

# 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<sup>A</sup>T<sub>E</sub>X

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

De Vereeniging, Nijmegen

©2018 github link

Book: louis onrust

Cover: Erkan & louis

Contact: `clin28@cls.ru.nl`

# 1 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.

## **1.1 Committees**

### **1.1.1 Organising Committee**

### **1.1.2 Support**

## **1.2 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

## 2 Sponsors

### 2.1 Gold

/instituut voor  
de Nederlandse  
taal/

### 2.2 Silver

**LEGAL**INTELLIGENCE  
Beter zoeken. Zeker weten.

**textkernel**  
Machine Intelligence for People and Jobs

### 2.3 Bronze



## Programme

8:30 - 9:30

**Registration and Coffee**  
Centrale Hal

Session1 9:30 - 10:50

**Emotion and Sentiment**

Kleine zaal

**Semi-supervised emotion lexicon expansion with label propagation**

Mario Giulianelli and Daniël de Kok

**Using common sense to detect irony on Twitter**

Cynthia Van Hee, Els Lefever and Véronique Hoste

**OMG het was zooo lekker! Can fine-grained sentiment analysis benefit from coreference resolution?**

Orphée De Clercq and Veronique Hoste

**Language generation**

Leeuwzaal

**Automating lexical simplification in Dutch**

Bram Bulte, Leen Sevens and Vincent Vandeghinste

**Pictograph-to-Text Translation for Augmented and Alternative Communication**

Leen Sevens, Vincent Vandeghinste, Lyan Verwimp, Ineke Schuurman, Patrick Wambacq and Frank Van Eynde

**Content Generation in German and Dutch from Structured Data**

Andreas Madsack and Johanna Heininger

**Semantics**

Keizer Karel Foyer

**Towards Wide-coverage Compositional Semantics for Dutch**

Johan Bos, Lasha Abzianidze, Hessel Haagsma and Rik van Noord

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

Mathias Coeckelbergs

**IDEA: Idiom Detection – an Easy Approach**

Anna Feldman, Jing Peng and Katsiaryna Aharodnik

**IR and Knowledge Extraction**

Sociëteitskamer

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

Tjerk de Greef and Pieter van Bortel

**User profiling with personal graphs**

Suzan Verberne

**Methods For Automatically Generating a Legal Thesaurus and discussing NLP in the Legal Domain**

Hugo de Vos

**Shared Task and Discussion on Open data**

Annazaal

**TBA**

TBA

**TBA**

TBA

**Open Data, Social Media and other User Generated Text: some topics for a discussion session**

Ineke Schuurman, Leen Sevens, Vincent Vandeghinste

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



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

**Comparing and Integrating Machine Learning and Rule Based Microtext Classification**

Ali Hürriyetoglu, Nelleke Oostdijk and Antal van Den Bosch

**Transfer Learning on Cyberbullying Classification**

Long Tran

**Relevant Information Detection on Twitter**

Ali Hürriyetoglu, Nelleke Oostdijk, Antal van Den Bosch and Mustafa Erkan Basar

**What can we learn from behavioral biometric data?**

Barbara Plank

**Microportrait detection for identifying stereotypes in Dutch News**

Antske Fokkens, Ruigrok Nel, Sarah Gagenstein, Wouter Van Atteveldt and Beukeboom Camiel

**Discourse lexicons for multiple languages and their use for gender profiling**

Ben Verhoeven and Walter Daelemans

**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

**Sequential Labelling with Active Learning to Extract Information about Disasters**

M. Erkan Basar

**145 ways to insert punctuation for speech translation**

Vincent Vandeghinste, Lyan Verwimp, Joris Pelemans and Patrick Wambacq

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

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

**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

---

## 2.4 Keynote

## 2.5 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 Heininger

*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 zo 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

## 2.6 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 Emmery, 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 Loyen 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

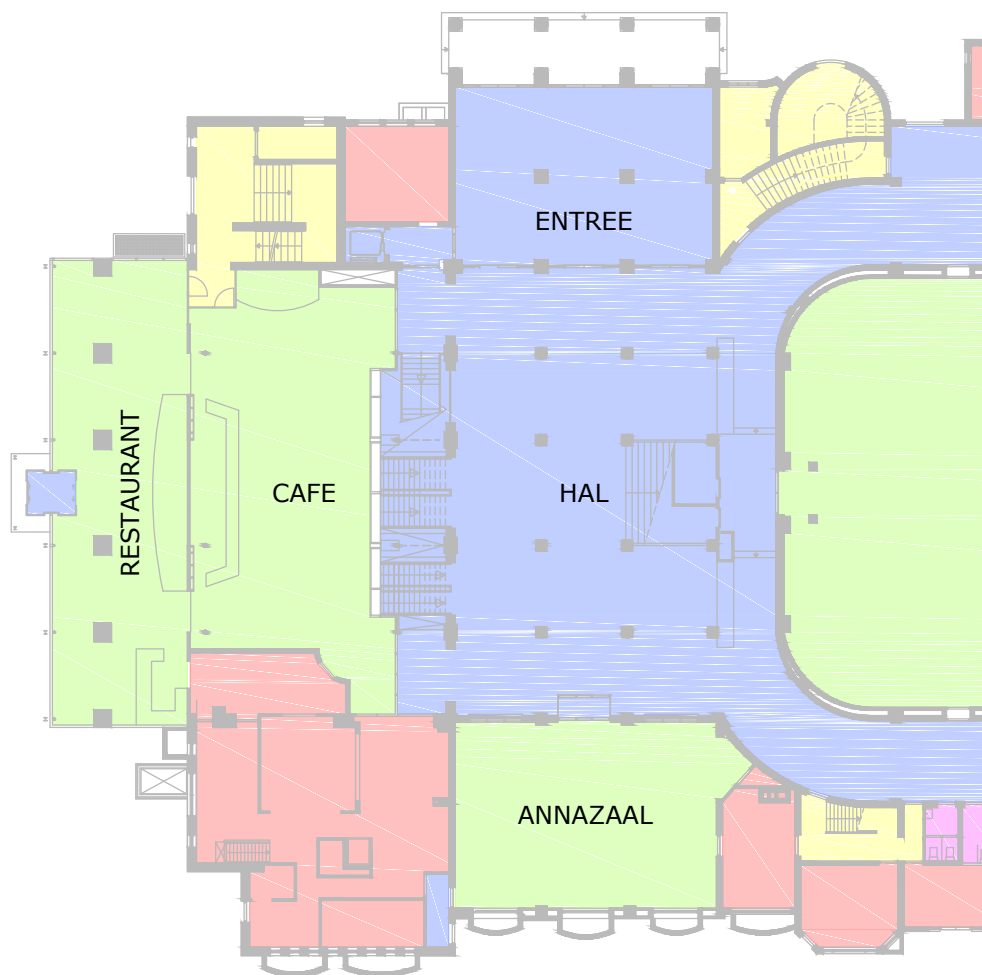
*Towards an automatic identification of tone of voice in online conversations*

Christine Liebrecht and Charlotte Van Hooijdonk

*Using Location and Dialect to Classify Twitter Users from the Netherlands and Flanders*

Ali Hürriyetoglu and Piet Daas

## BEGANEGROND



## EERSTE VERDIEPING

