Variations in Action: a MUltilingual Approach

Chu-Ren Huang¹  Laurent Prévot²

¹Faculty of Humanities, The Hong Kong Polytechnic University
²Laboratoire Parole et Langage, Aix Marseille Université

Variamu First Workshop @ Second Asia Pacific Corpus Linguistics Conference, Hong Kong

[Logos of Aix-Marseille Université, The Hong Kong Polytechnic University, INVESTISSEMENTS D’AVENIR, A*Midex]
What is VariAMU?

- Co-Funded collaboration project between Aix Marseille Université, The Hong Kong Polytechnic University and University of Macau.

- **Objective:** Establish a perennial international research structure between the 3 institutions focusing linguistic variations from cross-linguistic and interdisciplinary perspective

- **Interdisciplinarity**
  - Humanities and Social Sciences
  - Information and Computer Sciences
  - Life Sciences

First phase of the project (2 years): Workshops, Summer schools, Short/mid term mobilities
Variatmu is about variations

Which sources of variations

1. Dialectal variation
2. Social variation
3. **Language impairment**
4. Proficiency variation
5. **Communicative situation variation**
6. Inter-individual and speaker state variations
A cross-linguistic perspective

The nice things of it

- Able to address the variation patterns across languages
- Which sources of variations are robust across languages

The challenges ahead

- Extreme care needed while constituting the datasets
- Control all the sources of variation otherwise difficult to draw conclusion from a cross-linguistic perspective
Structure of Variamu

Task 1 - Management

Task 2 - Specific Events

Task 3 - Dissemination

Axis 1 - Variation and natural language processing

Data Acquisition

Data Processing and Linguistic resources

Variation studies across sites

Axis 2 - Variation: Linguistics and Phonetics

*Project Example: Vowel Variation*

Axis 3 - Variation and language pathologies

*Project Example: Parkinson French / Portuguese variations*
Axis 1: Language Resources and Technologies

- Data acquisition and management
  - Control conditions of production (Experimental Linguistics Methodology)
  - Importance of meta-data and data management (Experience acquired through Open Resources and Tools for Language)

- Natural Language Processing and Language Resource Creation
  - Spoken: Segmentation, transcription or ASR, Forced-alignment
  - Written or once transcribed: Tokenization, POS-tagging, chunking, parsing?

- Quantitative Variation studies across sites
  - Require approach that is generic enough to apply access languages
  - More processing performance rather than competence structures (performance grammar) [Blache and Rauzy, 2014]
Axis 2: Phonetics and Linguistics

- Most commonly: Dialectal and/or language contact induced variations
  - gender and regional dialect [Clopper and Smiljanic, 2011]
  - Accent-independent adaptation to foreign accented speech [Baese-Berk et al., 2013]
- Focus on specific phenomena: speech reduction across communicative situations according to lexical factors [Meunier and Espesser, 2011, Yao and Meunier, 2014]
- Corpus Driven approach to Light Verb Variations in World Chineses [Lin et al., 2014]
- Methodology: Automatically extract (speech processing) acoustic and/or prosodic [Hirst, 2007] parameters, duration of phones, of tokens... (SPPAS [Bigi, 2014])
Axis 3: Variations and Language Pathologies

- Motor aspects of speech impairment related diseases
  - Dysarthria across Languages in Parkinson Disease [Pinto and Nunes, 2014]
  - Automatically extract (speech processing) acoustic and/or prosodic parameters
  - Hypothesis: Associated with degree of disease severity, but robust across languages
  - How this alter speech intelligibility? (that results in psychosocial decline)

- Cognitive aspects of language impairments
  - Lack of flexibility on lexical production tasks for Asperger children (HFA) [Duvignau et al., 2013]
  - Discourse reference problems with schizophrenic patients [Champagne-Lavau et al., 2009]

- Clinical Applications
  - Early diagnosis
  - Evaluating impact of treatment to help design treatment program me [Leung, 2014]
NLP and CALL (Computer-Aided Language Learning)

   error grammar (robust parsing + specific rules) [Foster and Vogel, 2004]

2. Phonetics - Phonology: Pronunciation correction

3. However, “processing completely free production input,
   allowing any number and type of errors is not tractable” [Amaral and Meurers, 2006]

Clinical Computational Linguistic

Computational dialectology [Nerbonne, 2003]

More standardly, tools and techniques adaptation to communicative situation (written vs. read speech vs.
spontaneous speech) [Rauzy et al., 2014]

Speaker identification rely on inter-individual variability

Speaker trait identification (incl. deceptive speech) relies on variation as well
Thank you!


Do patients with schizophrenia attribute mental states in a referential communication task?

Effects of gender and regional dialect on prosodic patterns in american english.

Tâches et énoncés complexes chez l’enfant: Jugements sémantiques et flexibilité dans le développement typique et le syndrome d’asperger.


A comparable corpus driven, multivariate approach to light verb variations in world chineses.
In Proceedings of Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), Variations in Action Workshop, Hong Kong.

Vowel reduction in conversational speech in french: The role of lexical factors.

Linguistic variation and computation.
In Proceedings of Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), Variations in Action Workshop, Hong Kong.

In Proceedings of Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), Hong Kong.

ReCALL, 13(1):110–120.

In Proceedings of Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), Variations in Action Workshop, Hong Kong.