

Curriculum Vitae

Dr. Thomas Schmidt
SFB ‚Multilingualism‘
University of Hamburg
Max Brauer-Allee 60
D-22765 Hamburg
☎ +49 (40) 42838-6425
thomas.schmidt@uni-hamburg.de
<http://www.exmaralda.org>
<http://www.kicktionary.de>

Last update: 03 December 2009

I. Education	2
II. Professional Appointments.....	3
III. Publications.....	4
IV. Talks (selection)	7
V. Published Software and Data Resources	9
VI. Teaching.....	9
VII. Other Research Activities and Qualifications.....	10

I. Education

2005-2006	Post-Doc visiting researcher at the International Computer Science Institute in Berkeley, CA (with a grant from the German Academic Exchange Service) Host project: <i>FrameNet</i> (principal investigator: Charles Fillmore)
July 2004	PhD (<i>summa cum laude</i>) in German linguistics (text technology) from the University of Dortmund Thesis title: <i>Computergestützte Transkription – Modellierung und Visualisierung gesprochener Sprache mit texttechnologischen Mitteln</i> (supervisor: Angelika Storrer, published as Schmidt 2005a)
1999-2000	Linguistics at the Free University of Berlin and the Université Paris VIII (with a SOCRATES grant) European Master's Degree in October 2000 Thesis title: <i>The Acquisition of German NP structure in different acquisitional settings</i> (supervisors: Carol Pfaff, Clive Perdue)
1993-1996 1997-1998	General Linguistics, English Linguistics and Roman philology at the Johannes Gutenberg University Mainz M.A. in October 1998 Thesis title: <i>Erkennung und Analyse temporaler Adverbiale in deutschen Fußballberichten über Endliche Automaten</i> (supervisor: Dieter Seelbach)
1996-1997	Linguistics and Artificial Intelligence at the University of Edinburgh, Scotland
1992-1994	Mathematics and Computer Science at the University of Kaiserslautern and the University of Mainz Intermediate diploma in October 1994
1991	Abitur at the Stefan George-Gymnasium in Bingen. Majors: Mathematics, French, English

II. Professional Appointments

2006-present 2000-2005	Principal investigator (since 2008) / project assistant (until 2008, principal investigator: Jochen Rehbein) in the project 'Computer-assisted Creation and Analysis of Multilingual Data' of the Research Centre on Multilingualism at the University of Hamburg
2008-present	Freelance work (software development) for the FOLK project (principal investigator: Arnulf Deppermann) at the Institut für Deutsche Sprache in Mannheim
2007-2008	Research assistant in the project 'Digital Dictionary of the German Language' (principal investigators: Wolfgang Klein, Angelika Storrer, Manfred Bierwisch) at the Berlin Brandenburg Academy of Science
2006-2008	Associated member of the joint project 'Sustainability of linguistic data' (principal investigators: Erhard Hinrichs & Marga Reis) of the SFB 441 'Linguistic Data Structures' (Tübingen), the SFB 538 'Multilingualism' (Hamburg) and the SFB 632 'Information structure' (Potsdam/Berlin)
2005-2006	Member of the Berkeley FrameNet project (principal investigator: Charles Fillmore) at the International Computer Science Institute
2000	Project assistant at the European Language Resource Agency (ELRA), Paris
1999-2000	Software tutor at LM IT Services, Berlin
1998-1999	Language Resource Engineer at Philips Speech Processing, Aachen
1994-2002	Freelance work for CASIO and Texas Instruments as an author and translator
1997-1998	Student assistant at the Institute for General Linguistics and the Institute for Computer Science at the University of Mainz
1995	Translator (English and French) for Schott Musik International, Mainz

III. Publications

[C] – Conference contribution (anonymous reviewing), [I] – Invited contribution (book, journal or conference proceedings), [J] – Journal Contribution (anonymous reviewing), [W] – Working paper (non-anonymous reviewing), [B] – Book

2009 / in preparation

- [I] (2009a): Another extension of the stylesheet metaphor – Visualising multi-layer annotations as musical scores. To appear in: A. Witt & D. Metzger (eds.): Linguistic modelling of information and Markup Languages. Dordrecht: Springer.
- [I] (2009b) (with S. Duncan, O. Ehmer, J. Hoyt, M. Kipp, D. Loehr, M. Magnusson, T. Rose & H. Sloetjes): An exchange format for multimodal annotations. In: M. Kipp, J.-C. Martin, P. Paggio, and D. Heylen (eds.): Multimodal corpora. Dordrecht: Springer, 207-221.
- [J] (2009c) (with Silke Merkel): Korpora gesprochener Sprache im Netz - eine Umschau. In: Gesprächsforschung (10), 70-93.
- [I] (2009d): The Kicktionary – A Multilingual Lexical Resource of Football Language. In: Boas, Hans C. (ed.): Multilingual Framenets in Computational Lexicography. New York: de Gruyter, 101-134.
- [I] (2009e): Kicktionary. In: Burkhard, Armin / Schlobinski, Peter (eds.): Flickflack, Foul und Tsukahara. Der Sport und seine Sprache (Thema Deutsch 10). Mannheim: Duden-Verlag.
- [I] (2009f) Creating and working with spoken language corpora in EXMARaLDA. To appear in: Lyding, Verena (ed.): Proceedings of the Second Colloquium on Lesser Used Languages and Computer Linguistics (LULCL II). Bozen: EURAC, 151-164.
- [J] (2009g) (with Kai Wörner): EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. To appear in: Pragmatics.
- [I] (20010a) EXMARaLDA. To appear in: Oxford Handbook of Corpus Phonology, Oxford: University Press.
- [I] (2010b) Grundzüge von EXMARaLDA – einem System zur computergestützten Erstellung und Auswertung von Korpora gesprochener Sprache. To appear in: Rehbein, J. & Kameyama, S. (ed.): Bausteine diskursanalytischen Wissens, Berlin: de Gruyter.
- [J] (2010c) Der Fußballwortschatz im Kicktionary, To appear in: Der Deutschunterricht.

2008

- [C] (2008a) (with Alexander Geyken & Angelika Storrer): Refining and Exploiting the Structural Markup of the eWDG. In: Proceedings of EURALEX 2008.
- [J] (2008b) (with Jasmine Bennoehr): Rescuing Legacy Data. In: Language Documentation and Conservation 2(1), 109-129.
- [C] (2008c) The Kicktionary Revisited. In: Storrer, Angelika / Geyken, Alexander / Siebert, Alexander / Würzner, Kay-Michael (eds.): Text Resources and Lexical Knowledge. Berlin: Mouton de Gruyter, 239-252.
- [I] (2008d) The Kicktionary: Combining corpus linguistics and lexical semantics for a multilingual football dictionary. In: Eva Lavric / Gerhard Pisek / Andrew Skinner / Wolfgang Stadler (eds.): The Linguistics of Football (Language in Performance 38). Tübingen: Gunter Narr, 11-23.
- [C] (2008e) (with S. Duncan, O. Ehmer, J. Hoyt, M. Kipp, D. Loehr, M. Magnusson, T. Rose & H. Sloetjes): An exchange format for multimodal annotations. In: Proceedings of the Language Resource and Evaluation Conference 2008, Marrakech. Paris: ELDA.

2007

- [I] (2007a) (with Nicole Baumgarten, Annette Herkenrath, Kai Wörner & Ludger Zeevaert): Studying Connectivity with the Help of Computer-Readable Corpora: Some Exemplary Analyses from Modern and Historical, Written and Spoken Corpora. In: Rehbein et al. (eds.): Connectivity in Grammar and Discourse. Hamburg Studies in Multilingualism (5). Amsterdam: Benjamins.
- [J] (2007b): Transkriptionskonventionen für die computergestützte gesprächsanalytische Transkription. In: Gesprächsforschung (8), 229-241.

- [C] (2007c) The Kicktionary: A Multilingual Resource of the Language of Football. In: Rehm, Georg / Witt, Andreas / Lemnitzer, Lothar (eds.): Data Structures for Linguistic Resources and Applications. Tübingen: Gunter Narr, 189-196.

2006

- [J] (2006a) (with K. Rohlfing, D. Loehr, S. Duncan, A. Brown, A. Franklin, I. Kimbara, J. Milde, F. Parrill, T. Rose, H. Sloetjes, A. Thies & S. Wellinghoff): Comparison of multimodal annotation tools. In: Gesprächsforschung (7), 99-123.
- [C] (2006b) Interfacing Lexical and Ontological Information in a Multilingual Soccer FrameNet. In: Proceedings of OntoLex 2006 - Interfacing Ontologies and Lexical Resources for Semantic Web Technologies, Genoa, Italy, May, 24-26, 2006.
- [C] (2006c) (with C. Chiarcos, T. Lehmsberg, G. Rehm, A. Witt & E. Hinrichs): Avoiding Data Graveyards: From Heterogeneous Data Collected in Multiple Research Projects to Sustainable Linguistic Resources. Proceedings of the E-MELD workshop 2006, Michigan State University in East Lansing, Michigan.
- [C] (2006d) (with Stefanie Dipper, Erhard Hinrichs, Andreas Wagner & Andreas Witt.): Sustainability of Linguistic Resources. In: Erhard Hinrichs, Nancy Ide, Martha Palmer, and James Pustejovsky (eds.): Proceedings of the LREC 2006 Satellite Workshop on "Merging and Layering Linguistic Information", Genoa 2006.

2005

- [B] (2005a): Computergestützte Transkription - Modellierung und Visualisierung gesprochener Sprache mit texttechnologischen Mitteln. Frankfurt a. M.: Peter Lang.
- [I] (2005b): EXMARaLDA und die Datenbank "Mehrsprachigkeit" - Konzepte und praktische Erfahrungen. In: Dipper, Stefanie / Stede, Manfred (eds.): Heterogeneity in Focus: Creating and Using Linguistic Databases. ISIS (Interdisciplinary Studies on Information Structure), Working Papers of the SFB 632, Universität Potsdam.
- [C] (2005c): Modellbildung und Modellierungsparadigmen in der computergestützten Korpusanalyse. In: Fisseni, Bernhard / Schmitz, Hans-Christian / Schröder, Bernhard / Wagner, Petra (eds.): Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen. Beiträge zur GLDV-Tagung 2005 in Bonn. (Sprache, Sprechen und Computer/Computer Studies in Language and Speech 8). Frankfurt a. M.
- [J] (2005d): Datenarchive für die Gesprächsforschung. Perspektiven, Probleme und Lösungsansätze. In: Gesprächsforschung 6, 103-126.
- [W] (2005e): Time-based data models and the Text Encoding Initiative's guidelines for transcription of speech. In: Arbeiten zur Mehrsprachigkeit (Working Papers in Multilingualism), Serie B (62). Hamburg.
- [J] (2005f): (with Kai Wörner) Erstellen und Analysieren von Gesprächskorpora mit EXMARaLDA. In: Gesprächsforschung 6, 171-195.

2004

- [I] (2004a): EXMARaLDA - ein System zur computergestützten Diskurstranskription. In: Mehler, Alexander / Lobin, Henning (Hrsg.): Automatische Textanalyse. Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte. Wiesbaden: Verlag für Sozialwissenschaften, 203-218.
- [C] (2004b) (with B. MacWhinney, C. Martell, J. Wagner, P. Wittenburg & E. Hoffer): Collaborative Commentary: Opening Up Spoken Language Databases. In: Proceedings of the Language Resource and Evaluation Conference 2004, Lisbon, Paris: ELRA.
- [W] (2004c) (with J. Rehbein, B. Meyer, F. Watzke & A. Herkenrath): Handbuch für das computergestützte Transkribieren nach HIAT. Arbeiten zur Mehrsprachigkeit, Folge B (Nr. 56). Hamburg.
- [C] (2004d): Transcribing and annotating spoken language with EXMARaLDA. In: Proceedings of the LREC-Workshop on XML based richly annotated corpora, Lisbon 2004, Paris: ELRA.
- [C] (2004e): EXMARaLDA - ein Modellierungs- und Visualisierungsverfahren für die computergestützte Transkription gesprochener Sprache. In: Buchberger, E. (ed.): Proceedings of Konvens 2004, Schriftenreihe der Österreichischen Gesellschaft für Artificial Intelligence 5 Wien.

2003

- [W] (2003a): Visualising Linguistic Annotation as Interlinear Text. Arbeiten zur Mehrsprachigkeit, Serie B (46). Hamburg.
- [C] (2003b): Korpus "Skandinavische Semikommunikation" - ein mehrsprachiges Diskurskorpus auf XML-Basis. In: Seewald-Heeg (eds.): Sprachtechnologie für die multilinguale Kommunikation - Textproduktion, Recherche, Übersetzung, Lokalisierung (Beiträge der GLDV-Frühjahrstagung 2003). Sankt Augustin: gardez!., 421-427.

2002

- [J] (2002a): Gesprächstranskription auf dem Computer: das System EXMARaLDA. In: Gesprächsforschung 3, 1-23.
- [J] (2002b): Stellungnahme zu Wolfgang Schneiders Artikel "Annotate in Transkriptionen aus DV-technischer Sicht". In: Gesprächsforschung 3, 237-249.
- [W] (2002c): EXMARaLDA - ein System zur Diskurstranskription auf dem Computer. In: Arbeiten zur Mehrsprachigkeit, Folge B(34).

2001

- [C] (2001): The transcription system EXMARaLDA: An application of the annotation graph formalism as the Basis of a Database of Multilingual Spoken Discourse. In: Steven Bird et al. (eds.): Proceedings of the IRCS Workshop On Linguistic Databases, 11-13 December 2001. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania, 219-227.

IV. Talks (selection)

Invited talks are marked with an *asterisk.

2009

- Montpellier, Colloque international «Pour une épistémologie de la sociolinguistique»:
EXMARaLDA : un système pour la constitution et l'exploitation de corpus oraux
- *Helsinki, Networking Event for European Research Infrastructures:
Annotation Formats: AG/EAF/Exmaralda etc. (with Peter Wittenburg)
- University of Hamburg, Ringvorlesung 'Multilingualism in Theory and Practice':
Das Kicktionary. Ein mehrsprachiges computergestütztes Fußballwörterbuch.
- *Université de Nantes:
Formation EXMARaLDA.

2008

- *Bozen, Lesser Used Languages and Computer Linguistics conference II:
Creating and working with spoken language corpora in EXMARaLDA
- London, 8th members meeting of the Text Encoding Initiative:
Transcription tools, transcription conventions and the TEI guidelines for transcriptions of speech.
(with Andreas Witt)
- Berlin, KONVENS conference:
The Kicktionary revisited.
- *Technical University of Darmstadt, Workshop on Linguistic Processing Pipelines:
Processing Pipelines für Korpora gesprochener Sprache
- Forlí, 41st annual meeting of the Societas Linguistica Europaea:
EXMARaLDA – A System for Creating, Analysing and Sharing spoken language corpora
- Barcelona, 13th International EURALEX congress:
Refining and Exploiting the Structural Markup of the eWDG. (with Alexander Geyken and Angelika Storrer)
- Humboldt University Berlin, Corpus Linguistics Colloquium:
Das WDG: Vom Wörterbuch zur digitalen lexikalischen Ressource (with Axel Herold)
- Marrakech, Language Resource and Evaluation Conference:
An exchange format for multimodal annotations.
- *University of Hannover, Linguistischer Arbeitskreis:
Kicktionary
- *Mannheim, Arbeitstagung Gesprächsforschung:
EXMARaLDA
- Berlin-Brandenburgische Akademie der Wissenschaften:
Feinstrukturierung des elektronischen WDG

2007

- Göteborg, International Pragmatics Conference:
EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research.
- Hamburg, International Symposium on Bilingualism:
Creating and analysing bilingual spoken language corpora with EXMARaLDA (with Kai Wörner).
- Tübingen, GLDV conference:
KWIC-Konkordanzen für verschiedene Korpora und Fragestellungen – eine Fallstudie zur Schnittstelle zwischen linguistischen Ressourcen und deren Anwendung.
The Kicktionary – a multilingual lexical resource of the language of football.
- *Innsbruck, Workshop „Language & Football“:
The Kicktionary – A Multilingual Lexical Resource of Football Language.

2006

- Victoria, BC, 6th members meeting of the Text Encoding Initiative:

- From Project Data to Sustainable Archiving of Linguistic Corpora (with Andreas Witt).
- *IMS Stuttgart, Linguistics Colloquium:
Das Kicktionary. Ein mehrsprachiges elektronisches Wörterbuch der Fußballsprache.
- Lansing, MI, EMELD conference:
Avoiding Data Graveyards: From Heterogeneous Data Collected in Multiple Research Projects to Sustainable Linguistic Resources.
- *University of California, Santa Barbara, Linguistics Colloquium:
EXMARaLDA.
- *Albuquerque, 80th Annual Meeting of the Linguistic Society of America:
Transcription and Multimodal tools: Creating & Analyzing Spoken Language Corpora With EXMARaLDA.

2005

- Bonn, GLDV conference:
Modellbildung und Modellierungsparadigmen in der computergestützten Korpusanalyse.
- *Bielefeld, Linguistics Colloquium:
Computergestützte Erstellung und Auswertung von Korpora gesprochener Sprache mit EXMARaLDA.
- Lyon, Interacting Bodies Conference:
Multimodal Annotation Tools: EXMARaLDA (with Sandra Wellinghoff).

2004

- Lisbon, Language Resource and Evaluation Conference:
Collaborative Commentary: Opening Up Spoken Language Databases.
Transcribing and annotating spoken language with EXMARaLDA.
- Vienna, KONVENS conference:
EXMARaLDA - ein Modellierungs- und Visualisierungsverfahren für die computergestützte Transkription gesprochener Sprache.
- Potsdam, Workshop on Creating and Using Linguistic Databases:
EXMARaLDA und die Datenbank "Mehrsprachigkeit" - Konzepte und praktische Erfahrungen.
- Bielefeld, ZiF conference „Processing Text Technological Resources“:
Comparing approaches to XML-based discourse modelling – Transcription Graphs and EXMARaLDA

2003

- Munich, DGfS conference:
Computergestützte Transkription als Modellierung und Visualisierung gesprochener Sprache
- Köthen, GLDV conference:
Korpus "Skandinavische Semikommunikation" - ein mehrsprachiges Diskurskorpus auf XML-Basis.
- Dortmund, Workshop „Pragmatik und Sprachlehrforschung“:
Segmentieren von EXMARaLDA-Daten.

2002

- Paris, Journée d'Etude d'ATALA „Constitution et exploitation de corpus du français parlé“:
EXMARaLDA.
- Trier, GLDV workshop:
EXMARaLDA - ein System zur Diskurstranskription auf dem Computer.

2001

- Philadelphia, IRCS workshop on linguistic databases:
The transcription system EXMARaLDA: An application of the annotation graph formalism as the Basis of a Database of Multilingual Spoken Discourse.
- Freiburg, Arbeitstagung Gesprächsforschung:
Entwicklung eines Systems zur Diskurstranskription auf dem Computer

V. Published Software and Data Resources

- EXMARaLDA Partitur-Editor – www.exmaralda.org/en_partitureditor.html
- EXMARaLDA Concordance and Analysis Tool – www.exmaralda.org/en_exakt.html
- FOLKER (FOLK Editor) - agd.ids-mannheim.de/html/folker.shtml
- Kicktionary (dictionary, corpus and annotation software) – www.kicktionary.de
- Corpora of the SFB 'Multilingualism' – corpora.exmaralda.org

VI. Teaching

- Introduction to Computer-Assisted Transcription, Undergraduate seminar 2004/2005, University of Dortmund
- Introduction to Text Technology, Undergraduate seminar 2004/2005, University of Hamburg
- Introduction to Corpus Linguistics, Undergraduate seminar 2008/2009, University of Hamburg (with Timm Lehmborg)
- "Das Kicktionary. Ein mehrsprachiges computergestütztes Fußballwörterbuch." Lecture as a part of the „Ringvorlesung: Mehrsprachigkeit in Theorie und Praxis.“ 2009, University of Hamburg.
- Introductory and Training courses 'Spoken language transcription with EXMARaLDA':
 - Institute for German Language, Free University Berlin, 2001
 - Institute for German Language, University of Köln, 2002
 - Graduate school "Bildungsgangforschung", University of Hamburg, 2003
 - As a part of the seminar "Introduction to Linguistics" (Bernd Meyer), Institute for German Language, University of Hamburg, 2008
 - As a part of the seminar "Quantitative Sociolinguistics" (Aria Adli), Department of Romance Languages, University of Hamburg, 2008
 - Project "Plurilinguisme, Apprentissage, Acquisition", Université de Nantes, 2009
 - Department of Foreign Language and Education, Middle Eastern Technical University Ankara, 2009.
 - Herder-Institute, University of Leipzig, 2009 (with Wilfried Schütte).
 - Herbstschule "Sprachdiagnostik und Sprachförderung", University of Hamburg, 2009 (with Kai Wörner and Timm Lehmborg).
 - Workshop "New research methods and software for the humanities", University of Pula (Croatia), 2010 (with Peter M. Fischer).

VII. Other Research Activities and Qualifications

- Member of the DFG network 'Internetlexikographie' (proposal under way)
- March 2009 / November 2009: Visiting Researcher at Middle Eastern Technical University, Ankara with a TUBITAK grant from the Turkish government
- Since 2008: Member of CLARIN (Common Language Resource Infrastructure) and D-SPIN (Deutsche Sprachressourcen-Infrastruktur)
- Since 2008: Co-editor of 'Hamburg Working Papers in Multilingualism'
- May 2008: Member of the Organising Committee of the LREC Workshop "Sustainability of Linguistic Resources", Marrakech
- 2001 – 2003: Member of the NITE (Natural Interactivity Tools Engineering) Advisory Board
- Program committee member:
 - GSCL conference "Schritte zur Textbedeutung - Aspekte der Automatischen Verarbeitung", Potsdam 2009
 - International Workshop on Terminology and Lexical Semantics, Montréal 2009
 - KONVENS conference "Text Resources and Lexical Knowledge", Berlin 2008
 - International Symposium on Bilingualism, Hamburg 2007
 - ACL workshop on "Interoperability of linguistic resources", Sydney 2006
- Journal and Book reviewing:
 - Online-Zeitschrift zur Gesprächsforschung (GO)
 - Corpus Linguistics and Linguistic Theory (Mouton de Gruyter)
 - Anke Lüdeling & Merja Kytö (eds.) (2009): Corpus Linguistics. An International Handbook. (Series: Handbooks of Linguistics and Communication Science). Mouton de Gruyter, Berlin.
 - Kipp et al. (eds.) (2009): Multimodal corpora. Dordrecht: Springer.
 - Rehbein et al. (eds.) (2007): Connectivity in Grammar and Discourse. Hamburg Studies in Multilingualism (5). Amsterdam: Benjamins.
- Grants/Scholarships/Awards:
 - 2008-2011: German Research Council (DFG) grant (€ 319,800) for the project 'Computer-assisted Creation and Analysis of Multilingual Data' of the Research Centre on Multilingualism at the University of Hamburg
 - 2007: Nominee of the German Academic Exchange Service (DAAD) for the Junge Akademie, Berlin
 - 2005/2006: German Academic Exchange Service (DAAD) scholarship (€ 32,307) for a twelve-month stay at the International Computer Science Institute in Berkeley, CA
 - 1999/2000: SOCRATES grant (DM 5,000) for a five-month stay at the Université Paris VIII
- Languages:
 - German: native
 - English: excellent
 - French: very good
- Programming Skills:
 - Expert knowledge of Java (GUI development, multimedia), XML/XSLT and related technologies and AWK scripting
 - Good knowledge of RDB programming and web development