

Project proposal

Free open source spell checker for Papiamentu

Curacao, February 12, 2012

Fundashon Kòrsou Habri / Stichting Curacao Open

Preamble

Papiamentu is a relatively young Creole language spoken by 500.000 native speakers on the ABC islands of Aruba, Bonaire and Curacao, and their descendants in the Netherlands. There is a Papiamentu spell checker for MS Word 97, 2003 and 2007. Yet it works only on MS Windows. There is no spell checker for MS Office, for Mac or for MS Office 2010. Moreover, the existing spell checkers are not free. Their license price is about 100 USD. The MS Office suite is not free either. A large amount of the Papiamentu population is therefore not able to afford and or use a spell checker when writing. They are also excluded from the opportunity to profit from the latest technologies and the facility to use a spell checker based on the official orthography and the standardization process of the Papiamentu language during the last 15 to 20 years. Most native speakers are, as yet, not familiar with the new writing system of their language. Thus a Papiamentu spell checker will full fill a very urgent local need.

Project objectives

The aim of this project is to make and improve a free open source spell checker for Papiamentu, the language of the majority of the inhabitants of the ABC-islands (Aruba, Bonaire and Curacao), and their descendants living in the Netherlands. The project also includes a free open source office suite.

The main project goals can be summarized as follows:

- to improve the quality of written Papiamentu texts in schools, government, private organizations and homes
- to remove financial barriers by enabling everyone to work with a free office suite
- to improve the existent spell checker of Papiamentu used by open source office
- to provide a spell checker of Papiamentu for MS Windows, Apple Mac OSX and Linux, as well as on smart phones and Firefox.

Target groups

The beneficiaries of this project are:

- all children and youngsters in Curacao (recently Papiamentu became one of the main instruction languages in primary education and is a core subject in secondary education)
- The ± 80% of households in Curacao for whom Papiamentu is their mother tongue

- Teachers in all types of education, as Papiamentu has become one of the instruction languages
- Employers in governmental and NGO settings, as since Papiamentu is recognised as an official language the local government has chosen to produce more documents in this language.

Project description

On September 17th, 2011, our foundation 'Curacao Open' released the first open source spell checker for LibreOffice (www.opencuracao.com). The word list used was made available by Fundashon Planifikashon di Idioma (FPI), the local language planning institute. The technical part was realized by a local Open Source company (Suares & Co). Improvement of the existent spell checker is however needed, both in linguistic and technical aspects. Our foundation will work together with both aforementioned entities in order to realize these improvements.

The improvements in the spell checker are mostly of linguistic nature. The current word-list (about 18.000 words) is not quite adequate nor complete. An analysis of which word forms should be systematically listed, will be part of the project.

For example, the existent spell checker lists words, but other word forms derived from the main word or lexeme, are excluded. An adequate Papiamentu spell checker must allow all possible and/or standardized contractions. For example, all contractions or word forms used in combined phrases. In the English language for example, "I am" and it's contracted form "I'm" are both valid word forms. Thus in the case of Papiamentu both *palu di koko* and *pal'i koko*, the contracted form in the combined phrase, should be included .

There is also a large list of highly common nouns derived from verbs (V), which are not listed in the current spell checker, for example *V-do*, *V-dor* (agent of the action), *V-mentu* and *V-shon* (nominalized action).

Thirdly there are word forms that exist only in combination with other words, such as *partidó di karta* or *bentamentu di plaka afó*, which have been excluded in the first version of the spell checker. The spell checker has to look at the context for certain word forms.

Finally, Papiamentu has a lot of diacritic features, such as four types of accents, the ñ and the ç. There is a lot of linguistic and technical work involved to enable the spell checker to recognise them in a good way. If one types for example 'Ana' it should also suggest 'Aña'. The awareness of this linguistic problem, and the fact that context is important for a right choice of sound transcription, is not included in the existent spell checker.

Because Papiamentu is a Creole language in development, the standardisation

of one lexeme for more variants of a particular semantic concept, and or the creation of new words for existent or new concepts, are ongoing processes. In some cases there are therefore more possibilities for the written form of one particular word and the general public is not able to choose between the two word forms, for example *aseptá*, *apsetá*, *aksektá*, *apseptá*. Now that the language has become the language of instruction in primary and secondary education, we expect the creative process to accelerate.

The project will therefore include the possibility of national and international participation by means of an interactive web application, where the Papiamentu speaking population on the ABC islands and The Netherlands, are able to interact by proposing words for inclusion in the word list and/or debate the inclusion of a new or of current words, or even make a choice for a certain word form based on their particular local usage.

Our President, Dr. Marta Dijkhoff will lead the linguistic part related to this project. She is a linguist with international merit in creole morph-syntax, language policy and language planning. She will lead the research towards the improvements, and also supervise the input of new words through the interactive web application. Dr. Dijkhoff's expertise will be our foundation's share of the project costs.

The project will consist of three phases:

1. The research needed to determine the words to be added and the (technical) addition to the word list
2. The creation and maintenance of a public website, where Papiamentu native speakers suggest new words and or debate about existent word forms
3. Release of the improved spell checker for LibreOffice, which can also be used with Firefox and other Open Source Software.

Since Papiamentu is also spoken in Aruba, Bonaire and the Netherlands, our Foundation will share it's expertise when possible with representatives or responsible entities on these islands and Foundation SPLIKA in The Netherlands.

Activities

The project will consist of five main activities:

- generating words to complete the word list
- improve the spell checker on linguistic aspects
- technically develop the programme behind the spell checker to make it more complete and user friendly
- share and exchange experiences with other Papiamentu speaking communities in Bonaire, Aruba and the Netherlands
- gaining commitment and involvement of relevant stakeholders by presenting the project to a broader audience in a seminar

Time schedule

According to plan, the implementation of the project will start in June 2012. The project period will be 6 months.

Budget and financing plan

The costs of the project are estimated as follows:

	Estimated costs in USD
<i>Project coordinator</i>	
* 4 hours a week @ 26 weeks, hourly rate = USD 95	9.860
<i>Assistents for linguistic work</i>	
* 8 hours a week @ 26 weeks, hourly rate = USD 50	10.400
<i>Programming work (ICT)</i>	
* total estimated time needed 10 days, daily rate = USD 562	5.618
<i>Seminars for introduction and exchange</i>	
* 1 ticket NL - Curacao	1.067
* hotel stay min. 3 days @ US 98 p/night	295
* 2 tickets Aruba - Curacao (retour same day)	298
* 2 tickets Bonaire - Curacao (retour same day)	208
* venue for seminar	843
* catering food & drinks	843
* marketing - advertisements in newspapers, magazines, etc.	562
* printing costs	281
TOTAL COSTS	25.716

Therefor, USD 25.716 will be requested from UNESCO.

Fundashon Kòrsou Habri

Dr. Martha Dijkhoff