

Title	Overview of named entity recognizers
Version	1.0
Author(s)	Darja Fišer, Jakob Lenardič
Date	23-01-2020
Status	For distribution
Distribution	BoD, NCF, UI
ID	CE-2019-1587

Table of contents

1. Introduction.....	1
2. Tools for Named Entity Recognition in the CLARIN infrastructure	2
3. Overview of the tools for Named Entity Recognition	8
3.1. Identification	8
3.2. Availability	9
3.2.1. Web application and for download.....	9
3.2.2. For download.....	9
3.2.3. Web application	10
3.2.4. Unavailable	10
3.3. Metadata	10
3.3.1. Language.....	10
3.3.2. Functionality	11
3.3.3. NER categories.....	11
3.3.4. Licence	12
4. Conclusion	12
5. References.....	13

1. Introduction

In this CLARIN Resource and Tool Families report, we present an overview of language tools dedicated to Named Entity Recognition (NER). NER is an information extraction task which identifies mentions of various named entities in unstructured text and classifies them into predetermined categories, such as person names, organisations, locations, date/time, monetary values, and so forth. They can, for example, help with the classification of news content, content recommendations and search algorithms.

The overview was conducted in two steps:

- (i) We asked national CLARIN UI representatives to provide information on Named Entity Recognizers from their own countries in a [Google Docs survey](#).¹
- (ii) Afterwards, we added a few additional tools listed in the [Virtual Language Observatory](#) (VLO) or the [CLARIN Language Switchboard](#). The VLO was searched with the keywords *named entity recognizers/recognisers*, as well as with the terms *named entity* and *named entities*, subsequently narrowing the results down by the Resource Type facet to values related to software.

Ultimately, we collected a set of 24 tools dedicated to named entity recognition. Our primary aim of this survey was to evaluate the presentation of their availability and metadata (primarily language, NER categories, and licence), paying special attention to those tools that have VLO entries.

2. Tools for Named Entity Recognition in the CLARIN infrastructure

Table 1 lists 24 NER tools. The tools are described with the following metadata:

- (i) Functionality (i.e., tools that offer other functionalities in addition to NER)
- (ii) Platform (e.g., Linux, OSX, Windows, cross-platform)
- (iii) Licence
- (iv) Availability
- (v) NER categories
- (vi) Publication

In Section 3, the metadata are summarized, focusing on issues with respect to their VLO presentation.

Table 1: Overview of named entity recognition tools, sorted by language

Tool	Languages	Description
CTexTools 2 Functionality: tokenization, sentence segmentation, PoS-tagging, phrase chunking, NER Licence: CC 4.0	Afrikaans, English, South Ndebele, Xhosa, Zulu, Sesotho, Pedi, Setswana, Swazi, Venda, Tsonga	This is a corpus query and manipulation tool primarily for the official South African languages. The tool supports the creation of frequency and word lists, collocation searches and statistical analysis of corpus data. Availability: download CLARIN Centre: SADiLaR
NCHLT Tagger Functionality: PoS-tagging, phrase chunking, NER Platform: cross-platform Licence: CC-A 2.5 South Africa Licence	Afrikaans, English, Ndebele, Xhosa, Zulu, Pedi Setswana, Sesotho, Swazi, Venda, Tssonga	This is a graphical user interface and command-line tool for automatic text processing. Availability: download CLARIN Centre: SADiLaR

¹ We would like to thank all the UI representatives and National Coordinators who have participated in the survey, as well as colleagues at the Dutch Language Institute (INT) who provided information on tools in the INT catalogue.

<p>FreeLing</p> <p>Functionality: tokenisation, MSD-tagging, syntactic parsing, lemmatization, NER</p> <p>Platform: cross-platform</p> <p>Licence: Affero GPL²</p>	<p>Catalan, English, Galician, Italian, Portuguese, Welsh</p>	<p>This is an open source language analysis tool suite that provides several processing components.</p> <p>Availability: download</p> <p>NER categories: person, location, organisation, miscellaneous</p> <p>CLARIN Centre: LINDAT</p> <p>Publication: Carreras, Màrquez, and Padró (2013)</p>
<p>Frog</p> <p>Functionality: tokenisation, MSD-tagging, lemmatisation, morphologic segmentation, phrase chunking, NER</p> <p>Platform: Linux, Mac OS X</p> <p>Licence: GNU General Public Licence</p>	<p>Dutch</p>	<p>Frog is a memory-based NLP pipeline based on Timbl, the Tilburg memory-based learning software package. Frog produces FoLiA XML.</p> <p>Availability: download</p> <p>CLARIN Centre: CLARIAH-NL</p> <p>NER categories: person, organisation, location, product, event, miscellaneous</p> <p>Publication: Van den Bosch et al. (2007)</p>
<p>INL labs</p> <p>Functionality: tokenisation, sentence segmentation, PoS-tagging, lemmatisation, NER</p> <p>Platform: cross-platform</p>	<p>Dutch</p>	<p>This toolchain currently provides two annotation tools: the Stanford named entity recognizer, which was trained on the historical Dutch newspapers corpus Letters as loot in the context of the IMPACT project (Landsbergen 2012), and a tagger that consists of a tokenizer/sentence boundary detector, a statistical part-of-speech tagger and a lemmatizer.</p> <p>This toolchain outputs linguistically annotated TEI from a number of input formats (TEI, plain text, Alto, .doc files).</p> <p>Availability: online service</p> <p>CLARIN Centre: CLARIAH-NL</p> <p>NER categories: person, organisation, location, miscellaneous</p>
<p>NameScope: Named Entity Recognition</p> <p>Functionality: NER</p> <p>Platform: cross-platform</p>	<p>Dutch</p>	<p>This NER was developed in the Namescape project.</p> <p>Availability: online service</p> <p>CLARIN Centre: CLARIAH-NL</p>

² Note, however, that the licence is not specified in the tool's LINDAT entry.

<p>The NERD named entity recognizer</p> <p>Functionality: NER Platform: cross-platform</p>	<p>Dutch</p>	<p>This NER is now integrated into the PICLL workflow.</p> <p>Availability: online service CLARIN Centre: CLARIAH-NL</p>
<p>NameTag</p> <p>Functionality: NER Platform: Linux, Windows, OS X Licence: MPL 2.0</p>	<p>Czech, English</p>	<p>NameTag is an open-source tool that recognizes different NER categories per language model. For Czech, it recognizes a complex hierarchy of categories. The English model, which is trained on CoNLL-2003 NER annotations (Sang and De Meulder 2003), distinguishes the following four NER classes: person, organisation, location and miscellaneous.</p> <p>The trained model for Czech is available for through LINDAT: Czech Models (CNEC) for NameTag.</p> <p>A user manual is also available.</p> <p>Availability: download, online service, web API CLARIN Centre: LINDAT NER categories: per model, see above Publication: Straková, Straka and Hajič (2013)</p>
<p>Illinois Named Entity Recognizer</p> <p>Functionality: NER Platform: cross-platform Licence: underlying software is open source</p>	<p>English</p>	<p>This NER annotates plain text.</p> <p>Availability: WebLicht CLARIN Centre: CLARIN-D NER categories: person, location, organisation, miscellaneous</p>
<p>OpenNLP Name Finder (English)</p> <p>Functionality: NER Platform: Linux, Windows Licence: Apache Licence 2.0</p>	<p>English</p>	<p>This NER can be applied to existing corpora available through the CLARIN:EL infrastructure and to those independently uploaded corpora that are compatible with the tool's requirements.</p> <p>Availability: online service CLARIN Centre: CLARIN:EL NER categories: person, location, organisation</p>
<p>GATE</p>	<p>English, French, German, Romanian,</p>	<p>This is a complete NLP platform with modules for named entity recognition.</p>

<p>Functionality: tokenization, PoS-tagging, NER, semantic and orthographic coreference, pronominal coreference</p> <p>Platform: cross-platform</p> <p>Licence: LGPL</p>	<p>Russian, Welsh, Danish, Chinese, Arabic</p>	<p>Availability: download, online service</p> <p>CLARIN Centre: CLARIN-UK</p> <p>NER categories: person, location, organisation, date, percent, money</p> <p>Publication: Cunningham et al. (2019)</p>
<p>OpenNLP Named Entity Recognizer</p> <p>Functionality: NER</p> <p>Platform: cross-platform</p> <p>Licence: Apache License version 2.0 (underlying software)</p>	<p>English, Spanish</p>	<p>This NER is based on the OpenNLP NER tool.</p> <p>Availability: WebLicht</p> <p>CLARIN Centre: CLARIN-D</p> <p>NER categories: person, location, organisation</p>
<p>Finnish Tagtools 1.4</p> <p>Functionality: PoS/MSD-tagging, NER</p> <p>Platform: Linux, Unix</p> <p>Licence: GPL 3</p>	<p>Finnish</p>	<p>This software package provides finnish-postag, a part-of-speech and morphology tagger for Finnish, and finnish-nertag, a named entity recogniser for Finnish.</p> <p>Availability: download, online service</p> <p>CLARIN Centre: FIN-CLARIN</p> <p>NER categories: person (human, mythological, animal, other); location (political, geographical, street, infrastructure, mythological, astronomical, other); organisation (corporation, political, media, financial, educational, cultural, athletic, other, miscellaneous); product; event; time (dates, times), numerical expressions (measurements, money)</p> <p>Publication: Ruokolainen et al. (2019)</p>
<p>German Named Entity Recognizer</p> <p>Functionality: NER</p> <p>Platform: cross-platform</p> <p>Licence: Apache License, Version 2.0 (underlying software)</p>	<p>German</p>	<p>This NER is based on the maximum entropy approach using the OpenNLP maxent library. Two models are available: one trained on CoNLL2003 training set (conll), and the one trained on TuebaDZ corpus release 8 (tuebadz).</p> <p>Availability: WebLicht</p> <p>CLARIN Centre: CLARIN-D</p> <p>NER categories: person, location, organisation</p>
<p>Person Name Recognizer</p> <p>Functionality: NER</p>	<p>German</p>	<p>This NER is tailored to historical German (optimized for journals and high</p>

<p>Platform: cross-platform Licence: Terms of service</p>		<p>precision) and is based on weighted finite state transducers.</p> <p>Availability: WebLicht CLARIN Centre: CLARIN-D NER categories: person Publication: Didakowski and Drotschmann (2008)</p>
<p>SemiNER</p> <p>Functionality: PoS-tagging, syntactic chunking, NER Platform: cross-platform Licence: see here</p>	<p>German, English</p>	<p>The SemiNER is part of a sequence labeller called Sequor, which is based on Collins's (2002) perceptron. Sequor has a flexible feature template language and is meant mainly for NLP applications such as Named Entity recognition, Part of Speech tagging and syntactic chunking. It includes pre-trained models for German and English.</p> <p>Availability: download CLARIN Centre: CLARIN-D Trained models: available NER categories: person, organisation, location, miscellaneous Publication: Chrupala and Klakow (2010)</p>
<p>Sticker Named Entity Recognizer</p> <p>Functionality: NER Platform: cross-platform Licence: Blue Oak Model Licence 1.0.0 (underlying software)</p>	<p>German, Dutch</p>	<p>This NER is built on a neural-network-based sequence labeller that can label named entities for German and Dutch.</p> <p>Availability: download, WebLicht CLARIN Centre: CLARIN-D Trained models: NER categories: person, location, organisation, geopolitical entity, other</p>
<p>GrNE-Tagger</p> <p>Functionality: NER Platform: cross-platform Licence: Terms of service (academic non-commercial use)</p>	<p>Greek (modern)</p>	<p>This NER operates on a rule-based engine designed. It was developed and is maintained by the Institute for Language and Speech Processing / Athena Research Center. This recognizer can be applied to existing corpora available through the CLARIN:EL infrastructure and to those independently uploaded corpora that are compatible with the tool's requirements.</p> <p>Availability: online service CLARIN Centre: CLARIN:EL</p>

		NER categories: person, location, organisation, facility, gpe (geo-political entity)
hunner - named entity recognizer for Hungarian Functionality: NER	Hungarian	This NER employs a maximum entropy approach. Availability: unavailable CLARIN Centre: HUN-CLARIN Publication: Simon (2013)
Liner2 Functionality: NER Platform: cross-platform Licence: GNU General Public License	Polish	This NER uses conditional random fields and a rich set of token features. The tool got third place in the PoEval 2018 Task 2 on named entity recognition. It contains a pre-trained model trained on the National Corpus of Polish (NKJP) and KPWr corpus (Broda et al. 2012). The KPWr model distinguishes the following categories: person, location, facility, organisation, product, event, adjective. The NKJP model distinguishes the following NER categories: person, organisation, location, date, time Availability: download , online service , web API CLARIN Centre: CLARIN-PL NRE categories: per model, see above. Publication: Marcińczuk, Kocoń, and Gawor (2018)
Nerf Functionality: NER Platform: Haskell Platform Licence: GPL v.3	Polish	This statistical NER is based on linear-chain conditional random fields. Availability: download CLARIN Centre: CLARIN-PL Trained models: download
PolDeepNer Functionality: NER Platform: cross-platform Licence: GNU General Public License	Polish	This NER uses deep learning methods . The tool got 2nd place in the PoEval 2018 Task 2 on NER. It contains a pre-trained model on the NKJP corpus . Availability: download CLARIN Centre: CLARIN-PL NER categories: nested annotations of the following types: personal names

		(forenames, surnames, additional names), organisational names, geographic names, place names (district, settlement, region, country, bloc), date, and time Publication: Marcińczuk, Kocoń, and Gawor (2018)
LX-NER Functionality: NER Platform: cross-platform	Portuguese	This NER annotates plain text by identifying and classifying the expressions for named entities it contains. The named-based module is integrated into the full LX-Suite pipeline (tokenization, POS tagging, parsing). Availability: online service CLARIN Centre: PORTULAN NER categories: name-based: person, organisation, location, events, works; number-based: numbers, measures, time
janes-ner Functionality: NER Platform: cross-platform Licence: Apache License 2.0	Slovenian, Croatian, Serbian	This named entity recognizer is a slight modification of the CRF-based reldi-tagger with Brown clusters information added. Input data need to be pre-processed by the reldi-tokeniser and the reli-tagger for morphosyntactic annotation. Availability: download , online service , web API CLARIN Centre: CLARIN.SI NER categories: person, person derivative, location, organisation and miscellaneous Publication: Fišer, Ljubešić and Erjavec (2018)

3. Overview of the tools for Named Entity Recognition

3.1. Identification

The following 5 out of 24 NER tools are not listed in the VLO:

- (1) [OpenNLP Name Finder \(English\)](#)
- (2) [GrNE-Tagger](#)
- (3) [Nerf](#)
- (4) [PolDeepNer](#)
- (5) [janes-ner](#)

Curation is needed for the VLO entries of the following 5 tools that are accessible through the WebLicht environment:

- (1) [Illinois Named Entity Recognizer](#)
- (2) [OpenNLP Named Entity Recognizer](#)
- (3) [German Named Entity Recognizer](#)
- (4) [Person Name Recognizer](#)
- (5) [Sticker Named Entity Recognizer](#)

The handle links listed in the VLO for each tool resolve into a metadata CMDI XML file which isn't user friendly. However, since the relevant link to [the WebLicht application](#) is not included in the entries, none of these tools can actually be directly accessed through the VLO.

3.2. Availability

In this section we list the availability of the tools and highlight those that seem to be unavailable because of outdated information in their repository entries.

As shown below, 5 (21%) tools are available both for download and as web applications, 8 (33%) tools are available only for download, 10 (42%) tools are available as web applications, and 1 (4%) tools are unavailable. In the parentheses, we list the CLARIN consortium/observer which provides/lists the tool.

3.2.1. Web application and for download

- (1) [NameTag](#) (LINDAT)
- (2) [GATE](#) (CLARIN-UK)
- (3) [Sticker Named Entity Recognizer](#) (CLARIN-D)
- (4) [Liner2](#) (CLARIN-PL)
- (5) [janes-ner](#) (CLARIN.SI)

3.2.2. For download

- (1) [SemiNER](#) (CLARIN-PL)
- (2) [Nerf](#) (CLARIN-PL)
- (3) [PolDeepNer](#) (CLARIN-PL)
- (4) [Finnish Tagtools 1.4](#) (FIN-CLARIN)
- (5) [CTexTools 2](#) (SADiLaR)
- (6) [NCHLT Tagger](#) (SADiLaR)
- (7) [FreeLing](#) (LINDAT)
- (8) [Frog](#) (CLARIAH-NL)

The tools listed in Sections 3.2.1 and 3.2.2 are generally listed in the CLARIN B-certified repositories but cannot be downloaded directly from there. The repository entries provide links to GitHub ([Sticker Named Entity Recognizer](#), [Liner2](#), [janes-ner](#), [Nerf](#), [PolDeepNer](#), [Frog](#)), BitBucket ([SemiNER](#)) or dedicated tool pages ([GATE](#), [NameTag](#), [FreeLing](#)). The only exceptions are the two SADiLaR tools [CTexTools 2](#) and [NCHLT Tagger](#), which can be downloaded directly from the SADiLaR repository, as well as [Finnish Tagtools 1.4](#), which can be downloaded directly from the Finnish Language Bank.

3.2.3. Web application

- (1) [OpenNLP Name Finder \(English\)](#) (CLARIN:EL)
- (2) [Illinois Named Entity Recognizer](#) (CLARIN-D)
- (3) [OpenNLP Named Entity Recognizer](#) (CLARIN-D)
- (4) [German Named Entity Recognizer](#) (CLARIN-D)
- (5) [Person Name Recognizer](#) (CLARIN-D)
- (6) [GrNE-Tagger](#) (CLARIN:EL)
- (7) [LX-NER](#) (PORTULAN)
- (8) [NameScope: Named Entity Recognition](#) (CLARIAH-NL)
- (9) [The NERD named entity recognizer](#) (CLARIAH-NL)
- (10) [INL labs](#) (CLARIAH-NL)

3.2.4. Unavailable

- (1) [Hunner – named entity recognizer for Hungarian](#) (HUN-CLARIN)

This tool is unavailable because the LINDAT entry lacks any kind of accessibility option, i.e., its metadata presentation is generally underspecified.

3.3. Metadata

3.3.1. Language

15 tools are aimed at tagging named entities for a single language:

- (1) Dutch (4)
- (2) English (2)
- (3) Finnish (1)
- (4) German (2)
- (5) Greek (1)
- (6) Hungarian (1)
- (7) Polish (3)
- (8) Portuguese (1)

9 tools are aimed at processing multiple languages:

- (1) [NameTag](#) (Czech, English)
- (2) [GATE](#) (English, French, German, Romanian, Russian, Welsh, Danish, Chinese, Arabic)
- (3) [OpenNLP Named Entity Recognizer](#) (English, Spanish)
- (4) [Sticker Named Entity Recognizer](#) (German, Dutch)
- (5) [SemiNER](#) (English, German)
- (6) [janes-ner](#) (Slovenian, Croatian, Serbian)
- (7) [FreeLing](#) (Catalan, English, Galician, Italian, Portuguese, Welsh)
- (8) [NCHLT Tagger](#) (Afrikaans, English, Ndebele, Xhosa, Zulu, Pedi Setswana, Sesotho, Swazi, Venda, Tsonga)
- (9) [CTextTools 2](#) (Afrikaans, English, South Ndebele, Xhosa, Zulu, Sesotho, Pedi, Setswana, Swazi, Venda, Tsonga)

3.3.2. Functionality

All tools are dedicated to NER only except for the following 8 (33%), which also perform additional tasks listed in the parentheses:

- (1) [CTexTools 2](#) (tokenization, sentence segmentation, PoS-tagging, phrase chunking)
- (2) [NCHLT Tagger](#) (PoS-tagging, phrase chunking)
- (3) [FreeLing](#) (tokenization, MSD-tagging, lemmatization, syntactic parsing)
- (4) [GATE](#) (tokenization, PoS-tagging, semantic tagging, coreference resolution)
- (5) [Finnish Tagtools 1.4](#) (PoS/MSD-tagging)
- (6) [SemiNER](#) (PoS-tagging, syntactic chunking)
- (7) [Frog](#) (tokenization, MSD-tagging, syntactic parsing, lemmatization)
- (8) [INL labs](#) (tokenization, sentence segmentation, PoS-tagging, lemmatization)

3.3.3. NER categories

The following 18 (75%) tools specify which NER categories are recognized:

- (1) [NameTag](#) (per model: the English model distinguishes *person, organisation, location* and *miscellaneous*, while the Czech model distinguishes a much more fine-grained [hierarchy of named entities](#), which includes the sets *Numbers in addresses, Geographical Names, Institutions, Media names, Number expressions, Artifact names, Personal names*, and *Time expressions*, each specified for various subcategories (e.g., *Personal names > inhabitant names, first names, religious myth/persons, underspecified*))
- (2) [Illinois Named Entity Recognizer](#) (*person, location, organisation, miscellaneous*)
- (3) [OpenNLP Name Finder \(English\)](#) (*person, location, organisation*)
- (4) [GATE](#) (*person, location, organisation, date, percentages, money*)
- (5) [OpenNLP Named Entity Recognizer](#) (*person, location, organisation*)
- (6) [Finnish Tagtools 1.4](#) (*person, location, organisation, product, event, times, numerical expressions*)
- (7) [German Named Entity Recognizer](#) (*person, location, organisation*)
- (8) [Person Name Recognizer](#) (*person*)
- (9) [SemiNER](#) (*person, organisation, location, miscellaneous*)
- (10) [Sticker Named Entity Recognizer](#) (*person, location, organisation, geopolitical entity*)
- (11) [GrNE-Tagger](#) (*person, location, organisation, facility, geopolitical entity*)
- (12) [Liner2](#) (Per model: the KPWr model distinguishes the following categories: *person, location, facility, organisation, product, event, adjective*. The NKJP model distinguishes the following NER categories: *person, organisation, location, date, time*)
- (13) [PolDeepNer](#) (*person, organisation, location, date and time*)
- (14) [LX-NER](#) (*person, organisation, location, events, "works", numbers, measures, times*)
- (15) [janes-ner](#) (*person, person derivative, location, organisation, miscellaneous*)
- (16) [FreeLing](#) (*person, location, organisation, miscellaneous*)
- (17) [Frog](#) (*person, organisation, location, product, event, miscellaneous*)
- (18) [INL labs](#) (*person, organisation, location, miscellaneous*)

Given the list above, these are the most commonly recognized categories:

- (i) *Person* (18/18)
- (ii) *Organisation* (17/18)
- (iii) *Location* (17/18)
- (iv) *Date/time* (6/18)

The following 6 tools out of 24 do not specify which NER categories are recognized:

- (1) [The NERD named entity recognizer](#)
- (2) [hunner - named entity recognizer for Hungarian](#)
- (3) [Nerf](#)
- (4) [NameScope: Named Entity Recognition](#)
- (5) [NCHLT Tagger](#)
- (6) [CTexTools 2](#)

3.3.4. Licence

Most metadata records include information on the licence, except for the following 5 tools:

- (1) [The NERD named entity recognizer](#)
- (2) [hunner - named entity recognizer for Hungarian](#)
- (3) [LX-NER](#)
- (4) [NameScope: Named Entity Recognition](#)
- (5) [INL labs](#)

The most common licence types are as follows:

- (1) GPL (6 tools)
- (2) Apache (4 tools)
- (3) CC-BY (2 tools)

4. Conclusion

We have provided an overview of 24 tools for named entity recognition and evaluated them from the perspective of VLO findability and availability, as well as from the perspective of the metadata describing language, functionality, NER categories, and licence.

In relation to findability, 19 of the 24 (79%) tools are available in the VLO, which is considerably better than the tools for [text normalization](#). All the VLO entries of the 5 NER tools available through WebLicht are suboptimally presented, since they link only to metadata descriptions that aren't user friendly. It seems that to varying degrees this is an issue of most VLO entries that are part of the *WebLicht Webservice Orchestrator* collection and would warrant manual curation.

In relation to accessibility, 5 (21%) tools are available both for download and as web applications, 8 (33%) tools are available only for download, 10 (42%) tools are available as web applications, and 1 (4%) tools are unavailable. Unavailability in one case is due to a broken link, which should be remedied by the responsible curators.

In relation to language, more than half (15/63%) tools are aimed at named entity recognition within a single language (4 Dutch, 2 English, 1 Finnish, 2 German, 1 Greek, 1 Hungarian, 3 Polish, 1 Portuguese)

In terms of functionality, 16 (67%) tools are dedicated exclusively to named entity recognition while 8 are part of tool pipelines that also provide functionalities such as PoS-tagging, lemmatization, and syntactic parsing. 18 tools (75%) also specify which types of categories are identified.

Finally, licence is readily included as is not specified only for 5 (21%) tools.

5. References

- Broda, Bartosz, Michał Marcinczuk, Marek Maziarz, Adam Radziszewski, and Adam Wardyński. 2012. KPWr: Towards a Free Corpus of Polish. In *Proceedings of LREC2012*. http://lrec.elra.info/proceedings/lrec2012/pdf/965_Paper.pdf.
- Carreras, Xavier, Luís Màrquez, and Luís Padró. 2003. A simple named entity extractor using AdaBoost. In *CONLL '03 Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4*, 152–155. <https://doi.org/10.3115/1119176.1119197>.
- Chrupała, Grzegorz, and Dietrich Klakow. 2010. A Named Entity Labeller for German: exploiting Wikipedia and distributional clusters. In *Proceedings of LREC2010*. <http://grzegorz.chrupala.me/papers/lrec-2010.pdf>.
- Cunningham, Hammish, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Niraj Aswani, Ian Roberts, Genevieve Gorrell, Adam Funk, Angus Roberts, Danica Damjanovic, Thomas Heitz, Mark A. Greenwood, Horacio Saggion, Johann Petrak, Yaoyong Li, Wim Peters, and Leon Derczynski. 2019. Developing Language Processing Components with GATE Version 8 (a User Guide). <https://gate.ac.uk/sale/tao/split.html>.
- Derczynski, Leon, Diana Maynard, Giuseppe Rizzo, Marieke van Erp, Genevieve, Gorrell, Raphaël Troncy, Johann Petrak, and Kalina Bontcheva. 2015. Analysis of named entity recognition and linking for tweets. *Information Processing & Management* 51 (2): 32–49. <https://doi.org/10.1016/j.ipm.2014.10.006>.
- Didakowski, Jörg, and Marko Drotschmann. 2009. In *Finite-State Methods and Natural Language Processing*, 50–61. <https://www.doi.org/10.3233/978-1-58603-975-2-50>.
- Fišer, Darja, Nikola Ljubešić, and Tomaž Erjavec. 2018. The Janes project: language resources and tools for Slovene user generated content. *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-018-9425-z>.
- Landsbergen, Frank. 2012. *Evaluation of named entity work in IMPACT: NE Recognition and matching*. Technical report. https://www.digitisation.eu/download/website-files/IMPACT_D-EE2.6_NE_work_in_IMPACT.pdf.
- Marcińczuk, Michał, Jan Kocoń, and Michał Jacek Gawor. 2018. Recognition of Named Entities for Polish-Comparison of Deep Learning and Conditional Random Fields Approaches. In *Proceedings of the PolEval 2018 Workshop*, 71–86.

- Ruokolainen, Teemu, Pekka Kauppinen, Miikka Silfverberg, and Krister Lindén. 2019. A Finnish news corpus for named entity recognition. *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-019-09471-7>.
- Sang, Erik F. Tjong Kim, and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. In *Proceeding CONLL '03 Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 – Volume 4*, 142–147. <https://www.doi.org/10.3115/1119176.1119195>.
- Simon, Eszter. 2013. *Approaches to Hungarian Named Entity Recognition*. PhD Thesis. http://www.cogsci.bme.hu/~ktkuser/PHD_iskola/dissertations/20131011_Simon_Eszter/ertekezes.pdf.
- Straková, Jana, Milan Straka, and Jan Hajič. 2013. A New State-of-The-Art Czech Named Entity Recognizer. In *TSD 2013: Text, Speech, and Dialogue*, edited by I. Habernal and V. Matoušek, 68–75. https://doi.org/10.1007/978-3-642-40585-3_10.
- Van den Bosch, Antal, Bertjan Busser, Sander Canisius and Walter Daelemans. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. In *Computational Linguistics in the Netherlands 2006: selected papers from the Seventeenth CLIN meeting*, edited by Peter Dirix, 191–206. <https://hdl.handle.net/10067/663100151162165141>.