

Dr. Kim Luyckx

Curriculum vitae

Last update: December 12, 2012

Affiliations

- Project leader text mining @ biomina, Antwerp University Hospital
- Associate researcher @ CLiPS, University of Antwerp

Personalialia

Last name	Luyckx
First name	Kim
Work address	biomina Antwerp University Hospital (UZA) Wilrijkstraat 10 B-2650 Edegem, Belgium
Tel.	+32 (0)3 821 39 94
E-mail	kim.luyckx@uza.be (primary) kim.luyckx@ua.ac.be
Personal website	http://www.clips.ua.ac.be/~kim
Group website	http://www.biomina.be http://www.clips.ua.ac.be
Nationality	Belgian
Date of Birth	24 December 1982
Marital status	Married, two children

Research interests

Text mining | Machine Learning | Computational linguistics | Mining of clinical free text

1. Education

Oct 2005 - Dec 2010: *PhD in Computational Linguistics*

University of Antwerp

PhD dissertation "Scalability Issues in Authorship Attribution" (supervised by Walter Daelemans)

Oct 2002 - June 2004: *MA in Germanic Linguistics and Literature (Dutch/English) with majors in Linguistics and Computational Linguistics*

University of Antwerp

MA dissertation "Syntax-Based Features and Machine Learning Techniques for Authorship Attribution" (supervised by Walter Daelemans)

Oct 2002 - June 2004: *Academic Teacher Training (Dutch/English)*

University of Antwerp

Oct 2000 - June 2002: *BA in Germanic Linguistics and Literature (Dutch/English)*

University of Antwerp

Sept 1995 - June 2000: *Secondary Education: Latin and Modern Languages*

Instituut Berkenboom Humaniora, Sint-Niklaas

2. Experience

Since Jan 2012: Project leader text mining @ biomina

University Hospital Antwerp

The biomina interdisciplinary research unit is a joint initiative of the University Hospital Antwerp and the University of Antwerp. Our focus is on R&D in biomedical informatics. I develop text mining applications for the hospital, enabling intelligent access to and analysis of the hospital's textual data.

Oct 2011 - Sept 2012: Teaching replacement position (10%)

Université Libre de Bruxelles

STIC-B-425 Ingénierie Linguistique (computational linguistics course for MA and MSc students)

Oct 2010 - Sept 2011: Postdoctoral researcher and coordinator of AMiCA

University of Antwerp

AMiCA (Automatic Monitoring for Cyberspace Applications) was a one-year cross-university pre-project (funded by IWT-SBO) with primary societal goal. My responsibilities were to assess the technological feasibility of the project, assemble a user committee consisting of stakeholders from various types of organisations (social media providers, software development companies, government, law enforcement, mental health and social care organisations), discuss transfer between the project and the users, and write the full IWT-SBO project.

The four-year project was granted and starts on January 1, 2013.

Mar 2007 - Sept 2010: PhD student in computational stylometry

University of Antwerp

In the 'Computational Techniques for Stylometry for Dutch' project (funded by FWO), style characteristics are automatically extracted and analysed using automatic linguistic analysis and supervised and unsupervised machine learning techniques. Within the framework of this project, I wrote my PhD dissertation, supervised by Walter Daelemans.

Oct. 2004 - Sept. 2006: Junior researcher in speech recognition project

University of Antwerp

In the FLVoR (Flexible Large Vocabulary Recognition) project (funded by IWT-GBOU), a modular architecture for speech recognition which involves domain-specific lexical and syntactic knowledge was created. I was responsible for the development of the morpho-phonological language model.

Sept. 2004: Teacher of Dutch @ De Veerman secondary school

3. Projects as promotor

- IOF-POC project 'Development of an automated software platform to support and improve the quality of clinical coding'; Promotors: *Tim Van den Bulcke*, Koen Smets, Bart Goethals, Kim Luyckx, Leon Luyten - duration: Jan. 2013 - Dec. 2013 - budget: € 100.000.

4. Teaching and supervision

4.1 Teaching

2012-2013

- guest lecture on 'Biomedical and clinical text mining' in 'Data mining' course @ UA (Master in Biochemistry; course by Kris Laukens)

2011-2012

- 'Computational Linguistics' @ UA (MA course; with Walter Daelemans and Guy De Pauw)
- 'Computational Linguistics' @ ULB (MA/MSc course; with Guy De Pauw and Max De Wilde)

2010-2011

- 'You are how you write' session @ UA in the framework of Wetenschapsweek (<http://www.wetenschapsweek.be>), an annual initiative to acquaint high school students with research at universities in Flanders

2009-2010

- guest lecture on 'Natural language processing' in 'Computational Tools for Linguistics' @ UA (MA course by Steven Gillis and Thomas Crombez)

2004-2005

- guest lecture on 'Corpus linguistics using the Spoken Dutch Corpus (CGN)' in 'Natural Language Processing' @ UA (MA course by Steven Gillis)

4.2 (Co-)supervision

2012-2013

- E. Scheurwegs (Artesis Hogeschool): de-identification of clinical free text

2010-2011

- L. Douws (MA in Linguistics, UA, 2011): personality prediction
- L. Van laer (MA in Linguistics, UA, 2011): personality prediction

2009-2010

- L. Douws (BA in Linguistics and Literature, UA, 2010): authorship attribution

2008-2009

- S. Pollak (Advanced MA in Computational linguistics, UA, 2009): dissertation on text classification of press articles on Kenyan elections

5. Scientific activities

- Member of the editorial board of the Computational Linguistics in the Netherlands journal (<http://clinjournal.org>) [2011 - ...]
- Member of the programme committee of the PAN lab (Uncovering Plagiarism, Authorship, and Social Software Misuse) [2011 - ...]
- Reviewer for
 - Literary and Linguistic Computing [2011]
 - Authorship [2011]
 - Natural Language Engineering [2011]
 - Proceedings of the Royal Society A: Mathematical, physical, and engineering sciences [2011]
 - International Journal of Web Engineering and Technology [2011]
 - The annual Digital Humanities conference [2009 - ...]
 - The 2011 PAN lab
 - The 2011 Computational Medicine Challenge (NLP track)

6. Organization

- Local organization of ECML PKDD 2008 (with members of CNTS and ADReM): "European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", University of Antwerp, 15-19 September 2008.
- Organizer of ATILA 2008 (with Roser Morante): 12th ATILA Research Meeting, University of Antwerp, 14-15 October 2008
- Organizer of the CNTS Research Assembly 2008 (with Roser Morante): Poster session organized on the occasion of CNTS's 15th anniversary, University of Antwerp, 7 November 2008

- Organizer of CNTS/CLiPS colloquia (2006-2012): Series of invited talks by key researchers in Natural Language Processing

7. Awards

- 2008: Young Scholar Award of the Association for Literary and Linguistic Computing; 500 GBP

8. Memberships

- PASCAL Network (Network of Excellence funded by the European Union)
- Computational Linguistics in Flanders (CLIF)
- Association for Literary and Linguistic Computing (ALLC)
- ACL SIG on Language Learning (SIGNLL)

9. Skills

9.1. Technical skills

- Python, Shell, Unix, Bash, AWK, HTML, CSS, CGI, XML, XSL, SQL
- Machine Learning: TiMBL, WEKA, MaxEnt
- Statistics: Matlab/Pylab, R, Excel
- Latex, iWork, MS Office
- Operating systems: Unix, Linux, Mac, Windows
- Apache SOLR

9.2 Languages

Dutch: mother tongue

English: near-native

French: good working knowledge

German: good working knowledge

10. Varia

- Webmaster for research centres CNTS (2005-'09) and CLiPS (i.e. a merge of CNTS and CPL) (2009-2012) at University of Antwerp.
- Author of the Taal&TechnologieBlog (in Dutch; 2008-'09), a weblog powered by EOS, a popular science magazine.

Scientific output

- Google Scholar page: <http://scholar.google.be/citations?user=UqZRsAAAAJ>
- # citations: 98 (overall), 95 (since 2007)
- h index: 5

A. Publications

Dissertations (2)

- K. Luyckx. Scalability Issues in Authorship Attribution. PhD Dissertation, University Press Antwerp, ISBN 978-90-5487-8230, 2010.
- K. Luyckx. Syntax-Based Features and Machine Learning Techniques for Authorship Attribution. MA Dissertation, University of Antwerp, Unpublished, 2004.

Journal papers (3)

2012

- M. Kestemont, K. Luyckx, W. Daelemans, T. Crombez. Unmasking for cross-genre authorship verification. *English Studies*, 93(3), 340-356, 2012. [Listed in the Thomson Reuters Arts & Humanities Citation Index®]
- K. Luyckx, F. Vaassen, C. Peersman, and W. Daelemans, Fine-grained emotion detection in suicide notes: A thresholding approach to multi-label classification, *Biomedical Informatics Insights*, 2012(5, Suppl. 1), 61-69, 2012.

2011

- K. Luyckx and W. Daelemans. The Effect of Author Set Size and Data Size in Authorship Attribution. *Literary and Linguistic Computing*, 26(1), pp. 35-55, 2011. [Impact Factor: 0.431]

Book chapters (2)

2007

- K. Luyckx, H. Kloots, E. Coussé and S. Gillis. Klankfrequenties in het Nederlands. In: Sandra e.a. (eds.) *Tussen taal, spelling en onderwijs. Essays bij het emeritaat van Frans Daems*, pp. 141-154, Gent, Academia Press, 2007.

2005

- K. Luyckx and W. Daelemans. Shallow Text Analysis and Machine Learning for Authorship Attribution. In: van der Wouden (ed.) *Computational Linguistics in the Netherlands 2004: selected papers from the 15th CLIN Meeting*, Leiden, The Netherlands, pp. 149-160, 2005.

Conference proceedings (10)

2012

- M. Kestemont, C. Peersman, B. De Decker, G. De Pauw, K. Luyckx, R. Morante, F. Vaassen, J. van de Loo, and W. Daelemans. The Netlog Corpus: A resource for the study of Flemish Dutch Internet language. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pp. 1569-1572, 2012.

- M. Kestemont, K. Luyckx, W. Daelemans, and T. Crombez, Evaluating unmasking for cross-genre authorship verification. Proceedings of Digital Humanities 2012 (DH 2012), pp. 249-252, 2012.

2011

- K. Luyckx. Authorship Attribution of E-mail as a Multi-Class Task. Notebook for PAN at CLEF 2011. In V. Petras, P. Forner, P.D. Clough (eds.) CLEF 2011 Labs and Workshop, Notebook Papers, Amsterdam, The Netherlands, ISBN 978-88-904810-1-7, 2011.
- M. Kestemont, K. Luyckx, and W. Daelemans. Intrinsic Plagiarism Detection Using Character Trigram Distance Scores. Notebook for PAN at CLEF 2011. In V. Petras, P. Forner, P.D. Clough (eds.) CLEF 2011 Labs and Workshop, Notebook Papers, Amsterdam, The Netherlands, ISBN 978-88-904810-1-7, 2011.

2008

- K. Luyckx and W. Daelemans. Authorship Attribution and Verification with Many Authors and Limited Data. In Proceedings of the 20th Belgian-Netherlands Conference on Artificial Intelligence (BNAIC 2008) [compressed version of COLING 2008 paper], pp. 335-336, Enschede, The Netherlands, 2008.
- K. Luyckx and W. Daelemans. Authorship Attribution and Verification with Many Authors and Limited Data. In Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008), pp. 513-520, Manchester, UK, 2008.
- K. Luyckx and W. Daelemans. Using Syntactic Features to Predict Author Personality from Text. In Proceedings of Digital Humanities 2008, pp. 146-149, Oulu, Finland, 2008.
- I. Hendrickx, W. Daelemans, K. Luyckx, R. Morante and V. Van Asch. CNTS: Memory-Based Learning of Generating Repeated References. In Proceedings of the 5th International Natural Language Generation Conference (INLG 2008), pp. 194-195, Salt Fork, Ohio, USA, 2008.
- K. Luyckx and W. Daelemans. Personae: A Corpus for Author and Personality Prediction from Text. In Proceedings of the 6th International Language Resources and Evaluation Conference (LREC 2008), Marrakech, Morocco, 2008.

2006

- K. Luyckx, W. Daelemans and E. Vanhoutte. Stylogenetics: Clustering-based stylistic analysis of literary corpora. In Proceedings of the 5th International Language Resources and Evaluation Conference (LREC 2006), Workshop "Toward Computational Models of Literary Analysis", Genova, Italy, pp. 30-35, 2006.

Scientific book reviews (3)

2008

- K. Luyckx. Stylistics: Prospect and Retrospect. [Review of Hoover and 2008 Lattig (eds.), idem, Rodopi, 2007], Literary and Linguistic Computing, 23(4), pp. 501-502, 2008.
- K. Luyckx. Corpus Linguistics and the Web. [Review of Hundt, Nesselhauf and Biewer (eds.), idem, Rodopi, 2007], Literary and Linguistic Computing, 23(2), pp. 246-248, 2008.
- K. Luyckx. Corpus Linguistics Beyond the Word. Corpus Research from Phrase to Discourse. [Review of Fitzpatrick (ed.), idem, Rodopi, 2007], Literary and Linguistic Computing, 23(1), pp. 123-125, 2008.

B. Invited talks (3)

2012

- K. Luyckx. *Introduction to Computational Stylometry*. Lecture at the Digital Humanities Summer School @ KULeuven 2012, Leuven, Belgium, September 20, 2012.

2011

- K. Luyckx. *Text Analytics and Machine Learning for User Identification and Content Reliability*. Keynote talk at Benelearn 2011 (20th Annual Belgian Dutch Conference on Machine learning), The Hague, The Netherlands, May 20, 2011.

2009

- K. Luyckx. *Computational Stylometry: Predicting author characteristics from text*. Invited talk at Textkernel BV, Amsterdam, The Netherlands, April 24, 2009.

C. Conference presentations (26)

2012

- M. Kestemont, K. Luyckx, W. Daelemans, and T. Crombez. Evaluating unmasking for cross-genre authorship verification, Digital Humanities 2012 (DH 2012), 18 July, Hamburg, Germany.
- M. Kestemont, C. Peersman, B. De Decker, G. De Pauw, K. Luyckx, R. Morante, F. Vaassen, J. van de Loo, and W. Daelemans. The Netlog Corpus: A resource for the study of Flemish Dutch Internet language. 8th International Conference on Language Resources and Evaluation (LREC 2012), 24 May, Istanbul, Turkey.
- K. Luyckx, F. Vaassen, C. Peersman, and W. Daelemans, A thresholding approach to emotion detection in suicide notes. 22nd Meeting of Computational Linguistics in The Netherlands (CLIN 22), January 20, Tilburg, The Netherlands.
- M. Kestemont, K. Luyckx, W. Daelemans, and T. Crombez, Cross-genre authorship verification using unmasking. 22nd Meeting of Computational Linguistics in The Netherlands (CLIN 22), January 20, Tilburg, The Netherlands.

2011

- K. Luyckx, F. Vaassen, C. Peersman, and W. Daelemans, Fine-grained emotion detection in suicide notes: Multi-label classification with probability estimates. 15th ATILA Research Meeting, December 1, Antwerp, Belgium.
- K. Luyckx, F. Vaassen, C. Peersman, and W. Daelemans, Fine-grained emotion detection in suicide notes: Multi-label classification with probability estimates. 2011 i2b2/VA/Cincinnati Workshop on the i2b2/VA Track on Challenges in Natural Language Processing for Clinical Data. October 21, Washington DC, USA.
- K. Luyckx. Authorship Attribution of E-mail as a Multi-Class Task, PAN lab at CLEF 2011, Amsterdam, The Netherlands, September 21, 2011.
- M. Kestemont, K. Luyckx, and W. Daelemans. Intrinsic Plagiarism Detection Using Character Trigram Distance Scores, PAN lab at CLEF 2011, Amsterdam, The Netherlands, September 22, 2011.
- K. Luyckx and W. Daelemans. Combined Qualitative and Quantitative Error Analysis in Multi-Topic Authorship Attribution, 21st Meeting of Computational Linguistics in the Netherlands, Ghent, Belgium, February 11, 2011.

2010

- K. Luyckx and W. Daelemans. Experimental Design in Multi-Topic Authorship Attribution. 14th ATILA Research Meeting, Ostend, Belgium, September 23, 2010.

2009

- K. Luyckx. Computational Stylometry: Predicting Author Identity and Personality from Text. Graduate Conference in Linguistics, University of Antwerp, Belgium, April 1, 2009.
- K. Luyckx and W. Daelemans. TACTiCS, a Tool for Analyzing and Categorizing Texts using Characteristics of Style. 19th Meeting of Computational Linguistics in The Netherlands (CLIN), Groningen, The Netherlands, January 22, 2009.

2008

- K. Luyckx and W. Daelemans. Authorship Attribution and Verification with Many Authors and Limited Data.

- 22nd International Conference on Computational Linguistics (COLING 2008), Manchester, UK, August 19, 2008.
- 12th ATILA Research Meeting, Antwerp, Belgium, October 15, 2008.
- 20th Belgian-Netherlands Conference on Artificial Intelligence (BNAIC 2008), Enschede, The Netherlands, October 31, 2008.
- K. Luyckx and W. Daelemans. Predicting Author Identity and Personality from Text using Many Authors and Limited Data. Poster presented at the 3rd Computer Intelligence and Learning (CIL) PhD student day (collocated with ECML PKDD 2008), Antwerp, Belgium, September 19, 2008.
- K. Luyckx and W. Daelemans. Using Syntactic Features to Predict Author Personality from Text. Digital Humanities 2008, Oulu, Finland, June 26, 2008.
- I. Hendrickx, W. Daelemans, K. Luyckx, R. Morante and V. Van Asch. CNTS: Memory-Based Learning of Generating Repeated References. 5th International Natural Language Generation Conference (INLG 2008), Salt Fork, Ohio, USA, June 12, 2008.
- K. Luyckx and W. Daelemans. Personae: A Corpus for Author and Personality Prediction from Text. 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco, May 28, 2008.
- K. Luyckx. Authorship Attribution on a Large Set of Authors. 3rd Computational Linguistics in Flanders (CLIF) Symposium on Language and Speech Technology, Leuven, Belgium, February 5, 2008.

2007

- W. Daelemans and K. Luyckx. Predicting author personality from text. 18th Meeting of Computational Linguistics in The Netherlands (CLIN 2007), Nijmegen, The Netherlands, December 7, 2007.
- K. Luyckx and W. Daelemans. Work in progress in predicting personality from text. 11th ATILA Research Meeting, Gent, Belgium, November 14, 2007.

2006

- K. Luyckx and W. Daelemans. Stylometry research using syntax-based features and machine learning techniques. 1st Computer Intelligence and Learning (CIL) Doctoral School Research Day, ULB, Brussels, Belgium, September 25, 2006.
- K. Luyckx, W. Daelemans and E. Vanhoutte. Stylogenetics: Clustering Based Stylistic Analysis of Literary Corpora. 5th International Language Resources and Evaluation Conference (LREC 2006), Workshop "Toward Computational Models of Literary Analysis", Genova, Italy, May 22, 2006.

2005

- K. Luyckx. Modeling Dutch pronunciation variation. 9th ATILA Research Meeting, Corsendonk, Belgium, November 28, 2005.

2004

- K. Luyckx and W. Daelemans. Combining Shallow Text Analysis and Machine Learning Techniques for Authorship Attribution. 15th Meeting of Computational Linguistics in The Netherlands