The ALTA Institute CCC-NLIP PublicHealthMT

Research Staff Forum, 6 November 2020
Andrew Caines, NLIP Group

My background

- ❏ Linguistics PhD, 2006 to 2009~10
- ❏ True Knowledge (-> Evi -> Alexa), late 2009
- ❏ European Bioinformatics Institute, 2010-2013
- ❏ ALTA Institute since Oct 2013
- ❏ Computer Lab transfer Oct 2018
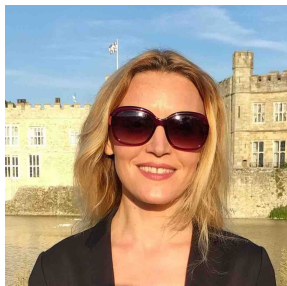- ❏ RSF since summer 2019

ALTA Institute

- ❏ Automated Language Teaching & Assessment
- ❏ Director: Prof Paula Buttery
- ❏ PIs: Paula, Profs Ted Briscoe, Andy Rice (CL), Mark Gales (CUED)
- ❏ Funding: Cambridge Assessment
- ❏ CL & CUED groups (+ Linguistics)
- ❏ AI for Education, Education Technology, Machine Learning, Natural Language Processing, Speech Processing

ALTA Institute

- ❏ At CL:
- ❏ RA / R.Assis: me, Øistein Andersen, Chris Bryant, Mariano Felice, Ekaterina Kochmar, David Strohmaier, Shiva Taslimipoor, Zheng Yuan
- ❏ PhD: Chris Davis, Mark Elliott, Youmna Farag, Sian Gooding, Russell Moore*, Gladys Tyen, Ahmed Zaidi*
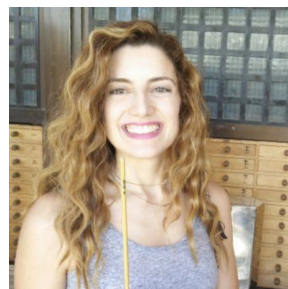
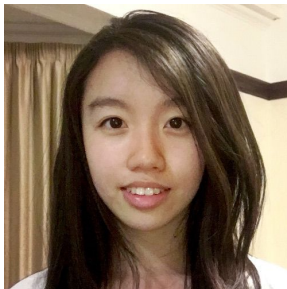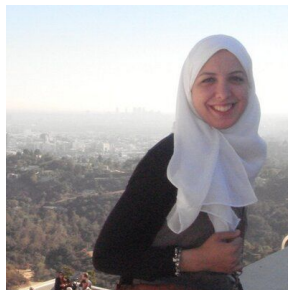# ALTA @ CL

Top: Paula, Andy, Ted

Middle: Theresa, Øistein, me, Ekaterina

Bottom: Chris, Mariano, Shawn, David, Shiva, Zheng

ALTA @ CL

Top: Chris, Mark, Youmna, Sian, Russell
Bottom: Gladys, Ahmed, Marek, Helen

ALTA Institute

- ❏ Our research strengths:
- ❏ Auto-marking written & spoken English exams
- ❏ Grammatical error detection & correction
- ❏ Complex word identification & text simplification
- ❏ Adaptivity (personalisation)

ALTA Institute

- ❏ New directions:
- ❏ Task creation & calibration (appropriate text generation at specific difficulty levels, good prompts of language display)
- ❏ Interpretable & actionable feedback
- ❏ Dialogue systems (chatbots) for language learning
- ❏ And more...

ALTA Institute

- ❏ Recent publications:
- ❏ Auto-marking speech with written transcriptions only
- ❏ Adaptive forgetting curves for spaced repetition of vocabulary flashcards
- ❏ Replication shared task: multilingual essay proficiency scoring
- ❏ Crowdsourced speech corpus
- ❏ Teacher-student chat corpus

ALTA Institute

❏ More about ALTA:
http://www.wiki.cl.cam.ac.uk/rowiki/NaturalLanguage/ALTA

CCC-NLIP

- ❏ Cambridge Cybercrime Centre X Natural Language & Information Processing
- ❏ Connection made thanks to Prof Alastair Beresford, summer 2017
- ❏ Alan Turing Institute grant: data science for cybersecurity
- ❏ Prof Paula Buttery, Drs Alice Hutchings & Sergio Pastrana

CCC-NLIP

- ❏ NLP approaches to **CrimeBB** analyses:
- ❏ Classifying function & intent of posts ('offer X', 'request X', comment, tutorial, vouch, etc)
- ❏ Characterising key actors ('characterizing Eve')
- ❏ Detecting aggressive language (quite friendly! more so than Wikipedia page edit comments)

CCC-NLIP

- ❏ Further funding from Alan Turing Institute:
- ❏ Apply NLP classifiers to all of HackForums
- ❏ New key actors (supply chain actors)
- ❏ Group-based trajectory modelling (+ Dr Ben Collier)

CCC-NLIP

- ❏ Continued co-supervision of projects...
- ❏ New EPSRC grant to Dr Alice Hutchings:
- ❏ CybercrimeNLP (CC-NLP): A natural language processing toolkit for the interdisciplinary analysis of underground online forums
- ❏ Profs Ross Anderson & Paula Buttery, Dr Ildiko Pete, Jack Hughes
- ❏ October 2020 -- April 2022

PublicHealthMT

- ❏ Global Challenges Research Fund:
- ❏ Machine translation of public health documents for low-resource languages
- ❏ Focus on African languages
- ❏ Low linguistics resources = challenging machine translation
- ❏ Attempting equitable dissemination of information for the *next* pandemic

Organisation
mondiale de la Santé

Thèmes de santé ∨   Pays ∨   Centre des médias ∨   Urgences sanitaires ∨   À propos ∨

Page d'accueil / Situations d'urgence sanitaire / maladies / Nouveau coronavirus (2019-nCoV) / Nouveau coronavirus (2019-nCoV): conseils au grand public

Maladie à coronavirus 2019 (COVID-19): conseils au grand public

Cette page est régulièrement mis à jour au fur et à mesure de l'évolution de l'épidémie et des nouvelles conclusions scientifiques. **Dernière mise à jour en date du 7 octobre 2020.**

Nouveau coronavirus (2019-nCoV)

Nouveau coronavirus (2019-nCoV): conseils au grand public ∨

Actualités ∨

Lignes directrices pour le nouveau ∨

En bonne santé à la maison

Quand et comment utiliser un masque

Soins à domicile pour les aidants

Violence à l'égard des femmes

Petits rassemblements publics

Les écoles et la COVID-19

En finir avec les idées reçues

Toutes les questions-réponses

PublicHealthMT

- ❏ Global Challenges Research Fund:
- ❏ Was due to begin 1st October
- ❏ Delayed due to COVID-19 delay in confirming GCRF funding
- ❏ £80K → 3 day R.Assis, 1 day for me
- ❏ 9 months, pump priming
- ❏ Apply for follow-on funding
- ❏ (as P.I.)
- ❏ 🤞

Thank you!

- ❏ Questions welcome
- ❏ apc38
- ❏ https://www.cl.cam.ac.uk/~apc38