SAGA

1. BASIC INFORMATION

1.1. Tool Name

SAGA: An automatic tool for grapheme-to-allophone transcription in Spanish and its dialectal variants

1.2. Overview and purpose of the tool

Saga is a rule-based automatic phonetic transcription system for Spanish, jointly developed by the Universitat Politècnica de Catalunya and the Universitat Autònoma de Barcelona in the framework of SAM-A European project (Esprit project 6819) and several project granted by the Spanish government (TIC-91-1488-C06-02,
TIC95-0884-C04-02 and TIC95-1022-C05-03).

The syntax of the rules has been designed to obtain a phonetic transcription of Spanish as it is pronounced in central Spain. The phonetic description is given in terms of SAMPA (Speech Assessment Methods Phonetic Alphabet, http://www.phon.ucl.ac.uk/home/sampa/index.html).
However, *Saga* allows the user to modify the transcription rules so that *Saga* can be adapted to dialectal variants. The tool as is distributed has the capacity to deal with the main variants of the Spanish spoken in America (Mexico, Caribbean region, Colombia, Peru, Chile and Argentina).

1.3. A short description of the algorithm

The transcriptor has hardwired general rules to provide an orthographic text with a phonetic transcription in SAMPA symbols corresponding to the standard Spanish spoken in central Spain. Additionally, a set of options and data files can specify alternative rules to alter the transcription, in such a way that dialectal variants can be accomplished. By the use of these files foreigner words can be substituted by an orthographic transcription that mimics the way they area pronounced by a Spanish speaker. The files are also used to specify the transcription of a word whose phonetic realization does not follow the general rules.

The tool is able to split the words into syllables and mark the prosodic stress.

When necessary, Saga can produce transcriptions in terms of triphone, a phonetic unit useful in speech recognition and speech synthesis.

2. TECHNICAL INFORMATION

2.1. Software dependencies and system requirements

The program is written in C++ and can be compiled to run in either Linux/Unix or Windows operative systems.

2.2. Installation

```
$ tar xzvf Saga.tgz
$ make
```

2.3. Execution instructions

```
$.work/Castilla . fileIn.txt fileOut.phn
```

2.4. Input/Output data formats

Text files

2.5. Input data formats

Text file with ortographic transcription.
Text files with commands to tailor the tool to dialectal variants

2.6. Output data formats

Text file with phonetic SAMPA transcription

3. CONTENT INFORMATION

3.1 A test input file

me gusta el pájaro español de buen agüero

3.2. the output file

```
m e / g 'u s - t a / e l / p 'a - x a - r o / e s - p a - J 'o l / d e / b
w 'e n / a - G w 'e - r o
```

(/ denotes the limits of words
- marks the border between syllables
' stands for stress)

3.3 approximation of the time necessary to process the test input file

less than 1 second

4. ADMINISTRATIVE INFORMATION

4.1. Contact person

Name: José B. Mariño
Address: Jordi Girona 1-3 Edifici D5, 08034 Barcelona, Spain
Affiliation: TALP research center. Universitat Politècnica de Catalunya
Position: Professor
Telephone: +34 93 401 6444
Fax: +34 93 401 6447
e-mail: jose.marino@upc.edu

4.2 Delivery medium (if relevant; description of the content of each piece of medium)

The resource will be uploaded on the MetaShare platform as an archive.

4.3 . Copyright statement and information on IPR

The resource has copyright. The Copyright belongs to Universitat Politècnica de Catalunya, Universitat Autònoma de Barcelona and Universitat de Barcelona. The resource is free, license-based, for research purposes of academic or research institutions; and fee license-based for companies and commercial purposes.

5. RELEVANT REFERENCES AND OTHER INFORMATION

[1] J. Llisterri, José B. Mariño,"Spanish adaptation of SAMPA and automatic phonetic transcription", Esprit Project 6819. Report SAM-A/UPC /001/V1 (February 1993).

[2] A. Moreno, José B. Mariño, “Spanish dialects: phonetic transcription”, Proc. ICSLP’98, pp. 189-192, Sydney, Australia (November 1998).