

The Grammix CD-ROM
A Software Collection for Developing Typed Feature
Structure Grammars

Stefan Müller
Freie Universität Berlin

Proceedings of the GEAF 2007 Workshop

Tracy Holloway King and Emily M. Bender (Editors)

CSLI Studies in Computational Linguistics ONLINE

Ann Copestake (Series Editor)

2007

CSLI Publications

<http://csli-publications.stanford.edu/>

Abstract

This paper describes a bootable CD-ROM that contains grammar development software for teaching and research.

1 Purpose

The CD-ROM contains the TRALE system and various other software components that are useful for grammar development (see Section 3). It can be used for research projects as well as for teaching syntax or grammar engineering courses. It contains example grammars for German that correspond to chapters in a textbook (Müller, 2007). The textbook motivates the complex feature geometry that is used in current HPSG publications. The book starts with basic head argument structures and head features and extends this feature geometry to more complex structures including a semantic representation, and features for nonlocal dependencies and relative clauses. Students who work with this textbook can use the CD to look at structures corresponding to their respective knowledge in detail. They can modify and extend the toy grammars and study the consequences of their changes. While the set of reference grammars in the HPSG framework is growing larger and larger, large scale grammars are not usable for grammar teaching for various reasons. For instance such grammars often use auxiliary features that are needed for technical reasons and that are not documented in the literature. Feature descriptions of large scale grammars have an enormous complexity which makes them unusable for teaching. The Grammix CD-ROM tries to fill the gap and provide a set of grammars with increasing complexity which allows to focus on certain features when they are introduced.

Apart from the textbook grammars the CD contains grammar fragments for German, Chinese, and Maltese. These grammars use a common core and can be seen as reference grammars for multi-lingual grammar engineering with the TRALE system.

In addition, the CD contains the textbook by Frank Richter on *Grammar Formalisms and Parsing* and the software that comes with it.

The CD-ROM is intended to be a reference installation of TRALE and the attached software components. The system is Unicode-enabled and displays grammar files and grammar output that use Simplified Chinese, ISO-Latin-1, and Maltese characters correctly. As a bootable, stand-alone CD including an operating system, this reference CD provides not only the software mentioned but also all appropriate fonts, configurations, and other collateral files.

The CD-ROM also contains an installation of the LKB system. Readers who have other textbooks, grammars, and/or software that they want to distribute together with TRALE or LKB, should contact the author of this paper.

2 Description

The Grammix CD-ROM is a bootable Knoppix-based CD-ROM that contains two complete grammar development systems (the TRALE system (Meurers, Penn and Richter, 2002; Penn, 2004) and the LKB system (Copestake, 2002)) together with the grammar profiling tool [incr tsdb()] and example grammars written for the TRALE system, which correspond to the respective chapters in the text book *Einführung in die Head-Driven Phrase Structure Grammar* (Müller, 2007). In addition, the CD-ROM contains the Babel System (Müller, 1996, 1999, 2004) and TRALE grammars for Chinese, Maltese, and German which have a common core and use Minimal Recursion Semantics (Copestake, Flickinger, Pollard and Sag, 2005) for semantic representations.

Users may use the software contained on the CD to:

- load grammars, display analyzes as trees or as feature structure with arbitrary level of detail
- display type hierarchies and signatures (types together with the features that are introduced by the respective types)
- change and extend existing grammars (and of course write new ones ...)
- to watch the Bottom-Up-Chart-Parser doing its work, visualize grammar rules, and explore newly developed grammars systematically.
- do systematic grammar testing and profiling with the profiling tool [incr tsdb()].

3 Grammar development software

The Grammix CD-ROM contains the following software (in addition to some of the software that usually comes with Knoppix):

- Main components:
 - TRALE
Developers and Project Coordinators: Gerald Penn (University of Toronto), Detmar Meurers (The Ohio State University), Frank Richter (Universität Tübingen), with help by Nick Pendar (University of Toronto), Thilo Götz (Universität Tübingen), Stephan Kepser (Universität Tübingen), Dale Gerdemann (Universität Tübingen), Frederik Fouvry (Universität Tübingen), Vanessa Metcalf (The Ohio State University), Markus Dickinson (The Ohio State University), Holger Wunsch (Universität Tübingen), Martin Lazarov (Universität Tübingen), Oliver Suhre (Universität Tübingen), Mike Daniels (The Ohio State University), and Stefan Müller (Freie Universität Berlin)

- Chart-Display
Author: Patric Stiffel (DFKI Saarbrücken)
TRALE- and Babel-Interface: Stefan Müller (Freie Universität Berlin)
- [incr tsdb()] distributed with the LKB
Author: Stephan Oepen (University of Oslo)
TRALE-Interface: Stefan Müller (Freie Universität Berlin) and Fred-
erik Fouvry (Universität des Saarlandes, Saarbrücken)
- utool: The Swiss Army Knife of Underspecification
Authors: Alexander Koller, Stefan Thater, and Michaela Regneri, with
help by Marco Kuhlmann
TRALE-Interface: Stefan Müller (Freie Universität Berlin)
- Grammars for TRALE
Author: Stefan Müller (Freie Universität Berlin)
- Babel with grammar
Author: Stefan Müller (Freie Universität Berlin)
- Other material:
 - MoMo
Authors: Katja Ovchinnikova (Universität Osnabrück), Frank Richter
(Universität Tübingen), Beata Trawiński (Universität Tübingen), Ash-
ley Brown and Levente Barczy
 - Textbook Grammar Formalisms and Parsing
Author: Frank Richter (Universität Tübingen)
 - Grammars for Grammar Formalisms and Parsing
Author: Frank Richter (Universität Tübingen), Manfred Sailer (Uni-
versität Göttingen), and Beata Trawiński (Universität Tübingen)

4 System Background and system Requirements

The CD-ROM is based on Knoppix (currently version 5.0.1). Knoppix is a Linux distribution that is based on Debian. Knoppix uses a compressed file system which makes it possible to store more than 2 Gbyte of software on a CD-ROM. Recent versions of Knoppix use the Union file system which makes it possible to “change” configuration and application files although they are stored on the non-writable CD-ROM. The changed files are stored in the RAM. The operating system looks at the modified files in the RAM rather than at the earlier versions of these files on the CD-ROM. The modifications in the RAM will be lost when the machine is switched off or rebooted. However, users can store such information permanently on the hard disc or on an USB stick. This makes it possible to modify example grammars that are delivered with Grammix and store these grammars permanently.

The graphical desktop that is delivered with Grammix is KDE. The system requirements correspond to the system requirements imposed by Knoppix:

- Intel-compatible CPU (i486 or later, including Intel Macs with firmware 1.0.1 or later),
- at least 128 MB,
- bootable CD-ROM drive,
- standard SVGA-compatible graphics card,
- serial or PS/2 standard mouse or IMPS/2-compatible USB-mouse.

If you want to work with more complex grammars (starting from Chapter 10 of the textbook), you need 256 MB memory, 512 MB is recommended. If you work with `[incr tsdb()]` to debug TRALE grammars, another TRALE process is started and you need the corresponding amount more memory. Since the LKB and `[incr TSDB()]` are more tightly integrated, this does not apply to the profiling of LKB grammars.

Since Knoppix uses open source drivers for hardware components, Grammix may not run on very new hardware. For instance the graphics chip may not be recognized which may result in failure to display anything or in a suboptimal display with low resolution. In such cases using virtualization software may help (see Section 8).

5 Performance Issues

Gammix contains a standalone version of TRALE, which consists of saved states. Saved states do not contain the SICStus Prolog compiler. Instead the TRALE code is interpreted, which slows down the system considerably. Users who want to use TRALE professionally should consider buying a SICStus license and installing a SICStus version on top of Grammix.

Note also that Grammix is based on a 32bit operating system. SICStus Prolog runs much faster (factor two in certain situations) on modern CPUs with 64bit architecture.

6 Localization

Since the CD-ROM is a supplement of a textbook in German, the default language for menus and desktop information is German. If you prefer English, you may boot the system and enter the language code as an option at the boot prompt and press return:

```
knoppix lang=en
```

This will select the respective keyboard driver and provide menus and icon names in KDE according to the language you selected. Since the localization files are very

big not all languages are included in the CD-ROM. If your language is missing, please send an email to the author. Please refer to the Grammix web page to find out which languages are supported.

7 Download

You need a good internet connection for the download, since the size of the CD image is approximately 500 MB.

The CD image is available at:

<http://hpsg.fu-berlin.de/Software/Grammix/>.

8 Grammix and Other Operating Systems

The Grammix CD contains an operating system, so no other operating system is required to run the software. However, users may find it more convenient to run their normal operating system while using Grammix. Thanks to some auxiliary files provided by Doug Arnold it is possible to use VM Player¹ on Windows to run Grammix in a virtual machine. It is then possible to switch between your normal Windows environment and Grammix. Grammix can use the network facilities used by the host operating system without any further configuration. Depending on personal preferences, users may run Grammix from the CD-ROM or from a copy of the ISO image on their hard disk. The necessary files for setting up VM PLayer can be downloaded on the Grammix page, which was given in the previous section. VM Player is provided free of charge. Mac users with Intel Macs can use VM Ware Fusion².

References

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Lecture Notes, No. 110, Stanford: CSLI Publications.

Copestake, Ann, Flickinger, Daniel P., Pollard, Carl J. and Sag, Ivan A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 4(3), 281–332. <http://lingo.stanford.edu/sag/papers/copestake.pdf>, 11.10.2006.

Flickinger, Dan, Koller, Alexander and Thater, Stafan. 2005. A New Well-Formedness Criterion for Semantics Debugging. In Stefan Müller (ed.), *The Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, pages 129–142,

¹<http://www.vmware.com/products/player/>

²<http://www.vmware.com/products/fusion/>

- Stanford: CSLI Publications. <http://cslipublications.stanford.edu/HPSG/6/>, 05.13.07.
- Koller, Alexander and Thater, Stefan. 2005. Efficient Solving and Exploration of Scope Ambiguities. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 9–12, Ann Arbor: Association for Computational Linguistics. <http://acl.ldc.upenn.edu/P/P05/P05-3003.pdf>, 04.09.2007.
- Meurers, Walt Detmar, Penn, Gerald and Richter, Frank. 2002. A Web-Based Instructional Platform for Constraint-Based Grammar Formalisms and Parsing. In Dragomir Radev and Chris Brew (eds.), *Effective Tools and Methodologies for Teaching NLP and CL*, pages 18–25, proceedings of the Workshop held at 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA. <http://www.ling.ohio-state.edu/~dm/papers/acl02.html>, 08.01.2004.
- Müller, Stefan. 1996. The Babel-System—An HPSG Prolog Implementation. In *Proceedings of the Fourth International Conference on the Practical Application of Prolog*, pages 263–277, London. <http://hpsg.fu-berlin.de/~stefan/Pub/babel.html>, 08.12.2007.
- Müller, Stefan. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Linguistische Arbeiten, No. 394, Tübingen: Max Niemeyer Verlag. <http://hpsg.fu-berlin.de/~stefan/Pub/hpsg.html>, 08.12.2007.
- Müller, Stefan. 2004. Continuous or Discontinuous Constituents? A Comparison between Syntactic Analyses for Constituent Order and Their Processing Systems. *Research on Language and Computation, Special Issue on Linguistic Theory and Grammar Implementation* 2(2). <http://hpsg.fu-berlin.de/~stefan/Pub/discont.html>, 08.12.2007.
- Müller, Stefan. 2007. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Stauffenburg Einführungen, No. 17, Tübingen: Stauffenburg Verlag. <http://hpsg.fu-berlin.de/~stefan/Pub/hpsg-lehrbuch.html>, 08.12.2007.
- Oepen, Stephan and Callmeier, Ulrich. 2000. Measure for Measure: Parser Cross-Fertilization. Towards Increased Component Comparability and Exchange. In *Proceedings of the 6th International Workshop on Parsing Technologies*, pages 183–194, Trento, Italy. <http://www.delph-in.net/itsdb/publications/fertilization.ps.gz>, 04.09.2007.
- Oepen, Stephan and Carroll, John A. 2000a. Ambiguity Packing in Constraint-Based Parsing—Practical Results. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*

- (NAACL'00), Seattle, WA, pages 162–169. <http://www.delph-in.net/itsdb/publications/packing.ps.gz>, 04.09.2007.
- Oepen, Stephan and Carroll, John A. 2000b. Parser Engineering and Performance Profiling. *Natural Language Engineering* 6(1), 81–97. <http://www.delph-in.net/itsdb/publications/parsing.ps.gz>, 04.09.2007.
- Oepen, Stephan and Flickinger, Daniel P. 1998. Towards Systematic Grammar Profiling. Test Suite Technology Ten Years After. *Journal of Computer Speech and Language* 12(4), 411–436, (Special Issue on Evaluation). <http://www.delph-in.net/itsdb/publications/profiling.ps.gz>, 04.09.2007.
- Ovchinnikova, Ekaterina and Richter, Frank. 2007. Morph Moulder: Teaching Software for HPSG and Description Logics. *Logic Journal of the IGPL*. <http://jigpal.oxfordjournals.org/cgi/content/abstract/jzm024v1>, 04.09.2007.
- Penn, Gerald. 2004. Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 239–246, Barcelona, Spain.
- Penn, Gerald and Carpenter, Bob. 1999. ALE for Speech: a Translation Prototype. In *Proceedings of the 6th Conference on Speech Communication and Technology (EUROSPEECH)*, Budapest, Hungary.
- Richter, Frank. 2006. A Web-based Course in Grammar Formalisms and Parsing, electronic Textbook. <http://milca.sfs.uni-tuebingen.de/A4/Course/PDF/gramandpars.pdf>, 20.09.2007.