Languages of India in the IMAGACT Multilanguage and Multimodal Ontology of Actions

Jawaharlal Nehru University of New Delhi

DILEF developed the IMAGACT Ontology of Action (designed and coordinated by the LABLITA Research Unit), a linguistic infrastructure which allows the exact translation of high frequency verbs by means of the universal language of images (scenes).

http://imagact.lablita.it

IMAGACT web application is an online dictionary that can be used for language teaching. The core Ontology can be also exploited for the development of enabling technologies in the fields of artificial intelligence and human-machine interaction (e.g. see MODELACT project: http://modelact.lablita.it).

After the annotation of a language is finished, data are automatically processed and integrated in the dictionary and the new language becomes available for translation, both as source than as target language.

Language and identity: spontaneous spoken Brazilian Portuguese resources in the Romance context

With 29 states, 1,721 spoken languages, 22 of which are official languages of the country, India needs to face a hard work to manage its linguistic variety. Although Hindi is the main Indian language (with over 400 millions native speaker), the majority of Indian people are mother-tongue speakers of other languages. Innovative teaching programs, reciprocal understanding, and technology support must be guaranteed for each of these languages.

IMAGACT extension to Indian languages, developed in cooperation with the Jawaharlal Nehru University of New Delhi, aims to spread the usage of the IMAGACT web dictionary to a wide part of Indian people by adding the support for the major Indian languages: currently Hindi and Urdu are fully implemented in the resource and other 7 languages are under development (Assamese, Bengali, Magahi, Manipuri, Odia, Tamil, and Sanskrit), covering about 700 millions of people.

IMAGACT4ALL Project provides the technological support for the IMAGACT linguistic annotation by native speakers.

The expected impact of this Project is high, given the huge number of potential users belonging to linguistic communities for which multilingual web resources are not easily available.

DILEF developed (through the LABLITA research unit, also in connection with national and EU projects) a wide series of resources collecting spoken Italian and spoken Romance languages in a standard framework for speech representation (within the agreement between UNIFI and UFMG, through the project of Universidade Federal de Minas Gerais, Belo Horizonte, Brazil), a large corpus of spoken Brazilian Portuguese has been released: C-ORAL-BRASIL (http://www.c-oral-brasil.org), which has been built on the model of LABLITA resources (and specifically of C-ORAL-ROM corpora of spoken Romance languages: French, Italian, Portuguese, and Spanish).


A part of C-ORAL-BRASIL corpus has been also extensively annotated and is a part of the CORAL-IPIC database (for Information Patterning Inter-linguistic Comparison), maintained by LABLITA.

http://www.lablita.it/app/dbipic/

C-ORAL-BRASIL is fully comparable with the Portuguese C-ORAL-ROM resource, and represents the specificities of the Brazilian Portuguese with respect to a substantial homogeneity between the Brazilian and the European written varieties. This work falls within the scope of the cooperation for development since it is aimed at offering the Brazilian autonomous identity.

Universidade Federal de Minas Gerais