

研究課題別評価

1. 研究課題名 :プログラミング言語としての自然言語～推論システムと人間の思考～

2. 研究者氏名 :戸次大介

3. 研究の狙い :

研究の目的として「我々自身の意識や思考の構造を知る」ということに興味を持っています。ただ、それらを直接観測することは難しいため、科学として取り組むには観測可能な「傍証」を対象とせざるを得ません。脳神経科学、心理学、認知科学、分析哲学、理論言語学等、多くの分野が異なる経路から同じ山に登っていると言えるかもしれません。

その中で、「言語」と対象とするアプローチは、以下の三つの点で有望であると考えています。

1. 思考のうち、言語化できる(文で表せる)範囲は大きい。もちろん非言語的思考も存在し重要であるが、たとえば故人の思考について我々が著作を通して知ることのできる情報量を鑑みれば、研究対象として十分である。
2. 新聞記事コーパス等、大規模なテキストデータが利用できるようになった。またここ数年のインターネットの発達により、膨大なテキストデータが WWW 上に蓄えられつつある。それらによって現在の言語学は、十年前には存在しなかった「客観的なデータ」を電子的に利用可能となっている。
3. 副作用として、自然言語処理技術(機械翻訳、自動要約、情報検索、自然言語インターフェース等)の基盤技術や新たな応用分野を提供できる。たとえば、チョムスキー階層がプログラミング言語理論に与えた影響等。

このような興味から、私は「言語」を対象としたアプローチすなわち理論言語学によって、意識や思考といった捉えがたい概念を捉えたいと考えています。ただし、私の考える理論言語学とは、必ずしも現在の理論言語学(生成文法、形式意味論)そのものではありません。それは、以下のような問題意識に依るものです。

元々「言語化された思考を通して思考を分析する」というアプローチは分析哲学のものです。分析哲学(およびそこから派生した形式意味論)そのものは、科学として捉えようとする反証可能性をもたないと考えています。すなわち、分析哲学では言語化された思考を論理式におきかえて分析しますが、言語化された思考についての我々の直観は、文に対する直観(文と文の間の演繹関係についての直観)であって論理式に対する直観ではありません。にも関わらず、文と論理式の間の変換規則が与えられていないならば、論理式上の議論は厳密には検証できないわけです。

一方、チョムスキーの生成文法では、理論は文と統語構造の可能なペアを計算し、(strong generative capacity を論じるときには)その両方が検証の対象となります。統語構造から論理式への変換については、生成文法の中では扱いません。ところが、我々は与えられた文の受容性についての直観はあっても、与えられた統語構造(理論の用語で記述されたもの)について、正しいかどうかの直観をもてないのではないかと私は考えています。したがって、ペアのうち文の受容性だ

けを検証する weak generative capacity についてはともかく strong generative capacity については理論を検証することができません。

しかしながら、分析哲学と生成文法を組み合わせることで、最低限の楽観的な仮定のもとで、直観による検証が可能な言語理論を構成することができます。すなわち、文と(統語構造よりも)一步踏み込んで)論理式の間の変換規則を生成文法や形式文法理論によって与え、論理式の間の変換規則を数学的に与えることで、二つの文の間の変換関係を計算する理論を構成できます。我々は二つの文の間の変換関係に対する直観を持っていると仮定できれば、この理論は直観により検証でき、なおかつその過程で統語構造について間接的に論じることもできるわけです。このように検証可能なデザインを取ると、理論が厳しく制約されるため、現在の生成文法が陥っている「ある現象に対する理論を、どうしても書ける」状態から脱却することができます。

これが私の考える理論言語学のデザインであり、現在これに共通する理論構成を取っている研究としては、範疇文法(Categorial Grammar)や、形式意味論の一部の研究がありますが、この立場はまだ理論言語学の主流をなすに至っておりません。

生成文法の研究対象は思考ではなく言語機能にあるのですが、私は生成文法およびそこから派生した HPSG、LFG、CG 等の形式文法理論は、分析哲学およびそこから派生した形式意味論と(組み合わせるのではなく)一体化することで、はじめて言語に対する科学的なアプローチを完成できると考えています。

4. 研究結果：

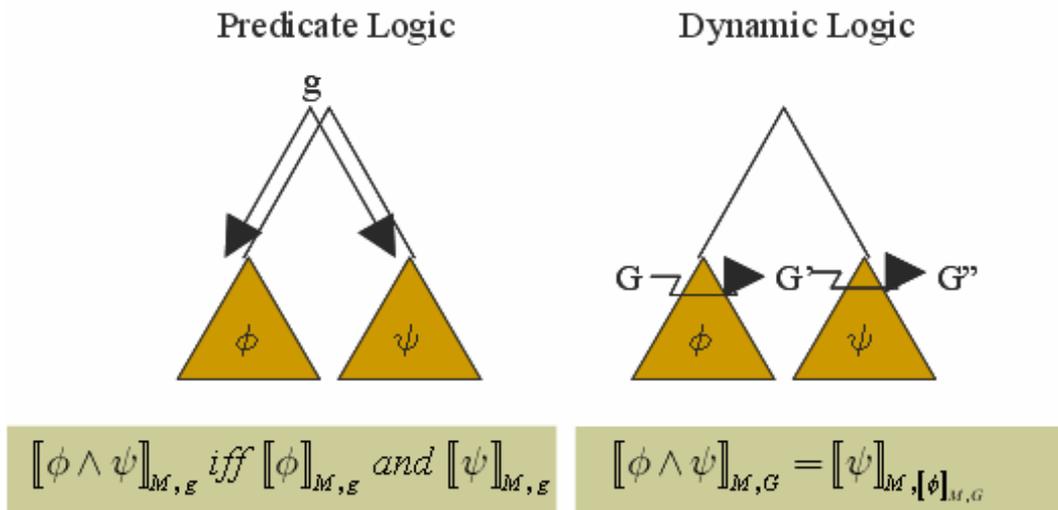
自然言語の意味のモデリングは、フレーゲ(Frege)以来の哲学的問題を数多く抱えており、実用化はおろか理論化すら見通しの立ちにくい領域です。意味解析の理論にとって特に問題となるのは、一階論理では記述できない次のような現象とされています。これらの現象は一般のテキストで頻出する現象であり、決して例外的現象ではないにも関わらず、自然言語処理では多くの場合処理を避けてきた現象です。

1. 量化の問題 Barwise and Cooper (1981)
2. 照応の問題 Heim (1982), Kamp (1984), Kamp and Reyle (1993)
3. 複数形の問題 Scha (1981), Link (1983)

これに対して、私はさきがけ研究において、「型付き動的論理」(Typed Dynamic Logic (TDL): Bekki (2000a)(2000b)他)という言語理論を独自に提案し、統語・意味理論を含む理論言語学・自然言語処理の統合的枠組みの構築を目指しました(TDL プロジェクト)。TDL は先行する理論言語学の諸理論(生成文法や HPSG 等)に対し、説明的妥当性において優位であることを、これまでの研究で示しています。特に、前述 1~3 の問題を統一的に、しかも統語理論と一体化した形で解決しています。

TDL は Groenendijk and Stokhof (1991)による動的意味論(Dynamic Semantics)と共通点を持ち、命題を、真偽値を表すものとしてではなく、文脈から文脈への関数(Information flow)として捉える点に特徴があります。

Figure 1: Typed Dynamic Logic (アイデア)



その他、TDL の特徴としては、以下が挙げられます。

- 型付き 計算に基づく三階論理である。
- 組合せ範疇文法(Combinatory Categorical Grammar)を統語理論として持ち、英語・日本語文法の差を最小限のパラメータで説明できる。
- TDL 表現間の証明論は、 計算の定理として健全性が証明される。

Figure 2: Typed Dynamic Logic (TDL)の定義

<i>atomic predicate</i> :	$r_{TIE \times \dots \times E}(x_1, \dots, z_1)$	$\equiv \lambda G_{T(EIII)} \lambda g_{EII}. Gg \ \& \ r(gx, \dots, gz)$
<i>conjunction</i> :	$\phi_{Prop} \wedge \dots \wedge \varphi_{Prop}$	$\equiv \lambda G_{T(EIII)} (\varphi \dots (\phi G))$
<i>disjunction</i> :	$\phi_{Prop} \vee \dots \vee \varphi_{Prop}$	$\equiv \lambda G_{T(EIII)} \lambda g_{EII}. (\phi G) g \mid \dots \mid (\varphi G) g$
<i>negation</i> :	$\neg \phi_{Prop}$	$\equiv \lambda G_{T(EIII)} \lambda g_{EII}. Gg \ \& \ \sim \exists h_{EII}. \phi Gh$
<i>distribution</i> :	$\Delta x_i [\phi_{Prop}]$	$\equiv \lambda G_{T(EIII)} \lambda g_{EII}. \exists d_e. \phi (\lambda h_{EII}. Gh \ \& \ hx=d) g$
<i>unary quantification</i> :	$uq(N_{TIN}, x_i)$	$\equiv \lambda G_{T(EIII)} \lambda g_{EII}. Gg \ \& \ N G/x $
<i>binary quantification</i> :	$bq(Q_{T(RTIE)(TIE)}, x_i, y_j)$	$\equiv \lambda G_{T(EIII)} \lambda g_{EII}. Gg \ \& \ Q(G/x, G/y)$
<i>copy</i> :	$x_i \triangleright y_j$	$\equiv \lambda G_{T(EIII)} \lambda g_{EII}. Gg \ \& \ \exists h_{EII}. Gh \ \& \ gy=hx$
<i>anaphora</i> :	$\phi_{Prop} \Vdash \varphi_{Prop}$	$\equiv \lambda G_{T(EIII)} \left[\begin{array}{l} \text{if } G/x = \phi G/x, \text{ then } \varphi G \\ \text{else if } G/x = (\lambda g_{EII}. true)/x, \text{ then } (\phi \wedge \varphi) G \\ \text{otherwise undefined} \end{array} \right]$

where $G_{T(EIII)}/x_i \equiv \lambda d_e. \exists g_{EII}. Gg \ \& \ gx=d$

Figure 3: Example of Syntactic Derivation

<p>“東京に”</p> <p>V_{FORM} / V_{FORM}</p> <p>$\lambda v. \lambda e. \lambda c. \lambda s. Tokyo(\zeta, \varsigma) \Vdash ve \left[\begin{array}{l} \lambda e'. \lambda s'. \left[\begin{array}{l} to(e', \zeta) \\ ce's' \end{array} \right] \right] s$</p>	<p>“行く”</p> <p>$V_{\text{終止連体}}$</p> <p>$\lambda e. \lambda c. \lambda s. \left[\begin{array}{l} going(e, s) \\ pres(e, s) \\ ces \end{array} \right]$</p>
<p>$V_{\text{終止連体}}$</p> <p>$\lambda e. \lambda c. \lambda s. Tokyo(\zeta, \varsigma) \Vdash \left[\begin{array}{l} going(e, s) \\ pres(e, s) \\ to(e, \zeta) \\ ces \end{array} \right]$</p>	

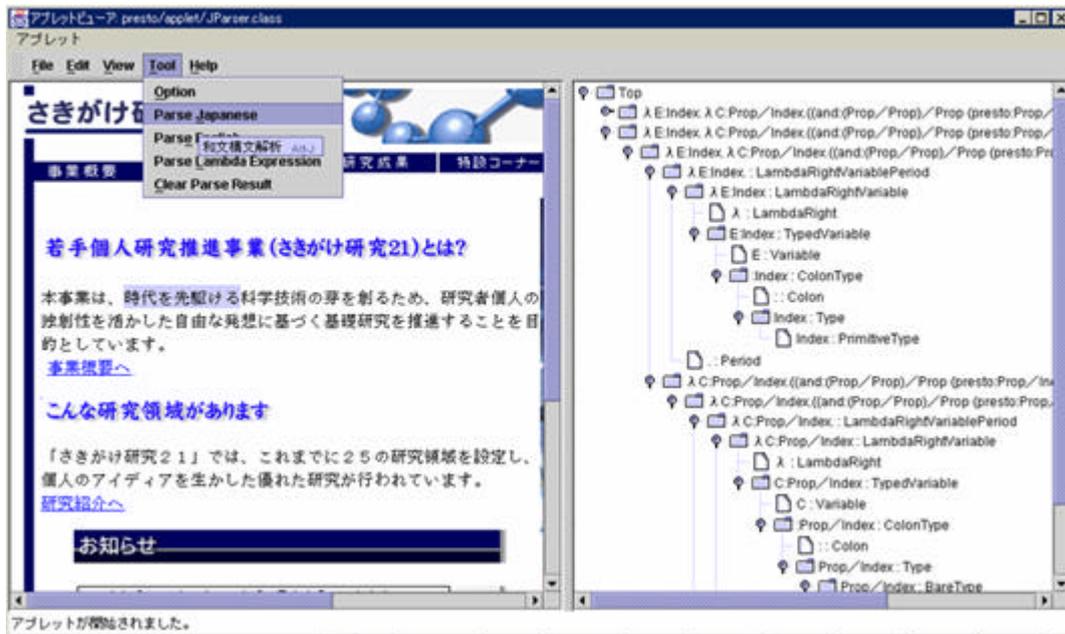
また、TDL は他の動的意味論に対して、複数形の扱い等様々な点で優位性を持っています (Bekki (2000))。また、TDL は型付き 計算 (Church (1951)) に基づいているため、独自の意味論を持つ動的意味論よりも計算機との相性がよいという性質もあります。

TDL プロジェクトではこの TDL の特長を生かし、自然言語処理のための推論システムをデザインしています。具体的には、

1. TDL による構文解析用の日本語・英語文法を完成し、実装する。
2. TDL の意味表現における、証明論 (= 推論システムの基礎) を完成させる。

の二つの副課題に取り組んできました。その結果として、研究開始時には予定になかった談話構造 (前提条件、フォーカス、照応の到達性など) についても、TDL による分析に利点があることが分かってきており、現在理論化を進めています。また、プロトタイプ実装としてはFigure 4の TDL-Parser があります。

Figure 4: TDL-Parser (JAVA 版)



TDL-Parser は、自然言語を入力として、文と文の間の演繹関係を計算できるシステムのプロトタイプです。これが完成すれば、情報検索でもより高度なクエリー、たとえば否定を含むもの(「ウディ・アレンの映画で、ウディ・アレンが出演していない作品」さらには「映画監督本人が出演していない作品」)などを用いることができます

5. 自己評価

研究の評価というもの是一般に研究目標に対する達成度ではかられる。工学的研究の場合は、社会的インパクトを持つ研究目標を掲げ、それを達成するための研究課題に細分化し、研究計画とする。

しかし自然科学の場合、「真理の探究、すなわち対象を理解すること」及び「我々が対象を理解する仕組みを理解すること」という二つ拮抗する活動自体を目標としており、これを工学的研究の評価スキーマにあてはめようとすると、問題が生じる。

その問題とは、研究目標というのはその時点での対象の理解に基づいて設定されるものであるが、自然科学の研究の場合はその理解そのものが進展していくのであり、従って研究の進展につれて当初の目標そのものが概念的に成り立たなくなる場合があることである。

例えば、多くの理論では言語に句構造があることを仮定しているが、その句構造を検証していきたいという目標設定をしたとする。ところが、本研究のように、句構造そのものは経験的な検証

の対象に成り得ず、真理条件の計算のための媒介的な理論概念にすぎないということが分かってくると、当初の「句構造を検証する」という問題設定そのものが成り立たないのである。しかしながら、「では句構造の検証はあきらめて、文の真理条件の関係を調べましょう、そのための構造的な理論を構築しましょう」と問題を再設定したときには、当初の目標が形を変えて包摂されているのである。

同様に、今回のプロジェクトの目標として、自然言語間の推論システムを構築するというのがあったわけだが、その目標は現在では形を変えつつあるため、目標に対する達成度という基準では評価できない。その目標設定の時点では、文から論理形式への変換というのは比較的自明なものであって単純な二項関係で表せるという仮定があったのであるが、研究を進めるにつれて、照応の参照先決定やフォーカスの構造、前提条件の解消といった問題が、文と論理形式の間において、言語機能にとっての「もう一つ」の推論レイヤーを要求していることが分かってきた。「もう一つ」というのは、単語から文を組み上げる派生そのものが人間にとっては無意識だが言語機能にとっては推論であり、また論理形式同士のいわゆる人間が能動的に行う推論が知られているが、その間にも同様な推論が必要なのだ、ということである。

こうなると我々が意識的に「これが推論である」という我々が命題の間で意識的に行うことと、単語から文を構成するといった無意識で行う推論といったもののどちらをもって推論システムと呼ぶのかを明らかにしない限り、当初の推論システムの構築という命題そのものが成り立たないといった意味で、当初の研究計画は文字通りには成り立たないのであるが、現在は当初の問題の理解よりも一段深いところで考察を続けているのである。

しかし、ここで述べたいのは、研究は大いに進展して多くの知見が得られたということである。研究目標の達成という意味では当初掲げた推論システムの構築ということには未だ至らなかったが、先に述べたような事情でその問題は再設定されているのである。また当初の目標には入っていなかった、様々な量化の現象や前提条件やフォーカスといった現象にまつわる伝統的な問題が、TDLのメカニズム及び文法によって大部分消滅するのだということを見出し、また示してきた。そういった意味でこの研究はまだまだ途上ではあるものの、大いに研究を進展させることができたと考えている。

6. 研究総括の見解

理論言語学の分野は、自然科学の中にあってもまだ未開の領域であり、特にわが国においては研究の土壌が不十分である。その中で戸次大介は、人間が、自然言語を用いて行う推論を処理できるシステムの構築を目指した。研究推進と共に、その目標はさらに理論的な研究へと形をシフトしたが、これは十分理解し得ることである。この中で、「型付き動的論理」(Typed Dynamic Logic(TDL))という言語理論を独自に提案し、統合・意味論を含む理論言語学・自然言語処理の統合的枠組みの構築を行い、従来の理論言語学の諸理論に対して、説明的妥当性において優位であることを実証した。同時にこれまでの既存の理論の中で伝統的な問題とされてきた事象をTDLのメカニズムおよび文法によって多くの部分が消滅できることを示した成果は評価できる。今後、「型付き動的論理」が言語現象全体の統一理論になりうるかどうかはまだ分からないが、更なる研究推進を推し進め、当研究分野発展への貢献を期待する。

7. 主な論文等

1. Bekki, Daisuke (2000b) "Typed Dynamic Logic for E-type link", In Proceedings of Third International Conference on Discourse Anaphora and Anaphor Resolution (DAARC2000), pp.39-48, Lancaster University, U.K.
2. Bekki, Daisuke (2000c) "Delayed Quantification for Cumulative Readings", A talk presented at SINN UND BEDEUTUNG V, The University of Amsterdam, Netherlands.
3. 戸次大介 (2002) 型付き動的論理(TDL)による日本語の量化・照応の分析, 情報処理学会研究報告 2002-NL-150, pp.69-76, 東北大学.
4. 戸次大介 (2003a) 型付き動的論理(TDL)によるイベント量化の分析, 情報処理学会研究報告 2003-NL-154, pp.23-30, 徳島大学.
5. 戸次大介 (2003b) 型付き動的論理(TDL)による前提条件と照応の統一的分析, 第 37 回言語・音声理解と対話処理研究会資料 SIG-SLUD-A203, pp.7-13, 立教大学.