7X-2

DP Matchingによる日本語禁則処理

来住 伸子　山内 長承

日本アイ・ビー・エム株式会社 サイエンス・インスティチュート

1 はじめに

レーザビームプリンターの普及にともない，オフィスで計算機を利用した印刷物の作成がなされるようになってきた。しかし，ワープロによる印刷は，品質や簡便さの点で活字印刷に一歩後れをとっている。より高品質にするには，高解像度の印刷，多種の書体の整備などさまざまな課題があるが，本研究では，日本語の禁則処理について検討し，DP matching を用いる手法を試みたので報告する。

2 日本語の禁則処理

2.1 禁則処理とは

日本語文書を組版する場合，ページの構成や見出し等の形式の統一についてのルールが必要になってくるが，ここで考える禁則処理は，それらは既に定まってからの処理をさす。すなわち，書体の種類や大きさは既に定められている条件の下で行頭禁則文字，行末禁則文字，分離禁止語を考慮して読みやすく文字を並べるために字間を調節する処理をさす。

2.2 従来の禁則処理

2.2.1 べた組み

活版印刷やワープロでよく使われる方法で，原則として，べた又は全角と呼ばれる一定の字幅にし，字間を変化させない。行頭禁則文字等の処理としては，

1. 行端からはみ出させる（ぶらさげ組み）。
2. 半角または4 分の1 角の空白や句読点を使う。
3. 行末に空白を残す。

がある。

2.2.2 追い込み／追い出し処理

写植とともに使われだした方法で，字間を拡げたり，詰めたりすることによって行長を調節する。調節量の定めかたとしては，

- 字間すべてを均一に調節する。
- 前後の字種によって字間に優先順位をつけて調節する。

がある。

Line Justification of Japanease Text Using DP Matching
Nobuko KISHI, Nagatsugu YAMANOUCHI
Science Institute, IBM Japan, Ltd.

3 DP Matching による禁則処理

3.1 背景

前述のように禁則処理の仕方はいろいろ存在するが，どの方式がよいかは，ユーザの好みや書体のデザインによって異なってくる。そこで計算機で禁則処理を行う場合には，どの方式も自由に選択できることが望ましい。また，和欧混合組や複数の書体の組み合わせがふえると字間の調節のルールが複雑になっていく可能性がある。そこで次のような理由からDP matching を日本語の禁則処理に使ってみることにした。

Stanford 大学のKnuth が開発したTEX は欧文清書システムで，行揃えの際にDP Matching を使って語間のばらつきがもっとも少なくなるように工夫されている。欧文にある様様な組み版ルールに対応するためにDP Matching に使うパラメータをユーザがかなり制御できる仕組みになっている。

2 バイトコードをフォント切り換えコマンドと1 バイトコードの組み合わせに変換するTEX のfrontend processor があれば，TEX を使って日本語文書が整形できることがわかった。

3.2 禁則処理に使われるDP Matching のあらまし

まず，間隔調節の対象となる各箇所に標準の間隔と標準からずれることがどれ位望ましくないかを示す評価関数(badness function) を与える。また，改行の候補になる各箇所には，そこで改行することがどれ位望ましくないかを示す評価点(penalty) を与えておく。

行揃えをするときには，必要な文字列を先頭から読み込んでいき，改行の候補になる箇所にくるたびに評価関数と評価点をもとに評価値(demerits) を計算し，評価値が一定の許容値以下であれば改行可能な箇所(active point) とする。この新しいactive point は，すでに存在しているactive point のうち，評価値が最小になる改行の仕方を与えるものと結びつけておく。行揃えの必要な文字列の終端にきたら，その時点で存在するactive point の中から，最小の評価値を与えるものを選び，それに結びつけられている箇所をたどっていって，改行箇所とする。

3.3 DP Matching in TEX

前述のような処理をしていれば，DP Matching をしていることになるが，評価関数，評価点，評価値の計算の仕方でさまざまなDP Matching が考えられる。TEX では，次のような機能を通じて評価関数等を制御している。

3.3.1 glue

語間のような調節可能な間隔はすべてglue として指定される。glue は原則として次のような3 つの数の組で表される。

<space> plus <stretch> minus <shrink>

<space> は標準となる幅，<stretch> は広げられる最大許容量，<shrink> は縮められる最大許容量をあらわす。たとえば，ある文字列の間隔を標準の幅以上に広げたいときは，各間隔の<stretch> に比例して広げる量をわりあてる。そこで，広げてもいいが狭めたくない箇所には，大き目の<stretch> と小さ目の<shrink> を割り当てるということになる。評価関数は，<stretch> または<shrink> に対する比の約3 乗にしてある。

3.3.2 penalty

改行が望ましくない箇所には，大きな評価点(penalty) をあたえて改行を禁止することができる。逆に負の評価点を与えて強制改行をさせることもできる。たとえば，評価点p をもつ箇所で改行をする場合の評価値d は　　次のように与えられる。

$$d = \begin{cases} (l+b)^2 + p^2, & \text{if } 0 \le p < 10000 \\ (l+b)^2 - p^2, & \text{if } -10000 \le p < 0 \end{cases}$$

ここで，b は，前述の評価関数を使って計算された値であり，l は，line penalty と呼ばれるパラメータで，全体としての改行の回数を調節する。

3.3.3 discretionary

ある単語が行末になって，ハイフンにより分割されることになると，'-' の分だけ文字の数が行中にある場合と異なってくる。そこで，行中にある場合と行間にまたがる場合との扱いをかえるために，discretionary というコマンドがある。たとえば，difficult という語のハイフンの仕方は次のようにして指示できる。

```
di\discretionary{f-}{fi}{ffi}cult
```

ここでf- はハイフンがおきたときに行末にくる文字列，fi は行頭にくる文字列，ffi はハイフンがおきなかったときに使う文字列をしめしている。

3.4 TEXを使った禁則処理

TEX のDP Matching を利用して，日本語の禁則処理を行うためにfront-end processor を作成し，日本語の禁則処理のいくつかを実現してみた。front-end processor は，2 バイトコードの変換にくわえて，2 バイトコードの行頭禁則文字，行末禁則文字，分離禁止語を見付けだして，不自然な改行が起きないように，TEX のコマンドで囲む処理をおこなう。1 バイトコードについては，殆ど何もせずにTEX 本体にわたすので，line justification, hyphenation などは，従来のTEXの方法がそのまま適用される。

更につぎのような各種の禁則処理に必要なTEX のコマンドはfront-end processor，または直接手で挿入して実験してみた。

3.4.1 追い込み／追い出し処理

これは，日本語の1 文字を欧文の1 語に対応させ，語間に使うglue を調節することで実現できる。追い込みだけで行を揃えたいときは，stretch を0 にし，shrink を大きくとればよく，追い出しだけにしたいときは，逆にすればよい。追い込みと追い出しの両方を使って調節したいときは，stretch とshrink の両方を使い，優先させたい方を大きくとる。

3.4.2 ブラ下げ組み

欧文では，本来，句読点を行端からはみださせない。そこで，ぶら下げ組みは特別な処理が必要になる。前述のdiscretionay コマンドを使って，行間にまたがるときにかぎって，はみだしに必要なコマンドが実行されるようにした。たとえば，

```
\discretionary {aa\kern-5pt}{}{aa}
```

としておくと，aa という文字列が行端にくると5 ポイント分はみださせることができる。はみださせる量を全角文字の幅と同じにすると，ぶらさげ組になるし，半角分にすると，行端の句読点を半角どりにしたことになる。

3.4.3 べた組

先程の追い込み/ 追い出し組の一種として，shrink やstretch の値を0 にするという方法はうまくいかない。行長を字の大きさと字間隔にあわせて正確に指定し，欧字を一切使わないというような制限を守らないかぎり，間隔の調節がかならず必要になるためである。ぶらさげ組の指定と組み合わせ，小さなshrink やstretch を使うと，ほぼべた組に見せることができる。

4 結論その他

TEX のDP Matching を利用することにより，DP Matching で日本語禁則処理の主だったものがほぼできることが確認できた。できた原因としては，

- 日本語に特有な処理(禁則文字及びその組み合わせの検出) をfront-end processor だけで行えた。
- DP Matching に使う評価関数をの制御をglue やpenalty の形でTEX が公開している。
- 行間にある文字列がかかるときの例外処理をTEX のdiscretionary コマンドで処理できた。

などがあげられる。DP Matching をつかった日本語禁則処理は，TEX の欧文および数式の整形機能とも並存するので，和欧混合組としては高品質な組み版を可能にする手法だと考えられる。

ちなみにこの原稿は，TEX の日本語front-end processor を使い，追い込み/ 追い出し処理をおこなっている。front-end processor の出力は，LATEX を使って処理した後，IBM 4250 electro-erosion printer を使って印刷した。