Elicitation vs. corpus data: how about positive evidence?

Katharina Haude, CNRS/SeDyL, Villejuif

In the description of a language, the aim is to rely as much as possible on natural discourse data rather than on elicitation. The reason is that elicitation data often yield artificial results which do not reflect speech as it is used in daily communication. For instance, speakers often consider ungrammatical constructions that are nevertheless found in spontaneous discourse; elicited negative evidence, therefore, does not say much about the actual grammar of the language.

However, what happens in the reversed case, when in elicitation, speakers produce or accept patterns that are not found in recorded spontaneous discourse? When describing the grammar of small, endangered and poorly documented languages, one is faced with a potential problem when relying on corpus data alone: the size of the discourse corpus is generally relatively small, which means that not all grammatically possible patterns may be attested. In order to find out about the existence of particular patterns without using translational equivalents is to use stimuli that can yield natural responses. However, it is not always possible to use stimuli with speakers of an endangered language community, who are typically old and few in number. In this case, translation equivalents in oral elicitation may be the only means to find out about possible, though unattested, grammatical patterns.

The example I will present is from Movima, an endangered linguistic isolate from lowland Bolivia spoken by a few hundred elderly speakers. Movima has two transitive constructions (so-called “direct” and “inverse”), and the research question was which factors determine the choice between the two in the third-person domain. A count of the text corpus seems to show clearly that animacy plays a central role in this choice: the inverse construction is always used when an inanimate entity acts on a human being, but it is never used when the agent is higher in the animacy hierarchy than the patient (e.g. a human acting on an animal). However, in elicitation, speakers accept the unattested patterns. In addition, the Movima corpus is rather small (< 140,000 words). Scenarios with an inanimate agent acting on a human third-person patient constitute less than 1% of the total of transitive constructions in this corpus, so that the exclusive use of the inverse construction here could be accidental. In order to find out if animacy is indeed a determining factor, the best option would be to use visual stimuli, such as animated videos or picture stories. However, due to the advanced age and poor health of the remaining speakers, there are not enough subjects to produce statistically valid results.

Therefore, we have to rely on a combination of limited spontaneous discourse data and elicitation. The absent evidence from discourse tells us that a given scenario is preferably expressed by one out of the two transitive constructions, while at the same time, positive evidence from elicitation tells us that animacy is not a factor that determines grammar. Thus, in the face of language attrition, elicitation is an indispensable tool for the investigation of possible linguistic structures.