

人間の文章誤り検出能力と誤り検出機能の効果に関する実験†

下 村 秀 樹** 並 木 美 太 郎**
中 川 正 樹** 高 橋 延 匡**

本論文では、人間が文章中の誤りを検出する能力と、計算機上の誤り検出機能が人間に与える効果について報告する。ワードプロセッサ等で書かれた文章の誤りを計算機で自動検出することについては高いニーズがある。われわれはその研究の一環として、誤り検出機能を実用化の際の基礎的データである人間の誤り検出能力、また現実の誤り検出機能が人間に与える効果の評価を試みた。本研究では、まず被験者に簡単な誤りを含む文章をできるだけ普段と同じように読んでもらい、人間の誤り検出能力を定量的に調べるとともに、人間の検出しにくい誤りを考察した。また、試作した2つの誤り検出機能について、それらを使用した場合の人間への効果、弊害を定量的・定性的に調べた。その結果、誤字、脱字、同音異義語誤りのような単純な誤りに対して人間は平均1/3程度しか検出できないこと、個人の検出能力差、検出しにくい誤りの傾向が明らかになった。また、誤り検出機能の効果だけでなく、検出機能の不完全さが生む弊害の存在も定量的に確認した。本論文では、これらの結果とともに、誤り検出機能の実用化に向けて本実験から得られた知見も報告する。

1. はじめに

ワードプロセッサ等で書かれた文章の誤りを計算機で自動検出することについては、高いニーズがある。これに対して、従来から多くの研究が行われている。CRITAC¹⁾、COMET²⁾では、形態素解析に基づき文章の誤りを検出する。われわれも、形態素解析処理時に単語に付けるコストを利用して誤りを検出する手法を提案している³⁾。推敲⁴⁾は、誤りというよりもむしろ問題となりそうな推敲対象箇所を、字面処理によって検出するので、前記の3つとは目的・手法とも異なった推敲支援ツールである。しかし以下では、推敲対象箇所も広い意味での誤りととらえて話を進める。

さて、これらの誤り検出機能は、文章中のすべての誤りを検出し、かつ、誤りでない箇所を1つも検出しないことが理想である。しかし現実には、検出もれ(以下、「第1種の検出誤り」と呼ぶ)や誤りでない箇所の誤検出(以下、「第2種の検出誤り」と呼ぶ)を完全になくすることはできない。特に、処理速度等の制約から字面処理や精度の低い形態素解析処理に基づいて誤り検出を行う場合は、検出の誤りが多く発生することが予測される。

このような現状に対して、誤り検出機能を実用化するために検出の精度を高める研究を行うことは重要で

ある。しかしそれとともに、人間自身の誤り検出能力に関して研究する必要もある。例えば、人間の誤り検出能力や発見しにくい誤りを知り、それを重点的に補助することによって、文章作成者に大きな効果をもたらす機能を提供することができる。また、どのような誤り検出機能も第1種・第2種の検出誤りを避けられないのであれば、そういった不完全な機能が人間にどのような効果・影響を与えるのかを知っておく必要がある。すなわち人間の誤り検出能力に関する研究は、誤り検出機能の実用化のための基礎的データとなりうる。これまで、誤り検出の精度を上げる研究は多く行われてきているが、この観点からの報告は見あたらない。

以上の問題提起から本研究では、被験者に誤りを含む文章をできるだけ普段と同じように読んでもらうという実験を行い、その誤り検出能力を定量的に調べるとともに、人間が検出しにくい誤りを考察した。また、検出誤りを犯す2種類の異なる特徴・能力を持った誤り検出機能について、それらの機能の使用によって人間の誤り検出能力がどのように変化するかを調べた。本稿では上記2点について報告するとともに、誤り検出機能の実用化について、本実験から得られた知見を述べる。

まず、2章では、実験の目的、方法の詳細、実施、結果について述べる。3章では誤り検出機能を用いない場合の人間の誤り検出能力を定量的、定性的に述べ、4章では誤り検出機能が人間に与える効果、悪影響を議論する。5章では誤り検出機能の実用に関して、本実験から得られた知見を述べる。

† An Experiment into the Human's Ability to Detect Errors in Text and the Effect of Error Detection Tools by HIDEKI SHIMOMURA, MITAROU NAMIKI, MASAKI NAKAGAWA and NOBUMASA TAKAHASHI (Department of Computer Science, Faculty of Technology, Tokyo University of Agriculture and Technology).

** 東京農工大学工学部電子情報工学科

2. 誤り検出能力調査実験

2.1 実験の目的と方針

第1章でも触れたが、本研究は文章の誤り検出機能の実用化のために、次の点を明らかにすることが目的である。

- (1) 普通に文章を読む場合の人間の誤り検出能力
- (2) 検出誤りを犯す文章誤り検出機能の人間への効果と悪影響

実験方法として、前者は人間（被験者）に誤りを含む文章を読ませ、そのうちいくつかの誤りを検出できるか、どのような誤りを検出しにくいかを調べればよい。後者については、誤り検出機能を使用した結果を付加した文章を被験者に読ませ、それを誤り検出機能を使用せずに読んだ場合の結果（すなわち、前者の実験の結果）と比較する必要がある。

しかし、この際全く同じ文章を同じ被験者が読んだのでは、文章についての学習がなされ、実験結果にノイズの乗ることが予測される。したがって、同一被験者が同一文章を2度は読まないような条件で実験を行うほうが好ましい。本実験では、実験用に誤り検出機能を2つ用意したので、それらを使わない場合と合わせて最低3種類の文章を用意する必要がある。また、3種類の文章と3種類の誤り検出機能の、ある1通りの組み合わせについてだけ実験を行ったのでは、そこで生じる差が文章間の性質の差によるものか、誤り検出機能使用の効果なのかの評価できない。そこで、被験者も3つのグループに分け、それぞれのグループに文章と誤り検出機能の異なる組み合わせを読んでもらい、その結果を比較した。

次に、用意した誤り検出機能、実験用文章、被験者、実験のローテーション（文章と誤り検出機能と被験者グループの組み合わせ）、実施について述べる。

2.2 用意した誤り検出機能の特徴

誤り検出機能の実例として、われわれが試作した特徴・能力の異なる2つのツールを用意した。この2つは、どちらも単純な字面処理に基づく手法を元としている。

(1) 字種分割照合ツール⁵⁾

文字種の変化に着目して文章を文字列パターンに分割し、それをあらかじめ用意した辞書（誤りのない文章から、同じ規則で切り出した文字列パターンを登録したもの）と比較して、辞書にないパターンを誤りとするものである。誤字や脱字など、偶発的な誤りによ

って発生する不自然な文字列パターンを検出することができる。しかしこのツールは、同音異義語の誤り（例えば、「対象」を「対称」と誤った場合）のように、辞書にあるパターンが誤った文脈で使われている場合には検出できない（第1種の検出誤り）。また、誤りのない文章から切り出された文字列パターンであっても、それが辞書の不備により登録されていなければ、誤りであるとして検出されてしまう（第2種の検出誤り）。

(2) 表記チェックツール

このツールは、あらかじめパターンファイルに登録してある文字列パターンに一致するものを、誤りとして検出する。同音異義語の誤り（「回答」と「解答」など）や送り仮名のゆれ（「行う」と「行なう」など）といった、間違いやすいパターンがあらかじめわかっている誤りの検出に有効である。パターンの記述では、活用型の指定や文字種（平仮名、片仮名など）に対応するワイルドカードの使用、ある文字（文字種）以外と一致する否定演算子の使用が可能である。

しかし、このツールでは文字列パターンとして予測できない偶発的な誤り（誤字や脱字など）に対処できない（第1種の検出誤り）。また、パターンファイルにあるパターンは、それがたとえ正しい文脈で使われていても、誤りであるとして検出してしまう（第2種の検出誤り）。

以上の異なる特徴を持つ2つのツールに加え、ツールを使わないことを形式上1つのツールと考えれば、ツールは3つ用意したことになる。以下では、それぞれツールを使わない場合をツールN、字種分割照合ツールをツールX、表記チェックツールをツールYと呼ぶ。

2.3 実験用文章

実験には、3種類の文章を用意した。以下それぞれをA、B、Cと呼ぶ。これらの元は、同一著者の似た内容（情報処理教育について）の文章^{6),7)}からそれぞれ一部分ずつ、A 4判4ページ程度を抜き出したものである。それらの文脈に合わせて、実際の工学系の論文およびその下書きに現れた誤りを、ほぼ同数（20個程度）埋め込んだ。ただし、文章の意味や内容に関する誤りは被験者ごとの解釈が違ふこと、また、語の呼応や係り受けなどの誤りはどの部分を誤りとするかの判断が難しいことなどから、本実験では対象としなかった。さらに、送り仮名や文末のゆれも個人ごとの基準によって誤りであるかどうかの判断が違ってくる

表 1 実験用文章中の誤り
Table 1 Errors in the texts used in this experiment.

分 類		個 数	例 (括弧内は正しいもの)
変換誤り	漢字同音異義語	18	解答 (回答) 現れる (表れる)
	漢字を平仮名表記	6	ちかい (近い) だいがく (大学)
	その他	3	理解使用と (理解しようと)
入力編集ミス	平仮名	19	だつという (だという) 偏見にに (偏見に)
	片仮名	7	キャレア (キャリア) プグラム (プログラム)
	英 字	7	teacheing (teaching) comptor (computer)
	その他	5	ない.. (ない..) 稽古 (傾向)
合 計		65	

ので、除外した。したがって、今回埋め込んだ誤りは、誤字、脱字や同音異義語の使用誤り、変換誤りなど比較的単純なものである。実験用文章に埋め込んだ誤りの個数とその例を表 1 に示す。

さて、考察では文章ごとの実験結果を比較するので、各文章ができるだけ同じ条件であることが望ましい。そこで、埋め込まれている誤りの個数も、それを各ツールが検出する割合 (以下「検出率 α 」と呼ぶ) も、第 2 種の検出誤りの割合 (以下 β と呼ぶ) も 3 つの文章ではほぼ同じになるように、文章やツール X 用の辞書やツール Y 用のパターンファイルを若干調整した。検出率と β の定義式は、次のとおりである。

$$\text{検出率} = \frac{\text{ツールの検出数}}{\text{埋め込んだ誤りの個数}} \times 100 (\%)$$

$$\beta = \frac{\text{誤りでないのに検出した文字数}}{\text{誤りでない部分の総文字数}} \times 100 (\%)$$

実験用文章それぞれの文字数、各ツールによる検出数、検出誤りの割合を表 2 に示す。ツール X は各文章ともに検出率 60% 前後、 β は 5% 前後、また、ツール Y は各文章とも検出率 20% 前後、 β は 1% 前後である。また、本実験で用いた誤り検出ツールを実際にこれらの文章に使った場合の出力例を図 1 に示す。

★なお、第 1 種の検出誤りの割合 (一般に α と呼ぶ) を用いると、検出率は、
検出率 $=100-\alpha$ (%)
であり、情報としては α も検出率も同じである。本稿では、以下の話を進める都合上、検出率を用いる。

表 2 実験用文章の統計的性質
Table 2 Statistical data of the texts used in this experiment.

項 目 \ 文 章	A	B	C
文 字 数	4086	4662	4882
埋め込んだ誤りの個数	22	21	22
ツール X の検出数 (率)	13(59.1)	12(57.1)	13(59.1)
ツール X の β	5.0	4.9	4.7
ツール Y の検出数 (率)	5(22.7)	4(19.0)	5(22.7)
ツール Y の β	1.0	1.1	1.0

注) β は、第 2 種の検出誤りの割合。

米国において、Dep. of Computer Science が、“Computer”という日本的な“物”のイメージがあっても何等のこだわりがない理由は、“Computer”が単なる機械でなく“自動機械”としての性格を持っていて、大きな知識の好奇心をそそる学問だからである。知識の対照にならない分野は、いずれにしても斜陽化する運命にある。

米国の ACM の CURRICURUM'78 でレコメンドしている Computer Science のカリキュラムは、Computer Science に関する知識の好奇心の対照である研究や開発などを進めるための専門的な能力を身に付けるためには、少なくともこのような科目を専門学科として教えてほしいという、最大公約的なものと思う。

ツール X の出力例

米国において、Dep. of Computer Science が、“Computer”という日本的な“物”のイメージがあっても何等のこだわりがない理由は、“Computer”が単なる機械でなく“自動機械”としての性格を持っていて、大きな知的好奇心をそそる学問だからである。知的好奇心の対照にならない分野は、いずれにしても斜陽化する運命にある。

米国の ACM の CURRICURUM'78 でレコメンドしている Computer Science のカリキュラムは、Computer Science に関する知的好奇心の対照である研究や開発などを進めるための専門的な能力を身に付けるためには、少なくともこのような科目を専門学科として教えてほしいという、最大公約的なものと思う。

ツール Y の出力例

図 1 誤り検出ツールの出力例

Fig. 1 Sample output from error detection tools.

ツールが検出した箇所は、反転表示になっている。

2.4 被験者

被験者は同じ研究室の教官、大学院生、学部4年生、研究生の計33名で、すべて日本人である。これを11人ずつ3つのグループに分けた。このとき、各グループの誤り検出能力に偏りが出ないように、学年別の人数ができるだけ均等になるように配分した。以下3つのグループをそれぞれグループ①、グループ②、グループ③と呼ぶ。

2.5 実験のローテーション

実験のローテーションは、次のような方針を持って作成した。

- (1) 同一被験者（グループ）が、同じ文章を2回以上読まないようにする
- (2) ツール使用の効果の測定に対して実験用文章間の差が影響しないようにする

この2点は、2.1節で議論したとおりである。また、これらを満たすために被験者を3つのグループに分けて実験を行うので、次の点も考慮した。

- (3) ツール使用の効果の測定に対してグループ間の能力差が影響しないようにする

このうち(2)、(3)に関しては、実験用文章の性質やグループ間の誤り検出能力が近くなるようにすでに考慮しているが、それでもすべてに差がないとはいえないので、このローテーションでさらに注意を払うことを意図している。

実際のローテーションを図2に示す。基本的には、被験者グループと使用するツールの組を文章ごとにずらして、3回実験を行うという考え方である。これによって、方針の(1)は満たされる。また、ツールごと

に(図2を縦に)見ると、どのツールもすべての文章に1回使われているし、また、どの被験者（グループ）にも1回使われることになる。もちろん、3回の実験のそれぞれでツール使用の効果が同じように現れることが理想ではあるが、被験者グループ間の能力差によって、ツール使用の効果が各回の実験でばらついて観測されることは当然予想される。しかしそうであってもツールごとに結果を集計すれば、その結果はツール・被験者・文章の偶然の相性を除いて同じ条件である。したがって、文章間の差、グループ間の差は基本的に打ち消され、ツール使用の効果を比較できると考える。

2.6 実験の実施

本実験は、図1に例示したものを紙に出力し、それを読んでもらう方法で行った。実験にさきだって、被験者には実験の方法を次のように伝えた。

- (1) 文章を読んでそのタイトルを推測し、それまでにかかった時間を記録する。
- (2) 実験はできるだけ速く行う。
- (3) 文章中に誤りを発見したら、それを円で囲み正しく直す。
- (4) 文章中の一部が反転表示になっているところは、あるツールによって誤りや注意すべきところを検出したものである。誤りを探すときの参考にしてよい。ただし、誤りでないところを反転したり、誤りを反転し忘れていているところがある。
- (5) 書式上の誤り（禁則など）や文字の全角半角の違いは、誤りとしなない。

実験の目的を「タイトルの推測と時間の計測」としたのは、被験者が普通に文章を読みチェックするという状態を再現したいからである。実験が誤り検出能力を調べるものであることを先に告げてしまうと、そのことに被験者の神経が集中してしまい、普段の誤り検出能力と違った結果がでるおそれがあると考えた。ただし、実験結果の集計では、付けられたタイトル・実験時間は無視した。また、できるだけ速く実験を行うという注意も、文章をチェックするときに、実験だからといって特別に内容を注意深く読む必要はないという意味である。「できるだけ速く」というだけでは表現が強いので、実際にそのように口頭でつけ加えた。ツールの能力については、(4)程度の記述にとどめた。これは、被験者がツールに対して過大な（あるいは過小な）期待を持たないようにするためである。

実験は、被験者の疲れによる能力の低下とノイズレ

実験回数	文章	使用するツール		
		N (なし)	X	Y
1回目	文章A	グループ①	グループ②	グループ③
2回目	文章B	グループ③	グループ①	グループ②
3回目	文章C	グループ②	グループ③	グループ①

グループ①	グループ②	グループ③
教官 : 1	教官 : 1	教官 : 1
大学院生 : 6	大学院生 : 6	大学院生 : 5
学部4年 : 4	学部4年 : 4	学部4年 : 4
		研究生 : 1

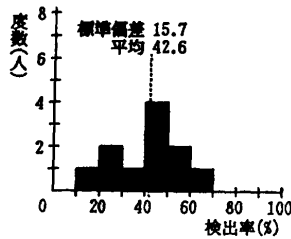
図2 被験者グループのうちわけと実験のローテーション
Fig. 2 Members of the three subject groups and the rotation of these groups in this experiment.

ベルの減少を考え、それぞれ1週間の間隔において、
同じ時間帯で行った。

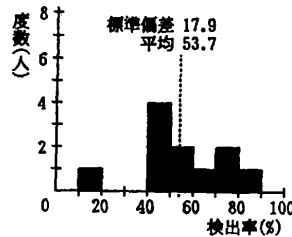
2.7 結 果

実験の結果を、図2のローテーションに対応させる
形で、図3に示す。図のそれぞれは、検出率の分布

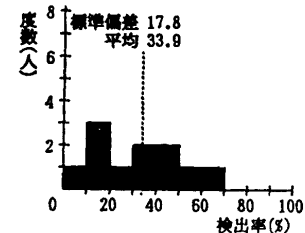
(10%ごと)であり、図中の点線は平均値を示している。
また、下の横1行(j, k, l)は、ツールN, X,
Y別の分布をすべて加算したものである。



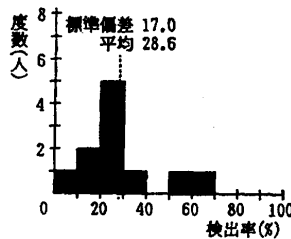
(a) テキストA, ツールN: グループ①
(a) Text A, tool N: Group①.



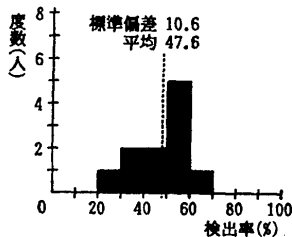
(b) テキストA, ツールX: グループ②
(b) Text A, tool X: Group②.



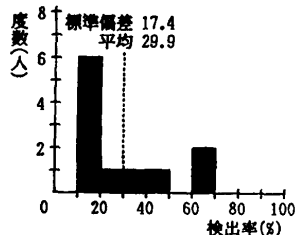
(c) テキストA, ツールY: グループ③
(c) Text A, tool Y: Group③.



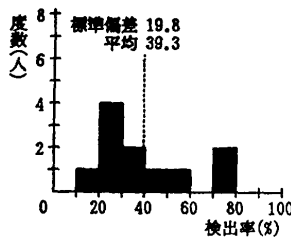
(d) テキストB, ツールN: グループ③
(d) Text B, tool N: Group③.



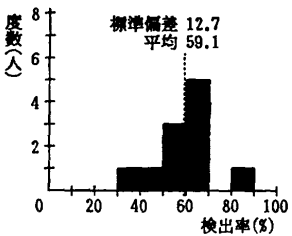
(e) テキストB, ツールX: グループ①
(e) Text B, tool X: Group①.



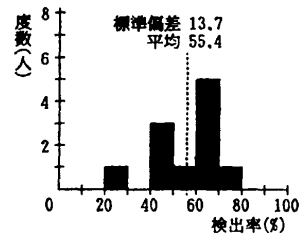
(f) テキストB, ツールY: グループ②
(f) Text B, tool Y: Group②.



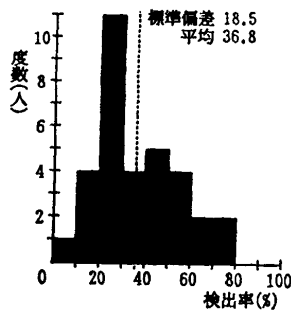
(g) テキストC, ツールN: グループ②
(g) Text C, tool N: Group②.



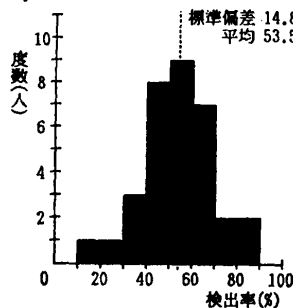
(h) テキストC, ツールX: グループ③
(h) Text C, tool X: Group③.



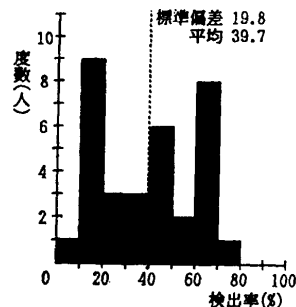
(i) テキストC, ツールY: グループ①
(i) Text C, tool Y: Group①.



(j) ツールN全体
(j) Whole tool N.



(k) ツールX全体
(k) Whole tool X.



(l) ツールY全体
(l) Whole tool Y.

図3 検出率の分布

Fig. 3 Distribution graphs of the human's error detection ratios.

3. 人間の通常の誤り検出能力

3.1 誤り検出能力の分布

人間の通常の誤り検出能力は、本実験でのツールを使わない場合（形式上ツールNを使用した場合、図3の左1列a, d, g, j）の結果によって考察する。

文章ごとの平均検出率は28.6%~42.6%であるが、どの分布も分散が大きく、個人の能力差が非常に大きいことがわかる。例えば、文章B（グループ③）の被験者では、検出率の最低が4.8%、最大が63.6%で、その差が60%近くもあった。

さて、この実験での文章間の平均値の差が有意だとすれば、その原因としては文章に埋め込んだ誤りの差、あるいはグループ間の能力差が考えられる。そこで、“それぞれの結果の平均値が等しい”ことを帰無仮説として、自由度20(11+11-2)でt検定（両側）を行った。しかし、図3のa d, a g, d gのどの結果間についても5%の危険率では棄却できなかった。したがって、文章間やグループ間の差は個人の能力差から見れば誤差の範囲内と考えると差し支えなく、図3のjは「ツールを使わなかった場合の人間の検出率の分布」と考えることができる。この図では、全体の平均検出率は36.8%となり、文章中に現れる単純な誤りに対してでさえも、人間の誤り検出能力では平均して1/3程度しか検出できないことがわかった。

3.2 人間の検出しにくい誤りとその原因

次に、人間（日本人）はどのような誤りを検出しにくいのかについて、本実験で得られた定性的な傾向を述べる。

実験結果を埋め込んだ誤り別に見て、被験者の検出率が低い（20%未満の）誤りを抜き出し、それを次の8つに分類した。分類に対応する実例は、表3にまとめる。

- (1) 同音で似た意味の言葉の誤り（4例）
- (2) 平仮名表記の同音異義語であるときの変換忘れ（2例）
- (3) 誤りであるが一般的に浸透してしまって誤りだと思わない語（1例）
- (4) 助詞や助動詞の重なり（5例）
- (5) 平仮名列中の脱字（3例）
- (6) 英単語の誤り（3例）
- (7) 番号の一貫性の誤り（1例）
- (8) そのほか（2例）

この中で(1)~(3)は、同音語の用法や表記の誤り

表3 被験者が検出しにくい誤り

Table 3 Difficult errors for subjects to detect.

分類	誤	正	延べ誤り数	検出数
(1)	解答	回答	11	0
	表す	現す	33	1
	答える	応える	11	0
	速い	早い	11	1
(2)	やすい	安い	11	2
	したがう	従う	11	0
(3)	関わらず	かかわらず	11	0
(4)	偏見に ことをを	偏見に ことを	11	2
	それにより	それにより	11	1
	これをを	ことを	11	2
	されたされた	された	11	2
			11	2
(5)	ようなる	ようになる	11	2
	しばし	しばしば	11	2
	もであり	ものであり	11	2
(6)	Compter	Computer	11	0
	carriculum	curriculum	11	2
	Introductoon	Introduction	11	2
(7)	1章	2章	22	2
(8)	なけらば	なければ	11	0
	ついてべる	ついて述べる	11	2
合計	——	——	264	27

注1) 分類(1), (2)については、文脈から誤りかどうかを判断した。

注2) 分類(2)の同音異義語は「～しやすい」「～にしたがって」という文脈に現れる。

である。この種の誤りは、同音語であることから、音読しても違いがわからない。また、たとえその箇所を見て文字を確認したとしても、用法や表記についての基準を持っていなければ、誤りだとはわからない。したがって、これらが見逃されるというのは、用法や表記について強い意識を持っている被験者が少なかったことを示している。

これに対して、(4)~(8)は、見つけさえすれば誤りであるとすぐにわかるものが多い。これが発見できないのは、人間（日本人）がこのような箇所をあまり確認していないことを意味している。特に、平仮名列が重なったり抜けたりすることについては、意外とも思えるほど無関心である。これに関する1つの仮説は、平仮名列を見る場合、前後の文脈と文字列を一見した印象から、適切な文字列を推定して読み流していることが考えられる。同じ平仮名の誤りでも、

「工学系のの学科」（6/11の検出率）

のように、漢字と漢字の間に現れた短い平仮名列の場合は、やや見つかりやすくなる。これは、誤りの前後が違う字種（漢字）に挟まれているので、推測した文字列に比べて文字数の多いことが一見してわかるためだと考えられる。

英単語については、1文字1文字を見ているわけではなく、単語を一見した印象と前後の文脈から適切な単語を推測してしまうことが見逃しやすい原因であろう。これも平仮名列の誤りを見逃す原因と非常によく似ている。

4. 誤り検出機能を使った場合の効果

4.1 誤り検出能力全体への効果

誤り検出機能を使わなかった場合と使った場合の検出率の差を、ツールX、ツールYのそれぞれについて議論する。

(1) ツールXの効果

まず、ツールを使わなかった場合（図3の縦左1列）とツールXを使った場合（図3の縦中央1列）の検出率分布を比較する。各文章ごと（図3のaとb, dとe, gとh）に見ると、どの場合もツールXを使用したほうが分布が右（検出率の大きいほう）へ移動していることがわかる。平均検出率で考えると10%~20%も高くなっており、文章や被験者グループによらず、ツールXを使用した効果が現れている。ツールごとに集計した結果（図3のjとk）を比較すると、ツールを使用しない場合に比べ、平均して17%程度の検出率向上となった。また、このjとkの分布に対して、帰無仮説を“ツールXを使用してもしなくても平均値は同じ”，対立仮説を“ツールXを使用したほうが平均値は大きい”として片側でt検定（自由度 $64=33+33-2$ ）を行うと、帰無仮説は有意水準0.1%で棄却される。したがって、統計的に見てもツールX使用の効果は極めて有意な結果となった。

(2) ツールYの効果

次に、ツールを使わなかった場合（図3の縦左1列）とツールYを使った場合（図3の縦右1列）の検出率分布を比較する。文章ごとの比較では、dとf, gとiでは、ツールYを使ったほうが検出率は高くなった。しかし、aとcではその関係が逆転し、Yを使用したほうが8.7%も低い検出率になっている。また、全体（図3のjとl）で比較すると、Yを使用したほうが平均2.9%の検出率向上となっているものの、分布の分散が大きいため統計的に有意とはならな

かった。したがって、検出率の平均値からは、ツールY使用の効果は観測されなかった。

この原因は、ツールYの検出率（20%）が低いので、人間の検出率全体に及ぼす影響が低かったことが考えられる。誤り検出能力の個人差は非常に大きく、標準偏差は今回のどの実験に関しても10%を超えている。また、20%の誤りをツールYが検出してもそれがすべて検出率の向上に直結するわけではない（すなわち、検出しても人間は誤りだと思わないものがある）。これらの原因から、本実験ではツール使用の効果が個人の能力の誤差に含まれてしまったと推測できる。

4.2 ツールが検出可能な誤りに関する効果

ツール使用の効果をさらに議論するために、誤り全体の中で各ツールが検出可能な誤りに着目し、それに対する検出率がツールを使わなかった場合と比べてどう変化したかを調べた。その結果を図4に示す。

ツールXが検出可能な誤りは、3つの文章で合計38（延べ418）あった。これについて、ツールを使わない場合は平均39.2%の検出率だったのに対し、ツールXを使うことによって69.0%となり、30%程度向上した。また、ツールYが検出可能な誤り、合計14（延べ154）については、23.6%が58.3%となり、35%も向上した。したがって、ツールの検出可能な誤りに対する効果は、ツールX、Yともに30%~35%の検出率向上として顕著に現れていることが確認された。また、その値の変化はツールYのほうが大きくなっていることから考えても、4.1節でツールYの効果があまり見られなかったのは、単にYの検出率が低かったためであることが検証された。逆にいえば、20%

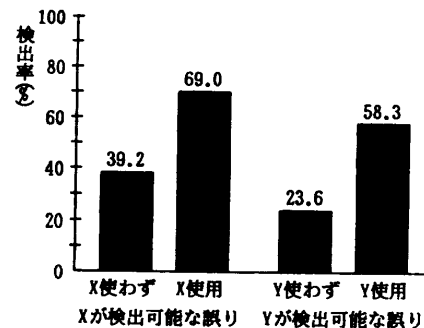


図4 ツールが検出可能な誤りに関するツール使用の人間への効果（全被験者）

Fig. 4 Effect of error detection tools on human error detection of errors detectable by tools (all subjects).

程度の検出率のツールでは、人間個々の誤り検出能力の誤差程度の効果しか得られないことがわかった。

4.3 第2種の検出誤りの弊害

さて、もし誤り検出ツールに第2種の検出誤りがなければ、人間はツールが検出したところをすべて誤りであると判断できるはずである。しかし、第2種の検出誤りがあるとツールの指摘を完全には信用できなくなるので、ツールが検出した箇所の中にある本当の誤りを見逃してしまうという弊害が予測される。実際に図4からもわかるように、本実験ではツールX（第2種の検出誤り5%前後）が検出した誤りのうち平均31.0%（100-69.0）、ツールY（第2種の検出誤り1%前後）が検出した誤りのうち平均41.7%（100-58.3）が見逃され、この弊害をある程度定量的にとらえることができた。

この値は第2種の検出誤りの割合が大きいほど高くなると考えるのが自然である。しかし、この実験では逆に、第2種の検出誤りの多いツールXのほうがこの弊害が小さいという結果になった。これは、ツールYの検出する誤りが同音異義語の誤りなど、人間が一見しても誤りであるかどうかを判断しにくいものであり、ツールが検出したとしても見逃しやすかったことが影響していると考えられる。

4.4 第1種の検出誤りによる弊害

前節では、第2種の検出誤りによる弊害を考察した。検出誤りが起こすもう1つの弊害として、ツールを使わなければチェックできた誤りが、ツールを使うことによって他に注意をそらされてしまい、検出できなくなることが予測される。実際に一部の被験者は、実験後の感想として、「ツールが指摘しない場所への注意が薄れる」ことを挙げていた。これは、第1種の検出誤りがなければ（すべての誤りを完全に検出できれば）もともと起こらない問題である。

このような弊害が起こっているかどうかを調べるために、前節とは逆に各ツールが検出できない誤りについて、ツールを使用した場合と使用しなかった場合の差を調べた。ツールXが検出できない誤りは合計27（延べ297）あり、ツールYが検出できなかった誤りは、合計51（延べ561）であった。結果を図5に示す。図からわかるとおり、ツールX、Yともにツールを使用したほうが平均検出率は若干低下している。個人の誤り検出能力差が大きいので、本実験ではこの差を統計的に有意であるとはいえないが、2つのツールの結果とも同じ傾向を示しており、値は小さいがこの

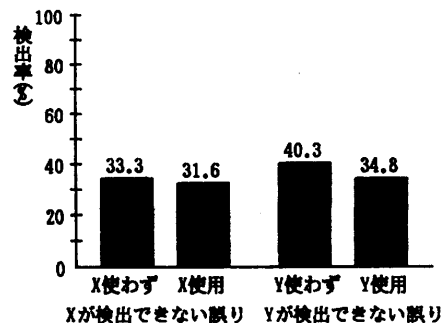


図5 ツールが検出できない誤りに関するツール使用の人間への影響（全被験者）

Fig. 5 Effect of error detection tools on human error detection of errors undetectable by tools (all subjects).

弊害が起こっている可能性は十分にあることがわかった。

4.5 誤り検出能力別に見た効果の差

次に、誤り検出能力が高い人間あるいは低い人間に対して、誤り検出機能を与える効果を議論する。

まず、ツールを使わなかった場合の誤り検出能力を見て、3つの被験者グループからそれぞれ最も検出率の高かった者2名ずつ（合計6名）を選び出した。これを誤り検出能力が高い被験者として、以下「上位6名」と呼ぶ。また、逆に各グループから最も検出率の低かった者2名ずつ（計6名）を選び出した。これを、「下位6名」とする。そして、上位6名、下位6名それぞれについて、ツール使用の効果を比較した。なお、すべての誤りに関する検出率で比較するとツールYの効果が議論できない（4.1節参照）ので、図4同様に各ツールが検出可能な誤りに対する検出率でツールの使用の効果を比較した。上位6名の平均検出率を図6に、下位6名の平均検出率を図7に示す。

図からわかるように、能力の上位下位にかかわらず、ツールの使用はプラス効果になっている。特に下位6名（図7）については、ツールXは46.0（61.8-15.8）%、ツールYは28.5（46.4-17.9）%の効果をあげており、上位6名との差をかなり縮める効果のあることがわかった。しかしながら、下位6名がツールX、Yを使用したとしても上位6名がツールを使用しない場合の水準にしかっていない。したがって、もともと誤り検出能力の低い人に対するツールの効果は大きいものの、個人の能力差もそれと同程度に大きいこともわかった。

一方、上位6名については、検出率の伸びが下位6

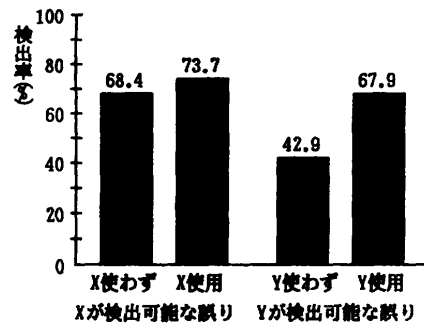


図 6 ツールが検出可能な誤りに関するツール使用の人間への効果 (能力上位 6 名)

Fig. 6 Effect of error detection tools on human error detection of errors detectable by tools (six subjects who have high error detection abilities).

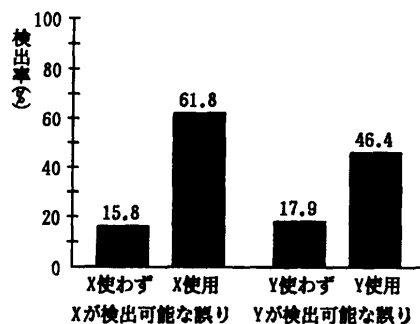


図 7 ツールが検出可能な誤りに関するツール使用の人間への効果 (能力下位 6 名)

Fig. 7 Effect of error detection tools on human error detection of errors detectable by tools (six subjects who have low error detection abilities).

名ほど大きくはない。特に、ツールXの効果はわずか 5%程度しか現れておらず、本実験での検出率の限界がこの近辺にあると推測できる。検出率の限界の原因としては、誤りの中にはツールが検出しても人間が誤りだとは思わないものがあること (4.6 節参照)、あるいは人間の注意力の限界などが挙げられる。したがって、これらの問題を解決しなければ能力上位者に対しても、ツール使用のこれ以上の効果は期待できない。

4.6 人間が検出にくい誤りに対する効果

誤り検出機能は、人間の検出にくい誤りに効果を示すことが最も重要である。そこで、3章で示した人間が検出にくい誤り 21 例のうち、X、Yどちらかのツールが検出した誤り 17 例に関して、ツールの使用によって人間の検出率がどのように変化したかを表 4 に示す。

表からわかるように、どの誤りに対してもツールの

表 4 被験者が検出にくい誤りに対するツール使用の効果
Table 4 Effect of using error detection tools on difficult errors for subjects to detect.

分類	誤	正	延べ誤り数	人間の検出数	
				ツールなし	ツール使用
(1)	解答表す	回答現す	11	0	6
	表す	現す	33	1	13
	答える	応える	11	0	1
	速い	早い	11	1	7
(2)	したがう	従う	11	0	0
(3)	関わらず	かかわらず	11	0	0
(4)	ことをを	ことを	11	1	11
	それにより	それにより	11	2	11
	これをを	ことを	11	2	10
	されたされた	された	11	2	11
(5)	ようなる	ようになる	11	2	6
	しばし	しばしば	11	2	2
	ものであり	ものであり	11	2	5
(6)	Compter	Computer	11	0	7
	carriculum	curriculum	11	2	3
	Introductoon	Introduction	11	2	8
(8)	なければ	なければ	11	0	4
合計	—	—	209	19	105

注) 分類 (7) は、ツールが検出した誤りが存在しなかった。

使用がマイナス効果になる例はない。全体では、5倍以上も多く誤りを人間が検出できるようになっており、ツールの使用は人間の検出にくい誤りに対して大きな効果をあげていることがわかる。特に、平仮名の重なりについては延べ 44 箇所のうち、ツールを使わなければ 7 箇所しか検出できないのに、ツールを使用すると 43 箇所とほぼ完全に検出できるようになった。

しかし一方で、次のような誤りについては、効果がほとんど現れていない。

- (a) 「答える/応える」
- (b) 「したがって/従って」
- (c) 「関わらず/かかわらず」
- (d) 「しばし/しばしば」
- (e) 「carriculum/curriculum」

この中で (b)、(c) に関しては、もともと多くの方がこの使い分けに無関心であり、ツールがいくら指摘しても誤りであると思わないことが原因であろう。しかしそのほかについては、ツールが検出しているのに誤りとして判断できない十分な理由が見あたらない。

現段階で推測するとすれば、誤りの出現位置（前後の文字列との関係、誤りの途中での改行など）の影響が原因として挙げられるが、詳細は今後の課題である。

5. 文章作成支援への考察

文章作成支援に誤り検出機能を取り入れる場合の指針について、本実験の結果から得られることを次に挙げる。

(1) 文章作成に関する教育、オンライン情報参照システムの必要性

本実験の結果から、誤り検出機能が指摘してもそれを人間が誤りであるとは思わずに見逃す場合のあることがわかった(4.3節参照)。この問題の根本的な原因は、第2種の検出誤りの存在である。しかし、文書作成者の知識不足もそれを助長している点が大いと思われる。特に、同音異義語の選択(例：回答/解答)、語の表記の誤り(例：関わず)については、その使い分けや表記に関しての教育を行えば多くが避けられるし、また仮に不注意で誤りを犯したとしても、誤り検出機能の指摘によって誤りを発見することは容易である。逆にいえば、このような教育を十分に行わなければ、誤り検出機能を実用化しても、効果は期待できない。

また、文書作成の知識を持っていたとしても、常にすべてに注意が払えるわけではないし、うっかり忘れてしまうこともある。それらに対しては、誤りに関する情報をオンラインで参照できる環境が必要であろう。例えば、検出機能が誤りと判断した理由、同音語の使い分け、間違いやすい語句の説明がオンラインで得られれば、文書作成者が誤りを検出する際の大きな助けになるとと思われる。

(2) 人間が検出しにくい誤りを検出する機能の強化

人間がチェックしにくい誤りについては、いくつかの傾向があることがわかった(3.2節参照)。すべての誤りに1つの手法やモデルで対応することも興味深い研究ではあるが、速度の重視により処理が字面処理程度の場合、第2種の検出誤りが多くなりやすく、現実的とはいえない。そこで、人間の不得意なところだけに特化し、第2種の検出誤りを抑えたいいくつかの機能を計算機上に用意し、それを組み合わせて文章の誤り検出に利用することも考えられる。例えば、平仮名列の誤りだけに対する機能や、日本語文章中の英単語に対して適用する spelling checker、見出し番号の一貫性だけをチェックする機能などを用意し、それを必

要に応じて使うという方法である。それぞれは比較的単純な機能でも、文章作成者の弱点に対応することによって、大きな効果をあげることが期待される。

6. おわりに

本研究では、人間の文章中の誤り検出能力を調べるとともに、検出誤りを犯す誤り検出機能を使用した際の、人間の誤り検出能力の変化に関する基礎的な実験を行った。その結果、次のことが明らかになった。

(1) 文章を普通に読むと、文章中に含まれる誤字、脱字、同音異義語誤り、変換誤りのうち、1/3強しか発見できない。また、その能力は個人差が非常に大きい。

(2) 同音異義語、平仮名列、英単語、番号の一貫性など、人間(日本人)が検出しにくい誤りには傾向がある。

(3) 誤り検出機能を使うと、その機能が検出した誤りだけに限って見れば、誤り検出機能を使用しない場合に比べて、30%以上も多くの誤りを検出できる。

(4) 約1%~5%の第2種の検出誤りがある弊害として、誤り検出機能が検出しても人間は見逃す箇所が、検出した誤り全体の30%~40%もある。

(5) 統計的に有意であるとはいえなかったが、誤り検出機能が検出できない誤りに関しては、誤り検出機能を使わないほうが人間は多くの誤りを検出できる可能性がある。

本研究は誤り検出機能の実用化に対して、これまでにない視点からの試みであり、小規模な実験ではあるが人間の誤り検出能力と誤り検出機能使用の効果を定量的、定性的に考察している。しかし、まだこれらを完全に明らかにできていない。また、今回の実験は他人の書いた文章を読んでいるが、自分自身が書いた文章の誤りを検出する場合には違った傾向が現れることも予測される。今後は、さらに大規模な実験に基づき、次の点を明らかにすることが課題である。

(a) 検出率、第2種の検出誤りの割合と人間の誤り検出能力の関係の定量化

(b) 誤り検出機能の効果が出ない誤りに関する原因の考察

(c) 自分自身で書いた文章に対する誤り検出能力と誤り検出機能使用の効果の考察

参考文献

- 1) 鈴木恵美子、武田浩一、藤崎哲之助：日本語文書校正支援システム CRITAC、情報処理学会日本

語文書処理研究会資料, 8-5 (1986).

- 2) 福島俊一, 大竹曉子, 大山 裕, 首藤友喜: 日本語文章作成支援システム COMET, 電子情報通信学会オフィスシステム研究会資料, OS 86-21 (1986).
- 3) 下村秀樹, 並木美太郎, 中川正樹, 高橋延匡: 最小コストパス探索モデルの形態素解析に基づく日本語誤り検出の方式, 情報処理学会論文誌, Vol. 33, No. 4, pp. 457-464 (1992).
- 4) 菅沼 明, 石田郎子, 倉田昌典, 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, 情報処理学会自然言語処理研究会資料, 68-8 (1988).
- 5) 下村秀樹, 酒井貴子, 並木美太郎, 高橋延匡: 字面処理による日本語誤り検出の方式, 第44回情報処理学会全国大会論文集, 5C-3 (1992).
- 6) 高橋延匡: 情報処理教育におけるメタ工学, 情報処理学会情報専門学科のコアカリキュラムシンポジウム論文集, pp. 17-24 (1991).
- 7) 高橋延匡: メタ情報工学の試み, 1991年情報処理学会夏のシンポジウム報告集, pp. 111-119 (1992).

(平成4年7月1日受付)

(平成4年9月10日採録)



下村 秀樹 (正会員)

昭和40年生。平成2年東京農工大学大学院修士課程(数理情報工学専攻)修了。同年同大学院博士後期課程進学, 現在に至る。日本語情報処理, 特に仮名漢字変換, 日本語文章作成支援環境の研究に興味を持つ。



並木美太郎 (正会員)

昭和59年東京農工大学工学部数理情報卒業。昭和61年同大学院修士課程修了。同4月(株)日立製作所基礎研究所入社。昭和63年より東京農工大学工学部数理情報助手。平成元年4月より電子情報助手。並列処理, 日本語情報処理のソフトウェア/ハードウェアアーキテクチャに興味を持ち, コンパイラ, オペレーティングシステムなどシステムプログラムの研究・開発に従事する。



中川 正樹 (正会員)

昭和52年東京大学理学部物理卒業。昭和54年同大学院修士課程修了。同在学中, 英国 Essex 大学留学 (M. Sc. in Computer Studies)。昭和54年東京農工大学工学部数理情報助手, 平成元年1月数理情報助教授, 同4月電子情報助教授。オンライン手書き文字認識, 日本語計算機システム, 文書処理の研究に従事。理学博士。



高橋 延匡 (正会員)

昭和8年生。昭和32年早稲田大学第一理工学部数学卒業。同年(株)日立製作所中央研究所入社, HITAC 5020 モニタ, TSS の開発に従事。昭和52年より東京農工大学工学部数理情報教授。平成元年電子情報教授。理学博士。オペレーティングシステム, 日本語情報処理, パターン認識の研究に従事。電子情報通信学会, ソフトウェア科学会, 計量国語学会, ACM 各会員。