

Multimodale bilinguale Korpora gesprochener Sprache: Korpuserstellung, -analyse und -dissemination in der TASX Umgebung

Zusammenfassung: Dieser Beitrag beschreibt die TASX-Korpusumgebung, ein XML-basiertes System zur Erstellung und Auswertung von großen Sprachkorpora. Das *Time Aligned Signal Data Exchange Format (TASX)* wurde speziell entwickelt für die Annotation zeitlich geordneter, multimodaler Sprachdaten. Anhand zweier exemplarisch vorgestellter multimodaler, multilingualer Korpora gesprochener Sprache werden die Anforderungen unterschiedlicher linguistischer Forschungsbereiche an die technische Unterstützung bei Korpusannotation, Korpuserstellung und Korpusdissemination diskutiert und die Lösungen in der TASX-Korpusumgebung präsentiert.

1 Einleitung

Beim Aufbau multimodaler, multilingualer Korpora ergeben sich eine Reihe von Schwierigkeiten, die auf den zum Teil erheblichen methodischen Unterschieden der linguistischen Teildisziplinen basieren. Dieser Beitrag beschreibt ein XML-basiertes System zur Erstellung und Auswertung von großen multilingualen, multimodalen Korpora, das versucht, diesen Unterschieden gerecht zu werden. Im TASX-System wird es durch einen theorieneutralen, texttechnologisch motivierten Modellierungsansatz möglich, Korpora mit unterschiedlicher Zielsetzung in einer einheitlichen Basistechnologie zu erfassen. Das so entstehende technische Rahmenwerk bietet einen Ansatz zur vereinheitlichten Beschreibung und Verarbeitung sprachlicher Daten. TASX wird derzeit in einer Reihe von Forschungsprojekten an der Universität Bielefeld verwendet, von denen zwei exemplarisch vorgestellt werden sollen. Bei den Beispielanwendungen handelt es sich einerseits um ein umfangreiches gesprächsanalytisches Korpus von natürlicher face-to-face-Interaktion zur Untersuchung der Wissenskonstruktion im bilingualen Schulunterricht, andererseits um ein experimentell erfasstes Korpus zur Analyse des Prosodieerwerbs von Fremdsprachen(L2)-Lernern.

2 Die Korpora

2.1 Ein multimodales gesprächsanalytisches Korpus zur Untersuchung der Wissenskonstruktion im bilingualen Schulunterricht

Im Rahmen des Projektes "Konstruktion von Wissen im bilingualen Schulunterricht"¹ wird ein multimodales Korpus von "natürlicher", d.h. nicht eigens für die Aufnahme-situation arrangierter, face-to-face Interaktion erstellt. Es verbindet u.a. die gesprächs-analytische² Aufbereitung und Transkription verbaler Daten (französisch, deutsch) mit einer Annotation visueller Aspekte (Mimik/Gestik/Raumverhalten etc.).

2.1.1 Ziele und Inhalte der Studie

Das Forschungsinteresse richtet sich auf die Untersuchung der interaktiven Verfahren, die Lehrer und Schüler zur gemeinsamen Erarbeitung, Weitergabe und Aneignung von Wissen im bilingualen Fachunterricht (hier: bilingual deutsch-französischer Geschichtsunterricht) verwenden. Dabei wird analysiert, wie die verschiedenen verfügbaren kommunikativen Ressourcen (Sprache, Mimik, Gestik, Blick- und Raumverhalten) von den Beteiligten in der Interaktion genutzt werden, und welche Rolle materielle Objekte (Texte, Tafelbilder etc.) in der Interaktion spielen. Ziel der Studie ist es, zum einen die "holistische Organisation" (Dausendschön-Gay & Krafft 2002; vgl. Goodwin 1996, Heath 1984) der Kommunikation in einem bilingualen Setting zu beschreiben, zum anderen anhand des empirischen Materials Kriterien zur Beurteilung des relativen Erfolgs oder Misserfolgs der Konstruktionsaktivitäten zu erarbeiten. (Pitsch (submitted)).

Die Untersuchung wird an jeweils 2 Interaktionsgruppen der Jahrgangsstufen 12 und 13 durchgeführt, deren Unterrichtsstunden mittels Audio und Video über einen zusammenhängenden dreiwöchigen Zeitraum aufgezeichnet werden. Dieses Setting lässt sich charakterisieren als:

¹ Hierbei handelt es sich um das Dissertationsprojekt von Karola Pitsch, das im Bereich Linguistik/Romanistik an der Universität Bielefeld entsteht. Kontakt: karola.pitsch@uni-bielefeld.de

² Gesprächsanalyse wird hier verstanden im Sinne Deppermanns (1999, 2000) als eine erweiterte Form der ethnomethodologischen Konversationsanalyse, d.h. als eine empirisch-qualitative Forschungsrichtung, bei der "die grundlegende[n] Konstitutionseigenschaften von Gesprächen" (Handlungscharakter, Methodizität, Sequenzialität, Interaktivität, Reflexivität) "methodisch gewendet und heuristisch fruchtbar gemacht werden" (Deppermann 1999:50).

- multispeaker – Unterrichtssituationen mit bis zu 30 Interaktanten,
- multilingual – im Rahmen sog. "bilingualer Bildungsgänge" bzw. "Abi-Bac"-Zweigen interagieren deutsche bzw. deutsch-französische Muttersprachler während des Fachunterrichts (hier: Geschichte) auf Französisch, d.h. für viele Teilnehmer: in ihrer 1. Fremdsprache; zeitweise wird aber auch das Deutsche mit einbezogen,
- multimodal – sowohl verbale als auch visuelle Aspekte sind Teil der Kommunikation.

2.1.2 Daten

Während der Feldforschungsphasen wurden große Mengen sehr unterschiedlicher Daten erhoben. Das Korpus umfasst derzeit insgesamt ca. 60 Stunden Videodaten, ca. 30 Stunden Audiomitschnitte, rund 400 Kopien der im Unterricht verwendeten Materialien (Texte, Overhead-Folien etc.) und der entstandenen Produkte (Tafelbilder, Hausaufgaben, Mitschriften etc.), Skizzen der Raumorganisation sowie eine Reihe textbasierter ethnografischer Informationen (Muttersprachen und Sprachkontakte der Teilnehmer, Schulprofil etc.) erhoben, die komplett in digitalisierter Form vorliegen. Wichtig für - auch die technisch unterstützte - Bearbeitung dieses Korpus ist, dass pro Unterrichtsstunde mit zwei getrennten Videokameras aus zwei verschiedenen Perspektiven gefilmt wurde und zusätzlich noch ein, z.T. auch zwei separate Tonmitschnitte erstellt wurden. Um nun an diesem Material die Interaktion der verschiedenen Teilnehmer, d.h. ihre aneinander orientierten Bewegungen, Gesten und Handlungen untersuchen zu können, ist ein gleichzeitiger und synchroner Zugriff auf die verschiedenen audiovisuellen Daten notwendig. Es wird also ein Programmsystem benötigt, das es erlaubt, mindestens zwei Videos parallel abzuspielen, ihre unterschiedlichen Startpunkte zu synchronisieren, in ihnen gleichzeitig und parallel zu navigieren, ihre Wiedergabegeschwindigkeit zu variieren und zwischen den verschiedenen Audiokanälen zu wechseln.

2.1.3 Annotation

Die Verschriftlichung und Annotation der beschriebenen audiovisuellen Primärdaten erfolgt entsprechend der spezifischen Methodik und Forschungslogik der Gesprächsforschung (vgl. Schmitt 2001, Deppermann 1999). Die Daten sind also so aufzubereiten, dass in der Analyse die Aktivitäten der Interaktionsteilnehmer als Lösungen für bestimmte konversationelle Probleme (in der Regel "unproblematische", weil gelöste "Probleme") rekonstruiert werden und unter den spezifischen Bedingungen ihrer Re-

kurrenz im Gesamtkorpus als methodische Verfahren der Handelnden herausgearbeitet werden können. Das bedeutet, dass sie zum einen als Grundlage einer detaillierten Sequenzanalyse, dem eigentlichen "Herzstück der Gesprächsanalyse" (Deppermann 1999:53) verwendet werden, zum anderen den größeren Interaktionsablauf wiedergeben sowie das Auffinden von Parallelstellen im Gesamtkorpus ermöglichen müssen. So entstehen als weitere Bestandteile des Korpus:

- 1 Annotationen zu Inhalt und Handlung der Interaktion, die einen Überblick über die Verlaufsstrukturen der Unterrichtsstunde ermöglichen,
- 2 Markierungen und Auszeichnungen von Passagen mit relevanten Phänomenen unterschiedlicher Länge, auf deren Basis ein systematischer Zugriff auf Vergleichspassagen im Gesamtkorpus erfolgen kann,
- 3 Transkripte der verbalen Äußerungen, angefertigt nach standardisierten Konventionen - hier: Schriftsatz des GAT (Selting et al. (1998) -, bei denen alle geäußerten Wörter und Silben in literarischer Umschrift inkl. Wortabbrüchen, Versprechern, Unterbrechungen, Intonationskonturen, Pausen etc. wiedergegeben werden,
- 4 Annotationen visueller Aspekte sowie
- 5 aus den Videodaten extrahierte, z.T. frame-genaue Bewegungsreihen zur Darstellung von Mimik/Gestik, die den entsprechenden Stellen aus Transkription und Annotation zugeordnet werden können (sog. "bebilderte Transkriptionen").

2.2 Das LeaP Korpus: Ziele und Methodik

Das LeaP-Korpus wird im Rahmen des LeaP (Learning Prosody) Forschungsprojekts³ erstellt, das den Erwerb der Prosodie von Fremdsprachenlernern untersucht, insbesondere die Intonation (Sprechmelodie), Sprachrhythmus und Betonung. Ziele des Projekts sind zum einen die Beschreibung fremdsprachlicher Prosodie mit Hilfe neuerer phonetischer und phonologischer Modelle und Theorien, und zum anderen deren Erklärung durch die Variation verschiedener Lern- und Lernerfaktoren. Zu diesem Zweck werden eine Reihe von unterschiedlichen Variablen erhoben und verschiedene Sprechergruppen bezüglich ihrer Sprachkompetenz verglichen. Die verschiedenen Untergruppen bestehen aus: Sprechern vor und nach einem längeren Auslandsaufenthalt, Sprechern vor und nach einem Prosodietrainingskurs, besonders fortgeschrittenen Sprechern, die nicht oder kaum von Muttersprachlern unterschieden werden können, Sprechern mit Anfänger- und mittlerer Kompetenz und Studierenden der Fremdsprache. Das dritte Ziel des Forschungsprojekts ist schließlich die Beschreibung des Lernverlaufs beim Prosodieerwerb, für das mehrere Experimentalstudien durchgeführt

³ www.spectrum.uni-bielefeld.de/LeaP/

werden. In ein- bis zweisemestrigen Intensivkursen erlernen und üben Teilnehmer die Prosodie der Zielsprache, wobei eine kursbegleitende Datenerhebung stattfindet. Die Intensivkurse beinhalten auch Tests zur Prosodieperzeption, d.h. den Fähigkeiten, Intonation, Rhythmus und Betonungsmuster der Zielsprache und deren Unterschiede zu anderen Sprachen wahrzunehmen. Des Weiteren soll im LeaP Projekt exemplarisch an einigen wenigen Sprechern auch das Zusammenspiel von Gesten und Prosodie bei Fremdsprachenlernern untersucht werden.

2.2.1 Daten

Zielsprachen der Sprachenlerner sind Deutsch und Englisch. Im LeaP Korpus werden vier verschiedene Sprachtypen erhoben: Wortlisten, gelesene Sprache, Nacherzählungen und spontane Sprache. Die Wortlisten bestehen aus 27 (Englisch) bzw. 32 (Deutsch) mehrsilbigen Nonsense-Wörtern, mit Hilfe derer die phonologischen Betonungsregeln der Fremdsprachenlerner erforscht werden. Der zu lesende Text ist eine kurze fabelähnliche Geschichte mit einfachem Vokabular. Bei der Nacherzählung der Geschichte müssen die Sprecher eigene Sätze formulieren, haben jedoch ein weitgehend vorgegebenes Vokabular. Das Interview erfordert völlig freie Sprache. Hier werden neben Fragen zum Alter, Geschlecht, Sprachenkenntnissen der Sprecher, Motivation, Sprachlerngeschichte, verschiedene Einstellungen und Interessen erfragt. Die Aufnahmen erfolgten in der Reihenfolge Interview, Lesetext, Nacherzählung, Wortliste. Zur Zeit besteht das Korpus aus 249 annotierten Aufnahmen von insgesamt 100 Sprechern (siehe Tabelle 1). Die Länge der einzelnen Aufnahmen variiert von 1,5 bis 35 Minuten, das Korpus insgesamt umfasst ca. 10 Stunden Sprache.

Zielsprache	annotierte Aufnahmen gesamt (ohne Wortlisten)	Sprecher	Aufnahmen von L2 Lernern	Anzahl Muttersprachen	Anzahl der Aufnahmen mit Muttersprachlern
Englisch	142	47	136	12	6
Deutsch	107	53	97	13	10

Tabelle 1: Gesamtzahl der annotierten Aufnahmen und Sprecher.

2.2.2 Annotation

Die phonetische Annotation der erhobenen Primärdaten im LeaP-Projekt erfolgt mit den Sprachanalyseprogrammen *esps/waves+* (Version 5.3.1) und *Praat* (Version 4.0.7) auf sechs verschiedenen Ebenen: Phrasen, Wörter, Silben, Segmente, Intonation und Tonumfang.

Auf der Ebene der Phrasen werden Beginn und Ende der Sprache der aufgenommenen Person und anderer Ereignisse annotiert. Andere Ereignisse sind beispielsweise Sprache des Interviewpartners, Sprechpausen, Geräusche, Atmen, Lachen und Hesitationsphänomene. Die Sprache der aufgenommenen Person wird entweder als flüssige oder als unterbrochene Phrase annotiert. Auf der Ebene der Wörter werden Beginn und Ende der einzelnen Wörter annotiert und in Orthografie (Kleinschreibung bei Deutsch) transkribiert. Auf der Ebene der Silben erfolgt die Annotation des Beginns und Endes jeder Silbe in phonetischer Transkription mit SAMPA. Die Annotation von Silben umfasst Resilbifizierungsprozesse der gesprochenen Sprache (Giegerich 1992, Kohler 1995). Auf der Ebene der Segmente werden Beginn und Ende aller vokalischen und konsonantischen Intervalle annotiert, wobei Vokale und postvokalische Halbvokale als Vokale und Plosive, Frikative, Nasale, Approximanten, Affrikate, prävokalische Halbvokale, Laterale, Trills und Retroflexe als Konsonanten klassifiziert werden. Die Determinierung von Vokalgrenzen erfolgt mit Hilfe des vom Sprachanalyseprogramm erzeugten Breitbandspektrogramms nach phonetischen Standardkriterien (z. B. Peterson & Lehiste 1960). Auf der Ebene der Intonation werden Töne nach einem ToBI-basierten Modell (Silverman et al. 1992, Grice, Baumann & Benz Müller 2001) annotiert. Dies umfasst Tonakzente und Phrasentöne. Auf der Ebene des Tonumfangs werden mit Hilfe der vom Sprachanalyseprogramm erzeugten F0-Kurve verschiedene Punkte wie höchster Punkt zu Beginn einer Phrase, Phrasenmittehochs und –tiefs und tiefster Endpunkt annotiert. Für eine Sprachaufnahme von einer Minute Länge werden so durchschnittlich 3000 Labels gesetzt.

3 Multilinguale Annotation und Korpuserstellung

Prinzipiell lässt sich der Aufbau eines multilingualen, bzw. multimodalen Korpus natürlicher Sprache in vier Phasen aufteilen:

1. Erfassung der sprachlichen Primärdaten, sowie der außersprachlichen Sekundärdaten,
2. Mehrebenen-Annotation der Primärdaten auf der Basis subjektiver Wahrnehmung unter Nutzung objektiver Messverfahren,

3. Analyse der annotierten Sprachdaten unter möglicher Einbeziehung der Sekundärdaten,
4. Aggregation der Daten zu einem Korpus und dessen (internetbasierte) Distribution.

Bei der konkreten Durchführung der Arbeitsschritte 1-4 zeigt sich schnell, dass die angedeutete lineare Abfolge der Arbeitsphasen nicht realistisch ist. Die Erstellung des Korpus lässt sich eher als ein *inkrementeller* Prozess beschreiben: Der Prozess des Annotierens ist eine ständig rekursive Aktivität, d.h. für die gleichen audiovisuellen Datenausschnitte werden zu unterschiedlichen Zeitpunkten neue Beschreibungsebenen und Verschriftlichungen hinzugefügt. Dabei müssen bereits vorgenommene Annotationen modifiziert und angeglichen werden können; das betrifft z.B. die Größe der segmentierten Einheiten und die Feingliedrigkeit ihrer Beschreibung.

Je nach linguistischem Forschungsfeld ergeben sich bei der Annotation unterschiedliche Anforderungen. In der gesprächsanalytischen Aufbereitung des Korpus sowie in der vom Menschen durchzuführenden Sequenzanalyse ist eine direkte Anbindung an die Primärdaten notwendig, d.h. - in bildlicher Formulierung - "das Hin- und Herspringen zwischen der erlebbaren Flüchtigkeit und der schriftsprachlichen Fixiertheit eines sozialen Geschehens" (Bergmann 1994:11); idealerweise geschieht dies so, dass die digitalisierten AV-Aufnahmen direkt aus dem Transkript und umgekehrt das Transkript direkt aus den Aufnahmen heraus steuerbar sind. Dabei ist eine geeignete grafische Repräsentation der Verschriftungen erforderlich. So werden z.B. die in 2.1.2 a) und b) beschriebenen Informationen traditionellerweise in einem sog. "Gesprächsinventar" in tabellarischer Form und chronologischer Reihenfolge unter Angabe des genauen Zeitpunktes ihres Auftretens und der beteiligten Akteure dargestellt. Ebenso existieren konventionalisierte Formen für die Gestaltung der Transkripte. Wichtig ist des Weiteren die Möglichkeit, die zur Validierung von gesprächsanalytischen Aussagen notwendigen Vergleichspassagen über Suchfunktionen zu identifizieren und dabei direkt auf die entsprechenden Stellen in den Primärdaten zuzugreifen.

Für die phonetische Annotation eines Korpus muss spezialisierte Software eingesetzt werden, die die Erzeugung von zur Annotation notwendigen Oszillogrammen, Spektrogrammen, pitch tracking und anderen Funktionen bietet. Üblicherweise werden die Annotationen dort von multiplen Annotatoren unter Verwendung unterschiedlicher Software erzeugt. Bei der Analyse phonetischer Phänomene sind automatische Werkzeuge notwendig, da die Datenmenge und erforderlichen mathematischen Kalkulationen für eine manuelle Auswertung ungeeignet sind. Für die Auswertung sprecherbezogener Metadaten ist eine Integration dieser Daten in das Gesamtkorpus erforderlich. Diese Daten ermöglichen z.B. die Gruppierung einzelner Sprecher und, daraus resultierend, die automatische Generierung abgeleiteter Subkorpora.

4 Datenformat

Aus den jeweiligen Fragestellungen der beschriebenen Projekte ergeben sich unterschiedliche Anforderungen an die zum Einsatz kommende Korpustechnologie. So sind beispielsweise die *Beschreibungseinheiten* sehr unterschiedlich, ebenso die zum Einsatz kommenden *Kodierungssysteme*, die *Genauigkeitsanforderungen* in der Segmentierung differieren, die *Länge* der Aufnahmen variiert zwischen den Projekten stark. Auch im *Arbeitsablauf* unterscheiden sich die Projekte: das erste Projekt steht prototypisch für eine Untersuchung im Rahmen eines Dissertationsvorhabens, LeaP dagegen wird von einer Vielzahl unterschiedlicher Mitarbeiter über einen längeren Zeitraum durchgeführt.

Um den genannten Anforderungen zu entsprechen, ist es notwendig eine geeignete informationstechnische Basis zu finden. Aktuell erscheint uns ein Ansatz, der auf den Technologien des XML-Umfeldes aufbaut, am vielversprechendsten. Mit XML werden sprachliche Daten in Form von Textdateien gespeichert, wobei diese Dateien intern baum-strukturiert sind. Eine separat vorliegende, formal definierte Dokumentgrammatik kann genutzt werden, um ein Dokument mit Hilfe eines validierenden XML-Parsers auf seine strukturelle Korrektheit hin zu überprüfen. Optional kann dabei das Dokument vom Parser in eine interne Repräsentation überführt werden. Auch für die Weiterverarbeitung XML-strukturierter Daten existieren Ansätze, die zum Teil durch das W3C oder die ISO standardisiert werden.

4.1 Das TASX-Format

Beide vorgestellten Projekte nutzen zur Repräsentation ihrer Daten das *Time Aligned Signal Data Exchange Format (TASX)*, ein XML-basiertes Format, welches speziell für die Annotation zeitlich geordneter, multimodaler Sprachdaten entwickelt wurde (Milde & Gut, 2001). Formal wird TASX durch eine DTD (Document Type Definition) definiert. Der folgende Ausschnitt der DTD zeigt die wesentlichen Element- und Attributdefinitionen von TASX:

```

<!ELEMENT tasx          (session+)>
<!ELEMENT session      ((meta*, layer)*)>
<!ELEMENT layer        ((meta*, event)*)>
<!ELEMENT event        (#PCDATA|meta)*>

<!-- Metadaten -->
<!ELEMENT meta          (desc*)>
<!ELEMENT desc          (name, val)>
<!ELEMENT name          (#PCDATA)>
<!ELEMENT val           (#PCDATA)>

<!--Attribute -->
<!ATTLIST event
  e-id CDATA #REQUIRED
  start CDATA #REQUIRED
  end CDATA #REQUIRED
>

```

Ein TASX-Korpus besteht demnach aus mindestens einer Aufnahme (*session*). Jede session besteht aus mehreren Beschreibungsschichten (*layer*), welche wiederum eine Menge von Ereignissen (*event*) enthält. Ein Event kann beliebige Textinformation tragen. Jede Session, jeder Layer und jedes Event kann optional Sekundärdaten (*meta*) enthalten. Für jedes Event werden zusätzlich eine Menge von obligatorischen XML-Attributen definiert. Mit diesen Attributen kann ein Event identifiziert (*e-id*) werden und das zeitliche Intervall des Events (*start*, *end*) festgelegt werden. Über die Attribute *start* und *end* wird also der Bezug zwischen der Annotation und der Primärdaten hergestellt.

Ein TASX-Korpus, das ausschließlich die direkte Koppelung an die Primärdaten über die *start/end* Attribute nutzt, wird als Korpus der TASX-Stufe 1 bezeichnet (*TASX-Level 1*). Die TASX-Stufe 1 löst ein wesentliches Problem der XML-basierten Annotation: die *Überlappung* von Beschreibungseinheiten. XML-Dateien sind baumstrukturiert und somit dürfen sich öffnende und schließende Elemente nicht kreuzweise überlappen, eine Beschränkung, die für eine Vielzahl linguistischer Anwendungsbereiche zu restriktiv ist. In TASX sind die Events zwar linear in der XML-Datei angeordnet, ihr Geltungsbereich wird jedoch über das in den Attributen kodierte Intervall beschrieben. Entsprechend können sich die Events zeitlich beliebig überlappen. Dieser Mechanismus ist auch dann nutzbar, wenn es sich bei den beschriebenen Primärdaten nicht um zeitlich angeordnete Daten handelt, bzw. deren genauer zeitlicher Verlauf nicht mehr rekonstruierbar ist. Um hier Überlappung von Ereignissen zu beschreiben muss für diese Daten lediglich eine linear geordnete Bezugslinie gewählt werden. Im einfachsten Fall kann dies die Folge der natürlichen Zahlen sein.

Ein TASX-Korpus, das neben der direkten Ankoppelung an die Primärdaten auch die Verlinkung von Ereignissen untereinander nutzt, wird als Korpus der TASX-Stufe 2 bezeichnet (*TASX-Level 2*). Das TASX-Stufe 2 erlaubt es, hierarchische Beziehungen zwischen den Beschreibungsebenen aufzubauen. Dies wird möglich, da die TASX-DTD keinerlei Beschränkung hinsichtlich der Werte der *start* und *end* Attribute definiert. Es können beliebige Zeichenketten verwendet werden, z.B. XPointer-Ausdrücke als Referenz auf eine andere Beschreibungsebene. Aus linguistischer Sicht kann dieser Beschreibungsansatz beispielsweise verwendet werden, um die Beziehung zwischen Wörtern und Phrasen auszudrücken. Diese ersten beiden Stufen von TASX sind in ähnlicher Weise auch mit den Ansätzen von Bird und Liberman (2001), Bird et.al. (2001), Kipp (2001), Carletta et. al. (2002) Dybkjaer (2001), Brugmann und Wittenburg (2001) und Schmidt (2001) darstellbar.

Das TASX-Format definiert eine *datenzentriertes Informationsmodell*, es fokussiert die Organisation der Daten, nicht jedoch deren interne logische Struktur. Im Modell von TASX werden die Informationen als Menge von zwei-dimensionalen Datenfeldern angeordnet, wobei eine Zeile eines Datenfeldes einer Beschreibungsschicht entspricht, in der jeweils die Ereignisse eingetragen werden. Die Ereignisbeschreibungen sind atomar und unterliegen keinerlei weiteren Beschränkungen.

Genau diese Flexibilität von TASX ermöglicht die verlustfreie Integration einer Vielzahl aktueller linguistischer Formate. Gleichzeitig stellt sie aber auch ein wesentliches Problem des datenzentrierten Ansatzes dar. Die interne Struktur der Daten ist aus Sicht von TASX arbiträr und damit innerhalb von TASX auch nicht formal beschreibbar. In der Folge bedeutet das, dass die strukturelle Integrität der repräsentierten linguistischen Daten nicht ohne weiteres überprüfbar ist. Damit gibt der TASX-Ansatz an dieser Stelle scheinbar zwei wesentliche Vorteile der Informationsmodellierung mit XML auf: die *strukturgesteuerte Datenerfassung* und die automatische *strukturelle Validierung* der Daten durch einen XML-Parser.

Mit TASX-Stufe 3 (*TASX-Level 3*) wird versucht, einen sinnvollen Kompromiss zwischen datenzentrierten und strukturzentrierten Ansatz zu finden. Zielsetzung ist es, eine Entkoppelung der XML-Elemente von TASX und den in TASX annotierten semi-strukturierten Daten zu erreichen. Hierzu wird ein XML-Namespace für TASX festgelegt. Ein XML-Parser kann so die TASX-Elemente von anderen eingebetteten XML-Elementen unterscheiden. Durch diesen Ansatz wird es möglich, XML-annotierte Baumstrukturen als Inhalt eines Ereignisses zu speichern. Dies ist dann linguistisch sinnvoll, wenn beispielsweise Ergebnisse eines syntaktischen Parsings in das Korpus integriert werden müssen. Das so aufgebautes TASX-Korpus unterliegt den Wohlgeformtheitsbedingungen von XML.

4.2 Der TASX-Annotator

Zur Annotation der multimodal vorliegenden Sprachdaten wird der TASX-Annotator verwendet (siehe Abb. 1). Das Programm ermöglicht die interaktive Annotation von Video- und Audiodateien. Dabei können im Videomodus mehrere Videos synchron bearbeitet werden. Hierdurch wird es beispielsweise möglich, Szenen aus unterschiedlichen Perspektiven aufzunehmen und diese dann parallel auszuwerten.

Eine Hauptzielsetzung bei der Entwicklung des Programms war es, das System möglichst bedienungsfreundlich zu gestalten, um so den Annotationsprozess optimal zu unterstützen und damit zu beschleunigen. Aus diesem Grund wurde einerseits eine grafische Oberfläche entwickelt, welche die Steuerung mit der Maus erlaubt, andererseits sind alle wesentlichen Funktionen auch durch die Tastatur ansteuerbar. Hierdurch wird das häufige Hin- und Herwechseln zwischen Maus und Tastatur vermieden. Schließlich ist es möglich das System mittels Fußschalter zu kontrollieren. Um die Bedienung des TASX-Annotators weiter zu vereinfachen, wurde ein kontextsensitives Hilfesystem integriert. Damit wird es für den Anfänger möglich, Elemente der grafischen Oberfläche auszuwählen und für deren Funktion beschreibende Hilfstexte anzufragen.



Abb. 1: Der TASX-Annotator ermöglicht die interaktive Erstellung multimodaler Korpora. Durch die unmittelbare Einbindung der Video- und Audiodaten in den Transkriptionsprozess vereinfacht sich die integrative Bearbeitung der betrachteten Teilmodalitäten. Die Abbildung zeigt einen Ausschnitt des bilingualen Unterrichtskorpus.

Der TASX-Annotator stellt eine Vielzahl von Funktionen zur Manipulation von Beschreibungsebenen und den darin kodierten Ereignissen zur Verfügung. So können z.B. Ebenen eingefügt, verschoben, gruppiert, kollabiert und expandiert werden, Er-

eignisse verschoben, deren Grenzen angepasst, an beliebiger Stelle aufgetrennt und miteinander verschmolzen werden. Das System erlaubt die Verwendung beliebiger Zeichensätze zur Auszeichnung der Primärdaten. Das System verwendet Unicode, so dass, wenn entsprechende Zeichensätze auf dem System installiert sind, in nahezu jeder beliebigen Sprache transkribiert werden kann.

Mit Hilfe der integrierten Suchfunktion, lassen sich komplexere Restrukturierungen der Daten durchführen, Transkriptionen können miteinander gemischt werden, das System erlaubt den Aufbau eines visuellen Index, der die Navigation in längeren Videos erheblich vereinfacht. Bilder aus Schlüsselszenen lassen sich extrahieren, um sie anschließend in der Korpusdokumentation zu verwenden.

4.3 Konversion, Integration und Dissemination der Korpusdaten

Um die mit anderen aktuellen Programmsystemen (z.B. ESPS/waves+, Praat, Multi-Tool, WavSurfer, CLAN, SyncWriter, Exmaralda etc.) annotierten Daten in der TASX-Umgebung wiederverwenden zu können, ist es notwendig, *Konversionswerkzeuge* zu entwickeln. Konzeptuell ähneln die meisten textbasierten Formate TASX-Level 1: es wird von einer Strukturierung in multiple Annotationsebenen ausgegangen, wobei die Einzelebene Ereignisse aufnimmt, welche über einen Zeitstempel an das primäre Signal gebunden sind. Die Konversionswerkzeuge arbeiten in der Regel in zwei Phasen. Zunächst werden die Textdaten durch ein Perl-Skript in eine einfache XML-annotierte Form gebracht, um anschließend mittels XSL-T nach TASX konvertiert zu werden. Die Rücktransformation erfolgt durchgängig durch XSL-T.

Neben der Konversion von Sprachdaten stellt die Integration der verteilt vorliegenden Informationen einen wesentlichen Schritt beim Aufbau eines linguistischen Korpus dar. Im Anwendungsfall LeaP besteht das Korpus z.B. aus über 2000 Einzeldateien. In der TASX-Umgebung wird zur Korpusintegration der Entitäten-Mechanismus von XML verwendet. Anstatt das Korpus in einer, möglicherweise sehr großen, Einzeldatei zusammenzufassen, wird lediglich eine zentrale *Steuerungsdatei* generiert, die Verweise auf die verteilt vorliegenden XML-Dateien der jeweiligen Einzelsessions enthält. Die generierte Steuerungsdatei definiert für jede Session eine eindeutige Entität, welche zu einer absoluten Pfadangabe expandiert. Das eigentliche Korpus besteht dann nur noch aus der Liste der so definierten Entitäten.

Die internetbasierte Dissemination der TASX-annotierten Korpusdaten ist Aufgabe des Korpusystems. Aufbauend auf einer Client/Server Architektur werden die Daten unter Nutzung von Standard-Softwarekomponenten verteilt. Das Korpusystem besteht aus zwei größeren Subkomponenten: dem *Informations-Pool* und der eigentliche *Korpus-Engine*. Der Informations-Pool nimmt die primären Sprachdaten und die TASX-annotierten Daten auf. Als aktive Komponente greift die Korpus-Engine auf Daten des Informations-Pools zu, bereitet sie auf und verteilt diese.

Der TASX-Annotator sowie alle weiteren TASX-konformen Software-Komponenten sind als OpenSource Software unter der Adresse <http://tasxforce.lili.uni-bielefeld.de> abrufbar.

5 Ausblick

Die TASX-Umgebung ist ein leistungsfähiges System für die Erstellung und Analyse sprachlicher Korpora, die auf zeitlich alignierten Sprachdaten basieren. Im Anwendungskontext der dargestellten Projekte konnte gezeigt werden, dass die konsequente Nutzung von XML-Technologie die Integration unterschiedlicher Datenquellen erleichtert, die Entwicklung von linguistisch motivierten Analysewerkzeugen vereinfacht, sowie die internetbasierte Verteilung der Ergebnisse gut unterstützt wird.

Der Nutzen für die gesprächsanalytische Arbeit - hier: die Untersuchung der Wissenskonstruktion im bilingualen Schulunterricht – und für die Erstellung von phonetisch annotierter Korpora – hier: die Untersuchung des Erwerbs der Prosodie von Fremdsprachenlernern - lässt sich auf sieben Ebenen lokalisieren. Das sind:

- die Integration und Manipulation der multiplen audiovisuellen Ausgangsdaten
- die erhöhte Flexibilität für die langfristige Nutzung der erstellten Korpora
- die Wiederverwendbarkeit bereits existierender Teilkorpora
- die erleichterte Navigation und Manipulation in einem umfangreichen Korpus
- eine flexible Datenrepräsentation: Es ist nicht mehr – wie z.B. in der Gesprächsforschung bislang mangels adäquater technischer Werkzeuge praktiziert und zu Recht als „theoriebeladen“ kritisiert – notwendig, sich mit Beginn der Verschriftung für *eine* der „*existenzielle[n]* Option[en]“ (Dittmar 2002:179) von Partiturnotation, Linearschreibweise oder Spaltenform zu entscheiden.
- die flexible grafische Präsentation der transkribierten Daten kann dem jeweiligen Analysegesichtspunkt angepasst werden, wobei die jeweils einzubeziehenden Ebenen aktuell ausgewählt werden können
- die Integration quantitativer Analysen: Auf der Basis der für die *qualitative* Analyse annotierten Daten können an gezielten Stellen *quantitative* Analysen einbezogen werden, z.B. bei der Berechnung von Redezeiten der verschiedenen Interaktanten oder der Häufigkeit der Verwendung von Muttersprache vs. Fremdsprache

Anvisierte Ziele bei der Weiterentwicklung von TASX fokussieren die Verbesserung der Korpusanfrage. Optimal wäre hier die Entwicklung einer Querysprache, die in der Lage ist, annotationenebenen-übergreifende Anfragen zu formulieren, ohne dabei die Komplexität einer allgemeinen Programmiersprache zu erreichen.

Literatur

- Bergmann, Jörg R. (1994). Ethnomethodologische Konversationsanalyse. In: Fritz, G./Hundsniischer, F. (ed.). *Handbuch der Dialoganalyse*. Tübingen, 3-16.
- Bird, Steven and Liberman, M. (1999). A Formal Framework for Linguistic Annotation, Department of Computer and Information Science, University of Pennsylvania, MS-CIS-99-01
- Bird, Steven and Kazuaki Maeda and Xiaoyi Ma. (2001). AGTK: the annotation graph toolkit, IRCS Workshop on Linguistic Databases, University of Pennsylvania, Philadelphia, USA, Steven Bird, Peter Buneman and Mark Liberman.
- Brugman, Hennie and Wittenburg, Peter. (2001). MPI tools for linguistic annotation, IRCS Workshop on Linguistic Databases, University of Pennsylvania, Philadelphia, USA, Steven Bird, Peter Buneman and Mark Liberman.
- Carletta, J. and McKelvie, D. and Isard A. (2002), Supporting linguistic annotation using XML and stylesheets., G. Sampson and D. McCarthy, *Readings in Corpus Linguistic*, Continuum International
- Dausendschön-Gay, Ulrich and Krafft, Ulrich (2002). Beobachtungen zur holistischen Organisation der Kommunikation. In: *Psychotherapie und Sozialwissenschaft* 4, 30-60.
- Deppermann, Arnulf (1999). *Gespräche analysieren. Eine Einführung in konversationsanalytische Methoden*. Opladen.
- Deppermann, Arnulf (2000). Ethnographische Gesprächsanalyse: Zu Nutzen und Notwendigkeit von Ethnographie für die Konversationsanalyse. In: *Gesprächsforschung* 1, 96-124 (www.gespraechsforschung-ozs.de).
- Dittmar, Norbert (2002): *Transkription. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien*. Opladen.
- Dybkjaer, L. and Moeller, M. B. and Bernsen, N. O. and Carletta, J. and Isard, A. and Klein, M. and McKelvie, D. and Mengel, A. (June 1999). The MATE Workbench, David Traum, *Proceedings of ACL'99, Demonstration Abstracts*. University of Maryland, , 12 - 13.
- Giegerich, H. (1992). *English Phonology*. Cambridge: Cambridge University Press.
- Goodwin, Charles (1996). Transparent vision. In: Ochs, E./Schegloff, E./Thompson, S. (ed.). *Interaction and Grammar*. Cambridge.
- Grice, M., Baumann, S. & Benz Müller, R. (2002). German Intonation in Autosegmental-Metrical Phonology. In: Jun, Sun-Ah (ed.) *Prosodic Typology*, Oxford University Press.
- Heath, Christian (1984). Talk and reciprocity: sequential organization in speech and body movement. In: Atkinson, M./Heritage, J. (ed.). *Structures of Social Action*. Cambridge, 247-265.
- Kohler, K. (1995). *Einführung in die Phonetik des Deutschen*. Berlin: Schmid (2nd edition).
- Kipp, Michael (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue, *Proceedings of the Eurospeech 2001, Aalborg*, 1367 - 1370.
- Milde, Jan-Torsten and Gut, U. B. (2001). The TASX-engine: an XML-based corpus database for time aligned language data, IRCS Workshop on Linguistic Databases, University of Pennsylvania, Philadelphia, USA, Steven Bird, Peter Buneman and Mark Liberman, 2001
- Peterson, G. and Lehiste, I. (1960): Duration of Syllable Nuclei in English. *Journal of the Acoustical Society of America* 32 (6), 693-703.

- Pitsch, Karola (submitted): Construction of knowledge in bilingual German-French history lessons: Interactive emergence of "common ground" (Tagung "Dialogues in and around multicultural schools).
- Schmidt, Thomas (2001). Gesprächstranskription auf dem Computer - das System EXMARaL-DA, Gesprächsforschung, <http://www.gespraechsforschung-ozs.de>.
- Schmitt, Reinhold (2001). Von der Videoaufzeichnung zum Konzept "Interaktives Führungshandeln": Methodische Probleme einer inhaltlich orientierten Gesprächsanalyse. In: Gesprächsforschung 2, 141-192 (www.gespraechsforschung-ozs.de).
- Selting, Margret et al. (1998). Gesprächsanalytisches Transkriptionssystem (GAT), In: Linguistische Berichte 173, 91-222.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Pierrehumbert, J. & Hirschberg, J. (1992). "ToBI: a standard for labeling English prosody". Proceedings, Second International Conference on Spoken Language Processing 2, Banff, Canada, 867-70.