

## Activity report 2011

### Sanskrit Associated Team

Centre de Paris-Rocquencourt, INRIA (France),

Dept. of Sanskrit Studies, Univ. of Hyderabad (India)

and the Sanskrit Library, Providence (USA)

## 1 Joint activities and exchanges in 2011

G erard Huet visited the University of Hyderabad from the 21st of february to the 5th of march. During this stay, a first interconnection of our Web services was successfully achieved. It is thus possible to invoke Amba Kulkarni’s parser from the selected tagged segmentations issued from G erard Huet’s Heritage Platform. This sequence of tools thus processes a continuous Sanskrit sentence into a set of segment sequences tagged by morphological features, further filtered by a crude semantic role constraints processor, further analysed by a deeper dependency parser, yielding a hopefully small set of dependency graphs solutions, conveniently displayed using the GraphViz technology. This seamless combination of heterogeneous Web services is a clear added value to both platforms. It gives the ability to process sandhied input to the UoH platform, while allowing deep structure generation for the Inria platform. The mutual installation of both platforms on two servers, under a variety of UNIX-like system installations (Linux, Mac OSX), necessitated a streamlining of various libraries and tools.

In April Peter Scharf, from Brown University, was nominated to a Blaise Pascal Chair for 2012. He will spend the year in Paris 7 UMR “Histoire des th ories linguistiques”, and will be able to collaborate cosely with the Rocquencourt site of the team.

In May the two site leaders met in Prague, Czech Republic, for a visit at the team of Prof. Vacek at Charles University. This Pandanus team does research on the semiotics of nature in Sanskrit. From May 11th to May 14th, the various groups exchanged on topics of potential collaboration, possibly in view of applying to an EC research call for proposals.

Amba Kulkarni then visited the Rocquencourt site from May 15h to June 3rd for further work on the platforms interface. It is now possible to communicate the morphological tags of a segment from the Heritage tagger to the UoH parser.

From June 1st to July 30th, Yoann Couillec, Master student at University Paris 6, worked in Rocquencourt on system aspects of the Heritage platform. He first worked on the streamlining of the installation procedure, replacing ad-hoc uses of old sed macros by Python programming. He then worked on porting the Web services on an Android platform, in view of delivering a hand-held

Sanskrit platform. This last part could not be finalized, due to difficulties in cross-compilation in Objective Caml.

On July 7th, Gérard Huet received the “Grand Prix Inria”.

In August, a significant revision of the *kāra*kas (semantic roles) processing is effected on the Heritage platform, in order to implement a treatment of nominatives consistent with Pāṇini’s grammar. Now the role of Agent in active voice (rep. Patient in passive voice) is denoted by the finite verb form, words in the nominative being mere attributes of the corresponding actant, possibly ellipsed. The precision of the parser is significantly improved.

From September 16th to 25th, Gérard Huet participated to an official delegation led by Michel Cosnard, President of Inria, visiting India. He gave a talk cum demonstration of the Sanskrit Heritage platform at the EFEO site in Pondicherry, at the invitation of François Grimal. On September 21st a plaque was inaugurated by the President of Inria and the Vice-Chancellor of UoH at the Sanskrit Studies Department of UoH, celebrating the joint team activity.

On October 1st Pawan Goyal is taking a post-doc position in the Sanskrit joint team at its site of Rocquencourt. He will work on a computer simulator for Pāṇini’s grammar, beyond doing joint work with Gérard Huet on the Heritage platform.

Gérard Huet will visit India again in December and January 2012. In December, he will participate with Amba Kulkarni and other scholars in the field to a Sanskrit Computational Linguistics Workshop in Karala from December 22nd to December 27th. The workshop will be hosted by the Chinmaya International Foundation Shodha Sansthan at Veliyanad in Kerala. The target of the workshop is to train School/college Sanskrit teachers in using the tools developed by the joint team.

In January 2012 will be held the 15 World Sanskrit Conference in Delhi. Gérard Huet will present a paper “Departing from PANini for good reasons” in the Linguistics panel “Models and Theories in Sanskrit grammar and linguistics” organized by Jan Houben and Pierre-Sylvain Filliozat.

## 2 Tentative work program for 2012

Next year will see a significant increase in the Rocquencourt/Paris work force, since Pawan Goyal will be associated as a post-doc to the Rocquencourt site, and Peter Scharf will be taking his Blaise Pascal Chair at the “Histoire des théories linguistiques” CNRS-Paris 7 UMR.

The main research goal will be to enhance the coordinated development of the Heritage Platform of Gérard Huet, the Sanskrit Computational Linguistics platform of Amba Kulkarni, and the Sanskrit Library developed by Peter Scharf and his collaborators. Here is a sketch of the main tasks envisioned for 2012:

1. Lexicon

- (a) Design of a unique naming scheme for lexical items of the 2 platforms plus the electronic Monier-Williams dictionary, with all inter-

- coercions
  - (b) Linking of the Heritage Platform morphological tags to the electronic Monier-Williams dictionary, in order to upgrade it with a truly bilingual French/English interface
2. Morphology
    - (a) Completion of the secondary conjugations paradigms
    - (b) Addition of kridanta generators in addition to the current participial ones
  3. Deep structure
    - (a) Better integration of the UoH dependencies constraints processor with the Heritage front end
    - (b) Improvements of the treatment of coordination and relative clauses
    - (c) Improvements of the dependency graph display
  4. Corpus processing
    - (a) Experiments in continuous corpus processing
    - (b) Evaluation of the precision/recall rates of the compound parser on various benchmark corpora
    - (c) Design of discourse structure operators, notably for anaphora resolution and topic tracking
  5. Links with the grammatical tradition
    - (a) Initial design of a Paninian simulator
    - (b) Synthesis of Paninian sūtras from the platform computations

### 3 Publications

The publications of Amba Kulkarni are listed at <http://sanskrit.uohyd.ernet.in/faculty/amba/>, those of Gérard Huet at <http://yquem.inria.fr/~huet/bib.html>, those of Peter Scharf at <http://www.brown.edu/Departments/Classics/people/facultypage.php?id=10044>

The series of proceedings “Sanskrit Computational Linguistics” published in the Springer Lecture Notes series is detailed at <http://yquem.inria.fr/~huet/EA/>.

The Sanskrit Library is accessible from <http://sanskritlibrary.org/>, the Sanskrit Heritage site at <http://sanskrit.inria.fr/>, the SCL platform of University of Hyderabad at <http://sanskrit.uohyd.ernet.in/scl/>.