LTAG 文法からの変換による HPSG 英文法の作成

建石由佳, 鳥澤健太郎, 牧野貴樹, 西田健二, 渕上正睦, 辻井潤一
東京大学大学院理学系研究科
{yucca, torisawa, mak, nishiken, masatic, tsujii}@is.s.u-tokyo.ac.jp

概要

Feature-based LTAG による文法（XTAG 文法）からの変換によって HPSG 英文法を作成し, LTAG の木構造および素性構造と HPSG の素性の関係を明らかにした. 文法は素性構造記述言語 LiLFeS 上に実装した.

Conversion of LTAG English Grammar to HPSG

TATEISI Yuka, TORISAWA Kentaro, MAKINO Takaki,
NISHIDA Kenji, FUCHIGAMI Masachika, TSUJII Jun'ichi
Graduate School of Science, University of Tokyo

Abstracts

An HPSG grammar for English was developed by conversion from XTAG grammar, an English grammar in feature-based LTAG. Relating the concepts of the two grammar theories, we converted the grammar by mapping LTAG structures to HPSG feature structures. The compiled grammar was implemented on feature structure manipulation language LiLFeS.

1 はじめに

本稿では我々の研究室で開発中の HPSG (Head-Driven Phrase Structure Grammar) [1] 英文法について報告する. 我々は実用規模の HPSG 文法を短期間で開発することを目標として University of Pennsylvania で開発されている XTAG システム [2] で用いられる文法 (XTAG 英文法) を HPSG に変換することを試みた. XTAG 英文法は Feature-Based Lexicalized Tree Adjoining Grammar (FB-LTAG) で書かれており語彙約 317,000 を有し Wall Street Journal の 50 %について正解を含む構文木セットを生成できる (1995 年 8 月現在).

XTAG 英文法はすでに半ば完成された文法であり, FB-LTAG という文法の記述形式で書かれた文法のインスタンスであるととらえるべきものである. 一方, 一口に HPSG 文法といった時には, さまざまなとらえ方があるが各研究者が [1, 3] に与えられた schema・principle を基に種々の拡張を加えたものを HPSG と呼んでいるのが一般的であろう. 本研究でも, [1, 3] に与えられた文法を基に若干の拡張を加えたものを HPSG と呼ぶ.

FB-LTAG で記述できる文法全体のクラスと, HPSG という名称の指す対象を最大限に広くとったクラス, つまり, 型つき素性構造で記述できる文法全体のクラスの間での変換を考えた場合には, 問題はむしろ単純になる. というのも, 1) 型つき素性構造は再帰的に無限の深さの入れ子を含むことができるのでスタックをエミュ

レートすることができ，2) 任意の FB-LTAG は CFG の書き替え規則にスタックの操作の記述を付加した Linear Indexing Grammar に翻訳することができる [4] ことが知られているからである．すなわち FB-LTAG の任意の文法から単に型つき素性構造で記述された文法を得るためには，一旦 FB-LTAG を Linear Indexing Grammar に翻訳し，その Linear Indexing Grammar をエミュレートできるように型付き素性構造で文法を記述してやればよい． しかし，以上のような方法で FB-LTAG のインスタンスである XTAG の翻訳を実現したとしても，例えば，XTAG の素性にこめられた言語学的意味を HPSG の枠組で再解釈したということにはならない．おそらく，得られた文法を言語学的視点から解釈する，あるいは，得られた文法と [1] に与えられた文法と比較するといった作業はかなりの困難をともなうであろう．

本研究では，以上のような方法をとらず，XTAG の素性，木構造などを [1] に可能な限り忠実な文法に変換するということをゴールとする．これにより，XTAG に現れる個々の素性が HPSG の枠組の中でどのように表現しうるかといった具体的な問題を検討できるようになり，実際に実世界の言語を扱う文法を実現するという観点から，各文法，あるいはその形式の優劣が比較できるようになる．具体的には XTAG 文法の木構造，素性構造と HPSG の素性構造との対応づけを確立した． 文法記述間の変換は得られた対応づけに従って行なわれ，得られた文法は我々の研究室で開発中の素性構造記述言語 LiLFeS[5] 上に実現している．

2 文法の概説

2.1 FB-LTAG

Tree-Adjoining Grammar(TAG) は文脈自由文法 (CFG) の拡張の１つで，記号列の代わりに木構造を操作することによって文の解析生成を行う文法である． TAG では，elementary tree と呼ばれる木構造から，substitution と adjoining という 2 種類の操作を用いて木構造同士を結合させて解析結果を得る (図 1)． ある木構造の葉を他の木構造で置き換える操作を substitution という．この操作は CFG における書き換えとみなすことができる． 一方 adjoining は,root のラベルが L である木構造 T_a の root から葉に至る枝の列を，他の木 T の任意のラベル L を持つ任意のノードと同一視することによって結合する (直観的には T の中途に T_a が割り込む形となる)． このとき，割り込む枝の先端は，L をラベルに持ち，特別にマーク (図 1中 * 印) された葉 (foot ノード) でなければならない．Elementary tree のうち，foot ノードを持つものを auxiliary tree といい，持たないものを initial tree という．

図 1で は,T_β の rootVP から VP_* に至る枝 (点線部) が T_{α_0} のノード VP_1 に割り込んで下の木を生成する[1]． TAG のうち，すべての elementary tree の葉の１つ以上のラベルが終端記号 (単語) であるものを Lexicalized TAG (LTAG) といい，その終端記号を anchor という．さらに,Feature-Based LTAG (FB-LTAG) は，木の全てのノードに素性のセットを持つものを

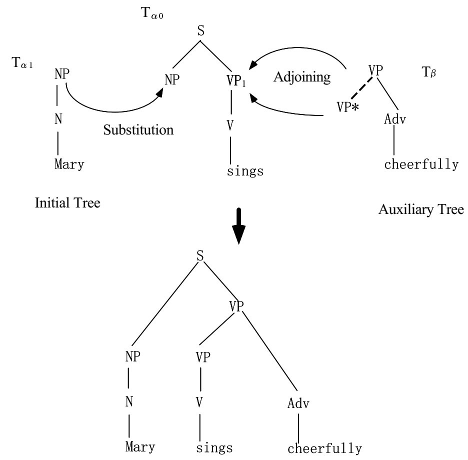

図 1: TAG における木の結合操作

[1] Adjoining は CFG の書き換え以上の能力を持った操作であり，例えば言語 $\{a^n b^n c^n\}$ を生成する文法は CFG では記述できないが TAG で記述することができる．

いう．XTAG 英文法ではノードに与えられる素性の値の差を除き 676 種類の elementary tree を持つ [2].

2.2 HPSG

HPSG は 1) 語彙項目 2) 書き換え規則（ID schema）3) 語順に対する制約規則（LP rule）4) 素性値の伝播に対する制約規則（principle）からなる．3) 以外はすべて素性構造で記述することができる．文法上の主な制約は語彙項目として記述し, また構文木の全てのノードは素性構造でその性質を表現する． 図 2に HPSG の語彙項目を示す．

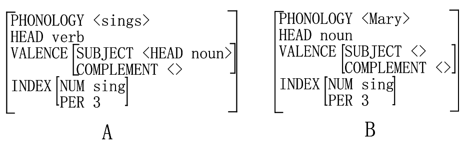

図 2: HPSG の素性構造

この素性構造は [1, 3] に挙げられる英語の素性構造の一部である．各素性の意味は付録 1 に挙げる． 図 2で A は `sings` が動詞の三人称単数形で主語として名詞をとること, B は `Mary` が名詞であり三人称単数であることを示す．

構文木は, 数種類の schema（ID schema と principle を単一化した素性構造）のうちの１つと子ノードの素性構造を単一化して親ノードを生成することによって１段ずつ生成する．すなわち, HPSG の schema は CFG の書き換え規則に対応し, 構文木の高さ１段に関わる制約を記述している[2]． また schema が適用される際素性値が構造共有を通じて子ノードから親ノードへと伝播する．文全体の解析は schema を繰り返し適用することによって行なう．

例えば, 図 3は, 動詞句と主語の結合に関わる schema(head-subject schema) を示す． 図中の四角で囲まれた数字は同じ番号で示される素性

[2]HPSG 文法は素性構造を複雑にすることによって CFG を越えた能力を持つことができる．

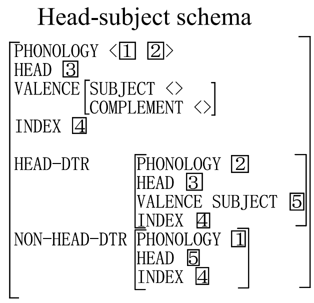

図 3: HPSG の schema

構造が構造共有 (論理型言語での変数の共有に相当) されることを示す．

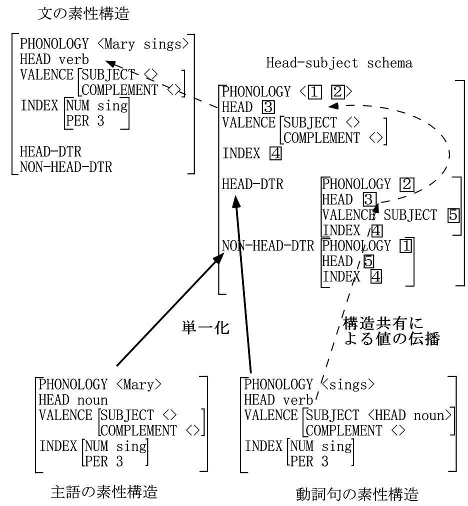

図 4: HPSG の schema を用いた解析

また, 図 4は, 図 3の schema を適用している様子を示している．

3 XTAG 文法の HPSG への変換

3.1 変換の方針

文法の同等性にはいくつかの定義があるが, ここでは「与えられた文を解析したとき, 元の FB-LTAG 文法で生成される構文木と等価な構文木を生成する」HPSG 文法を作成する．ただし, 両文法の思想の違いから，１対１に対応するがまっ

たく同形ではない木構造が少数存在する. このため, LTAG の木で語彙範疇 X から XP へ至る分岐のない枝の列を１つのノードに縮約したものと HPSG の木で head-complement schema の連続した適用によってできる多段の部分木を１段に縮約した木が同形であれば等価とする.

等価な構文木（に相当する素性構造）を生成するためには, 語彙項目に schema(木の分岐１段分に相当) を順次適用して

要請１ elementary tree と等価な素性構造が生成できる

要請２ LTAG 文法で許されていた結合 (substitution と adjoining) を模倣することができ, それ以外の結合はできない

要請３ FB-LTAG 文法の素性値の伝播が正しく模倣される

ようにできればよい.

我々は標準的な HPSG 文法 [1, 3] の文法規則である schema を保存し, また XTAG 文法の素性を保存した上で, 上の要請を満たすように XTAG 文法の elementary tree から語彙項目に当たる HPSG の素性構造を作成する. schema は [1] で使用される７つの schema（付録 1-1）を使用する. また素性構造は [1] から意味に関わる素性を除いたもの（付録 1-2）から出発し, XTAG 文法で使用される素性に対応する素性を追加する. 以下に述べる変換手順によって作成された HPSG 文法が XTAG 英文法と等価な構文木を生成することの証明はまだ終了していないが, 多段に渡る adjoining の適用を除けば等価な構文木が生成できることを確認している.

3.2 変換手順

以下では, 付録 1 で挙げた schema, および素性構造の骨格を前提とし, HPSG の語彙項目の素性構造を設定していく. 単に素性構造を決定するとすると自由度が高すぎるため, 問題を以下の手順で解く.

手順 1 XTAG の elementary tree と同じ形を持つ木構造が HPSG で生成できるように HPSG の語彙項目の素性構造を設定する.

手順 2 XTAG 側で substitution だけで組み上げられる構文木が HPSG 側で生成できるように素性構造を決定する.

手順 3 XTAG の elementary tree 内部での素性構造の伝搬が正しく HPSG 側で実現されるように素性構造を決定する.

手順 4 任意の initial tree に対して任意の auxiliary tree を一回 adjoin した時に, 木構造が保たれること, および素性構造の伝搬が XTAG と HPSG で同じになるように素性構造を決定する.

以下では HPSG での schema および素性の名称が言及されるが, これらについては付録１を参照されたい.

まず, 手順 1 の elementary tree の形の再現であるが, HPSG 側で適用されるべき schema の選択, および, XTAG の木構造中に現れるどのノードをその親の主辞とするかについて, 次のことを仮定する.

仮定１a XTAG の elementary tree の anchor から root に至る枝の列 (幹) 上にあるノードを主辞, その他の枝上にあるノードを非主辞とする.

仮定１b XTAG の elementary tree の各分岐は, HPSG 側では head-subject, head-complement, head-specifier のうち言語学的意味が妥当な schema の１つを適用することにより生成される.

名称からわかる通り, head-subject schema, head-complement schema, head-specifier schema はそれぞれ主辞が subject, complement, specifier をとる時に適用される schema であり, elementary tree の anchor を除いた葉を subject, complement, specifier のいずれかに分類すれば, 適用すべき schema が決定できる.

またこれらのschemaは，主辞のVALENCE素性の値にそれぞれ対応する素性を持ち，非主辞にそれらの値を単一化するという働きをする． したがって，elementary treeの各ノードで適用すべきschemaが決定できれば，VALENCE素性の値はelementary treeのanchorを除いた葉に対応する素性構造とすればよい．

また，仮定１aにより，elementary treeの幹はすべて主辞なので，語彙項目のHEAD素性の値は全てelementary treeのrootに相当するノードに伝搬される． 従ってelementary treeのanchorの文法的カテゴリーを示すラベルおよびanchorからrootに伝わる素性構造に対応するものをHPSG側でHEAD素性に付与する．

仮定1bに従って決定したschemaが意図した順序以外の順序で適用され，結果としてもとのelementary treeとは異なった形の木構造が生成されてしまう可能性はschemaの適用条件[1]により避けられる． 例えば，動詞を表すelementary treeではsubjectに至る枝はcomplementに至る枝よりも上位のノードから出ている．HPSG側でこの逆，つまり，complementに至る枝がsubjectに至る枝よりも上位にあるようなことがないことは，head-subject schemaがCOMPLEMENTS素性が空である場合にのみ適用されるという条件により避けられている．手順１をスケッチしたものが図5である．

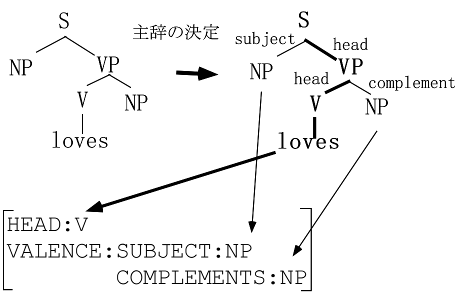

図 5: 木構造の変換

次に手順２であるが，手順１の時点でelementary treeのanchor以外の葉がVALENCE素性の値と単一化されるように設定した． したがって，XTAGでのsubstitutionがそのままVALENCE素性値の非主辞との単一化となるように他のelementary treeのrootに相当する素性構造を調整できれば，手順２は終了したことになる．これは前述のHEAD素性の調整などによって可能である．

手順３は，まず，すでに設定したHEAD素性，およびVALENCE素性以外の部分のHPSG素性の値の伝搬の仕方を各schemaごとに調べ（付録1-2），手順１および２が終了した段階でelementary tree内部でのXTAG素性値の伝搬の仕方とつき合わせることにより，XTAGの素性のHPSGの語彙項目中での位置を決定する．

手順４では adjoiningを受けたノードより上(rootに近い方)の構造が保たれ，XTAG文法の素性値の伝播がadjoiningによって変わる場合(つまりadjoinして来たauxiliary treeからその上の木構造へと素性値が伝播する場合)をも正しく模倣できなければならない． このために，まず仮定１abをfootノードのすぐ上の分岐以外に適用するように修正した上で手順１から手順３までを適用したのち，footノードについては，手順３まででHEAD素性とされた素性の値がanchorからrootへ伝わる場合と， footからrootへ伝わる場合に分け[3]，次のことを仮定する．

仮定２a HEAD素性がfootから伝わる場合：footが主辞となるようにhead-adjunctまたはhead-markerのschemaの一方を適用する． 両者のどちらを選択するかについては本来は意味素性の伝播に基づいて決定すべきであるが，XTAG英文法の素性は構文の制御のために使用され，意味に関わる素性がないので，anchorからrootに伝わる素性がある場合はhead-marker schema，そうでなければhead-adjunct schemaを適用する（図６）．

仮定２b HEAD素性がauxiliary treeのanchorから伝わる場合：initial treeの場合と

[3]なお，ここでHEAD素性とされた素性の値が両方から伝わる，またはanchorでもfootでもないノードから伝わる場合はXTAG文法には見られなかった

同様に元の手順 1 から手順 3 を適用する. ただし, foot ノードにあたる素性構造の VALENCE 素性の値も root の VALENCE 素性に伝播するように語彙項目を作成する（図 7）.

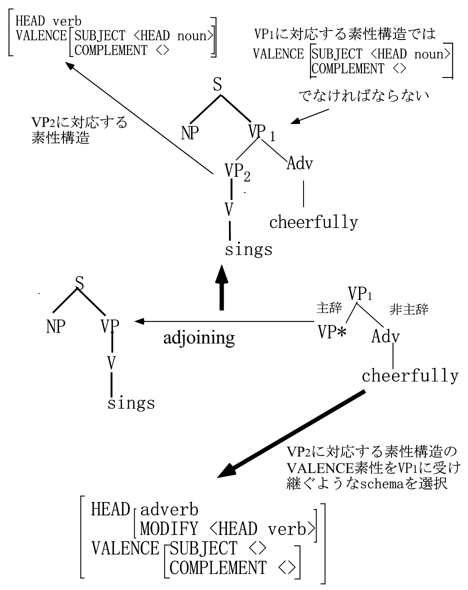

図 6: Auxiliary Tree の変換（仮定 2 a）

以下長距離依存に関わる SLASH 素性の作成について述べる. XTAG 英文法では, wh- 移動などの移動が起こった場合に対して, 移動の痕跡と移動先を含む elementary tree を持つ（図 8の $T1$）. 長距離依存は, この tree に他の tree（例えば図 8の T2）が adjoin することによって実現される.

一方 HPSG では, 痕跡を SLASH 素性の値として lexical rule で作成し, それを 1 段ずつ親へ伝播し, 移動先にあたる句の素性と SLASH 素性を head-filler schema により結合する. 前述の手順では head-filler schema を使用しないので痕跡（図 8中の ϵ）にあたる素性構造を SLASH 素性の値とすれば, $T1$, $T2$ に対応する HPSG 素性構造内部においても, また両者を adjoining により結合することに対応する schema において

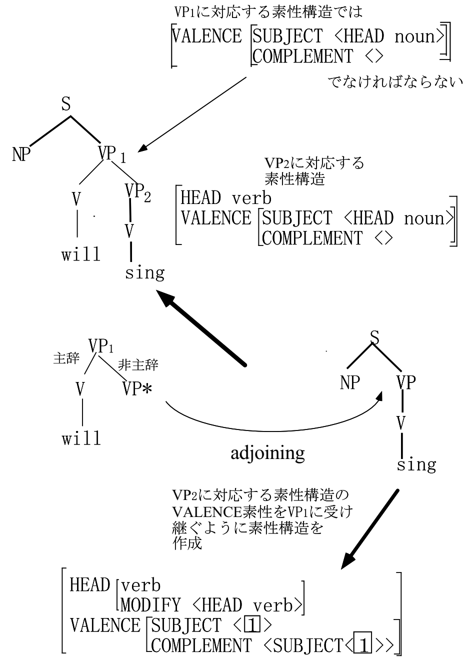

図 7: Auxiliary Tree の変換（仮定 2 b）

も root 側へ伝播され, 移動先の句を結合している分岐に head-filler schema を適用することができるので等価な構文木が生成できる.

4 実装

XTAG 英文法には, サブセットとして XTAG 文法の木構造の形 676 種類を網羅する約 600 語を選んだセットからなる文法がある. このサブセットを前節で述べた方法で HPSG に変換した文法を素性構造記述言語 LiLFeS 上に実装した.

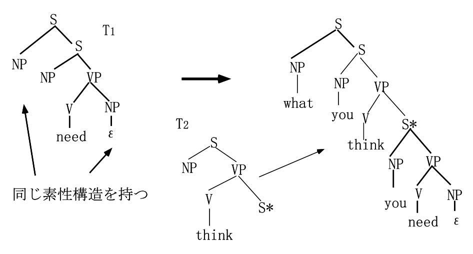

図 8: LTAG における長距離依存の扱い

図 9は LiLFeS 上で実装した文法に基づいて構文解析を行なった結果を表示したものである.

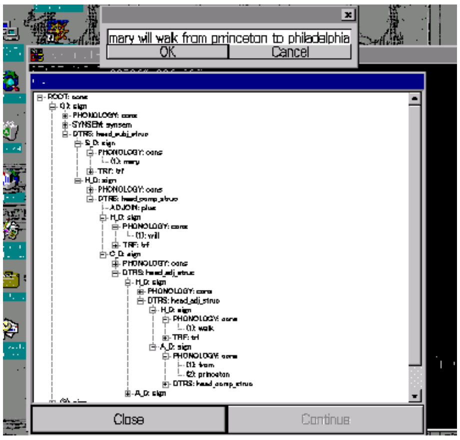

図 9: LiLFeS での入出力画面

現在 XTAG 文法のフルセットの変換を実装中であり, 名詞と動詞については完了している.

5 おわりに

実用規模の HPSG 英文法を作成するため, XTAG 英文法から HPSG へ変換する手法を開発した.

今後は XTAG 文法のフルセットの変換を完了し, また, 等価な構文木が生成されることの厳密な証明, LTAG の derivation tree に対応する素性構造の作成を行なう. 我々の行なった変換は, 英文法の FB-LTAG による表現 (XTAG 英文法) から HPSG による表現への変換となっている. これをさらに検討することにより, 自然言語を記述する枠組としての LTAG と HPSG の共通の制約・差異などを明らかにすることも行ないたい.

6 謝辞

XTAG 英文法を提供して下さり, また有益な助言を下さった Aravind Joshi 教授 (University of Pennsylvania) に感謝します.

参考文献

[1] C.Pollard and I.A.Sag, "Head-Driven Phrase Structure Grammar", University of Chicago Press, 1994.

[2] The XTAG Research Group, "A Lexicalized Tree Adjoining Grammar for English", IRCS Research Report 95-03, University of Pennsylvania, 1995.

[3] I.A.Sag and S.Rieheman, "Lecture Materials for Teaching HPSG", Stanford University , 1995 (available in `http://hpsg.stanford.edu/hpsg/lecture-materials.html`).

[4] Y.Schabes, "Stochastic Lexicalized Tree-Adjoining Grammars", Proc. Coling 92, pp. 425–432, 1992.

[5] T.Makino, K.Torisawa, and J.Tsujii, "LiLFeS—Practical Unification-Based Programming System for Typed Feature Structures", to appear in Proc. NLPRS 97, 1997.

[6] R.Kasper, B.Kiefer, K.Netter, K.Vijay-Shanker, "Compilation of HPSG to TAG", Proc. 33rd Annual Meeting of ACL, pp. 92–99, 1995.

付録１:HPSG の文法規則

1-1. ID schema

名称	略称	機能
head-subject schema	sb	主語と動詞句から文を作る
head-complement schema	cm	動詞句と主語以外の補語から動詞句を作る
head-subject-complement schema	sc	助動詞と主語補語から疑問文を作る
head-specifier schema	sp	名詞句と冠詞から名詞句を作る
head-adjunct schema	ad	修飾句と非修飾句から非修飾句と同じカテゴリーの句を作る
head-marker schema	mk	句に構文上のマーカーを付け加える
head-filler schema	fl	移動の痕跡を含む句と移動した先の句から，移動を含む句全体を作る

1-2. 素性構造とその伝播

素性名	schema ごとの素性値の伝播							意味
	sb	cm	sc	sp	ad	mk	fl	
PHONOLOGY	両者の接続 (concatenation)							表層の文字列
SYNSEM LOCAL CAT HEAD	H	H	H	H	H	H	H	主辞のカテゴリー
MODIFY	H	H	H	H	H	H	H	その語が修飾する語の主辞
VALENCE SUBJECT	U	H	U	H	H	H	H	主語
COMPLEMENTS	H	U	U	H	H	H	H	主語以外の補語
SPECIFIER	H	H	U	H	H	H	H	冠詞
MARKING	H	H	H	H	H	n	H	構文上のマーカー
CONTENT INDEX	A	H	A	A	n	H	H	人称・性・数
NONLOCAL INHER SLASH	B	B	B	B	B	B	A	移動の痕跡

H: 主辞から伝播 n: 非主辞から伝播 B: どちらの場合もある
A: 一致させる U: 主辞の対応する素性と非主辞が単一化

付録２：素性構造の対応

XTAG 素性	schema ごとの素性値の伝播						対応させた HPSG 素性
	sb	cm	sp	ad	mk	fl	
agr	A	H	A	H	H	H	SYNSEM:LOCAL:CONT:INDEX
assign-case	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
assign-comp	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
card	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:DET
case	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
comp	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:COMP
const	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:DET
decrease	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:DET
definite	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:DET
extracted	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
gen	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:DET
inv	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
mainv	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
mode	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
neg	H	H	H	n	n	H	SYNSEM:LOCAL:CAT:MARKING:ASP
passive	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
perfect	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
pred	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
progressive	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
pron	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
quan	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:DET
sub-conj	H	H	H	H	n	H	SYNSEM:LOCAL:CAT:MARKING:COMP
tense	H	H	H	H	H	H	SYNSEM:LOCAL:CAT:HEAD
trace	B	B	B	B	B	A	SYNSEM:NONLOCAL:INHER:SLASH
wh	H	H	H	n	n	H	SYNSEM:LOCAL:CAT:MARKING:ASP

H: 主辞から伝播 n: 非主辞から伝播 B: どちらの場合もある A: 一致させる

注）head-subject-complement schema については head-subject-schema に準ずる