Introduction to the Symposium

Gérard Huet

Thanks to all present. We are inaugurating a new scientific meeting, identifying an area where recent progress of various groups in the world makes it meaningful to try and share experience, with the hope that an effort at cooperation will leverage mutual success. We are facing a formidable task, turning toy prototypes into production tools for philological analysis of a daunting corpus.

If we want to have some hope of cumulative and durable process, we have to make a conscious effort at identifying clear interfaces with precise specifications so that our software components become inter-operable. This is the first rationale for a minimum of standardization. The second rationale is that software technology evolves at such a pace that if you do not follow the de facto standards, your software will quickly disappear as obsolete. You must imagine software as a living organism, with an incredibly fast rate of evolution.

The increase in computer power, and more importantly the enormous software development effort around the world is both a blessing - we may hope to solve problems the computational solution of which would have been unthinkable 10 years ago - and a curse - we have to follow an incredibly fast technological development, which makes any piece of software very quickly obsolete. This rate of change is strikingly out of phase with the rate of evolution of vyakarana, where we are still discussing fresh interpretations of Panini's grammar - a work composed 25 centuries ago !

Of course the technological aspect of our challenge should not either take precedence over the more conceptual problems of linguistic modelling. This is why we decided not to restrict ourselves to workshop sessions on standardization issues, but to profit of this occasion to have a scientific symposium where various theoretical issues could be presented and debated. We gathered a Program Committee which is fairly representative of the various problematics we are facing. And there was a significant response to the call for contributions, 20 submissions came in time, out of which we selected 10 composing a rather balanced program.

One of our challenges is the inter-disciplinarity of the topic. We have in this room western linguists, indologists, sanskritists, paninian experts, vyakarana pandits, philologists, computer engineers and mathematicians, each area with its own scholarly tradition, its concepts, methodologies and terminologies. Each of us has to make efforts at understanding others points of view, and this is certainly harder in such a meeting than in more mainstream disciplines, where colleagues look more like clones of each other.

Of course one could say the following. What we are doing is attempting to understand natural language by computer in the particular case that this language is Sanskrit. So Sanskrit Computational Linguistics should not be intrinsically different from say French Computational Linguistics or English or Hindi or Arabic or Chinese Computational Linguistics. Thus let us apply the best off-the-shelf technology of computer modeling of natural language to the specific parameters of Sanskrit, and leave the native locutors to their place as linguistic informants.

Unfortunately, this is a fallacy. The problems with this line of reasoning are the following.

First of all, there is no such off-the-shelf technology. There are myriads of linguistic models, grammar formalisms, there are structuralist and statistical approaches, and at the end of the day we have no fully convincing general framework easily applicable, even in English, where most efforts have been spent, so we must be modest and recognize that progress is slow on a long-range problematics. Worse, a lot of the technology applies with more or less success to languages with rigid syntax such as English or French, but is mostly irrelevant to languages with freer syntax such as Sanskrit, and where semantical analysis has priority.

Secondly, there is a wonderful native tradition of linguistic analysis which is a treasure chest. I was personally amazed to discover the sophistication of Panini's work. The Western linguist Jakobson in his lectures on General Linguistics in New York in 1942 explains the genesis of the notion of phoneme, attributing its origins in the work of the Polish linguist Baudoin de Courtenay in 1870. The crucial idea is discretization of phonetics in the minimal way which permits their discriminative power. But Sanskrit phonemes were already identified, and logically classified, by Panini in the 4th century before Christ. And a complete system of morpho-phonetical description was elaborated in a completely formal way. Not only Panini was by far the first linguist in recorded history, but I claim he was the first informaticien, 24 centuries before computers came into existence.

This makes Sanskrit quite unique among human languages. Whatever we may think of the prescriptive intentions of Panini, his grammar and the following traditions were de facto used as prescriptive. And Sanskrit could prosper all those centuries with very little variation compared to vernacular languages.

Thirdly there is a deep difference between scientific investigation in the Western and in the Indian tradition. In the West, knowledge is acquired principally according to principles of logic, resting ultimately upon mathematics. In Indian tradition, grammar - vyaakara.nam - is the vehicle of reasoning, of argumentation and of knowledge acquisition. So A.s.taadhyaayii, and the further work of the navya nyaaya tradition, should not be compared to "La grammaire de Port-Royal", but more to set theory or to primitive recursive arithmetic as a foundational framework for rigorous knowledge processing.

Thus, citing Michel Angot: "Un trait de mdecine est d'abord un ouvrage sanskrit sur la mdecine avant que d'tre un trait de mdecine en sanskrit : la consquence contemporaine, c'est que traduire et comprendre un texte sanskrit en

ignorant les conceptions des vaiyaakara.naas demeure impossible" ("A treatise of medicine is first a Sanskrit exposition of medicine before being a medicinal treatise in Sanskrit; the consequence is that translating and understanding a Sanskrit text without knowledge of the vaiyaakara.naas concepts is plainly impossible").

Thus we have to go deeper in understanding a Sanskrit text than just being able to give an English paraphrase of it. We must understand how it can be produced by the grammatical devices of the grammarians traditions.

This means that we cannot make the economy of understanding the work of Paa.nini and his successors. It is needed for doing the job right of computational modeling of Sanskrit, and not just out of historical interest.

Conversely, this also means that if we do the job right, then we shall ease the work of scholars studying the Sanskrit texts in a way which will be significantly more informational than a mere translation.