

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



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 focussing 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

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

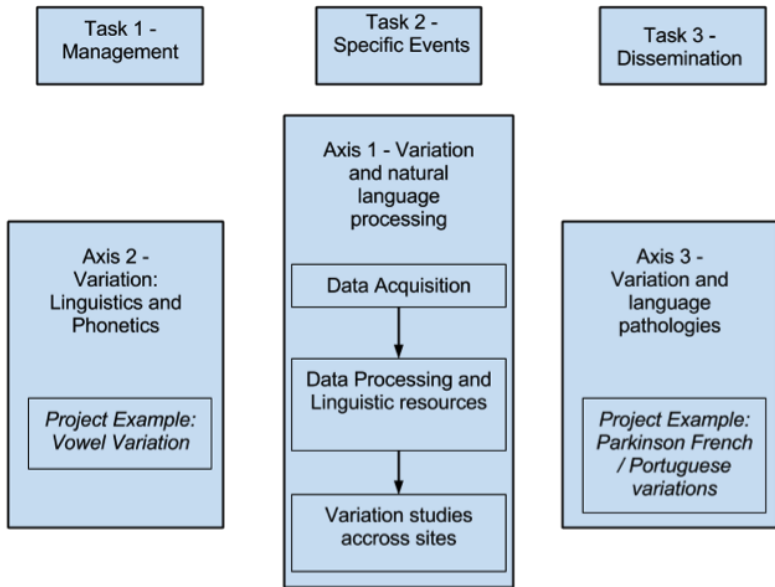
The nice things of it

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



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 across languages
 - More processing performance rather than competence structures (performance grammar) [Blache and Rauzy, 2014]

- 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)
 - ① Syntax: parsing with relaxed constraints [Vandeventer, 2001] ; 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!



Amaral, L. and Meurers, D. (2006).

Using foreign language tutoring systems for grammatical feedback.
Eurocall 2006 Granada. Integrating CALL into study programmes,
page 63.



Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013).

Accent-independent adaptation to foreign accented speech.
The Journal of the Acoustical Society of America, 133(3):EL174–EL180.



Bigi, B. (2014).

Automatic speech segmentation of french: Corpus adaptation.
In Proceedings of Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), Variations in Action Workshop, Hong Kong.



Blache, P. and Rauzy, S. (2014).

A chinese constraint grammar extracted from the chinese treebank.
In Proceedings of Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), Hong Kong.



Champagne-Lavau, M., Fossard, M., Martel, G., Chapdelaine, C., Blouin, G., Rodriguez, J.-P., and Stip, E. (2009).

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

Cognitive neuropsychiatry, 14(3):217–239.



Clopper, C. G. and Smiljanic, R. (2011).

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

Journal of phonetics, 39(2):237–245.



Duvignau, K., Bregeon, C., Garrigou, C., Maffre, T., and Raynaud, J. P. (2013).

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

Travaux de linguistique, (1):53–72.



Foster, J. and Vogel, C. (2004).

Parsing ill-formed text using an error grammar.

[Artificial Intelligence Review](#), 21(3-4):269–291.



Hirst, D. (2007).

A praat plugin for momel and intsint with improved algorithms for modelling and coding intonation.

In [Proceedings of the XVIth International Conference of Phonetic Sciences](#), volume 12331236.



Leung, M.-T. (2014).

Applying corpus study to the treatment of developmental dyslexia in chinese.

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



Lin, J., Jiang, M., and Huang, C.-R. (2014).

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.



Meunier, C. and Espesser, R. (2011).

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

Journal of Phonetics, 39(3):271–278.



Nerbonne, J. (2003).

Linguistic variation and computation.

In Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1, pages 3–10.
Association for Computational Linguistics.



Pinto, S. and Nunes, A. M. (2014).

Dysarthria across languages in parkinson's disease.

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



Rauzy, S., de Montcheuil, G., and Blache, P. (2014).

Marsatag, a tagger for french written texts and speech transcriptions.

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



Vandeventer, A. (2001).

Creating a grammar checker for call by constraint relaxation: a feasibility study.

ReCALL, 13(1):110–120.



Yao, Y. and Meunier, C. (2014).

Phonetic variation in french conversational speech.

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