ヒューマンインタフェース 45−12
情 報 メ デ ィ ア 8−12
（1992. 11. 13）

文章誤り検出ツールの実用化に対する基礎的実験

下村秀樹，酒井貴子，並木美太郎，中川正樹，高橋延匡
（東京農工大学 工学部 電子情報工学科）

文章の誤り検出ツールを実用化するための基礎的なデータである（1）普通に文章を読むときの人間の誤り検出能力，（2）誤り検出ツールの人間に対する効果・悪影響，を調べる実験を行った．実験では，被験者に誤りを含む文章をそのままで，あるいは試作した誤り検出ツールを使って読んでもらった．その結果，普通に文章を読んだ場合，誤字脱字のような単純な誤りでさえ 1/3 程度しか見つけられないことや，人間が検出しにくい誤りの傾向がわかった．また，誤り検出ツールは人間に大きな効果があるが，検出のミス，特に誤りでない箇所を誤りとして検出してしまうことは，その効果を著しく減少させることが明らかになった．

A Preliminary Experiment into The Practical Use of Tools for The Detection of Errors in Texts

Hideki SHIMOMURA, Takako SAKAI, Mitarou NAMIKI,
Masaki NAKAGAWA and Nobumasa TAKAHASHI

Dep. of Computer Science, Faculty of Technology
Tokyo University of Agriculture and Technology

This study examines, (1)human's ability to detect errors in text and, (2)the beneficial or detrimental effect of error detection tools to this ability. Japanese subjects were instructed to read text containing errors under usual conditions or using the error detection tools. From this experiment, it can be seen that Japanese can only detect 1/3 of errors, and that certain types of errors are difficult to detect. The beneficial effect of error detection tools and the detrimental effect due to their imperfection have also been revealed.

1．はじめに

文章の誤り検出機能を中心とする文章作成支援には高いニーズがあり，従来から研究が行われている[1,2,3]．さて，文章誤り検出機能は，検出もれ（以下「第１種の検出誤り」と呼ぶ）と，誤りでない箇所の誤検出（以下「第２種の検出誤り」と呼ぶ）のないことが理想である．しかし，現状の技術ではこれらの検出誤りを完全になくすことはできない．

誤り検出機能の実用化のために，第１種・第２種の検出誤りを減らす研究が必要であることは，一般的によく認識されている．一方，見落とされがちではあるが（１）人間の誤り検出能力，（２）検出誤りを犯す不完全な誤り検出機能が人間に与える効果・影響，を知ることも誤り検出機能の実用化のためには重要である．例えば，人間の検出しにくい誤りを知り，それに対する機能を強化することは，より効果のある機能を提供することにつながる．また，不完全な誤り検出機能を利用したとき，人間にどのような悪影響があるのかを知っておけば，予め対策を立てることも可能である．このように，人間の誤り検出能力に関する情報は，誤り検出機能の実用化の基礎データとなる．

本研究では，比較的単純な誤りを含む文章を被験者に読んでもらい，その誤り検出能力を定量的，定性的に調べた．また，誤り検出ツールを試作し，その使用が人間の誤り検出能力に与える効果と不完全さの悪影響を調べたので，以下に報告する．

2．人間の誤り検出能力に関する実験

2.1　実験の概要

この実験では，文章の誤り検出に関する（１）人間の能力，（２）不完全な誤り検出ツールの効果・悪影響，の２点を調べることが目的である．（１）は，被験者に誤りを含む文章を読ませ，その検出能力を調べればよい．（２）については，誤り検出ツールを使って文章を読んでもらい，（１）の結果と比較する．しかしこのとき，全く同じ文章を同じ被験者が２回以上読んだのでは，文章や誤りの位置を覚えてしまうことが予測される．したがって，文章を複数用意し，また被験者も複数のグループに分け，同一被験者が同じ文章を読まないように工夫する必要がある．次に，用意した誤り検出ツール，実験に使用した文章，被験者，実験のローテーション（同一被験者が同一文章を読まないための工夫），実験の実施を説明する．

2.2　用意した誤り検出ツール

本実験には，我々が試作した２つの誤り検出ツールを使用した．どちらも単純な字面処理に基づく手法を元にしている．

（１）字種分割照合ツール[4]

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，それを予め用意した辞書（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．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．

（２）表記チェックツール

このツールは，予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予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．

もちろん，２つのツールとも，第１種・第２種の検出誤りがある．以上の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．

表1　実験用文章中の誤り

分類	個数	例（括弧内は正しいもの）
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）
平仮名入力ミス	19	だとという（だという）　偏見にに（偏見に）
片仮名入力ミス	7	キャレア（キャリア）　プグラム（プログラム）
英字入力ミス	7	teacheing（teaching）　compter（computer）
その他	8	ない．．（ない．）　稽古（傾向）
合計	65	

2.3　実験用文章

同一被験者が同一文章を2回以上読まずに実験が行えるように，誤り検出ツールの数に合わせて3種類の文章を用意した．以下それぞれを文章A，B，Cと呼ぶ．これらは，質・量ができるだけ同じであることが望ましい．そこで，同一著者の似た内容（情報処理教育について）の文章[5,6]から一部分（A4版4ページ程度）を抜き出し，文脈に合わせて実際の工学系の論文やその下書きに現れた誤りを，ほぼ同数ずつ（20 個程度）埋め込んだ．また，それを誤り検出ツールが検出する割合（以下，ツールの「検出率*」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，次の通りである．

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

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

埋め込んだ誤りの分類，個数，例は表1に示す．誤りは，誤字，脱字や同音異義語の使用誤り，など比較的単純なものである．また，実験用文章それぞれの文字数，各ツールによる検出率，第2種の検出誤りの割合を表2に示す．

表2　実験用文章の統計的性質

文章	A	B	C
文字数	4086	4662	4882
埋め込んだ誤りの数	22	21	22
ツールXの検出率(%)	59.1	57.1	59.1
ツールXのβ(%)	5.0	4.9	4.7
ツールYの検出率(%)	22.7	19.0	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．

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，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，同一被験者が2回以上同じ文章を読むことはない．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，また，すべての被験者に，1回だけ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，ツール使用の効果はばらついて観測されるであろう．しかし，そうで

---

* 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 − α (%)

であり，情報としてはαも検出率も同じである．本稿では話を進める都合上，検出率を用いる．

あってもツールごとに結果を集計すれば，ツール・被験者・文章の偶然の相性を除いて同じ条件となる．

実験回数	文章	使用するツール		
		N	X	Y
1回目	A	G①	G②	G③
2回目	B	G③	G①	G②
3回目	C	G②	G③	G①

注）“G①”はグループ①を意味する．
“G②”，“G③”も同様である．

図1　実験のローテーション

2.6　実験の実施

本実験は，テキストを紙に出力し，それを読んでもらう方法で行った．誤り検出ツールが検出した箇所は，反転表示して出力した．被験者には実験前に，目的と方法を次のように指示した．

（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，誤りとしない．

実験の本当の目的を明らかにしなかったのは，被験者にできるだけ普段と同じように文章を読んでもらうためである．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注意深く読む必要はないという意味である．ただ，無理に速く読んでも不自然なので，普通に読めばよいことを口頭で付け加えた．誤り検出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ベルの減少を考え，それぞれ1週間の間隔をおいて，同じ時間帯で行った．

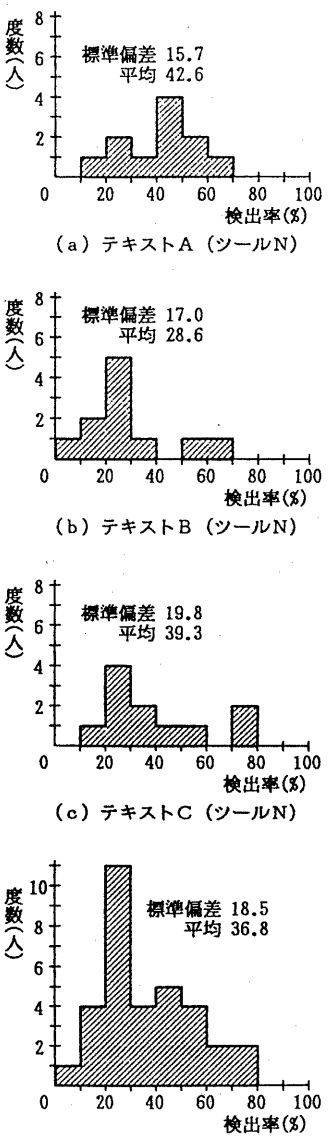


図2　誤り検出ツールを使わないときの人間の誤り検出能力の分布

3．人間の通常の誤り検出能力

3.1 誤り検出率の分布

誤り検出ツールを使わない場合（形式上ツールNを使用した場合）の人間の誤り検出率の分布を図2に示す．文章ごとの平均検出率（図のａ，ｂ，ｃ）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）では，被験者の最低検出率が 4.8%，最大検出率が 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意味を持つとすれば，文章の差，被験者グループ間の能力差が原因として考えられる．しかし，“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）を行った結果，図2のａｂ，ａｃ，ｂｃ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．これは，個人の能力差に比べて，文章やグループ間の差は誤差の範囲内であることを意味する．そこで，「人間の普通の誤り検出能力」として，3つの結果を加算した結果が図2の（ｄ）である．その結果，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という結果になった．

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）意味の似た同音語の誤り
（２）平仮名表記の同音異義語があるときの変換忘れ
（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
（４）助詞や助動詞の重なり
（５）平仮名列中の脱字
（６）英単語の誤り
（７）番号の一貫性の誤り
（８）そのほか

この中で（１）～（３）は，同音語の用法や表記の誤りである．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．

表３　被験者が検出しにくい誤り

分類	誤りの例（括弧内は正しいもの）	延べ誤り数	検出数
(1)	解答（回答）　表す（現す）　速い（速い）	66	2
(2)	やすい（安い）　したがう（従う）	22	2
(3)	関わらず（かかわらず）	11	0
(4)	偏見にに（偏見に）　されたされた（された）	55	9
(5)	ようなる（ようになる）　もであり（ものであり）	33	6
(6)	Compter（Computer）　Carriculum（curriculum）	33	4
(7)	１章（２章）	22	2
(8)	なけらば（なければ）　ついてべる（ついて述べる）	22	2
合計	―――――――――――――――	251	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」いう文脈に現れる．

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．ただし，同じ平仮名の誤りでも，「工学系ののの学科」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（6/11 の検出率）．これに関する1つの仮説としては，「長い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し読み流している」ことが考えられる．もしそうだとすれば，誤りの前後が違う字種（漢字）に挟まれている短い平仮名列の誤りが比較的発見しやすくなるのもうなずける．英単語の誤りが見つかりにくいのも同様の理由と思われる．

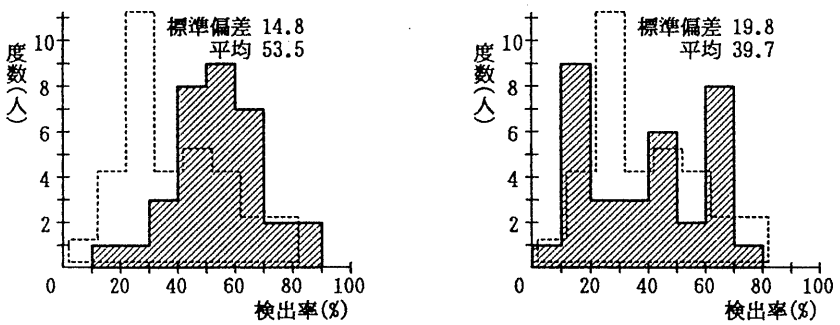
4．誤り検出機能を使った場合の効果と影響

4.1　誤り検出能力全体への効果

図3に，誤り検出ツールを使った場合の人間の誤り検出率の分布を示す．これは，3回の実験の結果をツールごとに加算したものである．実験のローテーションを工夫したことによって，偶然の相性を除いては，どれも同じ条件になっているので，ツールの効果が比較できる（2.5参照）．比較のために図中に点線で，ツールを使わなかった場合の検出率の分布（図2（d））を示す．

この結果から，ツールXを使った場合は，使わない場合に比べて分布が検出率の高い方へ移動していることがわかる．平均では，ツールを使わない場合に対して 17% も検出率が向上した．

一方，ツールYについては平均で 3% 程度しか人間の検出率が向上していない．この原因は，ツールYそのものの誤り検出率が 20% 前後と低いので，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．誤り検出能力は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，ツールが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．



（a）ツールX使用合計　　（b）ツールY使用合計

図3　誤り検出ツールを使ったときの人間の誤り検出率の分布
（破線はツールを使わない場合の検出率の分布）

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 391）あった．これに関しては，ツールＸを使用した結果，ツールを使わない場合に比べて，検出率が 30%（69.0-39.2）も向上した．また，ツールＹが検出可能な誤り，合計 14（延べ 151）については，35%（58.3-23.6）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の方が大きくなっていることから考えても，４.１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．

4.3 検出誤りの弊害

もし誤り検出ツールに第２種の検出誤り（誤りでないところを検出してしまう）がなければ，人間はツールの指摘を完全に信用することができる．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いので，ツールが検出した本当の誤りを見逃してしまうという弊害が起こる．実際に図４から計算すると，ツールＸ（第２種の検出誤り 5% 前後）については 31.0%（100-69.0）が，ツールＹ（第２種の検出誤り 1% 前後）については 41.7%（100-58.3）が，ツールが検出したにもかかわらず見逃されている．この結果から，わずか数%の第２種の誤りがあるだけで，誤り検出機能の効果は，30%～40% も失われることがわかった．

また，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」が挙げられる．一部の被験者は，この弊害を実感したと感想を述べていた．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，図４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検出率の差を調べた．ツールＸが検出できない誤りは合計 27（延べ 281）あり，ツールＹが検出できなかった誤りは，合計 51（延べ 521）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方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．しかし，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．

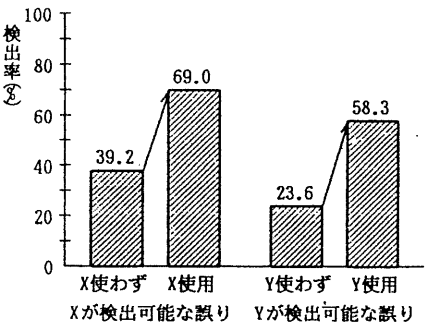


図４　ツールが検出可能な誤りに関するツール使用の人間への効果

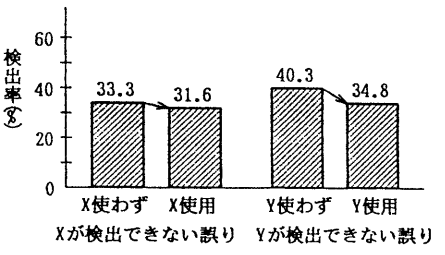


図５　ツールが検出できない誤りに関するツール使用の人間への影響

5．文章作成経験による誤り検出能力の差

文章の誤り検出能力は，個人差が大きい．この原因の一つとして，文章作成経験の差が考えられる．そこで，被験者を卒業論文や投稿論文を書いたことがある大学院生以上と，学部４年生に分けて結果を集計し，その誤り検出能力を調べた．結果を図６に示す．

ツールを使わない場合，文章作成経験の多い大学院生の方が 10% 程度検出率が高い．また誤り検出ツール使用の効果は大学院生以上，学部４年生のどちらにもよく現れているが，その差はツールを使っても全く縮まらないという結果になった．これは，誤り検出ツールを使ったからといって，文章経験差を補えるわけではないことを示している．もちろん，誤り検出ツールが完全であれば，ツール使用の効果の差を議論する必要はない．しかし，不完全なツールを使用せざるを得ない現状では，誤り検出能力は文章作成者の経験（あるいは意識）に依存する部分が大きく，ツールはあくまでもその補助でしかあり得ない．

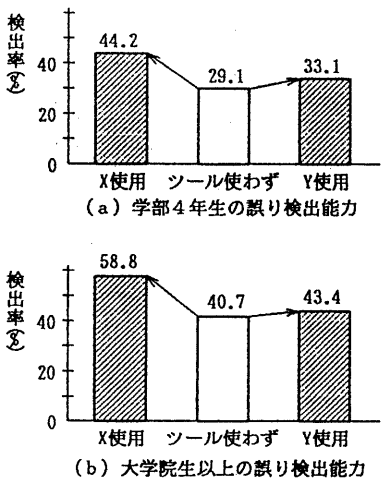


図６　文章作成経験による誤り検出能力の差

6．おわりに

本稿では，人間の文章中の誤り検出能力を調べるとともに，不完全な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の結果を報告した．

本研究の最終的な目標は，人間の誤り検出能力とともに誤り検出機能が人間に与える効果・影響を明らかにすることである．今回の実験はその第一段階であり，問題の大まかな性質をつかんだと位置付けられる．今後は，大規模な実験に基づき，より精密な評価，考察を行いたい．

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，大竹暁子，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，OS86-21 (1986)

[3]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)

[4]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文誤り検出の一方式，第44回情報処理学会全国大会，5C-3 (1992)

[5]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)

[6]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)