

Yusuke Miyao, Associate Professor, Digital Content and Media Sciences Research
Division

At the root of language is sure to be a set of rules peculiar to human beings

Despite being intelligent entities, monkeys lack a language like that of human beings. If one considers why only human beings can use language, one concludes that there exist some rules peculiar to human beings in their brains. I hope to discover the rules at the root of language, which we use every day without thinking.

Discovering of mathematical formulas when seeing an apple fall from a tree

All languages share some commonalities; for example, they have nouns, verbs, subjects, predicates, etc. This indicates that certain rules govern how overall meaning is constructed by ordering and combining words. On this point, one can follow the theories of grammar and attempt to describe them in some type of formula. To date, however, such theories have involved numerous contradictions. Organizing these into formal rules that can be understood by computers is clearly a formidable task.

The method of deriving meaning by statistically processing massive volumes of data has recently entered widespread use in the field of natural language processing. Still, however, computers remain incapable of learning everything automatically from data. Just as with Newtonian physics, to elucidate the rules of language regulation, I believe one must observe natural language as a natural phenomenon.

Applications of deep analysis as corroborative research

One can verify whether an apparently identified language rule is correct by running it through computer programs. The Enju parser I developed analyses the structure of English text to compute its meaning. Deep analysis is applied to compute the nature of each individual word and to identify the logical structure in which they are combined. This is distinct from—for example—computing superficial relationships such as the probability that a certain word is the object of a certain transitive verb. MEDIE and Info-PubMed, applications that use Enju to search molecular biology papers, differ from ordinary search engines in that they can pinpoint the specific content one is searching for, such as papers on interactions between protein A and protein B. They have now entered widespread use.

Weaving together parsing and meaning

If parsing is the warp of a textile, meaning is the weft. The challenge I face is how to weave these two threads together for the English, Japanese, and Chinese languages. What kind of framework should one begin with to describe meaning? Individual texts involve basic approaches on what needs to be communicated, including assumptions, assertions, and proposals, that precede their content. I believe the development of such a framework as the foundations of meaning will one day make it possible for computers to answer questions like "What is the author's point?" which appear so frequently in contemporary written language tests.

Interviewed and summarized by: Rue Ikeya