

特集 「ポスト経験主義の言語処理」

LFG 理論をめぐる言語処理研究の発展

The Progress of NLP Researches Based on Lexical Functional Grammar

大熊 智子
Tomoko Ohkuma

富士ゼロックス (株) 研究技術開発本部コミュニケーション技術研究所
Fuji Xerox Co., Ltd.
ohkuma.tomoko@fujixerox.co.jp

増市 博
Masuichi Hiroshi

(同上)
hiroshi.masuichi@fujixerox.co.jp

Keywords: lexical functional grammar, unification, syntax, semantics.

1. はじめに

1.1 背景

Lexical Functional Grammar (LFG) は自然言語文の分析を行うための文法理論であり, [Bresnan 78, Kaplan 82] によって提案された。

LFG では, 主語や目的語などの文法機能がいかなる言語の解析においても有効であることを前提としている。すなわち, 見掛けの表現は全く異なっていたとしても, 文法機能の概念はあらゆる言語において共通に存在すると仮定する。この特徴を踏まえて, 世界各国の言語を対象にこの LFG 理論に基づく分析が行われている。また, f-structure は自然言語文を表現する構造化データとして, 自然言語処理に極めて親和性が高い。このような背景から, LFG 理論は言語学だけでなく自然言語処理分野でも多くの研究が行われてきた。

LFG のもう一つの特徴として, 上記のような言語に依存しない普遍的な解析結果を言語に依存する情報を記述した文法規則で導出できる点があげられる。この仕組みは, 多国間で並行に文法開発を進めるときに大変有利である。Xerox PARC ではこの性質を利用して, LFG 理論に基づく多国語の文法を開発するプロジェクトを行っている。

1.2 本稿の構成

本稿では自然言語処理分野における LFG 理論に基づく研究を紹介する。2 章ではまず LFG 理論の概要について解説する。3 章では多言語の LFG 文法を開発することを目的とした ParGram プロジェクトの活動について紹介する。ParGram プロジェクトでは各言語の担当者が言語学的な知見に基づいて文法を記述していたが, 統計的な LFG パーザの構築に取り組んだ研究もある。4 章では統計的な手法とルールベースの parser の融合, f-structure 情報の自動アノテーションや LFG 文法の自

動獲得の研究を紹介する。

f-structure は構文構造を表現するものであるが, それを言語と意味の interface としてより深い処理を行う研究も行われている。5 章では f-structure を入力とした意味解析について紹介する。6 章では f-structure や f-structure を用いた意味解析結果を利用した応用の事例として, 翻訳, 要約, 含意判定などを紹介する。

2. LFG 理論の概要

2.1 LFG 理論とは

LFG は自然言語文の解析を行うための文法理論であり, [Bresnan 78] によってその概念が提唱された後, [Kaplan 82] によって現在の形の定式化が完成した。LFG に基づく解析では, 解析結果として c (constituent)-structure と f (functional)-structure と呼ばれる 2 種の構造を出力する。c-structure は自然言語文の構造を, 文の形態素を上位のフレーズへとまとめ上げることによって木構造として表現するものである。一方, f-structure は, 文法機能の概念に基づき, 文の格構造, 時制, 様相, 話法などの意味情報を属性-属性値のマトリックス構造で表現するものである。

LFG 理論では, SUBJ (ect) や OBJ (ect) などの文法機能がいかなる言語の解析においても有効であることを前提としている。すなわち, 見掛けの表現は全く異なっていたとしても, 文法機能の概念はあらゆる言語において共通に存在するという立場をとる。言語が異なれば同じ意味内容を表現する文であってもその句構造すなわち c-structure は大きく異なる一方で, 文法機能に基づく構造である f-structure の違いは多くの場合極めて小さいことが知られている [Dalrymple 01]。

この性質はつまり, 図 1 に示すように, 語順が異なっているために, 言語によって全く異なる c-structure で表現される構文も, f-structure では複数の言語間で共通の構造をもつ表現に置き換えられる可能性をもつ。

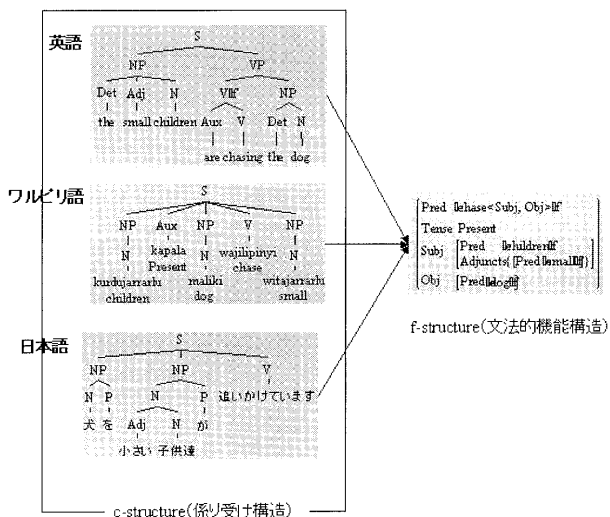


図 1 c-structure と f-structure

言語解析システムの出力結果が言語を問わず一定であればあるほど、多言語を対象とする言語処理システムの構築に要するコストは低減すると考えられる。ParGramに参加する各言語のLFG文法では、このf-structureがもつ言語普遍性の特徴を損なわないことを記述方針とする [Butt 99].

2.2 文法規則の概要

本節ではLFG理論における文法規則について説明する。

LFG規則は語彙規則と文法規則によって定義される。語彙規則では、f-structureの素性に対応するSUBJ(主語)やOBJ(目的語)などの引数やlexical category, 属性と属性値などを定義する。lexical categoryは品詞よりもより細かい分類になる場合が多い。(1)に語彙規則の例を示す。1行目は動詞「読む」にVというlexical categoryが割り当てられており、PRED(icate)「読む」はargumentにSUBJとOBJという二つの引数をもつことが記述されている。PREDは主辞であり、argumentは項である。

2行目には名詞「太郎」のlexical categoryがN(名詞)でありPREDが「太郎」であることが記述されて

いる。3行目には名詞「本」のlexical categoryがNでありPREDが「本」であることが記述されている。4行目には格助詞「が」のlexical categoryがPP(後置詞)であり、主格を示す属性NOM(inative)とその値「+」をもつこと、すなわち、「が」が主格を示す機能を有していることを表現している。

5行目には格助詞「を」のlexical categoryがPPであり、目的格を示す属性ACC(usative)とその値「+」をもつことが記述されている。

- (1)

読む	V	PRED= '読む < (↑ SUBJ) (↑ OBJ) >'
太郎	N	PRED= '太郎'
本	N	PRED= '本'
が	PP	(↑ NOM) = +
を	PP	(↑ ACC) = +

- (2)

S	→	NP	V
		{ (↑ SUBJ) = ↓ (↑ NOM) = _c + ↑ = ↓	
		(↑ OBJ) = ↓ (↑ ACC) = _c + }	
NP	→	N	PP
		↑ = ↓	

文法規則は、句構造規則とそれに付与される機能的注釈で表現される。注釈において用いられる「↑」という記号は句構造の一つ上の節点に存在するf-structureを、「↓」は現在の節点のf-structureを指す。(2)に(1)の語彙規則を含む文法規則を示す。1行目の句構造規則では、0個以上のNPとVがSになることを定義している。Vに付与されている注釈はVのPREDが一つ上、すなわち文全体のPREDであることを定義している。NPに付与されている注釈ではNOM属性の値が「+」であればNPの値が一つ上のSUBJに、ACC属性が「+」であればOBJになる条件を=_cという記号で定義している。2行目の句構造規則では、NとPPがNPになることを定義している。Nに付与されている注釈はPREDが一つ上の節点であるNPのPREDであることを定義している。(1)の語彙規則と(2)の文法規則を「太郎が本を読む」の解析に適用すると、図2に示すように句構造規則によってc-structureの木構造が形成されると同時に、機能的注釈によって属性と属性値で表現されるf-structureが生成される。

3. LFGによる多言語処理
—ParGramプロジェクト—

3.1 The Parallel Grammar Project

大規模な文法開発のプラットフォームはコストの負荷が大きく、その構築のためには多くの時間がかかる。また、そのようなプラットフォームに望まれるものは非常に範囲

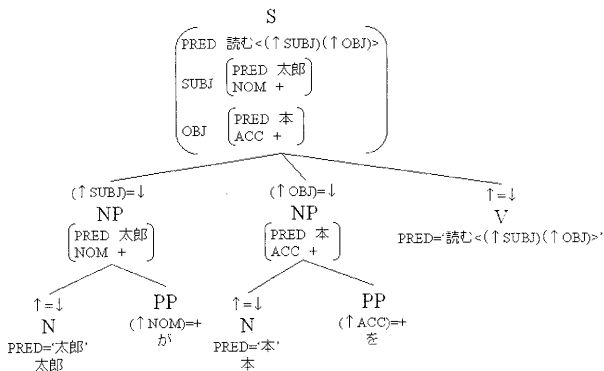


図 2 LFG 規則の適用例

が広い。文法開発プラットフォームはさまざまな種類の言語の文法を記述するために使われることが必要になるが、その要求は言語によって多岐にわたることが考えられる。Parallel Grammar Project (ParGram) は Xerox Linguistic Environment (XLE) を多くの言語のための文法開発プラットフォームとして採用する国際プロジェクトである [Butt 02]。

XLE は LFG 理論の仕様をほぼ完全に実装した parser であると同時に、LFG の文法記述を行う際のデバック環境であり、かつ、f-structure から自然言語文を生成する generator でもある。XLE がもつ最も重要な特徴はその解析速度であった。2・2 節で説明したように、LFG では、c-structure を構成するための句構造規則（文脈自由文法規則）と、f-structure を構成するために句構造規則の右辺の各構成素に付与する機能的注釈を同時に記述する。XLE は、句構造規則に基づく解析処理を実行した後に、処理が必要な機能的注釈を動的に決定する。すなわち、計算量の大きい機能的注釈の処理を選択的に行うことによって解析速度の向上を実現している [Maxwell 93]。また、解析結果の排他性・独立性を考慮した単一化手法 [Maxwell 91] も高い解析効率に寄与している。つまり、XLE を採用することによって、それまで行われてきた特定の言語事象に特化した小規模な文法ではなく大規模な LFG 文法に基づく解析を実時間で行うことが可能になったのである。

LFG はチョムスキーの提唱した言語普遍性という概念を前提としている。すなわち、すべての言語は類似した基本原理によって構造化されるというものである。LFG においては、f-structure は言語間に共通なレベルの分析結果を表現するものであり、このレベルの抽象度によって言語間の並行性をもつことができるようにする。しかしながら、c-structure の構造はそれぞれの言語の語順、形態素などによって異なる。

ParGram プロジェクトは LFG の言語普遍性とその限界について検証し、言語横断性を維持するための方法論を確立することを目指す。もし、それが可能なのであれば、それぞれの言語のための文法によって導出された f-structure は並行になる。これは、自然言語処理にとって大変有利である。もし f-structure の仕様が同じならば、違う言語に対してアプリケーションを適用したり翻訳に利用することが容易になるからである。

ParGram において、各言語の性質や各国の参加目的などに違いがあっても、各言語の f-structure の並行性は高いレベルで達成された。この並行性は構文構造的なものだけではなく、文法開発そのものにもあった。例えば、テンプレートや属性の共有化などである。また、ある言語の知識やテクニックがほかの言語に導入されることもあった。これによって、新たな言語が加わってもより少ない工数で LFG 文法を構築することが可能になった。

3・2 ParGram プロジェクトにおける日本語 LFG の研究

ParGram では、半年に一度全言語の文法記述担当者が集まってミーティング (ParGram Meeting) を開催し、複数言語間で f-structure の整合性を可能な限り高める、あるいは、矛盾を回避する機会を継続してもつ。すなわち、f-structure の構成から属性・属性値の用法やネーミングコンベンションに至るまでの詳細を議論し、ParGram の標準仕様を決定する作業を行っている。

[増市 03] では、この ParGram Project における日本語 LFG 文法の構築を行っている。日本語 LFG 文法における、肯定文、否定文、疑問文、受動文、並置表現などの基本的な構文に対して出力される f-structure は、この ParGram 標準仕様 [Butt 02] に準じている。ただし、この標準仕様は、f-structure の基本的な構成、属性・属性値については全言語でほぼ合意がなされているが、詳細な文法事項に関しては継続的に議論を行うと同時に各国の担当者が各言語の特性に沿った文法の分析を行っている。日本語においても、係助詞 [Ohkuma 03]、副助詞 [大熊 06]、形容詞 [大熊 08]、助数詞 [大熊 09] などについて独自の分析を行い、適切な f-structure を導出できるような語彙規則と文法の記述を行っている。

f-structure の属性と属性値のネーミングレベルの仕様に関しても、各言語間で注意深く言語事象の比較を行って協議している。例えば、当初全言語で性別を示す属性として使用していた「GEND」を、英語と日本語の LFG 文法では「GEND-SEM」とし、「GEND」と区別するよう変更を行った。英語と日本語の文法では代名詞に対して「GEND」属性を与えていたが、これはドイツ語、フランス語、ノルウェー語、ウルドゥー語の一般名詞がもつ性別の概念とは本質的に異なるためである。

c-structure に関しては、ノード (grammatical category) 名に対する緩やかなネーミングコンベンションの取決めはあるものの、どのような構造を定めるかは各言語の担当者に任されている。

3・3 Grammar Porting の事例

新しい LFG 文法の作成は、特に初期段階の立上げ時に大きな労力が必要になる。文法記述には LFG と対象言語に対する両方の知識が求められ、さらにそれを XLE 上で開発するための技術者としてのスキルが必要になるからである。

したがって、もし既存の文法をほかにポーティングできたら、この初期の段階において顕著な時間の節約が可能になる。これを実現するためには、二つの言語は類似していて文によって同じ述語項構造関係をもつ必要がある。日本語と韓国語はまさにそのような関係をもっているため、[Kim 03] では日本語文法から韓国語文法へのポーティングについて報告している。

日本語と韓国語は語順が似ている。日本語の動詞が末尾に来る構造、それを修飾する項や修飾成分の語順が比

較的自由であること、さらに主語のほうが目的語よりも先行する傾向があることも、韓国語にそのまま転用することができる。

語彙はもちろんかなり異なるが、語彙規則の部分は転用が可能である。例えば、日本語の格助詞「を」の語彙機能は韓国語の“ul”, “rul”にそのまま転用ができる。

この研究では約2か月の作業の結果、初期の立上げとしては非常に有望な結果を確認している。このように ParGram では類似したほかの言語の LFG 文法を参考にすることによって、自国語の文法開発を加速させる効用がある。

3.4 ほかの文法理論との比較

LFG と同じく、効率的な parsing が可能であると同時に多様な言語現象を表現できる高い記述力をもった文法理論として、Head-driven Phrase Structure Grammar (HPSG) と Combinatory Categorical Grammar (CCG) をあげることができる。生成文法が広く採用する『変形』の概念を計算機にそのまま実装することは容易でない。また、1970年代になり Richard Montague の意味論が言語学者の広く知るところとなり、『心理的実在性』への関心から、『変形』を仮定しない文法理論が提案されるようになった。1970年代末に Essex 大学の Gerald Gazdar によって、それまで変形を用いなければ説明できないとされてきた長距離の依存関係や等位構造などの言語現象を、統語範疇を素性と値の組合せによって表現することにより、句構造文法によって簡潔に記述できることが示された。LFG, HPSG, CCG は、いずれもこのアプローチの成果として発展した文法理論である。これらの文法理論に共通に見られる特徴は文法を制約充足と見る点と厳密な語彙主義である。HPSG および CCG では、下位範疇化フレームのような言語に依存する情報をすべてレキシコン（語彙）に記述する。一方で、文法規則による処理は言語に依存しない普遍性の高いものである。これに対して LFG では、言語に依存する解析結果と依存しない解析結果を c-structure と f-structure に分離するものの、言語に依存する情報を文法規則に記述する点において HPSG および CCG とは異なる。この特徴から、LFG では、文脈自由文法規則に対して文法機能に関わる制約を機能的注釈として付することによって統語的規則と文法機能の関係を簡潔に形式化することが可能となる。

以下、複数の言語を対象として並行的に深い言語解析システムの構築あるいはシステム構築のための文法記述を行うプロジェクトを紹介する。欧州では 1994 年から 1996 年まで、LRE (Linguistic Research and Engineering) プログラムのもとで、LS-GRAM と呼ばれるプロジェクトが複数の言語を対象とする研究活動を行った。LS-GRAM の対象言語は英語、ドイツ語、フランス語、スペイン語、ポルトガル語、デンマーク語、オ

ランダ語、イタリア語、ギリシャ語の 9 か国語である。ANLP (Advanced Language Engineering Platform) と呼ばれる共通のシステム上で単一化文法ベースの文法記述が行われた。ただし、後述のほかのプロジェクトと比較して文法のカバー率は低いとの報告がなされている。ドイツ Federal Ministry of Education の長期プロジェクトである Verbmobil では、PAGE と LKB という二つのシステム上で、ドイツ語、英語、日本語の HPSG 文法の記述が行われた。Verbmobil で開発された HPSG 文法は、比較的短い口語的発話を主な解析対象としている点に特徴がある。Stanford 大学 CSLI が中心となって研究が行われている LinGO Grammar Matrix では、主に英語、ドイツ語、日本語、スペイン語の文法開発を行っている。LinGO Grammar Matrix の英語文法は ERG (English Resource Grammar) と呼ばれ、Verbmobil プロジェクトの機械翻訳システムでも利用された実績をもつ。LinGO Grammar Matrix はオープンソースの形態で研究が行われ、比較的自由に新しい言語の文法記述を開始することが可能である。LinGO Grammar Matrix では、MRS (Minimal Recursion Semantics) と呼ばれるフォーマットを用いることによって、さまざまな言語の解析結果をできる限り統一的に表現しようという試みがなされている。

4. 統計処理によるアプローチ

4.1 f-structure の選択と c-structure の枝刈り

ParGram の文法記述のポリシーの一つに「言語学(統語)的に妥当な解析結果はすべて出力する」というものがある。XLE の出力には、可能性のあるすべての f-structure の集合を一つにまとめた packed f-structure という構造体がある。翻訳ではこの packed f-structure で曖昧性を保持しながら、目的言語に翻訳する手法が提案されている。

しかし、f-structure の応用先によっては、解の絞り込みが求められる局面がある。そこで [Riezler 02] は treebank コーパスを用いた識別学習を行い、複数の f-structure の中から正しい結果を判別する手法を提案した。学習にはコーパスの中でも特に 1000 以上の f-structure が出力されるような曖昧性の高い文の 10000 文を採用した。tree-bank には f-structure の情報はないため、ここで利用したのは係り受けの情報や NP, VP など構文木に付与されている句構造の情報であ

表 1 [Riezler 02] における実験結果 (抜粋)

	all	full	non-full
upper bound	84.1	88.5	73.4
stochastic	78.6	82.5	69.0
lower bound	75.5	78.4	67.7

表 2 [Cahill 05] における実験結果 (抜粋)

Pruning Level	None	4	5	6	7	8	9	10
upper bound	83.07	84.50	85.47	85.75	85.57	85.57	85.02	84.10
Time (CPU seconds)	288	100	109	123	132	151	156	182
non-full f-structure	23	39	36	31	29	27	27	24

る。素性には c-structure や f-structure の情報を用いている。表 1 に実験結果の抜粋を示す。lower bound は無作為に一つの f-structure を抽出したもの、upper bound は複数の出力結果から最も正解に近い f-structure を一つ選んだ値である。full は完全な f-structure、non-full は完全な f-structure が得られなかったときの部分解を示す。結果を見ると、この手法が特に完全な f-structure に対して有効であることがわかる。

[Cahill 05] では適用される句構造ルールの生起確率を元にして特定のルールを発火させないことで c-structure の生成を制御する手法を提案している。表 2 にこの手法の実験結果の抜粋を示す。Pruning level は同じフレーズの解析に採用されるルールのしきい値を示しており、Pruning level が小さくなるほど発火するルールの数は少なくなる。

表 2 の実験結果を見ると、適用されるルールが絞り込まれることで、完全解を導出できなかったケースは増加するものの、速度の大幅な向上が見られる。さらに、精度も向上することがわかる。

4.2 コーパスの自動構築と文法自動獲得

[Kim 03] のような手法が提案されてはいるものの、やはり文法の記述にはコストがかかるため、LFG parser の自動構築を目指す研究が行われた。しかし、文法を自動獲得するためには f-structure のコーパスが不可欠である。[King 03] では treebank から無作為に抽出した 700 文に対し、完璧な f-structure の情報を付与している。このように専門家によって精緻なコーパスを作成することは理想的であるが、多くの場合、専門的な知識を持った人物が大量のコーパスを作成することはほぼ不可能に近い。そこで、[Cahill 02] では treebank のような既存の大量コーパスを利用して、f-structure の情報をもったコーパスを作成する手法を提案している。この手法では treebank に f-structure の情報をアノテーションするルールを適用する。このようにして得られた f-structure の情報付きコーパスを用いて [Cahill 04] では文法の自動獲得を行っており、この parser の解析結果は人手で構築された LFG 文法による解析結果とほぼ同等の f-score (適合率と再現率の調和平均による性能指標) を得ている。しかし、この parser では単一化を行っていないため、この parser によって導出された f-structure は完全には言語学的に保証されない可能性がある。

4.3 LFG と統計手法の融合

4.1 節で紹介した [Riezler 02] は treebank から取得した情報を部分的な情報をもったラベルとして利用することで、曖昧性解消のための大規模コーパスを得ている。また、[Cahill 05, Cahill 08] では c-structure をあらかじめ限定することによって、単一化のコストを低減し、精度を維持したまま速度を向上させることに成功している。この句構造文法ルールの制限は単一化を行う前に c-structure を限定してしまうので、正しい f-structure が出力されない可能性はある。いずれにしても、ParGram で開発された精緻な文法を使っているため、言語学的な裏付けのある f-structure の出力が保証されている。

4.2 節で紹介した [Cahill 02] は treebank に f-structure の情報を自動的に付与する手法について報告している。また、[Cahill 04] はこれによって獲得された f-structure コーパスを学習し、LFG 文法規則の自動獲得を実現している。

両者は LFG の専門家が記述した精緻な文法を処理の前提としているか、(ほぼ) 自動で文法を構築しようとしているかという点に違いがある。後者は [Tounsi 09] や [O'Donovan 05] などにも適用され、文法の自動構築の効果が確認されているが、f-structure の妥当性はやはり専門家によって記述された文法によって導出したものには及ばない。単一化を前提としないパーザは時に一貫性が保証されていない f-structure を出力する可能性があるからである。

アプローチの違いがあるとはいえ、[Cahill 02] の研究においても f-structure の属性と属性名を ParGram と統一したり、分析の方法を取り入れたりするために、半年に一度の ParGram meeting には欠かさず出席を行っていた。このような形で LFG 理論の知見を取り入れることもまた、LFG 理論と統計手法の融合であるといえる。

5. f-structure からセマンティックへ

5.1 Glue 意味論

f-structure は構文構造を表現する機構であったが、それを入力としてさらに深い解析を行う理論が提案されている。その一つが Glue 意味論である。Glue 意味論は構文と意味論のインタフェースを提供し、意味的な合成を線形論理の導出として表現する [Dalrymple 01]。論理的な導出の前提は、文の構文的な項目に対応しており、文

から得られた前提の集合から複数の証明が導かれた場合、それは文の意味的な多義性、すなわち解釈の違いを表し、その際ほかの方法論で使われる特別な機構を必要としない。導出に線形論理を利用することで、意味的な完全性と一貫性が結果に保証されている。すなわち、すべての前提条件は充足され、使われなかった前提が線形論理の導出結果に残ることはない。Glue 意味論は元来、LFG に対して開発されたが、特定の文法フォーマリズムに適用が制限されているわけではなく、さまざまな構文的フォーマリズムや意味表現に適用することができる [Asudeh 01]。

Glue 意味論の解釈は二つのステップから構成される。最初のステップでは、与えられた文の構文的な項目から対応する意味構成要素を生成する。このステップでは、規則に基づいて *f-structure* を解析することで意味構成要素を得る。対応付けを行うための写像規則は、意味辞書で定義されている。

次のステップでは、意味構成要素を組み立てて、文全体の意味表現を導出する。このステップでは意味構成要素を導出する。このメカニズムは言語に関係なく共通に利用することができる。意味的な合成は意味構成要素の集合に対する線形論理の証明として行われ、意味項を合成し、生成可能なすべての意味表現を出力する。スコープの対象範囲などの意味的な曖昧性は、意味構成要素の同じ集合から成り立つ複数の証明で表されている。

Glue 意味論の解析器は XLE 内に組み込まれているが、言語に依存しないフレームワークを提供している。しかし意味辞書には、*f-structure* の項目を意味構成要素に写像する規則が定義されているので、言語に依存することになる。各写像規則は条件部分と本体部分に分かれている。条件部分にはチェックされることになる形態素ラベル、述語形式、ほかの *f-structure* との関係が記述されている。本体部分には、その条件が満たされた場合に提供される、意味構成要素およびスコープ制約条件が定義されている。

日本語と英語の *f-structure* は ParGram プロジェクトによってその並行性が担保されているため、英語の *f-structure* に適用するために構築された写像規則は日本語にも適用可能であることが [Umemoto 05] で報告されている。

5.2 Abstract Knowledge Representation (AKR)

Glue 意味論は線形論理によって厳格に導出されるものだった。そのため、意味表現としては非常に理想的な結果が得られるものの、カバー率を向上させることが難しく、実用的なシステムの意味解析エンジンとして採用するのは困難な面もあった。また、アプリケーションによってはこの意味表現の形式を扱い難かった。

そこで、より実用性の高い意味解析を目的として、[Crouch 05, Crouch 06] では *f-structure* を「書き換える

(Transfer)」機構を利用することで、抽象的な知識表現 (AKR) を出力する手法を提案している。元々、この Transfer という仕組みは ParGram プロジェクトで翻訳を行う際に用いられている XLE の組み込み機能である。

書換えは (3) のようなルールによって実現される。

```
(3)
PRED(%V, eat), SUBJ(%V, %S), OBJ(%V, %O),
-OBL(%V, %)
→
word(%V, eat, verb), role(Agent, %V, %S),
role(Theme, %V, %O).
```

このルールは *f-structure* の PRED が 'eat' で、OBL(斜格) は存在せず、SUBJ と OBJ を項にもつという条件を満たす場合には、PRED, SUBJ, OBJ をいったん削除したうえで、PRED を WORD に、SUBJ を意味役割の Agent に、OBJ を意味役割の Theme に書き換える。

既存の言語資源を利用することは大規模な意味解析システムを構築するうえで効率性が高い。[Crouch 06] では書換規則に WordNet や VerbNet の意味辞書が取り入れられている。その観点からもここで提案された手法は実用性が高いといえる。

6. LFG を用いたアプリケーション

6.1 機械翻訳

f-structure の応用先として最初に取り組みられていたのは機械翻訳だった。[Kaplan 89] では *f-structure* の枠組みを用いた翻訳手法について提案している。ここでは、元言語の文の *f-structure* を目的言語の *f-structure* に書き換えることによって翻訳を実現している。

[Frank 99] は、ParGram プロジェクトの初期の段階における英仏間の機械翻訳への取組みについて報告している。ここでは、曖昧性が表現された Packed *f-structure* を用いることで、元言語の曖昧性を解消せず、曖昧性を保持したまま目的言語に翻訳する手法を提案している。

[Oepen 05] で紹介されているプロジェクトでは英語 - ノルウェー語間の翻訳プロジェクトについて報告している。このプロジェクトは大変ユニークで、それまでの研究の経緯から、Head-driven phrase structure grammar (HPSG) と LFG の両方を用いたシステムを採用している。

6.2 含意判定

[Bobrow 07] では 5.2 節で解説した意味解析システムによって出力された AKR を利用して含意判定を行う手法について提案している。RTE タスクでは、入力されたテキスト (T) から仮説 (H) が推論できるか否かを判定する。ここでは 5.2 節のシステムに T と H のテキ

ストを入力し、その出力結果として二つの AKR を取得して比較を行う。この AKR には WordNet や VerbNet の情報も組み入れられているため、含意判定の際にはそれぞれの同義語、上位下位語などを参照することができる。RTE3 タスクのテストセットを使った評価結果では再現率は低いものの、極めて高い適合率が得られたことが報告されている。[梅基 08] ではこのシステムを日本語に適用した事例が紹介されている。日本語では WordNet や VerbNet の代わりに EDR 電子辞書や IPAL 辞書が用いられている。

6.3 要約

[Riezler 03] は LFG の曖昧性解消手法と packed f-structure を文圧縮 (sentence condensation) に適用する手法を提案している。このシステムではまず Maximum-entropy を用いて不要な PRED を削除することで f-structure の縮小を行い、生成の機構を使って圧縮された文を出力する。f-structure の縮小は元の文の f-structure と縮小された f-structure のペアを用いて学習される。

7. おわりに

以上のように、本稿では LFG 理論に基づく多言語処理、統計処理との融合、意味解析、応用システムの研究について紹介した。LFG 理論は非常に優れた文法理論であり各国で言語学的な研究が盛んに行われている。LFG 理論を利用することはこれらの言語学の成果の恩恵を受けることができるため、自然言語処理にとってもまた有効であることは間違いない。しかし、理論を理解するためには言語学的な知見が必要とされることや大規模な LFG 文法ルールを記述するには初期段階のコストがかかるため、一般に公開されている LFG のリソースはごくわずかではある。

しかし、4.2 節で紹介した文法自動獲得研究の事例では既存のコーパスをうまく使うことによって LFG 理論に基づいた実用的なパーザを実現している。また、統計処理だけではなく、5.2 節の意味解析システムでも WordNet や VerbNet さらには EDR 電子辞書を用いることによって大規模な意味解析システムを構築している。今後はこのように、LFG 理論の概念を取り入れながらも、ほかの手法や既存のシステムを用いた研究が考えられるかもしれない。

f-structure は構文構造を表現するための強力な「指針」と捉えることができる。言語学的にきちんと動機づけられた「指針」は自然言語処理にとって必要不可欠なものだと考えられる。これまでの研究で LFG 理論を自然言語処理にどのように取り入れてきたかを参考にすることによって、ポスト経験主義の言語処理に新たなヒントが得られることを期待したい。

◇ 参考文献 ◇

- [Asudeh 01] Asudeh, A. and Crouch, R.: Glue semantics for HPSG, *Proc. 8th Int. HPSG Conference* (2001)
- [Bobrow 07] Bobrow, D. G., Condoravdi, C., Crouch, R., Paiva, V. D., Karttunen, L., King, T. H., Nairn, R., Price, L. and Zaenen, A.: Precision-focused textual inference, *Proc. ACL-PASCAL Workshop on Textual Entailment and Paraphrasing* (2007)
- [Bresnan 78] Bresnan, J.: *Linguistic Theory and Psychological Reality*, chapter A realistic transformational grammar, pp. 1-59, The MIT Press (1978)
- [Butt 99] Butt, M., King, T. H. and Maria-Eugenia Nino, F. S.: *A Grammar Writer's Cookbook*, CSLI Publications (1999)
- [Butt 02] Butt, M., Dyvik, H., King, T. H., Masuichi, H. and Rohrer, C.: The parallel grammar Project, *Proc. COLING 2002 Workshop on Grammar Engineering and Evaluation*, pp. 1-7 (2002)
- [Cahill 02] Cahill, A., McCarthy, M., Genabith, J. V. and Way, A.: Automatic annotation of the penn treebank with LFG f-structure information, *Proc. LREC Workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data*, pp. 8-15, Las Palmas, Canary Islands (2002)
- [Cahill 04] Cahill, A., Burke, M., Genabith, J. V. and Way, A.: Long-distance dependency resolution in automatically acquired wide-coverage PCFG-based LFG approximations, *Proc. 42nd Meeting of the Association for Computational Linguistics (ACL)*, pp. 320-327 (2004)
- [Cahill 05] Cahill, A., King, T. H. and Maxwell, J. T.: Pruning the search space of a hand-crafted parsing system with a probabilistic parser, *Proc. 5th Workshop on Important Unresolved Matters*, pp. 65-72 (2005)
- [Cahill 08] Cahill, A., Maxwell, J. T., Meurrer, P., Rohrer, C. and Rosen, V.: Speeding up LFG parsing using c-structure pruning, *Proc. COLING 2008 Workshop on Grammar Engineering Across Frameworks*, pp. 33-40 (2008)
- [Crouch 05] Crouch, D. and King, T. H.: Packed rewriting for mapping semantics to KR, *6th Int. Workshop on Computational Semantics (IWCS-6)* (2005)
- [Crouch 06] Crouch, D.: Semantics via f-structure rewriting, *Proc. LFG06* (2006)
- [Dalrymple 01] Dalrymple, M.: *Lexical Functional Grammar*, Academic Press, New York (2001)
- [Frank 99] Frank, A.: From parallel grammar development towards machine translation, *Proc. MT Summit VII* (1999)
- [Kaplan 82] Kaplan, R.: *Lexical-Functional Grammar: A Formal System for Grammatical Representation*, pp. 173-281, The MIT Press (1982)
- [Kaplan 89] Kaplan, R. M., Netter, K., Wedekind, J. and Zaenen, A.: Translation by structural correspondences, *Proc. European Chapter of the Association for Computational Linguistics (EACL)* (1989)
- [Kim 03] Kim, R., Dalrymple, M., Kaplan, R., H. King, T., Masuichi, H. and Ohkuma, T.: Multilingual grammar development via grammar porting, 2003 *European Summer School in Logic, Language, and Information (ESSLLI03) Workshop "Ideas and Strategies for Multilingual Grammar Development"* (2003)
- [King 03] King, T. H., Crouch, R., Riezler, S., Dalrymple, M. and Kaplan, R. M.: The PARC 700 dependency bank, *Proc. 4th Int. Workshop on Linguistically Interpreted Corpora (LINC-03)*, pp. 1-8 (2003)
- [増市 03] 増市 博, 大熊 智子: 日本語 LFG による大規模構文意味解析システムの構築, 自然言語処理, Vol. 10, No. 2, pp. 79-109 (2003)
- [Maxwell 91] Maxwell, J. T. and Kaplan, R. M.: A method for disjunctive constraint satisfaction, *Proc. Int. Workshop on Parsing Technologies*, pp. 18-27, Kluwer Academic Publishers

- (1991)
- [Maxwell 93] Maxwell, J. T. and T. R. M. K.: The interface between phrasal and functional constraints, *Computational Linguistics*, Vol. 19, pp. 571-590 (1993)
- [O'Donovan 05] O'Donovan, R., Cahill, A., Genabith, van J. and Way, A.: Automatic acquisition of Spanish LFG resources from the Cast3LB treebank, *Proc. LFG05* (2005)
- [Oepen 05] Oepen, S., Dyvik, H., Flickinger, D., Lonning, J. T., Meurer, P. and Rosen, V.: Holistic regression testing for high-quality MT. Some methodological and technological reflections, *Proc. 10th Annual Conference of the European Association for Machine Translation* (2005)
- [Ohkuma 03] Ohkuma, T., Masuichi, H., Yoshimura, H. and Harada, Y.: The treatment of Japanese focus particles based on lexical-functional grammar, *17th Pacific Asia Conference on Language, Information and Computation* (2003)
- [大熊 06] 大熊智子, 増市博, 吉岡健: 日本語 LFG を用いた副助詞の多義性解消, 自然言語処理, Vol. 13, No. 1, pp. 27-52 (2006)
- [大熊 08] 大熊智子: 日本語 LFG における形容詞連用形の処理, 月刊言語, Vol. 8, pp. 74-81 (2008)
- [大熊 09] 大熊智子, 梅基宏, 三浦康秀, 増市博: 日本語 LFG にもとづく助数詞の処理, 自然言語処理, Vol. 16, No. 3, pp. 51-80 (2009)
- [Riezler 02] Riezler, S., King, T. H., Kaplan, R. M., Crouch, R., III, J. T. M. and Johnson, M.: Parsing the Wall Street Journal using a lexical-functional grammar and discriminative estimation techniques, *Proc. 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (2002)
- [Riezler 03] Riezler, S., King, T. H., Crouch, R. and Zaenen, A.: Statistical sentence condensation using ambiguity packing and stochastic disambiguation methods for lexical-functional grammar, *Proc. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL2003)* (2003)
- [Tounsi 09] Tounsi, L., Attia, M. and Genabith, van J.: Parsing Arabic using Treebank-Based LFG Resources, *Proc. LFG09* (2009)
- [Umemoto 05] Umemoto, H.: Implementing a Japanese Semantic parse based on glue approach, *20th Pacific Asia Conference on Language, Information and Computation (PACLIC20)* (2005)
- [梅基 08] 梅基宏, 杉原大悟, 大熊智子, 増市博: LFG 解析と語彙資源を利用した日本語含意関係判定, 情処自然言語処理研究会, No. 188 (2008)

2012年3月19日 受理

著者紹介



大熊 智子 (正会員)

1994年東京女子大学文理学部卒業。1996年慶應義塾大学大学院政策・メディア研究科修士課程修了。同年、富士ゼロックス株式会社総合研究所入社。現在同社研究技術開発本部コミュニケーション技術研究所研究主任。2009年より東京女子大学非常勤講師。博士(学術)。



増市 博 (正会員)

1991年京都大学大学院工学研究科修士課程修了。同年、富士ゼロックス株式会社入社。1998～2000年スタンフォード大学 CSLI 客員研究員および Xerox PARC コンサルタント研究員。現在、富士ゼロックス株式会社研究技術開発本部コミュニケーション技術研究所研究主幹。博士(工学)。