RaPID-4@LREC2022: Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments (<https://spraakbanken.gu.se/en/rapid-2022>; full day event, Saturday June 25th, 2022, Marseille, France)

Call for papers

RaPID-4 aims to be an interdisciplinary forum for researchers to share information, findings, methods, models and experience on the collection and processing of data produced by people with various forms of mental, cognitive, neuropsychiatric, or neurodegenerative impairments, such as aphasia, dementia, autism, bipolar disorder, Parkinson's disease or schizophrenia. Particularly, the workshop's focus is on creation, processing and application of data resources from individuals at various stages of these impairments and with varying degrees of severity. Creation of resources includes e.g. the annotation, description, analysis and interpretation of linguistic, paralinguistic and extra-linguistic aspects of such data (i.e. spontaneous spoken language, transcripts, eye tracking, wearable and sensor measurements, digital biomarkers, etc.). Processing of such data can be used to identify, extract, correlate, evaluate and disseminate various linguistic or multimodal phenotypes and measurements, which then can be applied to aid diagnosis, monitor the progression or predict individuals at risk.

A central aim is to facilitate the study of the relationships among various levels of linguistic, paralinguistic and extra-linguistic observations (e.g., acoustic measures; phonological, syntactic and semantic features; eye tracking. sensors, signs and multimodal signals). Submission of papers are invited in all of the aforementioned areas, particularly emphasizing multidisciplinary aspects of processing such data and the interplay between clinical/nursing/medical sciences, language technology, computational linguistics, natural language processing (NLP) and computer science. The workshop will act as a stimulus for the discussion of several ongoing research questions driving current and future research by bringing together researchers from various research communities.

Workshop Description

There is a growing interest among healthcare professionals and clinicians to apply non-invasive, time and cost-effective, easy-to-measure techniques as a complement to the battery of medical and clinical examinations currently undertaken for the early diagnosis or monitoring of brain and mental disorders.

Although many of the causes of cognitive and neuropsychiatric impairments are difficult to foresee and accurately predict, physicians and clinicians work with a wide range of factors that potentially contribute to such impairments, e.g., traumatic brain injuries, genetic predispositions, side effects of medication, and congenital anomalies. In this context, there is new evidence that the acquisition and processing of human language data (e.g., spontaneous story telling) and extra-linguistic and production measures (e.g., from eye tracking, wearable devices or sensors) could be used as a complement to the clinical diagnosis and also provide the foundation for future development of objective criteria to be used for identifying progressive decline or degeneration of normal mental and brain functioning.

An important new area of research in computational linguistics and Natural Language Processing (NLP) emphasizes the processing, analysis, and interpretation of such data. Current research in this field, based on linguistic-oriented analysis of text and speech

produced by such a population, compared to healthy adults, has shown promising outcomes. This is manifested in early diagnosis and prediction of individuals at risk, the differentiation of individuals with various degrees of severity forms of brain and mental illness, and for the monitoring of the progression of such conditions through the longitudinal analysis of language samples or other para and extra-linguistic measurements from various modalities.

Nevertheless, there remains significant work to be done to arrive at more accurate estimates for prediction and classification purposes in the future and more research is required in order to reliably complement the battery of medical and clinical examinations currently undertaken for the early diagnosis or monitoring of, e.g., neurodegenerative and other brain and mental disorders and accordingly, aid the development of new, non-invasive, time and cost-effective and objective (future) clinical tests in neurology, psychology, and psychiatry.

Organizing Committee (*alphabetic list*)
- Kathleen C. Fraser, National Research Council, Canada
- Dimitrios Kokkinakis, University of Gothenburg, Sweden
- Kristina Lundholm Fors, University of Lund, Sweden
- *Johan Skoog, University of Gothenburg, Sweden, <TBC>*
- Charalambos Themistokleous, Johns Hopkins University, USA
- Athanasios Tsanas, University of Edinburgh, UK

Topics of interest
The topics of interest for the workshop session include but are not limited to:

- Infrastructure for the domain: building, adapting and availability of linguistic resources, data sets and tools
- Methods and protocols for data collection
- Acquisition and combination of novel data samples; including digital biomarkers, continuous streaming, monitoring and aggregation of measurements; as well as self-reported behavioral and/or physiological and activity data
- Guidelines, protocols, annotation schemas, annotation tools
- Addressing the challenges of representation, including dealing with data sparsity and dimensionality issues, feature combination from different sources and modalities
- Domain adaptation of NLP/AI tools
- Acoustic/phonetic/phonologic, syntactic, semantic, pragmatic and discourse analysis of data; including modeling of perception (e.g. eye-movement measures of reading) and production processes (e.g. recording of the writing process by means of digital pens, keystroke logging etc.); use of gestures accompanying speech and non-linguistic behavior
- Use of wearable, vision, and ambient sensors or their fusion for detection of cognitive disabilities or decline
- (Novel) Modeling and deep / machine learning approaches for early diagnostics, prediction, monitoring, classification etc. of various cognitive, psychiatric and/or developmental impairments
- Evaluation of the significance of features for screening and diagnostics
- Evaluation of tools, systems, components, metrics, applications and technologies including methodologies making use of NLP; e.g. for predicting clinical scores from (linguistic) features

- Digital platforms/technologies for cognitive assessment and brain training
- Evaluation, comparison and critical assessment of resources
- Involvement of medical/clinical professionals and patients
- Ethical, gender bias and legal questions in research with human data in the domain, and how they can be handled
- Deployment, assessment platforms and services as well as innovative mining approaches that can be translated to practical/clinical applications
- Experiences, lessons learned and the future of NLP/AI in the area

Summary of the Call

RaPID-4 aims to be an interdisciplinary forum for researchers to share information, findings, methods, models and experience on the collection and processing of data produced by people with various forms of mental, cognitive, neuropsychiatric, or neurodegenerative impairments, such as aphasia, dementia, autism, bipolar disorder, Parkinson's disease or schizophrenia. Particularly, the workshop's focus is on creation, processing and application of data resources from individuals at various stages of these impairments and with varying degrees of severity. Creation of resources includes e.g. annotation, description, analysis and interpretation of linguistic, paralinguistic and extra-linguistic data (such as spontaneous spoken language, transcripts, eye tracking measurements, wearable and sensor data, etc.). Processing is done to identify, extract, correlate, evaluate and disseminate various linguistic or multimodal phenotypes and measurements, which then can be applied to aid diagnosis, monitor the progression or predict individuals at risk.

A central aim is to facilitate the study of the relationships among various levels of linguistic, paralinguistic and extra-linguistic observations (e.g., acoustic measures; phonological, syntactic and semantic features; eye tracking measurements; sensors, signs and multimodal signals). Submission of papers are invited in all of the aforementioned areas, particularly emphasizing multidisciplinary aspects of processing such data and the interplay between clinical/nursing/medical sciences, language technology, computational linguistics, natural language processing (NLP) and computer science. The workshop will act as a stimulus for the discussion of several ongoing research questions driving current and future research by bringing together researchers from various research communities.

Identify, Describe and Share your LRs!

- Describing your LRs in the LRE Map is now a normal practice in the submission procedure of LREC (introduced in 2010 and adopted by other conferences). To continue the efforts initiated at LREC 2014 about "Sharing LRs" (data, tools, web-services, etc.), authors will have the possibility, when submitting a paper, to upload LRs in a special LREC repository. This effort of sharing LRs, linked to the LRE Map for their description, may become a new "regular" feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.

- As scientific work requires accurate citations of referenced work so as to allow the community to understand the whole context and also replicate the experiments conducted by other researchers, LREC 2022 endorses the need to uniquely Identify LRs through the use of the International Standard Language Resource Number

(ISLRN, www.islrn.org), a Persistent Unique Identifier to be assigned to each Language Resource. The assignment of ISLRNs to LRs cited in LREC papers will be offered at submission time.

Estimated Audience
30-40

Duration of the workshop
Full Day session

Technical Requirements
- High speed wired connection. Many wireless internet connections should be sufficient to connect and be able to use Zoom, for participants that need to be virtual for remote video presentation
- Access to a computer, video wall display, electricity power (multiple sockets)
- 10 standing poster stands for a planned poster session

Invited speaker(s)
- <TBA>
- <TBA>

(confirmed) **Programme Committee (*alphabetic list)*
- Visar Berisha, Arizona State University, USA
- Gaël Dias, University of Caen Normandie, France
- Jon Andoni Duñabeitia, Nebrija University, Spain and The Arctic University, Norway
- Davida Fromm, Carnegie Mellon University, USA
- Valantis Fyndanis, Cyprus University of Technology, Cyprus
- Gloria Gagliardi, University of Bologna, Italy
- Leontios Hadjileontiadis, Khalifa University, United Arab Emirates
- Christine Howes, University of Gothenburg, Sweden
- Alexandra König, National Institute for Research in Computer Science and Control, France
- Saturnino Luz, University of Edinburgh, UK
- Christina Manouilidou, University of Ljubljana, Slovenia
- Ricardo Muñoz Sánchez, University of Gothenburg, Sweden
- Angus Roberts, King's College, UK
- Kairit Sirts, University of Tartu, Estonia
- Spyridoula Varlokosta, University of Athens, Greece
- Yasunori Yamada, IBM Research Tokyo, Japan
- *… **names to be added***