

日本人の英語の特性を考慮した誤綴訂正

2J-4

岡田岳之、古郡延治

電気通信大学

○はじめに

英文ワープロにおいて右ぞろえ、ハイフン付けとともに誤綴の検出と訂正はワードプロセッサの標準ソフトウェアとなった。これらは英語を書く必要のある日本人にとってはたいへん便利である。しかし、市販の誤綴訂正プログラムには不満な点も残っている。というのは、それらが英語を母国語としない日本人のおかす誤綴に対して必ずしも適切な対処をしないからである。

本報告は日本人の誤綴の特性を調査し、その結果を用いて市販の誤綴訂正プログラムの訂正効率の向上を図ろうとするものである。

○誤綴の調査

大学生を対象に日本人の英語の誤綴を調査した報告によると誤綴の約7割は英語と日本語の音韻の違いを反映したものである[1]。その特性をさらに追求するため、国語辞典に載っており、普通の大学生なら知っていると思われるカタカナ英語100語を選び、106人の大学生にもとの英語表記に直してもらった。ここにその結果から正解率の高かった10単語と低かった10単語を示す([表1])。

上位10単語	正解率(%)	下位10単語	正解率(%)
highway	92	hamburger	8
wonderful	89	barricade	7
champion	89	rehearsal	7
standard	86	ecstasy	7
demonstration	84	sponsor	7
lucky	82	terrace	5
captain	82	mattress	4
comedian	79	pamphlet	4
penalty	77	hyphen	3
bridge	76	bracelet	2

[表1]テスト結果の一部

○誤綴の特性

調査した100語のそれぞれについて2個以上の同じ誤り綴りが発見されたものをあげると、全部で658個であった。これらの誤綴を調べてみると、結局、音韻に起因した書記素の変換を行っているものが多い(例 charity→chality)。以下にその代表例と生起回数を示す([表2])。

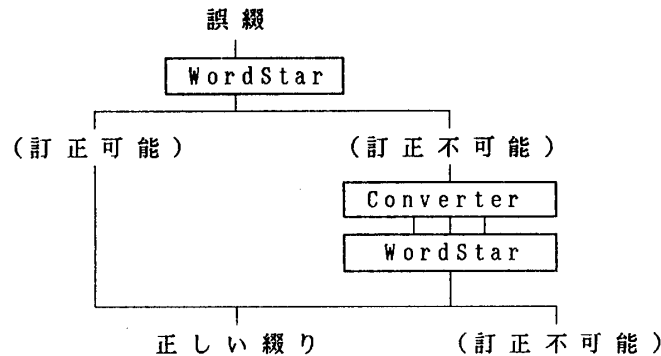
r -> l(466)	y -> i(153)
l -> r(371)	m -> n(135)
a -> u(251)	b -> v(117)
s -> c(208)	c -> s(107)
u -> a(181)	e -> u(104)
or -> er(174)	ph -> f(103)

[表2]変換の代表例

○ 誤綴訂正の改良

市販されているワードプロセッサの中で、誤綴の検出・訂正機能をもつものとしてMicroPro社のWordStarがある。これを使って前出の誤綴658個の訂正を試みたところ、その中の13個は他の正しい綴りとなって誤りであることがわからなかった(例 locker→rocker)。残りの645個のうち、392個(61%)の訂正が可能であった。

WordStarで訂正不可能であった253語の誤綴の訂正を行うために図1に示すようなコンバータを作って簡単な実験を試みた。



[図1]全体構成図

コンバータは日本人が誤って使ったと思われる書記素を、もとの書記素に置換するというものである。

たとえば mattress に対する誤綴 matless は、

m	a	t	r	e	s	s	{	l	->	r	}
m	u	t	l	e	s	s	{	a	->	u	}
m	a	t	l	e	c	s	{	s	->	c	}

などに変換される。

こうしてコンバータによって変換された誤綴をもう1度WordStarに入力し、それによって新たに訂正が可能となった誤綴の数を調べてみた。その結果、さらに111個が訂正可能となり全体の訂正率は78%(17%の改善)まで上がった。驚くべきことに、rとl, bとv, aとuのうちどれか1つを交換することによって、111個のうち102個が訂正可能になるということもわかった。

○ 考察

誤綴はタイポロジカルなものでない限り、書き手本人は正しいと思っているか、あるいは、大体の形はわかっているが細かいところで迷っていたりする場合が多い。特にbとv, rとlなどは日本人を大いに苦しめている。したがって、以上のような簡単な手段で誤綴の訂正率が上がれば、訂正プログラムの実用性はさらに高まるわけである。しかし、コンバータは現在のところad-hocな簡単なものである。もう少し大量の誤綴データを得ることによって、コンバータをもっと形式化し、訂正率をあげることを考察中である。

◇ 参考文献

[1]古郡,平沼:日本人の英語の特性とスペリング・エラーの検出と訂正,計量国語学, Vol.16, No.1, pp.16-27, 1987.