The Observer's Paradox: Some forms of skewing that can occur in language documentation, and some possible ways to mitigate them

In hard science, the observer's paradox states that the means of perception itself affects the perceived action. For example in physics, the light used for visible perception can affect the physical state of the phenomenon being observed.

Language documentation has as a primary goal a record of natural speech, springing from a desire for an accurate representation of the speech of a language community.

However, there is often a tension, long well-known to translators, between accuracy and naturalness.

Consider the differences between casual speech and slow speech – which is a more accurate record? Which is more natural?

The more the researcher manipulates the situation to achieve precisely the desired type of data, the more unnatural the data may be, in various ways.

In this paper I offer some practical suggestions mostly derived from my experience with language documentation on a small scale in several different ways for a few years.

Within a language documentation project, skewing can enter in during the preparation phase, wordlist elicitation, the recording sessions themselves, and also regarding the metadata annotation and the archiving.

The preparation phase includes aspects such as: choice of a language area for documentation, bibliographic research, and determining scope.

While eliciting the wordlist, the researcher can try to avoid the lexical gap, and to increase accuracy by using context and frames and by collecting more extensive phonetic data.

Other important considerations related to corpus selection are: emic vs. etic perspectives, generational differences, and discourse selection.

The technical quality of an audio or video recording can be improved by using several on-site techniques.

Metadata collection procedures can affect others' interpretation of the text, so it's best practice to be sure to maintain enough information about the speaker, the audience, and the text.

Other details include the usefulness of an oral interview for obtaining duly informed consent, and techniques to mitigate the effect of the researcher's mere presence.

Finally, considerations of appropriate access and some tips related to archiving and distributing conclude the paper.

This paper also includes a number of practical appendices: a metadata questionnaire on the speaker's personal data, a sociolinguistics survey, a basic interview for gaining oral informed consent, an overview of the "Dictionary Development Process" by Ron Moe, and an annotated bibliography of some field guides.