日本語文章推敲支援ツール『推敲』における否定表現の抽出法†

菅沼 明†† 倉田昌典†† 牛島和夫††

日本語文章推敲支援ツール『推敲』は日本語文章を字面だけで解析し，推敲に役立つ情報を書き手に提供することを目的として我々が開発したツールである．『推敲』には現在，受身，接続助詞「が」，指示詞「これ，それ，…」，とりたて詞（副助詞，係助詞の一部），否定表現などの候補を指摘する機能がある．文章中でそれらを使用していれば，『推敲』がそれを指摘し，書き手に推敲する手がかりを提供する．本論文では，このうち否定表現を抽出する方法と『推敲』への適用について述べている．否定表現の候補を抽出する手法は，日本語文章約70万字を実際に調査した結果を参考にしてプロトタイプを構築し，これを別の約300万字の文章に適用して評価し改訂した．構築した抽出法は文字についての簡単な条件をいくつか適用するだけの形になっている．これは，「指摘した候補を書き手が必ず吟味する」を『推敲』の開発方針としているために，「否定でない表現も否定表現の候補の中に含まれてしまう」という誤りをある程度許しているからである．実際に『推敲』で否定表現の候補を抽出すると，候補の中にいくつかの否定でない表現も含まれる．しかし，抽出精度（実際の否定の数／総指摘件数）は，9割以上である．『推敲』で字面解析を採用したのは「実用規模の文章を待ち遠しくない時間で処理して欲しい」ためである．パソコン（PC-9801 VX）上に実現した『推敲』で処理時間を測定すると，実用規模（1万字：論文誌 7〜8 ページの文字数）の文章からすべての否定表現の候補を1秒以内で抽出できる．さらにこの抽出法は，解析対象の文章を一度しか走査しないので，検索時間は文章の長さに比例する．

1. はじめに

日本語ワードプロセッサの普及に伴って，機械可読の日本語文章が増えつつある．しかし，ワードプロセッサは推敲作業を積極的に支援しているわけではない．日本語文章推敲支援ツール『推敲』[1),2)]は，機械可読の日本語文章（漢字かな混じりの科学技術文章を主な対象とする）から推敲に役立つ情報を抽出することを目的として我々が開発したツールである．

人間が行う推敲作業は，文章を読み返して問題となる箇所を探し，その部分を検討して，必要があれば書き直すという形で行う．この作業のうち，問題となる箇所を探す部分を計算機で支援できれば，人間は計算機が指摘したものを吟味して推敲を行うことができるので，人間の作業を軽減できる．このことから，『推敲』の開発にあたって次の2つの方針を設けた．

① 文書中に問題となりそうな箇所があればそれを指摘できればよい．（実際に推敲するのは書き手である．）

② 実用規模（1万字程度：図や表を含めると論文誌刷り上がり 7〜8 ページの文字数）の文章を待ち遠しくない時間で処理して欲しい．

日本語の文章は分かち書きされていない．そのため，日本語の文章を解析するには，まず辞書に基づいて形態素解析を行い，構文解析を行うのが普通である．しかし，それらを行っても必ずしも一意に解析できるとは限らない．さらに，辞書を使った文法処理を行うと，実用規模の文章では解析時間がかかりすぎて上記②の方針を満たさない．このため，我々は辞書を使わず文法処理も行わず字面だけで解析する方法を採ることにした．

文章を字面だけで解析するアプローチを採ったため，解析精度は文法解析を行う場合よりも低いことが予想できる．しかし，『推敲』の開発方針①から，『推敲』が指摘したものは書き手が一度は目を通すことになるので，文章の解析精度はこの開発方針を満たす程度でよい．つまり，ある程度の誤りを許している．文章から問題となる箇所を抽出する場合に犯す可能性がある誤りには，次の2つがある．

第一種の誤り：指摘に漏れがある．

第二種の誤り：指摘すべきでないものまで指摘してしまう．

これらの誤りのうち，『推敲』のような支援ツールでは第一種の誤りは原則として犯してはならない．しかし，第二種の誤りはある程度許容できる．字面だけの解析で第一種の誤りを犯さずに開発方針を満たす程度の精度が得られれば，『推敲』で採用する解析手法としては十分である．

『推敲』には現在，受身[3)]，接続助詞「が」[4)]，指示詞

† A Textual Analysis Method to Extract Negative Expressions in Writing Tools for Japanese Documents by AKIRA SUGANUMA, MASANORI KURATA and KAZUO USHIJIMA (Department of Computer Science and Communication Engineering, Faculty of Engineering, Kyushu University).

†† 九州大学工学部情報工学科

「これ，それ，…」，とりたて詞（副助詞，係助詞の一部），否定表現などを指摘する機能がある．文章中でそれらを使用していれば，『推敲』がそれを指摘し，書き手に推敲する手がかりを提供する．本論文では，このうち否定表現の指摘について，2章でその表現を指摘する意義，3章で字面だけの解析で否定表現の候補を抽出する方法について述べる．また，4章で約300万字の日本語文章に否定表現の候補の抽出法を適用して，それを評価する．さらに5章で，否定表現の候補の抽出法を『推敲』に組み込むことについて述べる．

2. 否定表現を指摘する意義

日本語の文章中に出現する否定は推敲の対象となる．例えば，二重否定を使うことで文章の意味がわかりにくくなったり，まわりくどくなったりすることがある．例えば，次の二重否定を使った文はまわりくどい．

- このツールでは Ada のパッケージを扱えないわけではない．
- これを最後まで残しておいたのは，これが重要でないからではない．

科学技術文章では明確な表現をする必要がある[5]．そのために，上のようなまわりくどい文は，次のようにはっきりと言い切る．

- このツールでは Ada のパッケージを扱える．
- これを最後まで残しておいたのは，これが重要だからである．

二重否定を含む文と含まない文とでは，かなりニュアンスが異なる．まわりくどい表現を採るのは，採らざるを得ない事情が背後にあるのであろう．この表現を使うことで，かえって読者にそれを見抜かれてしまうかもしれない．科学技術文章では二重否定のようなまわりくどい文章をできるだけ控えたほうがよいので，文章中から二重否定を抽出して指摘してくれると文章を推敲する上で役立つ．

別の表現として，「ように」＋否定表現の形のものがある．この表現は文章中に現れると意味があいまいになることがある．例えば，次のような表現である．

- FORTRAN は，Ada のように構造化されていない．

この表現は二通りの解釈ができる．

- FORTRAN は Ada が構造化されているようには構造化されていない．
- FORTRAN は Ada が構造化されていないのと同じように構造化されていない．

このように，様態の助動詞「ようだ」の連用形「ように」がかかる用言が否定されているときにあいまいな表現となりうる．このため，「ように」＋否定表現を指摘することは推敲の助けとなりうる．

さらに，「新幹線に乗って東京に行かなかった」のような表現もある．この表現は，「なかった」という否定を表す単語が何を否定するのかによって，複数の解釈ができる．

- 東京に行かなかった．
- 東京には行かなかった（けれど大阪には行った）．
- 新幹線を使わずに（新幹線以外の乗り物で）東京に行った．

否定表現を含む文を指摘することによって，このような表現を発見することができる．

3. 否定表現の抽出

前章に挙げた例はいずれも否定を含んでいる．そこで，これらの表現を抽出するためには，まず否定を表す単語の候補を抽出できなければならない．否定を表す単語には，形容詞および助動詞の「ない」，助動詞の「ぬ」「まい」の3つがある．

我々の研究室で蓄えている機械可読の日本語文章（卒論，修論，翻訳文，レポート，その他）683,867文字について，否定表現になりうる文字列（「ない」「ぬ」「まい」の活用形）の出現数を計数した．その結果を**表1**に示す．この結果を参考にして否定を表す単語の候補の抽出法を構築する．

3.1 形容詞および助動詞の「ない」の抽出

表1の精度（実際の否定の数を否定の候補の数で割って百分率で表したもの）をみると，否定を表す「ない」の活用形については，文字列を探すだけでもかなり高い精度で抽出できると言える．また，漢字表記する場合も表1の精度からみて，「無かろ，無かっ，無く，無い，無けれ」の文字列を探すだけで「無い」の活用形を高い精度で抽出できそうである．

この計数の結果に含まれる第二種の誤りのうち形容詞「少ない」，名詞の「行ない」を取り除くことができれば，否定表現の抽出精度を向上させることができる．一方，漢字＋「ない」で終る形容詞と名詞とを公用データベース日本語単語辞書[6]*で調べると，上記

* 見出し語総数約18万7千語からなる自立語だけの辞書を使用した．

表 1 否定を表す単語の調査結果
Table 1 A number of occurrences of the negative words.

否定を表す文字列	出現数	実際の否定	精度
なかろ	2	2	100.0%
なかっ	60	60	100.0%
なく	317	261	82.3%
ない	2,168	2,113	97.5%
なけれ	364	364	100.0%
漢字表記	8	7	87.5%
小計	2,919	2,807	96.2%
ず	600	198	33.0%
ぬ	227	221	97.4%
ね	174	149	85.6%
ん	326	2	0.6%
小計	1,327	570	43.0%
まい	28	4	14.3%
計	4,274	3,381	79.1%

$$\text{精度}=\frac{\text{実際の否定の数}}{\text{出現数}}\times 100(\%)$$

のほかに，「危ない，切ない」「…し損ない」の3つがある．これらもあわせて取り除くことにする．この公用データベース日本語単語辞書がすべての日本語の単語を含んでいるとは言えない．そのため，この辞書に含まれていない単語で，漢字＋「ない」で終る単語があるかもしれない．しかし，そのような単語があったとしても，それは第二種の誤りとして抽出してしまうのであって，『推敲』の開発方針『第一種の誤りは犯さない』は満たしている．そこで，以下を助動詞および形容詞「ない」を抽出するための判定条件とする．

判定条件：

「な」の1文字前は「少，危，切，行，損」でない．

ただし，「少」の1文字前が「少，多」の場合，「切」の1文字前が「一」の場合，「行」の1文字前が数字または，漢数字，「数」の場合は除く．

「多少ない」「一切ない」という表現に使われる「ない」は否定の「ない」である．これらの表現があるので，「ただし」以下の条件がなければ上の判定条件を使うことで第一種の誤りを犯す可能性がある．

3.2 否定の助動詞「ぬ」の抽出

(1) 連用形の「ず」

この計数調査で出現した否定の助動詞以外の「ず」を表2に示す．否定の助動詞「ぬ」の連用形「ず」を抽出するために下の4つの判定条件を設ける．

表 2 出現した否定でない「ず」
Table 2 A list of 'ず' which occurs as a non-negative.

誤り	数 (割合)
ま ず	118 (29.4%)
必 ず	94 (23.4%)
いずれ	77 (19.2%)
ずらす，ずらして	29 (7.2%)
その他	84 (20.8%)
計	402 (100.0%)

判定条件 1：[「ず」の1文字前の条件]

「ず」の1文字前が表3に示した文字のいずれかであれば，その「ず」は否定の候補である．

根拠：否定の助動詞「ず」の接続は動詞および，動詞型の活用をする助動詞の未然形である．動詞の未然形はア，イ，エ段の文字で終る．さらに，語幹と語尾の区別がない動詞もあるので，未然形が漢字で終る場合もある（「見る，似る，来る」など）．公用データベース日本語単語辞書[6]で調べた結果，表3に示した53種類の文字は動詞の未然形の最後の文字になりうる．また，動詞型の活用をする助動詞の未然形の最後の文字はすべて表3の平仮名に含まれている．

ここで，表3にある漢字について説明を付け加える．これらの漢字には「見る，似る，来る」のように比較的よく使う単語もある．しかし，「卑（下卑る），化（時化る），気（しょ気る），消（魂消る）」と，すぐには単語を思い浮かべることができないものもある．ここで，未然形が漢字で終るすべての動詞がこの表に含まれているか否かが問題になる．もし未然形が漢字で終る動詞でこの表に含まれていないものがあれば，判定条件1を使用することで第一種の誤りを犯す可能性がある．しかし，公用データベース日本語単語辞書に

表 3 否定の助動詞「ず」の1文字前にくる可能性がある文字
Table 3 A list of Japanese characters which can occur immediately before a negative auxiliary verb 'ず'.

平仮名	か，さ，た，な，ま，ら，わ，が，ば， い，き，ち，に，ひ，み，り，ぎ，じ，び， え，け，せ，て，ね，へ，め，れ，げ，ぜ，で，べ，こ
漢字	干，居，見，似，射，煮，着，卑， 宛，化，気，経，昏，出，消，寝，耽，貞，得，禿，来

登録されている単語はかなり広範囲であるので，ここでは，表3に示した漢字以外の漢字で未然形が終る動詞は極めてまれであるとみなす．

判定条件 2：[「ず」の1文字後の条件]

「ず」の1文字後が促音，撥音ならば，その「ず」は否定の候補でない．

根拠：ある文字（句読点など特殊な文字は除く）が否定の助動詞「ず」の後ろに続く場合，その文字は必ず単語の最初の文字である．そして，単語の最初の文字が促音「っ」であるものは存在しない．また，撥音「ん」で始まる口語の単語はそのどれもが否定の助動詞「ず」に続くことはない．

以上，「ず」の前後1文字に関する判定条件を設けた．しかし，表2に示した誤りのうち出現頻度の多かった「まず」「いずれ」はこれらの判定条件では取り除くことができない．そのために，「まず」「いずれ」を取り除くために下の2つの条件を付け加える．

判定条件 3：[「ず」の2文字前の条件]

「ず」の1文字前が「か，さ，た，な，ま，ら，わ，が，ば，ち，り，ぎ，じ，び，け，て，め，れ，げ，ぜ，べ」のいずれかの場合，「ず」の2文字前の文字が平仮名または漢字のときに限って，その「ず」を否定の候補とする．

根拠：判定条件3に示した文字が否定の助動詞「ず」の1文字前にあると，その文字は五段活用動詞の活用語尾かまたは，語幹と語尾の区別がある上一段，下一段活用動詞の活用語尾である．それらの動詞の語幹に現れる文字は漢字と平仮名だけである．

判定条件 4：[「ず」の2文字後の条件]

「ず」の1文字後が「れ」の場合，「れ」の1文字後が「い，き，こ，つ，っ，て，ん」のいずれかに限って，その「ず」を否定の候補とする．

根拠：否定の助動詞「ず」の1文字後に「れ」が続く場合，その「れ」は単語の最初の文字である．読みが「れ」で始まる単語を公用データベース日本語単語辞書で調べてみると，「れ」の次の文字は「い，き，こ，つ，っ，て，ん」の7種類の文字である．

（2） 終止形，連体形の「ぬ」

連用形の「ず」の候補の抽出では，否定の助動詞「ぬ」の接続関係を参考にして4つの判定条件を設けた．そのため，上で述べた4つの判定条件はそのまま終止形，連体形の「ぬ」にも適用できる．

（3） 仮定形の「ね」

仮定形の「ね」の場合も連用形の「ず」で設けた判定条件をそのまま使用することができる．ところが，「ね」の場合は，仮定形の接続を考慮した判定条件を設けるほうが効率よく第二種の誤りを取り除くことができる．この判定条件を使うと，調査したデータに限っては，否定でない「ね」をすべてふるい落とすことができる．

判定条件：

「ね」の1文字後は「ば」である．

根拠：口語の日本語では係結びがないために，文章が仮定形で終ることがない．また，仮定形に接続する助動詞もない．そのために，仮定形の後ろにくる単語は接続助詞の「ば」だけである．

（4） 終止形，連体形の「ん」

表1に示したように，文字「ん」は否定であるものの数が非常に少ない．実際に調査した文章中で「ん」は326個現れ，そのうち否定であったものは以下の2つであった．

- 使ったことがありません．
- ヒントがえられるかもしれません．

否定の「ん」を抽出するための判定条件も，「ず」の4つの判定条件と同じになる．この条件を用いると，抽出する「ん」の数は326個から49個に減る．しかし，実際に否定の助動詞「ん」であるものの数が少ないので，精度は4.1%と低い．けれども，否定の候補全体からみると，候補とする「ん」の数は少ないので，否定全体としての精度はそれほど下がらない．そのため，「ず」で設けた4つの判定条件のほかに判定条件を設けることはしない．

3.3 推量否定の助動詞「まい」の抽出

表1の文字列「まい」の出現頻度を見ると，否定を表す「ない」「ぬ」の活用形の出現頻度に比べて非常に少ない．そのために，文章中に出現する「まい」をすべて否定の候補としても否定全体としての精度はそれほど下がることはない．このことから，否定を表す「まい」を抽出する特別な判定条件は設けない．

3.4 否定表現の抽出法の精度

3.1〜3.3節で述べたすべての判定条件を適用して否定の候補を抽出する場合の精度を**表4**に示す．表1の出現数を参考にして，第二種の誤りを減らす方向で判定条件を構築してきたのだから当然であるが，表1と表4を比べれば，特に「ず」「ん」の出現数が大幅に減っていることがわかる．しかし，第二種の誤りの大部分は依然として「ず」と「ん」が占めている．

この計数結果を基にすると，『推敲』で基準として

表 4 否定の候補の抽出結果
Table 4 A result of extracting the candidates of the nagative words.

否定を表す文字列	候補	否定	精度
なかろ	2	2	100.0%
なかっ	60	60	100.0%
なく	270	261	96.7%
ない	2,123	2,113	99.5%
なけれ	364	364	100.0%
漢字表記	8	7	87.5%
小計	2,827	2,807	99.3%
ず	257	198	77.0%
ぬ	223	221	99.1%
ね	149	149	100.0%
ん	49	2	4.1%
小計	678	570	84.1%
まい	28	4	14.3%
計	3,533	3,381	95.7%

表 5 否定の候補の抽出手法の評価
Table 5 An evaluation of the textual analysis method to extract the candidates of the negative words.

否定を表す文字列	候補	否定	精度
なかろ	7	7	100.0%
なかっ	240	240	100.0%
なく	1,013	949	93.7%
ない	3,733	3,562	95.4%
なけれ	525	525	100.0%
漢字表記	23	17	73.9%
小計	5,541	5,300	95.7%
ず	897	616	68.7%
ぬ	306	295	96.4%
ね	396	396	100.0%
ん	1,047	0	0.0%
小計	2,646	1,307	49.4%
(「ん」を除く)	(1,599)	(1,307)	(81.7%)
まい	52	7	13.5%
計	8,239	6,614	80.3%
(「ん」を除く)	(7,192)	(6,614)	(92.0%)

いる1万字の文章当り平均 51.7 個の指摘があり，そのうちの平均 2.2 個が第二種の誤りとなる．1文中に複数の否定がある二重否定や，「ように」+否定表現の候補を抽出する場合には，それらの候補として抽出するものの数はさらに少なくなる．『推敲』の開発方針『実際に推敲するのは書き手である』からすると，第二種の誤りは人間が推敲する際に誤りであることを判定すればよいので，上で述べた字面解析手法は『推敲』に組み込む解析法として満足できる．

3.5 当然の表現

「ねばなら」「なければなら」「なくてはなら」などに否定が続く形のものは当然の表現と呼ばれる．この表現は慣用表現であり，否定を表す単語を含んでいるにもかかわらず，否定の意味を持っていない．このために，当然の表現は否定の抽出から外すことにする．今回調査した日本語文章には，当然の表現が 397 含まれており，これを除くと候補として抽出する数が 2,739 個，否定を表す単語の数は 2,587 個になる．そのため，否定を表す単語の抽出精度は 94.5% となり，精度は 1% 程度下がる．しかし，否定の意味を持つものの抽出精度という観点からは，当然の表現を含める場合より好ましい．

4. 字面解析手法の評価

上で述べた否定の候補を抽出する字面解析手法は，機械可読の文章を実際に調査して構築したものであり，一般の日本語文章にも有効であることを期待したい．この字面解析手法が一般の文章にも有効であることを確かめるために，調査に使用した文章とは別の文章で評価を行った．

評価には，JICST 科学技術文献ファイルの管理システム編（文献数 14,380，総文字数 2,842,062 文字）の表題と抄録の部分を使用した．評価方法は，否定表現の候補の抽出法に沿って候補を計算機で抽出し，その候補が正しいか否かを目視で確認する．さらに，各判定条件で取り除いたデータに対して，第一種の誤りを犯していないかどうかを調べる．

4.1 結果

否定の候補の抽出法に沿って抽出した結果を**表 5** に示す．助動詞，形容詞「ない」の活用形の候補の抽出については，ほぼ満足のゆく精度を示している．一方，助動詞「ぬ」の活用形の候補の抽出では，特に「ん」がすべて第二種の誤りとなっており，しかもその数が多いことが問題である．評価に用いた文章中には，文字「ん」が 2,352 個あり，4つの判定条件を適用した後でも 1,047 個が残る．これらを詳細に見てみると，常用漢字では「頻（頻度），缶，濫（濫用），瓶，然（依然），遷（変遷），旋（斡旋），関」など（総数 291），非常用漢字では「塵（粉塵），穿（穿孔），塡（充塡），燐，牽（牽引），癌，澱（沈澱）」など（総数

638) が平仮名で書かれていた．それらが漢字で書かれていれば，4つの判定条件を満たす「ん」の数は1,047から118になる．しかも，そのうち実際に否定であったものは1つもなかった．

否定の「ん」は「ぬ」の発音上の言い替えである．話言葉ではなく文章中で使われる否定の「ん」は，丁寧の助動詞「ます」に続いて「ません」という表現として主に出現する．しかし，『推敲』で主な対象と考えている科学技術文章を「です，ます」調で書くことはまれであるので，否定の「ん」が文章中に出現する確率は非常に小さい．

これらのことを考えて，否定表現を抽出する手法を『推敲』に組み込む際には，すべての「ん」を候補から外すことにした．ただし，こうすることで，第一種の誤りを犯す可能性がでてくる．そのため，現段階での解決策として，

①「ん」のすべてを否定の候補から外す．

②「ません」だけを否定の候補の中に含める．

③「ん」を否定の候補として抽出する．

の3つの場合を，否定表現を検索するコマンドにスイッチを付けて選択できるようにすることを考えている．

否定の候補（「ん」を含まない）として抽出した7,192のうち，誤って抽出してしまった578の内訳を**表6**に示す．表6にある「まず」は，文字「ま」の1文字前が漢字または平仮名で，判定条件3を満たすものである．

否定の「まい」だけで抽出精度を考えると，精度は非常に悪い．しかし，「ない，ぬ」が候補の大半を占めているので，「まい」は否定を表す単語全体としての精度にあまり影響を与えていない．これは，判定条件を構築する際の調査と同じ結果である．ただ「まい」の誤りで目立って多いのは，「あいまい（曖昧）」で，「まい」の誤りの半数以上を占めている．このことから，「まい」を抽出する際には「あいまい」を考慮した判定条件を追加することを検討している．

表6　出現した第二種の誤り
Table 6　Errors of the second kind appearing in the evaluation of the textual analysis method.

語　句	数（割合）
ま　ず（先ず）	115（19.9%）
な い し〔接続詞〕	106（18.3%）
わ ず か（僅か）	62（10.7%）
あ い ま い（曖昧）	26（4.5%）
たずさわる（携わる）	22（3.8%）
そ の 他	247（42.8%）
計	578（100.0%）

以上の調査で，最初に文字列照合だけで否定を表す文字列（「ない，ぬ，まい」の活用形）を抽出する際には，第一種の誤りを犯していない．次に，各判定条件でそこから取り除いた文字列をすべて人手で調べた結果，否定を表す文字列は含まれていなかった．このことから，3章で構築した判定条件を使用しても，第一種の誤りを犯していないと判断した．

5. 『推敲』への適用

5.1 「ように」+否定表現，二重否定の指摘

この節では，否定の候補を抽出する手法を使用して，「ように」+否定表現の候補や二重否定の候補を文章から抽出する方法について述べる．

「ように」+否定表現や二重否定を文章中から正確に抽出するためには単語間の係受けの解析をしなければならない．しかし，『推敲』では字面でしか解析しないので，単語の係受けを調べることは不可能である．そのため，『推敲』では，文字列「ように」と否定の候補が同じ文中にある文を「ように」＋否定表現の候補として指摘し，1つの文中に否定の候補が2つ以上ある文を二重否定の候補として指摘している．したがって，『推敲』の指摘の中には，

- 相対頻度は英文の**ように**安定せ**ず**，…
- 相互再帰呼び出しができ**なく**なることは**ない**．

のようなあいまいな文や，まわりくどい文もあるけれど，

- スタックがどの**ように**実現されているかは，ユーザは知ら**なく**てもよいことである．
- 辞書を使わ**ず**，文法処理も行わ**ず**，…

のような文も含まれる．

上のように，『推敲』の指摘は第二種の誤りを含んでいるので，指摘する数が多いと書き手が1つずつ吟味するのは大変である．しかし，判定条件を構築する際に使用した日本語文章約68万字では，全文章19,226のうち，「ように」+否定表現の候補として指摘する文の数は87であった．1万字当りでは平均1.3個になる．また，1つの文に複数の否定の候補が存在する文の数は297で，1万字当りでは平均4.3個である．このように少ない表現を文章全体から探し出すのは骨の折れる仕事である．『推敲』を使うと，その作業の範囲を少数の文章に絞り込むことができるという

点で『推敲』に組み込んだ「ように」+否定表現の候補の抽出や二重否定の候補の抽出は有用である．

「ように」+否定表現を抽出するには「ように」をキー文字列としたが，2章で述べた「新幹線に乗って東京に行かなかった」という表現を抽出する場合にはキーとなる文字列がない．そのため，『推敲』のような解析方法でこのような表現の候補の数を絞り込むことは困難である．しかし，『推敲』では第一種の誤りは犯さないので，「新幹線に乗って東京に行かなかった」という表現があれば，否定表現の候補の中に含まれているはずである．

5.2 応答時間

『推敲』は現在パソコン上に実現している[2]．ユーザがキーボードからコマンドを発すると，『推敲』はすべての候補を検索し，**図1**のように結果の先頭部分を画面に表示する．画面からあふれた結果に対して，ユーザは画面をスクロールさせて候補を1つ1つ吟味していく．このように，『推敲』を使用する際には，コマンドを発してから先頭部分の画面が表示されるまでの時間（応答時間）が問題となる．

『推敲』の開発にあたって，実用規模の文章を待ち遠しくない時間で処理してほしいという要求を課した．上で述べた字面解析手法がこの要求を満たしているかどうかを調べるために，パソコン上で実現した『推敲』で，その応答時間を測定した[7),8)]．測定には，PC-9801 VX (CPU i80286, CLOCK 10 MHz) を使用した．なお，パソコン版『推敲』は，すべての処理を主記憶上で行うので，二次記憶のアクセス時間の影響を受けない．

判定条件を構築する際に使用した文章で，応答時間の測定を行った．その結果を**図2**に示す*．図中の各点は測定値を表し，直線は最小二乗法で計算したものである．図2を見てわかるように，各測定値はほぼ直線上に並ぶ．これは，否定表現の抽出法が解析対象の文章を一度しか走査しないためである．図中Aの「ように」+否定表現の検索では，文字列「ように」を探すときと，否定表現の候補を探すときと二度文章を走査するので，図中Bの否定を複数含む文と比べて応答時間が約2倍になっている．また，図中Cの直線は検索時間を示しており，表示時間は含

```
   :FORTRANは，Adaのように構造化されていない。          —   98  •
  2:FORTRANはAdaが構造化されているようには構造化されていない。  —   93  •-
  3:FORTRANはAdaが構造化されていないのと同じように構造化されてい—   95  •-
  4:  このように，様態の助動詞「ようだ」の連用形「ように」がかか—   96  •••-
  5:  •相対頻度は英文のように安定せず，…              —  249  -
122:で役立つ。
123:  別の表現として，「ように」+否定表現の形のものがある。この表現は文章
124:中に現れると意味があいまいになることがある。例えば，次のような表現であ
125:る。
126:  •
   :FORTRANは，Adaの様に構造化されていない。
128:この表現は二通りの解釈ができる。
129:  •
130:FORTRANはAdaが構造化されているようには構造化されていない。
131:  •
132:FORTRANはAdaが構造化されていないのと同じように構造化されていない。
133:  このように，様態の助動詞「ようだ」の連用形「ように」がかかる用言が否
```

図1 「ように」+否定表現の検索結果の画面
Fig. 1 The output of the command to find sentences including 'ように'+negative word.

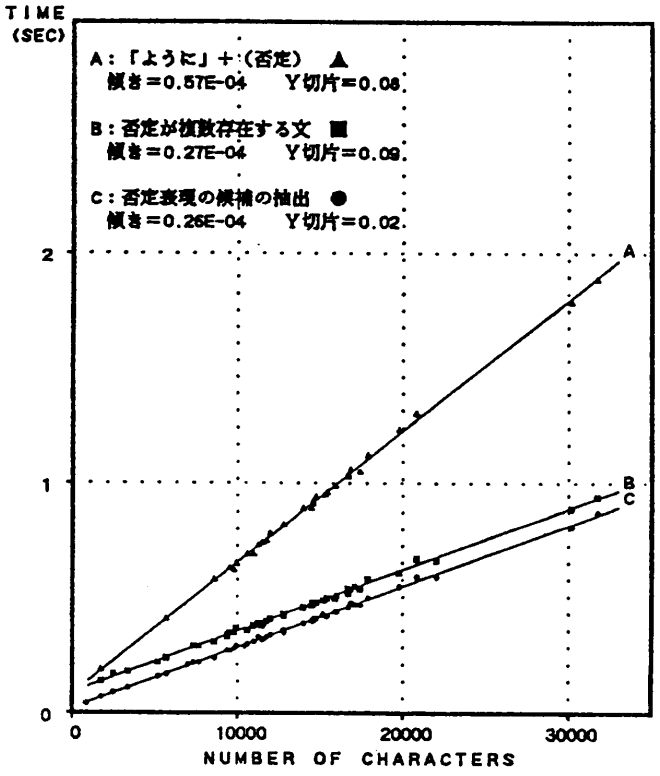

図2 応答時間の測定結果
Fig. 2 The response time of the textual analysis method.

* 応答時間は検索時間と表示時間の和となっている．『推敲』が1画面を表示するのにかかる時間は解析対象の文章の長さにかかわらず一定で，約 60〜70 ms である．検索の結果，候補が存在しない場合はメッセージを表示するだけで表示時間がほとんどかからない．そのため，「ように」+否定表現と否定の候補が複数存在する文とを検索するのに要する応答時間の測定値は，候補が存在したものだけを図中に示している．

んでいない.

図2から，入力文章が1万字程度であれば応答時間が1秒以内であることがわかる．さらに，3万字の文章でも応答時間は2秒程度である．この測定によって得た結果は，『推敲』に課した『実用規模の文章を待ち遠しくない時間で処理して欲しい』を十分満たしているといえる．

5.3 人間とコンピュータの共同作業

最近，文書の校正支援機能を持ったワードプロセッサが製品化され始めた[9]．分かち書きをしていない日本語文章を解析するためには，形態素解析を行い，構文解析をするのが普通である．そのため，それらシステムのほとんどが辞書を使った解析や構文解析を行っている．これらのシステムの処理速度は1秒当り20字程度とある[10),11)]から，1万字当りの処理時間は単純に換算して8分程度になる．しかし，これらのシステムは解析に時間がかかるけれど，抽出の精度や範囲を考えると『推敲』では指摘できないものも指摘できる．

ツールというものは人間の行う作業を代行または手助けすることで，人間の負担を軽くするものである．そのために人間との関わりは重要な意味を持っている．『推敲』では人間とコンピュータとの協調系を想定している．人間がコンピュータに待たされることなく作業を行うために，『推敲』では反応速度を重視して設計してきた．反応速度が速ければ，やり直しが苦にならないので，書き手は何度でも文章を『推敲』に処理させて，推敲作業を行うことができる．一方，上で述べたシステムでは文章全体を処理するには時間がかかるが，1回の処理でさまざまな情報を書き手に提供する．このシステムに文章を何回も処理させるような使い方はしないであろう．このように，『推敲』と上で述べたシステムとではそれらを使って推敲作業を行う形態が異なっている．我々は『推敲』と上で述べたシステムとは，推敲作業を支援するという目的は同じでも，使い方の全く違ったツールであると考えている．

6. その他の否定表現

本論文では，否定表現の抽出を形容詞，助動詞の「ない」「ぬ」「まい」の活用形に限定した．日本語文章には，接頭辞「非，不，未，無」を付けることでそれに続く単語の意味を否定する漢語的表現がある．この章では，それらの表現について述べる．

3章の計数調査に使用した日本語文章で「非，不，未，無」の4文字の出現数を調査した結果，「非」が176，「不」が228，「未」が51，「無」が127出現した．それらのうちの25個が否定の意味は持っているが接頭辞ではないもの（否定の「ない」の漢字表記，「有無」など）で，23個が「不思議，未来，未然形」などの否定の意味を失っている用法であった．さらに公用データベース日本語単語辞書[6]で「非，不，未，無」という文字を含む単語を調べた結果，「検非違使，不知火，未年，白無垢」など「非，不，未，無」の漢字が否定の意味を失っている単語があった．しかし，『推敲』で主な対象としている科学技術文章では，これらの単語の出現頻度は低いと考えられるので，否定を表す接頭辞を抽出する方法としては，「非，不，未，無」の4つの文字を検索するだけでよいであろう．

現在『推敲』に組み込んでいる否定表現の候補の抽出では，否定を表す接頭辞の候補は抽出の対象から外している．否定を表す接頭辞を含めた否定表現の候補を『推敲』で指摘することについては現在検討中である．

7. おわりに

文章を字面だけで解析する方法を採ることで，パソコンでも高速な処理を行えるようにできた．字面だけの簡単な解析法であるが，否定の候補の抽出精度は9割以上と十分満足できる値を得ている．否定の助動詞「ぬ」の活用形「ん」を無視していることを考えると，第一種の誤りを犯す可能性があるので，『推敲』の設計方針である『第一種の誤りは犯さない』を少し下方修正したことになる．しかし，我々が計数調査した科学技術文章約350万字の文章中には否定の「ん」は2つしか存在しなかった．科学技術文章に対しては，実際に第一種の誤りを犯す可能性はほとんどないであろう．

今回の評価に使用した文章は科学技術文章の抄録であり，日本語文章全体からみれば偏ったものである．さらに広範な日本語文章で評価を進めていきたい．

謝辞 本研究を進めるにあたり，JICST 廃棄テープの使用について姫路短大の田中康仁先生に便宜をはかっていただいた，また応答時間の測定には製品科学研究所の森川浩氏から提供していただいた打鍵データ収集システムを使用した．ここに記して謝意を表する．

参考文献

1) 牛島和夫，日並順二，尹 志熙，高木利久：日本語文章推敲支援ツール『推敲』のプロトタイピング，コンピュータソフトウェア，Vol. 3, No. 1, pp. 35-46 (1986).
2) 倉田昌典，牛島和夫：日本語文章推敲支援ツール『推敲』のパソコン上での実現と使用，第29回情報処理学会プログラミングシンポジウム報告集，pp. 45-54 (1988).
3) 牛島和夫，石田真美，尹 志熙，高木利久：日本語文章推敲支援ツールにおける受身形の抽出法，情報処理学会論文誌，Vol. 28, No. 8, pp. 894-897 (1987).
4) 菅沼 明，牛島和夫：日本語文章推敲支援ツール『推敲』における字面解析手法とその評価，自然言語処理研究会報告，No. 68, 68-8 (1988).
5) 木下是雄：理科系の作文技術，中公新書 (1981).
6) 吉田 将，日高 達，稲永紘之，田中武美，吉村賢治：公用データベース日本語単語辞書の使用について，九州大学大型計算機センター広報，Vol. 16, No. 4, pp. 335-361 (1983).
7) 倉田昌典，菅沼 明，牛島和夫：日本語文章推敲支援ツール『推敲』における応答時間，第37回情報処理学会全国大会論文集，6B-2 (1988).
8) 森川 浩：キーボードエミュレーションを行うとき望まれるBIOSの機能について，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，16-2 (1988).
9) 大用昌之：次世代ワープロの決め手となる校正支援／可読性評価ツール，日経バイト，3月1日号，No. 43, pp. 96-104 (1988).
10) 浅見直樹：添削支援機能を備えたワードプロセッサ・日立製作所とリコーが発売，日経エレクトロニクス，12月14日号，No. 436, pp. 116-117 (1987).
11) 浅見直樹：文法誤りを検出する文書処理ソフトウェア，日経エレクトロニクス，1月25日号，No. 439, pp. 212-214 (1988).

(平成元年4月24日受付)
(平成2年4月17日採録)

菅沼 明（正会員）

1961年生．1986年九州大学工学部情報工学科卒業．1988年同大学院工学研究科修士課程修了．現在同博士後期課程在学中．日本語処理，ユーザインタフェースに興味を持つ．日本ソフトウェア科学会会員．

倉田 昌典（正会員）

1965年生．1987年九州大学工学部情報工学科卒業．1989年同大学大学院工学研究科修士課程修了．同年ヤマハ(株)入社．ユーザインタフェース，オブジェクト指向，音楽情報処理，日本語処理に興味を持つ．

牛島 和夫（正会員）

1937年生．1961年東京大学工学部応用物理学科（数理工学）卒業．1963年同大学院修士課程修了．同年九州大学中央計数施設勤務．1977年九州大学工学部情報工学科教授（計算書ソフトウェア講座担当），現在に至る．1990年4月から九州大学大型計算機センター長を兼務．工学博士．著書「Fortran プログラミングツール」（産業図書）ほか．日本ソフトウェア科学会，電子情報通信学会，ACM 各会員．