

解 説



## 音声言語情報処理の現状と研究課題

3. 音声言語処理における頑健性<sup>†</sup>河原達也<sup>††</sup> 松本裕治<sup>†††</sup>

## 1. はじめに

従来の多くの音声・自然言語解析技術は、システムが持つ辞書や文法によって定義される適格な文を対象としてきた。しかし、現実の入力においては、様々な非文法的な表現や誤りが含まれる。特に、音声言語（話し言葉/自由発話）においては、人間が文法的に誤りと判断できるような現象も、自然に発せられる。そこで本稿では、これら不適格 (ill-formed) な発話に対する頑健性を指向したアプローチについて概観する。

解析システムが頑健であるとは、入力が不適格であったり、システムが完全な解析に失敗した場合でも、何らかの処理結果を返せる能力を備えていることである。

## 2. 音声言語における不適格性

音声言語処理の頑健性を論じる前に、音声言語に見られる種々の不適格な現象の分類を行う。これを表-1に示す。

まず解析のレベルに関する分類として、形態素・構文・意味・語用論のレベルを考える。なお本稿では、音響レベルの不適格性は扱わない。不適格性の種類については、情報の提示の仕方の観点から、情報の不足、過多、および言語的な誤りに分類できる。このうち、情報の不足については、一意に解釈することができない曖昧性・漠然性の問題と、必要な情報の省略に分けられる<sup>4)</sup>。

これとは別に、解析システムの能力不足のために入力を解析できない場合がある。これは、(システムに対して) 相対的な不適格性と呼ば

れる<sup>24)</sup>。

しかし、人間が逸脱と認めることのできる絶対的不適格性についても、話し言葉においては、普遍的に出現する現象が多数ある。特に、情報過多に分類される(間投語の挿入などの)現象や、(助詞などの)軽微な省略・誤りについては、頻度も多く、対処する必要性が大きいと考えられる。音声言語処理における頑健性とは、これらへの対処能力を含むものである。

ただし本稿においては、自然言語においてよく取り上げられる曖昧性・漠然性の問題については、最小限の議論しか扱わない。

ここで示した不適格性の分類は、本稿で扱う対象の範囲を示すためのもので、詳細な説明や議論は行わないが、最近、音声対話コーパスを対象にいくつかの分析がなされているので、それら<sup>1)-3)</sup>を参照されたい。

以下、3.では、主としてテキストを対象とした構文レベル以降の不適格性について扱う。4.では、3.の議論に基づいて、形態素レベルも含めて、音声言語を対象とした不適格性への対処について議論する。

## 3. 文法的不適格文の自然言語処理手法

文法的不適格文を処理するアプローチとして、不適格文の構造を予測し、適格文と同様に文法規則として記述するアプローチもある。しかし、常にすべての不適格性の可能性を考慮しながら解析を行うために、処理に無駄があり、大きなシステムの実現は困難である。

また、統計的な手法やパターンマッチング的な解析など、画一的な処理のみを行い、どのような入力に対しても何らかの結果を返すことを保証するアプローチがある。これは、言語をより浅いレベルで捉えることによって頑健性を増す方向の研

<sup>†</sup> On the Robustness of Spoken Language Processing by Tatsuya KAWAHARA (Kyoto University) and Yuji MATSUMOTO (Nara Institute of Science and Technology).

<sup>††</sup> 京都大学工学部情報工学教室

<sup>†††</sup> 奈良先端科学技術大学院大学情報科学研究科

表-1 音声言語の不適合性の分類

	曖昧性・漠然性 (情報不足)	省略 (情報不足)	重複 (情報過多)	言語的な誤り (エラー)	システムの 能力不足
形態素	同音異義語 区分化曖昧	省略語	間投語 繰返し 言い直し	言い誤り	未知語
構文	係り受け曖昧 数量詞・限量詞 否定の範囲	助詞省略 中止文・中断 断片文・箇条書	挿入句	数・人称不一致 助詞誤り 語順誤り・倒置	システムの 構文を逸脱
意味	複数語義 比較基準 比喩・換喩	必須格の省略	同格 言い換え	選択制限違反	語彙知識不足 世界知識不足
語用論	背景・視点依存 照応曖昧 間接的言語行為	文脈による省略 ゼロ代名詞	あいづち 相手の繰返し 相手の先回り	状況的誤り	文脈的要因

究であるが、本章では扱わない。

本章で対象とするのは、以下の考え方に基づくものである。

解析が失敗するのは、文法規則や単語における言語情報の記述の詳細度が不十分なためであり、より詳細な情報を与えることにより頑健性を改善する。ただし、システムが持つ言語的な制約を強めるのではなく、不適格文を処理するために緩和しなければならない制約を、いかに発見し処理するかを研究する。

この視点における文法的不適格文の処理のアプローチは、次のように分類できる。

**部分解析** 全体の解析に失敗した場合に、解析途中で得られた句や節などの断片を集めて、それを中心に意味の抽出を試みる<sup>9),11),12),19)</sup>。

**制約緩和** 適格文のみを解析する文法にしたがって解析を行い、失敗した場合に、エラー回復のため、制約を緩和する処理を行う。

**緩和規則法** 失敗時のエラー回復のため、制約を緩和する規則(メタ規則)を用意する<sup>17),24)</sup>。

**段階的緩和** 複数のレベルの制約を用意し、解析の失敗にともなって段階的に制約を緩和することにより制約違反の原因を同定する<sup>5),16)</sup>。

**チャート法の拡張** チャート法などの構文解析アルゴリズムを拡張して、構文的制約を緩和することによって正しい構文構造を得る<sup>8),13),14),20),21),23)</sup>。

**意味主導** 意味的な情報を構文情報より優先す

る。意味的情報を中心に考えて、それを実現する構文規則の選択を行う<sup>15),18)</sup>。

**連続的解釈** 意味的・構文的な制約を連続的なものと考え、適格文と不適格文を明示的に区別しない。制約は必ずしも満たされる必要はなく、優先度が高いものと考え<sup>6),7),22)</sup>。

**アブダクション** 文の解釈のための一般的な枠組を用意し、文の意味理解の過程として不適格文の理解を考える<sup>9),10)</sup>。

以下では、上記のうちで代表的なものの概要を述べる。

Jensenらは、ともかくすべての文に対して何らかの結果を返す Fitted Parse という考え方を提案している<sup>12)</sup>。適格文のみを記述した核文法(core grammar)を用いて文を上昇型並列に解析し、解析が成功しなかった場合、それまでに得られた構造の中から一番もっともらしい構造の列を選び、それをその後の処理の入力とする。この操作が fitting 処理と呼ばれる。この方法の特徴は、解析システムを、引き続き処理の前処理として位置付け、ともかく何らかのもっともらしい結果を与えることである。McDonald<sup>19)</sup>や Hobbs<sup>9)</sup>も、部分的に成功した解析結果を用いて、それを中心に文の復元を試みたり、その一部から有用な情報を抽出することを提案している。これらの方法は部分解析(partial parsing)と呼ばれる。

その他のアプローチの多くは、文の不適合性の原因を推定し、それが犯している制約を緩和(relaxation)することによって、その原因の同定と修正を行う。したがって、適格文を対象とし

た文法によって文の解析を行い、それが失敗した段階で、不適格文の同定の処理を行う。

Kwasny ら<sup>17)</sup>および Weischedel ら<sup>24)</sup>は、ATN によって入力文の解析を行い、失敗時の状況に応じた緩和条件を記述したメタ規則を起動する機構を導入している。メタ規則として不適格文の処理を統一的に取り扱ったことは優れているが、ATN を対象にしているため、下降型かつ左から右への後戻り解析を行うことになる。解析が失敗したときに、どの部分にメタ規則を適用すればよいか必ずしも明確ではない。また、同じメタ規則が繰り返して適用されることがあり得る。

適格文を解析する文法だけではなく、種々の不適格性を判断する不適格文用の文法を用意しておくのは素直な考え方であるが、それらをいつどのように用いるかは自明ではない。工藤ら<sup>16)</sup>は、複数のレベルの文法を段階的に用いることにより、不適格性を含む文の解析とその原因の同定を行う枠組を提案している。

緩和法を段階的に用いる方法は、Douglas ら<sup>5)</sup>によっても提案されている。PATR と呼ばれる単一化文法を対象にし、そこに記述する制約条件にいくつかのレベルを設けている。解析が失敗すると、制約を段階的に緩和し、不適格文の処理と不適格性の原因の同定を行う。制約条件を素性構造のような統一的なデータ構造で記述する点で、個別のメタ規則を持つ一般的な緩和法より優れているが、処理可能な不適格性の範囲は限られる。

構文的な制約の緩和については、Mellish<sup>21)</sup>によるチャート法を拡張した不適格文処理がよく知られている。ただし、彼の方法が扱っているのは、語の欠落と冗長な追加という構文レベルの不適格性だけである。基本的には上昇型チャート法によって文の解析が行われ、それが成功しない場合に下降型の規則適用と語彙カテゴリの挿入や削除が行われる。加藤<sup>13),14)</sup>は一般の上昇型解析が失敗した場合に、なるべく多くの部分解析結果が上昇型処理によって得られるように、文法規則の左隅以外のカテゴリからの活性弧生成や活性弧同士との結合を行うことを提案している。

Hobbs らの提唱するアブダクションに基づく文の解釈の研究<sup>9),10)</sup>は、不適格文の解析のために様々なレベルの知識を用いる可能性を与える考え方である。文の解釈は、アブダクティブな解釈過

程として考える。この考え方は、上で述べた構文的な不適格文の救済法や緩和法に基づく不適格文処理などの一般的な拡張と捉えられる。また、彼のアブダクションの枠組ではコストという概念が導入されており、これにより処理の順序が制御される。つまり、処理順序のヒューリスティックの情報があわせて記述される。コストの割当の方法やコストを動的に決定する仕組みがあまり考えられていないが、構文・意味・談話に関する処理をすべて同一の枠組で取り扱える点に特徴がある。

#### 4. 不適格な発話に対する音声言語処理手法

##### 4.1 音声に対する言語処理

頑健な音声言語処理について説明する前に、音声に対する言語処理とテキストに対する言語処理との違いについて簡単に述べる。

まず、テキスト入力の場合に加えて、音声認識誤りに起因する不適格性も扱う必要がある。また、音声認識結果の向上のためには、認識時に言語的制約の利用が不可欠であり、このテキストを得る段階を含めて全体として考える必要がある。これについては、4.4 で議論する。

次に、同じく音声認識の性能・効率上の要請から、連続音声を対象とする場合、テキスト入力の場合より、さらにタスクドメインや語彙を制限せざるをえない。

タスクドメインを限定することにより、タスクの知識を利用することで、曖昧性の多くは減る。また、意味の知識と構文の知識の区別はあまりなされず、タスクの知識も埋め込んだ統合的な知識表現(ネットワーク文法・テンプレート・書き換え規則など)が効率的になる。

##### 4.2 文法的不適格文の扱い

文法的に不適格な文の扱いについては、基本的に3.で述べたアプローチに分類されるが、その例を示す。また、タスクドメインが限定されることにより、文法の拡張やパターンマッチング的な手法も有力であると考えられるので、それらについても述べる。

**文法の拡張** 限られたタスクドメインにおいては、不適格な現象もほとんどすべて記述できるという考え方に基づく。ただし、解析時に生成される仮説数が膨大になるので、意味レベルでの仮説のマージが必要である<sup>25)</sup>。

**パターンマッチング** 格構造フレーム<sup>26)</sup>やテンプレート<sup>27),28)</sup>を用意しておいて、文中の断片とマッチングするアプローチである。テンプレートの方がよりドメイン依存であり、たとえば、航空券予約タスクでは日時や行き先などのスロットが用意される。これらは、しばしば意味主導のアプローチと呼ばれる。

**部分解析** ネットワークパーザが失敗した場合に、重要な句に対応するノードを集める方法<sup>29)</sup>や、チャートパーザが失敗した場合に、チャート内の断片をヒュースティックに結合する方法<sup>30),31)</sup>などが提案されている。

**制約緩和** 助詞の脱落や倒置などの音声言語固有の不適格性に関するヒュースティックを用いて段階的に緩和する方法<sup>32)</sup>や、キーワード優先のパーズを行う規則を付加する方法<sup>33)</sup>などが提案されている。

**確率的解釈モデル** 表層的な統計的言語モデルとは異なり、意味<sup>34),35)</sup>や対話レベルの知識<sup>36),37)</sup>を確率的に表現する。

なお、音声認識を行いながら文法的解析を進める場合、部分解析や制約緩和において、解析の失敗を決定的に判断するのは不可能であり、その判定(制約逸脱の検出/リジェクション)<sup>38)</sup>が課題となる。

#### 4.3 形態素レベルの不適格性の扱い

未知語や、間投語を含む言い淀みに対処するためには、任意の音声で学習された garbage モデル<sup>39)</sup>や、音節連鎖に基づく未知語モデル<sup>40),41)</sup>を導入するアプローチが一般的である。また、間投語の出現傾向には偏りがあることから、頻出する間投語を語彙に登録することも考えられる。これらを不適格な部分にマッチングさせることにより、適格な文法の最小限の修正で対処できる<sup>42),43)</sup>。

#### 4.4 音声言語理解システムの構成

不適格な発話を含む自由発話を対象とした、音声認識も含めた音声言語理解システム全体の制御方式(探索戦略)を分類すると以下のようになる。

##### 1 パス探索[文法の拡張]

言語的・タスクの知識を統合し、音声認識時に制約として用いる方式である。効率がよく、性能もよいと考えられるが、制約が画一的であるので、頑健性をめざすと制約が弱くなり認識精度や

処理効率が低下するというトレードオフに陥る。

##### 統計的言語モデル→文法的解析

単語や品詞の N-gram などの表層上の統計的言語モデルにあらゆる知識を反映させて、どんな入力に対しても、もっともらしい音声認識結果を得るようにする。米国では、大規模コーパスの整備にともない、本アプローチが主流である<sup>44)</sup>。その認識結果の(N-best)単語列候補を文法的解析に用いるのが普通であるが、中間結果をヒュースティックとしてパーズを行う方式も考えられる<sup>45)</sup>。

##### 緩い文法→文法的解析

助詞脱落や倒置などの不適格性を考慮した文法<sup>32)</sup>や、文節レベルの文法のみ<sup>46)</sup>を最初に用いて音声認識を行い、得られるテキストに対して解析を行う。

##### 適格用処理→不適格用処理[部分解析][制約緩和]

適格文用の処理が失敗したときにはじめて、不適格文用の処理が起動される。上記の方式のバックアップ処理と捉えられる。

##### スポッティング→解析[パターンマッチング]

最初から不適格な部分をスキップして、タスク内のキーワード<sup>47)</sup>や格に対応する句<sup>48)</sup>を音声から抽出していく。ただし、スポッティング時の言語的制約が弱くなるので、大語彙を対象とした場合に認識精度を維持するのが課題である。したがって、統計的言語モデルなどを適用してから、キーワードや句を抽出することが考えられる<sup>27),48)</sup>。

## 5. おわりに

本稿では、不適格な発話に対する頑健性を指向した種々の音声言語処理手法について概観した。

2. で述べたように、話し言葉における不適格性の範囲は必ずしも明確ではないので、話し言葉の分析・モデル化を今後さらに進めていくことが重要である。

処理システムにおいては、頑健性を安易に実現するために解析手法や知識表現がタスク依存・アドホックになりがちであるが、言語レベルのモデル化とタスクに依存した知識の利用を切り分けた見通しのよい方法論が必要である。

## 参考文献

- 1) 竹沢寿幸他: 音声言語データベースを用いた自然発話の言語現象の調査, 人工知能学会, SLUD-9403-3 (1995).
- 2) 上條俊一他: 音声対話データの分析と発話理解への応用, 情報処理学会研究報告, SLP-3-6 (1994).
- 3) 村上仁一, 嵯峨山茂樹: 自由発話音声認識における音響的および言語的な問題点の検討, 信学技報, SP 91-100 (1991).
- 4) 田中穂積他: 音声対話表現における多様性, 重点領域研究「音声対話」WG 資料 (1994).
- 5) Douglas, S. and Dale, R.: Towards Robust PATR, COLING, pp. 468-474 (1992).
- 6) Fass, D. and Wilks, Y.: Preference Semantics, Ill-Formedness, and Metaphor, Computational Linguistics, Vol. 9, No. 3-4, pp. 178-187 (1983).
- 7) Genthial, D. et al.: Contribution of a Category Hierarchy to the Robustness of Syntactic Parsing, COLING, pp. 139-144 (1990).
- 8) Goeser, S.: Chart Parsing of Robust Grammars, COLING, pp. 120-126 (1992).
- 9) Hobbs, J.R. et al.: Robust Processing of Real-World Natural-Language Texts, Applied Natural Language Processing, pp. 186-192 (1992).
- 10) Hobbs, J.R. et al.: Interpretation as Abduction, Artificial Intelligence, Vol. 63, No. 1-2, pp. 69-142 (1993).
- 11) Imaichi, O. and Matsumoto, Y.: Integration of Syntactic, Semantic and Contextual Information in Processing Grammatically Ill-Formed Inputs, IJCAI (1995).
- 12) Jensen, K. et al.: Parse Fitting and Prose Fixing: Getting a Hold on Ill-formedness, Computational Linguistics, Vol. 9, No. 3-4, pp. 147-160 (1983).
- 13) 加藤恒昭: 非文の解析—チャートに基づく新たな手法, 情報処理学会研究報告, NL-83-10 (1991).
- 14) Kato, T.: Yet Another Chart-Based Technique for Parsing Ill-Formed Input, Applied Natural Language Processing, pp. 107-112 (1994).
- 15) Kirtner, J. D. and Lytinen, S. L.: ULINK: A Semantic-Driven Approach to Understanding Ungrammatical Input, AAAI, pp. 137-142 (1991).
- 16) Kudo, I. et al.: Schema Method: A Framework for Correcting Grammatically Ill-Formed Input, COLING, pp. 341-347 (1988).
- 17) Kwasny, S. C. and Sondheimer, N. K.: Relaxation Techniques for Parsing Grammatically Ill-Formed Input in Natural Language Understanding Systems, Computational Linguistics, Vol. 7, No. 2, pp. 99-109 (1981).
- 18) Lytinen, S. L.: Semantic-first Natural Language Processing, AAAI, pp. 111-116 (1991).
- 19) McDonald, D. D.: An Efficient Chart-based Algorithm for Partial-Parsing of Unrestricted Texts, Applied Natural Language Processing, pp. 193-200 (1992).
- 20) Meknavin, S. et al.: Parsing Ill-Formed Input with ID/LP Rules, Natural Language Understanding and Logic Programming, pp. 158-171 (1993).
- 21) Mellish, C.: Some Chart-Based Technique for Parsing Ill-formed Input, ACL, pp. 102-109 (1989).
- 22) Schank, R. C.: Conceptual Information Processing, North Holland (1975).
- 23) Stock, O.: Parsing with Flexibility, Dynamic Strategies, and Idioms in Mind, Computational Linguistics, Vol. 15, No. 1, pp. 1-18 (1989).
- 24) Weischedel, R. M. and Sondheimer, N. K.: Meta-rules as a Basis for Processing Ill-Formed Input, Computational Linguistics, Vol. 9, No. 3-4, pp. 161-177 (1983).
- 25) 南 泰浩他: 自由発話音声認識における意味を考慮した2段LRパーザ, 音講論, 3-4-10 (Mar. 1993).
- 26) Hayes, P. J. et al.: Parsing Spoken Language: A Semantic Caseframe Approach, COLING, pp. 587-592 (1986).
- 27) Ward, W.: Understanding Spontaneous Speech: The PHOENIX system, ICASSP, pp. 365-367 (1991).
- 28) Jackson, E. et al.: A Template Matcher for Robust NL Interpretation, DARPA Speech & Natural Language, pp. 190-194 (1991).
- 29) Seneff, S.: Robust Parsing for Spoken Language Systems, ICASSP, pp. 189-192 (1992).
- 30) Stallard, D. and Bobrow, R.: Fragment Processing in the DELPHI System, DARPA Speech & Natural Language, pp. 305-310 (1992).
- 31) Stallard, D. and Bobrow, R.: The Semantic Linker—a New Fragment Combining Method, ARPA Human Language Technology, pp. 37-42 (1993).
- 32) 山本幹雄他: 自然発話の意味理解と対話システム, 情報処理学会研究報告, SLP-2-13 (1994).
- 33) Linebarger, M. C. et al.: A Portable Approach to Last Resort Parsing and Interpretation, ARPA Human Language Technology, pp. 31-36 (1993).
- 34) Pieraccini, R. et al.: A Speech Understanding System Based on Statistical Representation of Semantics, ICASSP, pp. 193-196 (1992).
- 35) Miller, S., et al.: Statistical Language Processing Using Hidden Understanding Models, ARPA Human Language Technology, pp. 278-282 (1994).
- 36) 川端 豪: 確率文法と話題マルコフモデルに

- 基づく音声認識のための話題制御, 信学論, Vol. J 77-DII, No. 10, pp. 1967-1972 (1994).
- 37) 永田昌明: 統計的な対話モデルの試みとその音声認識への応用, 人工知能学会, SLUD-9202-11 (1992).
- 38) 渡辺隆夫, 塚田 聡: 音節認識を用いたよう度補正による未知発話のリジェクション, 信学論, Vol. J 75-DII, No. 12, pp. 2002-2009 (1992).
- 39) Wilpon, J. G. et al.: Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models, IEEE Trans. ASSP, Vol. 38, No. 11, pp. 1870-1878 (1990).
- 40) Asadi, A. et al.: Automatic Modeling for Adding New Words to a Large-Vocabulary Continuous Speech Recognition System, ICASSP, pp. 305-308 (1991).
- 41) 伊藤克亘他: 連続音声認識における未知語の扱い, 信学技報, SP 91-96 (1991).
- 42) 武田一哉, 小沼知浩: 自由発話文理解のための garbage HMM の利用の検討, 信学技報, SP 92-127 (1993).
- 43) 甲斐充彦, 中川聖一: 日本語連続音声認識システム SPOJUS-SYNO の改良と評価, 信学技報, SP 93-20 (1993).
- 44) Charniak, E.: Statistical Language Learning, MIT Press, Cambridge (1993).
- 45) 河原達也他: 単語対制約をヒューリスティックとする A \* 探索に基づく会話音声認識, 信学論, Vol. J 77-DII, No. 1, pp. 1-8 (1994).
- 46) ATR 編: 自動翻訳電話, オーム社 (1994).
- 47) 坪井宏之他: キーワードスポッティングに基づく連続音声理解, 信学技報, SP 91-95 (1991).
- 48) 河原達也他: フレーズスポッティングに基づく頑健な音声理解, 情報処理学会研究報告, SLP-4-6 (1994).  
(平成7年3月1日受付)



河原 達也 (正会員)

1964年生。1987年京都大学工学部情報工学科卒業。1989年同大学院修士課程修了。1990年同大学院博士後期課程退学。同年京都大学工学部助手。1995年同助教授。音声認識・理解の研究に従事。京大博士(工学)。電子情報通信学会, 日本音響学会, 人工知能学会, IEEE 各会員。



松本 裕治 (正会員)

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