

NESSI: A multimodal corpus of New Englishes

Gaëtanelle Gilquin
University of Louvain, Belgium

NESSI = New Englishes Student Interviews

- Multimodal corpus of informal interviews with university students using institutionalized second-language varieties of English ('New Englishes')
- So far, data have been collected for the Hong Kong subcorpus
- Meant as a collaborative international project with subcorpora representing different regional varieties of New Englishes

Multimodal corpora of non-native English

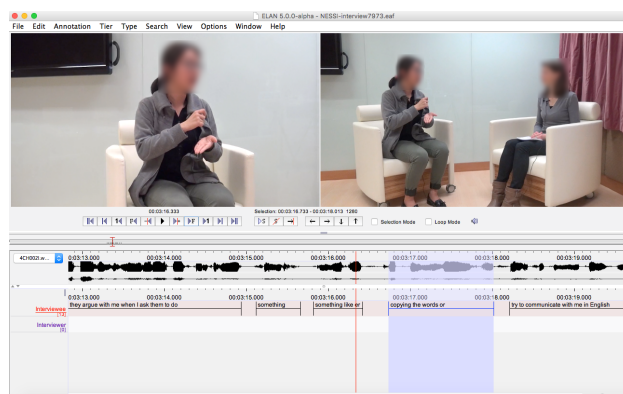
- *Multimedia Adult ESL Learner Corpus* (Reder et al. 2003), *City University Corpus of Academic Spoken English* (Lee 2008-2009), *Singapore Corpus of Research in Education* (Hong 2009), *Multimedia Learner Corpus of Basic Presentation* (Hashimoto & Takeuchi 2010), *Corpus of Academic Spoken English* (Diemer et al. 2016)
- Few such corpora, and few multimodal studies based on these

Data

- Each subcorpus contains 50 interviews, lasting c. 15 minutes each, for a total of some 15 hours
- Each interview is made up of three parts:
 1. Set topic
 2. Free discussion
 3. Picture description
- Interviewees are:
 - Non-native speakers of English
 - Born and raised in the country
 - University students in their second to fourth year, majoring in English studies
- Metadata include information about:
 - The interviewee (gender, L1, L2s, English proficiency, exposure to English, age, etc.)
 - The interviewer (gender, L1, L2s, degree of familiarity with the interviewee)
 - The interview (date, university, duration)

Data processing

- Orthographic transcription in ELAN
- Part-of-speech tagging + error-tagging
- Multimodal annotation
- CEFR rating



Advantages of NESSI

- Videos allowing for the study of body language in non-native English
- Rich metadata for detailed sociolinguistic studies
- Same design as the *Louvain International Database of Spoken English Interlanguage* (Gilquin et al. 2010) and the *Louvain Corpus of Native English Conversation* (De Cock 2004), which allows for reliable comparisons between New Englishes, learner English and native English

We are looking for international partners who would like to join the project and collect a subcorpus of their own (or another) regional variety of New Englishes.



WE WANT YOU!

Please get in touch with us!

gaetanelle.gilquin@uclouvain.be

Bibliography: DE COCK, S. 2004. Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literature New Series* 2: 225-246.
— DIEMER, S., BRUNNER, M.-L. & SCHMIDT, S. 2016. Compiling computer-mediated spoken language corpora: Key issues and recommendations. *International Journal of Corpus Linguistics* 21:3: 348-371. — GILQUIN, G., DE COCK, S. & GRANGER, S. 2010. *Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain. — HASHIMOTO, K. & TAKEUCHI, K. 2010. Multimedia learner corpus of foreigner's basic presentation in English with evaluations. *International Conference on Educational and Information Technology (ICEIT)*, Volume 2. — HONG, H. 2009. The compilation of the Singapore Corpus of Research in Education (SCoRE): A technical report for the Ministry of Education. Singapore: National Institute of Education, Nanyang Technological University. — REDER, S., HARRIS, K. & SETZLER, K. 2003. The Multimedia Adult ESL Learner Corpus. *TESOL Quarterly* 37.3: 546-557.
Acknowledgements: This project is funded by the National Fund for Scientific Research (F.R.S.-FNRS) through a Research Credit (CDR J:0221.16). Thanks are due to Meilin Chen (City University of Hong Kong), Michelle Huang (Hong Kong Baptist University), Qing Ma (Education University of Hong Kong) and Gerald Nelson (Chinese University of Hong Kong) for their help with the collection of the Hong Kong subcorpus of NESSI.
Website: <https://uclouvain.be/en/research-institutes/ilc/cecl/nessi.html>