

## Image-aligned annotation in language documentation and revitalization

In many language documentation projects, still-image data (e.g., scans of historical manuscripts or more recent handwritten field notes, published or unpublished typeset materials, photographs of signage and other instances of public language use, etc.) constitute one important source of linguistic and cultural information (Stebbins & Hellwig 2010, Baldwin et al. 2016, Lukaniec et al. 2021). While incorporating such materials may significantly broaden the scope of a documentary collection, drawing on their contents often requires documents to be manually transcribed, which can present a time-consuming and costly barrier to their use in documentation. At the same time, still-image materials also pose a number of challenges for current digital methods and tools in language documentation, which have tended to focus on working with audiovisual, rather than written, records of language practices and knowledge.

In this presentation, we share our experiences in the context of language documentation and revitalization with two freely available, cloud-based software tools that aim to improve the accessibility of still-image sources through image-aligned annotation. In both tools, collections of still images are presented through a web-based interface to invited contributors to transcribe, correct, and interpret, aligning text and commentary with the scanned image at either the page or sentence level. We discuss in practical terms what is required to use both systems (in terms of software requirements, document formats, and training); how features of both tools have assisted in recent collaborative projects involving historical handwritten Indigenous language materials; and how annotated documents can be shared and brought into other language initiatives. While collaborative, image-aligned transcription tools such as these are not yet common in language documentation and revitalization, we suggest that they may serve a valuable role in this context, helping both to open up such language materials' contents to a broader range of uses and to facilitate the involvement of a wider community of contributors in language work.

## References

- Baldwin, Daryl, David J. Costa & Douglas Troy. 2016. Myaamiaataweenki eekincikoonihkiinki eeyoonki aapisaataweenki: A Miami language digital tool for language reclamation. *Language Documentation and Conservation* 10. 394–410. <http://hdl.handle.net/10125/24713>.
- Lukaniec, Megan, Sally Cartwright, Martin Holmes, Erin Hashimoto, Emma Pearce-Marvell, Jamie Quibell & Bethany Scholfield. 2021. Using open-source, sustainable tools for language reclamation: Preliminary findings from building a digital corpus of Wendat. Paper given at the 7<sup>th</sup> International Conference on Language Documentation and Conservation (ICLDC7), University of Hawai'i at Mānoa, online, March 4–7, 2021. <https://youtu.be/1A0G-kG79m4>.
- Stebbins, Tonya N. & Birgit Hellwig. 2010. Principles and practicalities of corpus design in language retrieval: Issues in the digitization of the Beynon corpus of early twentieth-century Sm'algyax materials. *Language Documentation and Conservation* 4. 34–59. <http://hdl.handle.net/10125/4466>.