

MARK LIBERMAN

Education:

1965–1969	Harvard University	Linguistics and Applied Mathematics	
1972	M.I.T.	Linguistics	M.S.
1972–1975	M.I.T.	Linguistics	Ph.D.

Professional Experience:

Director, Linguistic Data Consortium: 1992–present

University of Pennsylvania:

(Primary Appointment):

Trustee Professor of Phonetics, Department of Linguistics: 1990–present

(Secondary Appointment):

Professor, Department of Computer and Information Science: 1992–present

Director, Institute for Research in Cognitive Science: 2001–present

Head, Linguistics Research Department, AT&T Bell Laboratories: 1987–1990

Visiting Assistant Professor, M.I.T.: 1978

Member of Technical Staff, AT&T Bell Laboratories: 1975–1987

Other Experience:

U.S. Army, 1969–1972

Research Interests:

Linguistics: Phonetics and phonology; tone and intonation; information structure and focus; modeling of pronunciation variation.

Speech and Language Technology: Linguistic databases; information retrieval and extraction; speech synthesis; speech recognition.

General: Machine learning and inductive methods; models of biological signal processing; neurobiology of language; language evolution (both biological and cultural); second language teaching and learning.

Current Projects: Standards and tools for linguistic annotation; fMRI imaging of spoken word processing; phonology, phonetics and morphology of tone, especially in West African languages; expression of focus and information structure; machine learning of morphology; documentation of endangered languages.

Selected Professional Activities:

Editorial Advisory Boards: Computer Speech and Language; Speech Communication; International Journal of Corpus Linguistics.

Committee on Undergraduate Education, Linguistic Society of America.

Chair, International Coordinating Committee on Speech Databases and Assessment (COCOSDA).

DARPA TIDES Advisory Committee.

U.S. Dictionaries Advisory Board, Oxford University Press.

Director, Linguistic Data Consortium.

The Linguistic Data Consortium is an open consortium of universities, companies and government research laboratories. It creates, collects and distributes speech and text databases, lexicons, and other resources for research and development purposes. The University of Pennsylvania is the LDC's host institution. The LDC was founded in 1992 with a grant from the Advanced Research Projects Agency (ARPA), and has been partly supported by grant IRI-9528587 from the Information, Robotics and Intelligent Systems division of the National Science Foundation.

The LDC has published more than 150 digital databases, which have been delivered to more than 750 universities, companies and government research laboratories. Many LDC databases are also available for interactive search and access on the internet (<http://www ldc.upenn.edu>).

Erdős number: 3

Teaching at Penn:

LING001: *Introduction to Linguistics* <http://www.ling.upenn.edu/courses/ling001>

LING105/CSE200/PHIL044/PSYCH107: *Introduction to Cognitive Science*

HUM100: *Human Nature* (with Wendy Steiner) <http://www.ling.upenn.edu/courses/hum100>

LING502/202: *Linguistic Field Methods* <http://www.ling.upenn.edu/courses/ling202>

LING525/CIS558: *Computer Analysis and Modeling of Biological Signals and Systems*
<http://www.ling.upenn.edu/courses/ling525>

LING520: *Introduction to Phonetics I* <http://www.ling.upenn.edu/courses/ling520>

LING521: *Introduction to Phonetics II*

LING620: *Topics in Prosody* <http://www.ling.upenn.edu/courses/ling620>

LING625: *Seminar in Morphosyntax*

Selected University Activities:

Current:

Faculty Advisory Board, Penn Humanities Forum
Executive Committee, African Studies Center
Search Committee, IRCS Director
Search Committee, Psycholinguistics (Psychology Department)
Co-organizer, IRCS Brain and Language Group (<http://www ldc.upenn.edu/brainlang>)

Past:

Penn Language Center Academic Advisory Committee
Penn Language Center Faculty Planning Committee (chair)
SAS Planning and Priorities Committee
SAS Personnel Committee
SAS Computing Committee
Distributed Computing Task Force Advisory Committee
ISC Academic Advisory Committee (chair)
University Computer Services Restructuring Task Force
University Computer Services Restructuring Steering Group
Internal Review Committee, Mathematics Department
Internal Review Committee, Philosophy Department

Research Grants (within past five years)

Linguistic Data Consortium (ARPA): P.I., 6/94-12/96, \$5,334,888.

Resources for Multilingual Large Vocabulary Speech Recognition (DoD): P.I., 9/93-12/97, \$4,366,506.

Multilingual Text and Speech Resources for Language Teaching (DoD): P.I., 4/94-4/96, \$314,000;
9/97-9/99 \$795,238.

Electronic Dictionaries (DoD): (co-P.I.) 9/1/94-8/31/97, \$592,183; 2/1/94-7/31/96; \$200,000.

Improved Speech and Text Data Resources (NSF): P.I., 9/95-9/99, \$2,567,000.

National Center for Sign Language and Gesture Resources (NSF): co-P.I., 10/98-10/2002, \$532,886.

Data Provenance (NSF Digital Libraries): co-P.I. 6/99-6/00 \$174,951.

Talkbank (NSF KDI and SBE): co-P.I., 9/99–9/2002, \$2.8M.

Multilingual Information Access and Management (NSF): P.I., 3/2000–3/2002, \$300K.

Resources for Multilingual Information Technology (DARPA): P.I., 1/2000–1/2003, \$1.3M.

Empirical Multilingual Processing (DARPA): co-P.I., 1/2000–1/2005, \$4.5M.

Multidimensional Exploration of Linguistic Databases (NSF): co-P.I., 2/2000 to 2/2002, \$152,445.

Academic Equipment Grant, Sun Microsystems: 1999, \$90K.

LDC Fund (Income from consortium memberships and data sales): 1993–present, \$5,941,718.

Publications

Books:

- (1979) *The Intonational System of English*, Garland Publishing.
- (1995) *Invitation to Cognitive Science*, ed. with Lila Gleitman, MIT Press.

Papers:

- (1974a) "On Conditioning the Rule of Subject-Auxiliary Inversion", NELS 5, pp. 77-91.
- (1974b) "Prosodic Form and Discourse Function", with I. Sag, CLS 10, pp. 416-427.
- (1975) "Intonational Disambiguation of Indirect Speech Acts", with I. Sag, CLS 11, pp. 487-497.
- (1977a) "The Geniohyoid and the Role of the Strap Muscles in Pitch Control", with D. Erickson and S. Niimi, Haskins Laboratories Quarterly Status Report 49, pp. 103-110.
- (1977b) "Studies of Metrical patterns", with J.P. Olive and P. Zukovsky, JASA 62 (S1).
- (1977c) "Further Work on Duration Modeling in Reiterant Speech", JASA 62 (S1).
- (1977d) "On Stress and Linguistic Rhythm", with A. Prince, Linguistic Inquiry 8 pp. 249-336.
- (1978a) "Use of Nonsense-Syllable Mimcry in the Study of Prosodic Phenomena", with L.A. Streeter, JASA 63 pp. 231-233.
- (1978b) "Modeling of Durational Patterns in Reiterant Speech", pp. 127-138 in D. Sankoff, Ed., *Linguistic Variation: Models and Methods*. Academic Press.
- (1978c) "Phonetic Transcription, Stress and Segment Durations from Spelled Proper Names", JASA 64.
- (1978d) "On the Nature of Normalization Functions for Durations in Speech", JASA 64.
- (1979a) "Text-to-Speech Conversion by Rule and a Practical Application," with P.B. Denes and J.P. Olive, *Proceedings of the 8th International Congress of Phonetic Sciences*, Copenhagen. Vol. I, p. 350.
- (1979b) "A Set of Concatenative Units for Speech Synthesis", with J.P. Olive, JASA 65.
- (1979c) "A Metric for the Height of Certain Pitch Peaks in English", with J.B. Pierrehumbert, JASA 66.
- (1979d) "The Intrinsic Pitch of Vowels in Sentence Context", with C.H. Shadle and J.B. Pierrehumbert, JASA 66.

- (1980) “Intelligibility of Consonants Produced by Dyadic Rule Synthesis”, with J.P. Olive, and K. O’Connor-Dukes, *JASA* 68.
- (1981a) “Speech Recognition by Computer”, with S.E. Levinson, *Scientific American*, April 1981, V. 244, p. 64-76.
- (1981b) “Effects of Linguistic Boundary and Stress Placement on Speech Dynamics: A Preliminary Study Using a New Method to Compare Articulatory Movement Across Utterances”, with O. Fujimura, *JASA* 69.
- (1982a) review of J.C. Simon, Ed. *Spoken Language Generation and Understanding*, in *JASA* 72 p. 1657.
- (1982b) “Test of an Automatic Syllable Peak Detector”, with M. Fleck, *JASA* 72.
- (1982c) “Modeling the Fundamental Frequency of the Voice”, with J.B. Pierrehumbert, *Contemporary Psychology* 27 pp. 690-692.
- (1983a) “In Favor of Some Uncommon Approaches to the Study of Speech”, pp. 265-274 in P. MacNeilage, Ed. *Speech Motor Control*, Springer Verlag.
- (1983b) “The Symmetric Time-warping Problem: from Continuous to Discrete”, with J.B. Kruskal, pp. 125-163 in D. Sankoff and J.B. Kruskal, Eds., *Time Warps, String Edits and Macromolecules*, Addison-Wesley. Republished 1999 in the David Hume Series, CSLI Publications.
- (1983c) “Neurobiology of Language Processes: a Linguistic Point of View”, pp. 7-16 in M. Studdert-Kennedy, Ed., *The Psychobiology of Language*, MIT Press.
- (1984a) “Words and Sounds”, pp. 157-173 in Kerr et al. Eds., *Science, Computers and the Information Onslaught*, Academic Press.
- (1984b) “Intonational Invariance under Changes in Pitch Range and Length”, with J.B. Pierrehumbert, pp. 157-234 in M. Aronoff and R. Oehrle, Eds., *Language Sound Structure*, MIT Press.
- (1985) “Text to Speech Work at Bell Laboratories”, with J.P. Olive, *JASA* 79.
- (1986a) “Connectionist Models of Natural Language”, Proceedings of the 24th Annual Meeting for the Association for Computational Linguistics.
- (1986b) “Synthesis of Falling Nuclear Pitch Accents”, with S. Steele, *JASA* 80.
- (1986b) “Stressing English Noun Compounds Correctly”, with Richard Sproat, *JASA* 80.
- (1987a) “The Shape and Alignment of Rising Intonation”, with S. Steele, *JASA* 81.
- (1987b) “Towards Treating English Nominals Correctly”, with R. Sproat, pp. 140-146 in Proceedings for the 25th Annual Meeting of the ACL.

- (1989a) “Speaker Independent Phonetic Transcription of Fluent Speech for Large Vocabulary Speech Recognition,” with S.E. Levinson and A. Ljolje, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing.
- (1989b) “The ACL Data Collection Initiative”, pp. 173-188 in Proceedings of DARPA Workshop on Speech and Natural Language Processing, Morgan Kaufman.
- (1990a) “Morphology and Rhyming: Two Powerful Alternatives to Letter-to-Sound Rules for Speech Synthesis”, with C.H. Coker and K.W. Church, pp. 83-87 in *Proceedings of the ESCA Workshop on Speech Synthesis*.
- (1990b) “A Finite-state Morphological Processor for Spanish,” with E. Tsoukermann, pp. 277-283 in Hans Karlgren, Ed., *COLING90*.
- (1991a) “Cryptographic Protection of Databases and Software,” with J. Feigenbaum and R. Wright, in Feigenbaum and Merrit, editors, *Distributed Computing and Cryptography*, AMS and ACM, v. 2, 161–172.
- (1991b) “The Trend Towards Statistical Models in Natural Language Processing”, in *Natural Language and Speech*, edited by E. Klein and F. Veltman, Springer Verlag, pp. 1-8.
- (1992a) “The Phonetics of Igbo Tone”, with J.M. Schultz, S. Hong and V. Okeke, ICSLP 92.
- (1992b) “Text Analysis and Word Pronunciation in Text-to-Speech Synthesis,” with K. Church, pp. 791–832 in Furui and Sondhi, Eds., *Advances in Speech Technology*, Marcel Dekker.
- (1992c) “The Stress and Structure of Modified Noun Phrases in English,” with R. Sproat, pp. 131–181 in *Lexical Matters*, Sag and Szabolsci, Eds. University of Chicago Press.
- (1992d) “The Structure and Intonation of Business Telephone Greetings”, with C. McLemore, pp. 68-83 in *Penn Review of Linguistics*, 1992.
- (1993) “The Phonetic Interpretation of Tone in Igbo” with J.M. Schultz et al., *Phonetica* 50(3) 147–160.
- (1994a) “Computer Speech Synthesis,” pp. 107–116 in D. B. Roe and J. G. Wilpon, Eds., *Voice Communication between Humans and Machines*, National Academy Press.
- (1994b) “Phonological Optionality in Latin Clitics,” *Penn Review of Linguistics*, 1994.
- (1994c) “UNIPEN: Data Exchange and Recognizer Benchmarks,” with I. Guyon et al., in *Proceedings, International Conference on Pattern Recognition*.
- (1995a) “On the Phonetic Interpretation of the Yoruba Tonal System” with Akin Akinlabi, in *Proceedings, International Congress of Phonetic Sciences*.
- (1995b) “The Sound Structure of Mawu Words,” pp. 55-86 in *Invitation to Cognitive Science*, L. Gleitman and M. Liberman, Eds., MIT Press.

- (1995c) “The Cognitive Science of Language,” with Lila Gleitman, in L. Gleitman and M. Liberman, Eds. *Invitation to Cognitive Science*, MIT Press, pp. xix–xxxvii.
- (1997) “Le Consortium de Données Linguistiques”, Journées Scientifiques et Techniques du Réseau Francophone d’Ingénierie de la Langue de l’AUPELF-UREF.
- (1998a) “The Creation, Distribution and Use of Linguistic Data,” with C. Cieri, *Proceedings, First International Conference on Language Resources and Evaluation*, Granada, 1998.
- (1998b) “Transcriber: a Free Tool for Segmenting, Labeling and Transcribing Speech,” with C. Barras, E. Geoffrois and Z. Wu, *Proceedings, First International Conference on Language Resources and Evaluation*, Granada, Spain, May 1998.
- (1998c) “Towards a Formal Framework for Linguistic Annotation”, with S. Bird, *Proceedings, International Conference on Spoken Language Processing*, Sydney, 1998.
- (1998d) “Annotation Graphs as a Framework for Multidimensional Linguistic Data Analysis”, with S. Bird, *Proceedings, Workshop on Standards and Tools for Discourse Tagging*. Association for Computational Linguistics.
- (1999a) “The TDT-2 Text and Speech Corpus”, with C. Cieri, D. Graff, N. Martey and S. Strassel, *Proceedings of DARPA Broadcast News Workshop*, 1999.
- (1999b) “Topic Detection and Tracking using IDF-Weighted Cosine Coefficient”, with M. Schultz, *Proceedings of DARPA Broadcast News Workshop*, 1999 (<http://www.itl.nist.gov/iaui/894.01/proc/darpa99/index.htm>).
- (2000a) “Issues in Corpus Creation and Distribution”, with C. Cieri, pp. 49-56, *Proceedings of LREC-2000*, Athens.
- (2000b) “Large Multilingual Broadcast News Corpora for Cooperative Research in Topic Detection and Tracking”, with C. Cieri, D. Graff, M. Nii and S. Strassel, pp. 925-930, *Proceedings of LREC-2000*, Athens.
- (2000c) “ATLAS: A Flexible and Extensible Architecture for Linguistic Annotation”, with S. Bird, D. Day, J. Garofolo, J. Henderson and C. Laprun, pp. 1699-1706, *Proceedings of LREC-2000*, Athens.
- (2000d) “The Tonal Phonology of Yoruba Clitics”, with A. Akinlabi, in B. Gerlach and J. Grijzenhout, Eds., *Clitics in Phonology, Morphology and Syntax*, John Benjamins (pp. 31-63).
- (2001a) “Transcriber: development and use of a tool for assisting in the creation of speech corpora”, with C. Barras, E. Geoffrois and Z. Wu, *Speech Communication* 33.1-2 (pp. 5-22).
- (2001b) “A Formal Framework for Linguistic Annotation”, with S. Bird, *Speech Communication* 33.1-2 (pp. 23-60).
- (forthcoming) “Tonal Complexes and Tonal Alignment”, with A. Akinlabi, in NELS 31.

(forthcoming) “Towards a ‘Universal Dictionary’ for Multi-Language Information Retrieval Applications”, with M. Schultz, in J. Allan, J. Carbonell and J. Yamron, Eds., *Topic Detection and Tracking: Event-based Information Organization*, Kluwer Academic Press.

(manuscript) “Focus in Manding”, with M. Bamba.