

TR-I-0337

ASURA における音声認識のための構文規則の拡張

Expansion of Syntactic rules for Speech Recognition in ASURA

谷戸文廣

衛藤純司⁺

Fumihiko Yato

Junji Etoh

1993年3月

内容梗概

日英・日独音声翻訳システムASURAにおいてモデル会話に加えてモデル文600文を音声認識可能とするため音声認識部に適用する構文規則を拡張した。本報告では拡張した文法について報告する。また、作成した文法が日本語音声認識にどの程度有効であるかを調査した結果を報告する。最後に、音声認識結果に基づいて今後どのような改良が必要かを論ずる。

ATR 自動翻訳電話研究所

ATR Interpreting Telephony Research Laboratories

+ 日本アイアール株式会社

目次

1	はじめに	3
2	文節に基づき文を組み立てる規則	4
2.1	感動詞句	4
2.2	接続詞句	4
2.3	名詞句	5
2.3.1	引用句	5
2.3.2	並列句	5
2.3.3	動詞の名詞化	6
2.4	動詞句	7
2.4.1	引用の動詞句	7
2.4.2	いる、ある、etc.	7
2.5	連体修飾句	8
2.5.1	連体修飾句の連接	8
2.5.2	引用の動詞の連体形文節	9
2.5.3	いる、ある、etc. の連体形文節	9
2.5.4	並列句の連体修飾句	10
2.6	副詞句	10
2.6.1	引用の動詞を含む副詞句	10
2.6.2	いる、ある、etc. を含む副詞句	10
2.7	年・月・日・時	11
2.8	ファックス番号	13
3	文節を構成する規則	14
3.1	名詞句	14
3.1.1	名詞と助詞の連接	14
3.1.2	副詞的名詞	16
3.1.3	並列句	16
3.1.4	動詞の名詞化	17
3.1.5	年・月・日・時	18
3.2	動詞句	19
3.2.1	引用の動詞句	19
3.2.2	いる、ある、etc.	20

3.2.3	住所	21
3.3	連体修飾句	21
3.3.1	副詞的名詞 + の	21
3.3.2	引用の動詞の連体修飾文節	21
3.3.3	いる、ある、etc. の連体修飾文節	22
3.3.4	ノ型形容名詞	22
3.4	副詞句	22
3.4.1	動詞句と接続助詞の接続	22
3.4.2	いる、ある、etc. と接続助詞の接続	23
3.4.3	「の」を伴う副詞	24
3.5	接頭語・接尾語	24
3.6	その他の規則	25
4	おわりに	27
4.1	構文情報の音声認識への応用	27
4.1.1	構文規則の応用効果	27
4.1.2	構文規則の応用効果の比較	30
4.2	誤認識の傾向と展望	31
4.2.1	誤認識の傾向	31
4.2.1.1	「サ行」の音、「しゃ、しゅ、しょ」の音	31
4.2.1.2	「も」と「の」「を」	31
4.2.1.3	副詞節の「と」と引用節の「と」	32
4.2.1.4	終助詞「よ」と「ね」	32
4.2.1.5	準体助詞による動詞の名詞化	33
4.2.2	展望	33
付録 A	文法	34
付録 B	形態素調整ルール	52
付録 C	文節区切りの規則	55
参考文献		60

第 1 章

はじめに

音声言語日英自動翻訳電話実験システム S L - T R A N S の日本語音声認識部では、文節を入力単位とした文の認識を行っている。そのための音声認識用文法は、文節を認識する文節内文法と、文節に基づいて文を組み立てる文節間文法の 2 段階の構成になっている。

音声認識用文法の役割は、次の音素を予測したり、容認性の低い音素列を排除したりして、出来るだけ日本語として容認性の高い音素列を得ることである。そのために、誤認識されやすい言語現象を扱う規則に制約を加えて、それらの規則が受け入れてしまう音素列を可能なかぎり絞り込むようにする。つまり、品詞を細分化し、規則をより詳細なものにして、個々の規則が扱う言語現象の範囲を狭く限定するのである。このように規則を異化・分化すればするほど、誤認識は少なくなり、音声認識率は高くなる。しかし、それとひきかえに、規則が細分化されればそれだけ規則の数が増え、その結果、音声認識にかかる時間が長くなってしまふ。このことは、リアル・タイムで応答しなければならない自動翻訳電話のようなシステムにとって好ましいことではない。したがって、あまりに特殊な制約を積み重ねて規則数をむやみに増やすことは避けなければならない。音声認識率を高めるために本質的に特殊なものである制約を加えるということと、その制約を出来るだけ一般的なものにするという、相反する要請のもとに文法を構築しなければならないのである。

これまでに、モデル会話 A, B, 1-5, 6-10 を対象にして、バージョン 1.1、バージョン 1.2 の文法を構築し、TR-I-0193 「S L - T R A N S における音声認識のための構文規則の概要」[1]と TR-I-0241 「S L - T R A N S における音声認識のための構文規則の拡張」[2]で報告した。本稿では、新たに 600 文のモデル文（以後、モデル会話 A, B, 1-10 を M セット、モデル文 600 文を E セットとする）を音声認識するために拡張したバージョン 1.3 の文法について報告する。第 2 章と第 3 章では、バージョン 1.2 で扱っていなかった文および文節を認識するための規則を説明する。第 4 章では、バージョン 1.3 が、日本語音声の認識にどの程度有効であるかを調べる。評価には音声認識率を使う。さらに、認識結果を基に、今後どのような改良が必要か検討する。

第 2 章

文節に基づき文を組み立てる規則

2.1 感動詞句

モデル文

いいえ、使用に対する制限はありません。

えーと、登録費のことなのですが。

感動詞が文頭にあり、述部が最低一つあるものである。方向性があり、「を」格をとれる述語を含むものと (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) に対応して、次の規則を設ける。

(〈SS〉 <--> (〈INTERJ1〉 〈NVS〉))

(〈SS〉 <--> (〈INTERJ1〉 〈NVS-DIR-OBJ〉))

2.2 接続詞句

モデル文

じゃあ また ご連絡します

文頭に接続詞が二つ連続する文である。「じゃあ、じゃあ、じゃあ」のように、ループになるのを避けるために、3 文節から成るフラットな規則を設けることにした。

(〈SS〉 <--> (〈CONJ1〉 〈NVS〉))

new (〈SS〉 <--> (〈CONJ1〉 〈CONJ1〉 〈NVS〉))

モデル文

四月に また お電話します

接続詞が文頭にではなく、文の途中に現れる文である。これも、3 文節から成るフラットな規則にする。

(〈SS〉 <--> (〈M-NN〉 〈CONJ1〉 〈NVS〉))

2.3 名詞句

2.3.1 引用句

モデル文

あしたは そちらに 伺えと 思います

引用の助詞「と」を伴う引用句は、これまで、他の補語と同じ扱いをしていたが (M-NN, M-NN-DIR-OBJ)、新たに M-NN-qn, M-NN-DIR-OBJ-qn という特別なカテゴリーを設けて、引用の述語文節 (後述のように VS-qn, NnVS-qn というカテゴリーを与える) との共起関係を制約することにした。

(〈M-NN〉 <--> (〈NQN〉))

new (〈M-NN-qn〉 <--> (〈NQN〉))

(〈M-NN-DIR-OBJ〉 <--> (〈NQN-DIR-OBJ〉))

new (〈M-NN-DIR-OBJ-qn〉 <--> (〈NQN-DIR-OBJ〉))

2.3.2 並列句

モデル文

会議で 使われる 言語は 英語と 日本語です

並列の助詞「と、や」などを伴う並列句も、他の補語と同じ扱いをして、特に並列関係を解析することはしていなかったが、新たに NN-H という特別なカテゴリーを与え、後続の名詞句ないし動詞句との並列関係を制約することにした。

(〈M-NN〉 <--> (〈NN〉))

(〈M-NN-DIR-OBJ〉 <--> (〈NN-DIR-OBJ〉))

new (〈M-NN〉 <--> (〈NN-H〉 〈NN〉))

new (〈M-NN〉 <--> (〈NN-H〉 〈NN-H〉 〈NN〉))

new (〈M-NN-DIR-OBJ〉 <--> (〈NN-H〉 〈NN-DIR-OBJ〉))

new (<NVC> <--> (<NN-H> <NVC>))

new (<NN-H> <--> (<np-h>))

N.B. Eセットのモデル文には、次のような並列構造がある。

例文 1 最寄りの駅は京都駅あるいは奈良駅です。

例文 2 お支払いは銀行振込もしくは現金でお願いします。

このような名詞単独の文節を含む並列構造を解析するための構文規則を、例えば次のように規定することができる。

(<NVC> <--> (<NN-H> <CONJ-H> <NVC>))

(<NN-H> <--> (<np-h>))

(<np-h> <--> (<n-proper>))

(<M-NN> <--> (<NN-H> <CONJ-H> <NN>))

(<NN-H> <--> (<np-h>))

(<np-h> <--> (<n-hutu>))

このうち、固有名詞は数が限られており、誤認識の可能性が小さいので、そのまま登録してあるが、普通名詞単独の文節は誤認識の大きな原因となったので、現在はコメント・アウトしてある。

2.3.3 動詞の名詞化

モデル文

会議で 発表を 聞くのを 楽しみに しています

動詞の連体形に準体助詞の「の」が後続し、さらに格助詞が後続するものは、これまで副詞句とされていたが、動詞の補語となることに変わりはないので、やはり後置詞句とみなして、次のように改める。

old (<VADV-H-P> <--> (<v-no-p>))

old (<VADV-H-P-O> <--> (<v-no-p-o>))

new (<M-NN-H> <--> (<v-no-p>))

new (<M-NN-H> <--> (<v-no-p-o>))
 new (<M-NN> <--> (<M-NN> <M-NN-H>))
 new (<NnM-NN-H> <--> (<M-NN> <M-NN-H>))
 new (<M-NN> <--> (<M-NN> <NnM-NN-H>))

2.4 動詞句

2.4.1 引用の動詞句

モデル文

ご提案の 主旨を 拝見した 上で 決めたいと 思います

引用の助詞「と」を伴う文節 (M-NN-qn, M-NN-DIR-OBJ-qn) を補語としてとる述語文節に、新たに VC-qn というカテゴリーを与え、両者の共起関係を次の規則で制約する。

new (<NVS> <--> (<NnVS-qn>))
 new (<NnVS-qn> <--> (<M-NN-qn> <VS-qn>))
 new (<VS-qn> <--> (<VC-qn>))

2.4.2 いる、ある、etc.

モデル文

いま 宣伝の ための 記事を 書いている
 委員会には 十一人の メンバーが いる

アスペクト要素の「いる」は必ず「動詞の連用形 + て」に後続し、一つの述語文節を作る。この接続規則は、文節内文法で次のように規定されている。

(<vaux> <--> (<v-te> <aspc-rentai>))

ところが「いる」は存在を表す動詞でもあって、そう解釈されると「動詞 + て」と「いる」が二つの文節に分割されてしまう。そこで、動詞の「いる」を含む文節に対して VC-iru というカテゴリーを与え、これには接続助詞を伴う副詞

句が係らないようにした。

```
new (<NVS> <--> (<VS-iru>))  
new (<VS-iru> <--> (<VC-iru>))  
new (<VS-iru> <--> (<ADV-s> <VS-iru>))  
new (<VS-iru> <--> (<VADV-H-COORD> <VS-iru>))
```

```
new (<NVS> <--> (<NnVS-iru>))  
new (<NnVS-iru> <--> (<M-NN> <VS-iru>))  
new (<NnVS-iru> <--> (<M-NN> <NnVS-iru>))  
new (<NnVS-iru> <--> (<ADV-s> <NnVS-iru>))  
new (<NnVS-iru> <--> (<ADV-c> <NnVS-iru>))
```

アスペクト要素と動詞との二義性を持つ他の語についても同様に扱う。すなわち、それらの動詞を含む述語文節に VC-iru というカテゴリーを与える。次のようなものがある。

ある おる いく いただく もらう みる

N.B. 文節を入力単位とする現在の方式では、アスペクト要素と動詞とを混同することはない。したがって、この規則はなくてもよいものである。しかし、将来、文単位連続音声を対象とする段階になれば、この規則が効果を発揮するようになるだろう。

-iru という限定辞よりも、たとえば -aspc のような限定辞の方がいいかもしれない。

2.5 連体修飾句

2.5.1 連体修飾句の連接

モデル文

公立の 機関の 方の 登録料は 割引に なります
会議で 発表される 論文の 内容を 教えて下さい

連体修飾句には、連体詞文節や所有格助詞「の」を伴う文節 (NM)、動詞・形容詞・形容動詞の連体形文節 (VM, NnVMS) があるが、例文のように、所有格助詞「の」を伴う連体修飾句が、さらに連体形文節によって修飾される規則は作っていなかった。今回、新たに次の規則を設けた。

```
new (<MOD-N> <--> (<VM> <NM>))
new (<MOD-N> <--> (<NnVMS> <NM>))
```

2.5.2 引用の動詞の連体形文節

モデル文

参加者は 二千人以上だと いう 事実を ご存知ですか

引用の補語をとる動詞の連体形文節に対して、新たに、VM-qn というカテゴリーを与える。そして、引用の補語 (M-NN-qn) との共起関係を次の規則で規定する。

```
new (<VMS-qn> <--> (<VM-qn>))
new (<NnVMS> <--> (<M-NN-qn> <VMS-qn>))
```

2.5.3 いう、ある、etc. の連体形文節

モデル文

観光ツアーが ある ことを 知りませんでした

アスペクト要素との二義性を持つ動詞の連体形文節に、新たに VM-iru というカテゴリーを与える。

```
new (<VMS> <--> (<ADV-s> <VM-iru>))
new (<NnVMS> <--> (<M-NN> <VMS-iru>))
new (<VMS-iru> <--> (<VM-iru>))
```

2.5.4 並列句の連体修飾句

モデル文

受付と 会場の 間に 講師の 部屋が あります

前述したように、並列助詞の「と」を伴う文節に NN-H というカテゴリーを与え、並列関係を扱うことにした。そこで、モデル文のように、並列句に所有格助詞「の」を伴う連体修飾句が後続する表現のために、次の規則を設ける。

new (<MOD-N> <--> (<NN-H> <NM>))

2.6 副詞句

2.6.1 引用の動詞を含む副詞句

モデル文

通訳料が 高いと 思われるかも しれません

引用の補語をとる述語に接続助詞がついて副詞句となる場合、引用の補語をとる述語を VC-qn としたのに対応して、新たに VADV-qn というカテゴリーを与える。

new (<ADVPH> <--> (<ADV-c-qn>))

new (<ADV-c-qn> <--> (<NnVADVS-qn>))

new (<NnVADVS-qn> <--> (<M-NN-qn> <VADVS-qn>))

new (<VADVS-qn> <--> (<VADV-qn>))

2.6.2 いる、ある、etc. を含む副詞句

モデル文

通訳が いますので ご安心下さい

アスペクト要素との二義性を持つ動詞に接続助詞がついて副詞句となる場合、述語句を VC-iru としたのに対応して、新たに VADV-iru というカテゴリーを与える。

new (<ADVPH> <--> (<ADV-c-iru>))
 new (<ADV-c-iru> <--> (<NnVADVS-iru>))
 new (<NnVADVS-iru> <--> (<M-NN> <VADVS-iru>))
 new (<NnVADVS-iru> <--> (<M-NN> <VADVS-iru>))
 new (<VADVS-iru> <--> (<VADV-iru>))

2.7 年・月・日・時

モデル文

次回の 会議は 千 九百 九十 三年に 開かれる 予定です

西暦の年号は、金額と同様、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。そこで、次の規則を設けて対処した。

new (<NN> <--> (<NM-ANUAL-SEN> <NM-ANUAL-HYAKU> <NM-ANUAL-ZYUU> <NN-ANUAL>))
 new (<NM-ANUAL-SEN> <--> (<anual-num-sen>))
 new (<NM-ANUAL-HYAKU> <--> (<anual-num-hyaku>))
 new (<NM-ANUAL-ZYUU> <--> (<anual-num-zyuu>))
 new (<NN-ANUAL> <--> (<np-anual>))

モデル文

6 月 30 日までに 原稿の 送付を お願いします
 会議は 8 月 22 日から 25 日まで 京都国際会議場で 開催されます

従来の文法では、「6 月」「30 日までに」や「8 月」「22 日から」「25 日まで」がそれぞれ単独に「お願いします」や「開催されます」に係るように構築されていた。しかし、意味的には「6 月 30 日までに」「8 月 22 日から 25 日まで」が全体として「お願いします」「開催されます」に係るようにすべきである。そこで、今回、次のような一連の規則を設けて対処することにした。

new (<NN> <--> (<NM-MONATE> <NN-TAG>)) ; 6 月 30 日までに
 new (<NN> <--> (<NM-2> <NN-TAG-2>)) ; 8 月 22 から 25 日まで

new (<NM-2> <--> (<NM-MONATE> <NN-TAG-2>)) ; 8月22日から
new (<NM-MONATE> <--> (<monate-rentai>)) ; 6月 8月
new (<NN-TAG> <--> (<np-tag>)) ; 30日までに
new (<NN-TAG-2> <--> (<np-tag-2>)) ; 22日から、25日まで

モデル文

8月 27日以後の 取り消しに対する 払い戻しは できません

これも、従来の文法では「8月」と「27日以後の」を結合することはしていなかった。今回、次の規則で一つにまとめることにした。

new (<NM> <--> (<NM-MONATE> <NM-TAG>))
new (<NM> <--> (<NM-TAG>))
new (<NM-TAG> <--> (<np-no-tag>))

N.B. 「取り消しに対する」は NPM というカテゴリーが与えられるので、これに連体修飾句に係る構文のために、次の規則を設ける。

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

モデル文

午前 9時までに 受付に 来て下さい

ヴァージョン1.2の文法では「午前」「9時までに」がそれぞれ単独で「来て下さい」に係るように構築されていた。今回、次の規則を設けて「午前9時までに」が全体として「来て下さい」に係るようにした。

new (<NN> <--> (<NM-SLOT> <NN-DATUM>)) ; 午前9時までに
new (<NN> <--> (<NN-DATUM>)) ; 9時に
new (<NN> <--> (<NM-ZI> <NN-DATUM>)) ; 9時50分に
new (<NN-DATUM> <--> (<np-datum>))
new (<NM-SLOT> <--> (<slot-tag>))
new (<NM-ZI> <--> (<zi-tag>))

2.8 ファックス番号

モデル文

そして ファックスの 番号は 一三〇八ですね

例文3 のような電話番号では、三桁の文節を TEL1、四桁の文節を TELVC として、次の規則で解析している。

例文3 電話番号は三三一の二五二一です。

(〈TEL-VC〉 <--> (〈TEL1〉 〈TELVC〉))

ファックス番号にもこの規則を用いるが、モデル文のように、三桁の文節を欠く文のために、次の規則を設ける。

new (〈TEL-VC〉 <--> (〈TELVC〉))

第 3 章

文節を構成する規則

3.1 名詞句

3.1.1 名詞と助詞の連接

助詞は音節数が少なく、誤認識されやすいもののひとつである。この傾向は語彙が増えるにつれて顕著になる。Eセットの文法でも、助詞の誤認識がきわめて多く、このことが文認識率の低下を招いていた。そこで、名詞と助詞の連接関係を厳格に制限してみることにした。すなわち、それぞれの助詞に異なったカテゴリー名を与え、各種の名詞との連接規則を詳細に規定したのである。助詞の数が限られているので、規則の体系はそれほど複雑なものにはならない。以下にその一部を示す。

格助詞 :	p-kaku-ga	が
	p-kaku-o	を
	p-kaku-ni	に
	p-kaku-e	へ
	p-kaku-de	で
	p-kaku-no	の
	p-kaku-kara	から
	p-kaku-made	まで
	p-kaku-yori	より
	p-kaku-toshite	として
	p-kaku-nitsuite	について
	p-kaku-nioite	において
	p-kaku-nitoqte	にとって
	p-kaku-niyote	によって
	p-kaku-nikanshite	に関して
	p-kaku-totomoni	とともに

係助詞 :	p-k-wa	は
	p-k-mo	も

p-k-qte	って
p-k-demo	でも

取立助詞 :	p-f-nado	など
	p-f-hodo	ほど
	p-f-nomi	のみ
	p-f-sae	さえ
	p-f-nanka	なんか
	p-f-dake	だけ
	p-f-bakari	ばかり
	p-f-kurai	くらい
	p-f-gurai	ぐらい
	p-f-shika	しか
	p-f-koso	こそ
	p-f-zutu	ずつ
	p-f-ka	か

並列助詞 :	p-h-to	と
	p-h-ya	や
	p-h-ka	か
	p-h-toka	とか

引用助詞 :	p-i-to	と
	p-i-toka	とか
	p-i-qte	って

規則の例 :

($\langle np \rangle$ <--> ($\langle n-hutu \rangle$ $\langle p-kaku-ga \rangle$))
($\langle np \rangle$ <--> ($\langle n-proper \rangle$ $\langle p-kaku-ni \rangle$))
($\langle np \rangle$ <--> ($\langle n-name \rangle$ $\langle p-f-nado \rangle$))
($\langle np \rangle$ <--> ($\langle n-zahl \rangle$ $\langle p-kaku-de \rangle$))
($\langle np \rangle$ <--> ($\langle n-hutu-time \rangle$ $\langle p-kaku-kara \rangle$))
($\langle np \rangle$ <--> ($\langle n-suffix \rangle$ $\langle p-kaku-kara \rangle$))
($\langle np-o \rangle$ <--> ($\langle n-hutu \rangle$ $\langle p-kaku-o \rangle$))
($\langle np-e \rangle$ <--> ($\langle n-hutu \rangle$ $\langle p-kaku-e \rangle$))
($\langle np-h \rangle$ <--> ($\langle n-hutu \rangle$ $\langle p-h-to \rangle$))

(<np-no> <--> (<n-hutu> <p-kaku-no>))
 (<np-adv> <--> (<n-adv> <p-k-wa>))
 (<wh-np> <--> (<wh-pro-iku> <p-k-demo>))
 (<np-keisiki> <--> (<n-keisiki> <p-kaku-ga>))
 (<np-money-man> <--> (<n-money-man> <p-kaku-de>))
 (<v-no-p> <--> (<nomina-p> <p-k-wa>))

N.B. 助詞の誤認識で最も多いのは「が」と「は」、「の」「を」と「も」の誤認識である。これらは、新しい文法でも解消されていない。

3.1.2 副詞的名詞

モデル文

お忙しい 中 申し訳ありません
 会議を 開催する たびに 参加者が 増えています

助詞を伴わずに副詞的に使われることのできる名詞で、従来 n-exc というカテゴリーを与えていたものである。副詞的という意味を明示するように n-adv と変更し、次のような規則を作った。ただし、n-adv-2 は、助詞を伴うことのないものである。

new (<np-adv> <--> (<n-adv>))
 new (<np-adv> <--> (<n-adv> <p-kaku-ni>))
 new (<np-adv> <--> (<n-adv> <p-k-wa>))
 new (<np-adv> <--> (<n-adv> <p-k-mo>))
 new (<np-adv> <--> (<n-adv-2>))

3.1.3 並列句

モデル文

受付と 会場の 間に 講師の 部屋が あります

並列の助詞「と、や、か、とか」を伴う名詞句に np-h というカテゴリーを与え、次の規則を設けた。

new (<np-h> <--> (<n-hutu> <p-h-to>))
 new (<np-h> <--> (<n-hutu> <p-h-ya>))
 new (<np-h> <--> (<n-hutu> <p-h-ka>))
 new (<np-h> <--> (<n-hutu> <p-h-toka>))

次の例文 4、例文 5 のように、名詞単独で並列句のひとつの項となることがある。

例文 4 八月 五日の 午後に 清水寺 金閣寺 龍安寺などを 見学
 します

例文 5 学生 および 大学関係者には 登録料が 割引に なります

このうち、例文 4 のような固有名詞単独の文節のみを次の規則で扱うことにした。普通名詞単独の文節は誤認識の大きな原因になるからである。

new (<np-h> <--> (<n-proper>))

3.1.4 動詞の名詞化

モデル文

日本へ 行くのは はじめてです

アスペクト要素との二義性をもつ動詞の連体形に nomina-iru というカテゴリー名を与える。また、これに準体助詞「の」が後続して名詞化したものに nomina-p-iru というカテゴリー名を、さらに格助詞が後続するものに、v-no-p-iru というカテゴリー名を与える。

new (<v-no-p-iru> <--> (<nomina-p-iru> <p-k-wa>))
 new (<nomina-p-iru> <--> (<nomina-iru> <p-j>))
 new (<nomina-iru> <--> (<v-5dan-u-iru>))
 new (<v-5dan-u-iru> <--> (<iku-5dan-ki> <flex-5-ku>))

3.1.5 年・月・日・時

モデル文

次回の 会議は 千 九百 九十 三年に 開かれる 予定です

西暦の年号を表す数詞のそれぞれの位を次の規則で扱う。

```
new (<anual-num-sen> <--> (<anual-num-1000>))
new (<anual-num-hyaku> <--> (<anual-num-1> <anual-num-100>))
new (<anual-num-zyuu> <--> (<anual-num-1> <anual-num-10>))
new (<np-anual> <--> (<datum-anual> <p-kaku-ni>))
new (<datum-anual> <--> (<monate> <anual-suf>))
new (<anual-suf> <--> (n e =))
```

N.B. <monate> は、もともと、一月・二月・三月 … など、月につく数詞を指すカテゴリである。今回これを流用した。

モデル文

6月 30日までに 原稿の 送付を お願いします
会議は 8月 22日から 25日まで 京都国際会議場で 開催されます
8月 27日以後の 取り消しに対する 払い戻しは できません

従来の文法では、「6月」と「30日までに」の係り受け関係を解析していなかった。今回、「6月」が「30日ごろに」に係るようにするために、次の規則を作った。

```
new (<monate-rentai> <--> (<datum-monate>))
new (<np-tag> <--> (<datum-tag> <p-kaku-ni>))
```

また、「8月22日から25日まで」をひとまとまりの句として解析するために、次の規則を作った。

```
new (<np-tag-2> <--> (<datum-tag> <p-kaku-kara-2>))
new (<np-tag-2> <--> (<datum-tag> <p-kaku-made-2>))
```

「8月」と「27日以後の」の係り受けを解析するために、次の規則を作った。

```
new (<np-no-tag> <--> (<datum-tag> <p-kaku-no>))
```

モデル文

午前 9時までに 受付に 来て下さい
9時 50分から 基調講演が 行われます

「午前」と「9時までに」の係り受けを解析するために、次の規則を作った。

```
new (<slot-tag> <--> (<tag-slot>))
```

```
new (<tag-slot> <--> (g o z e =))
```

```
new (<tag-slot> <--> (g o g o))
```

```
new (<tag-slot> <--> (a s a))
```

```
new (<np-datum> <--> (<n-datum> <p-kaku-made> <p-kaku-ni>))
```

```
new (<n-datum> <--> (<tag-zi>))
```

「9時」と「50分から」の係り受けを解析するために、次の規則を作った。

```
new (<zi-tag> <--> (<tag-zi>))
```

3.2 動詞句

3.2.1 引用の動詞句

モデル文

ご提案の 主旨を 拝見した 上で 決めたいと 思います
コンピューターに関する 国際会議が あると 聞いているんですが
長尾先生は 出席なさると 思いますか

引用の補語をとる動詞を含む動詞句に `vaux-qn` というカテゴリー名を与える。
また、最後に「が、けれども」のような接続助詞がつくものに `vaux-s+f-qn`、「
か、ね」のような終助詞がつくものに `vaux-sfp-qn` を与える。

```

new (<VC-qn> <--> (<vaux-qn>))
new (<VC-qn> <--> (<vaux-s+f-qn>))
new (<vaux-s+f-qn> (<nomina-cop-qn> <v-cop-rentai> <sp-rentai+f>))
new (<nomina-cop-qn> <--> (<nomina-qn> <p-j>))
new (<nomina-cop-qn> <--> (<nomina-qn> <p-j-n>))
new (<VC-qn> <--> (<vaux-sfp-qn>))
new (<vaux-sfp-qn> <--> (<vaux-qn> <sfp>))

```

3.2.2 いる、ある、etc.

モデル文

委員会には 十一人の メンバーが いる
 会議の 間に 市内観光が あるそうですが
 会場に スライド用の スクリーンが ありますか
 どうしても 見たい お寺が 一箇所 あるのですが

アスペクト要素との二義性を持つ動詞を含む動詞句に *vaux-iru* というカテゴリー一名を与える。また、最後に「が、けれども」のような接続助詞がつくものに *vaux-s+f-iru*、「か、ね」のような終助詞がつくものに *vaux-sfp-iru* を与える。また、連体形に準体助詞の「の」が後接し、さらにコブラの「です」が後続するものに *vaux-cop-iru* を与える。

```

new (<VC-iru> <--> <vaux-iru>))
new (<VC-iru> <--> <vaux-s+f-iru>))
new (<vaux-s+f-iru> <--> (<vaux-iru> <sp-rentai+f>))
new (<VC-iru> <--> <vaux-sfp-iru>))
new (<vaux-sfp-iru> <--> (<vaux-iru> <sfp>))
new (<vaux-sfp-iru> <--> (<vaux-cop-iru> <sfp>))
new (<VC-iru> <--> <vaux-cop-iru>))
new (<vaux-cop-iru> <--> (<nomina-cop-iru> <v-cop>))
new (<nomina-cop-iru> <--> (<nomina-iru> <p-j>))
new (<nomina-cop-iru> <--> (<nomina-iru> <p-j-n>))

```

3.2.3 住所

モデル文

ご住所は 東京都ですか

モデル会話文では、住所に「だ、です」が後続し、それに終助詞がつくものは扱っていなかった。Eセットでは上記の文があるので、次の規則を作った。

```
new (<AVC> <--> (<adre-vaux-cop-sfp>))
new (<adre-vaux-cop-sfp> <--> (<adre-vaux-cop> <sfp>))
```

3.3 連体修飾句

3.3.1 副詞的名詞 + の

モデル文

ぜひ 編集の ための 部会に 出席してください

副詞的用法を持つ名詞に所有格助詞の「の」がついたものに、np-adv-no というカテゴリー名を与える。

```
new (<np-adv-no> <--> (<n-adv> <p-kaku-no>))
```

3.3.2 引用の動詞の連体修飾文節

モデル文

参加者は 二千人以上だと いう 事実を ご存知ですか

引用の補語をとる動詞の連体形による連体修飾文節に vaux-nom-qn というカテゴリー名を与える。

```
new (<VM-qn> <--> (<vaux-nom-qn>))
new (<vaux-nom-qn> <--> (<v-5dan-u-qn>))
new (<v-5dan-u-qn> <--> (<v-5dan-w-qn> <flex-5-wu>))
```

```
new (<v-5dan-u-qn> <--> (<v-5dan-ki-qn> <flex-5-ku>))
new (<v-5dan-u-qn> <--> (<v-5dan-r-qn> <flex-5-ru>))
```

3.3.3 いる、ある、etc. の連体修飾文節

モデル文

観光ツアーが ある ことを 知りませんでした
どうしても 見たい お寺が 一箇所 あるんですが

アスペクトを表す補助動詞との二義性を持つ動詞を含む動詞句の連体修飾文節に `vaux-nom-iru` というカテゴリー名を与える。

```
new (<VM-iru> <--> (<vaux-nom-iru>))
new (<vaux-nom-iru> <--> (<v-5dan-u-iru>))
new (<vaux-nom-iru> <--> (<v-renyol-iru> <optt-rentai>))
new (<vaux-nom-iru> <--> (<v-mizenl-iru> <negt-rentai>))
```

3.3.4 ノ型形容名詞

モデル文

公立の 機関の 方の 登録料は 割引に なります

連体形が「…の」の形になる形容名詞のために、新たに `kdo-no` というカテゴリー名を設け、次の規則を作った。

```
new (<kdo-rentai> <--> (<kdo-no> <flex-kdo-rentai-no>))
```

3.4 副詞句

3.4.1 動詞句と接続助詞の接続

接続助詞も格助詞と同様に誤認識の多いもののひとつである。そこで、それぞれの語にカテゴリー名を与えて、動詞句との接続を詳細に規定することにした。

接続助詞：	sp-rentai-ga	が
	sp-rentai-to	と
	sp-rentai-kara	から
	sp-rentai-naraba	ならば
	sp-rentai-node	ので
	sp-rentai-tomo	とも
	sp-renyo-te	て
	sp-renyo-temo	ても
	sp-renyo-tewa	ては
	sp-renyo-tara	たら
	sp-renyo-tomo	とも
	sp-renyo-nagara	ながら
	sp-renyo-shidai	次第
	sp-katei-ba	ば

規則の例： (<vaux-s> <--> (<vaux> <sp-rentai-node>))
 (<vaux-s> <--> (<v-renyo2-t> <sp-renyo-tara>))
 (<vaux-s> <--> (<v-katei> <sp-katei-ba>))

3.4.2 いる、ある、etc. と接続助詞の連接

モデル文

最低 十人は いないと ツアーは 中止に なります
 分からない ことが ありましたら 事務局に お尋ね下さい
 いま 案内書を 見ながら 電話を かけています

アスペクトを表す補助動詞との二義性を持つ動詞を含む副詞句に vaux-s-iru というカテゴリー名を与える。

```
new (<vaux-s-iru> <--> (<vaux-iru> <sp-rentai-to>))
new (<vaux-s-iru> <--> (<v-5dan-i-iru> <masu-renyo> <sp-renyo-tara>
)
new (<vaux-s-iru> <--> (<v-renyo2-t-iru> <sp-renyo-nagara>))
```

3.4.3 「の」を伴う副詞

モデル文

最初の 部分の 字が 読めません

しかし 団体の 方々には 若干の 割引が ございます

「の」を伴って連体修飾文節を作りうる副詞に、新たに adv-no というカテゴリー一名を与える。そして、この「の」は、ノ型形容名詞につく「の」と同じだと考え、次の規則で解析する。

```
new (<kdo-rentai> <--> (<adv-no> <flex-kdo-rentai-no>))
```

3.5 接頭語・接尾語

Eセットでは、次の語を新しく接頭語・接尾語として認定した。

接頭語： お（御） ご（御） 各 約 およそ

接尾語： 回 字 ページ 版 以上 目 程度

用 前 後 中 内 さん 先生

このうち、音声認識部では、「約」「さん」「先生」だけを実装した。

```
new (<n-zahl-num> <--> (<num-prefix> <n-zahl>))
```

```
new (<num-prefix> <--> (y a k u))
```

```
new (<name-suf> <--> (s a =))
```

```
new (<name-suf> <--> (s e = s ei))
```

それ以外のものは、誤認識の大きな原因となったので、今回は規則化を断念した。

3.5 その他の規則

もっぱら音声認識率を上げるためにのみ作ったアド・ホックな規則がいくつかある。

モデル文

それほど 大規模な 会議だとは 知りませんでした
参加者は 二千人以上だと いう 事実を ご存知ですか
そうだとは 思いません

Eセットの文では、「…だ」が単独で文節となることはなく、必ず「…だと」または「…だとは」という文節で使われている。これらの文節は、次の規則で解析されていた。

```
(<n-quote> <--> (<n-vaux-cop> <p-i>))  
(<n-vaux-cop> <--> (<n> <cop-syusi>))  
(<cop-syusi> <--> (d a))
```

```
(<quote> <--> (<vaux-cop> <p-i>))  
(<vaux-cop> <--> (<adv> <cop-syusi>))
```

ところが、n-vaux-cop と vaux-cop は文節カテゴリーのひとつなので、音声認識の際に「…だ」と誤認識される例がきわめて多かった。そこで、次のように改めた。

```
(<n-quote> <--> (<n-vaux-cop-da> <p-i-to>))  
(<n-quote> <--> (<n-vaux-cop-da> <p-i-to> <p-k-wa>))  
(<n-vaux-cop-da> <--> (<n-hutu> <cop-syusi>))  
(<n-vaux-cop-da> <--> (<n-zahl> <cop-syusi>))  
  
(<quote> <--> (<adv> <cop-syusi> <p-i-to> <p-k-wa>))
```

モデル文

ビデオを そちらに 運ぶとか いう ことは できますか

はじめ、この文を解析するために、次の規則を作った。

(`<quote> <--> (<vaux> <p-i-6>))`)

ところが、`vaux` を生成する規則は非常に多く、この規則が副作用を及ぼして「…とか」と誤認識される例が頻出した。そこで、「運ぶとか」に限定するために、次のように改めた。

(`<quote> <--> (<v-5dan-b> <flex-5-bu> <p-i-toka>))`)

モデル文

ここからは その建物が 見えません

代名詞の「ここ」も、頻繁に誤認識の原因となった。とくに、形式名詞の「こと」と混同されることが多かった。Eセットでは、「ここ」が使われるのは上記の文だけであるので、特別に `pro-exp` というカテゴリー名を与え、次の規則で制限した。

(`<np> <--> (<pro-exp> <p-kaku-kara> <p-k-wa>))`)

(`<pro-exp> <--> (k o k o)`)

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 バージョン 1.3 のサイズを、バージョン 1.2 のそれと比較して 表 4.1 に示す。文節内文法の終端規則数は語彙数を表し、文の文法の終端規則数は文節カテゴリーの数を表している。

文節内文法と文文法をマージしたもののパープレキシティは、バージョン 1.2 のそれは 3.32、バージョン 1.3 のそれは 4.29 である。

表 4.1 文法のサイズ

	文節内文法		文の文法	
	Vers. 1.2	Vers. 1.3	Vers. 1.2	Vers. 1.3
規則数	1927	2732	511	397
終端規則数	716	1654	138	111

N.B. 文節内文法の規則数から終端規則数を引いた数、文の文法の規則数、および文の文法の終端規則数を見ると、Vers.1.3の方がVer.1.2よりも少ない。その理由は、音声認識率を上げるために、Eセットの文を解析するのに必要な規則以外を削除しているからである。

文節を組み立てる規則及び文節に基づき文を組み立てる規則がどのくらい音声認識に貢献しているか、また、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 音声認識を行った。

4.1.1 構文規則の応用効果

バージョン 1.3 が、どの程度音声認識に有効であるかを調べる。実験対象は Eセットのモデル文 600 文である。1 人の話者が 200 文ずつ発話したものが

6通りあり、それぞれ FAK、FKN、MTK、FMT、FYU、FYK としている。（実際には、ある種の並列構文など規則化していないものや、文節の区切り方が規則とずれているものなどを除き、561文のみを実験対象とした）

表 4.2 実験対象

	FAK	FKN	MTK	FMT	FYU	FYK
総文数	194	188	181	179	192	
総文節数	770	700	697	696	761	
1文中の平均文節数	4.0	3.7	3.9	3.9	4.0	
1文中の最高文節数	7	6	8	8	7	

実験は、グローバルビーム幅 100、ローカルビーム幅 18 の組と、グローバルビーム幅 250、ローカルビーム幅 18 の組の 2 通りで行った。表 4.3 に文節認識率、表 4.4 に文認識率を示す。

表 4.3 文節認識率

Rank	累積認識率（単位：％ ビーム幅：100:18）					
	FAK	FKN	MTK	FMT	FYU	FYK
1	80.0	75.5	78.8	79.0	75.3	
2	88.2	88.7	87.9	88.9	87.4	
3	91.4	92.1	90.5	91.1	91.2	
4	93.2	92.9	91.8	92.1	92.5	
5	93.9	93.3	92.4	92.8	92.8	

Rank	累積認識率（単位：％ ビーム幅：250:18）					
	FAK	FKN	MTK	FMT	FYU	FYK

1	80.2	76.5	79.0	79.5	75.5	
2	88.4	90.0	88.1	89.6	87.6	
3	91.5	93.6	90.5	92.2	91.7	
4	93.6	94.4	91.8	93.3	93.0	
5	94.5	95.1	92.6	94.1	93.5	

表 4.4 文認識率

Rank	累積認識率（単位：％ ビーム幅：100:18）					
	F A K	F K N	M T K	F M T	F Y U	F Y K
1	46.39	43.55	43.02	44.69	38.54	
2	60.31	55.38	57.54	58.66	51.56	
3	66.49	63.98	60.34	62.01	54.69	
4	68.04	68.82	63.69	65.36	57.81	
5	72.16	70.21	62.98	67.04	63.02	
:	:	:	:	:	:	
10	74.23	72.87	66.85	69.83	66.15	

Rank	累積認識率（単位：％ ビーム幅：250:18）					
	F A K	F K N	M T K	F M T	F Y U	F Y K
1	47.94	44.09	43.89	45.19	40.31	
2	61.86	56.99	58.33	61.24	54.45	
3	69.07	65.05	62.22	65.73	59.16	
4	71.65	72.04	65.00	69.66	63.35	
5	74.74	74.19	65.56	71.35	67.02	
:	:	:	:	:	:	
10	78.35	77.96	68.89	74.72	71.73	

4.1.2 構文規則の応用効果の比較

規則を追加したことが、音声認識にどのような影響を及ぼすかを調べる。バージョン1.2の文法規則を使った時の認識率と、バージョン1.3の文法規則を使った時の認識率の比較を表4.4に示す。実験対象はモデル会話A, B, 1-10の260文である。また、グローバルビーム幅は100、ローカルビーム幅は18である。

表4.4 文節認識実験結果 -比較-

Rank	累積認識率（単位：%）	
	バージョン1.2	バージョン1.3
1	92.3	89.4
2	98.7	95.6
3	99.3	96.9
4	99.3	96.9
5	99.4	97.2

表4.5 文認識実験結果 -比較-

Rank	累積認識率（単位：%）	
	バージョン1.2	バージョン1.3
1	85.38	77.22
2	92.69	83.40
3	95.00	87.26
4	96.92	89.19
5	98.08	90.35

4.2 誤認識の傾向と展望

4.2.1 誤認識の傾向

Eセットのモデル文を対象とした音声認識実験で、正解を得られなかったものについて、誤認識の原因となった規則がどのようなものであったかを調べてみた。そのために、モデル文を正しく解析するのに必要な規則と、誤認識された文を生成した規則の頻度を比較した。そして、前者の頻度に比較して後者の頻度が際立って高かったもの、または際だって低かったものを調べた。

以下に述べるのは、その中でも特に目に立った例で、ある種の傾向と言いうるものである。

4.2.1.1 「サ行」の音、および「しゃ、しゅ、しょ」の音

語彙のレベルで誤認識が目立ったものに、「サ行」の音や「しゃ、しゅ、しょ」の音を含むものがある。

出席
セッション
最終的

最初
事情
新製品

「出席」「セッション」「最終的」などは、Global beam == 100 のときは一度も、Global beam == 250 のときもわずか一度もしくは二度しか正しく認識されず、「最初」「事情」「新製品」などと誤認識された。

これらは、構文規則で制限することはむずかしく、根本的には音声認識の性能の向上を待つほかはない。

4.2.1.2 「も」と「の」「を」

(<np> <--> (<n-hutu> <p-k-mo>))

例文 6 今 宣伝も ためも 記事を書いている

例文 7 先生も 講演も 順番も 一番目から 三番目に 直したいのですが

正しくは、「宣伝の」「ための」「先生の」「講演の」「順番を」である。このように、「の」や「を」が「も」と誤認識される例が目立って多かった。

統語的な規則だけでこれらの誤認識を防ぐことはむずかしい。何らかの形で意味的な整合性をチェックする機構を導入すべきだろう。たとえば、例文 6 では、「書く」の必須格補語が「が格」「を格」であり、「が格」には人間しかなりえないから、排除されることになる。また、例文 7 では、「直す」の必須格補語が「が格」「を格」であり、助動詞「たい」の主格が話者でなければならないので、「先生も」は排除できるし、「講演も」より「講演の」を、「順番も」より「順番を」を優先的に選ぶことができる。

あるいは、1 文に 2 つ以上の「も格」が現れるのを制限することも考えられる。ただし、この場合は、並列の「…も…も」を排除しないようにしなければならないが。

4.2.1.3 副詞節の「と」と引用節の「と」

($\langle \text{VADV} \rangle \langle \text{--} \rangle (\langle \text{vaux-s} \rangle)$)
($\langle \text{vaux-s} \rangle \langle \text{--} \rangle (\langle \text{vaux} \rangle \langle \text{sp-rentai-to} \rangle)$)

例文 8 長尾先生が 説明されると 泊まります

例文 9 先生が 会議の 詳細を 知りたいと 載せました

文末の述語文節が、正しくは「思います」「おっしゃいました」である。これらは、いずれも引用節をとる動詞で、現在の文法では共起関係を制限している。たぶん、会話文では、接続助詞の「と」よりも引用助詞の「と」の方が圧倒的に頻度が高いであろうから、確率文法をうまく使えば、この種の誤認識を排除することができるであろう。

4.2.1.4 終助詞「よ」と「ね」

($\langle \text{vaux-sfp} \rangle \langle \text{--} \rangle (\langle \text{vaux} \rangle \langle \text{sfp} \rangle)$)
($\langle \text{vaux-sfp-iru} \rangle \langle \text{--} \rangle (\langle \text{vaux-iru} \rangle \langle \text{sfp} \rangle)$)

(<sfp> <--> (<sfp-2>))
(<sfp> <--> (<sfp-3>))
(<sfp-2> <--> (y o))
(<sfp-3> <--> (n e))

例文 1 0 それでは 早速 飛行機の 予約を しますよ

例文 1 1 登録料と 論文集代の ほかに 費用は いりますね

「しましょう」「いりません」が正しい。このように「… しましょう」が「… ますよ」と、「… ません」が「… ますね」と誤認識される例が目立った。この種の誤認識は、統語的な制約でも意味的な制約でも排除することはむずかしいだろう。正しく認識するためには、会話における質問・応答の流れや発話意図に関する情報のような、語用論的な情報を使わなければならない。

4.2.1.5 準体助詞による動詞の名詞化

(<nomina-cop> <--> (<nomina> <p-j-n>))
(<nomina> <--> (<v-te> <aspc-rentai>))
(<nomina> <--> (<v-5dan-u>))

例文 1 2 ちょうど 三名分だけ パンフレットが 残ってるんですした

例文 1 3 早く 論文を 提出していただかないと 印刷に 間に合うんですよ

正しくは「残っていました」「間に合いません」である。このように、「… いました」が「… るんですした」と、「… ません」が「… るんですした」と誤認識される例も目立った。これも、単純な統語的な制約や意味的な制約では排除することがむずかしい。やはり、根本的には、談話の流れ・発話意図・プランなどの語用論的な情報が必要になる。

4.2.2 展望

音声認識に制約として使っている構文規則は、もっぱら統語的な制約だけを文脈自由文法で記述している。しかし、Eセットのモデル文の音声認識実験で分か

ったことは、単なる統語的な制約だけでは、あまり認識率を向上させることはできないだろうということである。もっと、意味的な制約や、語用論的な制約を積極的に導入する必要がある。

現在の枠組みで意味的な制約を導入するとすれば、たとえば、動詞を格パターンに従って分類し、名詞の意味素性をも含めた格のカテゴリーを設定して、精密な共起制約を規則化することが考えられる。ただし、この規則はきわめて複雑なものになるだろうから、規則体系の全体を容易に見渡すことができ、規則相互の無矛盾性を検証できるようなデバッグツールが必要である。

また、語用論的な制約を導入するとすれば、文末述語の助動詞・補助動詞の連接を、発話意図の表現パターンに従って語彙化することが考えられる。すなわち、実際の会話に現れる文末表現のタイプは数が限られている可能性があるので、現在のような、助動詞・補助動詞の連接可能性を考慮して作った精密な承接規則の代わりに、発話意図の表現パターンの全体を語彙として登録するわけである。一般に音節数が多い語ほど認識率が高いので、この方法はうまくいくと認識率を高めるのに役立つかもしれない。ただし、あくまで文末述語のタイプの数が本当に限られているということが前提であって、一度コーパスを調査してみなければならないが。

;;; ***** BUNSETU COMPOSITION GRAMMAR 1.3 version *****

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

;;=====

;; 初期記号

;;=====

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

;;=====

;; 感動詞を含む文

;;=====

(<SS> <--> (<INTERJ1>))
(<SS> <--> (<INTERJ1> <NVS>))
(<SS> <--> (<INTERJ1> <NVS-DIR-OBJ>))
(<SS> <--> (<EXCEP>))

;;=====

;; 接続詞を含む文

;;=====

(<SS> <--> (<CONJ1> <NVS>))
(<SS> <--> (<CONJ1> <NVS-DIR-OBJ>))
(<SS> <--> (<CONJ1> <CONJ1> <NVS>))
(<SS> <--> (<M-NN> <CONJ1> <NVS>))

;;=====

;; 述語を含む文

;;=====

(<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>))

(<NVS> <--> (<VS-iru>))

(<VS-iru> <--> (<VC-iru>))

(<VS-iru> <--> (<ADV-s> <VS-iru>))

(<VS-iru> <--> (<VADV-H-COORD> <VS-iru>))

(<NVS> <--> (<NnVS-iru>))

(<NnVS-iru> <--> (<M-NN> <VS-iru>))

(<NnVS-iru> <--> (<M-NN> <NnVS-iru>))

(<NnVS-iru> <--> (<VADV-H-COORD> <NnVS-iru>))

(<NnVS-iru> <--> (<ADV-s> <NnVS-iru>))

(<NnVS-iru> <--> (<ADV-c> <NnVS-iru>))

(<VS-qn> <--> (<VC-qn>))

(<VS-qn> <--> (<ADV-s> <VS-qn>))

(<VS-qn> <--> (<VADV-H-COORD> <VS-qn>))

(<NVS> <--> (<NnVS-qn>))

(<NnVS-qn> <--> (<M-NN> <VS-qn>))

(<NnVS-qn> <--> (<M-NN-qn> <VS-qn>))

;;=====

;; 方向性がなく目的語を取らない述語を含む文

;;=====

(<SS> <--> (<NVS-DIR-OBJ>))
 (<NVS-DIR-OBJ> <--> (<VS-DIR-OBJ>))
 (<VS-DIR-OBJ> <--> (<ADJ-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-O
 BJ>))
 (<NnVS-DIR-OBJ> <--> (<M-NN-DIR-OBJ> <NnV
 S-DIR-OBJ>))

(<NVS-DIR-OBJ> <--> (<NnVS-DIR-OBJ-qn>))
 (<NnVS-DIR-OBJ-qn> <--> (<M-NN-DIR-OBJ-q
 n> <VS-qn>))

(<NVC> <--> (<NN-H> <NVC>))

(<ADJ-DIR-OBJ> <--> (<vaux-adj>))
 (<ADJ-DIR-OBJ> <--> (<vaux-adj-sfp>))
 (<ADJ-DIR-OBJ> <--> (<vaux-adj-sfpq>))
 (<DA-FORM> <--> (<NVC>))
 (<DA-FORM> <--> (<MOD-N> <NVC>))
 (<NnVC> <--> (<MOD-N> <NVC>))
 (<DA-FORM> <--> (<MOD-N> <NnVC>))

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

```

(<ADRE-VC> <--> (<ADRE-VC-NUM> <NVC>))
(<ADRE-VC-NUM> <--> (<ADRE1> <AVC-NUM>))
(<TEL-VC> <--> (<TEL1> <TELV C>))
(<TEL-VC> <--> (<TELV C>))

```

```

;;=====
;; 副詞連文節
;;=====

```

```

(<ADVPH> <--> (<ADV-s>))
      (<ADV-s> <--> (<ADV1>))
      (<ADV-s> <--> (<ADV1> <ADV-s>))
(<ADVPH> <--> (<ADV-c>))
      (<ADV-c> <--> (<VADVS>))
            (<VADVS> <--> (<VADV>))
            (<VADVS> <--> (<VADV-H-COORD>))
            (<VADVS> <--> (<VADV-H-O>))
            (<VADVS> <--> (<ADV-s> <VADVS>))
            (<VADVS> <--> (<ADV-c> <VADVS>))
            (<VADVS> <--> (<ADVPH> <VADVS>))
      (<ADV-c> <--> (<NnVADVS>))
            (<NnVADVS> <--> (<M-NN> <VADVS>))
            (<NnVADVS> <--> (<M-NN> <VADV-H>))
            (<NnVADVS> <--> (<M-NN> <VADV-H-iru>))
            (<NnVADVS> <--> (<MOD-N> <VADV-H-O>))
            (<NnVADVS> <--> (<M-NN> <VADV-H-O>))
            (<NnVADVS> <--> (<M-NN> <NnVADVS>))
            (<NnVADVS> <--> (<ADV-c> <NnVADVS>))

(<ADVPH> <--> (<ADV-c-iru>))
      (<ADV-c-iru> <--> (<VADVS-iru>))
            (<VADVS-iru> <--> (<VADV-iru>))
      (<ADV-c-iru> <--> (<NnVADVS-iru>))
            (<NnVADVS-iru> <--> (<M-NN> <VADVS-iru>))

```

)

(<NnVADVS-iru> <--> (<M-NN> <NnVADVS-iru
>))

(<ADVPH> <--> (<ADV-c-DIR-OBJ>))
(<ADV-c-DIR-OBJ> <--> (<VADVS-DIR-OBJ>))
(<VADVS-DIR-OBJ> <--> (<NKVADV>))
(<VADVS-DIR-OBJ> <--> (<MOD-N> <NKVA
DV>))

(<VADVS-DIR-OBJ> <--> (<ADV-s> <NKVA
DV>))

(<ADV-c-DIR-OBJ> <--> (<NnVADVS-DIR-OBJ>))
(<NnVADVS-DIR-OBJ> <--> (<MOD-N> <NK
VADV-KEI>))

(<NnVADVS-DIR-OBJ> <--> (<M-NN-DIR-O
BJ> <VADVS-DIR-OBJ>))

(<NnVADVS-DIR-OBJ> <--> (<ADV-s> <VA
DVS-DIR-OBJ>))

(<NnVADVS-DIR-OBJ> <--> (<ADV-s> <Nn
VADVS-DIR-OBJ>))

(<NnVADVS-DIR-OBJ> <--> (<ADV-c> <VA
DVS-DIR-OBJ>))

(<NnVADVS-DIR-OBJ> <--> (<M-NN-DIR-O
BJ> <NnVADVS-DIR-OBJ>))

(<VADV-H> <--> (<VADV-H-COORD>))

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

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

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

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

;;=====

;; 後置詞句連文節

;;=====

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

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

(<M-NN> <--> (<MOD-N> <M-NN>))

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

(<NQN> <--> (<QN>))

(<NQN> <--> (<NKQN>))

(<NQN> <--> (<MOD-N> <NKQN>))

(<NQN> <--> (<ADV-s> <NQN>))

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

(<NQN> <--> (<ADVPH> <NQN>))

(<M-NN> <--> (<NN-H> <NN>))

(<M-NN> <--> (<NN-H> <NN-H> <NN>))

(<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>))

;;=====

;; 連体修飾連文節

;;=====

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

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

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

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

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

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

(<NM> <--> (<NM-adv> <NPM>))

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

(<MOD-N> <--> (<ADV-s> <n-rentai>))

```

(<MOD-N> <--> (<MOD-N> <NM-KEI>))
(<MOD-N> <--> (<NPM>))
(<MOD-N> <--> (<NM> <NPM>))
(<MOD-N> <--> (<NVM-c>))
    (<NVM-c> <--> (<VMS>))
        (<VMS> <--> (<VM>))
            (<VMS-iru> <--> (<VM-iru>))
            (<VMS> <--> (<ADV-s> <VM>))
            (<VMS> <--> (<ADV-c> <VM>))
            (<VMS> <--> (<ADV-s> <VM-iru>))
        (<NVM-c> <--> (<NnVMS>))
            (<NnVMS> <--> (<M-NN> <VMS>))
            (<NnVMS> <--> (<M-NN> <VMS-iru>))
            (<NnVMS> <--> (<M-NN-qn> <VMS>))
            (<NnVMS> <--> (<ADV-s> <NnVMS>))
            (<NnVMS> <--> (<VADV-H-P-K> <VM-SA>))
        (<NVM-c> <--> (<NnnVMS>))
            (<NnnVMS> <--> (<M-NN> <NnVMS>))

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

```

```
;;=====
```

```
;; 文節グループ
```

```
;;=====
```

```
;;-----
```

```
;; 名詞文節
```

```
;;-----
```

```

(<NN> <--> (<np>))
(<NN-H> <--> (<np-h>))
(<NN> <--> (<MOD-N> <np-adv>))
(<NN> <--> (<NP-MONEY>))
(<NN> <--> (<wh-np>))

```

(<NN> <--> (<ka-douka>))

(<NN> <--> (<wh-n>))

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

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

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

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

(<NN> <--> (<NP-ADRE>))

(<NN> <--> (<NP-ADRE-O>))

(<NN> <--> (<NP-ADRE-E>))

(<NN> <--> (<NN-H> <CONJ-H> <NN>))

(<NN> <--> (<NN> <CONJ-H> <NN>))

(<NN> <--> (<NN-H> <NN>))

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

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

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

(<NN-DIR-OBJ> <--> (<NP-ADRE>))

(<NN-DIR-OBJ> <--> (<ka-douka>))

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

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

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

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

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

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

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

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

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

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

(<NKQN> <--> (<vaux-adj-sfpq>))

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

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

;;-----
 ;; 連用修飾文節
 ;;-----

(<ADV1> <--> (<adv-p>))
 (<ADV1> <--> (<adv-ph>))
 (<VADV> <--> (<adj-s>))
 (<VADV> <--> (<vaux-katei>))
 (<VADV> <--> (<vaux-s>))
 (<VADV-iru> <--> (<vaux-s-iru>))
 (<VADV> <--> (<vaux-s+f>))
 (<VADV-iru> <--> (<vaux-s+f-iru>))
 (<VADV-qn> <--> (<vaux-s+f-qn>))
 (<VADV> <--> (<vaux-h>))
 (<VADV> <--> (<vaux-ni>))
 (<VADV> <--> (<vaux-sa-s>))
 (<VADV> <--> (<vaux-sa-h>))
 (<VADV> <--> (<vaux-naru-s>))
 (<VADV> <--> (<vaux-naru-s+f>))
 (<VADV> <--> (<vaux-kdo-coord>))
 (<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-h>))
 (<VADV-H-COORD> <--> (<vaux-coord>))
 (<VADV-H-COORD> <--> (<vaux-sa-coord>))
 (<VADV-H-P> <--> (<n-ren-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>))

 (<VADV-H-P> <--> (<v-no-p>))
 (<VADV-H-P-iru> <--> (<v-no-p-iru>))

(<M-NN-H> <--> (<v-no-p>))
 (<M-NN-H> <--> (<v-no-p-o>))
 (<NnM-NN-H> <--> (<M-NN> <M-NN-H>))
 (<M-NN> <--> (<M-NN> <NnM-NN-H>))

(<NKVADV> <--> (<n-vaux-katei>))
 (<NKVADV> <--> (<n-vaux-s>))
 (<NKVADV> <--> (<n-vaux-s+f>))

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

(<CONJ1> <--> (<conj>))
 (<CONJ-H> <--> (<conj2>))
 (<NKVADV> <--> (<n-vaux-coord>))
 (<NKVADV> <--> (<N-VAUX-COORD-MONEY>))
 (<NKVADV> <--> (<n-vaux-h>))
 (<NKVADV> <--> (<N-VAUX-H-MONEY>))
 (<NKVADV> <--> (<N-VAUX-KATEI-MONEY>))
 (<NKVADV> <--> (<N-VAUX-S-MONEY>))
 (<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>))

;;-----
 ;; 述語文節
 ;;-----

(<VC> <--> (<vaux>))
 (<VC> <--> (<vaux-fin>))
 (<VC-iru> <--> (<vaux-iru>))
 (<VC-qn> <--> (<vaux-qn>))

(<VC> <--> (<vaux-cop>))
 (<VC-iru> <--> (<vaux-cop-iru>))
 (<VC> <--> (<vaux-s+f>))
 (<VC-iru> <--> (<vaux-s+f-iru>))
 (<VC-qn> <--> (<vaux-s+f-qn>))
 (<VC> <--> (<vaux-sfp>))
 (<VC-iru> <--> (<vaux-sfp-iru>))
 (<VC-qn> <--> (<vaux-sfp-qn>))
 (<VC> <--> (<vaux-sa>))
 (<VC> <--> (<vaux-sa-s+f>))
 (<VC> <--> (<vaux-sa-sfp>))
 (<VC> <--> (<vaux-sa-cop>))
 (<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-cop>))
 (<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-s+f>))
 (<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux-sfp>))
 (<VS-KEI-DIR-OBJ> <--> (<n-keisiki-vaux>))
 (<VS-SA> <--> (<vaux-sa>))
 (<VS-SA> <--> (<vaux-sa-s+f>))
 (<VS-SA> <--> (<vaux-sa-sfp>))
 (<VS-SA> <--> (<vaux-sa-cop>))

;;-----

;; 連体修飾文節

; ; -----

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

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

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

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

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

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

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

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

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

(<VS-ka> <--> (<VC-ka>))

(<NnVS-ka> <--> (<M-NN> <VS-ka>))

(<NN> <--> (<NnVS-ka> <conj2> <NnVS-ka>))

(<ka-douka> <--> (<M-NN> <ka-douka>))

(<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>))

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

(<NN-DIR-OBJ> <--> (<NM-MONEY> <NP-MONEY>))

(<NM-TAG> <--> (<np-no-tag>))

(<NM> <--> (<NM-TAG>))

(<NM> <--> (<NM-MONATE> <NM-TAG>))

(<NM-adv> <--> (<np-adv-no>))

;;=====

;; 住所

;;=====

(<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>))

(<ADRE1> <--> (<ADRE11> <ADRE12> <ADRE14> <ADRE15>))

(<ADRE11> <--> (<adre-ken>))

(<ADRE12> <--> (<adre-si>))

(<ADRE13> <--> (<adre-ku>))

(<ADRE14> <--> (<adre-tyo>))

(<ADRE15> <--> (<adre-no>))

(<ADRE16> <--> (<adre-num-ch>))

(<ADRE17> <--> (<adre-num-ban>))

(<NP-ADRE> <--> (<ADRE1> <adre-num-np>))

(<AVC> <--> (<adre-vaux-cop>))

(<AVC> <--> (<adre-vaux-cop-sfp>))

(<AVC-NUM> <--> (<adre-num>))

;;=====

;; 電話番号

;;=====

(<TELVC> <--> (<tel-vaux-cop>))

(<TELVC> <--> (<tel-vaux-cop-sfp>))

(<TEL1> <--> (<tel-no>))

;;=====

;; 金額

;;=====

(<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> <--> (<HYAKU> <N-VAUX-S-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>))

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

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

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

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

;;=====

;; 例外

;;=====

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

付録 B 形態素調整ルール

音声認識部と構文解析部は、それぞれ独立に開発されてきたので、語彙の区切り方で食い違っているものがある。そこで、形態素レベルでインターフェースを行う場合、音声認識部の出力ラティスの語彙区切りを調整して、構文解析部が受け入れられるようにする必要がある。形態素調整ルールはそのための規則で、次のファイルに記述されている。

`vocabsort.d` : 音声認識結果のローマ字音韻表記を仮名漢字表記に変換するための辞書。音声認識部の品詞名、ローマ字表記、仮名漢字表記、データベースの形態素名の順で記述されている。

例: `adre-si-si seikachou` 精華町 固名詞
 `adv-b choQto` ちょっと 副詞
 `n-hutu iiN` 委員 普名詞
 `v-5dan-r oku` 送 動詞語幹

注意: 文節内文法では、「ん」の音を "n"、「っ」の音を "q" と表記しているが、音声認識の出力結果はそれぞれ "N" "Q" と表記してある。したがって、`vocabsort.d` の表記も "N" "Q" と変更しなければならない。

音声認識のための文節内文法の辞書から自動的に生成するプログラムがある。

`token.d` : `vocabsort.d` に記述されているデータベースの形態素名を構文解析部の品詞名に変換するための規則。

例: `%token` 普名詞 `n-com`
 `%token` 固名詞 `n-prop`
 `%token` 副詞 `adv`
 `%token` 動詞語幹 `vstem`

注意: データベースの形態素性名と構文解析部の品詞名

は 一 対 一 に 対 応 す る わ け で は な い の で 、 変 換 さ れ
た 品 詞 名 を そ の ま ま 構 文 解 析 に 使 う こ と は で き な
い。

rule1.d : 音 声 認 識 部 の 品 詞 名 を 、 vocabsort.d に 記 述 さ れ て い る デー
タ ベース の 形 態 素 名 に 変 換 す る た め の ルール。

例 : ({ 固 名 詞 } <= { nil adre-si-si })
({ 普 名 詞 } <= { nil n-hutu })
({ 副 詞 } <= { nil adv-b })
({ 動 詞 語 幹 } <= { nil v-5dan-r })

rule2.d : 活 用 語 の 活 用 ルール。 現 在 は 使 っ て い な い。

rule3.d : 構 文 解 析 部 で は 一 語 登 録 さ れ て い る が 、 音 声 認 識 部 で は 分 割
さ れ て い る 語 を 、 構 文 解 析 部 に 合 わ せ て ひ と つ に ま と め る た
め の ルール。

例 : ({ 接 続 詞 } <= { " そ れ " pro }
+ { " で " p-kaku-de }
+ { " は " p-k-wa })

({ 補 動 詞 } <= { " な " negt-kyo }
+ { " け れ " flex-kyo-katei }
+ { " ば " sp-katei-ba }
+ { " な " v-5dan-r }
+ { " り " flex-5-ri }
+ { " ま せ " masu-mizen }
+ { " ん " nai-n-rentai })

rule4.d : 構 文 解 析 部 で は 分 割 さ れ て い る が 、 音 声 認 識 部 で は 一 語 登 録
さ れ て い る 語 を 、 構 文 解 析 部 に 合 わ せ て 分 割 す る た め の ルー
ル。

例： ({"お名前" n-hutu} => {"お" 接頭語}
+ {"名前" 普名詞})

({"せる" caus-seru-rentai} => {"せ" 助動詞}
+ {"る" 活語尾}))

付録 C 文節区切りの規則

音声認識のための構文規則は、『形態素情報利用解説書』に規定された文節を想定して組み立てられている。しかし、もっぱら音声認識という観点から、それとは異なる文節区切りをしている規則もある。

さらに、Mセットのモデル会話文およびEセットのモデル文の音声データには、文節の区切り方に若干の揺れがある。

ここでは、『形態素情報利用解説書』に規定されている文節区切りの基本原則を要約し、音声認識のための構文規則でそれとは異なっている規則について述べる。さらに、発話の際に誤りやすい点について述べる。

データベースの文節区切りの原則は、「形態素情報利用解説書」の「付録4」に記述されている。それによると、文節には次の3通りの組み合わせがある。

1. 接頭辞* + 自立語 + 接尾辞* + 付属語*

例：お 名前 は

2. 接頭辞* + 接尾辞* + 付属語*

例：何 人 くらい

3. 接頭辞* + 普通名詞* + 接尾辞* + 付属語*

例：口座 番号 は

注1：「*」は0個以上複数の連続を許すことを表している。

自立語とは以下の品詞を指す。

名詞（固有名詞、普通名詞、サ変名詞、形容名詞、数詞、代名詞）

本動詞

形容詞

副詞

連体詞

接続詞

感動詞

間投詞

付属語とは以下の品詞を指す。

補助動詞

助動詞

助詞（格助詞、係助詞、副助詞、並立助詞、接続助詞、終助詞、準体助詞）

注2：複合語や数詞を含む表現には、この規則には当てはまらないが、一つの文節にするものがある。

5 万 円 前後 に
数詞 接尾 普名 格助

第 一 文 科 会 の
接頭 数詞 普名 普名 格助

3 0 0 人 収容 の
数詞 接尾 普名 格助

4 、 5 0 名
数詞 記号 数詞 接尾

予定 参加 者 数 は
普名 普名 接尾 普名 係助

出国 申請 書 作成 に
普名 普名 接尾 普名 格助

次の表現は2文節とする。

お ひと り / 5 万 円 と
接頭 数詞 接尾 数詞 接尾 格助

約 / 5 0 0 名 の
副詞 数詞 接尾 格助

第 二 、 / 第 三 セッション の
接頭 数詞 記号 接頭 数詞 普名 格助

(「4、50名」と違って、「第二セッションと第三セッション」の意味なので)

音声認識のための構文規則もおおむね以上の基準にしたがって文節区切りをしているが、例外的に、次の表現はひとつの文節としている。

1. 格助詞相当語を含む表現。

会議 に 関する
普名 格助 動詞

天候 に よっ て は
普名 格助 動詞 接続助 係助詞

2. 「かどうか」を含む表現

出席 できる か どう か
サ名 補動詞 取立助 副詞 取立助

参加 される か どう か は
サ名 助動詞 取立助 副詞 取立助 係助

3. 断定の助動詞「で」＋係助詞「は」に、否定の意味の動詞「ございません」や形容詞「ない」が続く表現。

そう で は ございませ ん
副詞 助動詞 係助 本動詞 助動詞

そう で は ない と
副詞 助動詞 係助 形容 格助

4. 「かもしれません」は、全体を助動詞として一語登録してある。したがって、「…かも／しれません」と区切らずに、一文節とする。

思わ れる か も しれ ませ ん

本動詞 助動詞 係助 係助 本動詞 助動詞 助動詞

5. 「どうしても」は、全体を副詞として一語登録してある。したがって、「どう／しても」と区切らずに、一文節とする。

どう し て も
副詞 本動詞 接続助 係助詞

音声データの収録の際に誤りやすかったものに、形式名詞や副詞的名詞を含む表現、「て」にアスペクトや授受を表す補助動詞が続く表現、禁止・許可を表すムード表現、ある種の連体詞を含む表現があった。これらは、次のように文節区切りをする。

1. 形式名詞は自立語とみなしている。したがって、連体修飾する語句の次で文節区切りをする。

記入 し た / はず です
サ名 補動詞 助動詞 形式名 助動詞

登録 し て おく / こと は
サ名 補動詞 接続助 補動詞 形式名 係助

2. 副詞的用法を持つ名詞もそれだけで一文節をつくる。

依頼 する / ため
サ名 補動詞 普名詞

送っ て くれる / よう に
本動詞 接続助 補動詞 普名詞 格助

3. 接続助詞「て」に、アスペクトを表す「いる」や、授受を表す「いただく」などが続く表現は、全体を一文節とする。また、「て」に「ほしい」が続く表現も一文節とする。

受け付け て い ます

本動詞 接続助 補動詞 助動詞

送っ て くれる / よう に
本動詞 接続助 補動詞 普名詞 格助

送っ て ほしい の です が
本動詞 接続助 助動詞 準体助 助動詞 接続助

4. 「… なければなりません」「… てもよい」など、禁止・許可などを表す表現は、全体でひとつの意味を表すので一文節としやすいが、二つの文節に区切る。

送ら なけれ ば / なり ませ ん
本動詞 助動詞 接続助 本動詞 助動詞 助動詞

お 送り すれ ば / よい です か
接頭 本動詞 助動詞 接続助 形容詞 助動詞 終助

5. 「この」「その」「どの」などの連体詞は自立語なので、それだけで一文節となりうる。

この / 場 で は
連体詞 普名 格助 係助

その / テーマ に つい て
連体詞 普名詞 格助 本動詞 接続助

参考文献

- [1] 保坂, 竹沢: S L - T R A N S における音声認識のための構文規則の概要,
A T R テクニカルレポート T R - I - 0 1 9 3 (1 9 9 1)
- [2] 永田, 衛藤, 保坂: 音声認識のための構文規則 (ガイドブック),
A T R テクニカルレポート T R - I - 0 2 4 0 (1 9 9 2)
- [3] 保坂, 竹沢: S L - T R A N S における音声認識のための構文規則の概要の
拡張, A T R テクニカルレポート T R - I - 0 2 4 1 (1 9 9 2)
- [4] 北: H M M - L R ユーザーズ・マニュアル, A T R テクニカルレポート
T R - I - 0 2 4 6 (1 9 9 2)
- [5] 篠崎, 水野, 小倉, 吉本: 形態素情報利用解説書, A T R テクニカルレポー
ト T R - I - 0 0 7 7 (1 9 8 9)