

# Curriculum Vitae ~ Maarten Janssen

## Personal Data:

Name: Maarten Janssen  
Address: Praça Dr. Bern. Ant. Gomes 180-2dt  
1100-209 Lisboa  
Portugal  
Nationality: Dutch  
Date of Birth: January 28, 1971  
Place of Birth: Nijmegen, The Netherlands

## Academic Preparation

**PhD in Linguistics/Artificial Intelligence**, Utrecht University, 2002

Concentrations: Multilingual Lexical Databases, Lexical Semantics

Dissertation: *SIMuLLDA: a Multilingual Lexical Database Application using a Structured Interlingua*

Advisors: Prof. H.J. Verkuyl, Prof. A. Visser

**MA in Cognitive Artificial Intelligence**, Utrecht University, 1996

Concentrations: Event Semantics, Aspect, Philosophy of Language, Categorical Grammar

Thesis: *The Parsons Project*

Advisors: Prof. H.J. Verkuyl, Prof. A. Visser, Prof. M. Moortgat

**BA in General Linguistics**, Utrecht University, 1995

**BA in Cognitive Artificial Intelligence**, Utrecht University, 1993

**BA in Chemistry**, Utrecht University, 1991

## Research Skills:

- Extensive Programming experience in Perl, Java, PHP, C, Actionscript
- Some programming experience in PROLOG, Java, C++, Pascal

## Languages:

- Fluent in English and Dutch
- Near fluent in German, French, and Portuguese
- Can read some Italian and Russian

## Professional Experience:

**Post Doc**, 2004-2006

Instituto de Linguística Teórica e Computacional

Topics: Morphological databases, Neologism

**Post Doc**, 2003

Equipe de Recherche de Syntax et Semantique, Université de Toulouse-leMirail

Topic: Automatic knowledge extraction from corpora using lexicosyntactic patterns

**Junior Teacher**, 1997-2002

Institute of Cognitive Artificial Intelligence, Faculty of Philosophy, Utrecht University

- Courses:
- Introduction to Cognitive Science
  - Introduction to the Philosophy of Language
  - Nonmonotonic Reasoning
  - Foundations of Knowledge Representation
  - Capita Selecta Knowledge Representation

**Teaching Assistant**, 1996-1998

Institute of Cognitive Artificial Intelligence, Faculty of Philosophy, Utrecht University

- Courses:
- Introduction to the Philosophy of Language

- Montague Grammar

**Linguistic Programmer**, 1996, van Dale Lexicografie, Utrecht  
 Perl programmer for preparation of linguistic data  
**System Operator**, 1995, VARA broadcasting company  
 Macintosh/UNIX support group

### Papers published and in print:

*Terminologie Traductive et Representation des Connaissances: l'usage des relations hyponymiques*. Maarten Janssen and Marc van Campenhoudt. To appear in *Langage*.  
*Multilingual Lexical Databases, Lexical Gaps, and SIMuLLDA*. To appear in *International Journal of Lexicography*.  
*Some Words on Word*. Maarten Janssen and Albert Visser. Artificial Intelligence Preprint Series, vol. 30. Utrecht University, April 2002. To appear in *La Nuova Critica*.  
*The Codification of Usage*. Henk Verkuyl, Maarten Janssen, and Frank Jansen. In: P. van Sterkenburg (ed.) *A Practical Guide on Lexicography*. Amsterdam: John Benjamins. 2003.

### Papers presented at conferences:

*Between Inflection and Derivation: Paradigmatic Lexical Functions in Morphological Databases*. Em: East West Encounter: second international conference on Meaning - Text Theory. Moscow, Russia. August 2005.  
*Lexical vs. Dictionary Databases: design choices of the MorDebe system*". In: Papers in Computational Lexicography - COMPLEX 2005. Budapest, Hungary. June 2006.  
*Open Source Lexical Information Network*. Third International Workshop on Generative Approaches to the Lexicon. Geneva, Switzerland.  
*Lexical Translation and Conceptual Hierarchies*. 5<sup>th</sup> International Tbilisi Symposium on Logic, Language and Computation, Tbilisi, Georgia, October 2003.  
*Differentiae Specificae in EuroWordNet and SIMuLLDA*. ONTOLEX workshop of the LREC 2002 conference, Las Palmas, Spain, May 2002.  
*Bilingual Dictionaries and Lexical Gaps*. Computational Lexicography Conference (COMPLEX), Birmingham, England, April 2001.  
*Differences between Dictionaries*. Tenth Computational Linguistic in the Netherlands (CLIN) meeting, Utrecht, The Netherlands, December 1999.  
*A Multilingual Lexical Database using a Structured Interlingua*. 3<sup>rd</sup> International Tbilisi Symposium on Language, Logic, and Information. Batumi, Georgia. September 1999.  
*Nominal Anaphora and Taxonomy*. 2<sup>nd</sup> Discourse Anaphora and Reference Resolution Conference (DAARRC), Lancaster, England, July 1998.

### Presentations without proceedings:

*MorDebe – uma base de dados morfológica*. With Carla Viana. Presented at Associação Portuguesa de Linguística 2005, Porto, Portugal, August 2005.  
*SIMuLLDA – une base de donnée multilangue*. Presented at the ERSS seminar, Toulouse, France, February 2003.  
*NeoTrack – semiautomatic neologism detection*. Presented at Associação Portuguesa de Linguística 2005, Porto, Portugal, August 2005.  
*Sticking to Shallow Lexicons*. Presented at *Constructing Lexica* workshop, Leiden, the Netherlands, September 2002.

### Academic software developed:

**JaLaBA** – an online tool for the graphic presentation of Formal Concept Analysis contexts – based on the LatDrawWin tool written by Ralph Freese with specific additions for FCA purposes. Can export graphics to LaTeX source code. Developed in 2003 at the OTS institute in Utrecht, written in Perl and Java.  
 URL: <http://maarten.janssenweb.net/jalaba/JaLaBA.pl>

**YakwaSI** – an online query tool for searching aligned corpora or monolingual corpora, aimed specifically at the

use with lexicosemantic pattern extraction. Developed in 2003 at the ERSS in Toulouse, written in Perl.  
URL: <http://maarten.janssenweb.net/yakwasi/YakwaSI.pl>

**Simpleconc** – a light-weight online query tool for searching corpora, in essence a stripped down version of the YakwaSI tool. Developed in 2005 at ILTEC in Lisbon, written in PHP.  
URL: <http://www.iltec.pt/handler.php?action=concord>

**MorDebe** – an online system for the MorDebe morphological database – consisting of an online consultation tool with many advanced search options, as well as a web-based administration tool for maintaining and updating the database. Developed in 2004-2005 at ILTEC in Lisbon, written in PHP and MySQL.  
URL: <http://www.iltec.pt/mordebe>

**NeoTrack** – an online system semi-automatic detection of neologisms in corpora. The system extracts neologism candidates automatically and presents them one by one for manual verification to the user, together with their context and several reference corpora. Fully integrated with the MorDebe database. Developed in 2004-2005 at ILTEC in Lisbon, written in PHP and MySQL.  
URL: <http://www.iltec.pt/neotrack>

**Spock** – an online system for the consultation of spoken corpora. Presents a orthography-based query environment for the corpus, and presents the results together with their original sound data. Developed in 2006 at ILTEC in Lisbon, written in PHP.  
URL: <http://www.iltec.pt/spock>