

パターンのラティス下の疑似エラー補正を通じた意味構築

Meaning construction through simulated parallel error correction under pattern lattice

黒田 航
Kow Kuroda

情報通信研究機構 知識創成コミュニケーション研究所
National Institute of Information and Communications Technology
kuroda@nict.go.jp

Abstract

This article presents a memory-based model of meaning construction, called “integration through simulated parallel error correction,” that works under pattern lattice and accommodates noncompositionality straightforwardly.

Keywords — simulated parallel error correction, meaning construction under pattern lattice

1. はじめに

文 $s = w_1 \cdot w_2 \cdots w_n$ の意味はどうやって決まるのか? 一般的な想定では, s を構成する語の集合を $W(s) = \{w_1, w_2, \dots, w_n\}$ とし, s の意味を $M(s)$ で表わす時, $M(s) = f(M(w_1), M(w_2), \dots, M(w_n))$ と考える. $M(w_i)$ は w_i の意味を返す関数, f は語の意味の組みに適用される関数だとする (f が何を返すかは不問にする). このモデル化は意味構築への構成的アプローチの本質である. 更に仮定を詳細化し「 f は s の統語構造によって決まる」とも仮定される.

構成的アプローチには直観的な妥当性がある一方, 難点も知られている. 本稿は (i) コトバの意味構築の非構成性を示す現象を追加し, (ii) その事実を記述する計算論的に十分に制約されつつも一般性のあるモデルを提案しながら, それが (iii) 意味構築が根本的に非構成的である可能性を示す.

2. バイアス源としての超語彙的パターン

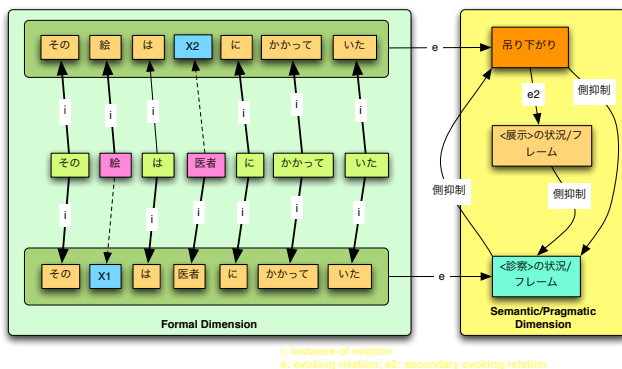


図 1 (2a) で生じる「かかる」の語義の競合のモデル: i は instance-of か = 関係を, e は evoke 関係を表わす (e_2 は派生的)

非構成性の証拠となる新たな現象とは, (1) と (2) の四つの文意味理解の際の超語彙的パターン (superlexical patterns) の影響である:

- (1) a. その(病んだ)絵は壁にかかっていた .
b. その(病んだ)男は医者にかかっていた .
- (2) a. ?*その(病んだ)絵は医者にかかっていた .
b. ?*その(病んだ)男は壁にかかっていた .

非構成性の新たな証拠となるのは, (2) の二文が逸脱していると感じられる他に, (特に(1)を読んだ後には)「かかっていた」の語義の競合によって, 解釈が得られるまでに時間がかかるという事実である¹⁾²⁾. この現象のもっとも簡潔かつ効果的な説明は, (2) では〈吊り下げ〉の状況の喚起と〈(不)定期受診療〉の状況の喚起が図 1 のように側抑制を通じて競合関係にある文理解のバイアス源だと考えることである.

これが非構成性の証拠になっていると考えられるのは, 解釈が構成的なプロセスであるならば, 問題の語義の競合は理論的に発生する余地がないからである. 注目すべきは〈吊り下がり〉や〈(不)定期受診療〉の状況の喚起は, 「その絵(は)」「その男(は)」「壁に」「医者に」「かかっていた」という個々の句による喚起の効果の積ではないという点である. ここで常套句の原則[4]という独立の記述的一般化を考慮に入れた一般化により, 次の一般的な意味構築のモデルを得る:

- (3) 事例記憶に基づく意味解釈: 表現 e に類似する (=部分一致する) 事例の集合 $E'(e) = \{e'_1, e'_2, \dots, e'_n\}$ があり, E' のうちで互いに矛盾しない最大の部分集合の集合を $\Sigma(E') = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$ とする. e の解釈は, $\Sigma(E')$ から (最適性の基準を満足するように) 選ばれた s_i の部分集合となっている事例の集合の意味の統合で与えられる.

以下では, Pattern Lattice (PL) [8] という記述モデル上でこれを実現する.

¹⁾ この「事実」は筆者の直観に基づくもので, 実験的に確認されたものではない. 妥当性の検証は今後の課題である.
²⁾ 「病んだ」があった方が効果が明確に出る.

3. Simulated Parallel Error Correction

表現 e の pattern lattice, $PL(e)$, とは instance-of の関係の下でのパターンの半順序集合である。PL の底 (bottom) は変項を一つも含まない具体事例, 頂点 (top) は定項を一つも含まない抽象的パターンであるとする。定項の数が1の時にパターンは語彙的パターンと, 数が2以上の時にパターンは超語彙的パターンと呼ぶ。詳しくは[8]を参照されたい。

§2. の語義の競合は, 底 T の表現 e に対する $PL(e)$ 上のパターンからのバイアスの強さは, T への近さ (= 変項の数の少なさ) に比例すると仮定すれば説明できる。ここで記述の対象を拡張し, すべての事例間の変項単位の編集距離を考え, それに基づいて類似度に基づく複雑ネットワークを考える。証明はないが, 変項単位の編集距離が最小の文相互類似度が最大の文であるのは明らかである³⁾。

表現 e について, 変項の実現値が異なるだけで, 残りが同一であるような事例の集合 $E' = e'_1, e'_2, \dots, e'_n$ は変項単位の編集距離が最小であるような集合である。フレーム意味論[1]に従い, E' が(3)の $\sigma_1, \sigma_2, \dots, \sigma_n$ に対応すると想定する。

語彙的パターンが特定の意味に結びついているのは明らかであるが, 超語彙的パターンはどうか? 提案したモデルでは, 底に近い超語彙的パターンも, 事例記憶に基づいた類似性の認識により, PL 上で相対的に上位の語彙的バイアス源の意味指定に還元できない形で意味に結びつく。超語彙的バイアス源の意味指定が語彙的バイアス源の意味指定に還元不能であるならば, 非構成性は無条件に成立する。逆を言えば, これが成立しないなら非構成性が成立する理由は外在的なものとなる。

以上の想定の下では k 個の定項をもつ表現 e の意味構築を次の形で定式化できる:

- (4) Simulated Parallel Error Correction (SPEC): [1] k 個の定項 x_1, x_2, \dots, x_k をもつ表現 e を基に, それぞれの定項がマスクされた k 個の超語彙的パターン p_1, p_2, \dots, p_k を生成し, e をそれらの論理積とする。[2] p_i のそれぞれについて互いに独立に (共起頻度を利用するなどして) マスクされた値の修復を試みる。結果として得られた補正を c_1, c_2, \dots, c_k とする。[3] c_j ($1 \leq j \leq k$) の論理和 (= 斉一化) を e の意味とする (c_i と c_j の指定に不一致があった場合にどう処理するかは別に定義する)。[4] p_i の修復が成功しない場合 (つまり推定される値が NULL の場合) に, p_i を $k-1$ 個の超語彙的パターン q_1, q_2, \dots, q_{k-1} に分解し, 手順を再帰的に繰り返すより, 処理がラティスの上方, つまり語彙的パターン側へ伝播する

³⁾ここでは単純化のため, 語の意味の単位での類似度の影響は無視している。

(4) がヒトの脳が言語表現の意味を構築する際の実際の処理であるとするれば, これは構文効果[7]に対する, 独立性の高い説明になる。強調に値すると思われるのは, 超語彙的パターンによる状況喚起が互いに独立している点である (そうでなければ語義の競合が起こる理由がない)。これが意味することは (i) 意味構築が並列分散的であり, かつ (ii) 統語構造の指定が冗長だということである。(i) は新規な見解ではなく, 80年代に明らかになっていたこと[3, 5]である。

4. 考察

以上のモデル化が妥当であるならば, (超)語彙的バイアス源の超語彙的パターンによる状況の喚起の実体は, (ノイズで攪乱された信号の復元と同じタイプの) エラー補正と同じ原理で働くものであり, ヒトの場合, それはおそらくパターン補完が基礎になっていると考えられる。だが, それには (例えば[2]が論じているような) 十分に豊かな事例記憶が不可欠である。

「膨大な事例記憶」に基づく言語処理システムの本質的条件は, 事例を覚えておくための記憶容量よりも, 類似度の高い事例を想起し, かつ類似度の低い事例を想起しないような効果的な想起の仕組みの有無である。それがないと (並列計算を仮定しても) 候補の絞りこみで計算資源が嵩み, システムは効果的に動作しない。脳が行なっている事例集合の特定=想起を単純な情報検索として理解するのは難しい。より見こみがあるのは, 恒常的な抑制の一時的緩和が想起を実現するという見方 (例えば[6]) であるように思われる。

参考文献

- [1] C. J. Fillmore, C. R. Johnson, & M. R. L. Petruck (2003). Background to FrameNet. *International Journal of Lexicography*, 16(3): 235–250.
- [2] J. Hawkins, and S. Blakeslee (2004). *On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*, Times Books.
- [3] C. K. Riesbeck (1986). From conceptual analyzer to direct memory access parsing: An overview. In N. E. Sharkey, ed., *Advances in Cognitive Science 1*. Ellis Horwood.
- [4] J. M. Sinclair (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- [5] S. Small (1983). Parsing as co-operative distributional inference: Understanding through memory interaction. In M. King, ed., *Parsing Natural Language*. Academic Press.
- [6] 月元敬 (2008). 抑制に基づく記憶検索理論の構成. 風間書房.
- [7] 中本 敬子, 李在鎬, & 黒田 航 (2006). 日本語の語順選好は動詞に還元できない文レベルの意味と相関する: 心理実験に基づく日本語の構文研究への提案. *認知科学*, 13: 334–352.
- [8] 黒田 航 & 長谷部 陽一郎 (2009). Pattern Lattice を使った (ヒトの) 言語知識と処理のモデル化. In 言語処理学会第15回大会発表論文集, pp. 670–673.