

Aweti Language Documentation Project

Zwischenbericht über das erste Jahr der DOBES-Hauptphase (April 2002—Februar 2003)

— Zur Information der Volkswagenstiftung und der DOBES-Gruppen —

Sebastian Drude, Hans-Heinrich Lieb, Sabine Reiter
März 2003

Einleitung

Der vorliegende Zwischenbericht soll über die Aktivitäten und Ergebnisse des *Aweti Language Documentation Projects* (kurz: Aweti-Projekts) im ersten Jahr der DOBES-Hauptphase informieren. In dieser Einleitung machen wir allgemeine Angaben zum Projekt, insbesondere zur internen Arbeitsteilung und seinen Außenbeziehungen. Genauere Angaben zum Projekt finden sich in den Anträgen zur Pilot- und Hauptphase sowie insbesondere im Abschlußbericht zur Pilotphase.

Die DOBES-Hauptphase begann für das Aweti-Projekt im April 2002. Zwischen dem Ende der Pilotphase (Ende November 2001) und diesem Zeitpunkt wurde für nur ca. zwei Monate die Möglichkeit einer Überbrückungsfinanzierung genutzt. In diesem Zeitraum wurden einige offengebliebene Aufgaben (Eingabe von handschriftlichen Daten in Datenbanken, Überarbeitung von Metadaten, Berichte an brasilianische Behörden) zu Ende geführt sowie u. a. die Abschlussfinanzabrechnung vorbereitet. Diese Zwischenphase wird im vorliegenden Bericht nicht weiter thematisiert.

Mit Beginn der Hauptphase sind im Aweti-Projekt nun zwei wissenschaftliche Mitarbeiter (1/2 BAT II a) als Forscher angestellt: wie bisher Sebastian Drude, der seit 1998 Feldforschungen bei den Aweti durchführt, und nun auch Frau Sabine Reiter, die bereits in der Pilotphase über Werkverträge am Projekt mitgearbeitet hat. Der Antragsteller, Prof. Lieb, ist nach wie vor engagiert bei allgemeinen, grundsätzli-

chen und theoretischen Fragen am Projekt beteiligt und begleitet die Projektdurchführung aufmerksam.

Prinzipiell führen nun beide Forscher im Projekt vergleichbare Aufgaben durch, wobei die umfassende Planung und die allgemeinen Aspekte der Projektarbeit, insbesondere die technische Konzipierung des Arbeitsablaufs und die Vertretung des Projekts nach außen in besonderem Maße Sebastian Drude obliegen. Frau Reiter ist jedoch gleichermaßen in die Arbeit mit den sprachlichen und kulturellen Daten (darunter Transkribierung und Übersetzung und Eingabe in digitale Datenformate) eingebunden. Darüber hinaus war Sebastian Drude neben der Arbeit am Projekt im April und Mai von dem Abschluss seiner Promotion und von Oktober bis Februar von der Lehre (6 Semesterwochenstunden) an der Freien Universität Berlin beansprucht, weswegen er den Hauptanteil seiner Arbeit während der restlichen Zeit, insbesondere während seines zweieinhalbmonatigen Feldaufenthaltes leistete und Frau Reiter im Herbst und Winter die technische Bearbeitung der Daten und Metadaten schwerpunktmäßig verantwortete. Die Aufnahme von Audio- und Videodaten erfolgte auch im Jahr 2002 fast ausschließlich durch Sebastian Drude.

Die Aweti selbst unterstützen das Projekt weiterhin nachdrücklich. Dies wurde im vergangenen Jahr eindrucksvoll dadurch dokumentiert, dass der Häuptling selbst sich als Informant zur Verfügung stellte und dafür mit einem seiner Söhne und ei-

nem weiteren Informanten für zwei Wochen nach Belém reiste.

Das Awetí-Projekt arbeitet im Rahmen des DOBES-Programms besonders eng mit zwei anderen Projekten zu Xingú-Sprachen (Kuikuro und Trumai) zusammen. Insbesondere zielen die Projekte auf die Erstellung von analogen Korpora, die als Grundlage zum Sprachvergleich dienen können. Dies ist insbesondere deswegen sinnvoll, weil die Sprachen einerseits nicht miteinander verwandt sind, andererseits aber z.T. seit Jahrhunderten in engem Kontakt miteinander stehen und in einem zu weiten Teilen ähnlichen kulturellen Kontext stehen. Auch stimmen sich die drei Projekte bei ihrer Herangehensweise im Feld und anderen Aspekten der Projektarbeit ab. Dies ist der Hintergrund dafür, dass auch auf dem jetzigen DOBES-Treffen in Frankfurt im April die drei Projekte nur durch eine Stellvertreterin vertreten sein werden, während die anderen aufgrund von anderen Aufgaben, insbesondere Feldforschungen, verhindert sind.

Die Präsentation des Awetí-Projekts in der Öffentlichkeit

Die Tätigkeiten im DOBES-Programm im Allgemeinen und im Awetí-Projekt im Besonderen stoßen auf ein reges und wachsendes Interesse in der wissenschaftlichen und breiten Öffentlichkeit. Daher ergaben sich verschiedene Gelegenheiten, über die Arbeit des Projekts und von DOBES Auskunft zu geben.

In der Form von Vorträgen vor einem Fachpublikum gab es vier größere solcher Gelegenheiten für Sebastian Drude:

- bei der Konferenz zu Language Resources and Evaluation (LREC) in Las Palmas im Mai 2002,
- im September in Belém do Pará, Brasilien, an der Bundesuniversität,

- im Februar mit einem Vortrag im Rahmen einer Ringvorlesung zu Sprachbedrohung und –dokumentation an der Universität Bielefeld und
- ein weiteres Mal im Februar durch die Teilnahme an einem Workshop zu Lexicon Schemas, Registries and Repositories in München.

In allen diesen Fällen standen Fragen der linguistischen Annotation von Sprachdaten im Vordergrund, insbesondere das Glossierungsformat „Advanced Glossing“, das Prof. Lieb und S. Drude im Rahmen von DOBES während der Pilotphase entwickelt hatten, und seine Umsetzung mit dem Shoebox-Computerprogramm.

Auch vor einer breiteren Öffentlichkeit hat sich das Awetí-Projekt wiederholt vorgestellt, so insbesondere im Mai bei einer Pressekonferenz von Prof. Lieb und S. Drude, die eigens von der Freien Universität Berlin anberaumt worden war. Hierauf, aber auch unabhängig, gab es mehrere Berichte in Zeitungen und im Radio, bei denen das Awetí-Projekt im Mittelpunkt stand.

Allgemeine Ergebnisse des Awetí-Projekts 2002

Auch im ersten Jahr der Hauptphase hat das Awetí-Projekt aktiv an der Arbeit im DOBES-Konsortium teilgenommen, auch wenn erwartungsgemäß dieser Aspekt der Projektarbeit im Vergleich zur wegbereitenden Pilotphase in den Hintergrund getreten ist.

Frau Reiter und S. Drude haben im Mai am DOBES-Workshop teilnehmen können, bei dem die Vorgehensweise und die Resultate der Pilotphase vorgestellt und die neu hinzugekommenen Teams in den DOBES-Rahmen eingeführt werden sollten. Auch hier hat S. Drude das Advanced-Glossing-Format zur potentiell maximalen Glossierung von sprachlichen Daten und seine

entsprechende Shoebox-Konfiguration vorgestellt.

Das Treffen diente auch zu Besprechungen, so mit dem TIDEL-Team am MPI Nijmegen, wo Fragen des Workflows, der Datenorganisation und insbesondere ein Feedback zu den von TIDEL entwickelten Computerprogrammen im Vordergrund standen. Ebenso wurde der Anlaß genutzt, vor den jeweils ersten Feldaufenthalten der Hauptphase weitere Details der Kooperation zwischen den ‚Xingú-Projekten‘ (s.o.) abzusprechen.

Insgesamt hat sich die Organisation der Arbeit aus der Pilotphase auch im ersten Jahr der Hauptphase bewährt. Es waren wenige Veränderungen, meist Erweiterungen, an der Konfiguration der verwendeten Programme notwendig. Dem TIDEL-Team wurden bei verschiedenen Gelegenheiten teilweise detaillierte Vorschläge für die Verbesserung der dort entwickelten Software gemacht, insbesondere im Sinne einer höheren Anwenderfreundlichkeit.

Im Sommer wurde ein Entwurf zu einem UNESCO-Grundsatzpapier zur Sprachbedrohung an die DOBES-Mitglieder verschickt. Die drei Xingú-Projekte haben daraufhin gemeinsam eine detaillierte Liste mit Änderungsvorschlägen und Bemerkungen ausgearbeitet.

Die Arbeiten an im Rahmen des Projekts entstandenen Ausarbeitungen, wie u. A. einer ethnographischen Beschreibung der Awetí, einer Beschreibung ihrer Verwandtschaftsterminologie, sowie einer überarbeiteten Fassung des Sprachdokumentationsformats „Advanced Glossing“ sind zwar fortgeschritten, aber im ersten Jahr der Hauptphase in keinem Fall zu einem Abschluss gekommen.

Dokumentation des Awetí: Ergebnisse 2002

Während des Feldaufenthalts 2002 wurde eine enorme Fülle an Audio- und Video-

Daten sowohl sprachlicher als auch kultureller Art hinzugewonnen. Das aufgenommene Material ist durchweg von hoher Aufnahmequalität und umfasst insgesamt 33 Minidisks und 24 Digital-Video-Kassetten.

In sprachlicher Hinsicht konnten verschiedene Textsorten (Autobiographien, Tätigkeitsbeschreibungen von Frauen, Mythen, Dialoge unter Männern und Kindern, Jawarí-Beschimpfungen, Gesänge) dokumentiert werden. Hinzu kam in Abstimmung mit den anderen beiden brasilianischen Projekten eine neue Wortliste (Fischnamen).

In kultureller Hinsicht handelt es sich vor allem um die umfassende Dokumentation des ersten Jawarí-Festes im Awetí-Dorf seit ca. 30 Jahren, die sämtliche Vorbereitungen (Herstellung der Strohuppe und der Jawarí-Pfeile), ritualisierte Sprechhandlungen (Auswahl des Abgesandten zur Einladung der Gäste, Empfang des Abgesandten bei seiner Rückkehr, rituelle Beschimpfungen der Jawarí-Strohuppe), Tänze und Gesänge miteinschließt. Gegen Ende des Feldaufenthalts konnte ein Kwarup-Fest des Yawalapiti-Stammes dokumentiert werden, zu dem die Awetí als Gäste geladen waren. Zusätzlich wurden frauenspezifische Alltagstätigkeiten zur Illustration der Tätigkeitsbeschreibungen gefilmt.

Im Feld wurde außerdem der bereits existierende Überblick über die demographische Situation im Awetí-Dorf aktualisiert und um einige soziolinguistische Angaben ergänzt.

Von den gewonnen Sprachdaten konnte ein Teil bereits mithilfe der Awetí-Lehrer, des Häuptlings und einer Sprecherin selbst im Feld transkribiert werden. Eine Übersetzung dieser Texte fand größtenteils in Belém am Museu Goeldi statt, wo sich der Häuptling, sein Sohn und einer der Lehrer zwei Wochen lang aufhielten. Während dieser Zeit entstanden weitere Transkriptionen neuerer und auch älterer Felddaten.

Eine summarische Auflistung der Daten

findet sich in der folgenden Tabelle.

	Sessions	Audio-dateien	Dauer	Video-dateien	Dauer
Sprache					
<i>Elizitation</i>	2	3	1:34	1	0:08
<i>Gespräche</i>	5	6	1:11	2	0:16
<i>Biographien</i>	4	11	1:27	10	1:25
<i>Erzählungen</i>	14	57	17:50	9	3:10
<i>Vorgangsbeschr.</i>	2	6	0:31	6	0:32
Kultur					
<i>Alltagskultur</i>	17	0	0:00	27	4:22
<i>Jawari-Fest 2002</i>	54	27	8:28	37	9:28
<i>Kuarup-Fest 2002</i>	11	2	1:10	11	2:46
	109	112	ca. 33h	103	ca. 22h

In Bezug auf Daten aus vorangegangenen Feldaufenthalten lässt sich festhalten, dass inzwischen zu sämtlichen Wortlisten Shoebox-Datensätze mit Transkriptionen und Übersetzungen angelegt worden sind. Dasselbe gilt für verschiedene Texte. Daten aus den Feldaufenthalten Ruth Monserrats in den 1970ern konnten bisher noch nicht ins Projektkorpus integriert werden, da es dazu weiterer Absprachen mit Frau Monserrat bedarf, mit der im September 2002 eine diesbezügliche Vorbesprechung stattfand.

Die Arbeit am Lexikon ist insofern vorangeschritten, als dass eine weitere Datenbank zur Erfassung sämtlicher bisher zu identifizierender Affixe und grammatischer Hilfsörter entstanden ist. Bereits existierende lexikalische Datenbanken sind weiter ergänzt worden.

Eine MPI-Korpusstruktur ist entwickelt worden, in der die bereits erstellten Sessions aus den Pilotphasen-Daten ihre jeweiligen Positionen erhalten haben. Zur Zeit werden im Max-Planck-Institut die neuen Rohdaten digitalisiert und die zugehörigen Sessions mithilfe der Metadaten-Angaben geschnitten. Diese können dann anhand der Angabe ihres jeweiligen Dateipfades in der Metadatei in das MPI-Korpus eingeordnet werden.

Zeitlicher Überblick über die Projektarbeit seit Beginn der DobeS-Hauptphase

April 2002

Vorbereitung des Feldaufenthalts 2002 (Anschaffung von Geräten, Impfungen, Visabeschaffung), Ausarbeitung eines Beitrags zur LREC-Konferenz

Mai 2002

Teilnahme am DobeS-Workshop in Nijmegen (Vortrag zu *Advanced Glossing*), Koordinationstreffen der drei brasilianischen Xingú-Teams in Nijmegen, Rechnungsabschluss für die DobeS-Pilot-Phase, Promotion S. Drude, Abreise S. Drude nach Brasilien

Juni 2002

Berlin: Fertigstellung der Metadaten für das Datenmaterial aus 2001, Brasilien: Dokumentation des Jawari-Festes der Awetí, Anfertigung weiterer Transkriptionen/ Übersetzungen im Feld, Ausbildung der Awetí-Lehrer (Befähigung zur eigenständigen Anfertigung von Transkriptionen), Aufnahme verschiedener Autobiographien älterer Personen

Juli 2002

Abreise S. Reiter nach Brasilien, Aktualisierung des demographischen/soziolinguistischen Überblicks, Aufnahme frauenspezifischer Tätigkeitsbeschreibungen (Verarbeitung von Maniok/Salzgewinnung) sowie Aufnahme einzelner beschriebener Tätigkeitsschritte zur Illustration, Anfertigung von Transkriptionen/ Übersetzungen neuer Daten, Elizitierung einer Wortliste (Fischnamen)

August 2002

Dokumentation des Kwarup-Festes der Yawalapiti, Aufnahmen von Awetí-Mythen mit Talakwaj im Kamayurá-Dorf (S. Drude), Anfertigung von Transkriptionen/ Übersetzungen mit den Awetí-Lehrern (S. Reiter), Abreise nach Belém, Fertigstellung des Verlängerungsantrags für die CNPq-Forschungsgenehmigung, Übertrag handschriftlicher Feld-Annotationen ins *Shoebox*-Format

September 2002

Ankunft der Awetí (Jakumin, Waranaku, Matirapa) in Belém, Transkription und Übersetzung neuerer und früherer Texte, Ausbildung von Matirapa am Computer (*Transcriber*)

Oktober 2002 bis Januar 2003

Rückkehr nach Berlin, Organisatorisches (Reisekosten-Abrechnung, Erstellung von Übersichtstabellen über Felddaten 2002), Erstellung von Sicherheitskopien der Rohdaten, Segmentierung der Rohdaten zu Sessions, Erstellung von Metadaten zu den Felddaten 2002, Überarbeitung von Transkriptionen und Übersetzungen

Februar 2003

Technische Besprechung am MPI in Nijmegen, Fertigstellung der Metadaten, Einordnung sämtlicher Sessions in die endgültige Korpusstruktur, Vorbereitung des Feldaufenthalts 2003

Bewertung

Im ersten Jahr der Hauptphase stand die Erhebung und erste Verarbeitung (Transkription, Übersetzung, Eingabe in den Computer) von einer möglichst umfangreichen Menge von Daten im Vordergrund der Projektarbeit. Insgesamt kann bei der Erhebung mit ca. 24 Stunden neuen Video- und ca. 45 Stunden neuen Audio-Aufnahmen, die ca. 110 neue „Sessions“ mit über zweihundert Medien-Dateien ergeben werden, von einem großen Erfolg gesprochen werden.

Was die erste Verarbeitung anbelangt, so ist es gelungen, über eineinhalb Stunden neues Material zu transkribieren und (vorläufig) zu übersetzen, und auch von dem alten Material konnte ca. eine weitere halbe Stunde so bearbeitet werden. (Ein großer Anteil davon ist bereits in digitalen Datenbanken eingegeben.) Da diese Arbeiten bisher nur in der Zusammenarbeit von Feldforschern und Awetí-Sprechern durchgeführt werden (können), ist dies angesichts der kurzen Zeit, in denen eine solche Zusammenarbeit möglich war, kein schlechtes Ergebnis. (Eine Minute Aufnahme braucht jetzt ca. eine bis anderthalb Stunden zur Transkription zusammen mit einer vorläufigen Übersetzung.) Allerdings ist dies immer noch unzureichend angesichts der Aufgabe einer umfassenden Dokumentation, die die unterschiedlichsten Genres beinhaltet (bisher sind fast nur monologische, standardisierte Texte bearbeitet worden), und vor allem angesichts vieler Stunden unbearbeiteter Aufnahmen von teilweise essentieller kultureller Bedeutung (mündliche Literatur wie Überlieferungen und Mythen), zu denen dieses Jahr weitere 18 Stunden gekommen sind.

Diese Diskrepanz dokumentiert eindrücklich, dass die eigenständige Teilnahme von Awetí-Sprechern dringend erwünscht ist. Die Ausbildung von Awetí zu Mitarbeitern des Projekts, die auch während der Abwesenheit der Forscher die Transkription und Übersetzung von Sprachdaten vorantreiben, hatte und hat daher eine hohe Priorität im

Aweti-Projekt. Dieses Jahr werden die Aweti neben anderen Ausrüstungsgegenständen ein robustes Laptop zum vorrangigen Zweck der Transkription erhalten.

Auch im ersten Jahr der Hauptphase gilt wieder, dass die organisatorische Verwaltung und rein technische Bearbeitung der Rohdaten, insbesondere die Erstellung von Metadaten, einen großen Anteil (bei Frau Reiter in den letzten fünf Monaten seit der Rückkehr aus Brasilien ca. 70%) der Arbeitskraft in Anspruch nimmt. Dieser Anteil ist höher als veranschlagt und sollte durch geeignete Maßnahmen reduziert

werden. Ein Beitrag von außen dazu wäre die Herstellung einer neuen Version insbesondere des Metadata-Editors durch das TIDEL-Team, die sich durch noch größere Benutzerfreundlichkeit und Stabilität auszeichnet (bisher ließ sich Datenverlust nicht ausschließen). Dies ist im Übrigen auch bereits vom MPI Nijmegen für die nächste Zeit in Aussicht gestellt worden.

Insgesamt kann festgestellt werden, daß das DOBES-Programm einen hervorragenden Rahmen zur Dokumentation des Aweti bietet, und daß die Ziele, die im Aweti-Projekt für 2002 angestrebt wurden, im Großen und Ganzen erreicht wurden.

INTERIM-REPORT: ENGLISH SUMMARY

Sabine Reiter, March 2003

This is an English summary of the interim report by S. Drude, H. Lieb and S. Reiter to inform about the results of the first year of the Dobes-Main-Phase (04/02 – 02/03)

Division of Labour/Cooperation

- Prof. Dr. Hans-Heinrich Lieb as coordinator of the project and consultant in general and theoretical matters
- since 04/02 there are 2 researchers with part-time positions (50% each): Sebastian Drude (as before), Sabine Reiter (who worked in the pilot phase of the project as a freelance linguist)
- Both researchers are involved in the processing of the linguistic and cultural data (i.a. transcription, translation, entering into digital formats) and conceptual aspects of the project (planning of work flow, administrative tasks)
- The task of representing the project in public is primarily carried out by S. Drude who also collected most of the data during his 2002 field stay of 2 ½ months
- S. Reiter who stayed less time in the field was responsible for the technical processing of the data and creation of metadata files after the return from Brazil
- The Awetí people continue to lend their support to the project, which could be discerned from the chief's readiness/willingness to work as a consultant and to travel to Belém together with his son and one of the Awetí teachers
- The Awetí-Project closely cooperates with the 2 other Xingú-Projects (Kuikuro, Trumai). The 3 Projects, working with unrelated languages of indigenous populations being in contact with each other for centuries and consequently sharing many cultural aspects, try to come up with analogous corpora which in the future may serve as a linguistic basis for comparative studies. They also strive for adjustment/agreement in other project matters (e.g. general approach towards field work). For this reason it is possible to have only one representative for the 3 projects at this Dobes-Meeting (while the other researchers could not come due to field work starting early this year).

Representation of the Awetí-Project in Public

Due to the increasing interest of a scientific/ non-scientific audience in the Awetí-Project and Dobes in general there were several opportunities to inform about Dobes and the work of the project.

- The Awetí-Project made a contribution at the Conference on Language Resources and Evaluation (LREC) in Las Palmas in May 2002 (paper read by Adrienne Dwyer).
- S. Drude held a lecture in September 2002 at the Universidade Federal do Pará in Belém/Brazil and in February 2003 at the Universität Bielefeld in a lecture programme on endangered languages and language documentation.
- In February 2003 he also participated in a workshop on lexicon schemas, registries and repositories in Munich.

At all these occasions linguistic annotation of language data and in particular the annotation format "Advanced Glossing", developed and proposed by Lieb/Drude in the Dobes context and its Shobox implementation were the focus of interest.

- In addition there was a press conference in May 2002 at the Freie Universität Berlin in which Prof. Lieb and S. Drude presented the Awetí-Project to a non-scientific audience.
- There were also several press publications and a radio programme about the project.

Documentation of Awetí: Result 2002/2003

During the field stay of 2002 a large amount of audio (33 hours) and video data (22 hours) of linguistic and cultural value was collected.

- The linguistic data include different genres of texts (autobiographies, task descriptions by women, myths, dialogs among men and (male) children, ritual swearing at Jawari, singing) and a new list of words (fish-names).
- The cultural data include the documentation of the first Jawari-celebration at the Awetí village after about 30 years and a Kwarup-celebration at the Yawalapiti village to which the Awetí had been invited. In addition the women's tasks of processing manioc and producing salt have been filmed in the various stages to illustrate the task descriptions.
- Part of the new linguistic data could already be transcribed in the field with the assistance of the two Awetí-teachers, the chief and one of the female consultants. These texts were translated when three Awetí came to visit the Museu Paraense Emílio Goeldi in Belém for two weeks. During their stay new and older field material was further transcribed.
- Progress was also achieved with regard to the lexicon work. A new database for all the affixes and grammatical (auxiliary) words so far identified, was developed. The existing lexical database was further supplemented.
- The sessions of the pilot-phase have been integrated in an MPI-Corpusstructure.
- At the moment the new raw data are being digitized and cut into sessions at the MPI.

Evaluation of Progress

- The goal set for the first year of the main-phase, to collect as diverse and as much material as possible and preliminarily process it, could be reached.
- More than two hours of transcription and preliminary translation is a good result considering the fact that there was little time and that the cooperation between the consultants and the researchers is an essential prerequisite but it is not enough if one takes into consideration that much of the data of primary importance to language and culture (esp. myths, historical narratives) is not yet processed.
- Therefore the instruction of Awetí consultants in order to make them capable of transcribing and translating data without assistance must be more emphasized in the future.
- This year the Awetí, of whom several are already familiar with the Transcriber programme, are going to receive their own laptop for this purpose.
- Another problem to the work flow of the project was the fact that administrative and purely technical tasks take a lot more time than scheduled with the result that there is little time left for the linguistic analysis. The creation of metadata, in particular, was very time-consuming. In this matter we have already consulted the TIDEL-team who are currently testing/developing a new version of the IMDI-Editor which will be more stable and get a more user-friendly surface.