

TR-I-0193

SL-TRANS における
音声認識のための構文規則の概要
Syntactic Rules for Speech Recognition
in SL-TRANS

保坂順子 竹沢寿幸
Junko HOSAKA Toshiyuki TAKEZAWA

1991年2月14日

要旨

音声言語日英自動翻訳実験システム SL-TRANS における音声認識部に適用される構文規則を作成した。構文規則は、文節を単位にして文脈自由文法の枠組に沿って構築した。本報告では、文節を組み立てる規則と、その文節を基に文を組み立てる規則について述べる。

ATR自動翻訳電話研究所

ATR Interpreting Telephony Research Laboratories

©ATR自動翻訳電話研究所 1991

©1991 by ATR Interpreting Telephony Research Laboratories

もくじ

1	はじめに	6
1.1	背景	6
1.2	経過	7
1.3	文法開発へ向けて	8
1.4	規則の記述	10
2	文節に基づき文を組み立てる規則	12
2.1	文とは	13
2.2	述語句	15
2.2.1	方向性のある述語句	15
2.2.2	方向性のない述語句	17
2.3	副詞句	20
2.3.1	副詞からなる副詞句	21
2.3.2	方向性のある述部を含む副詞句	22
2.3.3	方向性のない述部を含む副詞句	23
2.4	名詞句	24
2.4.1	方向性のある名詞句	25
2.4.2	方向性のない名詞句	27
2.5	連体修飾句	27
2.5.1	助詞「の」を伴うもの	28
2.5.2	連体詞	28
2.5.3	活用語の連体形	29
2.5.4	複合格助詞	29
2.6	前処理を行なうもの	30
2.6.1	住所	30
2.6.2	金額	30
2.7	文節のタイプ	33

2.7.1	感動詞文節	34
2.7.2	接続詞文節	34
2.7.3	述語文節	34
2.7.4	副詞文節	37
2.7.5	名詞文節	43
2.7.6	連体修飾文節	44
2.8	助詞の「で」「なら」と助動詞の「で」「なら」	47
2.9	NADINE の CFG 規則	49
3	文節を構成する規則	50
3.1	名詞	50
3.1.1	固有名詞	50
3.1.2	サ変名詞	52
3.1.3	形容名詞	53
3.1.4	普通名詞	54
3.1.5	数詞	57
3.1.6	代名詞	61
3.2	動詞	62
3.2.1	本動詞	63
3.2.2	補助動詞	65
3.3	形容詞	70
3.4	副詞	71
3.5	連体詞	72
3.6	接続詞	72
3.7	間投詞	72
3.8	感動詞	72
3.9	助動詞	73
3.10	助詞	76
3.10.1	格助詞	77
3.10.2	準体助詞	77
3.10.3	係助詞	77
3.10.4	副助詞	77
3.10.5	並列助詞	78
3.10.6	接続助詞	78
3.10.7	終助詞	79

3.11 接頭辞	79
3.12 接尾辞	79
3.13 例外規則	82
4 おわりに	84
4.1 構文情報の音声認識への応用	84
4.2 展望	86
A HMM-LR 文法記述用音素対応表	87
B 文法	89
B.1 バージョン 1.0	89
B.2 バージョン 1.1	91
参考文献	99

図一覧

1.1	一文節中の単語数	8
1.2	一文中の文節数	9
4.1	構文規則の制約としての強さ	85

表一覧

4.1	文法のサイズとパープレキシティ	84
4.2	実験対象	85
4.3	文節認識実験結果	86
4.4	文認識実験結果	86

第 1 章

はじめに

日本語の音声認識は、孤立発声単語の処理から文単位連続音声の処理へ研究対象が発展しつつある。この中間段階として現在文節単位で発声された連続音声の処理が試みられている。そのような連続音声認識を実現するには、次の音素の予測のために構文規則を適用することが有効である。本稿では、そのために現在開発中である文節内の単語の連鎖制約、および文節を基に文を組み立てる規則について報告する¹。

1.1 背景

文節は橋本により導入された概念である [1]。橋本は、文を分解して最初に得られる単位は、単語ではなく、いくつかの単語が結合したものであることを指摘し、これを文節と名付けている。さらに、この文節は音韻的に一定の特徴を持っていることを説いた²。音声認識において、文単位の処理へ進む前段階として、文節が取り上げられるのは文節という概念が学校文法で扱われているのに加え、これが音韻的特徴を表しているためであろう。

さらに橋本は連文節 [3] という考えを導入し、文節を基にした文の構成の説明を試みた。

日本語を対象とした自然言語処理でも文節単位の処理が行なわれている。その場合、文の処理には文節間の係り受けを使うことが多い [4]。これは、日本語では文節出現順序が比較的自由なため、係り受け関係のように、意味を中心に処理を進める方が有利だと考えられているからである。最近では、ICOT で開発されている談話理解システム DUALS [5] の日本語解析部でも文節を基にした処理が取り入れられている³。これらはいずれも書き言

¹文節内の単語の連鎖制約、および文節を基に文を組み立てる規則は、いずれもモデル会話 AB, 1-5 に対処している。

²服部は文節の音韻的特徴は必ずしも一定でないことを指摘している [2]。

³DUALS 3.5 バージョンでは活用語の活用部も含め約 400 単語の語彙登録があり、10 文を扱えるようになった。

葉を対象に、主に文法作成者の直感に頼り開発されてきた。

音声認識では、認識率向上のため、文法的制約を用いることがある。日本語音声認識においては、語順の制約がきつい文節内は構文規則を使い、文節間には係り受けを用いて文の整合度を調べるという処理が行なわれている [6]。また、最近では、文節を認識単位としない構文的制約の利用も試みられている [7]。

ATR では、文節単位の音声認識処理を行なっている。そこで、文節内の連鎖制約規則と、文節を基に文を組み立てる制約規則という二段構えで構文規則を開発している。この構文情報は音声認識において、次の音素の予測および制約として利用されるので、タスクに合わない文節または文の生成は極力避けるのが望ましい。これは、従来のように文法作成者の直感だけに頼っていたのでは、非常に達成し難い。そこで、我々は、対話データベース [8] を利用して理想的な構文規則の構築を目指している。さらに、対話データベースには、書き言葉に近いキーボード会話と、話し言葉である電話会話が収録されているので、対話データベースを利用することにより、書き言葉と話し言葉との違いが明確になることが期待される。

1.2 経過

音声認識パーザ HMM-LR[9] の音素予測には、文脈自由文法を LR テーブルに展開したものを使っている。日本語の文脈自由文法としては、ICMT⁴で開発された文法 [10] に手を加えたものが利用されていた。ICMT の文法は、日本語と英語で行なわれる医者と患者の会話を機械翻訳で処理するために開発されたものである。これは語彙機能文法 (LFG)[11] に似た記述を許しており、文を単位にしている。ATR では、この文法の句構造規則を抽出し、文節単位の処理ができるように手を加えて使用していた。この文法を文節単位の音声認識に応用した結果、以下のような文節候補が得られた:

認識結果 1 *kibou-sa-re-re-re-ru*

認識結果 2 *oku-ra-se-re*

上記のような文節は、言葉で遊んでいる時、または、吃る時以外には、ほとんど発声されないと思われる。音素の予測、または、制約として構文規則を用いる場合には、通常使用されている言い回しを扱うべきであろう。

ている。

⁴International Center for Machine Translation の略称。

また、入力が構文的に正しい機械翻訳と異なり、音声認識では入力の正確さが保証されていないので、制約を厳しくする必要があると思われる。

1.3 文法開発へ向けて

言葉は無限に生成される可能性を秘めている。しかし、実際の運用を考えた場合、声帯の酷使または思考の限界などの制約により、無限に続く文は発声されない。一方、音声認識では、構文規則は制約として使うのであり、その規則が与えられたタスク以上に文を過剰生成することは望ましくない。

ATR では「国際会議に関する問合せ」をドメインとして自動翻訳の研究を進めている。また、このドメインに関する会話を収集して、データベース化している。この対話データベースで使われている一文節中の形態素数(単語数)、および一文中の文節数 [12] は、それぞれ図 1.1、図 1.2 のような分布になっている。頻度は正規化を行なっている。

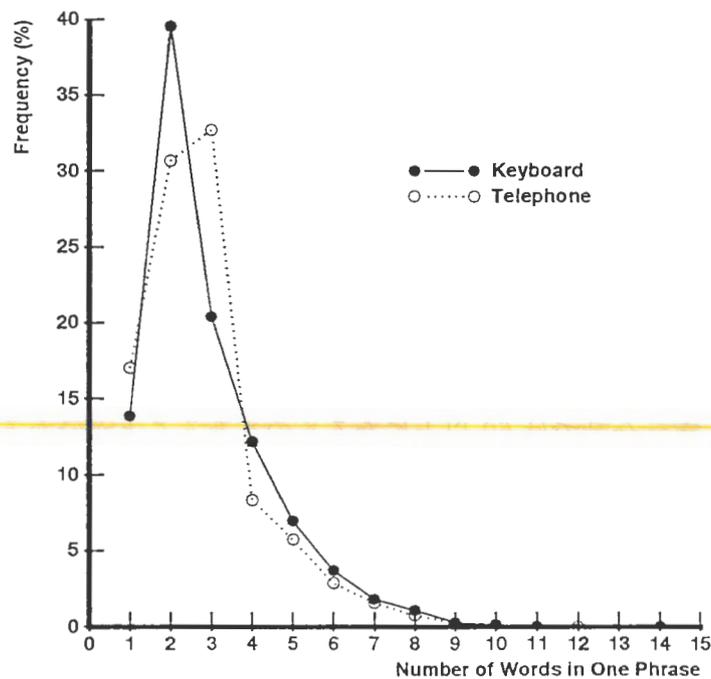


図 1.1: 一文節中の単語数

図 1.1 から、一文節中の形態素数は、最高5つまで扱えば、かなりの coverage が得られることが分かる。この範囲を扱うには、文節内の接続規則

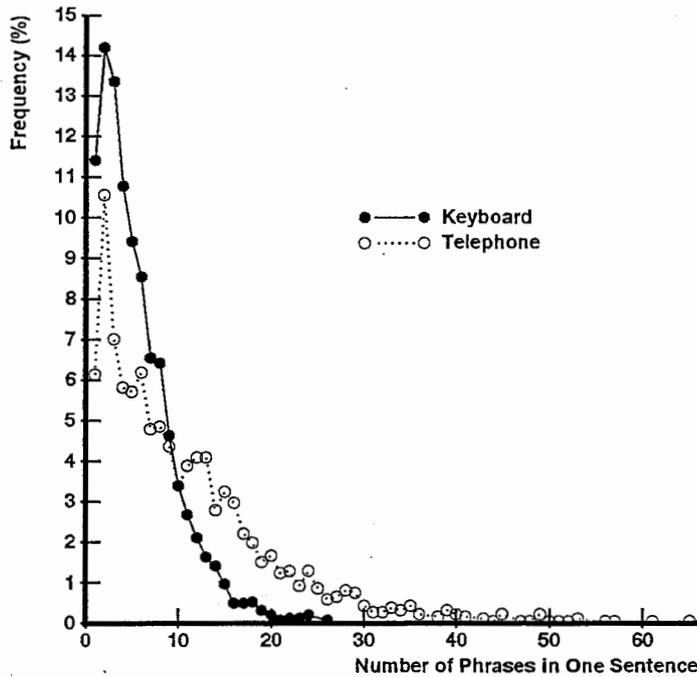


図 1.2: 一文中の文節数

は、文脈自由文法で記述せず、付属語の表現パターンを用意しておき、自立語を適当な位置に挿入するという方式も考えられる。また、この統計では、名詞句や副詞句とその他の活用語を含む文節を分けずに処理しているが、活用語を含む文節、特に文末表現を別に処理してみる必要があると思われる。

図 1.2 から、一文を形成する文節数は、キーボード会話では約 20 までであるのに、電話会話では 60 以上になるものもあることが分かる。これは、言い直し、いい淀みなども一文節として扱われているためであろう。文節数の多い電話会話の一例を以下に示す⁵:

[あの] 実は、[えー](らい)[え] このほど [あのー] (確か) 確か来週ですか、アメリカのダラスの方で開催されます、アイトリブルイーの学会に [あの] 先生の方がですね、[えー] 参加されるという風に実は [あの] (こ) 国内の (企画) [えー] (こ) 国際コンピューター会議企画委員の先生から聞いたんですけども [えー] 実はですね、[あのー] 非常にご足労をかけるんですが、まず [あの] 私共 (い) 現在 [あの] 国際コンピューター会議の [え] ファース

⁵[] は問投詞を示し、() は言い間違いを示す。

トアナウンスメント及びポスター等ですね、そういった広報物を [あの一] も
う作成いたしまして、 [ま] 国内及びそれに関連する団体にですね [あの] 配
布中なんです。

図 1.1 および図 1.2 の統計からも、無限の生成可能性を秘めている日本語
も、実際にはその運用がかなり限られていることが分かる。

我々は、音声認識のための構文構築にあたり、実際に使用されている日
本語の言い回しを扱うことを目指した。また、認識結果 1、2⁶ のような音
声認識を防ぐため、リカージョンの使用には特に注意した。

1.4 規則の記述

第 2 章以降で説明する構文規則は、文脈自由文法の枠組で書かれている。
以下、「名前が」という文節を生成する規則を例に、文脈自由規則記述用
フォーマットの説明をする。

- 1 (<start> <--> (<_start>))
- 2 (<_start> <--> (q1 <bunsetu> q2))

- 3 (<bunsetu> <--> (<np>))
- 4 (<np> <--> (<n> <p>))
- 5 (<n> <--> (<n-hutu>))
- 6 (<p> <--> (<p-kaku-ga>))
- 7 (<n-hutu> <--> (n a m a e))
- 8 (<p-kaku-ga> <--> (ng a))

規則 1 は、HMM-LR パーザからスタートシンボルとして指定されている。
双方向の矢印 <--> は、ここでは左辺のものが右辺のものに展開され
るという意味を持っている。

規則 2 の右辺で使われている q1 及び q2 は無音を表している。この無音
には含まれているところから、構文規則の本体が始まる。ここでは、
bunsetu という非終端記号を与えている。非終端記号は < > で囲ま
れる。

規則 3 は、非終端記号 bunsetu が np という非終端記号に展開されるとい
うことを表している。

規則 4 は、非終端記号 np が、非終端記号 n と p に展開されることを表し
ている。

⁶ 1.2 節、1 ページ参照。

規則 5 は、非終端記号 n が、非終端記号 n-hutu に展開されることを表している。

規則 6 は、非終端記号 p が、非終端記号 p-kaku-ga に展開されることを表している。

規則 7 は、非終端記号 n-hutu が、終端記号 n a m a e に展開されることを表している。終端記号は、< > で囲まれない。

規則 8 は、非終端記号 p-kaku-ga が、終端記号 ng a⁷ に展開されることを表している。

文節を基に文を組み立てる規則では、その終端記号は、文節の種類の名前である。一方、上記のような文節を組み立てる規則で使われている終端記号は、辞書登録である。規則 7, 8 から分かるように、辞書には、音素の列が登録されている⁸。

文節に基づき文を組み立てる規則と文節を構成する規則は、それぞれ、以下の bun7.gra と dp10-re72.gra である:

```
atr-dp:/ifg1/SI/Grammar/bun7.gra
```

```
atr-dp:/ifg1/SI/Grammar/dp10-re72.gra
```

⁷“ng” は鼻音化された “g” を表す。

⁸この音素表記は、川端モデルと呼ばれている HMM-LR 音素モデルで扱っている音素表記に準じている。付録 A 参照。

第 2 章

文節に基づき文を組み立てる規則

日本語では、文節の出現位置が比較的自由であり、文節の省略も頻繁に行なわれる。CFG 規則で、これらの可能性をすべて列挙すると、文節の出現位置についての制約を全く与えないのと同等の効果しか得られない。音声認識において、構文情報は容認性の低い音素列を排除するために使われるのであり、認識された文の構造がどのように解析されるかは、現段階では自然言語処理の構文解析ほどの重要性を持たない。我々は、構文情報利用の効果が発揮されるよう、音声認識において特に誤認識されやすい文法要素に制約を加えることにした。

誤認識されやすい文法要素を調べるため、まず制約の緩い規則、バージョン 1.0¹ を使って実験した。このバージョン 1.0 の中で、文として考慮しているのは、感動詞からなるもの、接続詞で始まり述語文節で終るもの、接続詞を伴わず述語文節で終るものの 3 種類である。終端記号としては、前もって構築された文節内の規則を利用した。文節タイプとしては、名詞文節、副詞文節、連体修飾文節、接続助詞で終る述語文節、文末の述語文節、引用を扱う文節など約 25 種類がある。この文法は、文節単位の音声認識で候補になった文節のすべての組合せを許したもののうち 80 % 以上を文として容認した。

この実験の結果得られた文候補を、特に、文として容認され難いものにしてほしいと思われる要素のうち、文脈自由文法で制約を加えられそうなものを以下に示す。

- 接続助詞終止の文末文節
- 文頭の述語文節
- 二重対象語構文²

¹付録 B.1 参照。

²日本語では、一文中で二つ以上の対象語をとらないと言われている [13]。

- 数詞を含む文節
- 形式名詞を含む文節
- 格助詞「へ」を含む文節
- 並列の助詞「や」を含む文節

以下紹介する、文節を基に文を組み立てる規則バージョン 1.1³では、接続助詞終止の文末文節、文頭の述語文節、数詞を含む文節、形式名詞を含む文節、格助詞「へ」を含む文節に対処している。また、この改良に伴い、文節内の規則にも手を加えている。

規則の中で使われている大文字のカテゴリー名は非終端カテゴリーを表し、小文字のカテゴリー名は、終端カテゴリーを表している。終端カテゴリーは、文節内の規則とのインターフェイスであり、文節のタイプを表している。

これからの課題 1 ダブル O (を) の制御を加える。

これからの課題 2 並列助詞の「や」を含むものに制約を加える。

2.1 文とは

感動詞文節または述語文節で終る文のうち、以下の 6 つの型を文として扱っている。

1. 「はい」「いいえ」などの感動詞を扱うもの。
2. 「それでは、登録用紙をお送り下さい」のように、接続詞が文頭にあり、述部が最低一つあるもの。
3. 接続詞が文頭にあり、述部が最低一つあるもののうち、動詞が方向性を持たず、かつ目的語を持つものと共起できないもの。
4. 接続詞がなく、述部が最低一つあるもの。
5. 接続詞がなく、述部が最低一つあるもののうち、動詞が方向性を持たず、かつ目的語を持つものと共起できないもの。

例文 1 太郎は勉強をした。

例文 2 太郎は英語を勉強した。

例文 3 * 太郎は英語を勉強をした。

³付録 B.2 参照。

6. 例外を含むもの。

この6つの型は以下のような規則で表される。

(〈SS〉 <--> (〈INTERJ1〉))
(〈SS〉 <--> (〈CONJ1〉 〈NVS〉))
(〈SS〉 <--> (〈CONJ1〉 〈NVS-DIR-OBJ〉))
(〈SS〉 <--> (〈NVS〉))
(〈SS〉 <--> (〈NVS-DIR-OBJ〉))
(〈SS〉 <--> (〈EXCEP〉))

これからの課題3文は、感動詞文節または述語文節で終ることを前提としている。従って、以下のように名詞を含む文節で終るものは考慮していない。

例文4 玄関を掃除してくれたのは誰？

例文5 私。

話し言葉を対象とする場合、名詞文節で終る文の扱いも考慮するべきであろう。

6つの文型〈SS〉を構成する、非終端カテゴリーについて説明する。

INTERJ1では、感動詞を扱っている。感動詞は、一文節で一文を構成するとしている。現在、「はい」、「いいえ」、「もしもし」、「さようなら」、「ありがとう」を扱っている。

- これからの課題4話し言葉では、感動詞が文中で使われることがある。特に「はい」は、文中のいろいろな箇所が使われる：

例文6 こちらは、はい、事務局ですが。

感動詞が文中のどの位置で使われるか、その傾向を調べる必要がある。

CONJ1では、接続詞を扱っている。例文7の、「それでは」のように、文頭で使われるものを扱っている。

例文7 それでは、登録用紙をお送り下さい。

これからの課題5話し言葉では、例文7の文節の出現順序を変えて、例文8、9のように言うこともできる：

例文8 登録用紙を、それでは、お送り下さい。

例文9 登録用紙をお送り下さい、それでは。

これらの3文からも分かるように、接続詞は、文のいろいろな位置に出現可能である。感動詞と同様、文中のどの位置で使われるか、その傾向を調べる必要がある。

NVS, NVS-DIR-OBJ では、共に述語を含むカテゴリーを扱っている。

音声認識予備実験では、助詞「へ」が文節候補になる割合が高かった。続いて、助詞「を」を含む文節も候補になる割合が高かった⁴。ここでは、方向性をあらわす「へ」と、目的格をあらわす「を」の両者と相容れないものを NVS-DIR-OBJ として、助詞の「へ」や「を」を伴える NVS と区別している。

これからの課題 6 現在、このグループで扱っているのは *copula* だけであるが、他の述語の検討が必要であろう。また、方向性を表すものとは、相容れないが、目的格を表す「を」とは共起するもの (*ex.* NVS-DIR) 等、いくつかのグループを作るべきであろう。そのためには、述語の細分類が不可欠である。

EXCEP では、対話データベース形態素解析の文節切りの規則と、音声入力単位の整合がとれていないものを、例外として扱っている。モデル会話 AB, 1-5 の範囲では、例文 10 と 11 の 2 文で文節切りが不整合である⁵：

例文 10 どうすればよろしいですか？ (会話 B)

例文 11 それは、別に問題ありません。 (会話 5)

例文 10 の「どうすれば」は、「どう」と「すれば」に文節分けされるが、音声入力では、「どうすれば」と、ひと続きである。また、例文 11 の「問題ありません」も「問題」と「ありません」の 2 文節からなっているが、音声入力では、1 文節として、発声されている。

2.2 述語句

文末で使われる述語句を扱う。述語句は、方向性があり、「を」格をとれる述語を含むもの (NVS) と方向性がなく、「を」格をとらない述語を含むもの (NVS-DIR-OBJ) とに分けている。これらは、さらに名詞句を含むものと含まないものとに分けている。

2.2.1 方向性のある述語句

NVS には、名詞句を含まない述語句 (VS) と、名詞句を含む述語句 (NnVS) とがある。

⁴12 ページ参照。

⁵3.13 節、3.13 ページ参照。

(<NVS> <--> (<VS>))
(<NVS> <--> (<NnVS>))

名詞句がないもの

VS は、例文 12 のように、自動詞を含むもの、または、他動詞でも、名詞句を持たない場合を扱っている (VC)。

例文 12 わかりました。

この VS は、例文 13 の「すぐ」のような副詞、または、例文 14 の「もう少し勉強すれば、」のように、接続助詞などを使い副詞的に働くもの (ADVPH) を伴える。

例文 13 すぐ、わかりました。

例文 14 もう少し勉強すれば、わかるでしょう。

(<VS> <--> (<VC>))
(<VS> <--> (<ADVPH> <VS>))

名詞句があるもの

NnVS⁶では、例文 15、16 のように、「用紙を」「そちらに」などの名詞句 (M-NN) を一つ以上伴うものを扱っている。

例文 15 用紙を送ります。

例文 16 そちらに用紙を送ります。

例文 17、18 など、助詞を含む文節のうち、「読みは」「書きも」のように動詞の連用形に係助詞が後続するもの (VADV-H-P-K) は、「する」(VS-SA) とだけ共起することを許している。

例文 17 読みはしない。

例文 18 書きもしない。

現在、係助詞として、「は、も、でも」を扱っている。「でも」に関しては、出現頻度が低いと思われるが、音声認識に支障をきたしていないので、「は、も」と同様の扱いをしている。「でも」の扱いは、今後の認識傾向により、検討する。

⁶名詞句の数により非終端カテゴリーの細分類をする場合は、小文字の n を変数として、その位置に名詞句数を記入する予定である。

名詞句を伴わない VS 同様、例文 19 の「すぐに」のように、副詞相当句を伴うことができる。また、VS の規則と組み合わせると、例文 20 のように、文節の出現位置の違うものも扱うことができる。

例文 19 すぐに、用紙を送ります。

例文 20 用紙をすぐに送ります。

(<NnVS> <--> (<M-NN> <VS>))
(<NnVS> <--> (<VADV-H-P-K> <VS-SA>))
(<NnVS> <--> (<ADVPH> <NnVS>))
(<NnVS> <--> (<M-NN> <NnVS>))

2.2.2 方向性のない述語句

NVS-DIR-OBJ では、方向性を持たず、かつ「を」を目的語に持たないものを扱っている。具体的には、助詞「へ」や「を」を含む文節と述語文節との共起を許していない。NVS 同様、名詞句を含まないもの (VS-DIR-OBJ) と含むもの (NnVS-DIR-OBJ) とを区別している。

(<NVS-DIR-OBJ> <--> (<VS-DIR-OBJ>))
(<NVS-DIR-OBJ> <--> (<NnVS-DIR-OBJ>))

名詞句がないもの

VS-DIR-OBJ では、方向性を持たず、名詞句を持たない述語句を扱っている。例文 21 のように、「です」または、「だ」を含む文節を扱う (DA-FORM)。

このうち、例文 22 のように、住所を含むもの (ADRE-VC) と、例文 23 のように、電話番号を含むもの (TEL-VC) には、独自の文節カテゴリーを与え、別に扱っている。これは、数字、特に電話番号の読み方は、音声認識の際、誤って認識されやすいためである。

例文 21 事務局ですか。

例文 22 大阪市北区茶屋町 2 の 3 です。

例文 23 07749 (ゼロナナナナヨンキュウ) の 5 (ゴ) の 1311 (イチサンイチイチ) です。

また、例文 24 のように、「来月お申し込みになれば」のような、副詞相当句を伴える。

例文 24 来月お申し込みになれば、4 万円です。

(<VS-DIR-OBJ> <--> (<DA-FORM>))
(<VS-DIR-OBJ> <--> (<ADRE-VC>))
(<VS-DIR-OBJ> <--> (<TEL-VC>))
(<VS-DIR-OBJ> <--> (<ADVPH> <VS-DIR-OBJ>))

これからの課題 7 現在ここで扱っているのは、Copula を含むものだけだが、今後、動詞の細分化を行ない、あるタイプの動詞を含むものは VS-DIR-OBJ のように扱うべきであろう。

だ / です DA-FORM では、例文 25 のような Copula 文を扱っている。また、例文 26 の「会議の」のような名詞部分の修飾 (MOD-N) も DA-FORM で扱っている。

例文 25 事務局です。

例文 26 会議の事務局です。

(<DA-FORM> <--> (<NVC>))
(<DA-FORM> <--> (<MOD-N> <NVC>))

住所 ADRE-VC では、住所に関する「だ」「です」を伴う句を扱っている。

モデル会話 AB, 1-5 には住所に関する文が以下の 3 例ある:

例文 27 住所は大阪市北区茶屋町 23 です。(会話 A,1)

例文 28 住所は大阪市東区徳井町 1 の 2 です。(会話 B)

例文 29 住所は大阪市東区玉造 2 丁目 27 の 7 です。(会話 4)

モデル会話で使われているのは、いずれも、Copula 文であるので、バージョン 1.1 では、Copula を伴うものだけを扱う。

住所に関する文の最後の文節 (AVC) は、「2 です」のように、本数詞に「です」を伴う、または、「5 号です」のように本数詞に助数詞を伴い、さらに「です」を伴う、としている。最後から 2 番目までの文節は、まず、あるまとまりとしてまとめあげ (ADRE1)⁷、それを、文節組み立て規則で基礎とするカテゴリーとして、住所を扱う最後の文節と共に使われるように定義している。

(<ADRE-VC> <--> (<ADRE1> <AVC>))

⁷2.6.1節、30 ページ参照。

これからの課題 8 例文 30 のように、Copula を伴わず、名前が続くもの、また、例文 31、32 のように助詞が後続するものも考慮すべきであろう。

例文 30 送り先は大阪市東区城見 2 の 1 の 61 渡辺明です。(会話 7)

例文 31 住所は東京都豊島区東池袋 3 丁目 2 番 5 号でよろしいですね。(会話 8)

例文 32 大阪市北区茶屋町 23 が私の住所です。

電話番号 TEL-VC では、電話番号に関する「だ」「です」を伴うものを扱っている。

モデル会話 AB, 1-5 には電話に関する文が以下の 2 例ある:

例文 33 372 (サンナナニ) の 8018 (ハチゼロイチハチ) です。(会話 4)

例文 34 372 (サンナナニ) の 8018 (ハチゼロイチハチ) でございますね。(会話 4)

さらに、会話 10 では、以下の文が使われている:

例文 35 電話番号は 331 (サンサンイチ) の 2521 (ニゴニイチ) です。(会話 10)

モデル会話で扱っている文を考慮して、電話番号に関しては、三桁の局番に四桁の番号を想定している。四桁の部分 (TELVC) は、Copula を伴い、三桁の部分 (TEL1) は助詞の「の」を伴う、としている。

(〈TEL-VC〉 <--> (〈TEL1〉 〈TELVC〉))

名詞句があるもの

NnVS-DIR-OBJ⁸ では、例文 36 のように、方向性を持たないもののうち目的語以外の名詞句を伴うものを扱っている。共起する名詞句は、助詞「へ」や「を」で終わらない M-MN-DIR-OBJ である。

例文 36 そちらは、事務局ですか。

「もの」「こと」「かた」などのいわゆる形式名詞は、連体修飾を受けずに使われることはめったにない。そこで、形式名詞を含み方向性や「を」格を伴わないもの (VS-KEI-DIR-OBJ) は、例文 37 の「会議に申し込んだ」(MOD-N) のように修飾を受けて、初めて文として認識している。

⁸名詞句の数により非終端カテゴリーの細分類をする場合は、小文字の n を変数として、その位置に名詞句数を記入する予定である。

例文 37 会議に申し込んだものですが。

例文 38 の「昨日」のような副詞相当句も伴える。

例文 38 昨日会議に申し込んだものですが。

(<NnVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <VS-DIR-OBJ>))
(<NnVS-DIR-OBJ> <--> (<MOD-N> <VS-KEI-DIR-OBJ>))
(<NnVS-DIR-OBJ> <--> (<ADVPH> <NnVS-DIR-OBJ>))
(<NnVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <NnVS-DIR-OBJ>))

これからの課題 9 連体修飾句では、名詞句が使われることが多いと思われるので、名詞句を持たない VS-DIR-OBJ では、形式名詞を扱っていない。しかし、これは直感に頼った判断なので、対話データベースを基に調査する必要がある。また、修飾を受けた形式名詞句を新たなカテゴリーとして扱うことも考えられる。

2.3 副詞句

副詞的に働くものと、名詞的に働くものの区別は明確でない⁹。将来的には、以下のように、まず、副詞句と名詞句を動詞に対して副詞的に働くものと考え (ADVS)、その次の段階で格要素となり得るものと、方向性を持つものなどに分類することが考えられる。

(<ADVS> <--> (<ADVPH>))
(<ADVS> <--> (<M-NN>))
(<ADVS> <--> (<M-NN-DIR-OBJ>))

しかし、このように中間段階をおくことは今のところ実質的な意味を持たないので、バージョン 1.1 では、直接、名詞相当句と、副詞相当句に分けている。

ADVPH には、「たくさん」「とてもたくさん」のように、副詞だけを扱うもの (ADV-s) と、「今、会議に申し込めば」のように、名詞要素を含むもの (ADV-c, ADV-c-DIR-OBJ)¹⁰とがある。

(<ADVPH> <--> (<ADV-s>))
(<ADVPH> <--> (<ADV-c>))
(<ADVPH> <--> (<ADV-c-DIR-OBJ>))

これからの課題 10 例文 39 では、副詞の「ほぼ」が名詞の「直線」を修飾しているが、ADVPH では、このように名詞を修飾する副詞は、考慮していない。

例文 39 これで、ほぼ直線だ。

⁹ 「ナイフで切る」の「ナイフで」は、副詞句として扱うべきであろうが、ここでは、他の格要素を含むものと同様に扱っている。

¹⁰ 小文字の s と c は、それぞれ simple と complex の略である。

2.3.1 副詞からなる副詞句

ADV-s では、「まず」「全然」のような副詞句 (ADV1)、または「もうちょっと」のように、副詞を重ねたものを扱っている。

(〈ADV-s〉 <--> (〈ADV1〉))
(〈ADV-s〉 <--> (〈ADV1〉 〈ADV-s〉))

これからの課題 11 ここでは、副詞はいくつ続いてもいいということにしているが、副詞に修飾される副詞と、されない副詞、または、副詞を修飾できる副詞とできない副詞という分類を導入し、連続回数も限定するべきであろう。

副詞句

ADV1 では、副詞句を扱っている。一般的な副詞、または副詞句 (adv-p) と連体詞を含むもの (adv-ph) がある。

(〈ADV1〉 <--> (〈adv-p〉))
(〈ADV1〉 <--> (〈adv-ph〉))

これからの課題 12 副詞を修飾できる副詞、「全然 ... ない」などある程度かかり先の決まっている呼応の副詞など、制約として有効な副詞の細分類が必要であろう。

一般的な副詞句 adv-p では、「まだ」「あらかじめ」のように、それだけで副詞として機能するもの、または、「ちょっと」「ちょっとだけ」のように、それだけで、または、さらに助詞を伴っても副詞として機能するものがある。

連体詞を含むもの adv-ph では、連体詞を含むものを扱っている。

モデル会話に以下の文がある:

例文 40 参加料はどのようにお支払いしたらよいのですか？ (会話 2)

例文 40 の「どのように」は、対話データベースの形態素解析¹¹では、以下のように解析される:

データベース

どの ように
連体詞 助動詞 (連用形)

¹¹参考文献 [14]、39 ページ参照。

しかし、連体詞の扱いに関する方針が定まっていないので、バージョン 1.1 では、仮に、「どのように」を一文節、副詞要素として扱っている¹²。

2.3.2 方向性のある述部を含む副詞句

ADV-c では、「今申し込めば」のように、述語を含み、基本的に文末以前に使われる副詞句を扱っている¹³。文末で使われる述語句 NVS 同様、名詞句を含まないもの (VADVS) と最低一つの名詞句を含むものがある (NnVADVS)。VADVS と NnVADVS は、述語句を扱っている NVS を書き換えたカテゴリー VS、NnVS とほぼ同じ構成になっている。

(<ADV-c> <--> (<VADVS>))
(<ADV-c> <--> (<NnVADVS>))

名詞句がないもの

VADVS では、「申し込めば」のように名詞句を含まないもの (VADV) を扱っている。「今申し込めば」の「今」 (ADV-s) のような副詞句と共に使うことも可能である。

(<VADVS> <--> (<VADV>))
(<VADVS> <--> (<ADV-s> <VADVS>))

名詞句があるもの

NnVADVS¹⁴ では、「そちらが申し込めば」の「そちらが」のように、最低一つの名詞句 (M-NN) を含むものを扱う。例文 41 の「読み」のように使われる動詞の連用形、または、例文 42 の「読みに」のように、連用形に助詞がついたもの (VADV-H) は、単独では使われないと仮定している。また、例文 43 の「読みを」のように、動詞が名詞的に働いているもの (VADV-H-0) は、連体修飾を受けて使われる (MOD-N) と仮定している。従って、これらのカテゴリーは、名詞句のない VADVS では扱っていない。

例文 41 本を読み、感想を書いた。

例文 42 本を読みに来た。

例文 43 この読みを教えてください。

¹² 2.5.2 節、28 ページ参照。

¹³ 述語を含むものは、文頭、文中、または、文末で使われるカテゴリーに分けている。2.7.4 節、37 ページ参照。

¹⁴ 名詞句の数により非終端カテゴリーの細分類をする場合は、小文字の n を変数として、その位置に名詞句数を記入する予定である。

「そちらが会議に申し込めば」の「そちらが」「会議に」のように、名詞句が繰り返し使われたり、「今そちらが申し込めば」の「今」のように、副詞句が共に使われることも可能である。

(<NnVADVS> <--> (<M-NN> <VADVS>))
(<NnVADVS> <--> (<M-NN> <VADV-H>))
(<NnVADVS> <--> (<MOD-N> <VADV-H-O>))
(<NnVADVS> <--> (<M-NN> <NnVADVS>))
(<NnVADVS> <--> (<ADV-s> <NnVADVS>))

活用語の連用形 VADV-H では、活用語の連用形を重ねて、並列を表すもの (VADV-H-COORD)、活用語の連用形に助詞を付属したもの (VADV-H-P)、活用語の連用形のあとに対比の助詞「は」 (VADV-H-P-K) さらに「する」の連用形 (VADV-SA) が使われるものを扱う。

(<VADV-H> <--> (<VADV-H-COORD>))
(<VADV-H> <--> (<VADV-H-P>))
(<VADV-H> <--> (<VADV-H-P-K> <VADV-SA>))

活用語に「を」が後続するもの VADV-H-O では、活用語の連用形に助詞の「を」が続くもの、活用語の連体形に準体助詞の「の」が続き、さらに、助詞の「を」が続くものを扱っている (VADV-H-P-O)。

(<VADV-H-O> <--> (<VADV-H-P-O>))

2.3.3 方向性のない述部を含む副詞句

ADV-c-DIR-OBJ では、「会議だったら」、「ATR の会議なので」など、方向性を持たず、目的語も持たないものを扱う。方向性や「を」格と共起しない述語句 NVS-DIR-OBJ と、ほぼ同じ構成になっており、名詞句を伴うもの (VADVS-DIR-OBJ) と、名詞句を伴わないもの (NnVADVS-DIR-OBJ) がある。

(<ADV-c-DIR-OBJ> <--> (<VADVS-DIR-OBJ>))
(<ADV-c-DIR-OBJ> <--> (<NnVADVS-DIR-OBJ>))

これからの課題 13 このバージョンでは、Copula だけを扱っているが、本動詞の一部もここで扱うことが考えられる。

これからの課題 14 形容詞は、本動詞と同様の扱いをしているが、方向性、目的語に関する観点から見ると、連体修飾はされないが、将来 Copula のように扱うべきであろう。

名詞句がないもの

VADVS-DIR-OBJ では、例文 44 の「会議だったら」のように、方向性を持たず、名詞句を伴わないもの (NKVADV) を扱う。

例文 44 会議だったら、ASTI に行こう。

また、例文 45 の「ATR の会議なので」のように、連体修飾 (MOD-N) されるものも扱う。

例文 45 ATR の会議なので行ってみよう。

(<VADVS-DIR-OBJ> <--> (<NKVADV>))
(<VADVS-DIR-OBJ> <--> (<MOD-N> <NKVADV>))

名詞句があるもの

NnVADVS-DIR-OBJ¹⁵ では、方向性を持たないもののうち、例文 46 の「ASTI は」のように、名詞句を伴うものを扱う。

例文 46 ASTI は ATR の会議で、1989 年に行なわれた。

形式名詞を含む Copula 文節 (NKVADV-KEI) は、連体修飾される、と仮定している。連体修飾句では、名詞句が使われることが多いと思われるので、名詞句を持たない VADVS-DIR-OBJ では、形式名詞を扱っていない¹⁶。

(<NnVADVS-DIR-OBJ> <--> (<MOD-N> <NKVADV-KEI>))
(<NnVADVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <VADVS-DIR-OBJ>))
(<NnVADVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <NnVADVS-DIR-OBJ>))
(<NnVADVS-DIR-OBJ> <--> (<ADV-s> <NnVADVS-DIR-OBJ>))

2.4 名詞句

名詞句は、方向性を表し、かつ、助詞「を」をとり目的語になるか、ならないかにより分けている。これらは、それぞれ方向性のある述語を含む NVS と共起するもの (M-NN) と方向性のない述語を含む NVS-DIR-OBJ と共起するもの (M-NN-DIR-OBJ) である。さらに、名詞の種類により、普通名詞や代名詞を扱うものと形式名詞を扱うものに分けている。バージョン 1.1 では、引用句は方向性のある名詞句の中で扱っている。

¹⁵ 名詞句の数により非終端カテゴリーの細分類をする場合は、小文字の n を変数として、その位置に名詞句数を記入する予定である。

¹⁶ 2.2.2 節、20 ページ参照。

2.4.1 方向性のある名詞句

M-NN では、普通名詞、代名詞または住所以外の固有名詞を含む一般的な名詞句 (NN)、形式名詞を含む名詞句 (NN-KEI) と、引用句 (NQN, NQN-DIR-OBJ) を扱っている。NN は、「分からない点が」のように、連体修飾 (MOD-N) されることも可能である。形式名詞を含む名詞句 (NN-KEI) は、「書いたものを」のように、連体修飾されると仮定している。

(<M-NN> <--> (<NN>))
(<M-NN> <--> (<MOD-N> <NN-KEI>))
(<M-NN> <--> (<MOD-N> <M-NN>))
(<M-NN> <--> (<NQN>))
(<M-NN> <--> (<NQN-DIR-OBJ>))

一般的な名詞句

NN では、名詞に助詞が後続する、またはしないものを扱っている。普通名詞、代名詞または「来月」や「今年」などの時を表す名詞を含むものは np で扱っている。そのうち普通名詞には助詞が後続しなければならないが、代名詞や時を表す名詞には、助詞の省略も許している。「一万円」「20 ドル」などは、前もって金額を表すまとまりとして組み立てている (NP-MONEY)。疑問代名詞は文法規則作成の経過理由により、助詞が後続するもの (wh-np) と助詞が後続しないもの (wh-n) に分けている。さらに、方向性を表す助詞「へ」が後続するもの (np-e, NP-E-MONEY, wh-np-e) と「を」を伴う目的語になるもの (np-o, wh-np-o, NP-MONEY-O) は別のカテゴリーとして扱っている。

(<NN> <--> (<np>))
(<NN> <--> (<NP-MONEY>))
(<NN> <--> (<wh-np>))
(<NN> <--> (<wh-n>))
(<NN> <--> (<np-e>))
(<NN> <--> (<NP-E-MONEY>))
(<NN> <--> (<wh-np-e>))
(<NN> <--> (<np-o>))
(<NN> <--> (<wh-np-o>))
(<NN> <--> (<NP-MONEY-O>))

形式名詞を含む名詞句

NN-KEI では、「もの」「こと」などの形式名詞を含む名詞句を扱っている。np-keisiki では、形式名詞に助詞が後続する、または、しないもの

のを扱っている。方向性を表す助詞「へ」が後続するもの (np-keisiki-e) や「を」を伴い目的語になるもの (np-keisiki-o) は別のカテゴリーとして扱っている。

(<NN-KEI> <--> (<np-keisiki>))
(<NN-KEI> <--> (<np-keisiki-e>))
(<NN-KEI> <--> (<np-keisiki-o>))

引用句

例文 47 の「見てみたいと」のような引用の「と」を伴う句は、Copula 以外の活用語の終止形に助詞の「と」が後続するもの (NQN) と、Copula の終止形に助詞の「と」が後続するもの (NQN-DIR-OBJ) に分けている。

例文 47 東西が統合されたドイツを見てみたいと思う。

方向性のある名詞句 NQN では、「分からないと」のように、述語に引用を表す助詞「と」が使われている引用句を扱っている (QN)。構成は、述語句 NVS とほぼ同じである。これは、「ちょっと分からないと」の「ちょっと」のような副詞句 (ADV-s) と共に使われたり、「それが分からないと」の「それが」のような名詞句 (M-NN) と使われることができる。

(<NQN> <--> (<QN>))
(<NQN> <--> (<ADV-s> <NQN>))
(<NQN> <--> (<M-NN> <NQN>))

これからの課題 15 このバージョンでは、述語を含む副詞句 (ADV-c) や、「する」を別に扱う規則 VS-SA は考慮していない。また、方向性を表すものや、目的語をとらない NVS-DIR-OBJ に対応するものは、まだ扱っていない。

方向性のないもの NQN-DIR-OBJ では、引用を表す助詞「と」が使われている文節で、方向性を表すものや目的語を持つものと共起しないものを扱う。NVS-DIR-OBJ 同様、このバージョンでは、「会議なのですと」のような Copula を扱っている。これは、さらに形式名詞が使われていないもの (NKQN) と使われているもの (NKQN-KEI) に分けてある。形式名詞が使われていないものは、「ATR の会議なのですと」の「ATR の」のように連体修飾 (MOD-N) をすることもできる。一方、形式名詞を含む引用句 (NKQN-KEI) は、「書いたものなのですと」の「書いた」のように、連体修飾されると仮定している。

(<NQN-DIR-OBJ> <--> (<NKQN>))
 (<NQN-DIR-OBJ> <--> (<MOD-N> <NKQN>))
 (<NQN-DIR-OBJ> <--> (<MOD-N> <NKQN-KEI>))
 (<NQN-DIR-OBJ> <--> (<ADV-s> <NQN-DIR-OBJ>))
 (<NQN-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <NQN-DIR-OBJ>))

2.4.2 方向性のない名詞句

M-NN-DIR-OBJ では、「へ」や「を」を含まない名詞句を扱っている。これは、形式名詞以外の名詞を含むもの (NN-DIR-OBJ) と、形式名詞を含むもの (NN-KEI-DIR-OBJ) に分類している。形式名詞は、連体修飾 (MOD-N) を受ける。

(<M-NN-DIR-OBJ> <--> (<NN-DIR-OBJ>))
 (<M-NN-DIR-OBJ> <--> (<MOD-N> <NN-KEI-DIR-OBJ>))
 (<M-NN-DIR-OBJ> <--> (<MOD-N> <M-NN-DIR-OBJ>))

一般的な名詞句

NN-DIR-OBJ では、一般的な名詞句 NN で扱っているもののうち、名詞に助詞の「へ」や「を」が後続しないものを扱っている。

(<NN-DIR-OBJ> <--> (<np>))
 (<NN-DIR-OBJ> <--> (<NP-MONEY>))
 (<NN-DIR-OBJ> <--> (<wh-np>))
 (<NN-DIR-OBJ> <--> (<wh-n>))

形式名詞を含む句

NN-KEI-DIR-OBJ では、形式名詞を含む名詞句 NN-KEI で扱っているもののうち、名詞に助詞の「へ」や「を」が後続しないものを扱っている。

(<NN-KEI-DIR-OBJ> <--> (<np-keisiki>))

2.5 連体修飾句

MOD-N では、連体修飾句を扱っている。形式名詞以外の名詞または名詞と同じような振舞いをするものに、助詞の「の」が後続するもの (NM)、形式名詞に助詞の「の」が後続するもの (NM-KEI)、活用語の連体形 (NVM-c)、連体詞を含むもの (NM) を扱っている。形式名詞 (NM-KEI) は、連体修飾される (MOD-N) と仮定している。さらに、複合格助詞を含むもの (NPM) も扱っている。

(<MOD-N> <--> (<NM>))
(<MOD-N> <--> (<MOD-N> <NM-KEI>))
(<MOD-N> <--> (<NVM-c>))
(<MOD-N> <--> (<NPM>))

2.5.1 助詞「の」を伴うもの

NM では、名詞、代名詞に助詞の「の」が後続するものを扱っている。また、NM-KEI では、形式名詞に助詞の「の」が後続するものを扱っている。形式名詞を別に扱っているのは、形式名詞は修飾を受けることが多いためである。

(<NM> <--> (<np-no>))
(<NM> <--> (<wh-np-no>))
(<NM> <--> (<adv-p-no>))
(<NM> <--> (<vn-no>))
(<NM-KEI> <--> (<np-keisiki-no>))

2.5.2 連体詞

NM では、連体詞を含む連体修飾句も扱っている。

モデル会話では、文節「どのような」を含む例文 48 と例文 49 が使われている:

例文 48 どのようなご要件でしょうか? (会話 1)

例文 49 どのような手続きをすればよろしいのでしょうか。(会話 1)

対話データベースの形態素解析では、「どのような」は以下のように解析される:

データベース

どの ような
連体詞 助動詞(連体形)

ここでは、連体詞の扱いに関する方針が定まっていないので、仮に、「どのような」を一文節、連体詞として、名詞に助詞の「の」が後続するものと同じカテゴリーで扱っている¹⁷。

(<NM> <--> (<n-rentai>))

¹⁷2.3.1節、21 ページ参照。

2.5.3 活用語の連体形

NVM-c では、活用語の連体形を扱っている。述語句 NVS の名詞句を伴わない VS や名詞句を伴う NnVS と同様に、名詞句を伴わないものと (VMS) 名詞句を伴うもの (NnVMS) とに分けている。NVM-c では、述語句や副詞句でカテゴリー分けの基準に使っていた方向性または目的語などは、考慮していない。

(<NVM-c> <--> (<VMS>))
(<NVM-c> <--> (<NnVMS>))

名詞句がないもの

VMS では、「分からない」のような、活用語の連体形 (VM) だけからなるものと、「ちょっと分からない」のように、副詞句 (ADV-s) を伴うものを扱っている。

(<VMS> <--> (<VM>))
(<VMS> <--> (<ADV-s> <VM>))

名詞句があるもの

NnVMS¹⁸ では、「そちらに送った」の「そちらに」のように、名詞句 (M-NN) を伴うもの、「読みはした」のように、活用語の連用形に係助詞の「は」や「も」が後続し (VADV-H-P-K)、さらに「する」の連体形 (VM-SA) が後続するもの、また、「すでにそちらに送った」の「すでに」のように、副詞句 (ADV-s) を伴うものを扱っている。「太郎がそちらに送った」の「太郎が」「そちらに」のように名詞句が二つ以上使われることも可能である。

(<NnVMS> <--> (<M-NN> <VMS>))
(<NnVMS> <--> (<VADV-H-P-K> <VM-SA>))
(<NnVMS> <--> (<M-NN> <NnVMS>))
(<NnVMS> <--> (<ADV-s> <NnVMS>))

2.5.4 複合格助詞

NPM では、格助詞に活用語の連体形が後続し、複合格助詞のように振舞うものを扱っている。「取消に対する」の「に対する」は、対話データベースの形態素解析では、以下のように二文節に解析されるが、ここでは、

¹⁸ 名詞句の数により非終端カテゴリーの細分類をする場合は、小文字の n を変数として、その位置に名詞句数を記入する予定である。

便宜上複合格助詞として扱っている¹⁹。対話データベースの形態素解析との不整合を明示するため、「に対する」が後続するもの (NPM) は、活用語の連体形 (NVM-c) とは別に扱っている。

データベース	
に	対する
格助詞	本動詞

(<NPM> <--> (<np-special>))

2.6 前処理を行なうもの

2.6.1 住所

数字は音声認識の誤認識の対象になりやすい。住所では、番地を表すのに数字が使われるため、住所の言い回しは特に連鎖制約として制限している。住所に関する最後の文節は、「だ」「です」を伴うと仮定している²⁰。住所に関する最後から 2 番目までの文節は、ADRE1 として一つにまとめて、それが「だ」「です」を伴う文節と共に用いられるとしている。ADRE1 で規定している文節の出現順序とそれぞれの規則により認識される住所の例を以下に示す:

(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14>))

大阪市北区茶屋町

(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14> <ADRE15>))

大阪市北区茶屋町一の

(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14> <ADRE16> <ADRE15>))

大阪市北区茶屋町二丁目一の

(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14> <ADRE15> <ADRE15>))

大阪市北区茶屋町二の一の

(<ADRE1> <--> (<ADRE11> <ADRE12> <ADRE14> <ADRE16> <ADRE17>))

東京都豊島区東池袋三丁目二番

2.6.2 金額

金額は、文節発声音声入力では、それぞれの位ごとに別の文節として分けて発声している。モデル会話 AB, 1-5 で使われている例文 50 から 54 のうち、例文 50、51 の「三万五千元」と例文 54 の「八万五千元」は、それぞれ「三万 / 五千元」「八万 / 五千元」と文節ごとに発声している。

¹⁹3.13 章、83 ページ参照。

²⁰2.2.2 節、18 ページ参照。

例文 50 登録費としてお一人三万五千円が必要です。(会話 B)

例文 51 参加料は現在お一人三万五千円です。(会話 2)

例文 52 来月お申し込みになりますと四万円です。(会話 2)

例文 53 参加料は四万円です。(会話 4)

例文 54 すでに登録料の八万五千円を振り込まれておられますね。
(会話 5)

金額に関するものは、それぞれ別の独立した文節として扱わず、住所を一つのまとまりとしてとらえたのと同様、金額に関する一つのまとまりとして扱う。

例文 50 から 54 の例のようにモデル会話 AB, 1-5 で使われているのは、万、千の単位であるが、バージョン 1.1 では、それ以降のモデル会話で使われている百の単位も考慮している。

会話 9 と会話 10 には、金額に関する以下の文があった:

例文 55 京都駅からですとおよそ六千円かかります。(会話 9)

例文 56 北大路駅からですとおよそ九百円です。(会話 9)

例文 57 一人部屋の値段は一晩七千円から一万円です。(会話 10)

例文 58 二人部屋の値段は九千五百円から六万円です。(会話 10)

例文 59 では、京都プリンスホテルの七千円の一人部屋をお願いします。(会話 10)

例文 60 京都プリンスホテルの七千円の一人部屋ですね？(会話 10)

金額に関するものは、名詞文節として扱うものと、述語文節として扱うものに分けている。ここでは、名詞句 NP-MONEY を例に、現行の組み立て方式を示す。小文字で書かれているカテゴリーは、文節のタイプである。

NP-M-MAN, NP-M-SEN, NP-M-HYAKU では、例文 50 の「三万五千円が必要です」の「五千円が」や例文 54 の「八万五千円を振り込まれて」の「五千円を」など、金額に関するものを含む文節列の最後の文節を扱っている。

SEN, HYAKU では、例文 50 の「三万五千円が」や例文 54 の「八万五千円を」の「三万」や「八万」および例文 58 の「九千五百円から」の「九千」や「二万九千五百円から」の「二万九千」など、金額に関するものを含む文節列の最後から二番目までの文節を扱っている。このバージョンでは、「千」単位の前には「万」単位 (M-MAN) が、「百」単位の前には「万」、「万」と「千」 (M-SEN) の並び、または「千」単位が出現できるとしている。

M-MAN, M-SEN, M-HYAKU では、それぞれ「万」「千」「百」の位の数字を扱っている²¹。

np-money-man, np-money-sen, np-money-hyaku では、「一万円」「三千ドル」「五百マルク」など、それぞれの位を表す数字に貨幣単位の「円」、「ドル」または「マルク」などが後続したものを扱っている。また、さらに、助詞が後続するものも扱っている。

money-man, money-sen, money-hyaku では、それぞれ「二万」「二千」「二百」など、それぞれの位の数字を扱っている。これらには、貨幣単位や、助詞などは後続しない。

²¹ 百の位の数字 M-HYAKU は例文 61 のように十以下の位を扱うために用意している。しかし、金額組み立て規則の中には、まだ取り入れていない。

例文 61 五千三百二十円。

(<NN> <--> (<NP-MONEY>))

(<NP-MONEY> <--> (<NP-M-MAN>))

(<NP-MONEY> <--> (<NP-M-SEN>))

(<NP-MONEY> <--> (<SEN> <NP-M-SEN>))

(<NP-MONEY> <--> (<NP-M-HYAKU>))

(<NP-MONEY> <--> (<HYAKU> <NP-M-HYAKU>))

(<SEN> <--> (<M-MAN>))

(<HYAKU> <--> (<M-MAN>))

(<HYAKU> <--> (<M-MAN> <M-SEN>))

(<HYAKU> <--> (<M-SEN>))

(<NP-M-MAN> <--> (<np-money-man>))

(<NP-M-SEN> <--> (<np-money-sen>))

(<NP-M-HYAKU> <--> (<np-money-hyaku>))

(<M-MAN> <--> (<money-man>))

(<M-SEN> <--> (<money-sen>))

(<M-HYAKU> <--> (<money-hyaku>))

2.7 文節のタイプ

第2章で紹介している、文節に基づき文を組み立てる規則、バージョン1.1は、第3章で紹介する、文節を構成する規則を基に構築した。2.7節では、この二つの規則のインターフェイスとなるバージョン1.1の終端カテゴリを紹介する。名詞に関するものなど、一部はすでに紹介しているが、2.7節で、改めて終端カテゴリを列挙する。また、以下のように、終端カテゴリは、必ずしも一種類のカテゴリが展開されたものではない:

(<NN> <--> (<np>))

(<NN-DIR-OBJ> <--> (<np>))

ここでは、場合に応じて代表的なカテゴリだけを示す。大文字で書いてあるカテゴリ名は非終端カテゴリであり、小文字で書いてあるものは終端カテゴリである。

金額に関するものは、すでにある程度文節が組み上げられているので、大文字のカテゴリ名が与えられている。また、2.6.2節で概要を説明しているので、ここでは、非終端カテゴリとして列挙するにとどめ、詳細は省く。

2.7.1 感動詞文節

INTERJ1 では、感動詞、または、一文節で一文をなすものを扱っている。

interj 「はい」や「もしもし」などの感動詞や「ありがとう」のような決まった言い方。

(<INTERJ1> <--> (<interj>))

2.7.2 接続詞文節

CONJ1 では、接続詞を扱っている。

conj 「また」や「ところで」など。

(<CONJ1> <--> (<conj>))

2.7.3 述語文節

文末で使われる述語文節には、方向性や「を」格を伴えるもの (VC, VS-SA) と、伴えないものがある (NVC, VS-KEI-DIR-OBJ)。前者では、さらに、「する」を区別している。これは、「読みはする」のように、動詞の連用形に対照の「は」が後続するものを扱うためである。後者では、形式名詞を伴わないもの (NVC) と、伴うもの (VS-KEI-DIR-OBJ) とを区別している。

VC, VS-SA では、方向性や「を」格を伴う名詞句と共起する活用語を扱っている。

vaux 例文 62 の「慣れるでしょう」のような、述語句。

例文 62 そのうち静かな田舎の生活にも慣れるでしょう。

vaux-cop 例文 63 の「行きたいのです」のように、活用語に準体助詞の「の」が後接し、さらに「です」が続くもの。

例文 63 南極に行きたいのです。

vaux-s+f 例文 64 の「申し込みたいのですが」のように、文末でも使われる可能性の高い接続助詞で終る文節²²。

例文 64 会議に申し込みたいのですが。

²²例文 82、38 ページ参照。

vaux-sfp 例文 65 の「行きますか」のように、活用語に終助詞が続くもの。

例文 65 北極へ行きますか。

vaux-naru 例文 66 のように、「になる」を伴い、丁寧さを表すもの。

例文 66 お申し込みになります。

vaux-naru-s+f 例文 67 の「お帰りになりますか」のように、「になる」に、さらに文末でも使われる可能性の高い接続助詞が後接するもの²³。

例文 67 先生は夕方にはお帰りになりますか。

vaux-sa 例文 68 の「している」のように、「する」で始まる文節。VC と VS-SA の両方で使えるようにしているが、「する」を文頭で使うことはあまりないと思われるので、VS-SA でだけ扱うべきかも知れない。

例文 68 勉強をしている。

vaux-sa-s+f 例文 69 の「しますが」のように、「する」を含み文末でも使われる可能性の高い接続助詞が続くもの。

例文 69 勉強はしますが。

- (<VC> <--> (<vaux>))
- (<VC> <--> (<vaux-cop>))
- (<VC> <--> (<vaux-s+f>))
- (<VC> <--> (<vaux-sfp>))
- (<VC> <--> (<vaux-naru>))
- (<VC> <--> (<vaux-naru-s+f>))
- (<VC> <--> (<vaux-sa>))
- (<VC> <--> (<vaux-sa-s+f>))
- (<VS-SA> <--> (<vaux-sa>))
- (<VS-SA> <--> (<vaux-sa-s+f>))

「する」や「になる」に関しては、終助詞が後接するものは、このバージョンでは考慮していない。

NVC は、助詞の「へ」や「を」を伴う名詞句と共起しない述語文節である。

n-vaux 例文 70 の「事務局でございます」のように、名詞に Copula の連用形が続き、さらに丁寧体が続くもの。

²³例文 87、39 ページ参照。

例文 70 こちらは、事務局でございます。

n-vaux-cop 例文 71 の「事務局です」のように、名詞に直接 Copula が続き、文が終るもの。

例文 71 こちらは、事務局です。

n-vaux-s+f 例文 72 の「事務局ですが」のように、Copula に、さらに文末でも使われる可能性の高い接続助詞が後接するもの²⁴。

例文 72 こちらは、事務局ですが。

n-vaux-sfp 例文 73 の「事務局ですか」のように、名詞に「だ」「です」が続き、さらに終助詞が後続するもの。

例文 73 そちらは、事務局ですか。

(<NVC> <--> (<n-vaux>))
(<NVC> <--> (<N-VAUX-MONEY>))
(<NVC> <--> (<n-vaux-cop>))
(<NVC> <--> (<N-VAUX-COP-MONEY>))
(<NVC> <--> (<n-vaux-s+f>))
(<NVC> <--> (<N-VAUX-S+F-MONEY>))
(<NVC> <--> (<n-vaux-sfp>))
(<NVC> <--> (<N-VAUX-SFP-MONEY>))

VS-KEI-DIR-OBJ は、形式名詞に「だ」「です」が続く文節である。文節カテゴリーわけは、助詞の「へ」や「を」を伴う名詞句と共起しない述語文節 NVC と同様の構成になっている。

n-keisiki-vaux 例文 74 の「ものでございます」のように、形式名詞に「だ」の連用形が続き、さらにその丁寧体が続くもの。

例文 74 会議に申し込んだものでございます。

n-keisiki-vaux-cop 例文 75 の「ものです」のように、形式名詞に「です」が続き、文が終るもの。

例文 75 会議に申し込んだものです。

n-keisiki-vaux-s+f 例文 76 の「のですが」のように、形式名詞に、「です」が続き、さらに文末でも使われる可能性の高い接続助詞が後続するもの²⁵。

²⁴例文 104、41 ページ参照。

²⁵例文 109、43 ページ参照。

例文 76 会議に申し込んだものですが。

n-keisiki-vaux-sfp 例文 77 の「かたですか」のように、形式名詞に「です」が続き、さらに終助詞が後続するもの。

例文 77 会議に申し込まれたかたですか。

(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux>))
(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-cop>))
(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-s+f>))
(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-sfp>))

2.7.4 副詞文節

副詞文節には、副詞からなるもの (ADV1) と、述語を含み副詞的に働くものがある。述語を含む副詞文節は、接続助詞で終るもの以外は、基本的に文末では使われない。述語を含み副詞的に働く文節は、「だ」「です」で終らないもの (VADV, VADV-SA, VADV-H-COORD, VADV-H-P, VADV-H-P-K, VADV-H-P-O) と、終るもの (NKVADV, NKVADV-KEI) とを区別している。

「だ」「です」で終らないものは、さらに、文頭でも使われるもの (VADV) と、文頭で使われる可能性の少ないもの (VADV-SA, VADV-H-COORD, VADV-H-P, VADV-H-P-K, VADV-H-P-O) に分けている。

ADV1 では、副詞句を扱っている。

adv-p 「まだ」「あらかじめ」のような一般的な副詞、または「もうちょっと」のように副詞に修飾される副詞句。

adv-ph 「どのように」のように連体詞を含み、副詞句のように振舞うもの²⁶。

(<ADV1> <--> (<adv-p>))
(<ADV1> <--> (<adv-ph>))

VADV では、方向性を持ち、「を」格をとれる述語を含む副詞文節を扱っている。VADV は、文頭でも使われる。

vaux-katei 例文 78 の「必要なら」のように、形容名詞に助動詞の「だ」の仮定形が後続する文節。または、例文 79 の「良さそうなら」のように、形容詞に様態を表す助動詞「そうだ」の仮定形が後接する文節など。

²⁶2.3.1 節、21 ページ参照。

例文 78 必要なら、教授に手紙を書こう。

例文 79 もし良さそうなら、あの掃除機に買い替えよう。

vaux-s 例文 80 の「行くので」や例文 81 の「読んでいるなら」のように動詞や助動詞などの活用語に接続助詞が後続する文節。

例文 80 海外に行くので、パスポートをとった。

例文 81 この本を読んでいるなら、もう一流の言語学者だ。

vaux-s+f 例文 82 の「買ったが」のように、活用語に接続助詞が後続する文節。接続助詞は、文末でも使われやすいものに限る²⁷。

例文 82 新しいテニスのラケットは買ったが、まだ試していない。

vaux-te 例文 83 の「あるいては」のように、活用語に「て」、さらに係助詞の「は」、「も」などが後続するもの。

例文 83 一歩あるいては立ち止まった。

vaux-h 例文 84 の「書くなり」、「発表するなり」のように、活用語に「なり」のような並列の助詞が後続し、並列を表すもの。

例文 84 論文を書くなり、学会で発表するなり、研究成果をまとめなければいけない。

vaux-ni 例文 85 の「書くには」のように、活用語の連体形に助詞が後続するもの。

例文 85 論文を書くにはまだ早過ぎる。

vaux-naru-s 例文 86 の「お申し込みになりますと」のように、接頭語の「お」や「御」を伴うサ変名詞に、「になる」が後続し、さらに接続助詞が後続するもの。

例文 86 今お申し込みになりますと、割引があります。

vaux-naru-s+f 例文 87 の「御覧になりましたが」のように、接頭語の「お」や「御」を伴うサ変名詞に、「になる」が後続し、さらに接続助詞が後続するもの。接続助詞は文末で使われる可能性の高いものに限る²⁸。

²⁷ 例文 64、34 ページ参照。

²⁸ 例文 67、35 ページ参照。

例文 87 天皇陛下はこの自動翻訳システムを御覧になりましたが、何もおっしゃいませんでした。

vaux-kdo-coord 例文 88 の「確かに」のように、形容名詞に「だ」の連用形の「に」が後続し、副詞的に働くもの²⁹で、文頭に來られるもの。

例文 88 確かに参加料を受け取りました。

(<VADV> <--> (<vaux-katei>))
(<VADV> <--> (<vaux-s>))
(<VADV> <--> (<vaux-s+f>))
(<VADV> <--> (<vaux-te>))
(<VADV> <--> (<vaux-h>))
(<VADV> <--> (<vaux-ni>))
(<VADV> <--> (<vaux-naru-s>))
(<VADV> <--> (<vaux-naru-s+f>))
(<VADV> <--> (<vaux-kdo-coord>))

VADV-SA, VADV-H-COORD は、共に文頭では使われない。VADV-SAでは、動詞「する」で始まる文節を扱っている。また、VADV-H-COORDでは、活用語の連用形を扱っている。

vaux-sa-s 例文 89 の「しながら」のように、「する」に接続助詞が後続するもの。文末では使われる可能性の低い接続助詞を扱っている。

例文 89 勉強をしながら、テレビを見ている。

vaux-sa-s+f 例文 90 の「しますが」のように、「する」を含み文末でも使われる可能性の高い接続助詞が続くもの³⁰。

例文 90 勉強はしますが、少しも成績が上がらない。

vaux-sa-te 例文 91 の「しては」のように、「する」に「て」、さらに係助詞が後続するもの。

例文 91 授業に出席をしてはいますが、先生の言うことは何も聞いていません。

vaux-sa-h 例文 92 の「しろとか」のように、「する」に並列の助詞が後続するもの。

例文 92 早くレポートの提出をしろとかうるさい。

²⁹これは並列的に働くわけではないので、例えば、vaux-kdo-adv などに、文節名を変えるべきであろう。

³⁰例文 69、35 ページ参照。

voux-sa-coord 例文 93 の「し」のように、「する」の連用形で、並列を表すもの。

例文 93 掃除をし、洗濯をして週末が終った。

voux-coord 例文 94 の「洗い」のように、活用語の連用形で、並列を表すもの。

例文 94 顔を洗い、歯を磨いて寝た。

(<VADV-SA> <--> (<voux-sa-s>))
(<VADV-SA> <--> (<voux-sa-s+f>))
(<VADV-SA> <--> (<voux-sa-te>))
(<VADV-SA> <--> (<voux-sa-h>))
(<VADV-SA> <--> (<voux-sa-coord>))
(<VADV-H-COORD> <--> (<voux-coord>))

VADV-H-P, VADV-H-P-K, VADV-H-P-O は、文中で使われる述語文節である。文頭では使われない。

n-ren-p 例文 95 の「読みが」のように、動詞の連用形に格助詞の「が」が後続し、動詞が名詞化したもの。

例文 95 この漢字の読みが分からない。

v-no-p 例文 96 の「作るのが」のように準体助詞の「の」が後続し、動詞が名詞化しているものに、助詞が後続するもの。

例文 96 食事を作るのが面倒だ。

n-ren-p-k 例文 97 の「読みは」のように、動詞の連用形に、対照を表す係助詞の「は」が後続するもの。この後には、「する」が続くことが多い。

例文 97 ドイツ語の教科書を読みはするが、少しも頭に入らない。

n-ren-p-o 例文 98 の「読みを」のように、動詞の連用形に格助詞の「を」が後続し、動詞が名詞化したもの。格助詞の「を」が後続するものは、「を」を持つ目的語と共起しない述語を別扱いしているので、独自の文節カテゴリーをつくっている。

例文 98 この漢字の読みを教えてください。

v-no-p-o 例文 99 の「作るのを」のように準体助詞の「の」が後続し、動詞が名詞化しているものに、助詞が後続するもの。

例文 99 食事を作るのを忘れた。

(<VADV-H-P> <--> (<n-ren-p>))
(<VADV-H-P> <--> (<v-no-p>))
(<VADV-H-P-K> <--> (<n-ren-p-k>))
(<VADV-H-P-O> <--> (<n-ren-p-o>))
(<VADV-H-P-O> <--> (<v-no-p-o>))

NKVADV では、方向性や「を」格を伴わない述語を含み、文中で副詞的に使われる文節を扱っている。

n-vaux-coord 例文 100 の「研究所で」のように、「だ」の連用形が文節末に来るもの。

例文 100 ATR は研究所で、人里離れた山の中にある。

n-vaux-h 例文 101 の「ナシだとか」や「ブドウだとか」のように、並立助詞で終る文節。

例文 101 ナシだとか、ブドウだとか、秋は果物に困らない。

n-vaux-katei 例文 102 の「ドイツワインなら」のように、「だ」の仮定形で終る文節。

例文 102 彼がドイツワインなら、私は日本酒だ。

n-vaux-s 例文 103 の「参加料でしたら」のように、「だ」「です」に接続助詞が後続するもの。

例文 103 参加料でしたら、すでに払ってあります。

n-vaux-s+f 例文の「参加料ですが」のように「だ」「です」に接続助詞が後続するもの。接続助詞は、文末でも使えるものに限る³¹。

例文 104 参加料ですが、割引はありますか。

³¹ 例文 72、36 ページ参照。

(<NKVADV <--> (<n-vaux-coord>))
 (<NKVADV <--> (<N-VAUX-COORD-MONEY>))
 (<NKVADV <--> (<n-vaux-h>))
 (<NKVADV <--> (<N-VAUX-H-MONEY>))
 (<NKVADV <--> (<n-vaux-katei>))
 (<NKVADV <--> (<N-VAUX-KATEI-MONEY>))
 (<NKVADV <--> (<n-vaux-s>))
 (<NKVADV <--> (<N-VAUX-S-MONEY>))
 (<NKVADV <--> (<n-vaux-s+f>))
 (<NKVADV <--> (<N-VAUX-S+F-MONEY>))

NKVADV-KEI では、形式名詞に方向性や「を」格を持たない述語が続くものを扱っている。形式名詞は、連体修飾されると仮定しているので、一般的な名詞を扱う NKVADV とは別に扱っているが、その中のカテゴリ分けは同じである。

n-keisiki-vaux-coord 例文 105 の「方で」のように、形式名詞に「だ」の連用形が続くもの。

例文 105 こちらは自動翻訳研究所の方で、英語の生成の部分を担当していらっしゃいます。

n-keisiki-vaux-h 例文 106 の「ものだとか」のように、形式名詞に「だ」「です」が続き、並列の助詞で終る文節。

例文 106 冷蔵庫には、食べるものだとか、飲むものだとかをいつも絶やさない。

n-keisiki-vaux-katei 例文 107 の「ものなら」のように、形式名詞に、「だ」の仮定形が続く文節。

例文 107 食べるものなら、日本のお寿司に限る。

n-keisiki-vaux-s 例文 108 の「ものですし」のように、形式名詞に、「だ」「です」さらに、接続助詞が続くもの。

例文 108 ふつつかな者ですし、御迷惑をおかけすると思いますが、よろしく願います。

n-keisiki-vaux-s+f 例文 109 の「ものですが」のように、形式名詞に、「だ」「です」さらに、接続助詞が続くもの。接続助詞は文末でも使われやすいものに限る³²。

³²例文 76、37 ページ参照。

例文 109 会議に申し込んでいるものですが、会場近くでスキー
ができますか。

(<NKVADV-KEI> <--> (<n-keisiki-vaux-coord>))
(<NKVADV-KEI> <--> (<n-keisiki-vaux-h>))
(<NKVADV-KEI> <--> (<n-keisiki-vaux-katei>))
(<NKVADV-KEI> <--> (<n-keisiki-vaux-s>))
(<NKVADV-KEI> <--> (<n-keisiki-vaux-s+f>))

2.7.5 名詞文節

NN では、名詞文節を扱っている。

np 普通名詞、代名詞または「来月」や「今年」などの時を表す名詞を含む
もの。そのうち普通名詞には助詞が後続しなければならないが、代名
詞や時を表す名詞には、助詞の省略も許している。助詞は、「へ」「を」
以外のものに限る。

wh-np 疑問代名詞に助詞が後続するもの。助詞は、「へ」「を」以外のも
のに限る。文法規則作成の経過理由により、疑問代名詞は、助詞が後
続するものと、しないものに分けている。

wh-n 疑問代名詞。助詞は後続しない。

np-e 普通名詞、代名詞または「来月」や「今年」などの時を表す名詞に方
向性を表す助詞「へ」が後続するもの。

wh-np-e 疑問代名詞に方向性を表す助詞「へ」が後続するもの。

np-o 普通名詞、代名詞または「来月」や「今年」などの時を表す名詞に助
詞「を」が後続するもの。

wh-np-o 疑問代名詞に助詞「を」が後続するもの。

(<NN> <--> (<np>))
(<NN> <--> (<NP-MONEY>))
(<NN> <--> (<wh-np>))
(<NN> <--> (<wh-n>))
(<NN> <--> (<np-e>))
(<NN> <--> (<NP-E-MONEY>))
(<NN> <--> (<wh-np-e>))
(<NN> <--> (<np-o>))
(<NN> <--> (<wh-np-o>))
(<NN> <--> (<NP-MONEY-O>))

NN-KEI では、「もの」「こと」などの形式名詞を含む名詞文節を扱っている。

np-keisiki 形式名詞を扱っている。助詞は、後続しても、しなくてもいい。

助詞は、「へ」「を」以外のものに限る。

np-keisiki-e 形式名詞に方向性を表す助詞「へ」が後続するもの。

np-keisiki-o 形式名詞に助詞「を」が後続するもの。

(<NN-KEI> <--> (<np-keisiki>))

(<NN-KEI> <--> (<np-keisiki-e>))

(<NN-KEI> <--> (<np-keisiki-o>))

QN, NKQN, NKQN-KEI では、引用句を扱っている。

quote 引用を表す「と」が使われている文節。

n-quote 方向性を表すものや「を」格と共起しない述語で、引用の「と」を伴うもの。

n-keisiki-quote 形式名詞に「だ」「です」が後続するもので、引用の「と」を伴うもの。

(<QN> <--> (<quote>))

(<NKQN> <--> (<n-quote>))

(<NKQN> <--> (<N-QUOTE-MONEY>))

(<NKQN-KEI> <--> (<n-keisiki-quote>))

2.7.6 連体修飾文節

連体修飾するものは、活用語で連体形をとるもの (VM, VM-SA) と、助詞の「の」を伴うもの (NM, NM-KEI) がある。活用語で連体形をとるものでは、文末で使える述語句を、VC と「する」だけを扱う VS-SA とに区別したように、「する」で始まる文節 (VM-SA) を他の文節 (VM) と区別している。これは、「読みはした本」のように、活用語の連用形に対照の「は」や、「読みもしなかった本」の「も」などが後続すると、その後に来るのは「する」である、と仮定しているからである。

文末で使われる述語表現は、連体修飾表現とは違うように思われる。例文 110 の「会う」のような言い切りの形は、文末ではあまり使わず、例文 111 の「会います」のように、動詞に助動詞などを付加すると思われる。しかし、連体修飾する場合には、例文 112 の「会う人」のように言うのが

普通であり、例文 113 の「会います人」の方がむしろ不自然であろう。また、例文 114 の「会うのは」のように、準体助詞と共に使われると、不自然さはない。例文 115 の「会いますのは」にも「会います人」ほどの不自然さは感じられない。

例文 110 今日京都で太郎に会う。

例文 111 今日京都で太郎に会います。

例文 112 今日京都で会う人。

例文 113 今日京都で会います人。

例文 114 今日京都で会うのは太郎です。

例文 115 今日京都で会いますのは太郎です。

活用語の連体形と終止形は同じ形のものが多いが³³、例文 116-119 で使われている助動詞の「ようだ」や「そうだ」のように、形の異なるものもある。

例文 116 花子はオペラに行くようだ。

例文 117 花子はオペラに行くような人だ。

例文 118 太郎はやさしそうだ。

例文 119 太郎はやさしそうな人だ。

以上の考察に基づき、VM や VM-SA を、さらに分類した。

nomina 述語文節で準体助詞の「の」を伴い、名詞化されやすいもの。

voux-mod 連体形と、終止形のかたちが違う活用語の連体形が文節末で使われる述語文節。

voux-nom 連体形と、終止形のかたちが同じ活用語が連体修飾句の文節末で使われる述語文節。

nomina-sa 「する」で始まる述語文節で、準体助詞の「の」を伴い、名詞化されやすいもの。

³³ 活用語の連体形と終止形は語彙登録を分ける予定である。現在は、語彙登録を分けているものといないものが混在している。

vaux-sa-mod 「する」で始まり連体形と、終止形のかたちが違う活用語の連体形で終る文節。

vaux-sa-nom 「する」で始まり連体形と、終止形のかたちが同じ活用語が連体修飾句の文節末で使われる述語文節。

(<VM> <--> (<nomina>))
(<VM> <--> (<vaux-mod>))
(<VM> <--> (<vaux-nom>))
(<VM> <--> (<nomina-sa>))
(<VM> <--> (<vaux-sa-mod>))
(<VM> <--> (<vaux-sa-nom>))
(<VM-SA> <--> (<nomina-sa>))
(<VM-SA> <--> (<vaux-sa-mod>))
(<VM-SA> <--> (<vaux-sa-nom>))

これからの課題 16 活用語が準体助詞を伴う時の表現は、名詞を連体修飾する時の表現とは、違うように思われる。これらが、どの程度違うか対話データベースを基に検討する必要がある。

これからの課題 17 「する」については、準体助詞の「の」が後続する規則はまだ取り入れていないので、早急に対処するべきであろう。

これからの課題 18 バージョン 1.1 では、活用語の連用形に後続する助詞は、「は」「も」「でも」としているが、「読みさえしなかった本」のように、それ以外の助詞が後続することもあるので、さらに検討を要する。

NM, NM-KEI では、助詞の「の」で終る文節を扱っている。NM-KEI では、その中で、形式名詞を含む文節を扱っている。

n-rentai 「どのような」や「そのような」のような文節。連体詞の扱いは、まだ決定していないので、形態素情報利用解説書 [14] で、連体詞に助動詞が後続するとしているものを扱っている³⁴。

np-no 「花子の本」「私の本」のように、形式名詞以外の名詞や代名詞に格助詞の「の」が後続するもの。

wh-np-no 「これは、誰の本ですか」の「誰の」のように、疑問代名詞に「の」が後続するもの。

adv-p-no 「ちょっとの油断」の「ちょっとの」のように、副詞に「の」が後続する文節。

³⁴ 2.5.2 節、28 ページ参照。

vn-no 「今日限りの出勤」「遊びの精神」など、動詞の連用形に「の」が後続し、連体修飾要素となるもの。「この件に関しの解決策」「反則がありの試合」などは、あまり使われなれないと思われるので、vn-noでは、扱っていない。

np-keisiki-no 「ものの」「ことの」のように形式名詞に助詞の「の」が後続するもの。形式名詞は、連体修飾されることを仮定しているため、他の名詞とは別に扱っている。

np-special 「に対する」が文節末に使われているもの。「に対する」は、形態素情報利用解説書では、格助詞に本動詞の連体形が続くものとしているが³⁵、バージョン 1.1 では、便宜上一つの助詞として扱っている。

```
(<NM> <--> (<n-rentai>))
(<NM> <--> (<np-no>))
(<NM> <--> (<NP-NO-MONEY>))
(<NM> <--> (<wh-np-no>))
(<NM> <--> (<adv-p-no>))
(<NM> <--> (<vn-no>))
(<NM-KEI> <--> (<np-keisiki-no>))
(<NPM> <--> (<np-special>))
```

これからの課題 19 バージョン 1.1 では、名詞などに「の」がつくものは、すべて連体修飾要素として扱っている。例文 120 の「チョムスキーの書いた」のように、埋め込み文の中で使われる「の」は、主格の「が」を表すが、現在は、連体修飾要素と解される。

例文 120 チョムスキーの書いた本は、よく売れる。

また、「書いた」を飛び越して「本は」を修飾することは許されていないので、「チョムスキーの」は「書いた本は」を修飾することになる。

2.8 助詞の「で」「なら」と助動詞の「で」「なら」

助詞の「で」「なら」と助動詞「です」が活用した「で」「なら」との区別は難しい。音声認識では、正しい音素列を得ることに重点がおかれるので、「で」および「なら」は曖昧性を残したまま処理している。

例文 121 国際会議で発表した。

例文 122 国際会議ならいってみよう。

³⁵参考文献 [14]、59 ページ参照。

例文 121、122 の「で」および「なら」は、それぞれ格助詞、副助詞である。形態素情報利用解説書によると、「なら」が体言の後にくる時は、副助詞である³⁶。この「国際会議で」や「国際会議なら」は以下のように、「は」を含む文節や「が」格を持つ文節と共起し、複文で使われることがある。

例文 123 *ASTI* は国際会議でたくさんの人が発表した。

例文 124 *ASTI* が国際会議ならいってみよう。

「*ASTI* は国際会議で」の「で」は、「だ」の連用形、「*ASTI* が国際会議なら」の「なら」は、「だ」の仮定形であろう。さらに、話し言葉ということに重点をおくと、例文 126 や例文 127 のように文節の省略は頻繁に起きる。

例文 125 *ASTI* って何ですか？

例文 126 国際会議でたくさんの人が発表する予定ですよ。

例文 127 そうですか。国際会議ならいってみよう。

例文 126 の「国際会議で」または、例文 127 の「国際会議なら」は文単位で見ると、例文 121 や例文 122 と同じ使われ方をしている。しかし、これらは格助詞や副助詞ではなく、例文 123 や例文 124 のように解釈されるべきである。従って、それぞれ「だ」の連用形、「だ」の仮定形とするべきであろう。

バージョン 1.1 とセットになっている文節を構成する規則では、名詞に助詞の「で」や「なら」が後続するという規則に加えて名詞に助動詞「だ」の連用形「で」や「だ」の仮定形「なら」が後続するという規則がある。

```
(<n-vaux-coord> <--> (<n> <cop-renyo1>))  
(<n-vaux-katei> <--> (<n> <cop-katei>))
```

このように、文節組み立て規則では、「で」や「なら」を特別扱いしていないので、異なる規則から同じ文字列が生成される。また、*n-vaux-coord* と *n-vaux-katei* は、文節に基づき文を組み立てる規則では、助詞「へ」や「を」を含む名詞句と共起しない述語を含む副詞句 *NKVADV* として扱われるので、曖昧性が残る。

³⁶ 参考文献 [14]、54 ページ参照。

2.9 NADINE の CFG 規則

1990 年 9 月に日本語対話文の解析用文法に関して、語彙規則の中で、選言として扱われていたものの一部が、句構造規則に書き直された。この際、住所に関する規則は、以下のものになった。

N-ward -> (N-city N-ward)
大阪市北区
N-town -> (N-ward N-town)
大阪市北区茶屋町
N-chome -> (N-town N-chome)
大阪市北区茶屋町二丁目
N-banchi -> (N-chome N-num)
大阪市北区茶屋町二丁目二三

N-city は、市に関するもの、N-ward は、区に関するものという定義があるので、これを第一番目の N-ward の書き換え規則に適用すると、「大阪市北区」の他にも、「大阪市大阪市北区」「大阪市大阪市大阪市北区」「大阪市大阪市大阪市大阪市北区」などが生成される。また、このような CFG 規則の書き方が全体的に使われているので、1.2 節の、認識結果 1、認識結果 2 のような生成もされる。

このような CFG 規則は、入力が正しいものという仮定のもとで行なわれる解析、または非常に性能のいい音声認識を補助するものとしては十分であるが、現在の我々の音声認識システムに適用するには、危険性が高い。

第 3 章

文節を構成する規則

本構文規則の音声認識への応用は、音声認識の際に認識率を向上させることを目的としている。そのため、制約として有効に働くように規則を構築している。一方、ATR では、日英機械翻訳用に可能な限りの解析をする構文解析規則も開発している。これら 2 つの構文規則は、それぞれの目的は異なるが、少なくとも一部の処理は重複して行なっている。そこで、両者を融合し、この重複を避けようという動きがある。現在、その第一段階として、対話データベース形態素解析の基準である、形態素情報利用解説書 [14] を基に、二種類の構文規則の形態素単位を見直している。第 3 章では、この見直しの一貫として、対話データベース形態素解析で形態素として扱っているものに、音声認識のための文節を構成する規則 (*dp10-re72.gra*) で定義している形態素を対応させることを試みる。*dp10-re72.gra* で定義している形態素は、ほぼ対話データベース形態素解析の形態素を細分化したものに相当する。第 3 章の構成は、形態素情報利用解説書に準じる。

形態素をどのように文節に組み上げているかは、一部、第 3 章で紹介するが、詳細は以下の構文規則を読んで確認してもらいたい。

```
atr-dp:/ifg2/HMM-LR/Grammar/dp10-re72.gra
```

3.1 名詞

これからの課題 20 複合名詞は、分割せず一つの形態素として扱っている。構文規則の精密化へ向けて、規則化が望まれる。

3.1.1 固有名詞

固有名詞は、氏名に関するもの、住所に関するものとそれ以外の固有名詞という 3 種類のグループに分けて扱っている。

氏名、住所以外の固有名詞

n-proper 固有名詞。氏名、住所に関するものを除く。情報処理学会、京都国際会議場、清水寺。

氏名

氏名は、苗字と名前を区別している。さらに、ATR でターゲットとしている日本語と英語も区別している。これは、口頭で名前を言う場合、日本語と英語では苗字と名前の順序が異なるためである。苗字と名前が一緒に使われる場合は、日本語では苗字が名前の前に来るように、また、英語では名前が苗字の前に来るようにしている。英語のミドルネームなどは考慮していない。

n-nachname 日本語の苗字。宝、鈴木、清水。

n-nachname-others 英語の苗字。スミス、ワイベル、ブラウン。

n-vorname 日本語の名前。一太郎、真由美、太郎。

n-vorname-others 英語の名前。アダム、ジョージ、ケン。

住所

住所の構成は、地方により異なるようであるが、*dp10-re72.gra* ではモデル会話で使っている構成を扱っている。基本にしているのは例文 128 である。

例文 128 大阪市北区茶屋町 23 です。

x を変数とした時、*ken-name-x* では、「大阪府」に当たる部分を扱っている。また、*si-name-x* では、「大阪市」に当たる部分を、*ku-name-x* では、「北区」に当たる部分を扱っている。住所に関する固有名詞は、「県」「区」「町」などの接尾辞により分けている。

ken-name-to 接尾辞「都」が後続するもの。ここで扱うのは、「東京」だけであるが、「東京都」と使ったり、「東京」と使ったりするので、登録は、「東京」とする。

ken-name-dou 接尾辞「道」が後続するもの。「都」同様、ここで扱うのは、北海道だけである。「道」を省略し、「北海」と言うことはほぼないと思われるので、登録は、「北海道」とする。

ken-name-hu 接尾辞「府」が後続するもの。大阪、京都。

ken-name-ken 接尾辞「県」が後続するもの。青森。

si-name-si 接尾辞「市」が後続するもの。大阪。

si-name-ku 接尾辞「区」が後続するもの。豊島。

ku-name-ku 接尾辞「区」が後続するもの。北、東。

adre-tyo 茶屋町に当たる部分を扱っている。茶屋町、東池袋。

```
(<n-proper> <--> (z y o u h o u s h o r i n g a q k a i))
(<n-nachname> <--> (t a k a r a))
(<n-nachname-others> <--> (s u m i s u))
(<n-vorname> <--> (i c h i t a r o o))
(<n-vorname-others> <--> (a d a m u))
(<ken-name-to> <--> (t o u k y o u))
(<ken-name-dou> <--> (h o q k a i d o u))
(<ken-name-hu> <--> (o o s a k a))
(<ken-name-ken> <--> (a o m o r i))
(<si-name-si> <--> (o o s a k a))
(<si-name-ku> <--> (t o s h i m a))
(<ku-name-ku> <--> (k i t a))
(<adre-tyo> <--> (c h a y a m a c h i))
```

気を付けて 1 対話データベース形態素解析では、「大阪市」や「北区」をそれぞれひとまとまりの形態素として扱っている¹。しかし、*dp10-re72.gra* では、「市」や「区」を接尾辞として扱っているのが、形態素の扱いにずれが生じている。対話データベースで形態素としているものには、これらのものをまとめあげた *adre-ken-to* など、以下の 7 つが対応する。

```
(<adre-ken-to> <--> (<ken-name-to> <ken-name-suf-to>))
(<adre-ken-dou> <--> (<ken-name-dou>))
(<adre-ken-hu> <--> (<ken-name-hu> <ken-name-suf-hu>))
(<adre-ken-ken> <--> (<ken-name-ken> <ken-name-suf-ken>))
(<adre-si-ku> <--> (<si-name-ku> <si-name-suf-ku>))
(<adre-si-si> <--> (<si-name-si> <si-name-suf-si>))
(<adre-ku-ku> <--> (<ku-name-ku> <ku-name-suf-ku>))
```

3.1.2 サ変名詞

サ変名詞では、学校文法でサ変動詞の語幹としているものを扱っている。接頭辞を含むものもサ変名詞と同様に扱っている。

pre-v-sahen 接頭辞にサ変名詞が後続しているもの。お願い、御覧。

¹参考文献 [14]、13 ページ参照。

v-sahen サ変名詞。用意、翻訳。

(<pre-v-sahen> <--> (o n e n g a i))

(<v-sahen> <--> (y o u i))

気を付けて 2 接頭辞を含むサ変名詞は、対話データベース形態素解析と *dp10-re72.gra* では、扱い方が違う。例文 129 の「お願い」は、対話データベース形態素解析では 2 つの形態素として扱っているが、*dp10-re72.gra* では、一つの形態素として扱っている。

例文 129 お願いします。

データベース	
お	願い
接頭辞	サ変名詞

HMM-LR	
お願い	
pre-v-sahen	

これからの課題 21 サ変名詞に接頭辞が前接するものは、一つの形態素として扱っているため、これを二つの形態素として扱っている対話データベース形態素解析に合わせるには、サ変名詞の細分化が必要になる。接頭辞として「御」と「お」のみを考慮した場合、サ変名詞を少なくとも 4 つに分類するべきであろう。

- 接頭辞に「御 (ゴ)」をとるもの：希望、紹介
- 接頭辞に「お」をとるもの：申し込み
- 接頭辞を取り難いもの：失礼、関連
- 動詞の連用形に接頭辞がついたもの。これは、接頭辞なしでは、サ変名詞として「する」に後続され難い：お知らせ、お持ち、お伺い

3.1.3 形容名詞

形容名詞では、学校文法で形容動詞としているものの語幹を扱っている。

kdo 形容名詞。必要、広範。

(<kdo> <--> (h i t s u y o u))

これからの課題 22 「同じ」は例文 130 の「同じだ」のように「だ」や「です」を伴って使われる形容名詞である。形容名詞「必要」は、活用して連体形をとり、「必要な」と使われる。しかし、例文 131 が不自然なことからわかるように、「同じな」という連体形を使うことはできない。

例文 130 このリンゴとあのリンゴの色は同じだ。

例文 131 * 色が同じなリンゴがある。

これからの課題 23 「的」がつくものは、形容名詞として扱っている。また、例文 132 の「専門的な」のように、形容名詞は「だ」や「です」の活用形を伴うとしている。

例文 132 こちらでは専門的な質問にお答えできません。(会話 7)

「的」を伴うものは、例文 133 のように連用形「に」を伴うものと同じ意味で、例文 134 のように使うことができると思われる。しかし、「絶対的」は、例文 135 と同じ意味で例文 136 のように使うことはできない。

例文 133 比較的に自由だ。

例文 134 比較的自由だ。

例文 135 絶対的に自由だ。

例文 136 * 絶対的自由だ。

3.1.4 普通名詞

日本語では、名詞に後続する助詞が省略されることがある。しかし、これをそのまま構文規則に反映させ、すべての普通名詞の後に助詞が後続しなくてもいいという規則を与えると、音声認識の際、文節候補が非常に増える。助詞が省略されている普通名詞句を対話データベースで検索したところ、副詞的に働く、時を表すものが大半であった。そこで、*dp10-re72.gra* では、時を表す名詞を別に扱い、助詞が省略されてもいいという規則を与えている。また、いわゆる形式名詞は、連体修飾されずに使われることが少ないので、別に扱っている。さらに、月日に関するものは、対話データベースと形態素の切り方が異なる。

n-hutu 普通名詞。会議、登録用紙、手続き。

n-keisiki いわゆる形式名詞。連体修飾されずに使われることがめったにないもの。もの、こと、かた。

n-hutu-time 時をあらわす名詞で、接尾辞「いっぱい」に後続されにくいもの。現在、今回。

n-ippai 「いっぱい」が後続できる、時に関する名詞。月日に関するものを除く。「十月いっぱい」、「十月十日いっぱい」のように、月日には「いっぱい」が後続できる。n-ippai では、月日以外で、「いっぱい」が後続できる名詞を扱っている。来月、今年。

n-exc 助詞を伴わずに使われることが多いもの。時をあらわす名詞を除く。
場合。

(<n-hutu> <--> (k a i n g i))
(<n-keisiki> <--> (m o n o))
(<n-hutu-time> <--> (g e = z a i))
(<n-ippai> <--> (r a i n g e t s u))
(<n-exc> <--> (b a a i))

気を付けて 3 サ変名詞同様、接頭辞を含む普通名詞は、対話データベース形態素解析と *dp10-re72.gra* では、扱い方が違う。「お名前」や「御住所」は、対話データベース形態素解析では、それぞれ「お」と「名前」または「御」と「住所」の 2 つの形態素として扱っているが、*dp10-re72.gra* では、一つの形態素として扱っている。

これからの課題 24 接頭辞を伴う普通名詞は、一つの形態素として扱っている。一方、対話データベース形態素解析では、このような普通名詞を、二つの別の形態素として扱っている。形態素の切り方を合わせるには、接頭辞として「御」と「お」だけを考慮した場合でも、普通名詞を少なくとも 3 つに分類するべきであろう。

- 接頭辞に「御 (ゴ)」をとるもの：住所、要件
- 接頭辞に「お」をとるもの：名前、へや
- 接頭辞を取り難いもの：歓迎会費、意味、英語

月日に関する普通名詞

数詞を含むものは、基本的に数詞と接尾辞などに分けて規則を組み立てている²。「一月」「二月」などの月を表すものは、「一」と「月」または「二」と「月」に分けている。

「一日」や「三十一日」は、「一」と「日」または、「三十一」と「日」に分けている。「三十一」は、さらに数詞を組み立てる tag-ka や tag-niti のような規則により生成している。生成範囲は、「一日」から「三十一日」までである。また、接尾辞（助数詞）としては、「タチ」「カ」「ニチ」を考慮している。

²月日に関するものは、かなり有限なので、規則を使い生成することはやめ、語彙登録する方がいいのかも知れない。

(<tag-ka> <--> (<tag-10> <tag-ka-4>))
ジュウヨッカ、ニジュウヨッカ

(<tag-niti> <--> (<tag-10> <tag-1>))
ジュウイチニチからニジュウクニチまで(ジュウヨッカ、ニジュウヨッカ
を除く)

(<tag-niti> <--> (<tag-30> <tag-31>))
サンジュウイチニチ

monate 月を表す数詞。一、二。

monate-suf 月を表す数詞に伴う助数詞。月(ガツ)。

tag-erste 月の初めの日を表す数詞。一(ツイ)。

tag-suf-1 月の初めの日を表す数詞に伴う助数詞。日(タチ)。

tag-ka 月日の日を表すのに助数詞「カ」を伴う数詞。四(ヨッ)、六(ム
イ)、十(トウ)、二十(ハツ)。

tag-suf-ka 月日の日を表す助数詞「カ」。日(カ)。

tag-suf-niti 月日の日を表す助数詞「ニチ」。日(ニチ)。

tag-1 二桁以上の日を表す時に助数詞「ニチ」を伴う数詞。「一」から「九」
まで。「七」は「シチ」と「ナナ」の両方が可能。

tag-ka-4 二桁以上の日を表す時に助数詞「カ」に前接する数詞。四(ヨッ)。

tag-10 十(ジュウ)、二十(ニジュウ)。

tag-30 三十(サンジュウ)。

tag-31 三十一(サンジュウイチ)。

(<monate> <--> (i ch i))
(<monate-suf> <--> (g a ts u))
(<tag-erste> <--> (ts u i))
(<tag-suf-1> <--> (t a ch i))
(<tag-ka> <--> (h u ts u))
(<tag-suf-ka> <--> (k a))
(<tag-suf-niti> <--> (n i ch i))
(<tag-1> <--> (i ch i))
(<tag-ka-4> <--> (y o q))
(<tag-10> <--> (zy uu))
(<tag-30> <--> (s a = zy uu))
(<tag-31> <--> (i ch i))

気を付けて 4 対話データベース形態素解析では、「一月」「二月」などの月は、一つの形態素扱いをしている³。しかし、*dp10-re72.gra* では、数詞にあたるものは、別扱いをするという立場から、形態素分けをしている。また、「一日」「二十二日」などの日も、対話データベース形態素解析では、一つの形態素として扱っているが、*dp10-re72.gra* では、月の扱い同様、形態素分けをしている。datum-monate や datum-tag が、対話データベースの形態素に対応する。

```
(<datum-monate> <--> (<monate> <monate-suf>)  
(<datum-tag> <--> (<tag-erste> <tag-suf-1>))  
(<datum-tag> <--> (<tag-ka> <tag-suf-ka>))  
(<datum-tag> <--> (<tag-niti> <tag-suf-niti>))
```

3.1.5 数詞

数詞は、住所、金額、電話番号、人数の中で使われるものを扱っている。

住所

住所の中で使われる数詞には、「一番」のように助数詞を伴うもの、「23 の」のように助詞が後続するもの、または、「99 です」のように Copula が続くものがある。

dp10-re72.gra では、例文 137 や例文 138 のような住所を扱えるように規則を構築した。

例文 137 住所は東京都港区新橋一丁目一番三号です。(会話 10)

例文 138 住所は東京都港区新橋一丁目 23 の 99 です。

モデル会話には住所に関する文で助数詞を伴う例文 137 の「一丁目」、「一番」や「三号」がある。そこで、助数詞としては、「丁目」⁴、「番」と「号」を考慮している。「丁目」と「番」は、文節の終わりに使われ、「号」は、Copula と共に使われるとしている。数詞の部分は、ひとけただけを認めている。

さらに、例文 138 の「23 の」や「99 です」のように、助詞「の」を伴ったり、Copula が続くものを扱っている。その際、数詞は 1 から 99 までを扱っている。

数詞の一の位は、後続する助数詞の発音に合わせて語彙登録している。「五」に関しては、文節のはじめとそれ以降でそれぞれ (g o) と鼻音化し

³参考文献 [14]、9 ページ参照。

⁴*dp10-re72.gra* では、「丁目」は一つの形態素としている。一方、atr-dp 上の対話データベース形態素解析では、「丁」と「目」という二つの接尾辞が連鎖したものとして扱っている。参考文献 [14]、13 ページでは、「丁目」は一つの接尾辞として扱っているのだが ... ?

た (ng o) というように、発音が変わるということを考慮し、二つの語彙登録をしていたものの名残 (num-1-g) がある⁵。2 から 4 までと 6, 7, 9 は、num-1 で扱っている。そのうち、「四」に関しては、「ヨン」という発音を与えている。「シ」は母音の無声化が起こりやすく、誤認識の対象になりやすいので、扱っていない。同様に、「七」は「ナナ」という発音を与え、「シチ」は認めていない。また、「九」は「キュウ」を許し「ク」は認めていない。

num-1-spe 助数詞「番」や「号」が後続する時使われる「一」。一 (イチ)。

num-1-ch 助数詞「丁目」が後続する時使われる促音になる数詞。一 (イツ)、八 (ハッ)。

num-1 数詞の「二」から「七」までと「九」。二 (ニ)、三 (サン)。

num-1-g 数詞の「五」。五 (ゴ)。

num-1-b 助数詞「番」や「号」が後続する時使われる「八」。八 (ハチ)。

num-10 数詞の「十」。十 (ジュウ)。

```
(<num-1-spe> <--> (i ch i))
(<num-1-ch> <--> (i q))
(<num-1> <--> (ni))
(<num-1-g> <--> (g o))
(<num-1-b> <--> (h a ch i))
(<num-10> <--> (zy uu))
```

気を付けて 5 数詞は基本的に規則により生成している。住所に含まれる数詞に関しては、num が、対話データベース形態素解析の形態素に対応する。

```
(<num> <--> (<num-1>));;2-9
(<num> <--> (<num-10-1>));;11-19
(<num-10-1> <--> (<num-10> <num-1>))
```

金額

数詞と共に用いられる「百」「千」「万」などの助数詞は、対話データベース形態素解析ではその前の数詞と共に一つの形態素である数詞として扱っている。一方、dp10-re72.gra では、すべての数詞を語彙登録するという方法はとっていない。「二百」は、数詞の「二」と助数詞の「百」に形態素分けするため、対話データベースの形態素切りと違いが生じている。

⁵現在 (g) と (ng) の違いは、HMM-LR パーザで吸収している。

助数詞としては、「百」「千」「万」を扱っている。これらの助数詞は、数詞との組合せにより、発音が変わる。また、数詞の発音も前接または後接するものにより変わる。金額に使われる数詞を組み上げる規則の例を示す⁶。

```
(<money-hyaku> <--> (<money-num-hyaku-b> <money-hyaku-suf-b>))
(<money-sen> <--> (<money-num-sen-spe> <money-zen-suf>))
(<money-man> <--> (<money-num-man> <money-man-suf>))
```

金額に関する数詞、助数詞は説明を加えたものは列挙しない。以下、語彙登録を示す。

助数詞:

```
(<money-hyaku-suf> <--> (h y a k u))
(<money-hyaku-suf-b> <--> (b y a k u))
(<money-hyaku-suf-p> <--> (p y a k u))
(<money-sen-suf> <--> (s e =))
(<money-zen-suf> <--> (z e =))
(<money-man-suf> <--> (m a =))
```

数詞:

```
(<money-num-hyaku> <--> (n i))
(<money-num-hyaku-b> <--> (s a =))
(<money-num-hyaku-p> <--> (r o q))
(<money-num-man-spe1> <--> (i c h i))
(<money-num-man-1> <--> (n i))
(<money-num-man-10> <--> (z y u))
(<money-sen> <--> (s e =))
(<money-num-sen> <--> (i q))
(<money-num-sen-spe> <--> (s a =))
(<money-num-1-spe1> <--> (i c h i))
(<money-num-1> <--> (n i))
(<money-num-1-spe4> <--> (y o))
(<money-num-1-spe44> <--> (y o =))
(<money-num-10> <--> (z y u))
```

気を付けて6 対話データベース形態素解析では、「三万五千円」を「三万五千」と「円」にわけ、それぞれ数詞と接尾辞という形態素を付与している。一方、音声認識の入力では、「三万」と「五千円」という二つの文節に分けている。ここでは、音声認識の入力の文節切りに合わせているので、対話データベースの形態素切りとは異なる。文節を基に前処理を行なって得られる句⁷が、対話データベースの形態素に対応するであろう。

⁶ 「三万円」や「五千円」のように一つの位だけを扱っている場合には money-hyaku, money-sen, money-man が、対話データベース形態素に対応する。

⁷ 2.6.1節、30 ページ参照。

電話番号

3桁と4桁の電話番号を扱い、ゼロからキュウまでの数詞を規則により生成している。「四」「七」「九」には、それぞれ「ヨン」「ナナ」「キュウ」という発音を与えている。

tel-num ゼロからキュウまでの数詞。

電話番号は、以下のように生成している:

```
(<tel-3> <--> (<tel-num> <tel-num> <tel-num>))  
(<tel-4> <--> (<tel-num> <tel-num> <tel-num> <tel-num>))
```

```
(<tel-num> <--> (z e r o))
```

気を付けて7対話データベース形態素解析では、例文139の「サンナナニ」と「ハチゼロイチハチ」がそれぞれ一つの形態素、数詞、であるとしている。

例文 139 電話番号は 372 の 8018 です。

一方、*dp10-re72.gra*では、規則による生成をしている。これは、それぞれの数字の組合せを形態素として一つずつ登録したのでは、辞書の登録数を非常に増やすからである。従って、tel-3 と tel-4 が対話データベース形態素に対応する。

人数

「三人」「五人」のような人数を表すものは、住所と金額のように、数詞と接尾辞の組合せとして扱っている。但し、「ひとり」「ふたり」「おひとり」「おふたり」はそれぞれ一つの形態素として扱っている。

human-num-1-spe1 二桁以上の人数を表す時に接尾辞「人」に前接する「一」。

human-num-1-spe2 二桁以上の人数を表す時に接尾辞「人」に前接する、または「十」に前接する「二」。

human-num-1 「三」から「九」まで。「四」には「ヨ」、「七」には「ナナ」と「シチ」、「九」には「キュウ」と「ク」という発音を与えている⁸。

human-num-10 「十」(ジュウ)。

⁸human-num-1 では、一の位と十の位を扱っている。しかし、「四」には「ヨ」という発音を与えているので、規則により「四十」に対しては「ヨジュウ」が生成される。また、「九」には「キュウ」と「ク」という発音を与えているので、「九十」に対しては「キュウジュウ」の他に「クジュウ」も生成される。

(<human-num-1-spe1> <--> (i ch i))
(<human-num-1-spe2> <--> (n i))
(<human-num-1> <--> (s a =))
(<human-num-10> <--> (zy uu))

気を付けて 8 数詞は規則により生成しているが、対話データベース形態素解析では、「ニジュウイチ」「ロクジュウゴ」のように、数がまとまったところで、数詞という形態素を付与している。数詞を生成する以下のような規則 human-num が対話データベースの形態素に対応する。

(<human-num> <--> (<human-num-10> <human-num-1-spe1>))
(<human-num> <--> (<human-num-1> <human-num-10>))

気を付けて 9 「ひとり」と「ふたり」は、一つの形態素として扱っているが、対話データベース形態素解析では数詞と接尾辞に分けている。

(<count-nin> <--> (h i t o r i))
(<count-nin> <--> (h u t a r i))

気を付けて 10 「おひとり」と「おふたり」は、一つの形態素として扱っているが、対話データベース形態素解析では接頭辞、数詞と接尾辞に分けている。

(<count-nin> <--> (o h i t o r i))
(<count-nin> <--> (o h u t a r i))

これからの課題 25 人数を表す接尾辞としては、「人」(ニン)を扱っているので、「四」には「ヨ」という発音しか与えていないが、さらに「名」(メイ)を扱う場合には、「ヨン」という発音も与えなければならない。

3.1.6 代名詞

代名詞は、疑問代名詞とそれ以外の代名詞に分けている。疑問代名詞では、例文 140 の「ナニ」と同じように例文 141 の「ナン」が使われる。「ナニ」は、さらに例文 142 のようにも使われる。しかし、標準日本語では例文 143 のように「ナン」を使わないことから「ナン」は、別扱いをしている。

例文 140 ナニですか。

例文 141 ナンですか。

例文 142 ナニがいいですか。

例文 143 ナンがいいですか。

疑問代名詞以外の代名詞は、「ワタシドモ」の「ワタシ」のように、接尾辞「ドモ」を伴えるものと、「コレ」のように「ドモ」を伴い難いものとに分けている。「ワタシ」のような代名詞が、接尾辞「ドモ」を伴うものは、以下の規則により、他の代名詞と同じ扱いをしている:

(<pro> <--> (<pro1> <pro-suf1>))

wh-pro 疑問代名詞。「ナン」以外のもの。

wh-pro-N 疑問代名詞の「ナン」。

pro 代名詞。接尾辞の「ドモ」を伴わないもの。

pro1 代名詞。接尾辞の「ドモ」を伴えるもの。

(<wh-pro> <--> (i t s u))
(<wh-pro-N> <--> (n a =))
(<pro> <--> (k o r e))
(<pro1> <--> (w a t a s h i))

対話データベースの検索に基づき、後続する助詞により、疑問代名詞の細分化を試みた。その結果、疑問代名詞を以下の7つのカテゴリーに分けた。dp10-re72.gra では、このように細分化したものと、上記のものが共存している。

(<wh-pro-itu> <--> (i t s u))
(<wh-pro-itu> <--> (n a n i))
(<wh-pro-N> <--> (n a =))
(<wh-pro-dare> <--> (d a r e))
(<wh-pro-iku> <--> (i k u r a))
(<wh-pro-doko> <--> (d o k o))
(<wh-pro-doko> <--> (d o c h i r a))
(<wh-pro-don> <--> (d o n a t a))
(<wh-pro-dore> <--> (d o r e))

これからの課題 26 「ドモ」以外にも「これら」の「ラ」、「あなたがた」の「ガタ」などの接尾辞がある。接尾辞をどの程度詳細に扱うか、基本線を決めるべきであろう。

3.2 動詞

これからの課題 27 複合動詞は、分割せず一つの形態素として扱っている。対話データベース形態素解析では、個々の動詞 / 複合動詞に対処するためのアルゴリズムを

設けている⁹。翻訳用の日本語解析部との構文規則の統一を目指した時、もう少し一般化した判断基準が望まれるだろう。

3.2.1 本動詞

本動詞は活用形により一段動詞、五段動詞とサ変動詞に分けている。サ変動詞では「する」を扱っている。「勉強する」のようないわゆるサ変動詞は、サ変名詞に「する」の活用語尾が後続したもの¹⁰として扱っている。さらに、一段動詞では、可能の助動詞「れる」「られる」と共起しない可能動詞を別に扱っている。

一段動詞

一段動詞 下一段活用、上一段活用をするものは、語幹として、母音の「え」または「い」までを扱っている。「え」または「い」以降の活用は下一段、上一段ともに同じなので、一段活用動詞として扱っている。語幹の部分は、v-1dan に、語彙登録している¹¹。「申し上げる」のような複合動詞は、さらに細かく形態素分けすることはせずに、一つの形態素として扱っている。

```
(<v-mizen4> <--> (<v-1dan> <flex-1dan-mizen4>))
(<v-1dan> <--> (i r e))
(<v-1dan> <--> (m ou sh i a ng e))
(<v-1dan> <--> (m o ch i i))
```

可能動詞 例文 144 の「できる」や例文 145 の「払える」のような可能動詞を扱っている。これらは、使役の「せる / させる」を伴い「できさせる」や「払えさせる」、または、可能の「れる / られる」を伴い、「できられる」や「払えられる」のように使わない。

例文 144 今の大学生は英語がよくできる。

例文 145 資産家の息子なので、このくらい現金で払えるだろう。

語幹の部分は、deki-v-1dan に、語彙登録してある。語幹と活用部の接続を示す:

```
(<deki-rentai> <--> (<deki-v-1dan> <flex-1dan-rentai2>))
```

⁹参考文献 [14]、28 ページ参照。

¹⁰3.2.2 節、68 ページ参照。

¹¹活用語には、一段動詞のように語幹と活用部を分けているものがある。この場合、活用部は flex-*x* で表している。

可能動詞の一部は、サ変名詞に後続する補助動詞として働くこともある¹²。

(<deki-v-1dan> <--> (h a r a e))

(<deki-v-1dan> <--> (d e k i))

五段動詞

五段活用する動詞は、否定の助動詞「ない」(nai)を後続させた時の、「あ」(a)の直前の子音により分類している。「書かない」(ka k anai)は‘k’(‘ki’)型である。同様に、「持たない」(mo t anai)は‘t’型、「喜ばない」(yoroko b anai)は、‘b’型等。dp10-re72.graでは、「行く」と「ある」を別に扱っている。「行く」と同じ‘k’のグループに属する「書く」や「着く」は、助動詞の「た」を伴うと「書いた」および「着いた」になる。一方「行く」は、助動詞の「た」を伴うと「行った」になる。また、「ある」に助動詞の「ない」を後続させて「あらない」とは使えないためである。「行く」と「ある」に加え、‘ki’ ‘g’ ‘s’ ‘t’ ‘b’ ‘m’ ‘r’ ‘w’の8つのグループに分類している。「死ぬ」のように‘n’のグループに属するものはまだ扱っていない。「払い戻す」のような複合動詞は、一つの形態素として扱っている。語幹と、活用部分とは別に語彙登録している。

語幹の語彙登録を以下に示す:

(<v-5dan-ki> <--> (k a))

(<iku-5dan-ki> <--> (i))

(<v-5dan-g> <--> (n o r i t s u))

(<v-5dan-s> <--> (t o r i k e))

(<v-5dan-t> <--> (m o))

(<v-5dan-b> <--> (y o r o k o))

(<v-5dan-m> <--> (m o u s h i k o))

(<v-5dan-r> <--> (k a k a))

(<v-arū> <--> (a))

(<v-5dan-w> <--> (i))

対話データベース形態素解析で一形態素として扱っているのは、活用形を含むものである。以下に v-5dan-s を語幹として活用したものを示す。

他のグループの活用もほぼ同様に扱っている:

(<v-5dan-a> <--> (<v-5dan-s> <flex-5-sa>))

(<v-5dan-i-s> <--> (<v-5dan-s> <flex-5-si>))

(<v-5dan-u> <--> (<v-5dan-s> <flex-5-su>))

(<v-5dan-e> <--> (<v-5dan-s> <flex-5-se>))

(<v-5dan-o> <--> (<v-5dan-s> <flex-5-so>))

(<v-renyo2-t> <--> (<v-5dan-s> <flex-r2-t-s>))

¹²3.2.2 節、66 ページ参照。

「ある」の荘重な言い方である「ござる」は、現在はほとんど「ございます」という丁寧形で用いられているので、語彙登録は「ございます」である。

活用形は、以下のように組み立てている。

```
(<polt-v-rentai> <--> (<polt-v-aruru> <flex-polt-aruru-rentai>))  
(<polt-v-aruru> <--> (g o z a i m a))
```

気を付けて 11 対話データベース形態素解析では、「ございます」を、補助動詞と本動詞という二つの異なる形態素として扱っている。例文 146 のように感動詞に後続する場合には、補助動詞であり、例文 147 のように使われる場合には本動詞である。

例文 146 ありがとうございます。

例文 147 分からない点がございましたら、いつでもお聞き下さい。(会話 2)

dp10-re72.gra では、本動詞「ある」の丁寧形「ございます」に補助動詞としての働きも加え両者を区別していない。

サ変動詞「する」

動詞は、ほとんどのものを、語幹と活用部分とに分けて扱っているが、「する」は、助動詞同様、活用した時点で、語彙登録している¹³。

```
(<v-sa-mizen1> <--> (sh i))  
(<v-sa-mizen2> <--> (sh i y o))  
(<v-sa-mizen3-4> <--> (s a))  
(<v-sa-mizen5> <--> (s e))  
(<v-sa-renyo1> <--> (sh i))  
(<v-sa-renyo2-t> <--> (sh i))  
(<v-sa-rentai> <--> (s u r u))  
(<v-sa-katei> <--> (s u r e))  
(<v-sa-meirei> <--> (sh i r o))
```

3.2.2 補助動詞

補助動詞は、一段活用をする補助動詞、五段活用をする五段補助動詞、サ変活用をする「する」と形容詞型の活用をする補助動詞に分けている。一段活用をする補助動詞としては、可能の意味を含む可能補助動詞と「いる」を扱っている。サ変名詞に後続する「する」は、サ変動詞の活用部として扱っているが、対話データベース形態素解析では、補助動詞としてい

¹³ x を変数とした時、活用語では $v-sa-x$ のように、 x では活用形を扱っている。

るので、3.2.2 節で扱う。なお、カテゴリー名は、一部、参考文献 [15] を基にしている。

一段補助動詞

一段補助動詞 例文 148 の「い」のように使う、一段活用をする補助動詞を扱っている。「いる」は、活用した形を語彙登録している。

例文 148 会議に参加しています。

(<aspc-renyo> <--> (i))
(<aspc-rentai> <--> (i r u))

可能補助動詞 例文 149 の「できる」や例文 150 の「いただける」など、可能の意味を含む補助動詞は、可能動詞の「できる」や「いただける」と同じ活用をする¹⁴。また、例文 151 や例文 152 から分かるように、助動詞「れる」や「せる」とは共起しない。

例文 149 参加できますか？

例文 150 御参加いただけると、嬉しいのですが。

例文 151 * 参加できられますか？

例文 152 * 参加できさせますか？

可能補助動詞は、例文 153 のように、「て」に後続するもの (dont-v-deki) と「て」に後続しないもの (v-deki) とを分けている。「いただける」は、活用語に助詞の「て」がつくものに後続して例文 153 のような文節 / 文を作ることがある。一方、「できる」は、例文 154 から分かるように、補助動詞としては「て」に後続できない。例文 154 は、例文 155 の「自分の仕事」が省略されたものと解釈できるが、その場合は、「できる」は本動詞である。

例文 153 参加していただけますか？

例文 154 * 参加してできますか？

例文 155 * (その会議に) 参加して自分の仕事ができますか？

(<dont-v-deki> <--> (i t a d a k e))
(<v-deki> <--> (d e k i))

¹⁴3.2.1 節、63 ページ参照。

気を付けて 12 対話データベース形態素解析では、サ変名詞に補助動詞の「できる」や「いただける」が後続したものは、サ変名詞と活用した補助動詞という二つの形態素として扱っている。一方、*dp10-re72.gra* では、サ変名詞に補助動詞の「できる」などの語幹が後続した時点で、他の可能動詞 (deki-v-1dan) と同じ扱いをしている。

```
(<deki-v-1dan> <--> (<v-sahen> <v-deki>))
(<deki-v-1dan> <--> (<pre-v-sahen> <v-deki>))
(<deki-v-1dan> <--> (<v-sahen> <dont-v-deki>))
(<deki-v-1dan> <--> (<pre-v-sahen> <dont-v-deki>))
```

五段補助動詞

補助動詞で、五段活用するものは、否定の助動詞「ない」(nai) を後続させた時の、「あ」(a) の直前の子音により分類している。現在、「いただかない」(itada k anai) のように 'k' ('ki') 型のもの、「くださらない」(kudasa r anai) のように 'r' 型のもの、「いたさない」(ita s anai) のように 's' 型のものの三種類を扱っている。「なる」や「なさる」は、発音だけを考慮した場合、「くださる」のように 'r' 型に属する。このうち、「なる」は例文 156 のように「になる」と使われることが多い。また「下さる」は例文 157 のように、「て」に後続して使われるが、「なさる」は例文 158 のように「て」に後続して使われることはない。このため、「なる」と「なさる」は別に扱っている。

例文 156 案内書は御覧になりましたか？

例文 157 申込書に記入して下さいますか？

例文 158 * 申込書に記入してなさいますか？

```
(<dont-5dan-ki> <--> (i t a d a))
(<dont-5dan-r> <--> (k u d a s a))
(<dont-5dan-s> <--> (i t a))
(<dont-naru> <--> (n a))
(<dont-5dan-r-te> <--> (n a s a))
```

これらの語幹には、以下のように活用語尾が後続する。「なる」や「なさる」も「いただく」などの組み立て規則に準拠するので、一例ずつ示す。対話データベース形態素解析では、このように活用語尾がついたものを一つの形態素として扱っている。

(<dont-5dan-a> <--> (<dont-5dan-ki> <flex-d-5-ka>))
 (<dont-5dan-i> <--> (<dont-5dan-ki> <flex-d-5-ki>))
 (<dont-5dan-u> <--> (<dont-5dan-ki> <flex-d-5-ku>))
 (<dont-5dan-e> <--> (<dont-5dan-ki> <flex-d-5-ke>))
 (<dont-5dan-o> <--> (<dont-5dan-ki> <flex-d-5-ko>))
 (<dont-renyo2-t> <--> (<dont-5dan-ki> <flex-d-r2-t-ki>))
 (<dont-naru-a> <--> (<dont-naru> <flex-d-n-a>))
 (<dont-5dan-a-te> <--> (<dont-5dan-r-te> <flex-d-5-ra>))

「くださる」や「なさる」の命令形「くだされ」や「なされ」は、現在ほとんど使われない。日常会話では、「ください」や「なさい」が主に使われている。この二つの例に関しては、命令形に活用した形で以下のように語彙登録している。

(<dont-5dan-e-kougo> <--> (k u d a s a i))
 (<dont-5dan-e-kougo-te> <--> (n a s a i))

また、「おる」と「ある」は、活用した形で語彙登録している。

(<aspc-pej-mizen> <--> (o r a))
 (<aspc-pej-renyo> <--> (o r i))
 (<aspc-pej-rentai> <--> (o r u))
 (<aspc-renyo> <--> (a r i))
 (<aspc-rentai> <--> (a r u))

「する」

サ変名詞に後続する「する」は、サ変動詞の一部であるという立場から、活用部として扱っている。一方、対話データベース形態素解析では、サ変名詞に後続する「する」は独立した形態素、補助動詞、として扱っている。

(<flex-sahen-mizen1> <--> (sh i))
 (<flex-sahen-mizen2> <--> (sh i y o))
 (<flex-sahen-mizen3-4> <--> (s a))
 (<flex-sahen-mizen5> <--> (s e))
 (<flex-sahen-renyo1> <--> (sh i))
 (<flex-sahen-rentai> <--> (s u r u))
 (<flex-sahen-katei> <--> (s u r e))
 (<flex-sahen-meirei> <--> (sh i r o))

参考までに、サ変名詞に「する」の活用した形が後続するものは、以下のように組み立てられる。

(<v-mizen1> <--> (<v-sahen> <flex-sahen-mizen1>))
 (<v-rentai> <--> (<v-sahen> <flex-sahen-rentai>))
 (<v-rentai> <--> (<pre-v-sahen> <flex-sahen-rentai>))

気を付けて 13 *v-mizen1*, *v-rentai* は、歴史的経緯から、下一段又は上一段動詞が活用した形に与えられているカテゴリ名と同じなので、注意を要する。サ変名詞に「する」が後続した *v-rentai* は、対話データベース形態素解析では、二つの形態素に対応するが、以下のように組み立てられる一段動詞の *v-rentai* は、一つの形態素に対応する。

(*<v-rentai>* <--> (*<v-1dan>* *<flex-1dan-rentai2>*))

さらに、*pre-v-sahen* では、サ変名詞に接頭辞がついたものだけでなく、本動詞の連用形に接頭辞がついたものも扱っていることにも注意しなければならない。

気を付けて 14 *x* を変数とした時、*flex-x* は、基本的に活用語の活用部である。しかし、*flex-sahen-x* は、対話データベース形態素解析の補助動詞に対応する。

気を付けて 15 「する」は例文 159 の「買いはしたが」のように、活用語の連用形に係助詞の「は」や「も」が後続したものにさらに続くことがある。

例文 159 この本を買いはしたが、難しくてとても読めない。

対話データベース形態素解析では、この「買いはしたが」は一文節として扱っている。

活用語の連用形には、「は」や「も」の他にも以下のように、「に」「が」「を」などの助詞が後続することがある。

例文 160 ゲーテの本を買いに行きましょう。

例文 161 この字の読みがわかりません。

例文 162 この字の読みを教えてください。

上記の「買いに」「読みが」「読みを」は、対話データベース形態素解析ではそれぞれ一文節として扱っている。一方、*dp10-re72.gra* では、活用語の連用形に後続する助詞に属するものとして格助詞の他に、「は」や「も」も同様に扱っている。従って、「買いはしたが」は「買いは」と「したが」という二文節からなるとしている。このため、「したが」に含まれる「する」は、本動詞として扱っている。

「する」は、活用したものも含めて、音声認識の妨げになりやすいので、文節を基に文を組み立てる規則では、他の動詞を含むもの (*vaux*) とは区別して扱っている (*vaux-sa*)。

(*<vaux>* <--> (*<v-renyo1>* *<masu-rentai>*))
(*<vaux-sa>* <--> (*<v-sa-renyo1>* *<masu-rentai>*))

形容詞型補助動詞

形容詞と同じ活用をする補助動詞として、「ほしい」を扱っている。対話データベース形態素解析では、「わかってほしい」の「ほしい」は、「わかっている」のように使われる補助動詞の「いる」と同じタイプだと考え、補助動詞にしている¹⁵。dp10-re72.graでも、「ほしい」は例外的に扱っているので、語幹と活用形を分けずに、活用したものを語彙登録している。

```
(<exp-mizen> <--> (h o s h i k a r o))  
(<exp-renyo1> <--> (h o s h i k u))  
(<exp-renyo2> <--> (h o s h i k a q))  
(<exp-rentai> <--> (h o s h i i))  
(<exp-katei> <--> (h o s h i k e r e))
```

3.3 形容詞

形容詞は、終止形の最後から二番目の音素により分類している。「詳しい」(kuwash i i)は‘i’のグループ(kyo)に属し、「近い」(chik a i)は‘a’のグループ(kyo-a)、「良い」(y o i)は‘o’のグループ(kyo-yoi)に属する。これは、‘i’のグループの「よろしい」が連用形をとった時は、(yoroshi u)のように‘u’をとるが、‘a’のグループの「ありがたい」は、(aringat ou)のように‘ou’をとるためである¹⁶。また、助動詞「そうだ」が後続する時のかたちは別に扱っている(kyo-pres)。さらに、「わかりやすい」「わかりにくい」のように使われる接尾辞「やすい」や「にくい」¹⁷も、活用する段階では、形容詞として扱っている。

kyo 終止形の活用語尾「い」の直前が「い」になるもの。よろしい(yoroshi)、くわしい(kuwashi)。

kyo-a 終止形の活用語尾「い」の直前が「あ」になるもの。ありがたい(aringat)、ない(n)、ちかい(chik)。

kyo-yoi 終止形の活用語尾「い」の直前が「お」になるもの。よい(yo)。

kyo-ii 「よい」と同意語の「いい」。「これはよい本だ」と同じように「これはいい本だ」ということもできる。しかし「これはよくない」と同じように、「これはいくない」とはいわない。「いい」は連体形のみを扱っている。いい(i)。

¹⁵参考文献[14]、32ページ参照。

¹⁶‘i’のグループと‘o’のグループの活用は同じだと思われる

¹⁷3.12節、80ページ参照。

kyo-pres 助動詞「そうだ」を伴う時のかたち。「そうだ」は形容詞の語幹につくが、「よそうだ」とか「なそうだ」とは使わない。その代わりに、他の語幹「よさ」や「なさ」が使われ、「よさそうだ」または「なさそうだ」となる。

```
(<kyo> <--> (y o r o sh i))
(<kyo-a> <--> (a r i ng a t))
(<kyo-yoi> <--> (y o))
(<kyo-ii> <--> (i))
(<kyo-pres> <--> (n a s a))
```

対話データベース形態素解析では、形容詞の活用したものを一つの形態素として扱っている。dp10-re72.gra では、kyo-pres 以外で、対応するのは、規則により生成される以下のようなカテゴリーであろう。

```
(<kyo-renyo1> <--> (<kyo-a> <flex-kyo-a-renyo1>))
(<kyo-renyo2> <--> (<kyo-a> <flex-kyo-a-renyo2>))
(<kyo-renyo3> <--> (<kyo-a> <flex-kyo-a-renyo3>))
(<kyo-rentai> <--> (<kyo-a> <flex-kyo-a-rentai>))
(<kyo-katei> <--> (<kyo-a> <flex-kyo-a-katei>))
```

3.4 副詞

副詞は、後続する助詞により分類している。助詞を伴わないもの (adv-b)、係り助詞を伴いやすいもの (adv-k-s, adv-k-d)、格助詞の「に」を伴うもの (adv-ni) と数量関係の副助詞を伴うもの (adv-num) がある。

adv-b 助詞を伴わずに単独で使われることが多い副詞。まだ、ぜんぜん。

adv-k-s 係助詞の「は」や「も」を伴いやすい副詞。そう。

adv-k-d 係助詞の「は」は伴いにくい、「でも」¹⁸を伴いやすい副詞。さらに格助詞の「に」に係助詞の「も」や「でも」が後続するものを伴いやすいもの。どう。

adv-ni 格助詞の「に」が後続してもしなくても副詞的に働く副詞。ここでは、「に」以外の助詞の後続は認めていない。

adv-num 副助詞の「ずつ」「だけ」「ばかり」「くらい」などが後続しやすい副詞。ちょっと。

¹⁸対話データベース形態素解析では、副助詞、または格助詞「で」に係助詞「も」が後続するものとして扱っている。

以上の副詞は、それぞれ助詞を伴ったり、助詞を伴わずに単独で使われたりする。単独で使われる場合は、advとしてまとめている。対話データベース形態素解析の副詞には、adv-b, adv-k-sなどの個々のカテゴリー、または、このadvが対応するであろう。

```
(<adv-b> <--> (m a d a))
(<adv-k-s> <--> (s ou))
(<adv-k-d> <--> (d ou))
(<adv-ni> <--> (sh i ky uu))
(<adv-num> <--> (ch o q t o))
```

気を付けて16「至急」は、対話データベース形態素解析では、普通名詞としている。普通名詞は、“格助詞「が」を付けて文の主語になる”ものと定義されている¹⁹。しかし、「至急が」とは、あまり使わない。また、「至急に」と副詞的に使われやすいため、dp10-re72.graでは副詞として扱っている。

3.5 連体詞

rentai 連体詞を扱っている。どの、その。

モデル会話で使われている表現を考慮している。

```
(<rentai> <--> (d o n o))
```

3.6 接続詞

conj 接続詞を扱っている。また、それでは、ところで。

```
(<conj> <--> (m a t a))
```

3.7 間投詞

音声認識部では扱っていない。

3.8 感動詞

interj 感動詞を扱っている。感動詞は、それだけで一文を構成するとしている²⁰。はい、もしもし、ありがとう。

¹⁹参考文献 [14]、1 ページ参照。

²⁰2.1 節、14 ページ参照。

対話データベース形態素解析では、「独立性がある」ということを感動詞の認定基準としている。「ありがとう」には独立性があり、感動詞であるため、例文 163 は以下のように解析している。

例文 163 ありがとうございます。

データベース	
ありがとう	ございます
感動詞	補助動詞

対話データベース形態素解析を考慮して、「感動詞に補助動詞が後続して文節が形成される」という規則を与えると、例文 164 や例文 165 のような文節が生成される。

例文 164 * はいございます。

例文 165 * もしもしございます。

例文 164 や例文 165 のような文節の生成を防ぐため、*dp10-re72.gra* では、感動詞は、単独で使われるという規則を与えている²¹。また、例文 163 のように使われる「ありがとう」は形容詞の連用形として扱っている。形容詞の連用形は、助動詞、補助動詞などを伴って文節を作るという規則により、感動詞との二重の生成は避けている。

HMM-LR	
ありがとう	
interj	

HMM-LR	
ありがとう	ございます
kyo-renyo3	polt-v-rentai

(<interj> <--> (h a i))

3.9 助動詞

助動詞は、働きごとに分類し、それをさらに単語および活用形により分類している。同じ活用形に属しても、発音の異なるものは、1, 2 などの

²¹例文 164 や例文 165 のような文節の生成を防ぐには、感動詞の細分化も考えられる。

番号により分けている。助動詞の活用形は、すべて扱っているわけではなく、まだ不完全である。

「です」「ようだ」「そうだ」「た」は、語彙登録が重複している。以下、これらの語彙に説明を加える。

です 「です」には、例文 166 のように名詞と共に使うもの (v-cop) と例文 167 のように丁寧さを添えるために形容詞と共に使うもの (polt-aux) がある。

例文 166 この本は三百円です。

例文 167 この本は安いです。

ようだ 「ようだ」には、例文 168 や例文 169 のように、比況の助動詞として扱っているもの (meta) と、例文 170 のように、推量の助動詞として扱っているもの (conj) がある。なお、比況の助動詞としては、例文 168 や例文 169 のように連体修飾句、または連用修飾句として使われる、「ような」や「ように」を扱っている。

例文 168 どのようなご用件でしょうか？

例文 169 参加料はどのようにお支払いしたらよいのですか？

例文 170 風邪で熱があるようです。

そうだ 「そうだ」には、例文 171 のように、伝聞の助動詞として扱っているもの (evid) と、例文 172 のように、様態の助動詞として扱っているもの (pres) がある。

例文 171 もうすぐレポートを書き終るそうだ。

例文 172 もうすぐレポートを書き終りそうだ。

た 「た」には、例文 173 のように本動詞に直接後続するもの (past-ta) と、例文 174 のように、助動詞に後続するもの (ta) がある。

例文 173 書いた。

例文 174 書きました。

助動詞の連用形は、例文 175 のように「た」を伴うことがある。また、例文 176 のように連用形を使い、並列の関係を表すこともある。この場合、連用形は文節の終りに使われるので、他の連用形と区別している (*x*-coord)。

例文 175 レポートが書き終りそうだった。

例文 176 レポートは書き終りそうで、なかなか書き終らない。

(*<pres-renyo2>* *<-->* (*s ou d a q*))
(*<pres-renyo-coord>* *<-->* (*s ou d e*))

活用語の連体形と終止形は、別に語彙登録する予定である。別登録に関しては、現在、まだ整理中であるが、例文 177 の「そうだ」と例文 178 の「そうな」のように終止形と連体形 (*x*-mod) が異なるものに関しては暫定的なカテゴリーを作っている。

例文 177 今にも木の枝が折れそうだ。

例文 178 今にも折れそうな枝だが、まだ折れずにいる。

(*<pres-rentai>* *<-->* (*s ou d a*))
(*<pres-rentai-mod>* *<-->* (*s ou n a*))

mizen, *renyo*, *rentai* などの活用形を *x* とした時の、助動詞のカテゴリー²²及びそのカテゴリーで扱っている助動詞の対応を示す。

²² カテゴリー名は、一部、参考文献 [15] によっている。

intn	う
v-cop- <i>x</i>	です
cop- <i>x</i>	だ
meta- <i>x</i>	よう(だ)
polt-aux- <i>x</i>	です
evid- <i>x</i>	らしい, そうだ, そうです
pres- <i>x</i>	そうだ, そうです
conj- <i>x</i>	ようだ, ようです
negt- <i>x</i>	ない
nai-n- <i>x</i>	ん
past-ta	た
past-da	だ
ta- <i>x</i>	た
masu- <i>x</i>	ます
caus-seru- <i>x</i>	せる
caus-saseru- <i>x</i>	させる
deac-reru- <i>x</i>	れる
deac-rareru- <i>x</i>	られる
optt- <i>x</i>	たい

3.10 助詞

助詞は、対話データベース形態素解析の分類を基に、細分化している。これは、音声認識率をあげるために、詳細な助詞の接続規則を構築しなければならないからである。さらに詳細な接続規則が必要であるが、助詞は、数に限りがあるので、それぞれにカテゴリー名をつけることも考えられる。

3.10 節では、対話データベース形態素解析で定義している助詞、格助詞、準体助詞、係助詞、副助詞、並列助詞、接続助詞、終助詞の順に、*dp10-re72.gra* で用意している助詞のカテゴリーと、その語彙登録の対応を試みる。

3.10.1 格助詞

p-kaku-ga	が
p-kaku-o	を
p-kaku-ni	に
p-kaku-de	で
p-kaku-e	へ
p-kaku3	の
p-kaku4	と
p-kaku5	から、まで
p-kaku7	より
p-kaku10	として、について、において、にとって、にて
p-kaku0	における、にせよ、にしる、でもって、をもって
p-i-6	と

気を付けて 17 複合格助詞の定義は明確でないが、*dp10-re72.gra* では「に対する」は、複合格助詞 *p-special* として一つの形態素の扱いをしている。一方、対話データベース形態素解析では、格助詞「に」と本動詞の活用形「対する」の二つの形態素として扱っている²³。

3.10.2 準体助詞

p-j	の
p-j-n	ん

3.10.3 係助詞

p-k-w	は
p-k-m	も

3.10.4 副助詞

p-f-1	ほど、のみ、ごと、さえ、など、なんか
p-f-num	だけ、ばかり、くらい、ぐらい
p-f-0	でも、しか、こそ、なら、なんて
p-f-z	ずつ
p-f-2	か

²³ 3.13節、83 ページ参照。

3.10.5 並列助詞

p-h-1	と
p-h-2	や
p-h-toka	とか
p-h-yara	やら
p-h-nari	なり

3.10.6 接続助詞

接続助詞は、前接する活用語の活用形により、分類している。また、連用形に後続する接続助詞は、音が濁るものもあるので、濁音 ($x-d$) か清音 ($x-t$) かでさらに分類している。

活用語の連体形と終止形は、別に語彙登録する予定である。しかし、まだ連体形と終止形が同じもので、連体形だけ語彙登録しているものがある。そこで、連体形だけに接続するもの ($x-only$) と、終止形と連体形のどちらにでも接続するもの ($x-both$) とに分けている。

我々は、話し言葉を対象にしているため、文節に基づき文を組み立てる規則、バージョン 1.1 では、接続助詞終止の文も認めている²⁴。すべての接続助詞に対して、接続助詞終止を許すと、音声認識の誤認識の対象が増えるので、対話データベースの検索に基づき、文末で使われやすい接続助詞 ($x-f$) を別に扱っている。

形容詞の連用形に後続する接続助詞は、歴史的経緯から、他の活用語に後続する接続助詞と別に扱っている。

²⁴第2章、12 ページ参照。

sp-rentai-both	けれども、けれど、けども、けど、と、から、し
sp-rentai-both+f	が
sp-rentai-only	ので、んで、のに
sp-rentai-katei	なら
sp-renyo-t	て
sp-renyo-t-k	ちゃ
sp-renyo-t-2	たら
sp-renyo-t-h	たり
sp-renyo-d	で
sp-renyo-d-k	じゃ
sp-renyo-d-2	だら
sp-renyo-d-h	だり
sp-katei	ば
sp-adj-renyo1	て
sp-adj-renyo1-k	ちゃ
sp-adj-renyo2	たら
sp-adj-renyo2-h	たり
te-te	て
te-de	で

3.10.7 終助詞

sfp-1	か
sfp-2	よ
sfp-3	ね

3.11 接頭辞

dp10-re72.gra では、接頭辞は、形態素として独立に扱っていない。サ変名詞や普通名詞の一部として語彙登録している²⁵。

3.12 接尾辞

接尾辞は、モデル会話で扱っているものに、その都度対処している。

金額を表す数詞に後続する単位は、数詞の発音により分けている。「四

²⁵3.1.2 節、52 ページ、および 3.1.4 節、54 ページ参照。

円」の「四」は「ヨ」と発音されるが、マルクやドルなどの単位と共に使われる場合は、「ヨン」と発音されるためである。

wh-suf 「何人」の「ニン」のように wh-pro-N に後続するもの。

count-human-suf 「三人」「四人」のように、人を数える時に使う接尾辞。

pro-suf1 「私ども」の「ドモ」のように「わたし」や「わたくし」に後続してさらに代名詞を形成するもの。

name-suf 「鈴木真由美様」の「サマ」のように、固有名詞または代名詞などに後続するもの。 *dp10-re72.gra* では、固有名詞に後続するものだけを扱っている。

suf-ippai 「来月いっぱい」の「イッパイ」のように期限をあらわす名詞に後続するもの。

suf-kyo 「この本は読みにくい」や「もっと理解しやすく書いてほしい」の「にくい」や「やすい」のように動詞の連用形や、サ変名詞に連用形の活用語尾がついたものに後続して使われるもの。

num-suf-ch 「一丁目一番三号です」のように住所に使われる「丁目」²⁶。

num-suf-b 「一丁目一番三号です」のように住所に使われる「番」。

num-suf-g 「一丁目一番三号です」のように住所に使われる「号」。

money-suf 「五千円です」の「円」のように数詞の後ろに使われる単位。
円。

money-suf-n 「五千ドルです」の「ドル」のように数詞の後ろに使われる単位。ドル、マルク。

²⁶ atr-dp 上の対話データベース形態素解析では、「一丁目」の「丁目」を二つの接尾辞「丁」と「目」からなるものとしている。参考文献 [14]、13 ページでは、「丁目」は一形態素として扱っている。

(<wh-suf> <--> (n i =))
 (<count-human-suf> <--> (n i =))
 (<pro-suf1> <--> (d o m o))
 (<name-suf> <--> (s a m a))
 (<suf-ippai> <--> (i q p a i))
 (<suf-kyo> <--> (n i k u))
 (<num-suf-ch> <--> (ch ou m e))
 (<num-suf-b> <--> (b a =))
 (<num-suf-g> <--> (g ou))
 (<money-suf> <--> (e =))
 (<money-suf-n> <--> (d o r u))

気を付けて 18 「京都府」「相楽郡」などの県名、地名などは、対話データベース形態素解析では固有名詞として一つの形態素として扱っている。一方、*dp10-re72.gra* では、「府」や「郡」に当たる部分は接尾辞として扱っている。

(<ken-name-suf-to> <--> (t o))
 (<ken-name-suf-dou> <--> (d ou))
 (<ken-name-suf-hu> <--> (h u))
 (<ken-name-suf-ken> <--> (k e =))
 (<si-name-suf-ku> <--> (k u))
 (<si-name-suf-si> <--> (sh i))
 (<ku-name-suf-ku> <--> (k u))

気を付けて 19 「わかりやすい」や「わかりにくい」などで使われている「やすい」や「にくい」は、対話データベース形態素解析では、本来の意味を失い、動詞に添えられているものなので、接尾辞として扱っている。*dp10-re72.gra* では、語彙登録では接尾辞にしているが、その後、動詞の連用形(v-n)に後続して(kyo-s)からは、形容詞の活用をそのまま受け継いでいるので、形容詞扱いをしている。以下に、規則構成を示す。

(<suf-kyo> <--> (y a s u))
 (<kyo-s> <--> (<v-n> <suf-kyo>))
 (<kyo-renyo1> <--> (<kyo-s> <flex-kyo-renyo1>))

従って、例文 179 の「使いやすくて」という文節は、対話データベース形態素解析と *dp10-re72.gra* では、扱いが異なる。

例文 179 スイスアーミーナイフは使いやすくて便利だ。

データベース		
使い	やすく	て
本動詞	接尾辞	接続助詞

HMM-LR		
((使いやす)	く))
		て

これからの課題 28 対話データベース形態素解析では、「円」や「ドル」が数詞に後続せず、単独で用いられる場合は、これらを普通名詞として扱っている。dp10-re72.graでは、単独使用は考慮していないので、例文 180 のような文は扱えない。

例文 180 御支払いは円をお願いします。

3.13 例外規則

文節内の規則は、人工的に作成されたモデル会話 AB, 1-5 さらに 6-10 の一部を扱えるように構築している。このモデル会話は、人手により文節切りが行なわれており、音声入力者は、それに従い発声している。この人手による文節切りと、対話データベース形態素解析の指針に従い行なった文節切りと一部ずれが生じている。文節内の規則では、このずれを吸収するため、以下のような例外規則を設けている。

どうすれば (会話 B)

対話データベース形態素解析では、「どうすればよろしいですか。」を三文節からなる文として扱っている。一方、音声入力では、「どうすれば」を一文節にしているので、この文は二文節からなる文として扱っている。

データベース	
どう すれば	よろしいですか
副詞 本動詞 / 接続助詞	

HMM-LR	
どうすれば	よろしいですか
adv-s1	vk-s1

問題ありません (会話 5)

対話データベース形態素解析では、「問題ありません。」は、普通名詞の「問題」と本動詞の「ある」を含む二文節からなるとしている。一方、音声入力では、一文節として扱われている。

データベース	
問題	ありません
普通名詞 本動詞	

問題ありません

mondai

どういう (会話 7)

「どうすれば」と同様、「どういう」も、対話データベース形態素解析では、副詞の「どう」と本動詞の「いう」という二つの文節として扱っている。一方、音声入力としては、一文節として扱っている。

どう いう (内容のものなんですか)

副詞 本動詞

どういう (内容のものなんですか)

adv-s2

に対する (会話 5)

複合格助詞として何を認めるかという基準は明確ではない。対話データベース形態素解析では、「に対する」は名詞性の自立語に後続する格助詞「に」と本動詞の活用形という二つの形態素からなるという扱いをしている²⁷。一方、音声入力としては、複合格助詞として一つの形態素として扱っている。

取り消しに 対する

普通名詞 / 格助詞 本動詞

取り消しに対する

n-hutu/ p-special

(<adv-s1> <--> (d o u s u r e b a))

(<vk-s1> <--> (y o r o s h i i d e s u k a))

(<adv-s2> <--> (d o u i u))

(<mondai> <--> (m o = d a i a r i m a s e =))

(<p-special> <--> (n i t a i s u r u))

²⁷参考文献[14]、59 ページ参照。

第 4 章

おわりに

4.1 構文情報の音声認識への応用

本稿で報告した構文規則のおよそのサイズとパープレキシティを表 4.1 に示す [16]。文節内文法の終端規則数は語彙数を表し、文節を組立てる文法の終端規則数は異なる文節の種類数を表している。

表 4.1: 文法のサイズとパープレキシティ

	文節内文法	組立て文法
規則数	1,973	471
終端規則数	744	133
パープレキシティ	3.57 / 音韻	99.7 / 文節

文節内規則は、対話データベースに収集されているキーボード会話で使われている 7405 文節の 96 % を解析できることが確認されている¹。

文節に基づき文を組み立てる規則、バージョン 1.1 がどのくらい文を容認するか、また、音声認識の際に、フィルターとしてどの程度の効果があるか調べた。その際、文節内は、あらかじめ HMM-LR 音声認識を行なった。実験対象を表 4.2 に示す。

文節を基に文を組み立てる規則に関しては、バージョン 1.0 が 80 % 以上のものをアクセプトしている²。それに対し、バージョン 1.1 では、図 4.1 から分かるように³、文節数 1 から 3 の場合約 40 % のものをアクセプトするのに留まっている。文節数 4 から 7 の場合には、制約としての効果がさ

¹測定には、ATR で開発されたパーザが使われている [17]。

²第 2 章、12 ページ参照。

³文節内の音声認識には 2 種類の HMM 音素モデルを使った。川端モデル (B-model) と呼ばれているものと、花沢モデル (P-model) と呼ばれているものである。

表 4.2: 実験対象

総文数	137
総文節数	353
1文中の平均文節数	2.6
1文中の最高文節数	8

らに発揮されている。文節数 8 の文は、実験対象としたモデル会話 137 文の中に一文しかなかったため、この値は不安定である [18]。

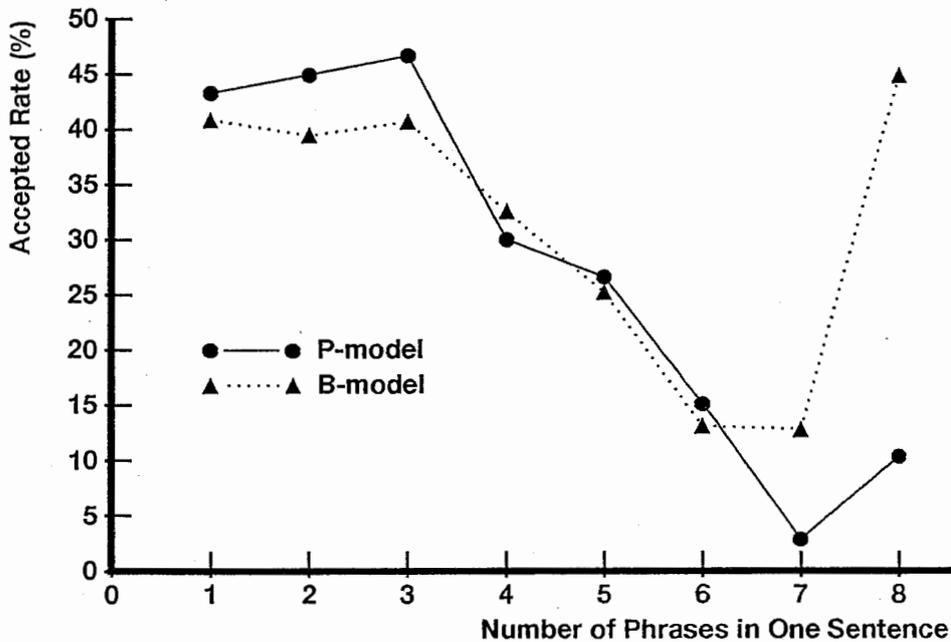


図 4.1: 構文規則の制約としての強さ

文節内規則を応用した HMM-LR[9] による音声認識実験結果⁴を表 4.3 に示す。

また、文節を文に組み上げる規則をフィルターとして利用した [18] 実験結果を表 4.4 に示す。

⁴文節内の音声認識には川端モデルと呼ばれている HMM 音素モデルを使った。また、実験条件は、グローバルビーム幅 100、ローカルビーム幅 10 とした。

表 4.3: 文節認識実験結果

Rank	累積認識率 (%)
1	68.3
2	87.8
3	91.8
4	94.1
5	95.5

表 4.4: 文認識実験結果

Rank	累積認識率 (%)	
	構文規則なし	構文規則あり
1	37.2	70.1
2	60.6	76.6
3	68.6	83.9
4	70.1	83.9
5	73.7	83.9

4.2 展望

現実の言語運用を反映させるため、対話データベースの利用を試みた。しかし、直感による仮定に基づき構築した規則もまだ多数ある。構文規則に客観性を与えるため、今後、対話データベース検索の成果を、さらに取り入れるべきであろう。

我々が対象としているのは、話しことばである。しかし、今まで、話しことばに関する研究はほとんど行なわれていなかった。これは、実データが十分になかったためである。ATR には、話しことばが使われている電話会話と、書きことばに近いキーボード会話とがある。これらを基に、話しことばの特徴を明らかにすることが期待できる。

さらに、その成果を音声認識に有効に働くよう構文規則に取り入れる予定である。

そのためにも、構文規則開発ツールとともに、対話データベース検索システムの向上が切に望まれる。

付録 A

HMM-LR 文法記述用音素対応表

1. “f u”, “j i” 等、この対応表にある表記以外を用いてはいけない。
2. 「が」行の表記は、原則的には発話の最初では “g”, 発話の途中では “ng” (鼻音化) を用いることとする。
3. 長母音の表記法は確定してない。

あ	a	い	i	う	u	え	e	お	o
か	ka	き	ki	く	ku	け	ke	こ	ko
さ	sa	し	shi	す	su	せ	se	そ	so
た	ta	ち	chi	つ	tsu	て	te	と	to
な	na	に	ni	ぬ	nu	ね	ne	の	no
は	ha	ひ	hi	ふ	hu	へ	he	ほ	ho
ま	ma	み	mi	む	mu	め	me	も	mo
や	ya			ゆ	yu			よ	yo
ら	ra	り	ri	る	ru	れ	re	ろ	ro
わ	wa	ゐ	i			ゑ	e	を	o
ん	=								
っ	Q								
が	ga nga	ぎ	gi ngi	ぐ	gu ngu	げ	ge nge	ご	go ngo
ざ	za	じ	ji	ず	zu	ぜ	ze	ぞ	zo
だ	da	ぢ	ji	づ	zu	で	de	ど	do
ば	ba	び	bi	ぶ	bu	べ	be	ぼ	bo
ぱ	pa	ぴ	pi	ぷ	pu	ぺ	pe	ぽ	po
きゃ	ky a			きゅ	ky u	きえ	ky e	きょ	ky o
しゃ	sh a			しゅ	sh u	しえ	sh e	しよ	sh o
ちゃ	ch a			ちゅ	ch u	ちえ	ch e	ちよ	ch o
にゃ	ny a			にゅ	ny u	にえ	ny e	にょ	ny o
ひゃ	hy a			ひゅ	hy u	ひえ	hy e	ひょ	hy o
みゃ	my a			みゅ	my u	みえ	my e	みょ	my o
りゃ	ry a			りゅ	ry u	りえ	ry e	りょ	ry o
ぎゃ	gy a ngy a			ぎゅ	gy u ngy u	ぎえ	gy e ngy e	ぎょ	gy o ngy o
じゃ	zy a			じゅ	zy u	じえ	zy e	じょ	zy o
ぢゃ	zy a			ぢゅ	zy u	ぢえ	zy e	ぢょ	zy o
びゃ	by a			びゅ	by u	びえ	by e	びょ	by o
ぴゃ	py a			ぴゅ	py u	ぴえ	py e	ぴょ	py o
ああ	aa	い	ii	う	uu	え	ei ee	お	ou oo

付録 B

文法

B.1 バージョン 1.0

;;; ***** BUNSETU-BASED SYNTACTIC RULES Version 1.0 *****

```
(<start> <--> (<_start>))
(<_start> <--> (q1 <SS> q2))

(<SS> <--> (<INTERJ1>))
(<SS> <--> (<CONJ1> <NVS>))
(<SS> <--> (<NVS>))
    (<NVS> <--> (<VS>))
        (<VS> <--> (<VC>))
        (<VS> <--> (<ADVS> <VS>))
    (<NVS> <--> (<NnVS>))
        (<NnVS> <--> (<M-NN> <VS>))
        (<NnVS> <--> (<ADVS> <NnVS>))
        (<NnVS> <--> (<M-NN> <NnVS>))
    (<NVS> <--> (<ADV-c>))

(<ADVS> <--> (<ADV-s>))
    (<ADV-s> <--> (<ADV1>))
    (<ADV-s> <--> (<ADV1> <ADV-s>))
(<ADVS> <--> (<ADV-c>))
    (<ADV-c> <--> (<VADVS>))
        (<VADVS> <--> (<VADV>))
        (<VADVS> <--> (<ADVS> <VADVS>))

    (<ADV-c> <--> (<NnVADVS>))
        (<NnVADVS> <--> (<M-NN> <VADVS>))
        (<NnVADVS> <--> (<M-NN> <NnVADVS>))
        (<NnVADVS> <--> (<ADVS> <NnVADVS>))

(<M-NN> <--> (<NN>))
(<M-NN> <--> (<MOD-N> <M-NN>))
(<M-NN> <--> (<NQN>))
    (<NQN> <--> (<QN>))
    (<NQN> <--> (<ADVS> <NQN>))
```

(<NQN> <--> (<M-NN> <NQN>))

(<MOD-N> <--> (<NM>))

(<MOD-N> <--> (<NPM>))

(<MOD-N> <--> (<NVM-c>))

 (<NVM-c> <--> (<VMS>))

 (<VMS> <--> (<VM>))

 (<VMS> <--> (<ADVS> <VM>))

 (<NVM-c> <--> (<NnVMS>))

 (<NnVMS> <--> (<M-NN> <VMS>))

 (<NnVMS> <--> (<M-NN> <NnVMS>))

 (<NnVMS> <--> (<ADVS> <NnVMS>))

;bunsetu grouping

(<NN> <--> (<np>))

(<NN> <--> (<wh-np>))

(<QN> <--> (<quote>))

(<ADV1> <--> (<adv>))

(<ADV1> <--> (<adv-p>))

(<ADV1> <--> (<adv-ph>))

(<VADV> <--> (<adj-s>))

(<VADV> <--> (<vaux-coord>))

(<VADV> <--> (<vaux-katei>))

(<VADV> <--> (<vaux-s>))

(<VADV> <--> (<vaux-te>))

(<VADV> <--> (<vaux-h>))

(<VADV> <--> (<vaux-ni>))

(<INTERJ1> <--> (<interj>))

(<CONJ1> <--> (<conj>))

(<VC> <--> (<vaux>))

(<VC> <--> (<vaux-cop>))

(<VC> <--> (<vaux-sfp>))

(<VM> <--> (<vaux-mod>))

(<VM> <--> (<vaux-nom>))

(<NM> <--> (<n-rentai>))

(<NPM> <--> (<np-special>))

;exceptions

(<VC> <--> (<mondai>));mondai arimasen

(<VADV> <--> (<adv-si>));dou sure-ba

B.2 バージョン 1.1

```
;;; ***** BUNSETU-BASED SYNTACTIC RULES Version 1.1 *****

;;; key symbols: N(meisi-youso)
;;;              V(dousi-youso)
;;;              ADV(hukusi-youso)
;;;              M(rentai-syuusyoku-youso)

(<start> <--> (<_start>))
(<_start> <--> (q1 <SS> q2))

(<SS> <--> (<INTERJ1>))
(<SS> <--> (<EXCEP>))
(<SS> <--> (<CONJ1> <NVS>))
(<SS> <--> (<CONJ1> <NVS-DIR-OBJ>))
(<SS> <--> (<NVS>))
    (<NVS> <--> (<VS>))
        (<VS> <--> (<VC>))
        (<VS> <--> (<ADVPH> <VS>))
    (<NVS> <--> (<NnVS>))
        (<NnVS> <--> (<M-NN> <VS>))
        (<NnVS> <--> (<VADV-H-P-K> <VS-SA>))
        (<NnVS> <--> (<ADVPH> <NnVS>))
        (<NnVS> <--> (<M-NN> <NnVS>))
(<SS> <--> (<NVS-DIR-OBJ>))
    (<NVS-DIR-OBJ> <--> (<VS-DIR-OBJ>))
        (<VS-DIR-OBJ> <--> (<DA-FORM>))
        (<VS-DIR-OBJ> <--> (<ADRE-VC>))
        (<VS-DIR-OBJ> <--> (<TEL-VC>))
        (<VS-DIR-OBJ> <--> (<ADVPH> <VS-DIR-OBJ>))
    (<NVS-DIR-OBJ> <--> (<NnVS-DIR-OBJ>))
        (<NnVS-DIR-OBJ> <--> (<MOD-N> <VS-KEI-DIR-OBJ>))
        (<NnVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <VS-DIR-OBJ>))
        (<NnVS-DIR-OBJ> <--> (<ADVPH> <NnVS-DIR-OBJ>))
        (<NnVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <NnVS-DIR-OBJ>))

(<DA-FORM> <--> (<NVC>))
(<DA-FORM> <--> (<MOD-N> <NVC>))
(<ADRE-VC> <--> (<ADRE1> <AVC>))
(<TEL-VC> <--> (<TEL1> <TELVC>))

;(<ADVS> <--> (<ADVPH>))
;(<ADVS> <--> (<M-NN>))
;(<ADVS> <--> (<M-NN-DIR-OBJ>))

(<ADVPH> <--> (<ADV-s>))
    (<ADV-s> <--> (<ADV1>))
    (<ADV-s> <--> (<ADV1> <ADV-s>))
(<ADVPH> <--> (<ADV-c>))
    (<ADV-c> <--> (<VADVS>))
```

(<VADVS> <--> (<VADV>))
 (<VADVS> <--> (<ADV-s> <VADVS>))
 (<ADV-c> <--> (<NnVADVS>))
 (<NnVADVS> <--> (<M-NN> <VADVS>))
 (<NnVADVS> <--> (<M-NN> <VADV-H>))
 (<NnVADVS> <--> (<MOD-N> <VADV-H-O>))
 (<NnVADVS> <--> (<M-NN> <NnVADVS>))
 (<NnVADVS> <--> (<ADV-s> <NnVADVS>))
 (<ADVPH> <--> (<ADV-c-DIR-OBJ>))
 (<ADV-c-DIR-OBJ> <--> (<VADVS-DIR-OBJ>))
 (<VADVS-DIR-OBJ> <--> (<NKVADV>))
 (<VADVS-DIR-OBJ> <--> (<MOD-N> <NKVADV>))
 (<ADV-c-DIR-OBJ> <--> (<NnVADVS-DIR-OBJ>))
 (<NnVADVS-DIR-OBJ> <--> (<MOD-N> <NKVADV-KEI>))
 (<NnVADVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <VADVS-DIR-OBJ>))
 (<NnVADVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <NnVADVS-DIR-OBJ>))
 (<NnVADVS-DIR-OBJ> <--> (<ADV-s> <NnVADVS-DIR-OBJ>))

 (<VADV-H> <--> (<VADV-H-COORD>))
 (<VADV-H> <--> (<VADV-H-P>))
 (<VADV-H> <--> (<VADV-H-P-K> <VADV-SA>))
 (<VADV-H-O> <--> (<VADV-H-P-O>))

 (<M-NN> <--> (<NN>))
 (<M-NN> <--> (<MOD-N> <NN-KEI>))
 (<M-NN> <--> (<MOD-N> <M-NN>))
 (<M-NN> <--> (<NQN>))
 (<NQN> <--> (<QN>))
 (<NQN> <--> (<ADV-s> <NQN>))
 (<NQN> <--> (<M-NN> <NQN>))

 (<M-NN-DIR-OBJ> <--> (<NN-DIR-OBJ>))
 (<M-NN-DIR-OBJ> <--> (<MOD-N> <NN-KEI-DIR-OBJ>))
 (<M-NN-DIR-OBJ> <--> (<MOD-N> <M-NN-DIR-OBJ>))
 (<M-NN-DIR-OBJ> <--> (<NQN-DIR-OBJ>))
 (<NQN-DIR-OBJ> <--> (<NKQN>))
 (<NQN-DIR-OBJ> <--> (<MOD-N> <NKQN>))
 (<NQN-DIR-OBJ> <--> (<MOD-N> <NKQN-KEI>))
 (<NQN-DIR-OBJ> <--> (<ADV-s> <NQN-DIR-OBJ>))
 (<NQN-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <NQN-DIR-OBJ>))

 (<MOD-N> <--> (<NM>))
 (<MOD-N> <--> (<MOD-N> <NM-KEI>))
 (<MOD-N> <--> (<NPM>))
 (<MOD-N> <--> (<NVM-c>))
 (<NVM-c> <--> (<VMS>))
 (<VMS> <--> (<VM>))
 (<VMS> <--> (<ADV-s> <VM>))
 (<NVM-c> <--> (<NnVMS>))
 (<NnVMS> <--> (<M-NN> <VMS>))

(<NnVMS> <--> (<VADV-H-P-K> <VM-SA>))
(<NnVMS> <--> (<M-NN> <NnVMS>))
(<NnVMS> <--> (<ADV-s> <NnVMS>))

(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14>))
(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14> <ADRE15>))
(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14> <ADRE16> <ADRE15>))
(<ADRE1> <--> (<ADRE12> <ADRE13> <ADRE14> <ADRE15> <ADRE15>))
(<ADRE1> <--> (<ADRE11> <ADRE12> <ADRE14> <ADRE16> <ADRE17>))

;bunsetu grouping

(<NN> <--> (<np>))
(<NN> <--> (<NP-MONEY>))
(<NN> <--> (<wh-np>))
(<NN> <--> (<wh-n>))
(<NN> <--> (<np-e>))
(<NN> <--> (<NP-E-MONEY>))
(<NN> <--> (<wh-np-e>))
(<NN> <--> (<np-o>))
(<NN> <--> (<wh-np-o>))
(<NN> <--> (<NP-MONEY-O>))

(<NN-DIR-OBJ> <--> (<np>))
(<NN-DIR-OBJ> <--> (<NP-MONEY>))
(<NN-DIR-OBJ> <--> (<wh-np>))
(<NN-DIR-OBJ> <--> (<wh-n>))

(<NN-KEI> <--> (<np-keisiki>))
(<NN-KEI> <--> (<np-keisiki-e>))
(<NN-KEI> <--> (<np-keisiki-o>))
(<NN-KEI-DIR-OBJ> <--> (<np-keisiki>))

(<QN> <--> (<quote>))
(<NKQN> <--> (<n-quote>))
(<NKQN> <--> (<N-QUOTE-MONEY>))
(<NKQN-KEI> <--> (<n-keisiki-quote>))

(<ADV1> <--> (<adv-p>))
(<ADV1> <--> (<adv-ph>))
(<VADV> <--> (<adj-s>))
(<VADV> <--> (<vaux-katei>))
(<VADV> <--> (<vaux-s>))
(<VADV> <--> (<vaux-s+f>))
(<VADV> <--> (<vaux-te>))
(<VADV> <--> (<vaux-h>))
(<VADV> <--> (<vaux-ni>))
(<VADV> <--> (<vaux-sa-s>))
(<VADV> <--> (<vaux-sa-te>))
(<VADV> <--> (<vaux-sa-h>))
(<VADV> <--> (<vaux-naru-s>))

(<VADV> <--> (<vaux-kdo-coord>))
 (<VADV> <--> (<vaux-naru-s+f>))
 (<VADV> <--> (<vaux-sa-s+f>))
 (<VADV-SA> <--> (<vaux-sa-s>))
 (<VADV-SA> <--> (<vaux-sa-s+f>))
 (<VADV-SA> <--> (<vaux-sa-coord>))
 (<VADV-SA> <--> (<vaux-sa-te>))
 (<VADV-SA> <--> (<vaux-sa-h>))
 (<VADV-H-COORD> <--> (<vaux-coord>))
 (<VADV-H-COORD> <--> (<vaux-sa-coord>))
 (<VADV-H-P> <--> (<n-ren-p>))
 (<VADV-H-P> <--> (<v-no-p>))
 (<VADV-H-P-K> <--> (<n-ren-p-k>))
 (<VADV-H-P-O> <--> (<n-ren-p-o>))
 (<VADV-H-P-O> <--> (<v-no-p-o>))
 (<NKVADV> <--> (<n-vaux-coord>))
 (<NKVADV> <--> (<N-VAUX-COORD-MONEY>))
 (<NKVADV> <--> (<n-vaux-h>))
 (<NKVADV> <--> (<N-VAUX-H-MONEY>))
 (<NKVADV> <--> (<n-vaux-katei>))
 (<NKVADV> <--> (<N-VAUX-KATEI-MONEY>))
 (<NKVADV> <--> (<n-vaux-s>))
 (<NKVADV> <--> (<N-VAUX-S-MONEY>))
 (<NKVADV> <--> (<n-vaux-s+f>))
 (<NKVADV> <--> (<N-VAUX-S+F-MONEY>))
 (<NKVADV-KEI> <--> (<n-keisiki-vaux-coord>))
 (<NKVADV-KEI> <--> (<n-keisiki-vaux-h>))
 (<NKVADV-KEI> <--> (<n-keisiki-vaux-katei>))
 (<NKVADV-KEI> <--> (<n-keisiki-vaux-s>))
 (<NKVADV-KEI> <--> (<n-keisiki-vaux-s+f>))

(<INTERJ1> <--> (<interj>))
 (<CONJ1> <--> (<conj>))

(<VC> <--> (<vaux>))
 (<VC> <--> (<vaux-cop>))
 (<VC> <--> (<vaux-s+f>))
 (<VC> <--> (<vaux-sfp>))
 (<VC> <--> (<vaux-sa>))
 (<VC> <--> (<vaux-sa-s+f>))
 (<VC> <--> (<vaux-naru>))
 (<VC> <--> (<vaux-naru-s+f>))

(<NVC> <--> (<n-vaux>))
 (<NVC> <--> (<N-VAUX-MONEY>))
 (<NVC> <--> (<n-vaux-cop>))
 (<NVC> <--> (<N-VAUX-COP-MONEY>))
 (<NVC> <--> (<n-vaux-s+f>))
 (<NVC> <--> (<N-VAUX-S+F-MONEY>))
 (<NVC> <--> (<n-vaux-sfp>))

(<NVC> <--> (<N-VAUX-SFP-MONEY>))
(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux>))
(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-cop>))
(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-s+f>))
(<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-sfp>))
(<VS-SA> <--> (<vaux-sa>))
(<VS-SA> <--> (<vaux-sa-s+f>))

(<VM> <--> (<nomina>))
(<VM> <--> (<vaux-mod>))
(<VM> <--> (<vaux-nom>))
(<VM> <--> (<nomina-sa>))
(<VM> <--> (<vaux-sa-nom>))
(<VM> <--> (<vaux-sa-mod>))
(<VM-SA> <--> (<nomina-sa>))
(<VM-SA> <--> (<vaux-sa-nom>))
(<VM-SA> <--> (<vaux-sa-mod>))

(<NM> <--> (<n-rentai>))
(<NM> <--> (<np-no>))
(<NM> <--> (<NP-NO-MONEY>))
(<NM> <--> (<wh-np-no>))
(<NM> <--> (<adv-p-no>))
(<NM> <--> (<vn-no>))
(<NM-KEI> <--> (<np-keisiki-no>))
(<NPM> <--> (<np-special>))

;;address

(<AVC> <--> (<adre-vaux-cop>))
(<ADRE11> <--> (<adre-ken>))
(<ADRE12> <--> (<adre-si>))
(<ADRE13> <--> (<adre-ku>))
(<ADRE14> <--> (<adre-tyo>))
(<ADRE15> <--> (<adre-no>))
(<ADRE16> <--> (<adre-num-ch>))
(<ADRE17> <--> (<adre-num-ban>))

;;telephone

(<TELVC> <--> (<tel-vaux-cop>))
(<TEL1> <--> (<tel-no>))

;;money

(<N-VAUX-MONEY> <--> (<N-VAUX-M-MAN>))
(<N-VAUX-MONEY> <--> (<N-VAUX-M-SEN>))
(<N-VAUX-MONEY> <--> (<SEN> <N-VAUX-M-SEN>))
(<N-VAUX-MONEY> <--> (<N-VAUX-M-HYAKU>))
(<N-VAUX-MONEY> <--> (<HYAKU> <N-VAUX-M-HYAKU>))

(<N-VAUX-COP-MONEY> <--> (<N-VAUX-COP-M-MAN>))
(<N-VAUX-COP-MONEY> <--> (<N-VAUX-COP-M-SEN>))

(<N-VAUX-COP-MONEY> <--> (<SEN> <N-VAUX-COP-M-SEN>))
(<N-VAUX-COP-MONEY> <--> (<N-VAUX-COP-M-HYAKU>))
(<N-VAUX-COP-MONEY> <--> (<HYAKU> <N-VAUX-COP-M-HYAKU>))

(<N-VAUX-COORD-MONEY> <--> (<N-VAUX-COORD-M-MAN>))
(<N-VAUX-COORD-MONEY> <--> (<N-VAUX-COORD-M-SEN>))
(<N-VAUX-COORD-MONEY> <--> (<SEN> <N-VAUX-COORD-M-SEN>))
(<N-VAUX-COORD-MONEY> <--> (<N-VAUX-COORD-M-HYAKU>))
(<N-VAUX-COORD-MONEY> <--> (<HYAKU> <N-VAUX-COORD-M-HYAKU>))

(<N-VAUX-KATEI-MONEY> <--> (<N-VAUX-KATEI-M-MAN>))
(<N-VAUX-KATEI-MONEY> <--> (<N-VAUX-KATEI-M-SEN>))
(<N-VAUX-KATEI-MONEY> <--> (<SEN> <N-VAUX-KATEI-M-SEN>))
(<N-VAUX-KATEI-MONEY> <--> (<N-VAUX-KATEI-M-HYAKU>))
(<N-VAUX-KATEI-MONEY> <--> (<HYAKU> <N-VAUX-KATEI-M-HYAKU>))

(<N-VAUX-S-MONEY> <--> (<N-VAUX-S-M-MAN>))
(<N-VAUX-S-MONEY> <--> (<N-VAUX-S-M-SEN>))
(<N-VAUX-S-MONEY> <--> (<SEN> <N-VAUX-S-M-SEN>))
(<N-VAUX-S-MONEY> <--> (<N-VAUX-S-M-HYAKU>))
(<N-VAUX-S-MONEY> <--> (<HYAKU> <N-VAUX-S-M-HYAKU>))

(<NP-MONEY> <--> (<NP-M-MAN>))
(<NP-MONEY> <--> (<NP-M-SEN>))
(<NP-MONEY> <--> (<SEN> <NP-M-SEN>))
(<NP-MONEY> <--> (<NP-M-HYAKU>))
(<NP-MONEY> <--> (<HYAKU> <NP-M-HYAKU>))

(<NP-NO-MONEY> <--> (<NP-NO-M-MAN>))
(<NP-NO-MONEY> <--> (<NP-NO-M-SEN>))
(<NP-NO-MONEY> <--> (<SEN> <NP-NO-M-SEN>))
(<NP-NO-MONEY> <--> (<NP-NO-M-HYAKU>))
(<NP-NO-MONEY> <--> (<HYAKU> <NP-NO-M-HYAKU>))

(<NP-E-MONEY> <--> (<NP-E-M-MAN>))
(<NP-E-MONEY> <--> (<NP-E-M-SEN>))
(<NP-E-MONEY> <--> (<SEN> <NP-E-M-SEN>))
(<NP-E-MONEY> <--> (<NP-E-M-HYAKU>))
(<NP-E-MONEY> <--> (<HYAKU> <NP-E-M-HYAKU>))

(<N-VAUX-H-MONEY> <--> (<N-VAUX-H-M-MAN>))
(<N-VAUX-H-MONEY> <--> (<N-VAUX-H-M-SEN>))
(<N-VAUX-H-MONEY> <--> (<SEN> <N-VAUX-H-M-SEN>))
(<N-VAUX-H-MONEY> <--> (<N-VAUX-H-M-HYAKU>))
(<N-VAUX-H-MONEY> <--> (<HYAKU> <N-VAUX-H-M-HYAKU>))

(<N-QUOTE-MONEY> <--> (<N-QUOTE-M-MAN>))
(<N-QUOTE-MONEY> <--> (<N-QUOTE-M-SEN>))
(<N-QUOTE-MONEY> <--> (<SEN> <N-QUOTE-M-SEN>))
(<N-QUOTE-MONEY> <--> (<N-QUOTE-M-HYAKU>))

(<N-QUOTE-MONEY> <--> (<HYAKU> <N-QUOTE-M-HYAKU>))
 (<N-VAUX-S+F-MONEY> <--> (<N-VAUX-S+F-M-MAN>))
 (<N-VAUX-S+F-MONEY> <--> (<N-VAUX-S+F-M-SEN>))
 (<N-VAUX-S+F-MONEY> <--> (<SEN> <N-VAUX-S+F-M-SEN>))
 (<N-VAUX-S+F-MONEY> <--> (<N-VAUX-S+F-M-HYAKU>))
 (<N-VAUX-S+F-MONEY> <--> (<HYAKU> <N-VAUX-S+F-M-HYAKU>))
 (<N-VAUX-SFP-MONEY> <--> (<N-VAUX-SFP-M-MAN>))
 (<N-VAUX-SFP-MONEY> <--> (<N-VAUX-SFP-M-SEN>))
 (<N-VAUX-SFP-MONEY> <--> (<SEN> <N-VAUX-SFP-M-SEN>))
 (<N-VAUX-SFP-MONEY> <--> (<N-VAUX-SFP-M-HYAKU>))
 (<N-VAUX-SFP-MONEY> <--> (<HYAKU> <N-VAUX-SFP-M-HYAKU>))
 (<SEN> <--> (<M-MAN>))
 (<HYAKU> <--> (<M-MAN>))
 (<HYAKU> <--> (<M-MAN> <M-SEN>))
 (<HYAKU> <--> (<M-SEN>))
 (<NP-MONEY-O> <--> (<NP-M-MAN-O>))
 (<NP-MONEY-O> <--> (<NP-M-SEN-O>))
 (<NP-MONEY-O> <--> (<SEN> <NP-M-SEN-O>))
 (<NP-MONEY-O> <--> (<NP-M-HYAKU-O>))
 (<NP-MONEY-O> <--> (<HYAKU> <NP-M-HYAKU-O>))
 (<N-VAUX-M-MAN> <--> (<n-vaux-money-man>))
 (<N-VAUX-M-SEN> <--> (<n-vaux-money-sen>))
 (<N-VAUX-M-HYAKU> <--> (<n-vaux-money-hyaku>))
 (<N-VAUX-COP-M-MAN> <--> (<n-vaux-cop-money-man>))
 (<N-VAUX-COP-M-SEN> <--> (<n-vaux-cop-money-sen>))
 (<N-VAUX-COP-M-HYAKU> <--> (<n-vaux-cop-money-hyaku>))
 (<N-VAUX-COORD-M-MAN> <--> (<n-vaux-coord-money-man>))
 (<N-VAUX-COORD-M-SEN> <--> (<n-vaux-coord-money-sen>))
 (<N-VAUX-COORD-M-HYAKU> <--> (<n-vaux-coord-money-hyaku>))
 (<N-VAUX-KATEI-M-MAN> <--> (<n-vaux-katei-money-man>))
 (<N-VAUX-KATEI-M-SEN> <--> (<n-vaux-katei-money-sen>))
 (<N-VAUX-KATEI-M-HYAKU> <--> (<n-vaux-katei-money-hyaku>))
 (<N-VAUX-S-M-MAN> <--> (<n-vaux-s-money-man>))
 (<N-VAUX-S-M-SEN> <--> (<n-vaux-s-money-sen>))
 (<N-VAUX-S-M-HYAKU> <--> (<n-vaux-s-money-hyaku>))
 (<NP-M-MAN> <--> (<np-money-man>))
 (<NP-M-SEN> <--> (<np-money-sen>))
 (<NP-M-HYAKU> <--> (<np-money-hyaku>))
 (<NP-NO-M-MAN> <--> (<np-no-money-man>))
 (<NP-NO-M-SEN> <--> (<np-no-money-sen>))
 (<NP-NO-M-HYAKU> <--> (<np-no-money-hyaku>))
 (<NP-E-M-MAN> <--> (<np-e-money-man>))
 (<NP-E-M-SEN> <--> (<np-e-money-sen>))
 (<NP-E-M-HYAKU> <--> (<np-e-money-hyaku>))
 (<M-MAN> <--> (<money-man>))

(<M-SEN> <--> (<money-sen>))
(<M-HYAKU> <--> (<money-hyaku>))
(<N-VAUX-H-M-MAN> <--> (<n-vaux-h-money-man>))
(<N-VAUX-H-M-SEN> <--> (<n-vaux-h-money-sen>))
(<N-VAUX-H-M-HYAKU> <--> (<n-vaux-h-money-hyaku>))
(<N-QUOTE-M-MAN> <--> (<n-quote-money-man>))
(<N-QUOTE-M-SEN> <--> (<n-quote-money-sen>))
(<N-QUOTE-M-HYAKU> <--> (<n-quote-money-hyaku>))
(<N-VAUX-S+F-M-MAN> <--> (<n-vaux-s+f-money-man>))
(<N-VAUX-S+F-M-SEN> <--> (<n-vaux-s+f-money-sen>))
(<N-VAUX-S+F-M-HYAKU> <--> (<n-vaux-s+f-money-hyaku>))
(<N-VAUX-SFP-M-MAN> <--> (<n-vaux-sfp-money-man>))
(<N-VAUX-SFP-M-SEN> <--> (<n-vaux-sfp-money-sen>))
(<N-VAUX-SFP-M-HYAKU> <--> (<n-vaux-sfp-money-hyaku>))

(<NP-M-MAN-O> <--> (<np-money-man-o>))
(<NP-M-SEN-O> <--> (<np-money-sen-o>))
(<NP-M-HYAKU-O> <--> (<np-money-hyaku-o>))

;;exceptions

(<EXCEP> <--> (<adv-s1> <vk-s1>))
(<EXCEP> <--> (<np> <adv-p> <mondai>))

参考文献

- [1] 橋本進吉: 国語法要説, 明治書院 (1934) (出典: 服部, 大野, 阪倉, 松村: 日本の言語学 - 文法 1 -, 大修館 pp.439-443 (1978, 1982³))
- [2] 服部四郎: 言語学の方法, 岩波書店 (1960)
- [3] 橋本進吉: 国文法体系論, 岩波書店 (1959) (出典: 服部, 大野, 阪倉, 松村: 日本の言語学 - 文法 1 -, 大修館 pp.444-472 (1978, 1982³))
- [4] 吉田将: “二文節間の係り受けを基礎とした日本語文の構文分析”, 電子通信学会論文誌 Vol.55-D, No.4, pp.238-244, (1972)
- [5] 内田, 吉岡, 杉村, 田中, 橋田, 向井: “第五世代コンピュータプロジェクトにおける自然言語処理システムの研究開発”, FGCS'88 (1988)
- [6] 柿ヶ原, 森元: “SL-TRANS における文節候補の削減 - 係り受け関係を用いた文節候補選択 -”, 情報処理学会第 39 回全国大会 4G-6 (1989)
- [7] 伊藤, 牧野, 木村, 城戸: “機能語予測 CYK 法による連続音声認識とその評価”, 信学技報 SP90-22 (1990)
- [8] 江原, 小倉, 森元: “電話対話データベースの構築”, 情報処理学会第 40 回全国大会 6F-1 (1990)
- [9] 北, 川端, 斉藤: “HMM 音韻認識と予測 LR パーザを用いた文節認識”, 日本音響学会昭和 63 年度秋期研究発表会講演論文集 2-P-29 (1988)
- [10] Kee, M., Watanabe, T., Tomita, M. : User's Guide for the Universal Parser, Version 6.9. Technical Memo (Center for Machine Translation, CMU) Pittsburgh (1986)
- [11] 郡司隆男: 自然言語の文法理論, pp.98-112 産業図書 (1987)
- [12] 江原暉将: 対話データベースからの頻度情報の抽出, ATR テクニカルレポート TR-I-0157 (1990)

- [13] 柴谷方良: 日本語の分析, 大修館 (1978, 1985⁴)
- [14] 篠崎, 水野, 小倉, 吉本: 形態素情報利用解説書, ATR テクニカルレポート TR-I-0077 (1989)
- [15] 吉本, 小暮: 句構造文法にもとづく日本語文の解析, ATR テクニカルレポート TR-I-0049 (1988)
- [16] Kita, K., Takezawa, T., Hosaka, J., Ehara, T., Morimoto, T.: "Continuous Speech Recognition Using Two-Level LR Parsing", Proc. of ICSLP-90, pp.905-908 (1990)
- [17] 坂野, 森元: "音声認識における正規文法活用の有効性", 信学技報 SP89-95 (1989)
- [18] 竹沢, 保坂, 北, 森元, 江原: "構文規則を用いた文音声認識", 信学技報 SP90-73/NLC90-45 (1990)