

Multi-lingual Noun Phrase Extractor (MuNPEX)

- [GATE Components](#)
- [MuNPEX](#)
- [Multi-Lingual](#)
- [Noun Phrase Chunking](#)

toc_collapse=0; Table of Contents

- [1. Overview](#)
- [2. Installation](#)
- [3. Version history](#)
- [4. License](#)
- [5. Feedback](#)
- [6. Acknowledgements](#)

1. Overview

The *Multi-Lingual Noun Phrase Extractor* (MuNPEX) is a noun phrase (NP) chunker component developed for the [GATE architecture](#), implemented in JAPE. It is fast, robust, customizable, well-tested and currently supports English, German, and French (with Spanish in beta). It provides detailed features for each NP annotation, with DET (determiner), MOD/MOD2 (pre/post-head modifiers), and HEAD noun slots, as well as (optional) text offset information.

MuNPEX requires a part-of-speech (POS) tagger to work and can additionally use detected named entities (NEs) to improve chunking performance. For English, MuNPEX works with the ANNIE (Hepple) tagger that comes as part of the ANNIE system with GATE. French, German, and Spanish are based on the TreeTagger. Please read the included documentation (or source code) for more details. Also note that you will need at least GATE version 5.x to run MuNPEX, as it makes use of the new Kleene operators for JAPE.

German NE+NPE run in 34.959 seconds

MuNPEX in GATE Developer: Screenshot of MuNPEX running on a German text

2. Installation

If you have GATE 7 or better, you can install MuNPEX directly from within GATE using the CREOLE Plugin Manager by selecting our Semantic Software Lab repository. The install package includes documentation, the main and supplementary JAPE grammars, the example pipeline, and some additional information. You can also download the [install package](#) manually (but the recommended way of installation is to use the GATE plugin manager through the GATE Developer GUI).

You can also check out the latest development version of MuNPEX from [our public GitHub repository](#). A continuous integration build of this repository is available on our [Jenkins server](#).

3. Version history

- 2.0: 26.07.2015. "Pronoun" feature added to NP annotations. Smoke tests for demo pipeline added. License updated from GPLv2 to LGPLv3.
- 1.2: 24.04.2012. Minor bugfix to English chunker
- 1.1: 12.02.2012. Repackaged for new GATE 7 Plugin Manager. Added example pipeline for English. Added wrappers to load MuNPEX grammars as PRs.
- 1.0: 16.08.2010. More robust on malformed input. Optional grammars for adding a HEAD_LEMMA slot. DET/MOD/HEAD/MOD2 now stored as strings instead of Content objects. Code cleanup and tweaks.
- 0.2: 03.03.2006. Preliminary Spanish support added. Renamed from "NPE" to "MuNPEX". Small cleanups. Number

transducer added.

- 0.1: 21.11.2005. Initial public release.

4. License

MuNPEX is distributed as free/open source software under the [GNU Lesser General Public License Version 3 \(LGPL3\)](#).

5. Feedback

For questions, comments, etc., please use the [Forum](#).

6. Acknowledgements

Thanks to [Michelle Khalifé](#) for helping out in developing the French version.



Except where otherwise noted, all original content on this site is copyright by its author and licensed under a [Creative Commons Attribution-Share Alike 2.5 Canada License](#).

Source URL (retrieved on 2019-06-18 00:26): <http://www.semanticsoftware.info/munpex>