The Observer's Paradox:

Skewing and some ways to mitigate it in language documentation

As derived from Heisenberg's Uncertainty Principle from the field of physics:

The observer's paradox states that the presence of the observer changes the condition of the observed.

Researcher as audience:

If a researcher serves as the audience for a speaker, and the researcher is a non-speaker of the target language, the speaker is very likely to speak slower, or possibly louder, or even interject phrases in a shared language.

There is an inherent tension between naturalness and precision. The more a researcher attempts to control for accurate data, the more the speaker tends to adjust for an unnatural situation.

This is often so in the field of language documentation. There are examples too numerous to mention.

The speaker may even use a restricted register. Any of this distorts the naturalness of the speech.
In the case of “researcher as audience”, there can be a relatively simple solution: have another speaker of the language present to serve in that role. The researcher is then free to be in a more inconspicuous location.

This may sound simple, but often it is experience that reveals a problem, and sometimes after all the data has already been collected. Learning the hard way can be costly.

Descriptions elicited from local photos should be fairly natural, right? Except that in every photo, a new participant is presented, so the syntax can be skewed!

The process itself can condition the results, affecting the reality of the results achieved by the investigator.

This disconnect, between the way people naturally speak and the results of a language documentation project, is what I'm calling “skewing”.

Awareness is key to addressing the issue.
There are many tips, like “be sure to have another speaker present”, but the scope of this presentation can't include them all.

Also, each situation is different, and it is not possible to account for all the variables beforehand.

So it is necessary to focus on **principles** that can generate the tips.

Four principles can help mitigate skewing; in the project plans, the data collection, and analysis and archiving.

- The researcher can adopt the attitude or role of a learner in the target language and culture (even if it is his/her own).

- The researcher can try to organize the documentation project including the **perspective of the insider**, not just from an outsider's perspective.
The researcher can “over”-document the metadata, to allow for lack of identification of relevant cultural concerns.

The researcher can leave complete records of the documentation with the language group concerned.

Attitude of a learner: This includes an initial study of the bibliography available on the language, if possible, or on the language family if not, and contact with others there.

Field worker John C. Street noted (1963), “it was only after studying related idiolects and dialects that a full awareness dawned on me of the vital importance in linguistic field work of some background in and feeling for the family of languages involved. (It is appalling how many hours one can waste working on problems already solved by others.)”

Within the community, if the researcher expresses willingness and ability to learn, perhaps even just a few simple greetings,
that can open the door for explanations from insiders in a way that is not likely to happen with a “teacher” or “scientist”.

The attitude of a learner can promote getting an insider's perspective:

To quote, “It's hard to fill a cup that is already full.”

The process of oral transcription (another speaker repeating slowly and phrase-by-phrase the words of the collected text) brought out a sociolinguistic distinction

which needed to be considered in corpus collection and metadata annotation.

A Zapotec man listening to the speech act of a woman, which he was supposed to repeat slowly and clearly phrase by phrase, said “I can't say that! Men don't speak that way.”

That was the evidence that indicated a gender distinction both in vocabulary and syntax, that would never have been in evidence naturally were I present.
Men speaking to men, with no women listening, use a third-person male pronoun for each other generally (regardless of age), that is only used by women when speaking of an adolescent male.

The parameters that condition a speech act can be quite complex, so including “extraneous” information in the metadata can help avoid leaving out something culturally relevant to the speaker.

Consider this scenario:
The researcher carefully notes the speaker and the audience for the metadata . . .

but misses the fact that the speaker's mother-in-law is within hearing distance of the speech act, say in the cooking area at the other end of the room.

If the language in question is one in which certain words are taboo in the presence of a man's mother-in-law, that is relevant information about the speech act.

Since it is difficult to know in advance all the relevant aspects related to distinctions in the speech acts,
therefore collecting extensive metadata can help to alleviate the biases created by the documentation process itself.

In video recording, simply panning the camera to record 360 degrees of the conditions, both before and after the speech act, can help in noting relevant details, even if the researcher is not aware of them at the time.

Even 360-degree video, however, cannot show if a woman is a man's mother-in-law, for example, without the researcher's prior knowledge of that relationship.

Leaving complete records of the documentation, besides being courteous, can also promote feedback, if people in the community perceive obvious errors or omissions.

An example of corrective feedback occurred when I was eliciting a wordlist in a relatively moribund language with only elderly speakers left,
and a passive speaker, overhearing the process, noted that the partially deaf speaker had just given me the word for “cat” instead when I had asked for “bat”!

For tips on mitigating skewing in specific areas of the language documentation process, including project framing and planning, wordlist elicitation, text collection, recording, metadata, etc., I refer you to the many field worker's guides, both ancient and modern, as well as to the full text of my paper.

At the beginning of the discipline of linguistics, field work was highly valued. Many helpful field manuals were prepared early on.

With an increased emphasis on universals, there was an increased emphasis on theory and a decreased emphasis on field work.

In recent years, that has changed again. So there are now a number of good field manuals recently published.
Also, in some cases, manuals directed to language learners can have very helpful tips about being a learner in another language and culture, if that's the case.


They point out that the personality and attitudes of the researcher inevitably enter into how the process is carried out, which is also well worth acknowledging in language documentation.

In conclusion, skewing to some degree will always be present in a language documentation project, but awareness on the part of the researcher, along with some specific principles to implement, can help to decrease the likelihood of that error being very significant.

Also, the more that an insider perspective is allowed to inform the whole process,
including but not limited to, the understanding of sociolinguistic distinctions which need to be covered in the data collection, the more that natural results which more clearly represent the actual language can be recorded.

Furthermore, sometimes the presence of the researcher and the documentation process can give prestige to what may be a deprecated language, and help to mitigate language decline, especially if the community is involved in the process and presented with tangible results.

To paraphrase Rubin and Thompson, “[Documenting] a language is a complex undertaking, but if you know more about the process and about your own approach to the task, it can be a very rewarding, enriching, and enjoyable experience. . .” (p. 117), both for the researcher and for the speech community.