



The Fifteenth Language Resources and Evaluation Conference (LREC 2026)

Main Conference Proceedings

Editors:

**Stelios Piperidis, Núria Bel, Henk van den Heuvel,
Nancy Ide, Simon Krek and Antonio Toral**

13–15 May, 2026

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)

©ELRA Language Resources Association (ELRA), 2026
These proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN: 978-2-493814-49-4

EAN: 9782493814494

ISSN: 2522-2686

Proceedings available at

<http://www.lrec-conf.org/proceedings/lrec2026/>.

Message from the General Chair

It is my great pleasure and honor to welcome you to the 15th International Conference on Language Resources and Evaluation in Palma, the capital of the wonderful island of Mallorca.

It is with great emotion that I am writing this message. Starting in 1998 in Granada, in its 15th edition this year, LREC has evolved into a well-established conference in our community. Running as an independent conference every two years, except for 2024 when it joined forces with COLING to stage the LREC-COLING 2024 in Turin, LREC, with the involvement of many members of our community in its design and implementation, has been gradually turning into a community-run conference. This change started already at LREC2022 and culminated into a fully community-run conference in this 15th edition. This is the right moment to sincerely thank Nicoletta Calzolari for her key role in coordinating the many colleagues, members of the LREC Program Committee, for over two decades, whose collective efforts have shaped the conference.

The organizational structure and the people behind LREC2026

In previous such messages we used to read about the novelties of each edition. This message will be no different. The first and most important novelty is that LREC2026 adopted a new structure, with 14 organisational roles undertaking different aspects of the conference. I wish to thank them all, one by one and all together for making LREC2026 not only a reality but also a big success.

The Program Chairs – my long-term colleagues and friends Nuria Bel, Nancy Ide, Simon Krek, Antonio Toral and Henk van den Heuvel – who have supervised every step of the scientific part of LREC2026, from the write-up of the Calls for Papers to the preparation of an exciting scientific programme. Nuria, Nancy, Simon, Antonio and Henk have overseen the double-blind reviewing process dealing with a complex structure involving 58 Senior Area Chairs, 233 Area Chairs and no less than 1800 scientific reviewers. They have replied to many hundreds of emails regarding different facets of the reviewing process, a process that has become overly demanding not only due to the constantly rising number of submissions but also due to the increasing questions and risks the use of AI models and LLMs introduces to both the paper writing and reviewing processes.

Our Ethics Chairs - Denise DiPersio, Thomas Margoni and Claudia Cevenini – together with the Ethics reviewers have diligently reviewed all those papers that were tagged as raising ethical concerns. Out of the 944 accepted papers, 28 of them were accepted on condition that the authors would resolve the ethical questions raised and the respective final submissions were therefore carefully scrutinised to ensure they have appropriately resolved them, and they meet the ethical standards of the conference.

The Publication Chairs – John McCrae and Gilles Sérasset – have in turn ensured the timely delivery of the conference proceedings totaling over 12.000 pages. John and Gilles processed the camera-ready papers attending to every detail, frequently interacting with many authors to secure the high-quality of every publication.

The Call for Workshops of LREC2026 was undebatably extremely successful. The Workshop Chairs – Darja Fiser, Luis Chiruzzo and Menno van Zaanen – undertook the responsibility of reviewing 55 very well-prepared workshop proposals. Darja, Luis and Menno put up a very compelling workshop programme with 45 workshops, spanning three days of the LREC week,

almost a conference in a conference. The Call for Tutorials was equally successful. The Tutorial Chairs – Francesca Frontini and Atul K. Ojha – reviewed 18 tutorial proposals to finally select 12 of them spanning a wide range of topics, many of them focusing on different aspects and challenges large language models present.

The Industry Day, which was introduced a few editions ago, was entrusted to Teresa Lynn and Natalie Schluter who did a tremendous job, from preparing a call for abstracts particularly addressed to industrial researchers, to reviewing and designing an exciting programme bringing together 19 presentations that aim to bridge research and real-world applications.

LREC2026 could never have been possible without the continuous, daily support of the LREC2026 Management Chairs – Khalid Choukri, Helene Mazo and Sara Goggi. Their experience with preparing and running LREC for so many years has been invaluable. Sara Goggi and Helene Mazo provided priceless help to the Program Chairs and Workshops Chairs respectively. Together with Fernanda González Campo, Alexandre Sicard, Valerie Mapelli and Caroline Merlieux Petit, from ELDA, the Management Chairs set up and maintained the LREC2026 website, attended to organizational details, communicated with and dealt with the many diverging requests of the conference participants especially during the last months. Caroline Kravec, together with the Conference Catalysts team, set up the LREC2026 conference app and helped with the hybrid aspects of the conference featuring videos and presentation materials of the oral and poster papers of the main conference as well as the details of all workshops and tutorials.

The LREC 2026 Local Organisation Committee Chairs – German Rigau, Maria del Mar Vanrell, Melania Sanchez Masià and Rafael Muñoz Guillena – together with UIB Congr s have done not only whatever it takes to solve the diverse questions the organization of a big conference like LREC raises at the local level, but also to prepare an exciting social programme.

The Sponsorship Chairs – Patrick Paroubek, Alexandros Poulis, Jimmy Kunzmann and Raquel L zaro Guti rrez – and the Exhibition Chair - Rafael Mu oz Guillena - have mobilised their networks and secured the support of Dataforce, Mozilla Data Collective, Hitz, OpenEuroLLM, Common European Language Data Space, ALT-EDIC, Multiverse Computing, translated and Linguist List, to whom we extend our special thanks.

The Publicity Chairs – Elena Volodina, Rafael Valencia Garc a and A. Seza Do ru z – have helped spread the word about LREC, and our Diversity and Inclusion Chair – Sakriani Sakti – has been ready to ensure the inclusivity of LREC and support diverse participation.

Our Advisor - Nicoletta Calzolari – shared her vast experience, in particular with me, at different times, from the start of the conference preparations to designing the program.

LREC2026 in numbers: breaking all records

This 15th edition of LREC has broken all previous records. 1786 papers were submitted to the main conference. The submissions were organized in 27 areas, the most popular being “Corpora, Treebanks and Annotation; Tools, Systems and Platforms” with 202 papers, “Evaluation, Validation, Quality Assurance and Benchmarking Methodologies” with 139 submissions, “Applications Involving LRs and Evaluation for any area/domain of language and speech processing” with 112 submissions and “Multilinguality, Machine Translation (including Speech-to-Speech) and Translation Aids” with 100 submissions.

Out of the 1786 submissions, 944 were accepted representing an acceptance rate of 52.85%.

The accepted papers is the result of work of almost 4300 authors from 87 countries. The results reflect the LREC principles of inclusiveness, multilingualism and language equality, and they re-affirm that the field of language resources and evaluation is constantly evolving, expanding and flourishing, even more so in the era of AI and Large Language Models.

Records have been broken at the workshop and tutorial level, too. Out of the 55 workshop proposals submitted, 49 were accepted and after a few mergers 45 made it to the conference programme. In addition, 18 tutorial proposals were submitted, while 12 of them have been selected for inclusion. The Industry Day Call attracted 21 abstracts, with 19 of them now in the programme.

Registrations have surpassed all expectations. At the time of writing, almost five weeks before the conference, over 2000 participants have registered for the main conference. Workshops organized in the framework of LREC2026 have at the same time attracted 1748 participants, while tutorials expect to get over 420 attendees.

Keeping the recently established tradition of hybrid conferences but also responding to the increasing requests for remote participation due to the international conflict, almost 300 participants are expected to join the conference remotely, while around 1700 will join onsite. Out of the current registrations, 834 colleagues are joining ELRA as new members, 382 of them being students; we welcome all of them to ELRA and are looking forward to their contributions.

A big “thank you” to ...

Organising a big conference like LREC is a demanding and challenging task. It requires detailed planning, clear roles and task allocation, 360° view and in tandem attention to detail at all stages, from the preparation of the organisational structure to the actual running of the conference. And it requires the involvement of many people, the expertise and motivation of whom are indispensable.

I would like to start this round by thanking all the members of the ELRA Board, the ELRA Presidents, Simon Krek and German Rigau, and the General Secretary, Khalid Choukri, for entrusting me with the role of LREC2026 General Chair. For this edition all members of the ELRA Board have assumed roles in the different conference committees, and I want to thank them for agreeing to be on board and for supporting as much as it was possible.

I want to warmly thank all the Committee Chairs: Program Chairs, Workshop Chairs, Tutorial Chairs, Industry Track Chairs, Management Chairs, Local Organisation Chairs, Publication Chairs, Sponsorship Chairs, Exhibition Chair, Publicity Chairs, Ethics Chairs, Diversion and Inclusion Chair, the Advisor for all they have done for this LREC, each in their own role. We have worked extremely well together, I believe we can be proud of the result.

Big thanks go to the 58 Senior Area Chairs, the 233 Area Chairs, the 1800 reviewers, you all did a tremendous job.

Extra special thanks are due to our keynote speakers, Nancy F. Chen, Dan Jurafsky and Nicolau Dols.

A big thank you to all workshop and tutorial organisers; your contribution to the conference is invaluable as it is also attested by the record number of participants you have attracted.

I wish to express my gratitude to the local authorities in Palma, the authorities and colleagues of the Universitat de les Illes Balears as well as the sponsors of LREC who have generously

helped with their support.

Thanks again to the ELDA staff, the UIB Congr  s, the Conference Catalysts staff, the volunteers and the Palau de Congressos de Palma for making it happen.

Above all, I wish to thank all conference participants, at all stages: the authors of the so many submissions, the proposers of the so many workshops and tutorials, the Industry Day participants; the enormous interest you have shown by submitting to LREC is the unique indication of a thriving field and of LREC's high relevance and importance.

Dear LREC participants, colleagues and friends,

Now the conference is in your hands. With your active participation in the 40 oral sessions and the 57 usually bustling with life poster sessions, the engaging Industry Day, the 45 workshops and the 12 tutorials you make LREC the unique conference it is. I hope LREC2026 will open new research paths, foster fruitful contacts and collaborations and offer memorable moments in Palma de Mallorca.

Enjoy LREC2026, enjoy Palma!

Stelios Piperidis

Athena Research Centre, Institute for Language and Speech Processing

LREC2026 General Chair

Message from the ELRA President

ELRA: an International Scientific Society within the Era of Large Language Models

In 2025, ELRA celebrated 30 years of existence and in 2026, we are organizing the fifteenth biannual LREC conference. These are two significant achievements that show the resilience and vitality of our mission and organization. Since our last gathering, ELRA has continued to evolve as a sustainable key player in the rapidly changing scientific landscape of language technology and language resources.

The impact of Large Language Models (LLMs) and generative Artificial Intelligence remains a defining force in our research field and beyond. As we convene for **LREC 2026** in Palma, this influence is clearly reflected in our keynote topics and the groundbreaking research presented in the accepted papers. ELRA continues to adapt to this era, ensuring we provide the necessary language data and support for all human languages.

At its inception, the primary mission of ELRA was to secure a self-sufficient platform for the preservation and distribution of language resources that were either developed by the LT industry, or resulted from various research projects supported by the European Commission. Currently, ELRA has evolved to become an international scientific society whose mission is to foster the progress of the science and technology of language with a particular interest in language data and evaluation.

Our transition into a truly international scientific society is now firmly established. Under our updated statutes, the ELRA acronym—standing for "ELRA Language Resources Association"—represents a global community of both institutional and individual members. We are proud to see our individual members actively participating in the General Assembly and shaping the future of the association.

As we look forward, it is vital that ELRA, as a senior international scientific organization in language resources, continues to coordinate and join forces with major international language technology scientific associations. These collaborations are essential to fostering progress in the science and technology of language.

Finally, I wish to thank everyone who has made this event possible. My gratitude goes to the general chair, the previous ELRA presidents, and of course, the ELRA Secretary General and the ELDA team, as well as the programme, workshop, and tutorial chairs, the publication and area chairs, and the dedicated local organizers and management teams.

To our new members: welcome to the association. To all participants: welcome to LREC 2026!

German Rigau

President of ELRA

Message of the ELRA Secretary General and ELDA CEO

LREC: A Longstanding Vision

The LREC conference has, over the years, established itself as a cornerstone event for the international community working on language resources and evaluation. It remains a unique forum where researchers, industry leaders, and public institutions come together to exchange ideas, share results, and collectively shape the future of language technologies.

It is worth recalling that the first edition of LREC took place in 1998, in Granada, Spain. This pioneering initiative was grounded in a strong conviction: progress in language technologies depends not only on algorithms, but also, and fundamentally, on the availability, quality, and proper use of data.

Nearly three decades later, this vision has proven to be remarkably accurate, as data has become the central driver of advances in language technologies.

A Transforming Landscape: The Rise of LLMs

In 2026, we meet in a context of profound transformation. The rapid evolution of Artificial Intelligence, and in particular the widespread deployment of Large Language Models (LLMs), is redefining not only technological capabilities but also the foundations of our field. These models are changing the way language data is produced, curated, accessed, and exploited. They are also raising new challenges related to data quality, legal compliance, evaluation methodologies, and sustainability.

A Fundamental Shift in the Role of Data

A more fundamental shift is also taking place in the way we relate to language data.

Over the past decades of Natural Language Processing, language data has primarily been used as a linguistic foundation to understand how language is structured and how it functions. It served as the basis for modeling linguistic phenomena and for developing applications such as Automatic Speech Recognition, Machine Translation, Information Extraction, Summarization, etc., typically relying on language/image and acoustic models, based on probabilistic approaches.

In this paradigm, data was produced independently of the systems that consumed it. Language data originated from human activities, including publications, media, conversations, and multimodal content, and was then collected, curated, and used to train or evaluate systems. The role of language technologies was essentially transformative: transcribing speech, translating text, extracting information, or generating limited and task-specific outputs.

While Natural Language Generation has existed for many years, its capabilities remained relatively constrained, often domain-specific, and typically dependent on structured inputs or predefined templates. It did not fundamentally alter the balance between data producers and technology providers.

With the advent of LLMs, this balance is shifting significantly.

Modern AI systems are now capable of generating high-quality content across multiple modalities, including text, speech, and increasing images and videos, often at a level that can approximate human-produced material. This represents a major shift: language technologies are no longer only transforming existing data but also producing new content at scale.

Data, Value, and Emerging Tensions

This evolution raises important questions for the entire ecosystem, and in particular for those who produce and own language data.

Publishers, media organizations, and content producers, including newspapers, broadcasters, and other information providers, are increasingly concerned about how their data is being used. On the one hand, their content constitutes a valuable source for training large models. On the other hand, these same models are now capable of generating content that may compete with their original productions.

As a result, these stakeholders are seeking ways to better control, protect, and monetize their data assets, while continuing to contribute to innovation. This includes exploring new licensing schemes, data-sharing agreements, and mechanisms for ensuring transparency and traceability in data usage in trainings.

Regulation and the Balance with Innovation

At the same time, this evolution is accompanied by a strong regulatory response. The European Union, along with other countries worldwide, is actively developing and enforcing legal frameworks aimed at ensuring fair, transparent, and lawful access to data. These include, in particular, regulations related to data protection, data governance, and artificial intelligence.

While such frameworks are essential to protect rights holders and data subjects, they also introduce additional complexity in the use and sharing of data. In some cases, they may slow down certain forms of innovation or require significant adaptation efforts from stakeholders.

From our perspective, the challenge is therefore to strike the right balance: enabling innovation to flourish while maintaining high standards of legal compliance, ethical responsibility, and respect for data ownership.

This situation highlights the need for a balanced and sustainable data ecosystem, where the interests of data producers, technology developers, and end users are aligned. It reinforces the importance of clear legal frameworks, robust data governance mechanisms, and trusted intermediaries capable of facilitating compliant and mutually beneficial data exchanges.

ELDA, with its long-standing expertise in data management, licensing, and compliance, is well positioned to contribute to this evolving landscape by supporting fair, transparent, and sustainable models for data sharing and reuse.

ELDA's Role: Data Production, Repurposing, and Automation

In this rapidly evolving landscape, the mission of ELRA and ELDA remains fully aligned with the needs of the community, while continuously adapting to new requirements.

A central dimension of our work today lies in data production and repurposing for LLMs. The development and fine-tuning of large-scale models require not only massive amounts of data, but above all high-quality, well-curated, and legally compliant datasets. This is where ELDA brings a long-standing expertise that has become even more critical in the current context.

Our activities cover the full lifecycle of language resources, from identification and acquisition, to cleaning, annotation, documentation, packaging, and distribution. Increasingly, we are also engaged in the repurposing of existing datasets, adapting them to new use cases such as fine-tuning, evaluation, or domain adaptation of LLMs. This includes transforming legacy datasets, aligning heterogeneous resources, and ensuring that they meet the technical and legal requirements of modern AI systems.

Real and Synthetic Data: Complementary Approaches

In this context, it is important to emphasize that today's data ecosystem relies on a combination of real-world data and synthetic data. Real data, collected from authentic sources, remains essential to capture the richness, variability, and cultural grounding of human language. At the same time, synthetic data generation has emerged as a powerful complement, enabling the extension of datasets, the balancing of under-represented phenomena, and the creation of controlled evaluation scenarios.

ELDA is actively working on both dimensions. We support the collection and curation of real data across languages, domains, and modalities, while also exploring methodologies for generating and validating synthetic data that can be reliably used for training and evaluation purposes. In all cases, our objective is to ensure that both real and synthetic datasets meet the highest standards in terms of quality, traceability, reproducibility, and compliance.

Automation and Compliance-by-Design

The scale and complexity of these operations require a shift in methodology. Manual processes alone are no longer sufficient. For this reason, ELDA is investing significant efforts in the automation and semi-automation of data processing pipelines. These efforts aim to increase efficiency, reproducibility, and reliability, while maintaining the high standards expected by the community.

A key aspect of this transformation is the integration of compliance-by-design approaches within our workflows. The datasets used to train and evaluate modern AI systems must comply with a growing number of legal and regulatory frameworks, both within Europe and internationally.

Within the European context, this includes in particular the General Data Protection Regulation (GDPR), the Data Governance Act, the Data Act, and the emerging AI Act, all of which define new requirements for data usage, sharing, and accountability.

Beyond Europe, similar regulatory developments are taking place worldwide, including privacy regulations such as the evolving regulatory landscape in the United States at the federal and state levels, India's Digital Personal Data Protection Act, Brazil's LGPD, and Japan's Act on the Protection of Personal Information (APPI). While these frameworks pursue objectives similar to those of the GDPR, they differ in scope, level of prescriptiveness, and enforcement mechanisms.

This global regulatory landscape requires a careful and systematic approach to ensure that datasets can be used across jurisdictions without legal uncertainty.

In this context, ELDA has developed a strong expertise in assessing and ensuring compliance throughout the data lifecycle. We are progressively embedding automated and semi-automated compliance checks into our data processing pipelines, covering aspects such as personal data detection, anonymization, licensing conditions, provenance tracking, and usage restrictions. These mechanisms are designed to provide guarantees to both data providers and data users, enabling trusted data sharing at scale.

Strategic European Initiatives

In parallel with these operational activities, ELRA and ELDA are actively contributing to major European strategic initiatives aimed at strengthening the ecosystem for language technologies.

In particular, we are involved in the ALT-EDIC4EU framework, as well as in projects such as LLMs4EU and OpenEuroLLM, which aim to support the development of large language models aligned with European strategic priorities, including multilingualism, openness, and compliance with regulatory frameworks.

These initiatives are essential to ensure that Europe maintains a strong position in the global AI landscape, while preserving its linguistic diversity and promoting responsible innovation. ELDA contributes to these efforts by bringing its expertise in data management, resource sharing, and legal compliance.

Rethinking Evaluation in the Era of LLMs

At the same time, we are reinforcing our commitment to another fundamental pillar of LREC: the evaluation of language technologies.

The emergence of LLMs challenges traditional evaluation methodologies, which were often based on well-defined benchmarks and static datasets. Modern systems are more dynamic, more complex, and often less transparent. This calls for new approaches to evaluation that can capture robustness, generalization, and real-world performance.

In this context, ELDA is involved in new initiatives such as the ARCHER project, Advancing Robust and Creative Human Language Technologies through Challenge Events and Research, funded by the European Defence Fund. ARCHER focuses on advancing Human Language Technologies through large-scale, challenge-based evaluation campaigns. Its scope encompasses several key tasks, including Automatic Speech Recognition (ASR), Machine Translation (MT), Optical Character Recognition (OCR), and Named Entity Recognition (NER), reflecting the diversity of modalities and applications addressed by modern language technologies. The project places a strong emphasis on European languages, while also considering a broader range of languages beyond Europe, in order to ensure robustness and adaptability in diverse linguistic and operational contexts.

By leveraging state-of-the-art artificial intelligence and machine learning, the project aims to improve the accuracy, adaptability, and robustness of these technologies across diverse environments and languages. Beyond its specific technical focus, ARCHER reflects a broader evolution toward coordinated evaluation frameworks, where data, metrics, and methodologies are designed to capture realistic and demanding use cases, bridging the gap between experimental settings and operational deployment. This approach is fully aligned with the historical mission of LREC and opens new perspectives for the evaluation of next-generation systems.

Core Challenges Ahead

The combination of data for LLMs and evaluation of complex systems highlights a central point: high-quality data remains at the core of our field, but its role is evolving. It is no longer only about producing datasets, but about ensuring that they are fit for purpose, trustworthy, reusable, and compliant within a complex ecosystem.

At the same time, we must not lose sight of one of the fundamental principles that has guided LREC since its inception: multilingual inclusivity. While large language models have made significant progress in covering multiple languages, a large number of languages, in particular sign languages, remain under-represented or insufficiently supported. It is our collective responsibility to ensure that technological advances benefit all languages and communities, including under-resourced and endangered languages.

Sustainability is another critical dimension. The development and deployment of large-scale models raise important questions in terms of energy consumption, environmental impact, and long-term maintenance of resources and infrastructures. Addressing these issues requires coordinated efforts across the community.

Acknowledgments and Outlook

LREC continues to play a central role in addressing these challenges. Through its scientific program, workshops, and community interactions, it remains a key platform to exchange knowledge, define best practices, and foster collaboration.

We would like to express our sincere gratitude to all those who contribute to making this conference possible: the General Chair, the various committees, reviewers, authors, sponsors, and participants. Their dedication and expertise are essential to the success and continuity of LREC.

We would like to warmly thank the LREC 2026 Industry Track Chairs for putting together a truly engaging and insightful program, bringing together diverse perspectives and showcasing innovative developments from across the field.

We would also like to extend our sincere thanks to the ELDA and ILC-CNR teams who, edition after edition, work tirelessly to ensure the seamless organization and consistently high quality of this conference.

Nearly three decades after its first edition in Granada, LREC continues to evolve with the field it helped structure. We are confident that, together, we will continue to advance language technologies in a way that is innovative, inclusive, and responsible.

We wish all readers and participants an inspiring and fruitful LREC 2026.

Khalid CHOUKRI
ELRA Secretary General and ELDA CEO

Message from the Program Chairs

As the Program Chairs of LREC 2026, we are delighted to welcome you to LREC 2026 in Palma de Mallorca, where the Mediterranean setting continues the LREC tradition of hosting the conference in locations that inspire both scientific exchange and meaningful connections. Palma is known for its rich history as a key Mediterranean hub, shaped over centuries by diverse cultures—from the Phoenicians, Genoese, and Pisans to North African sailors—and today it stands as a renowned tourist destination, welcoming over 13 million visitors each year. We feel sure that this environment will provide a unique and enjoyable backdrop for engaging discussions, fruitful collaborations, and memorable experiences throughout the conference.

LREC 2026 received 1786 submissions, of which 944 (52.85%) were accepted following a blind, four-tier peer review. At the core of this effort were the 1800 reviewers, whose careful evaluations formed the foundation of the final program. The 233 Area Chairs (ACs) guided the review discussions and provided metareviews synthesizing the initial assessments. This effort was overseen by 58 Senior Area Chairs (SACs), whose dedication and leadership ensured the smooth coordination of a complex and large-scale review cycle. Their final recommendations were then forwarded to the Program Committee Chairs, who finalized the program based on the SACs' input. We extend our sincere thanks to all of those whose substantial contributions helped shape this year's conference, enabling us to assemble an outstanding final program. In recognition, the names of all reviewers, ACs, and SACs are listed in the conference proceedings.

Among the papers accepted for presentation, 200 (21%) were selected for oral presentation and 744 (79%) for poster sessions. The result is a rich program spanning both established topics in the development, creation, use, and evaluation of language resources, along with emerging topics involving large language models (LLMs). We observed several trends in the topics addressed by this year's submissions. In line with LREC's longstanding focus, there was continued attention to multilingual and low-resource settings, with particular emphasis on bias (especially gender bias) and adapting or developing benchmarks for languages beyond English. As expected, work related to large language models (LLMs) was especially prominent. Many contributions centered on analyzing LLM reasoning capabilities, often through benchmarking existing models to examine reasoning behavior, consistency, and generalization, pointing to a clear shift toward understanding how LLMs reason rather than simply measuring end-task accuracy. We also observed a growing emphasis on grounded and robust reasoning, often situated in retrieval-augmented or knowledge-supported settings; this work explores issues such as faithfulness, hallucination, and evidence-aware inference, reflecting an increased focus on trustworthiness and reliability alongside raw performance. Overall, the LREC 2026 program accurately reflects the current state of language resource development, use, and evaluation, while also documenting the evolution—and relevance—of language resources in a rapidly changing technological landscape.

In addition to the oral and poster presentations, the program features talks by three distinguished keynote speakers: Dan Jurafsky (Stanford University, USA), who addresses the social limitations of language models; Nancy Chen (Agency for Science, Technology, and Research, Singapore), who discusses language resources as the backbone of AI; and Nicolau Dols, who gives an overview of the Balearic language landscape, following the LREC tradition of highlighting local languages and culture.

While we strongly encourage in-person attendance given its greater opportunities for discussion and networking, we also provided remote participation options for both presenters and attendees. Regrettably, the number of remote presenters grew substantially following the onset of a major international conflict during the planning of the conference. To ensure broad access,

videos and accompanying slides for all oral presentations, whether delivered in person or remotely, are available on the conference website. Similarly, videos and PDF versions of both in-person and remote poster presentations can also be accessed online.

We thank the entire LREC organizing team for their continuous assistance throughout:

Workshop Chairs: Luis Chiruzzo (Universidad del al República, Uruguay), Darja Fişer (CLARIN ERIC, The Netherlands), and Menno van Zaanen (North-West University, South Africa); Tutorial Chairs: Francesca Frontini (ILC-CNR, Italy), and Atul Kumar Ojha (University of Galway, Ireland); Industry Track Chairs: Teresa Lynn (Mohammed Bin Zayed University of Artificial Intelligence, UAE) and Natalie Schluter (Apple, Denmark); Management Committee Chairs Khalid Choukri (ELDA/ELRA, Paris, France) Sara Goggi (ILC-CNR, Pisa, Italy), and H         Mazo (ELDA/ELRA, Paris, France); Local Organization Committee Chairs German Rigau (University of the Basque Country, Spain), Maria del Mar Vanrell (Universitat de les Illes Balears, Spain), Melania S         Masi     (Universitat de les Illes Balears, Spain), and Rafael Mu         Guillena (Universidad de Alicante, Spain); Publication Co-Chairs: John McCrae (University of Galway, Ireland) and Gilles S             (Universit     Grenoble Alpes, France); Sponsorship Chairs: Patrick Paroubek (Universit     Paris-Saclay, France), Alexandros Poulis (Transperfect/Dataforce, Luxembourg), Jimmy Kunzmann (Retired, Amazon Alexa Hybrid Science), and Raquel L         Guti             (Universidad de Alcal    , Spain); Exhibition Chair: Rafael Mu         Guillena (Universidad de Alicante, Spain); Publicity Chairs: Elena Volodina (University of Gothenburg, Sweden), Rafael Valencia Garc         (Universidad de Murcia, Spain), and A. Seza Do           (Ghent University, Belgium); Ethics Chairs: Denise DiPersio (Linguistic Data Consortium, USA), Thomas Margoni (KU Leuven, Belgium), and Claudia Cevenini (Universit     di Bologna, Italy); Diversity and Inclusion Chair: Sakriani Sakti (Nara Institute of Science and Technology, Japan); and Conference Advisor: Nicoletta Calzolari (Italy).

We extend extra-special thanks to LREC 2026's General Chair, Stelios Piperidis (Athena Research Center/ILSP, Athens, Greece), for his tireless guidance--always offered with grace and unwavering cheerfulness--throughout every phase of the conference's planning and execution.

Finally, we extend our gratitude to the people who contribute the most to LREC's success--its participants. We look forward to welcoming you in Palma and sharing a rewarding and engaging experience together.

LREC 2026 Program Committee Chairs

N         Bel (Pompeu Fabra University, Barcelona, Spain)

Nancy Ide (Vassar College, Poughkeepsie, New York/Brandeis University, Waltham, Massachusetts, USA)

Simon Krek (Jo         Stefan Institute, University of Ljubljana, Ljubljana, Slovenia)

Antonio Toral (Universitat d'Alacant, Sant Vicent del Raspeig, Spain)

Henk van den Heuvel (Radboud Universiteit Nijmegen, Nijmegen, The Netherlands)

Message from the Workshop Chairs

We are excited to be able to welcome workshops to the Language Resources and Evaluation Conference 2026, incorporating traditionally recurring workshops as well as a number of totally new workshops on contemporary topics for our audience.

In total, we received 55 workshop proposals, of which we were able to accept 49 workshops for the conference (89%). After merging some workshops and two cancellations, the final number of workshops at the conference amounted to 45 plus the Latin American NLP (LANLP) associate event. We are very pleased to have received this high volume of interest in the conference and hope that this level of engagement will continue on site and online, as well as in future editions.

Based on this process, LREC 2026 hosted 1 one-and a half day workshop, 18 full-day workshops and 27 half-day workshops.

We would also like to take this opportunity to thank all the workshop organizers, as well as authors who submitted papers to the workshops, whom we congratulate for accepted papers and presentations in Palma de Mallorca. Finally, we would like to thank the amazing reviewers, who made the conference program what it is, and the various organizing committees, and in particular the General Chair, for their seemingly unending work in bringing order and making a reality out of an incredibly complex and large event.

Luis Chiruzzo, Darja Fišer, Menno van Zaanen

Workshop Chairs

Message from the Tutorial Chairs

In the fast-paced panorama of Language Resources and Technologies (LRT), we have endeavoured to provide the Fifteenth biennial Language Resources and Evaluation Conference (LREC-2026) attendees with a broad selection of tutorials covering both foundational and "hot" topics in the field. The resulting selection addresses established areas and cutting-edge research, as well as topics in adjacent research domains. We received 18 submissions, of which 12 were selected for presentation at the conference.

Tutorials related to Large Language Models (LLMs) are now an established presence. These are explored through diverse perspectives, including linguistic creativity, the use of LLMs as annotators, and the ongoing challenges of hallucination and bias. In line with the conference's commitment to diversity, several tutorials specifically address low-resource languages, focusing on inclusive corpus building and the deployment of multilingual and multimodal models "in the wild."

At the same time, several other proposals provide opportunities for a deep dive into formal and foundational topics, such as distributed compositionality, meaning representation, and metaphor processing. The program is further enriched by specialized sessions on Healthcare NLP, terminology-augmented generation, and a retrospective look at the 26-year journey of benchmark datasets.

Ultimately, we believe this offer is well-balanced, maintaining a strong emphasis on core LREC themes and placing practical training on language resource creation and evaluation at its very heart.

Francesca Frontini and Atul Kr. Ojha
LREC-2026 Tutorial Chairs

Message from the Industry Day Chairs

As Chairs of this year's Industry Day, we are delighted to bring together students, professionals, researchers, and innovators from across academia and industry to exchange ideas, showcase advancements in NLP and resource management, and foster meaningful collaborations. The purpose of this event is to bridge the gap between our ever-advancing fields of research and real-world applications, creating a platform where insights can move beyond theory into practice.

Our program has been designed to encourage both meaningful and diverse interactions with a line-up of 19 speakers who are ready to inspire us through a series of engaging talks, interactive poster presentations, and a thought-provoking panel session. Their aim is to highlight emerging trends, share practical experiences, and stimulate discussion on current challenges and opportunities within our field.

We anticipate that this year's Industry Day will serve not only as a forum for knowledge sharing but also as a springboard for new partnerships and future innovations. We encourage all participants - students, researchers, industry players and innovators - to actively engage, ask questions, and lean into the networking and learning opportunities throughout the day.

We look forward to a vibrant, insightful, and memorable experience for all.

Teresa Lynn and Natalie Schluter
Industry Day Chairs

Message from the Local Organisation Chairs

We are delighted to welcome the international communities of **LREC 2026** to the beautiful city of **Palma** in the isle of Mallorca. This journey began in December 2024, and while the path has been filled with unique challenges and uncertainty, it has also given us the opportunity to weave an extraordinary network of collaborators, colleagues, and friends.

Palma is a city that perfectly captures the spirit of the Mediterranean—historically rich, culturally vibrant, and architecturally stunning. As the capital of the Balearic Islands, it offers a unique blend of traditional charm and modern innovation, where *Mallorquí*, as the local Catalan variety is popularly known, thrives in a multilingual environment that serves as an ideal hub for our gathering. From its lively neighborhoods to its world-renowned regional food, the city provides a sober yet charming backdrop for our scientific and human exchange.

The Spanish scientific community has always been a strong presence among the participants and supporters of the LREC events. After Granada in 1998 and Las Palmas de Gran Canaria in 2002 this is the third time our community is organizing an LREC conference. We would like to specifically thank the Spanish Society for Natural Language Processing (SEPLN) for their unconditional support in bringing this conference to life. This community has supported our organizational work since the beginning, serving as chairs, workshop organizers, volunteers, and participants.

The organizing committee has held weekly meetings for more than a year, including in-person sessions with providers and tours of the venues, as well as many other behind-the-scenes efforts that sustain an event of this scale. We would like to thank the General Chair Stelios Piperidis and the ELRA Secretary General Khalid Choukri for keeping everything in line and running such a complicated machine like clockwork. We extend our thank you to Helène Mazo, Sara Goggi and Simon Krek and to the all LREC2026 committees, for their professional and fruitful collaboration.

Regardless of the many days and nights spent thinking about solutions to every issue, we could never have supported an event of this complexity by ourselves. We were accompanied by our professional conference organizers UIB Congr  s, and in particular Eva Calvo and Ver  nica Lillo. In addition, we were supported by the University of the Balearic Island (UIB), the Palma city council, el Consell de Mallorca and the Balearic government, helping us connect participants not only with the conference, but also with the place that hosts it.

Finally, our colleagues and students from UIB heeded the call, forming the contingent of orange-shirted volunteers who ensure the conference goes smoothly and pleasantly for everyone involved. It is often through their quiet dedication that organization becomes hospitality and effort becomes a fruitful experience. To them goes our biggest thank you.

Rafael Mu  oz Guillena

German Rigau Claramunt

Melania S. Masi  

Maria del Mar Vanrell Bosch

Acknowledgments The ELRA Language Resources Association and LREC 2026 Committees acknowledge with gratitude the support and sponsoring of the following institutions:

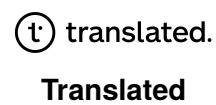
GOLD



SILVER



BRONZE



SUPPORTERS



Organizing Committees

General Chair

- Stelios Piperidis, Athena Research Center/ILSP, Athens, Greece

Program Committee Chairs

- Núria Bel, Pompeu Fabra University, Barcelona, Spain
- Henk van den Heuvel, Radboud University, Nijmegen, Netherlands
- Nancy Ide, Vassar College, Poughkeepsie, New York / Brandeis University, Waltham, Massachusetts, USA
- Simon Krek, Jožef Stefan Institute / University of Ljubljana, Ljubljana, Slovenia
- Antonio Toral, Universitat d'Alacant, Spain

Workshop Chairs

- Luis Chiruzzo, Universidad de la República, Uruguay
- Darja Fišer, CLARIN ERIC, The Netherlands
- Menno van Zaanen, North-West University, South Africa

Tutorial Chairs

- Francesca Frontini, ILC-CNR, Italy
- Atul Kumar Ojha, University of Galway, Ireland

Industry Track Chairs

- Teresa Lynn, Mohamed bin Zayed University of Artificial Intelligence, UAE
- Natalie Schluter, Apple

Management Committee Chairs

- Khalid Choukri, ELDA/ELRA, Paris, France
- Sara Goggi, ILC-CNR, Italy
- Hélène Mazo, ELDA/ELRA, Paris, France

Local Organisation Committee Chairs

- German Rigau, HITZ, University of the Basque Country, Spain
- Maria del Mar Vanrell, Universitat de les Illes Balears, Spain
- Melania Sánchez Masià, Universitat de les Illes Balears, Spain

- Rafael Muñoz Guillena, Universidad de Alicante, Sociedad Española Procesamiento del Lenguaje Natural, Spain

Publication Chairs

- John McCrae, University of Galway, Ireland
- Gilles Sérasset, Université Grenoble Alpes, France

Sponsorship Chairs

- Patrick Paroubek, Université Paris-Saclay, France
- Alexandros Poulis, Transperfect/Dataforce, Luxembourg
- Jimmy Kunzmann, Retired Amazon Alexa Hybrid Science
- Raquel Lázaro Gutiérrez, Universidad de Alcalá

Exhibition Chair

- Rafael Muñoz Guillena, Universidad de Alicante, Sociedad Española Procesamiento del Lenguaje Natural, Spain

Publicity Chairs

- Elena Volodina, University of Gothenburg, Sweden
- Rafael Valencia García, Universidad de Murcia, Spain
- A. Seza Doğruöz, Ghent University, Belgium

Ethics Chairs

- Denise DiPersio, Linguistic Data Consortium, USA
- Thomas Margoni, KU Leuven, Belgium
- Claudia Cevenini, Università di Bologna, Italy

Diversity and Inclusion Chair

- Sakriani Sakti, Nara Institute of Science and Technology, Japan

Advisors

- Nicoletta Calzolari, Advisor, Italy

Senior Area Chairs

Applications Involving LRs and Evaluation for any area/domain of language and speech processing

- Sophia Ananiadou (University of Manchester, United Kingdom)
- José M. Gómez Pérez (Expert.AI, Spain)

Bias, Offensive and Non-inclusive Language; Guardrails, filters

- Tommaso Caselli (Rijksuniversiteit Groningen, The Netherlands)
- Flor Miriam Plaza Del Arco (Leiden University, The Netherlands)

Corpora, Treebanks and Annotation; Tools, Systems and Platforms

- Amir Zeldes (Georgetown University, USA)
- Ines Rehbein (University of Münster/U-Mannheim, Germany)
- Deniz Zeyrek (Middle East Technical University, Turkey)
- Sandra Kuebler (Indiana University, USA)

Dialogue, Conversational Systems, Chatbots, Human-Robot Interaction

- Zoraida Callejas (University of Granada, Spain)
- Kallirroi Georgila (University of Southern California (USC), USA)

Digital Humanities, Cultural Heritage and Computational Social Science

- Petya Osenova (Sofia University "St. Kl Ohridski" and IICT-BAS, Bulgaria)
- Alexander Mehler (Goethe University Frankfurt, Germany)

Discourse and Pragmatics

- Massimo Poesio (Queen Mary Uni, United Kingdom)
- Yulia Grishina (Amazon, Germany)

Document Classification, Information Retrieval and Cross-lingual Retrieval

- William Hersh (Oregon Health & Science University, USA)
- Gareth J. F. Jones (Dublin City University, Ireland)
- Kirk Roberts (UTHealth Houston, USA)

Evaluation, Validation, Quality Assurance and Benchmarking Methodologies

- Elena Cabrio (Université Côte d’Azur, Inria, CNRS, I3S, France)
- Paolo Rosso (Univeridad Politécnica de Valencia, Spain)

Inference, Reasoning, Question Answering

- Yiming Cui (iFLYTEK Research, China)
- Yao Lu (University College London, United Kingdom)

Information Extraction and Text Mining

- Mark Finlayson (Florida International University, USA)
- Diana Inkpen (Dublin City University, Ireland)

Interpretability/explainability of language models and language and speech processing tools

- Fabio Massimo Zanzotto (Uni Roma 2, Italy)
- Giuseppe Attanasio (Instituto de Telecomunicações, Portugal)

Knowledge discovery/representation (knowledge graphs, linked data, terminologies, lexicons, ontologies, etc.)

- Paul Buitelaar, Galway (University of Galway, Ireland)
- Jorge Gracia (University of Zaragoza, Spain)

Language Modeling and LRs (including training, fine-tuning, representation learning, and generation of synthetic data)

- Lucia Passaro (Uni Pisa, Italy)
- Suzan Verberne (Universiteit Leiden, The Netherlands)

Less-Resourced/Endangered/Less-studied Languages

- David Ifeoluwa Adelani (McGill University / MILA, Canada)
- Antonios Anastasopoulos (George Mason University and Archimedes, Athena RC, USA)

Lexicon and Semantics

- Gerardo Sierra (UNAM, Mexico)
- Martha Lewis (University of Amsterdam, The Netherlands)
- Annemarie Friedrich (University of Augsburg, Germany)

Machine Learning Methods and Techniques for Language and Speech Processing, including efficient/low-resource methods

- Vincent Vandeghinste (KU Leuven / INT, Belgium)
- Bram Vanroy (KU Leuven / INT, Belgium)

Multilinguality, Machine Translation (including Speech-to-Speech) and Translation Aids

- Taro Watanabe (Nara Institute of Science and Technology, Japan)
- Houda Bouamor (Carnegie Mellon University in Qatar, Qatar)

Multimodality, Cross-modality (including Sign Languages, Vision and Other Modalities), Multimodal Applications, Grounded Language Acquisition

- Sarah Ebling (University of Zurich, Switzerland)
- Sriparna Saha (Indian Institute of Technology Patna, India)

Natural Language Generation and Summarization

- Marko Robnik Šikonja (University of Ljubljana, Slovenia)
- Antoine Doucet (University of La Rochelle, France)

Simplification, Plain Language and Assistive Technologies

- Horacio Saggion (Universitat Pompeu Fabra, Spain)
- Anna Matamala (Universitat Autònoma de Barcelona, Spain)

Opinion & Argument Mining, Sentiment Analysis, Emotion Recognition/Generation

- Luis Espinosa Anke (Cardiff University, United Kingdom)
- Haris Papageorgiou (ILSP/Athena RC, Greece)

Parsing, Tagging, Chunking, Grammar, Syntax, Morphosyntax, Morphology

- Magda Ševčíková (UFAL, Czechia)
- Parameswari Krishnamurthy (NITI, Hyderabad, India)

Policy and Legal Issues (including Language Resource Infrastructures, Interoperability, Standards for LRs, Metadata) - Ethics, Research Reproducibility and Replicability, and Environmental Issues

- Catia Cucchiari (DLU/ Radboud University, The Netherlands)

- Tommi Jauhiainen (U Helsinki, Finland)

Psycholinguistics, Cognitive Linguistics and Linguistic Theories

- Rui Mao (Nanyang Technological University, Singapore)
- Tan Yue (Peking University, China)

Social Media Processing

- Costanza Navarretta (UCPH, Denmark)
- Jose Camacho-Collados (Cardiff University, United Kingdom)

Speech Resources and Processing (including Phonetic Databases, Phonology, Prosody, Speech Recognition, Synthesis and Spoken Language Understanding)

- Christoph Draxler (LMU Munich, Germany)
- Isabel Trancoso (INESC-ID, Portugal)

Area Chairs

Listed in alphabetical order by Surname

[A] Gavin Abercrombie (Heriot Watt University, United Kingdom), Haithem Afli (ADAPT Centre, Munster Technological University, Ireland), Rodrigo Agerri (HiTZ Center - Ixa, University of the Basque Country EHU, Spain), Sina Ahmadi (University of Zurich, Switzerland), Jesujoba Alabi (Saarland University, Germany), Mehwish Alam (Telecom Paris, Institut Polytechnique de Paris, France), Hanan Aldarmaki (MBZUAI, United Arab Emirates), Hassan Alhuzali (Umm Al-Qura University, Saudi Arabia), Mohammad Aliannejadi (University of Amsterdam, Netherlands), Malihe Alikhani (Northeastern University, United States), Fernando Alva-Manchego (Cardiff University, United Kingdom), Dimosthenis Antypas (Cardiff University, United Kingdom), Catherine Arnett (EleutherAI, United States), Eleftherios Avramidis (Alangu; German Research Center for Artificial Intelligence (DFKI), Germany), Jeremy Barnes (University of the Basque Country EHU/UPV, Spain)

[B] Valerio Basile (University of Turin, Italy), Riza Batista-Navarro (Department of Computer Science, The University of Manchester, United Kingdom), Jonas Beskow (KTH Speech, music and hearing, Sweden), Jelke Bloem (University of Amsterdam, Netherlands), Alessandro Bondielli (University of Pisa, Italy), Giovanni Bonetta (FBK, Italy), Nadav Borenstein (Department of Computer Science, University of Copenhagen, Denmark), Claudia Borg (University of Malta, Malta), Emanuela Boros (University of La Rochelle, France), Fethi Bougares (LIUM- Le Mans Université, France), Zied Bouraoui (CRIL-CNRS & University of Artois, France), Richard Bowden (University of Surrey, United Kingdom), Chloé Braud (IRIT, CNRS, France)

[C] Pablo Calleja (Universidad Politécnica de Madrid, Spain), Rémi Cardon (HULAT - Universidad Carlos III de Madrid, Spain), Paula Carvalho (University of Aveiro, Portugal), Davide Ceolin (Centrum Wiskunde & Informatica, Netherlands), Amanda Cercas Curry (Bocconi University, Italy), Shubham Chatterjee (Missouri University of Science and Technology, United States),

Chung-Chi Chen (National Institute of Advanced Industrial Science and Technology, Japan), Guanyi Chen (Central China Normal University, China), Christian Chiarcos (Goethe-Universität Frankfurt am Main, Germany), Shammur Absar Chowdhury (Qatar Computing Research Institute, Qatar), Christos Christodouloupoulos (Information Commissioner's Office, United Kingdom), Çağrı Çöltekin (University of Tübingen, Germany), Danish Contractor (IBM Research IBM Research, United States)

[D] Daniel Dakota (Indiana University, United States), Sandipan Dandapat (IIT Hyderabad, India), Iria de-Dios-Flores (Universitat Pompeu Fabra, Spain), Arantza del Pozo (Vicomtech, Spain), Yuntian Deng (University of Waterloo, Canada), Lorenzo Desideri (Sigmund Freud University, Italy), Luis Fernando D'Haro (Speech Technology and Machine Learning Group, E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid, Spain), Milan Dojchinovski (FIT, CTU in Prague / InfAI, Germany, Czech Republic), Lucia Donatelli (Vrije Universiteit Amsterdam, Netherlands), Antoine Doucet (University of La Rochelle, France), Yupei Du (Utrecht University, Netherlands)

[E] Mo El-Haj (VinUniversity, Viet Nam), Liana Ermakova (HCTI EA-4249, Université de Bretagne Occidentale, France), Yannick Estève (LIA - Avignon Université, France), Thierry Etchegoyhen (Vicomtech, Spain)

[F] Yansong Feng (Peking University, China), Elisabetta Fersini (University of Milano-Bicocca, Italy), Giorgos Filandrianos (National Technical University of Athens, Greece), Thomas François (UCLouvain, CENTAL, Belgium), Stella Frank (Technical University of Denmark (DTU), Denmark)

[G] Petra Galuscakova (University of Stavanger, Norway), Alborz Geramifard (Linkedin, United States), Luke Gessler (Indiana University Bloomington, United States), Anastasia Giachanou (Utrecht University, Netherlands), Chen Gong (Soochow University, China), Carlos-Emiliano Gonzalez-Gallardo (LIFAT, Université de Tours, France), Kartik Goyal (Georgia Institute of Technology, United States), Deepak Gupta (National Library of Medicine, NIH, United States)

[H] Nizar Habash (New York University Abu Dhabi, United Arab Emirates), Xianpei Han (Institute of Software, Chinese Academy of Sciences, China), Annette Hautli-Janisz (University of Passau, Germany), Kai He (Saw Swee Hock School of Public Health, Singapore), Benjamin Heinzerling (RIKEN AIP & Tohoku University, Japan), Ryuichiro Higashinaka (Nagoya University/NTT, Japan), Hai Hu (City University of Hong Kong, China), Jimin Huang (The Fin AI, United States), Patrick Huber (University of British Columbia, Canada)

[I] Koji Inoue (Kyoto University, Japan), Federica Iurescia (Università Cattolica del Sacro Cuore Milano, Italy)

[J] Tianyu Jiang (University of Cincinnati, United States), Kristiina Jokinen (AIRC, AIST and University of Helsinki, Finland)

[K] David Kauchak (Pomona College, United States), Tatsuya Kawahara (Kyoto University, Japan), Casey Kennington (Boise State University, United States), Fazel Keshtkar (St. John's University, United States), Erik Ketzan (King's College London, United Kingdom), Felicitas Kleber (Department of Language Science and Technology, Saarland University, Germany), Katarzyna Klessa (Adam Mickiewicz University in Poznań, Poland), Elena Kochkina (JPMorgan Chase, United Kingdom), Kazunori Komatani (University of Osaka, Japan)

[L] Wai Lam (The Chinese University of Hong Kong, Hong Kong Special Administrative Region of China), Ivano Lauriola (Amazon AGI, United States), John Lee (City University of Hong Kong, Hong Kong Special Administrative Region of China), Annie Lee (Ontario Tech University, University of Toronto, Canada), Els Lefever (LT3, Ghent University, Belgium), Spyretta Leivaditi

(Unicorn Labs B.V., Netherlands), Gaël Lejeune (STIH, Sorbonne Université, France), Junhui Li (Soochow University, Suzhou, China), Jundong Li (University of Virginia, United States), Jindřch Libovický (Charles University, Czech Republic), Qika Lin (National University of Singapore, Singapore), Pierre Lison (Norwegian Computing Centre, Norway), Zhiwei Liu (University of Manchester, United Kingdom), Zhengyuan Liu (Agency for Science, Technology and Research (A*STAR), Singapore, Singapore), XIAO LUWEI (National University of Singapore, Singapore)

[M] Mounica Maddela (Bloomberg, United States), Pranava Madhyastha (City, University of London, United Kingdom), Ramesh Manuvinakurike (Intel labs, United States), Stella Markantonatou (ILSP/ATHENA RESEARCH CENTER, Greece), Ilia Markov (Vrije Universiteit Amsterdam, CLTL, Netherlands), Raquel Martínez (Universidad Nacional de Educación a Distancia (UNED), Spain), Katerina Mavrou (European University Cyprus, Cyprus), Xin Mei (Northwestern Polytechnical University, China), Maite Melero (AI Institute - Barcelona Supercomputing Center, Spain), Amália Mendes (University of Lisbon - Centre of Linguistics, School of Arts and Humanities, Portugal), Florian Metze (Meta Platforms Inc, United States), Alessio Miaschi (Institute for Computational Linguistics "A. Zampolli" (CNR-ILC), Pisa, Italy), Timothy Miller (Boston Children's Hospital and Harvard Medical School, United States), Jelena Mitrović (University of Passau, Germany), Jose G. Moreno (Paul Sabatier University - IRIT, France), Lourdes Moreno (Universidad Carlos III de Madrid, Spain), Yasmin Moslem (ADAPT Centre, Trinity College Dublin, Ireland), Josiane Mothe (Univ. Toulouse, INSPE UT2J, IRIT, CNRS CLLE, France), Shamsuddeen Hassan Muhammad (Bayero University, Kano, Nigeria)

[N] Mikio Nakano (C4A Research Institute, Inc./Univ. of Osaka/Nagoya Univ./Nagoya Inst. of Tech., Japan), Usman Naseem (University of Sydney, Australia), Anna Nedoluzhko (Charles University in Prague, Czech Republic), Isar Nejadgholi (National Research Council Canada, Canada), Debora Nozza (Bocconi University, Italy)

[O] Maciej Ogródniczuk (Institute of Computer Science, Polish Academy of Sciences, Poland), Antoni Oliver (Universitat Oberta de Catalunya, Spain), Anais Ollagnier (Université Côte d'Azur, Inria, CNRS, I3S, France), Juri Opitz (University of Zurich, Switzerland), Juan Orozco-Arroyave (Universidad de Antioquia, Colombia), John E. Ortega (Northeastern University, United States), Nedjma Ousidhoum (Cardiff University, United Kingdom), Saziye Ozates (Bogazici University, Turkey)

[P] Alexis Palmer (University of Colorado Boulder, United States), Natalie Parde (University of Illinois Chicago, United States), Panupong Pasupat (Google, United States), Carmen Peláez-Moreno (Universidad Carlos III Madrid, Spain), Siyao Peng (LMU Munich, Germany), Gabriele Pergola (University of Warwick, United Kingdom), Davide Picca (University of Lausanne, Switzerland), Flammie Pirinen (UiT Norgga Ártalaš universitehta, Norway), Lucie Polakova (Charles University, Czech Republic), Simone Paolo Ponzetto (University of Mannheim, Germany), Daniel Preotiuc-Pietro (Bloomberg, United States), Prokopis Prokopidis (ILSP/Athena RC, Greece), Piotr Przybyła (Universitat Pompeu Fabra, Spain)

[Q] Chatrine Qwaider (MBZUAI, United Arab Emirates)

[R] Heri Ramampiaro (Norwegian University of Science and Technology (NTNU), Norway), Leonardo Ranaldi (University of Edinburgh, United Kingdom), Priya Rani (Data Science Institute, University of Galway, Ireland), Kirk Roberts (University of Texas Health Science Center at Houston, United States), Lina M. Rojas Barahona (Orange Innovation Research, France), Elena Sofia Ruzzetti (University of Rome Tor Vergata, Italy)

[S] Tulika Saha (IIIT Bangalore, India), Hassan Sajjad (Dalhousie University, Canada), Keisuke Sakaguchi (Tohoku University / RIKEN, Japan), Felipe Sanchez-Martínez (Universitat d'Alacant, Spain), Kengatharaiyer Sarveswaran (University of Jaffna, Sri Lanka), Beatrice Savoldi (Fon-

dazione Bruno Kessler, Italy), Viktor Schlegel (Imperial College London, Singapore), Hanna Schmueck (University of Augsburg, Germany), Nathan Schneider (Georgetown University, United States), Annika Marie Schoene (Northeastern University, United States), Indira Sen (RWTH Aachen, Germany), Rico Sennrich (University of Zurich, Switzerland), Matthew Shardlow (Manchester Metropolitan University, United Kingdom), Shuju Shi (Indiana University Bloomington, United States), Alexander Shvets (Barcelona Supercomputing Center, Spain), Lucia Siciliani (University of Bari Aldo Moro, Italy), Edwin Simpson (University of Bristol, United Kingdom), Blerina Spahiu (University of Milan - Bicocca, Italy), Caroline Sporleder (Goettingen University, Germany), Joe Stacey (University of Sheffield, United Kingdom), David Stap (NXAI, Netherlands), Manfred Stede (University of Potsdam, Germany), Svetlana Stoyanchev (Toshiba Europe, United Kingdom), Saku Sugawara (National Institute of Informatics, Japan), Torbjorn Svendsen (NTNU, Norway)

[T] Raphael Tang (UCL, United Kingdom), Mariona Taule (University of Barcelona, Spain), Zhiyang Teng (Bytedance, Singapore), Arda Tezcan (LT3, Language and Translation Technology Team, Ghent University, Belgium), Carole Tiberius (Instituut voor de Nederlandse Taal, Netherlands), Nadi Tomeh (LIPN-CNRS, Université Sorbonne Paris Nord, France), Sara Tonelli (Fondazione Bruno Kessler, Italy), Maria Ines Torres (Universidad del Pais Vasco UPV/EHU, Spain), Pedro Torres-Carrasquillo (MIT Lincoln Laboratory, United States), David Traum (University of Southern California Institute for Creative Technologies, United States)

[V] Sowmya Vajjala (National Research Council, Canada), Michiel van der Meer (Leiden Institute of Advanced Computer Science, Netherlands), Daan van Esch (Google Research, Netherlands), Menno van Zaanen (South African Centre for Digital Language Resources, South Africa), Vasudha Varadarajan (Carnegie Mellon University, United States), Giulia Venturi (Institute of Computational Linguistics "Antonio Zampolli" (ILC-CNR), Italy), Serena Villata (Université Côte d'Azur, CNRS, Inria, I3S, France), Svitlana Volkova (Aptima Inc., United States), Ekaterina Vylomova (University of Melbourne, Australia)

[W] Abigail Walsh (ADAPT Centre / Dublin City University, Ireland), Jiaan Wang (WeChat AI, Tencent, China), Xiaozhi Wang (Tsinghua University, China), Tao Wang (Haihe Laboratory of Brain-Computer Interaction and Human-Machine Integration, China), Haochen Wang (Nantong University, China), Nigel Ward (University of Texas at El Paso, United States), Leonie Weissweiler (Leipzig University, Germany), Michael Wiegand (University of Vienna, Austria), Gijs Wijnholds (Leiden Institute of Advanced Computer Science, Netherlands), Rodrigo Wilkens (University of Exeter, United Kingdom), Andreas Witt (Leibniz Institute for the German Language, Germany), Shengqiong Wu (National University of Singapore, Singapore), Jialun Wu (Northwestern Polytechnical University, China)

[X] Qianqian Xie (Wuhan University, China)

[Y] Kailai Yang (University of Manchester, United Kingdom), Seid Muhie Yimam (University of Hamburg, Germany), Zhangyue Yin (Fudan University, China), Koichiro Yoshino (Institute of Science Tokyo / RIKEN GRP / Nara Institute of Science and Technology, Japan), Juntao Yu (Queen Mary University of London, United Kingdom), Dian Yu (Tencent AI Lab, United States)

[Z] Daniel Zeman (Charles University, Faculty of Mathematics and Physics, Czech Republic), Qi Zeng (Meta, United States), Chen Zhang (Peking University, China), Dong Zhang (Soochow University, China), Yuji Zhang (University of Illinois Urbana Champaign, United States), Meishan Zhang (Harbin Institute of Technology (Shenzhen), China, China), Xulang Zhang (Nanyang Technological University, Singapore), Luyao Zhu (National University of Singapore, Singapore), Heike Zinsmeister (University of Hamburg, Germany), Imed Zitouni (Meta, United States), Arkaitz Zubiaha (Queen Mary University of London, United Kingdom)

Scientific Committee

Listed in alphabetical order by Surname

[A] Ayomide Abayomi-Alli (IRIT, France), Mourad Abbas (University of Science and Technology Houari Boumediene, Algeria), Slim Abdennadher (German University in Cairo, Egypt), Asad Abdi (University of Derby, United Kingdom), Muhammad Abdo (Indiana University Bloomington, United States), Idris Abdulmumin (University of Pretoria, South Africa), Anne Abeille (LLF Université Paris Cite, France), Giuseppe Abrami (Goethe University Frankfurt, Germany), Halidandmu Abudukelimu (Xinjiang University of Finance & Economics, China), Lasha Abzianidze (Utrecht University, Netherlands), Cengiz Acarturk (Cognitive Science Department, Jagiellonian University, Poland), Judit Acs (HUN-REN Institute for Computer Science and Control, Hungary), Manish Adawadkar (Nike, United States), Eleri Aedmaa (Institute of the Estonian Language, Estonia), Stergos Afantenos (IRIT, Université de Toulouse, CNRS, Toulouse INP, Toulouse, France, France), Kaustubh Agarwal (Amazon, United States), Favour Aghaebe (University of Sheffield, United Kingdom), Zeljko Agic (Neo4j, Inc., Denmark), Manex Agirrezabal (University of Copenhagen, Denmark), Sheshansh Agrawal (Contextual AI, United States), Shyam Agrawal (KIIT College of Engineering, Gurgaon, Advisor, COE-AI- IGDTUW, Delhi, India), Lourdes Aguilar (UAB-Universitat Autònoma de Barcelona, Spain), Kathleen Ahrens (The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Arafat Ahsan (IIIT Hyderabad, India), Sanchit Ahuja (Northeastern University, United States), Salah Ait-Mokhtar (Naver Labs Europe, France), Simbiat Ajao (Masakhane | Business School of AI, Nigeria), Syed Sarfaraz Akhtar (Apple Inc, United States), Berfin Aktas (University of Technology Nuremberg, Germany), Khalid Al Khatib (Groningen University, Netherlands), Meesum Alam (Indiana University Bloomington, United States), Firoj Alam (Qatar Computing Research Institute, HBKU, Qatar), Itziar Aldabe (HiTZ Center - Ixa, University of the Basque Country (UPV/EHU), Spain), Izaskun Aldezabal (University of the Basque Country, Spain), Jan Alexandersson (DFKI GmbH, Germany), David Alfter (Gothenburg University, Sweden), Bashar Alhafni (MBZUAI, United Arab Emirates), Abdulaziz Alhamadani (Florida Polytechnic University, United States), Iqra Ali (Queen Mary University of London, United Kingdom), Ahmed Ali (HUMAIN, Saudi Arabia), Elham Alighardash (SJAU university, Iran (Islamic Republic of)), Anna Aljanaki (University of Tartu, Estonia), Hend Al-Khalifa (King Saud University, Saudi Arabia), Sakhar Alkhereyf (HUMAIN, Saudi Arabia), Jon Alkorta (University of the Basque Country, Spain), Emily Allaway (University of Edinburgh, United Kingdom), Miguel A. Alonso (Universidade da Coruña, Spain), Laura Alonso Alemany (Universidad Nacional de Córdoba, Argentina), Tanel Alumäe (Tallinn University of Technology, Estonia), Ana Alves (CISUC - University of Coimbra and Polytechnic Institute of Coimbra, Portugal), Patricia Amaral (Indiana University, United States), Raquel Amaro (CLUNL – Research Center for Linguistics at NOVA University Lisbon, Portugal), Maxime Amblard (Université de Lorraine, France), Jacopo Amidei (Universitat Oberta Catalunya (UOC), Spain), Saadullah Amin (Saarland University, Germany), Maaz Amjad (Texas Tech University, United States), Evelin Amorim (Porto University, Portugal), Bo An (The Institute of Ethnology and Anthropology, Chinese Academy of Social Sciences, China), Jisun An (Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, United States), Sophia Ananiadou (University of Manchester, United Kingdom), Dimitra Anastasiou (Luxembourg Institute of Science and Technology, Luxembourg), Rafael Anchieta (Federal Institute of Maranhão, Brazil), Mark Anderson (Norsk Regnesentral, Norway), Krasimir Angelov (University of Gothenburg and Chalmers University of Technology, Sweden), Miriam Anshütz (Technical University of Munich, Germany), Dimosthenis Antypas (Cardiff University, United Kingdom), Isuri Anuradha (Lancaster University, United Kingdom), Emilia Apostolova (Language.ai, United States), Angelina Aquino (Charles Darwin University, Australia), Nora Aranberri (University of the Basque Country (UPV/EHU), Spain), Oscar Araque (Universidad Politécnica de Madrid, Spain), Spela Arhar Holdt (University of Ljubljana (Faculty of Arts & Faculty of Computer and Information Science), Slovenia), Catherine Arnett (EleutherAI,

United States), Antti Arppe (University of Alberta, Canada), Ekaterina Artemova (Toloka.AI, Germany), Ron Artstein (USC Institute for Creative Technologies, United States), Lalaram Arya (IIT Dharwad, India), Masayuki Asahara (National Institute for Japanese Language and Linguistics, Japan), Lauriane Aufrant (Inria, France), Javier Aula-Blasco (Barcelona Supercomputing Center, Spain), Serena Auriemma (University of Pisa, Italy), Mathieu Avanzi (Université de Neuchâtel, Switzerland), Tania Avgustinova (Saarland University, Germany), Eleftherios Avramidis (Alangu; German Research Center for Artificial Intelligence (DFKI), Germany), Abinew Ali Ayele (Bahir Dar University, Ethiopia), Israel Abebe Azime (Saarland University, Germany), Tara Azin (Carleton University, Canada), Normaziah Aziz (Independent, Malaysia)

[B] Premjith B (Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India, India), Senthil Kumar B (Velammal Institute of Technology, India), Bogdan Babych (Heidelberg University, Germany, Germany), Carlos Badenes-Olmedo (Universidad Politécnica de Madrid, Spain), Sher Badshah (Dalhousie University, Canada), Adriana Badulescu (Dallas College, United States), Selene Baez Santamaria (University of Zurich, Switzerland), Mevlüt Bağcı (Goethe University Frankfurt, Department of Computer Science and Mathematics, Text Technology Lab, Frankfurt, Germany, Germany), Robert Bagheri (Department of Methodology and Statistics, Utrecht University, Netherlands), Parnia Bahar (Apple, Germany), Xiaoyu Bai (University of Potsdam, Germany), Smita Bailur (Amazon, United States), Alexandra Baiuk (Pushkin State Russian Language Institute, Hungary), JinYeong Bak (Sungkyunkwan University, Republic of Korea), Amir Bakarov (Vivun Inc., United States), Collin Baker (International Computer Science Institute, United States), Alexandra Balahur (European Commission, Belgium), David Bamutura (Chalmers University of Technology, Sweden and Mbarara University of Science and Technology, Uganda, Uganda), Ritwik Banerjee (Stony Brook University, United States), Piotr Banski (Leibniz-Institut für Deutsche Sprache, Mannheim, Germany), Xiaoyi Bao (The Hong Kong Polytechnic University, China), Jorge Baptista (University of Algarve, Portugal), Denilson Barbosa (University of Alberta, Canada), Eduard Barbu (Institute of Computer Science, Estonia), Verginica Barbu Mititelu (Research Institute for Artificial Intelligence, Romanian Academy, Romania), Ipek Baris Schlicht (Universitat Politècnica de València, Spain), Fabian Barteld (Heinrich Heine University Düsseldorf, Germany), Davis Bartels (National Institutes of Health, United States), Hemanta Baruah (Research Scholar, IIT Guwahati, India), Paulis Barzdins (University of Latvia, Institute of Mathematics and Informatics, Latvia), Guntis Barzdins (University of Latvia, Latvia), Erkan Basar (Wageningen University & Research, Netherlands), Pierpaolo Basile (Department of Computer Science, University of Bari Aldo Moro, Italy), Fernando Batista (INESC-ID and ISCTE-IUL, Portugal), Timo Baumann (Ostbayerische Technische Hochschule Regensburg, Germany), Rachel Bawden (Inria, France), Frederic Bechet (Aix Marseille Université - LIS/CNRS, France), Nicolas Béchet (Université Bretagne Sud, CNRS, IRISA, France), Maria Becker (Heidelberg University, Germany), Karin Becker (Informatics Institute -Federal University of Rio Grande do Sul, Brazil), Dorothee Beermann (Norwegian University of Science and Technology, Norway), Marijke Beersmans (KU Leuven, Belgium), Nuria Bel (Universitat Pompeu Fabra, Spain), Andrea Bellandi (Institute for Computational Linguistics - CNR, Italy), Patrice Bellot (Aix-Marseille Université CNRS (LIS), France), Meriem Beloucif (Uppsala University, Sweden), Himanshu Beniwal (Indian Institute of Technology Gandhinagar, India), Martin Benjamin (Kamusi Project International, Switzerland), Vladimir Benko (Slovak Academy of Sciences, L. Stur Institute of Linguistics, Slovakia), Gábor Berend (University Of Szeged, Hungary), Maria Berger (Ruhr University Bochum, Germany), Sabine Bergler (Concordia University, Canada), Bernhard Bermeitinger (University of St.Gallen, Switzerland), Thales Bertaglia (Utrecht University, Netherlands), Lorenzo Bertolini (European Commission Joint Research Centre (JRC), Italy), Laurent Besacier (Naver Labs Europe, France), Romaric Besancon (CEA LIST, France), Akshita Bhagia (Allen Institute for AI, United States), Varad Bhatnagar (Microsoft India, India), Abari Bhattacharya (University of Illinois Chicago, United States), Shuqing Bian (Tencent, China), Giovanni Maria Biancofiore (I.R.C.C.S. Saverio de Bellis, Italy), Hanno Biber (Austrian Academy of Sciences, Austria), Chris

Biemann (Universität Hamburg, Germany), Ann Bies (Linguistic Data Consortium, University of Pennsylvania, United States), Laura Biester (Middlebury College, United States), Madeline Bitner (National Institutes of Health, United States), Philippe Blache (LPL CNRS, France), Tatiana Bladier (Aix-Marseille University, France), Eduardo Blanco (University of Arizona, United States), Verena Blaschke (LMU Munich, Germany), Joanna Blochowiak (University of Zurich, Switzerland), Christian Boitet (Université Grenoble-Alpes (UGA, LIG-GETALP), France), Necva Bölücü (Csiro, Australia), Alessandro Bondielli (University of Pisa, Italy), Claire Bonial (Georgetown University, United States), Mayumi Bono (National Institute of Informatics, Japan), Georgeta Bordea (La Rochelle University, France), Claudia Borg (University of Malta, Malta), Hsuvas Borkakoty (Cardiff University, United Kingdom), Emanuela Boros (University of La Rochelle, France), Johan Bos (University of Groningen, Netherlands), Federico Boschetti (ILC-CNR, Italy), Cristina Bosco (Dipartimento di Informatica - Università di Torino, Italy), Stefan Bott (Universitat Pompe Fabra, Spain), Nadjat Bouayad-Agha (NLP Scientist, Spain), Florian Boudin (Inria, LS2N, Nantes Université, France), Fethi Bougares (LIUM- Le Mans Université, France), Pierrette Bouillon (UNIGE FTI, Switzerland), Gosse Bouma (University of Groningen, Netherlands), Andrey Bout (Yandex, Russian Federation), Karim Bouzoubaa (Mohammed V University in Rabat, Morocco), Marco Braga (Politecnico di Torino, Italy), António Branco (University of Lisbon, Portugal), Pavel Braslavski (HSE University, Russian Federation), Adrian Brasoveanu (Modul Technology GmbH, Austria), Daniel Braun (University of Marburg, Germany), Sofia Brenna (FBK, Unibz, Italy), Timo Breuer (TH Köln - University of Applied Sciences, Germany), Jacqueline Brixey (UW-Madison, United States), Thomas Brovelli (Meyer) (Google LLC, Switzerland), Henrique Brum (Sinch AB, Brazil), Caroline Brun (Naver Labs Europe, France), Dominique Brunato (Institute of Computational Linguistics "A. Zampolli" (ILC-CNR), Pisa, Italy), Marc Brysbaert (Ghent University, Belgium), Chenyang Bu (Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China, China), Ana-Maria Bucur (Università della Svizzera italiana, Switzerland), Alberto Bugarín-Diz (Univ. Santiago de Compostela, Spain), Paul Buitelaar (University of Galway, Ireland), Maja Buljan (Language Technology Group (LTG), University of Oslo, Norway), Harry Bunt (Tilburg University, Netherlands), Aljoscha Burchardt (DFKI, Germany), Susanne Burger (Solventum, United States), Felix Burkhardt (audEERING, Germany), Elisabeth Burr (University of Leipzig, Germany), Davide Buscaldi (LIPN, Université Sorbonne Paris Nord, France), Hendrik Buschmeier (Bielefeld University, Germany), Miriam Butt (University of Konstanz, Germany), Jan Buys (University of Cape Town, South Africa), Happy Buzaaba (Princeton University, United States), Happy Buzaaba (Princeton University, United States)

[C] Melania Cabezas-Garcia (University of Granada, Spain), Luca Cagliero (Politecnico di Torino, Italy), Yi Cai (South China University of Technology, China), Sylvie Calabretto (LIRIS-INSa Lyon, France), Basilio Calderone (Université Toulouse Jean Jaurès & CNRS, France), Hiram Calvo (Center for Computing Research, Mexico), Leonardo Campillos-Llanos (Consejo Superior de Investigaciones Científicas (CSIC, Spanish National Research Council), Spain), Fazli Can (Bilkent University, Turkey), Arnaldo Candido Junior (Sao Paulo State University (UNESP), Brazil), Pascual Cantos (University of Murcia, Spain), Jie Cao (University of Oklahoma, United States), Spencer Caplan (CUNY Graduate Center, United States), Cornelia Caragea (University of Illinois at Chicago, United States), Franco Alberto Cardillo (Istituto di Linguistica Computazionale, Consiglio Nazionale delle Ricerche, Italy), Rémi Cardon (HULAT - Universidad Carlos III de Madrid, Spain), Aida Cardoso (Centro de Linguística da Universidade Nova de Lisboa, Portugal), Michael Carl (Kent State University, United States), Jorge Carrillo-de-Albornoz (UNED, Spain), Lucien Carroll (Kedara, United States), Sara Carvalho (University of Aveiro, Portugal), Helena Caseli (Federal University of São Carlos, Brazil), Pierluigi Cassotti (University of Gothenburg, Sweden), Sheila Castilho (Dublin City University, Ireland), Thiago Castro Ferreira (Federal University of Minas Gerais, Brazil), Brittany Cates (Colorado State University, United States),

Damir Cavar (Indiana University, United States), Alexandru Ceausu (Court of Justice of the European Union, Luxembourg), Giuseppe Celano (Leipzig University, Germany), Fabio Celli (Maggioli Informatica R&D, Italy), Roberto Centeno (UNED, Spain), Dumitru-Clementin Cercel (University Politehnica of Bucharest, Romania), Alberto Cetoli (Private, United Kingdom), Mauro Cettolo (Fondazione Bruno Kessler, Italy), Claudia Cevenini (University of Bologna, Italy), Suchet Chachra (LinkedIn, United States), Bharathi Raja Chakravarthi (University of Galway, Ireland, Ireland), Somnath Chandra (Dept of Electronics & Information Technology , Govt of India, India), Kent Chang (UC Berkeley, United States), Anis Charfi (Carnegie Mellon University in Qatar, Qatar), Stergios Chatzikyriakidis (University of Crete, Greece), Chung-Chi Chen (National Institute of Advanced Industrial Science and Technology, Japan), Chien Hung Chen (Graduate Institute of Networking and Multimedia, National Taiwan University, Taiwan), Guanyi Chen (Central China Normal University, China), Kuang-hua Chen (National Taiwan University, Taiwan), Po-Chun Chen (Department of Computer Science & Information Engineering National Taiwan University, Taiwan), Lin Chen (Engineering Manager, Meta Platform Inc., United States), Hsin-Hsi Chen (National Taiwan University, Taiwan), Xuanang Chen (Institute of Software, Chinese Academy of Sciences, China), Guanhua Chen (Southern University of Science and Technology, China), Xinchu Chen (Fudan University, China), Kehai Chen (School of Computer Science and Technology, Harbin Institute of Technology, China), Miaomiao Cheng (Capital Normal University, China), Mikhail Chernoskotov (Krasovskii Institute of Mathematics and Mechanics, Russian Federation), Emmanuele Chersoni (Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Jonathan Chevelu (Univ Rennes, CNRS, IRISA, France), Davide Chiarella (National Research Council - Institute for Computational Linguistics "A. Zampolli", Italy), Adrian-Gabriel Chifu (Aix-Marseille Université, LIS, France), Jenny Chim (Queen Mary University of London, United Kingdom), Patricia Chiril (University of Chicago, United States), Luis Chiruzzo (Universidad de la Republica, Uruguay), Won Ik Cho (AI Center, Samsung Electronics, Republic of Korea), Eleanor Chodroff (University of Zurich, Switzerland), Seungtaek Choi (Hankuk University of Foreign Studies, Republic of Korea), Hee-Soo Choi (ATILF - LORIA, France), Key-Sun Choi (KAIST, Konyang University, Republic of Korea), Shammur Absar Chowdhury (Qatar Computing Research Institute, Qatar), Tahiya Chowdhury (Computer Science, Colby College, United States), Xiaomin Chu (Soochow University, China), Michal Chudoba (University of Stavanger, Norway), Jayeol Chun (Brandeis University, United States), Young-joo Chung (Rakuten Group, Inc., United States), Cristiano Ciaccio (Istituto di Linguistica Computazionale "A.Zampolli" (CNR-ILC), Italy), Jaka Čibej (University of Ljubljana, Slovenia), Simone Ciciliano (Free University of Bozen-Bolzano, Italy), Philipp Cimiano (Univ. Bielefeld, Germany), Daniela Claro (Federal University of Bahia - Institute of Computing - FORMAS Research Center on Data and NLP, Brazil), Vincent Claveau (AMIAD & CNRS, France), Maximin Coavoux (CNRS, Univ Grenoble Alpes, France), Mariona Coll Ardanuy (Barcelona Supercomputing Center (BSC-CNS), Spain), Jaime Collado-Montañez (University of Jaén, Spain), Çağrı Çöltekin (University of Tübingen, Germany), Gloria Comandini (Istituto Italiano di Studi Germanici, Italy), Elisabet Comelles (Universitat de Barcelona, Spain), Yan Cong (Purdue University, United States), Mathieu Constant (Université de Lorraine, CNRS, ATILF, France), Danish Contractor (IBM Research IBM Research, United States), Paul Cook (University of New Brunswick, Canada), Robin Cooper (University of Gothenburg, Sweden), Bonaventura Coppola (University of Trento, Italy), Anna Corazza (DIETI, Università di Napoli "Federico II", Italy), Francesco Corcoglioniti (Free University of Bozen-Bolzano, Italy), Ricardo Cordoba (Speech Technology Group. Information Processing and Telecommunications Center. Universidad Politécnica de Madrid, Spain), Rute Costa (CLUNL - FCSH - Universidade NOVA de Lisboa, Portugal), Mickael Coustaty (L3i laboratory - La Rochelle Université, France), Marta Cozzini (University of Bologna, Italy), Benoit Crabbé (University of Paris Cité, France), Mathias Creutz (University of Helsinki, Finland), Berthold Cysmann (CNRS, Laboratoire de linguistique formelle, France), Montse Cuadros (Vicomtech, Spain), Daniel Cunliffe (University of South Wales, United Kingdom), Francesco Cutugno (Department of Electrical Engineering and Information

Technologies University "Federico II", Naples, Italy), Vaclav Cvrcek (Czech National Corpus, Charles University in Prague, Czech Republic), Michael Cysouw (Philipps-Universität Marburg, Germany)

[D] Thenmozhi D. (SSN College of Engineering, India), Deborah Dahl (Conversational Technologies, United States), Xiang Dai (CSIRO Data61, Australia), Qin Dai (Tohoku University, Japan), Beatrice Daille (Nantes Université- LS2N, France), Dhairya Dalal (University of Galway, Ireland), Geraldine Damnati (Orange Innovation, France), Greta Damo (Université Côte d'Azur, France), Dana Dannélls (Språkbanken Text, Dept. of Swedish, University of Gothenburg, Sweden), Aswarth Abhilash Dara (Apple, United States), Ankita Das (Indian Institute of Technology Hyderabad, India), Dipankar Das (Jadavpur University, India), Mithun Das (Indian Institute of Technology Kharagpur, India, India), Debopam Das (Abo Akademi University, Finland), Sudeshna Das (Emory University, United States), Mihai Dascalu (National University of Science and Technology POLITEHNICA Bucharest, Romania), Christopher Davis (University of Cambridge, United Kingdom), Brian Davis (Dublin City University, Ireland), Suparna De (University of Surrey, United Kingdom), Luna De Bruyne (CLiPS, University of Antwerp, Belgium), Gaël de Chalendar (CEA, LIST, Paris-Saclay University, France), Orphee De Clercq (LT3, Ghent University, Belgium), Irene De Felice (University of Piemonte Orientale, Italy), Ona de Gibert (University of Helsinki, Finland), Franciska de Jong (CLARIN ERIC/ Utrecht University, Netherlands), Kordula De Kuthy (IWM Tübingen, Germany), Éric de la Clergerie (INRIA, France), Loic De Langhe (Ghent University, Belgium), Ernesto William De Luca (Otto-von-Guericke-University of Magdeburg, Germany), Lorenzo De Mattei (Aptus.AI, Italy), Valeria de Paiva (Topos Institute, United States), Brian de Silva (Amazon, United States), Florian Debaene (Ghent University, Belgium), Jasper Degraeuwe (Ghent University, Belgium), Barsha Deka (3i-Infotech, India), Angelo Mario Del Grosso (Cnr-Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR-ILC), Italy), Arantza del Pozo (Vicomtech, Spain), Jane Arleth Dela Cruz (Radboud University, Netherlands), Julien Delaunay (L3i, France), Agustín Delgado (UNED, Spain), Pietro Dell'Oglio (University of Pisa, Italy), Felice Dell'Orletta (ItaliaNLP Lab @ Institute for Computational Linguistics "Antonio Zampolli", ILC - CNR, Italy), Rodolfo Delmonte (Ca' Foscari University Venice now retired, Italy), Daryna Dementieva (Technical University of Munich, Germany), Yasuharu Den (Graduate School of Humanities, Chiba University, Japan), Naihao Deng (University of Michigan Ann Arbor, United States), Sourabh Deoghare (IIT Bombay, India), Lorenzo Desideri (Sigmund Freud University, Italy), Nina Dethlefs (University of Hull, United Kingdom), Luis Fernando D'Haro (Speech Technology and Machine Learning Group, E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid, Spain), Maria Pia di Buono (University of Naples "L'Orientale", Italy), Luigi Di Caro (University of Turin, Italy), Marion Di Marco (TUM, Germany), Giorgio Maria Di Nunzio (University of Padua, Italy), Alberto Diaz (Universidad Complutense de Madrid, Spain), Antonios Dimakis (National and Kapodistrian University of Athens; Archimedes, Athena Research Center, Greece, Greece), Anna Beatriz Dimas Furtado (University of Galway, Ireland), Tsvetana Dimitrova (Institute for Bulgarian Language, Bulgaria), Yuning Ding (FernUniversität in Hagen, Germany), Liang Ding (Alibaba | The University of Sydney, China), Ly Dinh (University of South Florida, United States), Anca Dinu (University of Bucharest, Romania), Denise DiPersio (Linguistic Data Consortium, United States), Stefanie Dipper (Ruhr University Bochum, Germany), Peter Dirix (Cerence, KU Leuven, Belgium), Amirbek Djanibekov (MBZUAI, United Arab Emirates), Nemanja Djuric (Aurora Innovation, United States), Anna Dmitrieva (Pushkin State Russian Language Institute, Russian Federation), Alphaeus Dmonte (George Mason University, United States), Simon Dobrisek (University of Ljubljana, Faculty of Electrical Engineering, Slovenia), Kaja Dobrovoljc (University of Ljubljana & Jozef Stefan Institute, Slovenia), Kohji Dohsaka (Akita Prefectural University, Japan), Kaja Dolar (Inalco (CREE), France), Miguel Domingo (Universitat Politècnica de València, Spain), Mingwen Dong (Amazon, United States), Minghui Dong (COLIPS, Singapore), Esra Dönmez (University of Stuttgart, Germany), Deborah Dore (Université Côte d'Azur, France), Jakub Dotlacil (Utrecht University, Netherlands), David Doukhan (Institut

national de l'audiovisuel (Ina), France), Christoph Draxler (Institute of Phonetics and Speech Processing, LMU Munich, Germany), Kasper Drazewski (KU Leuven Centre for IT & IP Law, Belgium), Yupei Du (Utrecht University, Netherlands), Yijun Duan (Kyoto Institute of Technology, Japan), Xiangyu Duan (Soochow University, China), Pablo Duboue (Internet Archive Canada, Canada), Fanny Ducel (LISN, Université Paris-Saclay, France), Nicolas Dugue (LIUM, Le Mans Université, France), David Dukić (University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia), Ewan Dunbar (University of Toronto, Canada), Magali Duran (Universidade de São Paulo, Brazil), Sujan Dutta (Rochester Institute of Technology, United States)

[E] Kurt Eberle (Lingenio GmbH, Germany), Abteen Ebrahimi (University of Colorado, Boulder, United States), Thomas Eckart (Saxon Academy of Sciences and Humanities, Germany), Jens Edlund (KTH Speech, Music and Hearing, Sweden), Markus Egg (Humboldt-Universität zu Berlin, Germany), Koji Eguchi (Hiroshima University, Japan), Yo Ehara (Tokyo Gakuji University, Japan), Maud Ehrmann (École Polytechnique Fédérale de Lausanne, Switzerland), Annerose Eichel (University of Stuttgart, Germany), Asif Ekbal (IIT Patna, India), Roxanne El Baff (German Aerospace Center, Germany), Yassine El Kheir (DFKI, Germany), Samhaa R. El-Beltagy (Newgiza University/Nawy Proptech, Egypt), Michael Elhadad (Ben Gurion University, Israel), Mo El-Haj (VinUniversity, Viet Nam), Tilia Ellendorff (University of Zurich, Switzerland), Desmond Elliott (University of Copenhagen, Denmark), Fadhl Eryani (University of Tübingen, Germany), Hugo Jair Escalante (INAOE, Mexico), Daniel Escobar-Grisales (University of Antioquia, Colombia), Carlos Escolano (Universitat Politècnica de Catalunya, Barcelona Supercomputing Center, Spain), Louis Escoufflaire (Massachusetts Institute of Technology, United States), Maria Eskevich (Huygens Instituut, KNAW, Netherlands), Luis Espinosa Anke (Cardiff University, United Kingdom), Miquel Esplà-Gomis (Universitat d'Alacant, Spain), Ainara Estarrona (HiTZ center, University of the Basque Country, Spain), Kilian Evang (Heinrich Heine University Düsseldorf, Germany)

[F] Pamela Faber (University of Granada, Spain), Fahim Faisal (George Mason University, United States), Mauro Falcone (Fondazione Ugo Bordon, Italy), Neele Falk (University of Stuttgart, Germany), Enfa Fane (University Of Arizona, United States), Yimai Fang (Apple, United Kingdom), Liri Fang (University of Illinois Urbana-Champaign, United States), Michael Färber (TU Dresden, Germany), Omar Farooq (Aligarh Muslim University, India), Mireia Farrus (Universitat de Barcelona, Spain), Christian Fäth (Augsburg University, Germany), Mehwish Fatima (National University of Science and Technology (NUST), Pakistan), Benoit Favre (Aix-Marseille University LIS/CNRS, France), Valeria Feltrim (Universidade Estadual de Maringa, Brazil), Zhaoxin Feng (The Hong Kong Polytechnic University, China), Javier Fernandez-Cruz (Universidad de Malaga, Spain), Adriano Ferraresi (University of Bologna, Italy), Giacomo Ferrari (Università Piemonte Orientale, Italy), Arnaud Ferre (INRAE, Paris-Saclay University, France), Vera Ferreira (CIDLeS - Interdisciplinary Centre for Social and Language Documentation, Portugal), Olivier Ferret (CEA-List, France), Oliver Ferschke (Solventum, United States), Elisabetta Fersini (University of Milano-Bicocca, Italy), Carol Figueroa (University of Texas at El Paso, United States), Giorgos Filandrianos (National Technical University of Athens, Greece), Maria Finatto (Universidade Federal do Rio Grande do Sul, Brazil), Marcelo Finger (University of Sao Paulo, Brazil), Mark Finlayson (Florida International University, United States), Beatrice Fiumanò (University of Bologna, Italy), Michael Flor (Educational Testing Service, United States), Antske Fokkens (VU Amsterdam, Netherlands), Alessandro Fonseca (Pratt & Whitney Canada, Canada), Robert Forkel (Max Planck Institute for Evolutionary Anthropology, Germany), Karen Fort (Université de Lorraine / LORIA, France), Meaghan Fowlie (Utrecht University, Netherlands), Enrico Francesconi (IGSG-CNR, Italy), Wellington Franco (Federal University of Ceará, Brazil), Katerina Frantzi (University of the Aegean, Department of Mediterranean Studies: Archaeology, Linguistics, International Relations, Greece), Greta Franzini (Eurac Research, Italy), Corinne Fredouille (LIA, Avignon University, France), Reva Freedman (Northern Illinois University, United States),

Simona Frenda (Heriot-Watt University, United Kingdom), Matteo Frigeri (KU Leuven, Belgium), Leon Froehling (GESIS, Germany), Francesca Frontini (Istituto di Linguistica Computazionale "A. Zampolli" - ILC Consiglio Nazionale delle Ricerche - CNR, Italy), Yoshinari Fujinuma (Cantina Labs, United States), Atsushi Fujita (National Institute of Information and Communications Technology, Japan), Kotaro Funakoshi (Institute of Science Tokyo, Japan), Monica Fürbacher (Independent Researcher, Germany)

[G] Guillaume Gadek (Airbus, France), Gloria Gagliardi (Alma Mater Studiorum - University of Bologna, Italy), Marco Gaido (Fondazione Bruno Kessler, University of Trento, Italy), Dimitrios Galanis (Institute for Language and Speech Processing, Athena Research Center, Greece), Ramiro H. Galvez (Universidad Torcuato Di Tella, Argentina), Pablo Gamallo (CITIUS, University of Santiago de Compostela, Spain), Federica Gamba (Charles University, Czech Republic), Sudeep Gandhe (Google Inc, United States), Suryakanth Gangashetty (KLEF Vaddeswaram, India), Neelamadhav Gantayat (IBM Research, India), Rena Gao (The University of Melbourne, Australia), Yanjun Gao (University of Colorado, United States), Angel Garcia Contreras (Guardian Robot Project, RIKEN, Japan), Leibny Paola Garcia Perera (Johns Hopkins University, United States), Miguel Ángel García-Cumbreras (University of Jaén, Spain), Iker García-Ferrero (Multi-verse Computing, Spain), Muskan Garg (Mayo Clinic, United States), Sarthak Garg (Meta, United States), Andrew Gargett (The Open University, United Kingdom), Juan-María Garrido (National Distance Education University, Spain), Federico Gaspari (Università San Raffaele di Roma, Italy), Albert Gatt (Utrecht University, Netherlands), Lorenzo Gatti (Human Media Interaction, University of Twente, Netherlands), Eric Gaussier (Univ. Grenoble Alpes, France), Tanja Gaustad (Centre for Text Technology (CTeXT), North-West University, South Africa), Maria Gavrilidou (Athena RC/ILSP, Greece), Michaela Geierhos (Universitaet der Bundeswehr Muenchen, Germany), Branislav Gerazov (Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia), Kim Gerdes (Paris-Saclay University, France), Sahar Ghannay (CNRS, LISN, France), Tirthankar Ghosal (Oak Ridge National Laboratory, United States), Soumitra Ghosh (Fondazione Bruno Kessler (FBK), Italy, Italy), Utshab Kumar Ghosh (Missouri University of Science and Technology, United States), Sucheta Ghosh (IWR University of Heidelberg, Germany), Maria Giagkou (ILSP/ATHENA RC, Greece), Dafydd Gibbon (Universität Bielefeld, Germany), Sylvie Gibet (IRISA, France), Daniela Gifu ("Alexandru Ioan Cuza" University of Iasi, Faculty of Computer Science & Romanian Academy - Iasi branch, Institute of Computer Science, Romania), Shlok Gilda (University of Florida, United States), Lee Gillam (Independent Researcher, United Kingdom), Michael Ginn (University of Colorado, United States), Voula Giouli (Aristotle University of Thessaloniki / ILSP, ATHENA RC, Greece), Emiliano Giovannetti (Istituto di Linguistica Computazionale "A. Zampolli" - CNR, Italy), Katerina Gkirtzou (ILSP/Athena Research Center, Greece), Lorraine Goeuriot (LIG, Université Grenoble Alpes, France), Jonas Golde (Humboldt-University of Berlin, Germany), Janis Goldzycher (University of Zurich, Switzerland), Luis Gomes (Faculdade de Ciencias da Universidade de Lisboa, Portugal), Carlos Gómez-Rodríguez (Universidade da Coruña, Spain), Hugo Goncalo Oliveira (CISUC, DEI, University of Coimbra, Portugal), Teresa Goncalves (Department of Informatics, University of Evora, Portugal), Carlos-Emiliano Gonzalez-Gallardo (LIFAT, Université de Tours, France), Jesús González-Rubio (WebInterpreter, Spain), Dhiman Goswami (George Mason University, United States), Isao Goto (Ehime University, Japan), Martijn Goudbeek (Tilburg University, Netherlands), Cyril Goutte (National Research Council Canada, Canada), Abhay Goyal (Google, United States), Natalia Grabar (CNRS STL UMR8163, Université de Lille, France), Arianna Graciotti (University of Bologna, Italy), Nora Graichen (UPF Faculty of Translation and Language Sciences, Spain), Tommaso Green (University of Mannheim, Germany), Nathan Green (Marymount University, United States), Lorenzo Gregori (University of Florence, Italy), David Griol (University of Granada, Spain), Cristina Grisot (University of Zurich, Switzerland), Morgan Grobol (Université Paris Nanterre, France), Cyril Grouin (LISN, France), Normunds Gruzitis (University of Latvia, Latvia), Jinghang Gu (The Hong Kong Polytechnic University, China), Yi Guan (School of Computer

Science and Technology, Harbin Institute of Technology, China), Jon Gudnason (Reykjavik University, Iceland), Anupam Guha (Indian Institute of Technology Bombay, India), Svenja Guhr (University of California at Berkeley, United States), Gael Guibon (Université Sorbonne Paris Nord - LIPN, LORIA, France), Bruno Guillaume (LORIA / Inria Nancy Grand-Est, France), Nuno Guimarães (INESC TEC, Portugal), Zhijin Guo (JP Morgan Chase, United Kingdom), Yinpeng Guo (Hong Kong Exchanges, Hong Kong Special Administrative Region of China), Fengyu Guo (Tianjin Normal University, China), Raghav Gupta (Google Inc., United States), Deepak Gupta (National Library of Medicine, NIH, United States), Rohit Gupta (Apple UK Ltd., United Kingdom), Joakim Gustafson (KTH, Sweden)

[H] Ivan Habernal (Ruhr-University Bochum, Germany), Janiça Hackenbuchner (LT3, Ghent University, Belgium), Udo Hahn (Friedrich-Schiller-Universitaet Jena, Germany), Thomas Haider (University of Passau, Germany), Eva Hajicova (Charles University, Czech Republic), Najeh Hajlaoui (Consultant at European Commission - European Parliament, Luxembourg), Kishaloy Halder (AWS AI Labs, United States), Ahmed Hamdi (IRIT, University of Toulouse, France), Injy Hamed (Mohamed bin Zayed University of Artificial Intelligence, Egypt), Harald Hammarström (Uppsala University, Sweden), Leon Hammerla (Goethe University (Frankfurt), Germany), Thierry Hamon (LISN, Université Paris-Saclay & Université Sorbonne Paris Nord, France), Na-Rae Han (University of Pittsburgh, United States), Kelvin Han (Independent, Singapore), Zhongyuan Han (FOSHAN UNIVERSITY, China), Ting Han (Money Forward, Inc., Japan), Thomas Hanke (University of Hamburg, Germany), Yanbin Hao (Hefei University of Technology, China), Kazuo Hara (Yamagata University, Japan), Sanda Harabagiu (University of Texas at Dallas, United States), Yannis Haralambous (IMT Atlantique & CNRS LabSTICC, France), Fouzi Harrag (Ferhat Abbas University, Algeria), Atticus Harrigan (Independent Scholar, Canada), Koiti Hasida (RIKEN AIP, Japan), Ali Hatami (University of Galway, Ireland), Ashutosh Hathidara (SAP Labs LLC, United States), Nabil Hathout (CLLE, CNRS & Université de Toulouse, France), Hans Ole Hatzel (Universität Hamburg, Germany), Akio Hayakawa (Universitat Pompeu Fabra, Spain), Yoshihiko Hayashi (Waseda University, Japan), Abigail Hayes (University of Mannheim, Germany), Amir Hazem (RCAST, The University of Tokyo, Japan), Guoxiu He (East China Normal University, China), Jianghui He (Kunming University of Science and Technology, China), Kai He (Saw Swee Hock School of Public Health, Singapore), Ruifang He (Tianjin University, China), Ulrich Heid (Universitaet Hildesheim, Germany), Johannes Heinecke (Orange Innovation, France), Benjamin Heinzerling (RIKEN AIP & Tohoku University, Japan), Lars Hellan (NTNU, Norway), Iris Hendrickx (Centre for Language Studies, Radboud University Nijmegen, Netherlands), Leonhard Hennig (German Research Center for Artificial Intelligence (DFKI), Germany), Fabiola Henri (University at Buffalo--SUNY, United States), Julio Hernandez (Dublin City University, Ireland), Chantal Pérez Chantal Hernández (Universidad de Málaga, Spain), Delia Irazu Hernandez Farias (Instituto Nacional De Astrofisica, Optica Y Electronica, Mexico), William Hersh (Oregon Health & Science University, United States), Hansi Hettiarachchi (Lancaster University, United Kingdom), Freya Hewett (University of Potsdam, Germany), Nulette Heyns (North West University, South Africa), Ryuichiro Higashinaka (Nagoya University/NTT, Japan), Derrick Higgins (Illinois Institute of Technology, United States), Tuomo Hiippala (Department of Languages, University of Helsinki, Finland), Mark Hill (King's College London, United Kingdom), Stefan Hillmann (Quality and Usability Lab, Technische Universität Berlin, Germany), Marcelo Yuji Himoro (Universidad Nacional de Educacion a Distancia, Spain), Tatsuya Hiraoka (MBZUAI, United Arab Emirates), Armin Hoenen (Empirical Linguistics, Goethe University, Germany), Eben Holderness (Brandeis University, United States), Takeshi Homma (Micoworks, Inc., Japan), Shi Yin Hong (Nanyang Technological University, Singapore), Mohammed Moshui Hoque (Department of Computer Science & Engineering, Chittagong University of Engineering & Technology, Bangladesh), Helmut Horacek (German Research Center for AI (DFKI), Germany), Ales Horak (Masaryk University, Czech Republic), Andrea Horbach (CAU Kiel / Leibniz Institute for Science and Mathematics Education, Germany), Yang Hou (Soochow University, China), Mael Houbre (Ministerial Agency

of Artificial Intelligence in Defense, France), Shu-Kai Hsieh (Graduate Institute of Linguistics, National Taiwan University, Taiwan), Songlin Hu (Institute of Information Engineering, CAS, China), Zhengyu Hu (Hong Kong University of Science and Technology, China), Yu-Shiang Huang (National Taiwan University, Taiwan), Quzhe Huang (Peking University, China), Zhihong Huang (University of Macau, Macau), Xiyang Huang (Wuhan University, China), Hen-Hsen Huang (Institute of Information Science, Academia Sinica, Taiwan), Chenyang Huang (Dept. of Computing Science, University of Alberta, Canada), Ming Huang (UTHealth, United States), Stéphane Huet (Université d'Avignon, France), Muhammad Humayoun (Karlstad University, Sweden), Syed Huq (City St George's, University of London, United Kingdom)

[I] Ignacio Iacobacci (ELM Europe, United Kingdom), Nancy Ide (Vassar College/Brandeis University, United States), Dmitry Idiatov (CNRS, France), Nikolai Ilinykh (University of Gothenburg, Sweden), Dmitry Ilvovsky (HSE University, Russian Federation), Sukairaj Imam (Bayero University Kano, Nigeria), Hamza Imam Amjad (Moscow Institute of Physics and Technology, Russian Federation), Kenji Imamura (National Institute of Information and Communications Technology, Japan), Joseph Marvin Imperial (University of Bath, United Kingdom), Balázs Indig (Eötvös Loránd University, Department of Digital Humanities, Hungary), Anton Ingason (University of Iceland, Iceland), Diana Inkpen (University of Ottawa, Canada), Koji Inoue (Kyoto University, Japan), Takashi Inui (University of Tsukuba, Japan), Radu Ion (Research Institute for Artificial Intelligence, Romanian Academy, Romania), Radu Tudor Ionescu (University of Bucharest, Romania), Maxim Ionov (University of Zaragoza, Spain), Elena Irimia (Research Institute for Artificial Intelligence, Romanian Academy, Romania), Mikel Irukieta (University of the Basque Country (UPV/EHU), Spain), Hitoshi Isahara (Otemon Gakuin University, Japan), Amy Isard (HCDS, University of Hamburg, Germany), Aminul Islam (University of Louisiana at Lafayette, United States), Kaoru Ito (Kyushu University, Japan)

[J] Kokil Jaidka (National University of Singapore, Singapore), Abhik Jana (IIT Bhubaneswar, India), Eugene Jang (Northeastern University, United States), Maarten Janssen (UFAL, Charles University, LINDAT/CLARIAH-CZ, Czech Republic), Inigo Jauregi Unanue (University of Technology Sydney, Australia), Tomáš Jelínek (Charles University, Faculty of Arts, Czech Republic), Sungho Jeon (Samsung Electronics, Republic of Korea), Zganec Gros Jerneja (Alpineon d.o.o., Slovenia), Sanjanasri Jeyamani Palanivel (Umeå University, Sweden), Elisabetta Jezek (University of Pavia, Italy), Sneha Jha (Imperial College, United Kingdom), Girish Jha (Jawaharlal Nehru University, India), Ping Jian (Beijing Institute of Technology, China), Junfeng Jiang (The University of Tokyo, Japan), Bohan Jiang (Arizona State University, United States), Jintao Jiang (AppTek, United States), Tianyu Jiang (University of Cincinnati, United States), Zhuoxuan Jiang (Shanghai Business School, China), Jiarong Jiang (AWS, United States), Yifan Jiang (Information Sciences Institute, University of Southern California, United States), Xiaotong Jiang (Suzhou University of Science and Technology, China), Weiwei Jiang (Beijing University of Posts and Telecommunications, China), Chao Jiang (Georgia Institute of Technology, United States), Salud María Jiménez Zafra (Universidad de Jaén, Spain), Antonio Jimeno Yepes (RMIT University, Australia), Liqiang Jing (University of Texas at Dallas, United States), Dewi Jones (Language Technologies Unit, Canolfan Bedwyr, Bangor University, United Kingdom), Pratik Joshi (Google DeepMind, United States), Zhu Junguo (Kunming University of Science and Technology, China)

[K] Tatiana Kachkovskaia (University of Turku, Finland), Kyo Kageura (University of Tokyo, Japan), Sylvain Kahane (Modyco, University Paris Nanterre & CNRS / Institut Universitaire de France, France), Kohei Kajikawa (Georgetown University, United States), Tomoyuki Kajiwarra (Ehime University / The University of Osaka, Japan), Sadia Kamal (Oklahoma State University, United States), Hiroyuki Kameda (Tokyo University of Technology, Japan), Hirotaka Kameko (Kyoto University, Japan), Hidetaka Kamigaito (Nara Institute of Science and Technology, Japan), Pawel Kamocki (Leibniz Institute for German Language, Germany), Jaap Kamps (University

of Amsterdam, Netherlands), Lis Kanashiro Pereira (NICT, Japan), Hiroshi Kanayama (IBM Research - Tokyo, Japan), Hemanth Kandula (Raytheon BBN Technologies, United States), Jenna Kanerva (University of Turku, Finland), Yoshinobu Kano (Faculty of Informatics, Shizuoka University, Japan), Takatomo Kano (NTT, Japan), Janakkumar Kapuriya (University of Galway, Ireland), Alina Karakanta (Leiden University Centre for Linguistics, Netherlands), Sarvnaz Karimi (CSIRO, Australia), Deepthi Karkada (Not Affiliated, United States), Anisia Katinskaia (Swiss Data Science Center, Switzerland), Yoshihide Kato (Nagoya University, Japan), Sophia Katenko (EcoVadis, Netherlands), Tatsuya Kawahara (Kyoto University, Japan), Daisuke Kawahara (Waseda University, Japan), Noriaki Kawamae (NTT Comware, Japan), Seiya Kawano (Kyoto Institute of Technology, Japan), Tolga Kayadelen (Google, United Kingdom), Hideto Kazawa (Google, Japan), Casey Kennington (Boise State University, United States), Katia Lida Kermanidis (Department of Informatics, Ionian University, Greece), Vlado Keselj (Dalhousie University, Canada), Erik Ketzan (King's College London, United Kingdom), Siti Oryza Khairunnisa (Tokyo Metropolitan University, Japan), Aditya Khan (University of Toronto, Canada), Raziullah Khan (HCL Tech, United States), Fahad Khan (Istituto di Linguistica Computazionale "Antonio Zampolli", CNR, Italy), Alizishaan Khatri (Software Engineer at Facebook, United States), Vivek Khetan (Accenture Labs, United States), Witold Kieraś (Institute of Computer Science, Polish Academy of Sciences, Poland), Zlata Kikteva (University of Passau, Germany), Inyoung Kim (Naver Labs Europe, France), Hansaem Kim (Yonsei University, Republic of Korea), Gyuwan Kim (University of California, Santa Barbara, United States), Derrick Eui Gyu Kim (Brandeis University, United States), Harksoo Kim (Konkuk University, Republic of Korea), Jin-Dong Kim (National Institute of Genetics, Japan), Vadim Kimmelman (University of Bergen, Norway), Yasutomo Kimura (Otaru University of Commerce, Japan), Tracy Holloway King (Adobe Inc., United States), Margaret King (Honorary Professor, University of Geneva, Switzerland), Leona King (KU Leuven, CiTiP, FWO, imec, Belgium), Christo Kirov (Google, United States), Denis Kiselev (Independent Researcher, Japan), Norihide Kitaoka (Toyohashi University of Technology, Japan), Christopher Klamm (University of Mannheim, Germany), Felicitas Kleber (Department of Language Science and Technology, Saarland University, Germany), Matej Klemen (University of Ljubljana, Faculty of Computer and Information Science, Slovenia), Alina Klerings (University of Mannheim, Germany), Katarzyna Klessa (Adam Mickiewicz University in Poznań, Poland), Roman Klinger (University of Bamberg, Germany), Daniil Kocharov (Speech and Cognition Research Group, Tampere University, Finland), Philipp Koehn (Johns Hopkins University, United States), Svetla Koeva (Institute for Bulgarian Language "Prof. Lyubomir Andreychin", Bulgarian Academy of Sciences, Bulgaria), Joachim Köhler (Fraunhofer IAIS, Germany), Dimitrios Kokkinakis (University of Gothenburg, Sweden), Athanasia Kolovou (Athena Research Center, Greece), Mamoru Komachi (Hitotsubashi University, Japan), Kazunori Komatani (University of Osaka, Japan), Kanako Komiya (Tokyo University of Agriculture and Technology, Japan), Maxim Konca (Goethe University Frankfurt, Germany), Grzegorz Kondrak (University of Alberta, Canada), Alexander König (CLARIN ERIC, Netherlands), Ioannis Konstas (Heriot-Watt University, United Kingdom), Maarit Koponen (University of Eastern Finland, Finland), Yannis Korkontzelos (Department of Computer Science, Edge Hill University, United Kingdom), Leila Kosseim (Concordia University, Canada), Maria Koutsombogera (Trinity College Dublin, Ireland), Olivier Kraif (Lidilem - Université Grenoble Alpes, France), Pavel Kral (University of West Bohemia, Dept. of Computer Science and Engineering, Czech Republic), Michael Kranzlein (Signify, United States), Matthias Kraus (Augsburg University, Germany), Lea Krause (Vrije Universiteit Amsterdam, Netherlands), Simon Krek (Jožef Stefan Institute, Slovenia), Brigitte Krenn (Austrian Research Institute for Artificial Intelligence, Austria), Marie-Pauline Krielke (Saarland University, Germany), Nikhil Krishnaswamy (Colorado State University, United States), Robert Krovetz (Lexical Research, United States), Canasai Kruengkrai (RIKEN, Japan), Joni Kruijsbergen (LT3, Ghent University, Belgium), Vladislav Kubon (Charles University, Czech Republic), Roland Kuhn (National Research Council of Canada, Canada), Gokhan Kul (University of Massachusetts Dartmouth, United States), Vivek Kulkarni (Samsung

Research, United States), Jimmy Kunzmann (Retired, Germany), Marc Kupietz (IDS Mannheim, Germany), Tatsuki Kuribayashi (MBZUAI, United Arab Emirates), Andrey Kutuzov (University of Oslo, Norway), Taja Kuzman Pungeršek (Jožef Stefan Institute, Slovenia), Sunjae Kwon (The University of Massachusetts Amherst, United States)

[L] Marie-Claude L' Homme (OLST, University of Montreal, Canada), Sofie Labat (Ghent University, Belgium), Cyril Labbe (Univ. Grenoble Alpes, France), Matthieu Labeau (Telecom Paris, France), Penny Labropoulou (ILSP / Athena RC, Greece), Clara Lachenmaier (University of Bielefeld, Germany), Jordan Lachler (University of Alberta, Canada), Mathieu Lafourcade (LIRMM, France), Kenneth Lai (Brandeis University, United States), Sobha Lalitha Devi (AU-KBC Research Centre, Anna University, India), David Langlois (LORIA, University of Lorraine, France, France), Mateusz Lango (Poznan University of Technology, Poland), Eric Laporte (Univ Gustave Eiffel, France), Ekaterina Lapshinova-Koltunski (University of Hildesheim, Germany), Staffan Larsson (University of Gothenburg, Sweden), Viesturs Lasmanis (University of Latvia, Institute of Mathematics and Computer Science, Latvia), Alberto Lavelli (FBK, Italy), Éric Le Ferrand (Boston College, United States), Sebastien Le Maguer (University of Helsinki, Finland), Alessandro Leanza (IDSIA USI-SUPSI, Switzerland), Gianluca Leboni (Ca' Foscari University of Venice, Italy), Benjamin Lecouteux (LIG/GETALP, France), Annie Lee (Ontario Tech University, University of Toronto, Canada), Roy Ka-Wei Lee (Singapore University of Technology and Design, Singapore), Hwanhee Lee (Chung-Ang University, Republic of Korea), Kiyong Lee (Korea University, Republic of Korea), Lung-Hao Lee (National Yang Ming Chiao Tung University, Taiwan), Jae Sung Lee (Chungbuk National University, Republic of Korea), Gyeongbok Lee (Meritz Fire & Marine Insurance Co., Ltd., Republic of Korea), Ji-Ung Lee (University of Saarland, Germany), Yu-Ting Lee (Apertis AI, United States), Jan Lehečka (University of West Bohemia, Czech Republic), Timm Lehmberg (Academy of Science and Humanities in Hamburg, Germany), Spyretta Leivaditi (Unicorn Labs B.V., Netherlands), Gaël Lejeune (STIH, Sorbonne Université, France), Arun Balajiee Lekshmi Narayanan (University of Pittsburgh, United States), Jens Lemmens (University of Antwerp, Belgium), Lothar Lemnitzer (Berlin-Brandenburgische Akademie der Wissenschaften, Germany), Jakob Lenardic (Institute of Contemporary History, Slovenia), Ladislav Lenc (University of West Bohemia in Pilsen, Czech Republic), Alessandro Lenci (University of Pisa, Italy), Pilar León-Araúz (University of Granada, Spain), Yves Lepage (Waseda University, Japan), Paul Lerner (Sorbonne Université, CNRS, ISIR, France), Gregor Leusch (eBay, Germany), Johannes Leveling (Fraunhofer IAIS, Germany), Lauren Levine (Georgetown University, United States), Gina-Anne Levow (University of Washington, United States), Barbara Lewandowska-Tomaszczyk (University of Applied Sciences in Konin, Poland), Dave Lewis (ADAPT Centre, Trinity College Dublin, Ireland), Jingxuan Li (Google, United States), Zhuoqun Li (Institute of Software, Chinese Academy of Sciences, China), Baoli Li (BoZhi Technology (Beijing) Co., Ltd., China), Binyang Li (University of International Relations, China), Dongyang Li (shiep.edu.cn, China), Bin Li (Nanjing Normal University, China), Jixing Li (City University of Hong Kong, Hong Kong Special Administrative Region of China), Shaobo Li (Harbin Institute of Technology, China), Junlin Li (The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Junhui Li (Soochow University, Suzhou, China), Dongfang Li (Harbin Institute of Technology, Shenzhen, China), Haonan Li (MBZUAI, United Arab Emirates), Chuyuan Li (The University of British Columbia, Canada), Piji Li (Nanjing University of Aeronautics and Astronautics, China), Bo Li (Peking University, China), Bobo Li (National University of Singapore, Singapore), Yu Xi Li (The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Qian Li (Beijing University of Posts and Telecommunications, China), Zhaolin Li (Karlsruhe Institute of Technology, Germany), Ruifan Li (Beijing University of Posts and Telecommunications, China), Ruixue Lian (University of Wisconsin-Madison, United States), Siyu Liang (University of Washington, United States), Yamei Liao (University of Colorado Boulder, United States), Jindřich Libovický (Charles University, Czech Republic), Mohamed Lichouri (University of Science and Technology Houari Boumediene, Algeria), Chaya Liebeskind

(Jerusalem College of Technology , Lev Academic Center, Israel), Anna Liednikova (ALIAE SAS, France), Melissa Lieffers (Computer Science, Indiana University - Bloomington, United States), Constantine Lignos (Brandeis University, United States), Tzi-Mi Lin (National Central University, Taiwan), Lucy Lin (Spotify, United States), Qika Lin (National University of Singapore, Singapore), Donghui Lin (Okayama University, Japan), David Lindemann (EHU University of the Basque Country, Spain), Krister Linden (University of Helsinki, Finland), Pasquale Lisena (EURECOM, France), Pierre Lison (Norwegian Computing Centre, Norway), Inari Listenmaa (Chalmers University of Technology, Sweden), Eleonora Litta (Università Cattolica del Sacro Cuore, Milano, Italy), Yixiao Liu (Nanyang Technological University, China), Zhengyuan Liu (Agency for Science, Technology and Research (A*STAR), Singapore, Singapore), Fengkai Liu (The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Pengfei Liu (Centre for Perceptual and Interactive Intelligence, Hong Kong Special Administrative Region of China), Bang Liu (University of Montreal, Canada), Guangliang Liu (Michigan State University, United States), Jiangming Liu (Yunnan University, China), Siyou Liu (University of Macau, Macau), Zhenghao Liu (Northeastern University, China), Yongbin Liu (School of Computer Science, University of South China, China), MeiLing Liu (Northeast Forestry University, China), Dongqi Liu (Saarland University, Germany), Chao-Hong Liu (ITRI, Potamu Research, Ireland), Danni Liu (Karlsruhe Institute of Technology, Germany), Ting Liu (Siena College, United States), Hexin Liu (Nanyang Technological University, Singapore), Nikola Ljubešić (Jožef Stefan Institute, Slovenia), Eduardo Lleida Solano (University of Zaragoza, Spain), Sharid Loaiciga (University of Gothenburg, Sweden), Christina Lohr (Universität Leipzig, Germany), Marketa Lopatkova (Charles University, Czech Republic), Lucelene Lopes (USP - ICMC, Brazil), Cédric Lopez (Emvista, France), Oier Lopez de Lacalle (University of the Basque Country, Spain), Pilar Lopez Ubeda (HT medica, Spain), Ehsan Lotfi (University of Antwerp, CLIPS, Belgium), Anastasia Loukina (Grammarly Inc, United States), Olga Lovick (University of Saskatchewan, Canada), London Lowmanstone (University of Minnesota, United States), Guangming Lu (LinkedIn Corp, United States), Yaojie Lu (Institute of Software, Chinese Academy of Sciences, China), Di Lu (Dataminr, United States), Xinyuan Lu (National University of Singapore, Singapore), Lu Lu (Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Wenpeng Lu (Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences); Shandong Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, China), Hui Lu (The Chinese University of Hong Kong, Hong Kong Special Administrative Region of China), Andy Luecking (Chemnitz University of Technology, Germany), Stephanie M. Lukin (DEVCOM Army Research Laboratory, United States), Alex Luu (University of Georgia, United States), Xiao Luwei (National University of Singapore, Singapore), Pedro Henrique Luz de Araujo (University of Vienna, Austria), Teresa Lynn (MBZUAI, United Arab Emirates), Chenyang Lyu (Alibaba Group, China)

[M] Wei-Yun Ma (Institute of Information Science, Academia Sinica, Taiwan), Yinglong Ma (North China Electric Power University, China), Tianyi Ma (University of Notre Dame, United States), Chunlan Ma (LMU Munich, Germany), Mateus Machado (University of São Paulo, Brazil), Petter Mæhlum (University of Oslo, Norway), Pierre Magistry (ERTIM - Inalco, France), Eimear Maguire (University of Dundee, United Kingdom), Sina Mahdipour Saravani (University of Utah, United States), Zola Mahlaza (University of Cape Town, South Africa), Wolfgang Maier (Mercedes-Benz AG, Germany), Mieradilijiang Maimaiti (Xinjiang University, China), Isa Maks (Vu university amsterdam, Netherlands), Aaron Maladry (Ghent University, Belgium), Shervin Malmasi (Amazon, United States), Sheikh Mannan (Colorado State University, United States), Ramesh Manuvinakurike (Intel labs, United States), Andrea Marchetti (IIT-CNR, Italy), Michal Marcinczuk (CodeNLP, Poland), Małgorzata Marciniak (Institute of Computer Science Polish Academy of Sciences, Poland), Thomas Margoni (CiTiP, Faculty of Law, KU Leuven, Belgium),

Alda Mari (<http://www.institutnicod.org/>, France), James H. Martin (University of Colorado Boulder, United States), Philippe Martin (LLF UMR 7110, UFRL, Université Paris Diderot, France), Matej Martinc (Jozef Stefan Institute, Slovenia), Raquel Martínez (Universidad Nacional de Educacion a Distancia (UNED), Spain), Eugenio Martínez Cámara (University of Jaén, Spain), Juan Martinez-Romo (Universidad Nacional de Educación a Distancia (UNED), Spain), Patricia Martin-Rodilla (University of A Corunna, Spain), Claudia Marzi (Institute for Computational Linguistics - CNR, Italy), Aleksandre Maskharashvili (University of Illinois Urbana-Champaign, United States), Michele Mastromattei (Humanitas University, Italy), Sarah Masud (LCS2, IIITD, India), Abu Kaisar Mohammad Masum (University of Louisiana at Lafayette, United States), Ditty Mathew (audEERING GmbH, Germany), Clint Solomon Mathialagan (Oracle, United States), Shigeki Matsubara (Nagoya University, Japan), Yuichiroh Matsubayashi (Tohoku University / Riken, Japan), Yuji Matsumoto (Riken Center for Advanced Intelligence Project, Japan), Kazuyuki Matsumoto (Tokushima University, Japan), Suguru Matsuyoshi (Tokyo University of Technology, Japan), Takuya Matsuzaki (Tokyo University of Science, Japan), Maximilian Maurer (GESIS Leibniz Institute for the Social Sciences and Heinrich Heine University Duesseldorf, Germany), Lina Mavrina (Bielefeld University, Germany), Candy Olivia Mawalim (Japan Advanced Institute of Science and Technology, Japan), Michael Maxwell (Retired from University of Maryland, United States), Philipp Mayr (GESIS - Leibniz Institute for the Social Sciences, Germany), David McDonald (Smart Information Flow Technologies (dba SIFT, LLC), United States), Lachlan McPheat (The Alan Turing Institute, United Kingdom), Susan McRoy (University of Wisconsin-Milwaukee, United States), Mouheb Mehdoui (University of Lille France, France), Xin Mei (Northwestern Polytechnical University, China), Philipp Meier (Heinrich-Heine-University, Germany), Maite Melero (AI Institute - Barcelona Supercomputing Center, Spain), Maite Melero (BSC, Spain), Yuanliang Meng (Microsoft Corporation, United States), Fanchao Meng (Misericordia University, United States), Katrin Menzel (Saarland University, Germany), Wolfgang Menzel (Hamburg University, Germany), Robert E. Mercer (The University of Western Ontario, Canada), Fabio Mercorio (University of Milano-Bicocca, Italy), Marie-Jean Meurs (UQAM, Canada), Stephan Meylan (Massachusetts Institute of Technology, United States), Ivan Vladimir Meza Ruiz (Insituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico), Antonio Valerio Miceli Barone (The University of Edinburgh, United Kingdom), Andrianos Michail (University of Zurich, Switzerland), Alexis Michaud (CNRS - LACITO, France), Margot Mieskes (University of Applied Sciences, Darmstadt, Germany), Victor Mijangos (National Autonomous University of Mexico (UNAM), Mexico), Marie Mikulová (Charles University, Czech Republic), Maja Milicevic Petrovic (University of Bologna, Italy), Jiří Milička (Institute of Comparative Linguistics, Faculty of Arts, Charles University, Prague, Czech Republic), Simon Mille (ADAPT Research Centre, Dublin City University, Ireland), Timothy Miller (Boston Children's Hospital and Harvard Medical School, United States), Corey Miller (self, United States), Alice Millour (Université Paris 8 Vincennes Saint-Denis, France), Junghyun Min (Georgetown University, United States), Koji Mineshima (Keio University, Japan), Sabino Miranda (Universidad Autónoma de la Ciudad de México, Mexico), Jiří Mírovský (Charles University, Czech Republic), Pruthwik Mishra (SVNIT Surat, India), Meladel Mistica (The University of Melbourne, Australia), Vibhu Mittal (Inflection AI, United States), Makoto Miwa (Toyota Technological Institute, Japan), Fengran Mo (Universite de Montreal, Canada), Mohammad Reza Modarres (University of Tehran, Iran, Iran (Islamic Republic of)), Sandip Modha (Università degli Studi di Milano-Bicocca, Italy), Ashutosh Modi (Indian Institute of Technology Kanpur, India), Silvia Molina Plaza (Upm, Spain), Guillermo Moncecchi (Facultad de Ingeniería - Universidad de la República, Uruguay), Massimo Moneglia (University of Florence, Italy), Helena Moniz (INESC-ID, University of Lisbon, EAMT, Portugal), Soto Montalvo (Universidad Rey Juan Carlos, Spain), Syrielle Montariol (UC Berkeley, United States), Arturo Montejo-Raez (Universidad de Jaen, Spain), Simonetta Montemagni (Istituto di Linguistica Computazionale "Antonio Zampolli", Italy), Manuel Montes (INAOE, Mexico), Johanna Monti ("L'Orientale" University of Naples, Italy), Elena Montiel-Ponsoda (Ontology

Engineering Group, Universidad Politécnica de Madrid, Spain), Richard Moot (CNRS, France), Rodrigo Morales-Sánchez (Universidad Nacional de Educación a Distancia, Spain), Steven Moran (University of Neuchâtel, Switzerland), Roser Morante (UNED, Spain), Jose G. Moreno (Paul Sabatier University - IRIT, France), Antonio Moreno-Sandoval (Universidad Autonoma Madrid, Spain), Julian Moreno-Schneider (DFKI, Germany), Alejandro Moreo (Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", CNR, Italy), Luis Morgado da Costa (VU Amsterdam, Netherlands), Yusuke Mori (The University of Tokyo, Japan), Shinsuke Mori (Kyoto University, Japan), Emmanuel Morin (LS2N UMR CNRS 6004, France), Yasmin Moslem (ADAPT Centre, Trinity College Dublin, Ireland), Larry Moss (Indiana University, Bloomington, United States), Jamshid Mozafari (University of Innsbruck, Austria), Hamdy Mubarak (Qatar Computing Research Institute, Qatar), Kadri Muischnek (associate professor, Estonia), Vandan Mujadia (IIIT-H, India), Philippe Mulhem (LIG-CNRS, France), Philippe Muller (IRIT, University of Toulouse, France), Markus Müller (Amazon, United States), Yohei Murakami (Ritsumeikan University, Japan), Koji Murakami (Rakuten Insutitute of Technology, United States), Yugo Murawaki (Kyoto University, Japan), Rudra Murthy (IBM India Research Limited, India), Benedetta Muscato (Scuola Normale Superiore, Italy), Sakthi Muthiah (Shiv Nadar University Chennai, India), Arianna Muti (Bocconi University, Italy)

[N] Nazia Nafis (The University of Sheffield, United Kingdom), Masaaki Nagata (NTT, Inc., Japan), Tetsuji Nakagawa (Google LLC, United States), Satoshi Nakamura (The Chinese University of Hong Kong, Shenzhen, China), Toshiaki Nakazawa (The University of Tokyo, Japan), Sangha Nam (Korea Telecom, Republic of Korea), Amani Namboori (Amazon, United States), Luca Nannini (CiTIUS USC, Spain), Fedelucio Narducci (Politecnico di Bari, Italy), Sudip Naskar (Jadavpur University, India), Alexis Nasr (Aix Marseille University, France), Arbi Haza Nasution (Universitas Islam Riau, Indonesia), Sripriya Natarajan (SSN College of Engineering, India), Rajeswari Natarajan (SASTRA Deemed to be University, India), David Nathan (Independent scholar and developer, Australia), Borja Navarro-Colorado (University of Alicante, Spain), Claire Nedellec (MaIAGE INRAE, France), Poli Nemkova (University of North Texas, United States), Friedrich Neubarth (OFAI, Austria), Clemens Neudecker (Staatsbibliothek zu Berlin, Germany), Aurélie Névéol (Université Paris Saclay, CNRS, LISN, France), Mariana Neves (German Federal Institute for Risk Assessment, Germany), York Hay Ng (University of Toronto, Canada), Axel-Cyrille Ngonga Ngomo (Paderborn University, Germany), Emmanuel Ngue Um (University of Yaoundé I, Cameroon), Dang Tuan Nguyen (Saigon University, Viet Nam), Kiem-Hieu Nguyen (Hanoi Univ of Science and Technology, Viet Nam), Huyen Nguyen (VNU University of Science, Hanoi, Viet Nam), Minh-Tien Nguyen (Hung Yen University of Technology and Education, Viet Nam), Kiet Nguyen (University of Information Technology, VNU-HCM, Viet Nam), Jan Niehues (Karlsruhe Institut of Technology, Germany), Sanni Nimb (Society for Danish Language and Literature (DSL), Denmark), Takashi Ninomiya (Ehime University, Japan), Noriki Nishida (RIKEN Center for Advanced Intelligence Project, Japan), Kosuke Nishida (NTT, Inc., Japan), Lydia Nishimwe (Inria, France), Sergiu Nisioi (Human Language Technologies Research Center, University of Bucharest, Romania), Joakim Nivre (Uppsala University, Sweden), Bill Noble (Massachusetts Institute of Technology, United States), Tadashi Nomoto (National Institute of Japanese Literature, Japan), Andreas Nuernberger (Otto-von-Guericke Universitaet Magdeburg, Germany), Pierre Nugues (Lund University, Sweden), Sérgio Nunes (INESC TEC and University of Porto, Portugal), Diana Nurbakova (National Institute of Applied Sciences of Lyon, INSA Lyon - University of Lyon - LIRIS, France), Brahmani Nutakki (Saarland University, Germany), Valentin Nyzam (Universite de Tours, France)

[O] Nicolás Benjamín Ocampo (Centrum Wiskunde & Informatica, Netherlands), Laura Occhipinti (University of Bologna, Italy), Alexander O'Connor (Autodesk, United States), Yusuke Oda (National Institute of Informatics, Japan), Antoni Ogonowski (.Cognizant, France), Yoo Rhee Oh (OulSoft, Republic of Korea), Soyoung Oh (Saarland University, Germany), Mayumi Ohta

(Heidelberg University, Germany), Kazushi Ohya (Tsurumi University, Japan), Atul Kr. Ojha (Data Science Institute, Unit for Linguistic Data, University of Galway, Ireland), Shu Okabe (TUM Heilbronn, Germany), Tsuyoshi Okita (Kyushu Institute of Technology, Japan), Alp Oktem (CLEAR Global, Spain), Antoni Oliver (Universitat Oberta de Catalunya, Spain), Esra Onal (Indiana University, United States), Juri Opitz (University of Zurich, Switzerland), Constantin Orasan (University of Surrey, United Kingdom), Siim Orasmaa (University of Tartu, Estonia), Maite Oronoz (HiTZ Center - Ixa, University of the Basque Country UPV/EHU, Spain), Juan Orozco-Arroyave (Universidad de Antioquia, Colombia), John E. Ortega (Northeastern University, United States), Jenny Ortiz-Zambrano (Universidad de Guayaquil, Ecuador), Yohei Oseki (University of Tokyo, Japan), Robert Östling (Department of Linguistics, Stockholm University, Sweden), Naoki Otani (Megagon Labs, United States), Joshua Otten (George Mason University, United States), Hiroki Ouchi (Nara Institute of Science and Technology, Japan), Mohamed Outahajala (Independent, Canada), Jessica Ouyang (University of Texas at Dallas, United States), Lilja Øvrelid (Dept of Informatics, University of Oslo, Norway)

[P] Win Pa Pa (University of Computer Studies, Yangon, Myanmar), Teresa Paccosi (KNAW Humanities Cluster DHTLab, Netherlands), Lluís Padro (Universitat Politècnica de Catalunya, Spain), Aline Paes (Institute of Computing, Universidade Federal Fluminense, Brazil), Patrizia Paggio (University of Copenhagen, Denmark), Vasile Pais (Research Institute for Artificial Intelligence, Romanian Academy, Romania), Santanu Pal (Wipro, United Kingdom), Lucia Palacios Palacios (Universidad Politécnica de Madrid, Spain), Chester Palen-Michel (Ebay, United States), Martha Palmer (University of Colorado, United States), Alessio Palmero Aprosio (University of Trento, Italy), Jingheng Pan (University of Hamburg, Germany), Youcheng Pan (Pengcheng Laboratory, China), Alexander Panchenko (S-NLP, Russian Federation), Mugdha Pandya (University of Edinburgh, United Kingdom), Ludovica Pannitto (LILEC - University of Bologna, Italy), Alessandro Panunzi (University of Florence, Italy), Haris Papageorgiou (Institute for Language and Speech Processing ILSP/ ATHENA R.C., Greece), Ivandre Paraboni (University of São Paulo, Brazil), Georgios Paraskevopoulos (National Technical University of Athens, Greece), Thiago Pardo (University of São Paulo, Brazil), Antonio Pareja-Lora (Universidad de Alcalá (UAH) / FITISPos (UAH) / ATLAS (UNED) / DMEG (UdG), Spain), Silvia Pareti (Independent researcher, Switzerland), Cecile Paris (CSIRO, Australia), ChaeHun Park (KAIST, Republic of Korea), Yannick Parmentier (Université de Lorraine, France), Luis Parra Gallego (University of Antioquia, Colombia), Nivranshu Pasricha (Data Science Institute, University of Galway, Ireland), Marco Passarotti (Università Cattolica del Sacro Cuore, Italy), Panupong Pasupat (Google, United States), Sapan Patel (Amazon Alexa, United States), Dhruva Patil (Amazon, Colorado State University, United States), Viviana Patti (University of Turin, Dipartimento di Informatica, Italy), Robert Patton (Oak Ridge National Laboratory, United States), Alexandre Pauchet (INSA Rouen Normandy - LITIS, France), Lucas Pavanelli (Stone, Brazil), Pavel Pecina (Charles University, Czech Republic), Bolette Pedersen (University of Copenhagen, Denmark), Stephan Peitz (Apple, Germany), Carmen Peláez-Moreno (University Carlos III Madrid, Spain), Valentin Pelloin (INA, France), Jing Peng (Montclair State University, United States), Bo Peng (The Hong Kong Polytechnic University, China), Xingyu Peng (State Key Laboratory of Complex & Critical Software Environment, China), José Pereira (Pontificia Universidad Católica del Perú, Peru), Martín Pereira-Farina (University of Santiago de Compostela, Spain), Naiara Perez (University of the Basque Country, Spain), Veronica Perez-Rosas (Texas State University, United States), Vynska Permadi (University of Sheffield, United Kingdom), Simon Petitjean (Carl von Ossietzky University of Oldenburg, Germany), Volha Petukhova (Saarland University, Germany), Sandro Pezzelle (University of Amsterdam, Netherlands), Scott Piao (Lancaster University, United Kingdom), Maciej Piasecki (Wrocław University of Science and Technology, Poland), Davide Picca (University of Lausanne, Switzerland), Massimo Piccardi (University of Technology Sydney, Australia), Silvia Piccini (Institute for Computational Linguistics, Italy), Ildiko Pilan (Norwegian Computing Center, Norway), Daniel Pimienta (Observatory of Linguistic & Cultural Diversity in the Internet -, France),

Aidan Pine (National Research Council Canada, Canada), Marcis Pinnis (Tilde, Latvia), Yuval Pinter (Ben-Gurion University of the Negev, Israel), Irina Piontkovskaya (Huawei Noah's Ark Lab, Russian Federation), Flammie Pirinen (UiT Norgga Árkatalaš universitehta, Norway), Lucia Pitarch (University of Zaragoza, Spain), Lidia Pivovarova (University of Helsinki, Finland), Laura Plaza (UNED, Spain), Alistair Plum (University of Luxembourg, Luxembourg), Brian Pluss (University of Dundee, United Kingdom), Thierry Poibeau (LATTICE (CNRS & ENS/PSL), France), Adam Poliak (Bryn Mawr College, United States), Marco Polignano (University of Bari, Italy), Senja Pollak (Jožef Stefan Institute, Slovenia), Kimberly Pollard (DEVCOM Army Research Laboratory, United States), Girishkumar Ponkiya (ISS STOXX, India), Rahul Ponnusamy (Data science institute, University of Galway, Ireland), Martin Popel (Charles University, Faculty of Mathematics and Physics, UFAL, Czech Republic), Octavian Popescu (IBM Yorktown Research, United States), Anisia Popescu (Université Paris 8 - Saint Denis, France), Andrei Popescu-Belis (HEIG-VD / HES-SO, Switzerland), Maja Popović (IU University, Germany), Francesca Portante (KU Leuven CiTiP, Belgium), François Portet (Univ Grenoble Alpes, Laboratoire d'Informatique de Grenoble, France), Gerasimos Potamianos (University of Thessaly, Greece), Céline Poudat (UMR7320 BCL, University Nice Côte d'Azur, France), Jakob Prange (University Medical Center Hamburg-Eppendorf, Germany), Damith Premasiri (Lancaster University, United Kingdom), Ajda Pretnar Zagar (Faculty of Computer and Information Science, University of Ljubljana, Slovenia), Laurent Prévot (Aix Marseille Université & CNRS, France), Prokopis Prokopidis (ILSP/Athena RC, Greece), Carlos A. Prolo (Department of Informatics and Applied Mathematics - Federal University of Rio Grande do Norte, Brazil), Emily Prud'hommeaux (Boston College, United States), Michal Ptaszynski (Kitami Institute of Technology, Japan), Giovanni Puccetti (information Science and Technologies Institute "A. Faedo", Italy), Robert Pugh (Indiana University, United States), Rajkumar Pujari (amazon.com, United States), Matthew Purver (Queen Mary University of London, United Kingdom), James Pustejovsky (Brandeis University, United States)

[Q] Zheng Qi (Amazon Web Services, United States), Kun Qian (Apple Inc., United States), Yiyue Qian (Amazon, United States), Le Qiu (Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Paulo Quaresma (University of Evora, Portugal), Solen Quiniou (Nantes University - LS2N, France), Valeria Quochi (Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale "A. Zampolli", Italy)

[R] Daniele Radicioni (Computer Science Department, University of Turin, Italy), Jan Radimsky (University of South Bohemia, Czech Republic), Arya Rahgozar (University of Ottawa, Canada), Nishat Raihan (George Mason University, United States), Lisa Raitel (Technische Universitaet Berlin, BIFOLD, DFKI GmbH, Charité-IKIM, Germany), Satyam Raj (Arizona State University, United States), Sai Krishna Rallabandi (Fidelity Investments, United States), Resmi Ramachandranpillai (Research Scientist, Northeastern University, United States), Heri Ramampiaro (Norwegian University of Science and Technology (NTNU), Norway), Gema Ramírez-Sánchez (Prompsit Language Engineering, Spain), Carlos Ramisch (Aix Marseille University, CNRS, LIS, France), Tharindu Ranasinghe (Lancaster University, United Kingdom), Surangika Ranathunga (Massey University, New Zealand), Anku Rani (Massachusetts Institute of Technology, United States), O. Ranum (University of Surrey, United Kingdom), Paul Rayson (Lancaster University, United Kingdom), Julia Rayz (Purdue University, United States), Livy Real (Federal University of Amazonas, Brazil), Traian Rebedea (University Politehnica of Bucharest & NVIDIA, Romania), Marianne Reboul (ENS de Lyon, France), Georg Rehm (DFKI, Germany), Arianne Reimerink (University of Granada, Spain), Ehud Reiter (University of Aberdeen, United Kingdom), Nils Reiter (University of Cologne, Germany), Yujie Ren (University of Hamburg, Germany), Feiliang Ren (Northeastern University, China), David Reymond (Université de Toulon, France), Ricardo Ribeiro (Iscte and INESC-ID, Portugal), Eugenio Ribeiro (ISCTE-IUL / INESC-ID, Portugal), Giuseppe Riccardi (University of Trento, Italy), Valentin Richard (Université de Montpellier Paul Valéry, France), Ayla Rigouts Terryn (Université de Montréal, Mila, Canada), Leonardo Rigutini (expert.ai,

Italy), Matiss Rikters (AIST, Japan), Albert Rilliard (Universite Paris Saclay, CNRS, LISN, France), Kyeongmin Rim (Department of Computer Science, Brandeis University, United States), Fabio Rinaldi (IDSIA USI-SUPSI, Dalle Molle Institute for Artificial Intelligence, Switzerland), Cristian Rios (Universidad de Antioquia, Colombia), Esther Rituerto-Gonzalez (University Carlos III of Madrid, Spain), Hammad Rizwan (Dalhousie University, Canada), Giulia Rizzi (Università degli Studi di Milano-Bicocca and Universitat Politècnica de Valencia, Italy), Giuseppe Rizzo (LINKS Foundation, Italy), Pattabhi RK Rao (AU-KBC Research centre, India), Kirk Roberts (University of Texas Health Science Center at Houston, United States), Mathieu Roche (CIRAD, TETIS, France), Alvaro Rodrigo (NLP and IR group at UNED, Spain), Ricardo Rodrigues (Polytechnic Institute of Coimbra & Centre for Informatics and Systems of the University of Coimbra, Portugal), Sandra Rodríguez Rey (Universidad de Santiago de Compostela, Spain), Luis Javier Rodríguez-Fuentes (University of the Basque Country UPV/EHU, Spain), Stian Rødven-Eide (University of Gothenburg, Sweden), Lina M. Rojas Barahona (Orange Innovation Research, France), Eva Rolin (UCLouvain, CENTAL, Belgium), Roland Roller (DFKI SLT Lab, Germany), Francesco Ronzano (Universitat Pompeu Fabra, Barcelona, Spain), Rudolf Rosa (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Czech Republic), Marco Rospoche (Università degli Studi di Verona, Italy), Robert Ross (Technological University Dublin, Ireland), Sophie Rosset (Université Paris-Saclay, CNRS, LISN, France), Paolo Rosso (Universitat Politècnica de València, Spain), Kay Rottmann (Hochschule der Medien, University of Applied Sciences, Germany), Jean-Luc Rouas (LaBRI CNRS UMR 5800 Univ. Bordeaux, France), Dmitri Roussinov (University of Strathclyde, United Kingdom), Jacqueline Rowe (University of Edinburgh, United Kingdom), Daniel Ruffinelli (University of Mannheim, Germany), Pablo Ruiz Fabo (LiLPa, Université de Strasbourg, Spain), Ramon Ruiz-Dolz (University of Dundee, United Kingdom), Josef Ruppenhofer (Fernuniversität in Hagen, Germany), Elena Sofia Ruzzetti (University of Rome Tor Vergata, Italy)

[S] Malliga S (Kongu Engineering College, India), Motaz Saad (University of Salento, Italy), Fatiha Sadat (UQAM, Canada), Monjoy Saha (National Institutes of Health, United States, United States), Sattvik Sahai (Amazon, United States), Nihar Sahoo (Indian Institute of Technology, Bombay, India), Nikhil Saini (Samsung Research, India), Patrick Saint-Dizier (Retired from IRIT-CNRS, France), Keisuke Sakaguchi (Tohoku University / RIKEN, Japan), Yusuke Sakai (Nara Institute of Science and Technology, Japan), Hiroki Sakaji (Hokkaido University, Japan), Saheed Abdullahi Salahudeen (Kaduna State University, Nigeria), Israfel Salazar (University of Copenhagen, Denmark), Jonne Saleva (Brandeis University, United States), Christian Sallaberry (UPPA, France), Tanja Samardžić (IDSIA USI-SUPSI, Switzerland), Fahime Same (trivago N.V., Germany), Sarah Samson Juan (Universiti Malaysia Sarawak, Malaysia), Iñaki San Vicente (Orai NLP Technologies, Spain), Haruko Sanada (Rissho University, Japan), Víctor M. Sánchez-Cartagena (Universitat d'Alacant, Spain), Felipe Sanchez-Martínez (Universitat d'Alacant, Spain), Branislava Šandrih Todorović (NLB DigIT, Serbia), Manuela Sanguinetti (University of Cagliari, Department of Mathematics and Computer Science, Italy), Eric SanJuan (Avignon University, France), Brenda Santana (UFPEL, Brazil), Diana Santos (University of Oslo, Norway), Injy Sarhan (Shell / Utrecht University, Netherlands), Prof. Shikhar Kumar Sarma (Gauhati University, India), Kengatharaiyer Sarveswaran (University of Jaffna, Sri Lanka), Yutaka Sasaki (Toyota Technological Institute, Japan), Felix Sasaki (SAP SE, Germany), Minoru Sasaki (Ibaraki University, Japan), Balint Sass (ELTE Research Centre for Linguistics, Hungary), Yo Sato (Satoama Language Services, United Kingdom), Andreas Säuberli (LMU Munich, Germany), Sunil Saumya (INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD, India), Agata Savary (Paris-Saclay University, France), Noemi Scarpato (San RaffaeleRoma Open University, Italy), Manuel Schaaf (Goethe-University Frankfurt am Main, Germany), Johannes Schaefer (Fundamentals of Natural Language Processing, Germany), Andreas Scherbakov (the University of Melbourne; Constraint Technologies Inc., Australia), Yves Scherrer (University of Oslo, Norway), Julian Schlenker (University of Mannheim, Germany), Helmut Schmid (CIS, Ludwig-Maximilians-

Universitaet, Germany), Vera Schmitt (Technische Universität Berlin, German Research Center for Artificial Intelligence, Germany), Nathan Schneider (Georgetown University, United States), Tyler Schnoebelen (Decoded AI, United States), Stephanie Schoch (University of Virginia, United States), Bernhard Schröder (University of Duisburg-Essen, Germany), Marijn Schraagen (Utrecht University, Netherlands), Marc Schulder (University of Hamburg, Germany), William Schuler (The Ohio State University, United States), Sabine Schulte im Walde (University of Stuttgart, Germany), Tanja Schultz (University of Bremen, Germany), Sebastian Schuster (University of Vienna, Austria), Ineke Schuurman (KU Leuven, Belgium), Didier Schwab (Univ. Grenoble Alpes, France), Wesley Scivetti (Georgetown University, United States), Pascale Sébillot (Univ Rennes, INSA Rennes, CNRS, Inria, IRISA - UMR 6074, France), Djamé Seddah (Inria, France), Parker Seegmiller (Dartmouth College, United States), Arulmozi Selvaraj (University of Hyderabad, India), Giovanni Semeraro (University of Bari "Aldo Moro", Italy), Nasredine Semmar (CEA LIST, France), Rico Sennrich (University of Zurich, Switzerland), Christophe Servan (AMIAD - CNRS, LISN, France), Ayışığı Sevdik Çallı (Bilkent University, Turkey), Anna Shadrova (Universität Leipzig, Germany), Deven Santosh Shah (LinkedIn, United States), Saeideh Shahrokh Esfahani (Applied Scientist, Amazon Web Services (AWS), United States), Samira Shaikh (University of North Carolina at Charlotte, United States), Anastassia Shaitarova (University of Zurich, Switzerland), Mehrnough Shamsfard (Faculty of Computer Science and Engineering, Shahid Beheshti University, Iran (Islamic Republic of)), Jiandong Shao (UESTC, China), Matthew Shardlow (Manchester Metropolitan University, United Kingdom), Prashant Sharma (Independent Consultant, Japan), Serge Sharoff (University of Leeds, United Kingdom), Zaid Sheikh (Carnegie Mellon University, United States), Gaofei Shen (Tilburg University, Netherlands), Tomohide Shibata (SB Intuitions Corp., Japan), Kazutaka Shimada (Kyushu Institute of Technology, Japan), Erica K. Shimomoto (National Institute of Advanced Industrial Science and Technology, Japan), Anastasia Shimorina (Orange, France), Mason Shipton (Ontario Tech University, Canada), Dimitar Shterionov (Tilburg University, Netherlands), Christopher Shulby (Elsa Speak, United States), Alexander Shvets (Barcelona Supercomputing Center, Spain), Shweta (Assistant Professor, Department of CSE, NIT Calicut, India), Gosuddin Kamaruddin Siddiqi (Microsoft, United States), Muhammad Hammad Fahim Siddiqui (University of Ottawa, Canada), Melanie Siegel (Hochschule Darmstadt - University of Applied Sciences, Germany), Ingo Siegert (Otto von Guericke University Magdeburg, Germany), Judith Sieker (Bielefeld University, Germany), Nádia Silva (University of São Paulo, Brazil), Vivian Silva (Federal University of Rio de Janeiro, Brazil), João Ricardo Silva (University of Lisbon, Portugal), Stefano Silvestri (ICAR-CNR, Italy), Michel Simard (NRC, Canada), Maria Simi (Università di Pisa, Italy), Gary Simons (SIL Global, United States), Edwin Simpson (University of Bristol, United Kingdom), Jyotika Singh (Oracle, United States), Thoudam Doren Singh (National Institute of Technology Meghalaya, India), Sanasam Ranbir Singh (Indian Institute of Technology, India), Mayank Singh (IIT Gandhinagar, India), Pranaydeep Singh (LT3, University of Ghent, Belgium), Rajalakshmi Sivanaiah (Sri Sivasubramaniya Nadar College of Engineering, India), Inguna Skadina (Tilde/ Institute of Mathematics and Computer Science, University of Latvia, Latvia), Pavel Skrelin (Saint Petersburg State University, Russian Federation), Vésteinn Snæbjarnarson (University of Copenhagen, Denmark), Marco Antonio Sobrevilla Cabezudo (Pontifical Catholic University of Peru, Peru), Sokratis Sofianopoulos (Researcher, Greece), Soma Soma (International Institute of Information Technology, Hyderabad, India), Pia Sommerauer (Vrije Universiteit Amsterdam, Netherlands), Wei Song (Capital Normal University, China), Huacheng Song (The Hong Kong Polytechnic University, China), Haiyue Song (NICT, Japan), Virach Sornlertlamvanich (Faculty of Data Science, Musashino University, Japan), Alexey Sorokin (Moscow State University, Russian Federation), Vilelmini Sosoni (Ionian University, Greece), Gerasimos Spanakis (Maastricht University, Netherlands), Manuela Speranza (FBK - Fondazione Bruno Kessler, Italy), Rachele Sprugnoli (Università Cattolica del Sacro Cuore, Italy), Hiranmai Sri Adibhatla (IIIT Hyderabad, India), Somayajulu Sripada (Arria NLG Plc and University of Aberdeen, United Kingdom), Felix Stahlberg (Google Research, Germany), Efsthios Stamatatos

(University of the Aegean, Greece), Vivian Stamou (Archimedes AI, Greece), Ranka Stankovic (University of Belgrade - Faculty of Mining and Geology, Serbia), Mark Steedman (University of Edinburgh, United Kingdom), Kenneth Steimel (Cisco Systems Incorporated, United States), Armando Stellato (University of Rome Tor Vergata, Italy), Egon Stemle (Eurac Research, Italy), Barbora Štěpánková (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Czech Republic), Z Steyn (SADiLaR, South Africa), Regina Stodden (Semantic Computing Group, University Bielefeld, Germany), Matthew Stone (Rutgers University, United States), Jana Straková (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Czech Republic), Marco Antonio Stranisci (University of Turin, Italy), Carlo Strapparava (FBK-irst, Italy), Jacob Striebel (Indiana University, United States), Kristina Striegnitz (Union College, United States), Tomek Strzalkowski (RPI, United States), Sebastian Stüker (Zoom Video Communications Inc., Germany), Sara Stymne (Uppsala University, Sweden), Qi Su (Peking University, China), Xuanyu Su (University of Ottawa, Canada), Katsuhito Sudoh (Nara Women's University, Japan), Saku Sugawara (National Institute of Informatics, Japan), Hiroaki Sugiyama (NTT Communication Science Labs., Japan), Chanchal Suman (NIT Warangal, India), Zengkui Sun (Beijing Jiaotong university, China), Wenjun Sun (La Rochelle University, France), Ling Sun (Indiana University Bloomington, United States), Chengjie Sun (Harbin Institute of Technology, China), Haibo Sun (Brandeis University, United States), Vijay Sundar Ram (AU-KBC, Anna University, Chennai, India), Thepchai Supnithi (National Electronics and Computer Technology Center, Thailand), Marek Suppa (Comenius University in Bratislava, Slovakia), Richard Sutcliffe (University of Essex, United Kingdom), Yu Suzuki (Gifu University, Japan), Torbjorn Svendsen (NTNU, Norway), Ekaterina Sviridova (Université Côte d'Azur, France), Colin Swaelens (Ghent University, Belgium), Sandesh Swamy (Amazon, United States), Daniel Swanson (Indiana University, United States)

[T] Santosh T.Y.S.S (Technical University of Munich, Germany), Anais Tack (KU Leuven; imec, Belgium), Marko Tadić (University of Zagreb, Faculty of Humanities and Social Sciences, Croatia), Marie Tahon (LIUM / Le Mans University, France), Dima Taji (Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Czech Republic), Tetsuro Takahashi (Kagoshima University, Japan), Shinnosuke Takamichi (Keio University, Japan), Hiroya Takamura (The National Institute of Advanced Industrial Science and Technology (AIST), Japan), Koichi Takeuchi (Okayama University, Japan), Toshiyuki Takezawa (Hiroshima City University, Japan), Ece Takmaz (Utrecht University, Netherlands), Kossay Talmoudi (ELDA, France), Fabio Tamburini (FICLIT - University of Bologna, Italy), Yunik Tamrakar (Colorado State University, United States), Akihiro Tamura (Doshisha University, Japan), Liling Tan (Apple, United States), Hongye Tan (School of Computer Science and Information Technology, Shanxi University, China), Qiaoyu Tang (Institute of Software, Chinese Academy of Sciences, China), Zheng Tang (University of Arizona, United States), Gongbo Tang (Beijing Language and Culture University, China), Xavier Tannier (Limics, Sorbonne Université, France), Mingxu Tao (Peking University, China), Shiva Taslimipour (University of Cambridge, United Kingdom), Mariona Taule (University of Barcelona, Spain), Mirko Tavosanis (Università di Pisa, Italy), Andon Tchechmedjiev (IMT Mines Alès, France), Elke Teich (Universität des Saarlandes, Germany), António Teixeira (DETI/IEETA, University of Aveiro, Portugal), Selma Tekir (Izmir Institute of Technology, Turkey), Irina Temnikova (Big Data for Smart Society Institute (GATE), Bulgaria, Bulgaria), Daniela Teodorescu (University of Alberta, Canada), Takehiro Teraoka (Takushoku University, Japan), Andrea Tettamanzi (Université Côte d'Azur, France), Arda Tezcan (LT3, Language and Translation Technology Team, Ghent University, Belgium), Medari Tham (St. Anthony's College Shillong, India), Thomas Thebaud (Johns Hopkins University, United States), Sarubi Thillainathan (Saarland University, Germany), Thulasi Tholeti (Northeastern University, United States), Craig Thomson (ADAPT / Dublin City University, Ireland), Camilo Thorne (Elsevier, Germany), Carole Tiberius (Instituut voor de Nederlandse Taal, Netherlands), Joerg Tiedemann (University of Helsinki, Finland), Tiago Timponi Torrent (Federal University of Juiz de

Fora, Brazil), Tarık Tıraş (Boğaziçi University, Turkey), Abhisek Tiwari (IIT Patna, India), Amalia Todorascu (LiLPa, University of Strasbourg, France), Maria Todorova (DCL, IBL, BAS, Bulgaria), Nasim Tohidi (Computer Science Department, Wayne State University, United States), Ryoko Tokuhisa (AICHI INSTITUTE OF TECHNOLOGY, Japan), David Tomás (University of Alicante, Spain), Natalia Tomashenko (Université de Lorraine, CNRS, Inria, LORIA, France), Nadi Tomeh (LIPN-CNRS, Université Sorbonne Paris Nord, France), Manuel Tonneau (University of Oxford, World Bank, United Kingdom), Yukio Tono (Tokyo University of Foreign Studies, Japan), Pedro Torres-Carrasquillo (MIT Lincoln Laboratory, United States), Juan-Manuel Torres-Moreno (LIA Avignon, France), Samia Touileb (University of Bergen, Norway), Thi Hong Hanh Tran (Arkhn, France), Thi Hong Hanh Tran (ARKHN, France), Isabel Trancoso (INESC-ID / IST Univ. Lisbon, Portugal), Diana Trandabat (University Alexandru Ioan Cuza of Iasi, Romania, Romania), Beata Trawinski (IDS Mannheim, Germany), David Tresner-Kirsch (Brandeis University, United States), Thorsten Trippel (Leibniz-Institut für Deutsche Sprache, Germany), Jan Trmal (AppTek.ai, Germany), Keith Trnka (Currently unaffiliated, United States), Cassia Trojahn (UGA, France), Benito Trollip (South African Centre for Digital Language Resources (SADiLaR), North-West University, South Africa), Jürgen Trouvain (Saarland University, Germany), Khiet Truong (Radboud University, Netherlands), Vivian Tsang (Quillsoft Ltd., Canada), Yuen-Hsien Tseng (National Taiwan Normal University, Taiwan), Michael Tseng (Google DeepMind, United States), Yu-Hsiang Tseng (Department of Linguistics, University of Tübingen, Germany), Prodrornos Tsiavos (OpenAIRE, Greece), Chara Tsoukala (Athena Research Center, Greece), Erlend Tøssebro (University of Stavanger, Norway), Masatoshi Tsuchiya (Toyohashi University of Technology, Japan), Jun'ichi Tsujii (Artificial Intelligence Research Centre at AIST, Japan), Evgenii Tsybalov (Amazon AGI, Germany), Dan Tufis (Research Institute for Artificial Intelligence, Romanian Academy, Romania), Gabriele Tuozzo (Università degli Studi di Salerno, Italy), Marco Turchi (Zoom Communications, Italy), Jordi Turmo (Universitat Politècnica de Catalunya (UPC), Spain), Na'im Tyson (New York University, United States)

[U] Kiyotaka Uchimoto (National Institute of Information and Communications Technology, Japan), Kosei Uemura (University of Toronto, Canada), Anuradha Uggi (IIT Hyderabad, India), Elaine Uí Dhonnchadha (Trinity College Dublin, Ireland), Stefan Ultes (University of Bamberg, Germany), Nureddin Unal (Bogazici University, Turkey), Zdenka Uresova (Charles University, Czech Republic), Elias Urios (Universitat Politècnica de València, Spain), Yasuyuki Usuda (Shizuoka Institute of Science and Technology, Japan), Masao Utiyama (NICT, Japan)

[V] Anna Vacalopoulou (ILSP / Athena R.C., Greece), Jurgita Vaicenoniene (Vytautas Magnus University, Lithuania), Avijit Vajpayee (Amazon, United States), Thomas Vakili (Department of Computer and Systems Sciences, Stockholm University, Sweden), Oto Vale (Universidade Federal de São Carlos, Brazil), Rafael Valencia-Garcia (Universidad de Murcia, Spain), Baiba Valkovska (IMCS, University of Latvia, Latvia), María Estrella Vallecillo-Rodríguez (Universidad de Jaén, Spain), Gisela Vallejo (The University of Melbourne, Australia), Henk van den Heuvel (CLS/CLST, Radboud University Nijmegen, Netherlands), Rob van der Goot (IT University of Copenhagen, Denmark), Chris van der Lee (Tilburg University, Netherlands), Marieke van Erp (KNAW Humanities Cluster, Netherlands), Daan van Esch (Google Research, Netherlands), Rik van Noord (University of Groningen, Netherlands), Jens Van Nooten (University of Antwerp, Belgium), Hannah VanderHoeven (Colorado State University, United States), Tamás Váradi (ELTE Research Centre for Linguistics, Hungary), Francielle Vargas (University of São Paulo, Brazil), Vasudeva Varma (IIIT Hyderabad, India), Amparo Varona (University of the Basque Country, Spain), Ioana Vasilescu (LISN CNRS, France), Juan Vasquez (Department of Computer Science, University of Colorado Boulder, United States), Pedro O.S Vaz-de-Melo (UFMG, Brazil), Gloria Vazquez Garcia (Universitat de Lleida, Spain), Eva Maria Vecchi (Universitat Stuttgart, Institut für Maschinelle Sprachverarbeitung, Germany), Yana Veitsman (Universität Göttingen, Germany), Aswathy Velutharambath (University of Stuttgart, Germany), Marianne Vergez-Couret

(Université Paul Valéry Montpellier, France), Marc Verhagen (Brandeis University, United States), Ghanshyam Verma (University of Galway, Ireland), Bhuvanesh Verma (Goethe University Frankfurt, Germany), Federica Vezzani (University of Padua, Italy), Renata Vieira (Évora University, Portugal), Gautham Vijay Kumar (Florida International University, United States), Jesús Vilares (Universidade da Coruña, Centro de Investigación en Tecnologías de la Información y las Comunicaciones, Spain), David Vilares (Universidade da Coruña, CITIC, Spain), Manuel Vilares Ferro (University of Vigo, Spain), Jesus Villalba (Johns Hopkins University, United States), Jeanne Villaneau (IRISA Université de Bretagne Sud, France), Esau Villatoro-Tello (Idiap Research Institute, Switzerland), Marta Villegas (Barcelona Supercomputing Center, Spain), Veronika Vincze (University of Szeged, Hungary), Spela Vintar (Jožef Stefan Institute, Slovenia), Marco Viviani (University of Milano-Bicocca, Italy), Dr. Aggeliki Vlachostergiou (National Technical University of Athens(N.T.U.A.), Researcher and Data Scientist, Niometrics Hellas, Project Manager at EXUS AI Labs, Greece), Carl Vogel (Trinity College Dublin, Ireland), Rob Voigt (University of California, Davis, United States), Elena Volodina (University of Gothenburg, Sweden), Pius von Däniken (Zurich University of Applied Sciences ZHAW, Switzerland), Piek Vossen (Vrije Universiteit Amsterdam, Netherlands), Ekaterina Vylomova (University of Melbourne, Australia)

[W] Glorianna Wagner-Jagfeld (LEP AG, Switzerland), Abigail Walsh (ADAPT Centre / Dublin City University, Ireland), Mingyu Wan (The Hong Kong Polytechnic University, China), Shouli Wang (Harbin Institute of Technology, China), Xiaozhi Wang (Tsinghua University, China), Ze-qiang Wang (University of Surrey, United Kingdom), Haochen Wang (Nantong University, China), Xuancong Wang (MOH Office for Healthcare Transformation, Singapore), Nan Wang (Sun Yat-sen University, China), Qingyun Wang (William & Mary, United States), Yu Wang (The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Xintong Wang (University of Hamburg, Germany), Baoxin Wang (Joint Laboratory of HIT and iFLYTEK(HFL), iFLYTEK Research, China), Xing Wang (TikTok, China), Jin Wang (Yunnan University, China), Tao Wang (Haihe Laboratory of Brain-Computer Interaction and Human-Machine Integration, China), Hao Wang (Sichuan University, China), Yumeng Wang (Leiden University, Netherlands), Zimu Wang (University of Liverpool, United Kingdom), Ping Wang (Stevens Institute of Technology, United States), Jiexin Wang (The School of Software Engineering, South China University of Technology, China), Guanghua Wang (Kean University, United States), Di Wang (Meta Inc., United States), Dan Wang (JP Morgan, United States), Xun Wang (Microsoft, United States), Xianzhi Wang (University of Technology Sydney, Australia), Hsin-Min Wang (Academia Sinica, Taiwan), Yunli Wang (National Research Council Canada, Canada), Nigel Ward (University of Texas at El Paso, United States), Christian Wartena (Hochschule Hannover - University of Applied Sciences and Arts, Germany), Nick Webb (Union College, United States), Bonnie Webber (University of Edinburgh, United Kingdom), Peipei Wei (Amazon, United States), Wei Wei (Huazhong University of Science and Technology, China), Chengkun Wei (Zhejiang University, China), Albert Weichselbraun (University of Applied Sciences of the Grisons, Switzerland), Simon Wells (Edinburgh Napier University, United Kingdom), Melvin Wevers (University of Amsterdam, Netherlands), David Wible (Tunghai University, Taiwan), Linda Wiechetek (UiT Norgga Árkttálaš universitehta, Norway), Matthew Wiesner (Johns Hopkins University, United States), John Wieting (University of Illinois; TTI-Chicago; CMU; Google, United States), Patrick Wilken (AppTek, Germany), Dexter Williams (University of Wisconsin-Madison, United States), Menzo Windhouwer (KNAW Humanities Cluster, Netherlands), Bartholomaeus Wloka (University of Vienna, Austria), Rosalee Wolfe (Institute for Language and Speech Processing, ATHENARC, Greece), Marcin Woliński (Institute of Computer Science, Polish Academy of Sciences, Poland), Diedrich Wolter (University of Lübeck, Germany), Christa Womser-Hacker (University of Hildesheim Information Science, Germany), Kam-Fai Wong (Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong, Hong Kong Special Administrative Region of China), Derek F. Wong (University of Macau, Macau), Dina Wonsever (Universidad de la República, Uruguay), Sylvia Woodrose Schwartz (George Mason University, United States),

Jonathan Wright (University of Pennsylvania, United States), Jialun Wu (Northwestern Polytechnical University, China), Shengqiong Wu (National University of Singapore, Singapore), Yaru WU (CRISCO/UR4255, Université de Caen Normandie, 14000 Caen, France, France), Xianchao Wu (NVIDIA, Japan), Shih-Hung Wu (Chaoyang University of Technology, Taiwan), Winston Wu (University of Hawaii at Hilo, United States), Amelie Wühl (IT University of Copenhagen, Denmark)

[X] Zhaomin Xiao (University of North Texas, United States), Tong Xiao (Northeastern University, China), Yunhe Xie (Harbin Institute of Technology, China), Kaige Xie (Georgia Institute of Technology, United States), Lu Xu (Sapienza University of Rome, Italy), Wang Xu (Tsinghua University, China), Ziwei Xu (Institute of Science Tokyo, Japan), Hao Xu (Harvard Medical School, United States), Yueshen Xu (Xidian University, China), Qihui Xu (Ohio State University, United States), Dongkuan Xu (North Carolina State University, United States), Wenduan Xu (Quantinuum/Cambridge Quantum, United Kingdom), Guangyue Xu (Target, United States), Wei Xue (Hong Kong University of Science and Technology, Hong Kong Special Administrative Region of China)

[Y] Srishti Yadav (University of Copenhagen, Denmark), Masaru Yamada (Rikkyo University, Japan), Masaya Yamaguchi (National Institute for Japanese Language and Linguistics, Japan), Chenwei Yan (University of International Business and Economics, China), Kailai Yang (University of Manchester, United Kingdom), Yuting Yang (Boston Children's Hospital, Harvard Medical School, United States), Yifei Yang (Shanghai Jiao Tong University, China), Zachary Yang (McGill | Mila | Ubisoft, Canada), Muyun Yang (Harbin Institute of Technology, China), Shu-wen Yang (National Taiwan University, Taiwan), Huichen Yang (CSIRO, Australia), Xin Yao (iFLYTEK Research, China), Yao Yao (The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Yusuke Yasuda (National Institute of Informatics, Japan), Mert Yazan (Hogeschool van Amsterdam, Netherlands), Bingyang Ye (Brandeis University, United States), Mohammad Yeghaneh Abkenar (University of Potsdam, Germany), Gerard Yeo (National University of Singapore, Singapore), Seid Muhie Yimam (University of Hamburg, Germany), Zhangyue Yin (Fudan University, China), Hikaru Yokono (Meisei University, Japan), Jiawei Yong (TOYOTA MOTOR CORPORATION, Japan), Hiyon Yoo (University Paris Cite, Laboratoire de Linguistique Formelle, France), Minoru Yoshida (Tokushima University, Japan), Koichiro Yoshino (Institute of Science Tokyo / RIKEN GRP / Nara Institute of Science and Technology, Japan), Masaharu Yoshioka (Hokkaido University, Japan), Xuemin Yu (Dalhousie University, Canada), Xinchun Yu (University of Arizona, United States), Dian Yu (Tencent AI Lab, United States), Liang-Chih Yu (Yuan Ze University, Taiwan), Hao Yu (Dalian University of Technology, China), Heng Yu (Beijing Normal University, China), Yi Yu (Kyoto University, Japan), Yu Yuan (College of Foreign Languages, Shanghai Maritime University, China), Jiahong Yuan (University of Science and Technology of China, China), Frances Yung (Saarland University, Germany), François Yvon (ISIR CNRS & Sorbonne Université, France)

[Z] Zdeněk Žabokrtský (Charles University, Czech Republic), Annie Zaenen (Stanford University, United States), Aleš Žagar (University of Ljubljana, Faculty of Computer and Information Science, Slovenia), Wajdi Zaghouani (Northwestern University Qatar, Qatar), Zbynek Zajic (University of West Bohemia, Czech Republic), Mahdi Zakizadeh (Tehran Institute for Advanced Studies, Iran (Islamic Republic of)), Roberto Zamparelli (Università di Trento, Italy), Marcos Zampieri (George Mason University, United States), Massimo Zancanaro (University of Trento & Fondazione Bruno Kessler (FBK), Italy), Marcelly Zanon Boito (NAVER LABS Europe, France), Alessandra Zarcone (Technische Hochschule Augsburg, Germany), Gian Piero Zarri (STIH Laboratory, University Paris-Sorbonne, France), Natalia Zawadzka-Paluekta (Institute of Computer Science, Polish Academy of Sciences, Poland), Mohammad Zeineldeen (RWTH Aachen University / AppTek, Germany), Weixin Zeng (National University of Defense Technology, China), Torsten Zesch

(Computational Linguistics, FernUniversität in Hagen, Germany), Rodolfo Zevallos (Universitat Pompeu Fabra, Spain), Yu Zhai (The Hong Kong Polytechnic University, China), Qujiaheng Zhang (Target, United States), Ran Zhang (University of Mannheim, Germany), Ziyang Zhang (Soochow University, China), Dong Zhang (Soochow University, China), Yuji Zhang (University of Illinois Urbana Champaign, United States), Yan Zhang (Peking University, China), Xulang Zhang (Nanyang Technological University, Singapore), Yuji Zhang (University of Illinois Urbana-Champaign, United States), Ningyu Zhang (Zhejiang University, China), Zhisong Zhang (Carnegie Mellon University, United States), Mike Zhang (University of Copenhagen, Denmark), Yang Zhang (Southwestern University of Finance and Economics, China), Xiaowang Zhang (Tianjin University, China), Zhuosheng Zhang (Shanghai Jiao Tong University, China), Chen Zhang (Peking University, China), Tianlin Zhang (The University of Manchester, United Kingdom), Xiaojun Zhang (Xi'an Jiaotong-Liverpool University, China), Qing Zhang (Beijing Key Laboratory of Key Technologies for AI+ Domain Applications, North China University of Technology, China), Wei Emma Zhang (The University of Adelaide, Australia), Xiaojing Zhao (The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China), Xiaoyan Zhao (The Chinese University of Hong Kong, China), Bing Zhao (SRI International, United States), Jin Zhao (Brandeis University, United States), Qinghua Zhao (Beihang University, China), Dequan Zheng (Harbin University of Commerce, China), Liu Zhifu (China Three Gorges University, China), Xian Zhong (Wuhan University of Technology, China), Zhenkun Zhou (School of Artificial Intelligence, Capital University of Economics and Business, China), Houquan Zhou (Soochow University, China), Zhengyang Zhou (Brandeis University, United States), Ying Zhou (Shandong University, China), Yongxin Zhou (Université Grenoble Alpes, France), He Zhou (The Hong Kong Polytechnic University, China), Junwei Zhou (Wuhan University of Technology, China), Dong Zhou (Guangdong University of Foreign Studies, China), Yifan Zhu (Brandeis University, United States), Charlotte Zhu (Indiana University Bloomington, United States), Luyao Zhu (National University of Singapore, Singapore), Muhua Zhu (Meituan Group, China), Lichao Zhu (Paris Cité University, France), Shaolin Zhu (Tianjin university, China), Angelo Ziletti (Bayer AG, Germany), Steven Zimmerman (Shell, Austria), Claus Zinn (University of Tübingen, Germany), Imed Zitouni (Meta, United States), Nassreddine Znaidi (Univ. Lille, ULR 4073 - GERiiCO, F-59000 Lille, France, France), Michael Zock (CNRS-LIS, France), Xinyu Zuo (Tencent, China), Pierre Zweigenbaum (LISN, CNRS, Université Paris-Saclay, France), Mark-Matthias Zymla (University of Konstanz, Germany).

Table of Contents

<i>Beyond Generic Responses: Target-Aware Strategies for Countering Hate Speech</i> Yen-Yu Chang, Daryna Dementieva and Alexander Fraser	1
<i>Topic-Initiator: A Proactive Chatbot with Personalized Topic RAG for Enhancing Willingness to Converse</i> Kazuya Matsuo, Atsushi Otsuka, Narichika Nomoto and Makoto Nakatsuji	19
<i>CoachLah: A Singlish–English Parallel Corpus of Health Coaching Conversations with Behavior Goal Annotations</i> Iva Bojic, Mathieu Ravaut, Stephanie Hilary Xinyi Ma, Doreen Tan, Andy Hau Yan Ho and Andy Khong	35
<i>Faithful Medical Dialogue Generation Using Homo-Heterogeneous Exemplar-based In-Context Knowledge Grounding</i> Priyanshu Priya, Hardik Goyal and Asif Ekbal	50
<i>Investigating Proactivity in Multimodal Task-Guidance Dialogues</i> Sofia Brenna, Elisabetta Jezek, Matthias Kraus and Bernardo Magnini	64
<i>Investigating How LLMs Propagate Female Stereotypes: Comparing What Models Say via Prompts with What They Represent in Their Embeddings</i> Andrea Valderrey Nuñez and Jelke Bloem	77
<i>Why So Separate: Analyzing In-Context Learning from a Vector Space Perspective</i> Tobias Kalmbach and Sandipan Sikdar	93
<i>Explaining Explanations: Interpretability Methods for Discourse Analysis of Transformer Attention Maps</i> Louis Escoufflaire, Jérémie Bogaert, Antonin Descampe, Cédric Fairon and Francois-Xavier Standaert	107
<i>TempPerturb-Eval: On the Joint Effects of Internal Temperature and External Perturbations in RAG Robustness</i> Yongxin Zhou, Philippe Mulhem and Didier Schwab	117
<i>Refusal Steering: Fine-grained Control over LLM Refusal Behaviour for Sensitive Topics</i> Iker García-Ferrero, David Montero and Roman Orus	128
<i>To Predict or Not to Predict? Towards Reliable Uncertainty Estimation in the Presence of Noise</i> Nouran Khallaf and Serge Sharoff	152
<i>An Extreme Multi-label Text Classification (XMTC) Library Dataset: What If We Took "Use of Practical AI in Digital Libraries" Seriously?</i> Jennifer D'Souza, Sameer Sadruddin, Maximilian Kaehler, Andrea Salfinger, Luca Zaccagna, Francesca Incitti, Lauro Snidaro and Osmo Suominen	169
<i>A Historical Database for the Study of Obstruent-Lateral Palatalization in Ibero-Romance</i> Andrea García Covelo	185

<i>Is Clinical Text Enough? A Multimodal Study on Mortality Prediction in Heart Failure Patients</i> Oumaima El Khettari, Virgile Barthet, Guillaume Hocquet, Joconde Weller, Emmanuel Morin and Pierre Zweigenbaum	194
<i>HistoriQA-ThirdRepublic: Multi-Hop Question Answering Corpus for Historical Research, Parliamentary Debates from the French Third Republic (1870-1940)</i> Aurelien Pellet, Marie Anna Puren and Julien PEREZ	207
<i>BURMESE-SAN: Burmese NLP Benchmark for Evaluating Large Language Models</i> Thura Aung, Jann Railey Montalan, Jian Gang Ngui and Peerat Limkonchotiwat	224
<i>Assessing the Political Fairness of Multilingual LLMs: A Case Study Based on a 21-Way Multi-parallel EuroParl Dataset</i> Paul Lerner and François Yvon	246
<i>AfriStereo: A Culturally Grounded Dataset for Evaluating Stereotypical Bias in Large Language Models</i> Yann Le Beux, Oluchi Audu, Oche David Ankeli, Dhananjay Balakrishnan, Melissah Weya, Marie Daniella Ralaiarinosy and Ignatius Ezeani	266
<i>Judging Instruction Responses in a Low-Resource Language: A Case Study on Basque</i> David Ponce, Harritxu Gete, Thierry Etchegoyhen, Irune Zubiaga and Aitor Soroa	281
<i>Appeal, Align, Divide? Stance Detection for Group-Directed Messages in German Parliamentary Debates</i> Ines Rehbein, Maris Leander Buttmann, Julian Schlenker and Simone Paolo Ponzetto	299
<i>Report-based Recommendations for Policy Making and Agency Operations: Dataset and LLM Evaluation</i> Aleksandra Edwards, Thomas Edwards, Jose Camacho-Collados and Alun Preece ..	319
<i>ConceptKT: A Benchmark for Concept-Level Deficiency Prediction in Knowledge Tracing</i> Yu-Chen Kang, Yu-Chien Tang and An-Zi Yen	333
<i>Open-access Dataset on Acceptability Ratings of Korean Clausal Constructions by Humans and GPT Models</i> Gyu-Ho Shin, Soo-Hwan Lee and Chanyoung Lee	344
<i>Talk2Ref: A Dataset for Reference Prediction from Scientific Talks</i> Frederik Yannick Broy, Maike Züfle and Jan Niehues	357
<i>MuSaG: A Multimodal German Sarcasm Dataset with Full-Modal Annotations</i> Aaron Robert Scott, Maike Züfle and Jan Niehues	372
<i>Icelandic Math Eval: A Competitive Mathematics Benchmark for Large Language Models</i> Hafsteinn Einarsson, Jökull Ari Haraldsson, Ívar Armin Derayat, Sigrún Helga Lund and Benedikt Steinar Magnússon	393
<i>MazeEval: A Benchmark for Testing Sequential Decision-Making in Language Models</i> Hafsteinn Einarsson	407

<i>J-ClinicalBench: A Benchmark for Evaluating Large Language Models on Practical Clinical Tasks in Japanese</i>	
Seiji Shimizu, Tomohiro Nishiyama, HISADA Shohei, Yamato Himi, Shoko Wakamiya, Yuki Yanagisawa, Masami Tsuchiya, Satoko Hori and Eiji ARAMAKI	419
<i>Is One Dataset Enough for Evaluation? Studying Generalizability of Automated Essay Scoring Models</i>	
Sohaila Eltanbouly, Marwan Sayed and Tamer Elsayed	431
<i>HiFi-KPI: A Dataset for Hierarchical KPI Extraction from Earnings Filings</i>	
Rasmus Thyge Aavang Jensen, Giovanni Rizzi, Rasmus Tjalk-Bøggild, Alexandre Iolov, Mike Zhang and Johannes Bjerva	441
<i>UniSkill: A Dataset for Matching University Curricula to Professional Competencies</i>	
Nurlan Musazade, József Mezei and Mike Zhang	456
<i>A Dataset for Evaluating ASR on Specialized Vocabulary</i>	
Emily Haubert Klering, Eduardo Gabriel Cortes, Tatjana Chernenko, Mariana Vargas Trarbach, Gabriel de Oliveira Ramos, Sandro José Rigo, Maitê Dupont, Ana Luiza Treichel Vianna, Gabriela Krause dos Santos, Vinicius Meirelles Pereira, Denis Andrei de Araujo and Rafael Kunst	470
<i>SommBench: Assessing Sommelier Expertise of Language Models</i>	
William Brach, Tomas Bedej, Jacob Nielsen, Jacob Pichna, Juraj Bedej, Eemeli Saarensilta, Julie Dupouy, Gianluca Barmina, Andrea Blasi Núñez, Peter Schneider-Kamp, Kristian Košťál, Michal Ries and Lukas Galke Poech	481
<i>CzechDocs: A Multiway Parallel Dataset of Formatted Documents for Minority Languages in Czechia</i>	
Josef Jon and Ondřej Bojar	498
<i>An LLM-Based Assistant for Debt Waiver Court Procedures</i>	
Lluís Padro, Daniel Ferrés, Roser Saurí and Mireia Artigot	505
<i>Enhancing Clinical Trial Analysis through Large Language Models for Multi-Evidence Natural Language Inference</i>	
Shobanapriyan Chandrasegaran and Amal Htait	515
<i>A Systematic Comparison of Large Language Models for Data Annotation in NER Tasks</i>	
Muhammad Uzair Ul Haq, Davide Rigoni and Alessandro Sperduti	524
<i>Toward Generalized Cross-Lingual Hateful Language Detection with Web-Scale Data and Ensemble LLM Annotations</i>	
Dang Hai Dang, Jelena Mitrović and Michael Granitzer	549
<i>Can LLMs Faithfully Explain Themselves in Low-Resource Languages? A Case Study on Emotion Detection in Persian</i>	
Mobina Mehrazar, Mohammad Amin Yousefi, Parisa Beygi and Behnam Bahrak	560
<i>Are LLMs Good Text Diacritizers? An Arabic and Yoruba Case Study</i>	
Hawau Olamide Toyin, Samar Mohamed Magdy and Hanan Aldarmaki	580

<i>Automatic Suggestions of Supplements in the Herculaneum Papyri: Language Models and RESTful API</i>	
Angelo Mario Del Grosso, Gabriele Giannessi, Simone Zenzaro and Federico Boschetti	590
<i>Designing LLM Agents for User-Centered Language Service Selection</i>	
Ryoichiro Ogawa, Donghui Lin and Fumito Uwano	599
<i>User Profiling for Specification-Sensitive Recommendations with Large Language Model Prompting</i>	
Chih-Yu Chien, An-Zi Yen, Hen-Hsen Huang and Hsin-Hsi Chen	609
<i>Comparing Traditional and LLM-based Approaches for Automated Scoring of Dutch Writing Products</i>	
Joni Kruijsbergen and Orphee De Clercq	619
<i>"Decode the Law": Towards Legal Text Simplification with Large Language Models</i>	
Mohammed Danish Rabbani, Subhadeep Roy, Sayantan Mitra and Tulika Saha	631
<i>CLASE: A Hybrid Method for Chinese Legalese Stylistic Evaluation</i>	
Yiran Rex Ma, Yuxiao Ye and Huiyuan Xie	642
<i>Neural Network-assisted Analysis of Tube Vocal Tract Models</i>	
Runhui Song, Johan Sjons and Axel G. Ekstrom	654
<i>Central Kurdish Text-to-Speech and Its Application in Speech-to-Text Translation</i>	
Mohammad Mohammadamini, Meysam Shamsi and Marie Tahon	664
<i>QuALA-NL: Question & Answer with Legal Attribution in Dutch</i>	
Romy A.N. van Drie, Roos M. Bakker, Daan L. Di Scala and Maaïke de Boer	674
<i>SouDeC: Source Detection and Classification in Czech</i>	
Jiří Mírovský and Barbora Hladka	685
<i>Frame Semantic Patterns for Identifying Underreporting of Notifiable Events in Healthcare: The Case of Gender-Based Violence</i>	
Lívia Dutra, Arthur Lorenzi, Lais Berno, Franciany Campos, Karoline Biscardi, Kenneth Brown, Marcelo Viridiano, Frederico Belcavello, Ely E. Matos, Olivia Guaranha, Erik Santos, Sofia Reinach and Tiago Timponi Torrent	694
<i>PrePPER: A Preference Pattern-based Profiling Framework for Explainable Recommendation</i>	
Taisuke Usumi, Akiko Masaki, Sanae Muramatsu, Akira Sakamoto and Takeharu Eda	705
<i>Evaluating the Impact of Source Diversity for RAG in Historical Research</i>	
Ruhi Mahadeshwar, Andreas van Cranenburgh, Tommaso Caselli and Malvina Nissim	716
<i>Automatic Essay Scoring and Feedback Generation in Basque Language Learning</i>	
Ekhi Azurmendi, Xabier Arregi and Oier Lopez de Lacalle	735
<i>Paragraph Segmentation Revisited: Towards a Standard Task for Structuring Speech</i>	
Fabian Retkowski and Alexander Waibel	747

<i>High-Order Question Generation in a Multilingual Educational Context</i> Suna Şeyma Uçar, Itziar Aldabe, Nora Aranberri and Orphee De Clercq	760
<i>From Print to Digital and beyond: The Retrodigitization of a Historical Dictionary of Italian as a Hybrid Lexical Resource</i> Marco Biffi, Sebastiana Cucurullo, Manuel Favaro, Elisa Guadagnini, Simonetta Montemagni and Eva Sassolini	770
<i>Learning through News: Bridging the Gap between Algorithmic Recommendation and Human Curation</i> Florian Debaene, Loic De Langhe, Orphee De Clercq and Veronique Hoste	778
<i>MaskedVerbalizer: Automatic Verbalizer Construction for Few-Shot Text Classification in Low-Resource Right-to-Left Languages</i> Faizad Ullah, Furqan Sikandar, Areeba Waqar, Faizan Ali, Muhammad Sohaib Ayub, Mubashar Mushtaq and Asim Karim	795
<i>RBR: RAG-Based Open-Domain Question Answering Using a Ranking Approach to Document Retrieval</i> Priyatam Sai Naravajhula and Vincent Ng	805
<i>Sentence-Level Back-Transliteration of Romanized Indian Languages: Performance Analysis and Challenges</i> Saurabh Kumar, Dhruvkumar Babubhai Kakadiya, Sanasam Ranbir Singh and Sukumar Nandi	818
<i>Cross-Corpus CEFR Classification through Artificial Learners Perplexities</i> bernardo stearns, John P. McCrae and Thomas Gaillat	828
<i>CorpusClues: Scalable Unsupervised Similarity Search for Historical Texts Using MinHash-LSH</i> Paulien Lemay, Klaas Bentein and Els Lefever	838
<i>BenCSSmark: Making the Social Sciences Count in LLM Research</i> Arnault Chatelain, Etienne Ollion, Qianwen Guan, Diandra Fabre, Lorraine Goeuriot, emile chapuis, Abdelkrim Beloued, Marie Candito, Nicolas Hervé and Didier Schwab	848
<i>Predicting Topic (Co-)Occurrence Using Topic Networks Built from the Project Gutenberg Corpus</i> Bhuvanesh Verma and Alexander Mehler	860
<i>AraHopeCorpus: Annotation Guidelines and Dataset for Hope Speech in Arabic Social Media Crisis Discourse</i> Esra'a Ahmad Sharqawi and Wajdi Zaghouani	870
<i>Cohesion-6K: An Arabic Dataset for Analyzing Social Cohesion and Conflict in Online Discourse</i> Aisha Ali Al-Athba and Wajdi Zaghouani	881
<i>Reference-free Evaluation at Inference for NER/NEL over OCRed Historical Texts</i> Tien-Nam Nguyen, Adam Jatowt, Ahmed Hamdi, Mickael Coustaty, Thi Hong Hanh Tran and Antoine Doucet	893
<i>Echoes of the Troubadours: A Corpus of Troubadour Poetry for Stylometric Analysis and Authorship Attribution</i> Loic De Langhe, Orphee De Clercq and Veronique Hoste	905

<i>Gretino: A Greek and Latin Dataset to Benchmark Retrieval Systems in Classical Languages</i> Hawau Olamide Toyin, Federico Iezzi, Elia Scapini, Giulio Federico and Giovanni Puccetti	919
<i>A Recipe for Adapting Multilingual Embedders to OCR-Error Robustness and Historical Texts</i> Andrianos Michail, Stylianos Psychias, Juri Opitz and Simon Clematide	929
<i>Phrase-Level Segmentation on Medieval Corpora for Aligning Multilingual Texts</i> Lucence Ing, Matthias Gille Levenson and Carolina Macedo	936
<i>RAGE: Roman and Greek Emotions</i> Frederick Riemenschneider, Jonathan D. Geiger, Thomas Kuhn-Treichel and Anette Frank	947
<i>From Variance to Invariance: Qualitative Content Analysis for Narrative Graph Annotation</i> Junbo Huang, Max Weinig, Ulrich Fritsche and Ricardo Usbeck	958
<i>A Dataset of Historical Medical Periodicals Annotated with Textual Genre</i> Vera Danilova and Sara Stymne	973
<i>Preserving Endangered Linguistic Heritage: Developing a Corpus for the Study of Contact- induced Changes in Corfioto</i> Giorgio Maria Di Nunzio and Georgios Vardakis	985
<i>To Eat and beyond: A FrameNet-Inspired Annotation of Food and Its Uses over Time</i> Teresa Paccosi, Gauri Bhagwat and Marieke van Erp	997
<i>To Overfit or Not to Overfit? An Evaluation of HTR Workflow on 17Th-18Th Century French Corpus</i> Marine Tiger	1009
<i>Automatic Segmentation of Classical Tibetan Texts into Autochthonous and Allochthonous Re- gions</i> Guy Bilitski, Lev Shechter, Sonam Jamtsho, Nir Marciano, Nicola Bajetta, Rebecca Sun- den, Omri Drori, Kai Golan Hashiloni, Orr Zwebner, Asaf Shina, Orna Almogi, Dorji Wangchuk and Kfir Bar	1017
<i>RespondeoQA: A Benchmark for Bilingual Latin-English Question Answering</i> Marisa Hudspeth, Patrick J. Burns and Brendan O'Connor	1031
<i>Transformer-Enabled Diachronic Analysis of Vedic Sanskrit: Neural Methods for Quantifying Types of Language Change</i> Ananth A. Hariharan and David R. Mortensen	1044
<i>Ithaca Revisited: Benchmarking a Domain-Specific Model for Epigraphy in the Age of LLMs</i> Alessandro Locaputo, Andrea Brunello, Nicola Saccomanno, Paraskevi Platanou and Giuseppe Serra	1054
<i>Beyond Literal Meaning: How LLMs Interpret Yemeni Proverbs</i> Nasser Thmer, Ali Al-Laith and Muhammad Shoalb	1071
<i>CEFR Level Prediction for Short Russian L2 Texts: Evaluating Classifiers and Instruction-Based LLMs</i> Anna Glazkova, Antonina Laposhina and Dmitry Morozov	1081

<i>Evaluation of Document-Level Text Simplification in Japanese</i> Iori Yamashita, Hikari Tanaka, Hajime Kiyama, Kexin Bian, Zhousi Chen and Mamoru Komachi	1092
<i>Parallel Corpus Filtering Based on Semantic Similarity and Surface Dissimilarity for Japanese Text Simplification with LLMs</i> Daisuke Maekawa, Tomoyuki Kajiwara and Takashi Ninomiya	1110
<i>A Multilingual Human Annotated Corpus of Original and Easy-to-Read Texts to Support Access to Democratic Participatory Processes</i> Stefan Bott, Verena Riegler, Horacio Saggion, Almudena Rascón Alcaina and Nouran Khallaf	1117
<i>Proffiliadur: Welsh Language Text Profiling Toolkit</i> Nicolás Gutiérrez-Rolón, Jonathan Davies, Tomos Williams, Dawn Knight and Fernando Alva-Manchego	1129
<i>Recovering Registers from Leveled Wordlists</i> Yo Ehara	1143
<i>Fill-in-the-Blanks: Automatic Generation and Evaluation of Language Models' Pseudonyms for English and Swedish Texts</i> Maria Irena Szawerna and Jacob Lee Suchardt	1155
<i>Integrating Services, Platforms and Resources into a National Infrastructure Cluster for FAIR Language and Cultural Data</i> Giulia Pedonese, Daniele Melaccio, Michele Mallia, Monica Monachini, Francesca Frontini, Valeria Quochi, Fahad Khan, Angelo Mario Del Grosso, Federico Boschetti and Riccardo Del Gratta	1170
<i>Common European Language Data Space: Development, Current Status, and Future Perspectives</i> Stelios Piperidis, Penny Labropoulou, Dimitrios Galanis, Khalid Choukri, Andrejs Vasiljevs, Mitos Deligiannis, Katerina Gkirtzou, Dimitris Gkoumas, Athanasia Kolovou, Leon Voukoutis, Kanella Pouli, Maria Giagkou, Maria Gavrilidou, Katrin Marheinecke, Elena Leitner, Simon Ostermann, Stefania Raccioppa, Kossay Talmoudi, Victoria Arranz, Valérie Mapelli, Helene Mazo, Fernanda González Campo, Shi Yu, Aivars Bērziņš, Andis Lagzdīns and Georg Rehm ...	1179
<i>Euskorpora: A Strategic Framework for Digital Sovereignty and Linguistic Inclusion of Basque in the Era of AI</i> Victoria Arranz, Sara Arregi, Leire Barañano and Aitor García-Pablos	1189
<i>Automating FAIRness: A FAIRification Tool within the Language Resources Infrastructure</i> Daniele Melaccio and Monica Monachini	1197
<i>FIBER: A Multilingual Evaluation Resource for Factual Inference Bias</i> Evren Ayberk Munis, Deniz Yilmaz, Arianna Muti and Cagri Toraman	1205
<i>EthiQuest: LLM-Powered Ethical Questionnaire Generation for Research Review</i> ishank kapania, Radhika Mamidi and Rahul Mishra	1216
<i>NegNLI-BR: A Brazilian Portuguese Benchmark for Negation in Natural Language Inference</i> Matheus Westhelle and Viviane Moreira	1226

<i>SWE-QA: A Dataset and Benchmark for Complex Code Understanding</i> Laila ELKOUSSY and Julien PEREZ	1236
<i>Augmenting LLM Reasoning with Dynamic Notes Writing for Complex MultiHop QA</i> Rishabh Maheshwary, Masoud Hashemi, Khyati Mahajan, Shiva Krishna Reddy Malay, sai rajeswar mudumba, Sathwik Tejaswi Madhusudhan, Spandana Gella and Vikas Yadav ..	1252
<i>Information Asymmetry across Language Varieties: A Case Study on Cantonese-Mandarin and Bavarian-German QA</i> Renhao Pei, Siyao Peng, Verena Blaschke, Robert Litschko and Barbara Plank	1280
<i>FRASE: Frame-based Structured Representations for Generalizable SPARQL Query Generation</i> Papa Abdou Karim Karou Diallo and Amal Zouaq	1303
<i>Representing Multimodality in Terminology Resources</i> Federica Vezzani	1320
<i>EPOP: A Benchmark Corpus for Assessing NLP Models on Structured Information Extraction in Plant Health</i> Claire Nedellec, Marine Courtin, Xinzhi Yao, Marie Grosdidier, Isabelle Pieretti, Sandy Duperier and Robert Bossy	1331
<i>ReTaT: A Unified Benchmark for Relation Extraction across Text and Table</i> Mohamed Ettaleb, Thibault Ehrhart, Nathalie Aussenac-Gilles, Yoan Chabot, Mouna Kamel, Véronique MORICEAU, Raphael Troncy and Fanfu Wei	1341
<i>LitTx: A New Treatment Relation Extraction Dataset</i> Yuhang Jiang, Md Sultan Al Nahian, Li Hao Richie Xu, Rani Chikkanna and Ramakanth Kavuluru	1352
<i>LegitimNarrate: A Dataset for Analyzing Legitimation Mechanisms in Crowdfunding Narratives</i> Asmaa Lagrid, Sebastien Fournier, Benedicte ALDEBERT, Ali Ghods, Daisy Bertrand and Gael Leboeuf	1361
<i>A Fine-tuned ASR Model for Historical American Dialect Recordings</i> Steven Coats	1372
<i>A Comprehensive Full-Form Lexicon for Arabic NLP and Speech Technology</i> Yannis Haralambous and Jack Halpern	1382
<i>MzansiText and MzansiLM: An Open Corpus and Decoder-Only Language Model for South African Languages</i> Anri M. Lombard, Simbarashe Mawere, Temi Aina, Ethan Wolff, Sbonelo Gumede, Elan Novick, Francois Meyer and Jan Buys	1394

HPLT 3.0: Very Large-Scale Multilingual Resources for LLMs and MT. Mono- and Bi-lingual Data, Multilingual Evaluation, and Pre-Trained Models

Stephan Oepen, Nikolay Arefyev, Mikko Aulamo, Marta Bañón, Maja Buljan, Laurie V. Burchell, Lucas Georges Gabriel Charpentier, Pinzhen Chen, Mariia Fedorova, Ona de Gibert, Barry Haddow, Jan Hajič, Jindrich Helcl, Andrey Kutuzov, Veronika Laippala, Zihao Li, Bhavitvya Malik, Vladislav Mikhailov, Amanda Myntti, Dayyán O'Brien, Lucie Polakova, Gema Ramírez-Sánchez, Janine Siewert, Pavel Stepachev, Joerg Tiedemann, Teemu Vahtola, Dusan Varis, Fedor Vitiugin
and Jaume Zaragoza 1409

Generation of Instruction and Preference Dataset for Improving Japanese Instruction Following in LLMs

Kei Moriyama, Takashi Kodama and Kouta Nakayama 1435

Adapting Pretrained Models to Endangered Languages in Japan: A Comparative Study on Ryukyuan and Ainu Speech Recognition

Kohei Matsuura, Takanori Ashihara and Tatsuya Kawahara 1455

Prerequisites for Advancing Automatic Speech Recognition in Breton

Morgan Grobol, Alice Millour, Wassim Zemouri, Yuna Drapier and Mélanie Jouitteau 1464

Integrating TEI, NER/NEL, Textometry, and Linked Data for a Semantically Enriched Interview Corpus

Ranka Stankovic, Tamara Vučenović, Biljana Rujević, Milica Ikonić Nešić
and Mihailo Škorić 1474

Uhura: A Benchmark for Evaluating Scientific Question Answering and Truthfulness in Low-Resource African Languages

Edward Thomas Bayes, Israel Abebe Azime, Jesujoba Alabi, Jonas Kgomo, Tyna Eloundou, Elizabeth Proehl, Kai Chen, Imaan Khadir, Naome A. Etori, Shamsuddeen Hassan Muhammad, Choice Mpanza, Igneciah Pocia IP Thete, Dietrich Klakow
and David Ifeoluwa Adelani..... 1485

Dialectal Filtering: Synthesizing Kurdish Corpora for Low-Resource Varieties by Utilizing "Noise" in Large Textual Data

Christian Schuler, Raman Ahmad, Ānrán Wáng, Daniil Gurgurov, Timo Baumann, Simon Ostermann and Josef van Genabith 1505

HybridCodeAuthorship: A Benchmark Dataset for Line-Level Code Authorship Detection

Luke S. Patterson, Li Wang and Adam Faulkner 1520

CorEGe-PT: Compiling a Large Corpus of Academic Texts in Portuguese

Tanara Zingano Kuhn, José Matos, Bruno Neves, Daniela Pereira, Elisabete Cação, Ivo Simões, Jacinto Estima, Delfim Leão and Hugo Goncalo Oliveira 1533

SLURP-TN : Resource for Tunisian Dialect Spoken Language Understanding

Haroun Elleuch, Salima Mdhaftar, Yannick Estève and Fethi Bougares 1544

From Semi-Digital Edition to Historical NLP Resource:Constructing and Annotating Historical Multilingual Parallel Text Collections on the TEITOK Platform

Maarten Janssen, Anna Jouravel and Piroska Lendvai 1553

Toward Conversational Hungarian Speech Recognition: Introducing the BEA-Large and BEA-Dialogue Datasets

Máté Gedeon, Piroska Zsófia Barta, Peter Mihajlik, Tekla Etelka Graczi, Anna Kohári and Katalin Mády 1562

Developing the German Medical Text Corpus (GeMTeX): Legal Compliance and Semantic Enrichment

Justin Hofenbitzer, Christina Lohr, Andrea Riedel, Rebekka Kiser, Aliaksandra Shutsko, Abanoub Abdelmalak, Peter Klügl, Jutta Romberg, Sarah Riepenhausen, Miriam Schechner, Jakob Faller, Frank Meineke, Luise Modersohn, Markus Löffler, Juliane Fluck, Udo Hahn, Stefan Schulz and Martin Boeker 1571

MaiChat: A Text-based Dialogue Corpus Rich in Conversational Features

Mai Hoang Dao, Catherine Lai and Peter Bell 1585

Saudi ASWAT: A Large-Scale Corpus of Spontaneous Saudi Arabic Speech

Abdullah I. Alharbi, Afrah A. Altamimi, Muneera Alhoshan, Amal Almazrua, Halah Munif Alharbi, Bayan M. Almuqhim, Hawra Aljasim, Abdulrahman Alosaimy, Yahya A. Asiri and Abdullah Alfaifi 1595

SciCiteVal: A Multi-Domain Dataset for Scientific Citation Verification

Qinyue Liu, Yongxin Zhou and Cyril Labbe 1603

RuznamceNER: A Named Entity Recognition Dataset for Ottoman Turkish

Esma Fatima Bilgin Tasdemir, Dilara Zeynep Güreler and Saziye Betül Ozates 1612

Scripting History: A Diachronic Urdu Text and Image Corpus from the 18Th to 19Th Centuries

Sana Shams, Sahar Rauf, Asad Mustafa, Muhammad Zeeshan Javed, Qurat-ul-Ain Akram, Sarmad Hussain and Miriam Butt 1622

IREKIER: An Easy Read Corpus for Basque and Spanish

Jesús Calleja and Thierry Etchegoyhen 1633

MekongPhon: A Large-Scale Parallel IPA Corpus for Lao and Khmer

Ammon Shurtz, Christian Richardson and Stephen D. Richardson 1650

CorSpell: Introducing a Semiautomatic Tool for Spelling Normalization in Brazilian Portuguese

Juliana Schoffen, Dennis Giovani Balreira, Elisa Marchioro Stumpf, Larissa Goulart, Tanara Zingano Kuhn, Rafael Oleques Nunes, Gabriel Ricci Pazzinato, Isadora Dahmer Hanauer, José Henrique de Souza Silva, Luiza Sarmento Divino and Marine Matte 1659

Meta4XNLI-ptBR: Brazilian Portuguese Extension of Meta4XNLI Corpus

Karina Johansson, Fernanda Assi, Isabella da Silva, Rafael Passador, Isabela Rodrigues, Aline Paes and Helena Caseli 1668

More than "Oh": Grounding Observable Events with Grunts in Multimodal Dialogue

Richard A. Brutti and James Pustejovsky 1677

COME-ALPs: Coreference Annotation with MErging HEuristics Using ALignment-based Projection in Parallel Corpora

gabriela nicole gonzalez saez, Mariam Nakhle, Illia Kholosha, Rachel Atherly and Marco Dinarelli 1688

<i>MEUR: A Benchmark for Evaluating Vision-Language Models on Multimodal Event Understanding and Reasoning</i>	
Zimu Wang, Yuqi Wang, Tong Chen, Changyu Zeng, Hongbin Na, Nijia Han, Fuyu Xing, Qi Chen, Qiufeng Wang, Anh Nguyen, Shuihua Wang, Ling Chen, Jionglong Su, Haiyang Zhang and Wei Wang	1696
<i>Building Collaborative Speech Corpora for Low-Resource Languages: The Galician Dataset in Mozilla Common Voice</i>	
Adina Ioana Vladu, Elisa Fernández Rei and María Pérez Lago	1710
<i>Frame-Guided Synthetic Claim Generation for Automatic Fact-Checking Using High-Volume Tabular Data</i>	
Jacob Devasier, Akshith Putta, Qing Wang, Alankrit Moses and Chengkai Li	1721
<i>A Bilingual Bimodal Benchmark for Arabic-English NLP across Grammatical Correction, Essay Scoring, Morphological Tagging, and Speech Recognition</i>	
Bashar Alhafni, Injy Hamed, Fadhil Eryani, David Palfreyman and Nizar Habash	1732
<i>Developing a Guideline for the Labovian-Structural Analysis of Oral Narratives in Japanese</i>	
Amane Watahiki, Tomoki Doi, Akari Kikuchi, Hiroshi Ohata, Yuki I. Nakata, Takuya Niikawa, Taiga Shinozaki and Hitomi Yanaka	1750
<i>German General Social Survey Personas: A Survey-Derived Persona Prompt Collection for Population-Aligned LLM Studies</i>	
Jens Rupperecht, Leon Froehling, Claudia Wagner and Markus Strohmaier	1761
<i>Slovene Morphological and Word Formation Segmentation: A Novel Dataset and Evaluation</i>	
Marko Pranjić, Boris Kern, Ines Voršič and Senja Pollak	1781
<i>GePaDeU - a Multi-layer Corpus of German Parliamentary Debates with Rich Semantic and Pragmatic Annotations</i>	
Ines Rehbein, Julian Schlenker, Lars Ostertag and Simone Paolo Ponzetto	1794
<i>What Are LLMs Doing to Scientific Communication? Measuring Changes in Writing Practices and Reading Experience</i>	
Filip Miletić and Neele Falk	1811
<i>GeneFRDebate: Generated French Debates from News Articles with Industrial-Expert Summaries</i>	
Rim Abrougui, Guillaume Lechien, Elisabeth Savatier and Benoît Laurent	1831
<i>AmbiCoRefVis: A Tool for Visualizing Coreferential Ambiguity</i>	
Patrick Paetzold, Lukas Beiske, Mark-Matthias Zymla, Massimo Poesio, Miriam Butt, Daniel Weiskopf and Oliver Deussen	1842
<i>Fables-DTR: A Corpus of Fables Annotated for Discourse and Temporal Relations</i>	
Purificação Silvano, António Leal, Maciej Ogrodniczuk, Aleksandra Tomaszewska, Joana Gomes, Luís Filipe Cunha, Evelin Amorim, Martyna Lewandowska, Anna Śliwicka and Alípio Jorge	1856
<i>A Benchmark Corpus for the Diagnostic Assessment of Content in L2 English Speech</i>	
Kosuke Doi, Justin Vasselli and Taro Watanabe	1869

<i>Insights from Romanized Manipuri Social Media Text: A Transliteration Corpus and Variation Analysis</i>	
Maisang Kamei Salice, Sanasam Ranbir Singh and Priyankoo Sarmah	1878
<i>MELD: Melding Diverse Multilingual and Multi-Domain Datasets for Named Entity Recognition Evaluation</i>	
Kevin Glocker and Marco Kuhlmann	1889
<i>FinER-ABSA: A Benchmark for Implicit and Explicit Entity Recognition and Aspect-Based Sentiment Analysis in Financial News</i>	
Pachara Akkanwanich, Pavorn Thongyoo, Mahannop Thabua, Konlakorn Wongpatikaseree and Natthawut Kertkeidkachorn	1904
<i>MUSIA: Multilingual Story Illustration Corpus for Cross-Cultural Alignment and Generation</i>	
Krishna Tewari, Supriya Chanda, Nirmal Patil and Sukomal Pal	1914
<i>MUDiC: A Dataset for Multi-User Dialogue and Collaboration in Chatbot Interaction</i>	
Nicolas Wagner, Cristina Luna Jimenez, Elisabeth Andre, Wolfgang Minker and Stefan Ultes	1925
<i>StoryCCDial: Collecting and Analyzing Human-Human Co-Creation Dialogues for Personalized Creative Support</i>	
Natsumi Ezure and Michimasa Inaba	1934
<i>DATASHI: A Parallel English–Tashlhiyt Corpus for Orthography Normalization and Low-Resource Language Processing.</i>	
Nasser-Eddine Monir and Zakaria Baou	1947
<i>Evaluating Social Intelligence in LLMs via Japanese Honorifics in Email Generation: A Social Semiotic System Perspective</i>	
Muxuan Liu, Tatsuya Ishigaki, Yusuke Miyao, Hiroya Takamura and Ichiro Kobayashi	1957
<i>Do Language Models Know Theo Has a Wife? Investigating the Proviso Problem</i>	
Tara Azin, Daniel Dumitrescu, Diana Inkpen and Raj Singh	1977
<i>Cross-Lingual and Cross-Cultural Transfer of Talk Move Classification to German Science Classrooms</i>	
Christian Wartena, Christian Schumburg, Andreas Nehring, Marcel Ebert, Friederike Korneck, David Schmitt, Marie Irmer and Birgit Neuhaus	1989
<i>IHPP: A Paragraph-Level Dataset for Investigating the Pragmatics of Hyperpartisan Italian News</i>	
Michele Joshua Maggini, Davide Bassi, Angelo Valente, Gaël Dias and Pablo Gamallo	1998
<i>Detecting Potentially Under-annotated Explicit Discourse Connectives in the Penn Discourse Treebank (PDTB-3) with LLMs</i>	
Yueh-Ting Chuang, Xixian Liao and Bonnie Webber	2012
<i>Can LLMs Understand Punchlines? LLMs’ Narrative Understanding Evaluation with Short-stories</i>	
Jiashi Cheng and Takehito Utsuro	2024

<i>Building the AURIS Corpus of Reference and Information Structure</i> Christian Chiarcos, Christian Fäth, Tabea Gröger and Quentin Alastair Frey	2035
<i>There Is No Spoon: Existential Presupposition in Large Language Models</i> Marie-Léontine Wörgötter, Shikai Lai and Sebastian Schuster	2048
<i>DiscoRAG: A Discourse-Aware Agent for Query-Based Summarization of Long Documents</i> Alexander Chernyavskiy, Lidiia Ostyakova and Dmitry Ilvovsky	2062
<i>In-Distribution Steering: Balancing Control and Coherence in Language Model Generation</i> Arthur Vogels, Benjamin Wong, Yann Choho, Annabelle Blangero and Milan Bhan ..	2076
<i>Improving Multilingual Language Models by Aligning Representations through Steering</i> Omar Mohamed Mahmoud, Buddhika Laknath Semage, Thommen George Karimpanal and Santu Rana	2090
<i>Explainable AI for Ethical Counter Speech Generation in Hate Speech Mitigation</i> Ashiful Islam Ridoy, Mohammed Faisal, Yogesh Kumar, Md Mamun-Ur Rashid, Marina Ernst and Frank Hopfgartner	2104
<i>Do Language Models Encode Semantic Relations? Probing and Sparse Feature Analysis</i> Andor Diera and Ansgar Scherp	2115
<i>The Sufficiency-Conciseness Trade-off in LLM Self-Explanation from an Information Bottleneck Perspective</i> Ali Zahedzadeh and Behnam Bahrak	2127
<i>Node-Level Uncertainty Estimation in LLM-Generated SQL</i> Hilaf Hasson and Ruocheng Guo	2145
<i>A Typologically Grounded Evaluation Framework for Word Order and Morphology Sensitivity in Multilingual Masked LMs</i> Anna Feldman, Libby Barak and JIng Peng	2154
<i>From Generation to Evaluation: A Resource for Error-Categorized Question Generation from Video Transcripts</i> Joshua Berger, Markos Stamatakis, Anett Hoppe, Ralph Ewerth and Christian Wartena	2166
<i>From Behavior to Geometry: A Causal and Geometric Analysis of LoRA-Based Domain Adaptation</i> Yizhe WANG, Liu He and Zhenhua Ling	2178
<i>Explainable Semantic Textual Similarity via Dissimilar Span Detection</i> Diego Miguel Lozano, Daryna Dementieva and Alexander Fraser	2190
<i>BIS Reasoning 1.0: The First Large-Scale Japanese Benchmark for Belief-Inconsistent Syllogistic Reasoning</i> Ha Thanh Nguyen, Hideyuki Tachibana, Chaoran Liu, Qianying Liu, Su Myat Noe, Koichi Takeda and Sadao Kurohashi	2211
<i>A Discourse-based Tool Series for Logical Validation of LLMs</i> Boris Galitsky and Dmitry Ilvovsky	2220

<i>COCOA: Creation and Exploratory Investigation of a COrpus of Claims frOm NLP Articles</i> Clémentine Bleuze, Fanny Ducel, Maxime Amblard and Karen Fort	2387
<i>SPOT: An Annotated French Corpus and Benchmark for Detecting Critical Interventions in Online Conversations</i> Manon Berriche, Célia Nouri, Chloé Clavel and Jean-Philippe Cointet	2400
<i>MedPT: A Massive Medical Question Answering Dataset for Brazilian-Portuguese Speakers</i> Fernanda Bufon Farber, Iago Alves Brito, Julia Soares Dollis, Pedro Schindler Freire Brasil Ribeiro, Rafael Teixeira Sousa and Arlindo R. Galvão Filho	2419
<i>Large Language Models for Citation Function Classification</i> Daniel Vodička, Pavel Kral, Christophe Cerisara and Jakub Šmíd	2430
<i>Small LLMs for Medical NLP: A Systematic Analysis of Few-Shot, Constraint Decoding, Fine-Tuning and Continual Pre-Training in Italian</i> Pietro Ferrazzi, Mattia Franzin, Alberto Lavelli and Bernardo Magnini	2440
<i>Analysing Lightweight Large Language Models for Biomedical Named Entity Recognition on Diverse Output Formats</i> Pierre Epron, Adrien Coulet and Mehwish Alam	2458
<i>WISTERIA: Weak Implicit Signal-based Temporal Relation Extraction with Attention</i> Duy Dao DO, Anaïs Haltermeyer and Thi Bich Hanh DAO	2471
<i>Dynamic Model Switching to Mitigate Outdated Knowledge in Large Language Models</i> Ramakrishna Pinninti, Sabyasachi Kamila, Ayan Mazumder and Mohammed Hasanuzzaman	2490
<i>Multi-Scale Model Compression via Nested Matrix Learning</i> Xiangjue Dong, Aditya Anantharaman, Hemant Pugaliya and Kai Zhong	2501
<i>Confabulations from ACL Publications (CAP): A Dataset for Scientific Hallucination Detection</i> Federica Gamba, Aman Sinha, Timothee Mickus, Raul Vazquez, Patanjali Bhamidipati, Claudio Savelli, Ahana Chattopadhyay, Laura A. Zanella, Yash Kankanampati, Binesh Arakkal Remesh, Aryan Ashok Chandramania, Rohit Agarwal, Chuyuan Li, Ioana Buhnila and Radhika Mamidi	2512
<i>MedInjection-FR: Exploring the Role of Native, Synthetic, and Translated Data in Biomedical Instruction Tuning</i> Ikram Belmadani, Oumaima El Khettari, Pacôme Constant dit Beauvill, Benoît Favre and Richard Dufour	2525
<i>The Impact of Tokenization Algorithms on Hungarian Language Model Performance</i> Mátyás Osváth, Máté Norbert Molnár, Roland Gunics and Noémi Ligeti-Nagy	2545
<i>FAME: Fictional Actors for Multilingual Erasure</i> Claudio Savelli, Moreno La Quatra, Alkis Koudounas and Flavio Giobergia	2557
<i>Detecting Risky Behavior Related to Alcohol and Drug Use within Adolescents' Private Messenger Conversations</i> Jaromír Plhák, Michaela Lebedíková, Ondřej Sotolar and David Smahel	2567

<i>Voices and Echoes in Fictional Dialogue: A Study of Linguistic Coordination in Literary Texts</i> Ioana-Roxana Boriceanu, Alina Iacob and Liviu P. Dinu	2581
<i>Bridging the Domain Divide: Supervised vs. Zero-Shot Clinical Section Segmentation from MIMIC-III to Obstetrics</i> Baris Karacan, Barbara Di Eugenio and Patrick Thornton	2594
<i>Reading Dynamics and Comprehension in Cognitive Aging: A Multimodal Language Resource</i> Claudia Marzi, Noemi Boni, Alice Todesco, Andrea Nadalini, Giorgia Albertin, Cristina Dolciotti, Paolo Bongioanni, Marcello Ferro, Fabio Tamburini, Gloria Gagliardi and Vito Pirrelli	2608
<i>Evaluating Style Embeddings for Machine-Generated Text Detection</i> Noé Durandard, Saurabh Dhawan and Thierry Poibeau	2619
<i>The Speech-LLM Takes It All: A Truly Fully End-to-End Spoken Dialog State Tracking Approach</i> Nizar El Ghazal, Antoine Caubrière and Valentin Vielzeuf	2629
<i>Off the Hamster Wheel: Rethinking Dialogue Research through a Meta-Analysis of the ACL Anthology 2024</i> Amandine Decker, Maxime Amblard and Ellen Breitholtz	2638
<i>VDAct 2.0: Scaling Video-Grounded Dialogue for Event-driven Activity Understanding with LLM-Assisted Filtering</i> Wiradee Imrattanastrai, Masaki Asada, Kimihiro Hasegawa, Ken Fukuda and Teruko Mitamura	2653
<i>Multi-dimensional Evaluation of Character-Authentic Dialogue Models Learned from Question-Answer Data</i> Atsushi Otsuka, Kazuya Matsuo, Kenta Hama, Masahiro Mizukami, Tsunehiro Arimoto, Hiroaki Sugiyama, Makoto Nakatsuji and Narichika Nomoto	2667
<i>Empathy in Greek Exam-Related Support Conversations: A Comparative Evaluation of LLM Responses</i> Panagiota Kyriazi and Prokopis Prokopidis	2682
<i>Evaluation of Two Leading Polish Language Models in a Real-world RAG Scenario</i> Szymon Bartanowicz and Krzysztof Jassem	2698
<i>A Mental State Extraction Dataset for Theory-of-Mind-based Reasoning in Emotional Support Conversations</i> Seulgi Kim and Harksoo Kim	2705
<i>Construction and Analysis of Japanese Parent-Child Dialogic Reading Corpus for Conversational Agents</i> Yuko Nakagi, Yuya Chiba, Sanae Fujita and Shoko Araki	2724
<i>ACLBot: A Knowledge Graph-Driven Assistant for ACL Anthology Research</i> Jan Buchmann, Steven Lynden and Kristiina Jokinen	2731

<i>This House Debates AI: Evaluating a Language Model in Oxford-Style Debates against Human Experts</i>	
Umberto Belluzzo, Kobi Hackenburg, Hannah Rose Kirk, Scott Hale and Paul Röttger	2742
<i>PAIR: A Pilot Dataset for Dual Perspective-based Video-Grounded Dialogue and Reconciliation</i>	
Lewis N. Watson, Carl Strathearn, Kenny Mitchell and Yanchao Yu	2760
<i>I Am Not Them: Persistent Outgroup Bias in Large Language Models Arising from Social Identity Persona Setting</i>	
Wenchao Dong, Assem Zhunis, Dongyoung Jeong, Hyojin Chin, Jiyoung Han and Meeyoung Cha	2772
<i>CONVERSE: Annotation Scheme and Dataset for Multimodal Conversational Engagement Analysis in Human-Human and Human-Robot Interaction</i>	
Ekaterina Torubarova, Oskar Ljung, Julia Uddén and André Pereira	2787
<i>FineDialFact: A Benchmark for Fine-Grained Dialogue Fact Verification</i>	
Xiangyan Chen, Yufeng Li, Yujian Gan, Arkaitz Zubiaga and Matthew Purver	2798
<i>Meta-Prompting Follow-Ups for Unsupervised Dialogue Evaluation Using Open-Source Large Language Models</i>	
Gaetano Cimino, Chuyuan Li, Giuseppe Carenini and Vincenzo Deufemia	2812
<i>HumaniCA: A Benchmark Resource for the Detection of Users' Ascription of Humanness to Conversational Agents</i>	
Sabrina Villata, Amon Rapp, Luigi Di Caro and Federica Cena	2825
<i>Towards Reliable Evaluation of Emotional Text Generation in LLMs: Human vs. Automatic Metrics</i>	
sadegh jafari, Els Lefever and Veronique Hoste	2836
<i>Question and Response Dynamics in Public Service Encounters</i>	
Wassiliki Siskou, Ingrid Espinoza, Laurin Friedrich, Steffen Eckhard and Annette Hautli-Janisz	2848
<i>Reasoning over Object Descriptions Improves Coreference Resolution in Task-Based Dialogue Systems</i>	
Oier Ijurco and Oier Lopez de Lacalle	2856
<i>Evaluating the Effect of Question Wording Variations on Answer Consistency in Large Language Models</i>	
Junya Takayama, Masaya Ohagi, Tomoya Mizumoto and Katsumasa Yoshikawa	2874
<i>Knowledge-Infused Hierarchy-Aware Emotion Recognition in Code-mixed Mental Health Counseling Conversations</i>	
Aseem Srivastava, Kushagra Mittal, Anusha Tiwari and Md. Shad Akhtar	2887
<i>A Corpus for Personalized Dialogue Breakdown Repair in Japanese Open-Domain Conversations</i>	
Kazuya Tsubokura, Yurie Iribe and Norihide Kitaoka	2899

<i>Conversational Assistants to Support Patients with Heart Failure: Comparing a Neurosymbolic Architecture with GPT</i>	
Anuja Tayal, Devika Salunke, Barbara Di Eugenio, Paula G. Allen-Meares, Eulalia P. Abril, Olga Garcia-Bedoya, Carolyn A. Dickens and Andrew D. Boyd.....	2913
<i>Disentangling Approaches to Conversation Disentanglement: Fine-Tune or Learn from Scratch?</i>	
Debaditya Pal, Anton Leuski, Ron Artstein, David Traum and Kallirroi Georgila	2927
<i>Evaluation of Failure Communication Strategies for Trust Repair in Human-AI Collaboration</i>	
Stina Klein, Alexandru Wurm, Elisabeth Andre and Matthias Kraus	2942
<i>Multi-Session Client-Centered Treatment Outcome Evaluation in Psychotherapy</i>	
Hongbin Na, Tao Shen, Shumao Yu and Ling Chen	2952
<i>Towards Reward Modeling for AI Tutors in Math Mistake Remediation</i>	
Kseniia Petukhova and Ekaterina Kochmar	2969
<i>HOTATE: A Japanese Dialogue Corpus Annotated with Responses of Private Thoughts and Public Statements</i>	
Yuko Toda, Daisuke Maekawa, Kota Manabe, Eito Yoneyama, Kanade Nonomura, Yuki Fujiwara and Tomoyuki Kajiwara.....	2987
<i>Mining Naturally Romanized Seed Corpora without Romanizations</i>	
Adrian Benton, Alexander Gutkin, Christo Kirov and Brian Roark	2996
<i>From Press to Pixels: Evolving Urdu Text Recognition</i>	
Samee Arif and Sualeha Farid	3013
<i>HalleluBERT: Let Every Token That Has Meaning Bear Its Weight</i>	
Raphael Schmitt.....	3022
<i>Sentiment Analysis and Language Models for Kwanyama</i>	
Ndapa Nakashole	3031
<i>TigerCoder: A Novel Suite of LLMs for Code Generation in Bangla</i>	
Nishat Raihan, Antonios Anastasopoulos and Marcos Zampieri	3044
<i>ViX-Ray: A Vietnamese Chest X-Ray Dataset for Vision-Language Models</i>	
Duy Vu Minh Nguyen, Chinh Thanh Truong, Trần Hoàng Phúc, Hung Tuan Le, Nguyen Van-Thanh Dat, Trung Hieu Pham and Kiet Van Nguyen	3055
<i>Creating Task-Specific Speech Recognition Datasets from Scratch for Low-Resource Languages: Assessing the Impact of Token Sequence Overlap</i>	
Adwoa Asantewaa Bremang, Dennis Asamoah Owusu, Victor Kow Quagrainne and Leanne M.M. Annor-Adjaye	3075
<i>Radio Haiti-Inter: A Large-Scale Annotated Corpus of Spoken Haitian Creole</i>	
William N. Havard, Rayan Ziane, Mélissa Menclé, Maximin Coavoux, Benjamin Lecouteux and Emmanuel Schang	3083
<i>Synthetic Function Demonstrations Improve Generation in Low-Resource Programming Languages</i>	
Nick McKenna, Xinnuo Xu, Jack Williams, Nicholas C. Wilson, Benjamin Van Durme and Christian Poelitz	3094

<i>PerHalluEval: Persian Hallucination Evaluation Benchmark for Large Language Models</i> Mohammad Hosseini, Kimia Hosseini, Shayan Bali, Zahra Zanjani and Saeedeh Momtazi	3107
<i>ADAB: Arabic Dataset for Automated Politeness Benchmarking - a Large-Scale Resource for Computational Sociopragmatics</i> Hend Al-Khalifa, Nadia Ghezaiel, Maria Bounnit, Hend Hamed Alhazmi, Noof Abdullah Alfear, Reem Fahad Alqifari, Ameerah Masoud Almasoud and Sharefah Ahmed Al-Ghamdi	3128
<i>GRDD+: An Extended Greek Dialectal Dataset with Cross-Architecture Fine-tuning Evaluation</i> Stergios Chatzikyriakidis, Dimitrios Papadakis, Sevasti Ioanna Papaioannou and Erofilis Psaltaki	3138
<i>Same-Language Subtitles for Low-resource Languages: A Case of Bundelkhandi</i> Anirudh Pradhan, Ayushi Pandey, Divyansh Kushwaha, Akshita Tiwary and Vivek Seshadri	3147
<i>The Chulalongkorn Corpus of Spoken Thai (CCOST)</i> Pittayawat Pittayaporn, Cathryn Yang, Sujinat Jitwiriyant and James Kirby	3154
<i>Nepal Script Text Recognition from Ancient Artifacts: Challenges and Opportunities</i> Swornim Nakarmi, Sarin Sthapit, Sahil Ratna Tuladhar, Arya Shakya, Bal Krishna Bal and Rajani Chulyadyo	3161
<i>LuxBorrow: From Pompier to Pompjee, Tracing Borrowing in Luxembourgish</i> Nina Hosseini-Kivanani and Fred Philippy	3171
<i>Ramsa: A Large Sociolinguistically Rich Emirati Arabic Speech Corpus for ASR and TTS</i> Rania Al-Sabbagh	3184
<i>DialectalArabicMMLU: Benchmarking Dialectal Capabilities in Arabic and Multilingual Language Models</i> Malik H. Altakrori, Nizar Habash, Teresa Lynn, Younes Samih, Abed Alhakim Freihat, Kirill Chirkunov, Muhammed AbuOdeh, Radu Florian, Preslav Nakov and Alham Fikri Aji	3199
<i>ForumOccitania: A Corpus of User-Generated Content for Multiple Occitan Varieties</i> Oriane Nédey, Juliette Janès, Rachel Bawden, Thibault Clérice and Benoît Sagot ...	3220
<i>A Dataset of Wolof Ajami Manuscripts for HTR and OCR</i> Oreen Yousuf, Elhadji Djibril Diagne, Christian Høgel, Beata Megyesi and Joakim Nivre	3234
<i>TDMulti: A Tunisian Dialect-Modern Standard Arabic Multitask Corpus with a Context-Aware Cross-Attention BERT Model</i> Roua Torjmen and Kais HADDAR	3240
<i>The Megrelian Language Corpus (MLC): Creation, Annotation, and Initial Steps toward a UD Treebank</i> Irina Lobzhanidze, Rusudan Gersamia and Tamar Gogia	3250

<i>Steering LLMs toward Korean Local Speech: Iterative Refinement Framework for Faithful Dialect Translation</i>	
keunhyeung park, Seunguk Yu and Youngbin Kim	3257
<i>LombardoGraphia: Automatic Classification of Lombard Orthography Variants</i>	
Edoardo Signoroni and Pavel Rychly	3270
<i>Meenz bleibt Meenz, but Large Language Models Do Not Speak Its Dialect</i>	
Minh Duc Bui, Manuel Mager, Peter Herbert Kann and Katharina von der Wense	3281
<i>Bootstrapping NLP for Sakha: Named Entity Recognition and Sentiment Analysis in an Extremely Low-Resource Setting</i>	
Mariia Everstova, Nikolai Efimov and Valerio Basile	3295
<i>Lightweight Cross-Lingual Federated Prompt Tuning for Low-Resource Languages</i>	
Ubaid Azam, Imran Razzak and Shoaib Jameel	3304
<i>A Parallel Corpus of the Parable of the Prodigal Son: Building a Resource for Documenting Language Varieties in Mainland France</i>	
Lucence Ing, Juliette Janès, Sven Ködel and Benoît Sagot	3317
<i>Developing Zila: A Spoken Language Resource for the Endangered Slovenian Gail Valley Dialect</i>	
Andrej Zgank, Gregor Donaj, Urh Kolaric, Usi Sereinig, Tatjana Koren-Zwitter, Sanja Boto, Sabina Zwitter-Grilc, Jasna Vidinic and Darinka Verdonik	3325
<i>Nawatl Context-Free Grammars for Natural Language Processing</i>	
Juan Jose Guzman Landa, Juan-Manuel Torres-Moreno, Graham Ranger, Miguel Figueroa-Saavedra, Ligia Quintana Torres, Carlos-Emiliano Gonzalez-Gallardo, Luis Gil Moreno Jimenez and Martha Lorena Avendaño Garrido	3333
<i>Physical Commonsense Reasoning for Lower-Resourced Languages and Dialects: A Study on Basque</i>	
Jaione Bengoetxea, Itziar Gonzalez-Dios and Rodrigo Agerri	3343
<i>Common Voice for Pakistan: Developing an Open Speech Corpus for Low-Resource Pakistani Languages</i>	
Meesum Alam and Francis Tyers	3355
<i>Amulwe Kimün: A Community-Grounded Demo, Resource, and ASR Baseline for Mapuzugun</i>	
Cristian Eduardo Ahumada Oliva and Fatiha Sadat	3360
<i>Development of Serbian QA Datasets through Prompt-Based Generation and Human Validation</i>	
Jovana Rađenović, Olivera Kitanović, Ranka Stankovic and Mihailo Škorić	3369
<i>An Enhanced Pipeline for the Manzini-Savoia Dialect Corpus</i>	
Achille Fusco, Greta Mazzaggio and Carlo Zoli	3379
<i>Are Language Models Borrowing-Blind? A Multilingual Evaluation of Loanword Identification across 10 Languages</i>	
Merilin Sousa Silva and Sina Ahmadi	3394
<i>Comparing Approaches to Automatic Summarization in Less-Resourced Languages</i>	
Chester Palen-Michel and Constantine Lignos	3402

<i>PsihoRo: Depression and Anxiety Romanian Text Corpus</i> Alexandra Ciobotaru, Ana-Maria Bucur and Liviu P. Dinu	3423
<i>Aligned Parallel Corpus of the Vedic Saṁhitās for Machine Translation</i> Yuzuki Tsukagoshi and Ikki Ohmukai	3434
<i>FormosanMT: A Multilingual Parallel Corpus of the Formosan Language Family</i> Hunter Scheppat, Joshua K. Hartshorne, Sema Koc, Éric Le Ferrand and Emily Prud'hommeaux	3445
<i>The Construction of a Mixe Variant Parallel Corpus</i> Ivan Vladimir Meza Ruiz, Delfino Zacarias Marquez, Martha Elba Ramírez Andrés, Victoriano Santiago Cayetano, Jonathan Santiago Antonio and Carlos Daniel Hernández Mena	3456
<i>Nepali Lemmatization with Multilingual Transformers: Intrinsic and Extrinsic Evaluation in a Low-Resource Setting</i> Sunil Regmi, Sundeep Dawadi and Bal Krishna Bal	3462
<i>Diacritic Restoration for Low-Resource Indigenous Languages: Case Study with Bribri and Cook Islands Māori</i> Rolando Coto-Solano, Daisy Li, Manoela Teleginski Ferraz, Olivia Sasse, Cha Krupka, Sharid Loaiciga and Sally Akevai Tenamu Nicholas ..	3470
<i>A Modern Online Learning Platform for 'Ōlelo Hawai'i Classrooms</i> Christian Castro, Keneth Martin, Winston Wu and William H. Wilson	3484
<i>Glossed Data in Northern Interior Salish</i> Anna Stacey	3490
<i>CEFR-Cymraeg: A Dataset and Baseline Models for Language Proficiency Assessment in Welsh</i> Eeshan Waqar, Jonathan Davies, Dawn Knight and Fernando Alva-Manchego	3496
<i>Singlish to English Translation with Precision: A Dataset and Language Detection-Driven Masked Modeling for Singlish to English Translation</i> Sujit Kumar, Gerome Kusuma Ang, Stephanie Hilary Xinyi Ma, Andy Hau Yan Ho and Andy Khong	3506
<i>LLMs in Ottoman Turkish: From MLM to NER</i> Enes Yilandiloğlu	3517
<i>SloPal: A 60-Million-Word Slovak Parliamentary Corpus with Aligned Speech and Fine-Tuned ASR Models</i> Erik Božík and Marek Suppa	3523
<i>SlovKE: A Large-Scale Dataset and LLM Evaluation for Slovak Keyphrase Extraction</i> Dávid Števaňák and Marek Suppa	3534
<i>Automatic Speech Recognition for Documenting Endangered Languages: Case Study of Ikema Miyakoan</i> Chihiro Taguchi, Yukinori Takubo and David Chiang	3547

<i>Adaptive Method for Self-Supervised Learning Models on Automatic Dialect Speech Recognition Based on Shared Knowledge of Japanese Dialects and Standard Japanese</i>	
Naoru Asakawa, Naoki Takahashi, Atsuhiko Kai and Seiichi Nakagawa	3556
<i>ATLAS: Article Tracking, Linking, and Analysis of Swedish Encyclopedias</i>	
Albin Andersson, Salam Jonasson, Fredrik Wastring and Pierre Nugues	3566
<i>Evaluating Embedding Models on Danish Historical Newspapers: A Corpus and Benchmark Resource</i>	
Alie Lassche, Pascale Feldkamp, Yuri Bizzoni, Katrine Baunvig, Kristoffer Nielbo and Johan Heinsen	3577
<i>Leveraging Linguistic Similarity for Low-Resource Speech Transcription</i>	
Valentina Fedchenko and Eric Jordan	3590
<i>A Corpus of Persuasion Techniques in Slavic Languages</i>	
Jakub Piskorski, Dimitar Iliyanov Dimitrov, Marina Ernst, Jacek Haneczok, Michal Marcinczuk, Arkadiusz Modzelewski and Roman Yangerber	3599
<i>GePaDeSE: A New Resource for Clause-Level Aspect in German Parliamentary Debates</i>	
Julian Schlenker, Ines Rehbein, Lilly Brauner, Florian Ertz, Ines Reinig and Simone Paolo Ponzetto	3617
<i>FrameNet Semantic Role Classification by Analogy</i>	
Van Duy Ngo, Stergos Afantenos, Emiliano Lorini and Miguel Couceiro	3633
<i>CEFR-Annotated WordNet: LLM-Based Proficiency-Guided Semantic Database for Language Learning</i>	
Masato Kikuchi, Masatsugu Ono, Toshioki Soga, Tetsu Tanabe and Tadachika Ozono	3645
<i>Towards a Gold Standard for Adjectival Hypernymy: Enriching the Open English WordNet with a Hybrid Approach</i>	
Lorenzo Augello, John P. McCrae and Marco Passarotti	3662
<i>REMOVE in LiLa: Integrating Latin Preverbed Motion Verbs with WordNet and VerbNet</i>	
Andrea Farina, Marco Passarotti, Francesco Mambrini, Matteo Pellegrini, Eleonora Litta and Giovanni Moretti	3672
<i>From Incidents to Framing: A Dutch and English Frame-semantic Corpus and Lexicon</i>	
Piek T.J.M. Vossen, Pia Sommerauer and Levi Remijnse	3684
<i>AI Safety Lost in Translation: Evaluating the Effectiveness of English-Italian Cross-Lingual LLM Safety Alignment</i>	
Alessio Wu and Martim Brandao	3697
<i>Semantic Label Drift in Cross-Cultural Translation</i>	
Mohsinul Kabir, Tasnim Ahmed, Md Mezbaur Rahman, Polydoros Giannouris and Sophia Ananiadou	3714
<i>Chain-of-Thought Reasoning Improves Context-Aware Translation with Large Language Models</i>	
Shabnam Ataei, Hugo Huart and Andrei Popescu-Belis	3725

<i>Adja-French Parallel Corpus: A New Resource for Machine Translation of a West African Under-Resourced Language</i>	
Josue Frejus Godeme and Rolando Coto-Solano	3742
<i>Goldfish: Monolingual Language Models for 350 Languages</i>	
Tyler A. Chang, Catherine Arnett, Zhuowen Tu and Benjamin Bergen	3750
<i>Dynaword: From One-shot to Continuously Developed Datasets</i>	
Kenneth Enevoldsen, Kristian Nørgaard Jensen, Jan Kostkan, Balázs Szabó, Márton Kardos, Kirsten Vad, Johan Heinsen, Andrea Blasi Núñez, Gianluca Barmina, Jacob Nielsen, Rasmus Larsen, Rob van der Goot, Peter Vahlstrup, Per Møldrup Dalum, Desmond Elliott, Lukas Galke Poech, Peter Schneider-Kamp and Kristoffer Nielbo	3782
<i>From Bones to Rocks: A Systematic Evaluation of Specialized Definition Generation for Portuguese</i>	
Rafael Oleques Nunes, Dennis Giovani Balreira and Joel Luís Carbonera	3794
<i>Bangla Key2Text: Text Generation from Keywords for a Low Resource Language</i>	
Tonmoy Talukder and G M Shahariar	3805
<i>Beyond Lemmas and Syntax: Comparing Human and LLM-Generated Scientific Abstracts</i>	
Sergei Bagdasarov and Diego Alves	3823
<i>Systematic Multi-Aspect Evaluation of Time Series-Based Report Generation: The Case of Financial Analysis from Stock Data</i>	
Elizabeth Fons, Elena Kochkina, Rachneet Kaur, Zhen Zeng, Berowne Hlavaty, Charese Smiley, Svitlana Vyetenko and Manuela Veloso	3833
<i>Towards Reliable AI Fairness: Challenges in Steering Features within Bias-Implicated Neurons</i>	
Ismael Garrido-Munoz, Arturo Montejo-Raez and Fernando Martínez-Santiago	3851
<i>From Body to Mind: Analyzing Gender Representation in Spanish Generative Language Models</i>	
Ismael Garrido-Munoz, Fernando Martínez-Santiago and Arturo Montejo-Raez	3861
<i>Incivility and Rigidity: Evaluating the Risks of Fine-Tuning LLMs for Political Argumentation</i>	
Svetlana Churina and Kokil Jaidka	3875
<i>EsBBQ and CaBBQ: The Spanish and Catalan Bias Benchmarks for Question Answering</i>	
Valle Ruiz-Fernández, Mario Mina, Júlia Falcão, Luis Antonio Vasquez Reina, Anna Salles, Aitor Gonzalez-Agirre and Olatz Perez-de-Viñaspre	3884
<i>ToxSyn-PT: A Synthetic Fine-Grained Dataset of Minority-Targeted Toxic Language in Portuguese</i>	
Iago Alves Brito, Julia Soares Dollis, Fernanda Bufon Farber, diogo fernandes and Arlindo R. Galvão Filho	3908
<i>AnswerCarefully: Creating a Dataset for LLM Safety in Japanese</i>	
Hisami Suzuki, Satoru Katsumata, Takashi Kodama, Tetsuro Takahashi, Kouta Nakayama and Satoshi Sekine	3921
<i>A Dutch Benchmark to Assess Social Bias in LLMs within a Hiring Decision Setting</i>	
Renate Burema, Anne Schuth, Christopher Spelt and Dong Nguyen	3932

<i>PBBQ: A Persian Bias Benchmark Dataset Curated with Human-AI Collaboration for Large Language Models</i>	
Farhan Farsi, Shayan Bali, Fatemeh Valeh, Parsa Ghofrani, Alireza Pakniat, Seyedkian Kashfipour and Amir H. Payberah	3944
<i>Contextualizing Toxicity: An Annotation Framework for Unveiling Pragmatics in Conversations of Online Discussion Forums</i>	
Yingxue Fu and Anais Ollagnier	3961
<i>How Far Can Bias Go? Tracing Bias from Pre-Training Data to Alignment</i>	
Marion Thaler, Abdullatif Köksal, Alina Leidinger, Anna Anna Korhonen and Hinrich Schütze	3975
<i>Robust Bias Evaluation with FilBBQ: A Filipino Bias Benchmark for Question-Answering Language Models</i>	
Lance Calvin Lim Gamboa, Yue Feng and Mark Lee	3996
<i>Uncovering Hidden Violent Tendencies in LLMs: A Demographic Analysis via Behavioral Vignettes</i>	
Quintin Myers and Yanjun Gao	4009
<i>Exploring Social Bias in Slovenia: The EEC-SL Dataset</i>	
Jaya Caporusso, Damar Hoogland, Boshko Koloski, Matthew Purver, Senja Pollak and Spela Vintar	4019
<i>The MISOMEM-Val Dataset for Identifying Human Values in Misogynistic Memes</i>	
Rakshitha Rao Ailneni and Sanda Harabagiu	4031
<i>ConGA: Guidelines for Contextual Gender Annotation. a Framework for Annotating Gender in Machine Translation</i>	
Argentina Anna Rescigno, Eva Vanmassenhove and Johanna Monti	4048
<i>University Speaking for Everyone: Assessing Changes in Italian Higher Education Statutes toward Gender-Inclusive Language</i>	
Sebastiano Vecellio Salto, Camilla Casula, Alessio Palmero Aprosio and Sara Tonelli	4058
<i>Breaking the Benchmark: Revealing LLM Bias via Minimal Contextual Augmentation</i>	
Kaveh Eskandari Miandoab, Mahammed Kamruzzaman, Arshia Gharooni, Gene Louis Kim, Vasanth Sarathy and Ninareh Mehrabi	4070
<i>TryggLLM: A Benchmark for Evaluating LLM Safety in Norwegian</i>	
Samia Touileb, Truls Pedersen and Isabell Stinessen Haugen	4093
<i>KOCOH: Korean Context-Dependent Hate Speech Dataset</i>	
Eunah Park and Sanghoun Song	4103
<i>Towards Fair Speech Recognition: Mitigating Demographic Bias in End-to-End ASR Systems</i>	
Maliha Jahan, Thomas Thebaud, Zsuzsanna Fagyal, Jesus Villalba, Mark Hasegawa-Johnson, Laureano Moro Velazquez and Najim Dehak	4115
<i>RuBIN: A Russian Benchmark for Evaluating LLMs with Cultural Insights</i>	
Polina Lazukova and Irina Piontkovskaya	4126

<i>Evaluating Phonetically Weighted and Unweighted Distance Measures in Dialectometry</i> Alfred Lameli	4141
<i>Piecing Together Cross-Document Coreference Resolution Datasets: Systematic Dataset Analysis and Unification</i> Anastasia Zhukova, Terry Lima Ruas, Jan Philip Wahle and Bela Gipp	4152
<i>Spotlights and Blindspots: Evaluating Machine-Generated Text Detection</i> Kevin Stowe and Kailash Patil	4173
<i>JAPAS: A Benchmark and Neural Approach for Japanese Patent Support Relation Extraction</i> Katsuki Chousa and Ryosuke Sugiura	4188
<i>A Teacher-Student Approach to Creating Verified Synthetic Clarification and Correction Dialogues for TableQA Tasks</i> Christian Poelitz and Nick McKenna	4199
<i>Persona-Aware Evaluation of Cognitive Bias in LLMs: From Benchmark to Applied Decision-Making</i> Katsumasa Yoshikawa, Junya Takayama and Takato Yamazaki	4213
<i>ArtistMus: A Globally Diverse, Artist-Centric Benchmark for Retrieval-Augmented Music Question Answering</i> Daeyong Kwon, SeungHeon Doh and Juhan Nam	4226
<i>MATA: Mindful Assessment of the Telugu Abilities of Large Language Models</i> Chalamalasetti Kranti and Sowmya Vajjala	4239
<i>Estonian Native Large Language Model Benchmark</i> Helena Grete Lillepalu and Tanel Alumäe	4257
<i>Indirect Question Answering in English, German and Bavarian: A Challenging Task for High- and Low-Resource Languages Alike</i> Miriam Winkler, Verena Blaschke and Barbara Plank	4268
<i>Benchmarking Large Language Models for Chinese and Japanese IMEs: Phonetic-to-Character Generation and Textual Error Correction</i> Yuchun Zou, Tedd Lee, Xiaodi Fan and Jun Li	4290
<i>DaLA: Danish Linguistic Acceptability Evaluation Guided by Real World Errors</i> Gianluca Barmina, Nathalie Carmen Hau Norman, Peter Schneider-Kamp and Lukas Galke Poech	4312
<i>KCIF: Knowledge-Conditioned Instruction Following</i> Rudra Murthy, Praveen Venkateswaran, Prince Kumar and Danish Contractor	4327
<i>GAIN: A Benchmark for Goal-Aligned Decision-Making of Large Language Models under Imperfect Norms</i> Masayuki Kwarada, Kodai Watanabe and Soichiro Murakami	4346
<i>Can LLMs Evaluate What They Cannot Annotate? Revisiting LLM Reliability in Hate Speech Detection</i> Paloma Piot, David Otero, Patricia Martin-Rodilla and Javier Parapar	4358

<i>PersianMedQA: Evaluating Large Language Models on a Persian-English Bilingual Medical Question Answering Benchmark</i>	
Mohammad Javad Ranjbar Kalahroodi, Amirhossein Sheikholeslami, Sepehr Karimi Arpanahi, Sepideh Ranjbar Kalahroodi, Heshaam Faili and Azadeh Shakery	4371
<i>HatePrototypes: Interpretable and Transferable Representations for Implicit and Explicit Hate Speech Detection</i>	
Irina Proskurina, Marc-Antoine Carpentier and Julien Velcin	4387
<i>Investigating Memorization in Language Models Trained via Knowledge Distillation</i>	
Maarten Mäcking and Michaela Regneri	4400
<i>Redefining Evaluation Standards: A Unified Framework for Evaluating the Korean Capabilities of Language Models</i>	
Hanwool Lee, Dasol Choi, Sooyong Kim, Ilgyun Jung, Sangwon Baek, Guijin Son, Inseong Hwang, Naeun Lee and Seunghyeok Hong	4414
<i>Cross-Lingual Stability and Bias in Instruction-Tuned Language Models for Humanitarian NLP</i>	
Poli Nemkova, Amrit Adhikari, Matthew Pearson, Vamsi Krishna Sadu and Albert V. Mark	4423
<i>Counting on Consensus: Selecting the Right Inter-Annotator Agreement Metric for NLP Annotation and Evaluation</i>	
Joseph H. F. James	4434
<i>Quadratic Weighted Kappa Is Not Enough for Evaluating Automated Essay Scoring Models</i>	
Salam Albatarni and Tamer Elsayed	4447
<i>Evaluating the Homogeneity of Keyphrase Prediction Models</i>	
Mael Houbre, Florian Boudin and Beatrice Daille	4457
<i>A Taxonomy of Safety: Harmonizing LLM Benchmarks in a Fragmented Landscape</i>	
Shadi Rastegar, Viktor Hangya, Fabian Kuech and Darina Gold	4470
<i>Consistency of LLMs to Comparative Statements in Mathematical Reasoning Tasks</i>	
Aidan W. San, Daniel Juyoung Son, Xiaodong Liu and Yangfeng Ji	4482
<i>PersianAnonymizer: Evaluating LLM-Labeled Training for Efficient NER-based Anonymization in Persian</i>	
Mohammad Hossein Shalchian, Mostafa Amiri and Amir Mahdi Sadeghzadeh	4497
<i>How Many Samples Do We Need? A Toolkit for Power-Aware Evaluation Design</i>	
Angelo Basile, Areg Mikael Sarvazyan and José Ángel González	4507
<i>Of Words and Meaning: A Grammatical and Semantic Benchmark for Faroese LLM Understanding</i>	
Iben Nyholm Debess, Barbara Scalvini and Bolette Pedersen	4514
<i>TURING: Evaluating Human Abilities to Identify AI-Generated Texts</i>	
Natalia Kalashnikova, Nicolas De Bufala, Sophie Fayad and Laurent Cervoni	4527
<i>JamC-QA: A Multiple-Choice Question Answering Benchmark for Japan-Specific Knowledge</i>	
Teruaki Oka, Tomohide Shibata and Nao Yoshida	4536

<i>Towards Dynamic Metaphor Identification: Evaluating GPT O-Series Models on Five Metaphoricity Cues in U.S. Trade Corpora</i>	
Berkay Bas, Jelke Bloem and Xiaojuan Tan	4547
<i>Evaluating Text Style Transfer: A Nine-language Benchmark for Text Detoxification</i>	
Vitaly Protasov, Nikolay Babakov, Daryna Dementieva and Alexander Panchenko ...	4560
<i>Irish-BLiMP: A Linguistic Benchmark for Evaluating Human and Language Model Performance in a Low-Resource Setting</i>	
Josh McGiff, Tung Khanh Tran, William Mulcahy, Dáibhidh Ó Luínín, Jake Dalzell, Róisín Ní Bhroin, Adam Burke, Barry O’Sullivan, Hoang D. Nguyen and Nikola S. Nikolov	4575
<i>EduBench: A Portuguese Benchmark for Open-Ended Discursive Question Answering</i>	
Pedro Henrique Paiola, Luís Gabriel Damiani Mendes, Bruno de Oliveira Monchelato, André da Fonseca Schuck, Gabriel Lino Garcia, Douglas Rodrigues, Helena de Medeiros Caseli and João Paulo Papa	4587
<i>SemBench: A Universal Semantic Framework for LLM Evaluation</i>	
Mikel Zubillaga, Naiara Perez, Oscar Sainz and German Rigau	4597
<i>EL-MIA: Quantifying Membership Inference Risks of Sensitive Entities in LLMs</i>	
Ali Satvaty, Suzan Verberne and Fatih Turkmen	4611
<i>Same Meaning, Different Scores: Lexical and Syntactic Sensitivity in LLM Evaluation</i>	
Bogdan Kostić, Conor Fallon, Julian Risch and Alexander Loeser	4626
<i>The Potential for Misleading Results in Text Sanitisation with Standard Evaluation Metrics</i>	
Dan Zhang and Mark Anderson	4638
<i>Mind the Language Gap: Assessing LLM Safety in Italian</i>	
Elena Marafatto and Roberto Navigli	4647
<i>Bulgarian Massive Multitask Language Understanding Benchmark</i>	
Svetla Peneva Koeva, Ivelina Stoyanova, Dimiter Georgiev, Svetlozara Leseva, Valentina Stefanova, Maria Todorova, Tsvetana Ivanova Dimitrova, Hristina Kukova, Mihaela Moskova and Tinko Tinchev	4658
<i>PHEB: An European Portuguese High School-Level LLM Benchmark</i>	
Diogo C. Tavares, Rafael Ferreira, Afonso Simplício, Gonçalo Vinagre, Ana Carolina Condez, Inês Calvo, Inês Vieira, David Semedo and Joao Magalhaes	4673
<i>S-GRADES – Studying Generalization of Student Response Assessments in Diverse Evaluative Settings</i>	
Tasfia Seuti and Sagnik Ray Choudhury	4684
<i>Who Benchmarks the Benchmarks? A Case Study of LLM Evaluation in Icelandic</i>	
Finnur Ágúst Ingimundarson, Steinunn Rut Fridriksdottir, Bjarki Ármannsson, Iris Nowenstein and Steinþór Steingrímsson	4702
<i>Is This Idea Novel? An Automated Benchmark for Judgment of Research Ideas</i>	
Tim Schopf and Michael Färber	4716

<i>Questionnaire Meets LLM: A Benchmark and Empirical Study of Structural Skills for Understanding Questions and Responses</i>	
Duc-Hai Nguyen, Vijayakumar Nanjappan, Barry O'Sullivan and Hoang D. Nguyen . .	4728
<i>Assessing the Effectiveness of LLMs in Delivering Cognitive Behavioral Therapy</i>	
Navdeep Singh Bedi, Ana-Maria Bucur, Noriko Kando and Fabio Crestani	4747
<i>Transcription Accuracy in the Icelandic Gigaword Corpus: Evaluating Automatic and Manual Annotation</i>	
Johanna Mechler, Lilja Björk Stefánsdóttir and Anton Karl Ingason	4757
<i>Benchmark Data Contamination in Underrepresented Languages: A Comprehensive Analysis Using Brazilian Data</i>	
Iriedson Souto Maior de Moraes Vilar, David Candeia Maia, João Brunet, Fabio Morais and Leandro Balby Marinho	4765
<i>TTSVowelViz: A Tool for Visualising Text-to-Speech Model Training via Vowel Spaces</i>	
Pasindu Udawatta, Jesin James, Balamurali B T, Catherine Inez Watson, Ake Nicholas and Binu Nisal Abeysinghe	4778
<i>A Sociophonetic Analysis of Racial Bias in Commercial ASR Systems Using the Pacific Northwest English Corpus</i>	
Michael Scott, Siyu Liang, Alicia Wassink and Gina-Anne Levow	4787
<i>ParliaBench: An Evaluation and Benchmarking Framework for LLM-Generated Parliamentary Speech</i>	
Marios Koniaris, Argyro Tsipi and Panayiotis Tsanakas	4797
<i>PARSEME 2.0 Multilingual Corpus of Multiword Expressions</i>	
Agata Savary, Manon Scholivet, Carlos Ramisch, Takuya Nakamura, Eric Bilinski, Sara Stymne, Voula Giouli, Stella Markantonatou, Vasile Pais, Maria Mitrofan, Louis Estève, Bruno Guillaume, Verginica Barbu Mititelu, Jaka Čibej, Roberto Díaz Hernández, Victoria Fendel, Polona Gantar, Olha Kanishcheva, Cvetana Krstev, Chaya Liebeskind, Irina Lobzhanidze, Aleksandra M. Marković, Gunta Nešpore-Bērzkalne, Adriana S. Pagano, Mehrnoush Shamsfard, Ranka Stankovic, Vahide Tajalli, Carole Tiberius and Aakanksha Padhye	4819
<i>Introducing PerMet 1.0: A Metaphor-Annotated Corpus for Persian</i>	
Mohammad Saeid Miri	4835
<i>Multi-SimLex for Dutch: Benchmarking Embedding- and Prompt-Based Model Performance on Semantic Similarity</i>	
Lizzy Brans and Jelke Bloem	4846
<i>MultiCoS: A Multilingual Dataset of Connective Semantics with Context–Sentence Compatibility</i>	
Anne Mucha, Ciyang Qing and Wataru Uegaki	4861
<i>Adverbs Revisited: Enhancing WordNet Coverage of Adverbs with a Supersense Taxonomy</i>	
Jooyoung Lee, Jader Martins Camboim de Sá and Cedric Pruski	4872
<i>KinyCOMET: Automatic Evaluation of Machine Translation Systems for Kinyarwanda–English</i>	
Prince Chris Mazimpaka, Jan Nehring, Samuel Rutunda and Cristina España-Bonet	4881

<i>Multiway Parallel Corpus in Forced Migration Domain for Multilingual Machine Translation</i> Fatemeh Azadi, Samuel Larkin and Chi-kiu Lo	4889
<i>Context-8: A Data Set for Evaluating Context Sensitivity in Machine Translation</i> Dongyue Wang and Kyo Kageura	4902
<i>AssamLegalTrans: A Parallel Corpus, Benchmark and Analysis for English-Assamese Machine Translation of Legal Judgments</i> Telem Joyson Singh, Hemanta Baruah, Sanasam Ranbir Singh, Anindita Talukdar, Nasrin Shahnaz, Okram Jimmy Singh, Priyankoo Sarmah, Pallav Kumar Dutta, Sukumar Nandi and Pranab Duara	4921
<i>Coordinate Structure Extraction for Patent Claims Using Multilingual LLMs</i> Tsukasa Ishimaru, Takehito Utsuro and Masaaki Nagata	4931
<i>Human Label Variation in Implicit Discourse Relation Recognition</i> Frances Yung, Daniil Ignatev, Merel Scholman, Vera Demberg and Massimo Poesio	4942
<i>Conversational Implicatures through the Lens of LLMs</i> Agnese Lombardi and Alessandro Lenci	4955
<i>The Emergence of the Pragmatic Dimension in Instructed-LMs</i> Davide Mazzaccara and Raffaella Bernardi	4967
<i>Distributed Partial Information Puzzles: Examining Common Ground Construction under Epistemic Asymmetry</i> Yifan Zhu, Mariah Bradford, Kenneth Lai, Timothy Obiso, Videep Venkatesha, James Pustejovsky and Nikhil Krishnaswamy	4974
<i>Grounded Misunderstandings in Asymmetric Dialogue: A Perspectivist Annotation Scheme for MapTask</i> Nan Li, Albert Gatt and Massimo Poesio	4988
<i>Assessing LLM Reasoning through Implicit Causal Chain Discovery in Climate Discourse</i> Liesbeth Allein, Nataly Pineda-Castañeda, Andrea Rocci and Marie-Francine Moens	5002
<i>AccurateRAG: A Framework for Building Accurate Retrieval-Augmented Question-Answering Applications</i> Linh The Nguyen, Chi Tran, Dung Ngoc Nguyen, Van-Cuong Pham, Hoang Ngo and Dat Quoc Nguyen	5015
<i>VideoEvent: Leveraging Relevance and LLMs for Video Question Answering</i> Chen-Chen Lin, Ming-Han Lee, KunRu Wu and Yu-Chee Tseng	5024
<i>MORQA: Benchmarking Evaluation Metrics for Medical Open-Ended Question Answering</i> Wen-wai Yim, Asma Ben Abacha, Zixuan Yu, Robert Doerning, Fei Xia and Meliha Yetisgen	5035
<i>LegalRikai: Open Benchmark – a Benchmark for Complex Japanese Corporate Legal Tasks</i> Shogo Fujita, Yuji Naraki, Yiqing Zhu and Shinsuke Mori	5055
<i>Integrating Arithmetic Learning Improves Mathematical Reasoning in Smaller Models</i> Neeraj Gangwar, Suma Bhat and Nickvash Kani	5078

<i>mSCoRe: A Multilingual and Scalable Benchmark for Skill-based Commonsense Reasoning</i> Nghia Trung Ngo, Franck Dernoncourt and Thien Huu Nguyen	5095
<i>A Binary Problem in Binary QA: Diverse LLMs or Diverse Question Interpretations? That Is the Ensembling Question</i> Rafael Rosales and Santiago Miret	5116
<i>ObfusQAte: A Proposed Framework to Evaluate LLM Robustness on Obfuscated Factual Question Answering</i> Shubhra Ghosh, Abhilekh Borah, Aditya Kumar Guru and Kripabandhu Ghosh	5129
<i>POLAR: A Corpus of Questions, Responses and Argumentation in Polish Political Radio Discourse</i> Daniel Ziembicki, Aleksandra Zwierzchowska, Ewelina Sobol and Katarzyna Anna Przerada	5146
<i>MHTS: Multi-Hop Tree Structure Framework for Generating Difficulty-Controllable QA Datasets for RAG Evaluation</i> Jeongsoo Lee, Daeyong Kwon, Kyohoon Jin, JunNyeong Jeong, Minwoo Sim and Minwoo Kim	5158
<i>CareMedEval Dataset: Evaluating Critical Appraisal and Reasoning in the Biomedical Field</i> Doria Bonzi, Alexandre Guiggi, Frederic Bechet, Carlos Ramisch and Benoit Favre ..	5169
<i>LongTailQA: Benchmarking LLMs and RAG Models on Disambiguated Long-Tail Entities</i> William Xion, Uwe Hadler, Tim Cofala, Maximilian Idahl, Soumyadeep Roy and Wolfgang Nejdl	5182
<i>CRaFT: An Explanation-Based Framework for Evaluating Cultural Reasoning in Multilingual Language Models</i> Shehenaz Hossain and Haithem Afli	5192
<i>HEAD-QA v2: Expanding a Healthcare Benchmark for Reasoning</i> Alexis Correa, Carlos Gómez-Rodríguez and David Vilares	5203
<i>Beyond MCQ: An Open-Ended Arabic Cultural QA Benchmark with Dialect Variants</i> Hunzalah Hassan Bhatti and Firoj Alam	5215
<i>Automatic Inter-document Multi-hop Scientific QA Generation</i> Seungmin Lee, Dongha Kim, Yuni Jeon, Junyoung Koh and Min Song	5232
<i>CRiT-QA: Evaluating Multi-hop Reasoning with Counterfactual Chains and Distractor Traps</i> Jungmin Yun, June Hyoung Kwon and Youngbin Kim	5246
<i>TARAZ: Persian Short-Answer Question Benchmark for Cultural Evaluation of Language Models</i> Reihaneh Iranmanesh, Saeedeh Davoudi, Pasha Abrishamchian, Ophir Frieder and Nazli Goharian	5256
<i>Benchmarking Mathematical Reasoning in a Low-Resource Language: Structured Prompting and Evaluation in Basque</i> Inigo Martinez-Criado, Aitor Soroa and Jeremy Barnes	5268
<i>Assessing the Difficulty of Inference Types in Natural Language Inference for Clinical Trials</i> Mathilde Aguiar, Pierre Zweigenbaum and Nona Naderi	5290

<i>Reasoning Graph-Structured Question Answering: Datasets and Insights from LLM Benchmarking</i>	
Khin Yone, Devasha Trivedi, Anish Pahilajani, Jincen Shuai, Samyak Rajesh Jain, Ryan Rossi, Nesreen K. Ahmed, Franck Dernoncourt, Yu Wang and Namyoung Park	5301
<i>JBE-QA: Japanese Bar Exam QA Dataset for Assessing Legal Domain Knowledge</i>	
Zhihan Cao, Fumihito Nishino, Hiroaki Yamada, Ha Thanh Nguyen, Yusuke Miyao and Ken Satoh	5317
<i>A Diagnostic Benchmark for Sweden-Related Factual Knowledge</i>	
Jenny Kunz	5328
<i>GeoBenchmark: Probing Large Language Models for Geo-Spatial Knowledge</i>	
Ayomide Abayomi, Jose G. Moreno, Karim Radouane and Lynda Tamine	5335
<i>FactORes: Fact-checking with an Evidence-based Open Resource in Spanish</i>	
Nagore Bravo, Jaione Bengoetxea, Iker García-Ferrero, Alba Bonet Jover, Estela Saquete and Rodrigo Agerri	5349
<i>Stands to Reason: Investigating the Effect of Reasoning on Idiomaticity Detection</i>	
Dylan Phelps, Rodrigo Wilkens, Edward Gow-Smith, Thomas M. R. Pickard, Maggie Mi, Marco Idiart and Aline Villavicencio	5367
<i>ESG-QA: Building a Dataset for Question Answering on Environmental, Social, and Governance Pillars</i>	
Gabriel Assis, Ayrton Surica, Pedro Kroll, Gabriela Aires Mendes, Darian Rabbani, Edson Bollis, Lucas Francisco Amaral Orosco Pellicer and Aline Paes	5377
<i>Enhancing and Evaluating Tabular Models on the Fly via Synthetic Question–Answer Generation</i>	
Jorge Osés Grijalba, Eugenio Martínez Cámara, L. Alfonso Ureñ-López and Jose Camacho-Collados	5389
<i>VIVID: A Culturally Grounded Benchmark Exposing the Figurative Language Gap in Vietnamese NLP</i>	
Tu Tran Do, Nhat Ngoc Nguyen, Tung Khanh Tran, Hoang D. Nguyen, Tu Minh Phuong and Long Hoang Dang	5414
<i>Assessing Logical Coherence of LLMs via Fine-Grained NLI</i>	
Jon Felix Apaolaza Larraya, Begoña Altuna, Aitor Soroa and Inigo Lopez-Gazpio ...	5431
<i>Counter-Hypothesis Generation: Towards Evaluating How LLMs Reason about Alternatives</i>	
Marzieh Abdolmaleki, Aaron Maladry, Veronique Hoste and Els Lefever	5445
<i>LFQA-HP-1M: A Large-Scale Human Preference Dataset for Long-Form Question Answering</i>	
Rafid Ishrak Jahan, FAHMID SHAHRIAR IQBAL and Sagnik Ray Choudhury	5450
<i>Orthographic Constraint Satisfaction and Human Difficulty Alignment in Large Language Models</i>	
Bryan E. Tuck and Rakesh Verma	5466
<i>LIT-RAGBench: Benchmarking Generator Capabilities of Large Language Models in Retrieval-Augmented Generation</i>	
Koki Itai, Shunichi Hasegawa, Yuta Yamamoto, Gouki Minegishi and Masaki Otsuki .	5482

<i>Investigating Reasoning with Hypotheses: The RIP2 Corpus</i> Ella Schad, Clara Seyfried and Chris Reed	5494
<i>Can Multimodal LLMs Generate Pedagogical Questions?</i> Thomas Gerald, Sahar Ghannay, Julie Lascar, Paul Lerner and Anne Vilnat	5506
<i>The Riddle of Reflection: Evaluating Reasoning and Self-Awareness in Multilingual LLMs Using Indian Riddles</i> Abhinav P M, Ojasva Saxena, Oswald C and Parameswari Krishnamurthy	5516
<i>Using Songs to Improve Kazakh Automatic Speech Recognition</i> Rustem Yeshpanov	5528
<i>Southern Kurdish Speech Recognition Resources and Benchmarking</i> Mohammad Mohammadamini and Marie Tahon	5538
<i>MASA: A Novel Multimodal Foundation Model for L2 Speaking Assessment in Picture-description Scenarios</i> Bi-Cheng Yan, Fu-An Chao, Hong-Yun H.Y. Lin and Berlin Chen	5545
<i>Tools for Estimating the Perceived Level of Phonetic Reduction</i> Nigel Ward, Javier Vazquez, Emma (Danny) R. Boushka and Oliver Niebuhr	5555
<i>FaLAR: A Large-scale Speaker-Annotated European Portuguese Speech Corpus of Parliamentary Sessions</i> Francisco Teixeira, Carlos Carvalho, Mariana Julião, Catarina Botelho, Rubén Solera-Ureña, Sérgio Paulo, Thomas Rolland, Ben Peters, Isabel Trancoso and Alberto Abad ...	5566
<i>English to Central Kurdish Speech Translation: Corpus Creation, Evaluation, and Orthographic Standardization</i> Mohammad Mohammadamini, Daban Jaff, Josep Crego, Marie Tahon and Antoine LAURENT	5578
<i>Automatic Prediction of Prominence and Boundary Strength from Text</i> Pauline Mas, Kévin Vythelingum, Jonathan Chevelu, Marion Ouédraogo, Damien Lolive and Olivier Rosec	5588
<i>SOMVOICE: A First Dataset to Study the Effects of Sleep Deprivation on Voice Characteristics of Healthy French Speakers</i> Vincent P. Martin, Jean-Luc Rouas, Colleen Beaumard and Pierre Philip	5597
<i>Automatic Prediction of Child Speech Fluency with Game-Based Data from German Preschoolers</i> Valentin Kany, Bernd Möbius and Jürgen Trouvain	5607
<i>Selective Augmentation: Improving Universal Automatic Phonetic Transcription via G2P Bootstrapping</i> Tobias Bystrich, Julia Maria Pritzen, Christoph Andreas Schmidt and Claudia Wich-Reif	5617
<i>AURORA Model of Formant-to-tongue Inversion for Didactic and Clinical Applications</i> Patrycja Strycharczuk and Sam Kirkham	5625

<i>Investigating the Role of Synthetic Data Augmentation and Training Strategies on Improving Low-Resource Language ASR</i>	
Yun Hao, Reihaneh Amooie, Wietse de Vries, Rik van Noord and Martijn Wieling ...	5633
<i>AutoRPT: A Tool for Bootstrapping Prosodic Annotation</i>	
Seth Heiney, Thomas Hicks, Sally Little, Fernanda Lourenco, Kai Retana, Eliana Stevens and Jonathan Howell	5640
<i>J-CHAT: Japanese Large-scale Spoken Dialogue Corpus for Spoken Dialogue Language Modeling</i>	
Wataru Nakata, Kentaro Seki, Hitomi Yanaka, Yuki Saito, Shinnosuke Takamichi and Hiroshi Saruwatari	5649
<i>ViMedCSS: A Vietnamese Medical Code-Switching Speech Dataset & Benchmark</i>	
Tung X. Nguyen, Nhu Vo, Giang Son Nguyen, Duy Mai Hoang, Chien Dinh Huynh, Inigo Jauregi Unanue, Massimo Piccardi, Wray Buntine and Dung D. Le	5657
<i>Towards Privacy-Preserving Fine-Tuning: Anonymization of Aphasic Speech for Effective ASR</i>	
Sebastian Hofstetter and Timo Baumann	5666
<i>ParlaSpeech 3.0: Richly Annotated Spoken Parliamentary Corpora of Croatian, Czech, Polish, and Serbian</i>	
Nikola Ljubešić, Peter Rupnik, Ivan Porupski and Taja Kuzman Pungeršek	5677
<i>LexiPhon: A Collection of Phonetically Transcribed Lexicons from Wikipedia</i>	
Amanda Doucette, Timothy J. O'Donnell and Morgan Sonderegger	5689
<i>ROG: A Multi-Layer Manually Annotated Corpus of Spoken Slovenian</i>	
Kaja Dobrovoljc Zor, Darinka Verdonik, Jaka Čibej, Peter Rupnik and Nikola Ljubešić	5701
<i>Building a Dataset for French Accent Classification Evaluation: Are We There Yet?</i>	
Diandra Fabre, Mathieu Avanzi and François Portet	5711
<i>M3-SLU: Evaluating Speaker-Attributed Reasoning in Multimodal Large Language Models</i>	
Yejin Kwon, TAEWOO KANG, Hyunsoo Yoon and Chang Ouk Kim	5722
<i>Medispeech: A French Reading and Spontaneous Speech Corpus for Sleepiness Estimation</i>	
Colleen Beaumard, Vincent P. Martin, Charles Brazier, Julien Coelho, Jean-Luc Rouas and Pierre Philip	5737
<i>StarDrinks: An English and Korean Test Set for SLU Evaluation in a Drink Ordering Scenario</i>	
Marcely Zanon Boito, Caroline Brun, Inyoung Kim, Denys M. PROUX, Salah Ait-Mokhtar, Nikolaos Lagos, Jean-Luc Meunier and Ioan Calapodescu	5749
<i>Audio-Lyrics Alignment Dataset for Italian Arias</i>	
Pushkar Jajoria, Arianna Graciotti, Giovanna Casali, Jesujoba Alabi, Rodolfo Delmonte, Angelo Pompilio, Rocco Tripodi, James McDermott and Dietrich Klakow	5757
<i>The Added Value of Metadata and Annotations: Evidence from Two Large-Scale, Naturalistic Corpus Studies</i>	
Anisia Popescu, Johanna Cronenberg, Ioana Vasilescu, Ioana Chitoran, Lori Lamel and Martine Adda-Decker	5767

<i>CS-YODAS: A Mined Dataset of In-the-Wild Code-Switched Speech</i>	
Brian Yan, Qingzheng Wang, Matthew Wiesner, Anuj Diwan, Olga Iakovenko, Alex Polok, Injy Hamed, Shuichiro Shimizu, Iris Emerman, Thomas Hain, David R. Mortensen, Peter Viechnicki and Shinji Watanabe	5776
<i>The Limits of Data Scaling: Sub-token Utilization and Acoustic Saturation in Multilingual ASR</i>	
Siyu Liang, Nicolas Ballier, Gina-Anne Levow and Richard Wright	5785
<i>AusKidTalk: Developing Transcription Guidelines for Continuous Australian English Child Speech</i>	
Tuende Szalay, Zheng Nan, Renata Huang, Mostafa Shahin, Sirojan Tharmakulasingam, Kirrie Ballard and Beena Ahmed	5794
<i>spINAch: A Diachronic Corpus of French Broadcast Speech Controlled for Speakers' Age and Gender</i>	
Simon Devauchelle, David Doukhan, Remi Uro, Lucas Ondel, Valentin Pelloin, Olympia Imbert-Brégégère, Véronique Lefort, Kévin Picard, Emeline Seignobos and Albert Rilliard	5805
<i>SALAN: A Massive ASR Dataset for the Languages of Niger</i>	
Mamadou K KEITA, Christopher Homan, Emily Prud'hommeaux, Abdoulaye SAKO and Seydou Diallo	5821
<i>Listening for Ideology: Automatic Analysis of Character Speech in Historical Nazi Propaganda Films</i>	
Nicolas Ruth, Manuel Burghardt and Andreas Niekler	5828
<i>Supplementary Resources and Analysis for Automatic Speech Recognition Systems Trained on the Loquacious Dataset</i>	
Nick Rossenbach, Robin Schmitt, Tina Raissi, Simon Berger, Larissa Kleppel and Ralf Schlüter	5839
<i>WhiteHouse: Translation of the Casablanca Corpus for Multi-dialectal Arabic Speech Translation</i>	
Fethi Bougares, Salima Mdhaffar and Yannick Estève	5849
<i>ToneSwiper: Facilitating Manual ToDI-annotation of Dutch Prosody</i>	
Matthijs Westera and Ariëlle Reitsema	5856
<i>IMaSC: A Malayalam Speech Corpus for High-Quality Text-to-Speech Synthesis</i>	
Deepa P. Gopinath, Thennal D K, Vrinda V. Nair, Swaraj K. S and Sachin G	5864
<i>Speak in Context: Multilingual ASR with Speech-Context Alignment via Contrastive Learning</i>	
Yuchen Zhang, Haralambos Mouratidis and Ravi Shekhar	5873
<i>Task-Lens: Cross-Task Utility Based Speech Dataset Profiling for Low-Resource Indian Languages</i>	
Swati Sharma, Divya V. Sharma and Anubha Gupta	5883
<i>Introducing MELI: The Mandarin-English Language Interview Corpus</i>	
Suyuan Liu and Molly Babel	5896

<i>PhonemeDF: A Synthetic Speech Dataset for Audio Deepfake Detection and Naturalness Evaluation</i>	
Vamshi Nallaguntla, Aishwarya R. Fursule, Shruti Kshirsagar and Anderson Raymundo Avila	5905
<i>How Much Data for Stable Formant Values? Pipeline for Convergence Detection Based on Read Speech</i>	
Kayla Sward, Johan Sjons and Axel G. Ekstrom	5916
<i>MUSCAT: MULTilingual, SCientific ConversATion Benchmark</i>	
Supriti Sinhamahapatra, Thai-Binh Nguyen, Yiğit Oğuz, Enes Yavuz Ugan, Jan Niehues and Alexander Waibel	5926
<i>Towards a Diagnostic and Predictive Evaluation Methodology for Sequence Labeling Tasks</i>	
Elena Alvarez-Mellado and Julio Gonzalo	5938
<i>Memorization or Lucky Guesses: Detecting Short Sequences from Copyrighted Dutch News in LLM Output</i>	
Joris Veerbeek, Kas Berendsen, Alessandra Polimeno and Antal van den Bosch	5960
<i>When Numbers Tell Half the Story: Human-Metric Alignment in Topic Model Evaluation</i>	
Thibault Prouteau, Francis Lareau, Nicolas Dugue, Jean-Charles Lamirel and Christophe Malaterre	5970
<i>Detecting Hallucinations in Authentic LLM–Human Interactions</i>	
Yujie Ren, Niklas Gruhlke and Anne Lauscher	5981
<i>Issue Detection and Category Classification in Domain-Specific Technical Logbooks</i>	
Afshin Karimi, Ingmar Hartl, Henrik Tuennermann and Anne Lauscher	5996
<i>Once upon a Kernel: Extracting Important Events from Narratives</i>	
Anshu Kiran Sharma, Miguel Castiblanco-Melendez, Alejandro Morales and Mark A. Finlayson	6009
<i>Temporal Expression Recognition in Legal Transcripts</i>	
Elizabeth J. Goldstein and Maria Berger	6022
<i>Multilingual, Multimodal Pipeline for Creating Authentic and Structured Fact-Checked Claim Dataset</i>	
Z. Melce Hüsünbeyi, Virginie Mouilleron, Leonie Uhling, Daniel Foppe, Tatjana Scheffler and Djamé Seddah	6038
<i>A Study on Building Efficient Zero-Shot Relation Extraction Models</i>	
Hugo THOMAS, Caio Corro, Guillaume Gravier and Pascale Sébillot	6056
<i>Beyond Catalogue Counts: The Dataset Visibility Asymmetry in Low-Resource Multilingual NLP</i>	
Zhiyin Tan and Changxu Duan	6068
<i>BLooP: Zero-Shot Abstractive Summarization Using Large Language Models with Bigram Looka-head Promotion</i>	
Varun Iyer and Cornelia Caragea	6080

<i>OasisSimp: An Open-source Asian-English Sentence Simplification Dataset</i>	
Hannah Liu, Murphy Tian, Iqra Ali, Haonan Gao, Qiaoyiwen Wu, Blair Yang, Uthayasanker Thayasivam, Annie En-Shiun Lee, Pakawat Nakwijit, Surangika Ranathunga and Ravi Shekhar	6103
<i>Fully Automated Identification of Lexical Alignment and Preference-Stage Shifts in Large Language Models</i>	
Thomas Stephan Juzek, Xiaoyang Ming and Jose A. Hernandez	6116
<i>How Much Noise Can BERT Handle? Insights from Multilingual Sentence Difficulty Detection</i>	
Nouran Khallaf and Serge Sharoff	6132
<i>Comparing Reading Behavior across Reader Expertise and Text Complexity: Insights from the French Eye-Tracking Corpus (FETA)</i>	
Oksana Ivchenko and Natalia Grabar	6144
<i>Scaling LLM Reasoning from Minimal Labels: A Semi-Supervised Framework with a Lightweight Verifier</i>	
Keizo Kato, Chenhui Chu, Yugo Murawaki and Sadao Kurohashi	6155
<i>PARL: Prompt-based Agents for Reinforcement Learning</i>	
Yarik Menchaca Resendiz and Roman Klinger	6166
<i>SPQ: An Ensemble Technique for Large Language Model Compression</i>	
Jiamin Yao and Eren Gultepe	6185
<i>FPSC: A Sustainable Pipeline for Building a Faroese Parliamentary Speech Corpus</i>	
Dávid í Lág, Barbara Scalvini, Carlos Daniel Hernandez Mena and Jon Gudnason	6196
<i>Efficient Dialect-Aware Modeling and Conditioning for Low-Resource Taiwanese Hakka Speech Processing</i>	
Peng An-Ci, Kuan-Tang Huang, Tien-Hong Lo, Hung-Shin Lee, Hsin-Min Wang and Berlin Chen	6206
<i>Construction of Japanese Prefectural Assembly Minutes Datasets across Three Electoral Terms: Comparative Analysis of 2011, 2015, and 2019 Four-Year Periods</i>	
Keiichi Takamaru, Hokuto Ototake, Yuzu Uchida and Yasutomo Kimura	6216
<i>EDDA-Coordinata: An Annotated Dataset of Historical Geographic Coordinates</i>	
Ludovic Moncla, Pierre Nugues, Thierry Joliveau and Katherine McDonough	6224
<i>Mental Health Disorder Detection beyond Social Media: A Systematic Review of Available Datasets</i>	
Sadiya Sayara Chowdhury Puspo, Ana-Maria Bucur, Stevie Chancellor, Özlem Uzuner and Marcos Zampieri	6235
<i>German Counseling Grounding-Act Corpus (GRACO)</i>	
Milena Belosevic	6251
<i>Presenting the Prague Discourse Treebank 4.0</i>	
Jiří Mírovský and Pavlína Synková	6262
<i>Evaluation of Co-Speech Gesture Tracking Techniques in Naturalistic Interactions</i>	
Victoria Ivanova and Naomi Harte	6277

<i>Voices across Decades: A Multimodal Diachronic Corpus of German Bundestag Debates (GerParlDi-MM)</i>	
Ingo Siegert	6289
<i>MultiWikiQA: A Reading Comprehension Benchmark in 300+ Languages</i>	
Dan Saattrup Smart	6298
<i>SALOMO: An Annotation Tool for Complex Annotation Tasks with a Large Number of Labels</i>	
Tim Menzner	6312
<i>VietJobs: A Vietnamese Job Advertisement Dataset</i>	
Hieu Pham Dinh, Hung Nguyen Huy and Mo El-Haj	6322
<i>A Resource on Dialogical Moves in Native and Non-Native Academic Writers of English</i>	
Giulia D'Agostino, Narjes Sheikh Asadi and Elena Musi	6332
<i>A Corpus-Based Profiling of Regional English Variants in Global Media: Insights from Olympic Journalism</i>	
Felix Mao	6340
<i>JFC-Recipe: A Dataset for Nutrient Estimation from Japanese User-Generated Cooking Recipes</i>	
Keisuke Shirai, Yoko Yamakata, Hirotaka Kameko, Akiko Sunto, Jun Harashima and Shinsuke Mori	6349
<i>Annotating Conversational Phases and Communication Techniques: A Corpus of German Teacher-Parent Counseling Conversations</i>	
Tobias Hallmen, Kathrin Gietl, Karoline Hillesheim, Annemarie Friedrich and Elisabeth André	6361
<i>RO-ABSA: A Romanian Dataset and Baselines for Aspect-Based Sentiment Analysis</i>	
Gheorghe Andreea Alina, Andrei Claudia, Ionescu Elena, Ruseti Stefan and Dascalu Mihai	6373
<i>The Moral Foundations Reddit Corpus</i>	
Jackson P. Trager, Alireza S. Ziabari, Elnaz Rahmati, Aida Mostafazadeh Davani, Preni Golazizian, Farzan Karimi-Malekabadi, Ali Omrani, Zhihe Li, Brendan Kennedy, Georgios Chochlakis, Nils Karl Reimer, Melissa Reyes, Kesley Cheng, Mellow Wei, Christina Merrifield, Arta Khosravi, Evans Alvarez and Morteza Dehghani	6383
<i>A Semi-Automatic Workflow for Transcribing and Annotating Broadcast News</i>	
Christoph Draxler, Sven Grawunder, Jürgen Trouvain and Felicitas Kleber	6408
<i>From Rosetta to Match-Up: A Paired Corpus of Linguistic Puzzles with Human and LLM Benchmarks</i>	
Neh Majmudar, Anne Huang, Jinfan Frank Hu and Elena Filatova	6418
<i>Tracing How Annotators Think: Augmenting Preference Judgments with Reading Processes</i>	
Karin Johanna Denton de Langis, William Walker, Khanh Chi Le and Dongyeop Kang	6428
<i>CodeClarity: A Framework and Benchmark for Evaluating Multilingual Code Summarization</i>	
Madhurima Chakraborty, Drishti Sharma, Maryam Sikander and Eman Nisar	6439

A Longitudinal, Multinational, and Multilingual Corpus of News Coverage of the Russo-Ukrainian War

Dikshya Mohanty, Taisiia Sabadyn, Jelwin Rodrigues, Chenlu Wang, Abhishek Kalugade and Ritwik Banerjee 6452

SKILL-IR-Discourse: A Large, Annotated Corpus of Argumentation and Domain Discourse on International Relations

Magdalena Wolska, Matti Wiegmann, Sassan Gholiagha, Mitja Sienknecht, Dora Kiesel, Irene Lopez Garcia, Patrick Riehmann, Bernd Fröhlich, Katrin Girgensohn, Jürgen Neyer and Benno Stein 6472

Building Multimodal Corpora Using Microtask Pipelines and Local Annotators

Helmiina Hotti, Raul Vazquez, Anna-Kaisa Jokipohja, Timo Kalliokoski, Henna Paakki, Rosa Suviranta and Tuomo Hiippala 6483

Beyond Fake News Detection: A Community-based Study of the Multicultural Nature of Information Disorder

Sara Gemelli, Giulia Di Cristina, Yiran Zhang, Md Azizul Hoque, Alberto De La Torre Solís, Mohamad Mojtoba Behboudi Eshkiki, Nikolai Efimov, Mariia Everstova, Caterina Maria Cappello, Maziar Kianimoghadam Jouneghani, Payam Latifi, Yashar Mahboudi, Farzaneh Mohseni, Dario Placenti, Tommaso Caselli, Manuela Sanguinetti, Aurora Scarpellini, Chiara Zanchi, Usman Naseem, Marco Antonio Stranisci and Simona Frenda 6496

FreeTxt-Vi: A Benchmarked Vietnamese-English Toolkit for Segmentation, Sentiment, and Summarisation

Hung Huy Nguyen, Mo El-Haj, Paul Rayson and Dawn Knight 6509

The Patrologia Graeca Corpus: OCR, Annotation, and Open Release of Noisy Nineteenth-Century Polytonic Greek Editions

Chahan Vidal-Gorène and Bastien Kindt 6519

National Library as Corpus: DeLiKo-2025@DNB – a Very Large Corpus of German-language Contemporary Literature

Marc Kupietz, Nils Diewald, Philippe Genêt and Andreas Witt 6528

Multi-party Conversational Corpus of L1 and L2 for Speech Alignment Research (Teams-SK): Methodological Approach

Stefan Benus, Viktor Gatjal, Erik György, Mária Hricková, Martin Kažimír, Zuzana Kozáčiková, Lucia Mareková, Róbert Sabo, Marian Trnka and Erik Vráb 6536

Is Semi-Automatic Transcription Useful in Corpus Creation? Preliminary Considerations on the KIParla Corpus

Martina Simonotti, Ludovica Pannitto, Eleonora Zucchini, Silvia Ballarè and Caterina Mauri 6547

Open Korean Historical Corpus: A Millennia-Scale Diachronic Collection of Public Domain Texts

Seyoung Song, Nawon Kim, Songeun Chae, Kiwoong Park, Jiho Jin, Haneul Yoo, Kyunghyun Cho and Alice Oh 6561

NAIST LIFE STORY: A Seven-Year Crowdsourced Dataset of Japanese Emotion-related Episodes

Kazuhiro Ito, Junko Hayashi, Hiroyuki Nagai, Shoko Wakamiya and Eiji ARAMAKI .. 6573

<i>Audience Engagement with Arabic Women's Social Empowerment and Wellbeing: A Decadal Corpus</i>	
Wajdi Zaghouani, Mabrouka Bessghaier, Md. Rafiul Biswas and Shimaa Amer Ibrahim	6585
<i>ArPoMeme: An Annotated Arabic Multimodal Dataset for Political Ideology and Polarization</i>	
Wajdi Zaghouani, Kais Attia, Md. Rafiul Biswas and Fadhl Eryani	6597
<i>Universal NER v2: Towards a Massively Multilingual Named Entity Recognition Benchmark</i>	
Terra Blevins, Stephen Mayhew, Marek Suppa, Hila Gonen, Shachar Mirkin, Vasile Pais, Kaja Dobrovoljc Zor, Voula Giouli, Jun Kevin, Eugene Jang, Eungseo Kim, Jeongyeon Seo, Xenophon Gialis and Yuval Pinter	6609
<i>JobArabi: An Arabic Corpus and Analysis of Job Announcements from Social Media</i>	
Wajdi Zaghouani, Shimaa Amer Ibrahim, Mabrouka Bessghaier and Houda Bouamor	6619
<i>ParaCLEAN: Improving Translation Quality through Systematic Parallel Data Cleaning</i>	
Audrey Mash, Ella Paulina Bohman and Maite Melero	6630
<i>DReUD: Discourse Relations in Universal Dependencies</i>	
Jiří Mírovský and Pavlína Synková	6641
<i>MultiGraSCCo: A Multilingual Anonymization Benchmark with Annotations of Personal Identifiers</i>	
Ibrahim Baroud, Christoph Otto, Vera Czehmann, Christine Hovhannisyan, Lisa Raithel, Sebastian Möller and Roland Roller	6647
<i>Structured Legal Document Generation in India: A Model-Agnostic Wrapper Approach with VidhikDastaavej</i>	
Shubham Kumar Nigam, Deepak Patnaik Balaramamahanthi, Noel Shallum, Kripabandhu Ghosh and Arnab Bhattacharya	6661
<i>PolyglotQL: A Pipeline for Multilingual Text-to-SPARQL Dataset Generation</i>	
Julio Perez, Fabio Barth and Georg Rehm	6674
<i>Building and Annotating a Large Comparable Corpus for Studying Semantic Quantification - Chinese, French, Japanese, Korean</i>	
raoul blin, Jinam Choi, WU qishen, Yuxin Zhang, Soonhee Hwang, Takahiro Morita, Alexander Delaporte, Ilaine Wang and Chang Liu	6685
<i>Towards the Generation and Application of Dynamic Web-Based Visualization of UIMA-based Annotations for Big-Data Corpora with the Help of Unified Dynamic Annotation Visualizer</i>	
Thiemo Dahmann, Julian Schneider, Philipp Stephan, Giuseppe Abrami and Alexander Mehler	6695

<i>The MultiPEYE Text Corpus: Towards a Diverse and Ever-Expanding Multilingual Text Corpus</i>	
Ramunė Kasperė, Anna Bondar, Sergiu Nisioi, Maja Stegenwallner-Schütz, Hanne B. Søndergaard Knudsen, Ana Matić, Eva Pavlinušić Vilus, Dorota Klimek-Jankowska, Chiara Tschirner, Not Battista Soliva, Deborah N. Jakobi, Cui Ding, Dima Abu Romi, Cengiz Acarturk, Matilda Agdler, Anton Marius Alexandru, Mohd Faizan Ansari, Annalisa Arcidiacono, Somayeh Bakhtiari, Elizabete Ausma Velta Barisa, Ana Bautista, Lisa Beinborn, Yevgeni Berzak, Nedeljka Bjelanović, Anna Isabelle Bothmann, Jan Brasser, Caterina Cacioli, Anila Çepani, Ilze Ceple, Adelina Cerpja, Dalí Chirino, Jan Chromý, Alessandro Corona Mendoza, Iria de-Dios-Flores, Nazik Dinçtopal Deniz, Ana Došen, Kristian Elersič, Inmaculada Fajardo, Zigmunds Freibergs, Angelina Ganebnaya, Shan Gao, Jéssica Gomes, Annjo Klungervik Greenall, Alba Haveriku, Miao He, Anamaria Hodiovanu, Yu-Yin Hsu, Amanda Isaksen, Andreia Janeiro, Kristine Jensen de López, Aleksandar Jevremovic, Vojislav Jovanovic, Hanna Kędzierska, Nik Kharlamov, Sara Kosutar, Nelda Kote, Vanja Kovic, Izabela Krejtz, Thyra Krosness, Oleksandra Kuvshynova, Eilam Lavy, Ella Lion, Marta Łockiewicz, Kaidi Lõo, Paula Luegi, Mircea Mihai Marin, Clara Martin, Svitlana Matvieieva, Diane C. Mézière, Xavier Mínguez-López, Valeriia Modina, Jurgita Motiejūnienė, Marie-Luise Müller, Tolgonai Nasipbek kyzy, Jamal Abdul Nasir, Johanne S. K. Nedergård, Ayşegül Özkan, Patrizia Paggio, Marijan Palmović, Maria Christina Panagiotopoulou, Alberto Parola, Helena Pérez, Klaudia Petersen, Anja Podlesek, Eva Pospíšilová, Marta Praulina, Mikuláš Preininger, Loredana Pungă, Diego Rossini, Špela Rot, Habib Sani Yahaya, Irina A. Sekerina, Anne Gabija Skadina, Jordi Solé-Casals, Lonneke van der Plas, Saara M. Varjopuro, Spyridoula Varlokosta, João Veríssimo, Oskari Juhapekka Virtanen, Nemanja Vračar, Mila Vulchanova, Ahmad Mustapha Wali, Peizheng Wu, Nilgün Yücel, Stefan Frank, Nora Hollenstein and Lena Jäger	6706
<i>Sanskrit Travelogue: A Large-Scale Unified and Annotated Corpus of Sanskrit Texts</i>	
Giacomo De Luca, Danilo Croce and Roberto Basili	6722
<i>The Foggia Occupator Corpus: Digitisation, Annotation, and Computational Analysis of an Occupation-Era Newspaper (1945-1946)</i>	
Michele Ciletti	6731
<i>SiDiaC-v.2.0: Sinhala Diachronic Corpus Version 2.0</i>	
Nevidu Jayatilleke, Nisansa de Silva, Uthpala Nimanthi Sooriya-Arachchi, Gagani Kasundhi Kulathilaka, Azra Safrullah and Johan Nevin Sofalas	6740
<i>ShAnEL-2: A Multilingual Benchmarking Dataset for Short-Answer Language Learning Exercises</i>	
Jasper Degraeuwe and Thomas Moerman	6764
<i>The Swedish Parliamentary Motions Corpus 1867-2024</i>	
Robert Borges, Fredrik Mohammadi Norén, Lotta Åberg Brorsson, Väinö Yrjänäinen, Hanna Bäck, Robert Klemmensen and Måns Magnusson	6772
<i>The Swedish Benchmark of Linguistic Minimal Pairs</i>	
Johan Sjons, Fredrik Heinat and Murathan Kurfali	6783
<i>Exploring the Transfer of Irony Explanation Generation from English to Dutch</i>	
Aaron Maladry, Els Lefever, Cynthia Van Hee and Veronique Hoste	6795
<i>DIDECO: An Annotated Dataset for Intent Detection in Digital Communications</i>	
Senaid Popovic, Damien Riquet, Maxime Meyer, Fabien Lauer and Yannick Parmentier	6808

<i>GUMBridge: A Corpus for Varieties of Bridging Anaphora</i> Lauren Levine and Amir Zeldes	6823
<i>Beyond Transcripts: Iterative Peer-Editing with Audio Unlocks High-Quality Human Summaries of Conversational Speech</i> Kaavya Chaparala, Thomas Thebaud, Jesus Villalba Lopez, Laureano Moro-Velazquez, Peter Viechnicki and Najim Dehak	6838
<i>SEEM-CZ: Annotation and Classification of Epistemic Markers in Czech</i> Barbora Štěpánková, Michal Novák, Tomáš Musil and Lucie Polakova	6856
<i>When Words Don't Mean What They Say: Figurative Understanding in Bengali Idioms</i> Adib Sakhawat, Shamim Ara Parveen, Md Ruhul Amin, Tahera Khatun, Shamim Al Mahmud and Md Saiful Islam	6870
<i>Human vs LLM in Conversational Repair Annotation: A New Resource and Comparative Study</i> Anh Ngo, Nicolas Rollet, Catherine Pelachaud and Chloé Clavel	6880
<i>GPT-NL Public Corpus: A Permissively Licensed, Dutch-First Dataset for LLM Pre-training</i> Jesse J. Van Oort, Frank Brinkkemper, Erik de Graaf, Bram Vanroy and Saskia Lensink	6893
<i>Estonian WinoGrande Dataset: Comparative Analysis of LLM Performance on Human and Machine Translation</i> Marii Ojastu, Hele-Andra Kuulmets, Aleksei Dorkin, Marika Borovikova, Dage Särg and Kairit Sirts	6904
<i>GENIUS Keylog Corpus - a German High School Student Corpus with Keystroke Logging Data</i> Nils-Jonathan Schaller, Thorben Jansen, Lars Höft, Hannah Pünjer and Andrea Horbach	6919
<i>OTA-BOUN: A Historical Turkish Dependency Treebank</i> Tarik Emre Tıraş, Nureddin Cüneyd Ünal, Ada Cengiz, Ece Yurtseven, Esma F. Bilgin Taşdemir and Saziye Betül Özates	6929
<i>TCMPHal: A Large-scale Dataset for Hallucination Detection in Traditional Chinese Medicine Pharmacy</i> Nijia Han, Zimu Wang, Ziwen Xie, Wei Wang, Jia Meng, John Moraros and Shuihua Wang	6939
<i>AraREQ: A Dataset and End-to-End System for Conflict Detection and Resolution in Software Requirements</i> Tymaa Hasanain Hammouda, Alaa Aljabari, Nagham Fahim Hamad and Mustafa Jarrar	6949
<i>MAD: A Corpus of Multilingual Argumentative Deliberation</i> Eimear Maguire, Ella Schad, Jacky Visser, Chris Reed and John Lawrence	6962
<i>Infox-QC: A Quebec-Focused French Corpus for Misinformation Detection and AI Robustness Assessment</i> Moetaz Doghmane, Hazem Amamou, Thiziri Sefsaf, Alan Davoust and Anderson Raymundo Avila	6979

<i>unarXive 2024: A Large-Scale Scientific Corpus for Citation-Aware Retrieval and Generation</i> Ines Besrour and Michael Färber	6990
<i>EPIC-EuroParl-UdS: Information-Theoretic Perspectives on Translation and Interpreting</i> Maria Kunilovskaya and Christina Polkläsener	6998
<i>FeedFetcher: A Resilient Web Feed Downloader for Corpus Construction</i> Ondřej Herman, Jan Kraus and Vit Suchomel	7014
<i>Human-in-the-Loop Mass Transcription and Ground Truth Annotation for Challenging Historical Documents</i> Norbert Fischer and Frank Puppe	7023
<i>CoMMA, a Large-scale Corpus of Multilingual Medieval Archives</i> Thibault Clérice, Simon Gabay, Malamatenia Vlachou-Efstathiou, Ariane Pinche and Benoît Sagot	7034
<i>Conversion of the Clark Hall Dictionary of Old English to TEI with RDF: An End-to-end Pipeline for Lexicographic Resource Retrodigitization</i> Sergei Stoliarov, Maxim Ionov, Fahad Khan, Marina Buzzoni and Francesca Frontini	7046
<i>AMORES: A Spanish Language Resource for an Extended Set of Moral Foundations</i> Oscar Araque, Daniel Molina, Anny D. Alvarez Nogales and Carlos A. Iglesias	7056
<i>The Moralization Corpus: Frame-Based Annotation and Analysis of Moralizing Speech Acts across Diverse Text Genres</i> Maria Becker, Mirko Sommer, Lars Tapken, Yi Wan Teh and Bruno Brocai	7069
<i>Targum — a Multilingual New Testament Translation Corpus</i> Maciej Rapacz and Aleksander Smywiński-Pohl	7092
<i>Trigger Warnings Are Grounded in a Shared Vocabulary: A Corpus Analysis with User-Generated Labels</i> Sebastian Heineking, Matti Wiegmann, Magdalena Wolska, Benno Stein and Martin Potthast	7106
<i>ENEIDE: A High Quality Silver Standard Dataset for Named Entity Recognition and Linking in Historical Italian</i> Cristian Santini, Sebastian Barzaghi, Paolo Sernani, Emanuele Frontoni, Laura Melosi and Mehwish Alam	7126
<i>YoNER: A New Yorùbá Multi-domain Named Entity Recognition Dataset</i> Peace Busola Falola, Jesujoba Alabi, Solomon O. Akinola, Folashade T. Ogunajo, Em- manuel Oluwadunsin Alabi and David Ifeoluwa Adelani	7137
<i>Linking Rationale to Decision on Internet Standards: A Retrieval-Based Approach Using Syn- thetic Data</i> Jie Bian and Michael Welzl	7149
<i>The GELATO Dataset for Legislative NER</i> Matthew Flynn, Timothy Obiso and Sam Newman	7163

<i>Controllable Sentence Simplification in Italian: Fine-Tuning Large Language Models on Automatically Generated Resources</i>	
Michele Papucci, Giulia Venturi and Felice Dell’Orletta	7178
<i>Evaluating LLM-based Text Simplification for German: Effects on Post-Editing Effort, Quality Ratings, and User Comprehension</i>	
Luisa Carrer, Andreas Säuberli, Martin Kappus, Lukas Fischer and Sarah Ebling	7192
<i>Reading Time in the Wild: An Assessment of Readability Predictors Based on Naturally-Observed Reading Times</i>	
Sijbren van Vaals, Rik van Noord and Malvina Nissim	7209
<i>Document-Level Text Simplification in Estonian Using Large Language Models</i>	
Meeri-Ly Muru and Eduard Barbu	7225
<i>A Human-in/on-the-Loop Framework for Accessible Text Generation</i>	
Lourdes Moreno and Paloma Martínez	7236
<i>Automatic Analysis of Collaboration through Human Conversational Data Resources: A Review</i>	
Yi Yu, Maria Boritchev and Chloé Clavel	7248
<i>Benchmarking Arabic Authorship Attribution and Style Transfer with Large Language Models</i>	
Injy Hamed, Bashar Alhafni, Nizar Habash and Thamar Solorio	7262
<i>ADHD-Lang: A Large-Scale Social Media Dataset for Verbal Behavior and Digital Phenotyping in Adult ADHD</i>	
Daniel Wiechmann, Elma Kerz, Edward Kempa and Yu Qiao	7279
<i>SynBullying: A Multi-LLM Synthetic Conversational Dataset for Cyberbullying Detection</i>	
arefeh kazemi, Hamza Qadeer, Joachim Wagner, hossein hosseini, Sri Balaaji Natarajan Kalaivendan and Brian Davis	7292
<i>The Multilingual Euphemism Benchmark: Datasets and Baselines for Pragmatic Language Understanding</i>	
Whitney Poh, Julia Sammartino, Jasper Andrew, Witold Kieraś, Natalia Zawadzka-Palucktau, Iryna Dilai, Libby Barak, Jing Peng and Anna Feldman	7307
<i>Advancing Retrieval-Augmented Generation for Persian: Development of Language Models, Comprehensive Benchmarks, and Best Practices for Optimization</i>	
Sara Bourbour Hosseinbeigi, Mohammad Hossein Shalchian, Sina Asghari, Mohammad Ali Seif Kashani and Mohammad Amin Abbasi	7320
<i>Corpus and Baselines for Distinguishing Authentic, AI-Generated, and AI-Enhanced Resumes</i>	
Andrea Loizidou, Anshu Kiran Sharma, Adrian Esquivel, Mark A. Finlayson and Mustafa Ocal	7331
<i>Mute Cods: A Multilingual Telegram Dataset with Benchmark Models for Conspiracy Theory Detection</i>	
Katarina Laken, Erik Bran Marino, Paloma Piot, Davide Bassi, Søren Kirkegaard Fomsgaard, Michele Joshua Maggini, Renata Vieira, Marcos Garcia and Sara Tonelli	7345

<i>Push and Pull: Training Sentence Encoders with Contrastive Losses for Distance-Based Multi-Label Text Classification</i>	
Jens Van Nooten and Andriy Kosar	7359
<i>PRIVaThe: An Annotated Dataset of Multi-Objectives Web Search Sessions</i>	
Claire Ibarboure, Ludovic Tanguy, Franck Amadieu and Josiane Mothe	7380
<i>Towards Safer Calls for Everyone: Designing a Benchmark Dataset for Evaluating Voice Phishing Detection Models</i>	
joeun kang, Gyuri Choi, Chanhhyuk Yoon, Yongbin Jeong, Younggyun Hahm, Shea Husband and Hansaem Kim	7391
<i>Learning Long-Document Embeddings via Chunk–Context Entailment</i>	
Waheed Ahmed Abro, Naïm Es-Sebbani and Zied Bouraoui	7405
<i>Scientific Article Section Classification (SASC) Dataset</i>	
Nicolau Duran-Silva, Julian Moreno-Schneider, César Parra-Rojas and Georg Rehm	7415
<i>JMTEB and JMTEB-lite: Japanese Massive Text Embedding Benchmark and Its Lightweight Version</i>	
Shengzhe Li, Masaya Ohagi, Ryokan Ri, Akihiko Fukuchi, Tomohide Shibata and Daisuke Kawahara	7423
<i>Construction of a Japanese RAG Benchmark Using Synthetic Documents on Non-existent Entities and Events</i>	
Shengzhe Li, Masaya Ohagi, Hayato Tsukagoshi, Akihiko Fukuchi, Tomohide Shibata and Daisuke Kawahara	7435
<i>C4: A Multilingual Benchmark for Retrieval-Augmented Generation Based on the Catechism of the Catholic Church and Its Compendium</i>	
Pius von Däniken, Mark Cieliebak and Jan Deriu	7446
<i>Contrastively Pre-trained Event Embeddings with Schema-free LLM Annotations</i>	
Frank Mtumbuka and Steven Schockaert	7457
<i>A Dataset of Psychiatric Hospital Notes with Temporal Information Annotations</i>	
Timothy A. Miller, Gaby Dinh, David Harris, WonJin Yoon, Spencer Thomas, Boyu Ren, MEIHUA HALL and Guergana Savova	7479
<i>Format Matters: A Critical Evaluation of Output Formats for Prompting LLMs in SLU and NER</i>	
Pierre Lepagnol, Sahar Ghannay, Thomas Gerald, Christophe Servan and Sophie Rosset	7485
<i>Identifying Imaging Follow-Up in Radiology Reports: A Comparative Analysis of Traditional ML and LLM Approaches</i>	
Namu Park, Giridhar Kaushik Ramachandran, Kevin Lybarger, Fei Xia, Özlem Uzuner, Martin Gunn and Meliha Yetisgen	7498
<i>Efficient Topic Extraction via Graph-Based Labeling: A Lightweight Alternative to Deep Models</i>	
SALMA MEKAOUI, Hiba Sofyan, Imane Benchrif, Imane Amaaz, Ilham Chaker, Arsalane Zarghili and Nikola S. Nikolov	7511

<i>From Noise to Signal: When Outliers Seed New Topics</i>	
Evangelia Zve, Gauvain Bourgne, Benjamin Icard and Jean-Gabriel Ganascia	7523
<i>Explore Political Discourse with Transformers. Emergent Paradigmatic and Syntagmatic Representations.</i>	
Laurent Vanni and Damon Mayaffre	7534
<i>The Growing Gains and Pains of Iterative Web Corpora Crawling: Insights from South Slavic CLASSLA-web 2.0 Corpora</i>	
Taja Kuzman Pungeršek, Peter Rupnik, Vit Suchomel and Nikola Ljubešić	7545
<i>MaritimEmails: A Synthetic Dataset for Maritime Chartering Correspondence</i>	
Kevin Bruendler and Simon Clematide	7556
<i>eSciBench: An Extensible Scientific PDF Extraction Benchmark</i>	
Noah Tremblay Taillon and Phillippe Langlais	7568
<i>Vrittanta-AS: Dataset Development and Benchmarking for Event Trigger Detection and Classification in Assamese</i>	
Chaitanya Kirti, Dhruvajyoti Pathak, Ashish Anand and Prithwijit Guha	7581
<i>From Facts to Hypotheses: Joint Detection of Biomedical Relations and Epistemic Commitment Using LLMs</i>	
Aleksandra Gabryszak, Phuc Tran Truong, Arne Binder, Nikola Milosevic, Felix-Sebastian Keese, Astrid Rheinländer and Philippe Thomas	7592
<i>SciLaD: A Large-Scale, Transparent, Reproducible Dataset for Natural Scientific Language Processing</i>	
Luca Foppiano, Sotaro Takeshita, Pedro Ortiz Suarez, Ekaterina Borisova, Raia Abu Ahmad, Malte Ostendorff, Fabio Barth, Julian Moreno-Schneider and Georg Rehm	7606
<i>CausalSense: Leveraging Common Sense Knowledge and LLMs for Joint Event Extraction and Relation Classification</i>	
Youssra REBBOUD, Pasquale Lisena and Raphael Troncy	7619
<i>Large Language Models Are Good Term Extractors: A Systematic Evaluation</i>	
Ayla Rigouts Terryn	7631
<i>A Large-Scale Dataset for Linking-Based Geocoding</i>	
Hibiki Nakatani, Yuichiro Yasui, Ryosuke Wakamoto, Masayuki Ishii, Tetsuhisa Suizu, Hiroki Ouchi and Taro Watanabe	7644
<i>FiNERVINER: Fine-grained Named Entity Recognition for Vulnerable Languages of India's North Eastern Region</i>	
Prachuryya Kaushik and Ashish Anand	7655
<i>APTFiNER: Annotation Preserving Translation for Fine-grained Named Entity Recognition</i>	
Prachuryya Kaushik, Adittya Gupta, Ajanta Maurya, Gautam Sharma, V. V. Saradhi and Ashish Anand	7668
<i>RelEx-PT: A Portuguese Sentence-Level Relation Extraction Dataset</i>	
Tomás Pinto, Catarina Silva and Hugo Goncalo Oliveira	7681

<i>Benchmarking Portuguese Open Information Extraction</i> Gabriel Silva, Mário Rodrigues, António Teixeira and Marlene Amorim	7692
<i>A Scalable Pipeline for Novelty Detection in Skill Extraction Using Large Language Models</i> Gian Seifert and Simon Clematide	7701
<i>Do LLMs Judge Distantly Supervised Named Entity Labels Well? Constructing the JudgeWEL Dataset</i> Alistair Plum, Laura Maria Bernardy and Tharindu Ranasinghe	7707
<i>From Articles to Premises: Building PrimeFacts, an Extraction Methodology and Resource for Fact-Checking Evidence</i> Premtim Sahitaj, Jawan Kolanowski, Ariana Sahitaj, Veronika Solopova, Max Upravitelev, Daniel Röder, Iffat Maab, Junichi Yamagishi, Sebastian Möller and Vera Schmitt	7719
<i>EpiGator: An Event-based Surveillance System for Infectious Disease Outbreaks</i> Yiheng Wu, Jue Hou, Trancasanchai Sathianpong, Lidia Pivovarova and Roman Yangarber	7732
<i>Relation Extraction across Entire Books to Reconstruct Community Networks: The AffilKG Datasets</i> Erica Cai, Sean Mcquade, Kevin Young and Brendan O'Connor	7744
<i>Vrittanta-EN: A Benchmark Dataset for Event Trigger Detection and Classification Advancing Event Understanding in English Narrative Discourse</i> Chaitanya Kirti, Ashish Anand and Prithwijit Guha	7755
<i>MUC-4 Revisited: Document-level Event Analysis beyond Span-based Arguments</i> Helene Bøsei Olsen, Erik Velldal and Lilja Øvrelid	7766
<i>Historical Medical Knowledge Graphs and Ontologies from the Medical History of British India Corpus (1850-1950)</i> Mehrdad Almasi and Tugce Karatas	7781
<i>Graph-TempCZ: A Graph Representation of Software Mentions for Predicting Software Usage in Scientific Publications</i> Congfeng Cao, Pengyu Zhang and Jelke Bloem	7791
<i>Automatic Suggestions Help Extending Eventive Ontology: A Case Study on SynSemClass</i> Jana Straková, Eva Fučíková, Zdenka Uresova and Jan Hajič	7804
<i>JPPB: Automatic Construction of a Soft-Labeled Japanese Patient Phrase Bank for Symptom Normalization</i> Tomohiro Nishiyama, Mana Kuramoto, Shoko Wakamiya and Eiji ARAMAKI	7816
<i>How I Met Your Snowclone: Unsupervised Discovery of Snowclone Patterns in Large Datasets</i> Julien Bezançon, Gaël Lejeune and Marceau Hernandez	7829
<i>HOME-KGQA: A Benchmark Dataset for Multimodal Knowledge Graph Question Answering on Household Daily Activities</i> Shusaku Egami, Aoi Ohta, Tomoki Tsujimura, Masaki Asada, Tatsuya Ishigaki, Ken Fukuda, Masahiro Hamasaki and Hiroya Takamura	7845

<i>Extending the Semantic Layer of the Compl-it Italian Lexicon: Traits, Semantic Types, and Definitions</i>	
Emiliano Giovannetti, Andrea Bellandi, Simone Marchi and Mafalda Papini	7857
<i>Integrating Knowledge Graph with Large Language Models for Multi-hop Question Generation</i>	
Yllias Chali and Al Hasib Mahamud	7867
<i>LocalGovPL: A Corpus of Speaker-Attributed Polish Local Government Transcripts</i>	
Dariusz Czerski and Maciej Ogrodniczuk	7883
<i>Amharic DBpedia Chapter: A Knowledge Graph for a Low-Resource Language</i>	
HLZkiel Mitiku Alemayehu, Tilahun Abedissa Taffa, Meti Adane Bayissa, Andargachew Asfaw Zewge, Hamada Zahera, Ricardo Usbeck and Axel-Cyrille Ngonga Ngomo	7894
<i>Cygnet: Refactoring the Open Multilingual Wordnet</i>	
Rowan Hall Maudslay and Francis Bond	7905
<i>Masrad: Arabic Terminology Management Corpora with Semi-Automatic Construction</i>	
Mahdi Nasser, Laura Sayah and Fadi Zaraket	7918
<i>SentiMalti: A Maltese Sentiment Analysis Dataset and Models</i>	
Ian Caruana, Matthew Vella, Fabio Zammit, Kurt Micallef and Claudia Borg	7927
<i>Multilingual Structured Sentiment Analysis for Environmental Sustainability</i>	
Muhammad Okky Ibrohim, Tommaso Caselli, Cristina Bosco and Valerio Basile	7937
<i>LLM-as-an-Annotator: Training Lightweight Models with LLM-Annotated Examples for Aspect Sentiment Tuple Prediction</i>	
Nils Constantin Hellwig, Jakob Fehle, Udo Kruschwitz and Christian Wolff	7955
<i>Extending Czech Aspect-Based Sentiment Analysis with Opinion Terms: Dataset and LLM Benchmarks</i>	
Jakub Šmíd, Pavel Priban and Pavel Kral	7973
<i>AnnoABSA: A Web-Based Annotation Tool for Aspect-Based Sentiment Analysis with Retrieval-Augmented Suggestions</i>	
Nils Constantin Hellwig, Jakob Fehle, Udo Kruschwitz and Christian Wolff	7985
<i>Zero-Shot to Full-Resource: Cross-lingual Transfer Strategies for Aspect-Based Sentiment Analysis</i>	
Jakob Fehle, Nils Constantin Hellwig, Udo Kruschwitz and Christian Wolff	7999
<i>LoveHate: Stance Detection and Generation for Multiple Topics in User-generated Comments in Russian and English</i>	
Natalia Evgrafova, Veronique Hoste and Els Lefever	8014
<i>From Trial by Fire to Sleep like a Baby: A Lexicon of Anxiety Associations for 20K English Multi-Word Expressions</i>	
Saif M. Mohammad	8026
<i>Entity-Level Sentiment Analysis with Sentence Relevance Detection</i>	
Egil Rønningstad, Roman Klinger, Lilja Øvrelid and Erik Velldal	8040

<i>Enhancing Multi-Label Emotion Analysis and Corresponding Intensities for Ethiopian Languages</i> Tadesse Destaw Belay, Dawit Ketema Gete, Abinew Ali Ayele, Olga Kolesnikova, Iqra Ameer, Grigori Sidorov and Seid Muhie Yimam	8056
<i>A Japanese Dataset for Aspect-based Sentiment Polarity Classification and Emotion Intensity Estimation</i> Kentaro Hanafusa, Kota Manabe, Yuki Maeda, Daisuke Maekawa, Tomoyuki Kajiwara, Hideaki Hayashi, Yuta Nakashima and Hajime Nagahara	8076
<i>Assessing the Persuasive Effect of AI-Generated Image Support of Arguments</i> Mackwyn Quadras, Manfred Stede and Henning Wachsmuth	8085
<i>CIARAM: Class Imbalance Aware Generative Framework for Relational Argument Mining</i> Nilmadhab Das, Sayan Pal, V. V. Saradhi and Ashish Anand	8096
<i>Surfacing Subtle Stereotypes: A Multilingual, Debate-Oriented Evaluation of Modern LLMs</i> Muhammed Yahia Gaffar Saeed, Muhammad Abdul-Mageed and Shady Shehata ...	8106
<i>Prompt-Based Stance Control in German: An Evaluation of LLMs for Experimental Research on Attitude Change</i> Florian Omiecienski, Cornelia Sindermann and Agnieszka Falenska	8122
<i>CoSt-BR: A Language Resource for Conversational Stance Detection</i> Felipe Penhorate Carvalho da Fonseca, Ivandre Paraboni and Luciano Antônio Digiampietri	8141
<i>Less Is More? The Role of Demographic Author Information in Emotion Classification of Ambiguous Text</i> Sabine Weber, Lynn Greschner and Roman Klinger	8147
<i>Big Five Personality Prediction through Emotion-Conditioned Representations and Learnable Psycholinguistic Mapping</i> Lorenzo Zangari, Antonin Schnyder and Davide Picca	8162
<i>SENSEI-ASG: A Challenging Dataset for Argument Summary Graph Parsing</i> Jonathan Clayton, Marco Damonte and Robert Gaizauskas	8174
<i>Categorical Emotions or Appraisals - Which Emotion Model Explains Argument Convincingness Better?</i> Lynn Greschner, Meike Bauer, Sabine Weber and Roman Klinger	8190
<i>Creation of the Estonian Subjectivity Dataset: Assessing the Degree of Subjectivity on a Scale</i> Karl Gustav Gailit, Kadri Muischnek and Kairit Sirts	8204
<i>Mitigating Misinterpretation in Policy Documents through Automated Language Understanding</i> Momojit Biswas, Anka Chandahas Tummepalli and Preethu Rose Anish	8217
<i>Sovereign AI-based Public Services Are Viable and Affordable</i> António Branco, Luis M. S. Gomes, Rodrigo Santos, Eduardo Santos, João Ricardo Silva, Nuno Marques and Madalena Rodrigues	8235
<i>A Typology of Synthetic Datasets for Dialogue Processing in Clinical Contexts</i> Steven Bedrick, A. Seza Doğruöz and Sergiu Nisioi	8246

<i>Text+: A National Hub Including Legacy Language Data</i>	
Florian Barth, Christoph Draxler, Jennifer Ecker, Stefan Fischer, Philippe Genêt, Alina Hemmer, Timm Lehmberg, Thorsten Trippel, Andreas Witt, Arden Zimmermann and Claus Zinn	8264
<i>Can NLP Tackle Hate Speech in the Real World? Stakeholder-Informed Feedback and Survey on Counterspeech</i>	
Tanvi Dinkar, Aiqi Jiang, Simona Frenda, Poppy Gerrard-Abbott, Nancie A. Gunson, Gavin Abercrombie and Ioannis Konstas	8276
<i>Towards Complex Debate Understanding: Predicting Claim Impact Scores through the Modelling of Claim Interactions</i>	
Maxime Brouat, Mihai Surdeanu, Srdjan Vesic and Eduardo Blanco	8291
<i>Is There Anything More Deceptive than an Obvious Fact? Investigating Implicitness in User-Generated Argumentative Text</i>	
Ekaterina Sviridova, Elena Cabrio and Serena Villata	8303
<i>Best-Worst Scaling of Hype in Biomedical Research: Building an Intensity Lexicon of Promotional Adjectives</i>	
Neil Millar, Dipesh Satav, Bojan Batalo, Erica K. Shimomoto and Ryosuke L. Ohniwa	8317
<i>Trust Me, I Can Convince You: The Contextualized Argument Appraisal Framework and the ContArgA Corpus</i>	
Lynn Greschner, Sabine Weber and Roman Klinger	8327
<i>Towards Clinical Applications of NLP: Detecting Emotion Regulation via Emotional Categories and Expression Modes in French Transcriptions</i>	
Salome Klein, Amalia Todirascu and Hélène Vassiliadou	8347
<i>R.U.Psycho? A Framework for Robust Unified Psychometric Testing of Language Models</i>	
Julian Schelb, Orr Borin, David Garcia and Andreas Spitz	8365
<i>Code-switching as a Bias Indicator in LLMs: "the Consequences Are Not the Same Para Nosotros"</i>	
Fanny Ducel, Aurélie Névél, Vidit Khazanchi, Loïc Leclerc, Arthur Pedrini, Léa Bouchet, Benjamin Caissial and Karen Fort	8387
<i>Exploration of How Hate Is Framed on Social Media</i>	
Rakshitha Rao Ailneni and Sanda Harabagiu	8400
<i>Are Social Biases in LLMs Consistent across Generative Tasks? A Case Study for Basque</i>	
Muitze Zulaika, Xabier Saralegi, Julia Shershneva, Lia Gonzalez and Arkaitz Fullaondo	8415
<i>Fine-grained Narrative Classification in Biased News Articles</i>	
Zeba Afroz, Harsh Vardhan, pawan bhakuni, Aanchal Punia, Rajdeep Kumar and Md. Shad Akhtar	8431
<i>A Shoal of Voices: Parallel Read Speech from Professional Swedish Narrators</i>	
Christina Tännander, Jim O'Regan and Jens Edlund	8446

<i>Deep Learning-Based Multi-Aspect Pronunciation Assessment for Individuals with Down Syndrome</i>	
David Fernández-García, César González-Ferreras, Valentín Cardeñoso-Payo and Mario Corrales-Astorgano	8455
<i>WikIPA: Integrating WikiPron and Lingua Libre for Multilingual IPA Transcription</i>	
Pierluigi Cassotti, Jacob Lee Suchardt and Domenico De Cristofaro	8465
<i>How Pragmatics Shape Articulation: A Computational Case Study in STEM ASL Discourse</i>	
Saki Imai, Lee Kezar, Laurel Aichler, Mert Inan, Erin Walker, Alicia Wooten, Lorna Cobban Quandt and Malihe Alikhani	8476
<i>Setting the Stage for Disfluency: Implications of Contextual Task Framing Effects for the Design of Listening Tasks</i>	
Ambika Kirkland and Jens Edlund	8491
<i>ACADData: Parallel Dataset of Academic Data for Machine Translation</i>	
Iñaki Lacunza, Javier Garcia Gilabert, Francesca De Luca Fornaciari, Javier Aula-Blasco, Aitor Gonzalez-Agirre, Maite Melero and Marta Villegas	8498
<i>A Single Model Ensemble Framework for Neural Machine Translation Using Pivot Translation</i>	
Seokjin Oh, Keonwoong Noh and Woohwan Jung	8520
<i>Gender Disambiguation in Machine Translation: Diagnostic Evaluation in Decoder-Only Architectures</i>	
Chiara Manna, Hosein Mohebbi, Afra Alishahi, Frederic Blain and Eva Vanmassenhove	8535
<i>Building a One-Million-Pair Bokmål–Nynorsk Translation Corpus: A Quality-First Harvesting and Cleaning Pipeline</i>	
Per E. Kummervold, Thea Tollersrud and Angelina Zanardi	8551
<i>New Trends for Modern Machine Translation with Large Reasoning Models</i>	
Sinuo Liu, Chenyang Lyu, Minghao Wu, Zifu Shang, Longyue Wang, Weihua Luo and Kaifu Zhang	8556
<i>MaitH 1.0: A Parallel Corpus and Baseline for Low-Resource Maithili-Hindi Translation</i>	
Kamanksha Prasad Dubey, Chandresh Maurya and Kumar Padmanabh	8567
<i>NRD: A Hybrid Disentanglement Framework for Mitigating Interference in Multilingual Machine Translation</i>	
Jiarui Zhang and Yifan Deng	8577
<i>Linguistic and Demographic Factors in an Online Free Translation Task</i>	
Tyler Lee, Irina Stenger and Tania Avgustinova	8587
<i>Biases in Translation: Assessing Opinion Distortion in Machine Translated Texts</i>	
Nazanin Shafiabadi and François Yvon	8596
<i>When Translations Surprise: Human Awareness of Predictability in Translations</i>	
Cristian García-Romero, Miquel Esplà-Gomis and Felipe Sanchez-Martinez	8615
<i>Bidirectional Chinese and English Passive Sentences Dataset for Machine Translation</i>	
Xinyue Ma, Pol Pastells, Mireia Farrus and Mariona Taule	8628

<i>CoTERM: A Consistency-Oriented Term Metric for MT System Evaluation</i> Amir Hazem and Kyo Kageura	8639
<i>SiniticMTErrors: A Machine Translation Dataset with Error Annotations for Sinitic Languages</i> Hannah Liu, Junghyun Min, Annie En-Shiun Lee, Ethan Yue Heng Cheung, Shou-Yi Hung, Elsie Chan, Shiyao Qian, RUNTONG LIANG, Kimlan Huynh, Wing Yu Yip, York Hay Ng, Tsz Fung Yau, Ka Ieng Charlotte Lo, You-Wei Wu and Richard Tzong-Han Tsai	8662
<i>Ancient Greek to Modern Greek Machine Translation: A Novel Benchmark and Fine-Tuning Experiments on LLMs and NMT Models</i> Spyridon Mavromatis, Sokratis Sofianopoulos, Prokopis Prokopidis and Maria Giagkou	8685
<i>Linguistic Knowledge-Infused Fine-Tuning for Mitigating Gender Bias in Machine Translation</i> Luis Ernesto Garcia Estrada, Audrey Mash, Carlos Escolano, Maite Melero and Christine Basta	8699
<i>What Triggers My Model? Contrastive Explanations Inform Gender Choices by Translation Mod- els</i> Janiča Hackenbuchner	8710
<i>ViKhoMT: A Vietnamese–K’Ho Neural Machine Translation Dataset and Evaluation for Commu- nity Health Communication</i> Tram Truong, Vinh Nguyen, Dang Van Thin and Ngan Nguyen	8725
<i>Hindsight Quality Prediction Experiments in Multi-Candidate Human-Post-Edited Machine Trans- lation</i> Malik Marmonier, Benoît Sagot and Rachel Bawden	8740
<i>PETra: A Multilingual Corpus of Pragmatic Explicitation in Translation</i> Doreen Osmelak, Koel Dutta Chowdhury, Uliana Sentsova, Cristina España-Bonet and Josef van Genabith	8756
<i>A Dataset for Probing Translationese Preferences in English-to-Swedish Translation</i> Jenny Kunz, Anja Jarochoenko and Marcel Bollmann	8767
<i>STAR-IL: A Dataset for Style-Aware Machine Translation of Product Reviews in Indian Lan- guages</i> Ketaki Shetye, Dipti Misra Sharma and Parameswari Krishnamurthy	8780
<i>Cultural and Knowledge Biases in LLMs through the Lens of Entity-Aware Machine Translation</i> Lu Xu, Luca Moroni and Roberto Navigli	8794
<i>Referenceless Evaluation of Machine Translation Models by Ranking Performance in Romanian to English Translate-train Settings</i> Mihail Feraru, Alexandra Diaconu and Bogdan Dumitru Alexe	8813
<i>Every Word Presented in Context: Syntactic Coverage as Objective for Low-Resource Machine Translation with Large Language Models</i> Samuel Frontull and Thomas Ströhle	8824

<i>Multilingual KokoroChat: A Multi-LLM Ensemble Translation Method for Creating a Multilingual Counseling Dialogue Dataset</i>	
Ryoma Suzuki, Zhiyang Qi and Michimasa Inaba	8838
<i>NepTam: A Nepali-Tamang Parallel Corpus and Baseline Machine Translation Experiments</i>	
Rupak Raj Ghimire, Bipesh Subedi, Balaram Prasain, Prakash Poudyal, Praveen Acharya, Nischal Karki, Rupak Tiwari, Rishikesh Kumar Sharma, Jenny Poudel and Bal Krishna Bal	8850
<i>Scoring the Translation: On Target Automatic Keyword-Based Evaluation of Machine Translation in the Sports Domain</i>	
Steinþór Steingrímsson and Einar Sigurdsson	8862
<i>Towards Improving Multimodal Machine Translation with LLMs: A Focus on Indic Languages</i>	
Amulya Ratna Dash, Chirag Wadhwa and Yashvardhan Sharma	8872
<i>Parallel Sentence Filtering for Low-Resource Language Pairs: A Case Study for Upper Sorbian, German, and Czech</i>	
Ruiyang Jiang, Shu Okabe and Alexander Fraser	8883
<i>OpenSubtitles2024: A Massively Parallel Dataset of Movie Subtitles for MT Development and Evaluation</i>	
Joerg Tiedemann and Hengyu Luo	8897
<i>CREST: Universal Safety Guardrails through Cluster-Guided Cross-Lingual Transfer</i>	
Lavish Bansal and Naman Mishra	8908
<i>Semantic Alignment across Ancient Egyptian Language Stages via Normalization-Aware Multi-task Learning</i>	
He Huang	8923
<i>Conditioning LLMs to Generate Code-Switched Text</i>	
Maite Heredia, Gorka Labaka, Jeremy Barnes and Aitor Soroa	8937
<i>Are the LLMs Capable of Maintaining at Least the Language Genus?</i>	
Sandra Mitrović, David Kletz, Ljiljana Dolamic and Fabio Rinaldi	8954
<i>Gender Bias in MT for a Genderless Language: New Benchmarks for Basque</i>	
Amaia Murillo, Olatz Perez-de-Viñaspre and Naiara Perez	8971
<i>Optimizing Multilingual LLMs via Federated Learning: A Study of Client Language Composition</i>	
Aleix Sant, Jordi Luque and Carlos Escolano	8985
<i>Multilingual Target-Stance Extraction</i>	
Ethan Leigh Mines and Bonnie J. Dorr	8997
<i>MUNICHus: MUltilingual News Image Captioning Benchmark</i>	
Yuji Chen, Alistair Plum, Hansi Hettiarachchi, Diptesh Kanojia, Saroj Basnet, Marcos Zampieri and Tharindu Ranasinghe	9008
<i>GlossMATE: Multi-Agent Translator Explanations for Glosses</i>	
Changbing Yang, Patrick Littell, Gabriel Bernier-Colborne, Yanfei Lu and Mengzhe Geng	9018

<i>Diagnosing Translated Benchmarks: An Automated Quality Assurance Study of the EU20 Benchmark Suite</i>	
Klaudia Thellmann, Bernhard Stadler and Michael Färber	9030
<i>Resource-Lean Lexicon Induction for German Dialects</i>	
Robert Litschko, Barbara Plank and Diego Frassinelli	9044
<i>FENCE: A Financial and Multimodal Jailbreak Detection Dataset</i>	
Mirae Kim, Seonghun Jeong and Youngjun Kwak	9051
<i>Evaluating Multimodal Large Language Models on Vertically Written Japanese Text</i>	
Keito Sasagawa, Shuhei Kurita and Daisuke Kawahara	9065
<i>ProMQA-Assembly: Multimodal Procedural QA Dataset on Assembly</i>	
Kimihiro Hasegawa, Wiradee Imrattana-trai, Masaki Asada, Susan E. Holm, Yuran Wang, Xuanang Zhou, Ken Fukuda and Teruko Mitamura	9082
<i>K-MIND: Korean Multimodal Interaction Data for Dyadic Conversation Analysis</i>	
Jae Hee Yang, Yuha Shin, Saim Shin, Je Woo Kim and Jin Yea Jang	9105
<i>Do Multimodal LLMs Understand Order? Measuring the Fragility of Multimodal Reasoning under Input Order Perturbations</i>	
Sheng-Lun Wei, Yu-Ling Liao, Hen-Hsen Huang and Hsin-Hsi Chen	9118
<i>Early Fusion with Contrastive Learning: A Lightweight Alternative for Multi-modal Classification</i>	
Felix Wernlein, Abhik Jana and Sandipan Sikdar	9129
<i>Multimodal Entrainment and Feedback in Online Group Meetings</i>	
Patrizia Paggio, Manex Agirrezabal, Giulia Di Cristina, Bart Jongejan and Costanza Navarretta	9139
<i>MMCIG: Multimodal Cover Image Generation for Text-only Documents and Its Dataset Construction via Pseudo-labeling</i>	
HYEYeon KIM, Sungwoo Han, Jingun Kwon, Hidetaka Kamigaito and Manabu Okumura	9150
<i>Multimodal Reference by Means of the Pronoun We and Hand Gestures in a Novel Corpus of Parliamentary Opening Debates</i>	
Costanza Navarretta	9162
<i>Multimodal Large Language Models for Low-Resource Languages: A Case Study for Basque</i>	
Lukas Arana, Julen Etxaniz, Ander Salaberria and Gorka Azkune	9172
<i>Real-Time Generation of Game Video Commentary with Multimodal LLMs: Pause-Aware Decoding Approaches</i>	
Anum Afzal, Yuki Saito, Hiroya Takamura, Katsuhito Sudoh, Shinnosuke Takamichi, Graham Neubig, Florian Matthes and Tatsuya Ishigaki	9188
<i>ARB: A Comprehensive Arabic Multimodal Reasoning Benchmark</i>	
Sara Ghaboura, Shubham Patle, Ketan More, Wafa Hamad Mohamed Alghallabi, Omkar Thawakar, Jorma Laaksonen, Hisham Cholakkal, Salman Khan and Rao Anwer	9202
<i>Event Chronography in Multi-modal Data: The BME Method for Quantitative Analyses</i>	
Anaïs Claire Murat, Maria Koutsombogera and Carl Vogel	9217

<i>CANVAS: A Multimodal Dataset of Chinese Textbook Images for Bias and Representation Analysis</i>	
Haotian Zhu, Kefan Yu and Min Li	9226
<i>MM-Conv: A Multimodal Dataset and Benchmark for Context-Aware Grounding in 3D Dialogue</i>	
Anna Deichler, Jim O'Regan, Fethiye Irmak Dogan, Anna Klezovich, Lubos Marcinek, Iolanda Leite and Jonas Beskow	9240
<i>Erase Persona, Forget Lore: Benchmarking Multimodal Copyright Unlearning in Large Vision Language Models</i>	
June Hyoung Kwon, Jungmin Yun and Youngbin Kim	9254
<i>DREAM: A Multicultural Multimodal Dataset Linking Dialogues and Realistic Image Sequences</i>	
Juan Mallo, Marcos Estecha-Garitagoitia, Ricardo Cordoba and Luis Fernando D'Haro	9266
<i>Multimodal Task Interference: A Benchmark and Analysis of History-Target Mismatch in Multimodal LLMs</i>	
Masayuki Kwarada, Tatsuya Ishigaki and Hiroya Takamura	9282
<i>Can Video LLMs See Through Illusions? Video-Illusion QA Benchmark Dataset</i>	
Souto Ohira, Toshio Hirasawa and Mamoru Komachi	9291
<i>To Skip, to Swap or to Not Swap? Identifying Step Transition Types in Instructional Manuals</i>	
Hsiu-Yu Yang, Michael Roth, Andreas Bulling and Carina Silberer	9301
<i>Fruitcakes and Cupcakes Emerging from Noise: The ComposiGen Dataset of Compounds and Their Compositionality</i>	
Jule Godbersen, Sinan Cem Kurtyigit, Emma Raimundo Schulz, Tonmoy Rakshit, Diego Frassinelli, Sabine Schulte im Walde and Carina Silberer	9321
<i>Large Language Models' Internal Perception of Symbolic Music</i>	
Andrew Shin and Kunitake Kaneko	9339
<i>Entity Image and Mixed-Modal Image Retrieval Datasets</i>	
Cristian-Ioan Blaga, Paul Suganthan G C, Sahil Dua, Krishna Srinivasan, Enrique Alfonso, Peter Dornbach, Tom Duerig, Imed Zitouni and Zhe Dong	9349
<i>Generating Sign Language Poses from HamNoSys and Natural Language Descriptions</i>	
Santiago Máximo and Luis Chiruzzo	9358
<i>Evaluating Discriminability of Vision-Language Models</i>	
Masayasu Muraoka and Naoaki Okazaki	9368
<i>Seeing the Other Side: Diagnostic Tasks for Viewpoint Reasoning in Vision–Language Models</i>	
Makoto Takenaka and Hitomi Yanaka	9386
<i>Multi-modal, Multi-task, Multi-criteria Automatic Evaluation with Vision Language Models</i>	
Masanari Oi, Masahiro Kaneko, Naoaki Okazaki and Nakamasa Inoue	9396
<i>Challenges in Image-Caption Association in Portuguese: Evaluating the CLIP Model on the FM30K Dataset</i>	
Vitória Colonetti Benedet, Gustavo Lopes Tamiosso, Rafael Oleques Nunes and Dennis Giovani Balreira	9409

<i>A Large-Scale Instruction-Tuning Dataset and Models for Slovenian Vision-Language Tasks</i> Matej Martinc and Domen Vreš	9420
<i>A Parallel Cross-Lingual Benchmark for Multimodal Idiomaticity Understanding</i>	
Dilara Torunoğlu-Selamet, Doğukan Arslan, Rodrigo Wilkens, Wei He, Doruk Eryiğit, Thomas Pickard, Adriana S. Pagano, Aline Villavicencio, Gülşen Eryiğit, Ágnes Abuczki, Aida Cardoso, Alesia Lazarenka, Dina Almassova, Amália Mendes, Anna Kanellopoulou, Antoni Brosa-Rodriguez, Baiba Valkovska, Beata Wojtowicz, Bolette Pedersen, Carlos Manuel Hidalgo-Ternero, Chaya Liebeskind, Danka Jokić, Diego Alves, Eleni Triantafyllidi, Erik Velldal, Fred Philippy, Giedre Valunaite Oleskeviciene, Ieva Rizgeliene, Inguna Skadina, Irina Lobzhanidze, Isabell Stinessen Haugen, Jauza Akbar Krito, Jelena M. Marković, Johanna Monti, Josue Alejandro Sauca, Kaja Dobrovoljc Zor, Kingsley O. Ugwuanyi, Laura Rituma, Lilja Øvrelid, Maha Tufail Agro, Manzura Abjalova, Maria Chatzigrigoriou, María del Mar Sánchez Ramos, Marija Pendevska, Masoumeh Seyyedrezaei, Mehrnoush Shamsfard, Momina Ahsan, Muhammad Ahsan Riaz Khan, Nathalie Carmen Hau Norman, Nilay Erdem Ayyıldız, Nina Hosseini-Kivanani, Noémi Ligeti-Nagy, Numaan Naeem, Olha Kanishcheva, Olha Yatsyshyna, Daniil Orel, Petra Giommarelli, Petya Osenova, Radovan Garabik, Regina E. Semou, Rozane Rebechi, Salsabila Zahirah Pranida, Samia Touileb, Sanni Nimb, Sarfraz Ahmad, Sarvinoz Sharipova, Shahar Golan, Shaoxiong Ji, Sopuruchi Christian Aboh, Srdjan Sucur, Stella Markantonatou, Sussi Olsen, Vahide Tajalli, Veronika Lipp, Voula Giouli, Yelda Yeşildal Eraydın, Zahra Saaberi and Zhuohan Xie	9434
<i>Which Way Does Time Flow? A Psychophysics-Grounded Evaluation for Vision–Language Models</i> Shiho Matta, Lis Kanashiro Pereira, Peitao Han, Shigeru Kitazawa and Fei Cheng ..	9449
<i>I Came, I Saw, I Explained: Benchmarking Multimodal LLMs on Figurative Meaning in Memes</i> Shijia Zhou, Saif M. Mohammad, Barbara Plank and Diego Frassinelli	9460
<i>DEJIMA: A Novel Large-scale Japanese Dataset for Image Captioning and Visual Question Answering</i> Toshiki Katsube, Fukuhara Taiga, Kenichiro Ando, Yusuke Mukuta, Kohei Uehara and Tatsuya Harada	9478
<i>CLEVR-3D-DeRef</i> Mary Lynn Martin, Martha Palmer and Maria Leonor Pacheco	9490
<i>Bridging Text-to-Sign Translation via Codebook-Oriented Pretraining</i> Ninlawat Phuangchoke and Chantri Polprasert	9504
<i>A Resource and Evaluation Method for Phonological Continuity in Japanese Sign Language</i> Jundai Inoue, Daisuke Hara and Makoto Miwa	9514
<i>Sentiment Analysis of German Sign Language Fairy Tales</i> Fabrizio Nunnari, Siddhant Jain and Patrick Gebhard	9525
<i>A Critical Study of Automatic Evaluation in Sign Language Translation</i> Shakib Yazdani, Yasser HAMIDULLAH, Cristina España-Bonet, Eleftherios Avramidis and Josef van Genabith	9535
<i>How Much Data Is Enough Data? A New Motion Capture Corpus for Probabilistic Sign Language Generation</i> Anna Klezovich, Johanna Mesch, Gustav Eje Henter and Jonas Beskow	9549

<i>Decomposing Sign Language Movements: A Multi-Band Visualization Method for Articulatory Analysis</i>	
Antonio F. G. Sevilla and José María Lahoz-Bengoechea	9559
<i>Implicit Bias in Peer Review: Through the Lens of Language Abstraction</i>	
Xulang Zhang, Rui Mao and Erik Cambria	9569
<i>The PARLO Dementia Corpus: A German Multi-Center Resource for Alzheimer's Disease</i>	
Franziska Braun, Christopher Witzl, Florian Hönig, Elmar Nöth, Tobias Bocklet and Korbinian Riedhammer	9581
<i>Lexical and Discourse Semantics in a Reading-time Corpus of English</i>	
Jakub Dotlacil, Laia Colina Fortuny, Li Kloostra and Johan Bos	9592
<i>Semantic Capacity in Language Learners and LLMs: A Case Study of Quantifier Scope</i>	
Shaohua Fang, Yue Li and Yan Cong	9602
<i>How Long Does a Quick Kiss Take? Studying Event Duration of Light Verb Constructions Using Explicit Word Embeddings</i>	
Lin de Huybrecht and Geraint A. Wiggins	9618
<i>Disambiguation of Emotion Annotations by Contextualizing Events in Plausible Narratives</i>	
Johannes Schaefer and Roman Klinger	9635
<i>Identifying Contexts of Distress in College Students' Reddit Posts: A Comparative Study of Classical NLP and Large Language Models</i>	
Carine Graff and Nikhil Krishnaswamy	9657
<i>TiC-MuFormer: Time-Aware Caption-Integrated Multimodal Transformers for User-Level Mental Health Modeling</i>	
Georgios Tsoumplekas, Yannis Spyridis and Vasileios Argyriou	9669
<i>Improving Neural Argumentative Stance Classification in Controversial Topics with Emotion-Lexicon Features</i>	
Mohammad Yeghaneh Abkenar, Weixing Wang, Manfred Stede, Mark A. Finlayson, Davide Picca and Panagiotis Ioannidis	9678
<i>Emotion Transcription in Conversation: A Benchmark for Capturing Subtle and Complex Emotional States through Natural Language</i>	
Yoshiki Tanaka, Ryuichi Uehara, Koji Inoue and Michimasa Inaba	9692
<i>SETUP: Sentence-level English-To-Uniform Meaning Representation Parser</i>	
Emma Markle, Javier Gutierrez Bach and Shira Wein	9710
<i>This One or That One? A Study on Accessibility via Demonstratives with Multimodal Large Language Models</i>	
Yu Wang, Emmanuele Chersoni and Chu-Ren Huang	9722
<i>AMR Parsing beyond English: An Experiment on Bulgarian, French, Hungarian and Ukrainian</i>	
Ivaylo Mitov, Tadzhat Marharian, Zsofia F. Hauk, Samba FALL, Maxime Amblard and Bruno Guillaume	9733

<i>Semantic Parsing for Evaluating Large Language Models: Separating Linguistic Abilities with YARN</i>	
Rémi DE VERGNETTE and Maxime Amblard	9745
<i>Two Ojibwe Constraint Grammars: Morphological Disambiguation and Dependency Parsing</i>	
Matthias Diederichsen and Christopher Hammerly	9756
<i>Multimodal LLMs Do Not Compose Skills Optimally across Modalities</i>	
Paula Ontalvilla, Aitor Ormazabal and Gorka Azkune	9767
<i>Code-Switching in End-to-End Automatic Speech Recognition: A Systematic Literature Review</i>	
Maha Tufail Agro, Atharva A. Kulkarni, Karima Kadaoui, Zeerak Talat and Hanan Aldarmaki	9790
<i>MUStReason: A Benchmark for Diagnosing Pragmatic Reasoning in VideoLMs for Multimodal Sarcasm Detection.</i>	
Anisha Saha, Varsha Suresh, Timothy Hospedales and Vera Demberg	9813
<i>Human-Centered Multimodal Fusion for Sexism Detection in Memes with Eye-Tracking, Heart Rate, and EEG Signals</i>	
Iván Arcos Gabaldón, Paolo Rosso and Elena Gomis Vicent	9830
<i>Nos_Brais-GL: A FAIR Galician TTS Corpus for Neural Speech Synthesis</i>	
Adina Ioana Vladu, Antonio Moscoso Sánchez, Carmen Magariños, María Perez Lago and Elisa Fernández Rei	9841
<i>DR-CUP: A Dataset on Real-time Commentary in U.S. Presidential Debates</i>	
Yu-Yu Chang, Huan-Wen Ho, Chung-Chi Chen and Ming-Hung Wang	9850
<i>Russian Generative Spelling, Punctuation and Capitalization Correction</i>	
Nikita Martynov, Danil Astafurov, Ulyana Isaeva, Ivan Vasil'yevich Maksimov, Joqsan Azocar, Dmitrii Kosenko and Alena Fenogenova	9861
<i>Using Multimodal and Language-Agnostic Sentence Embeddings for Abstractive Summarization</i>	
Chaimae Chellaf El Hammoud, Salima Mdhaffar, Yannick Estève and Stéphane Huet	9873
<i>Gradient-Controlled Decoding: A Safety Guardrail for LLMs with Dual-Anchor Steering</i>	
Purva Chiniya, Kevin Joseph Scaria and Sagar Chaturvedi	9884
<i>The Chronicles of RiDiC: Generating Datasets with Controlled Popularity Distribution for Long-form Factuality Evaluation</i>	
Pavel Braslavski, Dmitrii Iarosh, Nikita Sergeevich Sushko, Andrey Sakhovskiy, Vasily Konovalov, Elena Tutubalina and Alexander Panchenko	9893
<i>MeteoGalEus: An Iberian Multilingual Weather Dataset in Galician, Euskera, and Spanish</i>	
Ainhoa Vivel-Couso, Nella Zabrina Pramata, David Robredo, Aitor Soroa and Jose Maria Alonso-Moral	9905
<i>RadTimeline: Timeline Summarization for Longitudinal Radiological Lung Findings</i>	
Sitong Zhou, Meliha Yetisgen and Mari Ostendorf	9920
<i>InstructSum: A Benchmark to Evaluate Instruction-Following Capability of Large Language Models in Summarization</i>	
Kosuke Nishida, Kyosuke Nishida and Itsumi Saito	9940

<i>NOVELSUM: Evaluating Long-Form Summary Generation for Historical Scandinavian Novels</i> Ali Al-Laith, Alexander Conroy, Kirstine Nielsen Degn, Jens Bjerring-Hansen and Daniel Hershcovich	9953
<i>Evaluating Large Language Models for Text-to-Gloss Translation in Kazakh-Russian Sign Language: A Pilot Study</i> Zhanibek Kozhimbayev and Alfarabi Imashev	9964
<i>HotelCheckSpan: A Benchmark Dataset for LLM Faithfulness</i> Patricia Schmidtova, Ondrej Dusek and Saad Mahamood	9973
<i>Procrustes Analysis for Improving Language Model Merging</i> Olivier Ferret	9988
<i>MetaCORA: A Meta-Learned Curriculum for Adversarial and Contrastive Robustness in Speech Recognition</i> Yuqian Dai, Chun Fai Chan, Ying Ki Wong and Tsz Ho Pun	9999
<i>Insights from Transfer Learning Experiments with Word-in-Context and Word Sense Disambiguation Models</i> Alp Mujko and Dominik Schlechtweg	10009
<i>Joint Identification and Induction of Semantic Frames with Scalable Semi-Supervised Graph Clustering</i> Fabian Barteld, Steffen Remus, Saba Anwar, Julian Stawecki, Alexander Ziem and Chris Biemann	10020
<i>Low-Rank Compression of Language Models via Differentiable Rank Selection</i> Sidhant Sundrani, Francesco Tudisco and Pasquale Minervini	10031
<i>Self-supervised Data Augmentation for Text Classification in Low-Data Settings</i> Deyu Ding, Mengying Wang and Andreas Spitz	10046
<i>Distribution-aware Low-bitwidth Quantization for Large Language Models</i> Bao Tan Duy Huynh, Takashi Tsunakawa and Masafumi Nishida	10057
<i>TG-ASR: Translation-Guided Learning with Parallel Gated Cross Attention for Low-Resource Automatic Speech Recognition</i> ChengYeh Yang, Chien-Chun Wang, Li-Wei Chen, Hung-Shin Lee, Hsin-Min Wang and Berlin Chen	10071
<i>Harnessing Synergy in Context and Emoji for Joint Detection of Harmful Online Content in Multi-turn Conversations</i> Feiyan Hu, Ciara Anne Byrne, Jiang Zhou, Rena Maycock and Mark Langan	10082
<i>Dynamic Layer Selection for Efficient Tone Recognition in Self-Supervised Speech Models</i> Saint Germes B. BENGONO OBIANG, Norbert TSOPZE and Paulin MELATAGIA YONTA	10093
<i>Intent Recognition in Speech-to-Text Processing in the Context of Natural Interaction with Cognitive Assistive Systems</i> Behnam Ensan, Magnus Jung, Matthias Busch and Andreas Wendemuth	10102

<i>Merging Continual Pretraining Models for Domain-Specialized LLMs: A Case Study in Finance</i> Kentaro Ueda, François Portet, Hirohiko Suwa and Keiichi Yasumoto	10114
<i>Phonetic-based Ranking for Improved Pseudo-Labeling in Low-Resource ASR</i> Marco Matassoni, Roberto Gretter, Falavigna Daniele, Mohamed Nabih Ali Mohamed Nawar, Alessio Brutti, Matteo Negri, Mauro Cettolo, Marco Gaido, Sara Papi and Luisa Bentivogli	10130
<i>Privacy-Preserving Information Extraction with Local LLMs: A Comparative Study on Dutch Debt Collection Letters</i> Beyza Celep, Natalia Amat-Lefort and Joost Visser	10140
<i>Forewarned Is Forearmed: When Non-Sequential Embedding Turns into an Anomaly Detector</i> Elys Allesiardo, Antoine Caubrière and Valentin Vielzeuf	10150
<i>A Joint Detection Framework for Latvian Loanwords and Calques Using Monolingual Data</i> Yelingyun Zhang, Atis Kapenieks and Marina Platonova	10157
<i>Pantagruel: Unified Self-Supervised Encoders for French Text and Speech</i> Phuong-Hang Le, Valentin Pelloin, Arnault Chatelain, Maryem Bouziane, Mohammed Ghen- nai, Qianwen Guan, Kirill Milintsevich, Salima Mdhaffar, Aidan Mannion, Nils Defauw, Shuyue Gu, Alexandre Daniel Audibert, Marco Dinarelli, Yannick Estève, Lorraine Goeuriot, Steffen Lalande, Nicolas Hervé, Maximin Coavoux, François Portet, Étienne Ollion, Marie Candito, Maxime Peyrard, Solange Rossato, Benjamin Lecouteux, Aurélie Nardy, Gilles Sérasset, Vin- cent Segonne, Solène Evain, Diandra Fabre and Didier Schwab	10168
<i>Merge and Conquer: Instructing Multilingual Models by Adding Target Language Weights</i> Eneko Valero, Maria Ribalta i Albado, Oscar Sainz, Naiara Perez and German Rigau	10192
<i>SemiAdapt: Semi-Supervised and Efficient LoRA-Based Domain Adaptation for Low-Resource Irish Machine Translation with Transformers</i> Josh McGiff and Nikola S. Nikolov	10208
<i>Data Selection Effects on Self-Supervised Learning of Audio Representations for French Audio- visual Broadcasts</i> Valentin Pelloin, Lina Bekkali, Reda Dehak and David Doukhan	10221
<i>SENS-ASR: Semantic Embedding Injection in Neural-transducer for Streaming Automatic Speech Recognition</i> Youness Dkhissi, Valentin Vielzeuf, Elys Allesiardo and Anthony Larcher	10233
<i>Efficient Financial Language Understanding via Distillation with Synthetic Data</i> Wen-Fong (Xavier) Huang and Edwin Simpson	10242
<i>Rubric-Guided Fine-tuning of SpeechLLMs for Multi-Aspect, Multi-Rater L2 Reading-Speech Assessment</i> Aditya Kamlesh Parikh, Cristian Tejedor-García, Catia Cucchiarini and Helmer Strik	10255

<i>Leveraging Semi-Supervised Learning for Multimodal Hate Speech Data Annotation and Detection</i>	
Rathi Adarshi Rammohan, Zhao Ren, Dominik Puchała, Aleksandra Świderska, Dennis Küster and Tanja Schultz	10266
<i>Lexicalized Constituency Parsing for Middle Dutch: Low-resource Training and Cross-Domain Generalization</i>	
Yiming Liang and Fang Zhao	10276
<i>Quantifying the Accuracy and Cost Impact of Design Decisions in Budget-Constrained Agentic LLM Search</i>	
Kyle A. McCleary and James M. Ghawaly	10291
<i>Reason-to-Learn (R2L): Multi-Agent Knowledge Distillation for Lightweight LLMs in Sentiment Analysis</i>	
Le-Huy Tu, Quan Nguyen, Vincent NGUYEN, Johanna Bjorklund and Xuan-Son Vu	10301
<i>PRISM: Partial Ranking via Inter-layer Semantic Measurement for Efficient Fine-tuning of Language Models</i>	
Aldrin Kabya Biswas, Md Fahim, Md. Ashraful Amin, Amin Ahsan Ali and AKM Mahbubur Rahman	10313
<i>SEFL: A Framework for Generating Synthetic Educational Assignment Feedback with LLM Agents</i>	
Mike Zhang, Amalie Pernille Dilling, Léon Gondelman, Niels Erik Ruan Lyngdorf, Euan D. Lindsay and Johannes Bjerva	10324
<i>LGSE: Lexically Grounded Subword Embedding Initialization for Low-Resource Language Adaptation</i>	
Hailay Kidu Teklehaymanot, Dren Fazlija and Wolfgang Nejdl	10341
<i>A Cheap Lunch: Synthetic Annotation With Reduced Human Effort for Medical Text Mining</i>	
Shutao Chen and Piek T.J.M. Vossen	10353
<i>Supervised Contrastive Fine-Tuning for Active Few-Shot Learning</i>	
Zirui Zhang, Lei Ge and Shengyu Qiao	10365
<i>Simulating Student Interactions for Virtual Pretesting with In-Context Learning</i>	
Arthur Thuy, Luca Benedetto, Ekaterina Loginova and Dries F. Benoit	10376
<i>An Exploration-Analysis-Disambiguation Reasoning Framework for Word Sense Disambiguation with Low-Parameter LLMs</i>	
Deshan Koshala Sumanathilaka, Nicholas Micallef and Julian Hough	10390
<i>Building Effective Japanese Medical LLMs with an Open Recipe for Domain Adaptation through Continued Pre-training</i>	
Akiko Aizawa, Yuki Arase, Fei Cheng, Jiahao Huang, Zhiyi Huang, Junfeng Jiang, Teruhito Kanazawa, Daisuke Kawahara, Kazuma Kobayashi, Takashi Kodama, Sadao Kurohashi, Yusuke Oda, Yuma Tsuta, Zhen Wan, Zhishen Yang and Rio Yokota	10405

<i>New Encoders for German Trained from Scratch: Comparing ModernGBERT with Converted LLM2Vec Models</i>	
Julia Wunderle, Anton Ehrmanntraut, Jan Pfister, Fotis Jannidis and Andreas Hotho	10424
<i>Arabic ChartSumm: An English-to-Arabic Benchmark for Metadata-to-Text Summarization</i>	
Passant Elchafei and Amany Fashwan	10447
<i>Introducing a Bangla Sentence – Gloss Pair Dataset for Bangla Sign Language Translation and Research</i>	
Neelavro Saha, Rafi Shahriyar, Nafis Ashraf Roudra, Saadman Sakib and Annajiat Alim Rasel.....	10457
<i>Language Models as Semantic Augmenters for Sequential Recommenders</i>	
Mahsa Valizadeh, Xiangjue Dong, Rui Tuo and James Caverlee	10467
<i>Efficient Adaptation of English Language Models for Morphologically Rich and Underrepresented Languages: The Case of Arabic</i>	
Ahmed Samy Eldamaty, Mohamed Maher Zenhom Abdelrahman, Mohamed Mostafa Ibrahim Elbehery, Mariam Ashraf and Radwa Elshawi	10485
<i>GhostWriter: Hidden AI-Generated Texts over Multiple Languages, Domains and Generators</i>	
Manuel Schaaf, Kevin Bönisch and Alexander Mehler	10497
<i>Using LLMs to Extract Instances of Schematic Constructions from Unannotated L2 Learner Corpora</i>	
Jelena Kallas, Ahto Kiil, Heete Sahkai, Geda Paulsen and Kertu Saul.....	10517
<i>Corruption-Based Data Augmentation for Arabic Essay Scoring: A Preliminary Study on the Organization Trait</i>	
May Saed Bashendy and Tamer Elsayed	10525
<i>Structured Prompting for Arabic Essay Proficiency: A Trait-Centric Evaluation Approach</i>	
Salim Al Mandhari, Hieu Pham Dinh, Mo El-Haj and Paul Rayson.....	10532
<i>ManufactuBERT: Efficient Continual Pretraining for Manufacturing</i>	
Robin Armingaud and Romaric Besancon	10545
<i>Śmigiel Dataset: Laying Foundations for Investigating Machine-Generated Text Detection in Polish</i>	
Jakub Strebeyko, Alina Wróblewska and Piotr Przybyła.....	10556
<i>Extracting Medical Image-Related Entities from Spanish Electronic Health Records Using NER Methods</i>	
Alexander Platas, Marcos Merino, Elena Zotova, Montse Cuadros, Karen López-Linares, Mikel Pérez de Mendiola, María Gálvez, Cristina Barba and Antón Asla	10569
<i>A Novel Synthetic Dataset for Few-Shot Legal Relation Extraction in German</i>	
Shiva Banasaz Nouri, Elena Leitner, Julian Moreno-Schneider and Georg Rehm...	10579
<i>LLM-Based Data Generation and Clinical Skills Evaluation for Low-Resource French OSCEs</i>	
Tian Huang, Tom Bourgeade and Irina Illina	10592

Instruction-Tuned Urdu LLMs: Efficient Adaptation of Llama Models and Evaluation Resources for Urdu

Munief Hassan Tahir, Sana Shams, Sarmad Hussain and Miriam Butt 10603

Is Biomedical Specialization Still Worth It? Insights from Domain-Adaptive Language Modelling with a New French Health Corpus

Aidan Mannion, Cécile Macaire, Armand Violle, Stéphane Ohayon, Xavier Tannier, Didier Schwab, Lorraine Goeuriot and François Portet 10617

TildeOpen LLM: Leveraging Curriculum Learning to Achieve Equitable Language Representation

Toms Bergmanis, Ingus Jānis Pretkalniņš, Martins Kronis, Davis Nicmanis, Jelīzaveta Jelin-ska, Roberts Rozis, Rinalds Vīksna and Marcis Pinnis 10634

Common Sense vs. Morality: The Curious Case of Narrative Focus Bias in LLMs

Saugata Purkayastha, Pranav Kushare, Pragya Paramita Pal and Sukannya Purkayastha 10653

“Emphasizing the Commendable”: A Study of Homogenized Transitive Verb Constructions in Machine Generated Peer Reviews

Hing-Yuet Fung, Chi-kiu Lo and Samuel Larkin 10664

CoDAE: Adapting Large Language Models for Education via Chain-of-Thought Data Augmentation

Shuzhou Yuan, William LaCroix, Hardik Ghoshal, Ercong Nie and Michael Färber . 10677

Synthetic Instruction Generation for Low-Resource Nordic Languages: Viability and Limitations in LLM Instruction-Tuning

Mathias Stenlund, Annika Simonsen, Lars Bungum, Jan Ebert, Jiangtao Wang, Oleg Filatov, Hemanadhan Myneni, Morris Riedel and Hafsteinn Einarsson 10688

AYN: A Tiny Yet Competitive Indian Legal Language Model Pretrained from Scratch

Mitodru Niyogi, Eric Gaussier and Arnab Bhattacharya 10699

Low-Resource Dialect Adaptation of Large Language Models: A French Dialect Case-Study

Eeham Khan, Firas Saidani, Owen Van Esbroeck, Richard Khoury and Leila Kosseim 10723

Reformulate and Create, Don’t Translate: Creating Natural Prompts for Underserved Languages

Annika Simonsen, Mathias Stenlund, Lars Bungum, Marc Daniel Skipstað Volhardt and Hafsteinn Einarsson 10735

Generating High Quality Synthetic Data for Dutch Medical Conversations

Cecilia Kuan, Aditya Kamlesh Parikh and Henk van den Heuvel..... 10750

DeepICD-R1: Medical Reasoning through Hierarchical Rewards and Unsupervised Distillation

Tom Röhr, Thomas Maximilian Josef Steffek, Roman Teucher, Keno Bressemer, Alexei Figueroa, Paul Grundmann, Peter Troeger, Felix Alexander Gers and Alexander Löser 10764

SynthLLM: An LLM-based Scalable Synthetic Data Generation Pipeline for Low-Resource Languages

Solmaz Panahi, Vasudevan Nedumpozhimana and John Kelleher 10776

<i>Persona-Conditioned Generation of Patient Self-Reports from EHRs</i> Yuexin Wu, jianming wei and Vasile Rus	10792
<i>SocialStep: Fast Prediction of Social Determinants of Health</i> Paul Landes, Adam Richard Cross and Jimeng Sun	10802
<i>Dynamically Acquiring Text Content to Enable the Classification of Lesser-known Entities for Real-world Tasks</i> Fahmida Alam and Ellen Riloff	10815
<i>RILEC: Detection and Generation of L1 Russian Interference Errors in English Learner Texts</i> Darya Kharlamova and Irina Proskurina	10826
<i>Critical Foreign Policy Decision (CFPD) Benchmark: Measuring Diplomatic Preferences of Large Language Models</i> Benjamin Jensen, Ian J. Reynolds, Yasir Atalan, Michael Garcia, Austin Woo, Anthony Chen and Trevor Howarth	10838
<i>CrisisCL: A Domain Incremental Learning Benchmark for Crisis Management</i> Paul Le Van Kiem, Romain Meunier, Farah Benamara and Véronique MORICEAU ..	10853
<i>Towards Consistent Detection of Cognitive Distortions: LLM-Based Annotation and Dataset-Agnostic Evaluation</i> Neha Sharma, Navneet Agarwal and Kairit Sirts	10866
<i>LLMs as Annotators: Evaluating Model–Human Alignment in Detecting Contentious Language in Historical Corpora</i> Yahui Zhao, Clemencia Siro and Laura Hollink	10883
<i>Widespread Gender and Pronoun Bias in Moral Judgments across LLMs</i> Gustavo Lucius Fernandes, Jeiverson Santos and Pedro O.S Vaz-de-Melo	10897
<i>Frame2KG: A Benchmark and Evaluation Toolkit for Interpretable Frame-to-Graph Generation</i> Lewis N. Watson, Carl Strathearn, Kenny Mitchell and Yanchao Yu	10912
<i>Injecting Structured Biomedical Knowledge into Language Models:Continual Pretraining vs. GraphRAG</i> Jaafer Klila, Sondes Bannour Souihi, rahma boujelbane, Nasredine Semmar and Lamia Hadrach-Belguith	10927
<i>Linguistic Knowledge Graphs for Sense Prediction: A Case-study on Latin</i> Eleonora Ghizzota, Paola Marongiu, Pierpaolo Basile, Stefano Ferilli and Barbara McGillivray	10937
<i>ACID: On the Perception of Online Classism</i> Arianna Muti, Elisa Bassignana, Amanda Cercas Curry, Federica Durante, Dirk Hovy and Debora Nozza	10953
<i>The Spectrum of Sentiment: Optimistic, Pessimistic, and Neutral Voices in Online Depression Discourse</i> Stefana Arina Tabusca, Ana-Maria Bucur and Liviu P. Dinu	10970
<i>A Benchmark Dataset and Comparative Evaluation of Phonemized and Romanized Urdu for Text-to-Speech</i> M Kaab Bin Shahid and Muhammed Izharuddin	10982

S-VoCAL: A Dataset and Evaluation Framework for Inferring Speaking Voice Character Attributes in Literature

Abigail Berthe-Pardo, Gaspard Michel, Elena V. Epure and Christophe Cerisara ... 10994

BankMathBench: A Benchmark for Numerical Reasoning in Banking Scenarios

Yunseung Lee, Subin Kim, Youngjun Kwak and Jaegul Choo 11010

TR-TEB: Turkish Text Embedding Benchmark

Omer Arslan, Atalay Celik, Yusuf Aslan, Hasan Fatih Durkaya, Mustafa Furkan Zenginoglu, Musa Alperen Yilmaz, Merve Gul Kantarci and Mehmet Haklidor 11028

Simple Additions, Substantial Gains: Expanding Scripts, Languages, and Lineage Coverage in URIEL+

Mason Shipton, York Hay Ng, Aditya Khan, Phuong H. Hoang, Xiang Lu, A. Seza Doğruöz and Annie En-Shiun Lee 11045

SciClaimEval: Cross-modal Claim Verification in Scientific Papers

Xanh Ho, Yun-Ang Wu, Sunisth Kumar, Tian Cheng Xia, Florian Boudin, Andre Greiner-Petter and Akiko Aizawa 11060

Localizing Events in Space: Comparing Humans and AI Models

Derrick Eui Gyu Kim, Kenneth Lai and James Pustejovsky 11072

STRUDEL: Unrolling a Benchmark for Evaluating Vision-Language Models on Structured Diagram Understanding across Domains

Daniel Steinigen, Lucie Flek and Sebastian Houben 11085

VG-CoT: Towards Trustworthy Visual Reasoning via Grounded Chain-of-Thought

Byeonggeuk Lim, Kyeonghyun Kim, Jungmin Yun and Youngbin Kim 11108

VectorEdits: A Dataset and Benchmark for Instruction-Based Editing of Vector Graphics

Josef Kuchar, Marek Kadlcik, Michal Spiegel and Michal Stefanik 11119

ViWikiFC: Fact-Checking for Vietnamese Wikipedia-Based Textual Knowledge Source

Hung Tuan Le, Long Truong To, Manh Trong Nguyen and Kiet Van Nguyen 11125

Automated Extraction of Answer Candidates for Question Generation

Claudia Preda, Mihai Dascalu, Stefan Ruseti and Danielle S. McNamara 11141

Green Bots versus Red Bots: Evaluating Large Language Models for Simulating Persuasion Dynamics in Online Influence Campaigns

Majd Eddin Al Ali, Filip Mihai Muntean, Lucia Donatelli and Jurriaan van Diggelen.. 11152

Towards Expectation Detection in Language: A Case Study on Treatment Expectations in Reddit

Aswathy Velutharambath and Amelie Wühl 11172

Empathy Speaks in Metaphors: The Empathy-Metaphor Corpus of Figurative Language in Empathetic Text

Gyeonggeun Lee and Natalie Parde 11188

A Computational Diachronic Analysis of Gen Z Mental Health Discourse: A Large-scale Reddit Corpus Study from Pre- to Post-COVID

Felix Mao 11199

<i>"Oat Milk Vegan Chocolate Taste Great!": Monitoring the Food Transition Debate in Reddit</i> Greta Zella, Jan Willem Bolderdijk, Saskia Peels, Gerry Wakker and Tommaso Caselli	11209
<i>ClimateChat-300K: A Multi-Modal Facebook Dataset for Understanding Diverse Perspectives in Climate Communication</i> Wajdi Zaghouani, Md. Rafiul Biswas, Mabrouka Bessghaier, Shima Amer Ibrahim and George Mikros	11221
<i>HateMirage: An Explainable Multi-Dimensional Dataset for Decoding Faux Hate and Subtle Online Abuse</i> Sai Kartheek Reddy Kasu, Shankar Biradar, SUNIL SAUMYA and Md. Shad Akhtar	11230
<i>MindSET: Advancing Mental Health Benchmarking through Large-Scale Social Media Data</i> Saad Mankarious, Edward Kempa, Daniel Wiechmann, Elma Kerz, Yu Qiao and Ayah Zirikly	11241
<i>A Corpus of Misunderstood Irony on Turkish Social Media</i> Çağrı Çöltekin and Güliz Güneş	11252
<i>A Corpus of Joint EEG and Self-Paced Reading of Natural Dutch Texts</i> Sara Møller Østergaard, Lenneke Doris Lichtenberg, Laura Boon and Bruno Nicenboim 11260	
<i>Evaluation Drift in LLM Personality Induction: Are We Moving the Goalpost?</i> Prateek Kumar Rajput, Yewei Song, Iyiola Emmanuel Olatunji, Jacques Klein and Tegawendé Bissyande	11272
<i>A Multi-Dialectal, Longitudinal Corpus of Human-AI Hybrid Language Production</i> Qiao Gan, Jonathan Dunn, Andrea Nini and Benjamin Adams	11286
<i>Semantic Information: A Difference That Makes a Difference</i> J. Nathanael Philipp, Max Kölbl and Michael Richter	11300
<i>Modeling the Memory-Surprisal Trade-Off over Time: Communicative Efficiency Decreases with Lexico-Grammatical Change in Scientific English</i> Julius Steuer, Marie-Pauline Krielke, Stefania Degaetano-Ortlieb, Elke Teich and Dietrich Klakow	11309
<i>Mechanistic Interpretability Meets Cognitive Linguistics: Modelling Locative Image Schemas in the Circuit Framework</i> Mattia Proietti, Afra Alishahi, Grzegorz Chrupała and Alessandro Lenci	11320
<i>Variation Is the Norm: Embracing Sociolinguistics in NLP</i> Anne-Marie Lutgen, Alistair Plum, Verena Blaschke, Barbara Plank and Christoph Purschke	11332
<i>Appraisal Theory-Informed Emotion Prediction</i> Xiaowei Wang, Jayant Teotia, Rui Mao, Wandeep Kaur Ratan Singh, Sabrina Binti Tiun and Erik Cambria	11345

<i>The Evolution of Philosophy: A Metaphorical Cognition Perspective</i>	
Rui Mao, Dapeng Chen, Zihao Huang, Xulang Zhang and Erik Cambria	11359
<i>Predicting States of Understanding in Explanatory Interactions Using Cognitive Load-Related Linguistic Cues</i>	
Yu Wang, Olcay Türk, Angela Grimmering and Hendrik Buschmeier	11368
<i>Figurative Language in Alzheimer's Discourse: Linguistic and Neural Alignment in Clinical Narratives</i>	
Diana Kylymnyk, Vitória Hilgert Tomasel, Helena Caseli, Edward Watkins, Aline Villavicencio and Rodrigo Wilkens	11379
<i>Prompting Instruction-tuned LLMs for Semantic Similarity Values</i>	
Xander Akiko Snelder, Yunchong Huang and Jelke Bloem	11390
<i>Rethinking Evaluation in Retrieval-Augmented Personalized Dialogue: A Cognitive and Linguistic Perspective</i>	
Tianyi Zhang and David Traum	11404
<i>Evaluating Multimodal Large Language Model Narrative Interpretation through the Lens of Appraisal Theory</i>	
Jayant Teotia, Xiaowei Wang, Xulang Zhang, Rui Mao and Erik Cambria.....	11417
<i>Mapping Liberty Metaphors across Cultures and Time</i>	
Sidney Suen, Rui Mao, Kenneth Kwok and Erik Cambria	11427
<i>The Sensorimotor Norms for the Chinese Classifiers</i>	
Yimei Shao, Yu-Yin Hsu and Chu-Ren Huang	11439
<i>DeepQuestion: Systematic Generation of Real-World Challenges for Evaluating LLMs Performance</i>	
Ali Khoramfar, Ali Ramezani, Mohammad Mahdi Mohajeri, Mohammad Javad Dousti, Majid Nili Ahmadabadi and Hesham Faili.....	11451
<i>Pragmatic Modelling in Language Learning: Caregiver Question-Answer Feedback in Child-Directed Dialogue</i>	
Maryam Bala, Johannes Heim, Elspeth Edelstein and Arabella Sinclair	11461
<i>Modular Approach to Automating Morphological Components in Grammar Engineering</i>	
Ekaterina Voloshina and Krasimir Angelov.....	11479
<i>MorfFlex: Handling Rich Morphology</i>	
Jaroslava Hlaváčová, Marie Mikulová, Barbora Štěpánková, Milan Straka and Jan Hajič	11495
<i>Using Valency Inheritance in Building a Valency Lexicon</i>	
Václava Kettnerová, Veronika Kolářová, Jiří Mírovský and Michal Olbrich.....	11506
<i>From CHAT to Coded CoNLL-U: A Reproducible Pipeline for the Syntactic Annotation and Querying of Child Language Data</i>	
Achim Stein	11516
<i>TækTæk: Syntactic Analysis of Language Use on Danish TikTok</i>	
Thea Kristensen and Rob van der Goot	11524

<i>Adaptive Chunking: Optimizing Chunking-Method Selection for RAG</i>	
Paulo Roberto de Moura Júnior, Jean Lelong and Annabelle Blangero	11535
<i>Do Large Language Models Grasp the Grammar? Evidence from Grammar-Book-Guided Probing in Luxembourgish</i>	
Lujun LI, Yewei Song, Lama Sleem, Yiqun Wang, Yangjie Xu, Cedric LOTHRTZ, Niccolo' Gentile, Radu State, Tegawendé F. Bissyandé and Jacques Klein	11552
<i>Survey of Tools for Manual Linguistic Annotation: Supporting Diversity through Interactive Exploration</i>	
Ludovica Pannitto, Kaja Dobrovoljc Zor and Bruno Guillaume	11562
<i>TextLens & LeTTuce: Automated Corpus Annotation and Multilingual Tagging as a Service</i>	
Cynthia Van Hee, Jonas Doumen, Vincent Prins, Pranaydeep Singh, Vincent Vandeghinste and Els Lefever	11574
<i>The Corpus of Contemporary Polish — a New Reference Corpus with Rich Syntactic Annotations</i>	
Witold Kieraś, Małgorzata Marciniak, Marcin Woliński, Katarzyna Krasnowska-Kieraś and Marek Łaziński	11585
<i>Prague Dependency Treebank - Consolidated 2.0: Enriching a Complex Annotation Scheme</i>	
Marie Mikulová, Jiří Mírovský, Milan Straka, Pavlína Synková, Jan Štěpánek, Barbora Štěpánková and Jan Hajič	11593
<i>Meet UD_Czech-PDTC: A Large and Genre-Rich Treebank in Universal Dependencies</i>	
Marie Mikulová, Barbora Štěpánková, Daniel Zeman, Jan Štěpánek, Milan Straka and Jan Hajič	11606
<i>Encoding Logical Relations of Chinese Complex Sentences within the Universal Dependencies Framework</i>	
Hongpu Zhu and Hongzhi Xu	11620
<i>Unsupervised Labelling of Mutation Triggers in Welsh</i>	
Nicolás Gutiérrez-Rolón and Fernando Alva-Manchego	11631
<i>UzUDT: Uzbek Universal Dependencies Treebank</i>	
Sanatbek Gayratovich Matlatipov and Mersaid Aripov	11642
<i>BRAGD: Constrained Multi-Label POS Tagging for Faroese</i>	
Annika Simonsen, Barbara Scalvini, Uni Johannesen, Iben Nyholm Debess, Hafsteinn Einarsson and Vésteinn Snæbjarnarson	11650
<i>Syntactic Sugar for Syntactic Queries: Sequential Representations for Dependency Queries</i>	
Niklas Deworetzki and Arianna Masciolini	11669
<i>Context Is (Almost) Everything: Llama-3 on Structured Output and AMR Parsing</i>	
Maja Buljan, Stephan Oepen and Lilja Øvrelid	11679
<i>Towards the Morphological Annotation of North Markian (Low German)</i>	
Christian Chiarcos	11699
<i>Cross-Dataset Inconsistencies in Morphological Annotation: Evidence from Universal Dependencies</i>	
Vlasta Ohlídalová	11715

<i>Improving Latvian Morphosyntactic Parsing with Pretrained Encoders and Analyzer-Constrained Decoding</i>	
Arturs Znotins	11724
<i>CommonMorph: Participatory Morphological Documentation Platform</i>	
Aso Mahmudi, Sina Ahmadi, Kemal Maulana Kurniawan, Rico Sennrich, Eduard H. Hovy and Ekaterina Vylomova	11735
<i>Datasets for Verb Alternations across Languages: BLM Templates and Data Augmentation Strategies</i>	
Giuseppe Samo and Paola Merlo	11747
<i>A Large and Balanced Multi-Domain Arabic Corpus Annotated for Morphology, Syntax, and Readability</i>	
Khalid N. Elmadani, Adel Mahmoud Wizani, Hanada Taha Thomure and Nizar Habash	11761
<i>The DELPH-IN Grammar: A Curated Repository of Grammars and Treebanks</i>	
Francis Bond and Dan Flickinger	11776
<i>Morphemes without Borders: Evaluating Root–Pattern Morphology in Arabic Tokenizers and LLMs</i>	
Yara Yousif Alakeel, Chatrine Qwaider, Hanan Aldarmaki and Sawsan Alqahtani ...	11787
<i>APODICTUS: Automatic Processing of DICTIONary Update candidateS</i>	
Felix Blessing, Johannes S. Sax, Julian Kaufmann, Wei Zhao, Nikolay Arefyev and Dominik Schlechtweg	11800
<i>A Test Collection for Part-of-Speech Tagging and Word Sense Disambiguation</i>	
Robert Krovetz	11813
<i>Creating a Hybrid Rule and Neural Network Based Semantic Tagger Using Silver Standard Data: The PyMUSAS Framework for Multilingual Semantic Annotation</i>	
Andrew Moore, Paul Rayson, Dawn Archer, Tim Czerniak, Dawn Knight, Daisy Monika Lal, Gearóid Ó Donnchadha, Mícheál J. Ó Meachair, Scott Piao, Elaine Uí Dhonnchadha, Johanna Vuorinen, Yan Yabo and Xiaobin Yang	11822
<i>Scare Quotes as Markers of "Questionable" Word Usages and Misalignment in Conversation: An Annotation Study</i>	
Aina Garí Soler, Juan Carlos Zevallos Huaco, Matthieu Labeau and Chloé Clavel ..	11834
<i>Modeling Clinical Uncertainty in Radiology Reports: From Explicit Uncertainty Markers to Implicit Reasoning Pathways</i>	
Paloma Rabaey, Jong Hak Moon, Jung-Oh Lee, Min Gwan Kim, Hangyul Yoon, Thomas Demeester and Edward Choi	11852
<i>ArabDiscrim: A Decade-Long Arabic Facebook Corpus on Racism and Discrimination</i>	
Wajdi Zaghouani, Shima Amer Ibrahim, Mabrouka Bessghaier and Houda Bouamor	11874

<i>DAMETA: An LLM Benchmark for Danish Metaphor Interpretation with Systematically Varied Distractors</i>	
Nina Skovgaard Schneidermann, Sanni Nimb, Nathalie Carmen Hau Norman, Sussi Olsen and Bolette Pedersen	11885
<i>A New Semantic Artifact Based Framework for Studying and Documenting Algospeak and Related Phenomena</i>	
Fahad Khan, Elisa Gugliotta, Elisa Squadrito, Maura Tarquini and Francesca Frontini	11896
<i>Creating a High Quality Abstract Meaning Representation Dataset Automatically</i>	
Johannes Heinecke, Munshi Asadullah, Frédéric Herledan and Geraldine Damnati	11907
<i>Towards a Comprehensive English Wordnet-Wikidata Mapping</i>	
John P. McCrae, Johann Bergh and Krasimir Angelov	11916
<i>AmDi - Ambiguous Words Diachronic Dataset</i>	
Felix Thielen and Kai Kugler	11926
<i>GerVLPro: A CEFR-Graded Vocabulary List of L2 Learners' Productive Vocabulary in German</i>	
Noah-Manuel Michael, Anna Huelsing and Andrea Horbach	11942
<i>Building Bridges between Student and Curricular Language: Creating a Corpus of Abstract Meaning Representations for the Classroom</i>	
Kristin Wright-Bettner, Zheng Cai, zekun zhao, James H. Martin, Jeffrey Flanigan and Martha Palmer	11960
<i>Mu'jam Arriyadh: A Comprehensive Lexicon for Contemporary Arabic Language</i>	
Afrah A. Altamimi, Abdulrahman Alosaimy, Halah Munif Alharbi, Hawra Aljasim, Muneera Alhoshan, Amal Almazrui, Hanan Alharbi, Abdulrahman Saeed Alshehri, Bayan M. Almuqhim, Maryam H. Algarny, Yahya A. Asiri, Abdullah I. Alharbi, SALEH ZAIDAN ALBALAWI, Fawziah Mohammed Asiri, Sara Ali Alhifithi and Abdullah Alfaifi	11972
<i>The Romanian Corpus Annotated with Multiword Expressions. PARSEME-Ro Version 2.0</i>	
Verginica Barbu Mititelu, Mihaela Cristescu, Elena Irimia and Carmen Mîrzea Vasile	11979
<i>Missing Links: LLM-Augmentation of Event Triggers of State Changes in the OpenPI Dataset</i>	
Kyeongmin Rim and James Pustejovsky	11992
<i>VUPMC: A New Political Metaphor Corpus in Mandarin Chinese</i>	
Xiaojuan Tan	12007
<i>Not All Disneys Are the Same: Making Coreference Metonymy-Aware</i>	
Bingyang Ye, Jingxuan Tu and James Pustejovsky	12019
<i>JSTS-Neg: Japanese Semantic Textual Similarity Dataset for Evaluating Negation Understanding Ability</i>	
Reiko Yuasa, Yoshihide Kato and Shigeki Matsubara	12031
<i>Few-shot Prompting or Supervised Tuning? A Comparative Study of LLMs for Linguistically Distant Language Pairs in BDI</i>	
Deepen Naorem, Sanasam Ranbir Singh, Telem Joyson Singh and Priyankoo Sarmah	12042

When Structure Matters: Cross-Lingual Hyperbolic Embeddings for Chinese and English Word-nets

Mao-Chang Ku, Da-Chen Lian, Pin-Er Chen, Po-Ya Angela Wang, Wei-Ling Chen and Shu-Kai HSIEH 12054