

Korpora bedrohter Sprachen als eierlegende Wollmilchsau? Das Beispiel GENIE

Roland Marti/Bistra Andreeva/William Barry (Saarbrücken)

Abstract

The question of documentation is central to any work with corpora of threatened languages and has important consequences for the corpus structure. The 'Korpus für GEsprochenes NIEdersorbisch (engl.: Corpus for spoken Lower Sorbian)' (GENIE, www.coli.uni-saarland.de/genie/) is a case in point. GENIE focuses on *spoken* language. The corpus, made up of more than 60 hours of recordings, comprises a) recordings made by Sorbian Radio (1956–2006) with different forms of Lower Sorbian, b) dialect recordings from the Serbski kulturny archiw (engl.: Sorbian Culture Archive) (SKA) in Budysin/Bautzen (1951–1971), and c) new recordings, also of the dialect forms of Lower Sorbian (2005–2006).

These data support the analysis of

- the relationship of dialects to the standard form of Lower Sorbian (phonetic/phonological, diachronic/synchronic),
- the relationship of spoken to written Sorbian,
- the influence of German,
- the development of idiolectal forms.

The corpus is also of direct practical use in

- the WITAJ project for the revitalization of Lower Sorbian,
- decisions on codification (spelling, grammar).

Problems for analysis stem from the eclectic nature of the recordings, the considerable variation in recording quality and in the lack of annotation. The talk discusses the perspectives for and the problems facing spoken corpora of threatened languages.

1 Allgemeines

Sprachliche Korpora haben in erster Linie die Funktion, für linguistische Analysen eine klar definierte und abgegrenzte Basis zur Verfügung zu stellen. Historisch haben sie Vorläufer in der Erstellung von Textsammlungen für klassische Sprachen sowie in Texten, die für Konkordanzen bzw. Karteien exzerpiert wurden. Einen "Quantensprung" in der Nutzung von Korpora brachte die Digitalisierung. Einen weiteren Fortschritt stellt die "Vergesellschaftung" von Korpora dar, die im Netz für die allgemeine Nutzung zur Verfügung gestellt werden. Beide sind für den Aufschwung der Korpuslinguistik in den letzten Jahren verantwortlich.

Eine weitere Entwicklung gab es hinsichtlich der Form der Sprache, die in Korpora gesammelt wurde. Waren es zunächst ausschließlich Korpora geschriebener Texte, so kamen, vor allem durch die Dialektologie, im 19. Jahrhundert Korpora verschriftlichter gesprochener Texte dazu. Sie wurden zunächst schriftsprachlich, aber zunehmend mit speziellen Transkriptionssystemen¹ aufgezeichnet. Auch hier brachte die Digitalisierung grundlegende Veränderungen, indem Korpora gesprochener Sprache nun auch als Audio-Dateien angelegt werden konnten.² Ein spezifisches Problem ist dabei die Notwendigkeit der Verschriftlichung, da

¹ Beispiele dafür sind im europäischen Raum IPA, SUT/FUT (Suomalais-ugrilainen tarkekirjoitus, Finnisch-Ugrische Transkription), Teuthonista auf weitgehend lateinischer Grundlage, die russische Dialekt-Transkription auf kyrillischer Basis sowie die gemischte Ščerba-Transkription (Ščerba 1912: IX–X).

² Diese Entwicklung kann man etwa am Beispiel des Deutschen von den Wenker-Sätzen (1876–1881) über das Zwirner-Korpus (1932, 1955–1960) zum heutigen Archiv für gesprochenes Deutsch verfolgen (vgl. Institut für

auch auditive und akustische Analysen eine vorhergehende Transkription (zumindest orthographisch, möglichst auch phonologisch) voraussetzen.³

Korpora, die (auch) die Gegenwartssprache berücksichtigen, wurden ursprünglich hauptsächlich für "große" Sprachen angelegt, da nur hier die entsprechenden Mittel für ihre Erstellung zur Verfügung standen und nur hier die Nachfrage entsprechend groß war. Hier konnte es auch zum Entstehen paralleler Korpora kommen (vgl. etwa die verschiedenen Korpora für Englisch oder Deutsch). Die Digitalisierung, und hier insbesondere die Möglichkeit, gedruckte Texte einlesen zu können, änderte auch dies, so dass selbst für "kleine" Sprachen verhältnismäßig große Korpora entstanden. Der Unterschied liegt heute im Wesentlichen darin, dass für "große" Sprachen in der Regel mehr annotierte Korpora zur Verfügung stehen, während es sich bei den "kleinen" Sprachen eher um reine Textkorpora handelt.

Einen Sonderfall innerhalb der "kleinen" Sprachen stellen die "bedrohten" Sprachen dar, und das gilt auch für die Korpora.⁴ Während bei "großen" wie "kleinen" Sprachen die Dokumentationsfunktion der Korpora sekundär ist und bestenfalls für historische Sprachzustände eine Rolle spielt, steht sie bei bedrohten Sprachen (leider) immer im Mittelpunkt. Das hängt einerseits damit zusammen, dass bedrohte Sprachen relativ rasch zu ausgestorbenen Sprachen werden können und damit die entsprechenden Korpora zu den letzten Zeugnissen dieser Sprachen werden. Zum andern ist die sprachtragende Bevölkerung meist mehrsprachig, und damit sind die bedrohten Sprachen starken Einflüssen seitens der dominanten Sprache(n) ausgesetzt, was ihre Struktur verändern kann ("language attrition"). Diese Sonderstellung hat auch Konsequenzen für die Ausgestaltung von Korpora. Dies soll im Folgenden am Beispiel des Niedersorbischen und speziell des Korpus GENIE (GESprochenes NIEdersorbisch) (Marti 2006–) gezeigt werden.

2 Sorbisch

Sorbisch ist der heute geographisch am weitesten westlich liegende Teil des slavischen Sprachgebiets. Es ist heute eine Sprachinsel (genauer: Inselgruppe) innerhalb des Deutschen, die in der Ober- und Niederlausitz liegt und die den Rest eines ursprünglich viel umfangreicheren Gebiets darstellt, das durch Sprachwechsel, vielfach ausgelöst und befördert durch sprachenpolitische Maßnahmen, allmählich germanisiert wurde, ein Prozess, der auch heute noch anhält. Das Sprachgebiet lässt sich grob in Ober- und Niedersorbisch teilen, wobei es lediglich im obersorbischen Bereich, und zwar in den katholischen Gebieten, noch Dörfer gibt, wo Sorbisch die Umgangssprache ist (vgl. zu dieser Sprachform Scholze 2008); anderswo ist es nur noch Familiensprache bzw. die Sprache der älteren Generation(en). Die Zahl der sprachtragenden Bevölkerung, die Sorbisch aktiv beherrscht, kann nur geschätzt werden: die Schätzungen liegen zwischen 15'000 und 30'000 für Obersorbisch bzw. 5'000 und 10'000 für Niedersorbisch (vgl. Elle 2000: 18, Jodlbauer/Spieß/Steenwijk 2001). Sowohl Ober- als auch Niedersorbisch sind selbstständige Standardsprachen⁵ und als Minderheiten-

Deutsche Sprache). Die weitere Entwicklung, nämlich die Berücksichtigung von Video-Dateien, spielt im sorbischen Kontext (noch) keine Rolle.

³ "Most corpora keep well away from the problems of spoken language – with some honourable exceptions – and, for a corpus which in any way purports to reflect a 'state of the language', this is most unfortunate" (Sinclair 1991: 16). Der Grund für diese Zurückhaltung ist genau das Problem der Transkription, die faktisch immer noch manuell erfolgen muss. Dazu noch einmal J. Sinclair: "Even at that time [1961! RM, BA, WB], I was assured that an automatic transcription of speech was 'just around the corner'. It still is." Vgl. zu Korpora gesprochener Sprache allgemein Leech/Myers/Thomas 1995, Thompson 2005 und Menge 1993, zu phonetischen Korpora Hess 1993.

⁴ Im Gegensatz zu Flora und Fauna, wo die Begrifflichkeit relativ klar definiert ist, fehlen für "bedrohte" Sprachen, die nach dem Vorbild der Biologie kategorisiert werden, eindeutige Parameter. Es gibt zwar für Sprachen in den entsprechenden biologischen Listen vergleichbares *UNESCO Red Book on Endangered Languages* (Salminen 1993–1999), aber die Kriterien für die Aufnahme bestimmter Sprachen in das Verzeichnis bzw. die Anwendung vorgeschlagener Kriterien sind nirgends klar definiert, und dies trotz zahlreicher Versuche, objektiverbare Skalen zu entwickeln (vgl. Lewis 2005 und die dort zitierte Literatur). Das Niedersorbische, um das es im Folgenden geht, ist im *Red Book* als "endangered" aufgeführt, und Untersuchungen vor Ort bestätigen diesen Befund (Jodlbauer/Spieß/Steenwijk 2001). Wir gehen deshalb im Folgenden ohne weitere Diskussion davon aus, dass das Niedersorbische eine bedrohte Sprache ist. (Zur Terminologie: wir geben "endangered" mit "bedroht" und nicht mit "gefährdet" wieder, weil Ersteres im Deutschen die für diese Sprachen gebräuchliche Bezeichnung ist.)

⁵ Umstritten ist, ob sich die beiden Standardsprachen aus einem Kontinuum heraus gebildet haben oder ob es sich ursprünglich um (zumindest partiell) selbstständige Kontinua handelte, die später in Kontakt traten.

sprachen in Deutschland offiziell anerkannt, einerseits in den Verfassungen und entsprechenden Sorben- (bzw. Sorben/Wenden)-Gesetzen des Freistaats Sachsen und des Landes Brandenburg,⁶ andererseits über die Europäische Charta der Regional- oder Minderheitensprachen.

Sorbisch war seit Beginn der deutschen Herrschaft (d. h. seit dem 10. Jahrhundert) einem Assimilationsdruck ausgesetzt, doch wurde die Sprachenfrage erst im Gefolge der Reformation mit ihrer Hinwendung zu den Volkssprachen akut. Das Problem war, dass die Landeskirchen deutschsprachig waren, die sorbischsprachigen Untertanen aber in der Regel kein Deutsch verstanden. Dies führte zu verschiedenen Maßnahmen seitens der deutschsprachigen kirchlichen und staatlichen Obrigkeit, von Sprachverboten über zielgerichtete Germanisierungspolitik und gelegentliche (widerwillige) Toleranz bis hin zu (marginaler) Förderung des Sorbischen. Entscheidend für die heutige Situation sind die Entwicklungen im 19. und 20. Jahrhundert, insbesondere die sorbische Wiedergeburtbewegung, die planmäßigen Germanisierungsbestrebungen im Deutschen Reich und der Weimarer Republik, die im Verbot des sorbischsprachigen öffentlichen Lebens im Dritten Reich gipfelten, und die ebenso bewusste Förderung des Sorbischen in der DDR, die allerdings mit politischem Wohlverhalten erkaufte werden musste und deshalb bei Teilen der sprachtragenden Bevölkerung zu einer Verweigerungshaltung führte. Ein Nebenaspekt der Förderung in der DDR war die Zentralisierung: Der Aufbau und die Organisation des sorbischen Lebens wurde wesentlich von obersorbischem Gebiet aus betrieben und kontrolliert, was zu einer zusätzlichen Entfremdung der niedersorbischen Seite von der staatlichen Förderungspolitik führte. Seit der Wiedervereinigung wird das Sorbische vom Bund und den betroffenen Ländern über die Stiftung für das sorbische Volk weiter gefördert, wenn auch in geringerem Umfang als zu Zeiten der DDR.

Das Hauptproblem für das Sorbische ist das Schwinden der sprachtragenden Bevölkerung und das Schrumpfen des Sprachgebiets. Letzteres ist ein Phänomen, das seit dem 16. Jahrhundert verfolgt werden kann; beides schreitet seit Mitte des 19. Jahrhunderts beschleunigt voran, und weder die Wiedergeburtbewegung noch die Förderung in der DDR haben die Entwicklung aufhalten können. Gegenwärtig gibt es Sprachbewahrungs- bzw. Revitalisierungsmaßnahmen (insbesondere das sogenannte WITAJ-Projekt, vgl. Budar/Norberg 2006), die aber den sprachlichen Assimilierungsprozess bestenfalls verlangsamen können. Insbesondere im Niedersorbischen ist die Situation dramatisch, da die Weitergabe der Sprache von einer Generation zur nächsten ("intergenerational transmission") nicht mehr existiert und die Kinder über (partielle) Immersion an eine Art "sekundäre Muttersprachlichkeit" herangeführt werden.

Für das Niedersorbische existieren noch weitere spezifische Probleme. Die Wiederbelebung des sorbischen Lebens und seine Organisation nach dem zweiten Weltkrieg gingen, wie oben erwähnt, wesentlich von obersorbischem Gebiet und von obersorbischen Exponenten aus. Dies führte zu einer teilweise objektiv vorhandenen, teilweise aber auch eher gefühlten "Obersorbisierung" des kulturellen Lebens. Besonders stark wurde dies im sprachlichen Bereich empfunden. Hier führte die Rechtschreibreform von 1949–1952 zu einer Annäherung der niedersorbischen Orthographie an die obersorbische. Da in der Schule und in den Medien eine Aussprache gepflegt bzw. gefordert wurde, die sich am Schriftbild orientierte, hatte die Reform der Orthographie auch orthoepische Folgen. Der obersorbische sprachliche Einfluss verstärkte sich noch, weil Funktionäre sorbischer Organisationen und Lehrkräfte wegen des Mangels autochthoner niedersorbischer Fachleute häufig aus der Oberlausitz kamen und sich nicht immer auf die sprachlichen Besonderheiten des Niedersorbischen einließen. Im Ergebnis entstand in der Bevölkerung vielfach der Eindruck, die niedersorbische Standardsprache sei gar nicht "echtes" Niedersorbisch, sondern bestenfalls eine gesamt-sorbische Hybridsprache oder gar ein nur geringfügig dem Niedersorbischen angepasstes Obersorbisch. Dies mündete einerseits in einer Verweigerungshaltung gegenüber der staatlichen Förderung des Sorbischen und einem Rückzug der Sprache ins Privatleben und in

Deshalb finden sich in der Literatur sowohl Singular als auch Plural (bzw. in sorbischen Texten Dual!) bei der Sprachbezeichnung (vgl. die unterschiedlichen Titel bei Schuster-Šewc 1963 und Faßke 1994). Dieses Nebeneinander der beiden Vorstellungen findet sich auch in den Sorbengesetzen: Im sächsischen ist von den sorbischen Sprachen die Rede, im brandenburgischen nur von der sorbischen Sprache.

⁶ Die offizielle Bezeichnung in Brandenburg lautet "Sorben/Wenden" und "sorbisch/wendisch", da ein Teil der sprachtragenden Bevölkerung die Bezeichnung "Sorben" und "sorbisch" für die Muttersprachlichen ablehnt. Gemäß sprachwissenschaftlicher (slavistischer) Tradition wird hier nur die einfache Bezeichnung verwendet.

die Familie (bzw. sogar im bewussten Verzicht, die Sprache an die nächste Generation weiterzugeben), andererseits in dem Gefühl, man könne nicht richtig Sorbisch (eine Meinung, die bei Feldaufnahmen immer wieder zu hören ist). Daraus erklärt sich auch der Wunsch nach verstärkter Abgrenzung vom Obersorbischen, nachdem jetzt die staatliche Kontrolle über das kulturelle Leben weggefallen ist. Dies äußert sich etwa in der unterschiedlichen Terminologie ("wendisch" statt "(nieder)sorbisch"), in der Rücknahme von Teilen der Rechtschreibreform von 1949–1952 und in der Ablehnung puristischer Sprache, die als obersorbisch empfunden wird.⁷

Besonders betroffen von dieser Situation war und ist das gesprochene Niedersorbisch. Hier entwickelten sich nach 1945 mehrere "Sprachstile" bzw. Normen, die nebeneinander existieren. Zum einen ist dies das dialektale Niedersorbisch mit seinen regionalen Varianten, das, von wenigen Ausnahmen abgesehen, heute nur noch von der älteren muttersprachlichen Bevölkerung beherrscht wird. Zum andern ist es das in der DDR in Schule, Medien und öffentlichem Sprachgebrauch geforderte standardsprachliche Niedersorbische, das an der geschriebenen Sprache orientiert war (mit klassischen Beispielen von *spelling pronunciation*). Dieses "offizielle" Niedersorbische existierte in mehreren Realisierungsvarianten, je nachdem, ob es von niedersorbischen Muttersprachlichen, von ursprünglich obersorbischen Muttersprachlichen oder von ursprünglich Deutschsprachigen angewandt wurde. Nach der Wende gab es eine Rückbesinnung auf die dialektalen Wurzeln des Niedersorbischen, so dass die Standardsprache in gewissen Punkten "redialektalisiert" wurde und eine neue Norm entstand. Auch hier gibt es analog die drei Varianten, allerdings mit der Einschränkung, dass diejenigen, die heute standardsprachliches Niedersorbisch verwenden, meist nicht muttersprachlich sind und früher *spelling pronunciation* gelernt haben, was oft zu scheidialektalisierten Formen und zu einer gemischten Sprache führt. Deswegen ist das gesprochene Niedersorbisch sowohl diachron als auch synchron von großer Uneinheitlichkeit gekennzeichnet.

3 Korpora im Sorbischen

Die ältesten Korpora im Sorbischen wurden wohl zur Erstellung von Wörterbüchern bzw. Grammatiken angelegt, und zwar in der damals üblichen Form von Exzerpierungen, vgl. etwa entsprechende Hinweise im obersorbischen Wörterbuch von Pful (1866: XV–XIX) oder im niedersorbischen von Muka (1911–1926: V–VI). Das älteste erhaltene Korpus sind die sorbischen Wenker-Sätze (vgl. die Veröffentlichung Stone 2003).⁸ Im 20. Jahrhundert wurden erstmals Korpora mit Audioaufnahmen erstellt. Dabei entstanden gelegentlich auch Aufnahmen des Sorbischen, so für das Berliner Lautarchiv⁹. Spezifisch für das Sorbische angelegte Korpora gab es aber erst in der DDR. Sie wurden vom Institut für sorbische Volksforschung/Institut za serbski ludospyt in Bautzen/Budyšin aufgebaut. Zum einen handelte es sich um die Verzettelung des geschriebenen (Ober-)Sorbischen (auf Lochkarten), die für ein großes deutsch-sorbisches Wörterbuch (Jentsch/Michalk/Šěrak 1989: 5f.) und für die Grammatik des Obersorbischen (Faßke/Michalk 1981: 18) vorgenommen wurde. Zum anderen erhob das Institut für sorbische Volksforschung/Institut za serbski ludospyt für den *Sorbischen Dialektatlas/Serbski řečny atlas* sprachliche Daten auf Tonband, die dann verschriftlicht bzw. exzerpiert wurden.¹⁰ Das Korpus ist im Institut vorhanden (vgl. dazu Kap. 4). Daneben nahm die Akademie der Wissenschaften ebenfalls sorbische Dialekte auf (das sogenannte "Schall-Archiv"). Diese Aufnahmen befinden sich heute vermutlich auch im Institut. In beiden Fällen handelt es sich um Aufnahmen aus dem ober- und dem niedersorbischen Sprachgebiet.

⁷ Das führt dazu, dass lexikalische Germanismen (*lazowaś* statt *cytaś*, *hundert* statt *sto*), die umgangssprachlich immer verwendet wurden, jetzt auch im schriftlichen Gebrauch wieder erscheinen. Ähnlich ist es im grammatikalischen Bereich, etwa bei der Determination (gelegentliche Verwendung von bestimmtem und marginal auch unbestimmtem Artikel).

⁸ Obwohl Wenker eigentlich nur die *deutsche Sprache* in ihren Dialekten dokumentieren wollte, trug er ungewollt zu einer Dokumentation der *Sprachen in Deutschland* bei, da die Schulmeister in anderssprachigen Gebieten z. T. die Sätze übersetzen ließen oder selbst übersetzten. Die entsprechenden Fragebögen sind heute noch im Archiv der Forschungsstelle Deutscher Sprachatlas erhalten (Schmidt/Herrgen 2001–).

⁹ Cf. Hermann von Helmholtz-Zentrum für Kulturtechnik.

¹⁰ Die Verschriftlichungen erschienen in den Reihen *Sorbische Dialekttexte* bzw. *Studien zur sprachlichen Interferenz* (Michalk/Protze 1967, 1974), die Exzerpte gingen in die Bände des *Sorbischen Dialektatlas (SDA/ SRA)* ein.

Im Rahmen der Abwicklung der Akademie der Wissenschaften der DDR wurde auch das Institut aufgelöst und als Sorbisches Institut/Serbski institut wieder gegründet. Das Institut baut zwei Korpora des geschriebenen Sorbischen auf, je eines für Obersorbisch¹¹ und für Niedersorbisch¹². Die beiden Korpora werden laufend aktualisiert und nach Möglichkeit auch nach hinten ergänzt. Im Falle des Niedersorbischen ist vorgesehen, allmählich das gesamte gedruckt vorliegende niedersorbische Material zu erfassen. Für die Beschreibung der obersorbischen Umgangssprache des katholischen Gebiets wurde im Rahmen des Konstanzer SFB 471 ein Korpus mündlicher und schriftlicher Texte erstellt (Scholze 2008: 19). Weitere Korpora sind uns nicht bekannt; insbesondere gibt es keine neueren Korpora mit Audio-Dateien.

4 GENIE

Angesichts der prekären Lage des Niedersorbischen, die in einschlägigen Untersuchungen beschrieben wurde (Jodlbauer/Spieß/Steenwijk 2001, Norberg 1996) und die allgemein bewusst war, ließ sich absehen, dass "echte" Muttersprachlichkeit spätestens innerhalb einer Generation nicht mehr existieren würde. Dies war für die gesprochene Sprache besonders fatal, da sich "sekundäre Muttersprachlichkeit" (das durch die Revitalisierungsbemühungen angestrebte Maximalziel) gerade auf dem Gebiet der Aussprache am stärksten von der "echten" Muttersprachlichkeit unterscheiden wird.¹³ Insofern war es wichtig und außerordentlich dringend, das gesprochene Niedersorbische zu dokumentieren. Mit dieser Absicht wurde das Korpus GESprochenes NIEdersorbisch (GENIE) aufgebaut.¹⁴ Um das Korpus für die Forschung international nutzbar zu machen, sollte es im Internet zur Verfügung stehen. Aus urheber- und datenschutzrechtlichen Gründen konnte es aber nicht allgemein zugänglich gemacht werden; seine Nutzung wird auf Antrag primär für wissenschaftliche Zwecke bewilligt.

Die Ausgestaltung des Korpus sollte den oben dargelegten Besonderheiten der Situation des gesprochenen Niedersorbisch Rechnung tragen und soweit möglich auch die diachrone Ebene berücksichtigen.¹⁵ Es besteht aus drei Teilen, wobei die ersten beiden aus bestehenden Tonarchiven stammen.

Der erste Teil sind Aufnahmen aus dem Archiv des Sorbischen Rundfunks (heute Studio Cottbus/Chošebuz des Rundfunks Berlin-Brandenburg rbb, früher ORB, noch früher Rundfunk der DDR). Es handelt sich um 110 Aufnahmen aus den Jahren 1956 bis 2006. Vertreten sind sowohl dialektale als auch standardsprachliche Aufnahmen (von Personen mit muttersprachlich niedersorbischem, obersorbischem oder deutschem Hintergrund), Letztere in verschiedenen Formen der Standardsprache. Die Textsorten sind sehr unterschiedlich: Gespräch, Interview, Ansprache, Reportage usw.

Der zweite Teil stammt aus den Beständen des Sorbischen Kulturarchivs/Serbski kulturny archiw (SKA) des Sorbischen Instituts/Serbski institut. Es sind 135 Aufnahmen aus den Jahren 1951 bis 1971. Die Aufnahmen wurden vom Institut für sprachwissenschaftliche Zwecke erstellt, insbesondere für den sorbischen Sprachatlas. Es handelt sich dabei ausschließlich um Dialektaufnahmen in den klassischen Formen dialektologischer Erhebungen (Erzählung, Interview, Erfragen von Wörtern usw.).

Der dritte Teil sind Aufnahmen, die speziell für GENIE gemacht wurden. Es handelt sich um 100 Aufnahmen aus den Jahren 2005 und 2006. Es sind Gespräche, die Juro Frahnow, selbst Muttersprachler, mit meist älteren Personen führte, die in der Regel den jeweiligen örtlichen Dialekt sprechen.

¹¹ Cf. Sorbisches Institut/Serbski institut; im Internet nutzbar.

¹² Cf. Starosta et al. (2001–); zur Nutzung im Internet vorgesehen.

¹³ Der Grund dafür liegt v. a. darin, dass die Lehrkräfte im Revitalisierungsprojekt WITAJ, von wenigen Ausnahmen abgesehen, das Niedersorbische auch schon nicht mehr muttersprachlich beherrschen, sondern bestenfalls über sekundäre Muttersprachlichkeit verfügen.

¹⁴ Cf. Marti (2006–). Die Erstellung des Korpus wurde mit Mitteln der Forschungskommission der Universität des Saarlandes unterstützt, die für 2005 und 2006 bewilligt wurden. Unterstützt wurde das Unternehmen außerdem durch den Rundfunk Berlin-Brandenburg rbb und das Sorbische Institut/Serbski institut.

¹⁵ Da aus urheberrechtlichen Gründen die ältesten Sorbisch-Aufnahmen aus dem Berliner Archiv nicht übernommen werden konnten, handelt es sich allerdings nur um eine geringe diachrone Tiefe: Die Aufnahmen stammen aus den Jahren 1951 bis 2006.

Zu jeder Aufnahme existiert ein Datenblatt, das u. a. Angaben zu Textsorte, Inhalt, Aufnahmeort und -datum sowie zur Person (Geschlecht, Geburtsort, -jahr, Dialekt, sprachliche Sozialisierung) enthält, soweit sie ermittelt werden konnten. Des Weiteren finden sich im Datenblatt auch technische Daten zur Aufnahme.¹⁶ Alle Aufnahmen sind sowohl unkomprimiert (als .wav-Datei) als auch komprimiert (als .mp3-Datei) vorhanden.

Insgesamt stehen in GENIE über sechzig Stunden gesprochenes Niedersorbisch in seinen verschiedenen Ausprägungen zur Verfügung. Auch wenn die durch die Aufnahmen abgedeckte Zeitspanne nur von 1951 bis 2006 reicht, zeigt die Berücksichtigung der Geburtsdaten, dass die Diachronie wesentlich tiefer ist: Der älteste Sprecher wurde 1860 geboren (er war zum Zeitpunkt der Aufnahme 94 Jahre alt), die jüngste Sprecherin 1973. Sogar individuelle Diachronie ist verfolgbare, da einige Personen durch mehrere Aufnahmen vertreten sind, die zu unterschiedlichen Zeiten entstanden sind.

5 Beispiele aus GENIE

Aus der Beschreibung des GENIE-Korpus ist ersichtlich, dass die wissenschaftliche Auswertung des Materials mit verschiedenen Zielsetzungen verfolgt werden kann. Zum einen sind die reine Erfassung und der Vergleich struktureller Eigenschaften der vertretenen Dialektgebiete eine reizvolle Herausforderung. Auch wenn die linguistisch unkontrollierte Form der Aufzeichnungen keine erschöpfende grammatische Beschreibung zulässt, stellt das Material eine wertvolle Ergänzung zu den (nicht direkt zugänglichen) Dialektaufnahmen der DDR-Zeit dar. Eine weitere wichtige Fragestellung ist die nach der Form und Variation der gesprochenen Standardsprache, die in Abhängigkeit von der Herkunft der Sprecher eine dialektale und damit eigentlich niedersorbische, eine obersorbische oder eine deutsche Prägung annehmen kann. Fokus unserer ersten Analysen wird aber der Einfluss des Deutschen auf das gesprochene Niedersorbisch sein, ein Einfluss, der im Laufe des 20. Jahrhunderts ständig wuchs, der aber schon lange vorher vorhanden war. Vergleiche der Aufnahmen von älteren und jüngeren Personen können die Entwicklung sowohl im Ausmaß als auch in den betroffenen Merkmalen erhellen. Noch prägnanter ist der Vergleich von Aufnahmen derselben Person zu verschiedenen Aufnahmezeiten.

Bekannte lautliche Unterschiede zwischen dem Deutschen und – nach bestehenden Beschreibungen (vgl. Schwela 1906, Janaš 1984, Starosta 1991) – dem Niedersorbischen sind im Segmentellen u. a. die Vokalqualität und -quantität, der R-Laut, die Realisierung der Plosive hinsichtlich der Stimmhaftigkeit und der Aspiration sowie das Vorhandensein eines dunklen L bzw. eines [w] und der in slavischen Sprachen verbreiteten Palatalitätskorrelation. Im prosodischen Bereich sind u. a. die Besonderheiten der Intonation (Sprachmelodie) und der Wortbetonung aus meist impressionistischen Beschreibungen bekannt. Selten erwähnte, aber dennoch wichtige Unterschiede sind die Wortverkettungsmodi wie die Trennung von aufeinander treffenden Vokalen durch Glottalverschluss oder die Art der Stimmhaftigkeitsassimilation.

Als Beispiele der vorhandenen und zunehmenden Auswirkung deutschsprachigen Kontakts auf das Niedersorbische zeigen wir vier der eben erwähnten Phänomene in Äußerungen einer älteren (A, Jahrgang 1890) und einer jüngeren (B, Jahrgang 1960) Sprecherin.

Abbildung 1, eine Darstellung des Mikrophonsignals und des Spektrogramms der Äußerung *Chtož tu rolu wobžětajō* (dt. "Wer das Land bearbeitet"), weist gleich mehrere Aussprachebesonderheiten auf, die deutschen Einfluss belegen und von denen wir drei kommentieren:

1. Im Wort *rolu* wird das /r/ als uvularer Approximant [ʁ] realisiert (siehe I).
2. *wobžětajō* /'obzɛwajō/ beginnt mit einem harten (glottalen) Einsatz statt des zu erwartenden glatten Übergangs von *rolu* (siehe II) oder eines alternativ möglichen [h].
3. Das silbenfinale /b/ und das folgende silbeninitiale /z/ werden stimmlos realisiert (siehe III).

¹⁶ Länge der Aufnahme; Größe der .wav-Datei; Größe der .mp3-Datei; Abtastrate (Samplingrate) in Hz; Quantisierungsrate der Amplitude in Bits per Sample; Anzahl der Kanäle; Signal-To-Noise Ratio SNR (nur bei Dateien aus dem dritten Teil); Bitrate (.mp3-Datei) in kBit/s.

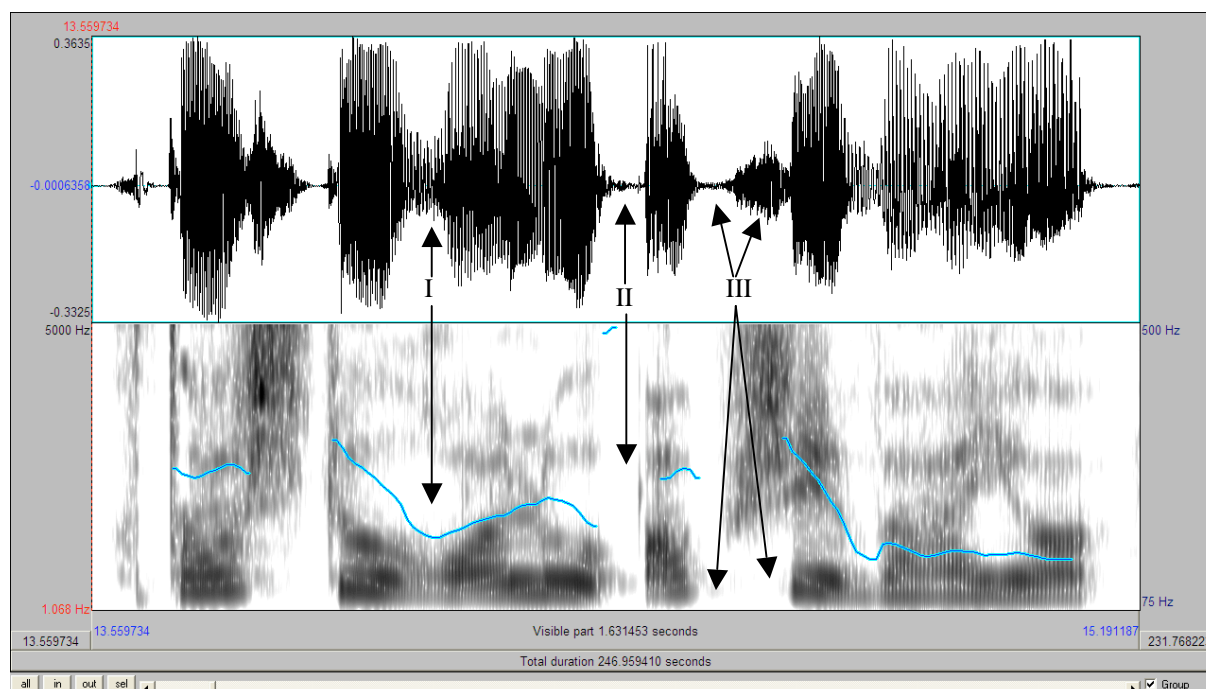


Abb. 1: Äußerung *Chtož tu rolu wobžělajo* (hier: [xtɔʃ tʊ ɛɔlu ʔɔʃɛvajo]) von Sprecherin B (Jahrgang 1960) mit (I) uvularem R, (II) hartem Vokaleinsatz (Glottalverschluss) und (III) Stimmtonverlust im Wortauslaut mit progressiver Entstimmung eines stimmhaften Initialfrikatives.

In Abbildung 2 mit dem Oszillogramm des akustischen Zeitsignals und dem Spektrogramm der Äußerung *tak daloko* (dt. "so weit") zeigen die stimmlosen Plosive /t/ (siehe I) und /k/ (siehe II) entgegen der Behauptung, dass im (Nieder-)Sorbischen stimmlose Plosive *unaspiriert* sind, klare Züge einer mittelstarken Aspiration (in beiden Fällen 26 ms). Die gemessene Aspirationsdauer ist im Vergleich zu dem, was bei monolingualen Deutschsprachigen häufig zu finden ist, relativ kurz. Es wäre deshalb zu überprüfen, ob sich allgemein im Sorbischen, bei dieser Generation oder allein bei dieser Sprecherin eine Zwischenform ähnlich der schwachen Aspiration bei kanadischen Französischsprechenden etabliert hat.

Die ältere Sprecherin (Jahrgang 1890) zeigt andere artikulatorische Muster. Bei ihr ist in Abbildung 3 (*To njejo tak dobre*, dt. "Das ist nicht so gut") zwar auch die Tendenz zur Aspiration zu erkennen: /t/ in *to* weist eine Aspirationsdauer von 37 ms auf (siehe I).

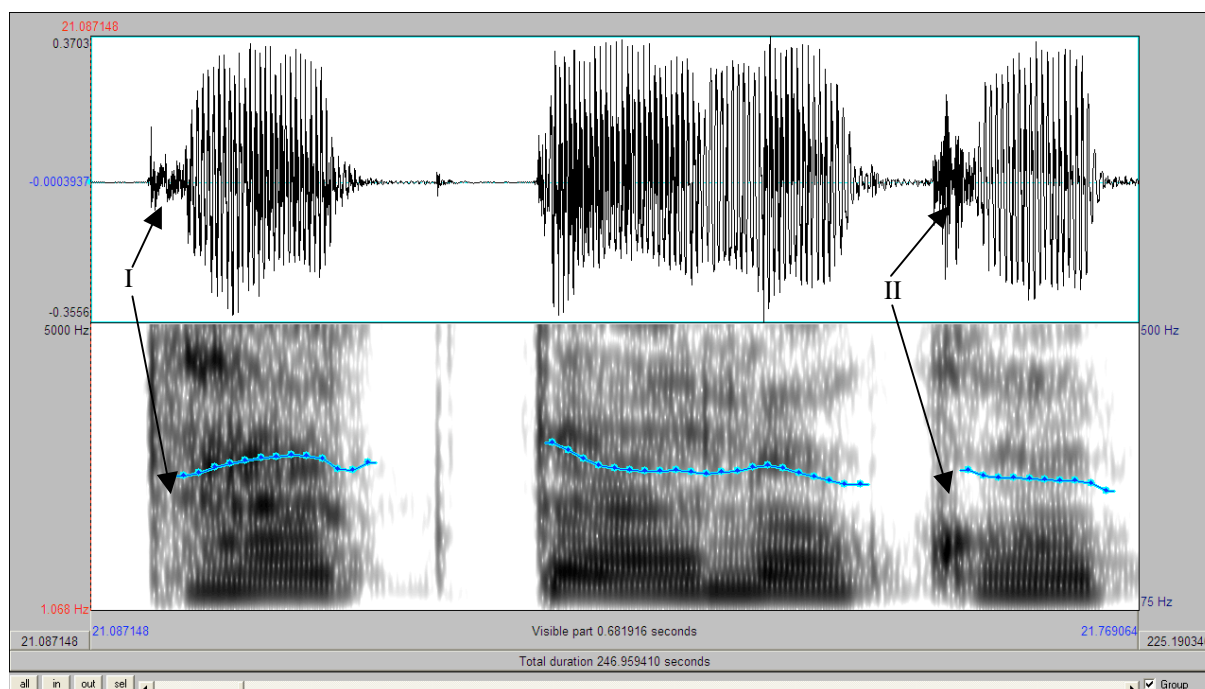


Abb. 2: Äußerung *tak daloko* (hier: [tʰak ɖalɔkʰɔ]) von Sprecherin B (Jahrgang 1960), in der die deutliche Aspiration (I) des /t/ und (II) des /k/ zu sehen ist.

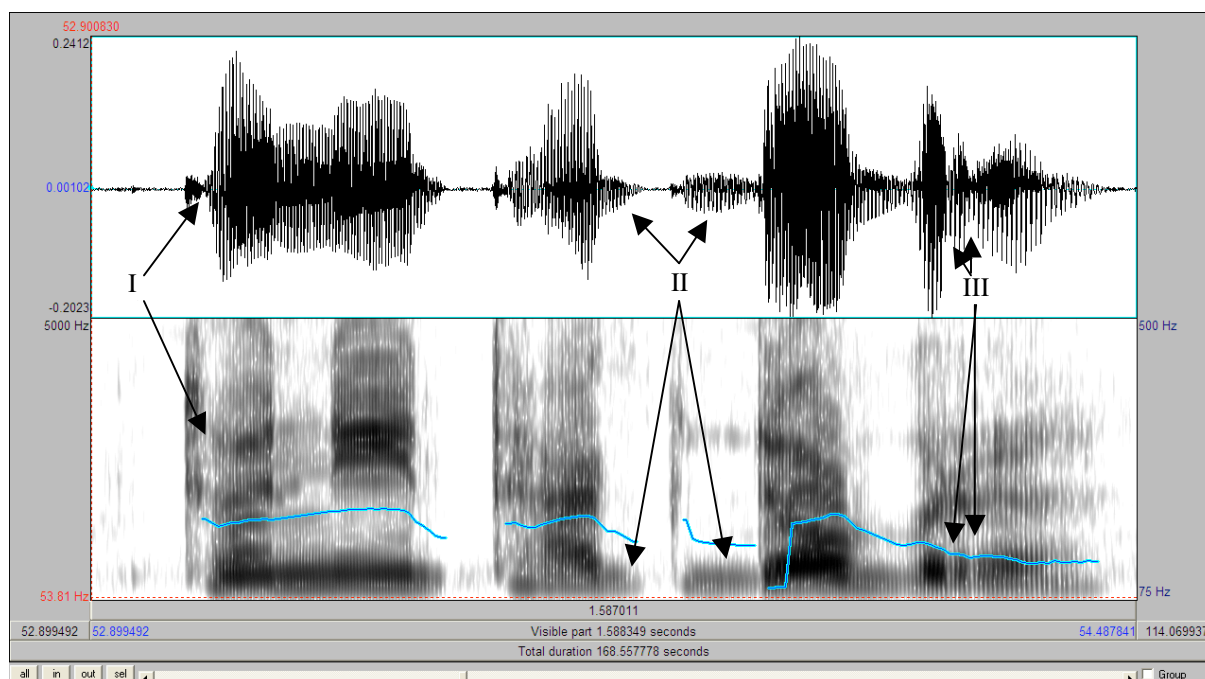


Abb. 3: Äußerung *To njejo tak dobre* (hier: [tʰɔ ne tʰak ɖɔbrɔ]) von Sprecherin A (Jahrgang 1890), in der (I) Aspiration des /t/, (II) ein voll stimmhaftes /d/ mit partieller Verstimmhaftung des vorhergehenden /k/ und (III) ein doppelschlägiges apikales /r/ zu sehen sind.

Andererseits produziert sie nach dem /k/ von *tak* ein voll stimmhaftes initiales /d/ in *dobre*, das regressiv auf das /k/ zurückwirkt und es verstimmlicht (siehe II). Das bedeutet, dass der Assimilationsprozess dem normalen deutschen Muster entgegengesetzt ist; er entspricht dem, was in anderen slavischen Sprachen üblich ist. Auch dem deutschen Standard-/r/, dem uvularen Frikativ [ʀ] entgegengesetzt ist das /r/ in *dobre*. Im Spektrogramm wie im Mikrophonesignal (siehe III) sind die zwei signaldämpfenden Schläge des apikalen /r/ zu sehen.

Was das vierte bei der jüngeren Sprecherin angesprochene Phänomen betrifft, die fehlende glatte Vokalverbindung über die Wortgrenze hinweg, lässt sich nicht behaupten, dass in früheren Zeiten glottale Konstriktion nach deutschem Muster nicht vorgekommen sei. In einer kurzen Äußerung (*a to ak*, dt. "und so wie") der Sprecherin A gibt es eine deutliche Glottalisierung am Äußerungsanfang *und* an der Wortgrenze zwischen *to* und *ak* (siehe I und II in Abbildung 4). Wie häufig solche Glottalisierungen bei ihr vorkommen, können wir noch nicht feststellen, weil die Aufzeichnung bis jetzt nur teilweise segmentiert und analysiert ist. Es ist auch nicht auszuschließen, dass sich slavische Sprachen wie andere "bindende" Sprachen (Französisch, Italienisch, Englisch usw.) und Mundarten (wie z.B. die alemannischen) verhalten, dass nämlich im emphatischen Kontext ein akzentuiertes Wort mit Initialvokal sehr wohl mit hartem Glottaleinsatz beginnen kann. Im Beispiel der jüngeren Sprecherin tritt die Glottalisierung in nicht emphatischem Kontext auf. Bei der älteren Sprecherin ist ein allgemein emphatischer "Wort-für-Wort"-Stil festzustellen. Die Äußerung ist damit zwar nicht ausgeprägt emphatisch, die Glottalisierung könnte aber dem allgemeinen Stil zuzuschreiben sein. Eine weitere Unsicherheit beim Vergleich der beiden Sprecherinnen ergibt sich aus den altersbedingten Unterschieden in der Stimmqualität, die eine Interpretation glottaler Phänomene zusätzlich erschweren.

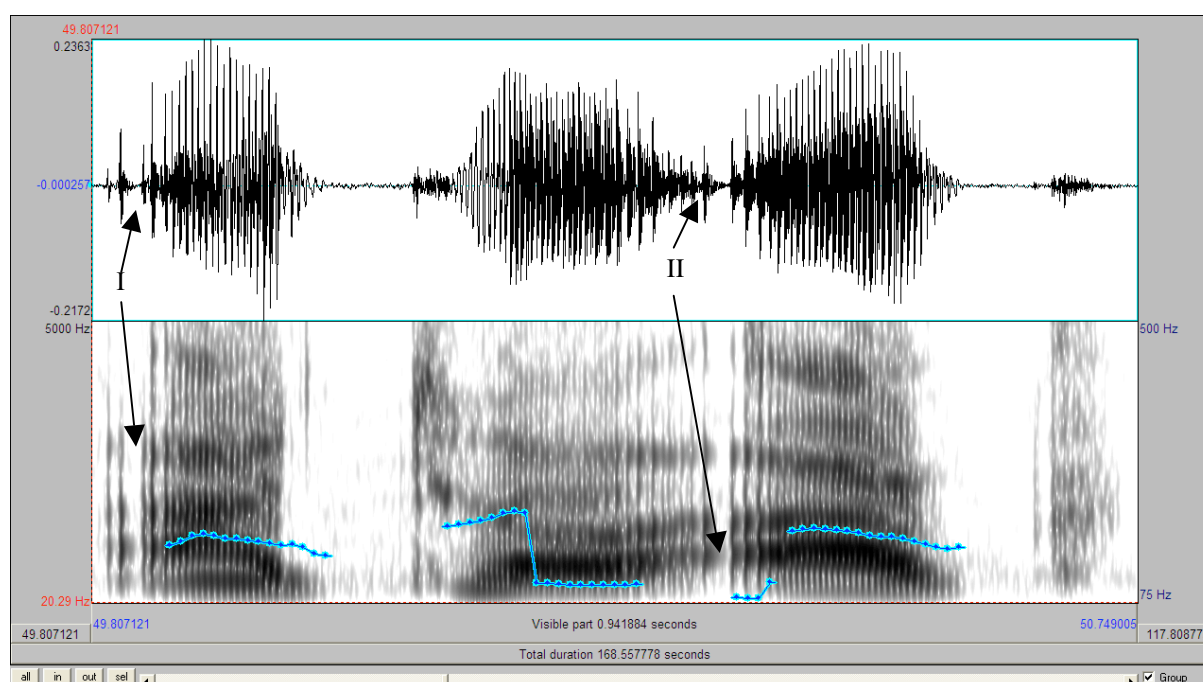


Abb. 4: Äußerung *a to ak* ([ʔa tʰɔ ʔak]) von Sprecherin A (Jahrgang 1890), in der Glottalisierung (I) am Äußerungsanfang und (II) an der Wortgrenze zwischen *to* und *ak* zu sehen ist.

6 Probleme von GENIE

Mit diesem Hinweis auf den kommunikativen Kontext, den Sprechstil und das Alter als zu berücksichtigende Faktoren bei der Suche nach den Ursachen etwaiger Unterschiede in der Aussprache sind wir bei den vielen Problemen angelangt, die in einem Korpus dieser Art zu bewältigen sind.

Ein Hauptproblem für jede Analyse ist die Uneinheitlichkeit des Materials. Unsere Beispiele stammen von zwei Sprecherinnen, die sich nicht nur hinsichtlich ihres Lebensalters unterscheiden. Die ältere ist muttersprachliche Dialektsprecherin, die zu Hause Niedersorbisch spricht. Die jüngere ist als Journalistin im Rundfunk Berufssprecherin der niedersorbischen Standardsprache, spricht aber zu Hause Deutsch. Das Ausmaß deutscher Merkmale bei der jüngeren Sprecherin (es waren auch andere neben den vier beschriebenen in der kurzen Äußerung zu beobachten) kann daher sehr wohl anderen Faktoren als dem Zeitpunkt und der Form des Spracherwerbs zuzuschreiben sein. Die Mehrdimensionalität der Sprecherkategorisierung macht jeden Vergleich schwierig und die Ursachen beobachteter Unterschiede schwer deutbar. Soweit uns Informationen zur Verfügung standen, wurde für die aufgenommenen

Personen ein reichhaltiges biographisches Profil bereitgestellt. Somit ist auf jeden Fall eine strenge Auswahl aus dem Korpus möglich.

Das Problem der Repräsentativität wurde oben schon erwähnt. Durch die Nutzung dreier unterschiedlicher Quellen (Rundfunksendungen, Dialektaufzeichnungen, eigene Aufnahmen) ist eine recht breite Streuung in mehreren Dimensionen (diachron, dialektal, thematisch) gewährleistet, was aber mit dem Nachteil erkaufte werden musste, dass pro Kategorie nur wenige Aufnahmen zur Verfügung stehen. Das birgt die Gefahr der "idiolektalen Verzerrung".

Das größte Problem von GENIE stellt allerdings die Bearbeitung des Korpus dar. Zwar ist insgesamt relativ viel Material gesammelt worden, aber die Ausnutzung dieser Datenmenge setzt eine vollständige, einheitliche und zuverlässige Annotierung voraus. Grundlegende Voraussetzung für jegliche linguistische Arbeit mit dem Korpus ist aber, dass eine orthographische Transkription des Materials vorhanden ist, und sie alleine stellt schon eine große Hürde dar. Sie und erst recht eine umfassende Annotierung erfordern erhebliche personelle und finanzielle Mittel, die leider nicht zur Verfügung stehen. Die Praxis verlangt deshalb ein stufenweises Vorgehen, bei dem für die jeweilige Fragestellung die nötige Annotierung vorgenommen wird. Eine weit schwierigere Aufgabe ist die phonetische Segmentierung, die für die gesamthafte Untersuchung der oben präsentierten Phänomene nötig ist. Eine vollständige phonetische Etikettierung des in GENIE vorhandenen Materials würde Jahrzehnte konzentrierter Arbeit erfordern. Auch hier gilt es, fragestellungsbezogen vorzugehen und auf der Basis der orthographischen Transkription Stellen zu identifizieren, die für die jeweilige phonetisch-phonologische Frage relevant sein können.

Die Hervorhebung von Problemen in dieser Diskussion soll aber nicht den Eindruck hinterlassen, dass keine wissenschaftlichen Erkenntnisse aus dem Korpus zu gewinnen seien. Zwar können nicht alle Fragen zum niedersorbischen Sprachsystem, zur Veränderung mit der Zeit oder zur regionalen Variation mit dem vorhandenen Korpus beantwortet werden. Durch sorgfältige Auswahl von Aufnahmen und konkreten Belegen sind jedoch viele neue objektivierte Erkenntnisse zu gewinnen, die auch zur Bestätigung oder Modifikation von bisher eher subjektiven Beobachtungen sowie zur Beantwortung mancher umstrittenen Frage beitragen können. Die Zugänglichkeit des Korpus und die Breite des Materials, zu dem eine große Menge spontansprachlicher Beiträge gehören, macht GENIE zu einem wichtigen Zeugnis des gesprochenen Niedersorbischen in allen seinen Varianten über einen Zeitraum von mehr als einem halben Jahrhundert bzw. indirekt sogar von anderthalb Jahrhunderten. Allerdings ist damit nicht die unverzerrte Projektion von der Performanz einer im Jahre 1950 aufgezeichneten Achtzigjährigen auf den Sprachzustand in den Jahren 1875–1880 gemeint, weil bei allen Angehörigen einer Sprachgemeinschaft im Laufe der Zeit eine gewisse Anpassung an die sich permanent verändernde Sprache stattfindet. Dabei bleiben aber, insbesondere in der individuellen Aussprache, in unterschiedlichem Maße ältere Zustände erhalten. Dies gilt insbesondere dann, wenn die Sprachgemeinschaft sehr klein ist und sich nicht mehr aus sich selbst heraus erneuert.

Die Darstellung wurde aus Platzgründen bewusst auf den lautlichen Aspekt des Niedersorbischen beschränkt. GENIE kann aber auch für die Analyse beliebiger anderer Aspekte gesprochener Sprache genutzt werden. Gerade weil es sowohl den öffentlichen als auch den privaten Sprachgebrauch dokumentiert, ist GENIE ein Korrektiv für das Korpus des geschriebenen Niedersorbisch und für die meisten Beschreibungen und Untersuchungen der Sprache, auf das nicht verzichtet werden sollte. Von besonderem Interesse ist hier natürlich die Untersuchung von Interferenzphänomenen auf allen sprachlichen Ebenen, da diese gerade im geschriebenen Niedersorbischen nach Möglichkeit ausgefiltert wurden.

7 Korpora bedrohter Sprachen – ein Sonderfall?

Nach der Präsentation eines konkreten Korpus einer bedrohten Sprache ist aus einer allgemein linguistischen Perspektive heraus zu fragen, ob sich Korpora bedrohter Sprachen oder in weiterem Sinne von Klein(st)sprachen grundlegend von Korpora anderer Sprachen unterscheiden und ob das Konsequenzen für ihre Planung, Zusammenstellung und Betreuung hat. Tatsächlich gibt es u. E. Unterschiede, aber sie sind nicht prinzipieller Art.

Ein wichtiger Unterschied betrifft die Aussagekraft oder, mit anderen Worten, die Repräsentativität von Korpora. Paradoxerweise sind Korpora bedrohter Sprachen gleichzeitig reprä-

sentativer und weniger repräsentativ als diejenigen anderer Sprachen. Der höhere Grad an Repräsentativität wird besonders bei schriftlichen Korpora deutlich. Nur bei Sprachen mit geringer schriftlicher Tradition kann ein hoher Anteil der gesamten schriftlichen Produktion in ein Korpus aufgenommen werden.¹⁷ Geringer ist die Repräsentativität aus zwei Gründen. Der eine ist außerlinguistisch: Bedrohte Sprachen sind entweder gar nicht oder dann durch verhältnismäßig kleine Korpora und nur in seltenen Fällen durch mehrere Korpora dokumentiert. Außerdem ist das, was ggf. vorhanden ist, in der Regel eher zufällig gesammelt und spiegelt nicht eine bewusste Auswahl wider. Der linguistische Grund ist darin zu sehen, dass die Norm bedrohter Sprachen weniger stabil ist und damit in ihnen größere Variabilität herrscht, die naturgemäß in den kleinen Korpora nur unvollkommen widergespiegelt wird. Es ist sogar möglich, dass idiolektale Dominanz in einem Korpus die sprachlichen Strukturen falsch abbildet.

Ein weiterer Unterschied betrifft den Korpusaufbau, die Bearbeitung und die Betreuung. Bei bedrohten Sprachen ist der Personenkreis, der sich für Korpora interessiert und sie erstellen kann, eher klein, und Ähnliches gilt für die finanziellen Möglichkeiten. Das hat zur Konsequenz, dass Korpora, wenn sie denn überhaupt zusammengestellt werden, nur wenig spezialisiert sein können (sie sind die sprichwörtliche "eierlegende Wollmilchsau") und nur zu einem geringen Teil, wenn überhaupt, annotiert werden. Die kontinuierliche Weiterbearbeitung, Aktualisierung und Dokumentierung ist auch nur in geringerem Maße gegeben.

Ein ganz gewichtiger Unterschied liegt schließlich in der Funktion der Korpora. Bei bedrohten Sprachen ist das Korpus nicht in erster Linie bzw. nicht nur ein linguistisches Arbeitsinstrument. Vielmehr ist es ein Denkmal mit ganz ausgeprägter kulturpolitischer Zielsetzung. Es soll dokumentieren, was noch ist und möglicherweise bald nicht mehr sein wird.¹⁸ Dies kann auch Konsequenzen für die Auswahl der aufzunehmenden Texte haben, wenn der "antiquarische" Gedanke obsiegt.

Korpora bedrohter Sprachen sind also durchaus ein Sonderfall. Dies ist sowohl von der Produzenten- als auch der Konsumentenseite her zu berücksichtigen. Die Hersteller müssen die einschränkenden Rahmenbedingungen und die zusätzlichen Funktionen in Rechnung stellen und dafür sorgen, dass solche Korpora trotz beschränkter Ressourcen betreut werden. Die Nutzer müssen Verständnis für die Besonderheiten solcher Korpora aufbringen, und sie sollten auch bereit sein, zu ihrer Optimierung aktiv beizutragen, indem sie z. B. Transkriptionen oder Annotierungen, die sie für ihre eigene Arbeit angefertigt haben, dem Korpus zur Verfügung stellen.

Literatur

- Budar, Ludmila/Norberg, Madlena (2006): "Les écoles sorabes après 1990". *Education et Sociétés Plurilingues* 20 (juin): 27–38.
- Faßke, Helmut (1994): "Der Weg des Sorbischen zur Schriftsprache". In: Fodor, István/Hagège, Claude (eds.): *Language Reform – History and Future* VI. Hamburg, Buske: 257–283.
- Faßke, Helmut/Michalk, Siegfried (1981): *Grammatik der obersorbischen Schriftsprache der Gegenwart. Morphologie*. Bautzen: Domowina.
- Hermann von Helmholtz-Zentrum für Kulturtechnik, Humboldt-Universität Berlin: *Lautarchiv*. www.sammlungen.hu-berlin.de/dokumente/125 (Stand: 10.01.09).
- Hess, Wolfgang (1993): "Phonetische Korpora". In: Lenders, Winfried (ed.): *Computereinsatz in der Angewandten Linguistik. Konstruktion und Weiterverarbeitung sprachlicher Korpora*. Frankfurt am Main etc.: Lang: 31–48. (= *Forum Angewandte Linguistik* 25).
- Institut für Deutsche Sprache (IDS): *Archiv für gesprochenes Deutsch*. agd.ids-mannheim.de (Stand: 10.01.09).
- Janaš, Pětr (1984): *Niedersorbische Grammatik für den Schulgebrauch*. Bautzen: Domowina.

¹⁷ Gerade im Falle des Niedersorbischen ist, wie erwähnt, vorgesehen, alle gedruckten Texte, soweit sie noch verfügbar sind, in das Korpus aufzunehmen. Ähnliche Vollständigkeit kann sonst nur noch bei Korpora "toter" Sprachen erreicht werden.

¹⁸ Es ist nicht zufällig, dass im "Archiv verschwundener Orte/archiw zgbjonych jsow" auch sorbische Sprachaufnahmen zu hören sind, um zu zeigen, was durch die "Devastierung" (Braunkohle-Tagebau) an Kulturgut verloren gegangen ist (cf. Stadt Forst (Lausitz)/Projekt Archiv verschwundener Orte ²2006).

- Jentsch, Helmut/Michalk, Siegfried/Šěrak, Irene (1989): *Deutsch-obersorbisches Wörterbuch I*. Bautzen: Domowina.
- Jodlbauer, Ralph/Spieß, Gunter/Steenwijk, Han (2001): *Die aktuelle Situation der niedersorbischen Sprache. Ergebnisse einer soziolinguistischen Untersuchung der Jahre 1993–1995*. Bautzen: Domowina. (= *Schriften des Sorbischen Instituts* 27).
- Leech, Geoffrey N./Myers, Greg/Thomas, Jenny (eds.) (1995): *Spoken English on Computer*. Harlow: Longman.
- Lewis, Paul (2005): *Towards a Categorization of the Endangerment of the World's Languages*. Ms. www.sil.org/silewp/2006/silewp2006-002.pdf.
- Marti, Roland (ed.) (2006–): *Korpus GENIE – GESprochenes NIEdersorbisch/Wendisch*. www.coli.uni-saarland.de/genie (Stand: 23.01.09).
- Menge, Heinz H. (1993): "Korpora gesprochener Sprache (Beispiel: Deutsch)". In: Lenders, Winfried (ed.): *Computereinsatz in der Angewandten Linguistik. Konstruktion und Weiterverarbeitung sprachlicher Korpora*. Frankfurt am Main etc., Lang: 15–30. (= *Forum Angewandte Linguistik* 25).
- Michalk, Siegfried/Protze, Helmut (1967): *Studien zur sprachlichen Interferenz I. Deutsch-sorbische Dialekttexte aus Nochten, Kreis Weißwasser*. Bautzen: Domowina.
- Michalk, Siegfried/Protze, Helmut (1974): *Studien zur sprachlichen Interferenz II. Deutsch-sorbische Dialekttexte aus Radibor, Kreis Bautzen*. Bautzen: Domowina.
- Muka, Ernst (1911–1926): *Słownik dolnosěrbskeje rěcy a jeje narěcow I*. Petrograd: RAN; Praha: ČAVU.
- Norberg, Madlena (1996): *Sprachwechselprozeß in der Niederlausitz. Soziolinguistische Fallstudie der deutsch-sorbischen Gemeinde Drachhausen/Hochoza*. Uppsala. (= *Acta Universitatis Upsaliensis. Studia Slavica Upsaliensia* 37).
- Pful, Krěšćan (1866): *Lužiski serbski słownik*. Budyšin: Maćica Serbska.
- Salminen, Tapani (1993–1999): *UNESCO Red Book of Endangered Languages*. www.tooyoo.l.u-tokyo.ac.jp/archive/RedBook/index.html (Stand: 10.01.09).
- Ščerba, Lev Vladimirovič (1912): *Russkie glasnye v kačestvennom i količestvennom otnoženii*. S.-Peterburg: Erlich.
- Schmidt, Jürgen E./Herrgen, Joachim (eds.) (2001–): *Digitaler Wenker Atlas (DiWA)*. Marburg: Forschungszentrum Deutscher Sprachatlas. www.diwa.info (Stand: 14.01.09).
- Scholze, Lenka (2008): *Das grammatische System der obersorbischen Umgangssprache im Sprachkontakt*. Bautzen: Domowina. (= *Schriften des Sorbischen Instituts* 45).
- Schuster-Šewc, Heinz (1963): "Die Geschichte der sorbischen Schriftsprachen. Ein Grundriß". *Slavjanska filologija* 3: 135–151.
- Schwela, Gotthold (1906): *Lehrbuch der Niederwendischen Sprache. Erster Teil: Grammatik*. Heidelberg: Ficker.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: OUP.
- Sorbisches Institut/Serbski institut: *Obersorbisches Textkorpus/Hornjoserbski tekstowy korpus*. www.serbski-institut.de/cms/de/44/Online-Publikationen → 80.153.223.118/cgi-bin/corpus (Stand: 10.01.2009).
- Stadt Forst (Lausitz)/Projekt Archiv verschwundener Orte (2006): *Dokumentation bergbaubedingter Umsiedlung in Forst (Lausitz), Ortsteil Horno*. www.forst-lausitz.de/sixcms/media.php/674/Broschuere_AVO_Aufl2.pdf (Stand: 23.01.09).
- Starosta, Manfred (1991): *Niedersorbisch schnell und intensiv 1*. Bautzen: Domowina.
- Starosta, Manfred/Hannusch, Erwin/Bartels, Hauke (2001–): *Deutsch-Niedersorbisches Wörterbuch*. Zweigstelle für niedersorbische Forschungen des Sorbischen Instituts/Wót-nožka za dolnosěrbske slěženja Serbskego instituta. www.dolnosěrbski.de/dnw/index.htm (Stand: 10.01.09).
- Stone, Gerald (2003): *Der erste Beitrag zur sorbischen Sprachgeographie. Aus dem Archiv des Deutschen Sprachatlas*. Bautzen: Domowina. (= *Lětopis* 50. Sonderheft).
- Thompson, Paul (2005): "Spoken Language Corpora". In: Wynne, Martin (ed.): *Developing Linguistic Corpora. A Guide to Good Practice*. Oxford, Oxbow Books: 59–70. ahds.ac.uk/linguistic-corpora/ (Stand: September 2009).