

## **Korpora in der Zweitspracherwerbsforschung. Internetzugang zu Daten des ungesteuerten Zweitspracherwerbs**

Man stelle sich folgendes Szenario vor: Ein Hochschullehrer im Bereich Linguistik macht im Rahmen der Vorbereitung seiner Lehrveranstaltungen zum Zweitspracherwerb (L2-Erwerb) eine Internet-Recherche, die zur Findung geeigneter Materialien dient. Er möchte seine Studenten nicht nur mit bestimmten Theorien vertraut machen, sondern ihnen auch Einsichten in die Eigenschaften der Primärdaten vermitteln, auf denen solche Theorien fußen. Er ist erfolgreich und findet Informationen, dass bereits transkribierte und teilweise analysierte Daten in einem Spracharchiv gespeichert, und via Internet zugänglich sind. Sie können direkt erreicht, d.h. angesehen bzw. heruntergeladen werden. An der gleichen Stelle finden sich auch alle relevanten Informationen über die Umstände und Zeitpunkte der Aufnahmen sowie über die Informanten (Alter, Kontaktdauer zur L2 etc.). Auch bereits publizierte Datenanalysen werden an dieser Stelle gelistet.

Dies ist kein Zukunftsszenario mehr, sondern eine bereits existierende Möglichkeit<sup>1</sup> für den Datenzugang via Internet. Die bereits in den 80-er Jahren erstellten multimedialen L2-Sprachkorpora, die im Internet zugänglich waren bzw. immer noch sind, waren mit den üblichen Web-Browsern (via Html-Kodes) erreichbar. Zu erwähnen sind in diesem Zusammenhang die CHILDES Datenbank<sup>2</sup> für Erst- und Zweitspracherwerbsdaten und die ESF-Datenbank<sup>3</sup> für Zweitspracherwerbsdaten. CHILDES' Daten und ihre Konzepte haben einen großen Einfluss auf die Spracherwerbsforschung ausgeübt und sind bis heute eine wichtige Quelle für Daten und Standards in der Spracherwerbsforschung (vgl. dazu Rutherford/ Thomas 2001 und Myles 2005). Die ESF-Daten wurden inzwischen in eine neue Umgebung (MPI-Archiv; s. weiter unten) übertragen. Die erwähnten Datensammlungen, die mit sehr viel Aufwand erstellt wurden, dienen bis heute der empirischen Zweitspracherwerbsforschung.

Die neue Korpusgeneration zeichnet sich dadurch aus, dass die Internet Browser, mit denen man auf die Daten zugreift, mit Java-Scripts<sup>4</sup> manipuliert werden. Dadurch ist die Benutzeroberfläche für den Zugang optimal an die Erfordernisse

- 
- 1 Voraussetzung dafür sind allerdings die Zugangsrechte; manche Archivdaten sind mit Password gesichert. Man muss die zuständigen Wissenschaftler erst um Erlaubnis bitten, bevor man die Daten ansehen bzw. laden kann. Die Informationen **über** die Daten (sog. Metadaten) sind frei zugänglich.
  - 2 Vgl. McWhinney 1991: ix und <http://childes.psy.cmu.edu/> [alle Internetadressen in diesem Artikel beziehen sich auf den Stichtag 15.September 2007].
  - 3 Vgl. Perdu (1993) und Perdu (1998).
  - 4 JAVA ist eine Plattform und Programmiersprache der SUN corporation, vgl. <http://java.sun.com/>.

der Benutzer angepasst. Die Funktionalität solcher Webseiten geht weit über die Möglichkeiten hinaus, die eine html-kodierte Seite bietet.

Der vorliegende Artikel stellt L2-Daten vor, die im MPI-Archiv gespeichert und verwaltet werden. Alle besprochenen Daten sind Teil des Medienarchivs am Max-Planck-Institut in Nijmegen, Niederlande (MPI-Korpus)<sup>5</sup>. Ihre Organisation und Zugangsmöglichkeiten werden hier präsentiert. Eine kurze Inhaltsangabe zu den einzelnen Archiven wird ebenfalls angehängt.

Außer Datensammlungen, die via Internet zugänglich sind, werden L2-Daten auf CD-ROMs<sup>6</sup> ausgetauscht. Eine unbekannte Anzahl von Daten befindet sich immer noch in den Schubladen von wissenschaftlichen Institutionen und wird über persönliche Kontakte vermittelt.<sup>7</sup> Diese Daten werden hier nicht behandelt.

### **MPI-Korpus: Standards und Technologien**

Die neue Korpus-Generation für L2-Daten zeichnet sich dadurch aus, dass sie mit öffentlich zugänglichen Metadaten angereichert wurde, die online und frei zugänglich zur Verfügung<sup>8</sup> stehen.

Eine Trennung von Daten (Medien, Transkripte, Annotationen) und Metadaten bietet den Vorteil, dass man sich ausgiebig über die Daten informieren kann, bevor man sie herunterlädt. Diese Trennung bietet via Metadaten und Info-Dateien alle für eine Datenanalyse wichtigen Informationen, dient aber zugleich dem Datenschutz. Die Metadaten des MPI-Archivs sind strukturiert und können mit spezialisierten Suchmaschinen gefunden werden (vgl. Wittenburg/Broeder, 2002). So sind z.B. Suchkombinationen möglich, die die L1 und L2 miteinander kombinieren, andere Charakteristika der Lerner (wie Alter und Geschlecht) spezifizieren und nur nach bestimmten Datentypen (z.B. Instruktionen) suchen. Die Suche kann anschließend noch verfeinert werden, indem die Annotationen nach bestimmten Formen und Funktionen abgefragt werden.

Im MPI-Archiv werden Metadaten im sogenannten IMDI-Standard gespeichert (vgl. Broeder et al. 2001: 48ff.) und <http://www.mpi.nl/ISLE>). Linguistische Analyseeinheiten, genannt *sessions* (z.B. ein Interview, eine Nacherzählung oder ein thematisch abgeschlossener Teil eines Gesprächs) werden in speziellen Dateien als Metadaten gespeichert. Sie enthalten Informationen über die Medien und Annota-

5 Zur Darstellung des gesamten MPI-Korpus vgl. auch Wittenburg/Skiba/Trilsbeek (2004). Die MPI-Korpora sind zu erreichen unter: [http://corpus1.mpi.nl/ds/imdi\\_browser/](http://corpus1.mpi.nl/ds/imdi_browser/). Viele hier nicht erwähnte Artikel zu technischen und inhaltlichen Aspekten der Spracharchive sind zu finden unter <http://www.lat-mpi.eu/papers>.

6 Z.B. das Projekt RUSIL zum Erwerb des Hebräischen und Deutschen durch russische Migranten, (vgl. dazu: [http://userpage.fu-berlin.de/~nordit/HP/Ditt\\_Korpora.html](http://userpage.fu-berlin.de/~nordit/HP/Ditt_Korpora.html)). Daten zum Erwerb des Italienischen sind unter <http://lettere.unipv.it/diplinguistica/pagina.php?&id=17> zu bestellen.

7 Z.B. die ZISA-Daten werden durch persönliche Kontakte vermittelt (vgl. Mika 2005: 13).

8 Auch die alten ESF-Formate und die CHAT-Formate enthalten Metadaten; sie sind jedoch Teil der Transkripte: Die sogenannten Transkriptionsköpfe enthalten Informationen über Sprecher, Ort, Datum und Thema der jeweiligen Aufnahmen/Transkriptionen.

tionsdateien, die zu einer *session* gehören, aber auch Informationen über Inhalte der *sessions*. Eine Metadateneinheit enthält also (xml-kodierte)<sup>9</sup> Informationen über:

- die verwendete(n) Sprache(n), Aufnahmeort und -datum der *session*,
- den Inhalt der *session* (unter anderem Thema, Genre, Datenart),
- die Beteiligten (z.B. Sprecher und ihre Charakteristika, Projektinformationen),
- die Originalaufnahmen (Name und Aufbewahrungsort der Rohdaten) und die Medieneinheiten (Ton-, Bild-, Video- und Textdateien), die mit der *session* assoziiert werden.

Alle Metadaten können jederzeit mit neuen Informationen (z.B. Publikationsangaben, Annotationsdateien etc.) ergänzt werden. Es existiert auch ein Metadaten-Editor<sup>10</sup>, mit dem die Metadaten eingegeben, validiert und gespeichert werden können. Zu sich wiederholenden Metadaten (z.B. zu Informanten, die in verschiedenen *sessions* auftreten, oder zur Projektbeschreibung) können Daten in assoziierten Dateien separat gespeichert und damit wieder verwendbar gemacht werden.

Eine Meta-Beschreibung fasst also alle linguistisch und technisch relevanten Informationen zusammen, die zu einer *session* gehören, und kann deshalb zum Aufbau eines gut beschriebenen Sprachkorpus verwendet werden. Die Korpusdaten werden auf diese Art leicht auffindbar.

Archivteile (Subkorpora) werden eingerichtet und verwaltet mit Hilfe eines Programms, mit dem Datenstrukturen (Datenbäume) definiert, und die Daten selbst in das Archiv geladen werden. Das Programm LAMUS<sup>11</sup> sorgt dafür, dass die Datenstrukturen erzeugt werden und die Daten an entsprechenden Plätzen gespeichert werden. Außerdem werden automatisch Sicherheitskopien der neuen Archivteile erstellt.

Das MPI Archiv arbeitet zusammen mit der DELAMaN-Organisation. Auf der Website von DELAMaN<sup>12</sup> findet man viele wertvolle Informationen zur Dokumentation von (bedrohten) Sprachen, zur Datensicherheit<sup>13</sup> und zu auf diesen Bereich spezialisierten Programmen. Die Organisation fördert die Zusammenarbeit von mehreren Spracharchiven und kümmert sich um die Einführung von Archivierungsstandards.

### **MPI-Korpus: Daten zum natürlichen Zweitspracherwerb**

Wie oben bereits ausgeführt, sind zum Datenzugriff auf das MPI-Archiv lediglich eine Internetverbindung und ein üblicher Internet-Browser notwendig.

9 XML (Extensible Markup Language) ist eine Auszeichnungssprache zur Darstellung hierarchisch strukturierter Daten in Form von Textdateien. Mehr dazu unter <http://www.w3.org/XML/>.

10 Frei erhältlich unter <http://www.mpi.nl/tg/j2se/jnlp/IMDI-BCTools/>.

11 Vgl. dazu Broeder et al. (2006) und Berck et al. (2005).

12 The Digital Endangered Languages and Musics Archives Network gegründet im Jahre 2003. Projektwebsite unter: <http://www.delaman.org/index.html>.

13 Vgl. dazu Schüller (2004).

Die L2 - Daten sind unter der folgenden Adresse zu erreichen: [http://corpus1.mpi.nl/ds/imdi\\_browser?openpath=MPI14765%23](http://corpus1.mpi.nl/ds/imdi_browser?openpath=MPI14765%23).

Das L2-Subkorpus enthält zur Zeit über 4000 *sessions* (d.h. Dateneinheiten, die aus Metadaten, Annotationen und Medien-Dateien bestehen).

Die bereits erwähnten Daten des **ESF**-Projekts sind bereits in das neue Archiv integriert. Das Korpus setzt sich zusammen aus L2-Daten zu folgenden Zielsprachen: Deutsch, Englisch, Französisch, Niederländisch und Schwedisch. Verschiedene Muttersprachen (L1) sind vertreten. Die Daten sind transkribiert und die meisten Mediendateien<sup>14</sup> wurden integriert. Der Zugang zum ESF-Korpus ist frei; die Daten sind zu erreichen unter: [http://corpus1.mpi.nl/ds/imdi\\_browser/?openpath=MPI353844%23](http://corpus1.mpi.nl/ds/imdi_browser/?openpath=MPI353844%23). Die wichtigsten Veröffentlichungen dazu sind: Dittmar/Klein (1979), Perdue (1993), Klein/Perdue (1992). Es handelt sich hierbei um ein Longitudinalkorpus.

Ebenfalls aus den 80-er Jahren stammen die Daten von **P-MoLL**<sup>15</sup>. Sie wurden nach dem Projektablauf in das MPI-Archiv integriert. Das Projekt untersuchte Zweitspracherwerb von polnischen und italienischen Lernern des Deutschen. Es handelt sich hierbei ebenfalls um Longitudinaldaten, die den ungesteuerten Zweitspracherwerb von acht Lernern über einen Zeitraum von zweieinhalb Jahren dokumentieren. Ergänzend wurden Kontrollaufnahmen von acht Muttersprachlern des Deutschen integriert. Die Daten sind mit einem Passwort gesichert. Die Informationen über die Zugangsmöglichkeiten sind integriert. Die Internet-Adresse lautet [http://corpus1.mpi.nl/ds/imdi\\_browser/?openpath=MPI23961%23](http://corpus1.mpi.nl/ds/imdi_browser/?openpath=MPI23961%23).

Die wichtigsten Publikationen zum Projekt sind: Dittmar/Terborg (1991), Dittmar/Reich (1993), Schumacher/Skiba (1992), Skiba/Dittmar (1992), Ahrenholz (1998), Dimroth/Dittmar (1998), Skiba et al. (im Druck), Rost-Roth (2003), Birkner et al. (1995). Alle anderen L2-Korpora im MPI-Archiv sind neueren Ursprungs.

Die **APN**-Daten<sup>16</sup> enthalten L2-Daten für Deutsch als Zweitsprache (L1-Polnisch) sowie Kinderspracherwerbsdaten (Deutsch). Das Hauptanliegen<sup>17</sup> dieser Sammlung ist der Vergleich der Diskursaufbauprinzipien zwischen den L1-Kinderdaten in verschiedenen Altersgruppen (4, 7 und 11 Jahre) und den L2-Daten. Vergleichbare Daten zum Französischen (L1 and L2) wurden bereits erhoben und werden in naher Zukunft integriert. Korpusanalysen findet man in: Benazzo et al. (2004).

Das **BARBSCH**-Korpus<sup>18</sup> besteht hauptsächlich aus tschechischen L2-Daten aus einem Projekt zum Erwerb der Simultaneität. Relativ wenige L2-Deutsch- und

14 Da diese Daten älteren Ursprungs sind, konnten lediglich ca 90% der Mediendateien integriert werden. Die übrigen Daten sind als Transkripte vorhanden.

15 Genauere Informationen zum Projekt in Dittmar et al. (1990).

16 Die Daten sind zugänglich unter: [http://corpus1.mpi.nl/ds/imdi\\_browser?openpath=MPI521039%23](http://corpus1.mpi.nl/ds/imdi_browser?openpath=MPI521039%23).

17 Weitere Informationen zum APN-Korpus findet man unter : [http://corpus1.mpi.nl/ds/imdi\\_browser?openpath=MPI521037%23](http://corpus1.mpi.nl/ds/imdi_browser?openpath=MPI521037%23).

18 [http://corpus1.mpi.nl/ds/imdi\\_browser/?openpath=MPI516826%23](http://corpus1.mpi.nl/ds/imdi_browser/?openpath=MPI516826%23).

L2-Englisch-Daten wurden mitgespeichert. Ergänzt wurde die Datensammlung durch Kontrolldaten von Muttersprachlern des Deutschen, Englischen und Tschechischen. Als Stimuli dienten dabei kurze Werbespots, die so ausgewählt wurden, dass dort mehrere Handlungen als parallel ablaufend dargestellt wurden. Die Informanten erzählen die Inhalte nach. Weitere Informationen dazu: Schmiedtová (2004).

**DaZ-AF**<sup>19</sup> ist ein Korpus, das zum Vergleich des Zweitspracherwerbs bei Kindern und Jugendlichen angelegt wurde. Die Erstsprache der Informanten ist Russisch. Die Zielsprache ist Deutsch. Die Longitudinalaufnahmen von zwei Lernerinnen erstrecken sich über einen Zeitraum von 1,5 Jahren. Inzwischen wurden mehrere Analysen des Korpus durchgeführt (vgl. z.B. Bast 2003, Dimroth 2007 und Pagonis 2007).

Das **Dimroth-Korpus**<sup>20</sup> enthält L2-Daten des Deutschen als Zweitsprache mit Erwachsenen der Ausgangssprachen Kroatisch, Russisch und Türkisch. Kontrollaufnahmen von Muttersprachlern des Deutschen ergänzen die Daten. Als Stimulus diente eine Bilder Geschichte<sup>21</sup>, die auf kontrollierte Art additive und kontrastive Information bereitstellt. Die Aufgabe der Informanten bestand darin, die Geschichte Bild für Bild nachzuerzählen (s. Dimroth 2002).

**Finiteness and Scope**<sup>22</sup> ist ein Korpus zum L2-Erwerb des Niederländischen mit den Erstsprachen Türkisch und Arabisch, und des Deutschen mit verschiedenen L1<sup>23</sup>. Die Daten wurden auf der Grundlage eines Videoclips erhoben, der darauf abzielt, Fokuspartikeln und die Negation in Äußerungen mit verschiedenen Informationsstrukturen zu elizitieren.

Im Aufbau befindet sich noch das **Julka-Korpus**<sup>24</sup>, in dem der bilinguale Spracherwerb eines Mädchens (L1 Deutsch und Polnisch) dokumentiert wurde. Die Dichte der Datenaufnahmen sowie der frühe Startpunkt der Dokumentation sind einmalig: Begonnen wurden die monatlichen Aufnahmen als die Probandin ein Jahr und drei Monate alt war. Die longitudinalen Daten sind im familiären Kontext aufgenommen und nur wenig strukturiert. Es existieren noch keine Analysen dazu.

### **MPI-Tools: Datenauswertungsprogramme**

Im Folgenden sollen kurz die Möglichkeiten der Exploration des **MPI-Korpus** dargestellt werden.<sup>25</sup>

19 Das Akronym steht für Deutsch als Zweitsprache-Altersfaktor. Zu erreichen sind die Daten unter: [http://corpus1.mpi.nl/ds/imdi\\_browser/?openpath=MPI529832%23](http://corpus1.mpi.nl/ds/imdi_browser/?openpath=MPI529832%23).

20 [http://corpus1.mpi.nl/ds/imdi\\_browser?openpath=MPI27087%23](http://corpus1.mpi.nl/ds/imdi_browser?openpath=MPI27087%23).

21 ebd.

22 [http://corpus1.mpi.nl/ds/imdi\\_browser/?openpath=MPI516911%23](http://corpus1.mpi.nl/ds/imdi_browser/?openpath=MPI516911%23).

23 Die L1 sind: Arabisch, Egyptisch, Basaa, Englisch, French, French, Ife (Ana), Französisch, Igbo, Kabardinisch, Kenyang, Krio, Liberianisches Pidgin, Pinyin, Portugisisch und Yoruba.

24 [http://corpus1.mpi.nl/ds/imdi\\_browser/?openpath=MPI519685%23](http://corpus1.mpi.nl/ds/imdi_browser/?openpath=MPI519685%23).

25 Vgl. auch Wittenburg et al. (2004) und (2005).

Die Metadaten dienen, wie schon oben erwähnt, dem Auffinden von Daten, die bestimmte Kriterien erfüllen. Die Suche kann viele Kriterien berücksichtigen, wie Alter, Geschlecht, Sprachen (L1/L2) etc. Das Programm zur Erstellung von Metadaten ist der **IMDI-Editor**<sup>26</sup>. Die Suchroutinen für Metadaten sind über einen Webbrowserschnittstelle in die Archivseiten integriert.

**ELAN**<sup>27</sup> ist ein Annotationswerkzeug zum Transkribieren und Annotieren von digitalisierten Daten (Audio und/oder Video). ELAN erlaubt es, Annotationszeilen (*tiers*) zu definieren: Die wichtigste davon ist die Sprachwiedergabezeile. Sie ist für die weitere Verarbeitung der Daten von großer Bedeutung (deshalb die verbreitete Bezeichnung - *main tier*). Diese Zeile setzt sich aus *tokens* zusammen und ist direkt mit dem Datenstrom verbunden (*time linking*). Die Annotationscodes in ELAN sind frei wählbar (möglich sind alle UTF-kodierten Zeichen, auch IPA.). ELAN verfügt außerdem über mehrere Konverter: So können z.B. CHAT-Dateien in ELAN verarbeitet werden, ohne ihre ursprünglichen Formate zu verlieren. Zu weiteren Möglichkeiten von ELAN vgl. Skiba (2006) und das ELAN-Handbuch (ebenfalls online). Das Programm verfügt auch über Möglichkeiten des Auswertens von Annotationen. Diverse Suchroutinen sind integriert; eine Suche über mehrere Dateien ist allerdings nicht möglich.

**Annex** ist eine webbasierte Suchmaschine und ein Viewer für Annotationen verschiedener Formate (z.B. CHAT, EAF). Annex ist in den Archiv-Daten-Browser<sup>28</sup> integriert. Es verfügt über elaborierte Suchmöglichkeiten. Es ist möglich, die Metadaten- und Annotationensuche zu kombinieren. So kann man z.B. zuerst im Browser nach Merkmalen von Metadaten (wie Sprache, Alter, Diskurstyp etc.) suchen, um danach eine Suche in den Annotationen selbst zu starten.

Bei Annex gestaltet man die Suche so, dass man die Annotationenwerte aus der Transkriptionszeile (*tokens*) mit anderen Annotationszeilen (*tiers*) mit Sprachanalyse, z.B. Annotationenwerte, wie Wortart, Kasus) miteinander kombinieren kann. Die Suchmaschine von Annex kann auch sogenannte *regular expressions* interpretieren und ausführen. Das macht das Programm zu einem leistungsfähigen Analysewerkzeug<sup>29</sup>.

**Lexus**<sup>30</sup> ist ebenfalls webbasiert. Es ist ein administratives Programm zur Erstellung und Verwaltung von Lexika. Lexus verfügt über Importmöglichkeiten von bereits erstellten Lexika, z.B. im CHAT-<sup>31</sup> oder shoebox-Format<sup>32</sup>. Es ist auch möglich, Medien in das Lexikon zu integrieren.

26 <http://www.lat-mpi.eu/tools/imdi/editor/>; dort auch das Handbuch.

27 Frei zu erhalten unter: <http://www.lat-mpi.eu/tools/elan/>; dort auch das Handbuch.

28 Für die L2-Acquisition: [http://corpus1.mpi.nl/ds/imdi\\_browser?openpath=MPI14765%23](http://corpus1.mpi.nl/ds/imdi_browser?openpath=MPI14765%23).

29 Vgl. dazu <http://www.lat-mpi.eu/tools/annex>.

30 Vgl. dazu <http://www.lat-mpi.eu/tools/lexus/>.

31 Vgl. McWhinney (1991).

32 *shoebox* ist ein Annotationsprogramm, das in der ethnolinguistischen Forschung verbreitet ist.

Vgl. dazu [http://www.ethnologue.com/tools\\_docs/shoebox.asp](http://www.ethnologue.com/tools_docs/shoebox.asp).

## Schluss

Im Vergleich zu traditionellen Herangehensweisen der empirischen Linguistik (selbst aufnehmen und selbst transkribieren/kodieren von Sprachdaten) sind die Vorteile der online Datenbanken deutlich: Man kann mehr Zeit für die Analysen der Daten aufbringen.

Diese Möglichkeiten des Zugriffs auf vorbereitete Daten stehen auch Wissenschaftlern offen, die selbst keine Daten aufnehmen (können) bzw. denjenigen, die nach Vergleichsdaten für ihre Studien suchen.

Um Datenaustausch zwischen Zweitspracherwerbsforschern zu ermöglichen, ist diese Konzentration von Daten von Vorteil. Die Zukunft sieht vielleicht noch rosiger aus. Wittenburg et al. (2006: 1) führen aus: „A number of serious reasons will convince an increasing amount of researchers to store their relevant material in centers which we will call ‘language resource archives’. They combine the duty of taking care of long-term preservation as well as the task to give access to their material to different user groups. Access here is meant in the sense that an active interaction with the data will be made possible to support the integration of new data, new versions or commentaries of all sorts.“

In den letzten Jahren kommen auch Lernerkorpora aus dem Bereich des gesteuerten Zweitspracherwerbs hinzu. Sie dienen der angewandten Unterrichtsforschung, können aber auch zur Hypothesenüberprüfung der Zweitspracherwerbsforschung dienlich sein. Eine gute Übersicht dazu bieten z.B. Myles (2005) und Sinclair (2004).

## Literatur

- Ahrenholz, Bernt (1998): *Modalität und Diskurs: Instruktionen auf deutsch und italienisch. Eine Untersuchung zum Zweitspracherwerb und zur Textlinguistik.* Tübingen: Stauffenburg Verlag.
- Bast, Conny (2003): *Der Altersfaktor im Zweitspracherwerb. Die Entwicklung der grammatischen Kategorien Numerus, Genus und Kasus in der Nominalphrase im ungesteuerten Zweitspracherwerb des Deutschen bei russischen Lernerinnen.* [<http://kups.ub.uni-koeln.de/>], Universität Köln.
- Benazzo, Sandra/Dimroth, Christine/Perdue, Clive/Watorek, Marzena (2004): *Le rôle des particules additives dans la construction de la cohésion discursive en langue maternelle et en langue étrangère.* In: *Langages*, 155, 76-105.
- Berck, Peter/Russel, Albert/Kemps-Snijders, Marc/Wittenburg, Peter (2005): *Advanced Web-based Language Archive Exploitation and Enrichment* [[http://www.lat-mpi.eu/papers/papers-2005/ltc\\_070\\_BERCK\\_2.pdf](http://www.lat-mpi.eu/papers/papers-2005/ltc_070_BERCK_2.pdf)].
- Birkner, Karin/Dimroth, Christine/Dittmar, Norbert (1995): *Der adversative Konnektor aber in den Lernervarietäten einer italienischen und zweier polnischer Lerner des Deutschen.* In: Handwerker, B. (ed.) *Fremde Sprache Deutsch: Grammatische Beschreibung – Erwerbsverläufe - Lehrmethodik*, Tübingen: Narr, 65-118.
- Broeder, Daan/Offenga, Freddy./Willems Don./Wittenburg, Peter (2001): *The IMDI Metadata set, its Tools and accessible Linguistic Databases.* In: *Proceedings of the IRCS Workshop on Linguistic Databases*, Philadelphia, 48-55 [[http://www ldc.upenn.edu/annotation/databse/papers/Broeder\\_etal/32.3.broeder.pdf](http://www ldc.upenn.edu/annotation/databse/papers/Broeder_etal/32.3.broeder.pdf)].

- Broeder, Daan/Claus, Andreas/Offenga, Freddy/Skiba, Romuald/Trilsbeek, Paul/Wittenburg, Peter (2006): LAMUS: the Language Archive Management and Upload System. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006) (pp. 2291–2294) [CD-ROM].
- Dimroth, Christine (2002): Topics, assertions and additive words: how L2 learners get from information structure to target language syntax. In: *Linguistics* 40, 891-923.
- Dimroth, Christine/Dittmar, Norbert (1998): Auf der Suche nach Steuerungsfaktoren für den Erwerb von Fokuspartikeln: Längsschnittbeobachtungen am Beispiel polnischer und italienischer Lerner des Deutschen. In: Wegener, H. (ed.): *Eine zweite Sprache lernen. Empirische Untersuchungen zum Zweitspracherwerb*. Tübingen: Narr, 217-241.
- Dimroth, Christine (2007): *Zweitspracherwerb bei Kindern und Jugendlichen. Gemeinsamkeiten und Unterschiede*. In: Anstatt, T. (ed.): *Mehrsprachigkeit bei Kindern und Erwachsenen*. Tübingen: Narr-Francke.
- Dittmar, Norbert/Reich, Astrid (eds.) (1993): *Modality in Language Acquisition*. Berlin: de Gruyter.
- Dittmar, Norbert/Reich, Astrid/Schumacher, Magdalene/Skiba, Romuald/Terborg, Heiner (1990): Die Erlernung modaler Konzepte des Deutschen durch erwachsene polnische Migranten. Eine empirische Längsschnittstudie. In: *Info DaF* 17 (2) , 125-172.
- Dittmar, Norbert/Terborg, Heiner (1991): Modality and second language learning: a challenge for linguistic theory. In: Hübner, T./Ferguson, C.A. (eds.): *Crosscurrents in Second Language Acquisition and Linguistic Theories*. Amsterdam: Benjamins, 347-384.
- Klein, Wolfgang/Dittmar, Norbert (1979). *Developing grammar: The acquisition of German syntax by foreign workers*. Berlin: Springer.
- Klein, Wolfgang/Perdue, Clive (1992): *Utterance Structure. Developing grammars again*. Amsterdam: Benjamins.
- Klein, Wolfgang/Perdue, Clive (1997): The basic variety (or: couldn't natural languages be much simpler?). In: *Second Language Research* 13, 4, 301-347.
- MacWhinney, Brian (1991): *The CHILDES Project: tools for analyzing talk*. Hillsdale, NJ [etc.]: Erlbaum.
- Mika, Egmont (2005): *Formeln und Routinen: Zum Genuserwerb italienischer, portugiesischer und spanischer Gastarbeiter mit Deutsch als Zweitsprache*. In: *Studia Germanistica Upsalienia*, 48. [[http://www.diva-portal.org/diva/getDocument?urn\\_nbn\\_se\\_uu\\_diva-5801-2\\_fulltext.pdf](http://www.diva-portal.org/diva/getDocument?urn_nbn_se_uu_diva-5801-2_fulltext.pdf)].
- Myles, Florence (2005): *Interlanguage corpora and second language acquisition research*. In: *Second Language Research*; 21; 373-391.
- Pagonis, Giulio (2007): *Der Einfluss des Alters auf den Spracherwerb. Eine empirische Fallstudie zum ungesteuerten Zweitspracherwerb des Deutschen durch russische Lerner unterschiedlichen Alters*. Dissertation, Universität Heidelberg.
- Perdue, Clive (1993): *Adult Language Acquisition. Vol 1: Field Methods*. Cambridge: University Press.
- Rost-Roth, Martina (2003): *Fragen - Nachfragen - Echofragen. Formen und Funktionen von Interrogationen im gesprochenen Deutsch*. In: *Linguistik online* 13, 1/03 [[http://www.linguistik-online.de/13\\_01/rostRoth.html](http://www.linguistik-online.de/13_01/rostRoth.html)].
- Rutherford, William/Thomas, Margaret (2001): *The child language data exchange system in research on second language acquisition*. In: *Second Language Research* 17, 195–212.
- Schmiedtová, Barbara (2004): *At the same time: The expression of simultaneity in learner varieties*. Berlin: Mouton de Gruyter.
- Schumacher, Magdalene/Skiba, Romuald (1992): *Prädikative und modale Ausdrucksmittel in den Lernervarietäten einer polnischen Migrantin. Eine Longitudinalstudie. Teil II*. In: *Linguistische Berichte*. 142, 451-475.



- Schüller, Dieter (2004): Safeguarding the Documentary Heritage of Cultural and Linguistic Diversity. In: Language Archives Newsletter, 5. 9-10. [[http://www.mpi.nl/LAN/issues/lan\\_03.pdf](http://www.mpi.nl/LAN/issues/lan_03.pdf)].
- Sinclair, John (2004): How to use corpora in language teaching. Amsterdam, Philadelphia: Benjamins.
- Skiba, Romuald/Dittmar, Norbert (1992): Pragmatic, Semantic and Syntactic Constraints and Grammaticalisation: A Longitudinal Perspective. In: Studies in Second Language Acquisition, 14, 323-349.
- Skiba, Romuald (2006): Computer Analysis. Corpus based language research. In: Amon, U./Dittmar, N./Mattheier, K./Trudgil, P. (eds.): Handbook "Sociolinguistics" (2nd edition). Berlin, New York: de Gruyter.
- Skiba, Romuald/Dittmar, Norbert/Bressem, Jana (im Druck): Planning, collecting, exploring and archiving longitudinal data of naturalistic L2 acquisition: Experiences from the Berlin Project on Modality (P-MoLL). In: Ortega, L./Byrnes, H. (eds.): The Longitudinal Study of Advanced L2 Learning. Mahwah, NJ: Lawrence Erlbaum.
- Wittenburg, Peter/Broeder, Daan (2002): Management of language resources with metadata. In: L. Romary, C. Galinski, N. Ide, & K.-S. Choi (eds.), Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002). Workshop on International Standards of Terminology and Language Resources Management. (pp. 49-53). Paris: European Language Resources Association [<http://www.mpi.nl/IMDI/documents/2002%20LREC/Management%20of%20Language%20Resources%20using%20Metadata.pdf>].
- Wittenburg, Peter/Broeder, Daan/Klein, Wolfgang/Levinson, Steven/Romary, Laurent (2006): Foundations of Modern Language Resource Archives. [<http://www.lat-mpi.eu/papers/papers-2006/general-archive-paper-v4.pdf>].
- Wittenburg, Peter/Skiba, Romuald/Trilsbeek, Paul (2004): Technology and Tools for Language Documentation. In: Language Archives Newsletter, 5. 3-4 [[http://www.mpi.nl/LAN/issues/lan\\_04.pdf](http://www.mpi.nl/LAN/issues/lan_04.pdf)].
- Wittenburg, Peter/Skiba, Romuald/Trilsbeek, Paul (2005): The Language Archive at the MPI: Contents, Tools, and Technologies. In: Language Archives Newsletter, 5, 7-9 [[http://www.mpi.nl/LAN/issues/lan\\_05.pdf](http://www.mpi.nl/LAN/issues/lan_05.pdf)].