# 型付き素性構造の差分演算

小 暮 潔[†]

本論文では，型付き素性構造の差分演算を提案し，形式的な定義を与える．型付き素性構造は，制約に基づく言語理論で用いられる素性構造の拡張された概念である．型付き素性構造の集合には，その表す対象の集合の包含関係を表す半順序関係が定義され，この半順序関係から2つの束演算，単一化（最大下界）と汎化（最小上界）が定義される．2つの型付き素性構造が与えられたときに，単一化が少なくとも一方が表す情報を集積し，集積結果の無矛盾性の判定などを行うのに対して，汎化は双方が表す共通情報を抽出する．本論文では，第3の演算として差分演算を定義する．この演算は，表す情報に包含関係が成り立つ2つの型付き素性構造が与えられたときに，情報の少ない方と単一化することにより情報の多い方を得ることができる型付き素性構造の中で情報が極小のものの1つを与える．本論文では，この差分演算を形式的に定義し，その存在を構成的に保証し，計算方法を示す．型付き素性構造の差分演算には，汎化演算との組合せによる型付き素性構造を用いた言語情報記述の階層化などの種々の応用がある．

# Typed Feature Structure Difference

KIYOSHI KOGURE[†]

A difference operation has been defined on the typed feature structures. The concept of typed feature structures augments the concept of feature structures. The set of typed feature structures has a partial ordering representing the subsumption relations between the set of objects they denote, and two lattice operations bassed on this partial ordering are defined: unification (greatest lower bound) and generalization (least upper bound). This paper proposes a new operation called difference. For any two typed feature structures $t_1$ and $t_2$ such that $t_1$ is less than or equal to $t_2$, the difference operation gives one of the maximal typed feature structures $t$ such that the unification of $t$ and $t_2$ yields $t_1$. This paper defines this operation formally, shows its properties, and gives an algorithm for calculating it. The difference operation in combination with the generalization operation has useful applications in the hierarchization of linguistic representations.

## 1. はじめに

制約に基づく言語理論[3),9),23),10),25)]では，対象の記述に論理的言語（素性論理）[22),14),20),26),6),11),4)]を用い，素性論理はモデルとして素性構造や型付き素性構造を用いる[☆]．有限論理式のモデルが有限に表現できるとき，論理式上での定理証明をモデルの演算で置換できることがあり，このようなとき，モデル演算に効率的アルゴリズムが存在すれば，定理証明よりモデル演算が有利なことが期待できる[25)]．素性構造の集合には，情報の包含関係に関する半順序関係と，これに基づく束演算―素性構造の持つ情報を集積する単一化（情報和[25)]）と，素性構造が共通に持つ情報を抽出する汎化（情報積）―が定義される．

素性構造に関する最も重要な演算は単一化である．素性構造を用いる言語処理の効率は単一化手法に大きく依存する．この効率向上のために多くの手法が提案されている[21),12),28),16),8),27)]．効率に関して，より深刻な問題は選言的情報である．これを簡単に取り扱う方法は，選言標準形[☆]に展開し，素性構造に対応する連言的素性記述の集合として取り扱う方法である．しかし，この方法は単一化で最悪の場合に元の選言数の指数時間を必要とする．この選言標準形への展開を遅延，回避する手法が提案されている[13),7),19)]．

素性構造を用いた言語処理を効率化するには，上述の選言的記述に関する効率的手法を有効に利用できる記述―階層的記述―を得る必要がある．しかし，階層的記述は，平坦な記述よりも作成がはるかに困難であ

† 日本電信電話株式会社基礎研究所
NTT Basic Research Laboratories

☆ 以下では，特にことわらないかぎり，素性構造で通常の素性構造と型付き素性構造の双方を表す．

☆ 選言標準形は，$\phi_1$ or $\phi_2$ or$\cdots$or $\phi_n$ の形式である．ここで，各 $\phi_i$ $(1\leq i\leq n)$ は選言を含まない．

る．記述の比較などを要求するからである．この困難を回避するには，平坦な記述から等価な階層的記述に変換する道具が必要である．この変換に必要なのは，複数の素性構造が共通に持つ情報を抽出する汎化演算と，素性構造の持つ情報の差を抽出する演算—差分—である．汎化演算は形式的に定義され，計算手法も提案されているが[18]，差分演算は定義すらない．

本論文では，型付き素性構造を用いた階層的記述を得るための基本演算としての差分演算の条件を分析する．この分析に基づき，差分（情報差）を定義し，その存在を構成的に保証し，計算方法を示す．そのために，$\psi$-型[1]を用い，形式的に取り扱う．

以下では，第2章で型付き素性構造について簡単に説明し，第3章で差分演算の必要性を具体的に示す．第4章で差分演算への要求を分析し，差分演算の非形式的定義を与える．第5章で形式的取り扱いを示す．

## 2. 型付き素性構造

型付き素性構造は簡潔な表現を可能にするための素性構造の拡張である．型付き素性構造は対象の種類を表す型記号と，対象の側面を表す素性—素性記号と値の対—の集合から構成され，この値も型付き素性構造である．型記号は包含関係の半順序関係 $\leq_{\mathcal{T}}$ による束の元である．例えば，人間，かつ，雄性であるならば，男であるという関係を，型記号 **human** と **male** の最大下界が **man** であるという形式で表せ，意味的選択制限などを取り扱うのに便利である．

型付き素性構造は**図1**のようなマトリクスで示される．この構造は，型記号 **syn** と後に続く“[”と“]”の間の素性を持つ．素性がない場合は，後の部分を省略する．埋め込まれた素性値の指定に，有限素性記号列である素性アドレスを用いる．素性アドレスを構成するための結合演算子を“・”で示す．例えば，図では，素性アドレス *agree・num* の値が型記号 **sg** を持つと言う．型記号 **agr** の前の X はタグ記号である．同一タグ記号の素性アドレス値は同一で，このようなとき，素性アドレスが共参照関係にあると言う．例えば，図では，*agree* と *subj・agree* が共参照関係にある．タグ記号の性質から同一タグ記号の2度目以降の出現では値を省略できる．以下では，型記号を **a**, **b**, … で，素性記号を $f, g, \cdots$ で，タグ記号を X, Y, … で表す．また，型付き素性構造を $t, t_1, t_2, \cdots$ などで示す．

型付き素性構造は**図2**のような有向グラフでも示される．有向グラフ表現では，型付き素性構造は型記号（とタグ記号の対）をラベルとして持つ節点で，素性は素性記号をラベルとして持つ弧で表される．共参照関係にある素性アドレス値は同一節点で表される．

型付き素性構造の集合は型記号集合の半順序関係を拡張した半順序関係を持つ．型付き素性構造 $t_1$ は矛盾（型記号 $\perp$）を含むか，あるいは，次の条件を満足するとき，かつ，そのときにかぎり，$t_2$ 以下であると言い，$t_1 \leq_t t_2$ と書く．すなわち，（1）$t_1$ の型記号が $t_2$ の型記号以下で，（2）$t_2$ に存在する各素性 $f$ が $t_1$ に存在し，$t_1$ の値 $t_{1,f}$ が $t_2$ の値 $t_{2,f}$ 以下で，かつ，（3）$t_2$ の各共参照関係が $t_1$ で成り立つ．関係 $\leq_t$ では，表す情報が多い型付き素性構造ほど小さくなる．関係 $\leq_t$ に基づき，$t_1$ と $t_2$ の最小上界 $t_1 \vee_t t_2$ と最大下界 $t_1 \wedge_t t_2$ を定義できる．最小上界は引数が共通に持つ情報を抽出する汎化で，最大下界は両方の引数が持つ情報を集積する単一化である．例を**図3**に示す．

## 3. 動　機

本章では，型付き素性構造の持つ情報の差を抽出する演算の言語処理での必要性について説明する．

**記述の階層化による効率化**

制約に基づく言語理論，特に語彙情報中心のものに基づき文法を構築する際，各語彙項目の独立な記述は，同様の記述の繰り返しから非効率である．実際は，テンプレートとそれとの差の形式で記述し，作業の効率化を図る．テンプレートは多くの場合，文法的原則などに従って定義されるが，既存の素性構造から

$$\mathbf{syn}\begin{bmatrix} \mathit{agree}: & \mathsf{X}:\mathbf{agr}\begin{bmatrix} \mathit{num}: & \mathbf{sg} \\ \mathit{per}: & \mathbf{3rd} \end{bmatrix} \\ \mathit{subj}: & \mathbf{syn}\begin{bmatrix} \mathit{agree}: \mathsf{X} \end{bmatrix} \end{bmatrix}$$

**図 1** 型付き素性構造のマトリクス記法の例
Fig. 1. Example of typed feature structure in matrix notation.

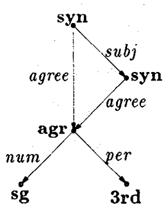


**図 2** 型付き素性構造の有向グラフ表現の例
Fig. 2 Example of directed graph representation of typed feature structure.

頻繁に共起する構造を抽出し，これをテンプレートとすることも考えられる．この際，複数の素性構造が共通に持つ情報を表す素性構造を計算する汎化と，元の素性構造と汎化の表す情報との差異を表す素性構造を計算する演算（差分）が必要になる．このような演算により，'$t_1$ or $t_2$' の形式を '$t_3$ and ($t_1'$ or $t_2'$)' の形式に変換できる．ここで，$t_3$ は共通情報を表す素性構造，$t_1'$, $t_2'$ は $t_3$ との組み合わせで $t_1$ と $t_2$ の情報を再現するのに必要最小限の情報を表す素性構造である．

選言的素性記述の効率的単一化手法の１つである Kasper の手法[13] はデータ構造 *feature description* を使用する．このデータ構造は確定部と不確定部から構成され，確定部は連言的情報を表し，不確定部は選言の集合で，各選言は *feature-description* 構造の集合である．確定部は素性構造と同一のデータ構造で表せ，単一化に素性構造のための手法を使用できる．Kasper の手法は確定部で情報を共有し，重複処理を回避するために，選言標準形を用いる手法より平均的にはるかに効率的である．

したがって，言語情報記述者は効率化のために確定部を最大，不確定部を最小にすることを要求される．例えば，多義語の各語義共通の情報（確定部に対応）と個々の語義固有の情報（不確定部の選言に対応）を分離して記述する必要がある．また，既に各語義の記述がある場合は，それらを１つの *feature-description* 構造で記述する必要がある．しかし，この作業は選言間の依存関係や共参照関係が絡み合うと人間には非常に困難になる．したがって，汎化と差分が必要になる．

$$\mathbf{syn}\left[ agree:\ \mathbf{agr}\left[ per:\ \mathbf{2nd} \right] \right]$$
$$\vee_t\ \mathbf{syn}\left[ \begin{array}{ll} agree: & \mathsf{X}:\mathbf{agr}\left[ \begin{array}{ll} num: & \mathbf{sg} \\ per: & \mathbf{3rd} \end{array} \right] \\ subj: & \mathbf{syn}\left[ agree:\ \mathsf{X} \right] \end{array} \right]$$
$$= \ \mathbf{syn}\left[ agree:\ \mathbf{agr}\left[ per:\ \mathbf{2nd\_or\_3rd} \right] \right]$$

$$\mathbf{syn}\left[ \begin{array}{ll} agree: & \mathsf{X}:\mathbf{agr} \\ subj: & \mathbf{syn}\left[ agree:\ \mathsf{X} \right] \end{array} \right]$$
$$\wedge_t\ \mathbf{syn}\left[ agree:\ \mathbf{agr}\left[ \begin{array}{ll} num: & \mathbf{sg} \\ per: & \mathbf{3rd} \end{array} \right] \right]$$
$$= \ \mathbf{syn}\left[ \begin{array}{ll} agree: & \mathsf{X}:\mathbf{agr}\left[ \begin{array}{ll} num: & \mathbf{sg} \\ per: & \mathbf{3rd} \end{array} \right] \\ subj: & \mathbf{syn}\left[ agree:\ \mathsf{X} \right] \end{array} \right]$$

**図 3**　型付き素性構造の汎化と単一化の例
Fig. 3　Examples of generalization and unification of typed feature structures.

る．

**制限素性記述との残差**

制約に基づく言語処理の効率化のために，曖昧性解消など処理の制御に貢献する素性とそれ以外の素性を分離し，２つの素性構造を用いる方法がある．例えば，日本語解析では，多くの場合，格助詞や活用語の形式，意味的選択制限を表す素性は解析候補の削減に貢献するが，意味構造を表す素性は貢献しない．そこで，候補削減に貢献する素性だけからなる素性構造とそれ以外の素性からなる素性構造に分割する．解析を行う際，最初に前者の素性構造だけを取り扱うことで入力表現全体の解析候補を得て，次にこの解析候補を構築した句構造に従って後者の素性構造を取り扱うことで必要な意味構造などを得る．このようにすると，不要な計算を回避できる．ここで，元の素性構造と処理制御用の素性構造が与えられたときに，残りの情報を表す素性構造を得るために差分が必要になる．

また，類似の応用に次のようなものもある．Shieber[24] は素性構造を用いたチャート解析[15]で予測手続きでの弧の無限生成を回避するために，予測弧の素性構造にフィルタを適用し，フィルタを通過した素性構造が同じ予測弧を同一視する方法を提案している．Shieber の場合，完成手続きで元の完全な素性構造の単一化を行う．しかし，差分によりフィルタ不通過の情報を表す素性構造を用意し，この素性構造だけを追加的に単一化することで効率化が期待できる．

以上で述べたように，情報の包含関係にある素性構造間の差異の計算には有用な応用がある．

## 4. 差分演算への要求条件

本章では，前章で有用性を示した差分を定義する上での問題点，要求を整理し，差分を規定する．

**暫定的非形式的定義（1）**

型付き素性構造 $t_a$, $t_b$ で，$t_b$ の表す情報が $t_a$ の表す情報に包含される（$t_a \leq_t t_b$）ものが与えられたとき，求めたい両者の情報の差異を表す型付き素性構造 $t$ は，$t$ と $t_b$ から $t_a$ を復元できる（$t \wedge_t t_b = t_a$）ものである．そして，そのようなものの中で表す情報が最も少ない（$\leq_t$ の意味で最大の）ものである．したがって，差分演算は次のように定義できそうである．

**暫定的非形式的定義 1**　無矛盾な型付き素性構造 $t_a$, $t_b$ で，$t_a \leq_t t_b$ であるものが与えられたとき，$t_a$ と $t_b$ の差分 $t$ は，$t \wedge_t t_b = t_a$ を満足する型付き素性構造の中で，関係 $\leq_t$ で最大の $t$ である．

しかし，上の定義は適格ではない．上の条件を満足する $t_a$, $t_b$ が与えられたとき，$t \wedge_t t_b = t_a$ を満足する（1）最大の型付き素性構造が存在せず，複数の極大のものが存在する場合や，（2）極大の型付き素性構造すら存在しない場合があるからである．

場合（2）は，循環構造を含む型付き素性構造が素性アドレスの無限集合を持つことによる．例えば，素性アドレス $f^i$ $(i=1,2,\cdots)$ を持ち☆，型記号がすべて ⊤ である構造 $t_l$, $t_m$ を考える．ここで，自然数 $l$, $m$ が存在し，$l$ が $m$ の整数倍で，任意の非負整数 $i$, $j$ に関して，$t_l$ では $f^i$ と $f^{i+lj}$ の間だけで，$t_m$ では $f^i$ と $f^{i+mj}$ の間だけで共参照関係が成り立つとする．すなわち，$t_m \leq_t t_l$ である．$t_l$, $t_m$ と同じ素性アドレス，型記号を持つもので，$m=\gcd(n,l)$ である $n$，任意の非負整数 $i$, $j$ に関して，$f^i$ と $f^{i+nj}$ の間だけで共参照関係が成り立つものを $t_n$ とする．このような $t_n$ は $t_n \wedge_t t_l = t_m$ である．$t_n$ は $n$ が整数倍の関係で大きくなるほど $\leq_t$ で大きくなり，$t_n \wedge_t t_l = t_m$ を満足する極大の $t_n$ は存在しない．例えば，**図4**では，6 の倍数以外の偶数 $n$ に関して，$t_n \wedge_t t_6 = t_2$ で，実際，$t_4 \wedge_t t_6 = t_2$, $t_8 \wedge_t t_6 = t_2$ である．

この問題は，循環構造を含む型付き素性構造では素性アドレスの集合と共参照関係が無限集合になり，これらの集合の包含関係に関する無限鎖が存在することから生じる．循環構造を含まない無矛盾なものに制限すると，その型記号，タグ記号，素性アドレスの集合の有限性から，無限鎖は存在しない．そこで，循環構造を含む型付き素性構造を対象外とする．

前述の場合（1）は，（a）型記号集合に関する問題と

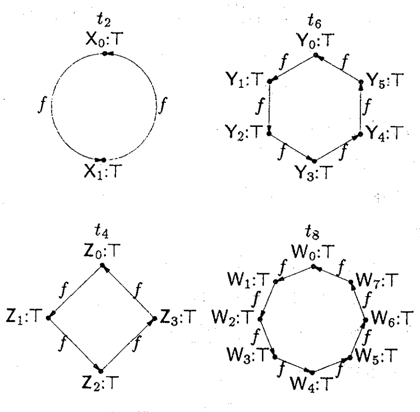


**図 4**　循環構造の取り扱いに関する問題
Fig. 4　Problem on cyclic structures.

☆ $f^i$ は長さ $i$ の素性記号 $f$ からなる列を示す．

（b）素性アドレス間の共参照関係に関する問題による．問題（a）は，一般の型記号集合の任意の要素 **a**, **b** で，$\mathbf{a} \leq_{\mathcal{T}} \mathbf{b}$ であるものに関して，$\mathbf{c} \wedge_{\mathcal{T}} \mathbf{b} = \mathbf{a}$ となる最大の型記号 **c** が存在しない問題である．また，問題（b）は，（i）単一化結果が共参照関係として，入力の共参照関係の単なる集合和ではなく，これを同値関係で，右不変であるように拡張したものを持つ問題と☆，（ii）共参照関係にある複数の素性アドレスが共有する部分構造の分割の仕方が複数ある問題である．

問題（a）は，型記号集合の Brouwer 束☆☆への拡張で回避できる☆☆☆．例えば，有限 Bool 束の表現定理[5]に基づき，型記号集合の原子元☆☆☆☆の巾集合を構成し，Bool 束に拡張できる[17]．すなわち，元の型記号をそれ以下の原子元の集合で表す．この Bool 束での半順序関係は集合の包含関係で，最小上界と最大下界は集合和と集合積で，補元は補集合である．**b** の補元を **b**′ で示すと，$\mathbf{c} \wedge_{\mathcal{T}} \mathbf{b} = \mathbf{a}$ となる最大の **c** は $\mathbf{b}' \vee_{\mathcal{T}} \mathbf{a}$ である．

問題（b）を例で説明する．問題（i）であるが，**図5**では，$t_1 \leq_t t_2$ で，$t_2$ との単一化が $t_1$ になるものに $t_3$ と $t_4$ がある．また，$t_3 \leq_t t$，あるいは，$t_4 \leq_t t$ で，$t_2$ と

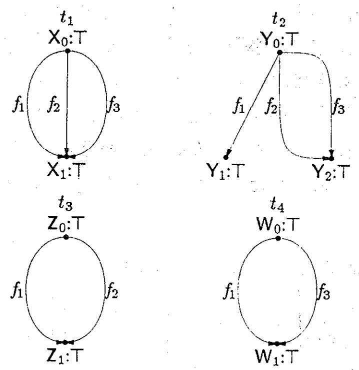


**図 5**　共参照関係の取り扱いに関する問題（1）
Fig. 5　Problem on coreference relations (1).

☆ 共参照関係が右不変であるとは，素性アドレス $p_1$ と $p_2$ の間に共参照関係が成り立つとき，素性アドレス $p_1 \cdot q$ がこの構造に含まれるならば，$p_2 \cdot q$ との間で共参照関係が成り立つことを言う．

☆☆ 束 $\langle L;\ \vee,\ \wedge,\ \leq \rangle$ で，任意の $a, b \in L$ に関して，$b \wedge x \leq a$ である最大の $x$ が存在するとき，$L$ を Brouwer 束と呼ぶ[2]．例えば，Bool 束は Brouwer 束である．この Brouwer 束は分配束である．

☆☆☆ $a$, $b$ が Brower 束の元で，$a \leq b$ であるとき，$x \wedge b \leq a$ を満足する最大の $x$ は，$a \wedge b = a$ であるから，$a \leq x$ であり，$x \wedge b = a$ となる．ゆえに，束が Brouwer 束であれば，差分が定義できる．

☆☆☆☆ 自分と ⊥ 以外に自分以下の元を持たない元．

の単一化が $t_1$ になる $t$ が存在しない．ゆえに，$t_3$ と $t_4$ は極大であり，$t\wedge_t t_2=t_1$ である極大型付き素性構造 $t$ が複数存在する．

また，問題(ⅱ)であるが，例えば，**図 6** では，$t_1\leq_t t_2$ である．ここで，$t_3$ は $t_\beta\wedge_t t_\gamma=t_\alpha$ であるかぎり，$t_3\wedge_t t_2=t_1$ で，$t_\beta$ と $t_\gamma$ が冗長でなければ，極大になる．すなわち，極大型付き素性構造が複数存在する．

**暫定的非形式的定義(2)**

上記の問題を回避するために，型付き素性構造の差分は次のように定義できるかもしれない．

**暫定的非形式的定義 2**　循環構造を含まない無矛盾な型付き素性構造 $t_a$, $t_b$ で，$t_a\leq_t t_b$ であるものが与えられたとき，$t_a$ と $t_b$ の差分 $T$ は，$t\wedge_t t_b=t_a$ を満足する $\leq_t$ で極大な型付き素性構造 $t$ の集合である．

循環構造を含まない無矛盾な型付き素性構造の場合，極大なものが存在するから，定義は適格である．

しかし，上の定義には問題(b)に関する実際的問題がある．極大型付き素性構造の数の組合せ的増大である．問題(ⅰ)に関しては，例えば，図5は2つの素性アドレス対からの1つの対の選択であるが，選択候補の対の数と選択する対の数が増えると，極大なものの数は組合せ的に増大する．また，問題(ⅱ)に関しても，例えば，図6で $t_\alpha$ の構造が大きくなるほど，分割の仕方の数が増える．したがって，極大型付き素性構造すべてを求めるのが実際的でなくなる．

また，階層的記述を得る目的には，1つの極大型付き素性構造だけで十分である．なぜならば，それだけで元の型付き素性構造を復元できるからである．

**非形式的定義**

暫定的非形式的定義2の $T$ の1つを得るとすると，

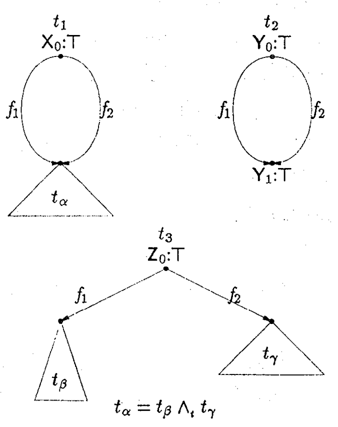


**図 6**　共参照関係の取り扱いに関する問題(2)
Fig. 6 Problem on coreference relations (2).

どのようなものを得るかが問題となる．主目的の言語処理の効率化を考えると，(1)最小の構造で表現でき，(2)単一化が効率的であるものが望まれる．

上の条件を考慮すると，特定の素性値に構造を可能なかぎり集中させ，部分構造（有向グラフでの節点と弧）を少なくすることが望まれる．例えば，図6で，$t_\beta=t_\alpha$，$t_\gamma=\top$ の（あるいは，その逆の）場合と，それ以外の場合を考えると，部分構造は前者の方が少ない．

また，3個以上の型付き素性構造を考えると，別の理由から特定の素性への構造の集中が望まれる．例えば，型付き素性構造 $t_1$, $t_2$, $t_3$ が与えられ，これらの汎化を $t_4$ とする．このとき，$t_1$ と $t_4$，$t_2$ と $t_4$ の差を表す構造 $t_1'$ と $t_2'$ の間には共通情報が存在する可能性があり，$t_1'$ と $t_2'$ の汎化を求め，階層を深めることが考えられる．この汎化に多くの情報を持たせるには，$t_1'$ と $t_2'$ が可能なかぎり多くの共通素性を持つ必要がある．以上の要請から，可能なかぎり特定の素性に構造が集中しているものを差分とする．

さらに，差分型付き素性構造の計算コストも考慮する必要がある．この側面を考慮すると，結果の部分構造を最小化すると同時に，結果の構造を一意に選択するために必要な計算コストが任意のどれか1つを選択するのと比較して相対的に小さいことが望まれる．この後者の要求を満足するために，次に述べるように素性記号の優先順序の辞書式順序に基づく単純な選択を用いて差分を定義する．

以上の要求を満足するために，暫定的非形式的定義2の $T$ から1つを選択するための優先順序関係を定義する．具体的には，(1)素性記号集合に全順序関係 $\leq_{\mathcal{F}}$ を与え，(2)関係 $\leq_{\mathcal{F}}$ に基づく辞書式順序 $\leq_{\mathcal{F}^*}$ を素性アドレス集合に与え，(3)暫定的非形式的定義2の $T$ 中の型付き素性構造に対して，全順序関係 $\leq_{\mathcal{F}^*}$ で小さい素性アドレスに $\leq_t$ で小さい部分構造を持つものが小さくなる半順序関係 $\preceq$ を定義する．この関係 $\preceq$ で最小の $T$ の元を差分と定義する．

**非形式的定義 1**　循環構造を含まない無矛盾な型付き素性構造 $t_a$, $t_b$ で，$t_a\leq_t t_b$ であるものが与えられたとき，$t_a$ と $t_b$ の差分 $t_b\rightarrow_t t_a$ は，$t\wedge_t t_b=t_a$ を満足する $\leq_t$ で極大の型付き素性構造 $t$ の集合の中で，$\preceq$ で最小の型付き素性構造である．

このような差分が存在することは，次節で証明する．差分の例を**図 7** に示す．ここで，素性記号間の順序関係 $agree\leq_{\mathcal{F}} subj$ が成り立つとする．

## 5. 差分演算の形式的定義

本章では，差分演算を形式的に定義し，差分結果の構造，および，差分の計算方法を示す．そのために，$\psi$-型[1]を用いる．$\psi$-型は型付き素性構造の表す情報を表せ，型付き素性構造と同様の束を構成するので，型付き素性構造の形式化となる．以下では，準備として $\psi$-型の束演算を定義した後，差分演算を定義する．

### 5.1 $\psi$-型

$\psi$-型は統語構造として $\psi$-項と呼ばれる項構造を持つ．これらを定義するために以下の語彙を使用する．

1. 型記号集合 $\mathcal{T}$．$\top$ と $\bot$ を含む．
2. $\mathcal{T}$ 上の包含通係の半順序関係 $\leq_{\mathcal{T}}$．$\mathcal{T}$ はBrouwer束である．$\top$ が最大，$\bot$ が最小である．任意の $\mathbf{a}$，$\mathbf{b}\in\mathcal{T}$ に関して，最小上界 $\mathbf{a}\vee_{\mathcal{T}}\mathbf{b}$，最大下界 $\mathbf{a}\wedge_{\mathcal{T}}\mathbf{b}$，および，$\mathbf{b}\rightarrow_{\mathcal{T}}\mathbf{a}$ が存在する．ここで，$\mathbf{b}\rightarrow_{\mathcal{T}}\mathbf{a}$ は $\mathbf{b}\wedge_{\mathcal{T}}\mathbf{c}\leq_{\mathcal{T}}\mathbf{a}$ となる最大の型記号 $\mathbf{c}$ である．
3. 素性記号集合 $\mathcal{F}$．
4. $\mathcal{F}$ 上の全順序関係 $\leq_{\mathcal{F}}$．
5. タグ記号集合 $\mathcal{V}$．

これらは次の意味を持つ．型記号は $\psi$-型で表そうとする対象の集合を示す．$\top$ は無情報を表し，この対象の集合全体 $\mathbf{U}$ を示し，$\bot$ は矛盾を表し，空集合を示す．他の型記号は $\mathbf{U}$ の部分集合を示す☆．関係 $\leq_{\mathcal{T}}$ はこれら集合間の包含関係を示す．素性記号は対象に素性記号が表す属性の値を割り当てる関数を示すが，すべての対象にこの属性が定義されるとはかぎらないため☆☆，この関数は $\mathbf{U}$ から $\mathbf{U}$ への部分関数である．関係 $\leq_{\mathcal{F}}$ は素性記号間の優先関係を示す．

$$
\mathbf{syn}\begin{bmatrix} \mathit{agree}: & \mathsf{X}:\mathbf{agr} \\ \mathit{subj}: & \mathbf{syn}\begin{bmatrix}\mathit{agree}: \mathsf{X}\end{bmatrix} \end{bmatrix}
$$
$$
\rightarrow_t \mathbf{syn}\begin{bmatrix} \mathit{agree}: & \mathsf{Y}:\mathbf{agr}\begin{bmatrix}\mathit{num}: & \mathbf{3g}\\ \mathit{per}: & \mathbf{3rd}\end{bmatrix} \\ \mathit{subj}: & \mathbf{syn}\begin{bmatrix}\mathit{agree}: \mathsf{Y}\end{bmatrix} \end{bmatrix}
$$
$$
= \top\begin{bmatrix}\mathit{agree}: \top\begin{bmatrix}\mathit{num}: & \mathbf{sg}\\ \mathit{per}: & \mathbf{3rd}\end{bmatrix}\end{bmatrix}
$$

**図 7** 型付き素性構造の差分の例
Fig. 7 Example of types feature structure difference.

☆ 例えば，型記号 **sign** が言語表現の集合を示すというようにである．

☆☆ 例えば，素性記号 *phon* が示す，言語表現に音韻属性値を割り当てる関数は，$\mathbf{U}$ の部分集合の言語表現の集合上で定義される．

これらの語彙で項を定義する．最初に素性記号を用い，項の骨格である項領域を定義する．

**定義 1** 素性記号集合 $\mathcal{F}$ 上の項領域 $\Delta$ は次の条件を満足する素性記号の有限列の集合で，空列 $\varepsilon$ を含む．

1. 接頭語で閉じている．すなわち，任意の $p\in\Delta$ に関して，$p=p_1\cdot p_2$ となる任意の $p_1$ に関して，$p_1\in\Delta$ である．
2. 有限分岐である．すなわち，任意の $p\in\Delta$ に関して，$\{f\in\mathcal{F}\,|\,p\cdot f\in\Delta\}$ が有限集合である．

項領域の元を素性アドレスと呼ぶ．

項は，項領域中の各素性アドレスに型記号とタグ記号を割り当てたもので，次で定義される．

**定義 2** 項は組 $\langle\Delta,\tau,\nu\rangle$ である．ここで，$\Delta$ は $\mathcal{F}$ 上の項領域，$\tau$ は $\mathcal{F}^*$ から $\mathcal{T}$ への全関数である型記号関数，$\nu$ は $\Delta$ から $\mathcal{V}$ への全関数であるタグ記号関数である．$\tau$ は $p\notin\Delta$ に関しては $\tau(p)=\top$ である．

**定義 3** 項 $t=\langle\Delta,\tau,\nu\rangle$ と素性アドレス $p\in\Delta$ に関して，$t$ の $p$ での部分項は項 $t/p:=\langle\Delta/p,\tau/p,\nu/p\rangle$ である．ここで，$\Delta/p,\tau/p:\mathcal{F}^*\rightarrow\mathcal{T}$，$\nu/p:\Delta/p\rightarrow\mathcal{V}$ は以下のとおりである．

$$\Delta/p:=\{q\,|\,p\cdot q\in\Delta\},$$
$$\tau/p(q):=\tau(p\cdot q),$$
$$\nu/p(q):=\nu(p\cdot q).$$

**定義 4** 項 $t=\langle\Delta,\tau,\nu\rangle$ は，部分項の集合 $\{t/p\,|\,p\in\Delta\}$ が有限であるときに正規である．

以下では，正規項だけを考える．

項 $t=\langle\Delta,\tau,\nu\rangle$ で同一タグ記号を持つ2つの素性アドレスは共参照すると言う．項の素性アドレスの共参照関係 $\kappa$ は項領域 $\Delta$ の上のタグ記号関数の核 $\kappa:=\mathbf{ker}(\nu)=\{\langle p,q\rangle\,|\,\nu(p)=\nu(q)\}$ である．関係 $\kappa$ は同値関係で，これが作る同値類を共参照同値類と呼ぶ．次に，共参照関係に関する無矛盾性を定義する．

**定義 5** 項は，各共参照同値類で，その中に含まれるすべての素性アドレスでの部分項が同一であるとき，参照無矛盾である．

参照無矛盾の項 $t=\langle\Delta,\tau,\nu\rangle$ は，上の定義により，右不変―すなわち，任意の $p_1$，$p_2\in\Delta$ に関して，$\nu(p_1)=\nu(p_2)$ であるとき，すべての $p_1\cdot q\in\Delta$ に関して，$t/(p_1\cdot q)=t/(p_2\cdot q)$ である．

参照無矛盾の項 $t=\langle\Delta,\tau,\nu\rangle$ は，$p,q\in\Delta$ が存在し，$p$ が $q$ の真の接頭語であり，$\nu(p)=\nu(q)$ であるとき，循環構造を含むと言う．

**定義 6** 整合項は参照無矛盾の正規項である．整合項

の集合を $\mathcal{WFT}$ で，循環構造を含まない整合項の集合を $\mathcal{WFT}_0$ で示す．

整合項は循環構造を含むとき項領域が無限集合となるが，含まれる型記号とタグ記号の集合は有限である．循環構造を含まない整合項は項領域も有限である．

整合項 $\langle\Delta,\tau,\nu\rangle$ は型付き素性構造と同様の有向グラフで表される．節点は素性アドレスの共参照同値類を表す．この節点はラベルとしてタグ記号と型記号の対を持つ．素性アドレス $p$ の同値類を表す節点には，それを始点とし，$p\cdot f\in\Delta$ の各素性記号 $f$ に関して，これをラベルとして持つ弧が1つ存在する．この弧の終点は素性アドレス $p\cdot f$ を含む共参照値類を表す．

タグ記号関数の役割は部分構造の同一性を示すことだけである．整合項の集合 $\mathcal{WFT}$ 中には項領域，型記号関数，共参照関係が同一で，タグ記号関数だけが異なる項が存在する．これらの項は同じ対象を示す．このような項を同一視する同値関係を定義する．

**定義 7** 任意の $t_1=\langle\Delta_1,\tau_1,\nu_1\rangle$, $t_2=\langle\Delta_2,\tau_2,\nu_2\rangle\in\mathcal{WFT}$ は，

$$\Delta_1=\Delta_2,$$
$$\mathbf{ker}(\nu_1)=\mathbf{ker}(\nu_2),\ \text{かつ},$$
$$\tau_1=\tau_2$$

であるとき，$t_1\alpha t_2$ である．

既に述べたように，型記号 $\perp$ は矛盾を示すので，$\perp$ を含む項*も矛盾と解釈する．このような矛盾を表す項を同一視するために，次の同値関係を定義する．

**定義 8** 任意の $t_1,t_2\in\mathcal{WFT}$ は，ともに $\perp$ を含むか，あるいは，$t_1=t_2$ であるとき，$t_1\Downarrow t_2$ である．

項が $\perp$ を含むとき，その項の $\alpha$-同値類中の他の項も $\perp$ を含む．ゆえに，定義8から，関係 $\simeq:=\alpha\cup\Downarrow$ も同値関係となる．この同値関係で $\psi$-型を定義する．

**定義 9** $\psi$-型 $[t]$ は，商集合 $\Psi:=\mathcal{WFT}/\simeq$ の元である．また，$\Psi_0:=\mathcal{WFT}_0/\simeq$ とする．$\psi$-型 $[t]$ の項構造 $t$ を $\psi$-項と言う．

型記号集合 $\mathcal{T}$ の包含関係に関する半順序関係 $\leq_{\mathcal{T}}$ を $\psi$-型の集合上に次のように拡張できる．

**定義 10** 任意の $t_1=\langle\Delta_1,\tau_1,\nu_1\rangle$, $t_2=\langle\Delta_2,\tau_2,\nu_2\rangle\in\mathcal{WFT}$ は，$t_1$ が $\perp$ を含むか，あるいは，

$$\Delta_2\subseteq\Delta_1, \tag{1a}$$
$$\mathbf{ker}(\nu_2)\subseteq\mathbf{ker}\nu_{(1)}, \tag{1b}$$

かつ，$\forall p\in\mathcal{F}^*$ に関して，

$$\tau_1(p)\leq_{\mathcal{T}}\tau_2(p) \tag{1c}$$

であるとき，かつ，そのときにかぎり，$t_1$ が $t_2$ に包含されると言い，$t_1\leq_t t_2$ と書く．$\Psi$ 上では，$t_1\leq_t t_2$ であるとき，$[t_1]\leq_t[t_2]$ である．

このような包含関係に関する束を構成できる．

**定理 1** $\mathcal{T}$ が束ならば，$\Psi$ も束である．

証明は Aït-Kaci[1] にあり，$t_1=\langle\Delta_1,\tau_1,\nu_1\rangle$ と $t_2=\langle\Delta_2,\tau_2,\nu_2\rangle$ を整合項とするとき，次の $t_3=\langle\Delta_3,\tau_3,\nu_3\rangle$ と $t_4=\langle\Delta_4,\tau_4,\nu_4\rangle$ が最小上界 $t_1\vee_t t_2$ と最大下界 $t_1\wedge_t t_2$ となることが示されている．

$$\Delta_3=\Delta_1\cap\Delta_2, \tag{2a}$$

$\nu_3:\Delta_3\to\mathcal{V}$ は，

$$\mathbf{ker}(\nu_3)=\mathbf{ker}(\nu_1)\cap\mathbf{ker}(\nu_2), \tag{2b}$$

かつ，$\forall p\in\mathcal{F}^*$ に関して，

$$\tau_3(p)=\tau_1(p)\vee_{\mathcal{T}}\tau_2(p). \tag{2c}$$

$$\Delta_4=C_{\mathrm{Dom-RI}}(\Delta_1\cup\Delta_2,\kappa^{[*]}), \tag{3a}$$

$\nu_4:\Delta_4\to\mathcal{V}$ は，

$$\kappa_4=\mathbf{ker}(\nu_4)$$
$$=\{\langle p,q\rangle\in\kappa^{[*]}\mid p,q\in\Delta_4\}, \tag{3b}$$

かつ，$\forall p\in\mathcal{F}^*$ に関して，

$$\tau_4(p)=\bigwedge\nolimits_{\mathcal{T}}\{\tau_i(q)\mid\langle p,q\rangle\in\kappa_4,\ i=1,2\}. \tag{3c}$$

ここで，$C_{\mathrm{Dom-RI}}(\Delta_1\cup\Delta_2,\kappa^{[*]})$ は $\Delta_1\cup\Delta_2$ の $\kappa^{[*]}$ での右不変閉包で☆，$\kappa^{[*]}$ は

$$\kappa^{[*]}=C_{\mathrm{Coref-RI}}(C_{\mathrm{Coref-Tr}}(C_{\mathrm{Coref-Refl}}(\kappa_1\cup\kappa_2,\ \Delta_1\cup\Delta_2)),\mathcal{F}^*)$$

である☆☆．さらに，ここで，

$$C_{\mathrm{Dom-RI}}(\Delta,\kappa)=\bigcup_{m\geq 0}C'_{\mathrm{Dom-RI}}(\Delta,\kappa,m)$$

$$C'_{\mathrm{Dom-RI}}(\Delta,\kappa,m)=\begin{cases}\Delta, & m=0\ \text{のとき},\\ C'_{\mathrm{Dom-RI}}(\Delta,\kappa,m-1)\cup\\ \{q\mid\langle p,q\rangle\in\kappa,p\in C'_{\mathrm{Dom-RI}}(\Delta,\kappa,m-1)\}, & m\leq 1\ \text{のとき},\end{cases}$$

$$C_{\mathrm{Coref-Refl}}(\kappa,\Delta)=\kappa\cup\{(\langle p,p\rangle\mid p\in\Delta\},$$

$$C_{\mathrm{Coref-Tr}}(\kappa)=\bigcup_{m\geq 0}C'_{\mathrm{Coref-Tr}}(\kappa,m)$$

---

* 項 $t=\langle\Delta,\tau,\nu\rangle$ は，$\tau(p)=\perp$ である $p\in\Delta$ が存在するとき，$\perp$ を含むと言う．

☆ $C_{\mathrm{Dom-RI}}$ は項領域 (Term Domain) の右不変 (Right-Invariant) 閉包演算子である．

☆☆ $C_{\mathrm{Coref-RI}}$, $C_{\mathrm{Coref-Tr}}$, $C_{\mathrm{Coref-Refl}}$ はそれぞれ共参照 (Coreference) 関係の右不変閉包演算子，推移 (Transitive) 閉包演算子，反射 (Reflective) 閉包演算子である．

$$C'_{\mathrm{Coref-Tr}}(\kappa, m) = \begin{cases} \kappa, & m=0 \text{ のとき}, \\ C'_{\mathrm{Coref-Tr}}(\kappa, m-1) \cup \\ \{\langle p, q\rangle \mid \langle p, r\rangle, \langle r, q\rangle \in C'_{\mathrm{Coref-Tr}}(\kappa, m-1)\}, \\ \quad m \geq 1 \text{ のとき}, \end{cases}$$

$$C_{\mathrm{Coref-RI}}(\kappa, \Delta) = \bigcup_{m\geq 0} C'_{\mathrm{Coref-Tr}}(\kappa, \Delta, m)$$

$$C'_{\mathrm{Coref-Tr}}(\kappa, \Delta, m) = \begin{cases} \kappa, & m=0 \text{ のとき}, \\ \{\langle p_1 \cdot q, p_2 \cdot q\rangle \in \Delta^2 \mid \\ \quad \langle p_1, p_2\rangle \in C'_{\mathrm{Coref-Tr}}(\kappa, \Delta, m-1)\}, \\ \quad m \geq 1 \text{ のとき} \end{cases}$$

である．このような項 $t_3$ と $t_4$ は最小上界と最大下界の条件を満足する．$\psi$-型に関しては，$[t_1] \vee_\iota [t_2] = [t_1 \vee_\iota t_2]$，$[t_1] \wedge_\iota [t_2] = [t_1 \wedge_\iota t_2]$ とする．

### 5.2 $\psi$-型の差分演算

次に，$\psi$-型に基づき，型付き素性構造（すなわち，$\psi$-型）の差分演算を定義し，差分結果の構成方法を示す．

最初に，包含関係 $\leq_\iota$ で比較不能な項の間に優先順序を与える．そのために，素性記号集合上の優先順序 $\leq_{\mathcal{F}}$ に基づく順序関係をいくつか定義する．まず，関係 $\leq_{\mathcal{F}}$ を素性記号の有限列の集合上に拡張する．

**定義 11** 関係 $\leq_{\mathcal{F}^*}$ は，$\mathcal{F}$ の上で定義される全順序関係 $\leq_{\mathcal{F}}$ に基づく，$\mathcal{F}^*$ の上の辞書式順序である．

この関係 $\leq_{\mathcal{F}^*}$ は $\mathcal{F}^*$ の上の全順序関係である．

この優先通係 $\leq_{\mathcal{F}^*}$ から素性記号列の対の間の優先関係と素性記号列の対の集合の間の優先関係を定義する．これらは，素性記号列の対の集合である共参照関係の間の優先関係を取り扱う際に用いる．

**定義 12** 関係 $\leq_{\mathcal{F}^* \times \mathcal{F}^*}$ は，$\mathcal{F}^*$ の上の優先関係 $\leq_{\mathcal{F}^*}$ を基にした $\mathcal{F}^* \times \mathcal{F}^*$ の上の辞書式順序である．

**定義 13** 集合 $S_1$ と $S_2$ を素性記号列の対のの有限集合とする．このとき，それぞれの要素を $\leq_{\mathcal{F}^* \times \mathcal{F}^*}$ の昇順で並べた列を $S'_1$ と $S'_2$ とするとき，列 $S'_1$ が関係 $\leq_{\mathcal{F}^* \times \mathcal{F}^*}$ の辞書式順序に関して $S'_2$ 以下であるとき，かつ，そのときにかぎり，$S_1 \leq_{2^{\mathcal{F}^* \times \mathcal{F}^*}} S_2$ である．

次に，循環構造を含まない整合項の間の優先関係を定義する．このような項の項領域は有限であるので，次のように定義される．

**定義 14** 項 $t_1 = \langle \Delta_1, \tau_1, \nu_1\rangle$ と $t_2 = \langle \Delta_2, \tau_2, \nu_2\rangle$ を無矛盾で，循環構造を含まない整合項とする．このとき，$p_0, p_1, \cdots, p_n$ を $\Delta_1 \cup \Delta_2$ 中の素性アドレスを $\leq_{\mathcal{F}^*}$ に関する昇順で並べた列とし，$i=1, 2$ と任意の $0 \leq j \leq n$ に関して，

$$S_{i,j} = \begin{cases} \{q \mid \langle p_j, q\rangle \in \kappa_i\}, & p_j \in \Delta_i \text{ のとき}, \\ \emptyset, & \text{その他のとき} \end{cases}$$

とする．型記号と素性アドレス集合の対の間の半順序関係 $\preceq'$ を，任意の型記号 $\mathbf{a}_1$, $\mathbf{a}_2$, 素性アドレス集合 $S_1$, $S_2$ に関して，$\mathbf{a}_1 \leq_{\mathcal{T}} \mathbf{a}_2$, かつ，$S_2 \subseteq S_1$ であるとき，かつ，そのときにかぎり，$\langle \mathbf{a}_1, S_1\rangle \preceq' \langle \mathbf{a}_2, S_2\rangle$ であると定義する．このとき，列 $\langle \tau_1(p_0), S_{1,0}\rangle$, $\langle \tau_1(p_1), S_{1,1}\rangle, \cdots, \langle \tau_1(p_n), S_{1,n}\rangle$ が関係 $\preceq'$ の辞書式順序に関して列 $\langle \tau_2(p_0), S_{2,0}\rangle$, $\langle \tau_2(p_1), S_{2,1}\rangle, \cdots, \langle \tau_2(p_n), S_{2,n}\rangle$ 以下であるとき，かつ，そのときにかぎり，$t_1 \preceq t_2$ である．$\psi$-型では，$[t_1]$, $[t_2] \in \Psi_0$ に関して，$t_1 \preceq t_2$ であるとき，$[t_1] \preceq [t_2]$ である．

以上の準備を用いて，$\psi$-型の差分を定義する．

**定義 15** $\psi$-型 $[t_1], [t_2] \in \Psi_0$ で，$[t_1]$ の $\psi$-項 $t_1$ が $\perp$ を含まず，かつ，$[t_1] \leq_\iota [t_2]$ であるものが与えられたとき，$[t_1]$ と $[t_2]$ の差分 $[t_2] \rightarrow_\iota [t_1]$ は，$[t] \wedge_\iota [t_2] = [t_1]$ を満足する $\leq_\iota$ で極大の $\psi$-型 $[t]$ の中で，$\preceq$ で最小のものである．

この差分 $\psi$-型の存在を次の定理で構成的に示す．

**定理 2** $\mathcal{T}$ が Brouwer 束であるとき，$\psi$-型 $[t_1]$, $[t_2] \in \Psi_0$ で，$[t_1]$ の $\psi$-項 $t_1$ が $\perp$ を含まず，かつ，$[t_1] \leq_\iota [t_2]$ であるものが与えられたとき，差分 $[t_2] \rightarrow_\iota [t_1]$ が存在する．

**証明．** 条件を満足する $\psi$-型の構造を示す項 $t_1 = \langle \Delta_1, \tau_1, \nu_1\rangle$ と $t_2 = \langle \Delta_2, \tau_2, \nu_2\rangle$ が与えられたときに，定義 15 の条件を満足する $\psi$-型の構造を示す項 $t_5 = \langle \Delta_5, \tau_5, \nu_5\rangle$ を構成できることを示す．

**共参照関係．** 差分 $\psi$-項の構成要素として，最初に，共参照関係を取り扱う．２つの項の最大下界の共参照関係は，(3b)が示すように，それぞれの共参照関係の集合和ではなく，反射律，推移律が成り立ち，右不変となるように拡張したものである．ゆえに，$t_5$ の共参照関係 $\kappa_5 = \mathbf{ker}(\nu_5)$ と $t_2$ の $\kappa_2 = \mathbf{ker}(\nu_2)$ の集合和のこのような拡張が $t_1$ の $\kappa_1 = \mathbf{ker}(\nu_1)$ でなければならない．また，$\kappa_5$ は，このような共参照関係の中で集合の包含関係で極小でなければならない．さらに，優先関係 $\preceq$ で最小になるために，このような任意の $\kappa$ に関して，$\kappa_5 \leq_{2^{\mathcal{F}^* \times \mathcal{F}^*}} \kappa$ でなければならない．

上の条件を満足する共参照関係 $\kappa_5$ を構成するために，暫定的共参照関係を $\{\langle \varepsilon, \varepsilon\rangle\}$ に設定し，$\kappa_1$ の構成に必要な素性アドレス対を優先順に追加する．集合差 $\kappa_1 - \kappa_2$ の要素を優先関係 $\leq_{\mathcal{F}^* \times \mathcal{F}^*}$ の昇順で並べたものを $\langle p_1, q_1\rangle, \cdots, \langle p_n, q_n\rangle$ とする．このとき，必要

な素性アドレス対は次に定義される $\kappa^{\langle n\rangle}$ で得られる．$\kappa^{\langle i\rangle}$ $(0\leq i\leq n)$ は，$i=0$ のとき，$\kappa^{\langle 0\rangle}=\{\langle\varepsilon,\varepsilon\rangle\}$ である．$i\geq 1$ のとき，$\kappa^{\langle i\rangle}$ は，$\langle p_i,q_i\rangle$ が必要ならば，この対を $\kappa^{\langle i-1\rangle}$ に追加し，この結果が共参照関係の条件を満足するためにさらに必要な対を追加したもので，不必要ならば，$\kappa^{\langle i\rangle}=\kappa^{\langle i-1\rangle}$ である．$\langle p_i,q_i\rangle$ が必要なのは，$\langle p_i,q_i\rangle$ が $\kappa^{\langle i-1\rangle}$ と $\kappa_2$ の集合和の反射，推移，右不変閉包に含まれないとき

$$\langle p_i,q_i\rangle\notin C_{\text{Coref-RI}}(C_{\text{Coref-Tr}}(C_{\text{Coref-Refl}}(\kappa^{\langle i-1\rangle}\cup\kappa_2,\Delta_1)),\Delta_1) \tag{4}$$

である．したがって，$\kappa^{\langle i\rangle}$ を次のように拡張する．

$$\kappa^{\langle i\rangle}=\begin{cases}\kappa^{\langle i-1\rangle}\cup\{\langle p_i,q_i\rangle\}\cup\{\langle q_i,p_i\rangle\}\\ \quad\cup\{\langle r,r\rangle\mid p_i=r\cdot r',\ \text{あるいは},\ q_i=r\cdot r'\},\\ \qquad\text{(4)を満足するとき},\\ \kappa^{\langle i-1\rangle},\ \text{その他のとき}.\end{cases}$$

すなわち，必要なときには対 $\langle p_i,q_i\rangle$ を追加し，対称律が成り立つように $\langle q_i,p_i\rangle$ を追加する．そして，対の中に含まれる素性アドレスの接頭語に関して反射律が成り立つように接頭語の対を追加する．

$\kappa^{\langle n\rangle}$ を共参照関係にするために，推移閉包を取り，既知の $\Delta_1$ での右不変閉包を取ったものを $\kappa$ とする．

$$\kappa=C_{\text{Coref-RI}}(C_{\text{Coref-Tr}}(\kappa^{\langle n\rangle}),\Delta_1). \tag{5}$$

$\kappa$ は，$\Delta_1$ 上で右不変になるように拡張したため，不要なものを含んでいる．不要なものを取り除くために，$\kappa$ を用いて計算される $\Delta_5$ 上に制限する．すなわち，

$$\kappa_5=\{\langle p,q\rangle\in\kappa\mid p,q\in\Delta_5\}. \tag{6}$$

この $\kappa_5$ は，構成方法からわかるように，必要な素性アドレス対だけを含む最小限のものであり，かつ，最優先の対から構成されている．また，この計算過程は，$\kappa_1$ の有限性から，各ステップの計算が有限であり，ステップの繰り返し回数も有限である．

**型記号関数**．型記号関数 $\tau_5:\mathcal{F}^*\to\mathcal{T}$ は，$t_1$ の項領域 $\Delta_1$ 中の素性アドレス $p$ に関して，その $\kappa_1$-類中の素性アドレスの $t_2$ の型記号と $t_5$ の型記号の最大下界が $\tau_1(p)$ と等しくなければならない．すなわち，

$$\textstyle\bigwedge_{\mathcal{T}}(\{\tau_2(q)\mid\langle p,q\rangle\in\kappa_1\}\cup\{\tau_5(q)\mid\langle p,q\rangle\in\kappa_1\})=\tau_1(p)$$

でなければならない．また，過剰な情報を含んではならないから，次式を満足しなければならない．

$$\textstyle\bigwedge_{\mathcal{T}}\{\tau_5(q)\mid\langle p,q\rangle\in\kappa_1\}=(\bigwedge_{\mathcal{T}}\{\tau_2(q)\mid\langle p,q\rangle\in\kappa_1\})\to_{\mathcal{T}}\tau_1(p)$$

さらに，この型記号関数を持つ項が $\preceq$ で優先されなければならない．以上の条件は次の $\tau_5$ で満足する．

$$\tau_5(p)=\begin{cases}(\bigwedge_{\mathcal{T}}\{\tau_2(q)\mid\langle q,p\rangle\in\kappa_1\})\to_{\mathcal{T}}\tau_1(p),\\ \quad p\text{がその }\kappa_1\text{-類中で }\leq_{\mathcal{F}^*}\text{で最小の要}\\ \quad\text{素 }r\text{ と }\langle p,r\rangle\in\kappa\text{ の関係にあるとき},\\ \top,\ \text{その他のとき}.\end{cases} \tag{7}$$

**項領域**．最後に，項領域を取り扱う．差分の項領域は，$t_2$ との最大下界を取ることにより $t_1$ の項領域を構成するのに必要な素性アドレスを含まなければならない．したがって，$t_1$ の項領域 $\Delta_1$ と比較して $t_2$ の $\Delta_2$ に不足しているものを含まなければならない．ただし，このような $p\in\Delta_1-\Delta_2$ であっても，$t_1$ の共参照関係 $\kappa_1$ から，$\langle p,q\rangle\in\kappa_1$ である $q\in\Delta_2$ が存在するものや，$\langle p,q\rangle\in\kappa_1, q\leq_{\mathcal{F}^*}p$ である別の $q\in\Delta_1$ が存在するものは不要である☆．すなわち，次を含まなければならない．

$$\begin{aligned}\Delta=\{p\in\Delta_1-\Delta_2\mid\ &\langle p,q\rangle\in\kappa_1,\ \text{かつ},\ q\in\Delta_2,\ \text{あるいは},\\ &\langle p,q\rangle\in\kappa_1,p\neq q,\ \text{かつ},\ q\leq_{\mathcal{F}^*}p\ \text{である }q\text{ が}\\ &\text{存在しない}\}\end{aligned}$$

また，型記号が $\top$ 以外である素性アドレスと，自分以外と共参照関係にある素性アドレスも含まなければならない．ただし，素性アドレス $p_1$ と $p_2$ が共参照関係にあれば，必然的に $p_1\cdot q$ と $p_2\cdot q$ $(q\neq\varepsilon)$ も共参照関係になるから，このような $p_1\cdot q, p_2\cdot q$ は必要ない．ゆえに，次の $\Delta'$ を含まなければならない．

$$\begin{aligned}\Delta'=\Delta\cup\{p\mid\tau_5(p)\neq\top\}\cup\ &\{p\mid p\neq q,\ \langle p,q\rangle\in\kappa\ \text{である }q\text{ が存在し，か}\\ &\text{つ，}r\neq\varepsilon,\ \langle p',q'\rangle\in\kappa,\ p=p'\cdot r,\ q=q'\cdot r\ \text{で}\\ &\text{ある }r,\ p',\ q'\text{ が存在しない}\}\end{aligned}$$

さらに，項領域の条件（定義1）を満足しなければならない．すなわち，空列 $\varepsilon$ を含み，接頭語で閉じ，有限分岐でなければならない．ここで，$\Delta_1$ が項領域で，有限分岐であるから，その部分集合の $\Delta'$ も有限分岐である．ゆえに，次の $\Delta''$ を含まなければならない．

$$\Delta''=\{\varepsilon\}\cup\{p\in\mathcal{F}^*\mid p\cdot q\in\Delta'\}$$

また，項が参照無矛盾でなければならない．これは，次の $\Delta_5$ のように右不変閉包を取ることで満足される．

$$\Delta_5=C_{\text{Dom-RI}}(\Delta'',\kappa) \tag{8}$$

以上をまとめると，式(8)で定義された項領域 $\Delta_5$,

☆ これは最大下界を取る際に，項を右不変にするために，このような素性アドレスが必然的に追加されるからである．

式(7)で定義された型記号関数 $\tau_5$ と，式(6)で定義された共参照関係 $\kappa_5$ を核として持つタグ記号関数 $\nu_5$ から構成される項 $\langle \Delta_5, \tau_5, \nu_5 \rangle$ の型は，確かに定義 15 の条件を満足する差分 $[t_2] \rightarrow_t [t_1]$ である． □

## 6. おわりに

本論文では，制約に基づく言語理論で中心的役割を担う型付き素性構造に関して，これを用いた言語記述を階層化するために差分演算を定義した．この差分演算は，表す情報の間に包含関係がある２つの型付き素性構造が与えられたとき，それらの情報の差を表す型付き素性構造を得る演算である．このような型付き素性構造は複数存在する可能性があるが，本論文ではその１つだけを選択するように差分演算を定義した．これは応用面からの要請で，共参照関係が多く，共参照関係にある素性アドレスの値が複雑になると，差を表す型付き素性構造の数が急増するからである．

この型付き素性構造の差分の性質を明確にするために，$\phi$-型を用いて，差分演算を定義した．また，その存在を構成的に保証し，計算方法を示した．

通常の素性構造は型付き素性構造の特殊な場合であり，本論文の定義は通常の素性構造にも有効である．通常の素性構造は原子素性構造か複合素性構造であるが，原子素性構造は原子元である型記号を持ち，素性を持たない型付き素性構造に対応し，複合素性構造はすべての素性を許す原子元の型記号を持つ型付き素性構造に対応する．このような型記号の集合は，$\top$，$\perp$，原子元の集合 $\mathcal{A}$ からなる束を構成するが，これは Brouwer 束ではない しかし，差分の考慮の対象となる包含関係にある対は，$\top$ と $\mathcal{A}$ の元の対と，同一元の対である．ここで，任意の $\mathbf{a} \in \mathcal{A}$ に関して，$\top \wedge_{\mathcal{T}} \mathbf{a} = \mathbf{a}$ であるから，$\mathbf{a}$ と $\top$ の差分を $\mathbf{a}$, $\mathbf{a}$ と $\mathbf{a}$ の差分を $\top$ とすることで十分である．このようにすれば，本論文の差分を通常の素性構造に応用できる．

型付き素性構造の差分演算には種々の応用がある．既に示したように，(1)記述の階層化，(2)情報共有による言語処理の効率化，(3)情報分割による言語処理の効率化，特に処理戦略の検討の際などに有用である．

本論文の差分演算は，単一化や汎化などと同様に増進的グラフ複製のアイデアを用いて計算できる．

**謝辞** 本論文を書くにあたり貴重なコメントをいただいたNTT基礎研究所対話理解研究グループの島津明氏，片桐恭弘氏，内藤昭三氏，川森雅仁氏，堂坂浩二氏，中野幹生氏に感謝の意を表する．

## 参考文献

1) Aït-Kaci, H.: An Algebraic Semantics Approach to the Effective Resolution of Type Equations, *J. of Theoretical Computer Science*, Vol. 45, pp. 293-351 (1986).
2) Birkhoff, G.: *Lattice Theory*, American Mathematical Society, 3rd edition (1967).
3) Bresnan, J. ed.: *The Mental Representation of Grammatical Relations*, The MIT Press (1982).
4) Carpenter, B.: *The Logic of Typed Feature Structures*, Cambridge University Press (1992).
5) Davey, B. A. and Priestley, H.: *Introduction to Lattices and Order*, Cambridge University Press (1990).
6) Dawar, A. and Vijay-Shanker, K.: A Three-Valued Interpretation of Negation in Feature Structure Descriptions, *Proc. 27th ACL*, pp. 18-24 (1989).
7) Dörre, J. and Eisele, A.: Feature Logic with Disjunctive Unification, *Proc. 13th COLING*, Vol. 2, pp. 100-105 (1990).
8) Emele, M.: Unification with Lazy Non-Redundant Copying, *Proc. 29th ACL*, pp. 325-330 (1991).
9) Gazdar, G., Klein, E., Pullum, G. and Sag, I.: *Generalized Phrase Structure Grammar*, Basil Blackwell (1985).
10) Gunji, T.: *Japanese Phrase Structure Grammar*, Reidel (1987).
11) Johnson, M.: *Attribute-Value Logic and the Theory of Grammar*, CSLI Lecture Notes No. 16, CSLI (1989).
12) Karttunen, L.: D-PATR—A Development Environment for Unification-Based Grammars, CSLI-86-61, CSLI (1986).
13) Kasper, R. T.: A Unification Method for Disjunctive Feature Descriptions, *Proc. 25th ACL*, pp. 235-242 (1987).
14) Kasper, R. T. and Rounds, W. C.: A Logical Semantics for Feature Structures, *Proc. 24th ACL*, pp. 257-266 (1986).
15) Kay, M.: Algorithmic Schemata and Data Structures in Syntactic Processing, CSL-80-12, XEROX Palo Alto Research Center (1980).
16) Kogure, K.: Strategic Lazy Incremental Copy Graph Unification, *Proc. 13th COLING*, Vol. 2, pp. 223-228 (1990).
17) Kogure, K.: Treatment of Negative Descriptions of Typed Feature Structures, *Proc. 14th COLING*, pp. 380-386 (1992).
18) 小暮　潔：増進的複製による型付き素性構造汎

化手法, 情報処理学会論文誌, Vol. 34, No. 9, pp. 1919-1930 (1993).

19) Maxwell, J. T., III and Kaplan, R. M.: A Method for Disjunctive Constraint Satisfaction, Tomita, M. ed., *Current Issues in Parsing Technologies*, pp. 173-190, Kluwer Academic Publishers (1991).

20) Moshier, M. D. and Rounds, W. C.: A Logic for Partially Specified Data Structures, *Proc. 14th ACM POPL*, pp. 156-167 (1987).

21) Pereira, F. C. N.: Stucture Sharing Representation for Unification-Based Formalisms, *Proc. 23rd ACL*, pp. 137-144 (1985).

22) Pereira, F. C. N. and Shieber, S. M.: The Semantics of Grammar Formalisms Seen as Computer Languages, *Proc. 10th COLING*, pp. 123-129 (1984).

23) Pollard, C. and Sag, I.: *An Information-Based Syntax and Semantics—Volume 1: Fundamentals*, CSLI Lecture Notes Number 13, CSLI (1987).

24) Shieber, S. M.: Using Restriction to Extend Parsing Algorithms for Complex-Feature-Based Formalisms, *Proc. 23rd ACL*, pp. 145-152 (1985).

25) Shieber, S. M.: Constraint-Based Grammar Formalisms—Parsing and Type Inference for Natural and Computer Languages, PhD thesis, Stanford University (1989).

26) Smolka, G.: A Feature Logic with Subsorts, LILOG Report 33, IBM Deutschland (1988).

27) Tomabechi, H.: Quasi-Destructive Graph Unification with Structure-Sharing, *Proc. 14th COLING*, pp. 440-446 (1992).

28) Wroblewski, D. A.: Nondestructive Graph Unification, *Proc. 6th AAAI*, pp. 582-587 (1987).

(平成5年10月22日受付)
(平成6年10月13日採録)

**小暮　潔**（正会員）

1957 年生．1979 年慶應義塾大学工学部電気工学科卒業．1981 年同大学院修士課程修了．同年日本電信電話公式武蔵野電気通信研究所入所．1986 年〜90 年 ATR 自動翻訳電話研究所出向．現在 NTT 基礎研究所情報科学研究部対話理解研究グループに所属．主幹研究員．自然言語処理の研究に従事．人工知能学会，電子情報通信学会，日本音響学会，日本認知科学会，言語処理学会，ACL 各会員．