# 「無意味つづり」はどのくらい無意味なのか？
# －「日本語らしさ」の評定による分析－

田中吉史・前田千浩
金沢工業大学情報フロンティア学部心理情報学科 (tanakay@neptune.kanazawa-it.ac.jp)

## 1. 目的

無意味つづり産出課題は、制限時間内に 2 から 10 文字のひらがなからなる無意味な文字の組み合わせ（無意味つづり）をできるだけ多く産出させる課題である。人間の産出した無意味つづりは、使用される文字の出現頻度や文字の組み合わせなどにおいて、実在する日本語の有意味語からの影響が強く見られること、またランダムな文字列や実在する有意味語よりも使用される文字に偏りが見られることが報告されている (田中,2007)。このような特徴がどのようにして生じるのかは興味深い問題と言える。

これまでの研究では、専ら文字の出現頻度など統計的な特性にのみ基づいて議論されてきた。しかし、人間が産出する無意味つづりは、どのような文字列が「日本語らしい (らしくない)」のかといった直感的な判断によっても影響されていると考えられる。

そこで本研究では、文字頻度の統計的な特性が完全にランダムなものや有意味語における文字頻度を反映する無意味つづりをシミュレーションによって生成し、人間の産出した無意味つづりは実際にはどの程度「日本語らしくない」と感じられるのか、またどのような特徴を持つ無意味つづりが「日本語らしくない」と見なされやすいのかを探索的に検討することを通して、人間の産出する無意味つづりの特徴について考察する。

## 2. 方法

### 2.1 シミュレーションによる無意味つづりの産出

次のような 3 種類の無意味つづりを生成するシミュレーション・プログラムを作成し、無意味つづりを産出させた。

(1) ランダムつづり：文字の出現確率がすべて均等であり、ランダムに文字が組み合わせられる。

(2) 低再現度つづり：文字の出現確率が、有意味語におけるそれに従う。ただし、つづりの中の位置（先頭か、末尾か、それ以外か）にかかわらずどの位置の文字もすべて同じ確率分布（何文字目かを無視した有意味語における文字の全般的な出現確率）に従って文字が出現する。

(3) 高再現度つづり：文字の出現確率が、有意味語におけるそれに従う。ただし、つづりの中の位置（先頭か、末尾か、それ以外か）によって確率分布が異なり、つづりの先頭文字は有意味語における先頭文字の出現確率に、つづりの末尾文字は有意味語における末尾文字の出現確率に、それ以外の位置の文字は (2) と同様の出現確率に従う。

### 2.2 無意味つづりの「日本語らしさ」の評定

上記の 3 種のシミュレーション・プログラムによって生成された無意味つづりに加え、大学生の被験者が産出した無意味つづり (田中,2007)、ひらがな表記した有意味語の 5 カテゴリーから各 100 個ランダムに選択した。つづり (語) の長さは全カテゴリーで同じになるよう揃えた。そして合計 500 個のつづり (語) を 250 個ずつ 2 つに分け、ランダムに並べて評定欄と共に印刷した冊子を 2 種作成した。これを大学生と教員（各冊子につき 10 人と 8 人）に呈示し、各つづり（や語）がどの程度「日本語らしい」か（実在する日本語に近いと感じられるか）を 7 段階で評定してもらった。

## 3. 結果

各種のつづりについての「日本語らしさ」の評定の平均値を図 1 に示した。分散分析の結果主効果が有意であり ($F(4, 499) = 174.730$)[1)]、多重比較 (Tukey の HSD 法) の結果、低再現度つづりと被験者による無意味つづりに差が無く、それ以外のすべてのつづり間に差が見られた。田中 (2007)

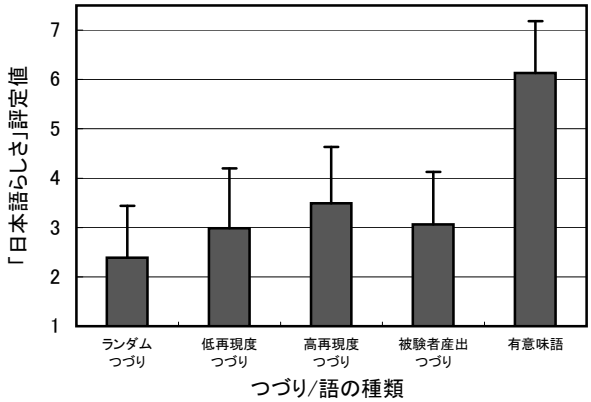


**図 1** 各種つづりに対する「日本語らしさ」の評定値

1) 以下の統計的検定では有意水準はすべて 5%に設定した。

によると、被験者の産出した無意味つづりは、先頭文字、末尾文字の出現頻度が、有意味語における先頭文字、末尾文字の出現頻度と類似していたので、高再現度つづりの方がより被験者の産出したつづりに近いと予想される。しかし、「日本語らしさ」の平均評定値を見ると、人間の産出した無意味つづりは、有意味語よりも先頭文字や末尾文字の出現頻度を考慮していない低再現度つづりの方に近かった。このことは、被験者が産出した無意味つづりには、明らかに「日本語らしくない」と感じさせるような特徴を持つものが、高再現度つづりよりも多く含まれていることを示唆する。

日本語の文字の組み合わせには一定の規則があり、そうした規則に違反する文字の組み合わせを含むつづりの評定値は低くなると考えられる。表1に、冒頭文字が「ん」「っ」「ー」「ゃ・ゅ・ょ」のつづりの頻度と平均評定値を示した。田中 (2003) の報告同様、こうしたつづりは被験者の産出したつづりでは殆ど見られなかった。

次に、「ゃ・ゅ・ょ」を含むつづりの頻度と、「ゃ・ゅ・ょ」の前の文字が拗音を形成しない「い」段以外の文字 (「くょ」「ばゃ」など) となっているかどうかを表2に示した。被験者の産出したつづりでは、「ゃ・ゅ・ょ」の用いられる頻度は少ないものの、ランダムつづりや低再現度つづりと同様、「ゃ・ゅ・ょ」が拗音を形成しないような組み合わせになっていることが多かった。

また「っ」が発音不可能な位置や文字の組み合わせになっているつづり（先頭や末尾、「うっん」などの組み合わせ）の数は、ランダムつづり (4個) と低再現度つづり (5個) でのみ見られ、他のカテゴリーでは見られなかった。

以上で検討した文字の組み合わせの規則に違反する各種つづりに対する平均評定値 (1.063〜1.656) は、$t$ 検定の結果、有意味語を除く平均評定値 (2.981) よりもいずれも有意に低いことが確認された。そこで、これらを除いた444個のつづり (語) について、改めてつづりの種類による日本語らしさの評定値の差について分散分析を行ったところ主効果が有意であり ($F(4, 499) = 157.834$)、多重比較の結果、低再現度つづり、高再現度つづりと被験者による無意味つづりに差が無く、それ以外のすべてのつづり間に差が見られた。このことから、被験者の産出したつづりの日本語らしさ評定値が低かったのは、こうした文字の組み合わせの規則に違反するつづりの存在が影響していたためと考えられる。

さらに有意味語を除き、日本語らしさの評定値とつづりの長さ、およびつづりに含まれる濁音・半濁音の比率との相関を見たところ、いずれも有意な負の相関が見られ、つづりが長いほど ($r = -.516$)、また濁音・半濁音の比率が高いほど日本語らしさの評定値が低下していた ($r = -.311$)。

**表 1**　冒頭が「っ/を/ん/ー/ゃゅょ」のつづりの頻度

| | っ | を | ん | ー | ゃゅょ | 合計 |
|---|---|---|---|---|---|---|
| ランダム | 2 | 1 | 3 | 2 | 6 | 14 |
| 低再現度 | 2 | 0 | 8 | 0 | 7 | 17 |
| 被験者産出 | 0 | 1 | 2 | 0 | 0 | 3 |

**表 2**　「ゃ・ゅ・ょ」の前の文字の種類と頻度

| | い段 | い段以外 | 合計 |
|---|---|---|---|
| ランダム | 1 | 14 | 15 |
| 低再現度 | 1 | 15 | 16 |
| 高再現度 | 3 | 1 | 4 |
| 被験者産出 | 1 | 6 | 7 |
| 有意味語 | 4 | 0 | 4 |

## 4. 考察

実験の結果、人間が産出したつづりは平均すると高再現度つづりよりも「日本語らしさ」が低いと評定され、またそれは文字の組み合わせ規則に違反するつづりが含まれるためであることがわかった。また、人間の産出したつづりでは、先頭のように目立つ位置の文字を操作するよりも、有意味語においては低頻度の「ゃ・ゅ・ょ」を使用する際に不規則に文字を組み合わせることによって「日本語らしくない」つづりが作られていることが示唆された。しかし、全体的に見るとそのような文字の組み合わせ規則に違反するつづりの頻度は少なかった。

また、つづりが長いほど、また濁音・半濁音が多く含まれるほど日本語らしさの評定値は低下していた。もし被験者がこれらの「日本語らしくない」つづりの特徴を意識していれば、より長いつづりを作ったり、濁音・半濁音をより多く使う傾向が生じるはずである。しかしそれとは逆に、人間の産出するつづりは全般的にはあまり長くなく、濁音・半濁音も有意味語よりも少ない傾向がある (田中,2003,2007)。このことは「日本語らしくなさ」の判断が、与えられた文字列の日本語らしさを判断する場合と、無意味つづりを産出する場合とで異なることを示唆すると考えられる。

ただし、今回検討したつづりや語は、各カテゴリーで100個と少なかった。今後さらに多くのつづりについて検討し、本報告の一般性を検証する必要がある。

## 文　献

田中吉史 (2003). 無意味つづり産出課題におけるヒントと反復産出の効果.『認知科学』, **10**(2), 223-243.
田中吉史 (2007). 無意味つづり産出課題における有意味語の影響. 『認知科学』, **14**(2), 206-216.