

# 語長変換を考慮したコーパス管理システム

浅原 正幸<sup>†</sup> 米田 隆一<sup>†</sup> 山下 亜希子<sup>†</sup>  
伝 康晴<sup>††</sup> 松本 裕治<sup>†</sup>

本論文では、現在開発している関係データベースを用いたコーパス管理システムについて述べる。日本語の辞書管理やコーパス管理において、一貫した語の単位（1語として認定される語の長さ）は大きな問題の1つである。この問題に対し、関係データベース上に辞書とコーパスとで語の単位を同期して管理する手法を提案する。また分野により必要とされる語の単位は異なり、1つの語の単位定義を異分野間で共有することは難しい。提案する手法では、1つの元コーパスデータに対し、複数の語の単位の情報を保持することが可能である。異なる語の単位定義間の関係を保持するために複合語の定義には構成語の依存構造を導入し、依存木による複合語の分類を可能にした。これにより、各分野で必要な語の単位を依存木の構造により区別することが可能になった。

## Maintenance Schema of Japanese POS Tagged Corpora

MASAYUKI ASAHARA,<sup>†</sup> RYUICHI YONEDA,<sup>†</sup> AKIKO YAMASHITA,<sup>†</sup>  
YASUHARU DEN<sup>††</sup> and YUJI MATSUMOTO<sup>†</sup>

In this paper, we present a use of a relational database for developing and maintaining linguistic resources. In languages that do not provide word delimitation in texts (e.g. Chinese and Japanese), consistent definition of words in the lexicon is a critical issue to build POS tagged texts. When we change the definition of word delimitation in the lexicon, we need to also modify the tagged corpora to make them consistent with the lexicon. We propose a use of relational database schema to perform these modifications in tandem. In addition, in the Japanese language, definition of word delimitation is by itself difficult, since it varies depending on the research fields. To accommodate more than one definition of word delimitation, we compose a compound word lexicon in the database. The compound word lexicon includes dependency structures of components.

### 1. はじめに

日本語および中国語はわかち書きせずに記述される言語である。このような言語に対して、品詞タグつきコーパスを構成する際、採用する語の単位（1つの語の長さ）によって1文中に認定される語境界が異なりうる。元テキストの表層に対し直接タグをつけると、語の単位の変更をとまなう場合、認定される語境界の変更をとまなうため、タグの修正作業が煩雑になる。また、分野によって使用する語の単位が異なり、語の単位そのものの共有ができない場合、1つのコーパス中に異なる語の単位の情報を持たせる必要が起こる。この場合、タグつけ領域が重なり合い、タグつけ作業

そのものが困難になる。

本研究では、これらの問題を解決するために、音声言語コーパス管理で広く導入されているスタンドオフアノテーション<sup>2)</sup>（以下「スタンドオフ」）と呼ばれる手法を導入する。スタンドオフは元のテキストとタグ情報を異なるファイルとして管理する手法のことをいう。あらゆるタグ情報を元のテキストに対するポインタとして記述する。語の単位の変更はポインタのさす先の変更のみで処理が可能になる。さらに異なる語の単位の情報のみならず、付与されるべき音韻の情報などのその他の情報が矛盾なく混在できるようになる。

品詞タグつきコーパス中の語の単位の一貫性は、そのコーパスを元にして構成する解析器の精度に重大な影響を及ぼす。コーパス中に出現する語の単位を揃えるために、コーパス中の単語と辞書とをリンクする手法を導入する。タグ情報を、コーパスに対するポインタと、辞書に対するポインタとの、2つのポインタによって表現する。これは、スタンドオフの自然な拡張

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科  
Graduate School of Information Science, Nara Institute  
of Science and Technology

<sup>††</sup> 千葉大学文学部

Faculty of Letters, Chiba University

であるといえる。

必然的に辞書が多層の単位を含むことになる。多層の単位について柔軟な定義を可能にするために、複合語と構成語との間の依存関係を持たせた複合語辞書を構成する。複合語の語の関係を二分木に抽象化し、2つの構成語へのポイントを各語が保持することにより構成を行う。

我々は、上に示した考え方にに基づき、関係データベースを利用した品詞タグつきコーパスと複合語辞書の管理手法を提案する。現在、モデルとして「RWCP テキストコーパス」<sup>10)</sup> および「IPADIC」<sup>15)</sup> に基づいて構成しているが、主たる考え方は「日本語話し言葉コーパス」<sup>9)</sup> および「UniDic」<sup>13)</sup> が求める必要要件(複数の語の単位をコーパスもしくは辞書上に管理する)を満たすことを想定している。

2章では UniDic における語の単位について概説する。3章では、関係データベース上におけるスタンドオフの構成手法について述べる。4章で、関係データベース上における複合語辞書の構成手法について述べ、最後にまとめと今後の課題について述べる。

## 2. UniDic における語の単位

『語の単位』とは認定する1つの語の長さのことをいう。本章では、コーパス管理をするうえでの『語の単位』の問題について述べる。

アプリケーションによって求められる語の単位は異なる。語形成力のある短い単位の接続により形成される、より長い複合語の単位は形態素解析器用の辞書に登録されない。これは辞書中に列挙しつくせないことによる。一方、音声情報処理の分野で求められている「読み」や「アクセント」の情報が付与しやすい単位は、上の形態素解析器用辞書に含まれる単位より長い。固有名詞などの単位はさらに長くなる。

以下では、本システムが背景に考えている UniDic における語の単位の分類について述べる。

### 2.1 単位による語の分類

UniDic 体系では、国語研短単位<sup>8)</sup>に準拠した単位設定を採用する。しかし、短単位は読みやアクセント情報がつけられる単位としては粒度が細かく、より粗い単位が求められている。また構文・意味レベルの分析・処理などのためにはさらに粗い必要がある。そこで短単位を基に、第0層～第3層からなる多層の単位設定を行う。以下、各単位について個別に説明する。

- 第0層：意味を持つ最小の単位を認定。
  - 第1層：(複合のない)単純語を認定。
- 国語研短単位に相当する。複合のない語として認

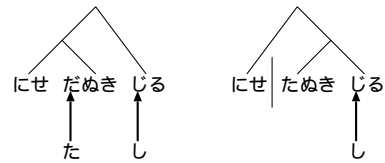


図1 依存構造と連濁

Fig.1 Dependency structures and “rendaku.”

定できる最小の単位。

- 第2層：複合語を認定。  
複合語を1語としてとらえる単位。読みやアクセント情報がつけられる単位。
- 第3層：長い名詞的表現を認定。  
情報検索の分野でいう固有表現を取り扱う。

第0層単位は「形態素」であり「語」ではないため品詞タグはつけられない。漢字は1文字で1単位を形成する。

第1層は単純語を認定する。複合語辞書の観点から見ると基底となる単位になる。他の単位は第1層の複合化により構成されると見なす。国語研短単位に相当するが、第1層単位の1次結合を第1層単位として認めないために、外来語などで国語研短単位より短い。

第2層は第1層の列から以下のように定義する：

- 以下の正規表現に適合するもの：  
(接頭辞)\*(名詞|形状詞|動詞|形容詞)+(接尾辞)\*
- かつ、すべての隣接要素対について、左要素が右要素に対して依存する関係にある。

この定義により、第2層語は内部構造として依存構造木を考えると、左分岐構造のみを持つものになる。

この依存構造に対する制約は、右分岐構造が連濁やアクセント変化を抑制するという知見<sup>1)</sup>に基づいて導入する。図1に依存構造と連濁の関係を示す。第2層単位では、左の例は1語として見なすが、右の例は2語として見なす。右の例の語の境界部分では連濁が行われぬ。このように第2層単位は読みやアクセントを認定するために適切な単位である。

第2層は音韻・形態論的な観点から設計されているとともに、第1層からまとめあげ解析するうえでチャンキングのタスクで利用されている BIO モデル<sup>5)</sup>との親和性がある。BIO モデルは語の列に対し以下のようなタグをつけることにより、より長い単位にまとめあげる手法である。以下に BIO の各タグの意味を示す：

- B：単位の開始
- I：単位の内部
- O：単位の外部

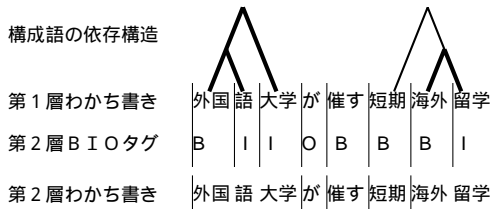


図2 第2層単位と分岐構造

Fig. 2 The second layer unit and dependency structures.

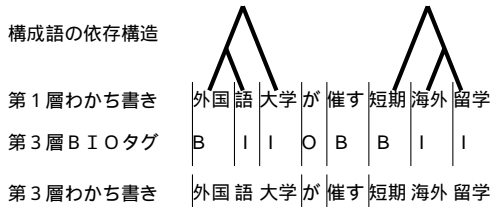


図3 第3層単位と分岐構造

Fig. 3 The third layer unit and dependency structures.

図2に第2層単位とBIOタグの関係を示す。第2層の定義にある正規表現に適合しないものにはOタグがつけられる。正規表現に適合するものはBタグもしくはIタグがつけられる。Bタグがつけられる要素は構成語の依存木を考えた場合に、ある部分木の最左要素(左枝)になる。Iタグがつけられる要素はある部分木の最右要素(右枝)になる。この意味でBIOモデルと第2層単位が求める要件とは親和性があり、第2層解析のためにBIOモデルによるチャンキングの手法が有効であると考えられる。

第3層は、意味・談話レベルの分析や情報検索などのアプリケーションなどを想定している。第3層では、第2層定義で述べた正規表現や語の依存構造を超える単位を想定しているためにBIOモデルによるタグと依存構造との関係が希薄になる。図3に第3層単位とBIOタグの関係を示す。

ここで定義した語の単位の分類は、形態素解析における語長の問題を解決するために導入する。さらに、形態素解析における未知語の問題に対して各層で異なる対処方法が必要であることが分かる。本システムは形態素解析時に各層に必要なコーパスおよび複合語辞書を構成することを目標としている。第2層語、第3層語は第1層語を基に派生が可能であるが、すべてを辞書に記述することは不可能である。複合語辞書は第2層語、第3層語解析の学習データという観点で構成される。

この体系に基づき品詞タグつきコーパスを作成する際に、上記、第1層～第3層についての個別のタグつ

```
<vp>恐れ<placename>入</vp>谷</placename>の  
鬼子母神
```

図4 オーバラップ問題

Fig. 4 Overlapping problems.

- 元テキスト “doc2.xml”

```
<p id="00">恐れ入谷の鬼子母神</p>
```

- “doc2.xml” に対するタグ情報ファイル “doc3.xml”

```
<vp xlink:href="doc2.xml#xpointer(  
  id(&apos;00&apos;);)[position=0] to  
  id(&apos;00&apos;);)[position=2]"/>  
<placename xlink:href="doc2.xml#xpointer(  
  id(&apos;00&apos;);)[position=2] to  
  id(&apos;00&apos;);)[position=3]"/>
```

図5 スタンドオフによるオーバラップ問題の解決

Fig. 5 Standoff annotation – the overlapping problems.

きコーパスを構成する必要がある。また、語の単位に関して多階層構造を持つことができる複合語辞書の形式化が求められている。

### 3. 関係データベースによるコーパス管理

本章では関係データベースによる日本語品詞タグつきコーパスの管理手法を例示する。最初にスタンドオフアノテーションについて述べ、次に関係データベース上における品詞タグつきコーパス管理手法を示し、最後に関連研究について述べる。

#### 3.1 スタンドオフアノテーション

複数の語の単位を1つのタグ情報ファイル中に管理するとオーバラップ問題が起こる。オーバラップとはタグつけする個所が一部重なる現象であり、図4のような問題のことをいう。XMLの枠組みでは、図のようなタグつけは妥当ではない。

GDA<sup>14)</sup>ではこの問題について、複合エレメントという枠組みで対処している。他の対処法としてスタンドオフアノテーション<sup>2)</sup>がある。スタンドオフは、元のデータに対するポイントとして付加情報を付与していくことにより、元テキストとタグ情報ファイルを分割して保持する手法である。図5にXpointer<sup>6)</sup>による解決例を示す。元テキスト “doc2.xml” に対し、タグ情報ファイル “doc3.xml” は分離されている。タグ情報ファイル中の最初のタグは、文字位置0から2の部分でvpであることを示している。

#### 3.2 日本語品詞タグつきコーパスに対するスタンドオフの導入

我々は、スタンドオフを日本語品詞タグつきコーパ



図 6 複数の語の単位によるスタンドオフ

Fig. 6 Standoff annotation—three definitions for word delimitation.

スを管理していくうえで生起する3つの問題を解決するために採用する．1つめは，1つの書き起こしデータに対し複数の語の単位定義を柔軟に管理すること，2つめは，他の音韻的情報についてのタグ情報との混在を無理なく許すこと，3つめは，書き起こしデータ中に存在するすべての単語の単位定義に一貫性を持たせるために辞書とコーパスをリンクすることである．

図6に日本語品詞タグつきコーパスに対するスタンドオフ手法を示す．元テキストに出現する各文字に文字位置を想定する．各タグファイルは，タグづけする部分の文字開始位置，文字終了位置，タグ情報の3つ組の情報を持つ．これらの情報は関係データベース上の異なるテーブルに保持する．

これにより，1つのテキストに対し，複数の語の単位についての品詞タグ情報の管理が可能になる．さらにスタンドオフは，コーパスに他の情報を付与することが想定されている場合に，他の情報に対して相互に制限を加えないという点で有効である．

コーパスを整備する際，語の単位の粒度を揃えることは難しい．この問題に対処するために，コーパス中に出現する単語はすべて，登録されている単語辞書とリンクを張る．図7に単語辞書へのリンクを導入したコーパス管理手法を示す．実際の単語辞書は図に示された「表記」と「品詞」以外に「活用形ID」，「語形」，「発音」と活用に関する情報を持つ．関係データベース上では，単語辞書の主キーを「単語ID」と「活用形ID」に関して設定し，タグ情報を単語辞書中のこの主キーに対するポインタで表現する．

最後にポインタの接地地点に関する注意点について述べる．書き言葉のコーパスの場合には，ポインタの接地に文字位置を利用するのに対し，音声言語コーパスの場合には，ポインタの接地に音声データにおける開

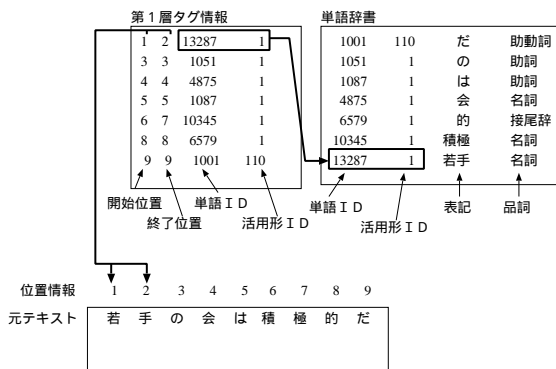


図 7 単語辞書へのリンク

Fig. 7 Standoff annotation—link between original text and lexicon.

始時間と終了時間で行う手法<sup>12)</sup>を用いる．

### 3.3 関連研究

スタンドオフは，SGMLの分野で形式化<sup>2)</sup>された．XMLの枠組みではXpointer<sup>6)</sup>として形式化されている．コーパスに対するタグづけの分野ではCES (Corpus Encoding Standard)<sup>3)</sup>に由来するXCES<sup>4)</sup>などで広く採用されている．文献7)は，この手法から発展して，タグ情報をグラフ構造として表現するアノテーショングラフという手法を提案している．

## 4. 複合語の取り扱い

前章では，1つのテキストに対し異なる語の単位についてのタグ情報を管理する手法について述べた．本章では，異なる語の単位間の関係の管理手法について述べる．

複合語中の構成語の依存構造を考えると依存木を作ることができる．まず，可能な依存構造に関する複合語の分類について示し，次に依存構造を関係データベース上で管理する手法について述べる．

### 4.1 複合語の分類

長い単位と短い単位との関係を管理するために，複合語と構成語間の依存構造に着目する．複合語に関して二分木の依存構造があると想定し，長い単位と短い単位との関係を二分木上の親子関係と見なす．依存構造の分類によって，複合語カテゴリのタグを，表1のように定義する．

以下，各タグをどのように利用するか，事例を示しながら解説する．

#### 4.1.1 複合語の二分木構造表現

ある語が2つの構成語に分割可能で，構成語の隣接要素が依存構造を持つ場合には，タグ「N」をつけ，複合語は2つの構成語に対するポインタにより，短い

表 1 複合語カテゴリ

Table 1 Categories for compound words.

複合語カテゴリ	分類
B	基底
N	複合化
P	複合化(並列)
C	複合化(縮約)
X	複合語の一部にしか出現しない単位
S	依存構造を持たない連語単位

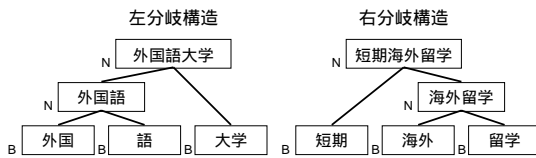


図 8 複合語の左分岐構造と右分岐構造

Fig. 8 Left and right branching structures of compound words.

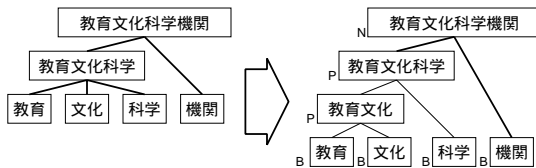


図 9 三分木構造の二分木化

Fig. 9 Binarization of ternary structure.

単位との関係を張る．構成語がさらに短い単位へと再帰的に分割可能な場合には，その構成語についても同様の操作を行う．

左右の構成語のどちらが再帰的に分割されているかにより，複合語が左分岐構造と右分岐構造のどちらを持つかを判別することが可能である．この判別は UniDic における第 2 層単位認定をする際に重要な情報となる．図 8 に，複合語の左分岐構造と右分岐構造の例を示す．

#### 4.1.2 複合語の三分木構造表現

並列などにより構成語の依存木中に三分木以上の枝分かれを持つ木が出現する場合がある．しかし提案する仕様では，複合語に対して二分木のみ定義することを許す．そこで仮想的に左分岐構造を持つ二分木に展開し，三分木以上の構造として展開されるノードにはタグ「P」をつける．一般に n 分木は，左分岐構造で表現することにする．図 9 に二分木構造の三分木化の例を示す．

#### 4.1.3 縮約形態

話し言葉特有の複合語形成として，縮約形態がある．縮約表現「ちやう」は，構成語「て」と「しまう」の 2 語が音韻的に縮退した形態である．提案した手法で

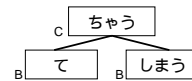


図 10 縮約形態「ちやう」

Fig. 10 Contracted word “chau.”

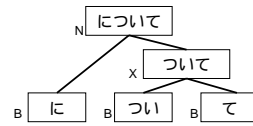


図 11 連語「について」

Fig. 11 Case particle collocation “ni-tuite.”

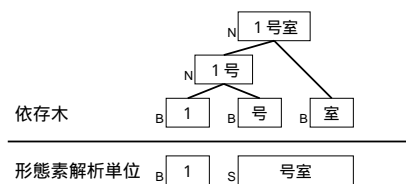


図 12 「号室」と依存構造

Fig. 12 Suffix “gou-shitsu” and dependency structure.

は，縮約形態についても縮約前の構成語を持たせることができる．このような縮約形態には「C」タグをつける．図 10 に縮約形態の例を示す．

#### 4.1.4 連語，イディオム，固有表現

連語やイディオム，固有表現などの場合，中間ノードに，すべての単位で認定する必要がない語が出現する場合がある．その場合，その認定しない中間ノードについて「X」タグをつける．図 11 は格助詞のように利用される連語「について」の例である．「ついて」はすべての単位で認定しない語として「X」タグをつける．

#### 4.1.5 依存構造を持たない連語単位

既存の辞書には，本来その語が持つべき依存構造の定義には合わない語が存在する．これらの語は形態素解析を精度を上げるために登録されている．

「号室」は本来数値をともなって，図 12 のような依存木を構成する．このため「号室」を単語として辞書登録し，その構成を記述することは，単語内の依存構造の定義と合わない．このように依存構造を持たず，かつ並列構造も持たない，解析の便宜のために登録される単位には「S」タグをつける．

#### 4.2 関係データベースによる複合語の取扱い

語の間の関係をポインタで表す手法について示した．関係データベース上では，構成語へのポインタをそのままテーブル上に保持することにより複合語を扱う．表 2 に関係データベース上における複合語の表現手法を示す．簡単のために表中では「単語 ID」により

表 2 関係データベース上における複合語表現手法

Table 2 The representation of compound word on relational databases.

単語 ID	複合語カテゴリ	表記	左構成語 ID	右構成語 ID
1	N	教育文化科学機関	2	7
2	P	教育文化科学	3	6
3	P	教育文化	4	5
4	B	教育	-	-
5	B	文化	-	-
6	B	科学	-	-
7	B	機関	-	-

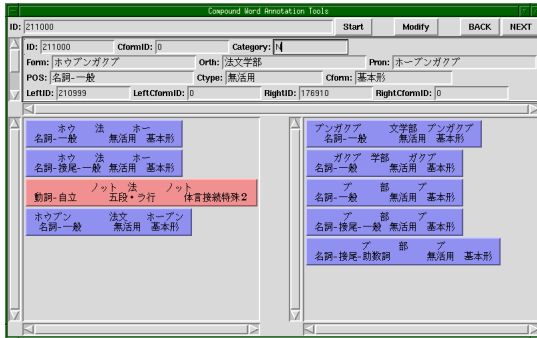


図 13 複合語整備 GUI

Fig. 13 GUI for compound word maintenance.

語に対するリンクを作成しているが、実際には「単語 ID」と「活用形 ID」の2つ組に対しリンクを張る。

このデータベース上での複合語整備のために図 13 のような GUI を構成しタグを付与している。データベースを検索することにより、左下のウィンドウにはプレフィックスとして合致する可能な構成語が、右下のウィンドウにはサフィックスとして合致する可能な構成語が列挙される。アノデータは列挙された語から最も良い組合せを選択し、ある単位の語をより細かい 2 つの語に分割する作業を再帰的に行う。

## 5. ま と め

関係データベース上で日本語品詞タグつきコーパスと辞書を管理する手法を提案した。本手法は UniDic が求める複数の語の単位保持を実現した。

スタンドオフアノテーションを導入することにより複数の語の単位を同時に管理することができる。コーパスと辞書とをリンクする手法により語の単位の粒度を揃えることが可能になった。また本手法では他のタグ情報と相互に干渉せずに混在させることについても想定している。

複数の語の単位間の関係を管理するために、複合語の依存構造を格納する電子化辞書の構成手法について提案した。これにより、語の単位の分類作業の煩雑さを軽減することができる。複合語は構成語に対するポ

インタを張ることにより構成している。この整備されたデータにより、より複雑な複合語解析が可能になると考えている。

本手法では XML データベースではなく、関係データベースでコーパスを管理することについて重点を置いて説明した。これは言語の表層が線形であることを重視していることによる。話し言葉の音声データは時間軸で表現される表層を持ち、書き言葉は文字位置（もしくはバイト位置）で表現される表層を持つ。提案する手法では、各タグは、この線形の表層上の長さ 0 を許すある区間を切り取り、その区間について情報を付与する。異分野のタグ間の重なりは、区間を共有するか否かにより識別され、これは関係データベース上での問合せで十分実現することが可能である。XML の枠組みは配布時のみ限定的に採用するが、コーパス管理の観点では上記理由で関係データベースを中心に置く。

## 参 考 文 献

- 1) Kubozono, H.: *The Organization of Japanese Prosody*, Studies in Japanese Linguistics, Vol.2, くろしお出版 (1991).
- 2) Thompson, H.S. and McKelvie, D.: Hyperlink semantics for standoff markup of read-only documents, *SGML Europe '97* (1997).
- 3) Ide, N.: Corpus Encoding Standard: SGML Guidelines for Encoding Linguistic Corpora, *Proc. 1st International Language Resources and Evaluation Conference* (1998).
- 4) Ide, N., Bonhomme, P. and Romary, L.: XCES: An XML-based Encoding Standard for Linguistic Corpora, *Proc. 2nd International Language Resources and Evaluation Conference*, Paris, European Language Resources Association (2000).
- 5) Sang, E.F.T.K. and Buchholz, S.: Introduction to the CoNLL-2000 Shared Task: Chunk-

長さ 0 の区間を許すことによりデータに表出しない現象(ギャップ)についてのタグ付けを許す。

- ing, *Proc. CoNLL-2000 and LLL-2000* (2000).
- 6) DeRose, S., Maler, E. and Daniel Jr., R.: XML Pointer Language (XPointer) Version 1.0 (2001). <http://www.w3.org/TR/xptr/>
  - 7) Bird, S. and Liberman, M.: A formal framework for linguistic annotation, *Speech Communication*, Vol.33, pp.23-60 (2001).
  - 8) 小椋秀樹：話し言葉コーパスの単位認定基準について，第1回話し言葉の科学と工学ワークショップ講演予稿集，pp.21-28 (2001).
  - 9) 国立国語研究所，通信総合研究所，東京工業大学：日本語話し言葉コーパス—モニター版 (2001)．
  - 10) 新情報処理開発機構 (RWCP) テキスト・サブ・ワーキンググループ：研究開発用知的資源：タグ付きテキストコーパス報告書 (1998).
  - 11) 伝 康晴，浅原正幸：リレーショナルデータベースによる統合的言語資源管理環境，第1回話し言葉の科学と工学ワークショップ講演予稿集，pp.77-84 (2001).
  - 12) 伝 康晴：音声談話データの記述・管理・分析，*日本語学*，Vol.20, pp.10-19 (2001).
  - 13) 伝 康晴，宇津呂武仁，山田 篤，浅原正幸，松本裕治：話し言葉研究に適した電子化辞書の設計，第2回話し言葉の科学と工学ワークショップ講演予稿集，pp.33-38 (2002).
  - 14) 橋田浩一：GDA 日本語タギングマニュアル草稿第0.65版 (2001). <http://www.i-content.org/gda/>
  - 15) 松本裕治，浅原正幸：IPADIC ユーザズマニュアル，Nara Institute of Science and Technology (2001). manuscript.
  - 16) 山田 篤，内元清貴，野畑 周，井佐原均：話し言葉コーパスへの音声情報付与，第2回話し言葉の科学と工学ワークショップ講演予稿集，pp.27-32 (2002).

(平成 13 年 11 月 15 日受付)

(平成 14 年 4 月 16 日採録)



浅原 正幸 (学生会員)

1998 年京都大学総合人間学部基礎科学科卒業。同年奈良先端科学技術大学院大学情報科学研究科博士前期課程入学。2001 年同大学博士後期課程進学。同年より日本学術振興会特別研究員，現在に至る。自然言語処理の研究に従事。言語処理学会学生会員。



米田 隆一

2001 年東京大学文学部言語学専修課程卒業。同年奈良先端科学技術大学院大学情報科学研究科博士前期課程入学，現在に至る。計算言語学の研究に従事。



山下亜希子

2001 年奈良女子大学理学部情報科学科卒業。同年奈良先端科学技術大学院大学情報科学研究科博士前期課程入学，現在に至る。自然言語処理の研究に従事。



伝 康晴

1988 年京都大学工学部電気工学第二学科卒業。1993 同大学院工学研究科博士後期課程研究指導認定退学。ATR 音声翻訳通信研究所研究員，奈良先端科学技術大学院大学情報科学研究科助教授を経て，2000 年 10 月より千葉大学文学部助教授。京都大学博士 (工学)。専門は，音声談話分析，心理言語学，計算言語学。日本認知科学会，人工知能学会，言語処理学会，社会言語科学会，日本心理学会各会員。



松本 裕治 (正会員)

1955 年生。1977 年京都大学工学部情報工学科卒業。1979 年同大学院工学研究科修士課程情報工学専攻修了。同年電子技術総合研究所入所。1984~1985 年英国インペリアルカレッジ客員研究員。1985~1987 年 (財) 新世代コンピュータ技術開発機構に出向。京都大学助教授を経て，1993 年より奈良先端科学技術大学院大学教授，現在に至る。京都大学工学博士。専門は自然言語処理。人工知能学会，日本ソフトウェア科学会，言語処理学会，認知科学会，AAAI，ACL，ACM 各会員。