

CHILDES 用の構文解析プログラム GRASP

宮田 Susanne・寒河江 健二 (南カリフォルニア大学)・

ブライアン・マクウィニー (カーネギー・メロン大学)

The Syntax Parser GRASP for CHILDES

Susanne MIYATA, Kenji SAGAE (University of Southern California),
and Brian MACWHINNEY (Carnegie Mellon University)

GRASP (Sagae et al, 2010) は CHILDES データベースの親子会話の解析のために開発された構文解析プログラムである。GRASP は別のプログラムで追加される形態素タグの情報を利用し、文内のそれぞれの単語の役割および単語間の依存関係を抽出する。学習プログラムである GRASP はどの言語にも利用することが可能であるが、その言語の完璧に解析されたモデルコーパスによって訓練させる必要がある。本稿では GRASP 用に日本語特有の文法関係や役割を定義したうえで、日本語版 GRASP を提案し、サンプルコーパスをもとに評価する。

Keywords : 文構造, シンタックス, パーシング, 日本語, 文法発達

Sentence Structure, Syntax, Parsing, Japanese, Grammar Acquisition

1. はじめに

コンピュータ時代になってから言語研究の可能性が広がった。大規模の言語データベースを利用し、多人数の参加者の実際の言語使用を辿ることができるようになった。しかし、データベースが大きくなるほど、自動的な解析ツールが不可欠になってくる。今までの獲得研究では単語単位で検索したり、手動でコーディングしたりしてきたが、負担が大きいため、処理できる量が比較的少なかった。自動的に基礎的な文法タグ (品詞情報、活用形の解析、文構造の情報など) を付けることが可能であれば、データベースで公開されているデータをより効率よく解析できるようになる。

本研究では国際発話データベース CHILDES (MacWhinney 2000, 宮田 2012b, 宮田・森川・村木 2004) のために開発された統語解析プログラム GRASP (Grammatical Relations Analysis for Spontaneous Protocols; MacWhinney, 2008; Sagae, Davis, Lavie, MacWhinney, & Wintner 2007, 2010; Sagae, Lavie & MacWhinney, 2005) の日本語版を紹介する。

GRASP は CHILDES の入力フォーマット CHAT (宮田 2012 を参照) に従っている発話ファイルを解析し、それぞれの発話行に統語情報専用の行 (%gra 行) を加える。統語解析で単語と単語の間の依存関係を記号化し、それぞれの役割を表す。例えば「同じ本見る?」という文の場合は「同じ本」という目的語が「見る」という述語に支配されている。目的語がさらに「同じ」という修飾語とその依存先 (ヘッド) の「本」に分析される。視覚的に括弧で表すと、この文の構造が [[[おなじ]MOD 本]OBJ 見る] となる。GRASP はこのような依存関係の情報を語順に沿った番号で表している。各単語に番号が振られてから、二番目の情報としてその依存先の番号が加えられる。三番目の情報として文法的役

割が表記される。それぞれの情報は「|」（縦棒）によって分けられ、**順番|依存先|役割**の形になる。例文の「同じ本見る?」（例1）を見ると、まずは単語に番号が振られる。この例の場合は「同じ」が1番、「本」が2番、「見る」が3番、そして最後の句読点（疑問符）が4番になる。次に依存する単語の番号が振られる。最初の単語「同じ」が「本」（2番）に依存しているので1|2|となり、役割は修飾語（MODifier）である。「本」が目的語（OBJect）として3番の動詞に依存しているので、2|3|OBJとなる。主動詞の「見る」は文の一番上のヘッド（ROOT）として依存値0が与えられる。文の最後にある句読点はROOTに依存しているPUNCTとして表記される（PUNCTは以下の例で省略する）。また、オンライン状態で%gra行をダブルクリックすると、文構造が図として表示される（図1）。

(1) *CHI: 同じ 本 見る ?
 %gra: 1|2|MOD 2|3|OBJ 3|0|ROOT 4|3|PUNCT

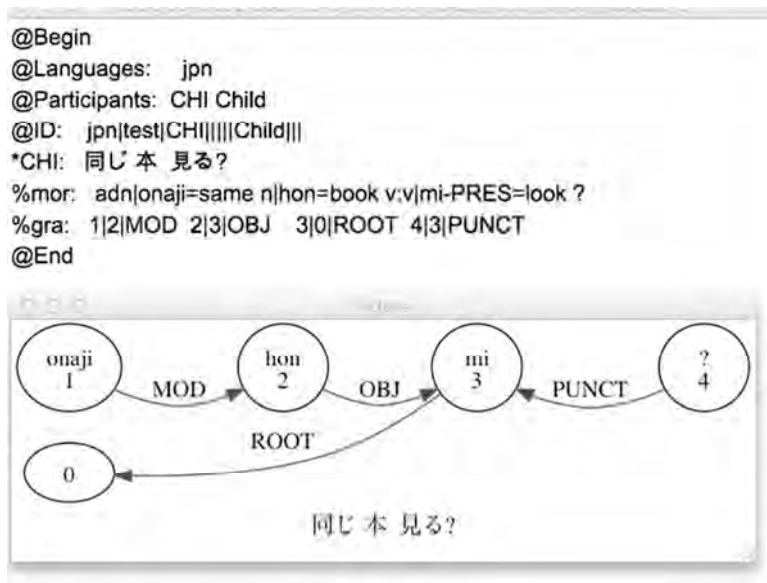


図1 GRASPによって抽出された依存関係の映像化

上記の解析結果を使って、多様な解析が可能になる。文構造パターンの研究や、ある動詞とともに現れる目的語や主語の一覧などを検索することができるようになる。さらに、CHILDESの入力フォーマットで文法の情報だけではなくスピーチアクトコード、エラーコード、意味論的な情報、音声解析など、様々な情報を含む行を加えることが可能であり、%gra行の統語的な情報を形態素情報（品詞や語尾）、スピーチアクトコードや意味論的な情報と組み合わせて使え、あらゆる視点からの解析が可能になる。

大規模のデータベースの解析法として活用するためには、信頼性の高い自動プログラムが必要となる。人の目で最終チェックが必要であっても、大部分を適切に解析できるプログラムであることが望ましい。その上、CHILDESの場合は子どもの言語が研究対象であるため、幼児とその保護者の特殊な言語使用に対応しなければならない。そのためにGRASPが固定された文法ルールの集まりではなく、学習プログラムという形になっている。

GRASPは完璧（“gold standard”）に解析されたサンプルコーパスを利用して訓練され、得られた解析ルールが補助ファイルの megrasp.mod にまとめられる。現在 CHILDES で配布されている日本語用のGRASPは標準語の2～3歳児とその母親のデータで訓練されているが、別の話者グループのデータで

訓練することも可能である。

GRASP プログラムを使う前提は次のようである。CHILDES のために開発されているため、CHILDES フォーマット (CHAT, MacWhinney 2000; 日本語の場合はそれに加えて Wakachi2002 v.4.0, 宮田 2002) に従っている発話ファイルが対象になるが、発話行のほかには形態素タグの情報も統語解析に必要である。そのため、GRASP を走らせる前に、MOR プログラムで形態素行の%mor を付ける必要がある。言語によって語形が異なるため、言語別に MOR の補助ファイルが必要であるが、日本語の場合は、12,000 語と 300 以上の活用語尾・接辞を含む辞書がすでに整備されている (中・宮田 1999; Miyata & Naka 2010; 詳しい説明は Miyata 2012 を参照)。

上記の例文に MOR プログラムによって%mor の行を付けると以下の形になる (例 2)。形態素タグの基本形は **品詞 | 語幹 - 語尾** である。例えばクッキーが名詞として n(oun) の品詞タグになる。母音動詞 (consonant verb) の食べるが v(erb):c(onsonant)の品詞になり、未完了の～るが-PRES(ent) として解析される。最後に意味論的解析等のための英語訳が付いている。

(2)	*CHI:	クッキー	食べる	?
	%mor:	n kukkii =cookie	v:v tabe-PRES=eat ?	

GRASP は品詞と語幹から文構造を推測する。動詞が一つしかないので、「食べる」が必然的に ROOT になるが、「クッキー」には格助詞が付いていないので、主語か目的語かはコンピュータにとって判断しにくいことである。学習訓練によって区別できるようになることもあるが、人の目でチェックする必要がある。

統語的役割には主語や目的語のような普遍的な役割もあるが、言語特有の品詞 (例えば連体詞) や役割もある。どのような文法関係が存在しているか、それぞれの品詞にはどのような役割があり得るかを、GRASP 用の文法で定義する必要がある。理想的には、親子会話で使われるすべてのパターンをカバーする「GRASP 文法」を定義することになる。

本論文ではこのような GRASP 用の統語関係と役割 (grammatical relations; GR) を実際の日本語発話データに基づいて定義することを目的とする。英語用の GRASP 文法 (Sagae, Davis, Lavie, MacWhinney, & Wintner 2007) を出発点として使い、必要に応じて日本語文法に新しく加える文法役割を提案する。日本語文法論としてはニュートラルな立場で、どのようなフレームワークでも使える文法表記をめざしている。なお、品詞の名称や定義は増岡・田窪 (1992) に従っているが、修正が必要と感じたところは明記してある。

以下は主動詞の ROOT をはじめに、文法役割を説明し、その表記を定義する。すべての文法役割タグとその例を含む一覧は<<http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/chapter08.html#08-2>> からダウンロードできる。

2. 日本語 GRASP の文法関連

2.1. 動詞が主ヘッドである文と動詞のない文

GRASP では文の依存関係の中で最上のヘッドが ROOT になる。時制を表す動詞 (例: 食べる) と形容詞 (美味しい) が典型的な ROOT である。従って「食べた?」という文は表 1-1 のように解析される。この典型的な動詞 ROOT のほかにはコピュラ (COPula) と述語名詞 (PREDicate) からなる ROOT がある。コピュラ「だ」・「です」は COPROOT として表記され、その前の名や形容名詞は述語名詞 (PREDicate) となる (表 1-2)。形容名詞 (ナ形容詞; 例: 奇麗です。) も同じ解析になる。

英語の場合は、時制を表す動詞やコピュラに限るが、日本語の場合はコピュラが省略される形も許

される（フォークは左手！）が、PREDROOT として解析される（表 1-3）。このような文型は最初から述語名詞がヘッドなのか、コンピュータが省略されたか（フォークは左だ）、それとも何らかの動詞が省略され、主語（フォークは左がいい）なのか、自由追加句（フォークは左で使う）なのかが明らかではないが、PREDROOT で無理な解釈を避けることができる。

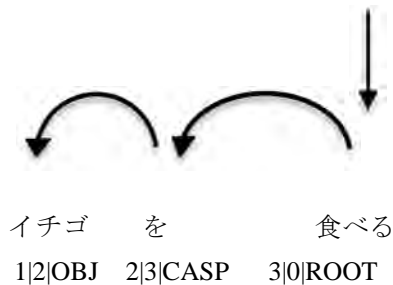
文のすべての項（主語、目的語、自由追加句、などの名詞句）が ROOT に依存している。ROOT は文全体が一番上のヘッドとして、IP ポジション（現在形や過去形などを含む屈折句；長谷川 1999:60）を含む。つまり、時制語尾を独立させないままで、動詞、形容詞、コンピュータが時制を含めた形で ROOT になる。

すべての項（主語もトピックも）が ROOT に依存しているように扱われる。同様に、文の外側（寺村 1982：II-219ff.）にある終助詞（例：食べるよ）やモダライザー（食べたいです）は ROOT に依存しているように扱われる。この省略的な表記は、より細かい正確な解析の支障にならない、シンプルで簡潔な表記である。

2.2. 格関係と格助詞

日本語は格関係を格助詞で表すことができるが、話し言葉では語順のみで表すことが多い。格助詞が利用された場合は、助詞がその項のヘッドになる。つまり、名詞が格助詞に依存し、格助詞が格を与える動詞等に依存していることになる（例 3）。

(3)



格の性質（どのような格関係を表すか）が名詞の役割で表されるので、格助詞がない場合は名詞が直接 ROOT に依存する（表 1-5）。

2.2.1. 格助詞の定義

日本語の格助詞の定義、どの助詞を格助詞として認めるかについては現在、一般的なコンセンサスが得られていない。「が」と「を」という二つの格助詞しか認めない学者（長谷川 1999；Shibatani 1990）から、3つ（「が」「を」「に」；Teruya 2007；Tsuji-mura 1996）、6つ（「が」「を」「に」「から」「と」「で」；仁田 1993:3）、8つ（「が」「を」「に」「で」「へ」「と」「から」「より」；町田・加藤 2001:77；Rickmeyer 1983:49）そして「まで」も含めて9つ（町田・加藤 2001:77；増岡・田窪 1992:49）の格助詞まで広げる学者がいる。その上に、「の」の位置づけについても議論があり、格助詞に含むか（町田・加藤 2001；長谷川 1999:66、Tsuji-mura 1996:134）、除くか（増岡・田窪 1992；Rickmeyer 1983；Teruya 2007）が未だにはっきりしない（「の」については 2.4.1 で述べる）。

格助詞の定義から考えると、格助詞は、ある名詞句（項 [argument]）が述語に対してどのような文法的な役割（格）を持っているかを表記す。どの格が可能なのが、その格を付与する動詞（または動作名詞、コンピュータや形容詞）で決まる。「食べる」のような他動詞は目的語を必要とし、対格 (accusative) を付与するので、「パンを」のように「を」が付く構成素が出現するが、「行く」などの自動詞が対格

を付与できない。

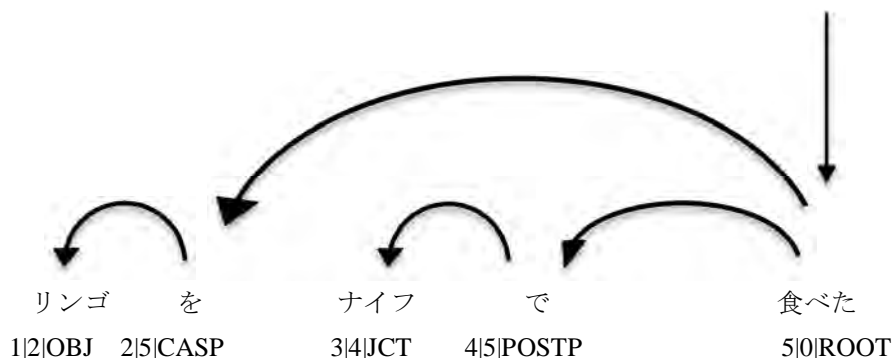
一方、自由にどの動詞にも加えられる構成素がある。副詞（例：ゆっくり）と同様に、場所（例：外で）や時間（例：朝から晩まで）、道具（例：フォークで）の構成素を自由に足せるが、加える必然性もない。これらの要素は「自由追加句」（free adjunct）と呼ばれ、それを表記する助詞は後置詞（post particle）と言う。

その後置詞と格助詞との間に線を引くのは難しい。極めて抽象的で格助詞らしいものから、具体的な意味を持ってより意味語に近いものがある。「が」「を」は格助詞であることには異論はないが、与格（dative）の「に」（例：ジョンに挙げた）、受け身文のときに行為者を表す起源格（ablative）の「に」（例：犬に食べられた）や対応格（symmetry）を表す「と」（例：リンゴと違う；ジャガ芋と合う）が存在する。それと比べ、場所的な関係を表す「で」、「から」、「より」、「まで」は意味が具体的で、文法的な性質は比較的弱い。つまり、格助詞と後置詞が連続帯（Payne 1997:100）になり、極端なもの、間に入るものがある。おそらく文法化によって、格助詞になろうとしている助詞であろう。

しかし、「が」、「に」と格助詞らしい「に」「と」の間には形態論的な差がある。ある構成素を提題にあげる時に「は」を使うが、格助詞の場合はその「は」が格助詞の代わりになり、格助詞が消える（例：を->は；パンは食べるが...）が、自由追加句の場合は格助詞も一緒に出現する（例：に->には；ジョンには挙げるが...）（長谷川 1999）。さらに動作名詞を含む節を名詞句に変えるとき、その項をマークする格助詞「が」「を」が「の」に変わるという現象があるが、「に」などの場合はこのような交換が不可能である（例：ミツバチを調査する=>ミツバチの研究；ジョンが調査した=>ジョンの調査；ジョンに贈呈する=>*ジョンの贈呈）。

従って GRASP では「は」とともに出現しない格助詞（CASP；「が」、「を」）と、それ以外の後置詞（POSTP；「に」、「で」、「へ」、「と」、「から」、「まで」、「より」）を分けることにする。その助詞を先行する名詞句は格助詞の場合は主格または対格（英語の GRASP と同様に SUBJ と OBJ を名付ける）、後置詞の場合は自由追加句（JCT）として解析される（例4）。

(4)



2.2.2. 格助詞や後置詞が省略された構成素の扱い

日本語の話し言葉の場合は、格助詞や後置詞を省略することがしばしばある。特に子どもに対する母親の言葉では助詞が使われないことが多い。それぞれの助詞の使用率に差があるが、0-5%程度（Rispoli 1995）でしか使われない対格の「を」から、ほとんど 100%程度で使われる後置詞の「で」まで、幅が広く、ともに使われる動詞や使用場面の影響もある（Miyata & Shirai 2008）。格助詞が省略

表1 日本語の GRASP 文法役割 (GR) の例一覧

番号	文法役割	例文
1	ROOT ルート (動詞)	食べた ? 1 0 ROOT 2 1 PUNCT
2	COPROOT ルート (コピュラ)	リンゴ だ 。 1 2 PRED 2 0 COPROOT 3 2 PUNCT
3	PREDROOT ルート (名詞)	フォーク は 左手 ! 1 2 TOP 2 3 TOPP 3 0 PREDROOT 2 1 PUNCT*
4	OBJ (+CASP) 目的語 (+格助詞)	イチゴ を 食べる 1 2 OBJ 2 3 CASP 3 0 ROOT
5		イチゴ 食べる 1 2 OBJ 2 0 ROOT
6	SUBJ (+CASP) 主語 (+格助詞)	イヌ が 食べた 1 2 SUBJ 2 3 CASP 3 0 ROOT
7	JCT (+POSTP) 追加語 (+後置詞)	家 に 帰る 1 2 JCT 2 3 POSTP 3 0 ROOT
8		家 帰る 1 2 JCT 2 0 ROOT
9	TOP (+TOPP) 提題 (+提題助詞)	イチゴ は 食べる 1 2 TOP 2 3 TOPP 3 0 ROOT
10		これ なに ? 1 2 TOP 2 0 PREDROOT
11		フォーク で は 食べない 1 2 JCT 2 3 POSTP 3 4 TOPP 4 0 ROOT
12	FOC (+FOCP) 取り立て (+取り立て助詞)	イチゴ しか 食べない 1 2 FOC 2 3 FOCP 3 0 ROOT
13		フォーク で しか 食べない 1 2 JCT 2 3 POSTP 3 4 FOCP 4 0 ROOT
14		サクランボ だけ が 高い 1 2 FOC 2 3 FOCP 3 4 CASP 3 0 ROOT
15	(PRED+) QUOTP (名詞的述語+) 引用助詞	淑子 と 言います 1 2 PRED 2 3 QUOTP 3 0 ROOT
16	(COMP+) QUOTP (動詞的述語+) 引用助詞	来る と 思った 1 2 COMP 2 3 QUOTP 3 0 ROOT
17	QUOTPROOT ルート (引用助詞)	行く って 。 1 2 COMP 2 0 QUOTPROOT 3 2 PUNCT
18	BQ ... EQ 引用マーク	「 アサガオ 咲いた 」 って 言った 1 3 BQ 2 3 SUBJ 3 5 COMP 4 3 EQ 5 6 QUOTP 6 0 ROOT

19	MOD (+ ATTP) 修飾語 (+ 修飾助詞)	パパ の イチゴ を 食べちゃった	1 2 MOD 2 3 ATTP 3 4 OBJ 4 5 CASP 5 0 ROOT
20	COORD (+COOP) 整合された名詞 (+ 並立助詞)	ナイフ と フォーク を 使って	1 2 COORD 2 3 COOP 3 4 OBJ 4 5 CASP 5 0 ROOT
21	CMOD 動詞的修飾節	洗った フォーク を 片付けた	1 2 CMOD 2 3 OBJ 3 4 CASP 4 0 ROOT
22		行く 予定 だ	1 2 CMOD 2 3 PRED 3 0 COPROOT
23	XMOD 連体詞の修飾語	いろんな おもちゃ が 出てる	1 2 XMOD 2 3 SUBJ 3 4 CASP 4 0 ROOT
24	COBJ 代名詞的「の」	洗った の を 片付けた	1 2 CMOD 2 3 COBJ 3 4 CASP 4 0 ROOT
25	CSUBJ 代名詞的「の」	いろんな の が 出てる	1 2 XMOD 2 3 CSUBJ 3 4 CASP 4 0 ROOT
26	CPRED 代名詞的「の」	行く の だ	1 2 CMOD 2 3 CPRED 3 0 COPROOT
27	JCT 追加語 (副詞)	ゆっくり 歩く	1 2 JCT 2 0 ROOT
28	XJCT 追加語 (動詞)	食べて 帰る	1 2 XJCT 2 0 ROOT
29		一人 で 食べる	1 2 PRED 2 3 XJCT 3 0 ROOT
30	COMP (+CPZR) 節 (+ 接続助詞)	来ない と だめ	1 2 COMP 2 3 CPZR 3 0 ROOT
31	XCOMP (+CPZR) 時制を含まない節	食べて から 帰る	1 2 XCOMP 2 3 CPZR 3 0 ROOT
32	(+ 接続助詞)	行く から 。	1 0 ROOT 2 1 CPZR
33	ZCPZR 接続詞	だから 帰った	1 2 ZCPZR 2 0 ROOT
34	PRED 名詞的述語	勉強 しよう	1 2 PRED 2 0 ROOT
35	SFP 終助詞	見た の よ	1 0 ROOT 2 1 SFP 3 2 SFP
36	SMDR ムード詞	帰った かしら	1 0 ROOT 2 1 SMDR
37	RDP 右外置	イチゴ 食べた よ ,, 熊さん	1 2 OBJ 2 0 ROOT 3 2 SFP 4 2 RDP 5 2 SUBJ

38	VOC (+VOCP) 呼びかけ (+マーク)	アキちゃん †	ご飯	食べた	か	1 0 VOC 2 1 VOCP 3 4 OBJ 4 0 ROOT 5 4 SFP	
39	ATTP 修飾助詞	パパ	の	イチゴ	を	食べた	1 2 MOD 2 3 ATTP 3 4 OBJ 4 5 CASP 5 0 ROOT
40	ATTP-OBJ (目的語省略)	パパ	の		を	食べた	1 2 MOD 2 3 ATTP-OBJ 3 4 CASP 4 0 ROOT
41	ATTP-SUBJ (主語省略)	パパ	の		が	落ちた	1 2 MOD 2 3 ATTP-SUBJ 3 4 CASP 4 0 ROOT
42	ATTP-JCT (追加語省略)	パパ	の		で	遊んだ	1 2 MOD 2 3 ATTP-JCT 3 4 POSTP 4 0 ROOT
43	ATTP-TOP (提題省略)	パパ	の		は	ない	1 2 MOD 2 3 ATTP-TOP 3 4 TOPP 4 0 ROOT
44	ATTP-PRED (名詞的述語省略)	パパ	の		だ		1 2 MOD 2 3 ATTP-PRED 3 0 COPROOT
45	ATTPROOT ルート (修飾助詞)	パパ	の		。		1 2 MOD 2 0 ATTPROOT
46	CASPROOT ルート (格助詞)	イチゴ	を		?		1 2 OBJ 2 0 CASPROOT
47	POSTPROOT ルート (後置詞)	どこ	で		?		1 2 JCT 2 0 POSTPROOT
48	TOPPROOT ルート (提題助詞)	イチゴ	は		?		1 2 TOP 2 0 TOPPROOT
49	FOCROOT ルート (取り立て助詞)	イチゴ	だけ		?		1 2 FOC 2 0 FOCROOT
50	PUNCT 句読点	食べた	?				1 0 ROOT 2 1 PUNCT

それでも格構造を明確にするために、格役割を格助詞ではなく、名詞で表記することにした。それによって、格助詞があっても（表1-4、6）なくても（例5）、ある名詞句が対格であるか、主格であるか、それとも追加句（7、8）であるかが分かる。この表記では助詞の有無と関係なく文の構造が明らかになり、今まで行われてきた手動の格コーディングも不要になる。

2.3. 提題・取り立て・引用

述語と項の格関係を表す格助詞のほかに、主題にフォーカスを当てる助詞があり、提題助詞 (TOPP ; 「は」等) と取り立て助詞 (FOCP; 「だけ」等) に区別できる。そのほかに、引用助詞 (QUOTP ; 「と」「って」等) もある。

2.3.1. 提題

日本語ではどの項でも、提題 (トピック) として強調することができる。文の先頭に移動させるこ

とが多いが、右外置も可能である（例：「難しいよ、これは。」；右外置については 2.10. を参照）。提題をマークする助詞として「は」が代表的であるが、「って」と「ったら」も提題助詞として使われることがある。GRASP では提題自体が TOP (topic)、提題助詞は TOPP (topic particle) として表記される (9)。

格助詞と同様に提題助詞も省略されることがあり、提題であるかどうか判断しにくい時もしばしばある。GRASP は提題助詞がなくても、次の二つの場合に提題として解析する：①指示詞と疑問詞を含む文（例：「これなに?」；表 1-10）、②「ゾウは鼻が長い」型の文（例：「僕お腹すいた」）。それ以外の場合は格関係を優先的に選択する。例えば「僕食べない」の場合は文脈によって主格も提題も解釈として可能であるが、GRASP は自動的に主格を選ぶ。従って、データの最終チェックで文脈を見ながら手動で直す必要がある場合もある。

また、後置詞と共に使われた場合は、TOPP が後置詞に依存するが、TOP の役割が表記されない (11)。

2.3.2. 取り立て

主題にせず、「ばかり」「だけ」「しか」等の取り立て助詞 (focus particle) で項に焦点を当てる言い方もある。GRASP では取り立てられた項が FOC の役割になる (12)。取り立て助詞は FOCP (focus particle) となる。後置詞とともに使われた場合は FOCP が POSTP を支配し、FOC が表記されない (13)。格助詞とともに現れる場合は FOCP が逆に CASP に依存し、取り立てられた名詞が FOC になる (14)。

2.3.3. 引用助詞

引用助詞の「と」「って」「ったら」は名詞と動詞をつなぐ意味では格助詞に似ているが、引用されるものは名詞句に限らず、形式が自由である。GRASP では引用される部分が時制を含むかによって、動詞等を含む修飾節 (finite COMPLEMENT ; 16 ; 以下 2.7. を参照) と、述語的名詞 (nominal PREDICATE ; 15 ; 以下 2.8. を参照) に区別している。引用助詞自体が QUOTP として表記される。述語が省略された場合は、QUOTPROOT として文のヘッドとして扱われる (17 ; 以下 2.12.2. を参照)。

さらに、引用が「」で囲むような文の場合は、「」が句読点として扱われ、BQ (Begin Quotation) と EQ (End Quotation) として表記される。BQ も EQ も引用文の ROOT に依存している。引用文の ROOT が COMP (時制含む) や PRED (時制含まない) として主文の引用助詞 (QUOTP) に依存することになる (18)。

2.4. 名詞句内の構造と修飾節

名詞句内の構造は引用と同様に修飾の部分が時制を含むかによって形式が異なる。修飾の部分が名詞であると修飾助詞の「の」(2.4.1.) やほかの整合助詞 (2.4.2.) でつなぐが、時制を含む動詞等の場合は直接名詞の前に置かれる (2.4.3.)。

2.4.1. 修飾助詞の「の」

名詞と名詞をつなぐ「の」は伝統的に格助詞として分類されるが、統語的な観点から見ると、ほかの格助詞と違って、述語から格を付与される訳ではないので、異質なものである。

欧米の文法研究を参考にすると、格が述語と名詞句の依存関係に限らず、名詞句内の構造を表す属格 (genitive) も定義に含むことが多いが (Trask 1993:35, Kroeger 2005 など)、それは属格が述語から出されることもあるからであろう。例えばコンピュータから付与される構造 (英: “this is yours / mine”) や、特定の動詞や形容詞とともに使われる属格がある (独: “sie klagt ihn des Diebstahls an” [彼女が彼を窃盗で責める]; 英語訳では前置詞 *of* が使われる: “she accuses him of theft”)。つまり純粋な格助詞と

して機能している場合もある。

面白いことに、日本語にも「の」の格助詞的な使用法が存在している。いわゆる「が/の」交替が起きている複文の場合は、「の」が複文の述語から付与された格助詞である（長谷川 1999；例：子どもの書いた字を見て、先生が微笑んだ）。また、動作名詞を含む節を名詞句化するときも格助詞が「の」に変わる（例：ミツバチを調査する => ミツバチの研究；ジョンがを調査した=>ジョンの調査）。

増岡・田窪（1992）は「の」の文構造の中の役割を重視し、接続助詞の一種として見ている。節と節を結ぶ接続詞（例：反意接続詞の「けど」「が」と、名詞と名詞を結ぶ接続助詞（例：「と」「や」「か」など）を区別し、後者に「の」を含む。しかし、「の」はその他の接続詞と違って相互的な関係ではなく、依存関係（所有、所属、特質等）を表す。形態的な面から見ると、「の」の場合は2番目の名詞を省略することが可能であるが、「と」などではそれが不可能である（例：パパのがあった；*ジョンとが遊んだ）ことから、「の」がやはり異質であると考えられる。寺村（1982：I-52）は「の」を連体助詞、「と」「や」「か」を並立助詞と名付け、分けているが、GRASPでも同じような解析にする。名詞句内の関係を表す「の」を「修飾助詞」(attributive particle; ATTP)とし、「と」「や」「か」を並立助詞(COOP；2.4.2.)とする。修飾助詞の場合は、最初の名詞が修飾語(modifier; MOD)、2番目の名詞(ヘッド)が名詞句全体の役割の記号になる(19)。

2.4.2. 並立助詞

並立助詞（「と」「や」「か」）は増岡・田窪（1992）の名詞と名詞を結ぶ「接続助詞」に該当するが、名詞と名詞をつなぐ助詞として、新しい名詞句を作るものである。修飾助詞と違って、その二つの対象物が平等な関係になるが、統語論的に見ると、最初の名詞が2番目の名詞に支配される。

GRASPでは並立助詞がCOOP (coordination particle)と表記される。最初の名詞を整合された部分としてCOORD (coordination)、2番目の名詞がその全体の名詞句の役割を表す(20)。

2.4.3 修飾節

修飾の部分が名詞ではなく、時制を含む修飾節の場合は、動詞句と名詞をつなぐ助詞が不要になり、時制を含む修飾節(Complementizing Modifier；CMOD)として直接名詞の前に置かれる(21、22)。CMODとして動詞の他に形容詞やコンピュータの「な」「だった」も使われ(例：美味しいご飯；綺麗な花；社長だった人)、時制は完了形の「～た」と未完了形の「～る」に限る。

また、時制を含まない、または活用語尾が付いていてもそれが化石化(morphologically fossilized)した不規則的な形容詞が連体詞(adnominal；例：小さな、細かな、いろんな、ろくな、たいした、いわゆる、同じ)と呼ばれる。特徴として、時制語尾の変化が不可能で、完了形(または未完了形)に変えることができない(例：たいした->*たいする；いろんな->*いろんだった、等)。時間の情報を含まないものとしてCMODではなく、XMODと表記される(23)。

2.5. 代名詞的「の」と「のだ」文

従属助詞を含む名詞句のヘッドをそのまま省略することが可能であるが(例：パパのイチゴを食べた->パパのを食べた；2.12.1.を参照)、動詞等を含む修飾節の場合はヘッドの代わりに「の」を入れる必要がある(例21、24)。この「の」は従属助詞の「の」と違って、名詞の代わりになるもので、「こと」「とき」と並び、「形式名詞」として解析されることがある(増岡・田窪 1992)。しかし、形式名詞は「意味的に希薄、修飾要素なしでは使えない」(p. 36)ものとして定義されているが、文法化途中の単語も含まれる。その場合はいくつかの使用法が平行し使われることがあり(例：わけがわからない；あいだが開く、等)、形式名詞と普通名詞の間に明白な線を引くのが難しい。一方、「の」は文法化中

の名詞ではなく、むしろ名詞の代わりになる「代名詞」的な助詞と考えられる。従って GRASP では「の」だけを「名詞的助詞」とし、「C～」でマークする。例えば目的語の名詞の代わりになった「の」は COBJ になる (24)。「の」の役割は格役割に応じて、CSUBJ (23、25)、COBJ、CJCT 等がある。なお、「のだ」文が上記の構造の一つとして考えられる。この場合は「の」が主語等ではなく述語名詞 (PREDicate) の代わりになるので CPRED として表記される。(例 22、26)。

2.6. 副詞と述語依存従属節

副詞 (例: ゆっくり、とても、等) は動詞等の修飾語として、そのヘッドの動詞や形容詞に依存する。GRASP では JCT (adverbial adJunCT) として解析される (例 27)。また、時制を含まない修飾語、例えば連用形 (テ形) の動詞や形容詞 (食べて帰る; 甘くて美味しい) が XJCT となる (例 28)。

GRASP ではこのような2つの主動詞を含む構造と、補助動詞 (subsidiary verbs) との構造 (食べている、食べてみる、等) を区別しない (しかし形態素解析の %mor 行では、補助動詞が v:sub として表記されるので検索が可能である)。そのほかの XJCT として、仮定形の「～たら」「～れば」(終わったら変える; 考えれば分かる)、交代の「～たり」(書いたりする) があげられる。さらに、形容詞を副詞形にさせる「～く」(美味しく食べる; 長くなる) も同じく XJCT として解析される。

時制を含まない修飾語にはテ形のコンピュータの「に」(学生になる)と「で」(「であって」の異形態) も含む (29)。後置詞の「で」と違って、コンピュータの「で」は「であって」や「でありながら」と置き換えることができる (例: 学生で (ありながら) 賢い; フォークで *(あって) 食べる; 学校で *(ありながら) 食べる)。

2.7. 接続助詞と複文

複文 (動詞的補語) と主文をつなぐのに接続助詞 (例: から、けど、のに、と、等) を利用する。複文の時制を担う動詞や形容詞 (complement; COMP) が接続助詞 (complementizing conjunction particle; CPZR) に依存する (30)。

ほとんどの接続助詞は完了形・未完了形の時制を含む述語の後に使われるが、時間を表す「から」と条件を表す「も」が接続形 (連用形) の「～て」を要する。動詞的補語の部分が時制を含まないので XCOMP として解析される (31)。また、終助詞的な使用法で、文が接続助詞で終わる場合は、終助詞と同様に述語の ROOT に依存する (32; 終助詞は 2.9.を参照)。逆に複文が省略され、主文が接続詞 (だから、だけど、でも、等) で始まる文もある。補語が zero なので ZCPZR (Zero-ComPlementiZeR) として表記される (33)。

2.8. 名詞的述語

名詞的述語 (PREDicate) は述語の代わりにコンピュータと共に使われる名詞である (例: ご飯だ)。基本形では名詞が PRED になり、コンピュータが COPROOT となる (2)。コンピュータの活用形によって名詞句の修飾節 (学生だった人) や動詞に補語節 (医者になる; 一人で食べる) になることがあるが、いずれの場合には名詞が PRED として解析される。また、コンピュータが省略された場合は、PRED が PREDROOT に変わる (3)。

また「ナ形容詞」(増岡・田窪 1992; 例: 綺麗だ) も同様に扱う。増岡・田窪と違い、「だ」「な」「に」などを語尾ではなく、独立したコンピュータとして考える。語幹自体は名詞の一種として「形容名詞」と名付け、上記の名詞+コンピュータと同様に解析する。

動作名詞 (例: 勉強、旅行、スタート) と形式動詞の「する」(増岡・田窪、1992:19) の組み合わせも同様に PRED と ROOT として扱う (34)。

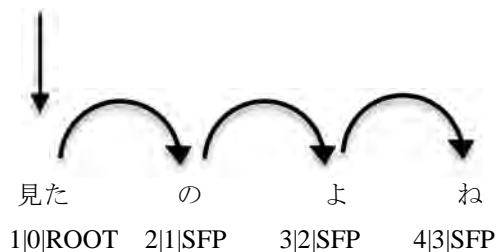
2.9. 終助詞とムード詞

話し言葉では文末にムードを表す単語を付けることによって、話し手の主観的な意見が態度をニュアンスとして表すことができる。形式の面から見ると終助詞と、それ以外の様々なムード詞に分けることができる。

2.9.1. 終助詞

話し言葉では文末に終助詞（ね、よ、か、さ、ぞ、の、等）を付けることが多く、いくつかの終助詞を組み合わせることもしばしばある。「ね」と「さ」は文内でも使うことができる（例：これはね、見たらね、変だったの）。両方の終助詞は SFP (Sentence Final Particle) として表記される。依存先が ROOT である。組み合わせの場合は先行する終助詞（例 5；表 1-35）、文内の場合は先行する名詞や助詞になる。

(5)



2.9.2. ムード詞

ムード詞は品詞の面から見ると、様々なものを含むが、話し手の態度を表すことと、終助詞に文法化しつつある単語である点で共通している。GRASP では SMDR (Sentence MoDalizeR) として表記され、終助詞の SFP と同様に ROOT に依存している (36)。

ムード詞の「です」はコンピュータの化石化したものとしてポライトネスを表す。文内終助詞としても使われるし（例：これはですね...）、時制を担っている形容詞のあとにも現れる（美味しかったです）。推定を表す「でしょう」、「だろう」、「じゃない」「じゃん」は SMDR として文末に限られるが、時制を表している形容詞だけではなく、動詞やコンピュータの後にも使われる（例：行くでしょう；これだったじゃん）。

不確かさを表す「みたい」と「らしい」はそのまま文末に現れている場合に SMDR として分類される（例：行くみたい；帰るらしい）。そのほかに「かもしれない」とその変形「かしら」と節のあとに使われる「っぽい」（帰ったっぽい）。

2.10. 右外置

話し言葉のもう一つの特徴は右外置である（“afterthought”；Martin 1973）。文の一部が述語の後に、つまり文の右外に置かれる形である（例：見たよ、あの人）。文のどの部分も外置できるし、連続でいくつかの要素を外置することもできる。右外置はロー・ピッチによってマークされているので、書き起こされた文章ではその情報が失われやすく、呼びかけや名詞句のヘッドとして間違えられる可能性がある（例：イチゴ食べた、ジョン）。このような誤解を防ぐため、CHAT フォーマットでは「,,」（ダブルコンマ）で表記される。GRASP では RDP (right dislocation) として句読点の一種として表記される。依存値は ROOT になる (37)。

2.11. 呼びかけ

呼びかけは CHAT フォーマットで「‡」（ダブルダガー）で呼格としてマークされ、他の文の構成素と区別できるようになっている。英語の場合は「‡」がコンマの代わりに使われ、文と呼びかけの区切り目を表すが、日本語の場合は呼びかけが文の後ろでも途中で使われることがあり、外置と混同区別する必要があるため、「‡」を必ず呼びかけの後ろに付けることになっている。

呼びかけ (VOC) が文構造と独立しているため、GRASP では依存値が 0 となる。呼びかけマーカのダブルダガー (VOCP) が句読点として、VOC に依存することになる (38)。

2.12. ヘッドの省略

日本語ではいろいろな省略が可能である。項全体を省略したり、助詞を省略したり、そして構成素のヘッドや主動詞なども省略することができる。主語や目的語の格助詞の省略、そして自由追加句の後置詞の省略については上記で触れたが (2.2.2.)、ここでは構成素のヘッドの省略の GRASP での扱いについて述べる。名詞句のヘッドの省略については 2.12.1.、述語の時制を持つ動詞などの省略については 2.12.2 で説明する。

2.12.1. 名詞句のヘッドの省略

2.4. で述べたように、名詞句には 2 種類があり、修飾の部分が名詞であるもの (例: パパの靴) と時制を含む節であるもの (例: 落ちた靴) に分かれるが、後者については別のところで扱う (代名詞的「の」; 2.5.)。

名詞と名詞は修飾助詞「の」で結ばれるが、2 番目の名詞を省略することができる。その場合はその名詞句の文の中の役割が残った「名詞+の」に移るが、GRASP の表記としては、修飾助詞に付ける。表 1 の 39 を考えると、「イチゴ」が目的語 (OBJ) の役割を担っているが、その「イチゴ」が省略されると、残った「パパの」の部分が OBJ の役割を持つことになる。表記として、OBJ が修飾助詞 (ATTP) に移り、ATTP-OBJ となる (40)。

文構造によって、ATTP-OBJ のほかに ATTP-SUBJ (41) そして自由追加句の ATTP-JCT (42) や提題の ATTP-TOP (43) もある。名詞的述語 (PRED; 2.8.) の場合も同じ移動が起き、「の」ATTP-PRED という役割になる (44)。さらにコピュラも省略されると、ATTPROOT に転化する (45; 2.12.2. も参照)。

2.12.2 述語のヘッド (動詞) の省略

述語として、時制を持っている ROOT (動詞や形容詞) と COPROOT (コピュラの場合) がある。コピュラが省略されると、その前の名詞や形容名詞が PREDROOT (2.1. を参照) になる。ここでは ROOT の動詞や形容詞の省略について述べる。また、「イチゴを食べる」(4) のような文で動詞を省略すると、ROOT の役割がその前の助詞に移り、格助詞の場合は CASROOT (46)、後置詞の場合は POSTROOT (47)、そして提題助詞の場合は TOPROOT (48) として表記される。すべての ROOT が依存値が 0 になり、文の一番上のヘッドとして解析される。

同様に取り立て助詞の場合は FOCROOT (49)、引用助詞の場合は QUOTROOT (17) となる。なお、助詞が付いていない場合 (例: イチゴ?) は、名詞自体が PREDROOT (3) になる。名詞の代わりに代名詞的「の」(2.5.) が使われている場合は、「の」が CPREDROOT となる (例: 馬の小さいの。)。一方、接続助詞 (CPZR; 2.7.) で終わる文は、終助詞 (SFP; 2.9.) と同じ扱いになり、述語の ROOT から依存しているものとして表記される (32)。

3. 信頼性

上記の文法構造で訓練された GRASP プログラムの信頼性を計るために、テストを行う。具体的には、二人の2歳台の子ども（平均発話長 3.0；宮田 2012a を参照）とその母親の発話を GRASP で解析し、手動で解析した結果と比較する。誤りのタイプとして役割（ラベル）の誤り（例えば 1|2|OBJ の代わりに 1|2|SUBJ として解析された場合）と依存先の誤り（例えば 2|5|CASP の代わりに 2|3|CASP、あるいは 2|3|POSTP になった場合）、という2つのパターンが区別できる（例5）。信頼性の指標として、ラベルの誤解率、依存先の誤解率、そして100発話当たりの誤解率を算出し、年齢に因る影響、およびエラーのパターンを分析する。

(5) *MOT: リンゴ を ナイフ で 食べた。
 %gra: 1|2|SUBJ 2|3|CASP 3|4|JCT 4|5|POSTP 5|0|ROOT 6|5|PUNCT ← test
 %grt: 1|2|OBJ 2|5|CASP 3|4|JCT 4|5|POSTP 5|0|ROOT 6|5|PUNCT ← gold

3.1 方法

訓練コーパスをコーディングし、GRASP プログラムを以下のように訓練した。訓練コーパスとしては Miyata Corpus (Miyata, 2004a, b) の Ryo (2;6~2;7 と 2;10~2;11) の9ファイル、および Tai (2;0, 2;6, 3;0) の4ファイル（合計 14,518 語、4,881 発話；表2）を選び、子どもおよび母親の発話に形態素解析プログラム JMOR06 (Miyata & Naka, 2010) で形態素タグ (%mor 行) を加えた。続いて、GRASP を走らせ、文法タグ (%gra 行) を加え、上記の文法構造に従って手動で訂正した。この完璧な (“gold standard”) の文法タグを %grt と名付けた。再び GRASP を走らせ、新しくできた %gra 行を gold standard の %grt 行と比較させ、GRASP を訓練した。

表2 訓練コーパスとして利用したデータのファイル名、子どもの年齢、発話数、発話数

ファイル名	子ども の年齢	子ども		母親	
		発話数	MLUm	発話数	MLUm
r20605.cha	2;06.05	197	2.6	122	4.5
r20612.cha	2;06.12	199	2.8	78	3.6
r20719.cha	2;07.19	88	2.6	45	4.2
r20725.cha	2;07.25	153	2.8	71	3.2
r21007.cha	2;10.07	237	3.3	156	4.7
r21010.cha	2;10.10	248	2.9	99	4.2
r21017.cha	2;10.17	135	3.1	48	4.5
r21116.cha	2;11.16	124	3.3	77	4.5
r21123.cha	2;11.23	141	3.4	82	4.5
tai940414.cha	2;0.4	95	3.8	220	5.9
tai940421.cha	2;0.11	431	2.4	608	4.3
tai941020.cha	2;6.10	404	3.5	663	4.1
tai950413.cha	3;0.3	69	4.8	91	6.1
合計		2,521		2,360	

評価は以下のように行った。テスト・コーパスとして、同じ両コーパスから 2;7 と 2;10 台のから訓練コーパスに含まれていないファイルを選び、約 300 発話ずつ（合計発話数：1,247；表 3）を抽出した。形態素タグ（%mor）と文法タグ（%gra）を加え、手で gold standard の %grt 行を作成した。続いて、訓練コーパスで訓練された GRASP でテスト・コーパスを解析し、出力された %gra 行を CLAN の trnfix プログラムによって gold standard の %grt 行と比較した。分析内容として、依存先の誤りの回数、そして役割（ラベル）の誤りの回数を抽出した。

表 3 テスト・コーパスのファイル名、子どもの年齢、発話数、発話数

年齢	ファイル名	子ども			母親		
		発話数	単語数	MLUm	発話数	単語数	MLUm
Ryo 2;7	r20704.cha, r20711.cha	224	728	3.3	93	308	4.0
Ryo 2;10	r21024.cha, r21101.cha	222	768	3.8	91	296	3.9
Tai 2;7	tai941110.cha	145	358	3.0	163	536	4.5
Tai 2;10	tai950216.cha	132	353	3.3	177	492	3.9
	合計	723	2,207		524	1,632	

3.2 結果

GRASP の出力を gold standard に比較した結果、3,839 単語のうち、45 単語（1.17%）の依存先が誤っていたことが分かった。さらに 64 単語（1.67%）の依存先が適切だったが役割ラベルが誤っていた（表 4）。合計として 109 単語（2.84%）の解析結果が誤っていた。発話数から見ると、1,247 発話のうち、75 発話（6.01%）が誤りを含んでいた。逆に言えば、単語の 97.16%、そして発話の 93.99% が適切な解析となっていた。

子どもと母親の結果を比較すると、子どもが 61 エラー（全単語の 2.76%）で、母親が 48 エラー（2.94%）になり、大きな差が認められなかった。エラーの種類別に見ると、役割エラー率は子ども（1.86%）と

表 4 子どもおよび母親の GRASP エラー（単語当たり、発話当たり）

子ども/年齢	役割エラー (%)	依存先エラー (%)	エラー単語数合計 (%)	エラー発話数 (%)
Ryo 2;7	18 (2.47)	1 (0.14)	19 (2.61)	18 (8.04)
Ryo 2;10	12 (1.56)	5 (0.65)	17 (2.21)	13 (5.86)
Tai 2;7	3 (0.84)	3 (0.84)	6 (1.68)	4 (2.76)
Tai 2;10	8 (2.27)	11 (3.12)	19 (5.38)	11 (8.33)
小計（子ども）	41 (1.86)	20 (0.91)	61 (2.76)	46 (6.36)
Ryo 2;7 (Mot)	4 (1.30)	0 (0.00)	4 (1.30)	3 (3.23)
Ryo 2;10 (Mot)	9 (3.04)	4 (1.35)	13 (4.39)	7 (7.69)
Tai 2;7 (Mot)	4 (0.75)	9 (1.68)	13 (2.43)	8 (4.91)
Tai 2;10 (Mot)	6 (1.02)	12 (2.64)	18 (3.66)	11 (6.21)
小計（母親）	23 (1.41)	25 (1.53)	48 (2.94)	29 (5.53)
合計	64 (1.67)	45 (1.17)	109 (2.84)	75 (6.01)

母親 (1.41%) の間にほとんど差が見られなかったが、依存先のエラー率は母親の場合のほうが高かった (子ども 0.91%、母親 1.53%)。さらに子どもの結果を年齢別に見ると、似た結果になった。つまり役割のエラー率は 1.93 (2;7) と 1.78% (2;10) でほとんど差がなかったが、依存先のエラー率が 0.37% (2;7) から 1.43% (2;10) に上がった。

さらに主な役割の精密性 (Sagae et al. 2010) を比較してみた。その比較の出発点として gold standard (%grt 行) のラベル (例えば SUBJ) を、GRASP の解析結果 (%gra 行) と比較し、正解率を計算した (表 5)。その結果、MOD (修飾語; 1.00)、ROOT (主動詞など; 0.99)、JCT (自由追加語; 0.97) と PRED (名詞的述語; 0.96) の精密性がかなり高かったが、SUBJ (主語; 0.88)、そして特に OBJ (目的語; 0.73) が低かったことが分かった。

表 5 5 種類のラベルの精密性 (gold standard から見た正解率)

役割	test ラベル数	gold ラベル数	test ラベル 誤解数	test ラベル 正解数	test ラベル精 密性
ROOT	456	456	4	452	0.991
SUBJ	177	166	10	156	0.881
OBJ	63	59	13	46	0.730
JCT	104	117	16	101	0.971
PRED	118	119	6	113	0.958
MOD	41	41	0	41	1.000

3.3 考察

今回提案した文法役割に従って訓練された GRASP の信頼性を 2;7 と 2;10 の子どもとその母親のテストデータをもとに調べた。GRASP が出力した解析結果を手動で解析したものと比較したところ、全単語の 97% が正確な文法的役割および依存先の解析になっていたことが分かった。エラーの種類として 2.84% の単語は依存先が間違っており、1.67% は役割のみが間違っていた。英語の結果と比較すると、やや良い正解率を得た (英語: 依存先 4.71%、役割 6.09%、Sagae, Davis, Lavie, MacWhinney, & Wintner 2010:721)。

日本語の話し言葉では色々なレベルの省略が許され、文法役割を表す格助詞・後置詞が表出されないことが多い。語順 (SOV) は項の役割解釈の手がかりになるが、項自体も省略されることがあるため、表出された項の役割が形態レベルで判断できないことが多い。日常会話では聞き手が文脈や意味論的な情報を使いながら解釈して行くが、コンピュータープログラムは文単位で形態素情報しか利用できないため、推測が不可能なときがある。例えば、「学校行く?」の場合は「学校」が場所として意味論的に追加語 (JCT) にしかなり得ないが、文構造的には「名詞句+動詞」であるため、主語と同じ形を取っている。訓練された GRASP では、この文がすでにレパートリーに含まれている場合は正しく出力されるが、別の地名が使われた場合は主語として解釈されることもあり得る。

また、全体の文脈からしか解釈できない文構造もある。例えば「蟻さん食べた?」の場合は「蟻さん」が生物であるため意味論的には主語にもなり得るし、餌として目的語にもなる可能性がある。この問題は訓練だけで解決できないものであろう。この問題はラベル別の精密度にも現れた。修飾助詞 (「の」) や後置詞 (「に」「で」「から」など) に比べ、格助詞「が」と特に「を」の省略の頻度が高いことが GRASP の精密度に影響したと考えられる。常に「の」によってマークされている修飾語が 1.00

程度で正確に MOD として解析されたが、格助詞が省略されやすい目的語の精密性が 0.73 で比較的に低かった。目的語のエラーパターンとして、特に SUBJ との履き違いが多く、役割エラーの 28% を占めていた。

子どもと母親のエラー率を見ると、大きな差が認められなく、依存先のエラー率だけが母親の場合に子どもに比べやや高い傾向があった。2;7 と 2;10 の子どもの結果にも同じ傾向が見られた。この結果は英語の結果に反している。英語の場合は、むしろ母親の解析結果が正確であったが、母親の発話には大きな変化がなく、バリエーションが少なかったことによると考えた (Sagae et al. 2010:274)。日本語の場合は文構造が複雑になることによって、曖昧さが増えることもある。形態素レベルで役割がマークされていない名詞句や外置を含む繰り返し (例: 「ワンちゃん [=ガ] ねそこにいたよ、今持ってたやつ [=ガ]」 Tai 2;6.10)、そして依存先が曖昧な副詞句 (例: 「もっと曲がったやつある?」 Tai 3;0.3) などを含む文が増える。この現象は親子ともに見られ、日本語の話し言葉の特徴でも言えよう。さらに、文が長くなると、スコープの問題も出て来る。例えば「エリちゃんとエリちゃんのママとエリちゃんのパパが行く」のような文で GRASP がそれぞれの「と」と「の」のスコープを正しく解釈できなく、名詞句を順番にまとめる傾向が見られた ([[[[エリちゃんとエリちゃんの]ママと]エリちゃんの]パパが])。この問題は英語でも指摘されている (Sagae et al. 2010: 724)。

4. 最後に

本研究では文法解析プログラム GRASP のための日本語文法役割を提案し、日本語のモデルコーパスで訓練した。訓練された日本語版の GRASP の信頼性を別のファイルをもとに確かめたところ、比較的低いエラー率 (2.84%) を得た。抽出される文構造の信頼性が高いと思われるが、盲信できるものではないので、自動解析のあとにチェックが必要である。特に SUBJ と OBJ の役割と、長い発話の文構造の確認が必要だと思われる。また、GRASP を別のコーパス、つまり違う話者の発話の解析に使う場合は、エラー率が上がると予測される (Sagae et al. 2010)。別のデータを解析する場合は、試しに数百発話を解析し、手動で直したものを GRASP の再訓練に使うことを勧める。

文献

- 長谷川信子 (1999) 『生成日本語学入門』大修館。
- Kroeger, P. (2005). *Analyzing Grammar: An Introduction*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Third Edition. Mahwah, N.J.: Lawrence Erlbaum Associates.
- MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. In H. Behrens (ed.), *Corpora in Language Acquisition Research: History, Methods, Perspectives*, Vol.6, 165-198. Amsterdam: Benjamins.
- 町田健・加藤重広 (2001) 『日本語学のしくみ』研究社。
- 増岡隆志・田窪行則 (1992) 『基礎日本語文法 — 改訂版』くろしお出版。
- 宮田 Susanne (2002) 『Wakachi2002 v.3.0 分かち書きガイドライン』 <<http://chilides.psy.cmu.edu/morgrams/Wakachi2002>>.
- 宮田 Susanne (2012a) 「日本語 MLU (平均発話長) のガイドライン: 自立語 MLU 及び形態素 MLU の計算法」『健康医療科学』2, 1-15. <<http://aska-r.aasa.ac.jp/dspace/bitstream/10638/5113/1/0039-002-201203-1-17.pdf>>.
- 宮田 Susanne (2012b) 『CHILDES 日本語版: 日本語用 CHILDES マニュアル 2012』

- <<http://www2.aasa.ac.jp/people/smiyata/CHILDESmanual/chapter01.html>>.
- 宮田 Susanne・森川尋美・村木恭子（編）（2004）『今日から使える発話データベース CHILDES 入門』ひつじ書房。
- Miyata, S. & Naka, N. (2010). *JMOR05.1: The Japanese Morphological Analysis Program Based on CLAN*. <<http://chilides.psy.cmu.edu/morgrams/Japanese/>>.
- 宮田 Susanne (2012a) 日本語 MLU (平均発話長) のガイドライン: 自立語 MLU および形態素 MLU の計算法 『健康医療科学』 2, 1-15. < <http://aska-r.aasa.ac.jp/dspace/bitstream/10638/5113/1/0039-002-201203-1-17.pdf> >
- Miyata, S. & Shirai, Y. (2008). Distributional vs. pragmatic effects in argument structure and case acquisition. Poster presented at *The 10th Annual International Conference of the Japanese Society for Language Sciences (JSLs2008)*, University of Shizuoka, July 12-13, 2008. Conference Handbook, 165.
- 中則夫・宮田 Susanne (1999) 「CLAN を用いた日本語形態素解析プログラム JMOR」 『日本教育心理学会フォーラムレポート No. FR-99-0001. 日本教育心理学会.仁田義雄 (1993) 「日本語の格を求めて」 仁田義雄 (編) 『日本語の格をめぐる』 くろしお出版 1-37.
- Payne, T.E. (1997). *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: Cambridge University Press.
- Rickmeyer, J. (1983). *Morphosyntax der japanischen Gegenwartssprache*. Heidelberg: J.Groos Verlag.
- Rispoli, M. (1995). Missing Arguments and the acquisition of predicative meanings. In: Tomasello, M. & Merriman, W. E. (eds.) *Beyond Names of Things*. Mahwah, N.J.: Lawrence Erlbaum Associates. 331-352.
- Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2007). High-accuracy Annotation and Parsing of CHILDES Transcripts. In *Proceedings of the ACL 2007 Workshop on Cognitive Aspects of Computational Language Acquisition*. Prague, Czech Republic. Association for Computational Linguistics. 25-32.
- Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language* 37, 705-729.
- Sagae, K., Lavie, A., & MacWhinney, B. (2005). Automatic measurement of syntactic development in child language. In: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, 197-204, Ann Arbor, Michigan, June. Association for Computational Linguistics.
- Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.
- 寺村秀夫 (1982-1991) 『日本語のシンタックスと意味 I-III』 くろしお出版。
- Teruya, K. (2007). *A Systemic Functional Grammar of Japanese*. London: Continuum.
- Trask, R. L. (1993). *A Dictionary of Grammatical Terms in Linguistics*. London: Routledge.
- Tsujimura, N. (1996). *An Introduction to Japanese Linguistics*. Cambridge Mass.: Blackwell Publ.