Ehud Reiter and the University of Santiago de Compostela: some notes and memories

Alejandro Ramos-Soto, Nikolay Babakov, Javier González-Corbelle, Jose M. Alonso-Moral, Alberto Bugarín-Diz, Senén Barro
CiTIUS-Centro Singular de Investigación
en Tecnoloxías Intelixentes
Universidade de Santiago de Compostela

1 From linguistic descriptions of data to Natural Language Generation

At the CiTIUS-University of Santiago de Compostela, we had been working for years on fuzzy logic, particularly on linguistic summaries—also known as linguistic descriptions of data. This line of research, based on the use of fuzzy labels and quantifiers to generate short, human-readable summaries of datasets, was (almost) unknown outside the fuzzy community. When Alejandro Ramos began his work as a PhD student in 2011, we had already started to glimpse the potential of Natural Language Generation (NLG) through the pioneering work of Janusz Kacprzyk (Kacprzyk and Zadrozny, 2010), who was among the first to recognize how data-to-text (D2T) approaches could enrich our field, and who often referenced Ehud Reiter's contributions.

At that time, our main challenge was to take a theoretical framework not originally designed for this purpose and make it work as part of a data-to-text system. This effort crystallized in the GaliWeather meteorological text generation system (Ramos-Soto et al., 2015), whose design—and especially whose evaluation—was deeply influenced by Ehud Reiter's work. His emphasis on human-centered evaluation resonated strongly with us and marked a shift in our way of doing research, introducing practices that were still uncommon in the fuzzy logic community.

This line of work led Alejandro and Alberto Bugarín-Diz to participate in the "1st Workshop on Data-to-text Generation" (Edinburgh, March 2015). It is fair to acknowledge here the vision of the Workshop organisers and the open-mindedness of the D2T community, since they aimed not only for (usual) papers, but also for summaries of research activities, work in progress, opinion pieces, ... We are pretty sure that, in a more conventional venue, our abstract would never have made its way to the workshop. There, we had the opportunity not only to present our ideas, but also to meet Ehud Reiter in person, along with many researchers whose work we had long admired. We still remember that first conversation with him: open, generous, and genuinely interested in what we were trying to do. Encouraged by that exchange, we asked about the possibility of a short research stay, which soon became a reality that same autumn after the Summer School on Natural Language Generation in Aberdeen.

That was the beginning of a collaboration that we look back on with great fondness. We carried out research stays in 2015 and 2016, working closely with Ehud and his group (Ramos-Soto et al., 2016), and gradually becoming part of a broader and welcoming community that included José Alonso, Kees van Deemter, Albert Gatt, Nava Tintarev, and many others. Another fine example of this inclusive and welcoming spirit is the fact that, during the 2016 INLG conference (Edinburgh again), following the Bid presented by Jose Alonso-Moral, the community chose Santiago de Compostela as the venue for the 2017 conference (Alonso et al., 2017). As a consequence of that, around 100 NLG researchers were visiting our city... just a couple of years after we first made contact with this community!. Ehud's guidance as General and Programme Co-Chair, as well as his Invited Tutorial on "Evaluating Natural Language Generation Systems," were key to the scientific and organisational success of this event (Alonso and Bugarín, 2017). During this time, Alejandro focused on the definition of fuzzy geographical referring expressions and their role in generating linguistic descriptions (Ramos-Soto et al., 2019), a line of work that continued into his postdoctoral research in 2017–2018, and later led him to a short but memorable visit with Albert Gatt in Malta in 2019.

Even as Alejandro later became involved in an entrepreneurial project (Inverbis Analytics), our

connection to this line of research—and, more importantly, to the people behind it—remained. Our collaboration with Ehud Reiter and the wider NLG community did not just influence our work; it changed how we understood our own research, helping us build a bridge between fuzzy logic and data-to-text generation that we had not initially imagined.

2 Evaluation

This shared commitment to rigorous, human-centered evaluation naturally paved the way for the next generation of researchers in our group. In 2019, Javier González-Corbelle joined CiTIUS, beginning a master's thesis focused on the evaluation of NLG systems and later becoming a doctoral student in 2021, on the topic of Improving Robustness and Reproducibility in Validation of Natural Language Generation Systems. Javier's objective was to develop strict validation guidelines centered on the most reliable assessment method: human evaluation. As the landscape shifted with the rise of Large Language Models (LLMs), Javier's work expanded to assess new challenges in the generation process, such as hallucinations and omissions (González-Corbelle et al., 2022). Ultimately, his goal has been to study how the selection of different parameters affects evaluation results, ensuring higher reliability and reproducibility (González-Corbelle et al., 2024).

As reproducibility became an increasingly critical topic, especially regarding human evaluation in NLP, our group's trajectory intersected with Ehud's once more. In 2022, we were a partner lab within the ReproHum project, led by Anya Belz and Ehud, and working with Craig Thomson, who managed the coordination across partner universities. Working hand in hand with the members of the project on NLP human evaluation reproduction experiments, we took away important lessons regarding the critical aspects required to conduct high-quality human evaluations (González-Corbelle et al., 2023; González Corbelle et al., 2024). This collaboration subsequently led to Javier completing a research stay with Anya's group at Dublin City University in 2025.

Also in 2025, Ehud visited the University of Santiago de Compostela and gave us a talk on "High Quality Human Evaluation of Generated Texts.' During the visit, Ehud was kind to dedicate time to discuss with us in detail several aspects of Javier's work on the evaluation of NLG systems. His valuable advice, insights, comments, and guidance were very helpful to clarify and improve this research, which is a core part of the forthcoming Javier's thesis.

3 Building and Explaining Bayesian Networks using Natural Language

This (last by the moment) visit to Santiago de Compostela was to attend Nikolay Babakov's PhD thesis presentation and defense, for whom he was co-supervisor, together with Alberto. Nikolay's thesis, on Novel Methods for Bayesian Networks Construction and Explanation using Natural Language, was one of the last results of the NL4XAI project (PI, Senén Barro, Deputy Coordinator Jose Alonso), an MSCA action on the interplay between Natural Language Technologies and Explainable Artificial Intelligence, which was a truly remarkable milestone for CiTIUS and our group in terms of international collaboration and external projection (Alonso, 2019). In his thesis, Nikolay improved the construction, inference explanation, and user trust of Bayesian Networks using Natural Language Techniques, including LLM (Babakov et al., 2025a,e). He addressed BN reusability, developed the CausalGraphBench benchmark for evaluating LLM-driven BN construction using structure learning (Babakov et al., 2025b), and validated the Factor-Argument-based framework for inference explanation in natural language, both in generic and specialised (medical) domains (Babakov et al., 2025c,d) Co-supervising Nikolay's thesis with Ehud was a real pleasure, since we three had complementary views, perspectives, and backgrounds, which produced fruitful discussions that enriched a lot Nikolay's research path.

4 Not an epilogue

After years of engaging with his work and maintaining close contact, Ehud's visits have always been deeply meaningful events, which reinforced the core principles of high standards in NLG modelling and evaluation he originally inspired in us from the very beginning.

Looking back, what stands out is not only the intellectual impact of Ehud's work but also his generosity, openness, and quiet encouragement. He made it easy for us to feel that our ideas mattered, even when we were stepping outside our original field. For that, and for the many conversations and

collaborations along the way, we remain deeply grateful.

But this does not stop here! Despite your retirement, Ehud, we will create new and prompt opportunities to host you at CiTIUS and to have you again in Santiago de Compostela very soon. Once again, as always,... it will be our great pleasure!.

Acknowledgments

The support of the Galician Ministry for Education, Universities and Professional Training, and "ERDF A way of making Europe" is also acknowledged through the grant "Centro de investigación de Galicia accreditation 2024-2027 ED431G-2023/04".

References

Jose M. Alonso. 2019. Interactive Natural Language Technology for Explainable Artificial Intelligence H2020-MSCA-ITN-2019 MSCA-ITN-ETN – European Training Network. Accessed = 2026-05-22.

Jose M. Alonso and Alberto Bugarín. 2017. 10th International Conference on Natural Language Generation (INLG2017) Website. Santiago de Compostela, September 4-7, 2017. Accessed = 2026-05-22.

Jose M. Alonso, Alberto Bugarín, and Ehud Reiter, editors. 2017. *Proceedings of the 10th International Conference on Natural Language Generation*. Association for Computational Linguistics, Santiago de Compostela, Spain.

Nikolay Babakov, Ehud Reiter, and Alberto Bugarín. 2025a. Scalability of Bayesian network structure elicitation with large language models: a novel methodology and comparative analysis. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 10685–10711, Abu Dhabi, UAE. Association for Computational Linguistics.

Nikolay Babakov, Ehud Reiter, and Alberto Bugarín-Diz. 2025b. CausalGraphBench: a benchmark for evaluating language models capabilities of causal graph discovery. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 240–258, Vienna, Austria. Association for Computational Linguistics.

Nikolay Babakov, Ehud Reiter, and Alberto Bugarín-Diz. 2025c. Explaining bayesian networks reasoning to the general public: Insights from the user study. In *Proceedings of the 2025 Workshop Multimodal, Affective and Interactive eXplainable AI MAI-XAI (part of ECAI2025)*, Bologna, Italy.

Nikolay Babakov, Elena Rezgova, Ehud Reiter, and Alberto Bugarín-Diz. 2025d. Evaluation of explainable ai by medical experts: a survey of the existing approaches. In *Proceedings of the 2nd Workshop on Explainable Artificial Intelligence for the Medical Domain EXPLIMED (part of ECAI2025)*, Bologna, Italy.

Nikolay Babakov, Adarsa Sivaprasad, Ehud Reiter, and Alberto Bugarín-Diz. 2025e. Reusability of Bayesian Networks case studies: a survey. *Applied Intelligence*, 55(6):417.

Javier González-Corbelle, Jose M. Alonso-Moral, and A. Bugarín-Diz. 2023. Some lessons learned reproducing human evaluation of a data-to-text system. In *Proceedings of the 3rd Workshop on Human Evaluation of NLP Systems*, pages 49–68, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Javier González-Corbelle, Jose M. Alonso-Moral, A. Bugarín-Diz, and J. Taboada. 2022. Dealing with hallucination and omission in neural natural language generation: A use case on meteorology. In *Proceedings of the 15th International Conference on Natural Language Generation*, pages 121–130, Waterville, Maine, USA and virtual meeting. Association for Computational Linguistics.

Javier González Corbelle, A. Vivel-Couso, J.M. Alonso-Moral, and A. Bugarín-Diz. 2024. ReproHum #0927-3: Reproducing the human evaluation of the DExperts controlled text generation method. In *Proceedings of the Fourth Workshop on Human Evaluation of NLP Systems (HumEval) @ LREC-COLING 2024*, pages 153–162, Torino, Italia. ELRA and ICCL.

Javier González-Corbelle, Jose M. Alonso-Moral, Rosa M. Crujeiras, and Alberto Bugarín-Diz. 2024. An empirical study on the number of items in human evaluation of automatically generated texts.

J. Kacprzyk and S. Zadrozny. 2010. Linguistic data summarization: A high scalability through the use of natural language? *Scalable Fuzzy Algorithms for Data Management and Analysis: Methods and Design*, pages 214–237.

Alejandro Ramos-Soto, Jose M. Alonso, Ehud Reiter, Kees Van Deemter, and Albert Gatt. 2019. Fuzzy-based language grounding of geographical references: From writers to readers. *International Journal of Computational Intelligence Systems*, 12(2):970–983.

Alejandro Ramos-Soto, Alberto Bugarín, Senén Barro, and Juan Taboada. 2015. Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data. *IEEE Transactions on Fuzzy Systems*, 23(1):44 – 57.

Alejandro Ramos-Soto, Nava Tintarev, Rodrigo de Oliveira, Ehud Reiter, and Kees van Deemter. 2016. Natural language generation and fuzzy sets: An exploratory study on geographical referring expression generation. In *IEEE World Congress on Computational Intelligence, 2016 IEEE International Conference on Fuzzy Systems,*, pages 587–594.