

Developing ASR for Indonesian-English Bilingual Language Teaching

Zara Maxwell-Smith^{1,3} Ben Foley^{2,3}

1. The Australian National University, Canberra, ACT, Australia

2. The University of Queensland, Brisbane, QLD, Australia

3. ARC Centre of Excellence for the Dynamics of Language (CoEDL), Australia

Zara.Maxwell-Smith@anu.edu.au, B.Foley@uq.edu.au

Usage-based analyses of teacher corpora and *code-switching* (Boztepe, 2003) are an important next stage in understanding language acquisition. Multi-lingual corpora are difficult to compile and a classroom setting adds pedagogy to the mix of factors which make this data so rich and problematic to classify. Using quantitative methods to understand language learning and teaching is difficult work as the ‘transcription bottleneck’ constrains the size of datasets. We found that using an automatic speech recognition (ASR) toolkit with a small set of training data is likely to speed data collection in this context (Maxwell-Smith et al., 2020).

For this study we used approximately 150 minutes of data from a project recording a single teacher’s speech in a second-year, tertiary Indonesian language program. Our methodological considerations addressed the following: which ASR tool to use, how to prepare training data for this tool, and how to best manage the bias of the training data inherent in all transcription processes.

We chose the Elpis ASR system, which combines user-friendly data processing scripts with a Kaldi HMM/GMM (Hidden Markov Model/Gaussian Mixture Model) recipe. Elpis generates transcripts as time-aligned ELAN files, which was a good fit with the broader project investigating Indonesian language teaching.

A team of transcribers established guidelines which reflexively responded to a range of methodological considerations. Indonesian diglossic variants exist in a highly diverse linguistic ecosystem (Djenar and Ewing, 2015; Sneddon, 2003; Goebel, 2010). This was highlighted by transcriber subjectivity in the teaching context. For example, the task of analyzing and choosing orthography to transcribe teacher speech into over-simplified, binary L1 versus L2 categories (1st language: English, 2nd language: Indonesian) is influenced by transcriber expectations of language norms in ‘high’ vs. ‘low’ varieties of Indonesian. Further, the goal

of modifying sociolinguistic norms which brings people to language classrooms precipitated a level of variance and unpredictability unusual in other speech contexts as teachers respond to student acquisition processes. We also provided examples of the development of a *Community of Practice* (Wenger, 1998) as another layer of complexity in the group classroom environment.

The dataset was transcribed using several ‘tiers’ to create parallel structures for storing data. While predominately working from a code-switching paradigm, the data structure allowed us to train multiple models for comparative evaluation. We trained three models, two of which included all training data and multi-lingual pronunciation lexicons, resonating with work on *translanguaging* in educational settings (Garcia and Wei, 2014). The third model was trained with Indonesian data only. Our preliminary result of 64% word error rate (WER) is high in comparison to mono-lingual ASR systems (Maxwell-Smith et al., 2020). However, WERs from code-switch bilingual data (Biswas et al., 2019) were more similar to our WER, especially given our small amount of training data.

By analysing the text spans in the machine transcription, we found a high incidence of resyllabification (word splitting), particularly with omission of initial or middle consonants. The analysis identified which model would include less disruptive errors than the others, and which would be more responsive to the addition of further training data.

The application of ASR tools is limited in this setting given the small set of training data, however using these tools has potential to expedite the transcription of teacher corpora. These tools could change workflow and decrease cognitive load for human transcribers by generating a draft transcript for revision. We highlight some of the benefits and risks of using these emerging technologies to analyze the complex work of language teachers, and in education more generally.

Acknowledgements

We acknowledge the contributions of Associate Professor Hanna Suominen (The Australian National University, Canberra, ACT, Australia, Data61/Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT, Australia and University of Turku, Turku, Finland) and Simón González Ochoa (The Australian National University, Canberra, ACT, Australia).

References

- Astik Biswas, Emre Yilmaz, Febe de Wet, Ewald van der Westhuizen, and Thomas Niesler. 2019. [Semi-supervised acoustic model training for five-lingual code-switched asr](#). In *Interspeech 2019*.
- Erman Boztepe. 2003. [Issues in code-switching: Competing theories and models](#). *Issues in CS: Competing Theories and Models*, 3(2).
- Dwi Noverini Djenar and Michael C. Ewing. 2015. [Language varieties and youthful involvement in Indonesian fiction](#). *Language and Literature*, 24(2):108–128.
- Ofelia Garcia and Li Wei. 2014. *Translanguaging: Language, Bilingualism and Education*. Palgrave Macmillan, New York.
- Zane Goebel. 2010. [Identity and social conduct in a transient multilingual setting](#). *Language in Society*, 39(02):203–240.
- Max Planck Institute for Psycholinguistics - The Language Archive: Nijmegen. 2018. [ELAN \(Version 5.2\) \[Computer Software\]](#). Retrieved from <https://archive.mpi.nl/tla/elan>.
- Zara Maxwell-Smith, Simón González Ochoa, Ben Foley, and Hanna Suominen. 2020. [Applications of natural language processing in bilingual language teaching: An Indonesian-English case study](#). In *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 124–134, Seattle, WA, USA. Online. Association for Computational Linguistics.
- James N. Sneddon. 2003. [Diglossia in Indonesian](#). *Bi-jdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia*, 159(4):519–549.
- Etienne Wenger. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge.