

# 言語処理の広報と啓蒙の試み

言語資源専門委員長

井佐原均

(情報通信研究機構)

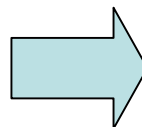
# 言語資源の適切な収集と利用に関する活動

言語資源と著作権

既存の言語資源についての情報の収集と広報

言語資源を用いた応用

言語ポータルグループ  
コーパスグループ  
応用グループ



言語ポータルグループ  
コーパスグループ

# 言語ポータルグループ

- 言語情報処理ポータル
  - 言語情報処理に関する様々な情報を提供するウェブサイト
  - 平成14年度より運営開始
  - [http://www.kc.t.u-tokyo.ac.jp/NLP\\_Portal/](http://www.kc.t.u-tokyo.ac.jp/NLP_Portal/)

# 言語情報処理 ポータル

Natural Language Processing Portal Site

[[English](#) | [Japanese](#)]

[JEITAによる調査研究報告書 ダウンロードおよびご購入  
当サイトについて/お問い合わせ](#)

(社)電子情報技術産業協会 知識情報処理技術委員会 | 47668

## What's New

- 10月26日(金)にJEITA主催の自然言語処理技術に関するシンポジウム2005が開催されます。詳しくはリンク先のページをご覧ください。(2005.9.12)
- 新刊案内を更新いたしました。(2005.9.26)

## コラム

韓国で開かれたIJCNLPから帰ってきました。そして、今、その韓国では韓国問題について、反日デモが起こっています。自然言語処理の技術が、国や文化の壁を越えた交流の助けになる日を夢見ながら、論文を読んでいます。(え)

[過去ログ](#)    ご意見は [nlp\\_portal@kc.t.u-tokyo.ac.jp](mailto:nlp_portal@kc.t.u-tokyo.ac.jp) まで

## 関連情報

- [掲示板](#)
- [言語情報処理 用語集](#)
- [講義資料リンク集](#)
- [論文DBリンク集](#)
- [人材募集](#)
- [JEITA委員会報告書・シンポジウム](#)

言語資源関係

- [アジアの言語資源・ツールのカタログ](#)
- [日本の言語資源・ツールのカタログ \(英文\)](#)

プロジェクト・研究機関・学会等

## 会議案内

- [第58回 音声言語情報処理研究会](#)  
(2005/10/21, 早稲田大学, 東京)
- [IWSLT2005 \(International Workshop on Spoken Language Translation 2005\)](#)  
(2005/10/24-25, Pittsburgh, PA)
- [JEITA 自然言語処理技術に関するシンポジウム2005 -ロボット・エージェントとの会話-](#)  
(2005/10/26, 主婦会館, 東京)
- [IBIS2005 \(第8回 情報論的学習理論ワークショップ\)](#)  
(2005/11/9-11, 早稲田大学, 東京)
- [第170回 自然言語処理研究会](#)  
(2005/11/21-22, 岡山大学, 岡山)
- [思考と言語研究会](#)  
(2005/11/25, 機械振興会館, 東京)
- [PACLIC 19 \(The 19th Pacific Asia Conference on Language, Information and Computation\)](#)  
(2005/12/1-3, Academia Sinica, Taipei)
- [The 5th NTCIR Workshop Meeting](#)  
(2005/12/6-9, Tokyo, Japan)
- [ALTW2005 \(2005 Australasian Language Technology Workshop\)](#)  
(2005/12/10-11, Sydney, Australia)
- [第7回 音声言語シンポジウム](#)  
(2005/12/21-22, 東京工業大学, 東京)
- [第171回 自然言語処理研究会](#)  
(2006/1/12-13, 国立国語研究所, 東京)
- [思考と言語研究会](#)  
(2006/1/12-13, 放送大学愛媛学習センター, 愛媛)
- [3rd International Conference of the Global Worinnet Association](#)  
(2006/1/22-26, Jeju Island, Korea)
- [IUI 2006 \(International Conference on Intelligent User Interfaces\)](#)

# 掲載情報 (前年度より継続)

- 定期更新
  - 会議案内, 製品ニュース, 新刊案内
- 不定期更新
  - 言語資源・ツールカタログ, 用語集,  
関連学会・関連機関リンク集,  
世界の言語イニシアティブ, 人材募集,  
講義資料リンク集, 掲示板

# 今年度の活動

- **ポータルサイトの新コンテンツ**
  - 論文データベースリンク集
  - コラム
- **言語処理の初心者向け紹介ビデオの作成**

# 論文データベースリンク集

- ウェブで公開されている論文DBへのリンク集
  - 研究者のサーベイ支援を目的とする
  - ダウンロード可能なサイト、論文情報(タイトル, 著者など)のみのサイトともに紹介
- 主な掲載サイト
  - 統合検索サービス (検索可能なサイト)
    - [CiteSeer](#), [PRESRI](#), [Google Scholar](#) など
  - 学会・研究会の論文データベース
    - [ACL Anthology](#), [情報処理学会電子図書館](#) など

# 言語情報処理に関連する論文データベースへのリンク集

言語情報処理に関する論文の情報を提供するサイトへのリンク集です。サーベイなどに御活用下さい。

英: 海外(英語)の論文

≡

日: 国内(主に日本語)の論文

DL: 論文をダウンロードできるサイト

## 統合検索サービス

複数の学会、研究会の論文を一括して検索できるサイトです。

- [CiteSeer](#) 英 DL

個人のサイトも含めてウェブで公開されている論文を自動的に収集し公開しているサイト。論文のキーワード検索や参照情報の閲覧ができる。

- [e-print archive \(ミラー\)](#) 英 DL

登録制の英語論文のアーカイブ。

- [PRESRI](#) 英 日

論文のキーワード検索が可能。また論文間の参照関係をグラフ表示できる。広島市立大の難波さんが運営。

- [自然言語処理 国内文献検索](#) 日

自然言語処理研究会(NL研)、情報学基礎研究会(FI研)、言語理解とコミュニケーション研究会(NLC研)、言語処理学会(論文誌、年次大会)のタイトルなどの検索ができる。長岡技科大の山本さんが運営。

- [松本研文献データベース](#) 日



# コラム

- **トップページにコラムを毎日掲載**
  - 10名のバラエティに富む執筆陣
    - 大学、公的研究機関、企業
  - 自由なエッセイ
    - 「ことば」に関連すること、それ以外のこと
- **ねらい**
  - ポータルサイトに日替わりの新鮮さを保つ
  - アクセス数の増加を期待

# 言語情報処理 ポータル

Natural Language Processing Portal Site

[JEITAによる調査研究報告書ダウンロードおよびご購  
当サイトについて/お問い合わせ](#)

## What's New

- 10月26日(金)にJEITA主催の自然言語処理技術に関するシンポジウム2005が開催されます。詳しくはリンク先のページをご覧ください。(2005.9.12)
- 新刊案内を更新いたしました。(2005.9.26)

## コラム

韓国で開かれたIJCNLPから帰ってきました。そして、今、その韓国では靖国問題について、反日デモが起こっています。自然言語処理の技術が、国や文化の壁を越えた交流の助けになる日を夢見ながら、論文を読んでいます。(え)

[過去ログ](#) ご意見は[nlp\\_portal@kc.t.u-tokyo.ac.jp](mailto:nlp_portal@kc.t.u-tokyo.ac.jp)まで

関連情報

日替わり  
過去ログも読める

- [IJCNLP2005 \(International Workshop on Spoken Lan](#)  
(2005/10/24-25, Pittsburgh, PA)
- [JEITA 自然言語処理技術に関するシンポジウム2005 -I](#)  
(2005/10/26, 主婦会館, 東京)
- [IBIS2005 \(第8回 情報論的学習理論ワークショップ\)](#)  
(2005/11/9-11, 早稲田大学, 東京)
- [第170回 自然言語処理研究会](#)  
(2005/11/21-22, 岡山大学, 岡山)
- [思考と言語研究会](#)  
(2005/11/25, 機械振興会館, 東京)
- [PACLIC 19 \(The 19th Pacific Asia Conference on La](#)  
(2005/12/1-3, Academia Sinica, Taipei)
- [The 5th NTCIR Workshop Meeting](#)  
(2005/12/6-9, Tokyo, Japan)

# NLP紹介ビデオの作成

- 言語処理技術の広報・啓蒙活動の重要性
  - 大学、大学院進学を考える学生  
研究者人口の増加
  - プロジェクトの審査員  
言語処理研究に対する資金獲得
  - わかりやすく伝えることが重要
- 言語処理技術を紹介するビデオを作成
  - ポータルサイトで公開、自由配布

# ビデオの内容

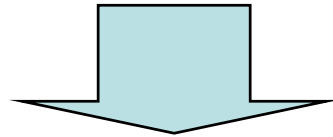
- 近未来の生活を再現
  - レストランで
    - 翻訳カメラによるメニューの翻訳
    - ウェイトレスの質問の音声翻訳
  - リビングで
    - 親子とロボットとの対話を再現
    - ロボット=エアコンの温度調整やウェブ検索を行うエージェント
- 5分程度

# 現在、鋭意製作中



# コーパスグループ

目的：言語処理技術の研究、製品開発のための  
言語資源利用と構築のガイドラインの策定



2002～2003年度

言語資源における著作権に関する調査

異なる立場の方へのヒアリングを通して、著作権に関する考え方、  
認識を確認する

2004年度

言語資源利用可能なコーパスを作るためのガイドライン

各種自由利用ライセンスの調査を通じ、言語資源として利用可能な  
コーパスを構築する際のライセンス

2005年度

言語資源利用可能なコーパスを流通させるための調査

各種の著作物流通活動・ライセンス体系についての調査を通じ、言  
語資源を流通させる際の問題点等の調査・整理

# 著作権法の解釈についての調査(2002)

- ヒアリング
  - 東芝 木村和広氏(GSK委員)
    - 「GSK著作権WG活動報告」
  - NTTアドバンステクノロジー 藤波進氏(著作権法学会)
    - 「不当と違法の狭間で」
  - 富士通 亀井正博氏(法務担当)
    - 「言語コーパスの著作権」
- 何がOK?何がNG?
  - 行為よりも結果で判断すべき

# 権利者の考え方についての調査(2003)

- ヒアリング
  - 読売新聞 川内友明氏
    - 新聞協会著作権関連委員会委員長
  - 小学館マルチメディア局 中村隆宏氏
    - 「言語資源化への取り組みと著作権処理の問題点」
  - 弁護士 小林明隆氏
    - 株式会社アルク顧問弁護士
    - 「言語資源の作成と利用をめぐる法律問題」
- 著作権侵害の有無よりも経済的側面を重視すべき



# 不特定多数の権利者との権利処理に 関しての調査(2004)

- 権利処理の上で入手可能なコーパスは、新聞等の権利者が集中しているものに限られる
  - 不特定多数の著作者のコーパスは流通困難。権利処理するには？ 調査
- 各種自由利用ライセンス(GFDL等)
  - コンテンツそのものの自由利用を許諾する体系であり、言語資源としての利用に比して幅広い許諾。
- 商用ウェブサイトでのライセンス条項
  - 運営者に対して幅広い許諾を与える条項ではあるものの、コーパスを第三者に利用許諾することを想定してはいない。

# 不特定多数の権利者との権利処理に 関しての調査(2004)

- ヒアリング
  - 弁理士 岡村信一氏
    - 権利処理のスキーム
    - 得るべき許諾内容
- 手持ちのコーパス(例えば運営している掲示板システムの過去ログ)を安心して公開するには?
  - 執筆時に許諾を求める。
  - 執筆者と公開者の免責。

 これらを盛り込んだ許諾体系案

## 2005年度の活動状況・予定

- 言語資源利用可能なコーパスを流通させるための調査
  - 各種の著作物流通活動・ライセンス体系についてのヒアリング調査
  - 言語資源を流通させる際の問題点等の調査・整理

# 応用グループ

- 自然言語処理の応用に関するユーザ調査 -

# アンケート実施の目的と概要

- 目的

- 産学協同で、長期的な観点で、ユーザニーズやユーザの期待に答える自然言語処理技術はなにか、どんな自然言語処理応用が重要なのか、を探る。

- 手法

- 自然言語処理を応用した製品に関するユーザシナリオを準備する。
  - ユーザシナリオとは、
    - アクターがいて、具体的なシーンがあって、時系列的なストーリーがあるもので、ユーザが直感的なイメージを得やすく理解しやすい。
    - 機能の記述や既存の応用の分類名ではない。それらの枠を離れられるという特徴がある。
- 上記のユーザシナリオに対して「知名度」「使い勝手」を調査するためのツリー状の質問を作成

# 2004年度に実施したアンケート調査

## 狙い

ビジネス向け用途だけではなく、より一般大衆向けの自然言語処理応用製品について模索する。今回は製品の認知度と使い勝手についての意識調査を行なう。

## 方法

ビジネスパーソン層と高齢者層と2グループに分けてアンケートを実施し、その結果を対比する。

パソコン関連に偏り過ぎず、かつ、現実の製品を題材としたユーザシナリオとする。また、自由記述の内容を個別に分析できるようフォーマットを工夫する。

## 実績

04年10月のCEATACで125件。(05年も追加実施、集計中)  
調査会社に依頼し、高齢者層へのアンケートを実施、179件を回収。

# ユーザシナリオ(1)~(3)

## 1. ペットロボット

- 飼い主の声を認識し、名前を呼ぶと反応したり、「写真を撮って」「ダンスして」というような要求に答えてくれる。

## 2. デジタル家電

- キーワードを登録すると、それに関連する番組を全て録画しておいてくれるハードディスクビデオレコーダ。

## 3. カーナビ

- ドライブで行き先を告げるとそこまでの道順を表示してくれる。途中、「近くの和風のレストラン」と言うと、もよりの店を探しそこまでの道順を示してくれる。

# ユーザシナリオ(4)～(6)

## 4. 携帯電話

- 携帯電話に向かって「鈴木さん」と言うと、鈴木さんの電話番号にかけてくれた。

## 5. 電訳機

- 「水を一杯ください」などの会話内容を選択すると対応する英語を発声してくれる電訳機が、海外旅行で役に立った。

## 6. テレホンガイド

- 粗大ゴミを捨てる時の手続きがわからず市役所のテレホンガイドに電話したら、自動音声応答により手続きの内容を詳しく書いたFAXを送信してもらえた。



# アンケートフォーマット

- 自由記述欄を重視(必須項目)
- ツリー状の選択肢により、自由記述欄を内容ごとに絞った分析が可能

(1) 【ペットロボット】飼い主の声を認識し、名前を呼ぶと反応したり、「写真を撮って」「ダンスして」というような要求に答えてくれる。

<p><b>Q1</b>：このような製品があることを知っていますか？</p> <p>→ <input type="checkbox"/> はい → <b>Q2</b>：このような製品を使ったことがありますか？</p> <p>→ <input type="checkbox"/> いいえ → <b>Q4</b>：このような製品を使いたいと思いますか？</p>	<p>→ <input type="checkbox"/> はい → <b>Q3</b>：その印象は？</p> <p>→ <input type="checkbox"/> 悪い → <b>Q5</b>：その理由は？ (必ずご記入ねがいます)</p>
<p>→ <input type="checkbox"/> 良い</p> <p>→ <input type="checkbox"/> 悪い</p>	<p>→ <input type="checkbox"/> 良い</p> <p>→ <input type="checkbox"/> 悪い</p>
<p>→ <input type="checkbox"/> はい</p> <p>→ <input type="checkbox"/> いいえ</p>	<p>→ <input type="checkbox"/> はい</p> <p>→ <input type="checkbox"/> いいえ</p>

# プロフィール集計結果

項目	ビジネスパーソン	高齢者
性別構成	男性が大部分を占める。	女性比率がやや多い。
年齢構成	30代から50代が中心	60代から70代が中心
携帯電話の用途	通話以外の用途も多い	通話用途が大部分
パソコンの用途	メール、インターネット、文書作成が半数以上	使用者は約10%で文書作成が中心

# 回答集計結果(例:シナリオ1)

	CEATEC			高齢者		
	Q1	Q2 / Q4	Q3	Q1	Q2 / Q4	Q3
パターン 1	(Q1-Y) 89人	(Q2-Y) 8人	(Q3-Y) 6人	(Q1-Y) 82人	(Q2-Y) 1人	(Q3-Y) 1人
パターン 2			(Q3-N) 2人			(Q3-N) 0人
パターン 3		(Q2-N) 80人			(Q2-N) 81人	
パターン 4	(Q1-N) 31人	(Q4-Y) 5人		(Q1-N) 97人	(Q4-Y) 16人	
パターン 5		(Q4-N) 23人			(Q4-N) 80人	

# 自由回答に対するコメント(1)

## 1. ペットロボット

飼い主の声を認識し、名前を呼ぶと反応したり、「写真を撮って」「ダンスして」というような要求に答えてくれる。

ビジネスパーソンにおける知名度が高い。さらに高齢者においては、他の製品と比べて比較的知名度が高い。半数が知っているという高率である。しかし、知名度が高い割には、使用経験者がほとんどいないことが際立つ。

知らない人の使いたくない率はどちらも同程度に高い。その理由の傾向もほぼ同じある。どちらのセグメントも、ペットにもロボットにも興味のない人が多い。

高齢者には、孫の代わり、ペット代わりというポジティブな意見もあったが、全体的に機械相手に対する違和感がある。

# 自由回答に対するコメント(2 - 1)

## 2. デジタル家電

キーワードを登録すると、それに関連する番組を全て録画しておいてくれるハードディスクビデオレコーダ。

ビジネスパーソンにおける知名度は他のシナリオの中でも一番高いが、高齢者における知名度は著しく低い。

高齢者で知らなかったと回答した人には、使いたくないという人が多い。その理由として多いのが、操作に対する不安(機械の使い方がわからない)でこれはビジネスパーソンの理由には見られないものである。両者ともに、録画しても見る時間がない、録画してまで見たいものがないというTVに対する本質的な問題に関する理由が見受けられる。

# 自由回答に対するコメント(2 - 2)

## 2. デジタル家電

キーワードを登録すると、それに関連する番組を全て録画しておいてくれるハードディスクビデオレコーダ。

ビジネスパーソンのみで見られる、知っていて使わない人の理由として、精度や仕様などに対する技術的な疑念がある。これはNLP技術に対する普遍的な疑念である。

ビジネスパーソンの使用経験者の印象は良い。必要とされている人(TVのヘビーユーザ)のニーズを適切にすくっている。

高齢者のパターン4の人(知らなかったが使いたい)の意見を見ると、そもそもビデオの録画設定が高齢者には難しいという問題があり、キーワードのみでそれができるのならば使ってみたいというものがあり、印象的であった。

# 自由回答に対するコメント(3)

## 3. カーナビ

ドライブで行き先を告げるとそこまでの道順を表示してくれる。途中、「近くの和風のレストラン」と言うと、もよりの店を探しそこまでの道順を示してくれる。

高年齢者における知名度が一番高いが「これはNLP技術によるカーナビを知っているか」という本来の質問がうまく機能していない可能性があり、単にカーナビを知っているという人が多いのではないかと思われる。

NLPに関する意見に焦点をあてると、ビジネスパーソンは曖昧な表現で音声操作できることに対するポジティブな意見も多いが、やはり性能に対する不満が多い。

使ってみたい人の意見として安全運転への期待が高い。高年齢者には「操作の手軽さへの期待」も見受けられる。

# 自由回答に対するコメント(4)

## 4. 携帯電話

携帯電話に向かって「鈴木さん」と言うと、鈴木さんの電話番号にかけてくれた。

プロフィールで見たように、高齢者の携帯電話所有率は低い  
ため、この機能を知らない人の割合はきわめて高い。

使いたくないと答えた人の割合は全体に高いのだが、ビジ  
ネスパーソンよりも高齢者の方が低い。つまり、高齢者でこ  
の機能に期待している人がある程度の割合で存在する。高  
齢者の使いたい理由として、現在の携帯電話はボタンが小  
さく押しづらい、機能が複雑である、ため声だけ済ませたい  
というものが多かった。

ビジネスパーソンの使いたくない理由は、やはり性能面  
での不安が多い。認識率が低ければキー操作の方が速くて  
確実である、などの意見が見受けられた。



# 自由回答に対するコメント(5)

## 5. 電訳機

「水を一杯ください」などの会話内容を選択すると対応する英語を発声してくれる電訳機が、海外旅行で役に立った。

高齢者は使うであろう機会(海外旅行など)が少ないためか知名度は低く、また使いたいと思う人に少ない。それに対してビジネスパーソンにおいては知らなかった人で使いたい人が多く、期待が感じられる。しかし、実用例文などのコンテンツ不足が不満として挙げられている。

# 自由回答に対するコメント(6)

## 6. テレホンガイド

粗大ゴミを捨てる時の手続きがわからず市役所のテレホンガイドに電話したら、自動音声応答により手続きの内容を詳しく書いたFAXを送信してもらえた。

知名度はどちらも高くない。高齢者においては一番低い。

高齢者においては知名度が極端に低いだけでなく、使いたい人の割合もビジネスパーソンと比べて低い。その理由は、直接聞いた方が早いといったものや、FAXの操作に対する不安などがある。また、インターネットで使いたい、ややこしそう、時間がかかりそうといったものが多かった。高齢者はそもそも市役所の開いている時間に行動できるため、テレホンガイドの必要性が低いと思われる。

一方、ビジネスパーソンにおいては利用経験者の評価が高いだけでなく、使いたい人の割合も高い。これは24時間使えるという利便性が大きい。

# まとめ

- IT技術に精通しているビジネスパーソンと高齢者という二つのセグメントを対象に、自然言語処理技術に対する認識・期待の差異を明確にすることを目的としアンケート調査、および、分析を実施。
- ビジネスパーソンは、提供する技術の精度を気にしている場合が多い。言語処理の精度が低ければ、使い勝手は却って減少する。また、全体に、操作が複雑なものに対しては音声などの技術でのちょっとしたフォローが望まれている。
- 高齢者は、技術よりも、製品・サービスをそもそも使う気になるかといった心理的な側面がまだ大きい。親しみやすさを狙って擬人化を試みても、情報弱者には全体的に機械相手に対する違和感が根強い。この壁を崩すための工夫が今後も必要。

# 言語情報処理ポータル

[http://www.kc.t.u-tokyo.ac.jp/NLP\\_Portal/](http://www.kc.t.u-tokyo.ac.jp/NLP_Portal/)