

Curriculum Vitae

Personal Details

Name	Harald Hammarström
Birthdate	Born 4th of August 1977 in Västerås, Sweden
Nationality	Swedish

Formal Education

2005-	PhD Student, Dept. of Comp. Sci., Chalmers University, Gothenburg, Sweden
2003-2005	Project Assistant on a VINNOVA Project, Chalmers University, Gothenburg, Sweden
2001-2002	Computer Science, University of Uppsala, Sweden
2001-2001	Master's Thesis, ILLC, Amsterdam, the Netherlands
2000-2001	Computer Science, University of Uppsala, Sweden
1999-2000	Computer Science, Facultad de Informática, Madrid, Spain
1997-1999	Computer Science, University of Uppsala, Sweden
1996-1997	Computer Science, University of Linköping, Sweden
1993-1996	High School (Natural Sciences focus), Västerås, Sweden
1984-1993	Elementary school, Västerås, Sweden

In parallel with the computer science studies 1997-2002, various less-than one semester courses were taken in Philosophy, Economics, Law, Nutrition, Linguistics, and various languages; See table 1.

Language	Credits or equiv.	Language	Credits or equiv.
Akkadian	15	Kurdish (Sorani)	10
Ancient Greek	50	Latin	20
Avestan	5	English	Fluent
Berber (Rif)	10 ^a	Old Church Slavonic	10 ^a
Biblical Aramaic	15	Pali	5
Biblical Hebrew	40	Pashto	20
Bulgarian	20	Persian	15
Catalan	5	Polish	15
Classical Arabic	40	Russian	15
Czech	10	Sanskrit	40
Dutch	10 ^b	Scottish Gaelic	10
Estonian	5	Somali	5 ^b
Finnish	15	Spanish	Fluent ^c
French	20 ^c	Sumerian	5
Ge'ez	5	Swahili	20
German	40 ^d	Swedish	Native
Hindi	40	Tamil	40
Hungarian	5	Turkish	30
Icelandic	5	Wolof	10 ^d
Japanese	10 ^e		

^a1 year course at Colegio Mayor Universitario "Nuestra Señora de Africa", Madrid

^bAt Gothenburg University

^c3 years in high school

^d5 years in high school

^eAt Gothenburg University

^aAt Gothenburg University

^bAt Gothenburg University

^c1 year stay in Madrid

^d1 year course at Colegio Mayor Universitario "Nuestra Señora de Africa", Madrid

Table 1: 20 old Swedish university credits = 30 ECTS credits = 1 semester full-time studies. 15 credits usually corresponds to the whole grammar of a language as well as a body of text material. > 15 means more text reading, study of dialects, and religion, history and politics etc of the applicable language area.

1 Experience

1.1 Councils and Committees

- Member of the Graduate Students' Council at GSLT (2005-2006)
- Member of the Board of Recruitment at the Dept. of IT, Uppsala Univ. 2001-2002
- Deputy member of the Board of Recruitment at Faculty of Languages, 2001-2002-2003
- Member of the Library Committee, 2001-2002
- Member of the Students' Council at the Dept. of Afro-Asiatic Languages, 2001-2002
- Member of the Students' Council at the Dept. of Romance Languages, 2002-2003

1.2 Conference/workshop Organization

- Assist: Swedish conferences in Language Technology (2005, 2005, 2006)
- Assist: Nordic Symposium of African Languages (2007)
- Assist: International Conference on Bantu Languages (2007)
- Assist: 14th Himalayan Linguistic Symposium (2008)
- Assist: GoTAL 6th International Conference on Natural Language Processing (2008)
- Assist: 23rd Scandinavian Conference of Linguistics (2008)
- Assist: Workshop on Language Documentation and Language Description (2008)
- Chair: Workshop on Unsupervised Methods in NLP (2008)
- Chair: Gothenburg Typological Seminars (seminar series 2007-)

1.3 Teaching

- Assist: Artificial Intelligence (2004, 2005, 2006, 2007, 2008)
- Assist: Databases (2004, 2005, 2006)
- Assist: Machine Learning (2007, 2008)
- Assist: Formal Languages and Automata Theory (2007, 2008)
- Assist: Graduate Course in Language Typology (2007)
- Co-Chair: Languages of the World (2008)

1.4 International Collaboration

- 4 weeks MPI EVA, Leipzig on Linguistic Typology (2007)
- 6 weeks IIT Bombay, Mumbai on Computational Morphology for Hindi and Marathi (2008)
- 2 weeks Dept. of African Languages, Helsinki (2008)
- 2+1 weeks Dept. of Linguistics, UNISA, South Africa on tools for exploitation of (transcribed) spoken Xhosa language data (2007+2009)

1.5 Work

Employer	Location	Duration	Task
Britsoft	Hertford, UK	2 weeks, Summer 1995	Assembly (IBM PC) Programming
ABB Industrial Systems	Västerås, Sweden	10 weeks, Summer 1996	Programming (Python) test tools for shared memory objects
ABB Industrial Systems	Västerås, Sweden	10 weeks, Summer 1997	Programming (Java) monitor tools for shared memory objects
ABB Industrial Systems	Västerås, Sweden	10 weeks, Summer 1998	Programming (C++) test tools for COM Objects
Dept. of Linguistics	Uppsala	10 weeks, 2002	Locating Resources for Computer Assisted Grammar Teaching

2 Grants

- In high school:
 - Scholarship (not applied for) of US\$2900 for skill of mathematics (esp. problem solving), 1996
- As an undergraduate:
 - Scholarship (applied for) of US\$3000 for general diligence in studies, 1998
 - Scholarship (applied for) of US\$1800 for general diligence in studies, 1999
 - Scholarship (applied for) of US\$12000 for general diligence in studies, 2000
- As a PhD Student:

- Many conference travel grants (2004-2008)
- Project Grant US\$8000 for 'Language Technology for Languages of the Central African Republic' from CLT (2007)

3 Publications/Presentations etc.

3.1 Popular Press

- 2009-01-11: Expressen Söndagsbilaga (the Sunday 'lifestyle' appendix to a tabloid-like national daily) had a 3-page interview with me since I have more university credits than anyone else in Sweden (according to official statistics gathered by them).
- 2010-05-06: The Times science magazine featured an article about numeral systems and number bases based on my work (with due mention of my name) <http://www.timesonline.co.uk/tol/news/science/eureka/article7112969.ece>.

3.2 Peer-Reviewed Books/Articles/Book Chapters etc

[Including submitted but not 'in prep.' work]

- Jens Allwood, Harald Hammarström, Rusandre Hendrikse, Mtholeni N. Ngcobo, Nozibele Nomdebevana, Laurette Pretorius 2010 Work on Spoken (multimodal) Language Corpora in South Africa, to appear in LREC-2010 Proceedings, Malta, May.
- Harald Hammarström 2010 A Full-Scale Test of the Language Farming Dispersal Hypothesis, 26pp + 340pp appendix, to appear in Diachronica
- Harald Hammarström 2010 The Status of the Least Documented Language Families in the World, 34pp., to appear in Language Documentation & Conservation
- H. Hammarström 2009 Poor man's word-segmentation: Unsupervised morphological analysis for Indonesian. Proceedings of the Third International Workshop on Malay and Indonesian Language Engineering (MALINDO). Singapore: ACL.

- H. Hammarström 2009 Whence the Kanum base-6 Numeral System? *Linguistic Typology* 13(2):305-319. H. Hammarström 2009 Rarities in Numeral Systems. In: Jan Wohlgemuth and Michael Cysouw (eds.) *Rara and Rarissima: Collecting and interpreting unusual characteristics of human languages [Empirical Approaches to Language Typology]*, pp. 7-55, MOUTON.
- H. Hammarström, Christina Thornell, Malin Petzell and Torbjörn West-erlund 2008 Bootstrapping Language Description: The case of Mpiemo (Bantu A, Central African Republic). (.pdf) In: *Proceedings of LREC-2008*, pp. 3350-3554. European Language Resources Association (ELRA). H. Hammarström 2009 Sampling and Genealogical Coverage in the WALS. *Linguistic Typology* 13(1):105-119, plus 198pp appendix.
- Harald Hammarström 2008 Automatic Annotation of Bibliographical References with Target Language, In: *Proceedings of MMIES-2: Multi-source, Multilingual Information Extraction and Summarization*, Manchester, 23 August, 2008
- H. Hammarström 2008 Counting Languages in Dialect Continua Using the Criterion of Mutual Intelligibility. *Journal of Quantitative Linguistics* 15(1), pp. 36-45.
- H. Hammarström 2007 A Survey and Classification of Methods for (Mostly) Unsupervised Learning of Morphology. In: *Proceedings of NODALIDA 2007, the 16th Nordic Conference of Computational Linguistics*, Tartu, Estonia, 25-26 May 2007, NEALT.
- H. Hammarström 2007 A Fine-Grained Model for Language Identification. In: *Proceedings of iNEWS-07 Workshop at SIGIR 2007*, 23-27 July 2007, Amsterdam, pp. 14-20, ACM..
- H. Hammarström 2007 Complexity in Numeral Systems with an Investigation into Pidgins, Pidgincreoles and Creoles. In F. Karlsson and M. Miestamo and K. Sinnemäki (eds.) *Language Complexity: Typology, Contact, Change*, pp. 287-304, JOHN BENJAMINS [Studies in Language Companion Series 94]
- H. Hammarström 2006 A Naive Theory of Morphology and an Algorithm for Extraction. (.pdf) In R. Wicentowski and G. Kondrak (eds.)

SIGPHON 2006: Eighth Meeting of the Proceedings of the ACL Special Interest Group on Computational Phonology, 8 June 2006, New York City, USA, pp. 79-88, Association for Computational Linguistics.

- M. Forsberg, H. Hammarström and A. Ranta 2006 Lexicon Extraction from Raw Text Data. In: Salakoski, T. and Ginter, F. and Pyysalo, S. and Pahikkala, T. (eds.) *Advances in Natural Language Processing: Proceedings of the 5th International Conference, FinTAL 2006* Turku, Finland, August 23-25, 2006, pp. 488-499, SPRINGER, LNCS 4139.
- M. Humayoun, H. Hammarström, and A. Ranta. Urdu Morphology, Orthography and Lexicon Extraction. CAASL-2: The Second Workshop on Computational Approaches to Arabic Scriptbased Languages, July 21-22, 2007, LSA 2007 Linguistic Institute, Stanford University. 2007.
- H. Hammarström 2007 Handbook of Descriptive Language Knowledge: A Full-Scale Reference Guide for Typologists. 312 pp. LINCUM Handbooks in Linguistics, LINCUM.
- H. Hammarström 2006 Poor Man's Stemming: Unsupervised Recognition of Same-Stem Words. In: Ng, Hwee Tou and Leong, Mun-Kew and Kan, Min-Yen and Ji, Donghong (eds.) *Information Retrieval Technology: Proceedings of the Third Asia Information retrieval Symposium, AIRS 2006*, Singapore, October 2006, pp. 323-337, SPRINGER, LNCS 4182.
- H. Hammarström 2006 A New Algorithm for Unsupervised Induction of Concatenative Morphology. In: *Finite State Methods in Natural Language Processing: 5th International Workshop, FSMNLP 2005*, Helsinki, Finland, September 1-2, 2005. Revised Papers, pp. 288-289, SPRINGER, LNCS 4002.
- H. Hammarström 2004 Deduction of Numeral Grammars. In Paul Egré and Laura Alonso i Alemany (Eds.) *Proceedings of the Ninth ESSLLI Student Session*, August 2004, Nancy, France.
- H. Hammarström 2004 Properties of Lower Numerals and their Explanation: A Reply to Pawel/ Rutkowski. *Journal of Universal Language* 5:2, 2004, pp. 1-20.

Submitted:

- Holman, Eric W., Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Hagen Jung, Dik Bakker, Pamela Brown, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Johann-Mattis List, and Dmitry Egorov 2009 Automated dating of the world's language families, Submitted to *Current Anthropology*
- Don Killian and Harald Hammarström 2009 Notes on the Morphosyntax of Uduk (Koman, Sudan) from mss. sources, submitted to the WOCAL 2009 proceedings
- Aymeric Daval-Rasmussen, Peter Bakker and Harald Hammarström 2008 On the Origins of the Malagasy Numeral System, 28 pp, Submitted to *Studies in Language*
- Harald Hammarström 2008 Two Hitherto Unnoticed Languages from Sobger River, Papua, Indonesia, 15pp., Submitted to *Bijdragen tot de taal-, land- en Volkenkunde*
- Harald Hammarström and Sebastian Nordhoff 2010 LangDoc: Bibliographic Infrastructure for Linguistic Typology, 13pp., Submitted to special issue of *Oslo Studies in Language*
- Harald Hammarström and Lars Borin 2010 A Survey of work on Unsupervised Learning of Morphology, 36pp., for *Computational Linguistics*
- Harald Hammarström 2010 A Survey of Computational Morphological Resources for Low-Density Languages, Submitted to *Language Resources and Evaluation Journal*

3.3 Other

- 21 lightly peer-reviewed conference abstracts
- 7 scholarly reviews
- 8 book notices

3.4 Theses

- PhD: H. Hammarström 2009 Unsupervised Learning of Morphology and the Languages of the World. (.pdf) Thesis for the Degree of Doctor of Engineering, Department of Computer Science and Engineering, Chalmers University. 284 pp.
- Licentiate: H. Hammarström 2007 Unsupervised Learning of Morphology: Survey, Model, Algorithm and Experiments. Thesis for the Degree of Licentiate of Engineering, Department of Computer Science and Engineering, Chalmers University. 91 pp.
- MA: H. Hammarström 2002 Shallow Reasoning on Open-Domain Texts. (.pdf), Master's Thesis, Department of Computer Science, Uppsala University, 2002.

Address

Harald Hammarström
c/o Marie-Louise Groen
Sint Annastraat 250 6525 HA Nijmegen
THE NETHERLANDS

Tel.: +31-646-747933

E-Mail: h.hammarstrom@let.ru.nl

Göteborg, Sweden, 11. Jan 2010