

文書校正検査によるエラー検出特徴の分析(1)

文書校正検査の開発

大橋 智樹

(宮城学院女子大学)

中村 順子

(宮城学院女子大学人文科学研究科)

【背景】

文書校正検査の適用可能性は、広いと考えられる。基礎心理学的適用においては、注意メカニズムの検討、文字認識過程の検討に用いることができる。そして、応用心理学的適用においては、産業現場等における個人特性測定用検査として、また学力テストに代わる新たな能力検査ツールとすることができる。しかし、精緻な条件操作を含む標準的な検査はない。

【目的】

広範な領域への適用を前提とし、かつ、厳格な条件操作に基づいて文書校正検査を開発することを目的とする。本研究で開発した検査を用いた実験については、研究(2)を参照のこと。

【対象となる元文章の選定】

文書校正検査に求められる特性として以下の5点を考えた。

- 適度な長さの文章であること
- 文章が中程度の難易度であること
- 文章が「正しい日本語」で書かれていること
- 政治的・宗教的内容を含まないこと
- (著作権に問題のないこと)

これらを踏まえ、高等学校の教科書に掲載されているエッセイ等を対象として選考。

尚学図書「新選現代文」(1999)掲載の中村光夫著「移動の時代」を対象とした(総文字数 2977 文字)。

【元文章への加筆・修正等】

加筆・修正箇所は、

「漱石」を「夏目漱石」と姓を加筆(「漱石」を検出対象エラーとしたため; 瀨石)

「おそらく」という単語を一箇所削除

以上2点。これら以外の漢字、仮名遣い、句読点は全て教科書に掲載されている原文のままとした。その上で文字の置き換え、脱落によるエラーを表現した。

【検査に含めるエラーの種類】

文字形状が似ているエラー(類似型)

- 偏や旁が異なるが、読みが同じ(ex. 振動 販動)
- 偏や旁が異なり、読みも異なる(ex. 技術 枝術)

文字の一部の長さが異なる(ex. 失う 矢う)
(上から類似1, 類似2, 類似3)

文字の順序が入れ替わるエラー(交換型)

おそらく おそくら
しえませんが しれせまんが
ex. 高まらざる 高まざる

(上から交換1, 交換2, 交換3)

一部の文字が脱落している(脱落型)

単語が行をまたぐ(ex. おぼつかない おぼ/かない)

単語が行をまたがない

(上から脱落1, 脱落2)

【エラーの設定について】

文章の内容を考慮した上で、ほぼ均等な字数となるように3パートに分け(1154文字, 877文字, 946文字)、それぞれに全条件で同数ずつのエラーを挿入。

校正検査成績がエラーの挿入箇所によって異なる可能性を検討するため、同じエラー(交換型; しれませんが しれせまんが)を全てのパートに挿入。

【パートごとのエラー数と全エラー単語】

パート1(1154文字)							
	音	形	条件名	出現数	エラー単語(文字)		
類似型	同じ	偏旁が異なる	類似1	3	障壁	販動	歎念
	異なる	偏旁が異なる	類似2	3	推新	枝術	政治
	異なる	長さが異なる	類似3	1	地球士		
交換型	おそらく		交換1	1	おそらく		
	しれませんが		交換2	1	しれせまんが		
	その他		交換3	2	ほどんと	高まざる	
脱落型	行またぎ		脱落1	1	できなかつ/という		
	その他		脱落2	1	なりつある		

パート2(877文字)							
	音	形	条件名	出現数	エラー単語(文字)		
類似型	同じ	偏旁が異なる	類似1	3	遠慮	前提	徑管
	異なる	偏旁が異なる	類似2	3	社会	象徴	待ち受けて
	異なる	長さが異なる	類似3	1	刀強く		
交換型	おそらく		交換1	1	おそらく		
	しれませんが		交換2	1	しれせまんが		
	その他		交換3	2	いれわて	まごつかない	
脱落型	行またぎ		脱落1	1	な/ている		
	その他		脱落2	1	おぼかない		

パート3(946文字)							
	音	形	条件名	出現数	エラー単語(文字)		
類似型	同じ	偏旁が異なる	類似1	3	構漢	特殊	累積
	異なる	偏旁が異なる	類似2	3	倫じて	瀨石	意味
	異なる	長さが異なる	類似3	1	突い		
交換型	おそらく		交換1	1	おそらく		
	しれませんが		交換2	1	しれせまんが		
	その他		交換3	2	ありまさ	よいより	
脱落型	行またぎ		脱落1	1	は/きり		
	その他		脱落2	1	言うでも		