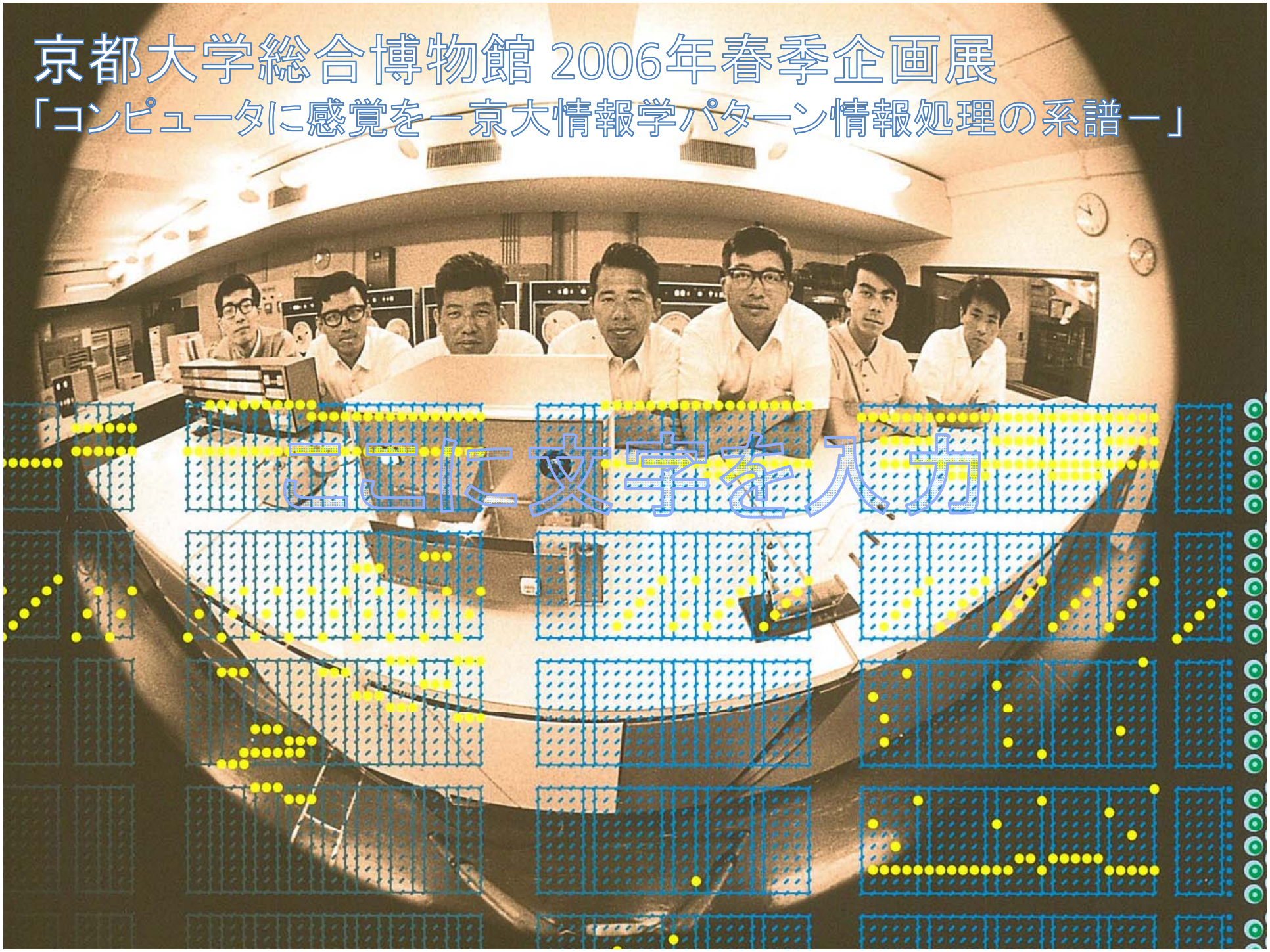
# 京都大学における
# 言語音声情報処理の産学連携

黒橋 禎夫　河原達也
京都大学情報学研究科

（第二回 産学デザインシンポジウム、2013/9/26、KRP）

# 京都大学総合博物館 2006年春季企画展

## 「コンピュータに感覚を－京大情報学パターン情報処理の系譜－」



# 音声認識ソフトウェア Julius

- オープンソースの大語彙連続音声認識ソフト
- 国内外で幅広く研究利用
- 様々なアプリケーションへ展開
  - ロボット、携帯端末、音声案内…

- 無償配布（月平均4000～5000ダウンロード）
  http://julius.sourceforge.jp/

# 音声対話システム

- 質問応答
- 情報推薦

- Wikipedia
- ニュース記事

- 一問一答で終わらない長いインタラクション（ナビゲーション／話し相手）

# 国会審議の音声認識

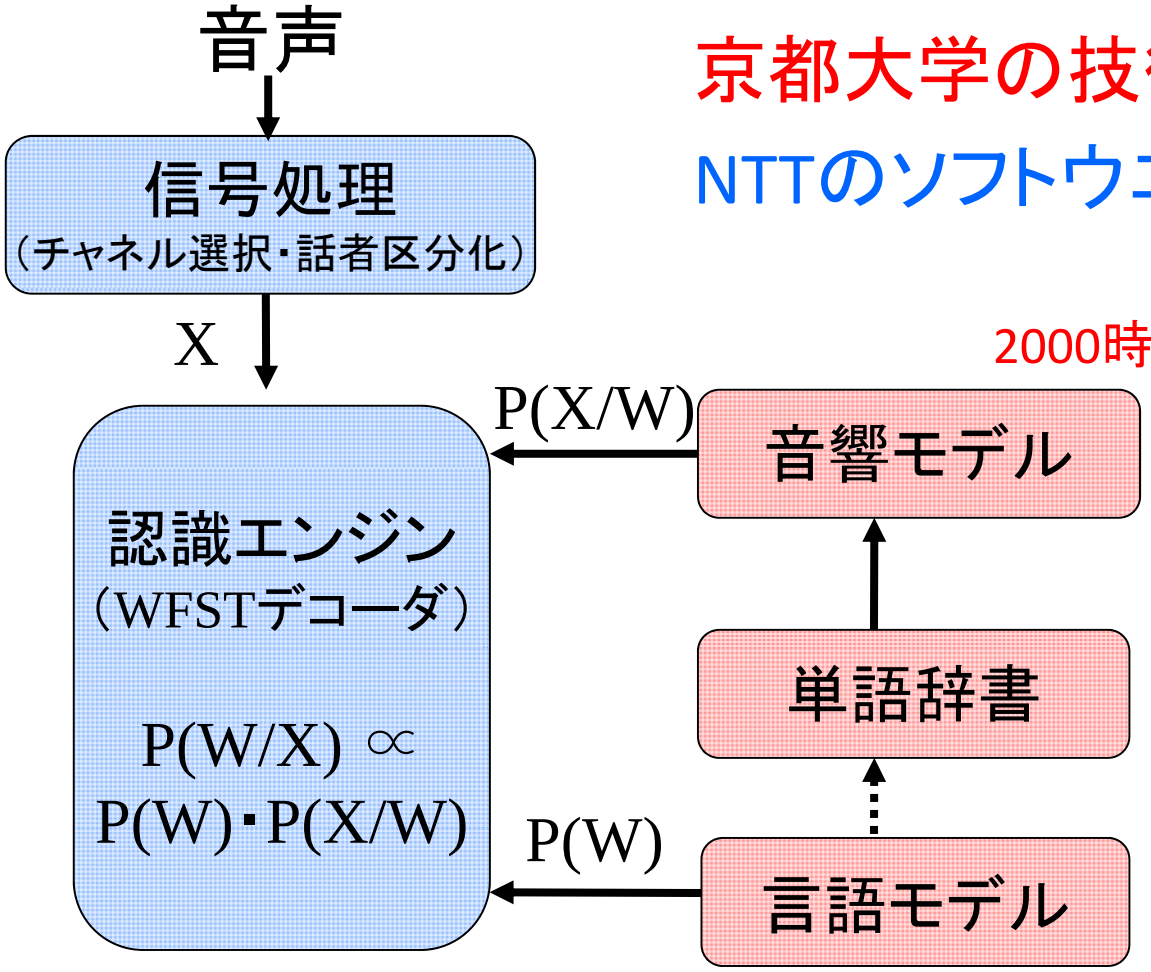
- 会議録作成支援
- 文字正解率 90%

- 2011年度から
  衆議院の全会議で運用
  （世界初・唯一）

http://www.ar.media.kyoto-u.ac.jp/diet/

# 音声認識システムの基本構成

京都大学の技術によるモデルを
NTTのソフトウエアに統合




2000時間もの会議音声データを活用

発声の怠けもモデル化
/a, i, u, e, o…/

発音の変形をモデル化
いう　y u u
いう　y u

口語的表現もモデル化
そう＋いう＋こと
これ＋が＋あー

NTT →

← 衆議院 ← 京都大学

地方議会向けパッケージも別途NTT東日本が商品化

# Smart Posterboard
# （JST CRESTプロジェクト→連携模索中）

マイク・カメラ群搭載の大型電子掲示板

- 音声・映像収録

↓

- 発話・視線検出

↓

- 興味・理解度推定
  （プレゼン支援）

http://www.ar.media.kyoto-u.ac.jp/crest/

# 言語解析の高度化⇔言語資源の構築

## 格フレーム構築

フィルタリング

構文解析

述語項構造データ

クラスタリング

超大規模ウェブコーパス
150億ユニーク文

格フレーム（4万述語 100万フレーム）

【再生する】
{ネットワーク、導入、意識、政策、人材、…}ガ
{教育、農業、医療、製造業、…}ヲ

## 同義句認識

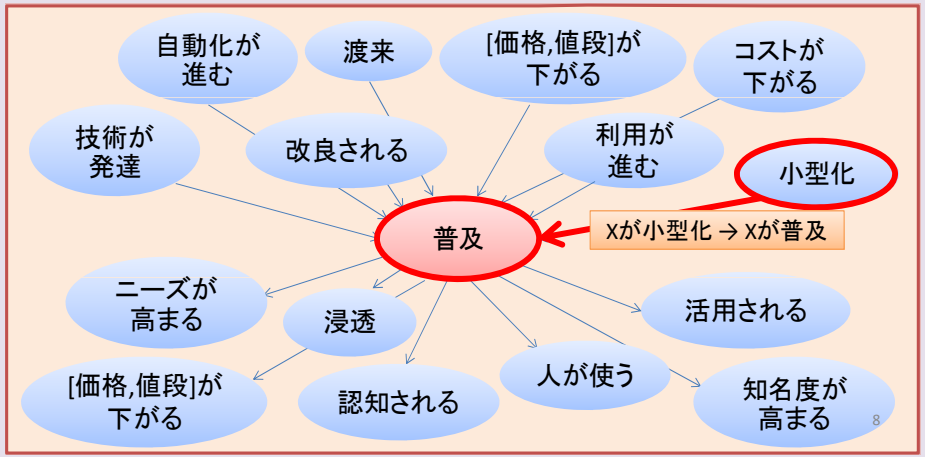
| | |
|---|---|
| 低迷し、景気が冷え込む<br>増税し、景気が冷え込む<br>弾けて、景気が冷え込む<br>・・・・ | 景気が冷え込み 株価が下がる<br>景気が冷え込み 金利を下げた<br>景気が冷え込み 困る<br>・・・・ |
| 低迷し、景気が悪化する<br>増税し、景気が悪化した<br>落ちて、景気が悪化した<br>・・・・ | 景気が悪化し株価が下落した<br>景気が悪化し厳しくなる<br>景気が悪化し金利を下げた<br>・・・・ |

文脈が類似する動詞句を抽出

**景気が冷え込む ＝ 景気が悪化する**

## 事態間知識

述語項構造ペア抽出
データマイニング（アプリオリ）
節間関係によるフィルタリング
格フレームによる項の対応付け



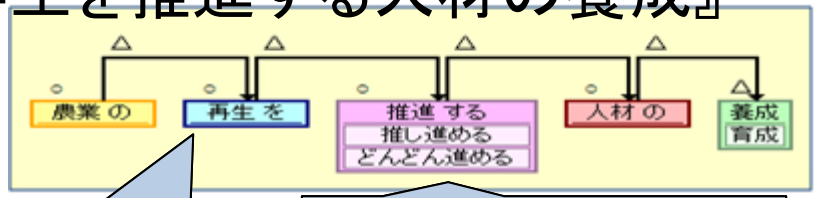

**基盤言語資源：**

京大テキストコーパス、形態素解析システムJUMAN、構文解析システムKNP

TSUBAKI による自然文検索

# 『農業の再生を推進する人材の養成』

## クエリの解析結果




| 【再生する】 | |
|---|---|
| ガ | ネットワーク, 導入, 意識, 政策, 人材, … |
| ヲ | 教育, 農業, 医療, 製造業, … |

| 【推進する】 | |
|---|---|
| ガ | 政府, 企業, 市民, 行政, 機関, … |
| ヲ | 活動, 整備, 連携, 再生, … |
| デ | 分野, 体制, 計画, 連携, 養成, … |

**農業の再生を推進する人材の養成** で検索された文書 50 件のうち 1 - 10 件目　クエリ解析結果の確認・修正

1 静岡県立磐田農業高等学校
http://www.shizuoka-c.ed.jp/iwata-ah/menu/18.html

2 「農商工連携推進」で提言書を発表（日商・東商）－日本商工会議所
http://www.jcci.or.jp/news/2010/0519180815.html

3 情報アグリビジネススクール｜農業経営IT活用支援推進事業再公募要領（情報アグリビジネススクール「ヴァ...
http://www.syskyo.or.jp/agr_school/info/16sai_kobo.html

4 地域再生人材創出拠点 ｜ 科学技術振興調整費
http://www.jst.go.jp/shincho/chikilink.html

5 豊橋技術科学大学

農業の再生を推進する人材を育成

農業再生 … 農商工連携を推進する人材の養成

食と農の再生プラン… 農業IT研修・IT指導人材の充実

新時代工学的農業クリエーター人材創出

農学的技術と思考力を有する「IT食農先導士」を養成

農業の再生を推進する人材を育成する。

育成＝養成

先進的農業地帯であるが …
生産額が停滞し、再生が切望され …

## 科研「情報爆発」（H19-23）

農業の再生を推進する人材の養成 - Google 検索

Google　農業の再生を推進する人材の養成　約 1,950,000 件 (0.09 秒)

日本語のページを検索

講座ガイド／昭和の町豊後高田観光再生協議会
www.kirari-bt.jp/guide/

[PDF] 重点分野雇用創造事業（平成23年度）
www.pref.nagano.jp/syoukou/roko/H20kikin/jyuuten/ken/03j-nourin.pdf

森林整備に直接支払、人材育成 森林・林業再生プランで最終報告｜農政 ...
www.jacom.or.jp/news/2010/12/news101210-11929.php

観光再生 … 人材育成

林業再生を推進する人材の育成

森林整備の人材育成

Google　農業　再生　推進　人材　養成　約 418,000 件 (0.22 秒)

Bag of Words

日本語のページを検索

講座ガイド／昭和の町豊後高田観光再生協議会

[PDF] スライド 1

[PDF] (3)経済・商工・観光・雇用

観光再生 … 人材育成

農業へのIT利活用 … 人材育成

ものづくりを担う人材の育成 … 農業の再生

# NTT・京大共同研究

「グローバルコミュニケーションを支える言語処理の研究」（H16～H20）

- 日本語形態素解析システムMecab
- 基本語カテゴリ・ドメイン辞書
- ブログコーパス

「次世代検索のための言語解析および言語推論の研究」（H21～H23）

「言語表現の多様性に頑健な情報抽出に関する研究」（H24～H25）

- 同義・含意・反意関係解析の評価データ

# ヤフー株式会社・京大共同研究

「日中・中日用例ベース機械翻訳の高度化・実用化のための開発研究」（H23～H24）

- Eコマースドメインでの中日機械翻訳の実用化を目指し、中日対訳コーパスの構築、翻訳エンジンの高度化を行った

「用例ベース機械翻訳の高度化・実用化のための開発研究」（H25）

- クラウドソーシングを用いた、対訳コーパスの修正、新たな言語資源の構築

# まとめ

- 京大にはパターン認識・言語音声情報処理の長い伝統があり、現在も活発な研究を展開
- デザイン学においても、成熟し始めた言語音声情報処理技術の活用が必須
- 産学連携の魅力
  - 言語資源の構築（広く公開することが難題）
  - 実データの利用、実問題への挑戦
  - フィードバック