

Lexical Functional Grammar に基づく 多言語テキストへのアクセス

大熊智子[†] 増市 博[†]

[†]富士ゼロックス(株) 中央研究所

1 はじめに

Lexical Functional Grammar(LFG) は自然言語文の解析を行うための文法理論であり, (Bresnan 1978) によってその概念が提唱された後, (Kaplan and Bresnan 1982) 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(onstituent)-structure と f(unctional)- structure と呼ばれる 2 種の構造を出力する. c-structure は構文木を表し, f-structure は, 文法機能 (grammatical function) の概念に基づいて, 文の格構造, 時制, 様相, 話法等の意味情報を属性-属性値のマトリックス構造で表現する.

LFG 理論では, SUBJ(ect) や OBJ(ect) 等の文法機能をいかなる言語の解析においても有効であると考え, 第一義の未定義要素としての地位を与えている. すなわち, 見か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, 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 (Butt et al. 1999a; Dalrymple 2001).

言語解析システムの出力結果が言語を問わず一定であることは, 多言語処理において極めて重要な特徴であるといえる. まず, 出力結果すなわち f-structure を中間言語とみなすことによって, 意味構造を勘案した高品質の機械翻訳システムの実現が期待できる. 例えば, (Frank 1999) では f-structure を中間言語とする LFG ベースの機械翻訳手法の提案を行っている. また, f-structure を取り扱う言語処理アプリケーションをある一つの言語を対象に構築すれば, それを他の言語を対象とするようポーティングすることが極めて容易に実現できる点が利点となる. さらに, 複数の言語が混在するドキュメントデータベースを単一のシステムで取り扱うことが可能となる点も利点として挙げることができる.

本稿では, 英語, ドイツ語, フランス語, ノルウェー

語, ウルドゥー語, デンマーク語, そして日本語の 7ヶ国語を対象として f-structure の言語普遍性を重視した LFG システムの構築を目指すプロジェクト (Parallel Grammar Project) の活動を紹介します. 同プロジェクトで開発中の日本語 LFG システムの実装内容について述べる.

2 Parallel Grammar Project

我々は, Parallel Grammar (ParGram) と呼ばれるプロジェクト活動 (Butt et al. 1999b, 2002) の中で日本語 LFG 文法の記述を行っている. ParGram は, LFG を共通の理論基盤として, 1 章で挙げた 7ヶ国語を対象とする言語解析システムの実現を目標としている. 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持つことにしている. 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. 我々の日本語 LFG 文法が基本的な構文に対して出力する f-structure は, この ParGram 標準仕様 (Butt et al. 1999a) に準じている.

3 システム構成

ParGram では共通のプラットフォームとして Xerox Linguistic Environment (XLE) と呼ばれる LFG の処理系を使用している. 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 (Kaplan and Wedekind 2000).

(Maxwell III and Kaplan 1991, 1993) に基づく高速化手法によって高い解析効率を実現している点が XLE の最大の特徴である. XLE を採用することによって, 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となった.

図 1 は, 日本語 LFG システムの構成図である. まず, 日本語入力文が形態素解析システムに渡され, 得

Tomoko Ohkuma[†], Hiroshi Masuichi[†]

[†]Corporate Research Center Fuji Xerox co., Ltd.

{Ohkuma.Tomoko, Hiroshi.Masuichi}@fujixerox.co.jp

名に対する緩やかなネーミングコンベンションの取り決めはあるものの、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。図4、図5は、上に挙げた日本語文と英語文をそれぞれ解析した結果得られるc-structureである。両者の構造は全く異なっていることが分かる。

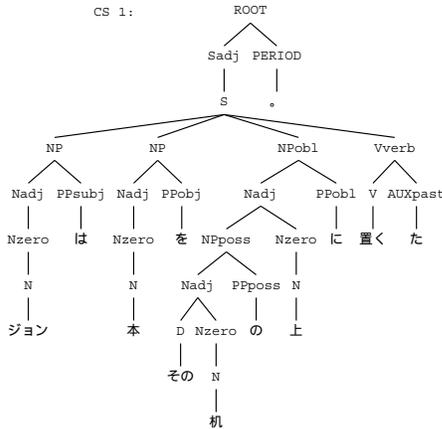


図 4: 日本語肯定文の c-structure 例

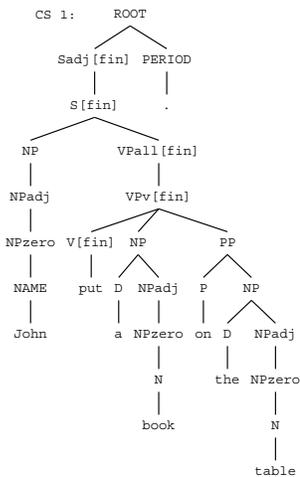


図 5: 英語肯定文の c-structure 例

図6は日本語命令文「その本を読んで下さい。」のf-structureである。図7は、英語命令文「Read the book.」を示すf-structureである。両者とも、SUBJのPREDには主語が表層に現われないことを示すnullを挿入している。また、構文の型を示すattributeのvalueには命令型(imp)を挿入する。またこの日本語文は「下さる」という補助動詞によって待遇表現になっているため、助動詞や補助動詞などの意味を示すattributeであるADDRESSにpoliteというvalueを入れている。

使役文はこれまでに挙げたmono-clausalな例文とは異なり、複数の主述構造を持つとみなす。LFGではこのようなbi-clausalな構造をXCOMP(外部から主語の

「その本を読んで下さる M_。」

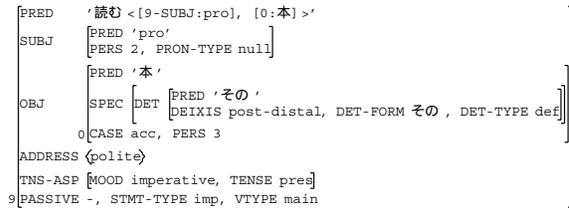


図 6: 日本語命令文の f-structure 例

"Read the book."

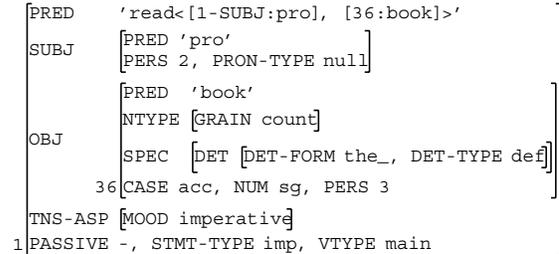


図 7: 英語命令文の f-structure 例

統制を受ける埋め込み句)によって表現する。日本語使役文「彼は彼女にその本を読ませた。」のf-structureと英語使役文「He let her read the book.」のf-structureを図8と図9に示す。これらを比較すると、日本語文において使役を担う「せる」と英語文における「let」が全体のPREDとして同じ位置に置かれていることが分かる。また、主語によって引き起こされた動作である「読む」'read'がXCOMPのPREDとなり、そのSUBJとして「彼女」と'she'、OBJとして「本」と'book'が、それぞれ代入される。

「彼は彼女にその本を読ませた。」

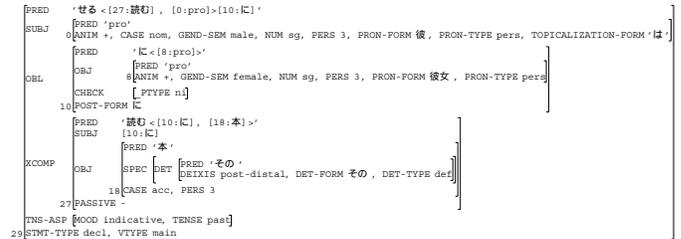


図 8: 日本語使役文の f-structure 例

"He let her read the book."

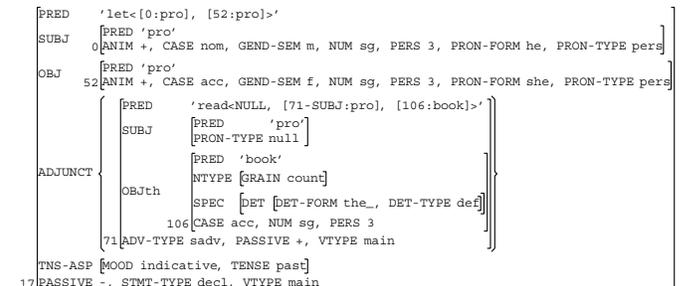


図 9: 英語使役文の f-structure 例

4.2 日本語文における曖昧性の処理

入力された構文が文脈や背景知識などによってのみ解消可能な意味の曖昧性を持つ場合、可能性のある意味全てに対応する f-structure を出力するという方針で文法記述を行っている (大熊, 増市 2002)。

例えば、形容詞連用形による連用修飾を含む構文「彼女は子供を優しく育てた。」という文では2つの解釈が可能である。1つは「子供が優しくなるように育てる」という意味である。つまり、「優しい」は「育てる」という動詞の帰結状態を示している (解釈 1)。もう1つは「子供の育て方」が優しいという意味である。つまり「優しい」は「育てる」という動詞の状況を示している (解釈 2)。

したがって、システムは上に掲げた文に対して2つの f-structure を出力する。図 10 は解釈 1 を示す f-structure である。OBL(ique) 中に格表示「ニ」、「ト」の省略を示す記号を挿入することにより、帰結状態の解釈であることを表現している。

図 11 は解釈 2 を示す f-structure である。解釈 2 に対応する英文の f-structure を図 12 に示す。両者を比較すると同じ構造を持っていることが分かる。

"その 女性 は 子供 を 優しい y__ 育てる た。"

PRED	'育てる <[0:女性], [10:子供], [18: 4]>'
	┌ PRED '女性'
SUBJ	SPEC [DET [DET-FORM その, DET-TYPE def]]
	└ 0 [CASE nom, TOPICALIZATION-FORM 'は']
OBJ	┌ PRED '子供'
	└ 10 [CASE acc]
OBL	┌ PRED '4 <[16:優しい]>'
	├ OBJ 16 [PRED '優しい <[10:子供]>']
	└ 18 [OBL-TYPE ni]
TNS-ASP	[MOOD indicative, TENSE past]
21	[PASSIVE -, STMT-TYPE decl, VTYPE main]

図 10: 解釈 1 を表現する f-structure

"その 女性 は 子供 を 優しい y__ 育てる た。"

PRED	'育てる <[0:女性], [10:子供]>'
	┌ PRED '女性'
SUBJ	SPEC [DET [DET-FORM その, DET-TYPE def]]
	└ 0 [CASE nom, TOPICALIZATION-FORM 'は']
OBJ	┌ PRED '子供'
	└ 10 [CASE acc]
ADJUNCT	┌ PRED '優しい <[16-SUBJ:pro]>'
	├ SUBJ [PRED 'pro']
	└ 16 [ADV-TYPE vpadv]
TNS-ASP	[MOOD indicative, TENSE past]
21	[PASSIVE -, STMT-TYPE decl, VTYPE main]

図 11: 解釈 2 を表現する f-structure

"The woman nursed children kindly."

PRED	'nurse<[11:woman], [86:child]>'
	┌ PRED 'woman'
SUBJ	NTYPE [SRAIN count]
	└ SPEC [DET [DET-FORM the_, DET-TYPE def]]
	└ 11 [CASE nom, NUM sg, PERS 3]
OBJ	┌ PRED 'child'
	├ NTYPE [SRAIN count]
	└ 86 [CASE acc, NUM pl, PERS 3]
ADJUNCT	┌ PRED 'kindly'
	└ 104 [ADEGREE positive, ADV-TYPE vpadv]
TNS-ASP	[MOOD indicative, TENSE past]
42	[PASSIVE -, STMT-TYPE decl, VTYPE main]

図 12: 解釈 2 を表現する英語文の f-structure

5 おわりに

以上のように、本稿では多言語間での整合性を保ちながら開発を進めている LFG に基づく大規模日本語文法の開発について述べた。今後は、1章で述べた多言語処理を実現するためのアドバンテージを得ることを目的として、f-structure の言語普遍性をより高める文法記述を継続すると共に、異なる言語間で f-structure の等価性を定量的に評価する手法の研究を行う予定である。

参考文献

- 大熊智子, 増市博 (2002). "日本語 LFG パーサによる助詞・助動詞の無形表示の統語意味解析." 情報処理学会研究報告 2002-NL-150.
- 情報処理振興事業協会技術センター (1987). 計算機用日本語基本動詞/形容詞辞書 IPAL (Basic Verbs/Adjectives) -辞書編. 日本情報処理振興協会.
- 松本裕治, 北内啓, 山下達雄, 平野善隆, 今一修, 今村友明 (1999). 日本語形態素解析システム「茶釜」 version 2.0 使用説明書. 奈良先端科学技術大学院大学松本研究室.
- Bresnan, J. (1978). "A realistic transformational grammar." In *Linguistic Theory and Psychological Reality*, pp. 1-59. The MIT press.
- Butt, M., Dyvik, H., King, T. H., Masuichi, H., and Rohrer, C. (2002). "The Parallel Grammar Project." In *Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002) Workshop 'Grammar Engineering and Evaluation'*.
- Butt, M., King, T. H., Nino, M. E., and Segond, F. (1999a). *A Grammar Writer Cookbook*. CSLI Publications.
- Butt, M., Dipper, S., Frank, A., and King, T. H. (1999b). "Writing Large-Scale Parallel Grammars for English, French, and German." In *Proceedings of the LFG99 Conference*, URL: <http://csli-publications.stanford.edu/>.
- Dalrymple, M. (2001). *SYNTAX and SEMANTICS Lexical Functional Grammar*. Academic Press.
- Frank, A. (1999). "From Parallel Grammar Development towards Machine Translation." In *Proceedings of MT Summit VII*, pp. 134-142.
- Kaplan, R. M. and Bresnan, J. (1982). "Lexical-Functional Grammar: A formal system for grammatical representation." In *The Mental Representation of Grammatical Relations*, pp. 173-281. The MIT press.
- Kaplan, R. M. and Wedekind, J. (2000). "LFG generation produces context-free languages." In *Proceedings of the 18th International Conference on Computational Linguistics*, pp. 425-431.
- Maxwell III, J. T. and Kaplan, R. M. (1991). "A Method for Disjunctive Constraint Satisfaction." In *Current Issues in Parsing Technology*, pp. 173-190. Kluwer.
- Maxwell III, J. T. and Kaplan, R. M. (1993). "The interface between phrasal and functional constraints." *Computational Linguistics*, 19(4), 571-590.