

# Dacco and qdacco

An Open Source Catalan-English Dictionary  
and its GUI



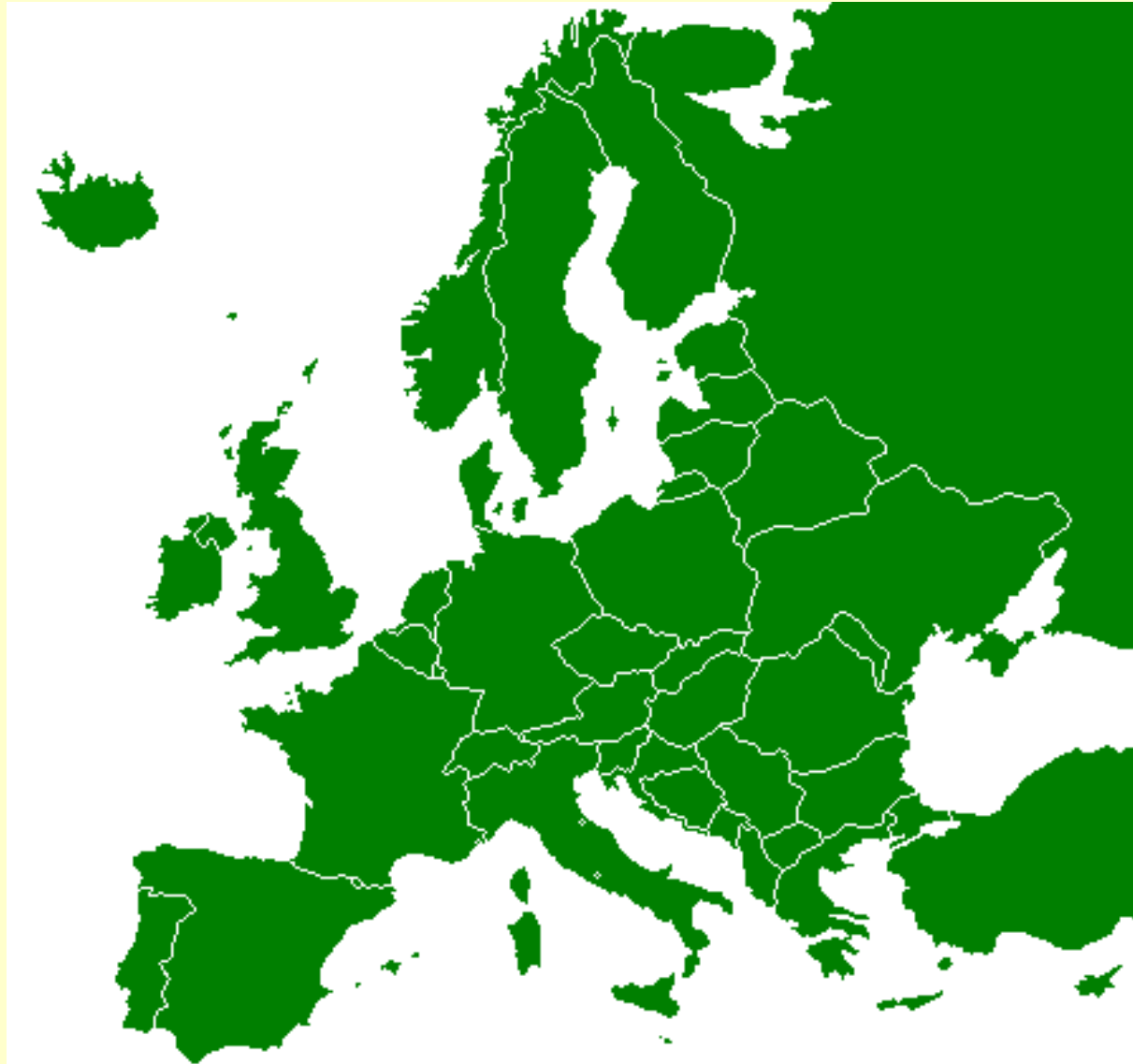
Fosdem 2008 in Brussels, Belgium

Carles Pina i Estany  
carles@pina.cat

# What is the history of Catalan and what is its future?

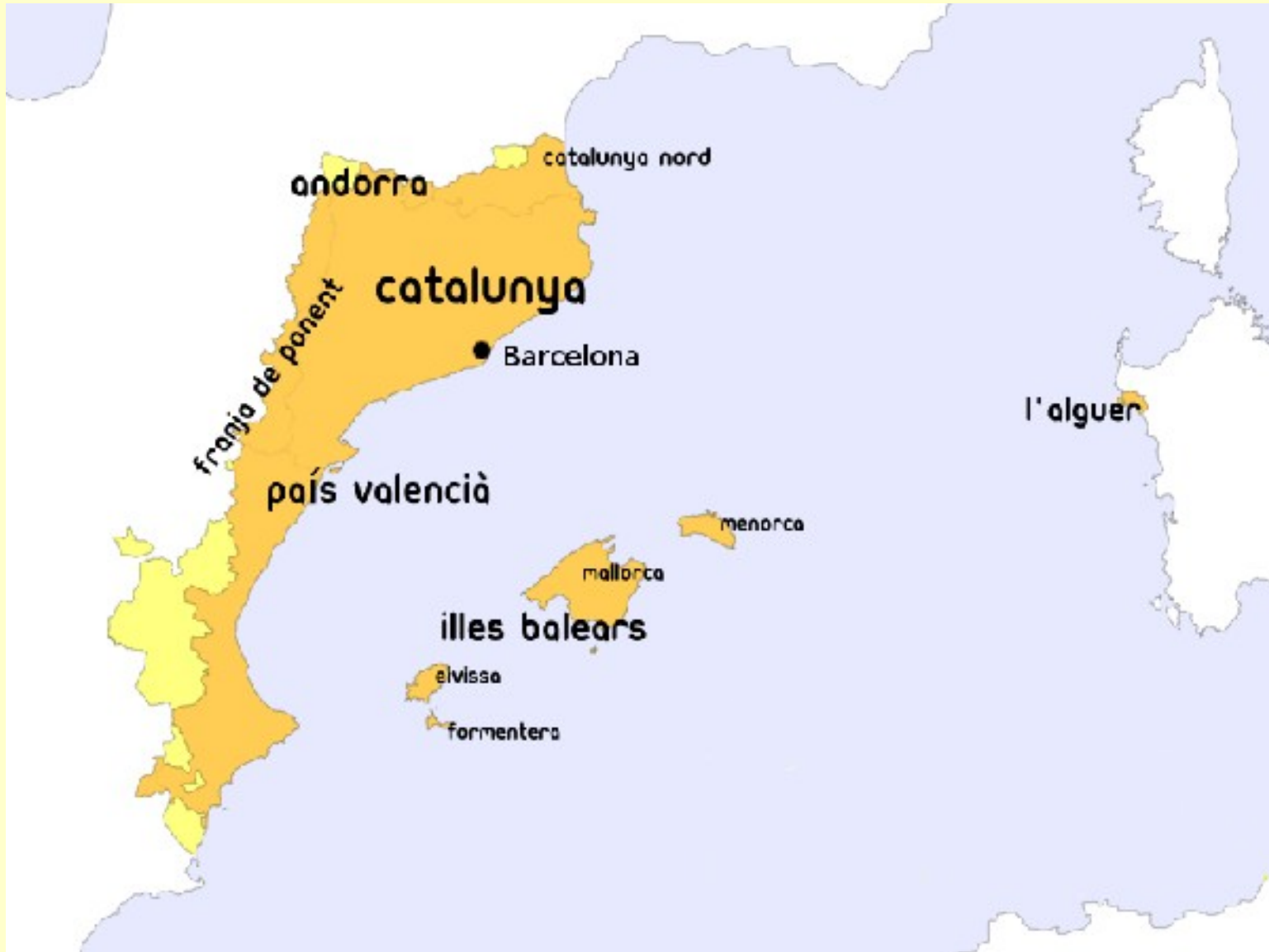
- It's a Romance language that derives from Latin, just like Spanish, Italian, French, Portuguese and Romanian
- The Catalan linguistic and cultural community has its own .cat Top Level Domain
- It's an official language of Catalonia, Andorra, Valencia and the Balearic Islands. It is also spoken in Alghero (Sardinia) and parts of southern France

# Map of Europe



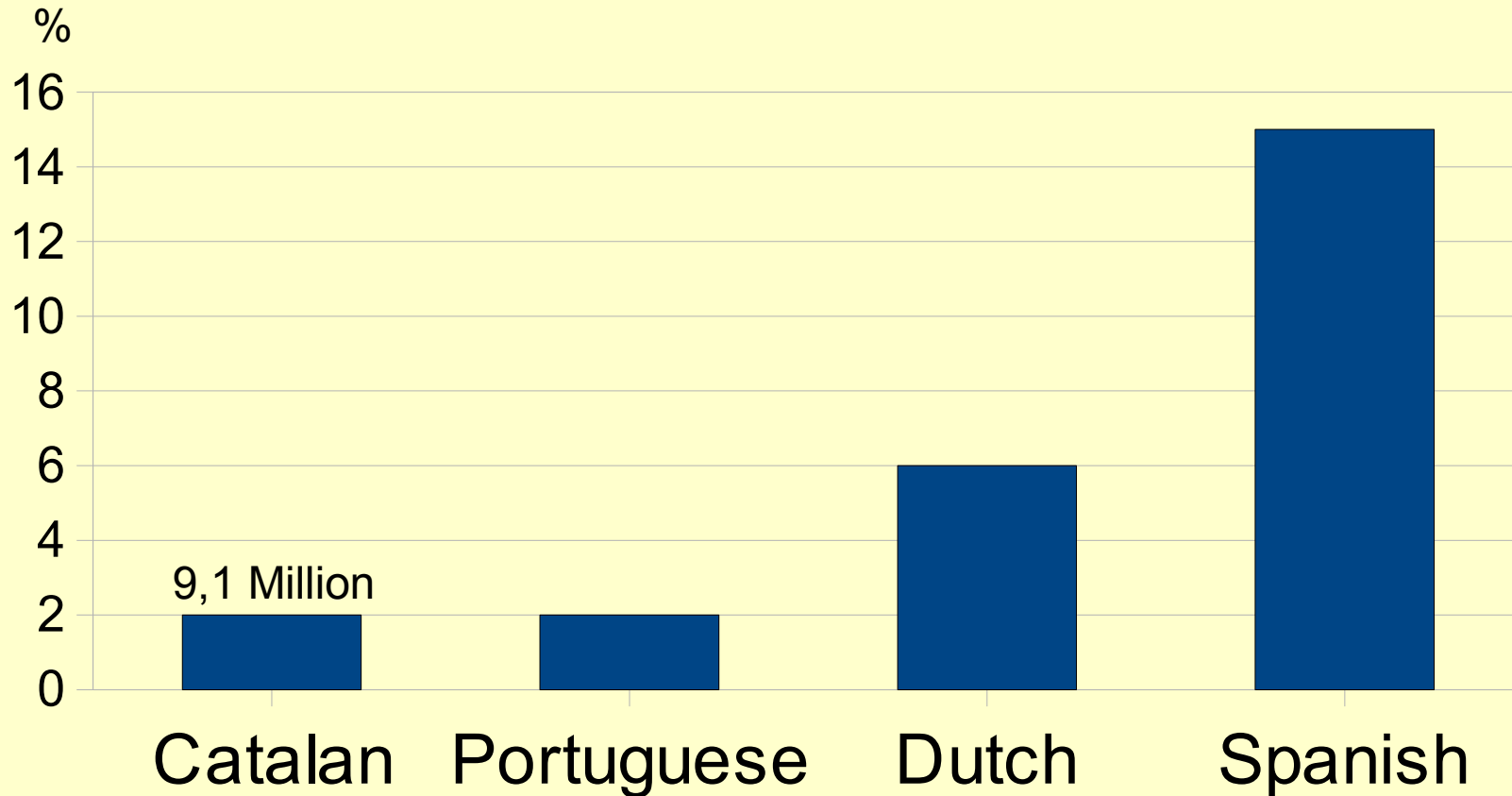
Author: Wikipedia

# Where is it spoken?



Author: Josep Gallart

# Speakers as % of EU population



Data source: [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_243\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_en.pdf)

# What is Dacco?

- English-Catalan dictionary tailored for speakers of both languages
- All entries are written in XML
  - Also exported to PDF for improved usability
  - Form-based web search at [www.catalandictionary.org](http://www.catalandictionary.org)
  - Plugins available for Firefox and IE
  - Gadget for iGoogle homepage
  - Catalan verb conjugator
  - qdacco (standalone application)

# Dacco's history

- An English student of Catalan found that there was a lack of Catalan dictionaries available which were comprehensive and aimed at English speakers
- This was confirmed by other students of Catalan through a survey she distributed during her PhD studies
- The collection of dictionary entries was begun in 2001. The first version of Dacco was released in 2003

# Who contributes?

- People from all walks of life: linguists but also mathematicians, geographers, IT professionals.



Lou Hevly



Josep M.  
López



James  
Macgill



Linda Oxnard



Leopold Palomo



Oriol Vilaseca

## **Others:**

- Joanathan Kaye
- Gill Martin
- David Gimeno
- Jaume Ortolà
- Max Wheeler
- Margarita Castañón

# The 'average' contributor

- Aged between 20 and 80
- English speaker living in UK, US, Canada or Australia
- Catalan speaker from Barcelona, Girona, Tarragona, Valencia or the Balearic Islands
- Common purpose: wanting to create a resource which is not only useful to language learners but is also free and open




# Where do we get our entries from?

- We do not copy entries from any dictionary
- Agreement with TermCat to incorporate their open source vocabulary lists into our dictionary
- Entries are often added by those who use the dictionary in their daily tasks when they search for a word and fail to find it
- Recently: we are matching some Catalan-Latin-English birds list to incorporate

# How does the project work?

- First step: find a new word
  - Mailing list: regular contributors send suggestions
  - Web form: any user can send suggestions
  - 'Missing word lists': created automatically from online search engine and manually from suggestions sent through qdacco
- Second Step: discuss the new word's translation equivalent
  - Different meanings according to geographic area, age, socioeconomic status, etc.
- Third Step: project admin adds word to the XML

# Dacco's Special Features (1/3)

- It's free: anyone can modify, copy or redistribute
  - Data (XML files): LGPL 
  - PDF files: CC Attribution-Share Alike 2.5 
  - qdacco: GPL 3 
- More than 15,000 entries in each side of the dictionary (over 200 DIN-A4 pages)

# Dacco's Special Features (2/3)

- 4 dictionaries:
  - Catalan-English dictionary
  - English-Catalan dictionary

} for Catalan speakers

  - Catalan-English dictionary
  - English-Catalan dictionary

} for English speakers
- Examples and usage notes tailored according to user's native language

# Dacco's Special Features (3/3)

- Links to images
- Usage notes
- Examples
- Word frequency counts (from Google)
- Semantic fields: apple  $\Rightarrow$  fruit  $\Rightarrow$  food

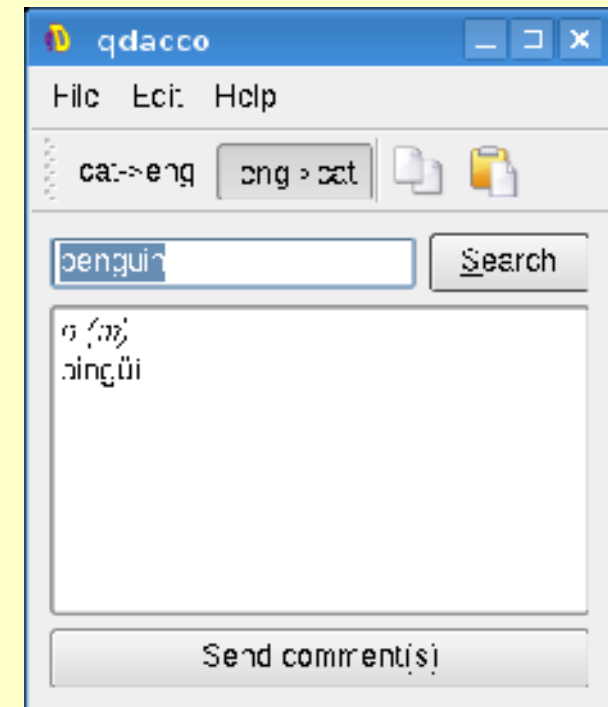
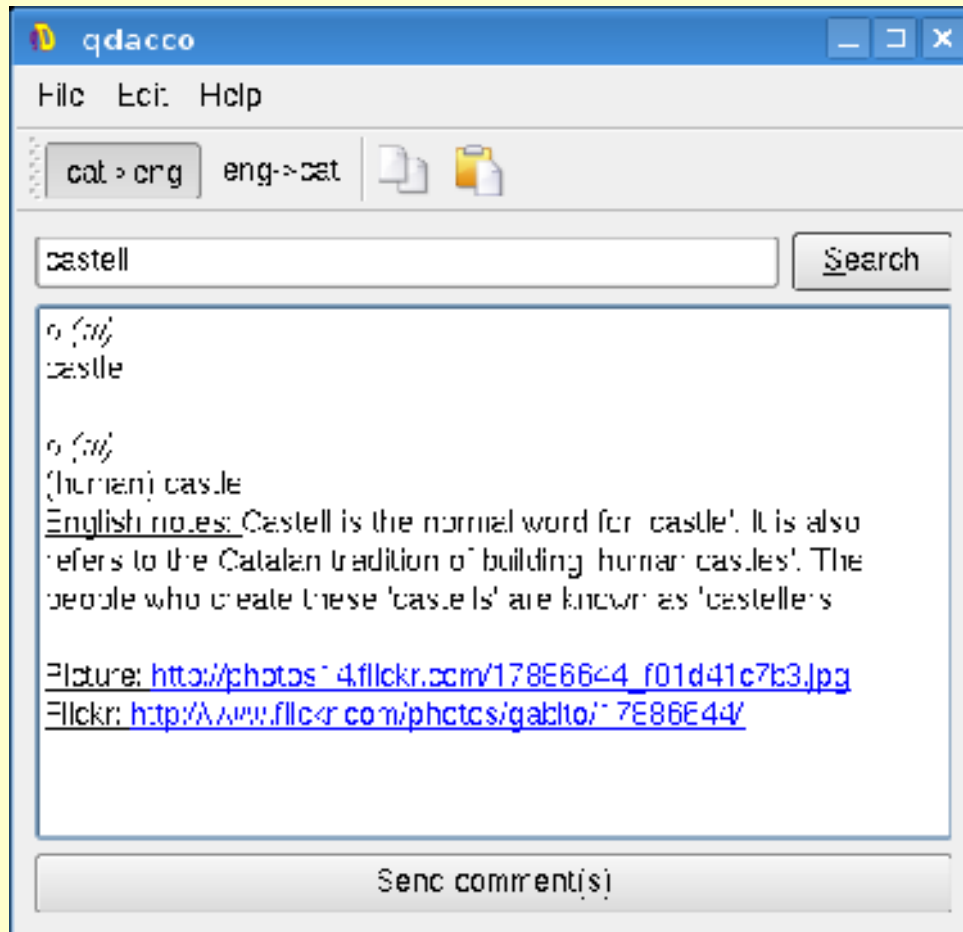
# Why is it so important that the dictionary be open source?

- Culture should be free
- Anybody can incorporate it into their own application or web site
- Anybody can suggest entries, though every suggestion is examined and discussed by team of contributors so that quality of dictionary is never compromised. Descriptive, not normative dictionary.
- We encourage others to create their own open source dictionaries and are happy to share the benefit of our experience

# What is qdacco?

- Multi-platform (Unix, Linux, Windows\*) standalone application
- Available in standard Debian repositories
- qdacco has been developed since 2005
  - Approximately two major releases a year

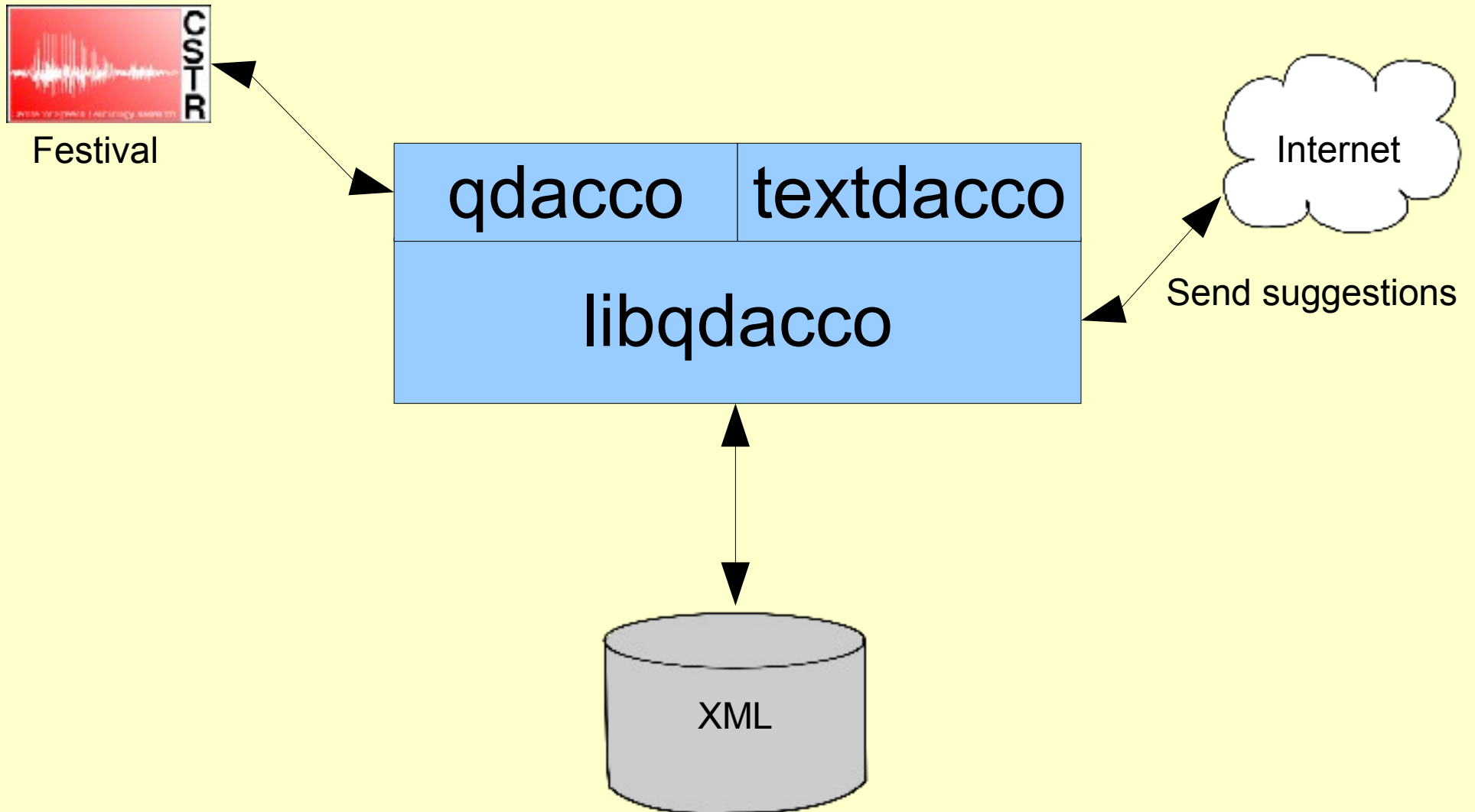
# Screenshots



# Why did we create qdacco?

- We wanted a reference application for Dacco
- Access to all Dacco resources (photos, links, examples, notes, etc.)
- Reports missing words to the Dacco Project.
- Integration with Festival (Speech synthesizer)
- Auto-completion and many other features
- Offline searching - faster than looking through a PDF file on your own.

# qdacco architecture



# Similar projects

- GPL German-Catalan dictionary:
  - GPL Deutsch-Katalanisches Wörterbuch
  - <http://www.aldeaglobal.net/diccionari/index.php>
- Wiktionary:
  - Catalan: <http://ca.wiktionary.org/wiki/Portada>
  - English: [http://en.wiktionary.org/wiki/Main\\_Page](http://en.wiktionary.org/wiki/Main_Page)

Thanks for your  
attention

Questions?

?

<http://www.catalandictionary.org>

Mail: [carles@pina.cat](mailto:carles@pina.cat)

# Creative Commons

This work is licensed under the Creative Commons Attribution 2.5 Spain License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/2.5/es/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.