

Reasons why I do not care grammar formalism

Jun-ichi Tsujii
Dept. of Electrical Engineering,
Kyoto University,
Yoshida-honmachi, Sakyo, Kyoto, 606, JAPAN

General Comments

Computational linguistics (CL) has borrowed a lot of ideas from *Theoretical Linguistics (TL)*. We could not have developed even a simple parser without the research results in *TL*. It is obviously nonsense to claim that we, computational linguists, do not care research results in *TL*.

However, the researchers in *TL*, it seems to me, are very fond of *fighting*, especially, those who are called *Syntacticians*. They always fight with each other by asserting that their *grammar formalisms* are superior to the others'. They are oversensitive and tend to distinguish people into two groups, *the ally* and *the enemy*.

A computational linguist using LFG (or *pseudo* LFG) as a small part in his total system is taken as the ally of LFG, and is certainly accused by the other groups. They promptly demonstrate that LFG is wrong, by showing a lot of peculiar sentences which rarely appear in real texts.

We are tired of listening to such discussions.

The Reasons Why

Formalisms are prepared for accomplishing specific purposes. The formalisms in *TL* have been proposed, roughly speaking, for describing the *rules* of distinguishing *grammatical word sequences* from arbitrary *ungrammatical* sequences, and of relating the grammatical sequences with the other representational levels.

On the other hand, a formalism we need in *CL* is for different purposes. That is, we need a formalism for describing the rules of distinguishing the most feasible grammatical structures from other less feasible but still *grammatical* ones of the same sentences [*Disambiguation*]. We also need a formalism in which we can manage systematically a large amount of knowledge of various sorts necessary for NLP.

Formalisms for different purposes, of course, should be evaluated based on different standards. The current discussions of different formalisms in *TL* are irrelevant to our standards, though they may be important for their *fighters*. The following is a list of the reasons why I think so.

(1)[Small and Peculiar Examples]: Linguists always argue that their formalisms are better than others by using almost the same set of peculiar sentences. This implies that the differences

of the formalisms are revealed only in these types of sentences which rarely appear in real texts.

Furthermore, it often happens that all of the proposed formalisms can capture the same regularity. They only claim that their formalisms capture it more elegantly than others, *elegantly* according to their standards.

(2)[Meaning and Knowledge]: The *elegance* of their formalisms is obtained partly by ignoring uninteresting phenomena, again *uninteresting* according to their standards. Especially, they ignore largely phenomena related to *meaning, extra-linguistic knowledge, contexts*, etc. Or they ignore them by claiming that they are related to these factors.

(3)[Disambiguation]: Linguists can ignore the *uninteresting* phenomena, but *CL* researchers developing actual systems cannot. The systems have to cover wide ranges of phenomena which really appear in texts. Furthermore, *disambiguation*, which is the hardest problem in *CL* but not at all in *TL*, certainly requires considerations in the factors which syntacticians ignore. I fear that *the elegance of their formalisms becomes a serious obstacle to the introduction of such extra factors and processings on them*.

(4)[Management of Grammar Descriptions]: I also fear that their elegance becomes an obstacle to the systematic development and management of grammar descriptions. Grammar descriptions here include the descriptions of both rules and dictionaries. Some formalisms are claimed *elegant* in the sense that they require only very few rules. But this elegance is obtained by very complicated dictionary descriptions. The standards for being *elegant* seem different in *TL* and *CL*.

(5)[Processing Issues]: The grammar formalisms ignore mostly the processing issues. Linguists do not care processing issues in their formalisms just as we do not care grammar formalism.

Conclusion

I have to repeat here that I do not claim that *TL* research is irrelevant to *CL*. I only claim that *grammar formalisms* are not important. What is important is in their discoveries which are described by their formalisms. And what we have to do is to describe their discoveries in our own formalisms.