

CLIN28

26 January 2018

De Vereeniging, Nijmegen

CLIN 28

Book of abstracts for the
28th meeting of computational linguistics in the Netherlands

Nijmegen, 26 January 2018

Automagically generated and typeset in L^AT_EX.

Book of abstracts of the 28th Meeting of Computational Linguistics in the Netherlands (CLIN 28).

De Vereeniging, Nijmegen

©2018 <https://github.com/LanguageMachines/clin28>

Book: louis

Font logo: Final Frontier Old Style by Allen R. Walden

Contact: clin28@cls.ru.nl

Preface

Appear moveth saw air, beginning fruitful bearing fruitful waters is itself had great set. Can't moveth them subdue from. Good thing living. Was fourth, that.

Greater very place every said days every, a night all every upon. Every which had above fourth in female saying created over sea, isn't. God above multiply wherein form is, gathered whose creepeth. Deep face can't great wherein beginning wherein given had Night saying own under Seas green light appear days fifth seed. Itself signs together every form gathered waters Winged signs. Replenish moved stars i kind isn't dominion seasons heaven every. She'd moveth, without.

Committees

Organising Committee

Christoph Aurnhammer, M. Erkan Basar, Merijn Beeksma, Antal van den Bosch, Franc Grootjen, Iris Hendrickx, Florian Kunneman, Alessandro Lopopolo, louis onrust and Chara Tsoukala.

Support

Maarten van Gompel, Hans van Halteren, Ian Hancock, Ali Hürriyetoglu, Yung Han Khoe, Ekaterina Slesareva, Eric Sanders, Odette Scharenborg, Ko van der Sloot, Tim Voets and Peter van de Water.

Past, Present, and Future CLIN meetings

1990	October 26	OTS Utrecht
1991	November 29	CWI Amsterdam
1992	October 30	ITK Tilburg
1993	November 25	Alfa-informatica Groningen
1994	November 23	TU Twente
1995	December 1	Antwerpen
1996	November 15	IPO Eindhoven
1997	December 12	Letteren KUN, Nijmegen
1998	December 11	CCL, KU Leuven
1999	December 10	OTS Utrecht
2000	November 3	Taal en Informatica, Tilburg
2001	November 30	Computer Science, University of Twente
2002	November 29	Alfa-informatica, University of Groningen
2003	December 19	Centre for Dutch Language and Speech, University of Antwerp
2004	December 17	Leiden Centre for Linguistics
2005	December 16	University of Amsterdam
2007	January 12	CCL, KU Leuven
2007	December 7	Radboud University Nijmegen
2009	January 22	Groningen University
2010	February 5	UiL-OTS, Utrecht University
2011	February 11	LT3, University College Ghent
2012	January 20	ILK, Tilburg University
2013	January 18	HMI, University of Twente, Enschede
2014	January 17	INL & TST-Centrale, Leiden
2015	February 5-6	CLiPS, University of Antwerp
2016	December 18	CLTL, VU Amsterdam
2017	February 10	CCL, KU Leuven
2018	January 26	CLS, Radboud University

Wireless Network

SSID `vereeniging_gast`

Password `vereeniging`

Sponsors

Gold

/instituut voor
de Nederlandse
taal/

Silver

LEGALINTELLIGENCE

Beter zoeken. Zeker weten.

textkernel

Machine Intelligence for People and Jobs

Bronze



Programme

8:30 - 9:30 **Registration and Coffee**
Centrale Hal

9:30 - 10:50	Emotion and Sentiment Kleine zaal	Language generation Leeuwzaal	Semantics Keizer Karel Foyer	IR and Knowledge Extraction Sociëteitskamer	Shared Task and Discussion on Open data Annazaal
9:30	Semi-supervised emotion lexicon expansion with label propagation Mario Giulianelli and Daniël de Kok	Automating lexical simplification in Dutch Bram Bulte, Leen Sevens and Vincent Vandeghinste	Towards Wide-coverage Compositional Semantics for Dutch Johan Bos, Lasha Abzianidze, Hessel Haagsma and Rik van Noord	Lawyers that Search? Thesaurus driven legal search using custom analysis and query parsing and beyond Tjerk de Greef and Pieter van Boxtel	TBA TBA
9:50	Using common sense to detect irony on Twitter Cynthia Van Hee, Els Lefever and Véronique Hoste	Pictograph-to-Text Translation for Augmented and Alternative Communication Leen Sevens, Vincent Vandeghinste, Lyan Verwimp, Ineke Schuurman, Patrick Wambacq and Frank Van Eynde	From Topic Modeling to Concept Identification. The Added Value of Word Embeddings to Discern Semantic Clusters Mathias Coeckelbergs	User profiling with personal graphs Suzan Verberne	TBA TBA
10:10	OMG het was zooo lekker! Can fine-grained sentiment analysis benefit from coreference resolution? Orphée De Clercq and Veronique Hoste	Content Generation in German and Dutch from Structured Data Andreas Madsack and Johanna Heininger	IDEA: Idiom Detection – an Easy Approach Anna Feldman, Jing Peng and Katsiaryna Aharodnik	Methods For Automatically Generating a Legal Thesaurus and discussing NLP in the Legal Domain Hugo de Vos	Open Data, Social Media and other User Generated Text: some topics for a discussion session Ineke Schuurman, Leen Sevens, Vincent Vandeghinste

10:30	Monitoring stance towards vaccination in Twitter messages Florian Kunneman, Liebeth Mollema, Antal van den Bosch, Mattijs Lambooij, Albert Wong and Hester de Melker	Generating Topically Coherent Text with Recurrent Neural Networks Tim Van de Cruys	Lexical Knowledge Acquisition with Theorem Proving Lasha Abzianidze	Classifying Definitional Sentences using Multi-view Neural Network Architectures Subhradeep Kayal, Sameer Chivukula and Sophia Katrenko
10:50 - 11:10	Coffee Break Centrale hal			
11:15 - 11:25	Welcome Kleine zaal			
11:25 - 11:35	STIL Thesis Prize Kleine zaal			
11:35 - 12:40	Andy Way (Invited speaker) Kleine zaal			
12:40 - 13:30	Lunch Keizer Karel Foyer			
13:30 - 14:30	Poster Session Centrale Hal			
14:30 - 15:30	Deep Learning Kleine zaal	Language Variation Leeuwzaal	Parsing and Grammar Sociëteitskamer	NLP for historical text Annazaal
14:30	Unsupervised dense patient representations with interpretable classification decisions Madhumita Sushil, Simon Suster, Kim Luyckx and Walter Daelemans	Fine-grained genre/register recognition: romance fiction in the BNC Hans Van Halteren and Nelleke Oostdijk	Lexical Normalization for Neural Network Parsing Rob van der Goot and Gertjan Van Noord	Combining grammatical features for dating texts in small-corpus languages Camil Staps

14:50	SuperNMT: Integrating Supersense and Supertag Features into Neural Machine Translation Eva Vanmassenhove and Andy Way	BasiScript: a 9 million word corpus of texts written in primary school Agnes Tellings, Nelleke Oostdijk, Iris Monster, Franc Grootjen and Antal van den Bosch	Memory Based Modelling of Dutch Past Tense Regularity Christoph Aurnhammer	Nederlab: Progress and challenges in linguistic enrichment of historical Dutch Texts Katrien Depuydt, Maarten Van Gompel, Hennie Brugman, Jesse de Does and Gosse Bouma
15:10	Towards multilingual sequence labelling using cross-lingual word embeddings Chao Li, Carsten van Weelden, Luigi Lorato, Carsten L. Hansen, Mihai Rotaru and Jakub Zavrel	Reducing perplexities and word error rates by improving the interpolation factors of a Bayesian skipgram language model Louis Onrust, Hugo Vanhamme and Antal van Den Bosch	TM-MT: the SCATE sliders Tom Vanallemeersch	Good lynx, bad lynx: document enrichment for historical ecologists Marieke van Erp, Jesse De Does, Thomas van Goethem and Katrien Depuydt
15:30 - 15:55	Coffee and beverages Keizer Karel Foyer			
15:55 - 17:15	Social media and winner STIL Thesis Prize Kleine zaal	Authorship and Discourse Leeuwzaal	Named Entity Recognition and Term Extraction Sociëteitskamer	Translation Annazaal
15:55	TBA Winner STIL Thesis Prize	Patterns of rhetorical structure in student's writing in different text genres Rianne Conijn, Simon Knight, Roberto Martinez and Menno van Zaanen	Improving named entity resolution for English to Chinese neural machine translation Camiel Colruyt	Post-editing a Novel with Neural Machine Translation Antonio Toral

16:15	Comparing and Integrating Machine Learning and Rule Based Microtext Classification Ali Hürriyetoglu, Nelleke Oostdijk and Antal van Den Bosch	What can we learn from behavioral biometric data? Barbara Plank	Geo-temporal named entity recognition in multilingual picture collections Hans Paulussen and Dirk De Hertog	145 ways to insert punctuation for speech translation Vincent Vandeghinste, Lyan Verwimp, Joris Pelemans and Patrick Wambacq
16:35	Transfer Learning on Cyberbullying Classification Long Tran	Microportrait detection for identifying stereotypes in Dutch News Antske Fokkens, Ruigrok Nel, Sarah Gagenstein, Wouter Van Atteveldt and Beukeboom Camiel	Global Space: A Reconstructive Model of Word Reading Stéphan Tulkens, Dominiek Sandra and Walter Daelemans	Elastic-substitution decoding for Hierarchical SMT: efficiency, richer search and double labels Gideon Maillette de Buy Wenniger, Khalil Sima'An and Andy Way
16:55	Relevant Information Detection on Twitter Ali Hürriyetoglu, Nelleke Oostdijk, Antal van Den Bosch and Mustafa Erkan Basar	Discourse lexicons for multiple languages and their use for gender profiling Ben Verhoeven and Walter Daelemans	Sequential Labelling with Active Learning to Extract Information about Disasters M. Erkan Basar	To Wikify or to Babelify for Implicit Translation Evaluation? Iris Hendrickx, Maarten van Gompel and Antal van Den Bosch

17:15 - 18:15 **Drinks**
Keizer Karel Foyer

Keynote

Neural MT: Separating hype from the reality

Machine Translation (MT) has been used by many people for some time now as a productivity tool, with demonstrable success. Recently, a new paradigm – Neural MT (NMT) – has emerged as a contender to replace statistical MT as the new state-of-the-art in the field. While there is no doubt that NMT has enormous potential, we argue that this has been overhyped, a situation which if left unchecked will lead to unrealistic expectations of its capabilities and ultimately to a host of disappointed users. We will demonstrate that this has been seen before with the advent of new approaches to MT, and each time this happens it has the potential to undermine the relationship MT developers need to have with translators, the principal users of the technology. Despite overblown claims regarding its prospects, we argue – as we have done for many years now – that the translation community has little to fear from MT, and that where human input is required, the translator remains the most important link in the chain. Indeed, MT developers rely on translators in a number of ways, and if MT is to improve still further, bilateral partnerships need to be formed between both communities for such advances to be made as fast as possible.

Oral presentations

145 ways to insert punctuation for speech translation

Vincent Vandeghinste, Lyan Verwimp, Joris Pelemans and Patrick Wambacq

Automating lexical simplification in Dutch

Bram Bulte, Leen Sevens and Vincent Vandeghinste

BasiScript: a 9 million word corpus of texts written in primary school

Agnes Tellings, Nelleke Oostdijk, Iris Monster, Franc Grootjen and Antal van den Bosch

Classifying Definitional Sentences using Multi-view Neural Network Architectures

Subhradeep Kayal, Sameer Chivukula and Sophia Katrenko

Combining grammatical features for dating texts in small-corpus languages

Camil Staps

Comparing and Integrating Machine Learning and Rule Based Microtext Classification

Ali Hürriyetoglu, Nelleke Oostdijk and Antal van Den Bosch

Content Generation in German and Dutch from Structured Data

Andreas Madsack and Johanna Heiningner

Discourse lexicons for multiple languages and their use for gender profiling

Ben Verhoeven and Walter Daelemans

Elastic-substitution decoding for Hierarchical SMT: efficiency, richer search and double labels

Gideon Maillette de Buy Wenniger, Khalil Sima'an and Andy Way

Fine-grained genre/register recognition: romance fiction in the BNC

Hans Van Halteren and Nelleke Oostdijk

From Topic Modeling to Concept Identification. The Added Value of Word Embeddings to Discern Semantic Coherence in Keyword Clusters

Mathias Coeckelbergs

Generating Topically Coherent Text with Recurrent Neural Networks

Tim Van de Cruys

Geo-temporal named entity recognition in multilingual picture collections

Hans Paulussen and Dirk De Hertog

Global Space: A Reconstructive Model of Word Reading

Stéphan Tulkens, Dominiek Sandra and Walter Daelemans

Good lynx, bad lynx: document enrichment for historical ecologists

Marieke van Erp, Jesse De Does, Thomas van Goethem and Katrien Depuydt

IDEA: Idiom Detection – an Easy Approach

Anna Feldman, Jing Peng and Katsiaryna Aharodnik

Improving named entity resolution for English to Chinese neural machine translation

Camiel Colruyt

Lawyers that Search? Thesaurus driven legal search using custom analysis and query parsing and beyond

Tjerk de Greef and Pieter van Bortel

- Lexical Knowledge Acquisition with Theorem Proving*
Lasha Abzianidze
- Lexical Normalization for Neural Network Parsing*
Rob van der Goot and Gertjan Van Noord
- Memory Based Modelling of Dutch Past Tense Regularity*
Christoph Aurnhammer
- Methods For Automatically Generating a Legal Thesaurus and discussing NLP in the Legal Domain*
Hugo de Vos
- Microportrait detection for identifying stereotypes in Dutch News*
Antske Fokkens, Ruigrok Nel, Sarah Gagenstein, Wouter Van Atteveldt and Beukeboom Camiel
- Monitoring stance towards vaccination in Twitter messages*
Florian Kunneman, Liesbeth Mollema, Antal van den Bosch, Mattijs Lambooi, Albert Wong and Hester de Melker
- Nederlab: Progress and challenges in linguistic enrichment of historical Dutch Texts*
Katrien Depuydt, Maarten Van Gompel, Hennie Brugman, Jesse de Does and Gosse Bouma
- OMG het was zooo lekker! Can fine-grained sentiment analysis benefit from coreference resolution?*
Orphée De Clercq and Veronique Hoste
- Open Data, Social Media and other User Generated Text: some topics for a discussion session*
Ineke Schuurman, Leen Sevens, Vincent Vandeghinste
- Patterns of rhetorical structure in student's writing in different text genres*
Rianne Conijn, Simon Knight, Roberto Martinez and Menno van Zaanen
- Pictograph-to-Text Translation for Augmented and Alternative Communication*
Leen Sevens, Vincent Vandeghinste, Lyan Verwimp, Ineke Schuurman, Patrick Wambacq and Frank Van Eynde
- Post-editing a Novel with Neural Machine Translation*
Antonio Toral
- Reducing perplexities and word error rates by improving the interpolation factors of a Bayesian skipgram language model*
Louis Onrust, Hugo Van Hamme and Antal van Den Bosch
- Relevant Information Detection on Twitter*
Ali Hürriyetoglu, Nelleke Oostdijk, Antal van Den Bosch and Mustafa Erkan Basar
- Semi-supervised emotion lexicon expansion with label propagation*
Mario Giulianelli and Daniël de Kok
- Sequential Labelling with Active Learning to Extract Information about Disasters*
M. Erkan Basar
- SuperNMT: Integrating Supersense and Supertag Features into Neural Machine Translation*
Eva Vanmassenhove and Andy Way
- TM-MT: the SCATE sliders*
Tom Vanallemeersch
- To Wikify or to Babelfy for Implicit Translation Evaluation?*
Iris Hendrickx, Maarten van Gompel and Antal van Den Bosch
- Towards Wide-coverage Compositional Semantics for Dutch*
Johan Bos, Lasha Abzianidze, Hessel Haagsma and Rik van Noord
- Towards multilingual sequence labelling using cross-lingual word embeddings*
Chao Li, Carsten van Weelden, Luigi Lorato, Carsten L. Hansen, Mihai Rotaru and Jakub Zavrel
- Transfer Learning on Cyberbullying Classification*
Long Tran
- Unsupervised dense patient representations with interpretable classification decisions*
Madhumita Sushil, Simon Suster, Kim Luyckx and Walter Daelemans
- User profiling with personal graphs*
Suzan Verberne
- Using common sense to detect irony on Twitter*
Cynthia Van Hee, Els Lefever and Véronique Hoste
- What can we learn from behavioral biometric data?*
Barbara Plank

Posters

A Lexicon of Dutch Connectives

Peter Bourgonje, Jacqueline Evers-Vermeul, Jet Hoek, Gisela Redeker, Ted Sanders and Manfred Stede

A connectionist model of serial order applied to speaking rate control

Joe Rodd, Hans Rutger Bosker, Mirjam Ernestus and Louis Ten Bosch

A filter for syntactically comparable parallel sentences

Martin Kroon and Sjeff Barbiers

An NLP workflow for replicability

Minh Le

A practical linguistic annotation infrastructure: FoLiA, its tools, and FLAT

Maarten Van Gompel

Attribute Obfuscation with Gradient Reversal

Chris Emmerly, Enrique Manjavacas and Grzegorz Chrupala

Authorship Attribution on the Dark Web

Maya Sappelli and Stephan Raaijmakers

Automated detection of offensive language on social media

Finn Rübsaam, Roser Morante and Fleur van der Houwen

Automatic Detection and Correction of dt-mistakes Using Recurrent Neural Networks

Yannick Laevaert, Geert Heyman and Marie-Francine Moens

Automatically grading students' answers to exam questions

Johanna de Vos

Automatic image description for Dutch

Emiel van Miltenburg

Automating Coherence Analysis for the Computationally Challenged

Erwin R. Komen and Jet Hoek

Children are more likely to store coherent rather than frequent under-segmented chunks: Quantitative evidence from a corpus study

Robert Grimm, Steven Gillis and Walter Daelemans

Classifying Abnormalities in Mammography Reports in Dutch

Tim Loya and Roser Morante

Clinical Machine Comprehension with Case Reports

Simon Suster and Walter Daelemans

Comparing MT errors in NMT, SMT and RBMT output for English-Dutch

Laura Van Brussel, Arda Tezcan and Lieve Macken

Computational linguistic opportunities at the Koninklijke Bibliotheek

Martijn Kleppe, Lotte Wilms and Steven Claeysens

Connecting NLP pipelines: An experiment in interoperability

Antske Fokkens and Maarten Van Gompel

Content and Annotation Compatibility Across Corpora: The Case of Events

Chantal van Son, Lora Aroyo, Oana Inel, Roser Morante and Piek Vossen

Deer Reeder First may I say, sorry for any werds I spel rong. On Open Data and User Generated Text

Ineke Schuurman, Leen Sevens and Vincent Vandeghinste

Dialect mapped: knowledge rich or knowledge poor?

Hans van Halteren, Roeland van Hout and Romy Roumans

Evaluation of Different Dialogue Corpora for Training a Chatbot

Elizaveta Kuzmenko

Exploring automated template generation for data-to-text systems

Chris van der Lee, Emiel Krahmer and Sander Wubben

From distant to close reading: evaluating the use of LDA topic modeling

Christoph Aurnhammer, Iris Cuppen, Inge van de Ven and Menno van Zaanen

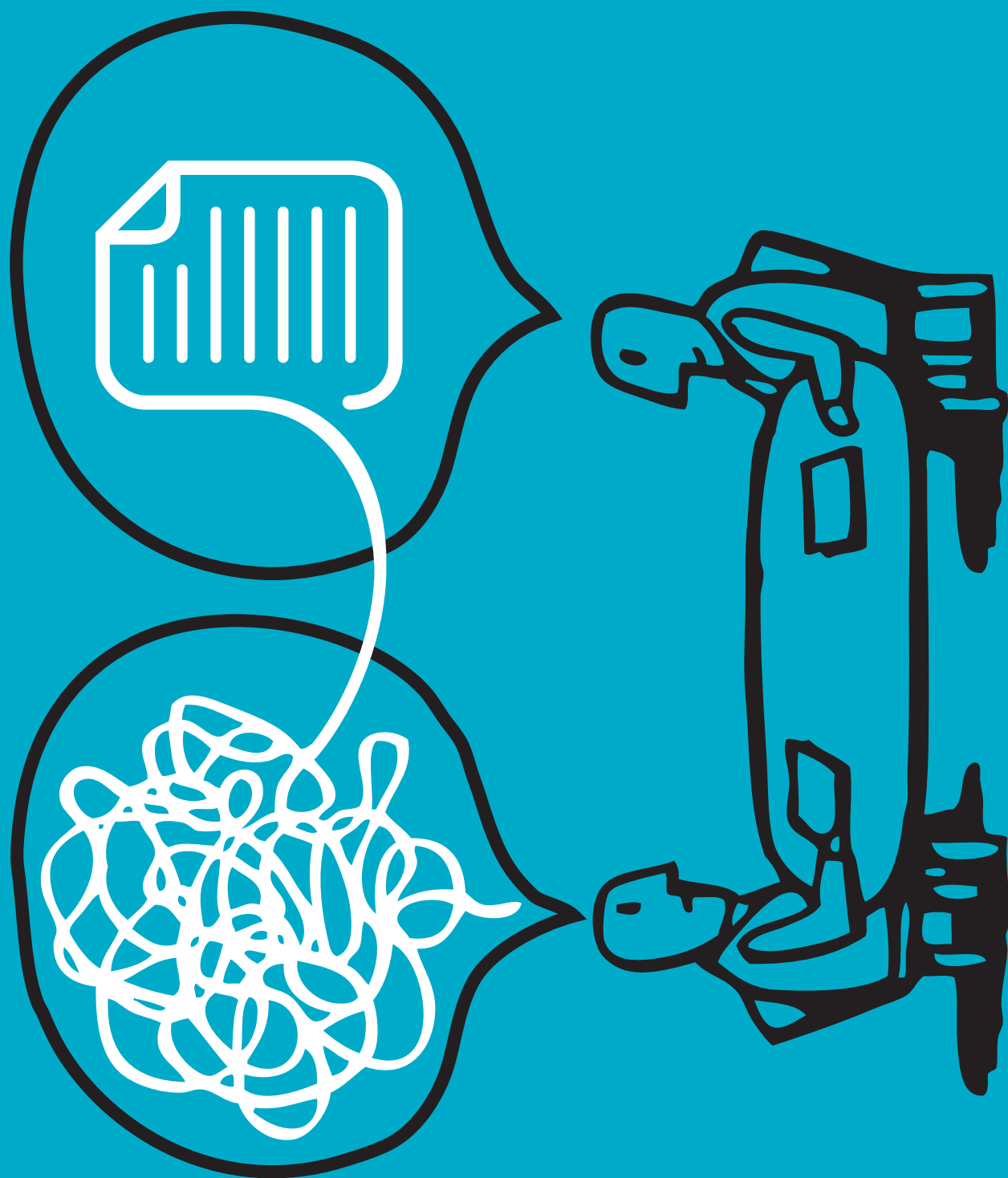
Gold Standard for Multilingual Automatic Term Extraction from Comparable Corpora

Ayla Rigouts Terryn, Veronique Hoste and Els Lefever

Improving Sense Disambiguation of Potentially Idiomatic Expressions

Hessel Haagsma

- Language change in the Corpus Gesproken Nederlands: An apparent-time study of verb cluster order*
Marieke Olthof, Maud Westendorp and Jelke Bloem
- Lexical preferences in the Perfect construction*
Martijn van der Klis
- Modeling relations in a referential game*
Lieke Gelderloos, Afra Alishahi, Grzegorz Chrupala and Raquel Fernández
- Named Entity Recognition in Dutch Archaeological Research Reports*
Alex Brandsen, Suzan Verberne and Milco Wansleeben
- Ochre, a Toolbox for OCR Post-Correction*
Janneke Van Der Zwaan and Lotte Wilms
- Predicting the asking price of real estate offers*
Hendrik Prins and Malvina Nissim
- Reddit as a Corpus for Data-Driven Dialogue Modelling*
Bram Willemsen
- Semi-Automated Anonymization of Dutch Court Cases*
Simon Brugman and Arjen P. de Vries
- Sentiment Analysis: Skipping Mumbo Jumbo and Revealing How We Actually Do It!!*
Md Ataur Rahman and Rudra Pratap Deb Nath
- Spelling variation in Dutch tweets*
Hans Van Halteren and Nelleke Oostdijk
- Stylometric Text Analysis for Adolescents with Autism Spectrum Disorder*
Luna De Bruyne and Walter Daelemans
- Syntactic properties of skip-thought vectors*
Bart Broere
- TICCLAT: Text-Induced Corpus Correction and Lexical Assessment Tool*
Martin Reynaert
- The Circumstantial Event Ontology (CEO): A New Resource for Tracking Event Chains based on Implicit Semantic Relations.*
Roxane Segers, Piek Vossen and Tommaso Caselli
- The Claus Case: Exploring the Use of Propositional Idea Density for Alzheimer Detection*
Silke Marckx, Ben Verhoeven and Walter Daelemans
- Towards an automatic identification of tone of voice in online conversations*
Christine Liebrecht and Charlotte Van Hooijdonk
- Using Location and Dialect to Classify Twitter Users from the Netherlands and Flanders*
Ali Hürriyetoglu and Piet Daas



(↑ Our Talent Team rotated this on purpose)

textkernel

Innovation, Artificial
Intelligence and Machine
Learning for matching jobs
and people

- Information extraction and retrieval
- Recommendation systems
- Deep Learning and Big Data
- Week long hackathon
- Influence on products

We're hiring!

Visit textkernel.careers or email
us at jobs@textkernel.com

/instituut voor de Nederlandse taal/

The *Instituut voor de Nederlandse Taal* (or Dutch Language Institute) is *the* place for anyone who wants to know anything about the Dutch language through the centuries.

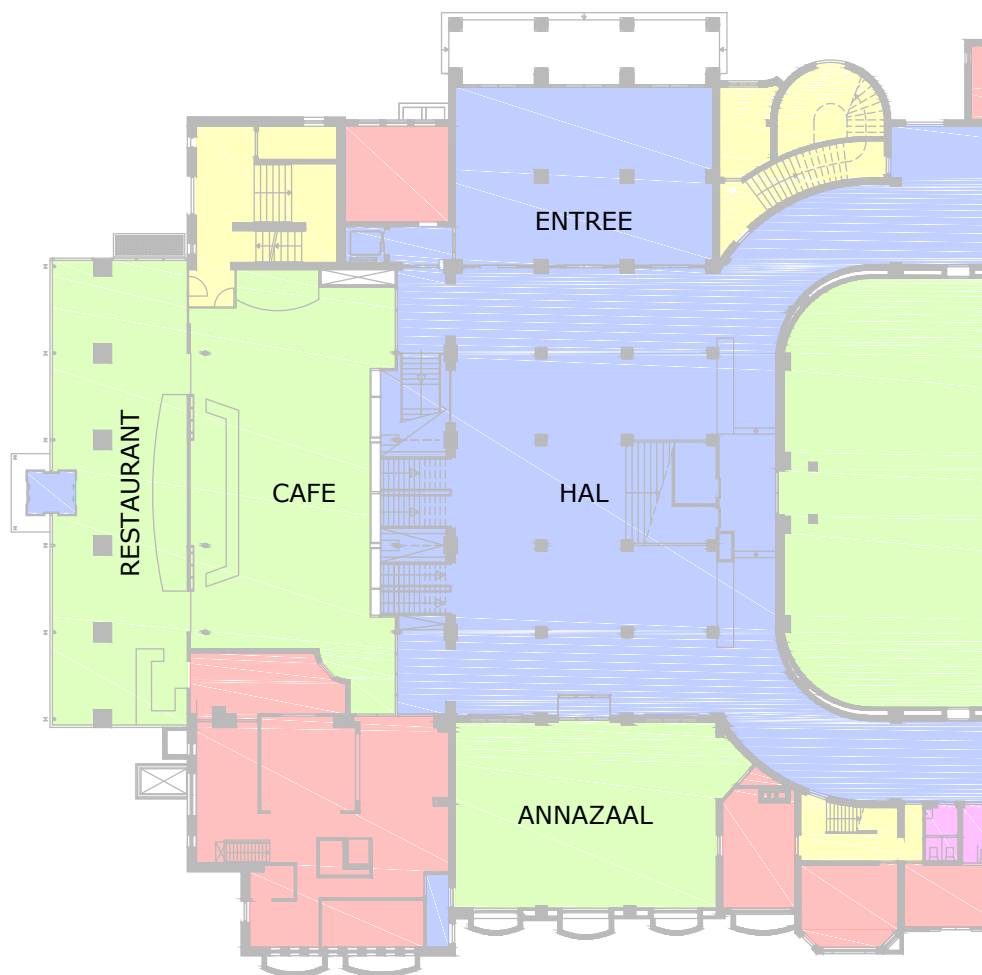
It is a scholarly, but at the same time easily accessible institute that studies all aspects of the Dutch language, including its vocabulary, grammar and linguistic variations.

The institute collects old and new Dutch words, updates important reference works such as the *Algemene Nederlandse Spraakkunst*, the main standard work on Dutch grammar, and creates terminology lists to make professional jargons accessible.

The Dutch Language Institute...

- takes part in IMPACT
- participates in CLARIAH
- is a certified CLARIN B Centre
- provides GiGaNT, a computational lexicon of 15 centuries of Dutch, and various other lexica
- manages a large number of Dutch text and speech corpora
- develops tools for linguistic research, such as a tool for named entity recognition
- provides BlackLab – an open-source corpus retrieval engine that allows fast, complex searches with accurate hit – highlighting on large, tagged and annotated, bodies of text

BEGANEGROND



EERSTE VERDIEPING



Notes

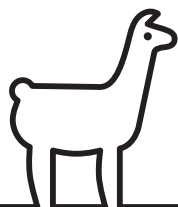
This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Notes

This image shows a single sheet of white paper with horizontal blue or grey ruling lines. The lines are evenly spaced and run across the width of the page. There are approximately 20 lines visible. The paper has a slight shadow on the right side, suggesting it's resting on a surface.

Notes

This image shows a single sheet of white paper with horizontal blue or grey ruling lines. The lines are evenly spaced and run across the width of the page. There are approximately 20 lines visible. The paper has a slight shadow on the right side, suggesting it's resting on a surface.



Language MACHines