- You want to get information or provide input about this research.

Please press “Next” below ONLY if you consent to participate in this study in order to begin the survey.

Background

In this survey, we are currently only targeting researchers who have or are working towards a graduate degree **and** have published on a system that was intended for use on either low-resource languages or in language documentation settings. **Please continue in this survey only if you meet these criteria**.

1. Do you currently have or are you working toward a graduate-level degree (e.g. MS, MA, PhD, …) in computer science, computational linguistics, NLP, or a related field? *

   Mark only one oval.

   - ◯ Yes
   - ◯ No

2. How many times have you published academic works in a venue for natural language processing? *

   These include, but are not limited to, those run by the Association for Computational Linguistics and other conferences such as NeurIPS

   Mark only one oval.

   - ◯ 0
   - ◯ 1-2
   - ◯ 3-5
   - ◯ 5-10
   - ◯ >10

3. If you did not answer "0" to the previous question, please list a couple of the conferences your work has appeared in. Abbreviations are fine.

4. Has any of your published work focused on serving “low-resource” languages as a primary goal? A language is “low-resource” if it has much less digital text data available than a language like English or Arabic. *

   Some other keywords that partially overlap with “low-resource” include “under-resourced”, “endangered”, “indigenous”, and “less-studied”.

   Mark only one oval.

   - ◯ Yes
   - ◯ No

5. Has any of your published work been primarily intended for application in the documentation of languages? For example, a system that is intended to help a linguist transcribe speech data that they collected in the field would qualify. *

   Mark only one oval.

   - ◯ Yes
   - ◯ No

Skip to question 6

General Information

Language documentation is the process of studying a language in order to produce products such as dictionaries, grammars, and documents. This work is typically conducted by a linguist in partnership with the language's community. The process involves many labor-intensive steps such as speech transcription, translation, and morphological segmentation and analysis.

In the context of this study, we are interested in a specific segment of NLP research, which we will call **"language documentation NLP" (LD-NLP)**. We use this to refer to NLP research that is intended or suitable for the challenges that arise in language documentation. Typical challenges in this domain include severe data constraints (a very large corpus might have a couple hundreds of thousands of tokens, and a more typical one would have tens of thousands) and computational resource constraints. Work in **low-resource NLP is therefore often relevant to LD-NLP, but work in other areas may also be relevant**.

6. In what year did you first publish work in LD-NLP? *

_____

7. In what year did you most recently publish work in LD-NLP? *

_____

8. Is LD-NLP a part of your ongoing research/work interests? *

Mark only one oval.

- ◯ Yes
- ◯ No

9. What country do you live in? (This question is optional.)

_____

10. What country were you born in? (This question is optional.)

_____

11. What is the name of the organization that you work for? (This question is optional.)

_____

12. What is your employment status? *

Mark only one oval.

- ◯ Graduate student
- ◯ Postdoc
- ◯ Tenure-track faculty member
- ◯ Non-tenure-track faculty member
- ◯ Other academic position
- ◯ Other non-academic position
- ◯ Unemployed

13. Which categories does your work in LD-NLP fall into? Select up to three. *

Check all that apply.

- [ ] Computational Social Science and Cultural Analytics
- [ ] Dialogue and Interactive Systems
- [ ] Discourse and Pragmatics
- [ ] Efficient/Low-Resource Methods for NLP
- [ ] Ethics, Bias, and Fairness
- [ ] Generation
- [ ] Information Extraction
- [ ] Information Retrieval and Text Mining
- [ ] Interpretability and Analysis of Models for NLP
- [ ] Linguistic theories, Cognitive Modeling and Psycholinguistics
- [ ] Machine Learning for NLP
- [ ] Machine Translation
- [ ] Multilinguality and Language Diversity
- [ ] Multimodality and Language Grounding to Vision, Robotics and Beyond
- [ ] NLP Applications
- [ ] Phonology
- [ ] Morphology and Word Segmentation
- [ ] Question Answering
- [ ] Resources and Evaluation
- [ ] Semantics: Lexical
- [ ] Semantics: Sentence-level Semantics, Textual Inference and Other areas
- [ ] Sentiment Analysis, Stylistic Analysis, and Argument Mining
- [ ] Speech recognition, text-to-speech and spoken language understanding
- [ ] Summarization
- [ ] Syntax: Tagging, Chunking and Parsing

14. Has your work in LD-NLP happened in the context of a relationship with a particular language community that has lasted for more than one publication cycle? *

Mark only one oval.

- ◯ Yes
- ◯ No

15. If yes, and you are comfortable sharing this information, please specify the language communities you have worked with in this way.

16. Have you ever worked with a documentary linguist (someone with a graduate degree in linguistics or a related discipline who has language documentation as a primary research interest) as a co-author on any of your LD-NLP work? *

Mark only one oval.

- Yes
- No

NLP in Language Documentation

Since at least 15 years ago, both documentary linguists and NLP researchers have noted the potential for NLP algorithms to expedite documentary linguists' work. For example, speech recognition systems can produce transcripts of recorded audio, and correcting these machine outputs could be quicker than transcribing audio manually. Similar interactions are also possible for most other tasks in the documentary process, such as translation, morphological analysis, and syntactic analysis.

17. Before this survey, were you aware of documentary linguists and the kind of work they do? *

Mark only one oval.

- Yes
- No

18. To the best of your knowledge, how much do you think NLP has already impacted LD work? *

Mark only one oval.

- NLP has had little impact potential
- NLP has had limited impact in certain settings
- NLP has had moderate impact in a few settings
- NLP has had sizeable impact in many settings
- NLP has had great impact in many settings

19. To the best of your knowledge, how would you describe your outlook on the potential of NLP to impact LD in the future? *

Mark only one oval.

- NLP has little potential
- NLP has limited potential in certain settings
- NLP has moderate potential in a few settings
- NLP has sizeable potential in many settings
- NLP has great potential in many settings

20. Has any of your work ever been applied in a LD project? *

Mark only one oval.

- Yes *Skip to question 21*
- No *Skip to question 31*

Past NLP System Use

For these questions, please consider **one** NLP system which you used in a project of yours, and answer all questions in relation to that system.

21. Please describe the system in a few sentences. *

22. Please describe the project the system was used in in a few sentences, in as much detail as you are willing to share about the language, the people involved in the documentation effort, and so on. *

23. Were you or your coauthors (for the work associated with the system) primarily responsible for practical operation of the system for the project? *

Mark only one oval.

- ◯ Yes
- ◯ No

24. If no, who was responsible for the operation of the system?

25. If you and your coauthors were responsible for the operation of the system, do you think others involved in the project would have been able to set the system up without you?

Mark only one oval.

- ◯ Yes
- ◯ No

26. What kind of software was being used for the primary language documentation activities of the project? *

Mark only one oval.

- ◯ Purpose-specific software such as FLEx, ELAN, Toolbox, or something else
- ◯ General-purpose software such as word processors or spreadsheet applications such as Excel
- ◯ No software, and instead an analog medium such as pen and paper

27. How did data flow between the NLP system and the software being used for the project? *

Mark only one oval.

- ◯ The software was able to send data NLP system and receive data back with minimal or no user intervention required
- ◯ The user had to manually provide the NLP system with some data (by copy pasting, uploading a file, or some other action), and then manually put the output back into the software system.
- ◯ Other: _____

28. From your perspective, how satisfied were the people on the project by the system relative to the initial and ongoing investment of resources into its use? *

Mark only one oval.

- They felt the system had held the project back significantly
- They felt the system had some merit but on the whole was not a good investment
- They felt the system provided benefit equal to the investment
- They felt the system provided moderate benefit in excess of what was invested
- They felt the system provided benefits well beyond what was invested
- I don't know

29. In terms of its ability to make the documentary work more efficient, how did your expectations for the system match up with reality? Please write a few sentences. *

30. How much work was it to set up the system compared to what you expected? Please write a few sentences, or write n/a if you were not involved in this part of the process. *

Skip to question 38

Prospective NLP System Use

31. Please choose one LD-NLP system you have created and describe the system in a few sentences. *

32. Please describe situations, either in terms of kinds of languages or circumstances which are relevant to language documentation, in which you think this system would best perform. *

33. How difficult do you think it would be for a documentary linguist to set your system up for use in their work? *

Mark only one oval.

- ◯ Quite difficult: it would require much technical knowledge and computational resources that they would not have
- ◯ Fairly difficult: it would require some technical knowledge that is beyond their training, and maybe require some computational resources beyond what they normally have
- ◯ Doable: technical knowledge required would be mostly within their training, and I do not expect they'd need additional computational resources
- ◯ Fairly easy: I'd expect them to have most of the technical knowledge they need already, and their current computational resources would be adequate
- ◯ Very easy: adequate technical knowledge and computational resources would certainly be available to them.

34. What kind of software system do you think potential users of your system would most likely be using in their core workflows? *

Mark only one oval.

- ◯ Purpose-specific software for language documentation such as FLEx, ELAN, Toolbox, or something else
- ◯ General-purpose software such as word processors or spreadsheet applications such as Excel
- ◯ No software, and instead an analog medium such as pen and paper

35. How do you expect data would flow between your NLP system and the software being used for the project? *

Mark only one oval.

- ◯ The software would send data to the NLP system and receive data back with minimal or no user intervention required
- ◯ The user would manually provide the NLP system with some data (by copy pasting, uploading a file, or some other action), and then manually put the output back into the software system.
- ◯ Other: _____

36. If a documentary project adopted your system, then on average, how much benefit would you expect the project to gain (in the form of time saved) relative to the investment required to set up and integrate your system into the project's workflows? *

Mark only one oval.

- ◯ The system would not be worth the investment
- ◯ The system would had some merit but on the whole would not be a good investment
- ◯ The system would provide a benefit roughly equal to the investment
- ◯ The system would provide moderate benefit in excess of what was invested
- ◯ The system would provide benefits well beyond the investment

37. If a documentary linguist approached you and was interested in a research collaboration where you would develop and maintain systems that assists you in their documentary work, how interested would you be in working with them? *

Mark only one oval.

- ◯ Not at all interested
- ◯ Somewhat uninterested
- ◯ Neutral
- ◯ Fairly interested
- ◯ Very interested

Follow-ups

38. We are also planning to conduct some short (30 to 60 minute), one-on-one interviews with survey respondents. Are you interested in participating? *

Mark only one oval.

- ◯ Yes
- ◯ No

39. May we contact you with clarificatory follow-up questions about some of your responses to this survey? *

Mark only one oval.

- ◯ Yes
- ◯ No

40. Please provide your email if you responded “Yes” to either question.

_____

41. Is there anything else you'd like to add to your responses?