Development of Diachronic Corpus

Project by the National Institute for Japanese Language and Linguistics, Japan, NINJAL: 2009–13, 4 year project.

Main purpose: Study of Japanese language
(sub) purpose: Study of Japanese (classic) literature

Parole

structure of language
visible
tangible

Langue

complex system
sequential reader
pattern prediction
dynamic presentation of diachronic language change

Figure 1: Corpus and Description, Langue and Parole:
The nature of language is dynamic and always changing while the phenomena of language might be static. We should consider the dynamic change of language as a component comprised of various elements. The feature of language we usually observe is a complex system and tangled with wide-ranging elements.

A case study: use of SAKURA (cherry blossoms) in Mt. Yoshino → Kokinshū (ca. 905) vs Shinkokinshū (1205)

Sakura (桜) and Yoshino (吉野), a place name in Nara prefecture
← Kokinshū (ca. 905)
Shinkokinshū (1205) → during 300 years differences.

Figure 2: Extraction of delta from each synchronic layer: A, B, C and D are arbitrarily-assigned synchronic layers on the time axis. Examination of linguistic transitions is achieved through the comparison of lexical items in each layer with those in other layers, and the discovery of common principles appearing in the delta of data extracted from both systems as well.

Figure 3: Serial comparison model; differential model of transitional linguistic elements of target texts; $A$ is a set of elements that occurred at Time $t_1$; $A'$ is a set of elements that occurred at Time $t_2$; $T$ is the time axis; $f(x)$ is a function for converting an element $x$ of $A$ into that of $A'$.

Future Task
- To define linguistic units suitable for each era
- To develop a dictionary for machine analysis
  → it allows us syntagmatic and paradigmatic analyses

Conclusion
- Addressed basic concepts and framework of diachronic corpus
- Illustrated the serial comparison model for historical analysis
  → Lexical differences between any two groups of texts.