

HAUKE BARTELS  
Cottbus

## MASSNAHMEN ZUR DOKUMENTATION DES NIEDERSORBISCHEN

### **Zur Notwendigkeit und Dringlichkeit einer umfassenden Dokumentation des Niedersorbischen**

#### 1.1. Sprachdokumentation

Für die meisten slavischen Sprachen, die wie das Polnische oder Russische als nationale Standardsprachen etabliert sind, hat der Begriff "Sprachdokumentation" eher geringe Relevanz. Für diese Sprachen gibt es Einzelphilologien, große und relativ stabile Sprachgemeinschaften und mithin wenig Anlass, jenseits etablierter sprachwissenschaftlicher Forschung besondere Maßnahmen zur Dokumentation der Sprache zu ergreifen. Zwar ist eine Dokumentation im Sinne der Erhebung primärer Sprachdaten vorgängiger und damit notwendiger Bestandteil jeder Sprachbeschreibung und wäre, als mehr oder weniger regelmäßige Tätigkeit, besonders für die Erforschung von Sprachwandel nützlich. Sprachdokumentation als besondere Aktivität von Linguisten erlebt jedoch seit nunmehr mindestens zwei Jahrzehnten im Zusammenhang mit der Sorge um die zahlreichen weltweit gefährdeten Sprachen<sup>1</sup> eine rasante Entwicklung und wird heute teilweise schon als eigene akademische Disziplin betrachtet [Woodbury 2011, Haig et al. 2011]. Die Dringlichkeit, eine Sprache oder bestimmte Varietäten einer Sprache zu dokumentieren, geht unmittelbar auf die akute Gefahr zurück, der Niedergang oder "Tod" einer Sprache könnte eine grundlegende Dokumentation ein für alle Mal unmöglich machen. Erst diese Ausgangslage führt in der Regel zu einer Prioritätensetzung, die die Sicherstellung von Daten bzw. die Erarbeitung einer für die (gegebenenfalls zukünftige) Forschung geeigneten Datenbasis<sup>2</sup> zumindest gleichberechtigt neben die Sprachbeschreibung selbst stellt.

Explizit auf die Dokumentation einer Sprache ausgerichtete Projekte enthalten sich häufig sogar bewusst einer traditionellen Sprachbeschreibung in Form von umfassenden deskriptiven Grammatiken und Wörterbüchern. Der Schwerpunkt liegt stattdessen auf der

---

<sup>1</sup> Vgl. z.B. Austin & Sallabank 2011.

<sup>2</sup> Vgl. Himmelmann [2006: 1]: "a language documentation is a lasting, multipurpose record of a language".

Sammlung und Aufbereitung von primären Sprachdaten, d.h. in der Regel von Audio- oder sogar Videoaufnahmen einer möglichst großen Anzahl möglichst vielgestaltiger Sprachhandlungen. Diese werden durch Metadaten und Annotationen ergänzt, die eine langfristige Interpretation der Primärdaten ermöglichen sollen [Himmelmann 2006: 7–17]. Diese gleichwohl sehr arbeitsintensive Konzentration auf die Erhebung und (linguistische wie technische) Aufbereitung von Daten ist vor allem den Zwängen geschuldet, denen Linguisten ausgesetzt sind, die sich mit stark bedrohten Sprachen beschäftigen, die oftmals nur noch von wenigen (einigermaßen) kompetenten Sprechern genutzt werden. Angesichts einer solchen Bedrohungslage müssen Detailstudien bzw. sorgfältig aufbereitete Grammatiken und Wörterbücher zurückstehen, und zwar in der Annahme oder Hoffnung, dass sie auf Grundlage der gesammelten Primärdaten später noch verfasst werden können.

Das Niedersorbische befindet sich in einer Lage zwischen dem um seine Zukunft nicht bangenden Polnischen und vor unseren Augen “sterbenden” Sprachen. Die niedersorbische Schriftsprache, deren Zukunft bekanntlich als unsicher gelten muss [Jodlbauer, Spieß & Steenwijk 2001], ist im Vergleich mit linguistisch gänzlich undokumentierten Sprachen in ihren Grundzügen relativ gut beschrieben; im Vergleich zur Forschungstiefe, die bei anderen slavischen Standardsprachen erreicht wird, gibt es gleichwohl zahlreiche Desiderata. Für die Schriftsprache existiert mittlerweile ein umfassendes Textkorpus [vgl. 2.2], das sogar die Erarbeitung eines nahezu vollständigen historisch-dokumentierenden Wörterbuchs ermöglichen würde [vgl. 3.]. Die primäre Existenzform des Niedersorbischen hingegen, die dialektal basierte einstige Alltagssprache auf den Dörfern des niedersorbischen Sprachgebiets, ist eine immer seltener genutzte, schon seit Jahrzehnten nicht mehr als Muttersprache weitergegebene und daher “sterbende” Varietät des Niedersorbischen [ebd.] und insofern vergleichbar mit den Sprachen, um deren Dokumentation man sich weltweit bemüht. Diese Form des Niedersorbischen ist noch nicht ausreichend dokumentiert, aber jetzt Gegenstand eines gesonderten Projekts [vgl. 2.3].

Die beschriebene Ausgangslage erlaubt es, den vorrangig auf die Sammlung und Aufbereitung von Primärdaten ausgerichteten Ansatz von Sprachdokumentation im engeren Sinne zu kombinieren mit Maßnahmen, die eher der traditionellen deskriptiven Sprachwissenschaft zuzuordnen wären. Insofern ist der im Titel gewählte Begriff von Sprachdokumentation ein weiterer. Er verbindet eine möglichst breite und dauerhafte Bereitstellung von Primärdaten mit laufender Datenanalyse [vgl. Himmelmann 2006: 20], zur Zeit vor allem in lexikografischen Vorhaben. Die im Folgenden kurz vorzustellenden Einzelprojekte stehen daher in engem konzeptionellen Zusammenhang und es ist im obigen Sinne gerechtfertigt, sie wie im Titel allesamt als „Maßnahmen zur Dokumentation des Niedersorbischen“ zusammenzufassen. Dabei handelt es sich nur in einem Fall, dem DoBeS-Projekt, um ein typisches, klar auf Sprachdokumentation im engeren Sinne zugeschnittenes Vorhaben. Für die Hervorhebung des Aspekts der Sprachdokumentation bei allen Vorhaben spricht vor allem die Bedrohungslage, in der sich das Niedersorbische befindet und damit die Perspektive, in der die Projekte konzipiert und angegangen wurden.

## 1.2. Sprecher als Informanten

Die Sprachwissenschaft steht bei der wissenschaftlichen Beschreibung des Niedersorbischen vor dem prinzipiellen Problem, diesseits der mittlerweile über 70jährigen Dialekt-

sprecher, die ja Gegenstand eines besonderen Dokumentationsprojekts sind [siehe 2.3], nur über eine sehr überschaubare Anzahl von für linguistische Untersuchungen zuverlässigen Informanten zu verfügen [Bartels 2009]. Dies ist darauf zurückzuführen, dass es sich in dieser Altersgruppe, bis auf vielleicht wenige Ausnahmen von vollkompetenten Sprechern (d.h. voller Spracherwerb und kein Sprachverlust), nur um Semi-Sprecher (semi-speaker) und so genannte Neu-Sprecher (neo-speaker) handelt [Grinewald & Bert 2011]. Als *S e m i - S p r e c h e r* werden solche Angehörigen der traditionellen Sprachgemeinschaft bezeichnet, die über gute rezeptive Fähigkeiten, jedoch in der Regel über eine nur eingeschränkte aktive Sprachkompetenz verfügen. Der Grad der aktiven Sprachbeherrschung kann dabei je nach individuellen Voraussetzungen und Umständen stark variieren. Die Beschränkung aktiver Sprachkompetenz kann dabei sowohl auf partiellen Sprachverlust (z.B. durch nur noch seltenen Gebrauch) als auch auf unvollkommenen Spracherwerb zurückzuführen sein [ebd.: 50]. Während der Begriff Semi-Sprecher impliziert, dass die jeweiligen Sprachkenntnisse in der Sprachgemeinschaft selbst durch überwiegend ungesteuerte Spracherwerbsprozesse erworben wurden – im Erst- oder, unter den gegebenen Umständen, häufig auch im Zweitspracherwerb –, geht man beim Typ *N e u - S p r e c h e r* davon aus, dass die Sprache erst durch (ggf. verschiedene Formen von) Unterricht erworben wurde. Es handelt sich hier also generell um Sprachlerner im Sinne eines gesteuerten Zweitspracherwerbs [auch: Fremdspracherwerb, vgl. Dietrich 2004], wie er im Rahmen von Spracherhalts- oder Sprachrevitalisierungsmaßnahmen stattfindet. Es ist klar, dass die Varianz der tatsächlich erworbenen Sprachkenntnisse hier mindestens so groß ist wie bei den Semi-Sprechern. Typisch für diese Kategorie ist aber auch, dass hiermit neue Sprecher sowohl aus der traditionellen Sprachgemeinschaft als auch von außerhalb zusammengefasst werden.

Drei Sprechertypen treten somit heute in der jungen und mittleren Generation der niedersorbischen Sprachgemeinschaft auf:

- (a) Semi-Sprecher I: ehemals vollkompetente Sprecher mit teilweise Verlust ihrer Sprachkompetenz (Existenz fraglich, falls vorhanden: sehr selten);
- (b) Semi-Sprecher II: Sprecher, die ihre Kenntnisse zwar überwiegend ungesteuert in der traditionellen Sprachgemeinschaft, jedoch nur unvollständig erworben haben;
- (c) Neu-Sprecher, die ihre Kenntnisse überwiegend gesteuert (teilweise auch autodidaktisch) erworben haben.

Bei den dominierenden Sprechertypen (b) und (c) ist es möglich, dass bei entsprechender Begabung, großem Lernanreiz und gutem Informations-Input ein hohes Kompetenzniveau erreicht wird, das dem eines Typus gleichkommt, der im Niedersorbischen ansonsten nicht mehr anzutreffen ist: dem jungen vollkompetenten Muttersprachler. In aller Regel haben wir es aber bei diesen beiden Typen mit Fossilierung, d.h. dem Abbruch des Spracherwerbsprozesses auf einer bestimmten Teilerwerbsstufe oder diversen Lernervariäten zu tun [Klein 2001: 611, 615f].

Der soeben erwähnte Begriff des Muttersprachlers (engl. native speaker) ist insofern von Belang, als in der Spracherwerbsforschung davon ausgegangen wird, „dass nur der Erstspracherwerb in der Regel zu perfekter Beherrschung führt“<sup>3</sup> [ebd.: 608] und damit ein

---

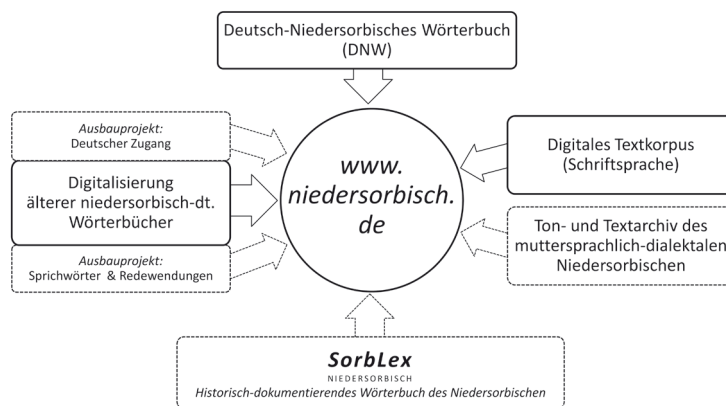
<sup>3</sup> Dies gilt daneben auch für den simultanen Erwerb mehrerer Erstsprachen: „Der simultane bilinguale Spracherwerb ist in gewisser Weise eine Sonderform des Mutterspracherwerbs, er bezeichnet den natürlichen Erwerb zweier Sprachen als Muttersprache(n), wobei normalerweise die Eltern des Kindes ihre jeweilige Muttersprache

Kompetenzniveau erreicht wird, das für sprachwissenschaftliche Informanten in der Regel, d.h. abgesehen etwa von speziellen Untersuchungen zu Spracherwerb, Sprachmischung usw., gefordert wird. Insofern ist es nicht verwunderlich, dass auch in der Literatur zur Sprachdokumentation häufig auf diesen Begriff Bezug genommen wird. Bei Mosel [2006: 75] etwa heißt es: “In the very beginning of fieldwork, the researcher has to rely on elicitation. Elicitation means getting linguistic data from native speakers by asking questions” [Hervorhebung H.B.].

Wegen der sich aus der Sprachsituation ergebenden Schwierigkeiten bei der Nutzung einer der wichtigsten sprachwissenschaftlichen Methoden, der Informantenbefragung, kommt dem Aufbau einer gut aufbereiteten und breit angelegten<sup>4</sup> Datenbasis auf Grundlage möglichst aller verfügbaren Informationen für zukünftige Untersuchungen zum Niedersorbischen eine entscheidende Bedeutung zu.

## 2. Maßnahmen zur Dokumentation des Niedersorbischen

Die folgende Grafik zeigt eine Übersicht über die Projekte, die im Folgenden kurz vorgestellt werden<sup>5</sup>. Schon länger laufende (DNW, schriftsprachliches Korpus) oder mittlerweile abgeschlossene Projekte (Wörterbuch-Retrodigitalisierung), d.h. Vorhaben mit schon öffentlich zugänglichen Ergebnissen, sind mit kompakter Linie umrahmt dargestellt. Eine gestrichelte Linie kennzeichnet solche Projekte, die erst vor kurzem begonnen wurden oder noch in Vorbereitung (*SorbLex*) sind.



als Input einbringen. Der simultane Bilingualismus unterliegt ähnlichen Steuerungsmechanismen wie der Mutterspracherwerb.“ [Wolff 2003: 835] – Mit „perfekt“ ist hier eine Sprachbeherrschung gemeint, die sich dadurch auszeichnet, dass sie sich in der jeweiligen sozialen Umgebung nicht auffällig unterscheidet [Klein 2001: 604f].

<sup>4</sup> “Ideally, then, a language documentation would cover all registers and varieties, social or local; it would contain evidence for language as a social practice as well as a cognitive faculty; it would include specimens of spoken and written language; and so on.” [Himmelman 2006: 2].

<sup>5</sup> Alle Vorhaben wurden oder werden von mehreren Personen bearbeitet. Informationen dazu finden sich auf den Internetseiten des Sorbischen Instituts unter <http://www.serbski-institut.de> > Abteilungen > Niedersorbische Forschungen > Projekte.

## 2.1. Deutsch-niedersorbisches Wörterbuch (DNW)

Das seit 2001 entstehende, derzeit bereits weit fortgeschrittene Deutsch-niedersorbische Internet-Wörterbuch (DNW) ist das bislang umfangreichste niedersorbische Wörterbuch. Es ist vornehmlich als aktives Wörterbuch konzipiert und versucht, den besonderen Anforderungen, die sich aus der aktuellen Sprachsituation ergeben, durch eine möglichst genaue Beschreibung des Wortschatzes gerecht zu werden. Besonders wichtig für Lerner, die das Niedersorbische auf sehr unterschiedlichem Niveau, aber eben fast ausschließlich nicht-muttersprachlich beherrschen, ist zunächst eine hinreichende Differenzierung unterschiedlicher Bedeutungen eines deutschen Ausgangswortes und ihrer jeweils möglichen niedersorbischen Entsprechungen. Der diesbezügliche Fortschritt wird durch einen direkten Vergleich des DNW mit einem Schulwörterbuch von 1990 deutlich:

Wörterbuch Deutsch-Niedersorbisch [Jannasch 1990]:

**erfüllen** społniś 79 p, dopołniś 79 p, społnjowaś 62, dopołnjowaś 62; **sich** ~ se dopołniś 79 p, se dopołnjowaś 62

DNW 2003ff:

## erfüllen

<b>1. (sich) (sich) verwirklichen</b>	dopołniś (se) p 79    dopołnjowaś (se) ip 62
Sie erfüllte ihm jeden Wunsch. ( <i>Usus</i> )	Wóna jo jomu kužde žyćenje dopołnjowała.
Endlich hat sich sein Wunsch erfüllt.	Skóńcnje jo se jogo žyceńko dopołniło.
<b>2. Genüge tun</b>	docyniś p 79    docynjaś ip 76
Er erfüllt seine Pflichten. ( <i>n. akt. Präs.</i> )	Wón docynijo / docynja swóje winowatosći.
<b>3. innerlich ausfüllen</b>	<i>allg; zufriedenstellen</i> wupołniś p 79    wupołnjowaś ip 62; <i>innerlich ganz einnehmen</i> napołniś p 79    napołnjowaś ip 62
Eine innere Unruhe erfüllte ihn.	Nutśikowny njeměr jo jogo wupołnił / napołnił.
Die neue Aufgabe erfüllt ihn ganz. ( <i>n. akt. Präs.</i> )	Nowy nadawk jogo krađu wupołnjujo / wupołnijo.
Neue Hoffnung erfüllte sie.	Nowa nažeja jo ju napołniła / wupołniła.
Er hatte ein erfülltes Leben.	Wón jo měi wupołnjone žywjenje.
Die Nachricht über den Tod ihrer Freundin erfüllte sie mit tiefer Trauer.	*Powěśc wó smjerši swójeje pšijašelki jo ju krađu roztužyla.
<b>4. räumlich einnehmen</b>	połniś ip 77    napołniś p 79    napołnjowaś ip 62
Der Duft von Weihrauch erfüllte die ganze Kirche.	Wóručowa wón jo napołniła celu cerkwju.

Dabei werden, wo möglich, Hinweise auf besondere Schwierigkeiten für Lerner mit Erstsprache Deutsch gegeben, z.B. Hinweise auf den Gebrauch des Verbalaspekts (hier *Usus* als Kommentar, der die Verwendung des imperfektiven Aspekts motiviert, oder *n[icht] akt[uelles] Präs[ens]* als Hinweis auf eine im Niedersorbischen übliche Gebrauchsweise des perfektiven Aspekts im Präsens). Für Lerner besonders wichtig sind aber die zahlreichen Beispielsätze, wovon es im Wörterbuch zehntausende gibt [mehr dazu in Bartels 2010b].

Beim DNW handelt es sich damit um das am stärksten unmittelbar sprachfördernde Vorhaben der Zweigstelle. Gleichzeitig enthält es zahlreiche bisher lexikografisch nicht registrierte lexikalische Einheiten [vgl. 2.5.2]. Der gesamte Datenbestand des DNW ist in XML kodiert und daher mit anderen Quellen [z.B. den digitalisierten niedersorbisch-deutschen Wörterbüchern; vgl. 2.4] systematisch verknüpfbar. Diese Datenbasis dient auch zur Generierung der Internet-Ausgabe in HTML:

Druckmanuskript:	XML-Modellierung des niedersorbischen Äquivalents:
<b>Schulbank</b> šulska ławka <i>f</i> 3	<pre> &lt;aequiv&gt;   &lt;formerg_vor typ="Adj"&gt;<sup>6</sup>     &lt;form&gt;šulska&lt;/form&gt;     &lt;paradigma&gt;48&lt;/paradigma&gt;   &lt;/formerg_vor&gt;   &lt;form&gt;ławka&lt;/form&gt;   &lt;genus&gt;f&lt;/genus&gt;   &lt;paradigma&gt;3&lt;/paradigma&gt; &lt;/aequiv&gt; </pre>

Seit 2003 steht das DNW als (noch nicht vollständig endredigierte) Vorab-Veröffentlichung im Internet zur Verfügung. Der dortige Artikelbestand wird seitdem laufend aktualisiert und erweitert. Seit Juli 2007 liegen für alle Lemmastrecken Wörterbuchartikel vor, sodass es sich bereits um ein (in diesem Sinne) vollständiges Nachschlagewerk zum aktiven Gebrauch des Niedersorbischen handelt. Derzeit sind bereits Informationen zu etwa 77.000 Stichwörtern im Internet verfügbar. Weitere Informationen zum Wörterbuch finden sich auf der Internetseite des DNW. Eine ausführliche Beschreibung des Projekts steht noch aus.

## 2.2. Elektronisches Textkorpus der niedersorbischen Schriftsprache

Schon bald nach der Gründung der Cottbuser Zweigstelle für niedersorbische Forschungen 1992 wurde mit dem Aufbau eines computerlesbaren Textkorpus der niedersorbischen Schriftsprache begonnen. Dabei spielte nicht nur die allgemeine Entwicklung in der Sprachwissenschaft eine Rolle, die der Korpuslinguistik methodologisch eine immer größere Relevanz zuschrieb. Vielmehr wuchs das Bewusstsein, dass in der bedrohlichen Lage, in der sich das Niedersorbische befindet, ein Korpus möglicherweise bald die wichtigste Informationsquelle zum Sprachgebrauch sein könnte. Der Ausbau des Textkorpus konnte dank umfangreicher Drittmittel in den letzten Jahren durch externe Digitalisierungsmaßnahmen stark vorangetrieben werden, sodass mittlerweile eine Größe von ca. 25 Mio. laufenden Wortformen (tokens) erreicht ist. Bedenkt man den relativ geringen Umfang des niedersor-

<sup>6</sup> Das Element <formerg\_vor> enthält einen Ausdruck, der linear vor dem Kopf des Gesamtausdrucks steht, in diesem Fall ein einfaches adjektivisches Attribut, das nach dem Muster 48 dekliniert wird.

bischen Schrifttums, so ist mit dieser Größenordnung schon ein wesentlicher Teil der überhaupt verfügbaren Schriften einbezogen [Bartels 2010a]. Da das Korpus in den letzten Jahren auch um digitalisierte Texte aus der Zeit vor Mitte des 19. Jahrhunderts, also um eine „historische Abteilung“ ergänzt wurde, repräsentiert es nun die niedersorbische Schriftsprache von den frühesten Zeugnissen (1. Hälfte des 16. Jahrhunderts) bis in die Gegenwart. Teile des Korpus sind seit Ende 2010 im Internet verfügbar<sup>7</sup>.

Im Hinblick auf die Nutzung des Korpus als wichtiger Datenbasis für Lexikografie und Lexikologie sowie Grammatikforschung unterliegt es bislang noch einigen wichtigen Beschränkungen:

(a) Ziel ist die Erstellung eines Vollkorpus des niedersorbischen Schrifttums, das alle noch verfügbaren Texte beinhaltet und daher die Gesamtheit der überlieferten Schriftsprache. Um dieses Ziel zu erreichen, müssen noch Lücken insbesondere in der Zeit nach 1945 geschlossen werden.

(b) Das Korpus besteht aus in verschiedenen Hinsichten formal uneinheitlichen und teilweise unkorrigierten Abschriften. Hier ist noch erhebliche Arbeit zu leisten.

(c) Die Texte sind bislang nur rudimentär annotiert. Für eine effektive Nutzung insbesondere für lexikografische Projekte wie *SorbLex* ist daher eine gründliche Analyse und Annotierung des Korpus notwendig, die etwa eine durchgängige Lemmatisierung ermöglicht [vgl. 3.]. Das Problem, dass es bislang kein Korpus des dialektalen Niedersorbischen gab, konnte inzwischen angegangen werden [vgl. 2.3].

### 2.3. Ein neues Text- und Tonarchiv des dialektalen Niedersorbischen

Ziel des Vorhabens ist der Aufbau eines bisher fehlenden und für Dokumentation und Forschung dringend notwendigen Korpus des muttersprachlich-dialektalen Niedersorbischen. Durch die Kombination von Sprachaufnahmen, Transkriptionen und Übersetzungen entsteht in diesem Projekt ein **Ton- und Textarchiv** des heute noch von älteren Sprechern in mündlicher Kommunikation verwendeten dialektal geprägten Niedersorbischen. Gegenstand ist damit die Sprache der heute in der Regel zwischen 70 und 90 Jahre alten muttersprachlichen Dialektsprecher, die in der ersten Hälfte des 20. Jahrhunderts als Kinder das Niedersorbische noch als Erstsprache erworben und verwendet haben. Es ist absehbar, dass das Niedersorbische in dieser Form in naher Zukunft aussterben wird. Da diese Sprachform sich in vielerlei Hinsicht stark von der Schriftsprache unterscheidet, muss eine Dokumentation und Beschreibung des Niedersorbischen ohne ihre Einbeziehung unvollständig bleiben.

Das seit Ende 2011 im Aufbau befindliche Ton- und Textarchiv wird aus neuen Audioaufnahmen sowie den zugehörigen Transkriptionen und Übersetzungen bestehen. Am aufwendigsten und damit Tätigkeitsschwerpunkt ist die Herstellung der Transkriptionen. Das Gesamtvorhaben ist auf drei bis vier Jahre angelegt. Eine ausführliche Beschreibung des Projekts findet sich in Bartels & Thorquindt-Stumpf 2013.

Die Ergebnisse des Projekts werden auf zweierlei Weise veröffentlicht. Zum einen werden die Sprachaufnahmen mitsamt den Transkriptionen und Übersetzungen in das DoBeS-

---

<sup>7</sup> URL: <http://www.niedersorbisch.de/korpus>



Archiv am Max-Planck-Institut für Psycholinguistik in Nijmegen<sup>8</sup> (bzw. in das derzeit entstehende „Language Archive“<sup>9</sup>) integriert. Diese Integration ermöglicht einen Zugriff primär über die Audiodateien. Zunächst in Form eines aus den Transkriptionen bestehenden Textkorpus werden die Ergebnisse des Projekts dagegen in das Informationsportal zum Niedersorbischen [www.niedersorbisch.de] eingebunden. Das Textkorpus des muttersprachlich-dialektalen Niedersorbischen wird somit das dort ebenfalls teilweise bereits zugängliche schriftsprachliche Korpus ergänzen. Die Audiodateien werden dort über eine Verknüpfung mit den Transkriptionen erreichbar sein. Die folgende Grafik zeigt einen Bildausschnitt aus der im Projekt benutzten Software ELAN<sup>10</sup>. Zu sehen sind die beiden Audiospuren, eine an die heutige schriftsprachliche grafische Norm angelehnte Transkription sowie eine Übersetzung ins Deutsche<sup>11</sup>:



#### 2.4. Digitalisierung niedersorbisch-deutscher Wörterbücher

In diesem vor kurzem abgeschlossenen Projekt [Laufzeit 2008–2011] 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, d.h. die Mikrostrukturen der vier Wörterbücher wurden analysiert und einzelne Informationspositionen so getaggt, dass auf sie mit Mitteln der elektronischen Datenverar-

<sup>8</sup> URL: <http://www.mpi.nl/DOBES>

<sup>9</sup> Siehe <http://www.mpi.nl/news/news-archive/official-opening-of-language-archive-in-berlin>

<sup>10</sup> <http://www.lat-mpi.eu/tools/elan/>

<sup>11</sup> Für einen Teil der Daten werden auch eine phonetische Transkription (IPA) sowie eine Übersetzung ins Englische angefertigt.



beitung systematisch und für alle Wörterbücher einheitlich zugegriffen werden kann<sup>12</sup>. So entstand nicht nur die Möglichkeit einer gemeinsamen Präsentation der Daten, sondern ebenso ein Art Datenbank der in den Wörterbüchern enthaltenen Informationen. Auf dieser Grundlage erstellte Internetversionen wurden Anfang 2012 über [www.niedersorbisch.de](http://www.niedersorbisch.de) [vgl. 4.] bereitgestellt. Mit dieser Maßnahme ist somit (a) einerseits eine nahezu vollständige, systematische und computerlesbare Erfassung aller in den Quellen vorhandenen Informationen erfolgt sowie (b) zusätzlich zur Inventarisierung aller Lemmata und der in Nestern enthaltenen Lexeme mitsamt ihrer jeweiligen Beschreibung durch die Erfassung aller Beispiel(sätze) ein weiteres (Teil)Korpus entstanden. Mit (a) wird eine Datengrundlage geschaffen, die sich nicht nur bei der Erarbeitung der Wörterbuchartikel für das geplante historisch-dokumentierende Wörterbuch *SorbLex* [vgl. 3.] effektiv einsetzen lässt, sondern die zugleich die in historischen Wörterbüchern üblichen Informationen zur „Buchung“ eines Lexems in der älteren Lexikografie liefert. Durch die technische Aufbereitung der Daten können als „Buchungen“ angebotene Informationen über traditionelle Lösungen weit hinausgehen, indem nicht nur die Buchung an sich angeführt wird (in der Regel als Sigle), sondern darüber hinaus per Hyperlink die gesamte Original-Beschreibung (als neu gestaltete Bildschirmdarstellung und als Faksimile). Um ein Beispiel zu geben: Angenommen, ein in *SorbLex* aufgenommenes Lemma *jaskolica* ‘Schwalbe’ sei schon in *ZWAHR* 1847 gebucht (u.a. in der Schreibvariante *jaskoliza*) und würde entsprechend gekennzeichnet. Die Sigle *Zw* würde als Hyperlink zum vollständigen Artikel in der digitalen Version von *Zwahr* 1847 führen:

Artikel **jaskoliza**, Quelle: Zwahr 1847.  
Hinweis: Es handelt sich um ein älteres Wörterbuch; die Informationen können teilweise veraltet sein.

▫ *jaskoliza, waskoliza, jaskolizka, waskolizka*

**jaskoliza**  
**waskoliza**  
**jaskolizka**  
**waskolizka**  
 ta

**die Schwalbe**

▫ *watschkotasch*

**watschkotasch**

**zwitfchern wie Schwalben**  
*ta jaskoliza watschkozo* — die Schw. zwitfchert

▫ *jaskolizny*

**jaskolizny** Flex.  
**Schwalben–**

<sup>12</sup> Je feinkörniger Analyse und Tagging vorgenommen werden, desto größer sind die späteren Zugriffsmöglichkeiten. So ergibt sich etwa – um ein weiter oben in anderem Zusammenhang (Kapitel 2.1) erwähntes Beispiel aufzugreifen – aus der Modellierung auch der internen Struktur des Ausdrucks *šulska lawka* die Möglichkeit eines Zugriffs nicht nur auf das gesamte Äquivalent, sondern auch auf einzelne Elemente (z.B. einem Substantiv vorangestellte Adjektive) und damit auf bestimmte Nominationstypen.

Von dort aus würde man weiter gelangen zum Faksimile des Artikels in der Original-Druckausgabe von 1847:

**Jaskoliza, waskoliza u. -zka, ta, die Schwalbe; watschkotasch, zwitschern wie Schwalben, ta j. watschkozo, die Schw. zwitschert, jaskolizny, a, e, Schwalben—.**

Auf diese Weise kann die bisherige lexikografische Tradition der in *SorbLex* erfassten Lemmata vollständig und transparent dargestellt werden, ohne dass der Nutzer die älteren Wörterbücher selbst zur Hand haben muss. Das Projekt liefert zugleich eine wertvolle Datenbasis für zukünftige Forschungen. Durch die völlige Neu-Modellierung der Daten im XML-Format wurden zahlreiche neue Recherchemöglichkeiten eröffnet<sup>13</sup>.

## 2.5. Ausbauprojekte zu den digitalisierten Wörterbüchern

Die im zuvor beschriebenen Projekt zur Digitalisierung niedersorbisch-deutscher Wörterbücher erarbeiteten Daten werden in zwei Anschlussprojekten weiterbearbeitet und ausgewertet.

### 2.5.1 Deutscher Zugang

Der Datenbestand wird so erweitert, dass möglichst viele Informationen auch über deutsche Suchbegriffe zugänglich werden. Dies betrifft nicht nur in den Wörterbüchern angeführte deutsche Äquivalente niedersorbischer Lemmata, sondern ebenfalls deutschsprachige semantische Definitionen, Kommentare und die Übersetzungen der niedersorbischen Beispiele. Dabei wird bei der Suche ein gezielter Zugriff auf in verschiedenen Mikrostrukturen enthaltene Autosemantika ermöglicht. So findet sich etwa *Bart* in Mukas Wörterbuch (neben vielen anderen Stellen) als direktes Äquivalent zu *broda*, als Teil der semantischen Paraphrase ‘der große, hässliche, ungepflegte, verwilderte Bart’ des Augmentativs *brožisko* oder als Teil eines Kommentars zum Äquivalent ‘Bauer’: ‘denn der wendische Bauer trägt auch heute noch zumeist keinen Bart‘ von *nébrodak* (neben ‘der Mann ohne Bart’). Zur Ermöglichung differenzierter Abfragen sollen weitere Klassifizierungen möglicher Suchwörter vorgenommen werden, so z.B. die gesonderte Kennzeichnung von Namen oder die unterschiedliche Erfassung von Ausdrücken in objekt- oder metasprachlicher Funktion [vgl. *Schimpfwort* als Äquivalent zu *gódlo* bzw. als Teil der semantischen Explikation ‘Schimpfwort für einen unbeholfenen Menschen’ zu *kóleno* (zweite Bedeutung neben ‘Knie’)]. Für den der Suchroutine zugrunde liegenden Index müssen außerdem flektierte Formen auf Grundformen zurückgeführt (*Frau des Pfarrers* à *Pfarrer*) und teilweise zusätzlich orthografisch modernisiert werden: (z.B. *mutwillig* statt *muthwillig*, *Karussell*

<sup>13</sup> Das Projekt wurde in groben Zügen beschrieben in Bartels 2012 und Szczepański 2012.

statt *Carouffel*). Der Nutzwert der Internet-Version niedersorbisch-deutscher Wörterbücher kann so weiter erheblich gesteigert werden.

### 2.5.2 Niedersorbische Phraseologie

Die digitalisierten Wörterbücher sowie diverse an verschiedenen Orten publizierte Listen oder Sammlungen (z.B. in Hauptmann 1761) enthalten eine große Anzahl von (unterschiedlich stark) idiomatischen Ausdrücken und erläuternden Informationen. Diese teilweise schwer auffindbaren Wendungen werden im Projekt systematisch erfasst und über eine XML-Modellierung so aufbereitet, dass ein mehrfacher und sicherer Zugriff ermöglicht wird. Daneben enthält auch das DNW zahlreiche bisher nicht lexikografisch erfasste niedersorbische Phraseologismen. Da es sich jedoch um ein deutsch-niedersorbisches Wörterbuch handelt, finden sich unter der dortigen Mikrostrukturposition „Wendungen“<sup>14</sup> in erster Linie deutsche (mehr oder weniger) idiomatische Ausdrücke. Die jeweiligen niedersorbischen Entsprechungen können ebenfalls idiomatisch sein (z.B. *aus dem Ärmel schütteln* = *z kapsy wutšěść*), häufig handelt es sich aber um nicht-idiomatische Ausdrücke (z.B. *die Zeit totschiagen* = *cas pšepóraš*). Daher finden sich viele der im DNW enthaltenen niedersorbischen Phraseologismen nicht im Abschnitt „Wendungen“, sondern – mit einem Stern<sup>15</sup> gekennzeichnet – unter den normalen Beispielen, wie im folgenden Abschnitt aus dem DNW illustriert wird:

#### Zeit

1. <i>Frist, Dauer</i>	<i>cas m 20</i>
Die Zeit vergeht.	Cas se minjo.
Die Zeit drängt.	Cas chwata / šišći.
Wir gehen schweren Zeiten entgegen.	My žomy šěžkim casam napšešiwu.
Er gab ihnen sehr viel Zeit.	Wón jo dał jim *cas a chylu.
in der Zeit zwischen Weihnachten und Neujahr	*mjazy gódami
Seit längerer Zeit schon beobachten wir dich.	Južo dlejšy cas / *Južo nadlej tebe wobglědujomy.
Bis dahin ist noch viel Zeit.	Až do togo jo hyšći wjele casa / *dlujka noga.
Es ist Zeit ins Bett zu gehen.	Jo cas do póstole (hyš). / *To se do pjerja zwóni.

Die auf diese Weise im DNW enthaltenen niedersorbischen Idiome (hier: *cas a chylu*, *mjazy gódami*, *dlujka noga byš*, *do pjerja se zwóniš*) sind bislang jedoch weder systematisch auffindbar (die Zuordnung folgt ja der deutschen Vorlage) noch lexikografisch beschrieben. Beides ist in einem deutsch-niedersorbischen Wörterbuch auch nicht üblich, soll aber jetzt gemeinsam mit dem anderen Material geleistet werden. Dazu wird der ge-

<sup>14</sup> Im DNW wird bewusst der weit gefasste Begriff „Wendungen“ genutzt, um deutsche Ausdrücke, die durch sehr unterschiedliche Grade von Idiomatizität gekennzeichnet sind, dort zusammenfassend mit möglichen niedersorbischen Äquivalenten aufführen zu können.

<sup>15</sup> Der Asterisk \* kennzeichnet, grob gesagt, Beispiele oder Teile von Beispielen, die keine reguläre Anwendung der zuvor genannten Äquivalente – im obigen Fall nur *cas* – darstellen. Näheres dazu siehe Bartels 2010b.

samte Bestand, unabhängig von der Präsentationsform in der jeweiligen Quelle (als Zitat, ganzer Satz, in Kurzform usw.), nach einheitlichen Prinzipien bearbeitet. Vorgesehen sind u.a. (a) eine standardisierte niedersorbische Grundform (z.B. *hyšći dlujka noga byś*), (b) eine niedersorbische Bedeutungsbeschreibung (z.B. *hyšći wjele casa byś*), (c) eine deutsche Bedeutungsbeschreibung (z.B. *noch viel Zeit sein*), (d) falls vorhanden, eine entsprechende deutsche Wendung in ebenfalls standardisierter Grundform, (e) eine wörtliche Übersetzung der standardisierten niedersorbischen Grundform, wenn es keine gleichartige deutsche Entsprechung gibt (z.B. „*noch ein langes Bein sein*“), (f) die in den Quellen vorgefundenen Originalbelege. Die schon in den anderen lexikografischen Vorhaben praktizierte Datenmodellierung im XML-Format wird vielfältige Zugriffsmöglichkeiten gewährleisten, sodass ein auf diese Weise erfasster idiomatischer Ausdruck mit hoher Wahrscheinlichkeit gefunden werden kann, und zwar unabhängig davon, ob man vom Deutschen oder vom Niedersorbischen ausgeht. Folgende Zugänge sind vorgesehen: (a) ein Zugriff über alle in der standardisierten niedersorbischen Grundform enthaltenen Substantive, Adjektive, Adverbien und Verben (z.B. *dlujki, noga, byś*); falls eine entsprechende deutsche Wendung existiert, auch über die dort enthaltenen Wörter; (b) ein Zugriff über alle in der niedersorbischen und deutschen Bedeutungsbeschreibung enthaltenen Substantive, Adjektive, Adverbien und Verben (z.B. *cas, byś, Zeit, sein*); (c) die in einigen Fällen vorhandene zusätzliche wörtliche Übersetzung der niedersorbischen Wendung erlaubt weitere Zugriffsmöglichkeiten (z.B. *lang, Bein*)<sup>16</sup>. Diese Zugriffsmöglichkeiten gehen weit über die herkömmlicher phraseologischer Wörterbücher hinaus.

Das in diesem Ausbauprojekt entstehende Internet-Wörterbuch, das 2013 bereitgestellt werden soll, versteht sich zunächst als eine systematisierende Materialsammlung und -aufbereitung. Eine vollständige wissenschaftliche Bearbeitung des Materials wird erst später, idealerweise im Rahmen des geplanten historisch-dokumentierenden Wörterbuchs erfolgen, wo phraseologische Ausdrücke zusätzlich zum Zugriff über in ihnen enthaltene autosemantische Wörter (Verweise) den Status einer eigenständigen lexikalischen Einheit (Lemma) haben werden.

### 3. *SorbLex* Niedersorbisch

Der Titel „*SorbLex* Niedersorbisch“ steht für ein geplantes langfristiges lexikografisches Vorhaben der Zweigstelle: die schrittweise Erarbeitung ein historisch-dokumentierenden Wortschatz-Informationssystems („Wörterbuchs“) der niedersorbischen Sprache. An dieser Stelle kann nur ein kurzer Einblick gegeben werden, eine ausführliche Beschreibung des Projekts erfolgt an anderer Stelle (Bartels 2013). Es ist zudem derzeit offen, ob das Gesamtvorhaben realisiert werden kann. Wahrscheinlich lassen sich aber zumindest Teile des Projekts umsetzen.

Das bereits erwähnte Niedersorbisch-deutsche Wörterbuch von Starosta [1999] war ein wichtiger Schritt auf dem Weg zu einer umfassenden lexikografischen Beschreibung des niedersorbischen Wortschatzes, musste aber trotz der etwa 45.000 Stichwörter, wie der Verfasser selbst hervorhebt, „unvollständig“ bleiben, da das verfügbare Schrifttum seiner-

<sup>16</sup> Selbstverständlich müssen diese Zugriffsvarianten in den späteren Suchabfragen unterschieden werden bzw. bei Bedarf unterscheidbar sein.

zeit nicht „umfassend“ lexikografisch ausgewertet werden konnte; außerdem wurden Wörter, deren Bedeutung z.B. aufgrund ihrer Wortbildungsstruktur leicht erschließbar sind, häufig nicht aufgenommen [ebd.: 5]. Auch die in Abschnitt 2.4 erwähnten älteren Wörterbücher, darunter insbesondere das von Mucke [1911–15/1926, 1928], sind unverzichtbare Vorarbeiten, weshalb ihre Retrodigitalisierung zugleich eine wichtige Vorbereitung von *SorbLex* darstellt. Keinem der Autoren stand aber das niedersorbische Schrifttum in der Vollständigkeit und vor allem in der Form zur Verfügung, wie sie durch die Digitalisierungsmaßnahmen der vergangenen Jahre erreicht werden konnte. Daher kann und soll *SorbLex* erstmals den gesamten dokumentierten Wortschatz von den ersten überlieferten Schriftzeugnissen aus der ersten Hälfte des 16. Jahrhunderts bis zur Gegenwart vollständig darstellen.

*SorbLex* ist nicht als Druck-Wörterbuch, sondern von vornherein als Wortschatz-Informationssystem im Internet konzipiert. Die Art der Datenverarbeitung und der lexikografischen Bearbeitung führen zu einem polyfunktionalen und polyakzessiven „Wörterbuch“: so sind neben der alphabetischen Zugriffsstruktur ebenso sprachimmanente Gliederungen des Wortschatzes vorgesehen, die entsprechende Zugänge auf die Daten ermöglichen. Dies betrifft in erster Linie eine systematische Analyse nach Wort(bildungs)familien. Daneben ist auch ein onomasiologischer Zugang zu den Daten geplant. Da *SorbLex* überdies die Eigenschaften eines historischen Wörterbuchs mit denen einer synchronen Beschreibung des heutigen Wortschatzes verbinden wird, werden bei der Benutzung des „Wörterbuchs“ und damit bei der Betrachtung des niedersorbischen Wortschatzes Wechsel der Perspektive möglich sein, so zum Beispiel von der lexikografischen Beschreibung eines einzelnen Lexems zur zugehörigen Wortfamilie, von dort zu den anderen Gliedern der Familie oder zu Teilwortfamilien usw. – und dies in synchroner oder diachroner Darstellung.

Um die für dieses Projekt notwendige komplexe lexikografische Bearbeitung effektiv gestalten zu können, ist zunächst eine gründliche Vorbereitung des Textkorpus [vgl. 2.2] notwendig. Dies beinhaltet u.a. die Korrektur einzelner Texte und die formale Vereinheitlichung zu unterschiedlichen Zeiten in das Korpus integrierter Texte, eine Segmentierung (Satzgrenzenerkennung, Tokenisierung), eine Analyse der historischen Schreibungen und die Entwicklung eines graphematischen Hyper-Systems, morphematische Analyse und Lemmatisierung. Allein diese notwendigen Maßnahmen sowie eine entsprechende Annotation des Korpus werden mehrere Jahre in Anspruch nehmen.

#### 4. [www.niedersorbisch|dolnoserbski.de](http://www.niedersorbisch|dolnoserbski.de)

Bis 2010 führte die Internet-Adresse [www.niedersorbisch.de](http://www.niedersorbisch.de) [oder [www.dolnoserbski.de](http://www.dolnoserbski.de)] einzig und ohne Umweg zum DNW. Nach Bereitstellung des Textkorpus im Internet Ende 2010 und der Veröffentlichung der Internet-Version älterer niedersorbisch-deutscher Wörterbücher Anfang 2012 wurde die Seite, wie schon lange geplant, zum “Niedersorbisch-Portal” ausgebaut. Von hier aus sind nun alle von der Zweigstelle erarbeiteten Internet-Ressourcen zum Niedersorbischen zugänglich, hier werden in Zukunft auch weitere Informationen bereitgestellt.

## 5. Resümee & Ausblick

Im Jahr 2003, vor also fast einer Dekade, haben die damals drei sprachwissenschaftlichen Mitarbeiter der Cottbuser Zweigstelle einen kurzen Beitrag über Stand und Perspektiven der Lexikografie des Niedersorbischen [*Stan i perspektywy leksykografii dolnołużyckiej*; Bartels, Spieß & Starosta 2004] für die Warschauer Zeitschrift *Zeszyty Łużyckie* verfasst. Pläne und Hoffnungen wurden seinerzeit vor allem durch „skromne środki finansowe“ und „njedostatek wykwalifikowanego personelu“ [S. 41] relativiert. Zwei große Aufgaben wurden damals benannt: Der dringend erforderliche Ausbau des Textkorpus und eine Datensammlung und -aufbereitung, die ihre effektive Nutzung für zukünftige Forschungen ermöglicht.

Nach zehn Jahren kann man sagen: die Fortschritte sind trotz der erwähnten Hemmnisse erfreulich. Das DNW ist seitdem gewaltig gewachsen [siehe 2.1], die Idee zu einem Projekt wie *SorbLex* lag damals schon in der Luft, allerdings schien ein solches Vorhaben eher unrealistisch. Das schriftsprachliche Korpus enthielt damals (mühsam erarbeitete) fünf Millionen Textwörter (tokens) und es hieß: „Większość piśmiennictwa dolnołużyckiego [...] dotychczas nie jest zawarta w istniejącym korpusie. To dotyczy między innymi głównego czasopisma Łużyc Dolnych, mianowicie tygodnika „Bramborski Serbski Casnik““ [S. 40]. Die Verhältnisse haben sich hier, wie in Abschnitt 2.2 dargestellt, völlig umgekehrt, das „dotychczas nie“ können wir streichen: So ist z.B. der erwähnte Casnik, der ja unter verschiedenen Namen bis heute erscheint, von 1848 bis 1970 bereits vollständig digitalisiert.

Diese Erfolge waren nur dank zusätzlicher Drittmittel möglich. Dadurch konnten nicht nur viele Texte digitalisiert, sondern vor allem auch zusätzliches Personal eingestellt und weitergebildet werden, sodass der in den *Zeszyty* beklagte Mangel an qualifiziertem Personal zwar generell in der Sorabistik weiter ein Problem bleibt, uns aber zur Zeit nicht ernsthaft behindert. Die derzeit laufenden Drittmittelprojekte [vgl. 2.3 und 2.5] fügen dem in den vergangenen Jahren aufgebauten Datenbestand zum Niedersorbischen wichtige ergänzende Komponenten hinzu und schaffen damit eine breitere Basis für zukünftige Forschungen. Eine möglichst umfassende Dokumentation und Beschreibung des Niedersorbischen wird weiter Hauptziel der Cottbuser Zweigstelle sein.

## Literaturverzeichnis

- Austin, Peter K. & Sallabank, Julia (eds.) (2011), *The Cambridge Handbook of Endangered Languages*, Cambridge, New York et al.: CUP.
- Bartels, Hauke (2013), *Ku koncepciji historisko-dokumentirujućego informaciskego sistema za dolnosorbški słowoskład. Plany k wótpóranju nuznego slěžeńskego deziderata*, [in:] “Lětopis” 60 (2013) 1, 16–26.
- Bartels, Hauke (2012), *Digitalizacija staršych dolnosorbško-nimskich słownikow a dokumentacija dolnosorbščiny*, [in:] Моторний, В.; Шольце, Д. (ред.): *Питання сорабістики. Матеріали XIII Міжнародного сорабістичного семінару 2010*. Львів-Будишин, 10–25.
- Bartels, Hauke (2010b), *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 (2010a), *Das (diachrone) Textkorpus der niedersorbischen Schriftsprache als Grundlage für Sprachdokumentation und Sprachwandelforschung*, [in:] Hansen, B. Grković-Major, J. (eds.), *Diachronic Slavonic Syntax. Gradual Changes in Focus*, München, Berlin, Wien: Otto Sagner [Wiener Slawistischer Almanach. Sonderbände 74], 7–18.
- Bartels, Hauke (2009), *Rěčna dokumentacija dolnosorbšćiny: pšigótowanje na cas bžez mamino-rěcnych?* [in:] “Rozhlad” 10/2009, 376–379.
- Bartels, Hauke, Spieß, Gunter & Starosta, Manfred (2004), *Stan i perspektywy dolnołužyckiej leksykografii*, [in:] “Zeszyty Łużyckie”, 37–38/2004, 33–43.
- Bartels, Hauke & Thorquindt-Stumpf, Kamil (2013), *Ein neues Ton- und Textarchiv des muttersprachlich-dialektalen Niedersorbischen*, [in:] “Lětopis” 60 (2013) 1, 39–60.
- Dietrich, Rainer (2004), *Zweitsprache – Fremdsprache*, [in:] Ammon, Ulrich et al. (Hrsg.): *Soziolinguistik. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*. 2., vollständig neu bearbeitete und erweiterte Auflage. 1. Teilband. Berlin, New York: Walter de Gruyter, 311–313.
- Grinevald, Colette & Bert, Michel (2011), *Speakers and communities*, [in:] Austin, Peter K. & Sallabank, Julia (2011), 45–65.
- Haig, Geoffrey; Nau, Nicole; Schnell, Stefan & Wegener, Claudia (2011), *Documenting endangered languages before, during, and after the DoBeS programme*, [in:] Dies. (eds.) (2011), *Documenting Endangered Languages. Achievements and Perspectives*, Berlin/Boston: de Gruyter Mouton, 1–14.
- Hauptmann, Johann Gottlieb (1761), *Nieder-Lausitzsche Wendische Grammatica. (Fotomechanischer Neudruck*, Bautzen: Domowina-Verlag 1984).
- Himmelmann, Nikolaus P. (2006), *Language documentation: What is it and what is it good for?*, [in:] Gippert, Jost; Himmelmann, Nikolaus P. & Mosel, Ulrike (eds.), *Essentials of Language documentation*, Berlin, New York: Mouton de Gruyter 2006, 1–30.
- Jannasch, Klaus-Peter / Janaš, Pěš (1990), *Wörterbuch Deutsch-Niedersorbisch*, Bautzen: Domowina-Verlag.
- Jodlbauer, Ralph; Spieß, Gunter & Steenwijk, Han (2001), *Die aktuelle Situation der niedersorbischen Sprache. Ergebnisse einer soziolinguistischen Untersuchung der Jahre 1993–1995*. Bautzen: Domowina-Verlag.
- Klein, Wolfgang (2001), *Typen und Konzepte des Spracherwerbs*, [in:] Helbig, Gerhard et al. (Hg.), *Deutsch als Fremdsprache. Ein internationales Handbuch*, Teilband 1. Berlin, New York: Walter de Gruyter, 604–617.
- Mosel, Ulrike (2006), *Fieldwork and community language work*, [in:] Gippert, Jost, Himmelmann, Nikolaus P. & Mosel, Ulrike (eds.), *Essentials of Language documentation*. Berlin, New York: Mouton de Gruyter 2006, 1–30.
- Mucke, Ernst / Muka, Arnošt (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 und 3 – Prag 1928: Verlag der böhmischen Akademie für Wissenschaften und Kunst. (Fotomechanischer Neudruck. Bautzen: Domowina-Verlag 2008)
- Starosta, Manfred (1997), *(Nje)wotkšyta słowna bogatosć*, [in:] “Pratyja” 1997, 89–90.
- 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), *Dolnoserbski-němski słownik*, Budyšin.
- Woodbury, Anthony C. (2011), *Language documentation*, [in:] Austin & Sallabank (2011), 159–186.

- 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 1989).

Hauke Bartels  
Sorbisches Institut, Zweigstelle für niedersorbische Forschungen  
August-Bebel-Str. 82, D-03046 Cottbus/Chóšebuz  
E-Mail: bartels@sorbisches-institut.de

HAUKE BARTELS

### Measures for language documentation of Lower Sorbian

#### Summary

Lower Sorbian is one of the most endangered European languages. The article states the necessity and urgency of a comprehensive documentation of this language and gives an overview of respective projects undertaken at the Lower Sorbian department of the Sorbian Institute. Apart from the building of text corpora representing the literary language as well as dialectal forms of Lower Sorbian, lexicographic projects are also described.

**Keywords:** Sorbian, Lower Sorbian, language documentation, lexicography, corpus linguistics