

ことばは形式意味を経て形になる /なっちゃってオロオロ!!

羽尻公一郎
立命館大学

August 20, 2002

Abstract

話し言葉や言語学の例文などの病的な文は、書き言葉に対して圧倒的に意味が倒錯している。しかし、倒錯していても、部分的には整合している。全体の倒錯と部分の整合。これが形式意味として病的な文を解析するときの問題である。本稿では係り受けなどの素朴な統語解析と、局所的な形式意味解析を用いて、文全体の形式意味を古典論理、モンタギュー論理、ファジー論理など様々なモデルに適用し、形式意味の値をトポロジーに変換し、そのトポロジー情報から構造を可視化するシステムを紹介する。可視化された形は形式意味の病的具合により、単純な図形からメビウスの環やクラインの壺、エッシャーの絵画の背後にあるペンローズの不可能図形などになる。その形をさらに直感的に理解することを促進するため、象形文字と品詞を対応させる。象形文字はシニフィアンとして用いるに過ぎず、任意の図形をあたえることが可能であるが、恣意性を縮退させるために用いているに過ぎない。

1 はじめに

ことばは本来線状的に表現される。これは話し言葉でも書き言葉でも同じである。ことばは最初、話し言葉が発現し、次に書き言葉が現れたとされている。話し言葉は空気を媒体とした音声で伝達されるため、1次元にしかコーディング出来ない。この制約がことばを線状的にしたのである。しかし書き言葉を獲得する過程で、人間は形を記号にするプロセスを経て、文字による表現を獲得した。その過程は通時言語学に詳しい。ヒエログリフやトンパ文字などはそのようなプロセスの副産物として現代に残っている。

ことばを直観的に理解したいという願望は、このようなプロセスの果てに現れた、実に皮肉なものである。そして、世の表現者たちは、言語の呪縛、記号の呪縛から解放されるための努力を今日も行なっている [3, 6]。言語学や論理学、計算言語学などは、図による推論や図と論理の結合 [14] などに腐心しているが、上記の要求に対して効果的な手段を提示するには至っていないのが現状であろう。

一方、形、トポロジーの数理的解釈のための記号化は、古くから行なわれてきた [8, 15, 19]。そして、画像理解などの分野において一定の成功を収めている [7, 16]。しかし、これらのアプローチはあくまで形の数理的解釈に留まっており、ことばと形を結び付けるまでには至っていない。

本稿ではこのような背景を踏まえて、文の形式意味をトポロジーに変換し、さらにトポロジーを形に変換する手法およびそのシステムの実装について述べる。本稿で提案される手法は、極めて単純なものであるが、ことばを形に変換することで得られる直観的理解に対して、ダイレクトなインパクトを与える。

2 アルゴリズム

本章では発話された音声を解析し、図形に変換するまでのアルゴリズムについて述べる。

2.1 音声認識

発話された音声は、標準的な HMM + n -gram を用いて、漢字仮名交じり文へと変換される。音声認識自体は現在確立されている中でも標準的かつ頑健な、一般的なものである。

2.2 パージング

音声から変換された漢字仮名交じり文は、形態素解析 [23] を経て、古典的な係受け解析を用いてパージングされる。パージングは IPA の辞書 [21] と係受けの制約 [10, 11, 9] を用いて行なわれる。ここで、品詞情報、係受け文構造情報が抽出される。これらの情報をもとにして、次の形式意味処理への情報が準備される。

2.3 形式意味解析

パージングによって得られた命題情報および統語構造に対して、語用論の立場からの情報を人間が提供した上で、古典二値論理、モンタギュ論理、ファジー論理の3種のうちのいずれかを選択し、入力された文の形式意味の値を算出する。古典二値論理の場合は真か偽かの2値、モンタギュ論理の場合は真偽の2値について、可能性を独立に、すべて列挙する。ファジー論理の場合は0から1までの閉区間の連続値をとる。

3つのいずれも、代表的な形式意味論であり、それらを動的論理、様相論理、時相論理、状況論理などに拡張することは平易である。また、文のみではなく、DRTなどを用いることで文脈の形式意味解析に拡張することも可能である。

2.4 病性解析

発話される文は常に well-formed であるとは限らない。むしろ、話し言葉などにおいては圧倒的に ill-formed である場合がほとんどである。しかし、発話された文が病的かどうかを判断するのは、あまり容易ではない。そこで、新たに病性解析のアルゴリズムを提案し、文の病性を判断する。

病的な文とは、言語学の例文に用いられるような非日常的な文や、言い間違いや言い直しなどを含む文である。例えば、

1. この文は偽である。
2. 私は嘘つきだ。
3. 男の子が、女の子が男の子を押している。
4. 象は鼻が長い。
5. 色のない緑の概念が猛烈に眠る。

などは形式意味論の世界の閉じた観点からは、全て病的であると判断される。(1)と(2)は「この」や「私」など、文の部分でありながら文全体を参照する矛盾を抱えた再的帰機能を有した語を含んでおり、その形式意味解釈が無限退行を引き起こす [4]。(3)は言い直しの一種で、合金 (Alga ma) と呼ばれるものであり、表層的には格助詞「が」によってマークされる主辞が2つ存在する [17]。(4)は「は」「が」の二つの助詞により、 $\forall$ と $\exists$ の対立が発生する [20]。(5)はナンセンス文の典型であり、文脈なくして、語の関係からは意味が成立しない。

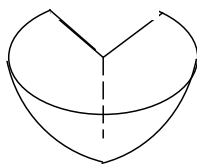


Figure 1: 形状の例

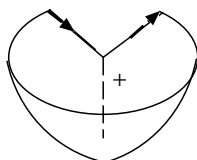


Figure 2: ラベルづけされた形状の例

2.5 トポロジ抽出

文献[8, 15, 19]などにより，形状を記述する形式が提案され，すでに広く様々な分野に浸透している．例えば図1は頂点や辺，面の制約(図3参照)から図2のようにラベルづけされ，式2.5のように記述できる．

$$a_k x_\beta + b_k y_\beta + z_\beta + c_k > 0$$

2.6 構造化

トポロジが決定されると，任意のパラメタを恣意的に与えれば，形状が形成できる．形状に対してどのように任意のパラメタを与えるかは，形状の見易さ，分かり易さなどを考慮することである程度絞り込める．例えば文献[25]では，不可能図形の知覚における違和感についての分析を行なっている．このような認知的分析を参考に，認知的負荷の少ない図形を形成するようにパラメタ決定に対してヒューリスティックなメタアルゴリズムを与える．

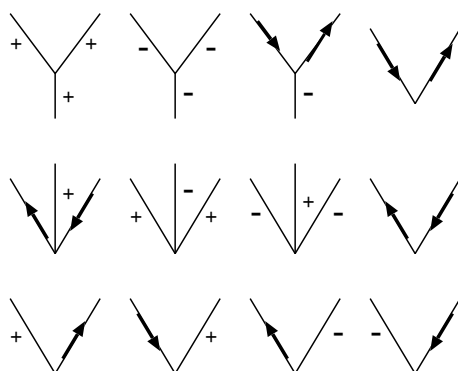


Figure 3: 可能な頂点の接合のリスト

2.7 形状操作

形状を形成すれば、あとはその形状を操作、例えば視点の変更や形状の表面処理、すなわち陰影処理やテクスチャの選択などの問題が残るだけである。これらは解釈者によって好みが分かれるところであるので、これらの操作はインタラクティブに行なうこととする。

3 システムの概要

前章で提案されたアルゴリズムは、計算機に実装されている。計算機自体は標準的なPC/ATマシンであり、OSにはLinuxを用いている。音声認識部には(株)日本IBMのViaVoiceを用いた。形態素解析および統語解析は奈良先端科学技術大学院大学で開発された「茶筌」「カボチャ」を拡張したものをを用いている。文解析結果である命題情報と統語構造をもとに、病性判断とトポロジ情報抽出のモジュールを新たに開発し、Objective-Cを用いて実装した。さらに形状形成にはPov-Ray システム [13] を用いており、インターフェイスはWebのcgiを用いてInternet ExplorerTM, Netscape NavigatorTMなどのブラウザを解して形成されており、インタラクティブな操作を可能にしている。

4 実装

4.1 パージング/係受け解析

音声認識・形態素解析を終えた発話は、パージングされ統語情報を抽出される。統語解析に際しては、CFGなどの計算クラスよりもより柔軟かつ日本語に適合している係受けの概念を用いる。係受けは単純な語同士の依存構造を解析する。命を主辞と述辞の2種類に分類し、主辞が述辞にたいしてどのように依存しているかのみを解析する。この解析はCFGのようなwell-formedな文よりも、話し言葉などのill-formedな文により向いている。係受け解析の詳細な説明は文献[10, 11, 9]などに譲る。

4.2 形式意味解析/局所性と大域性・病性検証

命題情報と係受けの統語情報をもとに、発話された文の形式意味解析を行なう。この際、文が病的かどうかについての判断も同時に行なう。発話された文が病的かどうかについて解析するにあたって、文献[24]の提案する日本語向けの論理体系を用いる。文献[24]によれば、「自分」は再帰詞として定義されており、ゆえに無限退行を引き起こす因子となっている。また、指示詞についても前方/後方/零照応などのトリヴィアルな機能以上の、文に対する病的な機能を有していることが指摘されている。こうした詳細な文献[24]の日本語に対する形式意味定義(付録参照)を用いることで、文の病性を検証する。

また、言い間違いや言い直しなどの病性に関しては、局所的意味整合と大域的意味不整合という部分と全体に関するメソスコピックな解析を行なう。これは文献[12]の提案するwell-formedness ruleに依拠する。なお、日本語における言い間違い、言い直しなどとwell-formedness ruleの関係については文献[18]を参照されたい。

病性検証については、図4のような手続きを経る。

4.2.1 論理選択/古典論理・モンタギュ論理・ファジー論理

形式意味解析および病性検証がおわると、その形式意味構造をどのような論理モデルにあてはめるかを選択する。論理モデルは多岐に渡るが、今回はその中でも基本的かつ代表的な、古典二値論理、モンタギュ論理、ファジー論理のいずれかを選択する。この選択は解釈者に任せる。

古典二値論理を構造的に表すには、真偽値が面の表と裏に対応させるのが相応である。そこで、古典(二値)論理は文の真偽をリボンの裏と表に対応させる。モンタ

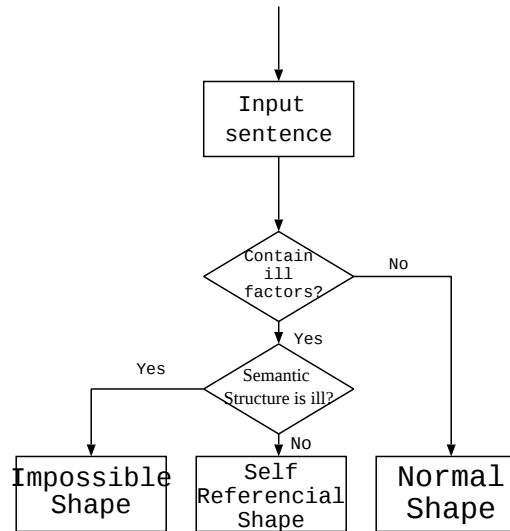


Figure 4: 文の病性解析の手続き

ギョ論理は古典論理の可能世界解釈への拡張なので、先のリボンが複数並列に並ぶものとする。また、ファジー論理は真偽値が0から1までの閉区間の連続値なので、構造的にはパイプもしくは円柱の表面の一点に真偽値を位置させるのが相応である。そこで、ファジー論理では3次元のパイプ構造に対応づける。モンタギョ論理とファジー論理を組み合わせる場合には、複数のパイプが複数並列に並ぶものとする。

4.3 トポロジ抽出/構造の制約検証

形式論理モデルに指定された発話の命題の真偽値などの情報と、パーズングの際に得られる品詞情報、および病性検証の結果を用いて、トポロジ情報を構成する。その際、構成されたトポロジ情報が構造的に実世界での制約を満たしているかどうか、構造の制約検証を行なう。構造の制約検証のアルゴリズムを図5に示す。この構造の制約検証によって、構成される形の特徴が決定される。例えば自己言及による無限退行が発生する場合、古典論理を選択すると、そのトポロジーはメビウスの環(図6)を形成する。ファジー論理を選択したら、クラインの壺(図7)となる。また、病性解析の結果局所的整合と大局的不整合が検出されたなら、そのトポロジーは局所的に整合しているが大局的に不整合な不可能図形(例えば図8)のようになる。

4.4 形状形成/恣意性と象形文字

トポロジの抽出と病性検証が終ると、トポロジの表現が得られる。これに任意に設定されたデフォルトの値を与えることで、形状を形成できる。形状の形成にはPov-Rayシステム[13]を用いる。さらに、形状の直観的理解を促進するために、テクスチャとしてトンパ文字[5]を用いる。トンパ文字は中国の地方で用いられている象形文字で、そのシニフィアンとしての機能は、極めて直観的である。このトンパ文字とパーズングの結果の品詞情報を対応づけることで、形状に対してテクスチャを張り付けることができる。

```

While E is not empty do
begin
  select and delete an element e from E;
  (let the two terminal junctions of e be s and t);
  if J(s) has elements that are not locally consistent with J(t) then
  begin
    delete all such elements from J(s);
    add to E all the lines other than e that emanate from s
    if they are not in E;
  end;
  if J(s) has elements that are not locally consistent with J(s) then
  begin
    delete all such elements from J(t);
    add to E all the lines other than e that emanate from t
    if they are not in E;
  end;
end;
end;

```

Figure 5: 構造の制約を検証するアルゴリズム

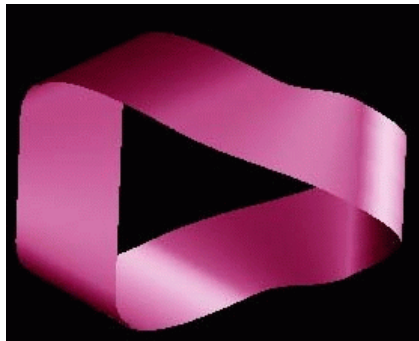


Figure 6: メビウスの環

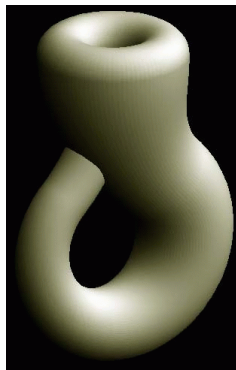


Figure 7: クラインの壺



Figure 8: 不可能図形

5 動作例

本章では先にあげた病的な文をシステムに入力して得られた形状およびその処理の中間過程を示す。

5.1 この文は偽である．

- パージング結果

```
この コノ この 連体詞
文 ブン 文 名詞-一般
は ハ は 助詞-係助詞
偽 ニセ 偽 名詞-一般
で デだ 助動詞 特殊・ダ 連用形
ある アル ある 助動詞 五段・ラ行アル 基本形
． ． ． 記号-句点
EOS
```

```
この-D
文は-D
偽である．
EOS
```

- 病性検証結果

```
ill
<because ill-functional word contains>
```

- 形状形成結果

To be shown in the presentation...

5.2 私は嘘つきだ．

- パージング結果

```
私 ワタシ 私 名詞-代名詞-一般
は ハ は 助詞-係助詞
嘘つき ウソツキ 嘘つき 名詞-一般
だ ダだ 助動詞 特殊・ダ 基本形
```

. . . 記号-句点
EOS

私は-
嘘つきだ.
EOS

- 病性検証結果

ill
<because ill-functional words contains>

- 形状形成結果

To be shown in the presentation...

5.3 男の子が、女の子が男の子を押している.

- パージング結果

男の子 オトコノコ 男の子 名詞-一般
が ガ が 助詞-格助詞-一般
, , , 記号-読点
女の子 オンナノコ 女の子 名詞-一般
が ガ が 助詞-格助詞-一般
男の子 オトコノコ 男の子 名詞-一般
を ヲ を 助詞-格助詞-一般
押し オシ 押す 動詞-自立 五段・サ行 連用形
て テ て 助詞-接続助詞
いる イル いる 動詞-非自立 一段 基本形
. . . 記号-句点
EOS

男の子が, -----D
女の子が---D
男の子を-O
押している.
EOS

- 病性検証結果

ill <because global semantics is inconsistent>

- 形状形成結果

To be shown in the presentation...

5.4 象は鼻が長い.

- パージング結果

象は鼻が長い.
EOS
象 ゾウ 象 名詞-一般

は ハ は 助詞-係助詞
 鼻 ハナ 鼻 名詞-一般
 が ガ が 助詞-格助詞-一般
 長い ナガイ 長い 形容詞-自立 形容詞・アウオ段 基本形
 . . . 記号-句点
 EOS

象は鼻が長い.
 象は---D
 鼻が-O
 長い.
 EOS

- 病性検証結果

ill <because plural functional words
 are making corrosion>

- 形状形成結果

To be shown in the presentation...

5.5 色のない緑の概念が猛烈に眠る.

- パージング結果

色 イロ 色 名詞-一般
 の ノ の 助詞-格助詞-一般
 ない ナイ ない 形容詞-自立 形容詞・アウオ段 基本形
 緑 ミドリ 緑 名詞-一般
 の ノ の 助詞-連体化
 概念 ガイネン 概念 名詞-一般
 が ガ が 助詞-格助詞-一般
 猛烈 モウレツ 猛烈 名詞-形容動詞語幹
 に ニ に 助詞-副詞化
 眠る ネムル 眠る 動詞-自立 五段・ラ行 基本形
 . . . 記号-句点
 EOS

色の-D
 ない-D
 緑の-D
 概念が---D
 猛烈に-O
 眠る.
 EOS

- 病性検証結果

ill
 <because the sentence is semantically nonsense>

- 形状形成結果

To be shown in the presentation...

6 考察

形の認知や形の意義などについては、文献 [2, 1, 25] など、数多の文献などで議論されている。それらは殆どが、形・図形とはなにか、そして人間はそれをどう捉え、操作しているのかといった観点からのものである。

形に対して構成論的に、しかも言語との関連に焦点を当てた研究は以外と少ない [14, 8]。このような背景から、本研究のユニークネスが窺える。また、タイポグラフィなどの分野 [22] との接点もあり、単なる理論およびシステムの枠を越えて分野横断的である。

実際にことばから形成された形状の認知を解析することは、文献 [25] の提案する手法で十分に可能である。これは今後の課題としたい。また、他の課題としてはシステムの拡張、文から文脈へ、さらに語用論モデルの導入やマルチモダリティへの対応などが挙げられる。

7 おわりに

本稿ではことばを形に変換する手法を提案し、計算機上に実装した。結果として、発話された文の病性などを考慮することにより、形式意味を解してトポロジカルに文の構造を可視化することができた。本稿で提案された手法は極めて平易なものであり、また様々な不確定要素の決定を解釈者に求めるが、それをインタラクテビティの観点から見ると、発話した Overt Speech と発話者のインタクショナループが形成されることになり、システムと人間のインタクシオンを豊かなものにしている。

今後の課題としては、前章で述べたように、実際にことばから形成された形状の認知を解析すること、他の課題としてはシステムの拡張、文から文脈へ、さらに語用論モデルの導入やマルチモダリティへの対応などが挙げられる。

付録：文献 [24] による日本語形式論理の語彙 (品詞) の定義

名詞	花子, 太郎, 先生, 生徒, 本, ...
再帰詞	自分
不定名詞	$X, Y, Z, \acute{X}, \acute{Y}, \acute{Z}, \dots$
1項動詞	笑った, ころんだ, ...
2項動詞	叱った, 読んだ, ...
3項動詞	あげた, ...
個体変数	$x_1, x_2, \dots, x_k, \dots$
量化子	$\forall, \exists, M, \nabla, r$
不定量化子	*
量化表現接続詞	と, か, ...
インデックス	$1, 2, 3, \dots, k, \dots$
修飾オペレータ	:
括弧	(,)
カンマ	,

References

- [1] かたちの科学. 朝倉書店, 東京, 1987.
- [2] 特集：形とはなにか. 現代思想. 青土社, 1992.
- [3] Kathy Acker. *Blood and Guts in High School*. Grove Press, New York, 1989.
- [4] Jon Barwise. *The Liar : An Essay on Truth and Circularity*. MIT Press, Cambridge, Boston, 1989.

- [5] Personal Media Corporation. 超漢字トンパ書体. On the web, 2002. <http://www.chokanji.com/cktoppa/>.
- [6] Raymond Federman. *Take it or Leave it*. Fc2/Black Ice Books, New York, 1997.
- [7] Martin Gardner. *Penrose Tiles to Trapdoor Ciphers ...and the return of Dr. MATRIX*. W. H. Freeman, New York, 1997.
- [8] D. A. Huffman. *Impossible Objects as Nonsense Sentences*, volume 6 of *Machine Intelligence*, chapter 19, pages 295–323. Edinburgh University Press, 1971.
- [9] T. Joachims. Making large-scale svm learning practical. In *Advances in Kernel Methods*, volume 11, pages 169–, 1999.
- [10] Taku Kudoh and Yuji Matsumoto. Japanese dependency analysis based on support vector machines. In *EMNLP/VLC 2000*, 2000.
- [11] Taku Kudoh and Yuji Matsumoto. Chunking with support vector machines. In *NAACL 2001*, 2001.
- [12] Willem J. M. Levelt. *Speaking*. ACL-MIT Press Series in Natural-Language Processing. MIT Press, London, England, 1989.
- [13] POV-Team. Pov-ray:documentation. On the web, 2000. <http://www.povray.org/documentation/>.
- [14] Atsushi Shimojima. The graphic-linguistic distinction –exploring alternatives. *Artificial Intelligence Review*, 13:313–335, 1999.
- [15] Kokichi Sugihara. *Machine Interpretation of Line Drawings*. The MIT Press, Cambridge, Massachusetts, 1986.
- [16] William P. Thurston. *Three-Dimensional Geometory and Topology*. Princeton University Press, Princeton, 1997.
- [17] 羽尻 公一郎, 岡田 美智男, and 小川 均. 漸次的発話産出における言い直しとそのシミュレーション. *人工知能学会 音声理解と対話処理研究会 (SIG-SLUD)*, 1997.
- [18] 羽尻 公一郎, 岡田 美智男, and 小川 均. 日本語漸次的発話産出における格助詞補完と文法的不整合の解消. *認知科学*, 5(3):87–99, 1998.
- [19] 杉原 厚吉. **不可能物体の数理**. 森北出版, 東京, 1993.
- [20] 三上 章. **象は鼻が長い**. くろしお出版, 東京, 1960.
- [21] 浅原 正幸 and 松本 裕治. Ipadic ユーザーズマニュアル. Technical report, 奈良先端科学技術大学院大学, 2002.
- [22] 寺山 祐策. writing space design 98/99. Technical report, 武蔵野美術大学視覚伝達デザイン学科研究室, 小平市, 東京, 1999.
- [23] 松本 裕治, 北内 啓, 山下 達雄, 平野 吉隆, 松田 寛, 高岡 一馬, and 浅原 正幸. 日本語形態素解析システム「茶筌」 version 2.2.1 使用説明書. Technical report, 奈良先端科学技術大学院大学, 2000.
- [24] 飯田 隆. 日本語と論理学. Technical report, 慶應義塾大学文学部, 3 2000. 平成 9–11 年度科学研究費補助金研究成果報告書.
- [25] 小森 良子. **絵画表現における知覚的制約違反と違和感-不可能図形と矛盾した空間表現を用いて**. PhD thesis, 九州大学大学院人間環境学研究院, 1999.