知識利用型文書画像理解システムに関する研究

1991年4月

黄　瀬　浩　一

序文

本論文は，筆者が大阪大学大学院工学研究科通信工学専攻，並びに大阪府立大学工学部電気工学教室において行った，知識利用型文書画像理解システムに関する研究を６章に分けてまとめたものである．

第１章は緒論であり，本研究の目的，位置付けと工学上の意義について述べる．

第２章では，文書画像理解システムの一般性および処理能力を高める一手法として，知識利用型のシステム構成を提案し，その設計思想について論じる．文書画像理解とは，最上位レベルの文書から最下位レベルの文字までの論理的な構成要素を文書画像から抽出することにより，文書の内容に関する構造化記述を生成する処理である．このためには，構成要素の大きさや配置などに関する知識 (レイアウト構造に関する知識)，および構成要素の記述内容の妥当性に関する知識 (論理的制約に関する知識) の活用が有効であると考えられる．ところが，これらの知識は対象文書に依存するため，システム内部に埋没させると一般性が損なわれるという問題が生じる．そこで，提案システムでは，知識を文書モデルと呼ぶ知識ベースに分離して蓄積し，処理部から利用するという構成をとることにより，この問題の解決を試みる．処理部では，画像特徴に基づいて構成要素を抽出する文書構造解析，記述内容に基づいて構成要素を抽出する構造化記述生成の２処理を実行する．文書構造解析は，記述内容の妥当性を考慮しない処理であることから，文書画像理解の枠組みにおいては仮説生成法として位置づけている．文書構造解析で生成された仮説は，構造化記述生成において記述内容の妥当性を検証して取捨選択される．本システムでは，以上のような仮説生成検証プロセスの導入により処理誤りに対して柔軟に対処し，システムのロバスト性を高めている．

第３章では，提案システムの知識ベース部である文書モデルについて詳述する．第２章で提案したシステムでは，対象文書に応じて知識を追加，変更するため，知識記述における表現能力，記述容易性，可読性がシステム全体の一般性に大きな影響を及ぼすと考えられる．文書モデルでは，従来手法の

ように数値的，平板的な記述法を用いず，知識を記号的，階層的に記述することにより，上記の要件を満たす知識ベースの作成を試みる．文書は構成要素の集合であることから，本手法では，構成要素を単位として知識を記述する．提案記述法の特徴は，画像の絶対的な物理量に依存しない知識記述を実現するためにレイアウト構造を構成要素間で相対的に規定すること，レイアウト述語を用いてレイアウト構造を記号的に記述すること，構成要素の記述内容に関する論理的制約を単語の接続性，単語列の整合性により表現することにある．さらに本章では，文書画像理解のために提案された既存の知識記述法との比較検討を通して，本手法の有効性を検証する．

第4章では，レイアウト構造に関する知識を活用する文書構造解析法について述べる．文書構造解析は，文書モデルに記述された構成要素を画像から抽出する処理である．構成要素は，構成要素名を属性として持つ矩形領域ととらえることができるため，実際には，画像から矩形領域を抽出して属性を付与することになる．本手法の特徴は，構成要素の階層構造に基づいて構成要素を抽出することにより処理効率を向上させること，レイアウト構造に関する知識を用いてトップダウン的に構成要素を抽出すること，レイアウト構造のバリエーションから構成要素の抽出に複数の可能性が残る場合には個々を仮説として生成することにある．加えて，構成要素の抽出履歴を基に仮説間の依存関係を記録することにより，仮説の無矛盾性管理を可能としている．さらに本章では，複雑なレイアウト構造を有する文書の一例として名刺を取り上げ，構造解析実験から本手法の有効性を確かめる．

第5章では，文書構造解析により得られた構成要素に基づき，文書の内容に関する構造化記述を生成する処理について述べる．本手法は，記述の生成を通して同時に仮説を取捨選択するものである．即ち，論理的に妥当な記述が得られない場合に，構成要素の仮説を棄却する．文書の記述内容は整合性を持つ単語列の集合であるとの観点から，本手法では，単語列生成処理，単語列整合処理の2処理により構造化記述を生成する．単語列生成処理では，単語の接続性に基づき，文字切り出し・認識候補から妥当な単語列を効率良く生成する．単語列整合処理では，構成要素の抽出履歴を逆にたどることにより，構成要素内に含まれる単語列を構造化する．このとき，文書モデルに記述された論理的制約を参照し，構造化された単語列の整合性を保証する．さらに，本章では，前章の実験で抽出された構成要素を対象に記述生成実験を行い，本手法が構造化記述生成および仮説検証に有効であることを示す．

第6章は結論であり，本研究で得られた諸結果を総括すると共に，今後の課題について述べる．

目次

第 1 章

緒論

パーソナルコンピュータ，ワードプロセッサの普及に代表されるように，大企業は無論のこと，個人生活にまで計算機による各種情報の生産・蓄積・処理が普及し，高度情報化社会がまさに開かれようとしている．このような状況下においても，社会的に必要とされる情報の大部分は文書の形式で蓄積，伝達されており，後に述べるような理由から今後もこの状態は継続すると考えられるため，文書に記録された情報 (以後，文書情報と呼ぶ) の計算機による処理は，高度情報化社会を形成していく上で，欠くことのできない要素といえるであろう．

現在，文書情報を計算機により処理可能とするために研究されている種々の対処法は，以下の 2 種類に分類できる．

ひとつは，現在までに紙に記録された文書情報を，計算機に入力する手段の開発である．対象となる文書情報が膨大であることから，人手による入力は，特殊な例を除けば時間，コストなどの面から実際的ではない．このため，我国では 20 年以上も前から文字認識の研究が盛んに行われている [1]．近年では，文書画像から文字を切り出すまでの前処理 [2, 3]，および文字認識誤りを訂正する後処理 [4, 5, 6] の研究成果が蓄積されるにつれ，それら 3 者を合わせた文書画像認識システムが論じられるようになってきた [7, 8]．

他のひとつは，新たに作成される文書情報を扱う手段の検討である．これについては，紙などの物理的メディアに記録するのではなく，最初から磁気記憶などの電子的メディアに記録し，計算機を介して利用するシステムが考えられる．新聞社など，日々大量の文書情報が作成される現場では，企業レベルでそのようなシステムの導入が図られている．また，ワードプロセッサや DTP システム (DeskTop Publishing System) の発達に伴い，個人レベルにおいても，文書情報の記録形式が電子化しつつある．近年では，オフィスにおける文書情報の記録形式を標準化し，効率的に検索，管理，通信しよう

とする動きにまで発展してきている [9, 10, 11]．今後，このような標準化は，オフィス文書にとどまらず，種々の文書に対してなされるであろう．

さて，将来，このような標準化が進めば，ペーパーレス・オフィスなどの言葉に代表されるように，紙が不要となる社会が実現可能との主張が見聞される．しかしながら，筆者は次の理由により，将来においても紙というメディアが不要になるとは考えない．電子的に記録された文書であっても，人間に提示される際には，多くの場合，紙に印刷されることになる．これは，情報処理システムの進展とともに，紙の消費量が年々増大していることからも裏付けられる．また，紙というメディアが人間に非常に適していることも一つの理由であろう．一度，紙に印刷された文書は，それ自体が情報のメディアとして他の人間に伝達される可能性は高いと考えられ，情報受信側では，印刷された文書がオリジナルな電子的記録から独立して流通するに至ると推察される．従って，紙が人間に適合したメディアである限り，印刷された文書から新たに文書情報を入力する必要があると考えられる．

それでは，従来の文書画像認識技術は，文書情報の入力手段として十分であろうか．図や表，写真などを除いた文書中の文字部分に着目すると，文書画像認識とは文書に含まれる文字の認識であると考えられる．すなわち，文書画像に含まれる個々の文字がすべて認識できれば，文書画像は認識された，少なくとも記号情報に変換されたことになる．ところが，前述の標準化形式にもあるように，文書情報は，論理構造と呼ばれる構造を持つものである．すなわち，文書は単なる文字の羅列ではなく，題名，著者名，章，節などから構成されるものである．前述の標準化形式では，文字を論理構造に基づき構造化することにより，高効率かつ高機能な検索，通信や変換を可能としている．従って，文書情報の有効利用という観点から考えると，個々の文字の認識に留まらず，文字の認識結果を論理構造に則って構造化することが必須となる．本論文では，認識結果の構造化までを含めた総合的な処理を，文書画像認識に対して文書画像理解と呼ぶ [12]．

文書画像理解においては，文書情報の構造化の際に，文書画像から得られる特徴量から論理構造を推定する必要がある．ここで，文字との対比により，論理構造の性質について考えてみる．まず第1に，文字が直接紙面に印刷されるのに対して，論理構造は紙面上には直接記述されないものであるといえる．論理構造は，文書のレイアウトや記述内容に間接的に反映されるに過ぎない．第2に，文字が文書情報のプリミティブであるのに対して，論理構造は文書情報の高次の構造であるといえる．前述の標準化形式では，木構造により論理構造を表現している．第3には，文字が対象文書に依存しないのに対して，論理構造は対象文書の記述内容，形式などに大きく依存したもので

あるといえる．書状，新聞，技術論文など文書の種類(以後，文書クラスと呼ぶ)が同一であれば，それぞれの論理構造を反映したレイアウトがそれぞれに固有な特徴を持つように，文書クラス内ではほぼ一定であると考えられる．ところが，文書クラスが異なればレイアウトも異なるため，論理構造についても同様に全く異なるといえるであろう．

以上のことを考えると，文書画像理解を実現するためには，文書画像認識とは異なったアプローチが必要となることがわかる．文書画像認識においては，文字に関する上記の性質から，文書クラスの区別なく使用可能な“辞書”を用いた認識処理が実行可能である．ここでいう辞書とは，各文字カテゴリと画像特徴量の対応を記録したデータベースを指し，それ自身が複雑な構造を持たないものである．ところが，文書画像理解では，論理構造が複雑な構造を持つことと同時に，文書クラスにより大幅に変化することから，対象文書に依存しない辞書的な基準を用いて論理構造を推定することが事実上，不可能となる．さらに，論理構造が紙面上に直接記述されないため，画像特徴量と論理構造の対応を，多様な側面から柔軟に決定しなければならない．このことから，文書画像理解においては，画像特徴量から論理構造を推定するための“領域知識”(domain specific knowledge；以後，単に知識と呼ぶ)が必要であると推察される．従って，一般的な知識ベースシステムと同様に，文書画像理解システムにおいても，使用する知識の種類と内容，記述法，および利用法が有効性を左右する大きな要因であると考えられる．

そこで筆者は，様々な文書クラスに対しても高い有効性を持つ文書画像理解システムを構築するためには，必要となる知識を明確に定義・記述し，積極的に利用するという知識工学的なアプローチが不可欠であると考え，本論文では，知識利用型文書画像理解システムについて検討する．

まず，第2章では，文書画像理解に関連する従来の諸研究を概観し，文書画像理解の種々の問題点を明らかにすると共に，知識利用型文書画像理解システムの概念設計を行う．本システムは，文書画像理解に必要となる知識が対象文書に依存することを考え，処理部から分離して蓄積するという構成を基本とするものである[13]．第3章では，第2章の議論を受け，本システムの知識ベース部である文書モデルの詳細について論じる．本手法は，文書の論理構造が，レイアウト構造，および記述内容と密接な関係をもつことに着目したものであり，知識としては，レイアウト構造と論理構造の関係記述(レイアウト構造に関する知識)，および記述内容と論理構造の関係記述(記述内容の論理的制約に関する知識，あるいは単に論理的制約に関する知識と呼ぶ)を提案する．レイアウト構造に関しては，知識を階層的，記号的，宣言的に記述することにより，高い表現能力，記述容易性，可読性を実現する[14, 15.

16]．一方，論理的制約に関しては，単語の接続性，および単語列の整合性という 2 種類の特徴による新しい知識記述法を提案する [17, 18, 19]．また，既存の知識記述法との比較検討を通して，本手法の有効性を検討する．第 4 章では，レイアウト構造に関する知識を使用することにより，文書画像の構造を解析し，論理構造を抽出する文書構造解析なる手法を提案する [20, 21, 22]．文書の論理構造は，記述内容を考慮して初めて決定できるという考えに基づき，本手法では，抽出結果を仮説として扱う [23, 24, 25]．第 5 章では，記述内容の論理的制約を用いることにより，文書構造解析において生成された仮説から妥当なものを選択する処理について述べる [18, 19, 26]．処理結果は，文書の論理構造により構造化された記述内容 (構造化記述) であるため，本論文では，この処理を構造化記述生成処理と呼ぶ．また，第 4 章，第 5 章において提案した手法を名刺に対して適用した実験結果から，本手法の文書画像理解への有効性を確かめる．

第 2 章

システムの概要

2.1 緒言

文書画像理解とは，文書画像という信号データを解釈し，構造化記述という記号列を生成する一連の処理であると考えられる．文書の内容に関する構造化記述を生成するためには，対象文書の論理構造と画像から得られた特徴量を対応付けるための知識が必要となる．

特定種類の文書を対象としたシステムを構築するのではなく，多種類の文書を処理可能なシステムの構築を目指す場合，以下の 3 点に留意する必要があると考えられる．

第 1 は，文書画像理解に必要となる知識に関する問題である．多種類の文書画像を安定して理解するためには，知識としてどのような種類のものを用意すべきであろうか．少なくとも，このような知識は，特定の文書にのみ考え得るものではなく，大多数の文書において設定可能でなければならない．また，知識の量と質に関しても，文書画像を理解するために十分でなければならない．

第 2 は，処理の不完全性に対処する方法である．文書画像理解は，信号レベルから記号レベルに至るまで，数多くの処理を実行することにより可能となる．多種類の文書を対象とする場合，文書画像における変動要素の種類および程度が共に増大することが予測されるため，個々の処理においては，誤りを皆無にすることが非常に困難となる．従って，各処理の結果は誤りを含む可能性があること，すなわち処理の不完全性を考慮する必要がある [27, 28]．

第 3 は，知識の対象依存性に対処する方法である．文書の論理構造は，対象とする文書に強く依存したものである．従って，文書画像理解においては，どのような種類の知識を用いるにせよ，論理構造との対応付けのための知識である限り，知識は対象文書に依存したものとなる．従って，文書画像理解

システムの構築を目指す場合，知識の扱い方が問題となる．もし，知識の変更が容易に行えないようなシステムであれば，多種類の文書を対象とすることは，事実上，困難となる．

そこで本章では，まず文書画像理解に必要な知識について考察する．つぎに，第2，第3の問題点に焦点をあて，従来の文書画像理解関連の研究を概観し，その結果，問題点の解決方法として，仮説生成検証法，および知識利用型システム構成を提案する．

2.2 文書画像理解のための知識

一般に，文書の論理構造は，文書の論理的な構成要素(以後，単純に構成要素と呼ぶ)の間に存在する階層構造として表現可能である．技術論文における論理構造の例を図 2.1に示す．文書画像理解とは，このような論理構造に則った形で文書の記述内容を構造化して出力することであると考えられる．なお，本論文では，構造化記述の生成という観点から，文書の標準化形式[11]における論理構造の概念を拡張し，階層の最下位レベルの構成要素として文字を考える．

文書画像理解には，文書の論理構造と入力画像から抽出された特徴との対応をとるための知識が必要である．文書には，2次元画像としての空間的側面，記述内容の論理的側面の2側面があることから，知識としては，前者に対応するものとしてレイアウト構造に関する知識，後者に対応するものとして記述内容の論理的制約に関する知識の2種類を考える．以下，各々について述べる．

2.2.1 レイアウト構造に関する知識

文書を空間的側面からとらえ，文書の基本となる領域として文字領域を考えると，一定方向に並ぶ文字領域を包含する領域として，文字列領域を定義することができる．さらに，複数の文字列領域を包含する領域として，ブロック領域を考えることができる．ブロック領域がいくつか集まることにより，さらに大きなブロック領域を構成し，最終的には文書領域に至る．すなわち，文書には領域の包含関係を基にした階層構造が存在するといえる．このような領域の階層構造は，論理構造に対して，レイアウト構造と呼ばれる[11]．なお，以後は，ブロック領域，文字列領域などレイアウト構造を規定する要素をレイアウト要素と呼ぶことにする．

さて，我々人間が文書を作成する際には，文書の構成要素を紙面にレイア

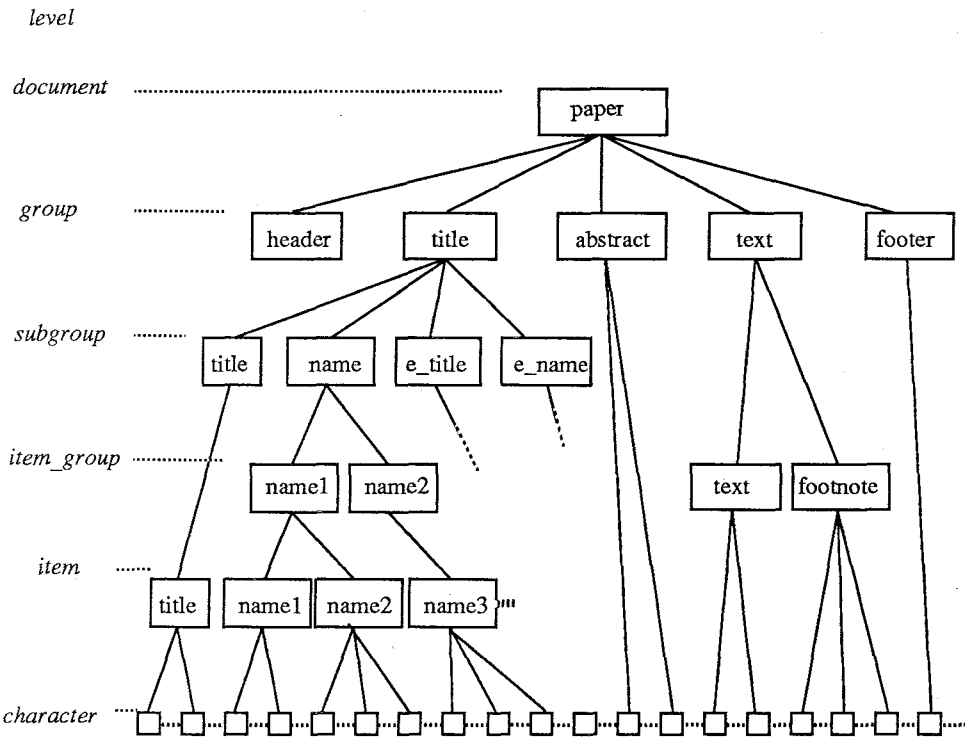


図 2.1: 技術論文の論理構造

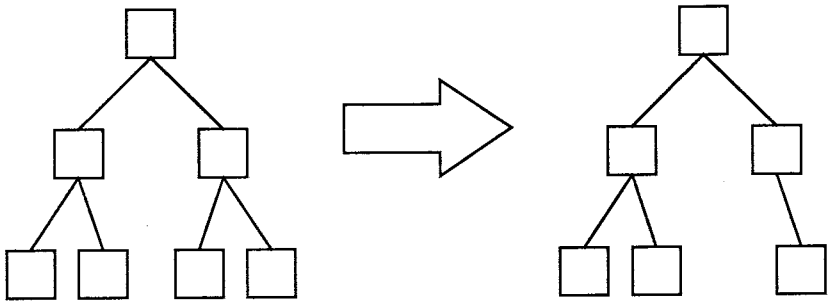


図 2.2: レイアウト構造に関する知識

ウトするプロセスを経ていると考えられる．このとき，個々の構成要素は，作成している文書の論理構造が視覚的に了解できるように，効果的にレイアウトされる．即ち，レイアウト構造と論理構造の間には，ある種の対応関係が存在すると考えられる．例えば，技術論文では，一般に“題名”はフロントページ上部に大きな文字で書かれており，ページ中で横方向にセンタリングされている．他の構成要素についても，同様に種々の規則に則り，レイアウトされているといえる．

以上のことから，文書のレイアウト構造と論理構造の対応関係に関する知識は，文書画像理解に有用であるといえる．このような知識を用いると，文書画像のレイアウト構造を解析することにより，論理構造を推定できると考えられる．本論文では，このような知識をレイアウト構造に関する知識と呼ぶ．レイアウト構造に関する知識を記述するためには，図 2.2に示すように，対象文書の論理構造，レイアウト構造，およびそれらの間に成立する対応規則を記述しなければならない．

レイアウト構造については，前述のレイアウト要素とそれらの間の階層関係により記述可能である．ただし，知識として記述するためには，レイアウト要素および階層関係について明確な定義が必要である．本論文では，以下に示すような定義を採用する．レイアウト構造は，図 2.3に示すような文字領域，文字列領域，ブロック領域，および文書画像全体の領域を表す文書領域の 4 種類のレイアウト要素により定義される．文字領域とは，個々の文字を

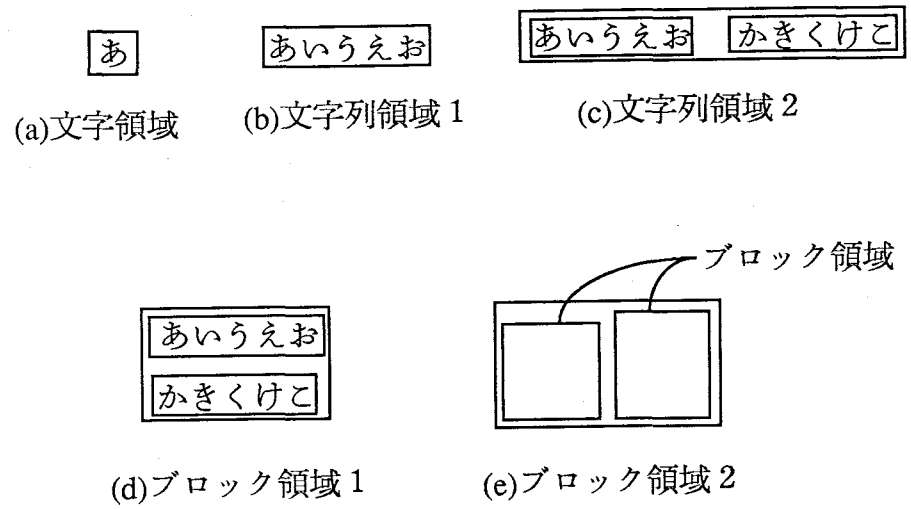


図 2.3: レイアウト要素

囲む領域である (図 2.3 (a)). 文字列領域とは，一定方向に並ぶ文字領域を囲む領域である (図 2.3 (b)). また，複数の文字列領域において，文字領域の並ぶ方向が同じであり，かつ文字列領域の並ぶ方向が文字領域の並ぶ方向と同じものであれば，それらを囲む領域も文字列領域であるとする (図 2.3 (c)). ブロック領域とは，一定方向に並ぶ文字列領域を囲む領域であり，かつ文字列領域ではないもの (図 2.3 (d))，あるいは，一定方向に並ぶブロック領域を囲む領域であるとする (図 2.3 (e)). 以上のようなレイアウト要素を採用すると，レイアウト構造の階層関係は木構造になる．レイアウト構造を表す木において，根は文書領域，葉は文字領域である．以後は，このような木により表される階層構造の各階層レベルを，単にレベルと呼び，レベルにおいて相対的に根に近い方を上位レベル，葉に近い方を下位レベルと呼ぶ．

本論文では，レイアウト要素の領域を矩形領域に制限する．また，あるレイアウト要素が，1 つ下位レベルのレイアウト要素を包含する際，下位レベルのレイアウト要素の並ぶ方向を縦方向あるいは横方向のいずれかに制限する．このような制限は，近年，DTP(DeskTop Publishing) システムとして注目されている TEX[29] においても，基本構造において採用されているものであり，大部分の文書に対して妥当であると考えられる．

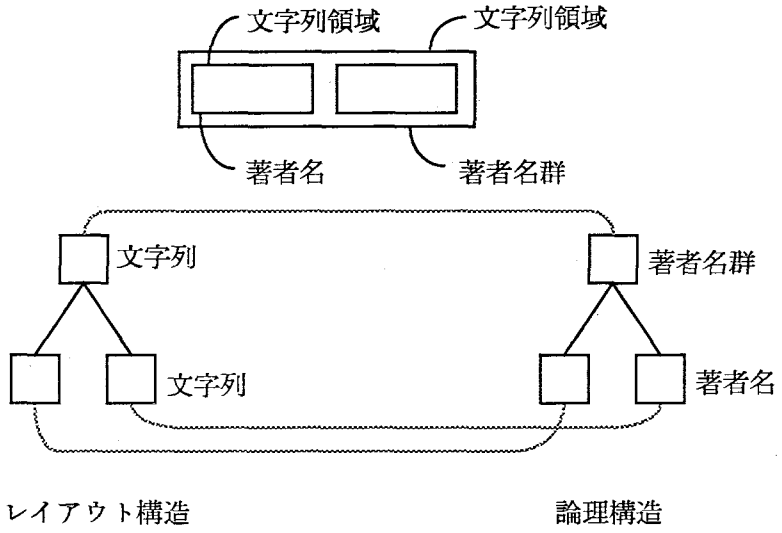
論理構造についても，前述の構成要素とそれらの間の階層関係により記述可能である．構成要素に関しては，対象とする文書，あるいは目標とする構造化記述の形式に応じて，任意に決定可能であるとする．また，階層構造としては，文書の標準化形式に採用されているように，木構造を考える．論理構造においても，レイアウト構造において定義した，レベル，レベル間の上位，下位という用語を同様の意味に用いる．

対応規則を記述するためには，その性質に関する分析が必要である．文書は論理構造が視覚的に了解できるようにレイアウトされるという原点から考えると，基本には，レイアウト構造と論理構造の階層関係は一致するものであると考えられる．実際，図 2.4 (a) に示すように，技術論文の“著者名”は，文字列領域に対応し，“著者名”から構成される“著者名群”は“著者名”を囲む文字列領域に対応するものである．ただし，図 2.4 (b) に示すように，段組などを表わすブロック領域や，“あらまし”における文字列領域などは，紙面の物理的な制限から必要となるレイアウト要素であり，対応する構成要素が存在しないことに注意する必要がある．このような場合，文字がどの段組領域，あるいはどの文字列領域に属しているかということには意味がなく，文字の読み順 (reading order) のみに意味があると考えられる．従って，レイアウト構造と論理構造の対応規則には，図 2.5に示すように，段組の領域や文字列領域を論理的なつながりに展開するための規則を記述する必要がある．

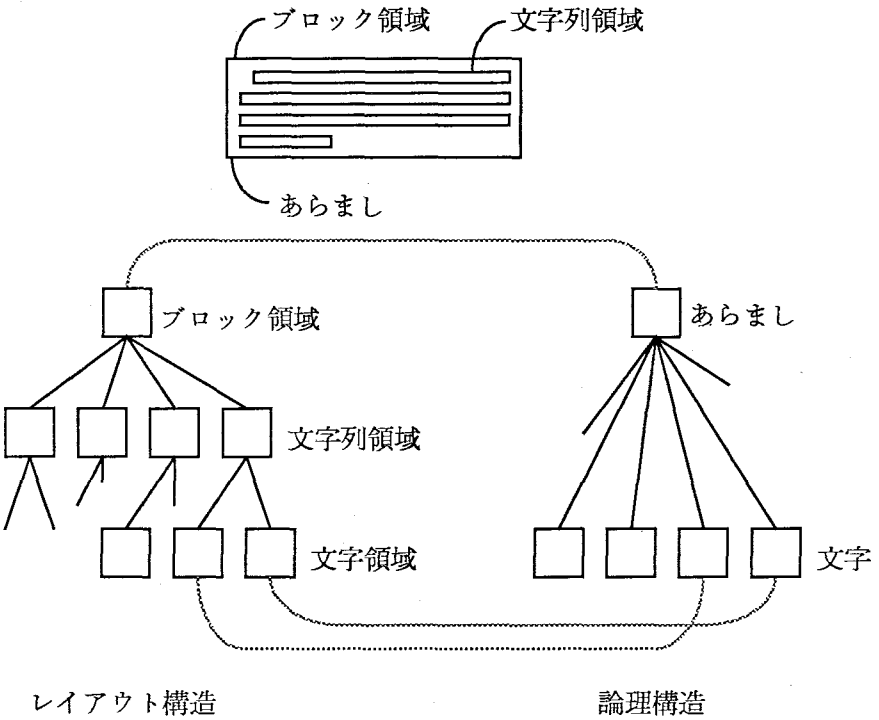
以上の点を考慮すると，文書の個々のインスタンスに対しては，レイアウト構造と論理構造の対応関係を記述可能であると考えられる．しかしながら，文書画像理解のための知識としては，インスタンスに関するものではなく，“技術論文”など文書のクラスに対する記述が求められる．従って，レイアウト構造，論理構造，対応規則ともに，クラスの記述に対応することが必要となる．

まず，論理構造のインスタンスに対する，対応規則およびレイアウト構造のクラス記述の必要性について述べる．この点を明らかにするため，もう一度，文書作成のプロセスについて考えてみる．文書の論理構造が定まり，それをもとに実際にレイアウトする場合，我々は様々な対応規則を考慮して，妥当な紙面を構成する．ただし，論理構造をレイアウト構造に対応させる際には，ある程度の自由度があるのもまた事実である．このような自由度には，

1. 個々のレイアウト要素の“形状”に影響を及ぼすもの
   例) 文字の大きさや行間の変動



(a) 著者名群と著者名



(b) あらまし中の文字列

図 2.4: 対応規則の例

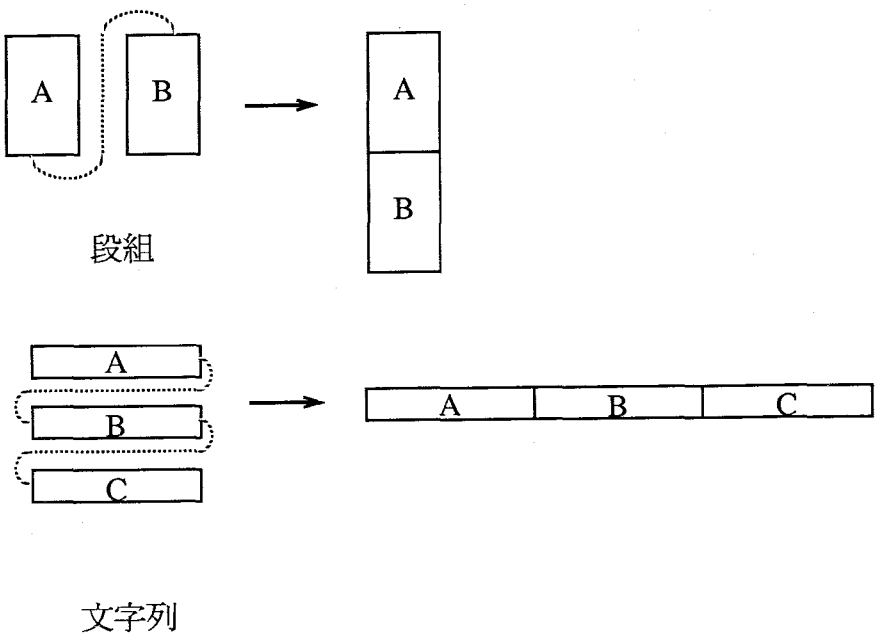


図 2.5: 展開規則

2. レイアウトの“構造”(階層関係) に影響を及ぼすもの
   例) 段組数の変動

が存在するといえる．本論文では，これらの変動をレイアウト構造のバリエーションと呼ぶ．また，前者を形状的バリエーション，後者を構造的バリエーションと呼ぶ．以上のことから，論理構造とレイアウト構造には，一般に，1対多対応の関係があるといえる．1の変動を許容した記述を実現するためには，レイアウト構造の記述において，レイアウト要素の変動許容範囲を明記する必要がある．また，2に対しては，レイアウト構造の階層関係，および対応規則の記述に変動を記述可能なことが求められる．

それでは，論理構造のクラスを記述するためには，どのような注意が必要であろうか．この点を明らかにするために，技術論文を例に，まず論理構造のクラス記述について考える．技術論文には，単一著者のもの，複数著者のものなど，構成要素の個数が変化する場合がある．さらに，題目に副題が付与されているもの，副題が存在しないものなど，構成要素の有無にかかわる変動もある．従って，知識記述に際しては，このような論理構造の変動を記述しなければならない．また，当然ながら，このような変動は，レイアウト構造，対応規則に次に示す影響を及ぼす．

1. 形状的バリエーション
   例) 文字数の変動による文字列領域の形状への影響

2. 構造的バリエーション
   例) 構成要素の有無によるレイアウトの階層構造への影響

従って，前述と同様の対処が必要となる．

以上の点に留意し，レイアウト構造に関する知識の記述が実現できれば，それを用いた処理により，文書画像理解を実現することが可能であると考えられる．ただし，文書中の構成要素には，技術論文の“参考文献リスト”における“著者名”，“論文名”などのように，レイアウト構造が不定のものが存在する．このような構成要素を含む文書を理解の対象とする場合には，別の側面，すなわち記述内容の側面からとらえた知識が必要となる．

2.2.2 論理的制約に関する知識

文書の論理的な構成要素は，それぞれが論理的に妥当な記述内容を持つものである．例えば，論文の“所属”には研究機関の名称を表す文字が含まれており，“著者名”には研究者の名前を表す文字が含まれている．このよう

な性質は，論理構造の階層におけるさらに上位の構成要素に対しても，同様に存在するものである．例えば，“著者名群”(複数の著者名からなる構成要素) には，“著者名”のみが含まれていなければならず，それぞれが個別の研究者に対応するものである．記述内容という側面から考えると，文書の構成要素は，論理的に妥当な記述内容を持つ下位レベルの構成要素の集合であるといえる．

以上のことを考えると，個々の論理的な構成要素に対して，妥当な記述内容を規定する規則を知識として蓄え，利用すると，文字をもとに論理構造を推定することが可能であるといえる．本論文では，このような知識を記述内容の論理的制約に関する知識と呼ぶ．

それでは，このような知識はいかに記述すべきであろうか．技術論文における参考文献リストなど，記述される事項がある程度，定まっているものについては，論理的制約は比較的容易に記述可能であると考えられる．しかしながら，“本文”のように，論理的制約を完全に記述すること，すなわち妥当な記述内容を完全に規定することが非常に困難な構成要素も存在する．これは，記述内容が自然言語文である構成要素に特有な性質ではなく，“題名”のように，項目的なものに関しても，同様の議論が成り立つ．

このような構成要素に対しては，妥当な記述内容を完全に規定するのではなく，妥当であるための必要条件を制約として記述するという方法を取らざるを得ない．具体的には，“本文”の記述内容に対して自然言語文としての妥当性を求めること，“題目”の記述内容に対して記述中に含まれる個々の単語の妥当性を求めることに対応する．従って，このように記述された論理的制約を満たす記述内容は，必ずしも構成要素の記述内容として妥当であるとは限らない．しかしながら，必要条件が記述されることを考えると，論理的制約を満たさない記述内容には，構成要素として妥当なものは含まれない．文書画像理解という立場からは，このような弱い制約の記述であっても，知識としての価値があると考えられる．これは，制約を満たさないものを不適当な記述内容として排除できるからである．

レイアウト構造に関する知識と論理的制約に関する知識の比較という観点から，これまでの議論をまとめると，以下のことがいえる．

- レイアウト構造に関する知識では，規定することが困難な構成要素が存在するが，このような構成要素に対しては，論理的制約に関する知識が有効である．

- 論理的制約に関する知識では規定することが困難な構成要素が存在するが，このような構成要素については，レイアウト構造に関する知識が有

効である．

従って，レイアウト構造に関する知識と，記述内容の論理的制約に関する知識は，相補的であると結論付けられる．我々人間が文書を理解するときには，両者の知識を有効利用して，高い精度を得ていると考えられる．従って，本手法においても，同様に両者の知識を使用することにより，高精度の文書画像理解の実現を目指す．

2.3 仮説生成検証プロセス

2.3.1 仮説生成検証プロセスの必要性

文書画像理解を実現するためには，ノイズ除去，領域分割，文字切り出し，文字認識，言語処理など，信号レベルから記号レベルに至るまで，種々の処理を実行しなければならない．従来，これらの処理は個々独立に研究されてきたという経緯から，言語処理などの後処理を除けば，提案されている処理手法のほとんどにおいて，入力されるデータが正しいという前提が敷かれている．例えば，文字認識の研究では，認識対象となる文字の画像が正しく切り出されているという前提があり，文字切り出し手法のほとんどは，正しく切り出された文字列を対象とするものである．

これらの従来手法が，対象とする文書に対して有効性が高く，また処理誤りの可能性が低い場合には，信号レベルから記号レベルに至る処理を順に接続することにより，文書画像理解を実現可能であると考えられる．実際，Akiyama らは，新聞を対象にしたシステムの開発例を報告している [7]．しかしながら，

- 複数種類の文書を対象とする場合
- 単一種類でもレイアウト構造に種々のバリエーションがある場合
- サンプルにより画像の品質が安定しない場合
- 様々な大きさやフォントの文字が使用されている場合

など，対象とする文書画像に著しい多様性が存在する場合には，上記のような処理形態を採用することは困難となる．このような文書を対象とする場合には，各処理の誤りが無視できない程度に発生するため，単純に接続しただけでは処理が進むにつれて誤りが累積してしまう．従って，一般的には，文書画像理解を構成する処理の間の接続形態に関して，処理誤りが起きること

を考慮した手法，すなわち，処理の不完全性に対処してシステム全体のロバスト性を向上させる手法が必要となる．

以上の問題点の解決を試みた先駆的研究としては，馬場口らの手法 [30] や村瀬らの手法 [31] を挙げることができる．このうち，馬場口らは，手書き文字列を対象に，文字切り出しと文字認識の協調的な動作を論じている．また村瀬らは，文字切り出しと文字認識に加えて言語処理を導入することにより，文字切り出しと文字認識のみでは対処が困難となる問題 — 分離文字“明”は“日”，“月”とも読める — の解決法を提案している．これらは共に手書きという困難な対象を用いた意欲的な研究である．しかしながら，文書画像理解という枠組みにおいては，必要となる処理の一部分に関するものであるため，提案されている枠組みを文書画像理解全体に，そのまま適用することは困難であると考えられる．

文書画像理解というトータルな立場からの検討としては，Wang らの手法 [32]，および Kubota らの手法 [33] がある．これらは共に，領域分割などの各処理の結果に評価値を与え，伝搬することにより，全体として一貫性のある解を求めるものである．しかしながら，このような評価値に基づく手法には，

- 各処理の結果に対する評価値の一般的な算出方法が明らかではない．
- 個々の処理に対して妥当な評価値が得られたとしても，一貫性のある解を得るための評価値の一般的な伝搬法を定義することは困難である．
- 様々な要因により引き起こされる各処理結果の不完全性を，評価値という“数値”に縮約してしまうため，全体として処理制御の見通しがよいとはいえない． さらに，処理に失敗した場合に，その原因を突き止めて対処することが困難である．

などの問題点があるため，必ずしも得策であるとはいい難い．

さて，人工知能や画像理解の分野においては，近年，仮説という概念に基づく推論の重要性が認識されつつある [34, 35, 36]．これは，“同様にもっともらしい選択肢からの選択”という仮説的状況を扱うことが，様々な問題解決の過程において共通して必要となるからである．仮説に基づく推論においては，仮説という選択肢が明示的に生成・記録され，また仮説間の依存関係についても推論の履歴等を基に記録される．従って，上記の評価値による対処法において問題となった点を解決する方法として優れていると考えられる．

そこで本手法では，文書画像理解における処理の不完全性を，

- 各処理においては，処理結果を一意に定めることは困難であっても，候補を挙げることは可能である．

- 候補中に正解が含まれていない可能性は十分低い.

という形でモデル化し，文書画像理解における処理を広義の推論ととらえることにより，仮説に基づく推論の導入を図る [28, 37].

2.3.2 仮説生成検証法の概要

前項において議論した仮説生成検証プロセスの基本概念に則り，ここでは，本システムにおいて採用する仮説生成検証法の概要について述べる.

文書画像理解に利用可能な知識には，2.2に述べたように，レイアウト構造に関する知識，記述内容の論理的制約に関する知識が存在する．ここで，後者の知識が文書中の文字を認識して初めて利用可能であることを考えると，文書画像理解の手順としては，

1. レイアウト構造に関する知識を基に，文書画像の構造を解析して構成要素を抽出し，最終的に文字切り出し・認識を実行する.

2. 記述内容の論理的制約を用いて1の結果を検査し，文書の内容に関する構造化記述を生成する.

というものが妥当であろう．本論文では，1の処理を文書構造解析，2の処理を構造化記述生成と呼ぶ.

文書構造解析では，記述内容を考慮せず，レイアウト上の特徴のみに基づいて構成要素を抽出する．従って，処理の不完全性，およびレイアウト構造のバリエーションの影響から，複数の可能な構成要素が抽出される．そのうち，いずれが適切であるかは記述内容を考慮してはじめて決定できると考えられる．そこで，本手法では，文書構造解析の段階では構成要素を一意に定めることを避け，複数の候補を仮説として許容する．すなわち，文書画像理解処理全体において，文書構造解析を仮説生成処理として位置づける．文書構造解析で生成された仮説は，構造化記述生成を通して取捨選択される．すなわち，構造化記述生成は仮説検証処理に対応するものである.

仮説生成検証法の概念を図 2.6に示す．仮説生成処理の目標は，レイアウト構造に関する知識を満たす構成要素を候補としてすべて生成することである．構成要素候補の生成は，

1. レイアウト構造に基づき，上位レベルで抽出された構成要素候補の領域中に含まれる可能なレイアウト要素を候補として抽出する.

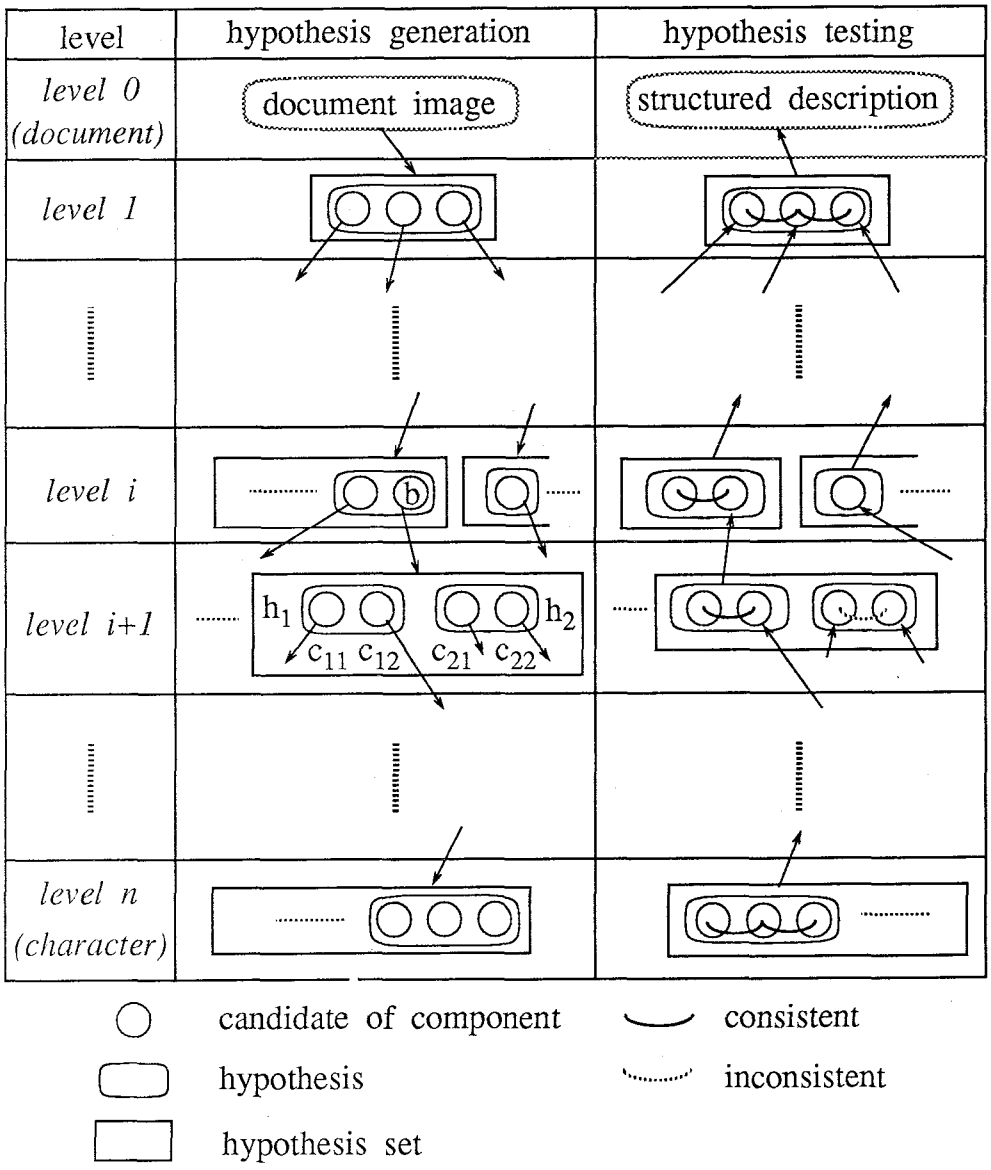


図 2.6: 仮説生成検証法の概念

2. レイアウト構造から論理構造への対応規則を用いて，レイアウト要素の候補を構成要素に対応付ける．

の2種類の処理からなる．1の処理は下位レベルにおける構成要素の領域を求める処理であるため，領域抽出と呼ぶ．また2の処理は，抽出された領域に構成要素名を属性として付与することと考えられるため，属性付与と呼ぶ．本手法では，レイアウト構造の階層性を利用して上位レベルから下位レベルへと処理を進めることにより，効率良くこの目標を満たす．なお，以後は，構成要素候補 c を領域 r と属性 a の組により，$c =< r, a >$ と表す．

各レベルにおける仮説生成処理では，上位レベルで生成された構成要素候補 (図 2.6の円に対応) を正しいものと仮定し，下位レベルで構成要素候補を生成する．ここで，正しいと仮定した構成要素候補を，下位レベルの構成要素候補に対するベースと呼び，その領域をベース領域と呼ぶ．

2.2.1において述べたレイアウト構造の階層性に関する性質を考えると，下位レベルにおける構成要素の領域は，ベース領域を横方向あるいは縦方向に分割したものであるといえる．ここで，レイアウト構造にはバリエーションが存在すること，および処理には不完全性が存在することを考えると，一般には，1つのベースから幾通りかの構成要素候補の組 (図 2.6の長方形に対応) が生成される．図 2.6の $level\ i$ と $level\ i+1$ の間では，ベース b から構成要素候補の組 (c_{11}, c_{12}) と (c_{21}, c_{22}) が生成された場合を表している．

図 2.7に示すように，上記の構成要素候補の組はベース領域の異なる分割パターンに対応するため，各々は同時には存在しえない (互いに矛盾する) ものであるといえる．本論文では，このような構成要素候補の組を仮説と呼ぶ．また，同一のベースから生成された仮説の集合を仮説集合と呼ぶ．さらに，ある仮説に複数の構成要素候補が含まれる場合，それらすべてが正しいときにのみ，その仮説を正しいものとみなすことができる．以上のような依存関係は，一般的に次のように記述できる．

$$b \Rightarrow H \tag{2.1}$$

$$H = \{h_1, ..., h_m\} \tag{2.2}$$

$$h_i = and(c_{i1}, \ldots, c_{in}) \tag{2.3}$$

ここで，b,H,h_i,c_{ij} はそれぞれベース，仮説集合，仮説，および構成要素候補を表す．また式 (2.1)，(2.2)，(2.3) はそれぞれ，

式 (2.1) b から仮説集合 H が生成されたこと

式 (2.2) 仮説集合 H が互いに矛盾する仮説 $h_1, \ldots, h_m$ から構成されること

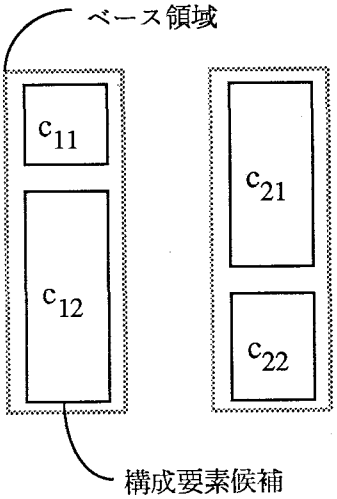


図 2.7: 仮説

式 (2.3) 仮説 h_i が構成要素候補 $c_{i1}, \ldots, c_{in}$ の連言として表されること

を示す．仮説生成処理とは，あるレベルで生成された構成要素候補を下位レベルのベースとすることにより，最上位の文書レベルから最下位の文字レベルまで，構成要素候補の生成を繰り返す処理である．

図 2.6に示すように，式 (2.1)，(2.2)，(2.3) により記録される仮説は木構造により表現できる．本論文では，仮説から構成される木を仮説木と呼ぶ．仮説木では，上位レベルの構成要素候補から下位レベルの仮説集合へのリンクは必ず1つである．また，一般に，仮説集合は複数の仮説を含み，仮説は複数の構成要素候補を含むものである．

一方，仮説検証処理の目標は，仮説集合から妥当な仮説を選択することである．仮説が式 (2.3) のように表わされることから，本論文で述べる仮説検証とは，記述内容の論理的制約という観点から，構成要素候補が式 (2.3) を満たすかどうかを検査することであるといえる．検査の結果，式 (2.3) 中の c_{ij} と c_{ik} $(j \neq k)$ が互いに矛盾する場合には，仮説 h_i を不適当として棄却できる．図 2.6に示すように，この処理を仮説木の葉から根へとたどりながら繰り返すと，上位レベルに移行するに従って仮説の数が減少し，最終的には，全レベルにおいて無矛盾な仮説を選択することが可能であると考えられる．

2.4 知識利用型システム構成

2.4.1 システムの適用性

文書画像処理，認識システムを評価する際の重要な尺度の１つに汎用性がある．システムを変更することなく処理可能な文書の種類が多ければ，それだけ汎用性の高いシステムであると考えられる．それでは，文書画像理解システムを評価する際にも，同様の尺度を用いることが可能であろうか．2.2にも述べたように，文書画像理解を実現するためには，対象とする文書に依存する知識が必要となるため，そのような知識をシステム内に保持しなければならない．このことは，システム全体を外側から眺めると，そのシステムがある特定の種類の文書に対するものであることを意味する．従って，文書画像理解システムを評価する際には，汎用性という尺度をそのまま用いることはできない．

そこで，本論文では，汎用性にかわる尺度として，適用性を考える．適用性とは，ある文書画像理解システムが，現在対象としている文書とは異なる種類の文書を対象とするための，システムの対応能力を意味するものである．

適用性には，適用容易性と適用可能性という２つの側面が考えられる．適用容易性とは，対象文書を変更する際の容易さを示す尺度である．また，適用可能性とは，知識の変更を許すという条件のもとで，処理可能な文書の種類の多さを示す尺度である．適用容易性の高いシステムであっても，適用可能性が低ければ，実際に処理可能な文書に制限が加わる．また，逆の場合においても，対象文書の変更に多大な労力を必要とするため，事実上，処理可能な文書に制限が加わることになる．有用な文書画像理解システムを構築するためには，適用容易性および適用可能性を高いレベルで実現する必要がある．

このような立場から，従来の文書画像理解関連のシステムを検討すると，必要となる知識の扱い方により，大きく非分離型 [33, 38, 39, 40, 41], 分離型 [42, 43, 44, 45, 46, 47] に分類されると考えられる．非分離型とは，知識と処理アルゴリズムが一体となった形式であり，知識は処理アルゴリズムそのもの，あるいは処理パラメータとして表現される．一方，分離型とは，知識が陽に表現され，処理機構と明確に分離された知識ベース型システムを指す．

適用容易性という観点から考えると，非分離型では処理アルゴリズムの変更，あるいは処理パラメータの再設定が必要となり，影響がシステム全体に及ぶのに対して，分離型では知識が分離されているため，知識のみを変更することが可能となり，基本的には処理機構の変更が不要となる．従って，十

分な適用容易性を得るためには，分離型のシステム構成が不可欠であると考えられる．ただし，高度な適用容易性を実現するためには，知識を分離するだけでは不十分であり，知識ベースにおいて，知識の記述容易性，および可読性が十分高いものでなければならない．加えて，適用可能性という観点から考えると，より多くの文書に対して知識を記述可能とするため，知識の記述形式が高い表現能力を有していること，および，処理機構が知識を有効に利用することにより，知識表現が可能な文書をすべて処理可能であることが要求される．

そこで，本論文では，高い適用性を実現するため，以上の点に留意したシステム構成を提案する．本システムでは，上記の要求を満たすために，分離型の構成を採用する．また，知識の記述容易性，可読性，表現能力，および処理機構の能力についても，上記の要求を満たすように，検討を加える．

2.4.2 各モジュールの役割

図 2.8に提案システムの構成を示す．本システムは大きく，文書モデル，処理ツール ，コアシステムに分割される．一般の知識ベースシステムと比較すると，文書モデルが知識ベース，コアシステムが推論エンジンに相当する．また，処理ツールは，コアシステムにより参照される外部手続き群である．以下に主要なモジュールの役割を示す．

1. 文書モデル
   文書のレイアウト構造および記述内容の論理的制約を蓄積する知識ベースである．2.2において考察した知識の性質を考慮し，従来手法に比べて高い表現能力，記述容易性，可読性を実現している．詳細については，第 3 章において述べる．

2. 文書構造解析部
   文書モデルに蓄積されたレイアウト構造に関する知識を基に，文書構造を解析し，構成要素に関する仮説を生成するモジュールである．本モジュールでは，外部手続きとして，基本画像処理部，文字認識部を起動する．処理の詳細については，第 4 章において述べる．

3. 構造化記述生成部
   文書モデルに蓄積された論理的制約を用いて，文書構造解析部により生成された仮説を検証するモジュールである．最終的には，対象文書の構造化記述を生成する．処理の詳細については，第5章において述べる．

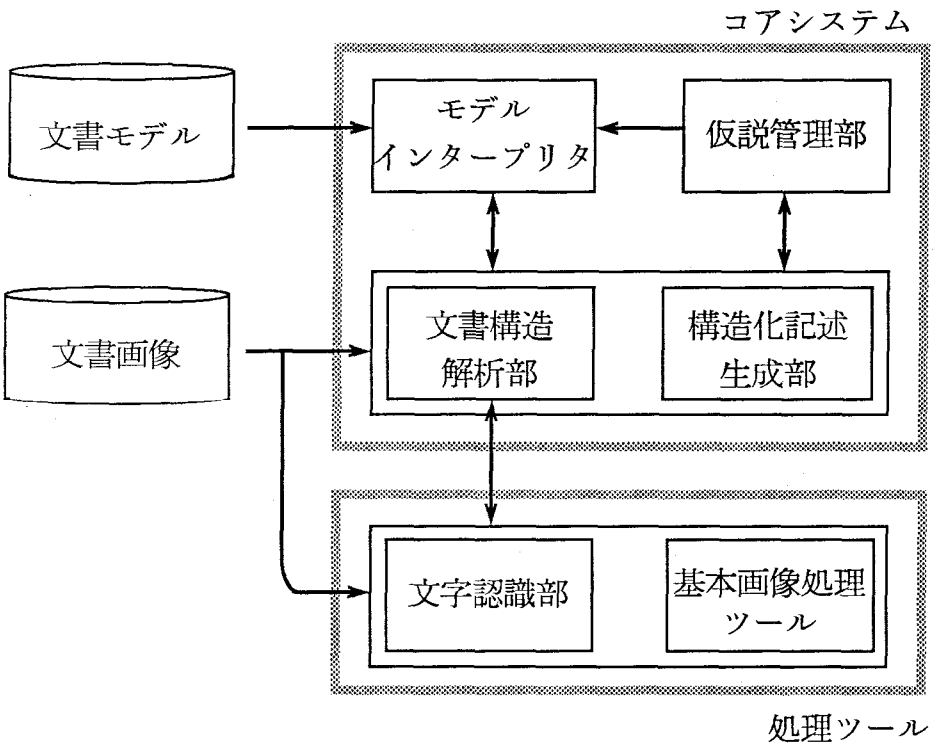

コアシステム
文書モデル
モデル
インタープリタ
仮説管理部
文書画像
文書構造
解析部
構造化記述
生成部
文字認識部
基本画像処理
ツール
処理ツール

図 2.8: システム構成

4. モデルインタープリタ
コアシステムと文書モデルのインタフェースの役割を果たすモジュールである．本モジュールは，知識の有効利用を図る目的で設定されたものであり，文書構造解析部や構造化記述生成部の要求に応じて，知識の検索，変換を担う [16]．知識の検索機能を用いることにより，コアシステムでは，知識の詳細な記述形式を考慮する必要がなくなり，知識と処理の分離が明確な形で実現できる．また，変換機能を用いることにより，知識ベースに陽には表現されていない知識をも利用可能となるため，知識の有効利用が図られる．

5. 仮説管理部
文書構造解析部により生成された仮説間の依存関係，および構造化記述生成部により発見された仮説間の矛盾を総合的に管理し，システムにおける処理を無矛盾に保つモジュールである．

6. 処理ツール
以下の2つのユニットを含む．

基本画像処理部 文書画像のノイズを除去し，文書構造解析に必要となる領域データを抽出する．

文字認識部 文書構造解析部において抽出された文字領域をもとに，領域内の文字を認識する．

2.5 結言

本章では，文書画像理解システムを構築する際に，留意しなければならない問題点として，文書画像理解に有効となる知識の種類と表現法，処理の不完全性への対処法，対象依存性への対処法の3点を指摘し，この観点から従来手法を検討した．また，その結果に基づき，知識利用型文書画像理解システムの概念設計を行った．本システムの特徴を以下に列挙する．

- 文書画像理解に有効な知識として，レイアウト構造に関する知識に加え，記述内容の論理的制約を用いること

- 処理の不完全性に対処し，システムのロバスト性を向上させるため，処理に仮説生成検証法を導入していること

- 文書画像理解に必要な知識が対象とする文書に強く依存することを考慮し，そのような知識を蓄える知識ベースとして文書モデルを設定し，コアシステムから分離することにより，システムの適用性を向上させていること

次章以降では，本章で提案した設計概念に則り，知識表現方法，文書構造解析法，構造化記述生成法の各々を明らかにするとともに，具体的な記述例，および処理実験結果の検討を通して，本システムの有効性を検証する．

第 3 章

文書モデル

3.1 緒言

知識分離型の文書画像理解システムにおいては，対象文書の種類を変更する際に，知識ベースの追加，変更が必要となる．従って，知識記述の表現能力，記述容易性，および可読性の 3 点がシステム全体の適用性を確保する上で重要となる．

レイアウト構造に関する知識に対しては，従来から様々な記述方法が提案されている．代表的なものとしては，矩形領域の座標を用いるもの [43, 48]，処理オペレータを用いるもの [45, 46]，ルールによるもの [42] などがある．しかしながら，これらの手法には，座標や処理パラメータを直接用いた数値的な記述であること，ルール形式の平板的な表現であるためレイアウト構造の階層性を陽に表現できないこと，処理オペレータによる記述であるため知識が手続き的であることなどから，知識の表現能力，記述容易性，可読性という観点からは十分であるとはいい難い．また，レイアウト構造におけるバリエーションの表現に関しては，一部の手法において，処理オペレータの適用順序による表現法が提案されているが [45, 46]，同様の観点から十分であるとはいえない．

記述内容の論理的制約に関する知識に対しては，図書目録カードを対象としたルール形式の記述法 [49] を挙げることができる．この手法は，図書目録カードにおける項目の境界が，特殊な区切り記号により区切られていることに着目したものである．しかしながら，一般には，そのような記号が存在するとは限らないため，知識の表現能力に疑問点が残る．

本章では，以上の問題点を考慮した知識ベースとして文書モデルを提案し，その構成および知識記述法について詳細に検討する．レイアウト構造については，従来の数値的，平板的，手続き的な記述方法を用いず，知識を可能な

限り記号的，階層的，宣言的に記述することにより，高い表現能力，記述容易性，可読性を持つ知識ベースの作成を試みる．論理的制約については，記述内容の構造を単語という立場から検討し，単語の接続性，単語列の整合性と呼ぶ2種類の特徴記述を用いて，構成要素としての妥当な記述内容の表現を試みる．

なお，本章以降では，対象文書を帳票，論文の表紙，図書目録カード，名刺などの項目主体の文書に限定する．この理由としては，新聞，本などの文章主体の文書に比べて，項目主体の文書には多種類の構成要素が存在するため，文書画像理解システムの有効性を評価するのに適した対象であること，計算機への自動入力が切望されている文書が多いことなどを挙げることができる．

本論文では，項目主体の文書のうち，具体例として縦書き名刺を用いて，議論を進めていく．これは，名刺には以下の様な性質，

- 氏名，社名，部署名などの構成要素が多数存在する
- 構成要素には，肩書のように個数が一定でないなど，レイアウト構造のバリエーションが豊富である
- 文字の大きさやピッチの変動が大きい

が存在することから，文書画像理解の基礎技術を確立するのに適した題材であると考えるからである．名刺には，縦書きと横書きのものがあるが，本論文では対象を縦書きのものに限定する．これは，縦書きのものが横書きに比べてレイアウト上の規則性が明確であるため，レイアウト構造に関する知識の記述形式を検討する第一段階の対象としては，適していると思われるからである [50].

3.2 論理構造

一般に，文書は階層的に構造化された複数の構成要素からなる．名刺における構成要素の例を図 3.1に示す (例示のための架空の名刺である)．最下位の文字レベルを除けば，名刺の構成要素には，文書 (document)，群 (group)，準群 (subgroup)， 項目群 (item_group)，項目 (item) の5つの階層レベルを考えることができる．名刺における論理構造を図 3.2に，また名刺中に含まれる項目レベル以上の全構成要素を表 3.1に示す．

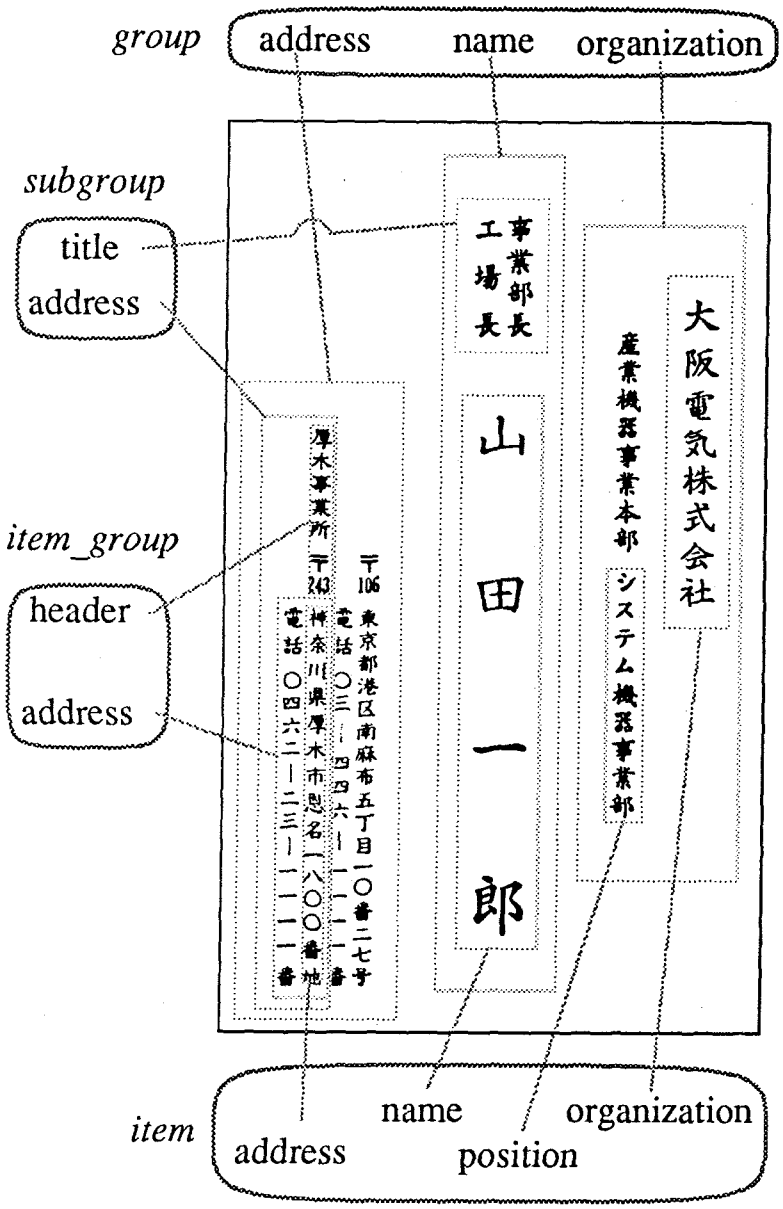


図 3.1: 名刺の構成要素例

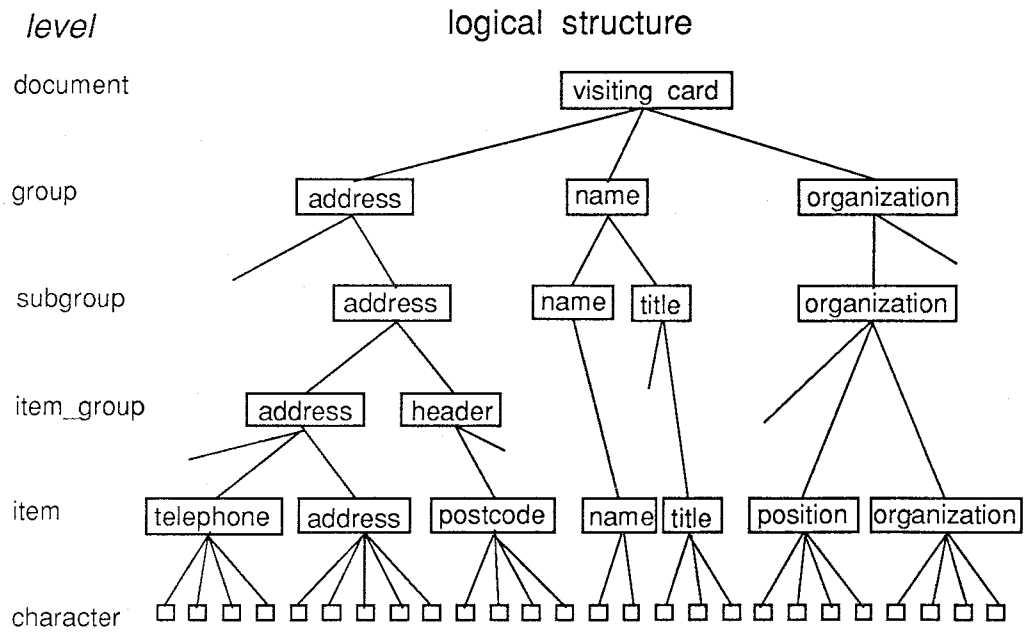

level
logical structure
document
group
subgroup
item_group
item
character
visiting_card
address
name
organization
address
name
title
organization
address
header
telephone
address
postcode
name
title
position
organization

図 3.2: 名刺の論理構造

表 3.1: 名刺の構成要素

レベル	構成要素
文書 (document)	名刺 (visiting_card)
群 (group)	住所群 (address group) 氏名群 (name group) 社名群 (organization group)
準群 (subgroup)	住所準群 (address subgroup) 氏名準群 (name subgroup) 肩書準群 (title subgroup) 社名準群 (organization subgroup)
項目群 (item_group)	見出し項目群 (header item_group) 住所項目群 (address item_group)
項目 (item)	見出し (header)[「本社」,「支社」など] 郵便番号 (postcode) 住所 (address) 電話番号 (telephone) ファックス (fax) テレックス (telex) 内線 (ext.) その他 (etc.)[ビル名など] 氏名 (name) 肩書 (title) 社名 (organization) 部署名 (position)

3.3 レイアウト構造の記述

レイアウト構造を知識として記述する際には，高い表現能力，記述容易性，可読性を得るために，以下の5点が求められると考えられる．

- レイアウト構造における階層性の明示的表現
- 処理とは独立なレイアウト構造の宣言的記述
- 記号的な記述
- 画像の絶対的な物理量に依存しない知識記述の実現
- レイアウト構造のバリエーションの簡潔な表現

以下，これらの点に着目した知識表現法について述べる．

3.3.1 記述形式

項目主体の文書においては，基本的に自然言語文の記述は存在しない．従って，2.2.1において述べたような，文字列や段組の展開規則は不必要であり，段組数の変動等による構造的バリエーションは存在しないと考えてよい．また，同様の理由により，レイアウト構造における階層関係(包含関係)の大部分は，論理構造における階層関係と一致すると考えられる．以上の点を考慮し，文書モデルでは，2.2.1の図 2.2に示したレイアウト構造に関する知識記述の概念を図 3.3に示すように単純化して考える．基本的な考え方は以下の通りである．

- 階層関係はレイアウト構造のものを記述し，論理構造のものと同一視する．
- 各レイアウト要素に構成要素に対するポインタを設けることにより，レイアウト構造と論理構造の対応規則を記述する．ただし，レイアウト要素には，論理構造の構成要素へのポインタを持たないものも許す．

このような基本概念に則り，以下では知識の具体的な記述形式について述べていく．本手法では，レイアウト構造の階層関係を明示的に記述するため，知識記述の枠組みとして，フレーム表現 [51] を採用する．図 3.4に示すように，個々のフレームは構成要素 (component) に対応するものであり，以下の5種類のスロットを介して，レイアウト構造を規定する．

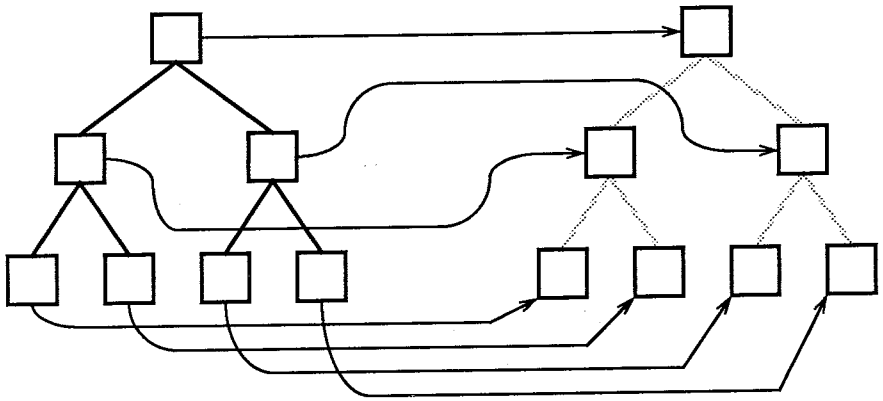


図 3.3: 項目主体の文書におけるレイアウト構造に関する知識

NAME 構成要素に割り当てられた名称 (構成要素名) を記述する

IS_A 構成要素と上位下位関係をなす上位フレーム名を記述する

SELF 各構成要素のみで定まる特徴を規定するための記述子を記述する

PART_OF 各構成要素と部分全体関係をなすサブフレーム名を記述する

SIMILARITY 各構成要素と類似差異関係をなすサブフレーム名を記述する

以下に PART_OF, SIMILARITY, IS_A のスロットに記述される部分全体, 類似差異, 上位下位の 3 種類の関係について説明する.

1. 部分全体関係
   部分全体関係は, 構成要素間に存在するレイアウト構造における階層関係 (包含関係) を規定するために設けられたものである. このとき, 部分と全体がどのようなレイアウト上の関係を持つのかという情報が重要であり, そうした情報を部分全体関係の意味記述としてもたせる必要がある. すなわち, 部分全体関係を単なるポインタとして表現するのではなく, 部分全体関係が他のフレームと同様にその内部にレイアウトを規

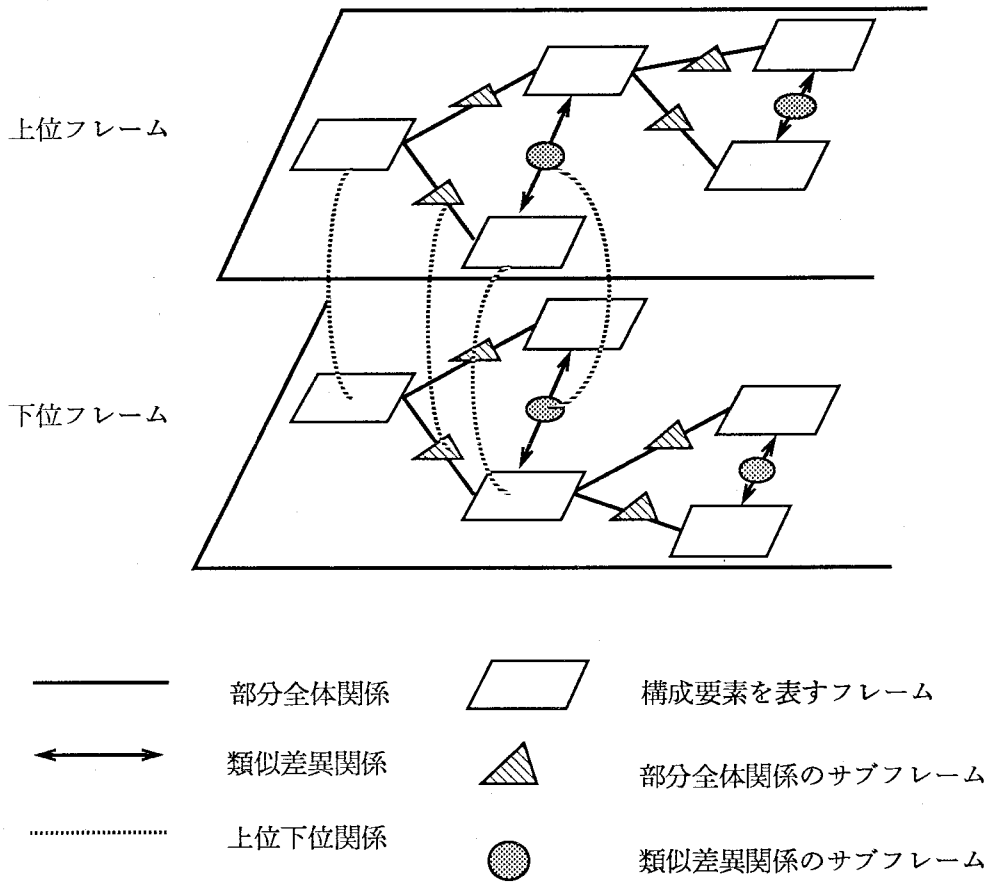


図 3.4: レイアウト構造のフレーム表現

定する記述子をもったデータ構造として表現されるべきである [52]. そこで本手法では，このようなデータ構造としてサブフレームと名付けたフレームを設け，サブフレーム名を PART_OF スロットの値として記述する．サブフレームには，以下のようなスロットを設ける．

NAME サブフレーム名を記述する

IS_A 各サブフレームと上位下位関係をなす上位サブフレーム名を記述する

RELATION レイアウトを規定する記述子を記述する

2. 類似差異関係
文書画像理解においては，文書画像から得られた特徴が何であるのかを識別・判断する基準が必要となる．そこで，部分全体関係に加え，異なる構成要素間の類似性や差異を表現する関係記述が望まれる [52]．本手法ではこのような関係として類似差異関係を導入し，独立した記述単位として部分全体関係と同様，サブフレームを用いて記述する．これにより，文書画像理解に必要な構成要素間の関係をコンパクトに表現することが可能となる．また，類似差異関係には，構成要素の配置を表現するため，相対位置を併せて記述する．

3. 上位下位関係
フレーム表現では，概念の上位下位関係に属性の継承という意味をもたせている．このことにより知識表現の重複が大幅に低減され，簡潔な表現が実現できる．上位下位関係を用いると，レイアウト構造のバリエーションを簡潔に表現可能であると考えられる．すなわち，レイアウト構造の各バリエーションに特殊な特徴を下位フレームに，共通の特徴を上位フレームに記述することにより，共通特徴の記述重複が回避できる．このような記述法は，2.2.1において述べた，レイアウト構造のクラス記述という概念とも整合性が高いと考えられる．

SELF スロットやサブフレームにおける記述子としては，次項で示すレイアウト述語を用いる．以上の記述形式をまとめると次のようになる．

```
(FRAME :NAME <構成要素名>
       :IS_A <上位フレーム名>
       :SELF <レイアウト述語>
       :PART_OF <部分全体関係をなすサブフレーム名>
       :SIMILARITY <類似差異関係をなすサブフレーム名>)

(SUBFRAME :NAME <部分全体関係または類似差異関係をなすサブフレーム名>
          :IS_A <上位サブフレーム名>
          :RELATION <レイアウト述語>)
```

3.3.2 レイアウト述語

文書の定型的なレイアウト構造には，センタリング，インデントなど記号的に表現可能な側面が多分に存在する．記述容易性，可読性という観点から考えても，知識が可能な限り記号的に記述されていると有利な点が多い．一方，文書画像理解においては，例えば構成要素領域の面積に関する制約など，記号化が困難な知識も必要となる場合がある．

本論文では，これらの相反する要求を満たし，かつ統一的な扱いが可能な記述子として，レイアウト述語を提案する．レイアウト述語は，構成要素の矩形領域に関する特徴量を規定する基本述語，およびレイアウト構造の記号的な側面を規定する複合述語からなる．基本述語を用いると，特徴量レベルでの記述が可能となるため，微視的なレイアウトを記述することが可能となる．一方，複合述語を用いると，センタリングなどの巨視的なレイアウトを簡潔に記述することが可能となる．複合述語は，基本述語を用いて定義される一種のマクロであるため，知識を利用する際には，マクロを展開することにより，統一的な扱いが可能となる．

レイアウト述語を定義する準備として，まず矩形領域の特徴量を定義する．本論文では，矩形領域を，図 3.5に示すように画像の原点を左上部とし，横方向にI軸，縦方向にJ軸をとる座標系を用いて，左上の座標 (I_s, J_s) と右下の座標 (I_e, J_e) の組で以下のように表す．

$$r = (I_s, I_e, J_s, J_e) \tag{3.1}$$

矩形領域の座標を用いると，面積等の基本的な特徴量 (以後，基本特徴量と呼ぶ) は，表 3.2に示すように定義できる．ここで正方形度は，矩形領域が正方形に近い度合を表す指標であり，矩形領域が正方形に近づくほど 0 に近

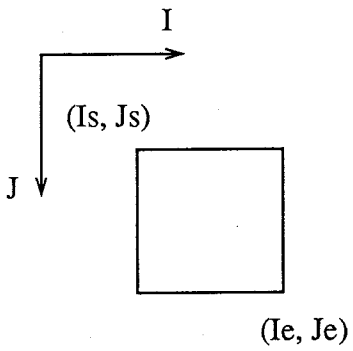


図 3.5: 矩形領域

づくものである．なお，正方形度が負の値をとるときは矩形が縦長であることを示し，逆に正の値をとるときには横長であることを示す．

表 3.2において正方形度を除く基本特徴量は，画像の大きさに依存しているため，このままではレイアウト構造の記述には適さない．絶対的な物理量に依存しない特徴量を定義するためには，基本特徴量を正規化する必要がある．まず，部分全体関係を規定する特徴量としては，より“部分”に近い矩形領域の基本特徴量を，より“全体”に近い矩形領域の基本特徴量により正規化するという方法が最も自然であろう．ここでは，そのような特徴量として正規化特徴量を定義する．下位レベルの矩形領域 r，上位レベルの矩形領域 b を，

$$r = (I_s, I_e, J_s, J_e) \tag{3.2}$$

$$b = (I_s^b, I_e^b, J_s^b, J_e^b) \tag{3.3}$$

ただし，b は r を包含する

としたときの定義を表 3.3に示す．

類似差異関係を規定する特徴量に対しては，類似性あるいは差異が表現可能であること，および正規化特徴量の場合と同様の理由から，画像の物理量に依存しないことが求められる．本手法では，このような特徴量として表 3.4に示す比較特徴量を定義する．ここで，比較対象の矩形領域 r_a,r_b を，

$$r_a = (I_{sa}, I_{ea}, J_{sa}, J_{ea}) \tag{3.4}$$

表 3.2: 基本特徴量

特徴名	記号	構成要素
横幅 horizontal_width	$D_i(r)$	$I_e - I_s + 1$
縦幅 vertical_width	$D_j(r)$	$J_e - J_s + 1$
面積 area	$A(r)$	$D_i \cdot D_j$
中心 center	$C(r)$	$(C_i(r), C_j(r))$
	$C_i(r)$	$(I_s + I_e)/2$
	$C_j(r)$	$(J_s + J_e)/2$
正方形度 square_degree	$S(r)$	$D_i(r) > D_j(r)$ のとき:$1 - (D_j/D_i)$ その他: $(D_i/D_j) - 1$

表 3.3: 正規化特徴量

特徴名	記号	構成要素
正規化横幅 normalized_horizontal_width	$D_{in}(r,b)$	$D_i(r)/D_i(b)$
正規化縦幅 normalized_vertical_width	$D_{jn}(r,b)$	$D_j(r)/D_j(b)$
正規化面積 normalized_area	$A_n(r,b)$	$A(r)/A(b)$
正規化中心 normalized_center	$C_n(r,b)$	$(C_{in}(r,b), C_{jn}(r,b))$
normalized_horizontal_center	$C_{in}(r,b)$	$2(C_i(r) - C_i(b))/D_i(b)$
normalized_vertical_center	$C_{jn}(r,b)$	$2(C_j(r) - C_j(b))/D_j(b)$
正規化座標 normalized_coordinate	$E(r,b)$	$(E_{is}(r,b), E_{ie}(r,b),$ $E_{js}(r,b), E_{je}(r,b))$
normalized_coordinate_E_{is}	$E_{is}(r,b)$	$\mid I_s - I_s^b \mid /D_i(b)$
normalized_coordinate_E_{ie}	$E_{ie}(r,b)$	$\mid I_e - I_e^b \mid /D_i(b)$
normalized_coordinate_E_{js}	$E_{js}(r,b)$	$\mid J_s - J_s^b \mid /D_j(b)$
normalized_coordinate_E_{je}	$E_{je}(r,b)$	$\mid J_e - J_e^b \mid /D_j(b)$

表 3.4: 比較特徴量

特徴名	記号	構成要素
比較横幅 compared_horizontal_width	$D_{ic}(r_a, r_b)$	$\frac{D_i(r_a)-D_i(r_b)}{max(D_i(r_a),D_i(r_b))}$
比較縦幅 compared_vertical_width	$D_{jc}(r_a, r_b)$	$\frac{D_j(r_a)-D_j(r_b)}{max(D_j(r_a),D_j(r_b))}$
比較面積 compared_area	$A_c(r_a, r_b)$	$\frac{A(r_a)-A(r_b)}{max(A(r_a),A(r_b))}$
比較中心 compared_center	$C_c(r_a, r_b)$	$(C_{ic}(r_a, r_b), C_{jc}(r_a, r_b))$
compared_horizontal_center	$C_{ic}(r_a, r_b)$	$2(C_i(r_a) - C_i(r_b))/D_i(b)$
compared_vertical_center	$C_{jc}(r_a, r_b)$	$2(C_j(r_a) - C_j(r_b))/D_j(b)$
比較座標 compared_coordinate	$P(r_a, r_b)$	$(P_{is}(r_a, r_b), P_{ie}(r_a, r_b),$ $P_{js}(r_a, r_b), P_{je}(r_a, r_b))$
compared_coordinate_P_{is}	$P_{is}(r_a, r_b)$	$\frac{2(I_{sa}-I_{sb})}{(D_i(r_a)+D_i(r_b))}$
compared_coordinate_P_{ie}	$P_{ie}(r_a, r_b)$	$\frac{2(I_{ea}-I_{eb})}{(D_i(r_a)+D_i(r_b))}$
compared_coordinate_P_{js}	$P_{js}(r_a, r_b)$	$\frac{2(J_{sa}-J_{sb})}{(D_j(r_a)+D_j(r_b))}$
compared_coordinate_P_{je}	$P_{je}(r_a, r_b)$	$\frac{2(J_{ea}-J_{eb})}{(D_j(r_a)+D_j(r_b))}$

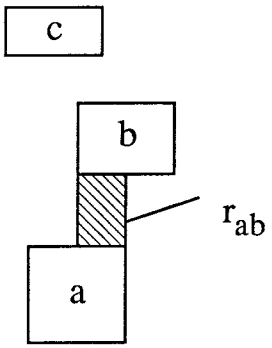


図 3.6: 相対位置

$$r_b = (I_{sb}, I_{eb}, J_{sb}, J_{eb}) \tag{3.5}$$

としている．

さらに，構成要素間の位置関係を表すために，相対位置を定義する．2.2.1でも述べたように，文書画像は，縦あるいは横方向に並ぶ矩形領域から構成されているとみなすことができる．このような前提のもとでは，一般の図形を対象とした場合のように，斜めの位置関係にある要素間の記述が意味を持つことは非常にまれであるといえる．そこで，レイアウト構造を記述するための相対位置として，UP，DOWN，RIGHT，LEFT の 4 方向のみを考える．以下，UP 方向を例に相対位置の定義を説明する．

定義 **1** *(相対位置)*
対象とする矩形領域の集合を R，R に属する 2 つの矩形領域 a，b を，

$$a = (I_{s1}, I_{e1}, J_{s1}, J_{e1}) \tag{3.6}$$

$$b = (I_{s2}, I_{e2}, J_{s2}, J_{e2}) \tag{3.7}$$

ただし，$J_{s1} > J_{e2}$

とする．$max(I_{s1}, I_{s2}) < min(I_{e1}, I_{e2})$ ならば，図 3.6に示すような，矩形領域 r_{ab} を考えること ができる． ここで，

$$r_{ab} = (max(I_{s1}, I_{s2}), min(I_{e1}, I_{e2}), J_{e2} + 1, J_{s1} - 1) \tag{3.8}$$

である．$r \in R$ なる全ての r に対して，矩形領域 r_{ab} と r が重複しないならば，

$$up(a, R) = b \tag{3.9}$$

とする．このとき，矩形領域 a を基準として，b を UP 方向にあるという．

図 3.6に示すように，矩形領域 b は矩形領域 a の UP 方向に位置するが，矩形領域 c は矩形領域 a の UP 方向には該当しない．なお，レイアウト構造に関する知識を記述する際には，対象とする矩形領域の集合 R は，同一レベルに属する構成要素の領域の集合となる．

以上のような矩形領域の特徴量の記述をする形式として，基本述語を設ける．基本述語は，プリミティブと指定子の組により定義されるものである．

プリミティブは，正規化特徴量，比較特徴量，正方形度などの矩形の特徴量(以後，矩形特徴量と呼ぶ)，構成要素の数を記述する構成要素数(number_of)，および相対位置(position)から構成される．プリミティブとスロットの対応を表 3.5に示す．

指定子には，方向指定子，数量指定子，程度指定子の3種類を設ける．相対位置(position)を記述する場合には方向指定子を，矩形特徴量を記述する場合には程度指定子を，構成要素数(number_of)を記述する場合には数量指定子をそれぞれ用いる．以下に，各指定子について説明する．

方向指定子 方向指定子は，相対位置を表すプリミティブ position に対して使用されるものであり，UP，DOWN，LEFT，RIGHT の4方向を指定するものである．

程度指定子 程度指定子は，矩形特徴量のプリミティブに対して使用されるものであり，プリミティブごとの矩形特徴量の値域を区間値 $[x, y]$ を用いて指定する．ここで，$[MIN, MAX]$ を対象とする矩形特徴量の値域とするとき，x,y は，

$$x \leq y, \quad MIN \leq x, \quad y \leq MAX \tag{3.10}$$

の条件を満たす実数である．程度指定子は，区間値内の実数すべてを表す．

数量指定子 数量指定子は，部分全体関係において構成要素数を記述するプリミティブ number_of に対して使用されるものであり，取り得る個数の範囲を区間値 $[m, n]$ を用いて指定する．ここで，m, n は，$m \leq n$ を満たす0以上の整数，あるいは $n = \infty$ である．数量指定子は，区間値内の整数すべてを表す．

表 3.5: プリミティブ

スロット	プリミティブ
SELF	square_degree number_of
PART_OF の RELATION	normalized_horizontal_width normalized_vertical_width normalized_area normalized_center normalized_horizontal_center normalized_vertical_center normalized_coordinate normalized_coordinate_E_{is} normalized_coordinate_E_{ie} normalized_coordinate_E_{js} normalized_coordinate_E_{je}
SIMILARITY の RELATION	compared_horizontal_width compared_vertical_width compared_area compared_center compared_horizontal_center compared_vertical_center compared_coordinate compared_coordinate_P_{is} compared_coordinate_P_{ie} compared_coordinate_P_{js} compared_coordinate_P_{je} position

表 3.6: 複合述語

スロット	複合述語
PART_OF	horizontal_centering vertical_centering right_edge left_edge upper_end bottom_end
SIMILARITY	horizontal_alignment vertical_alignment vertical_right_indention vertical_left_indention

基本述語の記述形式を以下にまとめる．

$$\begin{aligned} <\text{基本述語}> \quad ::== \quad & \{<\text{矩形特徴量}><\text{程度指定子}>\} \mid \\ & \{\text{position} <\text{方向指定子}>\} \mid \\ & \{\text{number\_of} <\text{数量指定子}>\} \end{aligned}$$

さて，人間は文書画像を見たとき，まず“センタリングされている”であるとか“文字列が揃っている”などのより本質的なレイアウトを意識していると考えられる．同じように，文書画像のレイアウト構造を記述する場合にも本質的なレイアウト構造の表現を許す記述形式とすることが望まれる．これにより，知識ベース作成の際には知識の記述が容易になると考えられる．そこで，レイアウトの本質を表すような述語を設ける．このような述語は，基本述語あるいはその否定の連言として定義されるものであるため，複合述語と呼ぶ．表 3.6にフレームのスロットごとに分類した複合述語の一覧を示し，個々の定義を表 3.7に示す． ここで，基本述語，および複合述語の否定を，

$$not(<\text{基本述語}>)$$

$$not(<\text{複合述語}>)$$

という形式により記述している．本論文では，以上の基本述語と複合述語をまとめてレイアウト述語と呼ぶ．

表 3.7: 複合述語の定義

複合述語	定義
horizontal_centering	{normalized_horizontal_center [-1/9,1/9]}
vertical_centering	{normalized_vertical_center [-1/9,1/9]}
upper_end	{normalized_coordinate_E_{js} [-1,-7/9]}
bottom_end	{normalized_coordinate_E_{je} [-1,-7/9]}
left_edge	{normalized_coordinate_E_{is} [-1,-7/9]}
right_edge	{normalized_coordinate_E_{ie} [-1,-7/9]}
horizontal_alignment	and({compared_coordinate_P_{js} [-0.4,0.4]}, {compared_coordinate_P_{je} [-0.4,0.4]})
vertical_alignment	and({compared_coordinate_P_{is} [-0.4,0.4]}, {compared_coordinate_P_{ie} [-0.4,0.4]}
vertical_right_indention	and({compared_coordinate_P_{js} [0, ∞]}, not(horizontal_alignment))
vertical_left_indention	and({compared_coordinate_P_{js} [$-\infty$,0]}, not(horizontal_alignment))

3.3.3 記述の解釈

記述された知識を実際に使用するためには，記述の解釈を定める必要がある．基本的には，レイアウト述語は真，偽いずれかの解釈を持つものである．解釈について具体的に説明する前に，準備として以下に示す事項を定義する．

1. 閉区間を表す区間値の表記法 $[a,b]$ に加え，開区間を表す表記法 (a,b) を導入する．

2. v を区間値とするとき，関数 $inf(v)$,$sup(v)$ を次のように定義する．v を $[a,b]$, $[a,b)$, $(a,b]$, (a,b) により表される区間値のいずれかとすると，

$$inf(v) = a, \quad sup(v) = b \tag{3.11}$$

である．

3. 区間値に対して，次のような演算を定義する．

$$[a_1,b_1] \cap [a_2,b_2] = [max(a_1,a_2), min(b_1,b_2)] \tag{3.12}$$

この演算は，2 つの区間の重複する部分を求めるものである．区間の表記法が混在する場合においても，同様の観点から定義することができる．例えば，

$$[a_1,b_1] \cap [a_2,b_2) = \begin{cases} [max(a_1,a_2), b_1] & b_1 < b_2 \text{ のとき} \\ [max(a_1,a_2), b_2) & \text{それ以外} \end{cases} \tag{3.13}$$

となる．

4. position 以外のプリミティブ *primitive* を持つ基本述語

$$BP = \{primitive\ v\} \tag{3.14}$$

において，$inf(v) > sup(v)$ のとき，BP を偽とする．また，区間値の表記に開区間が含まれるならば，$inf(v) = sup(v)$ のとき，BP を偽とする．

5. 基本述語 $BP_1, \ldots, BP_n$ の選言を以下のように記述する．

$$or(BP_1, BP_2, \ldots, BP_n) \tag{3.15}$$

この式は $BP_1, \ldots, BP_n$ のいずれかが真ならば真となる．

以上の準備のもと，基本述語から順に記述の解釈について述べる．

基本述語の解釈

基本述語は，画像から得られた特徴量に応じて，真偽いずれかの値を持つ．

1. プリミティブが position の場合
   矩形領域間の相対位置が方向指定子に記述された相対位置と一致する場合には真，それ以外は偽となる．

2. プリミティブが矩形特徴量の場合
   矩形領域の特徴量が程度指定子に記述された区間値を満たす場合には真，それ以外は偽となる．

3. プリミティブが number_of の場合
   構成要素の個数が数量指定子に記述された区間値を満たす場合には真，それ以外は偽となる．

また，基本述語 BP の否定 $not(BP)$ については，BP が真のとき偽となり，BP が偽のとき真となる．指定子を考慮すると，否定の解釈は以下のようになる．

1. プリミティブが position の場合

$$not(\{\text{position } p\}) = or(\{\text{position } q_1\}, \{\text{position } q_2\}, \{\text{position } q_3\}, \{\text{position } q_4\}) \quad (3.16)$$

ただし，$q_i \in \{\text{UP}, \text{DOWN}, \text{RIGHT}, \text{LEFT}, \text{NULL}\}$,
$q_i \neq p$, $i \neq j$ ならば $q_i \neq q_j$

ここで，NULL とは相対位置の関係が成立しない場合を指す．

2. プリミティブが矩形特徴量 *feature* の場合

$$not(\{feature\ v\}) = or(\{feature\ [MIN, inf(v))\}, \{feature\ (sup(v), MAX]\}) \quad (3.17)$$

3. プリミティブが number_of の場合

$$not(\{\text{number\_of } v\}) = or(\{\text{number\_of } [0, inf(v) - 1]\}, \{\text{number\_of } [sup(v) + 1, \infty]\}) \quad (3.18)$$

複合述語の解釈

複合述語は基本述語の連言として解釈される．また複合述語 CP を，

$$CP = and(BP_1, \ldots, BP_n) \tag{3.19}$$

とするとき，その否定 $not(CP)$ は，

$$\begin{aligned} not(CP) &= not(and(BP_1, \ldots, BP_n)) \\ &= or(not(BP_1), \ldots, not(BP_n)) \end{aligned} \tag{3.20}$$

と同値である．

基本述語が存在しない場合の解釈

複合述語を基本述語に変換後，基本述語の集合を考え，その集合に存在しない基本述語については，

1. プリミティブが position の場合
   {position NULL} と同値である．

2. プリミティブが矩形特徴量 *feature* の場合
   $\{feature \ \ [MIN, MAX]\}$ と同値である．すなわち，どのような値をとる場合にも真となる，ここで，$[MIN, MAX]$ は *feature* の値域である．

3. プリミティブが number_of の場合
   {number_of $[0, \infty]$} と同値である．すなわち，どのような値をとる場合にも真となる．

とする．

基本述語の重複に対する解釈

複合述語を基本述語に変換する際に，同一のプリミティブに対する記述が現われる場合がある．このときの解釈は，

1. プリミティブが position の場合

$$and(\{\text{position } v_1\}, \{\text{position } v_2\}) = \begin{cases} \text{偽} & v_1 \neq v_2 \text{ のとき} \\ \{\text{position } v_1\} & v_1 = v_2 \text{ のとき} \end{cases} \tag{3.21}$$

その他の場合は，一般の論理式と同様である．

$$\text{例)} \quad and(\{\text{position } v_1\}, or(\{\text{position } v_1\},\{\text{position } v_2\})) = \{\text{position } v_1\}$$

2. プリミティブが矩形特徴量 *feature* の場合

$$and(\{feature\ v_1\},\{feature\ v_2\}) = \{feature\ v_1 \cap v_2\} \tag{3.22}$$

3. プリミティブが number_of の場合

$$and(\{\text{number\_of } v_1\},\{\text{number\_of } v_2\}) = \{\text{number\_of } v_1 \cap v2\} \tag{3.23}$$

is_a relation の解釈

全ての複合述語を基本述語に変換した後を考えると，以下の 2 つの場合が存在する．

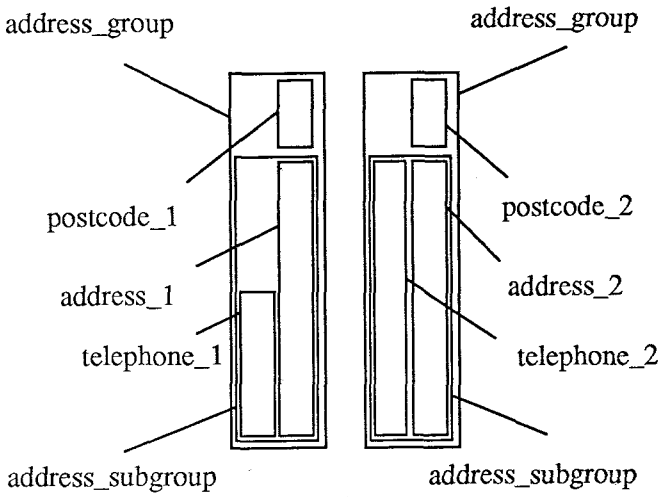
- あるプリミティブに対する基本述語が上位フレームに記述されており，下位フレームに記述されていない場合
  上位フレームの基本述語を継承する．
- 同一のプリミティブに対する基本述語が上位フレーム，下位フレーム双方に記述されている場合
  下位フレームの基本述語を優先する．

3.3.4 記述例と考察

ここでは，前述の形式により記述される知識の具体例を挙げ，その後，他手法との比較検討を行う．

記述例

図 3.7 (a) に示した address_1 と address_2 に対する記述を図 3.7 (b) に示す．以下に，各番号に対応する記述の意味を列挙する．




```
(FRAME    :NAME 《address_1》
          :SELF  {number_of [1,1]}                    1
                 {square_degree [-1,-1/3]}            2
          :PART_OF
           《address_1-par.-address_subgroup》
          :SIMILARITY
           《address_1-sim.-telephone_1》
           《address_1-sim.-postcode_1》 )

(SUBFRAME :NAME
          《address_1-par.-address_subgroup》
          :RELATION  right_edge                       3
                     bottom_end)

(SUBFRAME :NAME 《address_1-sim.-telephone_1》
          :RELATION
            {position right}                          4
           compared_vertical_width [1/3,5/9]}) 5

(SUBFRAME :NAME 《address_1-sim.-postcode_1》
          :RELATION
         {compared_horizontal_width [-1/9,1/9]}
         {position up})
```

```
(FRAME    :NAME   《telephone_1》
          :SELF    {number_of [1,1]}
                   {square_degree [-7/9,-1/9]}
          :PART_OF
           《telephone_1-par.-address_subgroup》
          :SIMILARITY
           《telephone_1-sim.-address_1》
           《telephone_1-sim.-postcode_1》 )

(SUBFRAME :NAME
           《telephone_1-par.-address_subgroup》
          :RELATION  bottom_end
                     left_edge)
```

```
(FRAME    :NAME   《postcode_1》
          :SELF    {number_of [1,1]}
                   {square_degree [-7/9,-1/9]}
          :PART_OF
           《postcode_1-par.-address_group》
          :SIMILARITY
           《postcode_1-sim.-address_1》
           《postcode_1-sim.-telephone_1》

(SUBFRAME  :NAME
            《postcode_1-par.-address_group》
           :RELATION  right_edge
                      upper_end)
```

図 3.7: レイアウト構造に関する知識の記述例

```
(FRAME    :NAME   《address_2》
          :IS_A   《address_1》                    6
          :PART_OF
           《address_2-par.-address_subgroup》
          :SIMILARITY
           《address_2-sim.-telephone_2》
           《address_2-sim.-postcode_2》 )

(SUBFRAME  :NAME
        《address_2-par.-address_subgroup》
           :IS_A
        《address_1-par.-address_subgroup》 )

(SUBFRAME  :NAME  《address_2-sim.-telephone_2》
           :IS_A  《address_1-sim.-telephone_1》
           :RELATION
            horizontal_alignment)

(SUBFRAME  :NAME
            《address_2-sim.-postcode_2》
           :IS_A
            《address_1-sim.-postcode_1》 )
```

```
(FRAME    :NAME   《telephone_2》
          :IS_A   《telephone_1》
                   {square_degree [-1,-1/3]}
          :PART_OF
           《telephone_2-par.-address_subgroup》
          :SIMILARITY
           《telephone_2-sim.-address_2》
           《telephone_2-sim.-postcode_2》 )

(SUBFRAME :NAME
           《telephone_2-par.-address_subgroup》
          :IS_A
           《telephon_1-par.-address_subgroup》 )
```

```
(FRAME    :NAME   《postcode_2》
          :IS_A   《postcode_1》
          :PART_OF
           《postcode_2-par.-address_group》
          :SIMILARITY
           《postcode_2-sim.-address_2》
           《postcode_2-sim.-telephone_2》

(SUBFRAME  :NAME
           《postcode_2-par.-address_group》
           :IS_A
           《postcode_1-par.-address_group》 )
```

図 3.7: レイアウト構造に関する知識の記述例 (つづき)

1. 構成要素 address_1 が 1 つ存在することを示す.

2. address_1 の縦幅が横幅の 3 倍以上であることを示す. これは, address_1 に 3 文字以上の文字が含まれていなければならないことを意味する.

3. address_1 は address_sub_group の右端に存在していることを示す.

4. address_1 は tel_1 の右側に存在していることを示す.

5. address_1 の縦幅が tel_1 の縦幅の約 2 倍であることを示す.

6. address_2 の上位フレームが address_1 であり address_1 から属性が継承可能であることを示す.

他手法との比較

以上のような記述手法は, これまでに提案された記述手法に比べ, 次のような特徴をもつ.

1. Niyogi らが用いているルール形式の記述 [42] では, 文書のレイアウト構造の特徴である階層性を明示的に表現できないが, 本手法ではフレーム形式を用いることにより, 階層性を明確に記述できる構造的な表現となる.

2. 東野ら [43] や中野ら [48] は矩形領域の記述に座標を直接用いているため, 記述が対象画像の物理的な大きさに依存しているが, 本手法では基本述語を用いることにより画像の大きさに依存しない記述が可能である. また, 複合述語を用いることにより記述容易性, 可読性を高めることが可能である.

3. 駱らの手法 [45] や Dengel らの手法 [46] では, 画像分割オペレータの種類, あるいは適用時のパラメータによりレイアウト構造を表現しているため, 記述容易性, 可読性に問題があると考えられる. また, 表現されているものは, オペレータの適用順序, すなわち画像分割手続きであるため, 本質的なレイアウト構造を表現しているとは考え難い. 本手法では, 処理とは無関係なレイアウト述語によりレイアウト構造を表現しているため, より記述容易性, 可読性が高く, 加えてレイアウト構造の本質を直接表現することが可能となる.

4. 上記の2つの手法では，AND/OR木により画像分割オペレータの適用順序をにより表現している．このとき，レイアウト構造のバリエーションはORノードにより表現される．このような手続き的な表現では，どの構成要素にバリエーションが存在するかを陽に表現することが困難となるため，知識の可読性に問題があると考えられる．また，AND/OR木の根に近い部分においてORノードが存在し，その下に続く部分木の大部分が同一の場合，知識記述が冗長となってしまう．一方，本手法では，構成要素に着目したバリエーションの表現方法を採用することにより，どの構成要素にバリエーションが存在するかを明示的に記述することが可能となる．また，フレームにおける上位下位関係や数量指定子を利用することによって，記述の重複を最小限にとどめ，簡潔な知識表現が可能となる．

以上の考察から，本手法は，従来手法に比べて，表現能力，記述容易性，可読性の高い知識記述法であると考えられる．

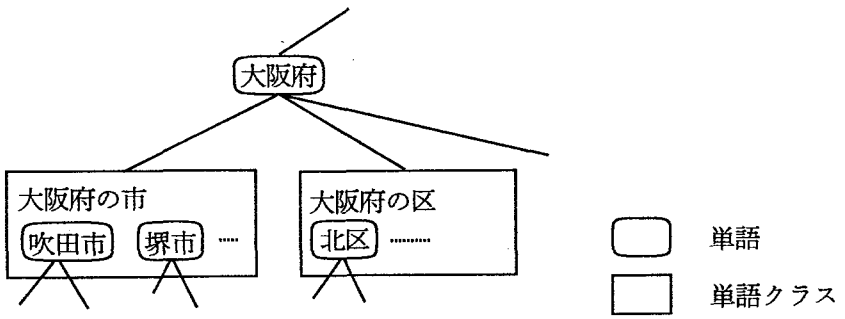
3.4 論理的制約の記述

3.4.1 論理的制約

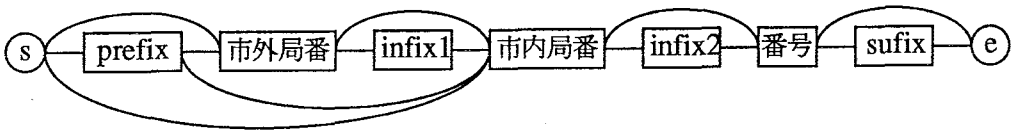
第2章でも述べたように，論理的制約とは文書の記述内容が妥当であるための必要条件である．本手法の対象とする文書は，新聞，本などの文章が主体となるものではなく，帳票，論文の表紙，名刺など項目が主体となるものである．項目の記述内容は文章ではないため，いわゆる自然言語の文法規則などを用いて妥当性を表現することが困難となる．項目主体の文書では，項目を意味的に接続する単語の並び(以後，単語列と呼ぶ)ととらえることができ，項目より上位レベルの構成要素を単語列の集合としてとらえることができる．本手法では，項目に対しては単語の接続性，項目より上位レベルの構成要素に対しては単語列の整合性により記述内容の妥当性を表現する．以下では，各々の詳細について述べる．

3.4.2 単語の接続性

項目中に存在する単語には，項目の属性に応じた偏りがある．例えば，名刺の住所には，地名，番地の単語のみが存在する．そこで本手法では，項目別に作成した単語辞書により，項目に対する記述内容の妥当性を表現する．項目別の各単語辞書には，項目に存在可能な単語を記述するほか，単語間に



(a) 単語ー単語クラス



prefix：「電話」
infix1：「ー」(ハイフン)
infix2：「ー」(ハイフン)
suffix：「番」

(b) 単語クラスー単語クラス

図 3.8: 単語間の接続性記述

接続性を記述することにより，項目として妥当な単語列を表現する．なお，特別な接続性として，項目の始端，終端に接続する単語については，その区別を記述する．

接続性の記述に際しては，項目に属する個々の単語すべてに接続性を記述すると非常に多数となり，効率的とはいい難い．ここで，単語間の接続性の性質について考えると，次のようなことがわかる．接続性は，単語間に無秩序に存在するのではなく，ある特定の種類の単語が別の種類の単語に接続するという傾向が存在する．そこで本手法では，そのような単語の種類を単語クラスにより表し，単語と単語クラス，あるいは単語クラスと単語クラスを対象とした接続性記述を許す．前者の例を図3.8 (a) に，後者の例を図3.8 (b) に示す．図中のリンクは接続性があることを示すものであり，接続性リンク

と呼ぶ．一般に住所の単語列は，単語と単語クラスの接続性により規定され，全体として木構造をなす．住所以外の項目に含まれる単語列は，単語クラスと単語クラスの接続性により規定される．

項目に含まれる単語列には，上記の単語クラスのいくつかを省略しても妥当となる場合がある．例えば，住所の単語列では，都道府県の単語クラスが省略される場合があり，また，電話番号でも，市外局番が省略される場合がある．このような単語列を規定するために，単語あるいは単語クラスには複数の接続性リンクの記述を許す．

また，単語クラスには，他の単語クラスの出現を制約するものがある．例えば，電話番号等において開き括弧「(」がある場合，必ず閉じ括弧「)」が出現しなければならない．そこで，本手法では，妥当な単語列となるための，上記のような制約条件を単語クラス間に記述する．

住所，社名，部署名などの項目では，単語辞書には具体的な単語が記述される．しかしながら，電話番号など数字が主体の項目では，番号部分にすべての可能な単語，すなわち数字列を記述するのは，効率の面から不適当であると考えられる．そこで，本手法では，数字列の記述に文字クラスという概念を導入する．具体的には，文字クラスとは，可能な文字カテゴリ，文字種を表現したものである．たとえば，電話番号の4桁の番号は，

「＜漢数字＞＜漢数字＞＜漢数字＞＜漢数字＞」

と表現され，また2桁の市外局番では，

「0＜0以外の漢数字＞」

のように記述される．ここで，<> は文字クラスであり，一文字に相当する．

3.4.3 単語列の整合性

項目より上位レベルの構成要素は，項目の集合であると考えられる．記述内容の性質としては，項目に含まれる単語列に関して，妥当な組合せが制限されることを挙げることができる．例えば，肩書準群(複数の肩書を含む構成要素)に，「部長」と「課長」という2つの肩書が同時に存在することはないまた，住所に含まれる単語列「大阪府吹田市山田丘」と妥当な組合せである郵便番号は，「565」のみといえる．本論文では，このような関係を単語列の整合性と呼ぶ．

単語列の整合性は，文書の記述内容に対してグローバルに定義できるものではなく，単語列がどの構成要素に属するかに応じて，変化する場合がある．

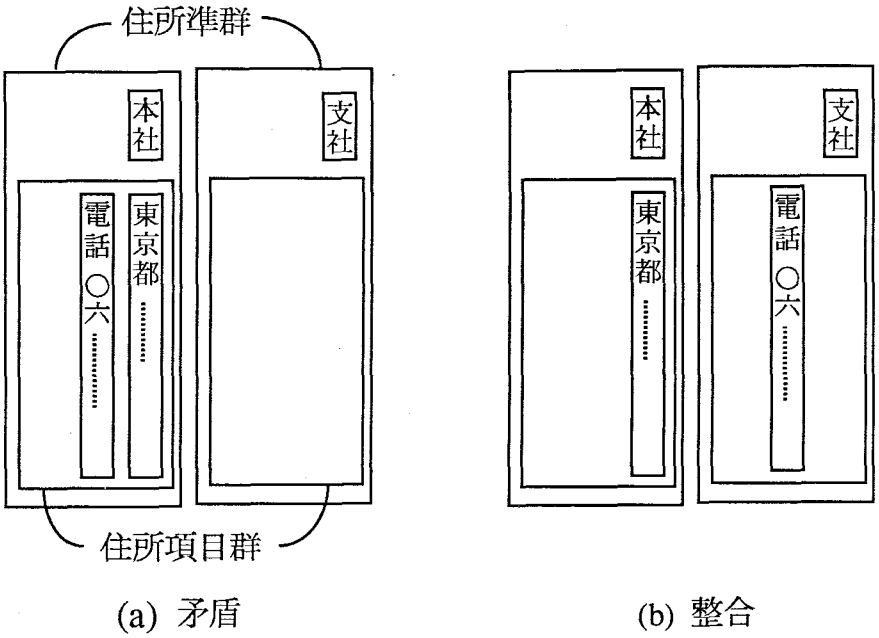


図 3.9: 単語列の整合性

例えば，図 3.9 (a) に示すように，同じ住所項目群に住所の「東京都」と市外局番の「〇六」が存在する場合は矛盾となるが，同図 (b) のように異なる住所項目群に存在する場合は，別の所在地に対するものであるため矛盾にはならない．そこで本手法では，単語列の整合性を構成要素ごとに記述する．

実際の記述に際しては，構成要素内に含まれる項目の対ごとに，整合を記述するか，逆に矛盾を記述するかを選択する．ある項目の対について，矛盾が記述されている場合には，記述に存在しないものは整合すると解釈され，また整合が記述されている場合には，記述に存在しないものは矛盾すると解釈される．このことから，矛盾の記述に比べて，整合の記述の方がより制約としては強いものであるといえ，文書画像理解にはより有効となる．ただし，実際には，住所－郵便番号間などの数例を除いては，整合する組合せを網羅することは困難であるため，整合の記述を採用している．また，整合性の記述が存在しない項目の対については，互いがどのような記述内容であっても整合するものとみなす．

表 3.8: 単語例

辞書名	単語クラス	単語例
社名	社名 1	「株式会社」
	社名 2	「東芝」
	社名 3	「エンジニアリング」
	社名 4	「株式会社」
部署名	部署名 1	「第一」
	部署名 2	「営業」，「システム」
	部署名 3	「第一」
	部署名 4	「部」，「課」
肩書	肩書 1	「システム」
	肩書 2	「部長」，「課長」
	肩書 3	「社長」
郵便番号	prefix1	「（」
	prefix2	「郵便番号」，「〒」
	番号 1	「<漢数字><漢数字><漢数字>」
	infix	「｜」（ハイフン）
	番号 2	「<アラビア数字><アラビア数字>」
	sufix	「）」

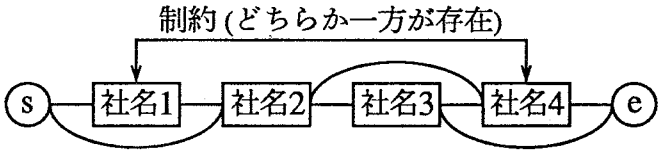
3.4.4 記述例と考察

レイアウト構造の場合と同様にまず記述例を示し，次に記述例を通して手法の有効性について考察する．

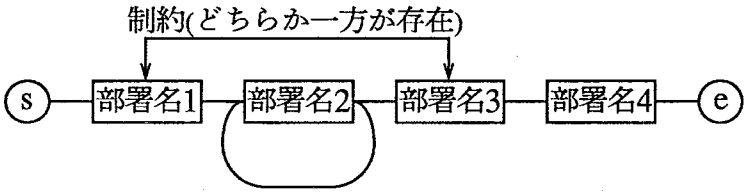
記述例

単語の接続性，単語列の整合性のそれぞれについて，例を以下に挙げる．

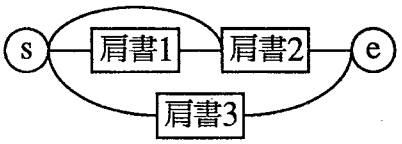
1. 単語の接続性
   単語，単語クラス，および接続姓により表現される単語辞書の構造を，代表的な項目について図 3.10に示す．また，単語クラスに属する単語例を表 3.8に示す．なお，住所，電話番号については，図 3.8を参照されたい．名刺の項目のうち，氏名に対する単語辞書は設定していない．これは，単語の接続性により妥当な氏名を表現することが，非常に困難



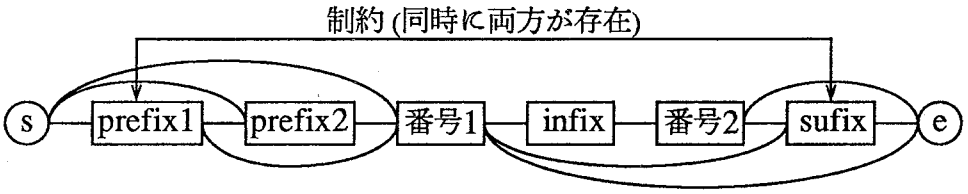
(a) 社名辞書



(b) 部署名辞書



(c) 肩書辞書



(d) 郵便番号辞書

図 3.10: 単語辞書の構造

表 3.9: 単語列の整合性

構成要素	対象項目		種類	例	
住所項目群	住所	電話番号	整合	「大阪府吹田市」	市外局番「０６」
	住所	郵便番号	整合	「大阪府吹田市山田丘」	「５６５」
	電話番号	郵便番号	整合	「５６５」	市外局番「０６」
肩書順群	肩書	肩書	矛盾	「課長」	「社長」
	部署名	部署名	矛盾	「本社」	「支店」
社名群	肩書	肩書	矛盾	(肩書順群と同様)	
	部署名	肩書	矛盾	「課」	「部長」
住所群	見出し	見出し	矛盾	「本社」	「本社」
名刺	部署名	肩書	矛盾	(社名群と同様)	

であるという理由による．従って，氏名に関しては，どのような単語列に対しても妥当であるとして扱う．

2. 単語列の整合性
単語列の整合性は，すべての構成要素に関して定義できるとは限らない．名刺において整合性が定義できる構成要素，記述対象となる項目の対，およびその例を表 3.9に示す．ここで，住所項目群に含まれる住所，郵便番号，市外局番は整合する組合せが限られているために整合の条件により記述しているが，その他は矛盾の条件により記述している．

考察

以上の論理的制約は，構成要素の記述内容が妥当であるための必要条件を規定したものに過ぎないため，論理的制約を満たす記述内容には，本来，名刺の記述内容としては妥当ではないものも含まれる．“部署名”に対する論理的制約を例にとると，接続性リンクをたどることにより，

「電子」-「情報」-「通信」-「システム」-....-「部」

など，実際には存在しない単語列を生成することが可能である．さらに，このような単語列は，“肩書”が「社長」など「部」に矛盾するものでなければ，他の単語列とも整合する．

このことは，単語の接続性および単語列の整合性が，名刺の記述内容の妥当性を完全に規定するには弱すぎることを示している．より精密に論理的制約を規定するためには，接続性，整合性の拡張，あるいは新たな規定方法が必要であると考えられる．

しかしながら，この制約は，妥当な記述内容の生成に用いるのではなく，文書画像理解に使用するものという立場に注意する必要がある．すなわち，文書画像理解においては，レイアウト構造に関する知識や論理的制約に関する知識，また文字の図形的特徴による認識など，複数の制約を同時に満たすものを結果として出力すればよく，論理的制約により完全に記述内容を規定できる必要はない．ただし，論理的制約の強さが，単語の接続性および単語列の整合性で表現できる範囲で十分かどうかは，文書画像理解という立場から実験的に検討する必要がある．

3.5 結言

本章では，レイアウト構造に関する知識，および記述内容の論理的制約に関する知識を記述する知識ベースとして文書モデルを提案し，記述例の考察，他手法との比較から，表現能力，記述容易性，可読性に関する本手法の有効性を示した．本手法の特徴を以下に列挙する．

1. レイアウト構造に関する知識の記述
    - フレーム表現を用いることにより，レイアウト構造の階層性を明示的に記述する
    - 基本述語を用いることにより，記号表現が困難なレイアウト構造を指定する
    - 複合述語を用いることにより，本質的なレイアウト構造を記号的に指定する
    - 上位下位関係の導入により，レイアウト構造のバリエーションをコンパクトに記述する
2. 論理的制約に関する知識の記述
    - 単語クラス，文字クラス，単語の接続性などの概念を導入することにより，項目として妥当な単語列をコンパクトに表現する
    - 単語列の整合性により，項目より上位レベルの構成要素に関しても，妥当な記述内容を表現する

次章以降では，本章で提案した知識ベースを有効利用する処理について述べ，実験から有効性を検証する．

第 4 章

文書構造解析

4.1 緒言

本論文における文書構造解析の目標は，レイアウト構造という側面から文書画像をとらえ，可能な構成要素の候補をすべて生成することにある．これは，項目以上のレベルにおいて，レイアウト構造に関する知識を満足する構成要素候補をすべて生成することに対応し，文字レベルにおいては，文字として妥当な領域の候補をすべて生成し，文字認識により属性を付与することに対応する．

第 2 章でも述べたように，項目レベル以上の文書構造解析については，様々な手法が提案されている．しかしながら，従来手法のほとんどは，構成要素候補の生成ではなく，構成要素を一意に抽出するものであるため，多数の候補が本質的に得られる場合においては，候補選択に関するヒューリスティックが処理に混在している．従って，上記の目標を実現するためには，従来手法において使用されている種々のヒューリスティックをすべて排除し，知識ベースである文書モデルのみを基準として候補を生成する手法が必要となる．加えて，候補を生成する場合には，候補間の依存関係を記述し，処理を無矛盾に保つ方法を考慮する必要がある．

文字切り出しについては，基本的には，文字列から文字を切り出すという従来手法と共通の枠組みによりとらえることができる．ただし，印刷文書に対する既存の文字切り出し手法は，文字の一定ピッチを仮定するものがほとんどであるため [2, 3]，名刺のように，文字の大きさやピッチが不定である文書には，適用が困難となる．この問題に対して，佐藤ら [53] は文字切り出しにおける人間の持つノウハウをアルゴリズムに反映させるという観点から統計的な手法を提案し，また遠城ら [54] は知識工学的観点からプロダクションシステムを導入し，解決を試みている．

本章では，以上の2点を考慮し，可能な候補をすべて生成するという観点から，新しい文書構造解析法を提案する．まず，項目レベル以上を対象とした候補生成処理では，文書モデルの記述を満たす候補をすべて生成する処理アルゴリズムを提案する．このとき，知識の階層性を利用し，生成対象をトップダウン的に仮定することにより，効率の高い処理を実現する．また，仮説の管理に重要となる依存関係の導出，記録方法についても述べる．

文字切り出し処理については，佐藤らの手法および遠城らの手法を融合することにより，より精度の高い文字切り出し処理を実現する [55, 56]．本手法は，人間を文字切り出しのエキスパートと考え，種々のプロダクションルールを用いることにより，人間の動作のシミュレートを目指すものである．プロダクションルールを用いることにより，画像の状態に応じた柔軟な処理が期待できることに加え，項目の属性に基づいて使用可能なルールを取捨選択することにより，項目の特性に応じた文字切り出し処理を実現する．

4.2 提案手法の概要

文書構造解析処理は，図 4.1に示すように，前処理，構成要素候補生成処理，文字切り出し・認識処理の3処理からなる．

前処理では，以後の処理に必要となるデータを文書画像から抽出する．このデータは，基本矩形と呼ぶ矩形領域であり，以後の処理では分割されることのない最小領域として扱われる．最小領域を設定することにより，以後の構成要素候補生成処理および文字切り出し・認識処理においては，基本矩形の統合のみにより，構成要素領域の候補を生成することが可能となる．

構成要素候補生成処理では，基本矩形をもとに，群レベルから項目レベルまでの構成要素候補を生成する．実際には，文書モデルに記述された構成要素間の部分全体関係に沿って，上位レベルから下位レベルへと構成要素候補を順次生成し，仮説木の形式により記録する．このような処理は，構成要素候補の生成と検査を繰り返すことにより，実行される．本論文では，前者をレベル間処理，後者をレベル内処理と呼ぶ．

あるレベルで構成要素候補生成処理が終了すると，生成された個々の構成要素候補をベースとして，レベル間処理により，次レベルの構成要素候補を生成する．レベル間処理において全てのベースから構成要素候補が生成されると，レベル内処理により，類似差異関係の知識に基づいて不適当な構成要素候補を除去し，文書モデルの全ての記述と整合性を保証する．レベル内処理が終了した後は，各構成要素候補をベースとして，レベル間処理によりさらに下位レベルの構成要素候補を生成する．

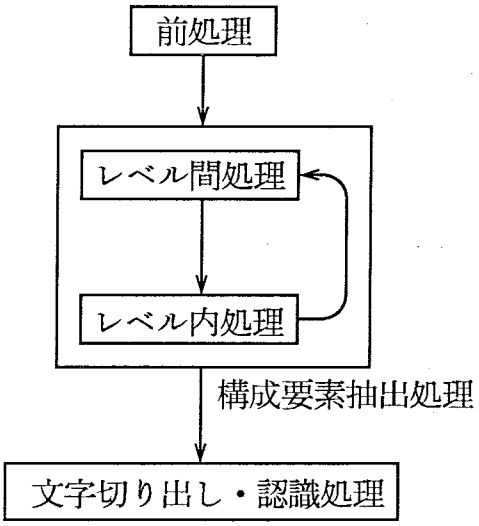


図 4.1: 文書構造解析の概要

文字切り出し・認識処理では，項目レベルの構成要素候補をもとに，文字レベルの構成要素を生成する．上位レベルの処理と同様に，文字切り出し・認識処理の結果も文字の候補として扱われる．以下では，各処理について詳しく述べていく．

4.3 前処理

文書構造解析において，対象となる構成要素のうち，領域の最も小さいものは文字である．従って，基本矩形としては文字領域と同等か，あるいは文字領域の一部分であるような矩形領域が適当である．

名刺のように，文字領域間の接触がほとんど存在しない文書では，基本矩形として黒画素領域を包含する矩形領域を考えることができる．このような矩形領域は，図 4.2に示すように，i 方向，j 方向の投影による画像分割を交互に可能な限り繰り返すことにより抽出することができる．ここで，画像分割の停止条件は，投影により，それ以上，画像を分割することが不可能となることである．

以上のような手続きにより得られた基本矩形には，構成要素領域の部分領

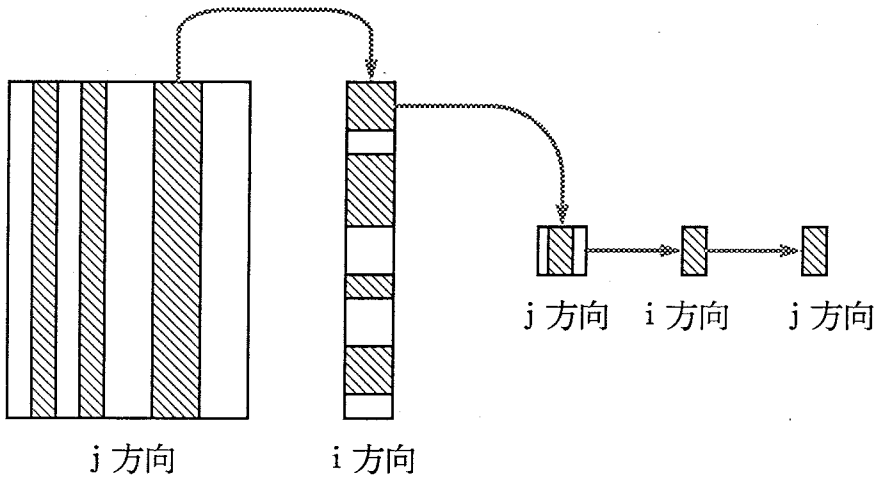


図 4.2: 基本矩形の抽出

域の他，ノイズが含まれている可能性がある．本手法では，矩形領域の座標，面積，相対位置により表される 3 種類の基準をもとにノイズを除去する．座標によるノイズ除去では，画像の枠の部分に存在するノイズを除去するほか，名刺中の左上部に存在するマーク (社名ロゴ等) を除去する．面積によるノイズ除去では，一定面積以下の小さな基本矩形をノイズとみなして除去する．また，相対位置に基づくノイズ除去では，文字列方向に対する孤立点をノイズとして除去する．すなわち，基本矩形を br，基本矩形の集合を BR とするとき，

- 縦書き文書の場合
  $up(br, BR) = \mathrm{NULL}$ かつ $down(br, BR) = \mathrm{NULL}$

- 横書き文書の場合
  $right(br, BR) = \mathrm{NULL}$ かつ $left(br, BR) = \mathrm{NULL}$

を満たす br をノイズとして除去する．名刺の住所部分に対する基本矩形の抽出例を図 4.3に示す．

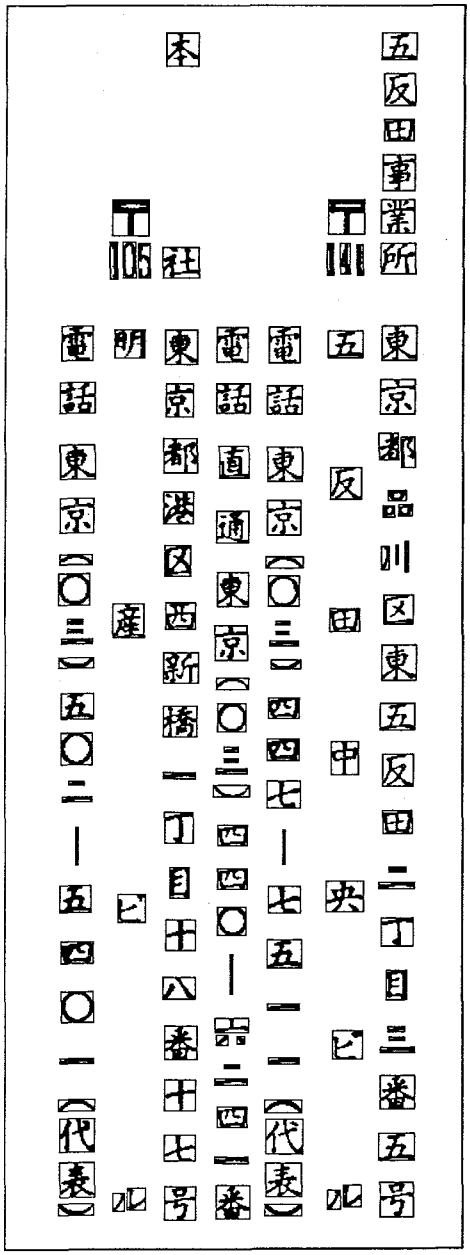


図 4.3: 基本矩形の抽出例

4.4 構成要素候補生成処理

構成要素候補生成処理のデータとして用いるのは，前処理で抽出された基本矩形である．基本矩形は，最小領域であるため，構成要素領域の候補を生成する際に，分割を考慮する必要はない．従って，本論文で述べる構成要素候補生成処理とは，基本矩形を構成要素領域へと統合し，適当な属性を付与することである．構成要素候補生成処理では，対象となる構成要素のレベルに応じて，文書モデルに記述されたレイアウト構造に関する知識を用いる．ただし，処理を実行する過程は，レベルに関係なく同一のものである．以下では，レベルとは独立に，模式例を用いて具体的な処理について説明する．

4.4.1 レベル間処理

あるレベルで処理が終了すると，生成された構成要素候補をベースとして，次レベルの構成要素候補を生成する．すなわち，ベース領域内に含まれる基本矩形を統合して構成要素候補の領域を生成し，それに属性として構成要素名を付与する．

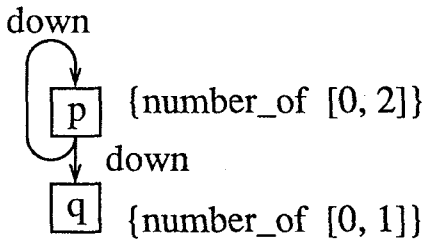
ベースとなる構成要素候補は，既に文書モデルに記述された構成要素と対応していることを考えると，ベース領域内に存在することが可能なすべての下位レベルの構成要素の集合，

$$F = \{f_1, \ldots, f_m\} \qquad \text{ただし，} f_i \text{ は構成要素名を表す} \tag{4.1}$$

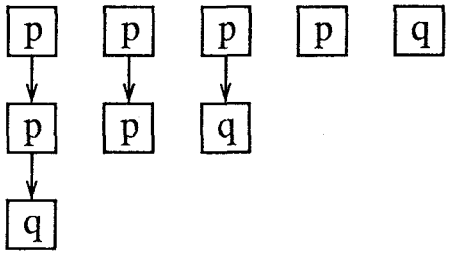
を，文書モデルの部分全体関係から求めることができる．

文書モデルでは，ベース領域内に包含される構成要素が，縦方向あるいは横方向のどちらか一定方向に並ぶように，部分全体関係を設定している．従って，F 内の構成要素の相対位置を調べることにより，構成要素の並ぶ方向を同定することができる．以下では，簡単のため，縦方向に構成要素が並ぶ場合を考える．

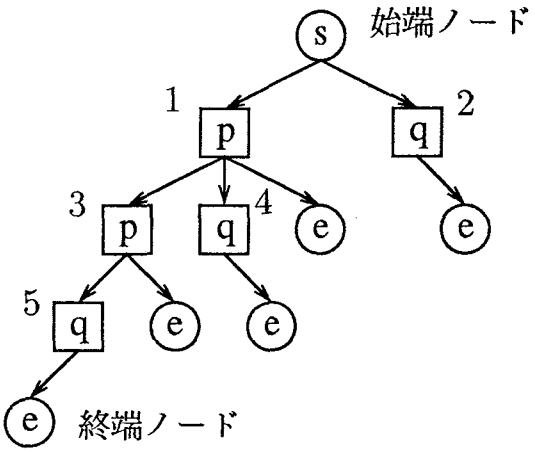
文書モデルに記述された相対位置と個数を参照すると，F の要素 f_i の可能な出現パターンを求めることができる．例えば，図 4.4 (a) に示す構成要素 p,q に対しては，同図 (b) の 5 種類の出現パターンを考えることができる．ここで，ベース内には構成要素が存在する必要があるため，p,q の個数が共に 0 である場合は考えない．本手法では，(b) に示す出現パターンを (c) に示す木構造により圧縮表現する．ここで，ノード s,e はそれぞれベース領域の上端，下端を表す．以後は，これらを始端ノード，終端ノードと呼ぶ．また，これら以外のノードを構成要素ノード，各種ノードから構成される木を構成



(a) 文書モデルの記述



(b) 出現パターン



(c) 構成要素木

図 4.4: 構成要素の出現パターン

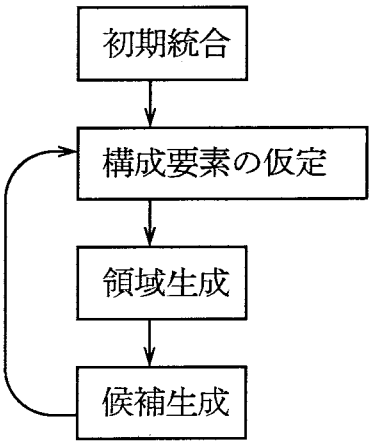


図 4.5: レベル間処理

要素木と呼ぶ．構成要素の候補を生成する際には，(b) に示す出現パターンではなく，(c) の構成要素木を用いる．

さて，このような構成要素木を考えると，レベル間処理とは，始端ノードから終端ノードへの各パスに適合するように，ベース内の基本矩形を統合することであるといえる．具体的には，図 4.5に示す 4 処理を実行する．以下，各処理について説明する．

1. 初期統合
   初期統合では，図 4.6に示すように，投影処理を用いてベース領域において構成要素が並ぶ方向とは逆の方向に基本矩形を統合する．この処理を施すことにより，以下の 3 処理において対象とする矩形領域を，一定方向に並ぶものに単純化することができる．すなわち，以下では，上下に隣接する矩形領域を統合するか，あるいは統合しないかを，モデル記述に応じて決定すればよいことになる．

2. 構成要素の仮定
   構成要素の仮定では，まず，領域生成の対象となる構成要素を構成要素木から求める．最初は，始端ノードに接続する構成要素候補が選択されることになる．ここで，文書モデルに記述された構成要素のレイアウト構造に構造的バリエーションが存在する場合には，複数の構成要素が得

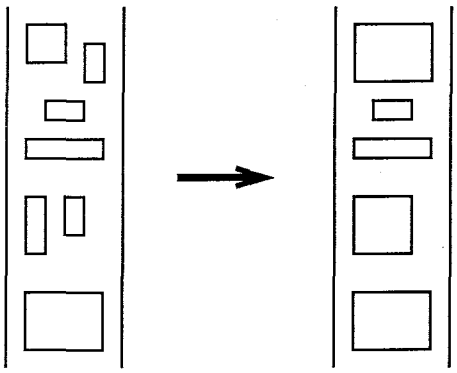


図 4.6: 初期統合

られることに注意されたい．以下では，このような構成要素の集合を，

$$G = \{g_1, \ldots, g_n\} \qquad \text{ただし，} G \subseteq F \tag{4.2}$$

と表す．前述の例の場合，処理の最初では，図 4.5 (c) の p (構成要素ノード 1)，q (構成要素ノード 2) の集合 $G = \{p, q\}$ が得られる．つぎに，集合 G 中の各構成要素 g_i を仮定して，3以降の処理を実行する．

3. 領域生成
領域生成では，g_i に関するモデル記述を満たすように，矩形領域を統合する．ここで，それまでに生成されたどの構成要素候補にも属していない矩形領域のうち，最も上側に存在するものを x とすると，具体的には，図 4.7に示すように，x から順に下方向に矩形領域を統合することになる．本手法では，矩形領域を 1 つ統合するごとに，モデル記述を満たすかどうかを検査するという生成・検査型の処理を実行する．処理の最初では，矩形領域 x がベース領域の上端に位置するものとなり，また満たすべきモデル記述は g_i とベースの間の部分全体関係となる．

関係を規定する特徴量が区間値により記述されていることを考えると，一般には，複数の矩形領域がモデル記述と適合する．以下では，この矩形領域の集合を，

$$R = \{r_1, \ldots, r_o\} \tag{4.3}$$

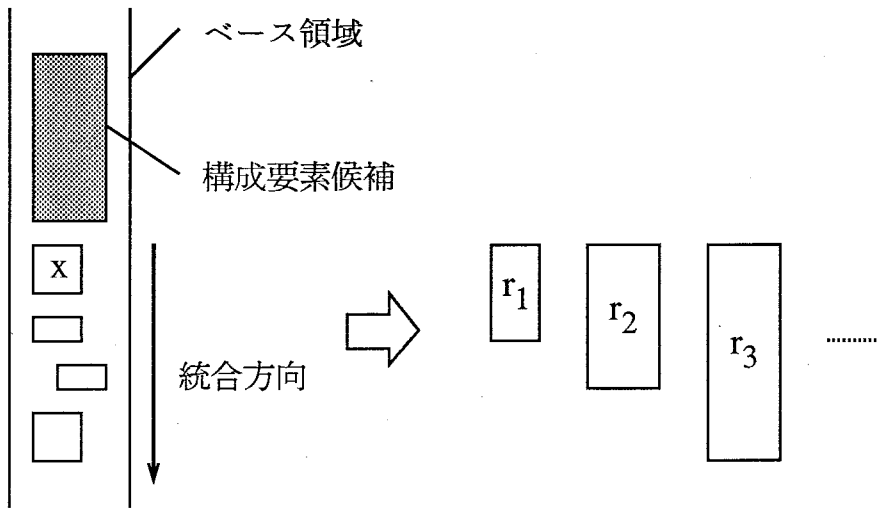


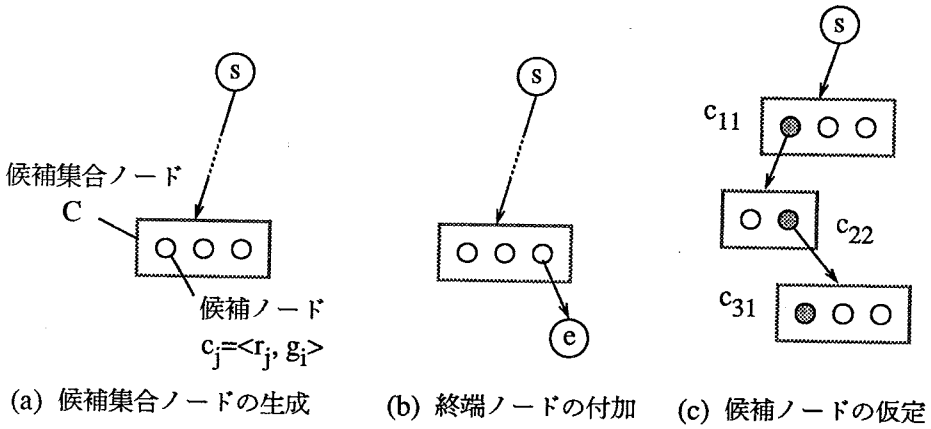
図 4.7: 領域生成

と表す．ここで，もし R が空集合ならば，2において仮定した構成要素 g_i を不適当と判断し，g_i に対する処理を中断する．なお，図 4.4 (c) における構成要素ノード 2のように，次に接続するノードが終端ノードの場合には，生成・検査型の処理を実行せず，矩形領域 x から下側の矩形領域すべてを統合する．このとき，結果として得られた矩形領域 r_1 がモデル記述と適合するならば，矩形領域集合 $R = \{r_1\}$ を生成する．

4. 候補生成
候補生成では，まず，3において生成された矩形領域に属性として g_i を付与し，構成要素候補を生成する．ここで，構成要素を領域 r と属性 a の組により $< r, a >$ のように表現すると，生成される候補の集合は，

$$C = \{c_1, \ldots, c_o\} \qquad \text{ただし，} c_j =< r_j, g_i > \tag{4.4}$$

となる．このようにして得られた候補集合 C は，構成要素木と類似した形式により，図 4.8 (a) のように記録される．以下では，候補を表すノードを候補ノード，候補集合を表すノードを候補集合ノード，候補集合ノードから構成される木を候補木と呼ぶ．



(a) 候補集合ノードの生成　(b) 終端ノードの付加　(c) 候補ノードの仮定

図 4.8: 候補木

c_k =< r_k, g_i >のうち，矩形領域r_kがベース領域の下端に位置するものについては，次のようなの処理を施す．もし，c_kに対応する構成要素g_iが，構成要素木において終端ノードに接続するならば，構成要素木と適合するため，図 4.8 (b) のように，候補ノードc_kに対しても終端ノードを接続する．一方，g_iに対応する構成要素ノードが終端ノードに接続しないならば，構成要素木と適合しないため，c_kを集合Cから除去する．

c_k =< r_k, g_i >のうち，上記の条件に該当しないものについては，それらの各々を正しい構成要素と仮定して再び2以降を実行し，c_kの下側に続く構成要素の候補を生成する．

4において構成要素候補c_k =< r_k, g_i >を正しいと仮定すると，2で求められる新たな対象は，構成要素木でg_iに接続する構成要素g_i'となる．例えば，g_iが図 4.4 (c) の構成要素ノード 1 に対応する場合には，次の処理対象g_i'としては，構成要素ノード 3，4 の構成要素p,qを得る．さらに，3における矩形領域xは，仮定した構成要素候補c_kの領域r_kの下側に位置する矩形領域となる．また，3において満たすべきモデル記述には，g_i'とベース間の部分全体関係のほか，これまでに仮定している構成要素候補とg_i'の間の類似差異関係が加わる．例えば，図 4.8 (c) に示す候補c_{31}を仮定している場合には，葉ノードc_{31}，および葉ノードから始端ノードsに至るまでのノードc_{22},

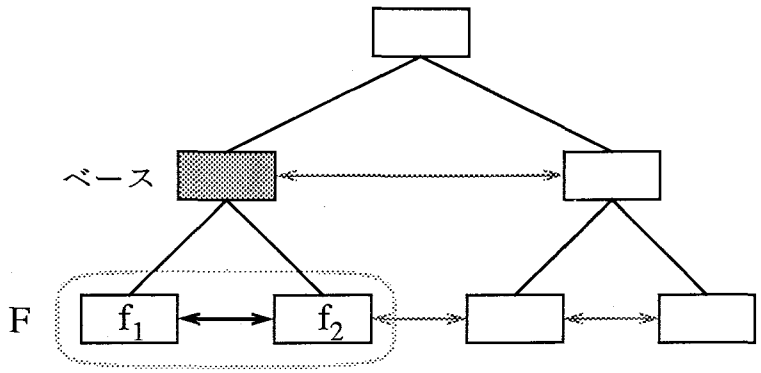


図 4.9: レベル間処理で使用される類似差異関係

c_{11} を考え，各ノードに対応する構成要素と g'_i の間の類似差異関係すべてを同時に満たす必要がある．最終的に，レベル間処理では，図 4.9に示すように，構成要素の集合 F の要素間に張られた類似差異関係をすべて検査することになる．

2，3，4の処理を繰り返して候補木を作成し，全ての候補ノードが終端ノードに接続すると，レベル間処理が終了する．完成した候補木において，始端ノード s から終端ノード e に至るパスを考え，その上に存在する構成要素候補 c_i を連言表現すると，式 2.3の形式の仮説が生成される．このとき，仮説集合 H は，始端ノードから終端ノードに至るすべてのパスに対応するものとなる． 図 4.10に処理例を示す．この例では 2 つの住所準群から，項目群の候補 18 個を含む 10 個の仮説が生成されている．

4.4.2 レベル内処理

レベル間処理は，個々のベースから独立に構成要素候補を生成する処理である．従って，異なるベースから生成された構成要素候補が，それらの間の類似差異関係を満たす保証はない．実際，図 4.11の s,t のようなベースにまたがる類似差異関係は，レベル間処理において用いられることはない．そこでレベル内処理では，このような未使用の類似差異関係を用いて，構成要素候補を検査する． 処理の具体的な説明に入る前に，構成要素候補，仮説，仮説集合の矛盾・整合について定義する．

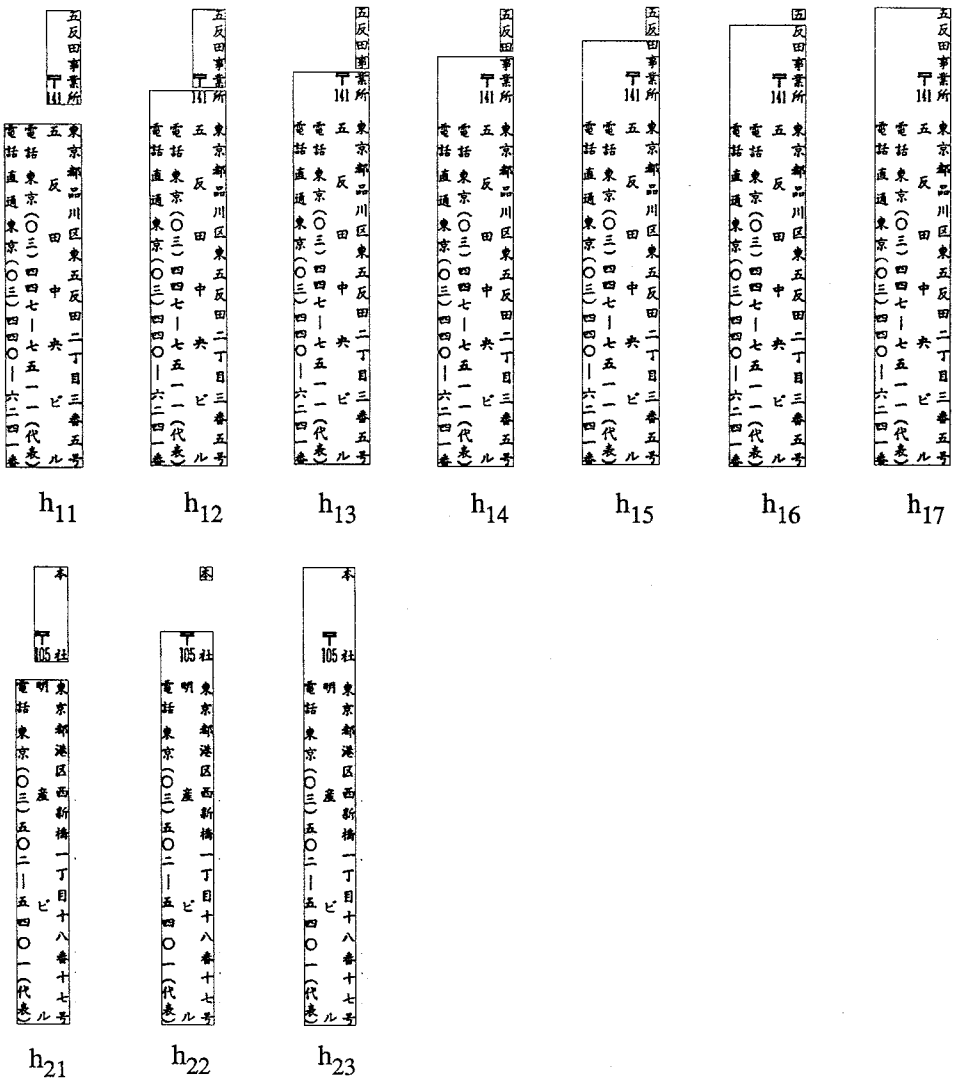


図 4.10: レベル間処理例

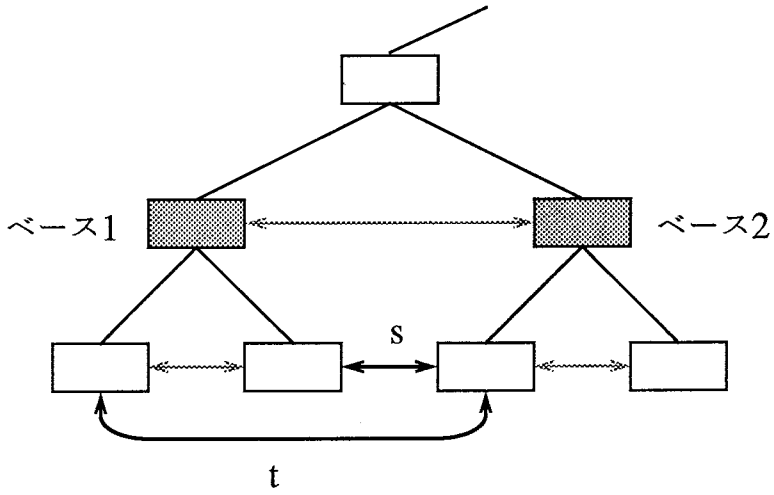


図 4.11: レベル内処理で使用される類似差異関係

矛盾・整合の定義

本論文では，構成要素候補，仮説，仮説集合の 3 種類のデータに関する矛盾・整合を 2 つに分類して考える．一方は，レベル間処理において仮説木を生成したとき，式 (2.1) ～ (2.3) により記録される矛盾，整合である．これをレベル間矛盾，レベル間整合と呼ぶ．もう一方は，レベル内処理を経て決定される矛盾，整合である．これを以下では，レベル内矛盾，レベル内整合と呼ぶ．

仮説木の性質から考えると，レベル間矛盾となるデータが整合することはない．また，レベル間処理を終了した段階では，未使用のモデル記述が存在するため，レベル間整合のデータはレベル内矛盾となる可能性がある．すなわち，レベル内処理とはレベル間整合の構成要素候補組のうち，レベル内矛盾するものを発見し，記録することであるといえる．以下に，構成要素候補，仮説，仮説集合に関するレベル間整合・矛盾，レベル内整合・矛盾を順に定義する．

I. レベル間整合・矛盾

定義 2 *(構成要素候補組 (c_1, c_2) のレベル間整合・矛盾)*

同一レベルに属する2つの構成要素候補を c_1, c_2 とする．仮説木において，構成要素候補組 (c_1, c_2) は，次の2つの場合に分類できる．

1. 同一の仮説に属する場合

$$h = and(\ldots, c_1, \ldots, c_2, \ldots) \tag{4.5}$$

2. 異なる仮説に属する場合

$$h_1 = and(\ldots, c_1, \ldots) \tag{4.6}$$
$$h_2 = and(\ldots, c_2, \ldots) \tag{4.7}$$
$$h_1 \neq h_2 \tag{4.8}$$

1の場合には，レベル間処理においてモデル記述と整合することが保証されている．このような構成要素候補組 (c_1, c_2) をレベル間整合するという．2の場合には，仮説組 (h_1, h_2) がレベル間整合であるかレベル間矛盾であるかにより解釈が異なる．すなわち，仮説組がレベル間整合するとき，構成要素候補組 (c_1, c_2) はレベル間整合するという．また，仮説組がレベル間矛盾のときには，構成要素候補組もレベル間矛盾となる．

定義 3 *(仮説組 (h_1, h_2) のレベル間整合・矛盾)*
同一レベルに属する2つの仮説を h_1, h_2 $(h_1 \neq h_2)$ とする．図 4.12に示すように，仮説木において，仮説組 (h_1, h_2) は1〜3の3種類の場合に分類できる．

1. h_1,h_2 が共通の仮説集合 H に属するならば，仮説集合の定義 (式 (2.2)) から仮説 h_1,h_2 は矛盾となる．

これ以外の場合には，共通の親仮説 h_c が現われるまで，仮説木の根の方向に向かって，それぞれの親仮説をたどることにより，矛盾するか否かを判断する．

2. h_1,h_2 の親仮説のうち，共通の親仮説集合 H_p に属するものを，それぞれ h_{1p},h_{2p} とするとき，$h_{1p} \neq h_{2p}$ ならば，仮説 h_1,h_2 は矛盾する．
3. 1，2以外の場合，すなわち共通の親仮説集合 H_p に属する親仮説 h_{1p}, h_{2p} に対して，$h_{1p} = h_{2p}(= h_c)$ ならば，仮説 h_1,h_2 は矛盾しない．

仮説組 (h_1, h_2) が，1，2に相当するとき，レベル間矛盾であるという．また，3の場合には，レベル間整合であるという．

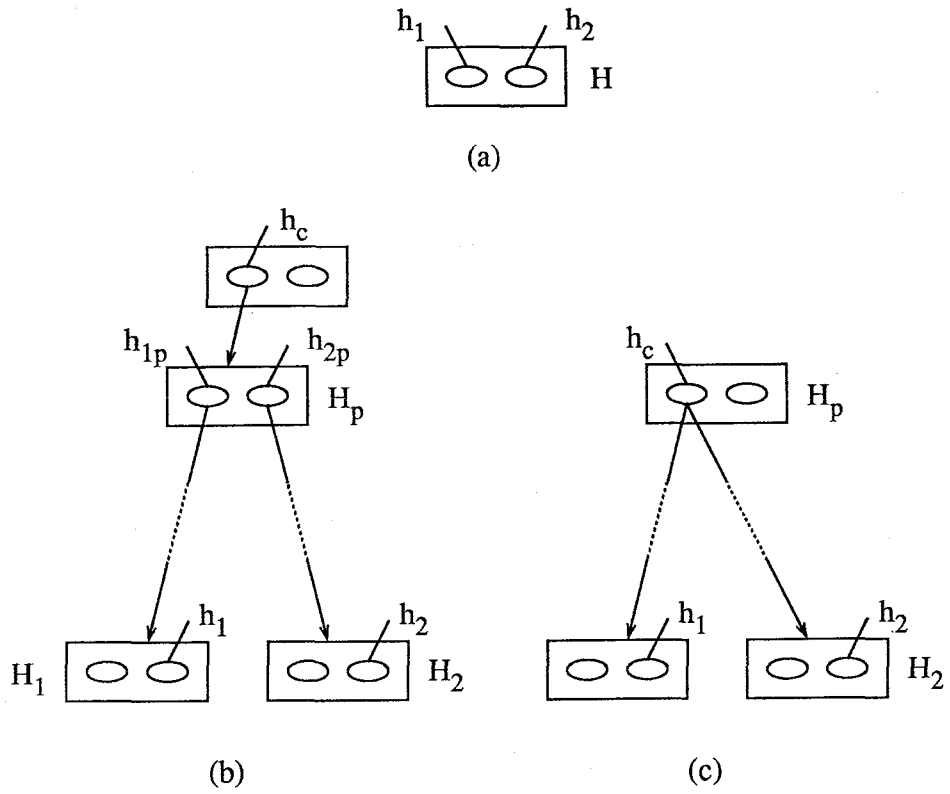


図 4.12: 仮説と仮説集合の分類

定義 4 *(仮説集合組 (H_1, H_2) のレベル間整合・矛盾)*
同一レベルに属する 2 つの仮説集合を H_1,H_2 とする．先ほどと同じ図 4.12 に示すように，仮説木における仮説集合組 (H_1, H_2) は，2，3の 2 パターンに分類できる．

仮説集合組 (H_1, H_2) が 2に相当するとき，レベル間矛盾であるという．また，3の場合には，レベル間整合であるという．

II. レベル内整合・矛盾

定義 5 *(構成要素候補組 (c_1, c_2) のレベル内整合・矛盾)*
異なるベースから生成された同一レベルの 2 つの構成要素候補を，

$$c_1 =< r_1, a_1 >, \quad c_2 =< r_2, a_2 > \tag{4.9}$$

とする．次の条件を満たす場合，構成要素候補組 (c_1, c_2) をレベル内整合であるという．

- 文書モデルに a_1 と a_2 の間の類似差異関係が記述されている場合:
  r_1 と r_2 が，類似差異関係に記述されたすべてのレイアウト述語を満たすもの
- 文書モデルに a_1 と a_2 の間の類似差異関係が記述されていない場合:
  すべての構成要素候補組

これらの条件に該当しないとき，(c_1, c_2) はレベル間矛盾であるという．

以後は，構成要素候補組 (c_1, c_2) がレベル内整合であることを，

$$consistent(c_1, c_2) \tag{4.10}$$

と表し，またレベル内矛盾であることを，

$$inconsistent(c_1, c_2) \tag{4.11}$$

と表す．

定義 6 *(仮説組 (h_1, h_2) のレベル内整合・矛盾)*
異なるベースから生成された同一レベルの 2 つの仮説を，

$$h_1 = and(c_{11}, \ldots, c_{1m}) \tag{4.12}$$

$$h_2 = and(c_{21}, \ldots, c_{2n}) \tag{4.13}$$

とする．ある i,j $(1 \leq i \leq m, 1 \leq j \leq n)$ について，

$$inconsistent(c_{1i}, c_{2j}) \tag{4.14}$$

が成立する場合，仮説組 (h_1, h_2) はレベル内矛盾であるという．それ以外の場合には，仮説組 (h_1, h_2) はレベル内整合であるという．

以後は，仮説組 (h_1, h_2) がレベル内整合であることを，

$$consistent(h_1, h_2) \tag{4.15}$$

と表し，またレベル内矛盾であることを，

$$inconsistent(h_1, h_2) \tag{4.16}$$

と表す．

定義 7 *(仮説集合組 (H_1, H_2) のレベル内整合・矛盾)*
同一レベルに属する仮説集合を，

$$H_1 = \{h_{11}, \ldots, h_{1m}\} \tag{4.17}$$
$$H_2 = \{h_{21}, \ldots, h_{2n}\} \tag{4.18}$$

として，これらの間の演算，

$$P = H_1 * H_2 \tag{4.19}$$
$$= \{p | p = (h_{1i}, h_{2j}) \in H_1 \times H_2, \text{ かつ} consistent(h_{1i}, h_{2j})\} \tag{4.20}$$

を考える．ここで，$H_1 \times H_2$ は，H_1 と H_2 の直積集合を表す．

この演算は，仮説集合 H_1, H_2 から，レベル内整合する仮説の組 (h_{1i}, h_{2j}) の集合 P を求めるものである．従って，$P \neq \emptyset$ ならば，1 つ以上のレベル内整合する仮説組が存在することになる．このとき，仮説集合組 (H_1, H_2) をレベル内整合であるという．一方，$P = \emptyset$ ならば，整合する仮説組が存在しないため，レベル内矛盾となる．

定義 8 *(仮説集合組 $(H_1, \ldots, H_l)$ のレベル内整合・矛盾)*
同様に，$P = H_1 * \ldots * H_l$ の値，すなわち，仮説集合 $H_1, \ldots, H_l$ から整合する仮説組が得られるかどうかにより判断する．ここで，$((h_1, \ldots, h_{l-1}), h_l)$

を $(h_1,\ldots,h_l)$ と表すと，$P = H_1 * \ldots * H_l$ は，

$$\begin{aligned} P &= H_1 * \ldots * H_l \\ &= (H_1 * \ldots * H_{l-1}) * H_l \\ &= ((H_1 * \ldots * H_{l-2}) * H_{l-1}) * H_l) \\ &\vdots \\ &= (\ldots(H_1 * H_2) * H_3)\ldots) * H_l) \end{aligned} \tag{4.21}$$

と変形できる．さらに，consistent に関する以下の性質，

> $consistent(h_1,\ldots,h_{l-1})$ が成立するとき，
> $consistent((h_1,\ldots,h_{l-1}),h_l)$ であることは，
> $and(consistent(h_1,h_l),consistent(h_2,h_l),\ldots,consistent(h_{l-1},h_l))$
> と同値である

を考えると，式 (4.21) から P を求めることができる．

集合 P は，前述の場合と同様に，$H_1,\ldots,H_l$ から求められたレベル内整合する仮説組の集合を表す．同様の観点から，$P \neq \emptyset$ ならば，仮説集合組 $(H_1,\ldots,H_l)$ はレベル内整合であるという．一方，$P = \emptyset$ ならば，レベル内矛盾であるという．

処理の構成

レベル内処理は，図 4.13に示すように 4 種類の処理から構成される．以下，各々について述べる．

1. 検査対象同定
   構成要素候補には，レベル内処理を施すまでもなく，あらかじめ矛盾するもの，すなわちレベル間矛盾なものが含まれる．また，レベル間整合する構成要素候補の組のうち，定義 2 におけるレベル間整合の 1 に相当するものは，すでにモデル記述との整合性が保証されている．そこで，検査対象としては，レベル間整合の 2 に相当する構成要素候補組を用いる．ただし，それまでのレベル内処理により矛盾が発見されている構成要素候補組は検査対象としない．

   図 4.14に示す例では，1 の組はレベル間整合の 1，3 の組はレベル間整合の 2，2 の組はレベル間矛盾に相当する．従って，もし 3 の組に矛盾が発見されていない場合には，これが検査対象となる．

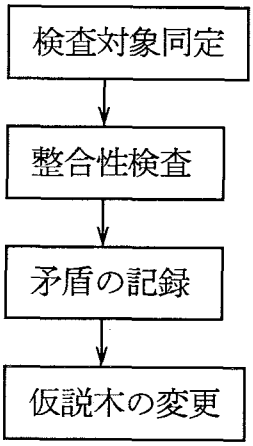


図 4.13: レベル内処理

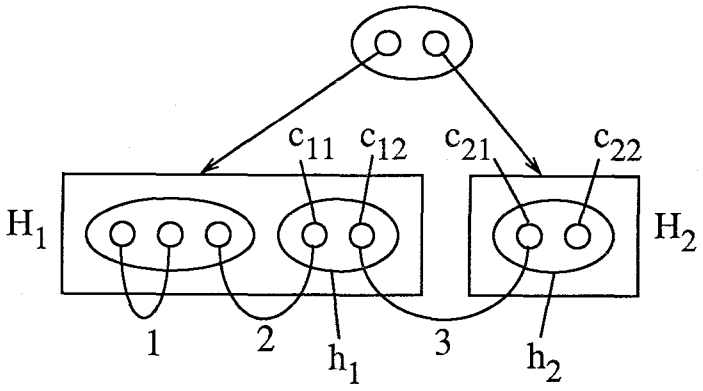


図 4.14: 検査対象

2. 整合性検査
検査対象が定まると，実際に類似差異関係を満たすかどうか，すなわち，構成要素候補組がレベル内整合するかどうかを検査する．

3. 矛盾の記録
類似差異関係を用いた検査により，新たにレベル内矛盾と判断された構成要素候補組は，以後の処理において同時に使用することはできない．本手法では，発見されたレベル内矛盾を仮説管理部の矛盾データベースに記録することにより，同時使用を抑制する．

ここで，前述の定義から，構成要素候補組がレベル内矛盾であることは，対応する仮説組のレベル内矛盾を意味する．例えば，図 4.14において，3 の組，すなわち構成要素候補組 (c_{12}, c_{21}) にレベル内矛盾が発見されたとすると，仮説 h_1，h_2 が，

$$h_1 = and(c_{11}, c_{12}) \tag{4.22}$$
$$h_2 = and(c_{21}, c_{22}) \tag{4.23}$$

と表されるため，仮説組 (h_1, h_2) もレベル内矛盾となる．ここで，式 (4.22)，(4.23) から，仮説 h_1，h_2 がレベル内矛盾することは，構成要素候補組 (c_{12}, c_{21}) に加えて，残りの 3 組，すなわち (c_{11}, c_{21})，(c_{11}, c_{22})，(c_{12}, c_{22}) もレベル内矛盾であることを意味する．

レベル内矛盾を構成要素候補組に関して記録するためには，上記のように，組合せ的な記録が必要となり，膨大な記憶領域が必要となる．そこで，本手法では，発見された構成要素候補組の矛盾を，仮説組の矛盾として記録する．実際には，仮説組 (h_1, h_2) が矛盾であるとすると，矛盾データベースに，

$$inconsistent(h_1, h_2) \tag{4.24}$$

を付加する．ここで，仮説 h_1，h_2 を，

$$h_1 = and(c_{11}, \ldots, c_{1m}) \tag{4.25}$$
$$h_2 = and(c_{21}, \ldots, c_{2n}) \tag{4.26}$$

とすると，式 (4.24) は，

$$inconsistent(c_{1i}, c_{2j}) \quad for\ all\ i, j\ (1 \leq i \leq m, 1 \leq j \leq n) \tag{4.27}$$

なる $m \times n$ 個の矛盾と同等である．以後の処理では，このような構成要素候補組 (c_{1i}, c_{2j}) が処理において同時に用いられることはない．

4. 仮説木の変更

仮説木の構造を考えると，仮説集合には必ず整合しなければならない組が存在する．例えば，図 4.12 (c) の仮説集合 H_1， H_2 は，必ず整合しなければならない．一般的には，図 4.15 (a) に示すように，共通の親仮説 h_c を持つ仮説集合 $H_1, \ldots, H_l$ は，必ず整合しなければならないといえる．

ここで，簡単な例を考えてみる．図 4.16に示すように，上位レベルの仮説 h_c から，次のような仮説集合が得られたとする．

$$h_c = and(c_1, c_2) \tag{4.28}$$

$$c_1 \Rightarrow H_1, \quad c_2 \Rightarrow H_2 \tag{4.29}$$

$$H_1 = \{h_{11}, h_{12}\}, \quad H_2 = \{h_{21}, h_{22}\} \tag{4.30}$$

このとき，式 (4.28)，(4.29) から，仮説集合 H_1， H_2 は必ず整合しなければならないと言える．これは，

$$\begin{aligned} &consistent(h_{11}, h_{21}), \quad consistent(h_{11}, h_{22}), \\ &consistent(h_{12}, h_{21}), \quad consistent(h_{12}, h_{22}) \end{aligned} \tag{4.31}$$

のいずれかが成立しなければならないことを意味する．もし，これらすべてが成立しないならば，仮説集合 H_1， H_2 がレベル内矛盾となり，式 (4.28)，(4.29) から，上位レベルの仮説 h_c は矛盾であると判断できる．従って，図 4.15 (b) に示すように，仮説 h_c を根とする部分木を，仮説木から除去し，以後の処理を無矛盾に保つ．

仮説集合 H_1， H_2 が整合する場合でも，仮説集合に属する仮説が矛盾となる場合が考えられる．例えば，前述の図 4.16の例において，

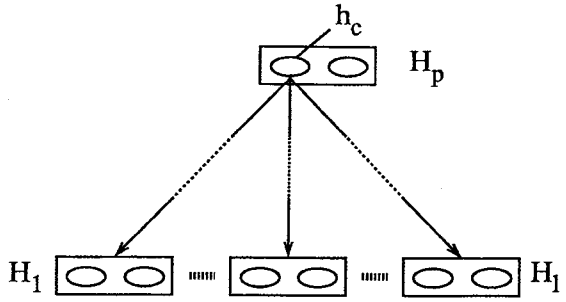
$$inconsistent(h_{11}, h_{21}) \tag{4.32}$$

$$inconsistent(h_{11}, h_{22}) \tag{4.33}$$

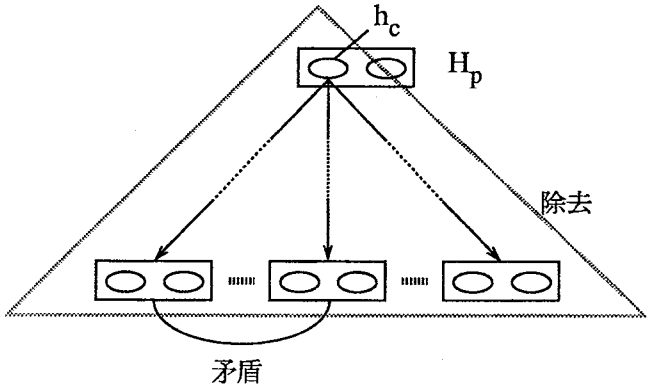
が矛盾データベースに記録されていると，仮説 h_1 は，整合する相手の仮説が存在しないため，仮説 h_c が正しい限り，矛盾となる．このような場合，本手法では，仮説管理部において式 (4.28) 〜 (4.33) から，

$$h_c \Rightarrow inconsistent(h_{11}) \tag{4.34}$$

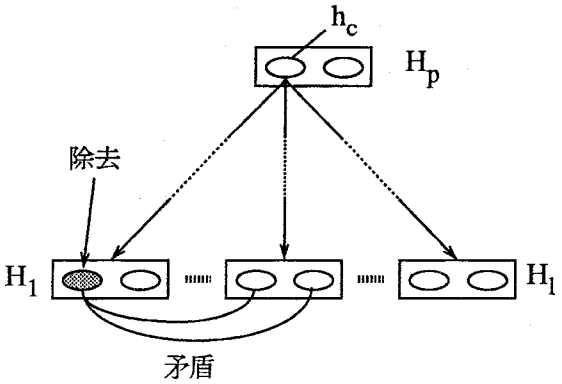
を推論する．この式は，h_c が正しい限り，仮説 h_{11} が必ず矛盾することを示している．一般的には，仮説の矛盾を発見するための推論規則を次のように表すことができる．



(a) 整合すべき仮説集合



(b) 部分木の除去



(c) 仮説の除去

図 4.15: 仮説木の変更

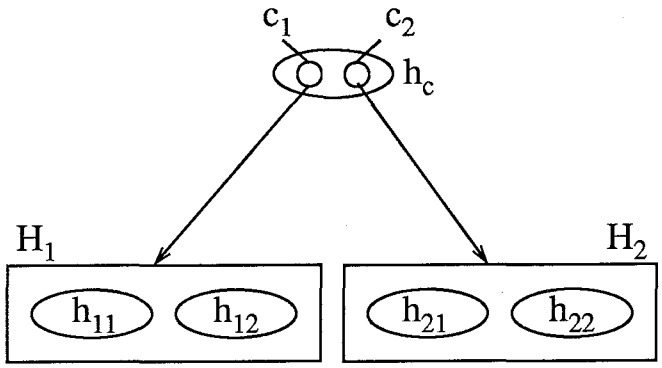


図 4.16: 仮説集合例

規則 1 *(仮説の矛盾発見)*
親仮説 h_c，親仮説を構成する構成要素候補 $c_1, \ldots, c_n$，構成要素候補 c_i から生成された仮説集合 H_i，および仮説集合 H_i を構成する仮説 $h_{i1}, \ldots, h_{in_i}$ を，

$$h_c = and(c_1, \ldots, c_l) \tag{4.35}$$

$$c_i \Rightarrow H_i \tag{4.36}$$

$$H_i = \{h_{i1}, \ldots, h_{in_i}\} \tag{4.37}$$

とするとき，ある仮説集合組 (H_i, H_j) に着目する．H_i に属する仮説 h_{ik} について，

$$inconsistent(h_{ik}, h_{jm}) \quad for\ all\ m\ (1 \leq m \leq n_j) \tag{4.38}$$

が存在するならば，

$$h_c \Rightarrow inconsistent(h_{ik}) \tag{4.39}$$

を結論付ける．

仮説管理部では，式 (4.39) のように単一の仮説からなる矛盾が記録されると，図 4.15 (c) に示すように，h_c を根とする仮説木の部分木から，h_{ik} を除去し，以後の推論を無矛盾に保つ．

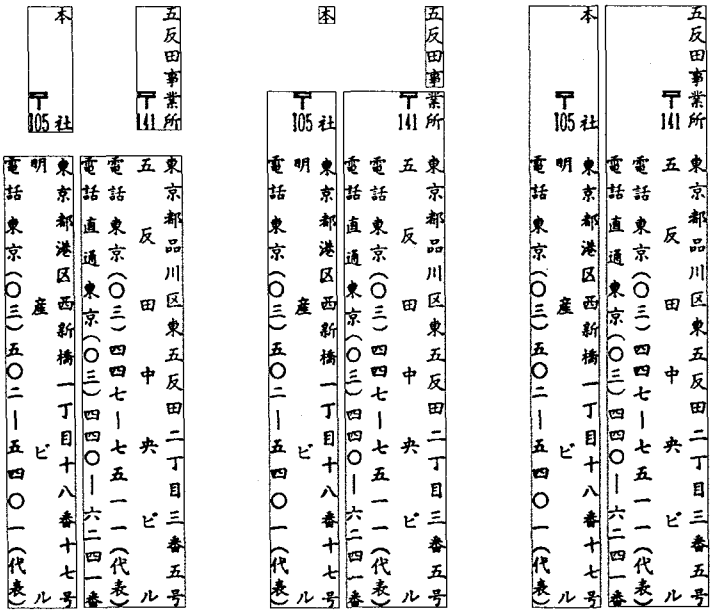


図 4.17: レベル内処理例

処理例

図 4.10の仮説に対して，レベル内処理を施した結果を図 4.17に示す．図 4.10では，レベル間処理により 2 つの住所準群から項目群の 18 個の候補を含む 10 個の仮説が生成されている．その結果に対して，

『住所項目群は上下が揃わなければならない』(h_alignment)

という住所準群間の類似差異関係を用いて検査すると，4 個の不適当な仮説を除去することができる．

特殊候補

構成要素候補生成処理は，以上に述べたようなレベル間処理とレベル内処理を上位レベルから下位レベルへと繰り返し実行することにより，仮説木を生成するものである．但し，住所項目群や社名群の文字列が複数の項目を含む場合については，項目を生成するためのレイアウト構造の特徴が存在しないため，仮説を生成することが困難となる．そこで，本手法では，ベース領

域から文字列領域を生成し，各々に可能な属性をすべて付与する．このような文字列は，複数の項目を含むものであるため，項目の特殊候補と呼び，他の項目の候補と区別する．記述形式としては，式 (2.1) ～ (2.3) のかわりに，

$$b \Rightarrow \{h\} \tag{4.40}$$

$$h = and(c_1, \ldots, c_n) \tag{4.41}$$

$$c_i = < r_i, \{a_1, \ldots, a_m\} > \tag{4.42}$$

を用いる．ここで，c_i は特殊候補，r_i は文字列領域，$\{a_1, \ldots, a_m\}$ は可能な属性の集合である．なお，特殊候補に対する構成要素候補の生成は，第 5 章において述べる単語列生成処理に委ねる．

4.5 文字切り出し・認識処理

項目レベルの構成要素候補が生成されると，それをベースとして文字切り出し・認識を実行し，文字レベルの候補を求める．以下，文字切り出し処理，文字認識処理の順に詳細について述べる．

4.5.1 文字切り出し処理

基本方針

多少の訓練を必要とするが，人間であれば，文字列領域内に分布する基本矩形を見て，かなり正確に文字領域を切り出すことができる．そこで，本手法では，人間を文字切り出しのエキスパートととらえ，人間の動作のシミュレートを目指す．

人間の文字切り出し過程には，以下の特徴点があると推察される．

1. 文字切り出しに際しては，切り出し対象となる矩形領域のみに着目するのではなく，周囲の矩形領域を考慮している．

2. 文字領域として確からしいものから順次切り出す．縦書きの文字列であれば，特に“｜”(ハイフン)，“二”，“三”など特徴的な形状の基本矩形から構成される文字領域を優先して切り出している．また，周囲の基本矩形から明らかに全角文字，半角文字と判断される場合には，それらを優先的に切り出している．

3. 半角文字が複数個並ぶ場合など，全角文字，半角文字の双方の解釈が可能となるときには，切り出し結果を一意に定められないことがある．しかしながら，可能性のある文字領域の候補を列挙することはできる．

本手法は，以上の点に着目した文字切り出し処理を実現するものである．1，2の特徴を有することにより，複雑に分布する基本矩形を対象とした場合にも，人間と同様に，高効率かつ高精度な文字切り出し処理が実現できると考えられる．また，3の特徴から，不確実な切り出し結果に対しては，後の処理に選択を委ねることが可能となる．

処理概要

本手法は，図 4.18に示すように，8段階の処理からなる．文字切り出しのデータとしては，構成要素候補生成処理と同様に基本矩形を用いる．まず初期統合では，投影処理を施すことにより，文字領域の並ぶ方向とは異なる方向に基本矩形を統合する．縦書きの場合は，横方向に統合されることになる．ラベル付与では，特徴的な形状の矩形領域に対して，4種類のラベルを付与し，他のものと区別する．接触文字除去では，接触文字を発見し，文字切り出し対象から除外する．優先切り出し1では，特殊な形状の基本矩形から構成される文字領域を優先的に切り出す．さらに，優先切り出し2においては，確からしい全角文字，半角文字に対する切り出しを行う．ただし，氏名など全角文字のみから構成される項目に対しては，半角文字に対する切り出し処理は行わない．補完切り出し処理，伝搬切り出し処理においては，優先的に切り出された文字領域をもとに，他の未切り出し領域の切り出しを試みる．以上の3処理において，未切り出しとなった矩形領域に対しては，候補生成において，全角文字，半角文字の双方に対する可能な切り出し候補を網羅する．一般の文字列に対しては，以上の8処理により，文字切り出しを終了する．ただし，電話番号，郵便番号，ファックス，テレックスに対しては，縦書き文字と横書き文字が混在する場合があるため，切り出された文字領域を再び文字列として，文字切り出し処理を再実行する．

使用特徴量および文字領域の定義

以下では，縦書きの文字列を対象とした場合を例に挙げて詳しく考察していく．まず，諸定義に使用する記号，関数，述語について述べる．本論文では，対象とする文字列領域を b，文字列領域内の矩形領域の集合を B，矩形領域 (B の要素) を r により表す．ここで，r は初期統合後の矩形領域を表すものである．また，関数 $merge$ を用いて統合された矩形領域を表す．矩形領

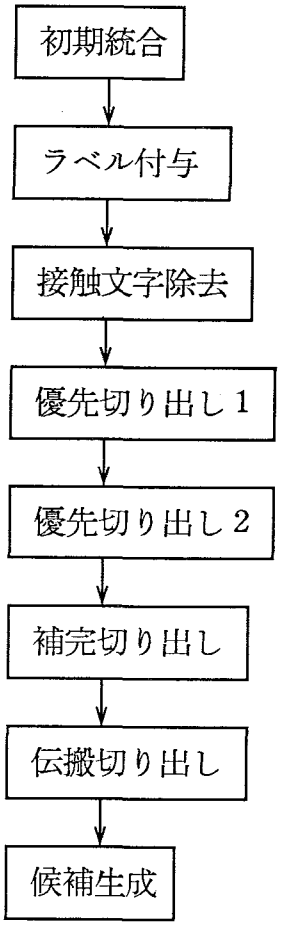


図 4.18: 文字切り出し処理

域 $r_i = (I_{si}, I_{ei}, J_{si}, J_{ei})$ に対して，$merge(r_1, \ldots, r_n)$ は，以下に示す矩形領域を表す．

$$\begin{aligned} merge(r_1, \ldots, r_n) &= (min(I_{s1}, \ldots, I_{sn}), max(I_{e1}, \ldots, I_{en}), \\ & \quad min(J_{s1}, \ldots, J_{sn}), max(J_{e1}, \ldots, J_{en})) \end{aligned} \quad (4.43)$$

これは，引数として与えられた全ての矩形領域を包含する，最小の矩形領域を表す．さらに，述語 $label(r, l)$ により，矩形領域のラベルを表現する．これは，矩形領域 r のラベルが l であるとき真，そうではないとき偽となる述語である．

次に，特徴量について述べる．縦書き文字列に対する文字切り出し処理において使用する矩形の特徴量は，縦幅比 H，縦方向の距離 d，正規化横幅 D_{in}，比較中心の i 座標 C_{ic}，比較縦幅 D_{jc}，比較横幅 D_{ic}，正方形度 S の 7 種類である．ここで，縦幅比とは，

$$H(r, b) = \frac{D_j(r)}{D_i(b)} \quad (4.44)$$

により定義される特徴量であり，全角文字の縦幅を文字列領域の横幅と仮定したとき，全角文字の縦幅に対する対象矩形領域の縦幅の比を表している．本手法では，縦幅比の値を図 4.19に示す 4 つの区間値に分類し，文字領域の定義に用いる．また縦方向の距離 d は，図 4.20に示すように，2 つの矩形領域，

$$r_1 = (I_{s1}, I_{e1}, J_{s1}, J_{e1}) \quad (4.45)$$

$$r_2 = (I_{s2}, I_{e2}, J_{s2}, J_{e2}) \quad (4.46)$$

に対して，

$$d(r_1, r_2) = J_{s2} - J_{e1} \quad (4.47)$$

により表されるものである．ただし，このとき $J_{s2} > J_{e1}$ でなければならない．その他の特徴量については，第 3 章において述べたものと同じである．以上の準備のもと，まず，ラベル付与の処理において付与されるラベルについて述べる．ラベルには，図 4.21に示すような小さい横棒 (shb)，大きい横棒 (lhb)，接触文字 (tc)，横方向に並ぶ小さい 2 つの点 (hsd) の 4 種類を用いている．

次に，2 つの矩形間の状態を表す述語 fcut，hcut を定義する．図 4.22に示す矩形領域 r_1，r_2 に対して，

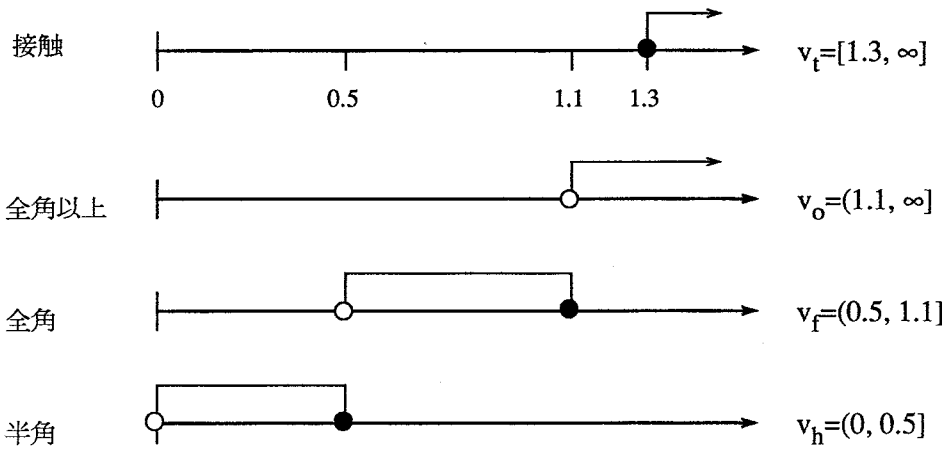


図 4.19: 縦幅比の分類

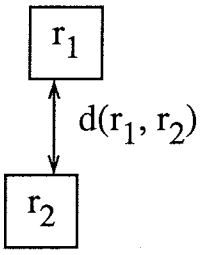


図 4.20: 縦方向の距離

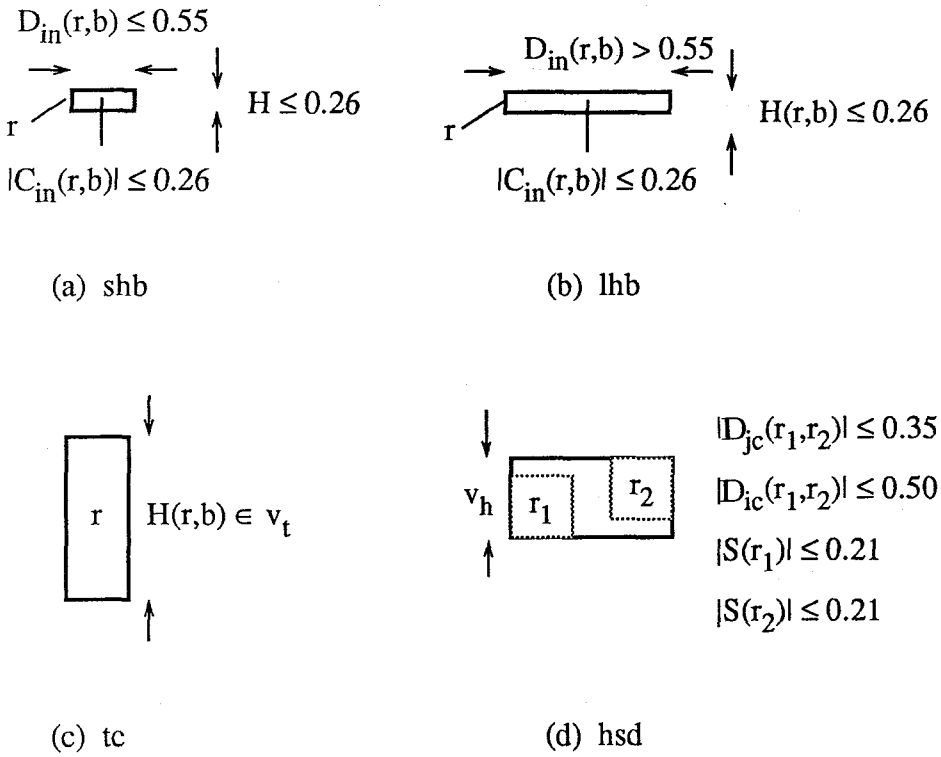


図 4.21: 矩形領域のラベル

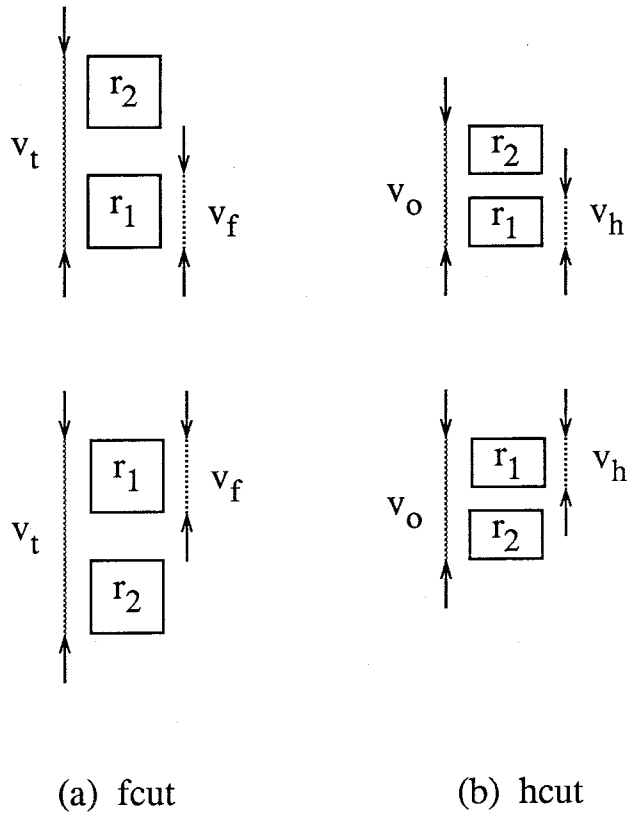


図 4.22: fcut, hcut の定義

$$fcut(r_1, r_2) = and(H(r_1, b) \in v_f, H(merge(r_1, r_2), b) \in v_t) \quad (4.48)$$

$$hcut(r_1, r_2) = and(H(r_1, b) \in v_h, H(merge(r_1, r_2), b) \in v_o) \quad (4.49)$$

ただし， $r_2 = up(r_1, B)$ または $down(r_1, B)$

となる．連言の第 1 項は，対象とする矩形領域 r_1 が全角あるいは半角の文字領域として妥当な縦幅比を持つことを表し，第 2 項は，隣接する矩形領域 r_2 と統合した場合に文字領域としては不適当な縦幅比を持つことを表す．全体としては，fcut が真であることは r_1 を全角文字として r_2 から分離可能なことを，hcut が真であることは r_1 を半角文字として r_2 から分離可能なことを表す．

最後に，上記の定義を用いて，種々の文字領域を定義する．対象とする文字領域は，“｜”(ハイフン)，連続する“一”，“二”，“三”，“六”，全角文字，半角文字の7種類である．図 4.23に示す矩形領域に対して，定義を表 4.1に示す． ここで，条件に示す項がすべて満たされるとき，文字領域の欄に示す領域を文字とみなす．なお，文字領域が文字列領域の上端(下端)に位置する場合には，表中の上端(下端)の列において○の付く条件を満たす必要はない．

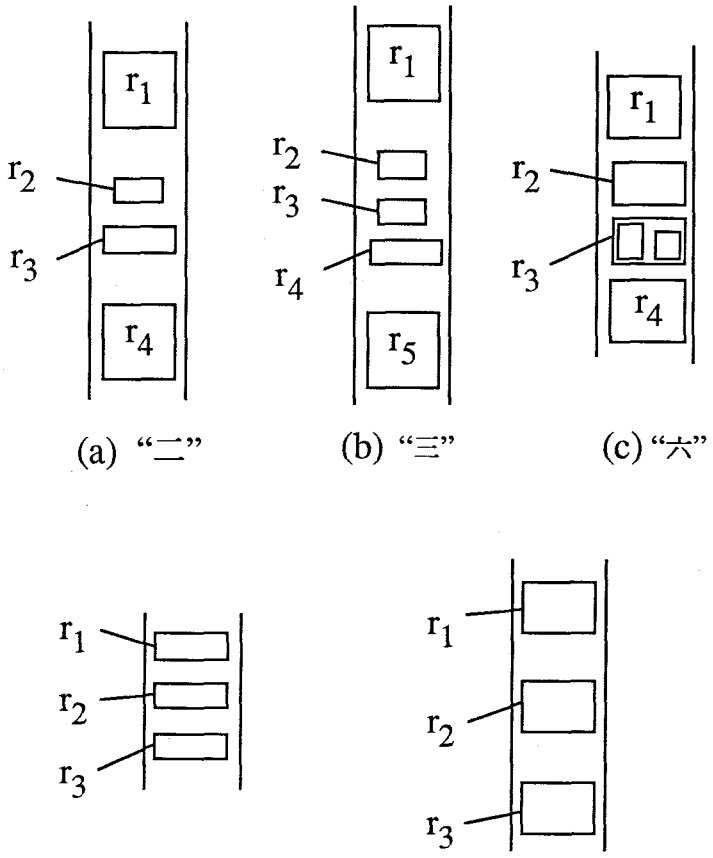
ルール形式による実現

本手法では，複雑に分布する矩形領域から柔軟に文字切り出しを実行するため，if-then形式のルールを用いて処理を実現する．このとき，文字領域のうち特徴的なものおよび確からしいものを優先して切り出すため，ルールに6段階の優先順位を付ける．優先順位ごとのルールの概要を表 4.2に示す．ここで，接触文字除去とは，接触文字とみなされた矩形領域を r とすると，切り出し対象の矩形領域集合 B を $B - \{r\}$ に変更する処理である．また，各ルールにより切り出された文字領域にはラベル chr を付与し，それ以後は切り出し対象としない．

補完切り出し，伝搬切り出し，候補生成の3処理について，図を用いてさらに詳しく説明する．補完切り出し処理は，図4.24に示す矩形領域に対して，表 4.3のルール comp により文字を切り出すものである． 上端，下端における考慮不要の条件については，文字領域の定義の場合と同様である．伝搬切り出し処理では，図 4.25 (a) に示す矩形領域に対して表 4.3のルール prop-down により文字を切り出し，(b) に示す矩形領域に対してルール prop-up により文字を切り出す．候補生成では，図 4.26に示すように，全角文字の縦幅比 v_f，半角文字の縦幅比 v_h を満たすように，全ての文字領域候補を網羅する．

処理結果の記録方法

一般に文字領域の候補は，他レベルの場合に比べて非常に多数となる場合がある．従って，式 (2.1) ～ (2.3) のように，あらかじめ候補の組合せを作成する形式により処理結果を保存することは実際的ではない．そこで本手法では，有向グラフを用いて文字領域の抽出結果を保存する．本論文では，この有向グラフを文字列グラフと呼ぶ．図 4.27 (a) の抽出結果に対する文字列グラフを同図 (b) に示す．ここで，ノードは文字領域，アークは文字の読み順を表す．なお，読み順とは，縦書きの場合はベース領域の上から下に，横書



(a) “二” (b) “三” (c) “六”

(d) 連続する“一” (e) 全角文字, 半角文字

図 4.23: 文字領域

表 4.1: 文字領域

文字領域名	文字領域	条件	上端	下端
ハイフン	r	$D_{in}(r,b) \leq 0.3$ $H(r,b) \in v_f$ $C_{ic}(r,b) \leq 0.1$		
二	$merge(r_2,r_3)$	$not(label(r_1,shb))$ $or(label(r_2,shb),$ $\quad label(r_2,lhb))$ $label(r_3,lhb)$ $or(fcut(merge(r_2,r_3),r_1),$ $\quad hcut(merge(r_2,r_3),r_1))$ $or(fcut(merge(r_2,r_3),r_4),$ $\quad hcut(merge(r_2,r_3),r_4))$ $d(r_1,r_2) > d(r_2,r_3)$ $d(r_2,r_3) < d(r_3,r_4)$	○ ○ ○	 ○ ○
三	$merge(r_2,r_3,r_4)$	$not(label(r_1,shb))$ $or(label(r_2,shb),$ $\quad label(r_2,lhb))$ $label(r_3,shb)$ $label(r_4,lhb)$ $or(fcut(merge(r_2,r_3,r_4),r_1),$ $\quad hcut(merge(r_2,r_3,r_4),r_1))$ $or(fcut(merge(r_2,r_3,r_4),r_5),$ $\quad hcut(merge(r_2,r_3,r_4),r_5))$ $d(r_1,r_2) > d(r_3,r_4)$ $d(r_3,r_4) < d(r_4,r_5)$	○ ○ ○	 ○ ○

表 4.1: 文字領域 (つづき)

文字領域名	文字領域	条件	上端	下端
六	$merge(r_2, r_3)$	$H(r_2, b) \in v_h$ $label(r_3, hsd)$ $or(fcut(merge(r_2, r_3), r_1),$ $\quad hcut(merge(r_2, r_3), r_1))$ $or(fcut(merge(r_2, r_3), r_4),$ $\quad hcut(merge(r_2, r_3), r_4))$	○	○
一	r_2	$label(r_1, lhb)$ $label(r_2, lhb)$ $label(r_3, lhb)$ $H(merge(r_1, r_2), b) \in v_f$ $H(merge(r2, r3), b) \in v_f$		
全角文字	r_2	$fcut(r_2, r_1)$ $fcut(r_2, r_3)$	○	○
半角文字	r_2	$hcut(r_2, r_1)$ $hcut(r_2, r_3)$	○	○

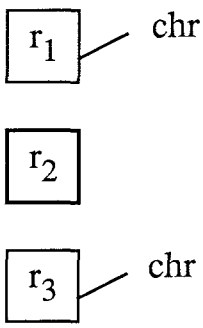
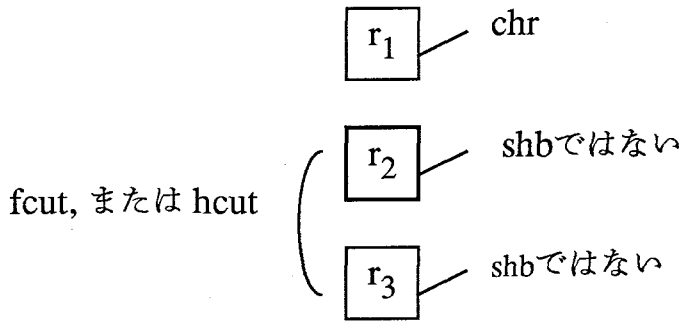


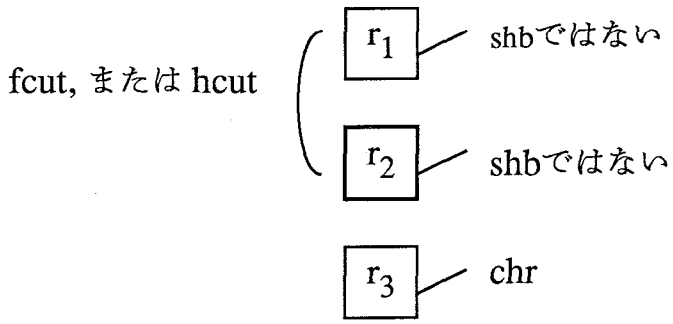
図 4.24: 補完切り出し処理

表 4.2: ルールの概要

処理名	優先順位	ルールの概要
接触文字除去	1	接触文字のラベル tc を持つ矩形領域を切り出し対象から 除外する
優先切り出し 1	2	“	”(ハイフン), “二”, “三”, “六”, 連続する “一” の定義を満たす文字領域を切り出す
優先切り出し 2	3	全角文字, 半角文字の定義を満たす矩形領域を文字 として切り出す
補完切り出し	4	切り出された文字領域 (chr のラベルを持つ矩形領域) の間に挟まれる未切り出しの矩形領域を文字として 切り出す
伝搬切り出し	5	切り出された文字領域 (chr のラベルと持つ矩形領域) をもとに, 上方向あるいは下方向に文字を切り出す
候補生成	6	未切り出しの矩形領域に対して, 全角文字, 半角文字 を仮定したときの全ての切り出しパターンを網羅する



(a) prop-down



(b) prop-up

図 4.25: 伝搬切り出し処理

表 4.3: 文字切り出しルール

処理名	ルール名	文字領域	条件	上端	下端
補完切り出し	comp	r_2	$label(r_1, chr)$ $label(r_3, chr)$	○	 ○
伝搬切り出し	prop-down	r_2	$label(r_1, chr)$ $not(label(r_2, shb))$ $not(label(r_3, shb))$ $or(fcut(r_2, r_3),$ $hcut(r_2, r_3))$		
	prop-up	r_2	$label(r_3, chr)$ $not(label(r_2, shb))$ $not(label(r_1, shb))$ $or(fcut(r_2, r_1),$ $hcut(r_2, r_1))$		

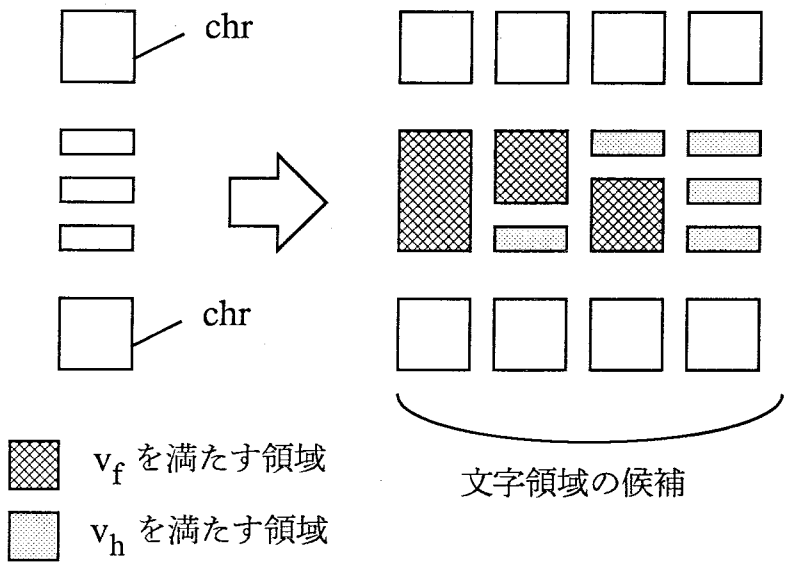
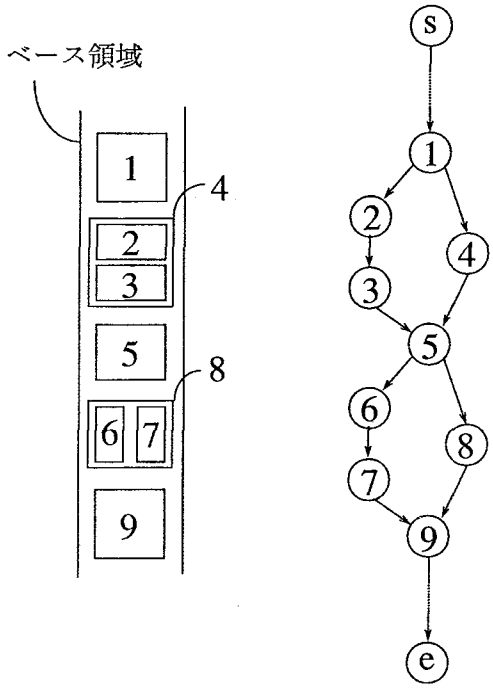


図 4.26: 候補生成



(a) 文字領域　　(b) 文字列グラフ

図 4.27: 文字切り出し結果の記録

きの場合には左から右に文字領域をたどるものである．文字列グラフにおいて，1つのノードから複数のアークが出ることは，互いに矛盾する文字領域の存在を意味する．例えば，ノード1から出る2つのアークは，ノード2，3とノード4の矛盾を表す．

4.5.2 文字認識処理

文字切り出し処理が終了すると，各文字領域について文字認識を施し，結果を属性として付与する．認識手法としては，パターン整合法的手法を採用する．認識辞書は，漢字，英数字，記号を合わせた3238カテゴリについて，教科書体，楷書体，ゴシック体，明朝体の4種類のフォントを考慮したものである．なお，本手法では，文字1カテゴリに対して教科書体，楷書体に対する辞書，およびゴシック体，明朝体に対する辞書の2辞書を設けている．

認識結果については，類似度をもとに以下の2つの基準により候補文字数を決定する．

基準1 $S_1 > \theta_1$ かつ $S_1 - S_2 > \theta_2$

基準2 $\frac{S_j}{S_1} > \theta_3 \ (j = 1, \ldots, n)$

まず，基準1が成立した場合には，1位候補の信頼性が高いと判断し，候補文字を一意に絞る．基準1が成立しないときには，基準2を満足する候補文字をn位を限度としてすべて認識結果とする．θ_1からθ_3までのしきい値およびnを決定する際には，以下の2条件を考慮しなければならない．

- 正解文字を可能な限り削減しない
- 誤認識文字を可能な限り削減する

これらは相反する条件であるため，実際には，正解文字の削減，誤認識文字の未削減をある程度許容せざるを得ない．

以上の候補文字数決定法は，使用する文字認識処理の性質，性能に大きく依存する．従って，本手法では，あらかじめ文字認識処理の性能評価実験を行い，次のような方針，

θ_1 誤って一意とされている候補文字の類似度の最大

θ_2 ある文字領域に対して誤って1位とされている候補文字と2位の候補文字との類似度の差の最大値

θ_3 ある文字領域に対する正解候補と，誤って 1 位とされている候補文字との類似度の比の最小値

n 検証処理が可能な数の最大

により，しきい値を決定する．

4.6 実験結果と検討

実験対象は，電機関連企業の縦書き名刺 100 枚である．使用した画像データは，解像度 400dpi のスキャナーにより得た 2 値画像である．また，文字認識における最大候補数 n を 20 とした．

文書構造解析処理の結果を評価する際には，結果が候補として与えられることから，抽出率，切り出し率，認識率などの尺度を用いることはできない．そこで，本論文では，以下のような尺度を用いて，結果を評価する．

平均候補数 N_{ave}: あるレベルにおいて，構成要素あたりに生成された候補数の平均

$$N_{ave} = \frac{N_e}{N_a} \tag{4.50}$$

N_e: あるレベルにおいて生成された全候補数

N_a: あるレベルにおいて抽出対象となった全構成要素数 (以後，構成要素数と呼ぶ)

信頼度 R: 生成された候補中に正解が含まれている割合

$$R = \frac{N_c}{N_a} \tag{4.51}$$

N_c: 候補中に含まれる正解数

なお，生成された候補が正解と判断されるのは，領域が正しく抽出されており，かつ属性が正しく付与されている場合である．ただし，項目の候補に関しては，4.4.2においても述べたように，文字列領域が複数の項目を含む領域として扱われることがある．この場合には，正解の領域が文字列領域に包含されており，かつ文字列領域に付与された属性に正解の属性が存在するとき，正解であると判断する．以上の尺度を用いると，平均候補数が少なく，かつ信頼度が高いとき，文書構造解析処理の有効性が高いと判断できる．

結果を表 4.4に示す．群，準群，項目群レベルでは，比較的少ない平均候

表 4.4: 文書構造解析結果

レベル	構成要素数	平均候補数		信頼度
文書	１００	−		−
群	３００	１．０		１００％
準群	５０７	１．４		９９．４％
項目群	１７０	３．９		９５．９％
項目	９７２	９．８		９９．２％
文字	８８１１	領域	１．３	９８．８％
		属性	８．３	９４．２％

補数で高い信頼度が得られている．特に群レベルでは，レイアウト構造に構造的バリエーションが存在しないため，100% の信頼度を得ることができた．準群レベルでは，ノイズ除去の失敗や画像の傾きが原因で3個の構成要素の生成に失敗した．また，レベルの失敗は，同様の原因によるもの5個，文書モデルにレイアウト構造が記述されていなかったもの2個の合計7個であった．

項目レベルでは，平均候補数が他のレベルに比べて多いものの，同様に高い信頼度を得ている．この主な理由は，氏名群，社名群中の項目について属性を決定するレイアウト構造上の特徴が存在しないために，同一の領域に数多くの属性が付与されていることである．領域に限って言えば，構成要素数の1.8倍程度の候補が生成されているにすぎない．なお，生成に失敗したものは8個であり，原因は項目群の場合と同じである．

文字レベルでは，領域と属性(認識結果)の結果を別々に示した．領域に関しては，他のレベルと同様に高い信頼度を得た．属性に関しては信頼度(平均8.3位までの累積分類率)が94.2% となり，必ずしも十分な結果とはいえない．ここで，文字認識処理の精度を評価するため，正解の文字領域を用いて文字認識実験を行った．その結果，文字認識率は78.7% にどどまることがわかった．この原因としては，

- 文字ごとに画像の濃度分布がかなり異なるため，固定しきい値による2値化では，文字のかすれ，つぶれ等が発生してしまうこと
- 画像の解像度が不十分であるため，住所群中の極端に小さい文字につぶれ等が頻繁に起こること
- 社名の特殊文字に認識辞書が対応していないこと

などが挙げられる．文字認識に失敗した画像例を図4.28に示す．

以上の結果から，文字認識を除けば十分な処理結果が得られており，文書構造解析処理の有効性が確認された．文字認識に対しても，CMB[57, 58, 59]などの適応的しきい値選択法を導入する，画像の解像度を上げる，認識辞書に特殊文字を登録するなどにより対処可能と考えられる．

4.7 結言

本章では，レイアウト構造に関する知識を援用する文書構造解析法を提案した．また，縦書き名刺を対象として仮説生成実験を行った結果，平均候補数，および信頼度の両面から，本手法の有効性を示した．

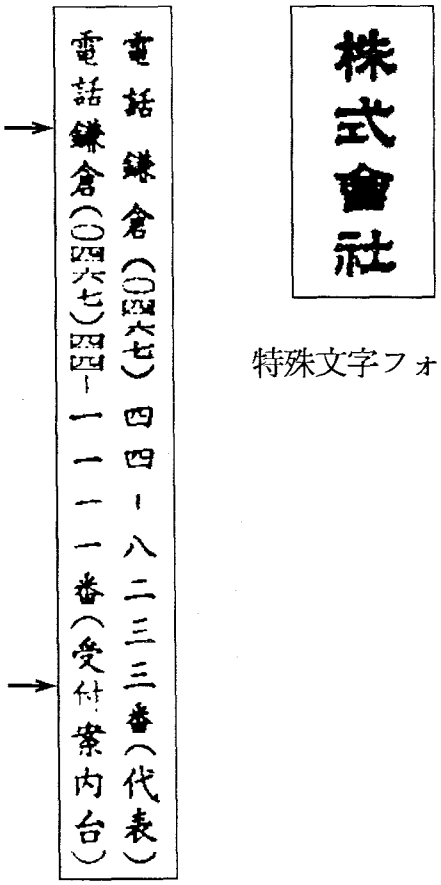

株式會社
特殊文字フォント
電話鎌倉(〇四六七)四四ー八二三三番(代表)
電話鎌倉(〇四六七)四ー一一一一番(受付案内台)
つぶれ・かすれ

図 4.28: 誤認識例

本手法の特徴としては，

1. レイアウト構造の部分全体関係，類似差異関係に着目したレベル間処理およびレベル内処理により，構成要素の候補を効率的に生成すること
2. 構成要素候補の生成過程から仮説間の依存関係を記録し，処理を無矛盾に保つこと
3. プロダクションルールを用いることにより，不定ピッチの文字列から文字を柔軟に切り出すこと

が挙げられる．一方，問題点としては，レイアウト構造に関する知識のみからでは，構成要素候補を完全に生成することが困難となる場合が存在することから，記述内容を考慮した処理の必要性が指摘される．

第 5 章

構造化記述生成

5.1 緒言

前章において述べた文書構造解析は，レイアウト構造に関する知識を用いて，構成要素候補を生成するものである．それに対して，本章では文書モデルに記述された記述内容の論理的制約に関する知識を用いて構成要素候補を生成し，最終的に構造化記述を生成する処理について述べる．

本手法は，文字レベルの候補からボトムアップ的に構造化記述を生成するものであり，基本的には，文書構造解析とは独立に考えることができる．すなわち，文字レベルの候補が与えられれば，それらをもとに順に上位レベルの候補を生成することが可能である．従来，同様の観点に立つ手法としては図書目録カードを対象とした手法がある [49]．しかしながら，第 3 章において述べたように，一般の項目主体の文書においては，図書目録カードの区切り記号に相当する項目間の明確な境界が存在しないため，論理的制約が弱いものとなり，単独で処理を実行した場合には，非常に多数の構成要素候補が生成される．その結果，候補の組合せ的な爆発から処理の継続が困難となると考えられる．

そこで，本章では，論理的制約を利用する処理に仮説検証という目的を与え，その目的を効率的に満たす手法を提案する．本手法における基本的な考えは，仮説を新たに生成せずに，すでに生成されている仮説の取捨選択のみを実行することにある．具体的には，文書構造解析において生成された仮説木，仮説間の矛盾，および構成要素候補の属性を有効利用することにより，論理的制約に関する知識の適用対象を最小限度に抑制する．また，本手法では，仮説木の変更，仮説の削除などを通して，処理の無矛盾性を保証する．

5.2 提案手法の概要

第 2 章でも述べたように，構造化記述生成とは，構成要素の記述内容を考慮することにより仮説を取捨選択し，最終的に対象文書に関する構造化記述を得る処理である．本手法では，対象とするレベルにより，構造化記述生成を単語列生成処理，単語列整合処理の 2 処理に分割して考える．

単語列生成処理では，単語辞書に記述された単語の接続性に関する知識を用いて，項目レベルの構成要素候補 (以後，項目候補と呼ぶ) に対する単語列を生成する．生成された単語列は，項目候補の記述内容に相当するものである．一般的には，単語列生成処理により複数の単語列候補が得られるため，本手法では，これらを項目に対する記述内容候補として扱う．仮説の検証という立場からは，妥当な単語列が得られるかどうかにより，項目候補を取捨選択するため，項目レベルの仮説を検証することになる．また，単語列の生成に際して，文字レベルの候補選択を同時に実行するため，文字レベルの仮説検証にも相当する．

単語列整合処理では，単語列の整合性に関する知識を用いて，項目群レベル以上の構成要素に対して，記述内容を同定する．単語列生成処理により得られた記述内容候補は，まず，項目群レベルにおいて整合性を検査される．項目群の記述内容は，項目の記述内容の組合せにより表現されることから，項目群の構成要素候補に対する記述内容候補は，項目候補の記述内容候補を単語列の整合性に関する知識を満足するように組み合わせることにより生成できる．このとき，仮説木，および矛盾データベースを参照し，組合せの対象をその段階では矛盾しないものに限ることにより，効率的な処理を実行する．このような処理を文書レベルまで実行することにより，対象文書に対する記述内容候補を得ることができる．なお，本手法では，一旦，矛盾と判断された構成要素候補を仮説ごと仮説木から除去することにより，構成要素候補の組合せ的な爆発を回避している．

5.3 単語列生成処理

文字レベルでは，文字列グラフにより仮説を表現しているため，単語列生成処理とは，分岐するアークおよびノードに記録された候補文字から，適切なものを選択することであるといえる．本手法では，文書構造解析で得られた項目候補を正しいと仮定して，項目候補に対する妥当な記述内容候補を生成する．実際には，項目候補に付与された属性から使用する単語辞書を決定し，単語列を生成するというトップダウン的処理を行う．

このようなトップダウン的な処理は，項目候補を用いて初めて実現できるものである．もし，項目候補ではなく，単なる文字列領域を対象とするのであれば，全単語辞書について単語列生成処理を実行しなければならず，処理量が膨大となる．従って，この点は，本手法の利点であると考えられる．

単語列生成処理の手順を図 5.1に示す．対象とする項目候補が特殊候補の場合には，4種類の処理すべてを実行する．特殊候補ではない場合には，単語列生成のみを実行する．以下では，各処理について詳しく述べる．

1. 始端の仮定
   項目候補が特殊候補の場合には，領域が複数の項目を含む文字列領域であるため，実際の項目の始端が文字列グラフの始端と一致するとは限らない．そこで，文字列グラフの先頭からアークに従って項目の始端を順に仮定し，以下の処理を実行する．このように始端を仮定すると，複数の項目を含む文字列に対しても，属性が正しければ単語列を得ることができる．

   項目候補が特殊候補ではない場合には，項目の始端と文字列グラフの始端が一致する．従って，文字列グラフの始端を項目の始端として，以下の処理を実行することになる．

2. 単語列候補生成
   文字列グラフを始端から終端へと探索しながら単語照合を繰り返すことにより，仮定した項目候補に対して妥当な単語列候補をすべて生成する．以下に単語照合の手順を説明する．

   (a) 辞書単語選択
   単語辞書から，照合に用いる辞書単語を選択する．選択の基準は，前段階の単語照合により得られた単語候補に接続することである．一般には，単語辞書から複数の辞書単語が選択される．ただし，文字列グラフの始端を対象とする場合には，辞書単語のうち始端に接続するものを選択する．次に，このようにして選択された辞書単語それぞれについて，文字列照合を実行する．

   (b) 文字列照合
   辞書単語長に対して $n_1\%$ までの未照合を許して，辞書単語の文字列と文字列グラフを照合し，単語候補を生成する．文字列グラフにおける照合開始位置は，図 5.2 (a) に示すように，前段階において得られた単語候補の最後の文字とアークで結ばれた文字であ

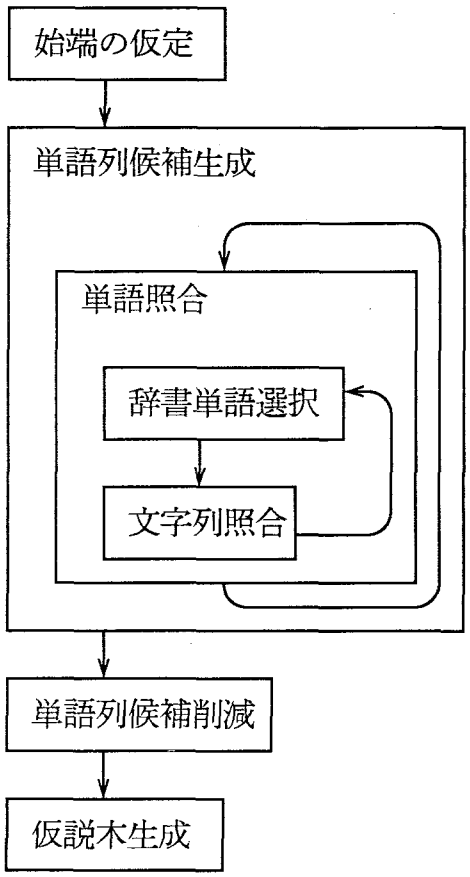


図 5.1: 単語列生成処理

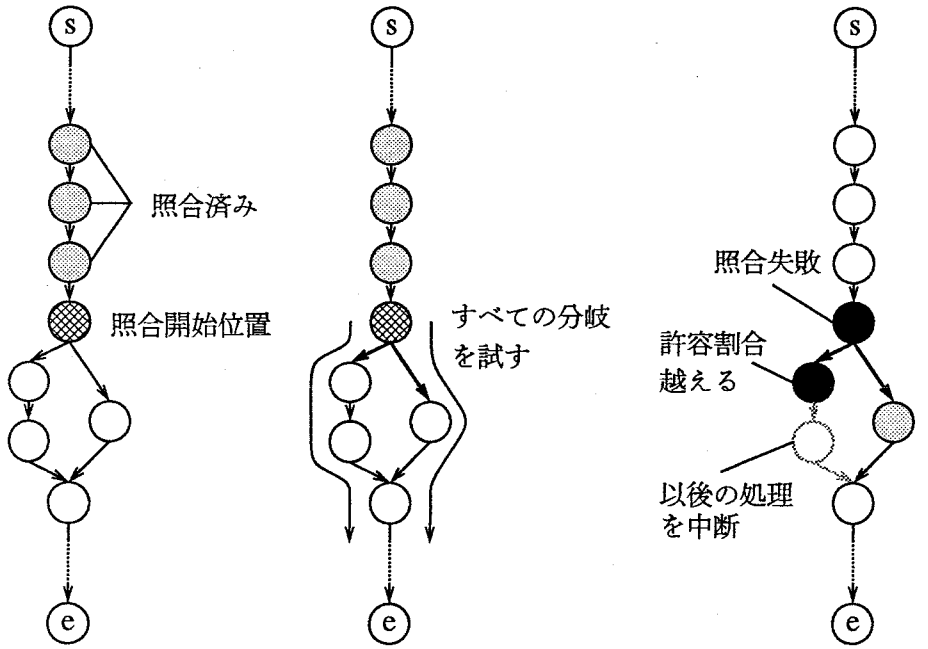

s
照合済み
照合開始位置
e
(a) 照合開始位置
s
すべての分岐
を試す
e
(b) 分岐への対処
s
照合失敗
許容割合
越える
以後の処理
を中断
e
(c) 文字列照合の中断

図 5.2: 文字列照合

る．ただし，始端に接続する辞書単語に対しては，文字列グラフの始端に接続する文字を照合対象とする．なお，図 5.2 (b) に示すように，文字列照合の途中において，アークの分岐が存在する場合には，分岐すべてを試すものとする．また，図 5.2 (c) に示すように，照合途中で未照合の許容割合を越えた場合には，以後の照合を中断する．

文字列照合を，(a) において得られた辞書単語すべてに対して実行すると，一般的には，複数の単語候補が得られる．(b) においてアークの分岐が存在する場合には，1つの辞書単語に対しても複数の単語候補が得られる可能性がある．複数の単語候補が得られた場合，そのうち，どれが妥当であるかを，この段階で判断することはできない．そこで本手法では，それぞれを正しい単語と仮定して，単語照合を繰り返す．

単語照合を繰り返していくと，以下に示す条件を満たす場合がある．

条件 1 終端に接続する辞書単語まで照合が成功していること

条件 2 文字列グラフの終端まで照合が成功していること

このとき，本手法では，始端から順に仮定してきた単語候補の列を単語列候補として生成する．ただし，項目候補が特殊候補である場合には，文字列グラフの終端が項目の終端と一致するとは限らないため，条件 2 を満たす必要はない．(a) において得られた可能な辞書単語，および文字列グラフにおけるアークの分岐をすべて考慮し，さらに未照合を許した単語照合を実行していることを考えると，一般的には，複数の単語列候補が得られる．ただし，本手法では，多数の不適当な単語列候補が生成されることを防ぐため，単語列として許容できる未照合文字の割合に，$n_2\%$ $(n_2 < n_1)$ という上限を設けている．

仮定した項目候補が特殊候補ではない場合には，得られた単語列候補集合を，項目候補に対する記述内容候補集合として記録し，処理を終了する．ここで，項目候補を c，単語列候補を s_i，記述内容候補を d_i，記述内容候補集合を D とすると，記録の形式は，

$$c : D = \{d_1, \ldots, d_n\} \quad \text{ただし，} d_i = and(s_i) \tag{5.1}$$

となる．ここで，D に属する各記述内容候補 d_i は，互いに矛盾するものである．ただし，単語列候補 s_i が得られないときは，仮定した項目候補に対して，妥当な記述内容候補が得られないことを意味する．従っ

て，本手法では，記述内容候補集合を記録するかわりに，仮説木において項目候補 c を含む仮説を特定し，その仮説を根とする部分木を仮説木から削除する．

一方，項目候補が特殊候補の場合には，次に述べる 3，4の処理を施した後，記述内容候補を生成する．

3. 単語列候補削減
単語列候補を生成する過程において，項目候補が特殊候補である場合には，複数の始端を仮定していることから，特殊候補ではない場合に比べて，多数の単語列候補が生成される．もし，このような多数の単語列候補を保持したまま，上位レベルにおいて単語列整合処理を実行すると，システムの処理効率を悪化させることになる．単語列候補の中には，単語列の整合性を考慮しなければ妥当性の判断が困難であるものも存在するが，整合性を考慮するまでもなく明らかに不適当なものも存在する．そこで本手法では，以下に述べる単純な基準により，明らかに不適当な単語列候補を発見し，単語列候補の集合から削減する．

始端を複数仮定していることを考えると，単語列候補集合の中には，他のものを包含するような単語列候補が存在する場合がある．ここで，単語列候補 s_1 が s_2 を包含するとは，単語列候補に含まれる文字候補 m_{ij} の集合を，

$$s_1 \ : \ M_1 = \{m_{11}, \ldots, m_{1n}\} \tag{5.2}$$

$$s_2 \ : \ M_2 = \{m_{21}, \ldots, m_{2l}\} \tag{5.3}$$

と表すとき，

$$M_1 \subset M_2 \tag{5.4}$$

を意味する．

例えば，単語列候補として“取締役社長”が得られた場合，“社長”が項目の始端，終端になり得るため，単語列候補として生成される．このとき，“社長”が正しい単語列候補である可能性は，非常に低いといえる．そこで，本手法では，単語列候補の内，他に包含されるものを候補から除去し，より長い単語列を採用する．

4. 仮説木の変更
単語列候補生成に際して，仮定した特殊候補 c_i を，

$$b \ \Rightarrow \ \{h\} \tag{5.5}$$

$$h = and(c_1, \ldots, c_n) \tag{5.6}$$

$$c_i = < r_i, \{a_1, \ldots, a_m\} > \tag{5.7}$$

により表されるものとする．また，c_i に対して生成された単語列候補集合から，3の処理により不適当な単語列を削減した後のものを S とする．もし，$S = \emptyset$ ならば，前述の場合と同様に，仮定した特殊候補 c_i を不適当と判断できる．従って，仮説 h を根とする部分木を，仮説木から削除する．一方，

$$S = \{s_1, \ldots, s_k\} \quad (\neq \emptyset) \tag{5.8}$$

ならば，4でも述べたように，特殊候補から構成要素候補を生成し，仮説木を変更する必要がある．

単語列候補 s_i の領域を r_{s_i}，s_i を生成するときに仮定した属性を a_i とすると，単語列候補 s_i に対応する項目候補 $c_{s_i} =< r_{s_i}, a_i >$ を考えることができる．このようにして得られた項目候補の集合を，

$$C_s = \{c_{s_1}, \ldots, c_{s_k}\} \tag{5.9}$$

とする．ここで，領域と属性が同じであれば，単語列候補が異なるものであっても，同一の項目候補とみなすことに注意されたい．項目候補 c_{s_i} に対応する単語列を $sc_1, \ldots, sc_p$ とすると，項目候補 c_{s_i} は，

$$c_{s_i} : D = \{d_1, \ldots, d_p\} \text{ただし，} d_i = and(sc_i) \tag{5.10}$$

なる記述内容候補を持つことになる．

このような項目候補の集合 C から，式 (2.1)～(2.3) の形式で表現される仮説を作成するためには，まず，

- 項目候補のうち，整合するものは同一の仮説に属する
- 項目候補のうち，矛盾するものは異なる仮説に属する

の2点に着目して，C_s から c_{s_i} を選択し，仮説作成のもとになる項目候補の組合せを生成しなければならない．項目候補の整合は，次の2条件から判断することができる．

条件1 領域が重複しないこと

条件2 可能な限り組み合わせること

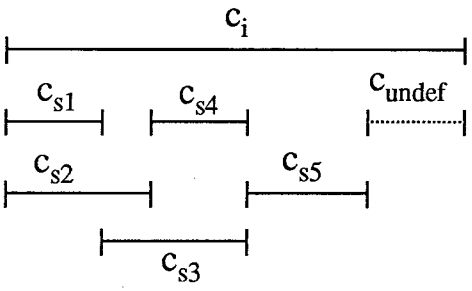


図 5.3: 選択基準

条件1は，「同一レベルの構成要素が重複する領域を持つことはない」という前提から導かれるものである．例として，図 5.3に示すような，構成要素が i 方向に並ぶ場合を考えてみる．この図では簡単のため，矩形領域の横幅のみを示している．例えば，領域 r_{s_1} と r_{s_2} は重複することから，構成要素候補 c_{s_1}, c_{s_2} は矛盾し，同一の仮説には属さない．

条件2は，「仮説は可能な限り領域を覆うものでなければならない」という考えに基づくものである．図 5.3に示す例では，条件1，2を満足する構成要素候補の組合せは，

$$\{c_{s_1}, c_{s_3}, c_{s_5}\}, \quad \{c_{s_2}, c_{s_4}, c_{s_5}\} \tag{5.11}$$

の2つである．$\{c_{s_1}, c_{s_3}\}$ などは，上記2つの組合せの一部分であるため，条件2を満たさず，不適当となる．

以上をまとめると，項目候補の選択基準は，次のように表すことができる．

規則 **2** *(項目候補の選択基準)*
C_s の部分集合 C_p を要素とする集合 $Q = \{C_p | C_p \subset C_s\}$ を考える．ただし C_s は，

(a) $c_{s_i} \in C_p$, $c_{s_j} \in C_p$, $i \neq j$ のとき，r_{s_i} と r_{s_j} は重複しない

(b) $C_p \in Q$, $C'_p \in Q$ のとき，$C_p \not\subset C'_p$ かつ $C_p \not\supset C'_p$

の2条件を満たすものである．

この基準を満たすように項目候補を選択すると，仮説の基となる組合せが得られる．なお，本手法では，文字認識，単語列生成処理の不完全性を考え，図5.3に示すように，どの項目候補にも属さない領域r_uを許容する．このような領域は，未定構成要素$c_{undef} =< r_u, undef >$として選択される．従って，式(5.11)の組合せが共に条件(a)，(b)を満たす場合には，

$$g_{i1} = (c_{s_1}, c_{s_3}, c_{s_5}, c_{undef}) \tag{5.12}$$

$$g_{i2} = (c_{s_2}, c_{s_4}, c_{s_5}, c_{undef}) \tag{5.13}$$

の2つの組が得られる．以後は，このような組g_{ij}を，特殊候補c_iに対する項目候補の基礎組と呼ぶ．また，基礎組の集合$G_i = \{g_{i1}, g_{i2}\}$を基礎集合と呼ぶ．ここで，同一の基礎集合に属する基礎組は，互いに矛盾することに注意されたい．

さて，式(5.6)により表される各c_iに対して，基礎集合G_iが得られると，各G_iから基礎組を一つづつ選択することにより，ベースbから得られる構成要素候補の組，すなわちベースbに対する仮説を生成することができる．ただし，このような構成要素候補組は，文書モデルに記述された項目の個数に関する条件を満たさなければならない．例えば，bが住所項目群のときを考えると，住所項目群に対する郵便番号は1つであるため，郵便番号を2つ以上選択することは許されない．具体的には，次のように仮説を生成することができる．

規則 3 *(特殊候補に対する仮説の生成)*
特殊候補c_iから得られた基礎集合をG_i，G_iに属する基礎組をg_{ij}とする．また，$g_{ij} = (c_{s_1}, \ldots, c_{s_m})$が，文書モデルに記述された項目の個数に関する条件を満たすことを，$consistent(c_{s_1}, \ldots, c_{s_m})$と表すこととする．ただし，未定構成要素は，すべてのモデル記述を満たすものとする．すると，

$$P = G_1 * \ldots * G_n \tag{5.14}$$

から，整合する構成要素候補組の集合Pを求めることができる．Pの要素p_kを，

$$p_k = (c_{s_1}, \ldots, c_{s_m}) \tag{5.15}$$

とすると，p_kの要素の連言は，

$$h_{sk} = and(c_{s_1}, \ldots, c_{s_m}) \tag{5.16}$$

なる仮説 h_{sk} を表す．P に含まれるすべての p_k について，同様に仮説を生成することにより，

$$H_s = \{h_{s_1}, \ldots, h_{s_q}\} \tag{5.17}$$

の形式の仮説集合を生成することができる．

最終的には，式 (5.5) ～ (5.7) を以下のように変更することができる．

$$b \Rightarrow H_s \tag{5.18}$$

$$H_s = \{h_{s_1}, \ldots, h_{s_q}\} \tag{5.19}$$

$$h_s = and(c_{s_1}, \ldots, c_{s_m}) \tag{5.20}$$

本手法では，これらの式に従い，仮説木を変更する．その後，新たに生成された構成要素候補に対して，第 4 章において述べたレベル内処理を実行する．

5.4 単語列整合処理

単語列整合処理では，項目群レベル以上の仮説を検証する．仮説が記述内容から妥当であるかどうかは，仮説に属する構成要素候補に着目し，記述内容という観点から式 (2.3) により表される連言が成立するかどうかにより判断できる．

整合性の検査対象としては，対象レベルに属するすべての仮説を考える．具体的な処理手法について述べる準備として，まず各種の整合・矛盾を定義する．

定義 **9** *(単語列候補組 (s_1, s_2) の整合・矛盾)*
項目候補 $c_1 =< r_1, a_1 >$， $c_2 =< r_2, a_2 >$ に対する単語列候補を s_1， s_2 とする．単語列候補組 (s_1, s_2) は，項目 a_1， a_2 の間に記述された単語列の整合性に関する知識を満たすときに整合するという．それ以外の場合は矛盾するという．ただし，a_1， a_2 の間に，単語列の整合性に関する知識が存在しないものに対しては，すべて整合するものとして扱う．

定義 **10** *(記述内容候補組 (d_1, d_2) の整合・矛盾)*
2 つの記述内容候補 d_1， d_2 が，

$$d_1 = and(s_{11}, \ldots, s_{1m}) \tag{5.21}$$

$$d_2 = and(s_{21}, \ldots, s_{2n}) \tag{5.22}$$

のように表されるとする．ここで，対象レベルが項目群のときは，$m = n = 1$ である．

$1 \leq i \leq m$，$1 \leq j \leq n$ なる i，j に関して，すべての単語列候補組 (s_{1i}, s_{2j}) が整合するとき，記述内容候補組 (d_1, d_2) は整合するといい，それ以外を矛盾するという．なお，以下では，整合する記述内容候補組を，

$$consistent(d_1, d_2) \tag{5.23}$$

と表す．

定義 **11** *(記述内容候補集合の整合・矛盾)*
2つの構成要素 c_1，c_2 に対応する記述内容候補集合，

$$D^1 = \{d_1^1, \ldots, d_m^1\} \tag{5.24}$$
$$D^2 = \{d_1^2, \ldots, d_n^2\} \tag{5.25}$$

について，

$$P = D^1 * D^2 \tag{5.26}$$

を考え，$P \neq \emptyset$ のとき D^1 と D^2 は整合するといい，それ以外を矛盾するという．

同様に，記述内容候補集合 $D^1, \ldots, D^l$ が整合するとは，

$$P = D^1 * \ldots * D^l \tag{5.27}$$

において，$P \neq \emptyset$ であることとする．

定義 **12** *(構成要素候補組 (c_1, c_2) の整合・矛盾)*
2つの構成要素候補 c_1，c_2 の記述内容候補集合をそれぞれ D^1，D^2 とすると，D^1 と D^2 が整合するとき，構成要素候補組 (c_1, c_2) は整合するといい，それ以外は矛盾するという．ただし，c_1，c_2 のいずれか，あるいは両方が，不定構成要素である場合には，すべて整合するとして扱う．

定義 **13** *(仮説の整合・矛盾)*
仮説を，$h = and(c_1, \ldots, c_l)$ とするとき，$c_1, \ldots, c_l$ に対応する記述内容候補 $D^1, \ldots, D^l$ が整合するとき，仮説 h は整合するといい，それ以外は矛盾するという．

以上の準備のもと，単語列整合処理について述べる．単語列整合処理は，仮説の整合性検査，記述内容候補集合生成の2処理からなる．以下，各々について説明する．

仮説の整合性検査では，仮説の整合性の定義に従って，対象レベルにおける仮説の整合性をすべて検査する．ここで，対象レベルにおける仮説集合 H，そのベース b，検査対象とする仮説 h_i を，

$$b \Rightarrow H \tag{5.28}$$

$$H = \{h_1, \ldots, h_l\} \tag{5.29}$$

$$h_i = and(c_{i1}, \ldots, c_{im}) \tag{5.30}$$

とすると，

$$c_{ij} : D^{ij} = \{d_1^{ij}, \ldots, d_n^{ij}\} \tag{5.31}$$

により表される D^{ij} に対して，仮説の整合性の定義から，

$$P^i = D^{i1} * \ldots * D^{im} \tag{5.32}$$

を求め，整合性を判断することになる．もし，$P^i = \emptyset$ ならば，仮説 h_i は矛盾するため，仮説 h_i を根とする部分木を仮説木から削除する．一方，

$$P^i = \{p_1^i, \ldots, p_k^i, \ldots, p_o^i\} \neq \emptyset \tag{5.33}$$

ならば，仮説集合は整合するため，正しい可能性がある．ここで，P^i の要素 p_k^i を，

$$p_k^i = (d_1, \ldots, d_t) \tag{5.34}$$

とするとき，

$$d_k^{bi} = and(d_1, \ldots, d_t) \tag{5.35}$$

を考えると，これは，ベース b に対する記述内容候補となる．

結局，仮説 h_i から求められた，ベース b に対する記述内容候補集合は，

$$D^{bi} = \{d_1^{bi}, \ldots, d_o^{bi}\} \tag{5.36}$$

と表すことができる．ここで，D^{bi} 中の記述内容候補は，互いに矛盾するものである．

記述内容候補集合生成では，各仮説 h_i から得られた D^{bi} を用いて，ベース b に対する記述内容候補集合 D^b を作成する．D^b は，

$$D^b = D^{b1} \cup \ldots \cup D^{bl} \tag{5.37}$$

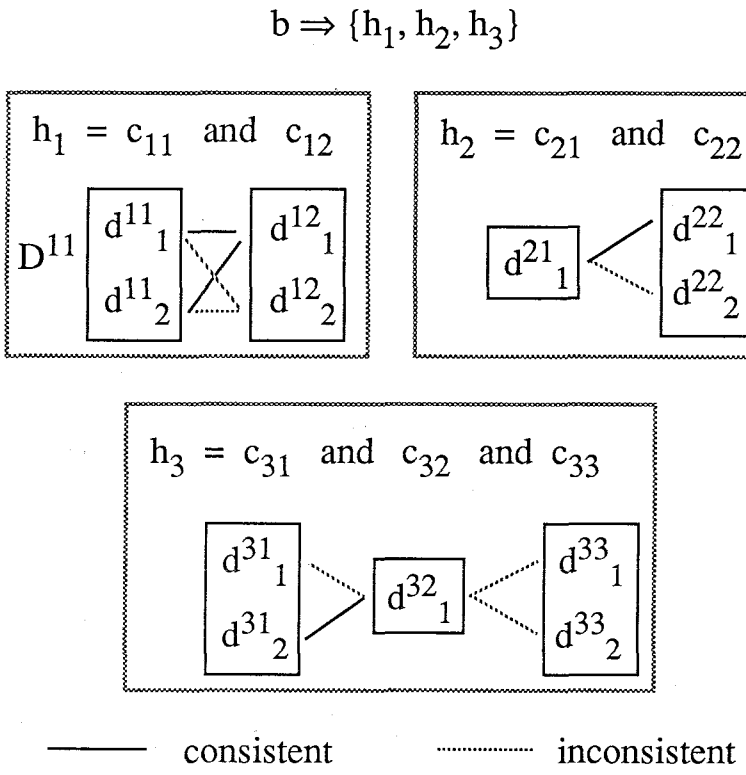


図 5.4: 単語列整合処理

の式により求められる．

次に，簡単な例を用いて，具体的に説明する．式 (2.3) が記述内容から妥当であるかどうかは，図 5.4に示すように，整合する記述内容候補の組合せが得られるかどうかにより判断できる．ここで，h_3 のように整合する組合せが存在しない仮説は，矛盾するものとして仮説木から除去できる．

一方，h_1，h_2 のように，整合する組合せが得られる仮説は，正しい可能性がある．しかしながら，どの仮説が正しいのか，また h_1 のように複数の組合せが得られる場合にどれが正しいものであるかを，現レベルで判断することはできない．但し，ベース b の記述内容は，以下の候補，

$$d_1^{b1} = and(d_1^{11}, d_1^{12}) \tag{5.38}$$

$$d_2^{b1} = and(d_2^{11}, d_1^{12}) \tag{5.39}$$

$$d_1^{b2} = and(d_1^{21}, d_1^{22}) \tag{5.40}$$

のいずれかであるといえる．そこで記述内容候補集合生成では，これらを用いて，ベース b に対する記述内容候補の集合，

$$D^b = \{d_1^{b1}, d_2^{b1}, d_1^{b2}\} \tag{5.41}$$

を作成し，処理を上位レベルに移行することにより，さらに整合性を検査する．以上の処理を文書レベルまで繰り返すと，名刺に対する無矛盾な記述内容候補が得られる．最終的に複数の記述内容候補が存在する場合には，各候補中の文字に着目し，文字認識時に得られた類似度の平均を求めることにより，最も信頼性の高いものを処理結果とする．ただし，未照合文字の類似度は 0 とする．その結果，図 5.5に示すような，名刺に対する構造化記述が得られる．

5.5 実験結果と検討

実験対象は，文書構造解析と同一のものである．また，構造化記述生成処理の対象とするデータは，文書構造解析処理により実験対象から得られた仮説木，文字列グラフである．

単語列生成処理における未照合文字の許容割合は，各単語に対して 50% 以内，単語列に対して 30% 以内とした．単語辞書については，住所辞書は対象の名刺に含まれる 9 都府県の全地名を用いて，その他の辞書は実験対象から単語を抽出して作成した．辞書中の総単語数は約 3 万語である．なお氏名については，単語列生成処理を行わずに文字認識候補第 1 位をそのまま採用した．一方，単語列整合処理では，約 1500 の整合性記述を使用した．

前章でも述べたように，構造化記述生成処理には，生成された仮説から正しいものを選択する，構成要素の記述内容を同定するという 2 通りの役割がある．仮説の選択については結果を表 5.1に示す．ここで，構成要素抽出率とは，仮説検証後に構成要素が正しく抽出された割合である．群レベルから項目群レベルまでは，信頼度と構成要素抽出率がほぼ等しいことから，仮説に含まれる正解の大半を正しく選択しているといえる．群レベル，項目群レベルにおいて構成要素抽出率が信頼度に比べて多少低下しているのは，ノイズ除去の失敗，文字のつぶれやかすれなどから記述内容の同定に失敗したためである．

項目レベルでは，101 個の抽出に失敗した．そのうち約 8 割は電話番号，郵便番号などの数字が主体の項目であった．単語列生成処理では，一般の辞

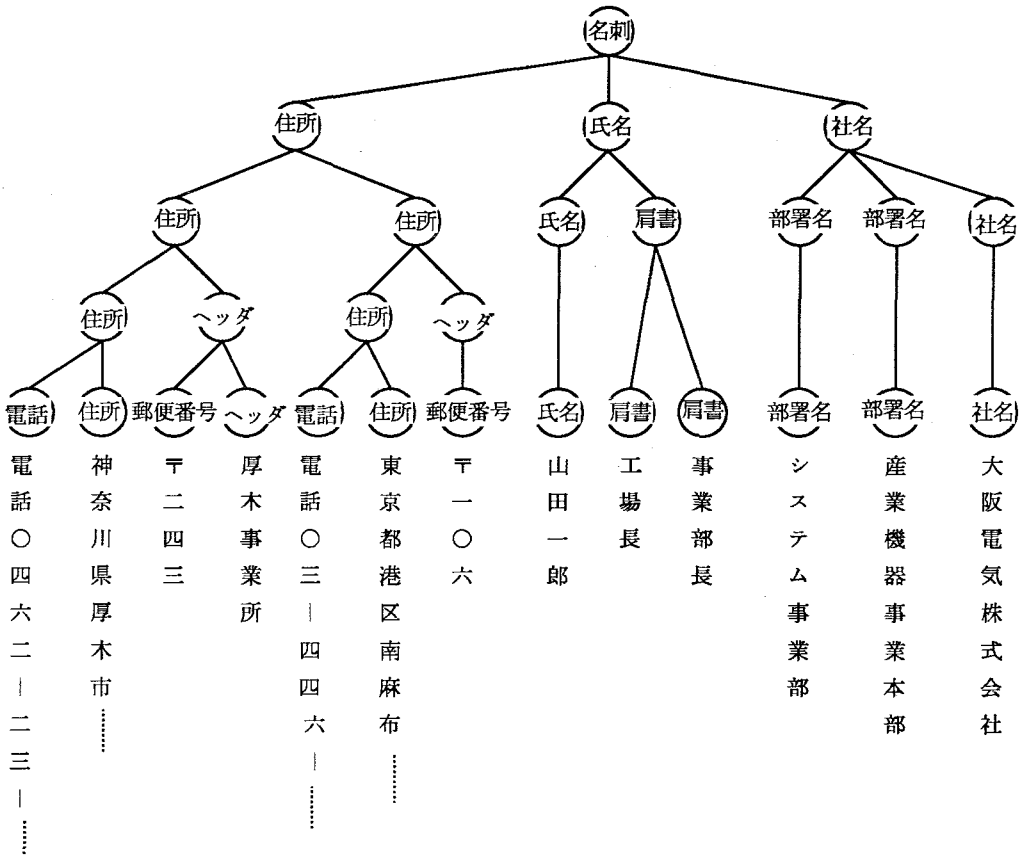

名刺
住所
氏名
社名
住所
住所
氏名
肩書
部署名
部署名
社名
住所
ヘッダ
住所
ヘッダ
電話
住所
郵便番号
ヘッダ
電話
住所
郵便番号
氏名
肩書
肩書
部署名
部署名
社名
電話〇四六二―二三―……
神奈川県厚木市……
〒二四三
厚木事業所
電話〇三―四四六―……
東京都港区南麻布……
〒一〇六
山田一郎
工場長
事業部長
システム事業部
産業機器事業本部
大阪電気株式会社

図 5.5: 構造化記述

表 5.1: 仮説検証結果

レベル	構成要素数	仮説生成			仮説検証
		平均候補数		信頼度	構成要素抽出率
文書	１００	−		−	−
群	３００	１．０		１００％	９９．０％
準群	５０７	１．４		９９．４％	９９．４％
項目群	１７０	３．９		９５．９％	９５．３％
項目	９７２	９．８		９９．２％	８９．６％
文字	８８１１	領域	１．３	９８．８％	９３．０％
		属性	８．３	９４．２％	

表 5.2: 記述生成結果

(a)社名および数字が主体の項目

項　目	社名	電話番号	ファックス	テレックス	郵便番号	内線	合計
構成要素数	100	142	10	25	119	5	401
同定率(%)	81.0	69.7	20.0	84.0	82.4	100	76.3

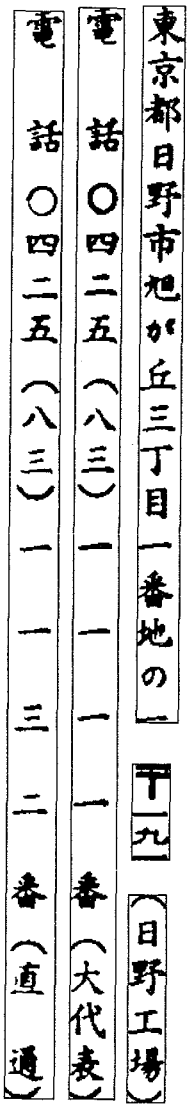
(b)その他の項目

項　目	部署名	肩書	氏名	見出し	住所	ビル名	合計
構成要素数	136	130	100	41	119	45	571
同定率(%)	94.9	98.5	89.0	87.8	93.3	100	94.2

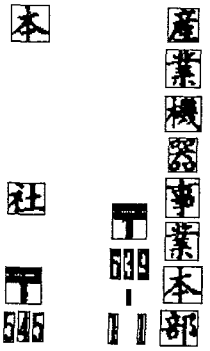
書単語に対しては単語照合時に未照合文字を許しているが，数字部分に対しては，意味をなさないために許していない．従って，数字部分の候補文字中に正解が存在しない場合には，単語列生成処理が失敗する．これが主な原因であると考えられる．図 5.6 (a) に住所項目群に対する成功例を示す．この例は，文字列の縦幅が同じであるにもかかわらず，右側の文字列のみが3つの項目を含むものである．電話番号の数字部分のように，文字間の距離が項目間よりも大きいものがあることを考えると，レイアウト構造の考慮のみでは，右側の文字列から項目を抽出することが困難であるといえる．本手法では，記述内容を考慮することにより，項目をすべて正しく抽出している．

文字レベルでは，78.7% の認識率が 93.0% まで向上した．この結果から，接続性を用いた単語列生成処理法の有効性が確かめられたと考えられる．なお失敗の主な原因は，数字，特殊文字の誤認識である．図 5.6 (b) に見出し項目群に対する成功例を示す．この例は，郵便番号部に縦書きと横書きが混在しており，また文字のつぶれが見受けられるにもかかわらず，すべての文字が正しく切り出され，かつ認識されている．

記述内容の同定については，表 5.2に項目別の同定率を示す．記述内容の同定とは，項目の属性が正しく抽出され，かつ項目内の文字がすべて正しく認識されることを意味する．処理結果の利用を考えると，記述内容の同定率



(a) 項目レベル



(b) 文字レベル

図 5.6: 構成要素抽出成功例

は，構成要素抽出率と並んで文書画像理解手法の有効性を判断する重要な指標であると考えられる．表 5.2 (a) に示す社名および数字主体の項目に対しては，構成要素抽出の場合と同様の原因から同定率が低下しているが，それらを除く項目に対しては，表 5.2 (b) に示すように高い同定率が得られている．項目全体で見れば，抽出率 89.6% に対して同定率が 86.8% となることから，正しく抽出された項目の大半に対して，記述内容が同定されているといえる．

文書構造解析，および構造化記述生成の実験結果から，本システムでは，レイアウト構造に関する知識および記述内容の論理的制約に関する知識の双方を考慮するため，画像の品質が低くかつ複雑な構造を有する名刺などの文書に対しても，安定して構造化記述を生成することが可能であるといえる．

5.6 結言

本章では，記述内容の論理的制約に関する知識を用いて，文書構造解析において生成された仮説木から妥当なものを選択する手法について述べた．また，実験結果から，本手法が仮説の選択および記述内容の同定の双方に対して有効であることを示した．

本手法の主な特徴は，

1. 項目候補を用いることにより，効率良く項目の記述内容候補を生成すること

2. 特殊候補に対しても始端の仮定などにより，記述内容候補を生成することが可能となる他，記述内容候補を用いて仮説木を変更することにより，後の処理の無矛盾性を確保すること

3. 仮説木および矛盾データベースを利用して記述内容候補の組合せに制限を加えることにより，上位レベルの構成要素に対する記述内容候補を効率よく生成すること

4. 発見された矛盾に基づき，仮説木から仮説を削除することにより，処理の無矛盾性を保ち，かつ後続の処理の効率を向上させること

の 4 点である．

得られた構造化記述は，本システム全体の出力でもあるため，本章で示した実験結果から本システムの有効性を検証できたと考える．

第 6 章

結論

文書画像理解を実現するためには，対象文書のレイアウト構造に関する知識，記述内容の論理的制約に関する知識の双方を有効利用する必要があるとの観点から，本論文では，知識利用型文書画像理解システムを提案した．本システムの主な特徴は，

1. 文書画像理解に必要な知識が対象文書に強く依存することを考え，処理を担うコアシステムから分離して蓄積することにより，システムの適用容易性を高めていること

2. レイアウト構造に関する知識，記述内容の論理的制約に関する知識の特性を考慮し，それぞれに応じた知識記述法を採用することにより，システムの適用可能性を高めていること

3. レイアウト構造のバリエーション，論理的制約の弱さ，および処理の不完全性に対処するため，処理戦略として仮説生成検証プロセスを導入し，システム全体のロバスト性を向上させていること

の３点である．これらの特徴により，本システムは従来システムに比べて，より高い適用性を持つことが可能となる．また，名刺を対象とした実験の結果から，本システムの有効性を検証した．

以下に本研究で得られた諸結果をまとめる．

第 2 章では，文書画像理解に関連した既存手法を概観し，文書画像理解システムを構築する際に留意しなければならない問題点として，文書画像理解に有効となる知識の種類，処理の不完全性への対処法，対象依存性への対処法の３点を指摘した．また，その結果に基づき，本システムの概念設計を行った．本システムは，知識としてレイアウト構造に関する知識に加え，論

理的制約に関する知識を用いること，対象に依存する知識を処理部から分離して蓄積するという分離型のシステム構成を採用すること，仮説生成検証プロセスの導入により処理の不完全性に対処することを特徴とするものである．

第 3 章では，対象を項目主体の文書に限定し，レイアウト構造に関する知識，および記述内容の論理的制約に関する知識の具体的な記述法について詳述した．レイアウト構造の記述法としては，フレーム表現の導入による階層構造の明示的な表現，レイアウト述語による記号的知識記述の実現，および上位下位関係の導入によるバリエーションのコンパクトな表現を特徴とする手法を提案した．また，論理的制約に関する知識の記述法としては，単語の接続性，単語列の整合性の 2 種類の特徴から，構成要素の記述内容の妥当性を表現した．さらに，縦書き名刺に対する記述例の考察，および他手法との比較検討から，知識の表現能力，記述容易性，可読性に関する本手法の優位性を示した．

第 4 章では，文書構造解析を仮説生成として位置づけ，レイアウト構造に関する知識を満足するすべての構成要素候補を網羅する手法を提案した．本手法は，レイアウト構造の階層性に着目することにより，処理効率を落とすことなく，候補を生成するものである．候補生成の際に，仮説間の依存関係や矛盾を記録することにより，処理の無矛盾性を保証するという特徴も合わせ持つ．また，文字切り出し法としては，エキスパートシステム的な観点から，不定ピッチの文字列に対しても柔軟に動作する手法を提案した．さらに，実験において平均候補数，信頼度なる 2 種類の評価基準により本手法の仮説生成能力を評価し，本手法が名刺のように複雑な構造を有する文書に対しても有効であることを示した．

第 5 章では，論理的制約に関する知識を用いて，構造化記述を生成しつつ仮説を検証する手法について述べた．本手法は，前章において記録された仮説間の依存関係，および矛盾を利用することにより，候補の組合せ的爆発を回避し，効率的に構造化記述を生成するという特徴を有する．また，構成要素候補に付与された属性に基づいて使用する知識を選択することにより，一層，処理効率を高めている．さらに，検証結果に基づいて不適当な仮説を削除することにより，処理の無矛盾性を保証している．実験において，各レベルにおける構成要素抽出率，および項目の記述内容の同定率という 2 種類の評価基準を用いて本手法を検討した結果，本手法の有効性を確かめた．

以上のように，文書画像理解に種々の知識を利用する本システムが，高い有効性を持つことを示した．最後に，本研究に関する今後の課題について述べる．

まず，第1は，処理効率の一層の向上である．本システムでは，必要に応じて対象レベルごとに知識を利用する方式を採用している．このような方式では，単純なモデルインタープリタにより，知識と処理の明確な分離が可能となる．しかしながら，同時に1レベルの知識しか考慮することができないため，人間には明らかに不適当と考えられる仮説を生成し，下位レベルにおいて棄却するという動作が見受けられる．筆者らは，対処法として，モデルインタープリタに高度の空間推論能力をもたせることにより，複数レベルを同時に考慮し，不適当な仮説生成を抑制する手法 [60] を開発中であり，今後，システムへの導入を図っていきたいと考えている．

第2は，処理速度の改善である．本システムは処理の必要に応じて知識を随時利用するため，画像理解システムの分類上ではインタープリタ方式に相当すると考えられる．このような方式では，知識ベースと処理部のやりとりが多いほど，処理速度が低下するという問題点が指摘されている．画像理解の分野では，対処法として，知識ベースから処理手続きを生成するというコンパイラ方式の研究がなされている [61]．本システムに対しても，同様に文書モデルから処理手続きを生成することが可能となれば，処理速度の大幅な改善が期待できる．

第3は，文書画像理解に必要となる知識の獲得方法である．本システムでは記述容易性の高い知識表現を用いていることから，従来手法に比べて対象文書の変更が容易であるといえる．しかしながら，知識を記述するのはあくまでも利用者であるため，十分な処理能力を得るには，対象文書の変更に際して複数の例を考慮しつつ知識を新たに記述し，さらに処理実験を重ねることにより知識を追加，変更する必要がある．現在，画像理解の分野においては，画像例および処理誤りから知識を獲得する研究 [62] が端緒についたばかりであり，今後の発展を期待すると共に，文書画像に適した手法を開発する必要があると考える．

第4は，文書中の文字の他，図，表などの表現する情報を考慮した文書画像理解への発展である．最近，図や表など個々の解析・理解については，研究成果が報告されている (例えば，[63])．今後は，個々の解析・理解の高精度化に加え，文字の情報とのかかわりを総合的な立場から検討する必要があろう．

現在，文書やシーンなどの画像，あるいは音声を対象としたパターン理解システムは，残念ながら実用に耐え得るレベルには到達しておらず，当然ながら人間の能力には程遠いと考えられる．今後，パターン理解システムを発展させるためには，人間の有する知識の質および量，推論・制御メカニズムに対する研究を積み重ねていく必要がある．本論文がそうした研究の発展の

一助となれば幸いである.

謝辞

本研究の全過程を通じ, 懇切なる御指導, 御鞭撻を賜った大阪大学工学部通信工学教室手塚慶一教授に心より御礼申し上げる.

本論文作成に際し, 御助言, 御教示を賜った大阪大学産業科学研究所北橋忠宏教授に深く感謝する.

大阪大学工学部, 同大学院において御指導, 御教示を賜った大阪大学熊谷信昭総長, 大阪大学工学部通信工学教室滑川敏彦名誉教授, 中西義郎名誉教授, 倉薗貞夫教授, 森永規彦教授に厚く御礼申し上げる.

本研究に関し, 直接御指導頂き, また終始有益な御助言, 御討論を頂いた大阪大学工学部馬場口登講師に深謝する.

本研究分野への最初の手ほどきを与え, また御指導下さった大阪大学経済学部真田英彦教授に心から感謝する.

筆者が大阪府立大学に勤務以来, 御指導, 御支援頂く大阪府立大学西田富士夫名誉教授, 米田正次郎教授, 日下浩次助教授, 高松忍講師に御礼申し上げる.

筆者の所属していた手塚研究室の中西暉助教授 (現, 大阪大学言語文化研究科教授), 岡田博美助教授, 山本幹助手, 後藤嘉代子技官, および和歌山大学経済学部内尾文隆講師, 神戸商船大学井上健助教授には, 適切な御助言を頂いた. 並びに, 神戸商船大学田中直樹助教授, シャープ株式会社北村義弘博士には, 研究遂行にあたり種々の面でお世話になった. 特に手塚研究室パターン情報処理グループの桜井善紀氏 (現, 野村総合研究所), 山田耕児氏 (現, 毎日放送), 杉山淳一氏 (現, 富士通研究所), 山岡正輝氏 (現,NTT データ), 百田賢一氏には, 本研究の細部にわたり熱心に御討論頂いた.

最後に, 筆者が手塚研究室に在籍中から同期生として種々の面でお世話になる大川剛直氏 (現, 大阪大学工学部助手), 金錫泰博士 (現, 釜山水産大学校工科大学講師), 小林真也博士 (現, 金沢大学工学部助手) に感謝したい.

ここに記して, 以上の方々に深甚なる感謝の意を表する.

参考文献

[1] 森俊二: “文字・図形認識技術の基礎”, エレクトロニクス文庫, オーム社 (1984).

[2] 豊田順一, 野口要治, 西村康: “日本語印刷文書における文字切り出し —新聞自動読み取りへの応用—”, 情報処理学会論文誌, **24**, 4, pp. 481–487(1983).

[3] 秋山照雄, 内藤誠一郎, 増田功: “非接触文字優先切出しによる印刷物からの文字切出し法”, 電子通信学会論文誌 D, **J67-D**, 1, pp. 1194–1201 (1984).

[4] 新谷幹夫, 梅田三千雄: “文字認識における複合後処理法の能力評価”, 電子通信学会論文誌 D, **J68-D**, 5, pp. 1118–1124(1985).

[5] 杉村利明: “候補文字補完と言語処理による漢字認識の誤り訂正法”, 電子情報通信学会論文誌 D(II), **J72-D-II**, 7, pp. 993–1000(1989).

[6] 高尾哲康, 西野文人: “日本語文書リーダ後処理の実現と評価”, 情報処理学会論文誌, **30**, 11, pp. 1394–1401(1989).

[7] T. Akiyama and N. Hagita: “Automated entry system for printed documents”, Pattern Recognition, **23**, 11, pp. 1141–1154(1990).

[8] D. G. Elliman and I. T. Lancaster: “A review of segmentation and contextual analysis techniques for text recognition”, Pattern Recognition, **23**, 3/4, pp. 337–346(1990).

[9] 佐藤和洋, 絹川博之, 大町一彦: “オフィス文書の標準化と文書データベースの研究動向”, 情報処理学会誌, **28**, 6, pp. 710–719(1987).

[10] 曽根原登: “ドキュメントアーキテクチャの標準化動向”, 情報処理学会, 文書処理とヒューマンインタフェース研究会報告, **17-1**, pp. 1–10(1988).

[11] 野口健一郎, 大谷真: “OSI の実現とその課題”, 情報処理学会誌, **31**, 9, pp. 1235–1244(1990).

[12] K. Kise, J. Sugiyama, M. Yamaoka, K. Momota, N. Babaguchi and Y. Tezuka: “Model based understanding of document images”, Proceedings of IAPR Workshop on Machine Vision Application '90, pp. 471–474 (1990).

[13] K. Kise, K. Yamada, N. Tanaka, N. Babaguchi and Y. Tezuka: “Visiting card understanding system”, Proceedings of 9th International Conference on Pattern Recognition, pp. 425–429(1988).

[14] 黄瀬浩一, 馬場口登, 手塚慶一: “名刺画像認識システムにおける文書モデル”, 昭和 63 年電子情報通信学会春季全国大会, D-480(1988).

[15] 山岡正輝, 黄瀬浩一, 馬場口登, 手塚慶一: “文書画像理解のためのモデル記述”, 1989 年電子情報通信学会春季全国大会, D-476(1989).

[16] 山岡正輝, 黄瀬浩一, 馬場口登, 手塚慶一: “文書画像理解におけるドメイン知識記述の一手法”, 電子情報通信学会技術研究報告 PRU89-75, 電子情報通信学会 (1989).

[17] 田中直樹, 黄瀬浩一, 馬場口登, 手塚慶一, 梶谷浩二: “名刺画像認識システムにおける単語処理”, 昭和 63 年電子情報通信学会春季全国大会, D-482 (1988).

[18] 杉山淳一, 黄瀬浩一, 馬場口登, 手塚慶一: “文書画像理解における単語情報と論理構造の援用法”, 電子情報通信学会技術研究報告 PRU89-90, 電子情報通信学会 (1990).

[19] 百田賢一, 黄瀬浩一, 馬場口登, 手塚慶一: “単語間の接続性に着目した文字認識後処理”, 1990 年電子情報通信学会秋季全国大会, D-363(1990).

[20] 黄瀬浩一, 杉山淳一, 馬場口登, 手塚慶一: “レイアウトモデルに基づく文書構造解析”, 電子情報通信学会論文誌D (II), **J72-D-II**, 7, pp. 1029–1039(1989).

[21] 杉山淳一, 黄瀬浩一, 馬場口登, 手塚慶一: “文書モデルを用いた文書画像の構造解析”, 1989 年電子情報通信学会春季全国大会, D-477(1989).

[22] 山岡正輝, 黄瀬浩一, 馬場口登, 手塚慶一: “知識ベース型文書構造解析システムの汎用性に関する一考察”, 電子情報通信学会技術研究報告 PRU90-119, 電子情報通信学会 (1991).

[23] 黄瀬浩一, 山田耕児, 田中直樹, 真田英彦, 手塚慶一: “名刺画像認識システム (1) — 項目仮説生成 —”, 情報処理学会第 35 回全国大会, 1H-8 (1987).

[24] 黄瀬浩一, 杉山淳一, 馬場口登, 手塚慶一: “名刺画像認識システムにおける項目仮説生成”, 昭和 63 年電子情報通信学会春季全国大会, D-481 (1988).

[25] 黄瀬浩一, 山田耕児, 田中直樹, 馬場口登, 手塚慶一: “名刺画像認識における項目仮説生成”, 電子情報通信学会技術研究報告 PRU87-88, 電子情報通信学会 (1988).

[26] 黄瀬浩一, 馬場口登, 手塚慶一: “文書画像理解のための仮説検証の一手法”, 1990 年電子情報通信学会秋季全国大会, D-386(1990).

[27] 黄瀬浩一, 馬場口登, 手塚慶一: “文書画像理解システムの制御戦略の一考察”, 1989 年電子情報通信学会春季全国大会, D-478(1989).

[28] 黄瀬浩一, 馬場口登, 手塚慶一: “文書画像理解における推論と制御の一提案”, 電子情報通信学会技術研究報告 PRU89-76, 電子情報通信学会 (1989).

[29] D. E. Knuth: “The TEXbook”, Addison-Wesley Publishing Company, Inc.(1983). (斉藤信男 監修, 鷺谷好輝 翻訳: “TEX ブック”, アスキー出版局).

[30] 馬場口登, 塚本正敏, 相原恒博: “認識処理の導入による手書き文字切り出しの一改良”, 電子通信学会論文誌 D, **J69-D**, 11, pp. 1774–1782 (1986).

[31] 村瀬洋, 新谷道夫, 若原徹, 小高和己: “言語情報を利用した手書き文字列からの文字切り出しと認識”, 電子通信学会論文誌 D, **J69-D**, 9, pp. 1292–1301(1986).

[32] C. H. Wang, P. W. Palumbo and S. N. Srihari: "Performance of a system to locate address blocks on mail pieces", Proc. AAAI-88, pp. 837–841(1988).

[33] K. Kubota, O. Iwaki and H. Arakawa: "Document understanding system", Proc. 7th ICPR, pp. 612–614(1984).

[34] 松山隆司: "知識型ビジョンの展望", 電子情報通信学会誌, **70**, 10, pp. 1045–1052(1987).

[35] 松山隆司: "画像理解における推論方式", 人工知能学会誌, **4**, 1, pp. 21–29(1989).

[36] 古川康一, 溝口文雄 (編): "知識プログラミング", 知識情報処理シリーズ 8, 共立出版 (1986).

[37] 黄瀬浩一, 馬場口登, 手塚慶一: "文書画像理解のためのモデル記述に基づく推論", 1990 年電子情報通信学会春季全国大会, SD-11-7(1990).

[38] 秋山照雄, 増田功: "周辺分布, 線密度, 外接矩形特徴を併用した文書画像の領域分割", 電子通信学会論文誌 D, **J69-D**, 8, pp. 1187–1195(1986).

[39] 田中昌也, 西武男, 富永英義: "文書画像の書式解析", 電子通信学会技術研究報告 PRU86-115, 電子通信学会 (1986).

[40] 辻本修一, 麻田治男: "文書画像理解による記事の自動抽出", 電子情報通信学会技術研究報告 PRU87-89, 電子情報通信学会 (1987).

[41] D. S. Yeh, S. Antoy, A. Litcher and A. Rosenfeld: "Address location on envelopes", Pattern Recognition, **20**, 2, pp. 213–227(1987).

[42] D. Niyogi and S. N. Srihari: "A rule-based system for document understanding", Proc. AAAI-86, pp. 789–793(1986).

[43] 東野純一, 藤澤浩道, 中野康明, 江尻正員: "矩形領域の集合表現に基づく知識表現言語 FDL と文書画像理解への応用", 電子通信学会技術研究報告 PRU86-31, 電子通信学会 (1986).

[44] 西村康, 高橋友一, 小林幸雄: "木構造モデルによる文書画像からの検索情報抽出", 電子情報通信学会技術研究報告 PRU89-34, 電子情報通信学会 (1989).

[45] 駱琴, 渡邊豊英, 吉田雄二, 稲垣康善, 斉藤隆夫: “知識ベースに基づいた図書目録カードの理解”, 情報処理学会論文誌, **31**, 12, pp. 1755–1767 (1990).

[46] A. Dengel and G. Barth: “Anastasil: A hybrid knowledge-based system for document layout analysis”, Proc. IJCAI-86, pp. 1249–1254(1989).

[47] F. Esposito, D. Malerba and G. Semeraro: “An experimental page layout recognition system for office document automatic classification: An integrated approach for inductive generalization”, Proc. 10th ICPR, pp. 557–562(1990).

[48] 中野康明, 藤澤浩道: “自動ファイリングのための文書理解の一方式”, 電子情報通信学会論文誌 D, **J71-D**, 10, pp. 2050–2058(1988).

[49] 長谷博行, 米田政明, 酒井充, 吉田順作: “図書目録カードの自動項目分類について”, 電子情報通信学会論文誌 D, **J70-D**, 8, pp. 1579–1588(1987).

[50] 桜井善紀, 黄瀬浩一, 村瀬宏一, 田中直樹, 真田英彦, 手塚慶一: “名刺画像認識システムに関する一考察”, 電子通信学会技術研究報告 PRL85-62, 電子通信学会 (1986).

[51] 上野春樹: “知識工学入門”, オーム社 (1985).

[52] 白井良明 (編): “パターン理解”, 知識工学講座 9, オーム社 (1987).

[53] 佐藤道弘, 木田博巳: “不定ピッチ文字列を含む印刷文書における文字切出し法”, 電子情報通信学会技術研究報告 PRU88-159, 電子情報通信学会 (1988).

[54] 松井正一, 岩城修, 木田博巳: “文書認識のためのプロダクションシステムの構成”, 電子通信学会技術研究報告 PRL84-76, 電子通信学会 (1984).

[55] 山田耕児, 黄瀬浩一, 田中直樹, 真田英彦, 手塚慶一: “名刺画像認識における文字切り出し・認識に関する一検討”, 昭和 62 年電子情報通信学会創立 70 周年記念総合全国大会, 1505(1987).

[56] 山田耕児, 黄瀬浩一, 田中直樹, 真田英彦, 手塚慶一: “名刺画像認識システム (2) — 文字矩形・文字仮説生成 —”, 情報処理学会第 35 回全国大会, 1H-9(1987).

[57] 馬場口登, 山田耕児, 黄瀬浩一, 手塚慶一: “コネクショニストモデルによる画像2値化の実験的検討”, 電子情報通信学会論文誌D (II), **J73-D-II**, 7, pp. 1281–1287(1990).

[58] N. Babaguchi, K. Yamada, K. Kise and Y. Tezuka: “Connectionist model binarization”, Proceedings of 10th International Conference on Pattern Recognition, pp. 51–56(1990).

[59] N. Babaguchi, K. Yamada, K. Kise and Y. Tezuka: “Connectionist model binarization”, International Journal of Pattern Recognition and Artificial Intelligence(to be published).

[60] 山岡正輝, 黄瀬浩一, 馬場口登, 手塚慶一: “文書画像理解におけるモデルインタープリタの一改良”, 1990年電子情報通信学会秋季全国大会, D-385(1990).

[61] 池内克史: “物体認識と認識プログラムの自動生成”, 人工知能学会誌, **4**, 1, pp. 30–42(1989).

[62] 松原仁, 坂上勝彦, 山本和彦, 山岸健太郎: “画像学習システムMIRACLE IVにおける機能的特徴と視覚的特徴の対応付け”, 情報処理学会論文誌, **31**, 9, pp. 1302–1311(1990).

[63] 森本正志, 有木康雄, 坂井利之: “図面の構成知識を用いた図面構造解析”, 電子通信学会技術研究報告 PRU86-12, 電子通信学会 (1986).