

## Łukasz Dębowski

---

**CONTACT INFORMATION** Institute of Computer Science *Voice:* (+48) 22 3800 553  
Polish Academy of Sciences *Fax:* (+48) 22 3800 510  
ul. Jana Kazimierza 5 *E-mail:* ldebowski@ipipan.waw.pl  
01-248 Warszawa, Poland *WWW:* www.ipipan.waw.pl/~ldebowski

**BIRTH:** 23.03.1975, Łódź, Poland

**RESEARCH INTERESTS:** information theory, stochastic processes with long memory, probabilistic language modeling, statistical laws of language, universal data compression

**SCIENTIFIC DEGREES:** DR. HAB. IN MATHEMATICS/COMPUTER SCIENCE  
Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland (01.2015)  
PH. D. IN MATHEMATICS/COMPUTER SCIENCE (*with honors*)  
Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland (10.2005)  
M. SC. IN THEORETICAL PHYSICS (*with honors*)  
Faculty of Physics, University of Warsaw, Warsaw, Poland (06.1999)

**EMPLOYMENT RECORD:** Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland  
ASSOCIATE PROFESSOR (POLISH: PROFESOR NADZWYCZAJNY) (02.2015–now)  
Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan  
VISITING PROFESSOR (09.2015–11.2015)  
Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland  
ASSISTANT PROFESSOR (POLISH: ADIUNKT) (12.2005–01.2015)  
Centrum voor Wiskunde en Informatica, Amsterdam, Netherlands  
POST-DOCTORAL FELLOW (01.2008–12.2009)  
Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland  
RESEARCH ASSISTANT (POLISH: ASYSTENT) (06.2000–11.2005)

**SCHOOLS:** Faculty of Physics, University of Warsaw, Warsaw, Poland (10.1994–06.1999)  
I LO im. M. Kopernika, Łódź (09.1990–06.1994)

**TEACHING EXPERIENCE:** Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland  
Information theory and statistics (monograph lecture) (02.2015–04.2015)  
Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland  
Information theory and statistics (monograph lecture) (02.2014–04.2014)  
Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland  
Information theory and statistics (monograph lecture) (10.2012–01.2013)

**PARTICIPATION IN GRANTS:** *Mierzenie stopnia zrozumiałości polskich tekstów użytkowych (pozaliterackich).* [Measuring readability of Polish non-literary texts.]  
Polish State Research Project, grant no. UMO-2011/03/B/HS2/05799 (08.2012–12.2015)  
*Learning when all models are wrong.*  
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Vidi P4431 (01.2008–12.2009)  
*Automatyczna ekstrakcja wiedzy lingwistycznej z dużego korpusu języka polskiego.* [Automatic extraction of linguistic knowledge from a large corpus of Polish.]  
Polish State Research Project, grant no. 3 T11C 003 28 (01.2007–07.2007)  
*Asynchronous Continuous Time Conditioning.*  
Australian Research Council, grant no. DP0210999 (08.2006–10.2006)

*Entropia nadwyżkowa jako teorioinformacyjna miara zależności.* [*Excess entropy as an information-theoretic measure of dependence.*]

Polish State Research Project, grant no. 1 P03A 045 28 (03.2005–02.2006)

*The Dynamics of Learning and the Emergence of Distributed Adaptation.*

Defense Advanced Research Projects Agency (08.2002)

*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.*]

Polish State Research Project, grant no. 7 T11C 043 20 (03.2001–06.2004)

- VISITS ABROAD: School of Computer Science and Engineering, University of New South Wales, Sydney, Australia  
POST-DOCTORAL FELLOW (08.2006–10.2006)
- Santa Fe Institute, Santa Fe, New Mexico, USA  
DOCTORAL STUDENT (08.2002)
- Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic  
DOCTORAL STUDENT (02.2001–05.2001)
- Summer School in Cognitive Science, Central and Eastern European Center for Cognitive Science, New Bulgarian University, Sofia, Bulgaria  
DOCTORAL STUDENT (07.2000)
- Vilem Mathesius Courses, Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic)  
DOCTORAL STUDENT (03.2000)
- MEMBERSHIPS: International Quantitative Linguistics Association (01.2008–now)
- AWARDS: prize of the IV Department of Polish Academy of Sciences (12.2012)  
invited lecture at the workshop Randomness, Structure, and Causality: Measures of Complexity from Theory to Applications, Santa Fe Institute, Santa Fe, USA (01.2011)  
travel grant for the IEEE International Symposium on Information Theory, Nice, France (06.2007)  
invitation for 14th European Young Statisticians Meeting, Debrecen, Hungary (08.2005)  
invited lecture at the workshop Interdisciplinary Applications of Ideas from Nonextensive Statistical Mechanics and Thermodynamics, Santa Fe Institute, Santa Fe, USA (04.2002)
- SCIENTIFIC COMMUNITY ACTIVITIES: member of the Scientific Council of the Institute of Computer Science, Polish Academy of Sciences (01.2007–now)  
member of the program committee of the 21st International Conference on Algorithmic Learning Theory, Canberra, Australia (10.2010)
- 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, R, bash, C, Prolog, SQL, L<sup>A</sup>T<sub>E</sub>X, Mathematica/Maxima
- OTHER SKILLS: driving licence (car, Poland)

March 19, 2018

---

Wyrażam zgodę na przetwarzanie podanych danych osobowych, w celu realizacji procesu rekrutacyjnego, zgodnie z ustawą z dnia 29 sierpnia 1997 r. o ochronie danych osobowych (t. j. Dz. U. z 2002 r. Nr 101, poz. 926 ze zm.).

## Publications of Łukasz Dębowski

---

### MONOGRAPHS:

1. Ł. Dębowski, (2013). *Information Theory and Statistics*. Institute of Computer Science, Polish Academy of Sciences.

### CHAPTERS IN MONOGRAPHS:

1. E. Charzyńska, Ł. Dębowski, W. Gruszczyński, M. Hadryan, (2015). Historia badań nad zrozumiałością tekstu. [History of research in text readability.] In: W. Gruszczyński, M. Ogrodniczuk, eds., *Jasnopis, czyli mierzenie zrozumiałości polskich tekstów użytkowych*. [*Jasnopis — measuring readability of Polish non-literary texts*.] Warszawa: Oficyna Wydawnicza ASPRA-JR. (pp. 11–38)
2. Ł. Dębowski, (2015). Konstrukcja nowych formuł analitycznych. [Construction of new analytic formulas.] In: W. Gruszczyński, M. Ogrodniczuk, eds., *Jasnopis, czyli mierzenie zrozumiałości polskich tekstów użytkowych*. [*Jasnopis — measuring readability of Polish non-literary texts*.] Warszawa: Oficyna Wydawnicza ASPRA-JR. (pp. 109–126)

### ARTICLES IN JOURNALS:

1. Ł. Dębowski, (2018). Maximal Repetition and Zero Entropy Rate. *IEEE Transactions on Information Theory*, vol. 64(4), pp. 2212–2219.
2. Ł. Dębowski, (2018). Is Natural Language a Perigraphic Process? The Theorem about Facts and Words Revisited. *Entropy*, vol. 20(2), pp. 85.
3. Ł. Dębowski, (2017). Regular Hilberg Processes: An Example of Processes with a Vanishing Entropy Rate. *IEEE Transactions on Information Theory*, vol. 63(10), pp. 6538–6546.
4. R. Takahira, K. Tanaka-Ishii, Ł. Dębowski, (2016). Entropy Rate Estimates for Natural Language—A New Extrapolation of Compressed Large-Scale Corpora. *Entropy*, vol. 18(10), pp. 364.
5. Ł. Dębowski, (2015). The Relaxed Hilberg Conjecture: A Review and New Experimental Support. *Journal of Quantitative Linguistics*, vol. 22, pp. 311–337.
6. E. Charzyńska, Ł. Dębowski, (2015). Empirical verification of the Polish formula of text difficulty. *Cognitive studies / Etudes cognitives*, vol. 15, pp. 125–132.
7. Ł. Dębowski, (2015). Hilberg Exponents: New Measures of Long Memory in the Process. *IEEE Transactions on Information Theory*, vol. 61, pp. 5716–5726.
8. Ł. Dębowski, (2015). A Preadapted Universal Switch Distribution for Testing Hilberg’s Conjecture. *IEEE Transactions on Information Theory*, vol. 61, pp. 5708–5715.
9. Ł. Dębowski, (2015). Maximal Repetitions in Written Texts: Finite Energy Hypothesis vs. Strong Hilberg Conjecture. *Entropy*, vol. 17, pp. 5903–5919.
10. Ł. Dębowski, (2014). Hilberg’s Conjecture—a Challenge for Machine Learning. *Schedae Informaticae*, vol. 23, pp. 33–44.
11. R. Ferrer-i-Cancho, A. Hernández-Fernández, J. Baixeries, Ł. Dębowski, J. Mačutek, (2014). When is Menzerath-Altman law mathematically trivial? A new approach. *Statistical Applications in Genetics and Molecular Biology*, vol. 13, pp. 633–644.
12. Ł. Dębowski, (2014). On Hidden Markov Processes with Infinite Excess Entropy. *Journal of Theoretical Probability*, vol. 27, pp. 539–551.
13. R. Ferrer-i-Cancho, Ł. Dębowski, F. Moscoso del Prado Martin, (2013). Constant conditional entropy and related hypotheses. *Journal of Statistical Mechanics: Theory and Experiment*, L07001.
14. Ł. Dębowski, (2012). On Bounded Redundancy of Universal Codes. *Statistics and Probability Letters*, vol. 82, pp. 2068–2071.
15. Ł. Dębowski, (2012). Mixing, Ergodic, and Nonergodic Processes with Rapidly Growing Information between Blocks. *IEEE Transactions on Information Theory*, vol. 58, pp. 3392–3401.

16. Ł. Dębowski, (2011). Excess entropy in natural language: present state and perspectives. *Chaos*, vol. 21, pp. 037105 (11 pages)
17. Ł. Dębowski, (2011). On processes with hyperbolically decaying autocorrelations. *Journal of Time Series Analysis*, vol. 32, pp. 580–584.
18. 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.
19. Ł. Dębowski, (2010). Variable-Length Coding of Two-Sided Asymptotically Mean Stationary Measures. *Journal of Theoretical Probability*, vol. 23, pp. 237–256.
20. Ł. Dębowski, (2009). Valence extraction using EM selection and co-occurrence matrices. *Language Resources and Evaluation*, vol. 43, pp. 301–327.
21. Ł. Dębowski, (2009). A general definition of conditional information and its application to ergodic decomposition. *Statistics & Probability Letters*, vol. 79, pp. 1260–1268.
22. Ł. Dębowski, (2007). On processes with summable partial autocorrelations. *Statistics & Probability Letters*, vol. 77, pp. 752–759.
23. Ł. Dębowski, (2006). On Hilberg’s law and its links with Guiraud’s law. *Journal of Quantitative Linguistics*, vol. 13, pp. 81–109.
24. A. Przepiórkowski, P. Bański, Ł. Dębowski, E. Hajnicz, M. Woliński, (2005). The IPI PAN Corpus. *Polish Academy of Sciences. Annual Report*, pp. 54–55.
25. 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.
26. 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.
27. 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.

ARTICLES IN  
CONFERENCE  
PROCEEDINGS:

1. Ł. Dębowski, (2016). Consistency of the Plug-In Estimator of the Entropy Rate for Ergodic Processes. In: *2016 IEEE International Symposium on Information Theory*. (pp. 1651–1655)
2. Ł. Dębowski, B. Broda, B. Nitoń, E. Charzyńska, (2015). Jasnopis – A Program to Compute Readability of Texts in Polish Based on Psycholinguistic Research. In: B. Sharp, W. Lubaszewski and R. Delmonte, eds., *Natural Language Processing and Cognitive Science. Proceedings 2015*. Libreria Editrice Cafoscara. (pp. 51–61)
3. Ł. Dębowski, (2015). A New Universal Code Helps to Distinguish Natural Language from Random Texts. In: A. Tuzzi, M. Benešova, J. Mačutek, eds., *Recent Contributions to Quantitative Linguistics*. Berlin: De Gruyter Mouton. (pp. 41–50)
4. Ł. Dębowski, (2015). Regular Hilberg Processes: Nonexistence of Universal Redundancy Ratios. In: J. Rissanen, P. Harremoës, S. Forchhammer, T. Roos, P. Müllimäke, eds., *Proceeding of the The Eighth Workshop on Information Theoretic Methods in Science and Engineering*. University of Helsinki, Department of Computer Science. Series of Publications B, Report B-2015-1. (pp. 7–10)
5. Ł. Dębowski, (2012). Empirical Evidence for Hilberg’s Conjecture in Single-Author Texts. In: I. Obradović, E. Kelić, R. Köhler, eds., *Methods and Applications of Quantitative Linguistics—Selected papers of the 8th International Conference on Quantitative Linguistics (QUALICO)*. Belgrade: Academic Mind. (pp. 143–151)
6. Ł. Dębowski, (2012). Information-theoretic models of natural language. In: *The Fifth Workshop on Information Theoretic Methods in Science and Engineering, WITMSE 2012, 27-30 August 2012, Amsterdam*. (3 pages)

7. 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)
8. Ł. Dębowski, (2009). Computable Bayesian Compression for Uniformly Discretizable Statistical Models. In: R. Gavalda et al., eds., *Proceedings of the 20th International Conference on Algorithmic Learning Theory, ALT 2009, Porto, Portugal, October 3-5, LNAI 5809*. (pp. 53–67)
9. 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)
10. 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)
11. 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)
12. 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)
13. 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)
14. 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)
15. 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)
16. 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)
17. 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)

ABSTRACTS IN  
CONFERENCE  
PROCEEDINGS:

1. Ł. Dębowski, R. Takahira, K. Tanaka-Ishii, (2016). Large Scale Entropy Entropy Rate Estimation: A New Law that Governs the Complexity of Language. In: *Cognitive Systems Modeling: 5th Peripatetic Conference. Zakopane, Poland, 6–8 October 2016*. (pp. 12–13)
2. R. Ferrer-i-Cancho, Ł. Dębowski, (2013). Constant entropy rate and related hypotheses versus real language. In: *35th Annual Conference of the Cognitive Science Society. Berlin, Germany*.
3. Ł. Dębowski, (2006). The ergodic decomposition of excess entropy. In: *XXVI European Meeting of Statisticians. Toruń, 24–28 July 2006*. (p. 36)
4. 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)

5. 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)
6. 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)
7. 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)
8. 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)
9. Ł. Dębowski, (2003). The role of Hilberg’s law in statistical language modeling. *4th Trier Colloquium on Quantitative Linguistics, 17–18 October 2003.*

CONTRIBUTIONS TO  
FESTSCHRIFTS:

1. Ł. Dębowski, (2016). Estimation of entropy from subword complexity. In: S. Matwin, J. Mielniczuk, eds., *Challenges in Computational Statistics and Data Mining* New York: Springer. (pp. 53-70)
2. Ł. Dębowski, (2011). Maximal Lengths of Repeat in English Prose. In: G. Altmann, S. Naumann, R. Vulanović and P. Grzybek, eds., *Synergetic Linguistics. Text and Language as Dynamic System.* Wien: Praesens Verlag. (pp. 23–30)
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)

TECHNICAL  
REPORTS:

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.

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, University of Warsaw. (supervised by dr hab. M. Napiórkowski)

March 19, 2018

---

Wyrażam zgodę na przetwarzanie podanych danych osobowych, w celu realizacji procesu rekrutacyjnego, zgodnie z ustawą z dnia 29 sierpnia 1997 r. o ochronie danych osobowych (t. j. Dz. U. z 2002 r. Nr 101, poz. 926 ze zm.).