The LiNCR (pronounced as ‘linker’) workshop aims to provide a venue to explore a new generation of language resources which link and aggregate cognitive behavioural, neuroimaging measurement data to a shared set of richly annotated linguistic data.

We plan to attract experts on brain and language, on language resources, and on big data and machine learning to look at the potential for collaboration towards building multiply linked and richly aggregated Linguistic and Neuro-Cognitive Resources (LiNCR) as well as methodologies for processing the big data integrated in LiNCR. The participants will work together to address format, methodology, as well as legal issues arising from this highly heterogeneous mix of linguistic, behavioral and physiological datasets (e.g. EEG, ERP, Eye-movement, electrodermal activity, fMRI, MEG, as well as behavioral norms). The issues will range from the ontology for aggregation of different types of neuro-cognitive data with linguistic facts to their usage for the evaluation of NLP tools.

In addition to providing a forum for presenting existing LiNCRs as well as innovative research based on integrated heterogeneous datasets, we also welcome project notes and discussions on our proposal that may address issues and challenges arising from new types of LiNCRs.

* MOTIVATION AND TOPICS OF INTEREST*

Language resources to-date can be described as collections of snapshots of language production. They are in vitro and ready to be tested but do not contain any direct information on the cognitive processes that produced them. That is, the in vivo perspectives of language are missing from them. On the other hand, studies on the neurobiological basis of language processing made significant progresses based on collected neurological, neuroimaging and behavioral datasets. But these experimental data typically focus on strictly controlled stimuli annotated with a single linguistic feature. Hence, the potential of linking richly annotated linguistic facts with
experimental data has yet to be realised. This workshop aims to bring together experts from computational, corpus, and neuro-cognitive linguistics to bridge this. We hope not only to herald in a new generation of language resources but also to open a new interdisciplinary frontier in the exploration of human cognition based on LiNCRs.

Recent NLP research demonstrates that the incorporation of behavioural data (e.g. eye-tracking) improves modelling on a variety of NLP tasks (Long et al. 2017, 2019) and these data provide a robust benchmark for the evaluation of essential components of several NLP systems (e.g. word embeddings: see Bakarov et al., 2018; Hollenstein et al., 2019). Similarly, cognitive neuroscience can benefit from richly annotated linguistic data to uncover the relationship between brain regions and different language subprocesses (Wehbe et al. 2014; Huth et al. 2016). The time is ripe to bring these two fields together, and this workshop aims to advance research in this new frontier.

* SUBMISSIONS *

We welcome contributions addressing any of the following aspects of a new generation of language resources that link and aggregate neurological behavioral measurement data to a shared set of richly annotated linguistic data.

- Corpus selection (Mono/Multi-lingual)
- Ontology/framework for linking annotations in different modalities
- Linking experimental results to linguistically annotated data
- Design for multiple neuro-cognitive experimental platforms to share same linguistic data set
- Aggregation and normalization of data between population with special cognitive conditions with normal, and across different linguistic backgrounds
- Stochastic models for knowledge aggregation.

Regular workshop submissions should consist of 4 to 8 pages, references excluded, and they should adopt the same format of the LREC main conference papers (the
submissions are NOT anonymous). The authors will be asked whether they prefer an oral or a poster presentation at submission time. We also accept as cross-submissions papers that might have been presented in other venues and that are relevant for the workshop topics. These will not appear in the conference proceedings and will be presented as posters. The cross-submissions should consist of abstracts of a maximum of 400 words (excluding references). Both workshop papers and abstracts should be uploaded as PDF files in the Softconf START system.

* IMPORTANT DATES *

- Submission Deadline: February 17, 2020
- Notification of Acceptance: March 11, 2020
- Camera-Ready Papers: April 2, 2020
- Workshop Date: May 12, 2020

* E-MAIL ADDRESS *

lincr2020@gmail.com

* ORGANIZING COMMITTEE *

Emmanuele Chersoni (The Hong Kong Polytechnic University)
Barry Devereux (Queen’s University Belfast)
Chu-Ren Huang (The Hong Kong Polytechnic University)

* PROGRAM COMMITTEE (TO BE COMPLETED) *
Kathleen Ahrens (The Hong Kong Polytechnic University)
Amir Bakarov (Higher School of Economics, Moscow)
Leonor Becerra-Bonache (Jean Monnet University)
Philippe Blache (LPL-CNRS)
Zhenguang Cai (Chinese University of Hong Kong)
Cristopher Cieri (University of Pennsylvania)
Steven Derby (Queen’s University Belfast)
Vesna Djokic (Goldsmiths University of London)
Stefan Frank (Radboud University of Nijmegen)
Diego Frassinelli (University of Stuttgart)
Francesca Frontini (Paul Valéry University of Montpellier)
John Hale (University of Georgia)
Nora Hollenstein (University of Zurich)
Shu-Kai Hsieh (National Taiwan University)
Yu-Yin Hsu (The Hong Kong Polytechnic University)
Elisabetta Jezek (University of Pavia)
María Dolores Jiménez López (Universitat Rovira i Virgili, Tarragona)
Ping Li (The Hong Kong Polytechnic University)
Andreas Maria Liesenfeld (The Hong Kong Polytechnic University)
Yunfei Long (University of Nottingham)
Karl David Neergaad (University of Macau)
Noel Nguyen (LPL-CNRS)
Mark Ormerod (Queen’s University Belfast)
Christophe Pallier (INSERM-CEA)
Ludovica Pannitto (University of Trento)
Adam Pease (Articulate Software)
Massimo Poesio (Queen Mary University of London)
Stephen Politzer-Ahles (The Hong Kong Polytechnic University)
James Pustejovsky (Brandeis University)
Giulia Rambelli (University of Pisa)
Anna Rogers (University of Massachusetts Lowell)
Marco Senaldi (Scuola Normale Superiore, Pisa)
Adrià Torrens Urrutia (Universitat Rovira i Virgili, Tarragona)
Marten Van Schijndel (Cornell University)
Cory Shain (Ohio State University)
Mingyu Wan (The Hong Kong Polytechnic University)

*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.