ComputEL-8 – March 4–5, 2025

User-friendly technology distribution and the language technology development pipeline

Lorena Martín Rodríguez[1] & Christopher Cox[1,2]
[1]Carleton University, [2]Tsuut'ina Gunaha Institute
lmaro@stanford.edu • christopher.cox@carleton.ca

Carleton University

TSUUT'INA GUNAHA NASTAGHA

Acknowledgments

- *At Tsuut'ina Nation:* **Bruce Starlight**, **Janelle Crane-Starlight**, **Hanna Big Crow**; Tsuut'ina Nation Elders, language teachers, and language learners
- *In academia:* **Antti Arppe** *(University of Alberta)*, **Han Sloetjes** *(MPI Nijmegen)*
- *Supporting institutions:* **Tsuut'ina Gunaha Institute** and **Carleton University**; **Social Sciences and Humanities Research Council of Canada** *(SSHRC, Partnership Grant Partnership Grant 895-2019-1012);* **Digital Research Alliance of Canada**; **Centre for Advanced Computing**;

Social Sciences and Humanities Research Council of Canada | Conseil de recherches en sciences humaines du Canada

Canada

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Centre for Advanced Computing

Roadmap

1. Introduction
2. ELAN recognizers
3. Challenges
4. Creating installers
5. Limitations and future work

Introduction

85% of the languages spoken in Canada are endangered

Loss of Indigenous knowledge, linguistic and cultural diversity

Efforts for language documentation, revitalization, and maintenance

Our contribution: Making speech technology resources more readily accessible to language workers and communities

Aligning priorities

Colonial and harmful practices in speech and language technology

- Data-centered approaches
- Gold standards and prescriptivism
- Access to knowledge: Universality

(Bird, 2020)

Aligning priorities

Colonial practices in language documentation projects

- Lack of access from the community
- Minimal community agency
- Conflicting ideas about access
- One-sided participation of community

(Henke & Berez-Kroeker, 2016)

“Self-determination is simply the idea that Native communities can effectively interpret their past and have a right to make decisions about the trajectory of their present and future.”

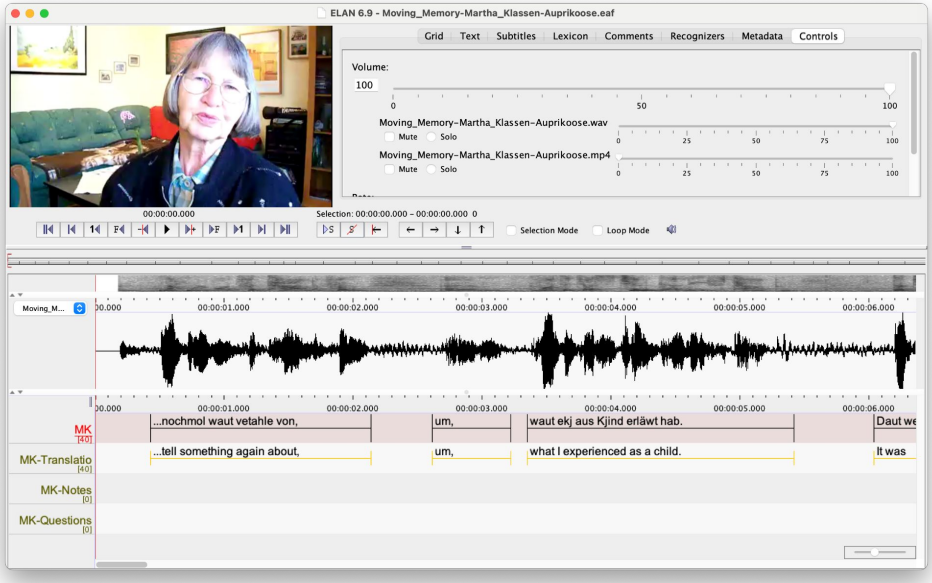
(Shepard, 2016, p. 460)

In collaboration with the **Tsuut'ina Gunaha Institute** *(Tsuut'ina Nation, Treaty 7)*:

- Active language revitalization, education, and documentation programs for **Tsuut'ina** (ISO 639-3: `srs`, Glottocode: `sars1236`) ⟶ 100+ hours of new language recordings each year
- *One goal:* Incorporate materials into a community language archive/database, so that language resources are kept safe and information is more accessible *(Martín Rodríguez, 2024)*
- **Issue:** Possible to draw on current speech technologies to help segment and transcribe language recordings for a database like this—but how to make those tools accessible to everyone involved?

Introduction: ELAN

https://archive.mpi.nl/tla/elan



Time-aligned annotation of **audio and video recordings**

- Current *de facto* standard for annotating audiovisual materials in language documentation—widespread adoption
- Open-ended annotation structure *(multiple tiers, annotation hierarchies, etc.)*
- Open-source, Java-based desktop application *(Windows, macOS, Linux)*

Closing the NLP-LDR tool gap

Our original contribution: Demonstrating the feasibility of distributing and installing current speech technologies in accessible ways—including integrating them into existing documentary software tools such as ELAN.

- **Voice activity detection** (VAD)
- **Automatic speech recognition** (ASR):
  - For Tsuut'ina *(via XLS-R)*
  - For English *(via both open- and closed-source services/models)*

Voice Activity Detection

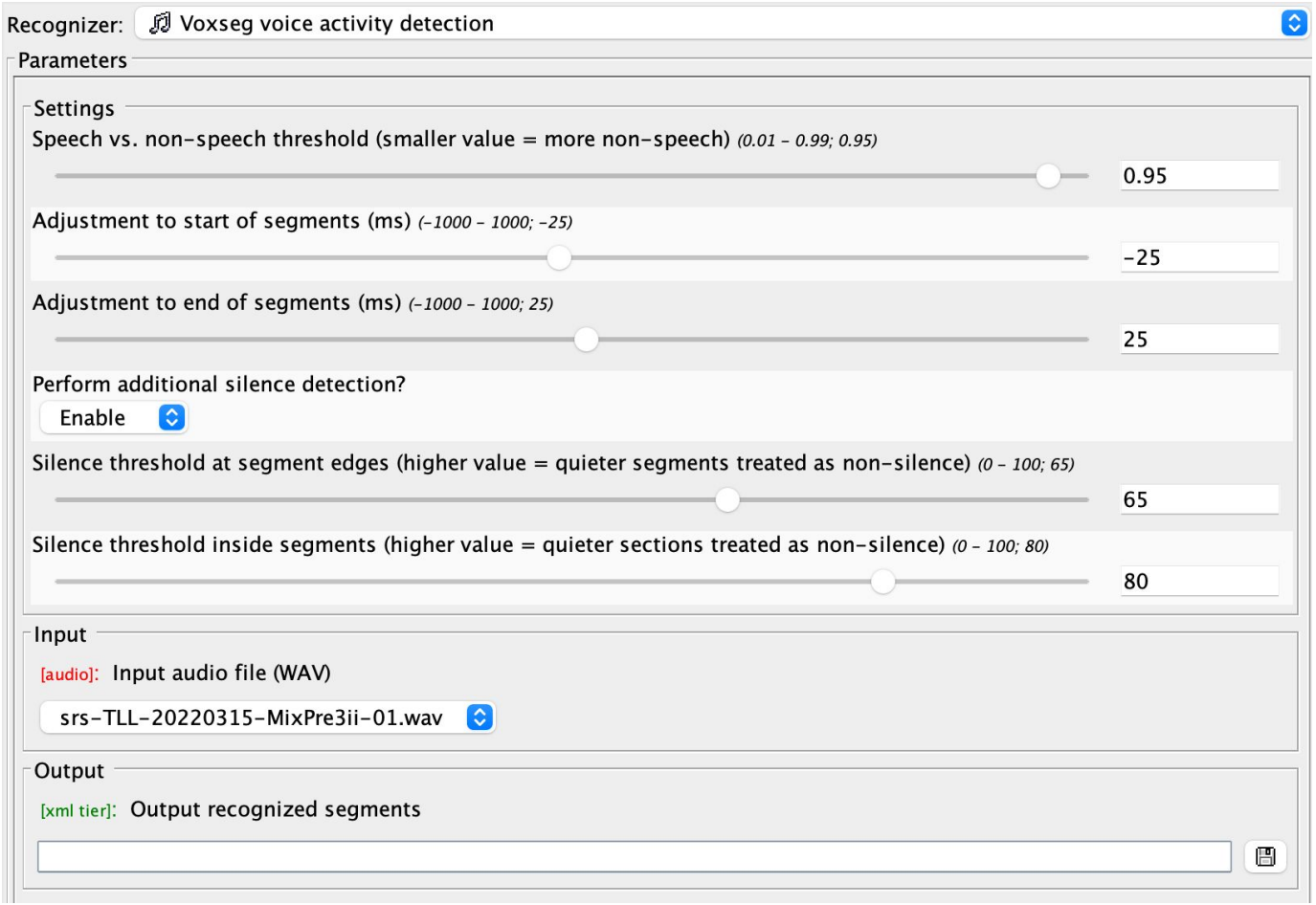
Automatic audio segmentation into speech utterances.

Integration of two DNN-based VAD models into ELAN:

1. **Voxseg-ELAN**: Voxseg voice activity detection
   https://github.com/coxchristopher/voxseg-elan
   (Wilkinson and Niesler, 2021)

2. **SileroVAD-ELAN**: Silero-VAD, pre-trained on multilingual corpora
   https://github.com/l12maro/SileroVAD-Elan
   (Silero Team, 2021)

Voice Activity Detection



User-friendly technology

"Technology that is easy to understand and use, requiring little or no training to be operated by the user."

(Isaias et al., 2019)

Existing work:

There exist approaches incorporate ASR to **user-friendly pipelines**

(Foley et al., 2018; Adams et al., 2018)

Incorporating other **speech technologies into ELAN**

(Cox, 2019, Partanen et al. 2020)

What about installation?

Challenges

Distribution using GitHub → Not user-friendly

Under macOS 12.6, the following commands can be used to fetch and install the necessary Python packages:

```
git clone https://github.com/coxchristopher/voxseg-elan
cd voxseg-elan

python3 -m virtualenv venv-voxseg
source venv-voxseg/bin/activate

git clone https://github.com/NickWilkinson37/voxseg.git
pip install ./voxseg
pip install pydub tensorflow
```

command line

Once all of these tools and packages have been installed, Voxseg-ELAN can be made available to ELAN as follows:

1. Edit the file `voxseg-elan.sh` to specify (a) the directory in which ffmpeg is located, and (b) a Unicode-friendly language and locale (if `en_US.UTF-8` isn't available on your computer).

2. To make Voxseg-ELAN available to ELAN, move your Voxseg-ELAN directory into ELAN's `extensions` directory. This directory is found in different places under different operating systems:

   - Under macOS, right-click on `ELAN_6.4` in your `/Applications` folder and select "Show Package Contents", then copy your `Voxseg-ELAN` folder into `ELAN_6.4.app/Contents/app/extensions`.
   - Under Linux, copy your `Voxseg-ELAN` folder into `ELAN_6-4/app/extensions`.
   - Under Windows, copy your `Voxseg-ELAN` folder into `C:\Users\AppData\Local\ELAN_6-4\app\extensions`.

Folder navigation

Challenges

Distribution using GitHub → Not user-friendly

Technical requirements (e.g. Python and Python packages) need to be installed by the user

Our aim: Using installers as user-friendly distribution methods for ELAN recognizers

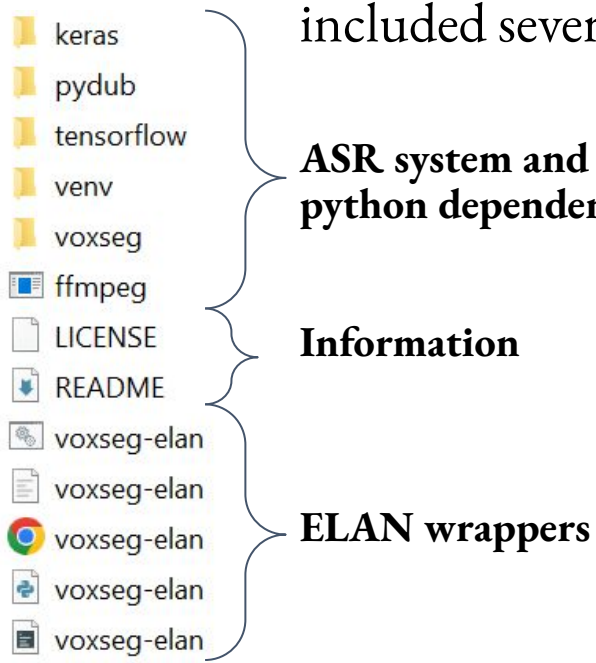
- **Voice activity detection** (VAD)

Creating Installers

Two main steps:

- **Wrapping recognizers as executable files (`.exe`)**
- **Distributing executables using installers**

Creating executable files



Original recognizers included several files:

ASR system and python dependencies

Information

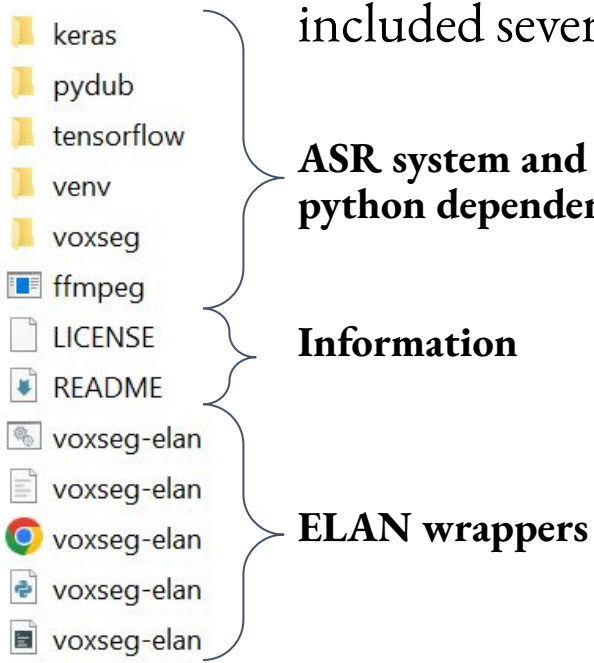
ELAN wrappers

Problem:

- Installation of all Python dependencies falls on user
- Requires some technical background for the installation of packages

Creating executable files

Original recognizers included several files:

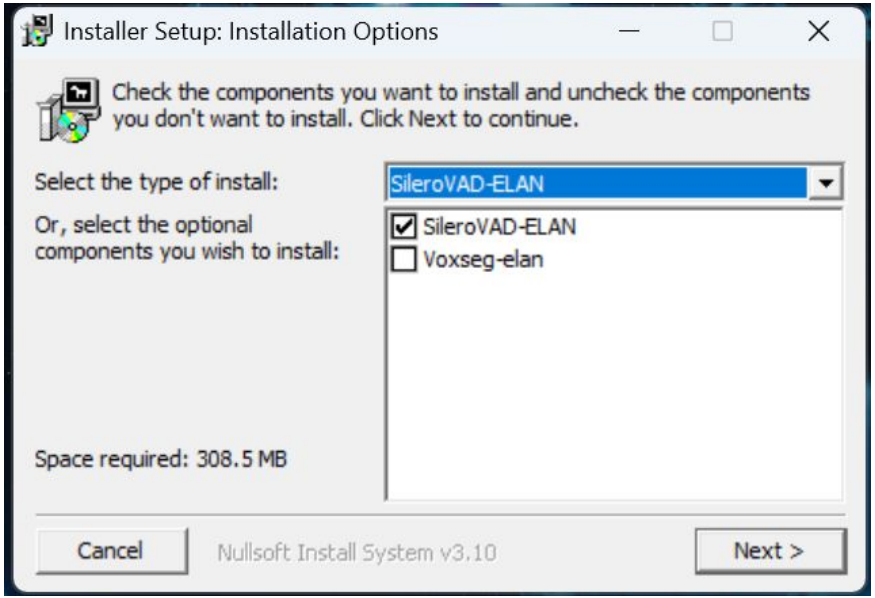


Solution:
Executable files that contain all files and Python dependencies

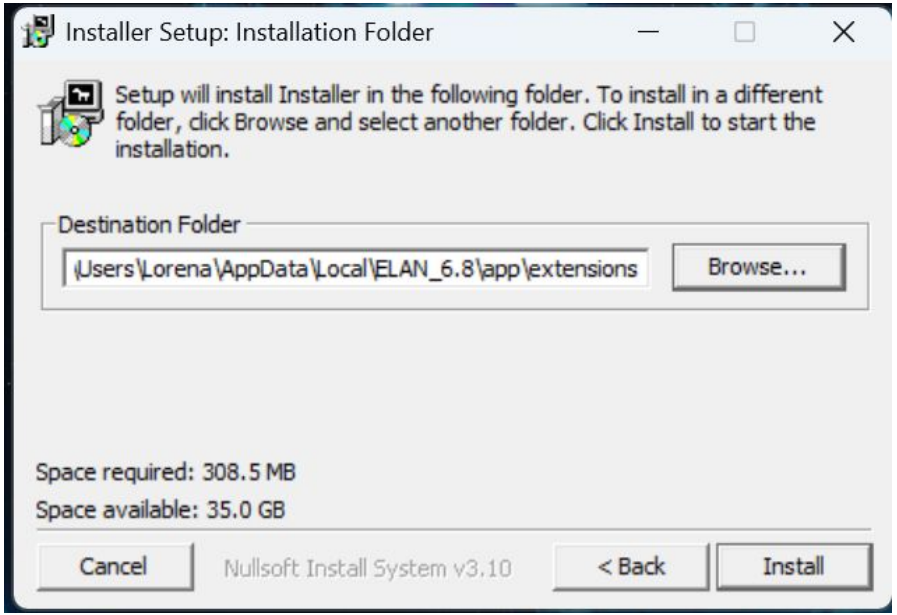
Windows distribution (NSIS)

Features:

1. **Multiple components**



2. **Detect ELAN location**



Current progress and future directions

Windows installers that make Silero VAD available in ELAN (6.4–6.8) are available for download on a Shared Drive

https://github.com/l12maro/SileroVAD-Elan

Expand user-friendly installation to

- Target **other OS** *(macOS, Linux)*
- Include **other speech technologies** *(e.g., automatic speech recognition, speaker and language diarization, etc.)*

References

Steven Bird. 2020. Decolonising Speech and Language Technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Christopher Cox. 2019. *Persephone-ELAN: Automatic phoneme recognition for ELAN users*. Version 0.1.2. https://github.com/coxchristopher/persephone-elan.

Ben Foley, Joshua T. Arnold, Rolando Coto-Solano, Gautier Durantin, T. Mark Ellison, Daan van Esch, Scott Heath, František Kratochvil, Zara Maxwell-Smith, David Nash, Ola Olsson, Mark Richards, Nay San, Hywel M. Stoakes, N. Thieberger, and Janet Wiles. 2018. Building speech recognition systems for language documentation: The CoEDL Endangered Language Pipeline and Inference System (ELPIS). In *Proceedings of the 6th International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU 2018)*, pages 205–209. Gurugram, India: ISCA. https://www.isca-archive.org/sltu_2018/foley18_sltu.pdf.

Pedro Isaias, Paula Miranda, and Sara Pífano. 2019. Higher education and Web 2.0: Barriers and best practices from the standpoint of practitioners. In Jean-Éric Pelet (ed.). *Advanced Web Applications and Progressing E-Learning 2.0 Technologies in Higher Education*. IGI Global. https://doi.org/10.4018/978-1-5225-7435-4.

Luke Gessler. 2022. Closing the NLP Gap: Documentary Linguistics and NLP Need a Shared Software Infrastructure. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 119–126. Dublin, Ireland: ACL. https://doi.org/10.18653/v1/2022.computel-1.15.

Ryan Henke and Andrea L. Berez-Kroeker. 2016. A Brief History of Archiving in Language Documentation, with an Annotated Bibliography. *Language Documentation & Conservation 10*. 411-457.

References

Lorena Del Rocío Martín Rodríguez. 2024. Collaborative Interface Design and Language Archiving: The Case of Tsuut'ina. Unpublished MA thesis. Ottawa, Canada: Carleton University. https://doi.org/10.22215/etd/2024-16145.

Brigitte Pakendorf, Nina Dobrushina, and Olesya Khanina. 2021. A typology of small-scale multilingualism. *International Journal of Bilingualism*, 25(4):835–859.

Niko Partanen, Mika Hämäläinen, and Tiina Klooster. 2020. Speech recognition for endangered and extinct Samoyedic languages. In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, pages 523–533, Hanoi, Vietnam: ACL.

Michael A. Shepard. 2016. The value-added language archive: Increasing cultural compatibility for Native American communities. *Language Documentation & Conservation, 10*, 458–479. http://hdl.handle.net/10125

Silero Team. 2021. Silero VAD: pre-trained enterprise-grade Voice Activity Detector (VAD), number detector and language classifier. https://github.com/snakers4/silero-vad.

Sebastian Tschöpel, Daniel Schneider, Rolf Bardeli, Oliver Schreer, Stefano Masneri, Peter Wittenburg, Han Sloetjes, Przemek Lenkiewicz, and Eric Auer. 2011. AVATecH: Audio/Video Technology for Humanities Research. In *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*, pages 86–89, Hissar, Bulgaria: ACL.

Nicholas Wilkinson and Thomas Niesler. 2021. A hybrid CNN-BiLSTM voice activity detector. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6803–6807.