

休止を単位とした発話の文法

秋葉友良 伊藤克亘
t-akiba@etl.go.jp

電子技術総合研究所

利用者の発話をできるだけ制限しない音声対話システムの枠組として、休止を処理の区切りとする方式を検討している。本方式で音声認識をおこなう際には、休止間音声の言語モデルが必要となる。本稿では、「文」認識用の文法から、なるべく精度の良い休止間音声用の文法を、自動的に獲得する方法について検討を行なった。また、実際の文法を用いて実験をおこない、精度の高い文法が獲得できることを確認した。

Constructing the Grammars for Interpausal Phrases

AKIBA Tomoyosi, ITOU Katunobu

Electrotechnical Laboratory

We proposed a dialogue processing method which used a pause as a period of an utterance, in order to construct a speech dialog system which made users restrict as little as possible. For the speech recognition along with this method, the language model for interpausal phrases is indispensable. In this paper, we studied a method to extract the grammars for interpausal phrases with good property (small perplexity), automatically from any grammar for the so-called "sentence." We actually have extracted grammars along with this method, and have confirmed that they really had small perplexities.

1 はじめに

「利用者が気楽に発話をつづけることができる」音声対話システムを実現する手法として、我々は休止を単位とした対話処理手法を検討している。本稿では、休止を処理の単位とした場合に問題となる休止間音声区間の言語モデル作成の方法について述べる。

2 自由な対話と発話の漸進的解釈

近年、さまざまな音声対話システムが研究されているが、音声言語の利点を十分に生かしたシステムの実現には、まだ多くの課題が残されている。音声言語の利点とは何であろうか。音声言語は、伝達の効率を考えた場合かならずしも良い手段とはいえない。それにもかかわらず、多くの人間が多くの時間を費やして音声言語で対話している。人間は、情報を伝える以外の側面を、話し言葉による対話に求めているのだと思われる。我々は、このような「利用者が気楽に発話をつづけることができる」ようなシステムのふるまいについて検討してきた。

このような視点に立った時、これまでの音声対話システムの重大な欠点の一つとして、発話のターンに関する強い仮定を採用していることを指摘することができる。すなわち、これまでのシステムでは、暗黙の内にユーザとシステムの発話のターンを明確に区別し、システムとユーザが交互に完全な「文」を発話するという前提のもとにシステムが設計されていることが多かった。

しかし人と人との対話を観察していくとわかるように、相手の発話に割り込みを行ったり、発話の途中でも言葉を濁して相手の発話を催促したりと、各人は発話のターンを自分の意のままに切り替えようと試みる。

このシステムの前提とユーザが音声言語に求めているものとの間に、大きなギャップがあることは、想像に硬くない。ユーザが自由な対話を試みようとしても、システムがそれに対応できなければ、ユーザは無理にシステムの望むターンの切り替えに合わせることを強いられることになり、音声対話に窮屈さを感じることになるであろう。逆に、システムは

割り込みなどの自由な発話を自ら積極的に行って、ユーザに窮屈さを感じさせずに発話してもらうよう試みなければならないであろう。

このような対話を実現するためには、システムは発話の任意の時点で応答を開始できなければならない、そのためにはユーザが「文」を発話することを仮定しないで、発話中の任意の時点でその内容を漸進的に解釈する能力が必要となる。

3 処理の単位としての休止

本稿で対話処理とよぶ処理は、普通のシステムでは、1. 音声認識 2. 発話理解 (意味処理など) 3. 発話生成、の三つの部分に分けられる。発話を漸進的に解釈するためには、これらの要素を統合もしくは同時に実行する必要がある。しかし、これらの要素は、従来、別々の要素技術として研究がすすめられてきた。そのため、ひとつにまとめようと思ってもなかなか難しい問題がある。

まず、音声認識部で文法を用いるアプローチの場合、その文法がそもそも、「文」を前提とする場合が多い。この手法の場合、発話理解部には、書き言葉の自然言語で用いられるような手法を用いることが多いので、そのあとの処理は「文」を単位に進められる。しかし、前節で述べたように、この「文」という単位は適切とはいえない。

そもそも「文」とはそれほど明確な単位ではない。書き言葉であっても、句点と読点は可換なものがあり、句点から句点までという定義すら形式上の意味しかない。話し言葉の場合、その句点すら明確でないため、文の認定自体が難しい問題であるといわれている [3]。

休止に着目して認識用の文法を構築する試みは、文献 [4] で述べられている。しかし、この文法は目標を翻訳電話においてあるため、機械相手の対話システムに対する前提とは異なる部分もあると考えられる。竹沢らの「部分木」という仮定を離れて、休止と言語現象との対応をまとめると、以下のようになる [2, 3]。

- いわゆる倒置、いいよどもい — いい足す前に休止が入ることが多い。
- つなぎ語 — 休止に囲まれることも多い。すくなくとも、前後どちらか片方には、休止がある。
- 語間 — 非活用語 (名詞, 副詞, 感動詞, 助詞) の後と助動詞列の後に休止が入る。だいたい、いわゆる「文節」の区切りだが、名詞と助詞の間にも入りうる。

このように、休止で区切られた部分は、構文的にも意味的にも、それなりに意味のあるまとまりであることがわかる。

したがって、音声認識と言語処理を統合するときには、それなりによい目安になる単位であると考えられる。

4 休止間音声の文法

休止間音声の言語モデルとして、音声認識用の文法を利用することを考える。一般に、音声認識用の文法は、発話を受け入れるだけでなく、認識過程において強い制約となることが期待される。すなわち、目的の発話を受理できることに加えて、目的以外の発話には現れない表現をなるべく排除するという、2つの性質を合わせ持っていないといけない。このような性質を持つ休止間音声用の文法を人手で記述することは、「文」の文法を記述するのとは異なり、大変難しい。

例えば、我々の実験 [1] では、休止間として次のような発話が観察された。

その建物には (休止) 他に何 (休止) がありますか

「他に何」という休止間音声は、どのような文法でどのように解析されるべきなのだろうか。構文的には、「他に」と「何」という語句は、「がありますか」を介して結びつけられ一つの構文木を構成する。したがって、「他に何」単独では一つの構文木としてまとまらず、「文」の文法を記述するには、この休止間音声を受理する文法を記述できない。

代替案として、文節程度のまとまりのある単位をモデルとして、そのモデルの任意個の接続を休止間音声の文法として記述することが考えられる。しかし、モデル化した単位の接続情報が欠落しているため、実際の発話には現れないような単位の連続も受理する文法となってしまう。

また、どのような単位を採用するかの判断が難しい。あり得ない表現を排除するためには、なるべく大きな単位を採用すべきだが、その分休止間音声として認識可能な表現は減少してしまう。例えば、単位として文節を採用すると、文節内での休止が許されなくなり上の例のような休止間音声認識できなくなる。しかし単語では、全ての単語列がモデル化されてしまい、実際にはない単語の連続を排除できない。

どのような単位が接続可能かという情報は、本来「文」のための文法に記されていることである。し

たがって「文」認識用の文法からその情報を引き出せれば、それを利用することができる。また、3節で述べたように、休止間音声は意味的・構文的にまとまった単位となるので、どのような単位がまとまり易いかということは、「文」用に記述した文法の構文情報を利用することができる。次節では、「文」の認識用に作成した文法から、自動的に休止間認識用の文法を抽出することを試みる。

5 文脈自由言語の部分列を受理する文脈自由文法

ある文脈自由言語について、そこに含まれる文の部分列から成る言語を考える。一般には、任意に選択した部分列の集合は文脈自由言語とはならない¹。そこでここでは、文脈自由文法を用いたパーザで使用することを前提に、文脈自由言語となるような部分列の集合について検討し²、それを受理する文脈自由文法を得ることを試みる。

5.1 単語列の最小単位

休止間区間として現れる、単語のまとまりの最小単位を考えることは重要である。活用語の活用を記述する、次のような規則を考える。

行かない → 行く 五段活用語尾未然形
 ない 形容詞終止形
 行く → i k
 五段活用語尾未然形 → a
 ない → n a
 形容詞終止形 → i

認識用文法では、このような音素レベルの接続可能性を記述することがある。実際の発話が、これらの規則の内部で(例では、i k と a の間で)、休止をはさむことは考え難い。

また、文法レベルでも、「電話番号」「3時」「御出発」など、複合名詞、名詞+接尾語、接頭語+名詞、などの間には、普通、休止ははさみにくいと考えられる。

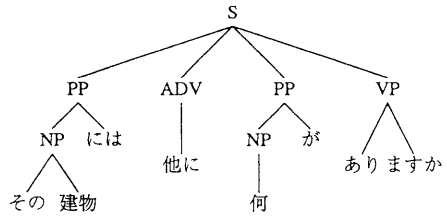
これらの語句の途中で切れるような文法が生成されることを防ぐために、文法 G に含まれる(非終端)記号の集合 C から、「その記号から生成される単語列内では休止をとらない」ものの集合 C_T を選び出す。 C_T が求まると、

$$G_T = \{r \in G | (r \text{の左辺}) \in C_T\}$$

$$G_N = G - G_T$$

¹ 特殊な場合として、文脈自由言語の(部分列ではなく文の)任意の部分集合は文脈自由言語とは限らない。

² 文脈自由言語の全ての部分列からなる集合は、5.2.6 節のクラス C の言語となり、文脈自由言語である。



NP → その 建物
 PP → NP には
 ADV → 他に
 NP → 何
 NP → NP が
 VP → あり ますか
 S → PP ADV PP VP

図1 「その建物には他に何がありますか」の構文木と文法

として、文法 G を二つの文法 G_T と G_N に分割することができる。そして、 G_T には手を加えず、 G_N に対してのみ次節以降の処理を施すことで、単語のまとまりの最小単位を固定することができる³。

以下の節では、こうして得られた文法 G_N を新たな文法 G とみて議論を進める。実際の認識用文法は G_T を加えることによって得られる。

5.2 部分列のクラス

次に、休止間区間として単語列がどのようにまとまるかを考察する。前述の「その建物には他に何がありますか」という文を考える(図1)。この文への休止の入り方としては、「その建物 には(休止)他に(休止)何が(休止)ありますか」や「その建物(休止)には他に何(休止)が(休止)ありますか」など、いろいろ考えられる。この2つを比べると、前者の方が休止の取り方としてより生じ易いと感じられる。前者は、休止間音声は構文的にまとまっている(いずれも文節)のに対し、後者の区間「には他に何」は、この部分だけで構文的に完結していない。

このように「構文的なまとまり具合」を休止間音声の候補としてとらえることは妥当であろう。「構文的なまとまり具合」を表す尺度として、構文木に対するその部分列の属する位置を考えることができる。図1を見てわかるように、「その建物は」「他に」などは、その部分列全体が一つの記号(それぞ

³ G_N は、新たに C_T を終端記号と考えた場合の文法に相当する。

れPP,ADV)を頂点とする木を過不足なく構成する。「他に何が」は、2つの隣接する記号(ADV PP)を頂点とする2つの木を構成する。「建物には」は、一つの木にはならないが、終端が部分列全体を包含するより大きな木(PPを頂点とするもの)で終了している(すなわち、PPを頂点とする部分列「その建物には」の接尾列となっている)。一方、「には他に何」は複数の木にまたがっている。

このような特徴に着目して、以下、言語 L の部分列からなる言語について、構文木との関係から、構文的まとまりとして考えられるいくつかのクラスを考えてみる。

5.2.1 クラス S

休止間区間の単語列がある記号を頂点とする単一の構文木を構成する場合。図1の例では、「その」「建物」「その建物」「その建物は」...など。言語 L_S を文法 G に含まれる記号 $c \in C$ が生成する単語列からなる集合とする。文法 G の開始記号を s とするとき、全ての C の要素 c について、 $s \rightarrow c$ を新たな生成規則として G に加えた文法 G_S が、 L_S を生成する。

$$G_S = G \cup \{s \rightarrow c | c \in C\}$$

5.2.2 クラス P

休止間区間の単語列が、隣接可能な複数の構文木を構成する場合。図1の例では、(クラスSに加えて)「その建物には他に (PP ADV)」「他に何が (ADV PP)」...など。言語 L_P をある規則の右辺に隣接して現れる複数の記号の列が生成する単語列からなる集合とする。例えば、ある規則 $c_0 \rightarrow c_1 c_2 c_3 c_4$ について、 $c_1 c_2$ 、 $c_1 c_2 c_3$ 、 $c_2 c_3$ 、 $c_2 c_3 c_4$ 、 $c_3 c_4$ が生成する部分列の集合を加える。 $s \rightarrow c_1 c_2$ 、 $s \rightarrow c_1 c_2 c_3$... を新たな生成規則として G_S に加えた文法 G_P が、 L_P を生成する。

$$G_P = G_S \cup \{s \rightarrow c_1 \dots c_j | c \rightarrow c_1 \dots c_n \in G, 1 < j - i < n - 1\}$$

5.2.3 クラス HT

休止間区間の単語列が、(純粹に単語列と見た場合の)「文」の接頭列あるいは接尾列である場合。図1の例では、「その」「その建物」...「その建物には他に何」...「には他に何ががありますか」...「がありますか」...など。言語 L_{HT} を言語 L の接頭列、接尾列から成る集合とする。

G	その建物には他に何ががありますか
G_S	何が
G_A	他に何が
G_{HT}	その建物には他に何
G_{HT-S}	建物には
G_{HT-A}	他に何
G_C	には他に何

表1 クラスと受理できる文の関係

言語 L の接頭列を生成する文法 G_H は、次のような手順で得ることが出来る。

1. G 中の全ての非終端記号 c について、 G に現れない新たな記号 c^h を作る。
2. G 中の全ての規則 $c \rightarrow c_1 \dots c_n$ について、
 - (a) $k = 1 \dots (n-1)$ について、規則 $c^h \rightarrow c_1 \dots c_k$ を生成する。
 - (b) $k = 1 \dots n$ について、 c_k が非終端記号であれば、規則 $c^h \rightarrow c_1 \dots c_k^h$ を生成する。
3. 開始記号 s について、規則 $s \rightarrow s^h$ を生成する。

同様に、言語 L の接尾列を生成する文法 G_T も構成できる。 G_H と G_T の規則を足し合わせることで文法 G_{HT} が得られる。

例えば、図1の文法について、接頭列を生成するために加える規則は次の通り。

$NP^h \rightarrow$ その
 $NP^h \rightarrow$ 何
 $PP^h \rightarrow NP$
 $PP^h \rightarrow NP^h$
 ...
 $S^h \rightarrow PP \text{ ADV } PP$
 $S^h \rightarrow PP \text{ ADV } PP^h$
 $S^h \rightarrow PP \text{ ADV}$
 $S^h \rightarrow PP$
 $S^h \rightarrow PP^h$
 $S \rightarrow S^h$

5.2.4 クラス HT-S

クラス HT は「文」(図1のS)の接頭列、接尾列であったが、さらに任意の記号が生成する単語列の接頭列、接尾列である場合。図1の例では、「建物には (PP の接尾列)」など。言語 L_{HT-S} を L_S の接頭列、接尾列から成る集合とする。文法 G_{HT} に、クラス S を得るための処理を施すことで、文法 G_{HT-S} が得られる。図1の文法では、さらに以下の規則を加えることになる。

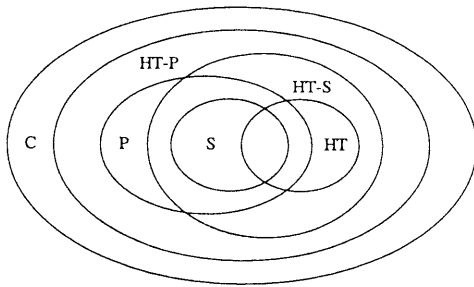


図 2 クラス間の関係

S → NP^h
 S → PP^h
 ...

5.2.5 クラス HT-P

HT-S に加えて、隣接可能な複数の構文木を構成する単語列の、接頭列、接尾列である場合。図 1 の例では、「他に何 (ADV PP の接頭列)」「建物には他に (PP ADV の接尾列)」「建物には他に何が (PP ADV PP) の接尾列)」など。言語 L_{HT-P} を L_A の接頭列、接尾列から成る集合とする。文法 G_{HT} に、クラス P を得るための処理を施すことで、文法 G_{HT-P} が得られる。例では、以下の規則を加える。

S → ADV PP^h
 S → PP' ADV
 S → PP' ADV PP
 ...

5.2.6 クラス C

言語 L の任意の部分単語列。図 1 の例では、その他の部分列「には他に何」など。言語 L_C を L の全ての部分列からなる集合とする。言語 L_{HT} に、言語 L_H の接尾列 (あるいは、言語 L_T の接頭列) を加えた集合に等しい。文法 G_H から、その接尾列を生成する文法 (5.2.3 節) を生成することで、文法 G_C が得られる。例では、次の規則を加える。

S → PP^h ADV PP'

以上のクラス間の関係を、図 2 に示す。

6 perplexity の推定

本手法を用いて、どの程度の複雑さをもつ文法が得られるかを調べるために、実際に「文」認識用文法を用意して、それぞれのクラスの文法を作成してみた。文法のドメインは「渋谷の道案内」である。

WOZ 実験 [1] のデータをもとに、被験者の典型的な発話をカバーするように作成した。規則数 1200、単語数 192 である。文法は、単語の意味・語用論情報を考慮し、動詞の格パターンを反映した記述をおこなった。例えば、動詞「行く」について、

S → PP_{locate}_KARA
 PP_{locate}._MADE DOU Ikuka
 S → PP_{locate}._HE Ikitai
 S → PP_{chikaku}._HE Ikitai

のように、モダリティによっても格パターンをそれぞれ用意した。

作成した文法に、前節で述べた手法により、休止間音声認識用の文法を 6 種類 ($G_S, G_A, G_{HT}, G_{HT-S}, G_{HT-A}, G_C$) 抽出した。規則数 1200 のうち、338 規則については最小単位を固定するために処理を行なわなかった (5.1 節)。単純に新たな規則を加えると冗長な規則が生成されるので、後処理により文法を最適化した⁴。

また、比較のため、人手によって簡単な休止間音声認識用の文法 G_{sim4}, G_{sim5} を作成した。これは、まず文節を定義し、休止間音声区間を、それぞれ 4.5 文節以内の文節列と定義したもので、おおよそ次のような文法である。

文 → 文末
 文 → 文節 文末
 文 → 文節 文節 文末
 文 → 文節 文節 文節 文末
 文末 → 文節
 文末 → 文末文節

以上の全ての文法について、その先頭に間投詞を認識するように手を加えた。

文法について、各文長 (単語数) ごとに生成される文の数を計算した。WOZ 実験 [1] で獲得した音声データから休止間音声に含まれる単語数の分布を求め、単語 perplexity を次の式で近似計算した [5](表 2)。

$$H(L) = - \sum_k \frac{p_k}{k} \log \frac{p_k}{N_k}$$

$$\text{Perplexity} = 2^{H(L)}$$

ただし、 p_k は単語数が k である確率、 N_k は単語数 k の文の総数である。

文法 G_{sim4}, G_{sim5} は、文節中でのポーズを許さない、許される文節数に制限があるなど、かなり制約が強い。それにも関わらず、自動生成した文法は、それと同等か、むしろより小さな perplexity を示している。

⁴最適化によって文法の生成力が落ちることはない。

文法	規則数	perp.	増加率
G	1200	18.10	1.0
G_S	1504	33.37	1.84
G_P	1833	36.95	2.04
G_{HT}	3870	57.64	3.18
G_{HT-S}	4035	59.22	3.27
G_{HT-P}	4538	59.40	3.28
G_C	4403	63.81	3.53
G_{sim4}	488	57.12	3.16
G_{sim5}	489	63.95	3.53

表 2 文法の比較

7 実際の発話との比較

では、実際にはどの文法を採用すればよいのであろうか。休止間音声として受理する表現の種類を多くするには、全ての部分列を認識する文法 G_C を採用すればよいが、文法が複雑になる分 perplexity が上昇し認識率が下がることになる。また、休止間音声は意味的にまとまる傾向があるため、「全ての部分列」という仮定は過度の一般化であろう。一方、文法 G_S を採用すると、perplexity が低い分、認識率は期待できる。しかし、休止間音声として、任意の文節の連続（「それは どこに」）、助詞から始まる列（「を教えてください」）などが受理できなくなってしまう。

実際の発話の、休止間音声⁵が5節のどのクラスに属するのかを調べてみた。WOZ 実験 [1] によって獲得した音声データのうち、43 対話における被験者の発話から、1294 の休止間音声区間を自動的に切り出した。そのうち、「文」の文法で認識できるものは 1064 (82%) であった。「文」の文法で認識できないもので、(休止間音声内での) 言い直し、いい淀み、倒置、中止の発話を除いた 120 区間について、どのクラスの文法で受理可能かを調べた (表 3)。

今回の調査では、真にクラス HT-A,C に属する音声区間は認められなかった。したがって、全ての休止間音声を受理した上で、Perplexity を低く押えるためには、文法 G_{HT-S} を採用すれば良い。

$L_{HT-S} - L_{HT}$ に含まれる 6 区間は次のようなものであった。

は<wa>? ($C L_S$) のつきあたりまでは
 中で とローマナというのは
 お店 とかオープの場所は

このうちの 2 番めを除く音声区間は直前に固有名称を発話しており、一定のパターンがあるように思われる。現在は、構文構造だけで休止間音声の候補を抽出しているが、より文法精度をあげるには、こ

言語	音声区間	割合 (%)
L_S	77	64
L_P	84	70
L_{HT}	114	95
L_{HT-S}, L_{HT-P}, L_C	120	100
L_{sim}	35	29
total	120	100

表 3 実際の休止間音声とクラスの関係

ういった個別のカテゴリの差も考慮する必要があるだろう。

また、120 区間のうちの約 95 がクラス HT に、約 64% がクラス S に含まれているなど、多くは制約の強いクラスに属することがわかる。将来的には、確率文脈自由文法などの尤度を扱えるモデルの表現形式を利用して、本稿のクラス階層を休止間音声認識用の言語モデルの優先度情報として用いることも可能であろう。

8 まとめ

休止間音声を認識するための文法を、「文」認識用の文法から自動抽出する手法について述べた。本手法を用いることにより、人手で安易に作成した文法に比べて、精度の良い文法が得られることを、実験により確認した。また、WOZ 実験で集めた発話例を調べ、本手法で自動抽出した 6 種類の文法のうち、どれを利用すればよいかについて考察した。

参考文献

- [1] 伊藤他. 音声対話システム構築のための実対話データ収録実験. 音声言語情報処理, 94-SLP-2, 1994.
- [2] 上條他. 休止を処理の単位にした自由発話理解. 情報処理学会第 50 回全国大会講演論文集, pp.3-91-3-92, 1995.
- [3] 国立国語研究所 (編). 話し言葉の文型 (1), 国立国語研究所報告, 第 18 巻. 秀英出版, 1960.
- [4] 竹沢他. 自然発話の言語現象と音声認識用日本語文法. 音声言語情報処理, 95-SLP-6-1, 1995.
- [5] 中川聖一. 情報理論の基礎と応用. 近代科学社, 1992.