

# LANGUAGE RESOURCES AND EVALUATION: INTERNATIONAL STRATEGY PANEL

Dr. Daniel Tapias (tapias\_d@tsm.es)  
Voice Products and Services Development Dept.

30<sup>th</sup> May 2002

# CONTENTS

➤ Introduction.

➤ RAE Initiatives.

➤ University Initiatives.

➤ Telefónica Initiatives.

**TME**

Dirección de Desarrollo de Productos y Servicios de  
Voz

*Telefonica*  
**Móviles**

# INTRODUCTION

## LANGUAGE RESOURCES PRODUCERS

- Institutions (Real Academia Española, Instituto Cervantes).
- Universities (Univ. Politécnica de Cataluña, Univ. de Vigo, Univ. del País Vasco, Univ. Politécnica de Madrid, etc.)
- Companies (ATLAS, IBM, Telefónica Group, etc.)

**EC HLT  
Programme**

**Ministry of  
Science and  
Technology**

**Regional  
Governments**

**Companies  
and  
Institutions**

**TME**

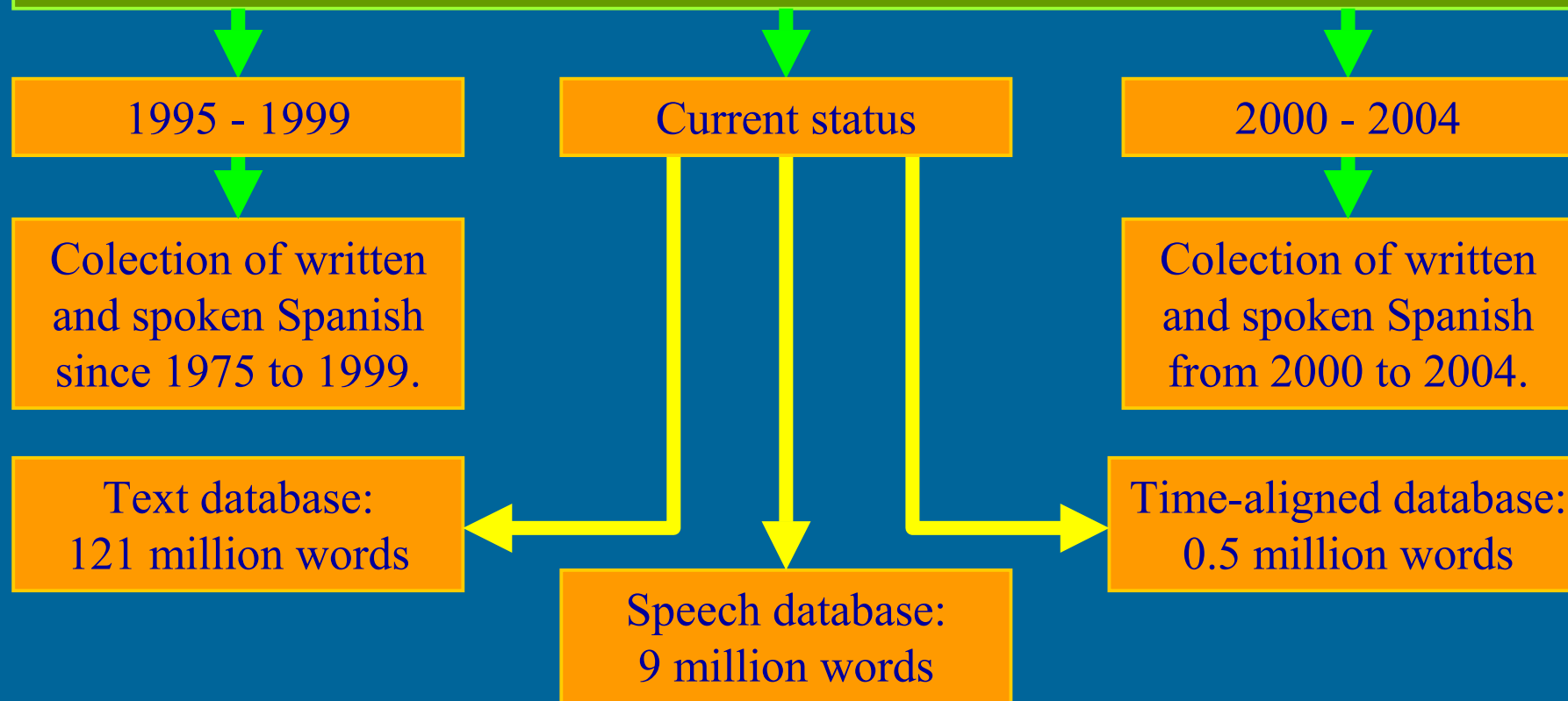
Dirección de Desarrollo de Productos y Servicios de  
Voz

*Telefónica*

**Móviles**

# RAE INITIATIVES: CREA PROJECT

CREA: Reference Corpus of the Present Spanish Language in Spain and Latinamerica (<http://www.rae.es>)



TME

Dirección de Desarrollo de Productos y Servicios de Voz

*Telefonica*

Móviles

# RAE INITIATIVES: CORDE PROJECT

CORDE: Reference Corpus of the Old Spanish Language (<http://www.rae.es>)

1997 - 1999

Collection of written and spoken Spanish since 1100 to 1974.

Current status

Text database:  
137 million words

2000 - 2004

Collection of written and spoken Spanish from 1100 to 1979.

Goal:

300 million words  
in 2003

Funding Organizations:  
**RAE, Ministry of Science and Technology**

**TME**

Dirección de Desarrollo de Productos y Servicios de  
Voz

*Telefónica*

**Móviles**

# UNIVERSITY INITIATIVES

Universitat Politècnica de Catalunya - UPC

EUROM1	Spanish
SpeechDat(M)	European Languages (Spanish & Catalan)
SpeechDat II	European Languages (Spanish)
SpeechDatCar	European Languages (Spanish)
Albaycin	Spanish
Sala I	Latinamer. Spanish & Brazilian Portuguese
Sala II	Lat. Span., Braz. Port., US Eng., US Span., Canadian French, Canad. English
Oriental	8 mediterranean languages

**TME**

Dirección de Desarrollo de Productos y Servicios de  
Voz

*Telefonica*

**Móviles**

# SPEECH DATABASES STRATEGY AT TELEFÓNICA

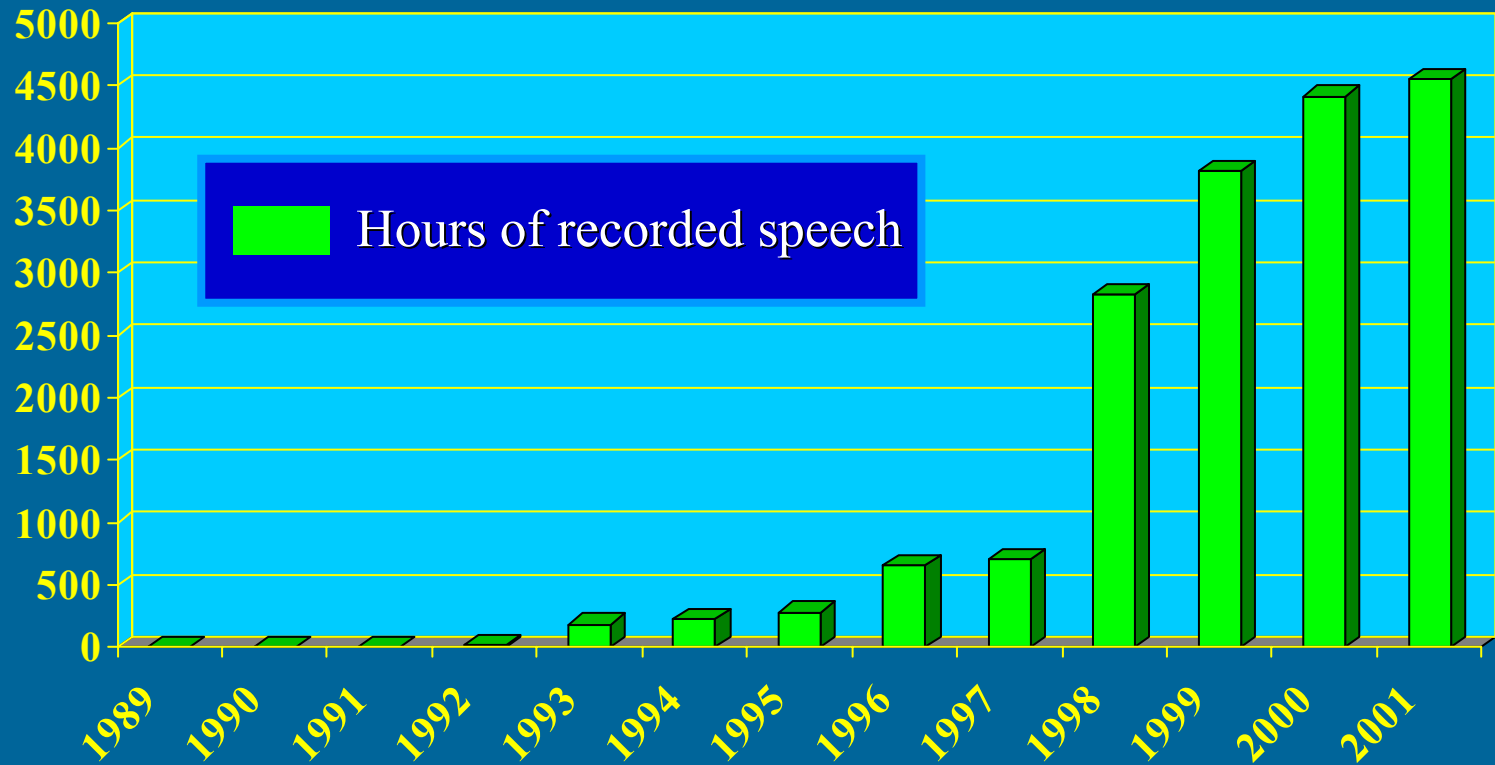
Current regions of interest for Telefónica: Europe, Latinamerica, and North Africa.

Produce the language resources when they are not available. For example:

- Castilian Spanish,
- Peruvian Spanish,
- Catalan,
- Galician,
- Vasque,
- Brazilian Portuguese.

Get the language resources through ELRA and LDC.

# SPEECH DATABASES AT TELEFÓNICA



TME

Dirección de Desarrollo de Productos y Servicios de Voz

*Telefónica*

Móviles



*Telefónica*

---

***Móviles***