

October 21, 2011

Dr. Łukasz Dębowski

Institute of Computer Science  
Polish Academy of Sciences  
ul. Ordona 21, 01-237 Warszawa  
Poland

Email: [ldebowsk@ipipan.waw.pl](mailto:ldebowsk@ipipan.waw.pl)

Homepage: [www.ipipan.waw.pl/~ldebowsk](http://www.ipipan.waw.pl/~ldebowsk)

## Curriculum vitae

- **Date and place of birth:**

23.03.1975 Łódź, Poland

- **Schools:**

10.1994–06.1999 Faculty of Physics, Warsaw University, Warsaw, Poland

09.1990–06.1994 I LO im. M. Kopernika, Łódź

- **Scientific degrees:**

10.2005 Ph.D. theoretical computer science (*with honors*)  
Institute of Computer Science, Polish Academy of Sciences  
Warsaw, Poland

06.1999 M.Sc. theoretical physics (*with honors*)  
Faculty of Physics, Warsaw University  
Warsaw, Poland

- **Interests:**

- information theory, source coding and stochastic processes with long-memory
- probabilistic modeling for computational and quantitative linguistics

- **Employment record:**

12.2005–now assistant professor (Polish: adiunkt)  
Institute of Computer Science, Polish Academy of Sciences  
Warsaw, Poland

01.2008–12.2009 post-doctoral fellow, Centrum voor Wiskunde en Informatica  
Amsterdam, Netherlands

06.2000–11.2005 assistant (Polish: asystent)  
Institute of Computer Science, Polish Academy of Sciences  
Warsaw, Poland

- **Participation in grants:**

01.2008–12.2009 Nederlandse Organisatie voor Wetenschappelijk Onderzoek,  
Vidi P4431, *Learning when all models are wrong*

- 01.2007–07.2007 Polish State Research Project, grant no. 3 T11C 003 28 *Automatyczna ekstrakcja wiedzy lingwistycznej z dużego korpusu języka polskiego* [Automatic extraction of linguistic knowledge from a large corpus of Polish]
- 08.2006–10.2006 Australian Research Council, grant no. DP0210999 *Asynchronous Continuous Time Conditioning*
- 03.2005–02.2006 Polish State Research Project, grant no. 1 P03A 045 28 *Entropia nadwyżkowa jako teorioinformacyjna miara zależności* [Excess entropy as an information-theoretic measure of dependence]
- 08.2002 Defense Advanced Research Projects Agency, *The Dynamics of Learning and the Emergence of Distributed Adaptation*
- 03.2001–06.2004 Polish State Research Project, grant no. 7 T11C 043 20 *Anotowany korpus pisanego języka polskiego z dostępem przez Internet — z uwzględnieniem zastosowań w inżynierii lingwistycznej* [An annotated Internet-accessible corpus of written Polish—for NLP applications]
- **Visits abroad:**
    - 01.2008–12.2009 Centrum voor Wiskunde en Informatica, Amsterdam, Netherlands (post-doctoral fellow)
    - 08.2006–10.2006 School of Computer Science and Engineering, University of New South Wales, Sydney, Australia (post-doctoral fellow)
    - 08.2002 Santa Fe Institute, Santa Fe, New Mexico, USA (visiting doctoral student)
    - 02.2001–05.2001 Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic (visiting doctoral student)
  - **Memberships:**
    - 01.2010–now Institute of Electrical and Electronics Engineers
    - 01.2008–now International Quantitative Linguistics Association
    - 01.2003–12.2003 Association for Computational Linguistics
  - **Awards:**
    - 01.2011 invited lecture at the workshop Randomness, Structure, and Causality: Measures of Complexity from Theory to Applications, Santa Fe Institute, Santa Fe, USA
    - 06.2007 travel grant for the IEEE International Symposium on Information Theory, Nice, France
  - **Activities for scientific community:**
    - 01.2007–now member of the Scientific Council of the Institute of Computer Science, Polish Academy of Sciences
    - 10.2010 member of the program committee of the 21st International Conference on Algorithmic Learning Theory, Canberra, Australia

- **Language skills:**

Polish (native), English (advanced — FCE, TOEFL, and TSE taken in 1993), Czech, Dutch, Esperanto, French, German, Russian (intermediate passive)

- **Programming experience:**

Linux, Perl, bash, C, Prolog, SQL,  $\LaTeX$ , Mathematica/Maxima

- **Other skills:**

driving licence (car, Poland)

## Publications

### A. Contributions to books:

- (1) Ł. Dębowski, (2011). Maximal Lengths of Repeat in English Prose. In: G. Altmann, S. Naumann, R. Vulcanović and P. Grzybek, eds., *Synergetic Linguistics. Text and Language as Dynamic System*.
- (2) Ł. Dębowski, (2010). A Link Between the Number of Set Phrases in a Text and the Number of Described Facts. In: P. Grzybek, E. Kelih, J. Mačutek, eds., *Text and Language: Structures, Functions, Interrelations. Quantitative Perspectives*. Wien: Praesens Verlag. (pp. 31–36)
- (3) Ł. Dębowski, (2006). Menzerath’s law for the smallest grammars. In: R. Köhler, P. Grzybek, eds., *The Exact Science of Language and Text*. Berlin: De Gruyter. (pp. 77–85)
- (4) Ł. Dębowski, (2004). Entropic subextensivity in language and learning. In: C. Tsallis, M. Gell-Mann, eds., *Nonextensive Entropy—Interdisciplinary Applications*. Oxford: Oxford University Press. (pp. 335–345)

### B. Articles in journals:

- (1) Ł. Dębowski, (2011). Excess entropy in natural language: present state and perspectives. *Chaos*, vol. 21, pp. 037105 (11 pages) (IF 2,081)
- (2) Ł. Dębowski, (2011). On processes with hyperbolically decaying autocorrelations. *Journal of Time Series Analysis*, vol. 32, pp. 580–584. (IF 0,747)
- (3) Ł. Dębowski, (2011). On the Vocabulary of Grammar-Based Codes and the Logical Consistency of Texts. *IEEE Transactions on Information Theory*, vol. 57, pp. 4589–4599. (IF 2,357)
- (4) Ł. Dębowski, (2010). Variable-Length Coding of Two-Sided Asymptotically Mean Stationary Measures. *Journal of Theoretical Probability*, vol. 23, pp. 237–256. (IF 0,750)
- (5) Ł. Dębowski, (2009). Valence extraction using EM selection and co-occurrence matrices. *Language Resources and Evaluation*, vol. 43, pp. 301–327. (IF 0,634)
- (6) Ł. Dębowski, (2009). A general definition of conditional information and its application to ergodic decomposition. *Statistics & Probability Letters*, vol. 79, pp. 1260–1268. (IF 0,386)
- (7) Ł. Dębowski, (2007). On processes with summable partial autocorrelations. *Statistics & Probability Letters*, vol. 77, pp. 752–759. (IF 0,386)
- (8) Ł. Dębowski, (2006). On Hilberg’s law and its links with Guiraud’s law. *Journal of Quantitative Linguistics*, vol. 13, pp. 81–109. (IF 0,645)
- (9) A. Przepiórkowski, P. Bański, Ł. Dębowski, E. Hajnicz, M. Woliński, (2005). The IPI PAN Corpus. *Polish Academy of Sciences. Annual Report*, pp. 54–55.
- (10) A. Przepiórkowski, P. Bański, Ł. Dębowski, E. Hajnicz, M. Woliński, (2003). Konstrukcja korpusu IPI PAN. [The design of the IPI PAN corpus.] *Polonica*, vol. XXII–XXIII, pp. 33–38.

- (11) Ł. Dębowski, J. Hajič, V. Kuboň, (2002). Testing the limits—Adding a new language to an MT system. *The Prague Bulletin of Mathematical Linguistics*, vol. 78, pp. 91-101.
- (12) Ł. Dębowski, (2002). Zipf's law against the text size: A half-rational model. *Glottometrics*, vol. 4 (To honor G. K. Zipf), pp. 49-60.

### C. Articles in conference proceedings:

- (1) Ł. Dębowski, (2009). Computable Bayesian Compression for Uniformly Discretizable Statistical Models. In: R. Gavaldà et al., eds., *Proceedings of the 20th International Conference on Algorithmic Learning Theory, ALT 2009, LNAI 5809*. (pp. 53–67)
- (2) Ł. Dębowski, M. Woliński, (2007). Argument co-occurrence matrix as a description of verb valence. In: *3rd Language & Technology Conference, October 5–7, 2007, Poznań, Poland*. (pp. 260-264)
- (3) Ł. Dębowski, (2007). On vocabulary size of grammar-based codes. In: *2007 IEEE International Symposium on Information Theory. Nice, France, July 25–29*. (pp. 91–95)
- (4) Ł. Dębowski, (2006). Excess entropy—a link between probabilistic and algorithmic approaches to mutual information. In: *Twenty-seventh Symposium on Information Theory in the Benelux. Noordwijk, The Netherlands, June 8–9, 2006*. (pp. 141–148)
- (5) Ł. Dębowski, (2004). Trigram morphosyntactic tagger for Polish. In: *Proceedings of New Trends in Intelligent Information Processing and Web Mining, Zakopane 2004*. Advances in Soft Computing. Berlin: Springer. (pp. 409–413)
- (6) A. Przepiórkowski, Z. Krynicki, Ł. Dębowski, M. Woliński, D. Janus, P. Bański, (2004). A search tool for corpora with positional tagsets and ambiguities. In: *The Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004*. (pp. 1235–1238)
- (7) Ł. Dębowski, (2003). A reconfigurable stochastic tagger for languages with complex tag structure. In: *Proceedings of the Workshop on Morphological Processing of Slavic Languages. 10th Conference of the European Chapter of Association for Computational Linguistics, Budapest 2003*. (pp. 63–70)
- (8) P. Bański, A. Przepiórkowski, A. Kupść, Ł. Dębowski, M. Marciniak, A. Mykowiecka, (2003). The design of the IPI PAN corpus. In: *Proceedings of PALC 2001: Practical Applications in Language Corpora*. Peter Lang – Europäischer Verlag der Wissenschaften. (pp. 225–232)
- (9) Ł. Dębowski, (2001). A revision of coding theory for learning from language. In: Geert-Jan Kruiff, Lawrence S. Moss, Richard T. Oehrle, eds., *Proceedings of Formal Grammar/Mathematics of Language Conference. August 10-12, 2001. Helsinki, Finland. Electronic Notes in Theoretical Computer Science, vol. 53*. Elsevier. (15 pages)

### D. Abstracts in conference proceedings:

- (1) Ł. Dębowski, (2006). The ergodic decomposition of excess entropy. In: *XXVI European Meeting of Statisticians. Toruń, 24–28 July 2006*. (p. 36)

- (2) Ł. Dębowski, (2005). Measure-theoretic information theory and its application to universal coding. In: *Warsaw Probability Meeting, 15–16 December 2005*. (3 pages in unnumbered leaflet)
- (3) Ł. Dębowski, (2005). Processes with absolutely summable partial autocorrelations. In: *Proceedings of the 14th European Young Statisticians Meeting, Debrecen, 22–26 August 2005*. (p. 19)
- (4) Ł. Dębowski, (2005). Excess entropy, ergodic decomposition, and universal codes. In: *Proceedings of the 30th Conference on Stochastic Processes and their Applications, Santa Barbara, 26 June–1 July 2005*. (p. 28)
- (5) Ł. Dębowski, (2004). On the sum of autocorrelations of a process with absolutely summable partial autocorrelations. In: *Proceedings of the 6th World Congress of the Bernoulli Society, Barcelona, 26–31 July 2004*. (pp. 88-89)
- (6) Ł. Dębowski, (2004). O sumie autokorelacji dla procesu stacjonarnego o bezwzględnie sumowalnej funkcji częściowej autokorelacji. [On the sum of autocorrelations of a stationary process with absolutely summable partial autocorrelations.] In: *VIII Konferencja z Probabilistyki. Będlewo 2004. Materiały konferencyjne*. (pp. 21–22)
- (7) Ł. Dębowski, (2003). The role of Hilberg’s law in statistical language modeling. *4th Trier Colloquium on Quantitative Linguistics, 17–18 October 2003*.

#### E. Technical reports and preprints:

- (1) E. Hajnicz, Ł. Dębowski, M. Wiech (2007). Przykładowe zastosowanie gradacyjnej analizy danych w badaniach lingwistycznych. [An example of application of grade data analysis in linguistic research.] IPI PAN Reports, nr 1005.
- (2) Ł. Dębowski, (2006). Ergodic decomposition of excess entropy and conditional mutual information. IPI PAN Reports, nr 993.
- (3) Ł. Dębowski, (2001). Tagowanie i dezambiguacja morfosyntaktyczna. Przegląd metod i oprogramowania. [Part-of-speech tagging and morphosyntactic disambiguation. A review of methods and software.] IPI PAN Reports, nr 934.

#### F. Theses:

- (1) Ł. Dębowski, (2005). Własności entropii nadwyżkowej dla procesów stochastycznych nad różnymi alfabetami. [Excess entropy for stochastic processes over various alphabets.] Ph.D. thesis, Institute of Computer Science, Polish Academy of Sciences. (supervised by dr hab. J. Mielniczuk)
- (2) Ł. Dębowski, (1999). Teoria funkcjonału gęstości dla monowarstw zaadsorbowanych na podłożu krystalicznym. [Density functional theory for monolayers adsorbed on a crystalline surface.] Master thesis, Faculty of Physics, Warsaw University. (supervised by dr hab. M. Napiórkowski)