

# Giovanni Cassani

## Curriculum Vitæ

Tilburg University, CSAI  
Room D337  
Warandelaan 2  
5037 AB Tilburg, The Netherlands  
☎ +32 (0) 471 697005  
✉ [cassani.giovanni@gmail.com](mailto:cassani.giovanni@gmail.com)  
🌐 <https://www.linkedin.com/in/giovincassani/>  
👤 GiovanniCassani

*The road to wisdom? — Well, it's plain and simple to express: Err, and err, and err again; but less, and less, and less. - Piet Hein*

## Professional Experience

2019

**Assistant Professor**, CSAI Department, Tilburg University, Tilburg, The Netherlands.

Assistant Professor in the Department of Cognitive Science and Artificial Intelligence at Tilburg University.

2019

2019

**Data Science Consultant**, Tooso AI, San Francisco, California.

I have been working remotely with software engineers, data scientists, and business experts on intent prediction for online consumers of multi-billion retailers. The project allowed me to contribute end-to-end to a real AI use case – working with real data loads on cloud infrastructure and complying with industry best practices.

## Education

2014

2019

**PhD in Computational Psycholinguistics**, Research Center for Computational Linguistics and Psycholinguistics (CLiPS) - University of Antwerp, Antwerp, Belgium.

SUPERVISORS: Prof. dr. Steven Gillis and Prof. dr. Walter Daelemans

THESIS: The bootstrapping toolbox: which cues are more useful to learn lexical categories and why

FUNDING: Research Council of the University of Antwerp, BOF/TOP grant (ID 29072)

**RESEARCH VISIT**: Seminar für Sprachwissenschaft - Eberhard Karls Universität Tübingen (Germany) under the supervision of Prof. dr. R. Harald Baayen.

2012

2014

**MSc Cognitive Science**, Center for Mind/Brain Sciences (CIMEC) - University of Trento, Trento, Italy.

Major in Language & Multimodal interactions

THESIS: Distributional semantics for child directed speech: A multimodal approach

ADVISORS: Prof. dr. Marco Baroni and dr. Elia Bruni

**ERASMUS PROGRAM**: Faculty of Arts, Department of Linguistics, Rijksuniversiteit Groningen (The Netherlands). Focus on computational semantics and computer-mediated communication.

**INTERNSHIP**: Press Office, University of Trento (5 months). Focus on science communication in an institutional context.

2009  
2012

**BA Historical, Philological, and Literary Studies**, *Faculty of Arts, University of Trento*, Trento, Italy.

Major in Modern Literary Studies

THESIS: Effects of congenital blindness on the semantics of verbs and nouns

ADVISORS: Prof. dr. Marco Baroni and Prof. dr. Alessandro Lenci

## Research Interests

**Language** Psycholinguistics, Language acquisition, Semantics, Morphology, Arbitrariness of the linguistic sign, Multimodality, Concept learning, Cross-situational learning

**Methodology** Computational modeling, Machine learning, Statistics, Behavioral experiments, Brain imaging

**Others** Computational Social Sciences, Categorization, Embodied Cognition, Cognitive Psychology, Cognitive Neuroscience, Dissemination

## Teaching Experience

2021

**University Teaching Qualification (UTQ)/Basis Kwalificatie Onderwijs (BKO)**, Tilburg, The Netherlands.

I obtained the UTQ/BKO certificate, necessary to teach university level courses in the Netherlands.

### Supervision

2020

**BSc in Cognitive Science and Artificial Intelligence**, *Tilburg University*, Tilburg, The Netherlands.

Supervision of bachelor theses on the relation between form and meaning in language acquisition.

2019

**MA in Data Science and Society**, *Tilburg University*, Tilburg, The Netherlands.

Supervision of approx. 30+ master theses per year on a variety of tasks in sequence modeling (clickstream data, EEG data, language).

### Teaching

2022

**Analysis of customer data**, *Tilburg University*, Tilburg, The Netherlands.

Master course on machine learning methods to analyse customer journeys on e-commerce websites, with a focus on intent prediction. Course co-taught with dr. Boris Čule and dr. Gonzalo Napoles.

2020

**Computational Linguistics**, *Tilburg University*, Tilburg, The Netherlands.

3rd year bachelor course on computational linguistics aimed at cognitive science and artificial intelligence students.

2019

**Natural Language Processing**, *Tilburg University*, Tilburg, The Netherlands.

Master course on NLP aimed at data scientists and social scientists.

2019  
2020

**Statistics for Cognitive Science and Artificial Intelligence II**, *Tilburg University*, Tilburg, The Netherlands.

2nd year bachelor course on mixed effect models aimed at cognitive science and AI students. Course co-taught with dr. Travis Wiltshire.

2015  
2016

**Computational Psycholinguistics**, *University of Antwerp*, Antwerp, Belgium.  
Teaching assistant to Prof. dr. Walter Daelemans and Prof. dr. Dominiek Sandra for the master course on Computational Psycholinguistics, aimed to Computational Linguistics students.

## Grants

2022

**SoDa Traineeship**, Utrecht, The Netherlands.

I co-supervise Aron Joosse together with Erik-Jan van Kesteren (Utrecht University) for 7 months in 2023. The project aims to derive various connotations (e.g., polarity, concreteness, and more) for Dutch words, proper names, and made-up words. To do so we leverage distributional semantic models (FastText) and Large Language Models (BERT) as well as sub-lexical regularities pertaining to sound symbolism. This will help predict whether people will associate, for example, a positive connotation to a given brand or person given their name, regardless of whether the name is a made-up word, a real word, or a proper name.

2021

**Research Traineeship, Tilburg School of Humanities and Digital Sciences**, Tilburg, The Netherlands.

Together with dr. Matteo Colombo, I obtained funding to hire one research trainee for 4 months to work on a project titled *"They are racists, but they are not bad."* – *How loaded language combines evaluation and description*. The project combines NLP, psycholinguistics and philosophy of language to study how loaded terms are conceptualised.

2020

**Research Traineeship, Tilburg School of Humanities and Digital Sciences**, Tilburg, The Netherlands.

Together with dr. Sara Van Den Bossche, I obtained funding to hire one research trainee for 10 months to work on a project titled *What's in a fictitious name? Form-meaning correspondences in character names in children's books*. The project (still ongoing) combines NLP, psycholinguistics, and cognitive literary criticism to study how character names are chosen by authors and perceived by readers, with a particular focus on how they convey intuitions about gender, polarity, and age.

## Articles

- [Fra+23] Marolla Francesco, Angelica Maria Maineri, Jacopo Tagliabue, and **Cassani, Giovanni**. "Voting, fast and slow: Ballot order and likability effects in the Five-Star Movement's 2012 online primary election". In: *Contemporary Italian Politics* (2023). DOI: 10.1080/23248823.2023.2175124.
- [CBM21] **Cassani, Giovanni\***, Federico Bianchi\*, and Marco Marelli. "Words with Consistent Diachronic Usage Patterns are Learned Earlier: A Computational Analysis Using Temporally Aligned Word Embeddings". In: *Cognitive Science* 45.4 (2021). ISSN: 0364-0213 1551-6709. DOI: 10.1111/cogs.12963.
- [CL21] **Cassani, Giovanni** and Niklas Limacher. "Not just form, not just meaning: Words with consistent form-meaning mappings are learned earlier". In: *Quarterly Journal of Experimental Psychology* (2021). DOI: 10.1177/17470218211053472.
- [Tag+21] Jacopo Tagliabue, Ciro Greco, Jean-Francis Roy, Bingqing Yu, Patrick John Chia, Federico Bianchi, and **Giovanni Cassani**. "SIGIR 2021 E-Commerce Workshop Data Challenge". In: *CoRR* abs/2104.09423 (2021). arXiv: 2104.09423. URL: <https://arxiv.org/abs/2104.09423>.

- [CCB20] **Cassani, Giovanni**, Yu-Ying Chuang, and R. Harald Baayen. "On the semantics of nonwords and their lexical category". In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46.4 (2020), pp. 621–637. DOI: 10.1037/xlm0000747.
- [Req+20] Borja Requena, **Cassani, Giovanni**, Jacopo Tagliabue, Ciro Greco, and Lucas Lacasa. "Shopper intent prediction from clickstream e-commerce data with minimal browsing information". In: *Scientific Reports* 10.1 (2020), pp. 1–23.
- [Gri+19] Robert Grimm, **Cassani, Giovanni**, Steven Gillis, and Walter Daelemans. "Children Probably Store Short Rather Than Frequent or Predictable Chunks: Quantitative Evidence From a Corpus Study". In: *Frontiers in Psychology* 10 (2019). DOI: doi.org/10.3389/fpsyg.2019.00080.
- [Cas+18] **Cassani, Giovanni**, Robert Grimm, Walter Daelemans, and Steven Gillis. "Lexical category acquisition is facilitated by uncertainty in distributional co-occurrences". In: *PLoS One* 13.12 (2018), e0209449. DOI: 10.1371/journal.pone.0209449.
- [Gri+17b] Robert Grimm, **Cassani, Giovanni**, Steven Gillis, and Walter Daelemans. "Facilitatory Effects of Multi-Word Units in Lexical Processing and Word Learning: A Computational Investigation". In: *Frontiers in Psychology* 8 (2017), p. 555. DOI: 10.3389/fpsyg.2017.00555.

---

## Conference Proceedings

- [Big+19] Luca Bigon, **Cassani, Giovanni**, Ciro Greco, Lucas Lacasa, Mattia Pavoni, Andrea Polonioli, and Jacopo Tagliabue. "Prediction is very hard, especially about conversion. Predicting user purchases from clickstream data in fashion e-commerce". In: *The Fourth International workshop on fashion and KDD*. 2019.
- [Cas+17] **Cassani, Giovanni**, Robert Grimm, Steven Gillis, and Walter Daelemans. "Distributional learning and lexical category acquisition: What makes words easy to categorize?" In: *Proceedings of the 39th annual meeting of the Cognitive Science Society*. Ed. by G. Gunzelmann, A. Howes, T. Tenbrink, and E. Davelaar. Austin, TX: Cognitive Science Society, 2017, pp. 216–221.
- [Gri+17a] Robert Grimm, **Cassani, Giovanni**, Walter Daelemans, and Steven Gillis. "Evidence for a facilitatory effect of multi-word units on child word learning". In: *Proceedings of the 39th annual meeting of the Cognitive Science Society*. Ed. by G. Gunzelmann, A. Howes, T. Tenbrink, and E. Davelaar. Austin, TX: Cognitive Science Society, 2017, pp. 433–438.
- [Cas+16] **Cassani, Giovanni**, Robert Grimm, Steven Gillis, and Walter Daelemans. "Constraining the search space in cross-situational word learning: Different models make different predictions". In: *Proceedings of the 38th annual meeting of the Cognitive Science Society*. Ed. by A. Papafragou, D. Grodner, D. Mirman, and J. C. Trueswell. Austin, TX: Cognitive Science Society, 2016, p. 1152.
- [Cas+15] **Cassani, Giovanni**, Robert Grimm, Walter Daelemans, and Steven Gillis. "Which distributional cues help the most? Unsupervised contexts selection for lexical category acquisition". In: *Proceedings of the 6th workshop on cognitive aspects of computational language learning*. Ed. by R. Berwick, A. Korhonen, A. Lenci, T. Poibeau, and A. Villavicencio. Lisbon, Portugal: Association for Computational Linguistics, 2015, pp. 33–39.
- [Gri+15] Robert Grimm, **Cassani, Giovanni**, Walter Daelemans, and Steven Gillis. "Towards a model of prediction-based syntactic category acquisition: First steps with word embeddings". In: *Proceedings of the 6th workshop on cognitive aspects of computational language learning*. Ed. by R. Berwick, A. Korhonen, A. Lenci, T. Poibeau, and A. Villavicencio. Lisbon, Portugal: Association for Computational Linguistics, 2015, pp. 28–32.

---

## Archival papers

- [Cas+19] **Cassani, Giovanni**, Robert Grimm, Steven Gillis, and Walter Daelemans. "What makes a distributional context useful? Lexical diversity is more important than frequency". 2019.
- [CL16] **Cassani, Giovanni** and Alessandro Lopopolo. "Multimodal Distributional Semantics Models and conceptual representations in sensory deprived subjects". 2016.

---

## Theses

- [Cas19] **Cassani, Giovanni**. "The bootstrapping toolbox: Which cues are more useful to learn lexical categories and why". PhD thesis. University of Antwerp, 2019.
- [Cas14] **Cassani, Giovanni**. "Distributional semantics for child-directed speech: A multimodal approach". MA thesis. Trento, Italy: University of Trento, 2014.

---

## Invited talks

2021

**Dante Symposium, Tilburg University**, Tilburg, The Netherlands.

*Dante's Linguistic Creativity: A Review of Computational Methods to Study Neologisms, Expressive Language and Sound Symbolism.*

2020

**CIMeC, University of Trento**, Rovereto, Italy.

*Rethinking the arbitrariness of the sign.*

2019

**Jheronimus Academy of Data Science**, 's Hertogenbosch, The Netherlands.

*Prediction is very hard, especially about conversion. Predicting user purchases from clickstream data in fashion e-commerce.*

---

## Selected conference presentations

2022

**Architectures and Mechanisms for Language Processing (AMLaP 2022)**, York, UK.

**POSTER** Sound symbolic associations via contrastive learning of text and images (joint work with Federico Bianchi and Giuseppe Attanasio)

2022

**Architectures and Mechanisms for Language Processing (AMLaP 2022)**, York, UK.

**POSTER** Computational Approaches to Study Systematicity in Word Learning (joint work with Willeke van der Varst)

2022

**Architectures and Mechanisms for Language Processing (AMLaP 2022)**, York, UK.

**POSTER** A bridge between form and meaning. Modeling the relation between names and attributes in fictional characters (joint work with Aron Joosse and Gökçe Kuscü)

2022

**Joint meeting of the European Society for Philosophy and Psychology and the Society for Philosophy and Psychology (ESPP-SPP 2022)**, Milan, Italy.

**POSTER** How thick concepts work in the moral, aesthetic and epistemic domain (joint work with Fenna Bloom, Chanae Smith, Emanuele Arcelli, Qianyu Cheng, and Matteo Colombo)

2021

**Architectures and Mechanisms for Language Processing (AMLaP 2021)**, Paris, France.

**LONG TALK** Not just form, not just meaning: Consistency in form-meaning mappings predicts age of acquisition beyond semantic and form neighborhood density (joint work with Niklas Limacher)

2020

**Architectures and Mechanisms for Language Processing (AMLaP 2020)**, Potsdam, Germany.

**POSTER** Words that shift are learned later: A study on the relation between diachronic semantic change and age of acquisition (joint work with Federico Bianchi and Marco Marelli)

2018

**11th International Conference on the Mental Lexicon**, Edmonton, Alberta, Canada.

**POSTER** On the semantics of non-words and their lexical categories

2017

**39th annual meeting of the Cognitive Science Society**, London, United Kingdom.

**TALK** Distributional learning and lexical category acquisition: What makes words easy to categorize?

2016

**38th annual meeting of the Cognitive Science Society**, Philadelphia, PA, United States of America.

**POSTER** Constraining the search space in cross-situational learning: Different models make different predictions

2015

**6th workshop on Cognitive Aspects of Computational language Learning (CogACLL)**, Lisbon, Portugal.

**POSTER** Which distributional cues help the most? Unsupervised context selection for lexical category acquisition

## Skills

**Research** Cognitive science, Computational Linguistics, Artificial Intelligence, Statistics & research design, Machine learning, Deep learning, Data analysis

**Programming** Python, R, Matlab, Unix

**Tools** RStudio, PyCharm, L<sup>A</sup>T<sub>E</sub>X, Git, Canvas

**Others** Scientific communication, Public speaking, Event organization, Improv theatre

## Honors

### ***Best Student Presentation Award***

11th International Conference on the Mental Lexicon. Edmonton, Alberta. 2018. POSTER TITLE: *On the semantics of non-words and their lexical categories.*

### ***Collegio di merito Bernardo Clesio***

Honor program of the University of Trento.

## Service

2022

**Lab coordinator, MindLabs.**

Coordinator of the Natural Language and Data Technologies at MindLabs. responsible for the management of the lab, the acquisition and management of equipment, the facilitation of research in computational linguistics and psycholinguistics.

2022

**Organization of the 2022 Symposium for Digital Humanities, Tilburg, The Netherlands.**

I co-organise the 2nd edition of the DH symposium organised by the Tilburg School of Humanities and Digital Sciences.

2021

**Organization of the 2021 SIGIR 2021 E-commerce workshop data challenge, Montreal, Quebec, Canada.**

I co-organised the 2021 data challenge of the E-commerce workshop co-located with SIGIR 2021, in particular on the set-up of the intent prediction task.

2020

**Science popularisation.**

Member and content coordinator of the Italian popular science YouTube channel and Spotify Podcast on AI created by the COMitato per l'INTelligenza ARTificiale (COMINTART) (<https://www.youtube.com/channel/UCMgjppN-5w50At67o2jzK7w>).

2018

**Reviewer.**

I've reviewed for *Deevelopmental Science*, *AACL-ICJNLP*, *Frontiers in Psychology*, the *Annual Meeting of the Cognitive Science Society (CogSci)*, the *Journal of Experimental Psychology: General*, the *Austrian Science Foundation*, *EACL*, *CoNLL*, *Language, Cognition and Neuroscience*, *EMNLP*, *CoNLL*, *AMLaP*, the *Journal of Child Language*, *Cognitive Science*, *Behavior Research Methods*, *ACL*, the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, the *Israeli Science Foundation*, *Linguistics*, the *Journal of Natural language Engineering*.

2015  
2018**Organization of the CLiPS research colloquia series, University of Antwerp, Antwerp, Belgium.**

I organized several research colloquia and seminars at CLiPS, involving professors and experts from other universities and companies.

2015

**Organization of the 2015 ATiA workshop, Antwerp, Belgium.**

I was in charge of the submissions and the program within the organization of the ATiA workshop, which involves the computational linguistics groups from the universities of Antwerp, Tilburg, Gent, and Nijmegen.

## Other Certificates

### Science Communication

*Scientific Storytelling*

*Writing well about science for popular audiences*

*How to create awesome scientific presentations and posters*

Let's Talk Science! Summer School, Gent, Belgium. 2016 – 2017.

### Methodology

*Introduction to Bayesian regression modeling with brms*

Fourth Summer School on Statistical Methods for Linguistics and Psychology, Postdam, Germany. 2020 (*online*).

*Multilevel modeling*

*Generalized Linear Models and model building*

FLAMES Summer School, Hasselt, Belgium. 2016.



*R workshop*

*Analysis of grouped and longitudinal data using Linear Mixed Models*

StatUA courses, University of Antwerp, Antwerp, Belgium. 2015 – 2016.

**English** *Test of English as a Foreign Language (TOEFL)* - 112/120 (C2)  
ETS. Bologna, Italy. 2014.

**French** *Diplôme d'Études de Langue Françaises (B2)* - 80%  
Alliances Française, Bologna, Italy. 2009.

**Dutch** *Nederlands niveaus 1 – 4*  
Linguapolis, University of Antwerp, Antwerp, Belgium. 2015 – 2017.

---

## References

**Prof. dr. R. Harald Baayen** Eberhard Karls Universität Tübingen  
Tübingen, Germany  
✉ harald.baayen@uni-tuebingen.de  
☎ +49 (0) 7071 29-73117

**Prof. dr. Walter Daelemans** University of Antwerp, CLiPS  
Antwerp, Belgium  
✉ walter.daelemans@uantwerpen.be  
☎ +32 (0) 3 2655222

**Prof. dr. Steven Gillis** University of Antwerp, CLiPS  
Antwerp, Belgium  
✉ steven.gillis@uantwerpen.be  
☎ +32 (0) 3 2655240

*Last updated on January 30, 2023*