3I1-R-9-9

オントロジーとHPSGを利用した日本語文の含意判定モデルの提案

A Suggestion of a Model for Recognizing Textual Entailment of Japanese Sentence using Techniques of Ontology and HPSG

松下 裕 Yu MATSUSHITA
山口 高平 Takahira YAMAGUCHI

慶應義塾大学
Keio University

This paper proposes a model for linguistic semantics, where the principle of compositionality is assumed between lexemes and the sentence, and lexical semantics is provided from data sources which are achieved to be sharable, context-rich or self-descriptive owing to the terms and techniques of Web Ontology. Studying on Recognizing Textual Entailment (RTE) tasks with Japanese sentences is intended to show how the semantics be composed through Head-Driven Phrase Structure Grammar, with the lexical-level semantic features and clearly specified syntactics, derived from ontologies.

1. はじめに

自然言語処理に統計的手法が用いられるようになって久しく，近年は元来の記号論的手法と相互に洗練し合うことで，形態素解析や構文解析，係り受け解析の技術では大変高い精度を得られるソフトウェアが多く開発されている．このような状況のもと日本でも，文や談話の意味論を大規模な言語資源に即して行う試みが，少しずつ始められている．

近年の意味論へのアプローチは，チャンク同士の依存関係(係受け)の分析結果に基づき，各項の予め定義された深層格タグの系列，さらには照応関係の解消も含めた述語項構造へのマッピング，あるいはラベリングをきっかけとしたものが主流である．ここから例えば，事態のオントロジー([乾 07])など高位の意味論を行う研究が日本では行われている．

本稿で述べる方法論は，こういった現在主流のアプローチとは異なる方向性を示すもので，深い意味解析に関する議論を構文論から極力切り離し，語彙の意味論，および文の意味論の構成を，別の場所で自由に議論する基盤を与えることを目的としている．このために本稿では，極力形式的な表現を保ちつつ意味を扱うプロセスに示唆のある素性文法，特に HPSG を再考することを考え，RDF グラフとオントロジーの技術を援用することで，意味論的な操作や議論の可能性を含みつつも意味を持たない，言わば意味論の素のようなものを構成する．

これは，例えば述語項構造解析やその扱いに対して直接的に貢献するものは何もないし，大規模な入力に対する処理速度や，頻出のパターンを多く含むサンプルの解析精度に関しては，統計的手法から発展した現在主流の手法，およびそれらの今後の発展形に敵うべくもない．しかしながらそういった手法の届かない，発話コンテクストや言語の生成的側面が大きく関るような言語現象に対し，より記号論の本質的な問題に対峙する姿勢，およびその土台となる，本稿で提案するようなモデルが必要となる場面が，今後必ず現れるだろう．

本稿では，Web の情報資源のメタ情報記述で利用されることを想定して設計された Web オントロジーの技術を，自然言語の語彙と語彙の意味の表現に援用し，これとの整合性を保ちつつ，意味論を HPSG の理論により構成するモデルを提案する．これを RTE(テキスト含意判定)タスクに対する実験を通して，モデルの有用性を検討する．

連絡先：松下裕，山口高平，慶應義塾大学理工学部
〒223-8522 神奈川県横浜市港北区日吉 3-14-1
TEL:045-566-1614
drowse@a8.keio.jp, yamaguti@ae.keio.ac.jp

2. Web オントロジーと語彙

Web オントロジーによって語彙意味論を表現する手法について，いくつかの要素技術の紹介を補足しながら示す．

2.1 RDF と OWL

RDF[1](Resource Description Framework)は，Web の情報資源に関するメタ情報を，形式的に明確な形で表現するための枠組みである．RDF は有向グラフによる意味ネットワークの形を取り，ある情報資源(以下リソース)と別のリソースとの関係をノードからノードへの有向リンク(プロパティと呼ぶ)で表現し，これを多対多で展開することができる．RDF においてリソースは，リテラルによる表現を除き全て URI[2]で名前付けされ，実用上の一意識別子が与えられる．

RDF は本質的に一階述語論理に依っており，この有向リンクはアリティ 2 の述語記号に対応する．加えて RDF は，リソースをクラスによってグループ化することができ，こちらは単項述語記号に対応する．これらは通常のリソースとしても扱うことができ，他のリソースとの関係の記述が可能である．

RDF に関して特筆すべきこととして，空白ノードの存在がある．これを含む表明を記述すると，空白ノードを変数とみなし，その周りの関係を制約として変数を束縛するような存在量子化を表現することができる．またプロパティ，クラス共にその外延(引数としたときグラフが真となるリソースの集合)の包含関係を記述することができ，これにより特定のスコープ内の全称量子化を記述することができる．これらの機能により，RDF によって状況意味論や DRT を直接表現することも可能である．

OWL[3]は RDF の構文上に構成されたオントロジー記述言語である．OWL は基本的に RDF の意味論を全て継承しながら，クラスを制約ベースで表明するための語彙や，リソース同士の(同一性ではなく)同義性を表現することで複数のオントロジーを連携させるための仕組みが追加されている．

RDF の枠組みでは，上位の言語を含むこれらの機能を全て保持したまま，クラスとプロパティを含む独自の語彙を定義することが可能であり，これを完全な形での再配布および公開する手段が提供されている．

[1] http://www.w3.org/TR/rdf-concepts/
[2] http://tools.ietf.org/html/rfc3986
[3] http://www.w3.org/TR/owl-ref/

なお以上に示したそれぞれの技術は，W3C[1]勧告によりその文法と意味論が規格化されており，少なくとも形式的な解釈は，常に一意的に与えられるようになっている．

2.2 オントロジーと語彙

以上で述べたオントロジーの技術には，これを再利用する手段があり，語彙の意味論を行うために，既存の資源を利用することができる．そして日本語には，いくつか既に語彙の面である程度の網羅性を持ったオントロジー資源が公開されているのである．代表的なものとして，オントロジーとしての整備にはまだ不足があるものの，日本語 WordNet[2]がまず考えられる．あるいは固有表現や即時性に対応するために，日本語 Wikipedia オントロジー[3][玉川 11]の階層関係を利用するのもよいだろう．もし必要な語彙を含み，様々なプロパティを擁するオントロジーがあれば，Web オントロジーの基礎的な機能である名前付けと上位下位関係のみを取捨選択し，利用してもよい．

3. 単一化文法，およびオントロジー意味論の連携モデルの提案

以下より，Web オントロジーによる意味論から，単一化文法，特に HPSG の方法論を通して文の意味論を構成していく方法を示す．

3.1 HPSG および JPSG

Head-Driven Phrase Structure Grammar (= HPSG)は，LFG や GPSG における，主辞の概念，語彙指向の素性表現と単一化，X-bar 理論の抽象化手法等をバランスよく継承した句構造文法の一種である．特筆すべきは句構造規則が，ある親範疇 M に対して，主辞 H とその補語 C を考えた時，以下の局所的な親と子の部分木に関する規則のみに抽象化されることと言える．

$$M \rightarrow C\ H$$

図 1 HPSG の句構造規則

この句構造規則のあり方は，単一化のプロセスを明解にさせると同時に，主辞に大きな役割が与えられていることからも分かる通り，文法に依存文法的な性質を与えることにも繋がっている．統語範疇の素性表現への還元を徹底することで，依存文法と同様に文の根を開始記号ではなく動詞範疇と見做して分析することもでき，実際これ以後の分析ではそれを採用する．

JPSG(Japanese Phrase Structure Grammar)は，HPSG を日本語の事情にあわせて変形させた文法理論であり，古いものでは[Gunji 87]，新しいものでは[大谷 2000]などで研究されている．JPSG では，日本語の自由なチャンク出現順序を考慮し，動詞やその他の下位範疇化素性を持つ語彙範疇に関して，この素性のデータ構造を配列でなく集合とすることで，補語の順序に関わらない係り受け関係の捕捉を行う．

3.2 意味素性のすり替えと集約

語彙の意味から文の意味を構成するにあたって，HPSG では意味素性を導入している．意味素性は，他の品詞などの素性と同様に語彙に埋め込まれ，単一化／下位範疇化の経路を経て文へと構成される．このプロセスは意味論にとって，少なくとも一つの明確な示唆を与えるものである．

ここで意味素性を，語彙のオントロジーから取得することを考える．基本的に素性文法の素性構造は，素性値として素性構造の埋めこみを含むあらゆる表現をとることができる．従って本稿では，オントロジーで表明されたリソースへの参照をここに用意することを提案する．こうすることで，単一化経路の恩恵を損なうことなく，オントロジーの意味論を文の構成的意味論に利用する土台を確保することができる．

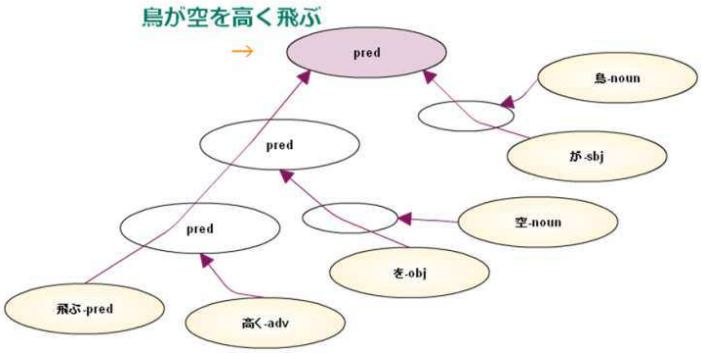

図 2 単一化経路の例

一方ここまでの議論だけでは，文の意味を構成する手法として不十分な点がある．すなわち，一つ一つの単一化の操作における，意味素性の操作，扱いが未だ明確に示されておらず，これを与える必要がある．

ここでの重要な目標は，最終的に文のレベルの意味表現を構成する際に，それを少なくとも RDF のモデル，意味論に整合可能なものとすることである．文の意味論に対して，オントロジーが供給した語彙の意味素を構造化した形での表現を与えることができてこそ，かつその構造が，RDF に整合的であればこそ，オントロジーに表明した意味関係を，意味論の議論に利用する準備ができたと言えるのである．ここでは，2 つの手法を示す．

(1) 状況意味論の一般的なアプローチ

まずは，状況意味論の研究から発展した構成意味論の一般的な手法をみる．詳細な説明およびスコープの曖昧性などの問題は省略し，簡単のため英文を例に取ると，例えば次の図 3 のようなものになる．

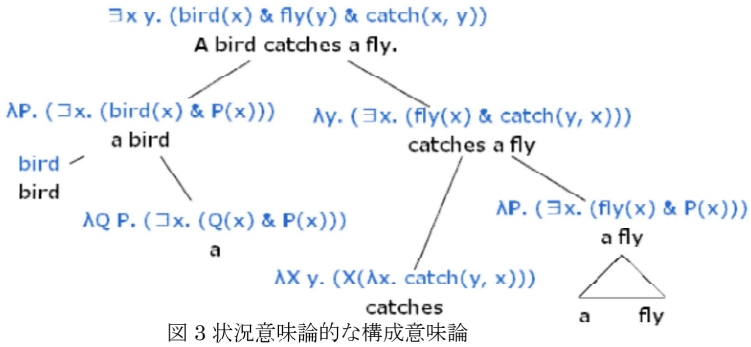

図 3 状況意味論的な構成意味論

構文木上の中間表現にλ演算式が現れているが，これは意味を関数適用の連続によって構成・表現するための，形式論理的な措置である．ここにオントロジーの意味論を導入するのは難しいが，この過程の計算を自動化するのは難しくないし，文全体の意味表現(計算結果)を RDF 上に展開するのも簡単である．

ここでむしろ問題なのは，このアプローチを採るために，文法的な語彙，すなわち冠詞や否定表現，あるいはもしかすると一部の助動詞に関して，その論理表現を予め知っておく必要があり，さらにこれには，オントロジーによる語彙提供が貢献するのは難しいという点にある．また，日本語に限らず，文法的な語彙

[1] http://www.w3.org/
[2] http://nlpwww.nict.go.jp/wn-ja/
[3] http://wikipedia-ont.sourceforge.jp/

は，その他の一般的な語彙にも増してその意味論的解釈が割れるところである．本稿ではこういった語彙に関しても，他の語彙と同様に，オントロジーへの参照に抽象化された意味素性を割り当て，構成のプロセスとは切り離された議論が行われるべきであると考える．

(2) RDF の意味論を利用したアプローチ

そこで，構文木全体を RDF グラフで表現する手法を考える．おおよそ表層格に相当するような構文論的な関係を表すプロパティを導入し，語彙リソース同士を接続する．一方，文法的語彙に関してもこのグラフに含め，オントロジーから提供することを考える．注意したいのは，この場合文法的語彙には他の語彙同様，識別子と，可能ならば階層関係さえあればよいということである．そしてここに，単一化を表現するための head プロパティを導入する．これらを用いると，HPSG の構文木構造全体を，構成はそのまま RDF グラフで表すことができる．

ここに意味素性の集約のプロセスを与えるために，単一化を示す head リンクに特殊な役割を与える．すなわち，対象ノードのタイプ同士が，クラス-サブクラス関係となることを保証する．先に述べたとおり，HPSG において文は全ての単一化操作の結果が流入するところの動詞範疇である．従って単一化のリンクを，クラスの継承関係に落としこめば，文の根のノードには理論的に，文の全てのノードの情報が集約されるはずである．これにより構成されたグラフの例が以下である．ただし語彙項目が直接記述されている部分は，状況意味論に従ってその語彙クラスのインスタンスであるとする．

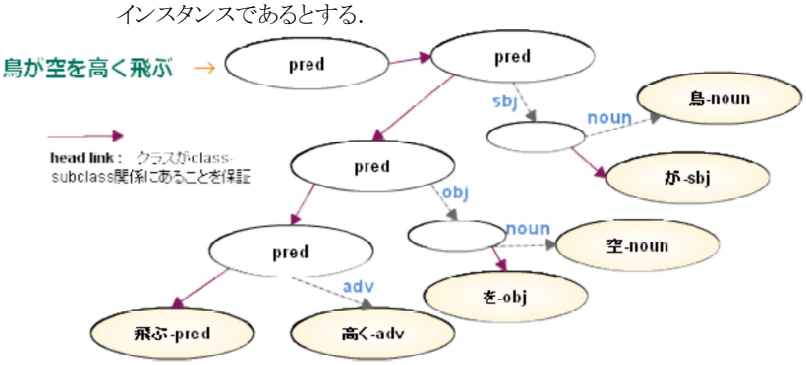

図 4 構文木の RDF グラフによる表現

あるいは，日本語の語順の自由性を更に尊重し，以下のようにしてもよいだろう．

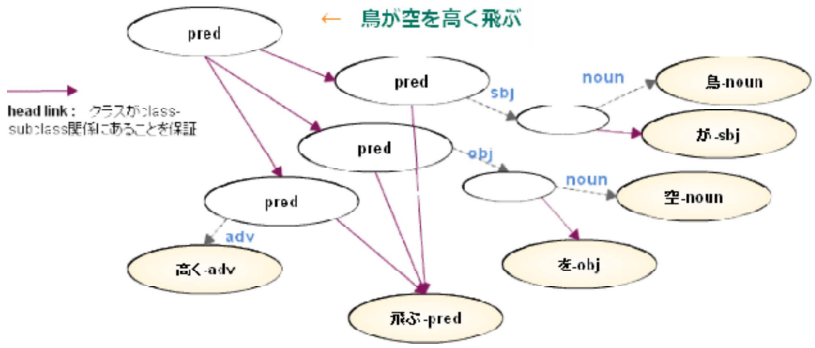

図 5 並列補語を尊重したモデル

この構文木の叙述構造におけるリソース同士の関係について，もう少し詳しく見てみる．提案モデルの構文木においては，通常プロパティ，あるいは二項述語記号によって記述される動詞の意味の表現が，ノードとして存在することが特徴的である．ここで，RDF では可能世界意味論に従い，リソース同士の関係性やリソースそれ自身の性質が，RDF によって何かを述べる前から決定しているということに注意する．即ち例えば図 5 における「鳥が飛ぶ」に相当するノード(文の根の直下の最も右側のノード)に関して，ここに head の他 sbj の文法的関係をもったノードが接続されているが，実のところこのノードは，sbj リンクの先のノードが表す概念をもともと内包しているものだということである．即ち，「飛ぶ」という動詞はもともと匿名の<飛ぶ者>というエンティティに対する<sbj>のインターフェースを持つスロットを含み，例のように実際にプロパティで関係付けられることで，これが具体化されるという解釈になる．そして文の根のノードは，こういった動詞範疇のクラスを全て継承するので，その意味でこのノードには，文の意味が集約されるということになる．この性質に留意することは，後の含意判定モデルについての議論においても，非常に重要である．

ところで，このグラフを通常の RDF パーサで解釈する場合，head プロパティで表現された関係が，クラス-サブクラス関係であることを保証するべく新たな意味論的条件，すなわち公理をシステムに加える必要がある．しかしながら，言語仕様において定められた語彙拡張のシステムを超えた範囲で，悪戯に公理を拡張することは望まれないため，通常の RDF(-S)の枠組みの中で用いられる“rdfs: subClassOf”を用いて，これを表現すべきである．とはいえ本稿において，head プロパティを用いて表現を簡略化することに支障はないため，以後のグラフ表現においても，これを用いていくこととする．

4. RTE による思考実験

以上の提案モデルが実際の意味論のタスクにどう貢献するか見るため，これをテキスト含意判定(Recognizing Text Entailment = RTE)に用いることを考える．テキスト含意判定は 2 つの文の含意関係，即ち，一方の文の意味内容がもう一方の文の意味内容より，語彙の役割や関係，あるいは常識を参照することで，推論可能かどうかを判定する課題である．この課題は単純ながら，人間の扱う「意味」に関する結論を急ぐことなく，一方で言語現象と意味との関わりを確かに問う，興味深い課題である．本稿ではこれに対して，提案モデルにおける，構文構造からの一次的な変換である RDF グラフでアプローチする可能性を論ずる．

提案モデルによる文の含意関係判定の手法について，文の根にあたるノードを比較することが有効だと考えられる．提案で述べた通り，このノードは理論上文中の全ての動詞による叙述構造を継承している．従ってこのノード同士の比較をしたとき，それらがクラス-サブクラス関係にあるならば，2 文間に含意関係があると認めてよいだろう．これらのノード間でクラス-サブクラス関係が成立する条件として，もっとも単純なのは，RDF グラフのどちらか一方がもう一方のグラフの部分グラフとなることである．空白ノードに関しては RDF の言語仕様によって，スコーレム化の操作を用いて，同じ制約をもった空白ノードを含む部分グラフ同士を等価とするよう，設定されている．

加えて大変重要な条件は，あるグラフの部分構造について他の構成が全く同一な条件で，ただ一箇所の語彙に関して異なるという状態があった場合，語彙を供給したオントロジーにおいてそれら語彙の関係が宣言されていれば，それが部分構造全体の関係に直接反映されるというものである．

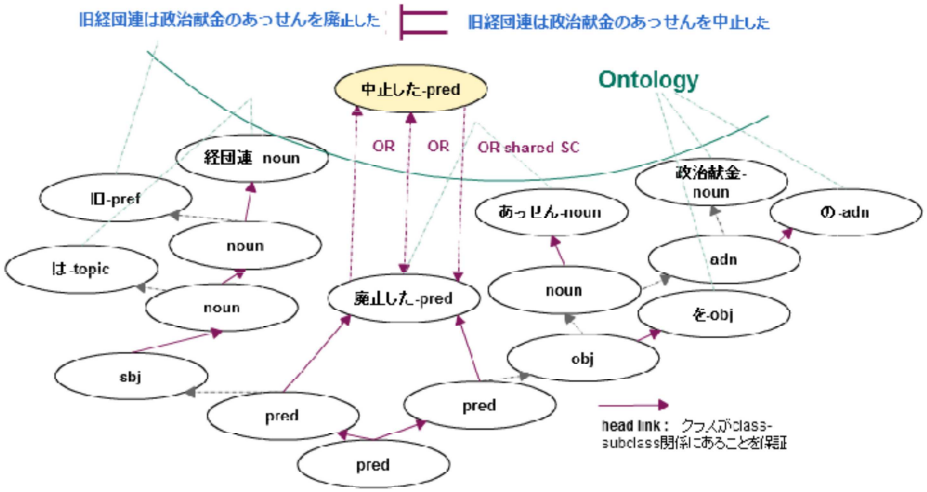

図 6 語彙の関係によって含意判定できるタスク(グラフは簡易形)

この条件により，例えば上図の例題で含意関係を検出することができる．例文は「旧経団連は政治献金のあっせんを廃止した。」および「旧経団連は政治献金のあっせんを中止した。」であり，ただしここでは，語彙「廃止する」が「中止する」のサブクラス，または逆あるいは等価なクラスであることを仮定している．

しかしながらこの手法を選択した場合，付加的な構造，および語彙のクラス関係による含意に関しては判定できても，付加の仕方を無差別化しても等価とならないような同義表現，あるいは言い換え表現については，このモデルだけでは判定不能となる．

加えて以下のように，誤った判定を与える場合もある．以下では「その鳥は飛べる。」および「その鳥は飛べた。」という 2 文に関して，このモデルを用いると，後ろの文は前の文を含意すると結論づけられるが，これは現実的な実感に合わない．

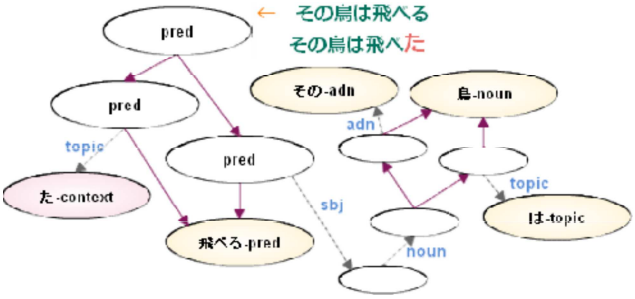

図 7 誤った含意判定を与えるケース

これは実際，特筆すべき事態である．まず，あるノードに head 以外のプロパティが付加する場合，それがもとのグラフを含意する場合としない場合が，現実的にはあることである．もう一つは，これが RDF グラフに通常仮定できる単調性の定理との矛盾であり，この問題を語彙レベルの関係の宣言の有無に帰着できないということである．このことから意味論の集約プロセスに関して，クラス・サブクラス関係に執着せず，より柔軟な扱いを導入することが望まれると言える．

5. おわりに

本稿では自然言語，特に日本語の意味論を語彙意味論より構成的に得るため，語彙の意味論的関係をオントロジーの技術によって柔軟な扱いを許す形で宣言し，これを素性文法の意味素性，特に素性の単一化プロセスについて充分な示唆のある HPSG におけるそれに埋め込むことで，構成性原理の下で文の意味論を扱うためのコンセプトを提案した．加えて，HPSG の方法論から意味論的集約のプロセスを得るため，構文木を RDF グラフで表現するモデルを提案した．そして，このモデルの性質を見るために，日本語文のテキスト含意判定タスクへの適用可能性について一見した．

本稿における語彙の意味についての扱いは，単語や形態素レベルの意味論から構成的に意味を構築する生成意味論[Pustejovsky 95]に近いものである．生成意味論は，語彙に関する知識獲得の困難，ならびに語彙から意味への集約プロセスに対する不信から，単一化文法と共に 90 年代に廃れていった経緯がある．本稿のモデルは今後，語彙についての扱いの自由度を生かして，あらゆる言語資源を，横断的に(意味論に限らず)有効利用する可能性についても展望する．すなわち，タグ付きコーパスなど今まで意味論で用いる事が困難だった言語資源を，意味論で何らかのタスクで利用する道を探ること，あるいは，系統的扱いが容易な語彙のレベルで，理論体系や実装手法の異なる研究者・研究グループ間での意味論的言語資源の共有を可能にすることなどである．

また，本稿で提案したモデルがさらに大きな意義を持つとすれば，意味の議論が，汎用の上位オントロジーの議論の文脈や，あるいは実世界指向のシステムにおける，センサ情報の記号化の文脈等，人間の意味処理に関する根本的な透察の場で行われる時であり，そういった領域への接近も当然，目指していくつもりである．

参考文献

[乾 07] 乾健太郎，竹内孔一，藤田篤，"含意関係計算のための事態オントロジーの開発に向けて"，電子情報通信学会，言語理解とコミュニケーション研究会，信学技法，Vol.106・No.518，pp.233-244 (2007)

[Pustejovsky 95] James Pustejovsky, The Generative Lexicon. The MIT Press (1995).

[玉川 11] 玉川 奨，森田 武史，山口 高平，"日本語 Wikipedia からプロパティを備えたオントロジーの構築"，人工知能学会論文誌 特集論文「近未来チャレンジ」Vol.26 No.4 pp.504-517 (2011)

[Gunji 87] Gunji, T (郡司隆男), Japanese Phrase Structure Grammar: A Unification-Based Approach (Studies in Natural Language and Linguistic Theory), Kluwer Academic Print on Demand (1987)

[大谷 2000] 大谷朗，宮田高志，松本裕治，"HPSG にもとづく実用日本語文法について"，自然言語処理，Vol7，No.5，pp.19-39，11 (2000)