

和欧混合組版機能を持つインテリジェントプリンタ第1版の実現

<sup>+</sup>  
関口治, 中川正樹, 高橋延匡

東京農工大学 工学部 数理情報工学科

<sup>\*)</sup>現在 日立製作所 機械研究所

本報告は、科学技術文書を対象とした日本語フォーマッタの開発に関するものである。試作したフォーマッタは、出力装置として解像度が240ドット/インチであり、A3サイズの用紙まで出力が可能なレーザービームプリンタ上に作成した。

このフォーマッタの特徴は、組版の中で最も難しいと言われる和欧混合組版を実現していることである。採用した組版規則は、写真植字(写植)で用いられている組版規則、すなわち行頭及び行末禁則処理の規則に完全に準拠している。行頭や行末禁則処理は、追込みを基本としており、追込みを行なう際の文字の選択やツメの量、そして、これらの優先順位も写植で用いられているものと同じである。英字の印字については、全角と半角の出力が可能であり、全角の英字が連なるときはプロポーションナルピッチを施している。

フォーマッタの入力ファイルは、JIS C 6237(日本語文書交換用ファイル仕様)に基づくものであるため、日本語ワードプロセッサで作成した文書を禁則処理をして右ぞろえをした形式で出力できる。なお、出力対象文字は、JIS第1, 第2水準の文字セットである。

“ An Intelligent Typesetting printer for Japanese Document including English words”  
(In Japanese)

<sup>+</sup>  
OSAMU SEKIGUCHI, MASAKI NAKAGAWA, NOBUMASA TAKAHASHI

Dept. of Information Science, Tokyo Univ. of Agriculture and Technology

<sup>\*)</sup>Mechanical Engineering Research Laboratory, Hitachi Ltd

This paper presents the first version of a Japanese Output Server with H0spitality (JOSH0). JOSH0 is an intelligent formatting machine for scientific Japanese documents. The system is composed of a dedicated MC68000, Mega byte frame memory and LBP. In the control memory of the MC68000, resides formatting software which perfectly obeys typesetting rules on Japanese documents with English words. When a special letter comes at a line head or tail, so-called run-on takes place. In such cases, where, how much and in what priority the run-on is made is determined according to Japanese photocomposition rules.

## 1. はじめに

日本語ワードプロセッサ（以下、ワープロ）の登場により、従来の文書作成あるいは文書の保管の方法が一新された。例えば、最近の情報処理学会の全国大会講演論文集を見ると、その大部分の論文はワープロやワープロソフトを用いて作成されていることが分かる。このような中で、手書き文字で書かれている論文を見ると思わず、“珍しい”と思ってしまう。数年前の論文集では、手書きのものが大半を占め、手書き以外のものは、英文フォーマッタを使用して作成されたと予想される英語の論文だけであったと思う。この当時は、計算機を用いて日本語の論文を書くということは皆無に近かった。

数年間でワープロが普及した最大の理由は、ワープロが単に筆記用具という性格だけでなく、文書の保管倉庫という性格も同時に持っているからである。すなわち、文書の再利用や文書の部品化という計算機で文書を扱う最大の利点が発揮されているのである。

しかし、便利なワープロに対して不満がないわけではない。今、ワープロを使用している人が一番強く抱いているニーズは、“活版印刷や写植機で実現されている印字品質としっかりした組版”ではないだろうか。つまり、現状のワープロ出力では飽き足らず、もう1ランク上の高印字品質や良質の組版を望んでいるわけである。現状のワープロで用いられている禁則処理の規則は、ぶら下がり組みと呼ばれるものである。この組版は、行頭禁則文字の中で句読点に限って、行末からはみ出して組む方法である。この方式では、満足できない人が多いに違いない。多くの人が、行末からはみ出したテンやマルを修正液で消して、自分の手でテンやマルを詰めて、自分のイメージした文書を作り上げて行ったことと思う。このような経験を持った人は、一般の書物で用いられている組版規則で出力が可能なワープロを待ち望んでいるに違いない。ワープロ出力に対するニーズや期待は大きく、出力の品質がワープロの雌雄を決する次の世代がそこまできていると言えよう。

## 2. 出力系への要求

くりかえし述べるが、ワープロのユーザが抱いている最大の要求は、一步進んだ“より美しい文書出力”である。この“より美しい文書出力”を実現するには、次のような重要な問題を解決しなければならない。

### (1) 印字品質の問題

現在のほとんどすべてのワープロの文字は、 $16 \times 16$ ないしは $24 \times 24$ のドットマトリックスから構成されるフォントである。我々は、このような文字フォントでは満足しない。より一步進んだ、活字や電算写植で実現されるような印字品質であるワープロを望んでいる。

### (2) 禁則処理の問題

禁則処理の中でワープロで実現されているものは、ぶら下がり組みによる禁則処理にしか過ぎない。このような一部の禁則処理しかできない文書作成機では、しっかりした本や論文は書けない。行頭禁則処理や行末禁則処理を実現して欲しいものである。

### (3) 和欧混合組版の問題

我々のようなソフトウェア技術者が作成する文書は、必ずといっていいほど英数字が登場し、日本語と英数字が混在したものとなる。我々の作成した文書を、我々の満足のいく組版で出力しようとするれば、現状のワープロでは、永遠にその出力を得ることができない。それは、組版規則に基づく和欧混合組版が不可能なためである。

ワープロでの和欧混合組版の仕方は、英数字を半角文字にして半角単位で文字を並べることによって行なう方法である。この方法では、英字のプロポーションピッチさえもできない。現状では、残念ながら、せっかくのビットマップが有効に生かされていない。

### (4) 出力可能な字種の問題

理工系の文書にとって、数式は必要不可欠なものである。数式を書く場合には、多種多様な文字が要求される。例えば、微分や積分などの記号、添字に使用される小さい英数字やギリシア文字、あるいは逆に強調したために使用する大きな文字やイタリック体などである。現在のワープロでは、これらの文字の限られたものしか出力できない。

このように出力系に対して高度の要求が発生するのは、それだけワープロ出力に対して目が肥えてきたこと、また仮名を漢字に変換して入力する方式が定着しつつあり、そして実際に使用して、ある程度満足しているからであろう。しかし見方を変えれば、人間が文字を覚えて書けるようになったが、自分の字のきたなさにいや気がさし、誰にもせよきれいだといわれるような文書を書きたいと願うそんな時期にあるかも知れない。

## 3. 我々の目指す日本語文書作成システム

我々は、自分自身を対象にした日本語文書作成システムの構築を進めている。このシステムは、特に、ソフトウェアドキュメントの作成から管理に至るライフサイクルを支援することを目標としている。

ソフトウェアの研究者は、さまざまな計算機システムにいろいろなソフトウェアを開発し、また、そのソフトウェアを運用・保守する必要がある。ここで、開発から保守に至るまでのライフサイクルを大きく左右するものは、そのソフトウェアシステムのドキュメンテーションがなされているか否かという点である。もし、そのシステムのドキュメントが十分に完備されていれば、システムの寿命は長期にわたるであろうし、そうでなければ、そのシステムを設計した研究者自身でさえも使えないシステムになってしまうであろう。すなわち、ソフトウェアには、ドキュメンテーションは不可欠なものであり、ドキュメントのないものはシステムとは言えないと言っても過言ではない。

ソフトウェアのドキュメントには、“理解の容易性”や“高い信頼性”が要求される。それは、あるシステムのドキュメントを読んでもそのシステムが何をやっているかが理解できなかったり、あるいは、まちがいはばかり記述してあつりしてはドキュメントの意味がないからである。質の良いドキュメントを作りたい、あるいは、得たいという願いは切実であり、現在までに多くの手法がソフトウェア工学の分野で提案され実行されてきた。ソフトウェアには、ハードウェアで使用されているような回路図に相当するものがないため、プログラムの構造や動作がわかりにくいと言われている。このことを解決するために、フローチャートやNSチャート、あるいはPADなどの図面が開発され使用されるようになった。また、プログラムの設計法についても、構造化プログラミング技法の利用などが叫ばれている。

しかし、このドキュメントの質を決定する本質的な問題は、ソフトウェアの研究者が日常使用している自然言語で手軽にかつ迅速にドキュメントを作成できる環境があるか否かという問題に帰着するのではないだろうか。つまり、我々日本人が思考し、情報を交換する手段である日本語でドキュメントを記述できる環境を構築することが不可欠である。しかし、ただ、日本語でドキュメントを作成するということではなく、システムの実現手段として用いたプログラムと同じ世界で日本語を使ってドキュメントが作成できなければ意味がない。それは、ソフトウェアは常に成長するものであり、プログラムのバージョンアップは必至だからである。このプログラムのバージョンアップにドキュメントが追いつかなければ、何のためのドキュメントかがわからない。これは、プログラムとその日本語ドキュメントを一体管理しなければならないことを示唆している。

我々は、以上のような認識を持って、我々自身の研究用計算機システム上に自らの手で日本語文書作成システムの構築を決意し、このシステムの開発計画をたてた。この計画の主目標は、

- (1) 素人でも容易に漢字仮名混じり文を入力できる方式の研究開発【入力法】
- (2) 日本語文書に向けた原稿編集方式の研究開発【処理法】
- (3) 高印字品質の文書出力が得られる文書出力システムの研究開発【出力法】

である。

本報告は、(3)の出力法に関するものである。この日本語文書出力システムの研究では、JOSHO(浄書:Japanese Output Server with H0spitality)計画[1][2][3]と題して、レーザビームプリンタを使用し、プリンタの解像度を最大限生かした formatting machine の研究開発を行なっている。つまり、単なるプリンタの解像度を考慮した日本語フォーマッタというソフトウェアの開発ではなく、ソフトウェア・ハードウェアを含めた計算機による印刷システムである formatting machine の開発を目標としているわけである。

JOSHO計画は、インテリジェント化したパフォーマンスの高い出力専用システムの構築を目指している。このため、ただ単に画面をコピーするといったプリンタの使い方ではなく、圧縮した情報(具体的には文字コード)に基づいた美しい文書出力をインテリジェントプリンタから出力したいと考えている。これは、機能分散・負荷分散の考え方に基づいている。今までの計算機システムにおけるプリンタの制御は、メインのプロセッサが主に行なってきたが、半導体価格の低下により、高性能のマイクロプロセッサや大容量のメモリを用いたインテリジェントデバイスの実現が可能になってきている。

#### 4. 日本語フォーマッタの設計と実現

今回我々は、レーザビームプリンタをインテリジェント化し、その上に和欧混合組版を実現したフォーマッタを作成した。このフォーマッタの入力ファイルは、JIS C 6237(日本語文書交換用ファイル仕様)という規格で定められたものである。このため、ワープロで作られた文書をそのまま入力ファイルとして利用できる。

##### 4.1 設計方針

我々自身が作成する文書を整形出力するフォーマッタの現にあたり、つぎのような目標を設定した。

- (1) 和欧混合組版を実現すること
- (2) 日本語独特の処理を考慮すること
- (3) 出力装置の機能を最大限に生かすこと
- (4) ワープロで作成した文書ファイルを入力ファイルとして扱えるようにすること

以下、それぞれの項目について説明する。

##### (1) 和欧混合組版を実現すること

我々が一番出力したい文書は、仕様書や論文などの文書である。これらの文書には、英字が欠かせない。この英字についても、日本語と同様に美しく出力したいと考えている。このことは、英字のプロポーショナルピッチを実現しなければならないことを意味している。その結果、組版の中で、最もむずかしいと言われる“和欧混合組版”に挑戦しなければならない。しかし、だからこそ intelligent formatting machine として構成する意味がある。

##### (2) 日本語独特の処理を考慮すること

高解像度のプリンタで文字が出力できるようになれば満足するというものではない。単に文字を美しく出力するのではなく、書かれている言語独自の特徴を生かした美しい出力を得たいのである。我々が出力したい文書の多くは日

本語文書である。この日本語文書を美しく出力したいのである。そのためには、日本語独自の組版手法を採り入れたものにしなければならない。日本語独自の組版手法により組版された美しい文書が得られるフォーマッタにすることを目標とする。具体的には、行頭・行末の禁則処理を実現することである。

現在のワープロ出力における禁則処理は、ぶら下げ組みによるものが一般的である。組版手法の中で、ぶら下げ組みによるものは、最も低いレベルのものである。我々は、このようなものでは満足せず、日本語独特の処理を生かすことができるフォーマッタの実現を目指す。

### (3) 出力装置の機能を最大限に生かすこと

出力装置として高解像度のレーザビームプリンタを我々は採用した。このプリンタを利用することにあたって、単なるディスプレイ上のハードコピーのような使用は行なわない方針とした。ディスプレイ上の文書をハードコピーしていたのでは、プリンタの解像度を生かした文字を出力できない。ディスプレイ上のハードコピー的な使用方法では、文字フォントのデザインがディスプレイ側で決定されてしまい、出力装置であるプリンタの解像度にあった文字フォントを利用することができなくなってしまう。本来、文字の2次元的な広がりである文字の大きさは、プリンタの解像度と文字フォントのデザインに依存するものである。プリンタの解像度と文字フォントの組み合わせで決定されたきれいな文字で文書を作成したいと我々は考えた。これを実現するには、プリンタへ文字コードで転送すれば良い。このことは、データ転送の面から考えても必要最小限のデータ量で済み、パフォーマンスの高いシステムと成りうる。

なお、我々が保有しているレーザビームプリンタの解像度は240ドット/インチ(9.45ドット/mm)であり、文字フォントは32ドット×32ドットのものを使用している。

### (4) ワープロで作成した文書ファイルを入力ファイルとして扱えるようにすること

ワープロの文書交換に関するJIS規格に、JIS C 6237(日本語文書交換用ファイル仕様)という規格が定められている。この規格は、あるワープロで作成された文書ファイルを交換用の5インチや8インチのフロッピーディスクへ出力するときの仕様を決定しているものである。入出力のインターフェースを決定することは重要であり、そのインターフェースは多くのシステムと共有できた方が望ましい。このような見地に立ち、この規格を満たす文書ファイルをフォーマッタの入力として取り扱えるような方針にした。

以上のように、ぶら下げ組みで組版されたディスプレイ上の文書をハードコピーのようなかたちでレーザビームプリンタから出力するのではなく、プリンタの解像度を生かし、そして日本語文書の特徴を引き出した文書出力が可能なシステムを目標としている。

## 4. 2 システム構成

我々は、高印字品質出力を得るために、レーザビームプリンタ日立SL-1000をインテリジェント化したシステムをformatting machineのプロトタイプとして構築した。レーザビームプリンタSL-1000の印字ドット密度は240ドット/インチであり、A3サイズの用紙まで印刷可能である。

レーザビームプリンタのインテリジェント化をマイクロプロセッサMC68000を用いて実現する予定であるが、現在はプロトタイプとして、MC68000とMC6809との2つのプロセッサを用いて実現している(図1)。SL-1000は、外部に対して、ビデオインターフェースを公開している。SL-1000をコントロールするシステムがMC6809コントローラである。MC6809コントローラとMC68000システムは、パラレルインターフェースで結ばれている。MC68000システムは、A3サイズに相当する2Mバイトのフレームメモリと24×24及び32×32文字フォント回路を保有している。2種類の文字フォント回路は、共にJIS第1、第2水準の文字セットを持つものである。また、英数字等については半角の文字も可能である。

浄書開発システム

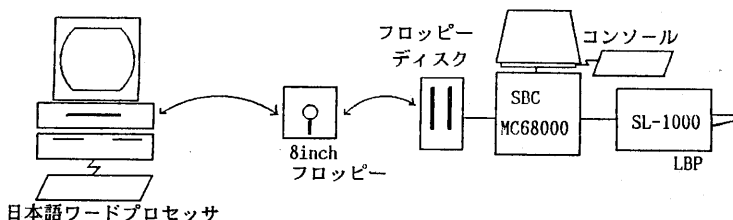


図1 システムの構成

各種ソフトウェアの開発は、MC68000システム上で行なう。MC68000システムのフレームメモリ上で描いた画像情報は、MC6809コントローラへパラレル転送で送られる。そしてMC6809コントローラからSL-1000ヘビデオデータが転送されてレーザビームプリンタ出力を得る。以上のようにしてプリントアウトされる。この結果、用紙一枚を出力するには約2分30秒の時間を要する。現在は、過渡的なシステム構成になっているためにプリンタアウトが遅いが、MC68000システムだけでSL-1000を直接コントロールすれば1枚出力するのに数秒はかからない。

なお、日本語入力システムとして文書変換システム（JIS C 6237規格を満たす機能を持つシステム）を持つ日本語ワードプロセッサ Word Pal 800（日立製作所）を利用している。

#### 4.3 入力仕様と出力対象文字

フォーマッタの入力ファイルとして JIS C 6237（日本語文書交換用ファイル仕様）に準拠したものを採用した。これは、現在までにワープロで蓄積してきた文書の財産を生かすことができるからである。また、文書ファイルを作成するとき、多くの場合はワープロを利用すると予想されるからである。

この規格では、文字コードはいわゆる JIS2 バイトコードである。英数字は1バイト、漢字は2バイトというような1バイト2バイトの混在コードではない。

フォーマッタの出力対象文字は、32×32のドットマトリックスで表される文字である。この32×32の文字フォントをSL-1000から出力すると、その大きさは約3.4mm角となる。これは、通常のワープロ出力の文字とほぼ同じ大きさである。特に、英数字などについては半角の文字も出力可能である。

#### 4.4 組版処理の流れとデータ構造

本フォーマッタは、和欧混合組版の実現を目標としている。良質の和欧混合組版を実現するには、英字についてはプロポーショナルピッチを行ない、かつ、行頭や行末において禁則文字が現われたときには禁則処理を行わなければならない。

このような組版処理を行なうときの問題は、“何をもちて行末とみなすか”ということである。プロポーショナルピッチを実現するために、英字と英字の字間を詰めなければならない。このことは、英単語が現われる行と、そうでない行とでは1行の長さが異なるということの意味しており、単に1行の文字数をカウントしただけで行末を検出できるとは限らないわけである。すなわち、プロポーショナルピッチをしたために、文字数だけで行末を判定してしまうと、その行長が通常の行長より短くなってしまふわけである。

本フォーマッタでは、1行の長さ（ドット数）が指定された行長以上になったときに行末としている。行末が判明すると、最後の文字が禁則文字ならば行末禁則処理を施す。そして、1文字先読みして、その文字が行頭禁則文字かどうかの判定をする。もし、禁則文字ならば、行頭禁則処理をする。これらの処理が終了した後、右ぞろえ（right justify）の処理を行なう。プロポーショナルピッチなどにより、右側がそろわないときがあるため、この処理により文章の右側をそろえるわけである。以上が、本フォーマッタの主な流れである。

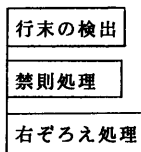


図2 組版処理の流れ

本フォーマッタの組版処理は、1行分のラインバッファを持って行単位に行なうこととした。このような処理方法にした目的は、禁則処理やプロポーショナルピッチなどを行ないやすくしようと考えたためである。行末まで文字を読み込んだとき、最後の文字が行末禁則文字ならば行末禁則処理をしなければならない。1行出力した後、読み込んだ文字が行頭禁則文字ならば、行頭禁則処理をしなければならない。これらの禁則処理を行なうとき、文書ファイルから読んだ文字をすぐにフレームメモリへ展開してしまうと、展開した何文字かの文字列を消去し、文字間をツメ、再び消去した文字列を展開しなければならなくなってしまう。これでは効率が上がらないばかりでなく、プログラムが複雑になってしまい、開発がむずかしくなる。このため、1行に相当するラインバッファを設けた。このラインバッファは、JISコード、フォントタイプ、x座標、y座標、字間、字種といった要素から成り立つものである。この構造を図3に示す。

JISコード						
フォントタイプ						
x座標						
y座標						
字間						
字種						
PRO						

count  →

図3 パツファの構成

JISコードには、文字のコードが入り、フォントタイプには、全角文字や半角文字の種類の別が入力される。x座標やy座標には、その文字をフレームメモリへ展開するときの絶対座標が入る。字間は、前の文字とのアキの量を示し、字種は、その文字が仮名か漢字か約物かなどを示すものである。PROとはプロポーショナルピッチをしているかどうかを示すフラグである。countは文字がどの位置まで詰まったかを示すものである。

#### 4. 5 英字の組み方

英字出力のされ方には、2つの場合があると考えられる。1つは英単語などの言葉として使用される場合であり、もう1つは主に1文字の記号として出力される場合である。すなわち、プロポーショナルピッチの処理を行なう場合とそうでない場合である。日本語文中に英字が連続して現われたときに、その英字の連なりを仮名や漢字と同じようにして組版を行なってしまうと、英字と英字のアキが大きく間が延びてしまい美しい組版とは言えない。このようなときには、プロポーショナルピッチで組版を行なうべきである。

本フォーマットにおける英字出力の対象文字は、32×32のドットマトリックスで示される全角文字と半角文字である。全角文字が英単語のように連なって出力される場合には、プロポーショナルピッチの処理を行なうことにした。もちろん、英字が日本語文中に1文字だけで現われたときには、他の仮名や漢字と同じように組版を行なう。半角の英字については、プロポーショナルピッチの処理を行なわず組版処理を行なうこととした。これは、入力ファイルがワープロ上で作成されるため、文書ファイルの作成者が意図的に半角文字を使用してワープロ上で組版を行なったと考えられるからである。もし、プロポーショナルピッチにするためにツメ処理を行なえば、ワープロ上で作成した文書イメージとは異なった文書出力になってしまい、文書作成者のイメージした文書は得られなくなってしまう。このようなことから、半角の英字文字については意図的に組版を行なったものと判断し、本フォーマットではプロポーショナルピッチの処理は避けたわけである。

#### 4. 6 禁則処理の実現

日本語文書において、文書をより美しく、読みやすくするために、特定の文字や記号・約物が行頭や行末に位置したり、2ページにまたがることをさける習慣がある。これが禁則処理である。禁則処理には、いろいろな規則に基づいた方法が複数あるが、どのような規則にしたがった禁則処理を行なうかが、その文書の美しさを左右する。それは、禁則処理の規則が組版の善し悪しをほぼ決定してしまうからである。そのため、禁則処理の規則を決めることが、組版規則を決定する上で最大の問題といえる。

禁則処理とは、次の2つのものをいう。1つは行頭禁則処理であり、もう1つは、行末禁則処理である。行頭禁則処理とは、ある特定の文字（約物や記号など）が行頭にきた場合、その特定の文字が行頭にこないようにして組み直すことをいう。同様に、行末禁則処理とは、ある特定の文字（約物や記号など）が行末にきた場合、その特定の文字が行末にこないようにして組み直すことをいう。これらの特定の文字をそれぞれ行頭禁則文字、あるいは、行末禁則文字という。

本フォーマットでは、これら行頭、行末禁則処理を実現する。行頭、行末禁則文字は表1に示すものとする。表1に示したものは、禁則文字として基本のものであるため採用した。

表1 禁則文字

行頭禁則文字		行末禁則文字
、	マル	(
・	ナカガロ	{
]	ブラケット	[
]	二重ヒッカケ	「
}	山パーレン	『
》	コロソ ;	◀
’	ダブルクォート	◁
—	ハイフン	“
	?	!

禁則処理を行なう手段には、“追出し”と“追込み”の2つの方法がある。“追出し”とは、文字と文字の間を広げることによって行末の端数処理を行ない、禁則文字を次の行に送り出す組み方で活版印刷に近い方法である。一方、“追込み”は、字間をベタよりも積めることによって、行末の端数処理をしたり、禁則文字を前の行に組み入れる方法である。追出しはアキを、追込みはツメを、それぞれ利用している。このどちらの方法を選択するかによって外観が違ってくるわけである。

我々が実現するフォーマッタは、基本的には、追込みで処理をすることとした。この追込みの方法は写植の世界などで用いられているものであるが、文字と文字の重ね合わせができなければこの手法の特徴が生かされない。それは、ときとして文字の部分と文字の部分を重ね合わせるものであるからである。我々のシステムは、フレームメモリを持っており、OHPのような重ね合わせが可能である。このような特徴を生かすために“追込み”の処理を採用するわけである。

追込みはツメ処理によって実現される。ツメ処理を行なう場合の問題は、『どの文字をどれだけ詰めるか』ということである。すなわち、ツメ処理を行なう文字の候補とツメを行なう量が問題である。このときの大原則は、ツメの目立たないところから行なうことである。

追込みを行なう文字の候補を次のように決定する。レベルとは、追込みを行なう際の優先順位である。

- レベル1 追込みを行なう行内に起こしと受けの約物がある場合には、これらの組を用いて均等に詰める。
- レベル2 追込みを行なう行内に約物が多数ある場合には、パーレンやカッコで詰めて、マルやピリオドでは詰めない。
- レベル3 追込みを行なう行内に半角約物があれば、半角約物（、や、など）で詰める。
- レベル4 追込みを行なう行内に約物がなければ、その行の行末から仮名と仮名、仮名と漢字、漢字と漢字の字間を1ドットずつ詰める。

この規則は、写植で用いられている基本規則[4][5]にほぼ準拠している。レベル4において、写植界では、さらに細かく決定されている。たとえば、まず、仮名と仮名との字間を詰め、その次に仮名と漢字との字間を詰めるというもの。本フォーマッタでは、レベル4までとした。これは、我々が作成する文書を考えてみると、レベル4により追込みを行なう確率が非常に少ないと判断したからである。

レベル1～4の規則に沿った禁則処理を実現するために図2で示したようなラインバッファにしたわけである。追込みを行なうために、このバッファの字種を見ながら、約物を詰めて禁則処理が可能かどうかを判断する。もし約物を詰めて可能であれば、約物を使用してツメを行ない、そうでないならば行末から仮名や漢字の間を詰めて1文字分追込むわけである。以上の処理で、もし、追込みきれないときには、そのまま出力することとした。

#### 4. 7 右ぞろえ処理の実現

右ぞろえ( right justify ) は、行の右端をたてにまっすぐそろえるものである。和文組版においては、特別にこの処理を行なわなくとも自動的に右端がそろいが、和欧混合組版の場合には、意図的に右ぞろえをしないと右端がそろわない。右ぞろえは、和欧混合組版を実現するために必要不可欠なものである。

和文に英文が混じった場合、1行の文字数が同じであっても行長が等しいとは限らない。これは、プロポーショナルピッチを行なうことにより英字と英字の字間が詰まるためである。また、本フォーマッタでは半角文字も取り扱い、半角文字が連続するときは字間をとらずに出力する（ベタ詰め）ため、半角文字が入っている行とそうでない行とでも行長が異なる。通常の行長と異なるこれらの行をそのまま出力すると、右端が不ぞろいになり美しい文書とは言い難くなってしまふ。このようなことを避けるために右ぞろえを行なうわけである。

右ぞろえの処理をしなければならぬものは、次のとおりである。

- (a) 通常の行長より長い行
- (b) 通常の行長より短い行

通常の行長より長い行とは、プロポーショナルピッチ等により1文字近くのスペースが空いたときに、1文字追込むためにできる行である。また、通常の行長より短い行とは、空いたスペースが1文字に満たないときにできる行である。どちらの場合も、プロポーショナルピッチをしたときか半角文字が連続して現われたときに起こるものである。

通常の行長より長い行については、ツメ処理を行なって右ぞろえをする。ツメ処理を行なうときの文字候補は、半角のブランクと約物である。まず、ブランクだけでツメを行なう。このとき、半角ブランクについて最大8ドットまで詰めるが、これでも通常の行長と一致しないときには、約物を均等に詰めて右ぞろえを行なう。

一方、通常の行長より短いときには、アキ処理によって右ぞろえをする。アキ処理は、プロポーショナルピッチを行なった文字列と半角文字が連なる文字列をアキ処理の候補とする。そして、これら文字列の両端を均等に空けることによって右ぞろえを行なうわけである。

以上の処理は、図2に示したフォントタイプ、字種、PROを検索することによって実現している。

#### 5. 今後の課題

フォーマッタの課題を一口で言えば、“フォーマッタの機能を増やすこと”になると思われる。本フォーマッタに対して我々の望む機能は、2つの異なるものがある。1つは、文書を整形する整形機能であり、そしてもう1つは、Writer's Work Bench の機能である。

整形機能として、望むものは、よりきめの細かい組版処理である。今回のフォーマッタで採用した組版規則は、写植界で用いられているものである。この組版規則は、原則として32×32ドットで表わされる全角文字を使用することを前提とする。我々の開発システム上には、全角文字でなく、一部の文字については半角のものもある。さらに、24×24ドットで表わされる文字もある。今後は、これら大きさの異なる文字を考慮して、より質の高い拡張組版規則とそのアルゴリズムを検討する必要があるだろう。そして、単に、文字組みに止まらず、図・表やグラフなどの浄書出力の実現を目指すことである。

フォーマッタに望まれる機能は、単なる組版処理機能だけでなく、文書作成者を援助する機能もある。この機能が、Writer's Work Bench の機能である。フォーマッタ第1版では、時間的な問題があり、この機能は全くサポートしていない。そのため、第2版以降では、目次の自動生成のような Writer's Work Bench の機能をサポートする必要があるだろう。この機能が加われば、より一層フォーマッタが使い易くなり、真の意味での formatting machine の実現になるであろう。

そして、日本文入力システム JOLIS [7] と、本研究室で同時に並行して開発しているオペレーティングシステム OS/omicon [8] との有機的な結合を実現し、総合的な文書作成システムの構築を目指すことである。

## 6. おわりに

我々は、写植で用いられている組版規則を応用したフォーマッタの第1版を試作した。このフォーマッタにより現状のワープロでは考えられない良質の文書出力を得ることができた。しかし、この成果は第1歩であり、我々の最終目標ではない。我々は、文書出力に対して多くの要求を持っている。本研究は、この要求の極一部を実現したに過ぎず、プロトタイプでしかありえない。今後、自分自身が満足し納得の行くような formatting machine の構築を進めていく予定である。なお、今回作成したフォーマッタのプログラムは言語Cで記述されており、そのサイズは、約2000行である。

## 謝辞

本開発システムの構築にあたり、レーザビームプリンタの方式設計に御協力を頂いた日立製作所多賀工場の佐々木道甫氏、榎本順一氏、堀内雄一氏の各位に深謝する。また、(株)写研の布施茂氏には写植のルールブックなどの資料を提供して頂いた。ここに感謝の意を表す。本 JOSH O 計画の基礎を築いてくださった池田裕治氏(現:キヤノン(株))、伊藤幸雄氏(現:(株)日立メディコ)に感謝の意を表す。最後に週1回のミーティングで討論に参加した情報基礎学講座の大学院生、学部生に感謝する。

## 参考文献

- [1] Takahashi, N. et al: "浄書: Japanese Output Server with Hospitality", Proc. of Intl. Conf. on Text Processing with a Large Character Set, Tokyo, 1983
- [2] 伊藤幸雄: "日本語文書作成システムに関する研究", 東京農工大学修士論文, 1983
- [3] 関口治: "日本語文書出力システムに関する研究", 東京農工大学修士論文, 1985
- [4] 写研・写植ルール委員会: 組みNOW, (株)写研, 1975
- [5] ルールブック編集委員会: 写真植字のための組版ルールブック, 日本写真製版工業組版連合会, 全日本印刷工業組合連合会, 日本軽印刷工業会, 1983
- [6] 藤森善貢: 出版編集技術, 日本エディタースクール, 1972
- [7] 中川正樹, 他: "オンライン手書き文字認識システムJOLIS-1の設計と試作", 情報処理学会日本文入力方式研究会資料3-1, 1982
- [8] 高橋延匡, 他: "MG68000用小型OS/OS/○の開発", 情報処理学会計算機システムの制御と評価研究会資料21-6, 1983