Let the Language Be the Guide: How can we capture the internal logic of a language?

Toshihide Nakayama
ILCAA, Tokyo University of Foreign Studies

Introduction

When we think about issues in documentation of a language, we tend to focus on what to document. However, the question of how to document (that is, what approach to take in documentation) is equally, if not more, important to consider. In this talk I will concentrate on this aspect of ‘how’.

The goal of language documentation is to capture a language as a systematic functional phenomenon. Language is not a simple collection of forms and structures. Linguistic forms and structures are interrelated organically to make a microcosm with its own internal ‘logic’ (i.e., grammar). Truly rich and accurate documentation of a language should capture this internal logic in full details. This requirement may appear to be obvious and simple, but in fact it is not so straightforward.

Every linguist would probably say that s/he is interested in capturing the ‘internal logic’ of a language. However, the analytical and descriptive approaches that are often taken in linguistic studies have characteristics that are unsuitable or even harmful for the objective of attaining comprehensive and rich documentation of the internal logic of a language. Common problems include:

- accounts of selected constructions, patterns, etc. that are removed from the structural environment of the language;
- overemphasis on mechanically systematic aspects of a language;
- analysis and description guided by language-external theoretical requirements/expectations

These characteristics are problematic in documentation of any language. However, the harm is much more serious when the language involved is an endangered language. Typically an endangered language is not studied by many, and whatever documentation done by one or a few will become practically the record of
the language. Furthermore, the language may not be around long enough to allow reexamination of the record by other researchers.

I will discuss the above problems in detail.

**Problem 1: Description/analysis that is removed out of structural environment**

Quite often a certain structural aspect is isolated and studied as such. It is not uncommon to run into studies that only deal with a particular construction or pattern without any consideration of other parts of the language. Needless to say, such focused studies have their place in linguistic research and can make a significant contribution. However, overemphasis on such focused studies can foster a modularized, compartmentalized view of language. The approach to a language that views a language as a collection of constructions and different modules (phonetic, phonological, morphological, etc.) fails to recognize the organic relationship among different structural patterns and domains. Then, it would be difficult, if at all possible, to attain an accurate understanding of the structural and functional role of a linguistic construction or pattern within the internal logic of the language.

For example, in the language I work on, Nuuchahnulth, there is a construction that looks like a passive. Although ‘passive’ constructions in various languages differ in form and function, it seems reasonable to characterize the prototypical passive as a construction for placing in the subject position a participant that would not otherwise be encoded as subject. Thus, the passive construction is essentially a strategy to manipulate encoding of the subject. Then, it makes a lot of sense to see the construction in a language like English where the grammatical relation of subject is well entrenched in the overall organization of grammar. However, unlike in English, ‘subject’ is not so strongly established in the grammar of Nuuchahnulth. If we take this structural context into account, presence of a ‘passive’ construction in Nuuchahnulth does not seem to be well motivated. Moreover, on careful examination of its use, the construction turns out to cover a much wider range of function, and the ‘passive’-like use is just a part of its overall function. Thus, in spite of the partial resemblance to the typical passive construction, it is not accurate within the overall organization of grammar to characterize the ‘passive’ construction in Nuuchahnulth as a passive construction.

As illustrated above, when a pattern or construction is taken out of context and the relationship with other parts of the grammatical environment is ignored, the
structural and functional role of the pattern/constructions within the language cannot be accurately captured. And we may in turn get a distorted picture of the system.

**Issue 2: Overly limited account of the internal logic (deal only with aspects that can be reduced to generative rules)**

The second problem that is often observed in linguistic studies is an overly limited account of the internal logic. Such a study only deals with an aspect of a linguistic phenomenon that can readily be reduced to a general rule or constraint. This tendency is particularly conspicuous in formalistic approaches to language. In these studies the internal logic of a language is essentially reduced to mechanical systematicity: that is, the only part of language that is considered worth documenting and explaining is the aspects that can be accounted for in terms of generative rules.

This attitude severely limits the scope of linguistic inquiry by leaving out from consideration exceptions to the rule and habits in language use. The main sources of synchronic exceptions are layers of historical changes and also clashes among various functional forces. Thus, the exceptions are a natural, integral part of a language as a historical product and as a complex system that is maintained within various cognitive and biological limitations inherent in human beings through social interaction. In this sense exceptions should be given as much attention as generalizable rules in description of the system of a language.

The speakers’ habits in language use are another thing that will be left out by insisting on mechanical systematicity. What people actually say in their language represents only a fraction of what could be said using the language. Of course that is why we need to look beyond actual language use to get the complete picture of the grammar. But this does not mean what is happening in real language use is trivial.

For various reasons (partly historical, partly stylistic, partly social, and partly aesthetic), speakers don’t use the full spectrum of grammatically acceptable utterances. And what is important is the fact that many aspects of such stylistic choices are conventionalized enough to characterize the habits of speech of the people. Such habits shape the speech behavior much more systematically than are often assumed to: native speakers have a fairly clear sense of what and how people would say. It is probably true that these patterns have emerged largely in historical accumulation of accidental choices, but they are nonetheless part of what makes the language system tick the way it does. If it is our goal to capture what makes people
talk in the way they talk, it is important to capture what people actually say and how people say it. Linguistic descriptions often put too much focus on what can possibly be said given the structural constraints. It is of course important to determine the ‘boundary’ of grammar (or grammatical constraints), but too often linguists do not seem to care about what the speakers actually say.

Besides, what speakers often do represents very productive and strong (in terms of suggestive power) pattern. These strong patterns drive and direct language change. In that sense, these frequent patterns are the essential part of what makes a language the way it is.

**Issue 3: Description guided by language-external logic/theoretical framework (theoretical calls)**

Often linguists, especially those with a strong interest in universal formalization of the human language capacity, come in with a strong preconceived idea about how the internal logic of a language should look like. For example, a human language would have subject and object as syntactic primitives; a sentence has hierarchical internal structure; etc. This is a deductive approach to documentation in the sense that one tries to apply presumably universal linguistic principles to documentation of an individual language.

Needless to say, this way of approaching language is not necessarily wrong. There is no denying that all human languages are realized on the same physical and cognitive foundation, and therefore that languages share a number of structural properties at very deep levels of linguistic structure. Then the deductive approach based on the universal linguistic principles and categories does not seem like such a bad idea, at least in theory.

The problem is that some linguists assume too much, either intentionally or unintentionally, in the ‘universal principles’. There are many structural features or principles ‘voted into’ the universal grammar on the basis of findings in European and other major languages. Consequently, application (or ‘imposition’ may be more appropriate) of these principles to analysis and description of a language with a very different structure is justified, psychologically and methodologically.

If one starts documentation of a language with this deductive attitude armed with a set of putatively universal principles, s/he would concentrate on looking for what s/he expects to find in the language, instead of observing what the language has
to offer. Then, the main part of documentation efforts would consist of collecting evidence for the categories and structures that are expected to exist on the basis of the ‘universal principles’. The research method that is used most widely in linguistics, i.e. elicitation, goes very well with this approach. Elicitation through translation and grammaticality judgment allows a researcher to collect a highly focused set of data and evidence quickly and efficiently.

Unfortunately, the deductive approach combined with the focused elicitation method is most likely to steer the researcher away from the objective of documenting the internal logic of the language. Having a load of preconception is a bad start, and too strong reliance on the elicitation method only compounds the problem. Elicitation is a very artificial task done out of natural context, and it is difficult to consider the data collected through such a task representative of how the language really works\(^1\). In elicitation it is difficult, if at all possible, to know how ‘good’ a language sample the requested sentence is. Especially in elicitation through translation, marginal sentences and representative sentences are collected without clear distinction. The most serious flaw of the elicitation method is that one can only get information that s/he thinks of asking. Thus, relying heavily on elicitation, the chance for the researcher to break away from his/her preconceptions is rather slim. As a result, such a researcher can easily miss the unique internal logic of the language. Essentially, what s/he is doing is making a beautiful structural system (based on his/her own theoretical assumptions) out of a language instead of getting a systematic description of a language. This is a form of intellectual imperialism, and it threatens endangered languages of their rightful places in the knowledge of human beings.

As I said, this problem is often observed in formalistic approaches to language, but linguists who do not participate in such approaches can be part of the problem, too. If we are not always critical about the concepts we apply to our analysis or description, and if we do not always make it clear to ourselves and others what we mean by the terms we use, we run the risk of committing the same mistake. We should always keep in mind of the fact that many of the standard terms and concepts we use in our everyday work in linguistics are highly ‘contaminated’ with European bias, being far from neutral and objective.

\(^1\) Note that I am talking about elicitation of sentences, and not elicitation of lexical items. These two are very different matters.
What should we do – listen and look first

The analytical and descriptive attitudes discussed above are problematic for documentation of any language. However, the harm is much more serious when it involves endangered languages. Endangered languages are not typically studied by many, and whatever documentation done by one or a few will practically become the record of the language. Furthermore, the language may not be around long enough to allow reexamination of the record by other researchers.

What, then, would be a better alternative? To approach the language with as little as possible preconception is a good start. And we should let the language do the talking, showing us what to look and where to go. Instead of starting with a theory-driven framework and preconceptions about how a language ‘should’ work, we must first carefully observe how the language behaves in the natural setting, i.e. natural discourse. We shouldn’t go in with too specific lists of what we should find.

Only then, we would be able to figure out, on the basis of the observation, the descriptive framework most suitable for capturing the unique internal logic of the language. Obviously, in order to get the complete picture of the internal logic of the language, we need to supplement observations obtained from discourse data with structured sets of data collected through elicitation and also with insights obtained in studies of other languages. However, the basic categories and framework of description should arise from what can be observed in natural discourse.

Some people seem to feel that text collection is something you do after you figure out the basic outline of grammar and something peripheral that you attach as an appendix in a grammar. But text is where you can actually see how speakers use the language. And therefore I think that textual data should play a central role in language description, and text collection should be done more and at much earlier point in linguistic research than it is typically done.

By taking this approach, we can address the problem of imposing an external logic to description of a language. Also, we can avoid the problem of drawing an artificial boundary around the phenomenon of language. Whatever that uniquely characterizes the way language is realized (used) should have a place in the record of the language.

It is this inductive approach that we linguists should take in documenting endangered languages. We need to document how the linguistic system works in its own terms, not in terms of this theoretical framework or that. Theoretical frameworks
come and go, and we cannot afford to lose the grounding of documentation, especially that of endangered languages. Different languages have their own different stories to tell. What we really need to do is to let the languages do the talking first and fully before the time runs out.