

TR-IT-0080

部分木を単位とする
音声言語解析のための言語現象の調査
An Examination of Linguistic Phenomena Based on
Partial Trees for Spoken Language Analysis by Machines

竹沢 寿幸 田代 敏久 衛藤 純司†
Toshiyuki TAKEZAWA Toshihisa TASHIRO Junji ETOH†

1994. 11

内容梗概

現在収録を進めている音声言語データベースの資料を利用して、音声言語解析研究の基礎となる言語現象の調査を行なった。「不適格文」、「ポーズ」、「並列句(「～から～まで」を含む)」、「複合語(名詞連続句を含む)」を中心に、部分木を単位とした分類調査を行なった。

ATR 音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories
†日本アイアール株式会社

© 株式会社 エイ・ティ・アール音声翻訳通信研究所
© 1994 by ATR Interpreting Telecommunications Research Laboratories

目次

1	まえがき	1
2	不適格文の調査	1
2.1	不適格文の定義	1
2.2	データの構成	2
2.3	形態素レベルの不適格文	2
2.3.1	言い淀み、その他の間投詞	2
2.3.2	言い直し	4
2.3.3	言い誤り	4
2.3.4	つづり	4
2.4	統語レベルの不適格文	4
2.4.1	格助詞の欠落	4
2.4.2	提題助詞の欠落	5
2.4.3	連体助詞の欠落	5
2.4.4	直示	5
2.4.5	助動詞の欠落	6
2.4.6	述語の欠落	6
2.4.7	サ変名詞述語と形容名詞述語	7
2.4.8	後置詞句・副詞句の名詞修飾	8
2.4.9	名詞句に接続する終助詞	8
2.4.10	同格	9
2.4.11	言い換え・繰り返し	9
2.4.12	語順の交替・倒置	10
2.4.13	挿入句	11
2.4.14	呼びかけ	11
2.4.15	断片文	11
2.4.16	中止文(主節の欠落・接続助詞終止)	11
2.4.17	連続文	12
2.4.18	特殊構文(箇条発話)	12
2.5	意味レベルの不適格文	12
2.5.1	格助詞の誤り	12
2.5.2	必須格の欠落	13
2.5.3	だ型文	13
2.5.4	終助詞の欠落	14
2.5.5	節関係の不整合	14
2.5.6	融合文	14
2.6	語用論レベルの不適格文	14
3	ポーズの調査	15
3.1	データの構成	15
3.2	ポーズの問題点と処理	15

4	並列句の調査	18
4.1	データの構成	18
4.2	並列句のタイプ	19
4.2.1	名詞句+名詞句+	19
4.2.2	名詞句+並立助詞/接続詞+名詞句 (+並立助詞)	19
4.2.3	名詞句+並立助詞/接続詞+動詞句	20
4.2.4	動詞句+動詞句	20
4.2.5	動詞句+並立助詞/接続詞+動詞句 (+並立助詞)	20
4.2.6	その他	21
4.3	「～から～まで」の調査	21
5	複合語・数量詞の調査	22
5.1	データの構成	22
5.2	複合語・数量詞のタイプ	22
5.2.1	名詞+接尾辞	22
5.2.2	形容詞+接尾辞	22
5.2.3	接頭辞+名詞	23
5.2.4	接頭辞+動詞	23
5.2.5	接頭辞+形容詞	23
5.2.6	名詞+名詞	23
5.2.7	動詞+動詞	23
5.2.8	複合語+接尾辞	23
5.2.9	接頭辞+複合語	24
5.2.10	名詞+複合語	24
5.2.11	複合語+名詞	24
5.2.12	複合語+副詞	24
5.2.13	複合語+副助詞	24
5.2.14	その他	24
5.3	名詞連続句	25
	謝辞	25
	参考文献	25
A	付録 1: 不適格文のタイプ別頻度	26
B	付録 2: 並列句のタイプ別頻度	28
C	付録 3: 複合語のタイプ別頻度	29
D	付録 4: 調査対象ファイル	31
D.1	不適格文の調査・並列句の調査・複合語の調査	31
D.2	ポーズの調査	31
E	付録 5: 調査結果ファイル	31

1 まえがき

日本語の音声認識は、文節を単位とする発話から、文を単位とするより自然な発話に研究対象を移しつつある。1993年度に、我々は文を単位とする発話を処理することができるような小規模な文法を開発した[4]。これは、

- (1) 文節の代わりに、ポーズ節(ポーズによって区切られたまとまり)を基本にする。
- (2) 音声認識用文法と言語解析用文法を統合するために一般的な句構造文法にする。

という2つの方針に基づいていた。

この文法を開発する過程で、我々はポーズ節という概念そのものがはらむ問題や、自然な発話にしばしば出現する文法的に不適格な文をどう処理するかという問題に直面することになった。

そこで、あらためて音声認識用文法を再考することにし、そのために、ATRで収録されている音声言語データベースを対象に、不適格な文やポーズについて調査することにした。さらに、音声翻訳システム ASURA 開発の時点から懸案になっていた並列句や複合語の調査もすることにした。

したがって、本報告書は、次の項目から成る。

- 不適格文の調査
- ポーズの調査
- 並列句の調査(付:「～から～まで」の調査)
- 複合語の調査(付:名詞連続句の調査)

2 不適格文の調査

これまでにATRで開発されてきた音声認識用文法および言語解析用文法は、すべて適格な文(well-formed sentence)を対象としてきた。しかし、自然な発話ではしばしば文法的に不適格な文(ill-formed sentence)が出現する。したがって、自然な発話を対象とする対話処理システムは不適格な文も処理できるような頑健(robust)なものでなければならない。

2.1 不適格文の定義

人間が自分の直感に照らしあわせて文法的な制約を逸脱していると判断する文がある。一方、たまたまシステムが持っている辞書や文法が限られていて処理できない文がある。前者は誰もが不適格と認めるという意味で絶対的な不適格文であり、後者はシステムが達成している文法の水準から見て不適格という意味で相対的な不適格文である[1]。

本調査では、対話処理システムを構築するという立場から、絶対的な不適格文だけでなく、相対的な不適格文も対象に含めた。ただし、その際に基準となるシステムはおおよそ昨年構築した音声認識用文法[4]を想定している。厳密に言うなら、音声認識用文法を実際に使用して適格・不適格を判断しなければ相対的な不適格文という定義に合わないことになるが、500語足らずの辞書と1,000個余りの規則しかない現在の文法ではほとんどの文が不適格ということになってしまいうだろう。そこで、ここでは現在の文法を自然に拡張したものを仮に想定して、その文法を基準に適格・不適格を判断した。

2.2 データの構成

データ例および各項目の説明をする。

会話番号：TBS12003

発話番号：110

不適格文：|例えば|自動車|で|旅行|で|す|とか|、|長期|滞在|で|す|
|とか|。

レベル：統語

タイプ：文末述語の欠落

適格文：|例えば|自動車|で|旅行|で|す|とか|、|長期|滞在|で|す|
|とか|、|有|り|ま|す|か|。

文脈：1文前

会話番号：音声言語データベースの会話番号

発話番号：音声言語データベースの発話番号

不適格文：何らかの不適格性を含む文(「|」は音声言語データベースの形態素解析)

レベル：不適格性が形態素レベル、統語レベル、意味レベル、語用論レベルのいずれであるかを記述

タイプ：不適格性の種類

適格文：不適格な部分を修正して適格文にしたもの

文脈：適格文に修正するに際して何文前を参照したかを記述

1つの文に不適格な部分が2つ以上ある場合は、それぞれの不適格性について別々のデータを作成した。

2.3 形態素レベルの不適格文

2.3.1 言い淀み、その他の間投詞

田窪行則は文献 [2] の中で、感動詞を言い淀み系と入出力制御系に分類し、それらをさらに次のように分類している。

● 言い淀み系

非語彙的形式：え、ええ、単語末母音の長音化

語彙的形式：

内容計算：ええ、(っ)と、ううんと

形式検索：あの(ー)、その(ー)、この(ー)

評価：ま(あ)、なんというか、なんか、やっぱり

● 入出力制御系

応答系：ああ、はい、ええ、うん、んん、ふん(下降イントネーション)

問返し系：は、はあ、え、ええ、へえ、ふん(上昇イントネーション)

驚き：あっ、えっ、はっ、ふんっ

迷い：ううん(高平長)

納得：ふうん、へええ、ええ(低平長)

本調査でも、間投詞を分類するに当たって、これを参考にした。ただし、「言い淀み系」は上記のように細分化せず、一括して「言い淀み」とタイプづけしている。

間投詞のうちもっとも多かったのは、「言い淀み」、すなわち、言おうとしてためらったり、言葉に詰まったりしたときに発する声であった。全部で336例、そのうち文頭にあるものが265例、文中にあるものが71例あった。

[えー] わたくし一人だけです。
近代美術館のすぐ隣りに[あの] になりますので。

間投詞の中には、「気づき」と名づけたものがあった。これは、話者が自分で何かに気づいたときに思わず発する声で、次のようなものである。

[あっ] チケットの受け取りの方法はどうなってるんです。
次の列車までは... [あ] 四分しかないですね。

「気づき」とは違って、相手の発話に対して驚いたときに発する声がある。「驚き」と名づけておいた。

[えっ] 九時からですか。

以上のような「言い淀み」とも「気づき」とも「驚き」とも違うもので、自分の感情を表現する発話に伴って発する声がある。例として次のような「いやー」「いや」のみであるが、「感動」と名づけておいた。

[いやー] 助かりましたよ。

相手の発話に対して「納得」したときに発する声がある。

[ふーん] 値段も予算内だし、良さそうだな。

どう判断したり、答えたりしたらいいか分からなくて、考え込んでいるときなどに発する声がある。次のような例で、「迷い」と名づけておいた。

[うーん] じゃあ四日のビー席でいいです。

最後に、相手の発話に対する「応答」と言ってよいものがある。これには、「はい」のように、現在の文法で感動詞として語彙登録しているものに近いものも含まれているが、[] でくくられているので、間投詞としておいた。

[ああ] 八時からですか。
[ええ] フォードアの小型車だったら何でもいいんですけど。
[うん] 今頃は一人なんだけど、あとで妻と合流して、そのあとに観光に行こうっていう感じなんだ。
[はい] ホテルの紹介を幾つかしてほしいんです。

2.3.2 言い直し

言い間違っ、あらためて言い直すケースである。今回の調査では、次の2例のみであった。

それと、そのホテルの電話番号を教えて(いたた)いただけますか。
九時十五分のバスに乗られましたら、(は)夜の八時までには着きますが。

2.3.3 言い誤り

単語の言い誤りである。次の例のみである。

はい、お使えになります。

「お使いになれます」が正しい。「可能」の意味を表すのに、「使う」の可能動詞「使える」を用いたために、このような発話になったのであろう。

2.3.4 つづり

ここで「つづり」と言っているのは、電話で名前のつづりを説明するときに使うアルファベットのことである。文法的には不適格ではないが、現在の文法にはこのような辞書項目はないので、相対的不適格文として調査に加えておいた。

スペルの方は、エムエーアールワイ、ピーエイチアイエルエルアイピーエスです。

2.4 統語レベルの不適格文

2.4.1 格助詞の欠落

何らかの格助詞が欠落した文は、話し言葉ではしばしば出現する。音声言語データベースでは、「が」「を」「に」「まで」「につき」の5つの格助詞が欠落しているケースがあった。そのうち、最も多いのは「を」で53例、その次が「が」で14例、「に」「まで」「につき」は少なく、それぞれ2例、4例、2例であった。

なにかお好みの条件(が)有りますか。
お名前とお電話番号のほう(を)、頂けますでしょうか。
そのほか(に)なにかございますか。
九時から五時(まで)、毎日営業しております。
一日(につき)、料金はどのぐらいですか。

2.4.2 提題助詞の欠落

提題助詞の「は」が欠落していると考えられるものである。今回の調査では27例あった。

わたくし(は)、担当の鈴木と申します。
朝食の予約(は)、できますか。

2.4.3 連体助詞の欠落

連体助詞の「の」が欠落していると考えられるものである。ただし、これらの文は、連体助詞の「の」を補うと統語的にも意味的にも現在の文法で解析できるという理由で、このように考えたものである。

それで、お席の方は普通席とグリーン席(の)、どちらになさいますか。
九月十日(の)、正午ぐらいに京都に着きたいんですが。

2.4.4 直示

名詞句が助詞を伴わずに出現しているものであるが、これまでの3つのタイプのように「助詞の欠落」と考えることができないものである。話者が聴者とのあいだに共有している知識を直接指し示していると考えて、「直示」と名づけておいた。

ご予約どうもありがとうございました。
じゃ、伏見桃山宇治コースのほう、予約をお願いいたします。

同じ「直示」でも、次のように、すぐ後に代名詞が続いているものがある。これは、変形文法で「左方転位」として分析されていたものである。

最後のニューヨークシティホテル、その電話番号を教えてくださいませんか。
ええ、できたらマツダのジオメトロ、それが欲しいんだけど、有ります。

音声言語データベースには、以上の2つのタイプがある。

2.4.5 助動詞の欠落

音声言語データベースに出現するのは、コンピュータの助動詞「だ・です」が欠落しているものがほとんどである。

年齢は四十歳(です)。
ファーストクラスは七十五ドル(で)、普通は五十ドルです。

後者は、「ファーストクラスは七十五ドル」と「普通は五十ドル」が並列していて、文末の「です」で包括されると考えることもできるが、ここでは上記のように「で」が欠落していると考えた。

次の例は「わたくし」と「鈴木」が同格関係にあるとも考えられるが、ここでは「担当はわたくしです」と「わたくしは鈴木と申します」という2つの文が融合したと考え、助動詞の「です」が欠落したと考えた。

担当はわたくし(で)、鈴木と申します。

また、次の例は、「朝食は洋食」が「二泊とも～でよろしいですね」の補文であると考え、助動詞の「だ」が欠落していると考えた。しかし、「だ」を補うと、文としてかえって不自然になる。

二泊とも、朝食は洋食(だ)ということでもよろしいですね。

2.4.6 述語の欠落

まず、文末の述語が欠落している文がしばしば出現するが、その多くは、次のような会話に多用される慣用句と言っているものであった。

お気をつけて。
ご用件をどうぞ。
どうぞこちらへ。
どうもご親切に。
どうぞ良いご旅行を。
いつでも、どうぞ。

その他は、何らかの方法で述語を復元しなければならないものである。そのうち、前の文脈を参照して述語を復元できる例として、次のようなものがあった。

なにかそのほかにございますか。
自動車旅行ですとか、長期滞在ですとか(ございますか)。

それでしたら、四つのコースがございます。
五千円の三時間のコースが二コース (ございます)。

[あの] きょうは何時まで開いていますか。
残念ですが、きょうはお休みなんです。
水曜日ですから。
[あ] 水曜日がお休みなんですか。
ええ、あしたは十一時から九時まで (開いています)。

二枚ございますので、ご確認ください。
はい、確かに (ございます)。

次の例は、前の文脈を参照しても復元できないものである。

それで、ご予算の方は (いくらですか)。

以上のように文末の述語ではなく、文中の述語が欠落していると考えられるものが少数あった。

周りの景色を見ながらでしたら、三日ほど (掛かると) お考えください。
消費税が三パーセント (掛かり)、それと一泊一万五千元以上の場合には三パーセントの特別地方消費税というのも掛かります。
東大寺に (行くに) は近鉄奈良駅からバスで十分、それから歩いて三分ですね。

2.4.7 サ変名詞述語と形容名詞述語

サ変名詞は「する」「なさる」などの補助動詞を伴って述語となるが、その補助動詞が欠落していて、単独で文の述語の役割を果たしているものがある。

八月十日の二〇一便を キャンセル。
ニューヨークへはいつご 宿泊 の予定でしょうか。

動詞性を持つ接尾辞からなる語が述語となる次のような文も、これと同じ現象である。

この便ですと、ジェーエフケー空港を十五時 発 になります。

これらは、補語や副詞がサ変名詞に係るような規則を作れば処理できるようになる。音声言語データベースの構文木データではそのようにしている。

形容名詞は「だ」「です」という助動詞を伴って述語となるが、その助動詞が欠落しているものが、次の1例だけあった。

これは好みの問題ですが、一番乗り継ぎが楽という点ではバスはいかがですか。

これも、補語が形容名詞に係るような規則を作れば処理できるようになる。

2.4.8 後置詞句・副詞句の名詞修飾

後置詞句と副詞句は、いずれも述語を修飾するはずのものである。しかし、次の例のように直後の名詞を修飾するものがある。

- 後置詞句

午前十時から午後四時までは十五分に一本出ています。
六時間で六千九百円という車が有りますが、それでよろしいですか。
車はガソリンを満タンでお返してください。

- 副詞句

もう一つはツインのお部屋をお一人でお使いいただいて、一万四千元というのが有ります。
[そうですね] 歩いて三十分っていうところですかね。

例文を見れば分かるように、いずれも修飾される名詞は数量を表すものである。数量表現の名詞は、副詞的な用法もあるように、他の名詞とは異なる性質を持っている。

次の文は、「禁煙席で」が連用形で、それが数量表現ならざる普通名詞を修飾している例である。

じゃあ、禁煙席で窓側はどうでしょうか。

2.4.9 名詞句に接続する終助詞

終助詞の「か」が名詞に接続した次のような文は、完全に適格な文であり、話し言葉ではごくありふれた表現である。しかし、現在の文法では終助詞は述語にしか接続しないことになっており、もし、名詞に接続するような規則を作ると、音声認識では曖昧性が増える原因になる。

[あー] あそこか。
[えー] 値段は一泊百ドルぐらいかな。

2.4.10 同格

次のような同格表現も完全に適格な文であるが、現在の文法には実装していない。このような名詞が連続する句を規則にすると、やはり曖昧性が増える。

それでは、あした十日 お願いします。
いえ、すいません、「えーと」 あとに出るやつ、「のぞみ」 お願いします。
変更手続きのほうは、わたくし鈴木 が承りました。

次の例は代名詞「これ」の特殊な用法である。

一万八千円 っていうのは これ、合計でこうなるんですか。

「一万八千円っていうのは、これは、合計でこうなるんですか」と言い換えることができるだろうし、そうすると2つの後置詞句が同格関係にあると考えることができる。

ちなみに、『大辞林』には、「漢文における「是」「之」「惟」などの訓読みから生じた文語的な言い方」として、次のような例文を挙げている。

人間は、これ 本来無一物である。
思想および良心の自由は、これを 侵してはならない。

2.4.11 言い換え・繰り返し

「言い換え」というのは、一度言ったことを表現を換えて言い換えたものである。「繰り返し」というのは、言いさして中断し、改めて言い直したものである。

- 言い換え

市街地、交通の便利 なところがいいですね。
東大寺へ ですけども、近鉄奈良駅 からバスで十分、その後 歩いて徒歩 三分です。

- 繰り返し

(十時) 朝の十時 からですんで、夜の七時、これで九時間ですか。
基本料金ですと、(一日) 一日 につき三十六ドルになります。

2.4.12 語順の交替・倒置

日本語では、文の諸要素の語順は比較的自由である。したがって、語順の適格・不適格という判断は相対的なものである。ここでは、標準的な語順から逸脱して異和感のあるものというぐらいの基準で集めておいた。

主題は文頭に来るのが普通であるが、次の文では文の途中にある。

たくさん フェリーは 出てるんですか。
何時ごろ 京都には いらっしゃいますか。

補語と修飾語、補語同士、修飾語同士の前後関係は自由度が高い。次の文は、修飾・被修飾関係の緊密な語同士が離れているので異和感が強い。

外は とっても お天気も良さそうなので、きっと 歩いても楽しいと思いますよ。
なにかお役に ほかに 立つことはございますか。
はい、興福寺ですと近鉄奈良駅前から 約 徒歩で五分ぐらいになります。

次の文は、主節が従属節によって分断されているものである。命題関係の解析で曖昧性が生じるだろう。

お値段のほうが お一人様ですので、七千三百円になります。
もしできましたら、何か連絡の必要があるかもしれませんので、お電話番号のほう、お願いいたします。
祇園周辺で、一泊 できましたら 三万円までで泊まりたいんですが。

最後の文は、「三万円まで」というのを特に希望していて、それを強調するために「できましたら」を直前に置いていると考えられる。したがって、意味的にはその方が適格だという気がする。

「述語」は文末に来るのが普通であるが、次の文では述語の後ろに修飾語が来ている。このような例は特に「倒置」としておいた。

ありがとうございました、どうも。
よく分かんないんですけども、二万から三万円くらいですかね、一人一泊で。
午後のコースのほうがいいですね、金閣寺には前に行ったことがありますんで。

2.4.13 挿入句

文の途中にあって、その文とは別の次元から導入された表現である。

特に無いんですけども、申し訳無い、ニューヨークシティホテルの電話番号だけ、もう一度お願いできませんか。
でも一人当たり八千円の予算と思ってたんで、竹ですか、それをお願いします。

後の文は、「竹ですか」を「それ」が受けて、文の構造の中に自然にはめこまれている。その意味ではまったく別の次元とも言えない。

2.4.14 呼びかけ

対話の相手に直接呼びかける語で、これも文とは別次元の表現である。音声言語データベースでは、相手の名前を呼びかける例がほとんどであった。

鈴木和子様、スペルをお願いいたします。
はい、ジョン・フィリップス様、当日は四名様でお越しですね。
それでは早速案内図を送りましょう、フィリップスさん。
お客様、申し訳ありません、到着の時間を教えていただけますか。

2.4.15 断片文

文としての結構を持たない、断片的な表現だけからなる発話である。

じゃあ、また。
でもなあ。
それでは、鈴木和子様。
八月十一日の二〇七便。

2.4.16 中止文(主節の欠落・接続助詞終止)

接続助詞を伴う述語で終わる文、すなわち従属節だけの文である。これには、前の文脈から主節を補うことのできるものと、そうでないものがある。後者は、接続助詞が語調を和らげるために用いられていると考えてよいもので、文の意味としてはそれで完結している。それぞれ「主節の欠落」「接続助詞終止」と区別した。

- 主節の欠落

チケット持ってるんですけども、都合が悪くて(、行けないんです)。
水曜日ですから(、お休みです)。
ホテルのわきからすぐ乗れますので(、バスをお薦めします)。
いいえ、旅行ですので(、急いではいません)。

- 接続助詞終止

ご予約のほう、確認いたしますので。
いずれも税、サービス料は別になりますので。
はい、二日前まででしたら、キャンセルできますので。

2.4.17 連続文

2つの文が、休止なしに1度に発話されたものである。

そうですか、鴨川屋ですか。
分かりました、それはすごく良さそうですね。

2.4.18 特殊構文(箇条発話)

ホテルや列車などの予約の際に、予約申込用紙の記入項目に従って読み上げるものである。「箇条書き」にならって「箇条発話」と名づけておいた。いわば外部の状況がそのまま言語に反映されているといったもので、規則化するのはむずかしい。

鈴木和子様、八月の十日から十二日まで、シングルルームシャワー付き二泊
ですね。
九月十日、土曜日、二名様、「のぞみ九号」博多行き、東京から京都まで、
禁煙席ですね。

2.5 意味レベルの不適合文

2.5.1 格助詞の誤り

格助詞の使い方が間違っているものである。動詞が要求する格助詞とは違った助詞が使われている。

はい、いつが(を)ご希望でしょうか。
その京都国際観光旅館と旅館金閣と、どちらを(に)先に泊まりますか。
パワリー通りはユニオンスクエアに(の)東側に在りまして、一番近いと思
いますよ。

2.5.2 必須格の欠落

日本語ではしばしば必須格が省略される。特に、話者と聴者が必須格になる場合はほとんど常に省略されると言ってもよい。

(こちらは) ニューワシントンホテルでございます。
(私は) 部屋の予約をしたいと思って (あなたに) 電話したんですけど。
:
それと (私は) (あなたから) 滞在期間もお知らせ願いたいんですけども。

しかし、本調査ではこのようなケースは取り上げなかった。取り上げたのは、話者と聴者以外の必須格の場合である。

チケット持ってるんですけども、(便を) 変更したいんですけども。
それからお水取りは無料で、(拝観日が) 三月十三日、翌日ですね。
まず、東大寺ですが大仏殿は (拝観料が) 五百円、八時から十七時までが拝観時間です。
五月三日は (チケットが) ありますでしょうか。
ただし、ガソリンのほうは満タンで (車を) 返却していただくこととなります。

最後の文は、使った分のガソリンを返却するという意味に解釈することもでき、その場合は「ガソリンのほうは」が必須格ということになる。

2.5.3 だ型文

普通に「だ型文」として問題になるのは、助動詞「だ・です」が何らかの動詞を代用しているときとみなされる場合である。例えば、次のような文である。

百四十ドルのツインルームを二日間ですね (= 予約なさいますね)。
次の列車は、何時ですか (= 何時に出発しますか)。

本調査では、この他に、コピュラ文ではあるが、「AはBだ・です」のAとBが意味的に不整合であるケースも取り上げた。

まず、京都国際観光旅館が、(電話番号は) 四四三の一七零零です。

2.5.4 終助詞の欠落

疑問を表す文末の終助詞「か」が欠落している次のような文がある。

で、お席の方は、普通席、グリーン席、どちらになさいます(か)。
ええ、できたらマツダのジオメトロ、それが欲しいんだけど、有ります(か)。

もちろん、これらは、実際の音声では上昇型のイントネーションになっていて、それが「か」の表す疑問の意味を伝えているのである。

2.5.5 節関係の不整合

接続助詞で結ばれた主節と従属節の意味的な関係が不適格な文である。音声言語データベースでは、次の1例のみであった。

[そうですね] 交通の便が良くて、市街の中心地がいいですね。

この文を解析可能な文に直すとすると、次のような文になるだろう。

交通の便が良くて市街の中心地にあるホテルがいいですね。
交通の便の良い市街の中心地がいいですね。

2.5.6 融合文

前半と後半で違ったことを言っているように見える文である。2つの文が1つの文に無理に押し込められたといった感じのする文である。たぶん、2つの事柄を一度に言ってしまうおうとしたので、そのようになったのであろう。

八時間コースは今の説明いたしました二つのコースと、昼食のほうが新都ホテルでご用意いたしております。

この文は、「八時間コースは今説明した二つのコースに昼食がついたものだ」と「昼食は新都ホテルで用意してある」という2つの事柄を一度に言おうとして、このようなぎくしゃくした構文になったものである。この種の文が、今のところ、最も複雑な現象である。

2.6 語用論レベルの不適格文

語用論レベルの不適格文は、今回の調査の対象である音声言語データベースでは、次の例のみであった。

[そうですねー] 新幹線があればだから．．．夕方六時ごろです。

「照応のための情報不足」としておいたが、この「あれ」は厳密な意味の照応というよりも、ただ漠然と言っているにすぎない。実際の対話では「新幹線があれだから」という発話は聞き逃されるところである。

3 ポーズの調査

我々は1993年度にポーズによって区切られたポーズ節を基本単位とする音声認識用文法を作成した[4]。その過程で、ポーズの位置と統語的・意味的な切れ目が必ずしも一致しないケースが問題になった。また、統語的・意味的な切れ目に必ずポーズが置かれるわけではなく、ポーズが現れるか否かをあらかじめ予測することができないことも指摘された。

ポーズはもともと息継ぎのために置かれるもので、発話の過程にとってはあくまでも外的なものである。しかし、いったんその過程に入り込むと、統語的・意味的構造の制約を受けないわけにはいかない。話し手は、どこでも息継ぎをしてよいわけではなく、言葉の文法的な区切りに合わせて息継ぎをしなければならない。

一方、聞き手の方は、ポーズで区切られたまとまりを手がかりにして発話を理解している。ポーズをすべて取り除くと聞き手はまったく発話を理解することができなくなることを立証した実験もある[3]。ポーズは話し手だけでなく聞き手にとっても欠くことのできないものであり、情報処理のために必要な時間である。

以上のように、ポーズは、文法にとって外的でありながらそれと密接な関係もあるという微妙な性格を持っている。そこで、音声言語データベースを対象にして改めて調査し、ポーズという概念を再考するとともに、音声認識用文法にポーズをどう取り込むべきか、等について再検討するための資料を作成することにした。

3.1 データの構成

データ例および各項目の説明をする。

会話番号	: TAS13002
発話番号	: 160
文	: あいにく一万円のお部屋は予約で満室になっております。
ポーズ位置	: あいにく一万円のお部屋は▼予約で満室になっております。
部分木リスト	: ADV/PP ▼ VP

会話番号	: 音声言語データベースの会話番号
発話番号	: 音声言語データベースの発話番号
文	: 音声言語データベースの文
ポーズ位置	: ▼で長いポーズ、▽で短いポーズを示す
部分木リスト	: ポーズで区切られた部分の最上位の統語カテゴリー (例のように、1つにまとめられない場合は、さらに部分木に分け、スラッシュで区分した)

3.2 ポーズの問題点と処理

前述したように、ポーズで問題になるのは、統語的・意味的な切れ目とポーズの位置が一致しないケースである。これには、次の2つのケースがある。

- 1つのポーズ節に2つ以上の句が含まれるケース

あいにく／一万円のお部屋は▼予約で満室になっております。

● 1つの句がポーズによって分割されるケース

(どのような▽お部屋)をご希望でしょうか。

次の例のように、2つのケースが1つの文に混在していることもある。

今／(ニューヨークシティーホテルの▼六百二号室)に滞在しています。

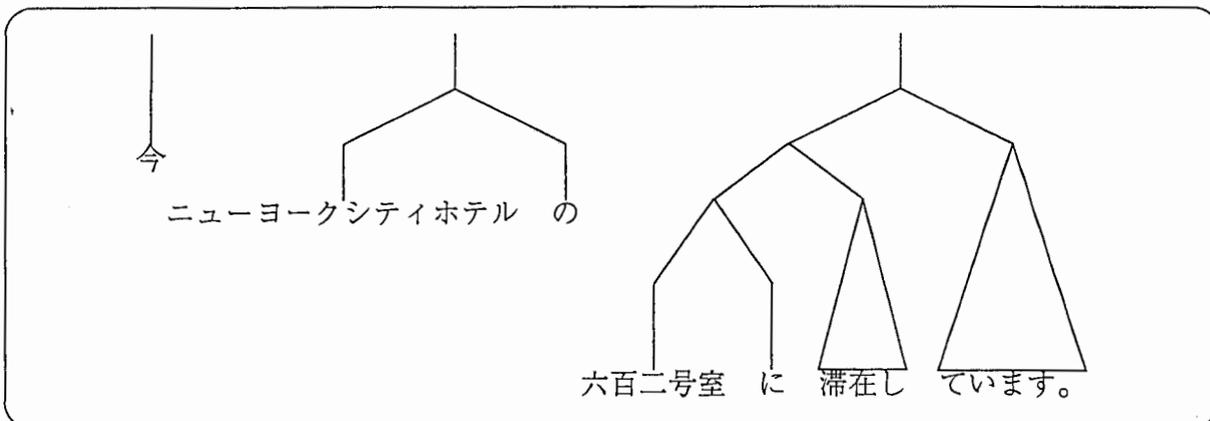
これらの例のようにポーズの位置と統語的・意味的な切れ目が必ずしも一致せず、またポーズの出現を予測することができないとすれば、ポーズ節を音声認識用文法の基本単位とすることは疑わしい試みだと思われるかもしれない。しかし、ポーズを音声認識用文法に取り込むことには、次のような理由がある。

- (1) 音声区間が長くなればなるほど、音声認識率が低下する。文全体を一度に認識することが現在の音声認識技術で難しいのであれば、ポーズによって区切られた比較的短い区間を認識単位にすることは、十分意味がある。
- (2) 音声言語データベースの調査結果を見る限り、ポーズの位置は文の統語的・意味的な切れ目と比較的よく一致する。一致しないケースの割合は約18%あった。

問題は、ポーズの位置が統語的・意味的な切れ目と一致しないケースをどう処理するかということであるが、それを次の例をモデルにして検討してみよう。

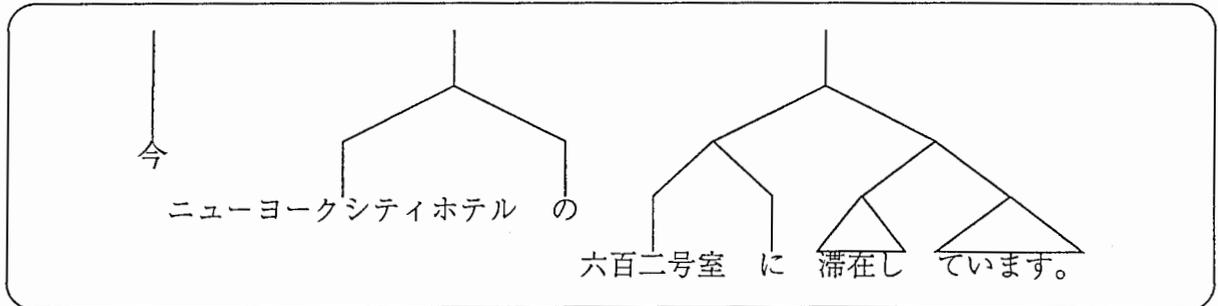
今ニューヨークシティーホテルの▼六百二号室に滞在しています。

「今ニューヨークシティーホテルの」という部分は1つの句にまとめることができず、「今」と「ニューヨークシティーホテルの」という2つの部分木に分割される。一方、「六百二号室に滞在しています」という部分は1つの動詞句にまとめられる。そこで、全体は次のような部分木の集まりとなる。



問題は、部分木「ニューヨークシティーホテルの」が、部分木「六百二号室に滞在しています」の中に深く埋め込まれた「六百二号室」に係るようにしなければならないことである。その後で、部分木「今」が「ニューヨークシティーホテルの六百二号室に滞在しています」に係ることになる。

このように、ポーズによって区切られた文では、ポーズの前にある部分木が、ポーズの後にある部分木に深く埋め込まれた語句に直接係るようなケースがしばしば起こる。このような操作を容易にするには、以前に議論したように、構文木を右枝分かれにすることが有効なのではなかろうか。



あるいは、部分木にまとめた段階で統語解析をやめ、それ以降は意味解析に委ねるということも考えられる。例えば、上の文のそれぞれの部分木が次のように意味解析されるとしよう。

今:

```
[[sem time]
 [lex 今]]
```

ニューヨークシティーホテルの:

```
[[sem possession]
 [lex の]
 [[sem facility]
 [lex ニューヨークシティーホテル]]]
```

六百二号室に滞在しています:

```
[[sem stay]
 [lex 滞在]
 [[sem location]
 [lex に]
 [[sem room]
 [lex 六百二号室]]]]]
```

すると、“sem 素性”の結合規則を定めておけば、文全体の意味解析として次の表現が得られる。

今ニューヨークシティーホテルの六百二号室に滞在しています。

```
[[sem stay]
 [lex 滞在]
 [[sem location]
 [lex に]
 [[sem room]
 [lex 六百二号室]
 [[sem possession]
 [lex の]
 [[sem facility]
 [lex ニューヨークシティーホテル]]]]]
[[sem time]
 [lex 今]]]
```

自由な発話になればなるほど変則的な部分木に分かれることが多くなるので、無理に統語解析を押し進めるよりも、その先を部分木に基づく柔軟な意味解析に委ねるとするのはよいかも知れない。

今回の調査の中で、ポーズの位置が最も変則的なのは次の文であった。

ファーストネーム▽がジョン様で▼ラストネームが▽フィリップス様で▽ご
ざいますね。
今ニューヨーク▽シティーホテルにお泊まりで▼電話番号のほうが、二零三▼
四四三▼一七零零でございますね。

「ファーストネーム▽が・・」では、名詞と格助詞の間にポーズがある。「フィリップス様で▽ございますね」では、述語を構成する助動詞と助動詞の間にポーズがある。「今ニューヨーク▽シティーホテルにお泊まりで」では、単語そのものの中にポーズがある(この例では、複合固有名詞を構成する語と語の間)。いずれも、部分木に基づく構文解析にとっては難しい問題である。

4 並列句の調査

並列句は、規則化するのが最も困難なもののひとつである。これまで、音声認識用文法でも言語解析用文法でも、名詞と名詞が並列助詞の「と」や「や」で結合されたものしか扱ってこなかった。しかも、並列助詞を連体助詞の「の」と同じ扱いにするという便法の域を出なかった。新しい文法では、並列句を可能なかぎり一般的な規則で扱いたい。そのための基礎資料として、音声言語データベースに出現する並列句の調査をした。

なお、並列句ではないが、「～から～まで」という表現も、並列句と同じように複雑な構造を持つことがあるので、この機会に調査した。

4.1 データの構成

データ例および各項目の説明をする。

会話番号：TAS12001

発話番号：450

文：|それから|サービス|料金|と|税金|は|別|に|な|る|の|で|、
|よろしく|お願いいたします|。|

並列句：(サービス料金)と(税金)

統語構造：NP-to-NP

会話番号：音声言語データベースの会話番号

発話番号：音声言語データベースの発話番号

文：音声言語データベースの文

並列句：並列句の各項を括弧で示す。項と項を結ぶ語句を括弧の外に記述。

統語構造：並列句の各項を統語カテゴリーで示す。項と項を結ぶ語句は、語句自体をアルファベットで記述。

4.2 並列句のタイプ

4.2.1 名詞句+名詞句+・・・

接続助詞なしに、名詞が並列されるタイプである。名詞の数は、最大5個であった。

こちらの朝食の方は、和食、洋食、どちらも千二百円で、ご注文いただけますよ。

ホテル、旅館、民宿などいろいろがございますが。

見学していただくところは東寺、三十三間堂、清水寺、平安神宮です。

そのあと、楽水苑、伏見桃山城、万福寺、平等院、醍醐寺三宝院を見学いたします。

最後の名詞の前に「後は」「最後に」のような語が接続する例がある。

京阪電車、ジェイアール、あとはバスと行き方がございます。

午前中のコースですが、二条城、金閣寺、竜安寺、最後に手工芸センターを見学いたします。

4.2.2 名詞句+並立助詞/接続詞+名詞句(十並立助詞)

名詞が並立助詞によって結ばれているものである。並立助詞には「と」「や」「か」「とか」がある。最も多いのは「と」であった。

名前とクレジットカードの番号をお知らせ願えませんか。
オートマティックとマニュアルとございますが。
そこではアンティークや珍しい物を売っていますけれども、最近は人気が出ています。
じゃあしょうがないから四日か五日でもいいですよ。
片道か往復かどちらでしょうか。
アメリカ街とか五番街はご存じですか。

接続詞の「または」「もしくは」で結ばれているもの。

電車のほうですと京阪電車またはジェイアールになります。
八月の十日までにエアパシフィックの支店、もしくは空港のチケットカウンターで変更の手続きを行ってください。

並立助詞と接続詞が併用されているものもある。

それからお寺の見学時間や拝観料、それから交通についても教えてもらえませんか。
当日は国際免許証とパスポート、それにビザカードを必ずお持ちください。
東大寺の大仏とそれから、お水取りを見たいんです。

4.2.3 名詞句＋並立助詞／接続詞＋動詞句

名詞と動詞が並立助詞または接続詞で結ばれているものである。

料金の方はマイルと週のいつにお使いになるかで違ってきます。
そうしますと、お客様のお年とほかに運転する方がいらっしゃるかどうかをお知らせ願えますか。
いつと、それからどこまでお使いになるか、教えていただけますか。

4.2.4 動詞句＋動詞句

並立助詞や接続詞なしに、動詞句が並置されているものである。

片道ですか、往復ですか。

4.2.5 動詞句＋並立助詞／接続詞＋動詞句（＋並立助詞）

動詞句が並立助詞または接続詞で結ばれているものである。

家族と相談してから、京阪に乗るか、ジェイアールに乗るか決めます。
片道ですか、それとも往復をお買いになりますか。

4.2.6 その他

並列句の各項が、後置詞句と名詞句の複合句から成っているものが4例あった。今回の調査では、最も扱いにくそうなものである。

でしたら、三時間のコースが二コースと八時間のコースが一コースござい
ます。
すいませんが、一泊目は一人で、二泊目と三泊目は二人でお願いしたい
んですが。
すいませんが一泊目は一人で、あと二泊目と三泊目は二人でお願いしたい
んですが。
例えば自動車で旅行ですとか、長期滞在ですとか。

4.3 「～から～まで」の調査

音声認識用文法では、「～から」「～まで」という後置詞句は、他の後置詞句と同じよ
うに、それぞれ単独で述語に係るようになっていく。しかし、統語的にも意味的にも両方が
ひとまとまりになって機能することが多く、その点で並列句と似ている。そこで、「～から
～まで」という表現を特に取り上げて調査した。

「から」「まで」によって結合される名詞は、次のタイプに限られた。

- 時間を表す名詞

フェリーは午前十時から午後四時までの間は、三十分おきに出ています。

- 場所を表す名詞

ラスベガスからロサンゼルスまで、おとな片道四十四ドルになります。

- 金額を表す名詞

一泊百五十ドルから、二百ドルぐらいまでだな。

「～から～」のように、「まで」を伴わない例もある。

たぶん、八時から九時ぐらいになると思うんですけども。
ラスベガスからロスアンゼルスの間は、その時間にもよりますけれども、
六時間ぐらいで着きます。
二万円から三万円ぐらいがいいんですが。

「まで」の代わりに「へ」が用いられている例。

すみません、東京から京都への新幹線の切符が欲しいんですけども。
バスのほうでしたらお乗り換えのほうは電車からバスへと一回だけになりますので。

5 複合語・数量詞の調査

複合語や数量詞も規則化することが困難なもののひとつである。特に、音声認識においては、短い接頭辞や接尾辞は誤認識の原因となるので、複合語や数量詞はいくつかのものを除いて一語登録せざるをえなかった。

新しい音声言語データベースの形態素体系では、接頭辞や接尾辞として採用されるものが限定されているので、音声認識用文法でも、これに準じて複合語や数量詞を規則化したいと考えている。そのための基礎資料として、音声言語データベースに現れる複合語と数量詞を調査した。

5.1 データの構成

データ例および各項目の説明をする。

複合語 : お | 楽 | し | み

統語構造 : (<接頭辞> (<形容詞><接頭辞>))

頻度 : 8

複合語 : 複合語・数量詞

統語構造 : 複合語・数量詞の統語構造 (活用語は、語幹と活用語尾を合わせて「動詞」「形容詞」等とした)

頻度 : 音声言語データベースに出現する頻度

5.2 複合語・数量詞のタイプ

統語構造の簡単な複合語から複雑な複合語まで、例を挙げる。

5.2.1 名詞+接尾辞

(<普通名詞><接尾辞>) : 朝食 | 代

(<数詞><接尾辞>) : 三 | 泊

(<日時><接尾辞>) : 一時 | 発

(<固有名詞><接尾辞>) : ロサンゼルス | 行き

(<人名><接尾辞>) : 田中 | 様

(<代名詞><接尾辞>) : わたくし | ども

(<サ変名詞><接尾辞>) : 予約 | 係

5.2.2 形容詞+接尾辞

(<形容詞><接尾辞>) : 長 | 過ぎ(る)

5.2.3 接頭辞十名詞

- (＜接頭辞＞＜普通名詞＞) : お | みやげ
(＜接頭辞＞＜代名詞＞) : お | いくら
(＜接頭辞＞＜サ変名詞＞) : ご | 連絡
(＜接頭辞＞＜形容名詞＞) : ご | 親切

5.2.4 接頭辞十動詞

- (＜接頭辞＞＜動詞＞) : お | 受け取り

5.2.5 接頭辞十形容詞

- (＜接頭辞＞＜形容詞＞) : お | 近い

5.2.6 名詞十名詞

- (＜普通名詞＞＜普通名詞＞) : ファックス | 番号
(＜普通名詞＞＜数詞＞) : エム | 六
(＜普通名詞＞＜サ変名詞＞) : 現地 | 集合
(＜数詞＞＜普通名詞＞) : 三 | 種類
(＜日時＞＜日時＞) : 九時 | 半
(＜日時＞＜サ変名詞＞) : 八時 | 開演
(＜固有名詞＞＜普通名詞＞) : 日本 | 文化
(＜固有名詞＞＜固有名詞＞) : 近鉄 | 奈良駅
(＜固有名詞＞＜サ変名詞＞) : 京都 | 見物
(＜サ変名詞＞＜普通名詞＞) : 発車 | 時刻
(＜サ変名詞＞＜サ変名詞＞) : 予約 | 変更
(＜形容名詞＞＜普通名詞＞) : 有効 | 期限

5.2.7 動詞十動詞

- (＜動詞＞＜動詞＞) : 言い | 忘れ

5.2.8 複合語十接尾辞

- ((＜普通名詞＞＜普通名詞＞)＜接尾辞＞) : フリーマイリッジ | 料金 | 制
((＜固有名詞＞＜固有名詞＞)＜接尾辞＞) : 近鉄 | 奈良駅 | 前
((＜日時＞＜日時＞)＜接尾辞＞) : 九時 | 五十分 | 発
((＜人名＞＜人名＞)＜接尾辞＞) : 田中 | 健史 | 様
((＜サ変名詞＞＜サ変名詞＞)＜接尾辞＞) : 免責 | 補償 | 料
((＜数詞＞＜接尾辞＞)＜接尾辞＞) : 四 | 時間 | 後
((＜サ変名詞＞＜接尾辞＞)＜接尾辞＞) : 予約 | 便 | 名
((＜接頭辞＞＜普通名詞＞)＜接尾辞＞) : お | 客 | 様

5.2.9 接頭辞十複合語

(〈接頭辞〉(〈数詞〉〈接尾辞〉)) : お | 一 | 人
(〈接頭辞〉(〈形容詞〉〈接尾辞〉)) : お | 楽し | み

5.2.10 名詞十複合語

(〈普通名詞〉(〈数詞〉〈接尾辞〉)) : 徒歩 | 五 | 分
(〈普通名詞〉(〈普通名詞〉〈接尾辞〉)) : 全室 | バス | 付き
(〈普通名詞〉(〈接頭辞〉〈普通名詞〉)) : マイル | 無 | 制限
(〈普通名詞〉(〈普通名詞〉〈普通名詞〉)) : 山地 | 標準 | 時間
(〈数量詞〉(〈普通名詞〉〈接尾辞〉)) : ワンブロック | 北 | 側
(〈固有名詞〉(〈数詞〉〈接尾辞〉)) : のぞみ | 九 | 号
(〈固有名詞〉(〈普通名詞〉〈普通名詞〉)) : 太平洋 | 標準 | 時間
(〈サ変名詞〉(〈数詞〉〈接尾辞〉)) : 合計 | 百 | 六 | ドル
(〈形容名詞〉(〈普通名詞〉〈普通名詞〉)) : 特別 | 地方 | 消費税

5.2.11 複合語十名詞

((〈普通名詞〉〈接尾辞〉) 〈普通名詞〉) : バス | 付き | ツイン
((〈数詞〉〈接尾辞〉) 〈普通名詞〉) : 六 | 時間 | 単位
((〈サ変名詞〉〈接尾辞〉) 〈普通名詞〉) : 搭乗 | 者 | 傷害
((〈接頭辞〉〈普通名詞〉) 〈普通名詞〉) : お | 部屋 | 番号
((〈接頭辞〉〈サ変名詞〉) 〈普通名詞〉) : ご | 予約 | 内容
((〈普通名詞〉〈サ変名詞〉) 〈普通名詞〉) : 市内 | 循環 | バス
((〈数詞〉〈接尾辞〉) 〈サ変名詞〉) : 一 | 人 | 使用
((〈接頭辞〉〈普通名詞〉) 〈サ変名詞〉) : 全 | 保険 | 保証
((〈普通名詞〉〈日時〉) 〈サ変名詞〉) : 午前 | 十一時 | 開演

5.2.12 複合語十副詞

((〈数詞〉〈接尾辞〉) 〈副詞〉) : 百 | ドル | ちょうど

5.2.13 複合語十副助詞

((〈数詞〉〈接尾辞〉) 〈副助詞〉) : 一 | 時間 | ごと

5.2.14 その他

((〈数詞〉〈接尾辞〉) (〈数詞〉〈接尾辞〉)) : 四 | ドル | 五 | 十 | セント
((〈数詞〉〈普通名詞〉) (〈普通名詞〉〈接尾辞〉)) : 一 | ブロック | 北 | 側
(〈接頭辞〉 (〈数詞〉〈接尾辞〉) 〈接尾辞〉) : 約 | 二 | 十 | 分 | 間
(〈普通名詞〉 (〈日時〉〈日時〉) 〈接尾辞〉) : 午前 | 零時 | 十五分 | 発
(((〈固有名詞〉〈普通名詞〉) 〈サ変名詞〉) 〈普通名詞〉) : 京都 | 市内 | 観光 | バス
(((〈接頭辞〉〈普通名詞〉) 〈接尾辞〉) 〈動詞〉) : お | 子 | 様 | 連れ
(〈形容詞〉 ((〈接頭辞〉〈普通名詞〉) 〈接尾辞〉) 〈動詞〉) : 小さ | い | お | 子 | 様 | 連れ

5.3 名詞連続句

名詞や複合語がいくつも連続してひとまとまりの句を構成することがある。特に、ホテルや列車などの予約項目について質疑・応答するときに、必ず現れる。これを名詞連続句として、複合語とは別に調査した。

連続句 : 一 | 泊 | 二 | 食 | 付き | 、 | お | 一 | 人 | 様 | 二 | 万 | 円

統語構造 : (一泊) (二食付き) (お一人様) (二万円)

カテゴリ : (数量詞) (名詞) (数量詞) (金額)

連続句 : 九月 | 十日 | 九時 | 五十六分 | 東京 | 発 | のぞみ | 九 | 号 | 博多 | 行き

統語構造 : (九月十日九時五十分) (東京発) (のぞみ九号) (博多行き)

カテゴリ : (日時) (固有名詞+接尾辞) (固有名詞) (固有名詞+接尾辞)

名詞連続句は非常に不規則な現れ方をするので、統語規則に収束させることはむずかしい。今回の調査では、用例を収集するのにとどめた。

謝辞

本調査を進めるにあたり、適切な助言や支援をいただいた ATR 音声翻訳通信研究所 第四研究室 森元 逞 室長、石崎 雅人 主任研究員に感謝します。

参考文献

- [1] 松本裕治: “頑健な自然言語処理の研究動向と課題”, 情報処理学会研究報告, 94-SLP-1-2 (1994-05).
- [2] 田窪行則: “音声対話の言語学的モデル — 談話管理標識としての感動詞の分析 —”, 情報処理学会研究報告, 94-SLP-1-3 (1994-05).
- [3] 杉藤美代子: “談話におけるポーズとイントネーション”, 日本語と日本語教育 第2巻 — 日本語の音声・音韻, 明治書院 (1993).
- [4] 竹沢寿幸, 衛藤純司: 「ポーズ節に基づく音声認識用日本語文法」, ATR テクニカルレポート, TR-IT-0046 (1994-03).

A 付録 1: 不適格文のタイプ別頻度

- 265 : 言い淀み (文頭)
- 131 : 気づき (文頭)
- 84 : 助動詞の欠落
- 71 : 言い淀み (文中)
- 53 : 格助詞の欠落 (を)
- 47 : 特殊構造 (箇条発話)
- 46 : 応答 (文頭)
- 37 : 述語の欠落
- 36 : 連続文
- 32 : 中止文 (接続助詞終止)
- 27 : 提題助詞の欠落
- 25 : 連体助詞の欠落
- 23 : だ型文
- 22 : つづり
- 20 : 直示
- 19 : サ変名詞述語
- 18 : 後置詞句の名詞修飾
- 18 : 融合文
- 17 : 同格
- 14 : 格助詞の欠落 (が)
- 14 : 中止文 (主節の欠落)
- 13 : 呼びかけ
- 13 : 語順の交替
- 10 : 驚き (文頭)
- 10 : 必須格の欠落
- 9 : 迷い (文頭)
- 8 : 感動 (文頭)
- 6 : 気づき (文中)
- 6 : 断片文
- 6 : 倒置
- 6 : 納得 (文頭)
- 4 : 格助詞の欠落 (まで)
- 4 : 副詞句の名詞修飾
- 4 : 連用修飾句の名詞修飾
- 4 : 並立助詞の欠落
- 3 : 格助詞の誤り
- 3 : 終助詞の欠落
- 3 : 挿入句

- 2: 格助詞の欠落 (に)
- 2: 格助詞の欠落 (につき)
- 2: 繰り返し
- 2: 形容名詞述語
- 2: 言い換え
- 2: 言い直し
- 2: 単位の欠落
- 2: 中止文 (述語の中断)
- 2: 名詞句に接続する終助詞
- 1: 格助詞の余剰
- 1: 照応のための情報不足
- 1: 節関係の不適合
- 1: 単語の言い誤り
- 1: 否定
- 1: 補助動詞の欠落
- 1: 連体助詞の余剰

B 付録 2: 並列句のタイプ別頻度

- 118 : (名詞句) と (名詞句)
- 25 : (名詞句) (名詞句)
- 20 : (名詞句) (名詞句) (名詞句)
- 19 : (名詞句) か (名詞句)
- 7 : (名詞句) (名詞句) (名詞句) (名詞句)
- 7 : (名詞句) と (名詞句) と
- 6 : (名詞句) と、それから (名詞句)
- 5 : (名詞句) や (名詞句)
- 4 : (名詞句) ですとか (名詞句) ですとか
- 4 : (名詞句) または (名詞句)
- 3 : (名詞句) か (名詞句) か
- 3 : (名詞句) と (名詞句) それに (名詞句)
- 3 : (動詞句) それとも (動詞句)
- 2 : (名詞句) (名詞句) (名詞句) など
- 2 : (名詞句) (名詞句) とも
- 2 : (名詞句) もしくは (名詞句)
- 2 : (動詞句) (動詞句)
- 1 : (名詞句) (名詞句) (名詞句) (名詞句) (名詞句)
- 1 : (名詞句) (名詞句) (名詞句) (名詞句) など
- 1 : (名詞句) (名詞句) (名詞句) 最後に (名詞句)
- 1 : (名詞句) (名詞句) 後は (名詞句)
- 1 : (名詞句) それから (名詞句)
- 1 : (名詞句) と (名詞句) と (名詞句)
- 1 : (名詞句) と (動詞句)
- 1 : (名詞句) と、ほかに (動詞句)
- 1 : (名詞句) とか (名詞句)
- 1 : (後置詞句 / 名詞句) ですとか (名詞句) ですとか
- 1 : (後置詞句 / 名詞句) と (後置詞句 / 名詞句)
- 1 : (後置詞句 / 後置詞句) (後置詞句 / 後置詞句)
- 1 : (後置詞句 / 後置詞句) あと (後置詞句 / 後置詞句)
- 1 : (動詞句) か (動詞句) か

C 付録 3: 複合語のタイプ別頻度

373 : (<接頭辞><動詞>)	例: おすすめ
319 : (<数詞><接尾辞>)	例: 一マイル
312 : (<接頭辞><普通名詞>)	例: お料理
148 : (<接頭辞><サ変名詞>)	例: ご確認
101 : (<普通名詞><接尾辞>)	例: ホテル内
99 : ((<数詞><接尾辞>) <接尾辞>)	例: 一時間前
94 : (<普通名詞><普通名詞>)	例: フリーマイレッジ方式
59 : (<サ変名詞><普通名詞>)	例: サービス料金
52 : ((<接頭辞><普通名詞>) <接尾辞>)	例: ご飲食代
36 : (<サ変名詞><接尾辞>)	例: サービス料
35 : ((<人名><人名>) <接尾辞>)	例: 田中健史様
32 : (<固有名詞><接尾辞>)	例: 博多行き
26 : (<接頭辞> (<数詞><接尾辞>))	例: 約十分
22 : ((<日時><日時>) <接尾辞>)	例: 十二時二分着
18 : ((<接頭辞> (<数詞><接尾辞>)) <接尾辞>)	例: お二人様
16 : ((<数詞><接尾辞>) <普通名詞>)	例: 一万円以下
15 : ((<固有名詞><固有名詞>) <接尾辞>)	例: 近鉄奈良駅前
14 : (<接頭辞><代名詞>)	例: おいくら
13 : (<人名><接尾辞>)	例: 田中様
13 : (<日時><接尾辞>)	例: 一時発
12 : (<形容名詞><普通名詞>)	例: 特別料理
10 : (<固有名詞> (<数詞><接尾辞>))	例: のぞみ九号
9 : ((<数詞><接尾辞>) (<数詞><接尾辞>))	例: 四ドル五十セント
8 : (<接頭辞> (<形容詞><接尾辞>))	例: お楽しみ
7 : (<普通名詞> (<数詞><接尾辞>))	例: 徒歩五分
6 : ((<数詞><普通名詞>) (<普通名詞><接尾辞>))	例: 一ブロック北側
6 : (<固有名詞><固有名詞>)	例: 近鉄奈良駅
6 : (<普通名詞> (<サ変名詞><接尾辞>))	例: 税サービス料
6 : (<普通名詞><サ変名詞>)	例: 自動車旅行
5 : (<固有名詞><普通名詞>)	例: 日本建築
5 : (<普通名詞><数詞>)	例: エム六
5 : ((<普通名詞> (<日時><日時>)) <接尾辞>)	例: 午前九時十五分発
5 : (<日時><サ変名詞>)	例: 八時開演
4 : ((<接頭辞><普通名詞>) <普通名詞>)	例: お部屋番号
4 : (<形容名詞> (<普通名詞><普通名詞>))	例: 特別地方消費税
4 : (<接頭辞> (<動詞><助動詞>))	例: お待たせ
3 : ((<接頭辞><サ変名詞>) <普通名詞>)	例: ご到着時間
3 : (<数詞><普通名詞>)	例: 三種類

- 2 : ((<サ変名詞><サ変名詞>) <接尾辞>)
 2 : ((<サ変名詞><接尾辞>) <接尾辞>)
 2 : ((<サ変名詞><接尾辞>) <普通名詞>)
 2 : ((<数詞><接尾辞>) <副助詞>)
 2 : ((<普通名詞><サ変名詞>) <普通名詞>)
 2 : ((<普通名詞><日時>) <サ変名詞>)
 2 : (<サ変名詞> (<数詞><接尾辞>))
 2 : (<固有名詞> (<普通名詞><普通名詞>))
 2 : (<固有名詞><サ変名詞>)
 2 : (<数量詞> (<普通名詞><接尾辞>))
 2 : (<接頭辞><形容詞>)
 2 : (<代名詞><接尾辞>)
 2 : (<日時><日時>)
 2 : (<普通名詞> (<普通名詞><接尾辞>))
 2 : (<普通名詞> (<普通名詞><普通名詞>))
 1 : (((<固有名詞><普通名詞>) <サ変名詞>) <普通名詞>)
 1 : (((<接頭辞><普通名詞>) <接尾辞>) <動詞>)
 1 : (((<形容詞> ((<接頭辞><普通名詞>) <接尾辞>))) <動詞>)
 1 : (((<数詞><接尾辞>) <サ変名詞>))
 1 : (((<数詞><接尾辞>) <副詞>))
 1 : (((<接頭辞><普通名詞>) <サ変名詞>))
 1 : (((<普通名詞><接尾辞>) <普通名詞>))
 1 : (((<普通名詞><普通名詞>) <接尾辞>))
 1 : (<サ変名詞><サ変名詞>)
 1 : (<形容詞><接尾辞>)
 1 : (<数詞><接尾辞>) <接尾辞>)
 1 : (<接頭辞><形容名詞>)
 1 : (<動詞><補助動詞>)
 1 : (<日時><普通名詞>)
 1 : (<普通名詞> (<接頭辞><普通名詞>))
- 例: 免責補償料
 例: 予約便名
 例: 搭乗者保険
 例: 一時間ごと
 例: 市内循環バス
 例: 午前十一時開演
 例: 合計百六ドル
 例: 太平洋標準時間
 例: 京都見物
 例: ワンブロック北側
 例: お近い
 例: わたくしども
 例: 九時半
 例: 全室バス付き
 例: 山地標準時間
 例: 京都市内観光バス
 例: お子様連れ
 例: 小さいお子様連れ
 例: 一人使用
 例: 百ドルちょうど
 例: 全保険保証
 例: バス付きツイン
 例: フリーマイリッジ料金制
 例: 予約変更
 例: 長過ぎ(る)
 例: 十分前
 例: ご親切
 例: 言い忘れ(る)
 例: 十二時過ぎ
 例: マイル無制限

D. 付録 4: 調査対象ファイル

D.1 不適格文の調査・並列句の調査・複合語の調査

TAS12001	TAS12002	TAS13001	TAS13002	TAS22001	TAS22002
TAS23001	TAS23002				
TBS12001	TBS12002	TBS12003	TBS12004	TBS13001	TBS13002
TBS13003	TBS13004	TBS22001	TBS22002	TBS22003	TBS22004
TBS23001	TBS23002	TBS23003	TBS23004		
TGS12001	TGS12002	TGS13001	TGS13002	TGS22001	TGS22002
TGS23001	TGS23002				
THS12001	THS12002	THS12003	THS12004	THS13001	THS13002
THS22001	THS22002	THS22003	THS22004	THS23001	THS23002
TKS12001	TKS12002	TKS13001	TKS13002	TKS22001	TKS22002
TKS23001	TKS23002				
TOS12001	TOS12002	TOS12003	TOS12004	TOS13001	TOS13002
TOS13003	TOS13004	TOS13005	TOS13006	TOS22001	TOS22002
TOS22003	TOS22004	TOS23001	TOS23002	TOS23003	TOS23004
TOS23005	TOS23006				
TRS12001	TRS12002	TRS13001	TRS13002	TRS22001	TRS22002
TRS23001	TRS23002				
TSS12001	TSS12002	TSS13001	TSS13002	TSS22001	TSS22002
TSS23001	TSS23002				

D.2 ポーズの調査

TAS12001	TAS12002	TAS13002	TAS22001	TAS22002	TAS23002
TAS32002	TAS33002				
TBS12001					
TGS12001					
THS12001	THS12002	THS13001			
TRS12001					
TSS12001	TSS12002				

E 付録 5: 調査結果ファイル

/data/as10/HMM-LR/JIRCO/heiretu.jis	並列句の調査
/data/as10/HMM-LR/JIRCO/hukugou.jis	複合語
/data/as10/HMM-LR/JIRCO/ill-form.jis	不適格文の調査
/data/as10/HMM-LR/JIRCO/karamade.jis	からまで
/data/as10/HMM-LR/JIRCO/pause.jis	ポーズ位置
/data/as10/HMM-LR/JIRCO/renzoku.jis	名詞連続句
/data/as10/HMM-LR/JIRCO/report.tex	本解説書