

## Computational Portuguese: the state of the art (?)

Diana Santos

www.portugues.mct.pt

## What is it?

There are at least three goals in the computational (processing of) Portuguese:

- create computer programs for a Portuguese-speaking audience
- study the language (for scientific / humanities research)
- automate information capabilities and communicative skills

Three different paradigms: business, academic world, and engineering

2

## Different goals

- Improve existing products with new functionalities
- Write papers, understand how language works
- Create systems that help some user(s)

There's no functional correspondence among these goals...

3

## Our project is meant as a bridge

We devoted ourselves mainly to two activities

- Observation
  - mainly cataloguing
  - some evaluation
- Resource creation
- (Debate)

4

## Observation: we list

- the companies that offer products for Portuguese
  - and their products
- the publications
  - and the research groups
- the systems and resources that have been developed

### NOTE

These are actually three very different kinds of information

5

## Comments from users

- Only developers and individual users have contributed with some feedback
- Researchers don't send publications
- Companies only contact us to publicize their job offers

Still, there have been a lot of visits to our site (aprox. 80 000 since July 1998)

6

## Our comments

- We tried to be as little evaluative as possible
  - Names as provided in the Web pages
  - Only added information about kind of availability
  - A very light classification (Corpora/lexicons/raw material/tools)
- Further work on refining classifications and evaluation
- Created a search system for our area
  - Publicized four weeks ago, only 63 searches so far (from 19 different computers)

7

## A quantitative overview

Corpora	23
Lexicons	145
Tools	77
Teaching material	61
Raw material	130
Companies	35
Publications	445



8

## Some ?original? views

- People with different backgrounds could gain from visiting our site and seeing different approaches to their “own” field
- Other people might use the site as raw material for other kinds of guided tours, for example for teaching purposes
- Suggestions for improvement, and discussion of the contents, would be useful to many

9

## Qualitative remarks

- Hardly any advanced tools
  - basic text processing facilities
  - basic resources
- Very few companies
  - most are publishers
  - big companies want groups of languages
- Few international publications

10

## Some measures which have a low value

- Number of publications per group / per researcher
- Number of Web pages per group
- Number of projects per institution
- Number of available results/products
- Evaluation material for each system / area

11

## Other indicators

- (Low) presence in discussion lists
- (Low) cross citation between groups
- (Low) intergroup collaboration
- (Doubtful) existence of a common forum
  - Meetings: EPLP/PROPOR? APL (Portugal), InPLA (Brazil)
  - Journals?

12