

**LFG AS A FRAMEWORK FOR
DESCRIPTIVE GRAMMAR**

Paul Kroeger

GIAL and SIL Intl.

Proceedings of the LFG07 Conference

Miriam Butt and Tracy Holloway King (Editors)

2007

CSLI Publications

<http://csli-publications.stanford.edu/>

Abstract:

LFG has a number of features that make it an attractive and useful framework for grammatical description, and for translation. These include the modular design of the system, the literal representation of word order and constituency in c-structure, a typologically realistic approach to universals (avoiding dogmatic assertions which make the descriptive task more difficult), and a tradition of taking grammatical details seriously.

(This talk was presented as part of a panel entitled “Directions of LFG: Many Paths”. Each of the six panel members was asked to “explain how LFG has related to their past work and to what they are doing now in their careers.”)

Since leaving Stanford a lot of my work has been focused on training people to do field linguistics, so I have not actually been using the full LFG formalism in my daily work most of the time. But I have found that the conceptual structure of LFG provides a very good framework for grammatical description, and that is what I would like to talk about today.

A number of people have asked me at various times whether SIL still teaches Tagmemics to its field workers.¹ The short answer is “no”; there is a fair bit of variety from one place to another, but none of the major training programs currently use Tagmemics as their basic model. For those of you too young to remember Tagmemics, I might summarize it by saying it was Kenneth Pike’s attempt to extend the methods of structural phonemic analysis to morphology and syntax. The phonologists among us may be eager to point out that structural phonemics was not a very satisfactory model of phonology either; but in its day it was considered a great triumph, the envy of the other social and behavioral sciences. And for all its limitations, Tagmemics was very successful in one important respect: using this framework, Pike was able to get people with fairly limited training in linguistics to study and describe languages that had never been studied before.

This kind of primary fieldwork on previously undescribed languages is a difficult thing to do, and I do not believe it has gotten nearly enough respect in American linguistics during the past 40 years or so; so I do not want to minimize in any way the contribution of Pike and his students and colleagues. But I would

¹ “SIL” originally stood for “Summer Institute of Linguistics”, but now only the initials are used. For information about the organization, see www.sil.org.

have to admit that many Tagmemic grammars are frustrating and difficult to read. One problem is that concepts like “contrast” and “minimal pair” do not apply as neatly to phrase or sentence patterns as they do to phonemes. The more fundamental issue is that the goal of a Tagmemic grammar was to list the inventory of contrastive units of various types, just as a major goal of phonemic analysis was to state the inventory of consonants and vowels. So Tagmemic grammars tend to be essentially lists of clause patterns, sentence patterns, word patterns, etc. What is often lacking is any statement about the generalizations, i.e. the rules that constitute the grammar of the language.

Even when our goal is to write a purely descriptive grammar, I feel it is important to adopt a rule-based perspective. We may never express these rules in formal syntactic notation, but the rules of the grammar are an important part of what we are trying to describe.

Of course, the focus on grammar as a system of rules is the defining characteristic of generative linguistics. But my impression is that a lot of people who do field linguistics and descriptive grammar have given up on generative syntax. In fact, a fair number seem to have given up on syntax in general, either ignoring it or adopting the view that most apparent syntactic regularities can be reduced to semantic and/or pragmatic generalizations. I suspect that a major reason for this is that the models of formal syntax that they have been exposed to seem so far removed from observable reality and impractical for descriptive purposes.

I believe that LFG offers a much better framework for descriptive grammar than recent transformational models. The modular design of the system means that c-structure representations are a direct and literal representation of the word order of the sentence as it is actually pronounced (WYSIWYG), and of constituency in the classical sense. These are basic facts that any descriptive grammar needs to spell out. (My impression is that a lot of current work in formal syntax, and even some more functional approaches, largely ignore these issues.)

Moreover, the modular design of the system means that problems or novel solutions in one area of grammar do not necessarily lead to complications in other parts of the analysis. For example, I have been interested for some time in the problem of “symmetric(al) voice”, as it has become known in Austronesian syntax. (The term was first used, I believe, by Bill Foley at LFG98 in Brisbane.) Essentially, this means a voice alternation (change in the assignment of the SUBJ function) without demotion. In a language like Balinese, there are two different transitive clause types: one in which the agent is SUBJ, and another in which the patient is SUBJ. Both of these are fully transitive, meaning that agent and patient are both terms (direct core arguments); neither gets demoted to oblique or

adjunct status, unlike familiar voice alternations such as passive and anti-passive. In a language like Tagalog, there are several additional voice options (dative, instrumental, locative) but again these are all fully transitive; in none of them is the agent demoted to oblique or adjunct status.

Now obviously this is a problem for the original, classical form of the Lexical Mapping Theory (as developed by Bresnan & Kannerva 1989, Alsina & Mchombo 1993, among others). When I arrived at Stanford in the late 1980s, there were several different theories of linking being developed, and these facts seemed to be a problem for all of them. But for LFG, they were a problem **ONLY** for the linking theory. Once you allow a non-canonical linking pattern for these languages, the rest of the system functions pretty normally. Wayan Arka (2003) later developed a model of LMT, adapted from Alex Alsina's model, that works for Balinese, and I am sure this could be further adapted for Tagalog. But having solved this problem does not force major changes to other aspects of the analysis.

I would like to contrast this with a very influential analysis of symmetric voice proposed within the Government and Binding framework by Guilfoyle, Hung and Travis (1992). They proposed that the agent is internal to VP at D-structure, specifically that it occupies [SPEC, VP]. Depending on the voice morphology of the verb, one argument will fail to get Case in its D-structure position and be forced to move into [SPEC, IP], the structural subject position. No theta roles get deleted or absorbed, so the agent is free to remain in [SPEC, VP] when it is not selected as subject. This is a very elegant model of non-demoting voice alternation. However, because the change of GFs is expressed in terms of phrase structure, it makes incorrect predictions about things like word order and long-distance extraction in Tagalog. When all information is represented in the same way, a change in any part of the system affects every part of the analysis. In contrast, the modular design of LFG allows us to address separately issues which are in fact independent of each other.

Wayan Arka (2003) also presented a fair bit of evidence for the claim that in Balinese Undergoer Voice clauses, in which the patient is the subject, the agent and verb form a very tight constituent which he labels V-bar. This is an interesting and typologically unusual claim. Similar claims have been made in other Western Malayo-Polynesian languages, e.g. Toba Batak (Schachter 1984) and West Coast Bajau (Miller 2007). In transformational theories that adopt the Uniform Theta Assignment Hypothesis (UTAH, Baker 1988), this hypothesis is virtually ruled out on theoretical grounds. Within LFG, however, because GFs and theta-roles are represented separately from constituent relations, it is simply an empirical issue: the analysis can follow the facts of the language.

Another thing that I appreciate about LFG is that there is a healthy respect for the degree to which languages may differ from each other. Universals that get

proposed within LFG tend to be fairly well motivated typologically. Last October Peter Sells and I participated as “respondents” in a workshop on Austronesian syntax at UCSD. All of the papers were written within the Minimalist framework, and most of them adopted Kayne’s “anti-symmetry” hypothesis — essentially the claim that underlying phrase structure for all languages is strictly binary and right-branching — and Cinque’s recent proposals about universal D-structure positions for various types of adverbs. Whatever the merits of these proposals as theories of universal grammar, they impose an immense (and to my mind, intolerable) descriptive burden on languages whose surface word order is very different. Anti-symmetry seems like the most inconvenient possible analysis for a language like Malagasy.

Most of my students are not planning to do descriptive linguistics for its own sake. Many of them hope to use it to support their work in Bible translation, adult literacy, bilingual education, etc. The relevance of syntactic analysis to translation depends heavily on your philosophy of translation. Many years ago I read a review article about George Steiner’s book *After Babel*. I have not been able to find that article again, but as I recall the author was a Marxist literary critic who believed that a good translation was one that preserved the foreignness of the source text; the translation should feel almost as strange, difficult, and off-putting as the original would be for someone who does not speak the source language. For this type of translation, relatively little linguistic analysis is required since the form of the source text is largely preserved in a fairly literal way.

SIL has traditionally preferred a different model of translation. Local circumstances sometimes require a somewhat literal approach, but where possible the ideal is generally seen as a translation that is as faithful as possible to the meaning of the source text, but as natural as possible within the linguistic system of the target language. For this type of translation linguistic analysis is quite important. Any difference between the grammars of the source language and target language is a potential area where translators may be unconsciously influenced by the form of the source text, resulting in a loss of accuracy, clarity, and/or naturalness. Every detail of the grammar is important; the distinction between “core” and “periphery” is not too helpful in this context.

Ken Pike used to tell a story about a missionary that he met on one of his trips to Africa. This man had spent a year or so studying the local language and then got up to preach his first sermon. He told the people: “Jesus said, ‘I am the way, the truth and the life; no one comes to the Father but by me.’” The people replied, “Then we will worship you.” The man was horrified. He said, “No, wait, you don’t understand; *Jesus* said ‘I am the way, the truth and the life.’” The people answered, “If Jesus said it, we believe it; we will worship you.”

Obviously the people were interpreting a direct quote as indirect speech. I believe that the language in question was Bariba. Pike devotes several pages in his 1966 book on African languages to the factors which determine the choice between direct vs. indirect speech in Bariba, involving such things as the relative prominence of the speech act participants on the person-animacy hierarchy. He does not spell out what the grammatical difference between the two is, but in an obscure footnote he mentions that indirect speech is preferred and expected for reporting statements that are believed to be true.

Reported speech has been a source of difficulty in a number of other African languages as well. In some cases the use of the logophoric pronoun has caused problems. Other languages have a very productive and widely used category of “semi-direct” speech. Some translators have been reluctant to use this “semi-direct” form in their translations, even in contexts where they would always use it in natural speech, because they feel obligated to match the direct or indirect speech of the source text. The point is, every detail of grammar is potentially significant. The LFG implementation of large scale grammars for English, French, German, etc. has (I believe) instilled an ethos which takes the details seriously; they all have to be accounted for.

My own first attempt at translation came roughly two months after I began to study the Kimaragang language. I was asked to return to the state capital to help teach a seminar on translation principles. I was lecturing in the mornings, and in the afternoon the participants were practicing on short passages from the book of Acts, using a simplified Indonesian version as the source text. One of the participants turned out to be a Kimaragang; he spoke a dialect significantly different from the one I had been studying, but I could make out a fair bit of what he said.

One of the assigned passages was a somewhat bizarre story from a very early period in church history, when the small band of believers were practicing a kind of voluntary communism (or communalism), sharing their possessions with each other. A certain couple decided to sell a piece of land that they owned, keep part of the money and donate the rest; but to pretend that they were donating the full price. When the husband brought the money to St. Peter, Peter asked him: “Is this the full amount you got for that land?” The husband said “Yes.” Peter said, “Why are you doing this? The land was yours, and after you sold it you were free to do whatever you like with the money. You are not lying to me, but to God.” Immediately the husband fell down dead, and the young men wrapped up his body and carried him out to bury him. Some time later the wife came in, not knowing what had happened; the same conversation was repeated, with the same result.

As I looked over the first draft of the Kimaragang version of this story, I felt that the translator was doing a pretty good job until he got to the part where the husband falls down dead. At that point I realized that he had switched into a literal, nearly word-for-word rendering of the Indonesian. The two languages are very different in their structure, and the result seemed very unnatural and confusing to me. So I volunteered to fix it up, based on my two-months acquaintance with a related dialect. When the man read what I had written, he agreed: “Yes, that does sound better. Of course, if we say it that way people will think that the husband fell down dead, then he picked himself up and wrapped up his own body, carried himself out and buried himself.”

Some years later, in the process of writing a dissertation about Tagalog, I think I finally figured out what went wrong. I had used pro-drop expecting it to signal subject continuity. But instead it was interpreted as indicating agent continuity. In these languages, as in most languages, agents tend to be highly topical in narrative. But because in Kimaragang (as in many Philippine-type languages) the agent is frequently not the grammatical subject, topic continuity is often not reflected by subject continuity.

Topic and focus are important issues in translation. I once heard a man named Roger van Otterloo talk about his initial attempts at translation in Kifuluru, a Bantu language of Zaire/Congo. They were beginning with one of the passages that talks about defending the rights of widows, orphans and foreigners. Roger proposed wording that simply said, “Don’t steal from widows,” and all the men with him began to laugh. He was afraid that he had perhaps gotten a tone wrong and said something improper, but he discovered that the problem was more interesting. Basic word order in Kifuluru, as in most Bantu languages, is SVO. But the immediate post-verbal position is also a structural focus position. So when he said “Don’t steal from widows,” the people heard “Don’t steal from widows” (implying: “anyone else is fair game”).

Nowadays everyone talks about topic and focus, but LFG was one of the first syntactic frameworks to integrate these pragmatic functions into the formal rule system. Joan Bresnan in particular was one of the pioneers in this area. I think this is a significant contribution to the field as a whole.

I once spent several months as an advisor to a committee of translators from one of the Land Dayak languages of western Borneo. They had been producing very literal renditions of the English Good News Bible, and I wanted to give them some sense of what a more natural style would look like; so I asked one of them to tell the story of Jonah and the whale in his own words. When he got to the point where the sailors ask Jonah how they can save themselves from the storm, he said (in Land Dayak): “Jonah told the sailor to throw himself into the sea.” I was surprised and asked who exactly ended up in the water, but it turned

out the meaning was correct; it was Jonah who went overboard. That was my first hint that the language allowed long-distance reflexives.

I told this story to K.P. Mohanan during one of my visits to Singapore, and Mo said something like, “Of course, what else would he say?” It seemed perfectly natural; just like Malayalam. But at that time I had not heard of long-distance reflexives in any closely related language, so I was not expecting it.

Now suppose that this had been a translation instead of a spontaneous story, and that the translator (following the English source text) had written: “Jonah told the sailor to throw him into the sea.” I believe that this would have been interpreted in Land Dayak as meaning that some third person was to be thrown overboard, which of course is incorrect. But if I had not accidentally learned this fact about reflexive binding in Land Dayak, I would never have thought to check it, because it looks like a perfectly accurate and natural translation. This is an example of a “blind error”, an error that no one would have caught without specific knowledge about that aspect of Land Dayak grammar.

Let me mention one final grammatical issue in translation. Malay/Indonesian shows a strong preference for the passive voice (specifically the *di-* passive) to encode main-line events in a narrative, especially where there is a series of actions by the same actor. Now in these contexts, it is clear that the actor is highly topical. In some functionalist approaches, such clauses cannot be analyzed as passives because the passive by definition is a construction that topicalizes patients. But syntactically the *di-* construction is clearly a passive: the patient has all the syntactic properties of a subject, and the agent has all the syntactic properties of an oblique argument. LFG takes both the syntactic relations and the pragmatic functions very seriously, but recognizes them as being distinct and logically independent of each other. Thus it is possible to ask, “What is the pragmatic function of the passive in language X” in a meaningful way, because the construction is not defined in terms of its pragmatic functions.

This is just one instance of a broader principle: languages can use the same syntactic constructions for quite different purposes. In a number of mainland Southeast Asian languages, the passive is used only for unfortunate events — the so-called ADVERSATIVE PASSIVE. In Biblical Hebrew, as in the Greek New Testament (probably due to Semitic influences), an agentless passive is often used as a way of describing God’s actions without using any name to refer to God. Clearly the functions of the passive in Hebrew, Vietnamese, Malay and English are quite different from each other. I will not spell out the details here, but with a bit of imagination you can see that a translation from any one of these languages to any of the others which mechanically preserves active for active and passive for passive can lead to confusing (and sometimes hilarious) results.

In summary, I appreciate LFG's combination of precision and flexibility, its practical and realistic approach to syntactic analysis, and its attention to detail. All of these are important features of a good descriptive grammar.

References

- Alsina, Alex and Sam A. Mchombo. 1993. Object asymmetries and the Chichewa applicative construction. In Sam Mchombo, ed., *Theoretical aspects of Bantu grammar*. Stanford, CA: CSLI Publications.
- Arka, I Wayan, 2003. *Balinese morphosyntax: a Lexical-Functional approach*. Canberra: Pacific Linguistics.
- Baker, Mark. 1988. *Incorporation: a theory of grammatical function changing*. University of Chicago Press.
- Bresnan, Joan and Jonni M. Kanerva. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry* 20.1:1-50.
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: a Cross-linguistic Perspective*. Oxford: Oxford University Press.
- Guilfoyle, Eithne, Hung, Henrietta, and Travis, Lisa. 1992. Spec of IP and Spec of VP: two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10:375-414.
- Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA, MIT Press.
- Miller, Mark. 2007. A grammar of West Coast Bajau. PhD dissertation, University of Texas at Arlington.
- Pike, Kenneth L. 1966. *Tagmemic and matrix linguistics applied to selected African languages*. Ann Arbor, MI: University of Michigan Center for Research on Language and Language Behavior; Office of Education, U.S. Dept. of Health, Education and Welfare.
- Schachter, Paul. 1984. Semantic-role-based syntax in Toba Batak. In Schachter (ed.), *Studies in the structure of Toba Batak*. UCLA Occasional Papers in Linguistics No. 5. Los Angeles: Department of Linguistics, UCLA, pp. 122-149.