



Beyond Borders, Beyond Words: A New Multimodal Corpus of L2 Academic English (UNNC-CAWSE)

Yu-Hua Chen, Simon Harrison, Michael Stevens
University of Nottingham Ningbo China



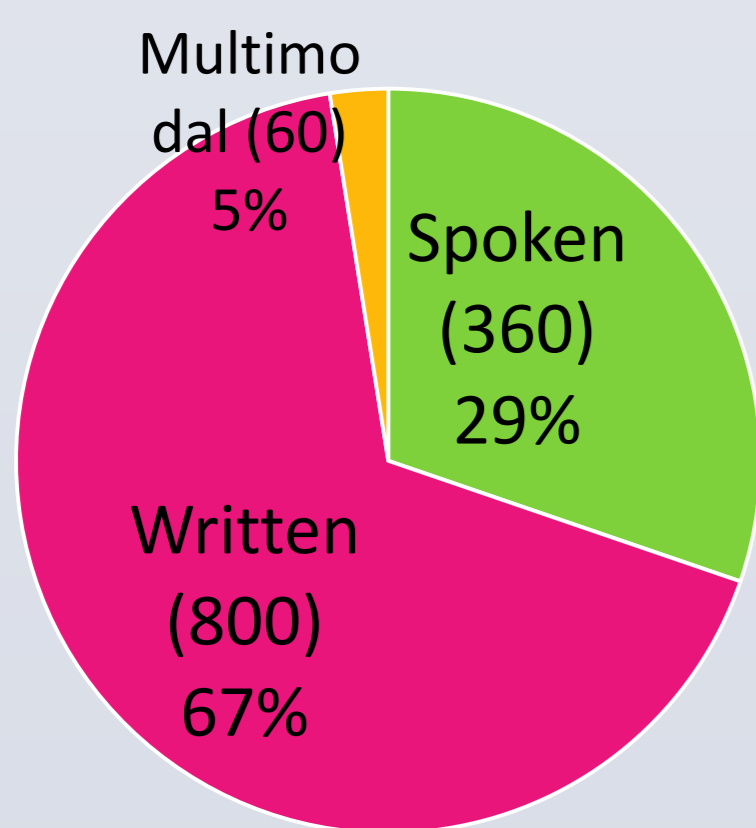
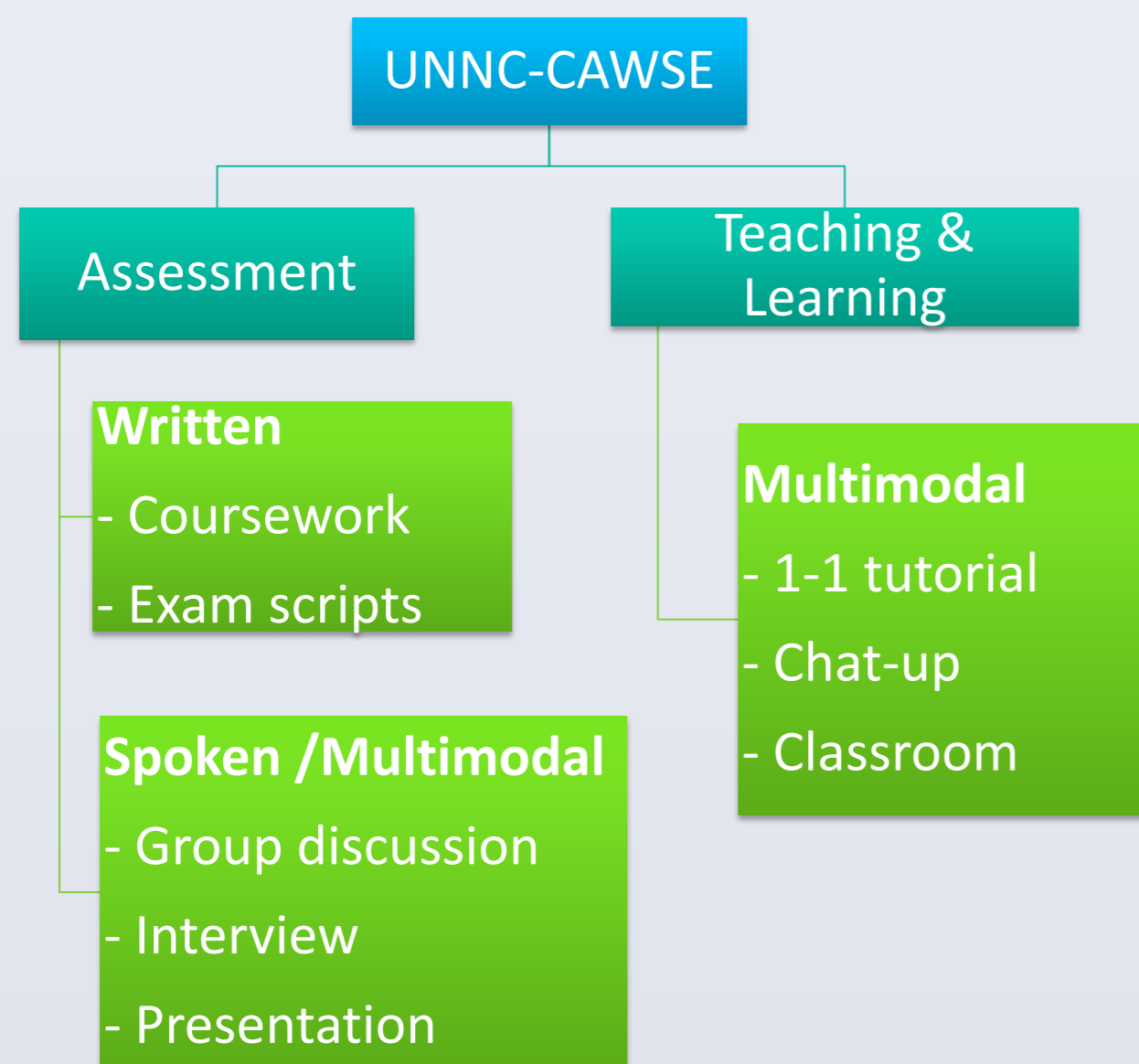
Corpus of Chinese Academic Written & Spoken English (CAWSE)

What makes CAWSE unique?

- ✓ EAP/ESP data from Preliminary Year at an EMI (English-as-medium-of-instruction) university
- ✓ Open access
- ✓ Multimodality
- ✓ Audio/video data from students (different from BASE, MICASE)
- ✓ Assessment data across all grades (Class I, II-1, II-2, III, fail)



Corpus Design



• Estimated Data to be Collected (n=scripts/speech events)

Written

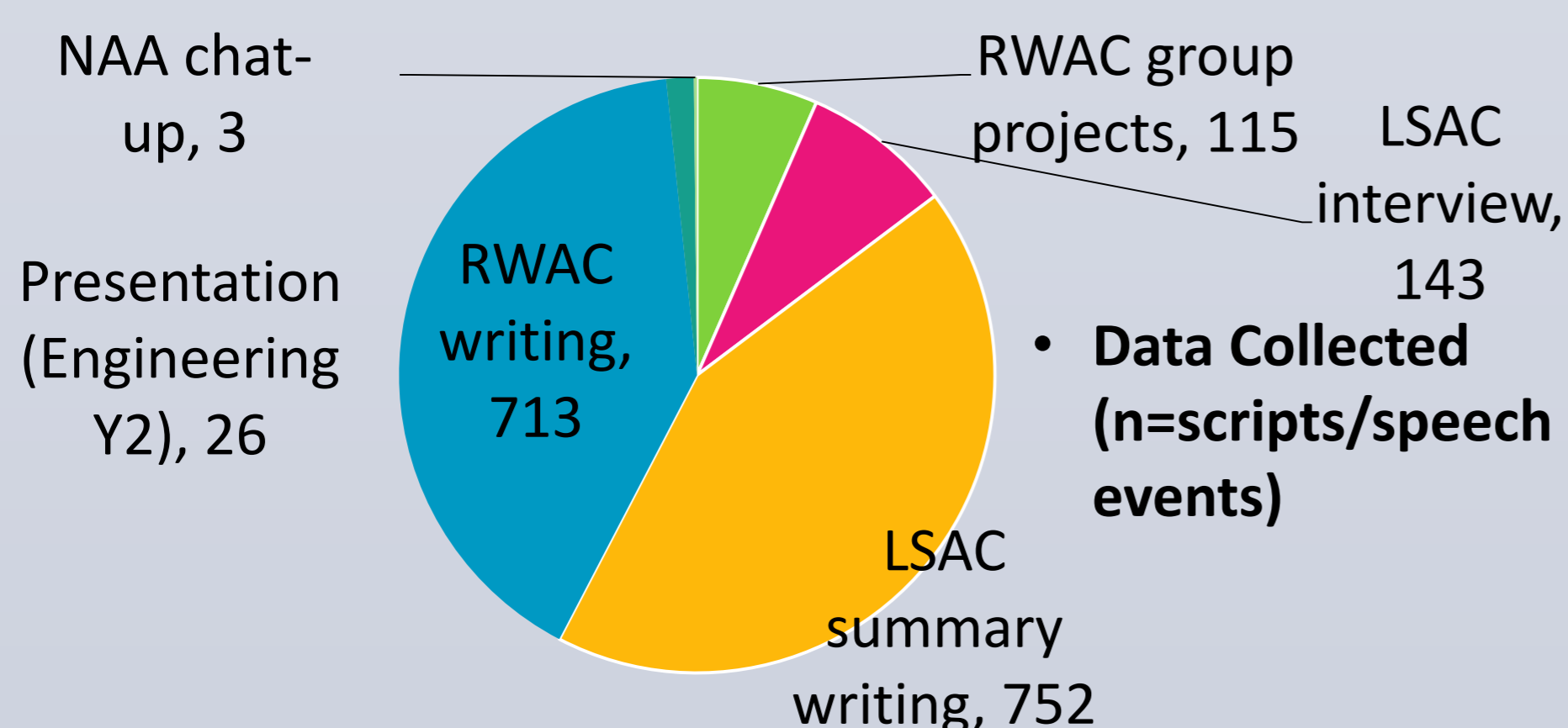
- RWAC group project, 200
- LSAC summary, 200
- ESAC exam script, 200
- ESAC IWA coursework, 200

Spoken

- LSAC Individual Interview, 120
- AOP Individual presentation, 120
- ESAC Group discussion, 120

Multimodal

- NAA chat-up, 10
- Classroom, 10
- 1-1 tutorial, 10

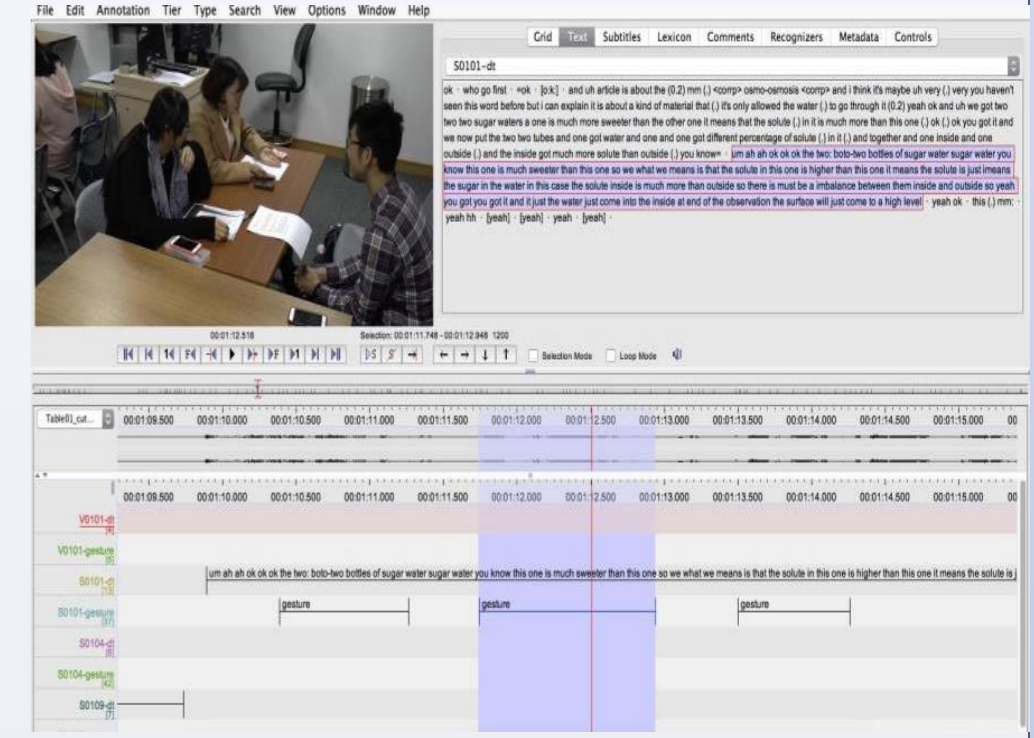


• Data Collected (n=scripts/speech events)

Challenges

Ethics

- Consent from students & staff
- Video recordings & open access - additional consent required



Transcription

- L2 English (mostly CEFR B1)

Data Type	No	No of transcription completed	Working time per event/script	Working time required
LSAC Individual Interview	143	53 (37%)	2 hours	286 hrs
LSAC summary writing (exam script)	752	85 (11%)	10 mins	126 hrs
RWAC (exam script)	713	22 (31%)	20 mins	237 hrs
NAA Chat-up	3	2 (67%)	2 hours	6 hrs
Subtotal				81.8 days
RWAC GWP (group essay)	115	-	-	-
Classroom interaction	1	-	-	-
Presentation (Engineering Year 2)	26	-	-	-

Solution

Ethics

- CELE (which delivers the Preliminary Year Program) as gatekeeper

Transcription

- Crowdsourcing (e.g. paid and voluntary interns, student dissertation projects)
- But *training and quality monitoring* required

Machine (NOT USEFUL)

- OCR or speech recognition tools accuracy rate <30%

Team



Principal investigator

Dr. Yu-Hua Chen
(School of English)



Dr. David Oakey
(Centre for English Language Education)



Dr. Simon Harrison
(School of English)



Dr. Shanru Yang
(Centre for English Language Education)



Assistant Research Fellow
Qianqian Zhou



PhD student
Michael Stevens

Acknowledgement

Ningbo Talent 3315 (CNY 1 mill = €131,587)



UNNC RKE Matched Funding (CNY 200K = €26,317)

We welcome any researchers or practitioners who wish to use CAWSE for a project and who may be able to offer some transcription in return. Please contact Qianqian Zhou (qianqian.zhou@nottingham.edu.cn).