

# 音声データ利用の連携 音声資源コンソーシアムの現状

板橋 秀一<sup>1,2</sup>, 山川 仁子<sup>1</sup>

<sup>1</sup> 国立情報学研究所 (NII)

<sup>2</sup> 産業技術総合研究所 (AIST)

# 今、なぜ、音声コーパスなのか？

- 音声情報処理: 認識・分析・合成等
- 世界の諸言語の音声研究・方言研究
- 計量的な言語研究

## 各種・大量の音声データが必要

- 単語音声・連続音声・多言語・方言など  
各種データベースの作成
- 危機的言語の保存

# 音声コーパスにおける問題点

- プロジェクト用・実験用に作成
- 非公開
- 作成の人的・金銭的負担

音声データの作成・収集・蓄積・配布・共有のための  
共通の枠組みが必要

各種音声データの収集・管理・配布を行う機関

音声資源コンソーシアム(NII-SRC)

# 音声資源コンソーシアム (NII-SRC)

音声コーパスの収集・配付・研究事業

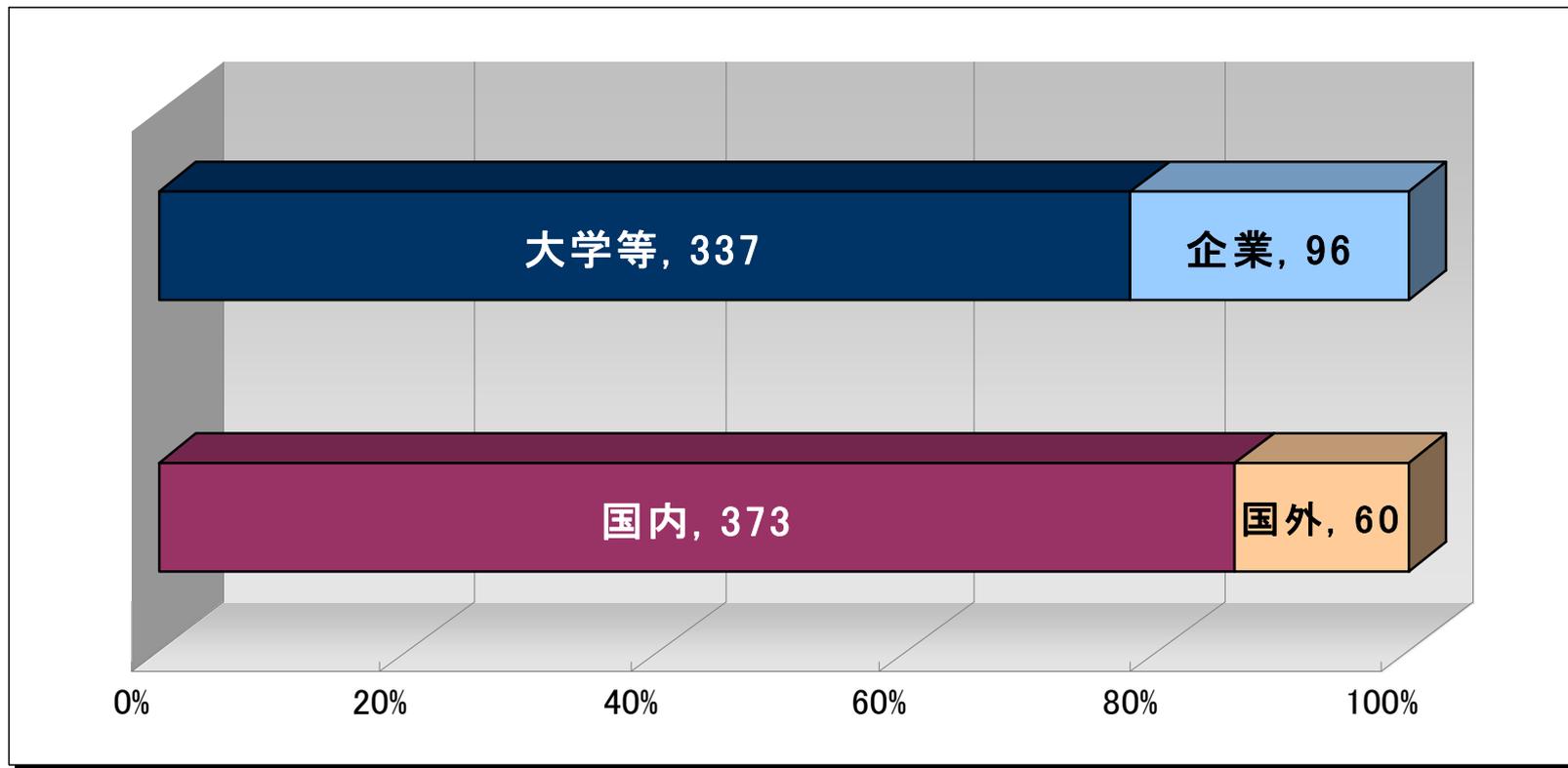
情報メディア特に音声メディアの未来価値創成に向けて  
国立情報学研究所内に設置

2006年6月サービス開始

<http://research.nii.ac.jp/src/>

# コーパス配布状況 (2006.9~2009.3)

申込み件数：433件 / 配布コーパス数：1046



# 日本における音声・言語コーパスの取り扱い

言語資源協会

国立国語研究所

## 国立情報学研究所

情報検索システム評価用  
テストコレクション構築プロジェクト

音声資源コンソーシアム

情報学研究用データリポジトリ

## 情報通信研究機構

言語グリッド

高度言語情報融合  
フォーラム

## 企業

(株)ATR-Langue

NTT-AT(株)

日外アソシエーツ(株)

等