oktober/tla> (letzter Zugriff: 24. 1. 2013).

### 3. Bedeutung des Projekts

Mit dem beschriebenen Vorhaben kann eine Lücke in den bisherigen Bemühungen um eine möglichst vollständige Dokumentation des Niedersorbischen geschlossen werden. Die Einbeziehung der heute noch gesprochenen niedersorbischen Dialekte (bzw. des muttersprachlich-dialektalen Niedersorbischen) stellt nicht nur einen wichtigen Schritt in Richtung einer Gesamtdokumentation dar, sondern eröffnet zugleich neue Möglichkeiten zur Erforschung der niedersorbischen Sprache. Die Datengrundlage für auf die Dialekte oder auf einen Vergleich zwischen dialektalem und schriftsprachlichem Niedersorbisch gerichtete Untersuchungen war bisher unbefriedigend und kann durch die im Projekt erstellten Aufnahmen erheblich erweitert werden. Sie wird für Forschungen sowohl über „The Language Archive“ als auch über [niedersorbisch.de](http://niedersorbisch.de) weltweit zur Verfügung stehen und ist durch die Transkriptionen zudem deutlich effektiver nutzbar als reine Audiokorpora. Dabei bleibt die Möglichkeit für Untersuchungen, die sich primär auf die Tondateien stützen, zum Beispiel zur dialektalen Phonetik oder Intonation, vollumfänglich erhalten. Wichtig ist auch die durch eine Verknüpfung mit dem schriftsprachlichen Textkorpus ermöglichte gemeinsame Nutzung des gesamten Datenbestandes. Dies ist insbesondere für die im SORBLEX-Projekt geplante umfassende Beschreibung des niedersorbischen Wortschatzes von Bedeutung.

### Literaturverzeichnis

- BARTELS, HAUKE (2010): The German-Lower Sorbian Online Dictionary, in: DYKSTRA, A.; SCHOONHEIM, T. (eds.): *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*. Ljouwert: Afûk-Fryske Akademy, 1450–1462.
- BARTELS, HAUKE (2012a): Digitalizacija staršych dolnosěrbsko-nimskich słownikow a dokumentacija dolnosěrbsčiny, in: МОТОРНИЙ, В.; ШОЛЬЦЕ, Д. (ред.): *Питання сорабістики. Матеріали XIII Міжнародного сорабістичного семінару 2010*. Львів-Будишин, 10–25.
- BARTELS, HAUKE (2012b) [im Druck]: Maßnahmen zur Dokumentation des Niedersorbischen, in: *Slavia Occidentalis* 69 (2012).
- BARTELS, HAUKE (2013) [im Druck]: Zur Konzeption eines historisch-dokumentierenden Wortschatz-Informationssystems des Niedersorbischen. Pläne zur Behebung eines drängenden Forschungsdesiderats, in: KEMPGEN, S.; FRANZ, N.; JAKIŠA, M.; WINGENDER, M. (Hrsg.): *Deutsche Beiträge zum 15. Internationalen Slavistenkongress, Minsk 2013*. München-Berlin: Verlag Otto Sagner (Die Welt der Slawen. Sammelbände) – [Eine inhaltlich identische niedersorbische Fassung ist veröffentlicht in: *Lětopis* 60 (2013) 1, 16–26.]
- BRIJNEN, HÉLÈNE B. (2004): *Die Sprache des Hanso Nepila. Der niedersorbische Dialekt von Schleife in einer Handschrift aus der 1. Hälfte des 19. Jahrhunderts*. Bautzen: Domowina-Verlag (Schriften des Sorbischen Instituts; 35).
- CARSTENSEN, KAI-UWE ET AL. (2010): *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Heidelberg-Berlin: Spektrum Akademischer Verlag.
- DOBES-BROSCHÜRE (2004): *Dokumentation bedrohter Sprache. Eine Reise zu den weltweiten Projekten einer Förderinitiative*. Hannover: VolkswagenStiftung. [<http://www.volkswagenstiftung.de/fileadmin/downloads/publikationen/Dobes.pdf>]
- DRUDE, SEBASTIAN; BROEDER, DAAN; TRILSBEEK, PAUL; WITTENBURG, PETER (2012): The Language Archive – a new hub for language resources, in: CALZOLARI, N. (ed.):

- Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, May 23<sup>rd</sup>–25<sup>th</sup>, 2012.* European Language Resources Association (ELRA), 3264–3267.
- FASSKE, HELMUT (1964): *Die Vetschauer Mundart*. Bautzen: Domowina-Verlag (Schriftenreihe des Instituts für sorbische Volksforschung; 19).
- FASSKE, HELMUT (Hg.) (1998): *Serbšćina*. Opole: Uniwersytet Opolski – Instytut Filologii Polskiej.
- FASSKE, HELMUT (2003): Dialekte und Entwicklung des Sprachterritoriums, in: *Der Niedersorben Wendisch. Eine Sprach-Zeit-Reise*. Bautzen: Domowina-Verlag, 43–50.
- FASSKE, HELMUT, MICHALK, SIEGFRIED (1965): *Sorbische Dialekttexte III. Schmogrow, Kreis Cottbus*. Bautzen: Domowina-Verlag.
- FASSKE, HELMUT, MICHALK, SIEGFRIED (1969): *Sorbische Dialekttexte VII. Mattendorf, Kreis Forst*. Bautzen: Domowina-Verlag.
- GRINEVALD, COLETTE; BERT, MICHEL (2011): Speakers and communities, in: AUSTIN, PETER K.; SALLABANK, JULIA (eds.): *The Cambridge Handbook of Endangered Languages*. Cambridge-New York et al.: CUP, 45–65.
- HAIG, GEOFFREY; NAU, NICOLE; SCHNELL, STEFAN; WEGENER, CLAUDIA (2011): Introduction: Documenting endangered languages before, during, and after the DoBeS programm, in: DIES. (eds.): *Documenting endangered languages. Achievements and perspectives*. Berlin-Boston: De Gruyter Mouton, 1–14.
- JAHN, PETER M. (2010): *Vom Roboter zum Schulpropheten. Hanso Nepila (1766–1856). Mikrohistorische Studien zu Leben und Werk eines wendischen Fronarbeiters und Schriftstellers aus Rohne in der Standesherrschaft Muskau. Mit einer Übersetzung der Handschriften*. Bautzen: Domowina-Verlag (Schriften des Sorbischen Instituts; 52).
- JODLBAUER, RALPH; SPIESS, GUNTER; STEENWIJK, HAN (2001): *Die aktuelle Situation der niedersorbischen Sprache. Ergebnisse einer soziolinguistischen Untersuchung der Jahre 1993–1995*. Bautzen: Domowina-Verlag (Schriften des Sorbischen Instituts; 27).
- KLEIN, WOLFGANG (2012): Alle zwei Wochen verschwindet eine Sprache, in: GÜNTER STOCK (Hrsg.): *Die Akademie am Gendarmenmarkt 2012/13*. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 8–13.
- MUCKE, ERNST (1911–15/1926, 1928): *Wörterbuch der nieder-wendischen Sprache und ihrer Dialekte*. Bd. 1. St. Petersburg 1911–1915, Prag 1926: Verlag der russischen und tschechischen Akademie der Wissenschaften. Bd. 2–3. Prag 1928: Verlag der böhmischen Akademie für Wissenschaften und Kunst (Fotomechanischer Neudruck. Bautzen: Domowina-Verlag 2008).
- SCHOLZE, LENKA (2008): *Das grammatische System der obersorbischen Umgangssprache im Sprachkontakt. Mit Grammatiktafeln im Anhang*. Bautzen: Domowina-Verlag (Schriften des Sorbischen Instituts; 45).
- SSA 1965–1996. *Sorbischer Sprachatlas*, 15 Bände. Bautzen: Domowina-Verlag.
- STAROSTA, MANFRED (1982): *Niedersorbische Orthographie und Interpunktion. Regeln*. Bautzen: Domowina-Verlag.
- STAROSTA, MANFRED (1999): *Niedersorbisch-deutsches Wörterbuch*. Bautzen: Domowina-Verlag.
- SZCZEPAŃSKI, MARCIN (2012): Modelowanie danych do słownika elektronicznego na przykładzie artykułów słownika A. Muki, in: МОТОРНИЙ, В.; ШОЛЬЦЕ, Д. (ред.):

*Питання сорабістики. Матеріали XIII Міжнародного сорабістичного семінару 2010. Львів-Будишин, 26–61.*

ŠWJELA, BOGUMIŁ (1961): *Dolnoserbско-němski słownik*. Budyšin: Ludowe nakładnistwo Domowina.

WOLFF, DIETER (2003). Gesteuerter Zweitspracherwerb, in: RICKHEIT, GERT; HERRMANN, THEO; DEUTSCH, WERNER (Hrsg.): *Psycholinguistik. Ein internationales Handbuch*. Berlin-New York: Walter de Gruyter, 833–844.

ZWAHR, JOHANN GEORG (1847): *Niederlausitz-wendisch-deutsches Handwörterbuch* (Fotomechanischer Neudruck. Bautzen: Domowina-Verlag 1989).