

患者向け説明文書の可読性判定

酒井由紀子（慶應義塾大学信濃町メディアセンター）

yukiko@lib.keio.ac.jp

1. 背景

患者には病気や治療に関わる様々な情報が必要とされ、医療関係者は一般市民である患者にも理解しやすい情報を提供することが求められている。このため、米国では患者向け文書の「理解のしやすさ」を計る可読性判定が盛んに行われてきた。しかし、国民の一般的な読み書き能力（Literacy）が総じて高く、患者向けの提供情報の開発が遅れている日本では、同分野での客観的な研究は未見である。

2. 目的と研究概要

本研究の最終目的は、日本語の一般向け健康関連文書の理解のしやすさを判定する評価ツールの開発である。今回は、米国で最も多く利用されてきた、多音節の語数に着目した SMOG (Simplified Measure of Gobbledygook)の日本語文書への応用可能性を探るため、連続する漢字の数に着目し、患者向け説明文書と医療専門家向け文書の比較分析を実施した。

3. 可読性に関する先行研究と判定ツール

移民が多く英語による一般的読み書き能力の平均が低い米国では、健康関連文書の可読性は6年生レベル（日本の小学校6年生相当）が望ましいとされている¹⁾。しかし、実際には9年生や10年生など（高校生から大学生レベル）高いレベルの文書が多いという判定結果が報告されている²⁾³⁾。なお、学術的な科学論文は公式教育年数の指標で、17から20（大学院卒業レベル）である⁴⁾。

健康関連文書の可読性判定で最も多く使われている手法は、理解の難しい語として多音節の語数を数える SMOG (Simplified

Measure of Gobbledygook)である⁵⁾。その他に FOG index⁶⁾、Fry Readability Scale⁷⁾、Flesch-Kincaid Reading Grade Level⁸⁾などがあり、いずれも音節数に文の長さを加えた指標を用いている。

3. SMOG の特徴と日本語文書への応用

SMOG の最大の特徴は、以下の簡易な手順で文書の可読性 (Readability) を示す学年レベル (Grade Level) を算出することができることである。1) まず、文書の最初、中間、最後の3箇所から10ずつ計30の連続する文章を抽出する。2) 次に30の文章から3つ以上の音節を持つ多音節語の数を数える。3) 最後に多音節語の数の平方根を求め、3を加える。また、全文書をペーストするだけで値が得られる解析ツールも Web 上で提供されている⁹⁾。この方法のもととなっている公式とその根拠は次のとおりである。

$$1.0430 \sqrt{\text{number of polysyllables} \times \left(\frac{30}{\text{number of sentences}} \right)} + 3.1291$$

a) 多音節語への着目は、先行の FOG index に依拠している; b) 音節数と3以上の多音節語数は関係がないので、Flesch のようにすべての音節の数は数えない; c) 文章が長くなるほど多音節語も多くなるので、文章の長さは測らない; d) 試行錯誤の結果30の文で判定できることがわかった e) 公式そのものは、読解標準テストの各設問に正解した人の「学年の平均」を被説明変数とした回帰分析によって、最も相関係数の高い数式を選んだものである。

日本語文書にもこの簡易な SMOG を応用できないか検討した結果、多音節語のかわりに、漢字が出現するごとに連続する文字数別にセットとして数えることとした。

多音節語ではなく漢字に着目したのは、まず、日本語の表記は語が分かれていないので、一語の音節数を客観的に計測するのが困難である。また、漢字の熟語は、特に未知の専門用語などはなじみがなければ読みにくく、文書全体の読みやすさを阻む要素となると推定されるためである。「読めないと次へはすすめない」という指摘は、別の医療薬の患者向け説明文書の試作品の有用性調査¹⁰⁾でも報告されている。また、連続文字数別に漢字セットを数えたのは、漢字が多いほど読めない、あるいは理解できない可能性が大きくなると考えたためである。これについては、熟知性(なじみの程度)と相関する「主題的表記頻度」(どの表記が普通か)の分析¹¹⁾で、漢字数が多い熟語ほど「漢字表記が普通」と判断する人が少なくなるという結果とも合致する。

4. 調査方法

分析の対象とした文書は、医療薬の患者向け服薬説明書「くすりのしおり®」と、薬事法で個々の医薬品に添付が義務付けられている「医薬品添付文書」(以下、「添付文書」)の2種類で、いずれもWeb公開されている¹²⁾¹³⁾。また、いずれも製薬企業が提供する文書だが、前者は「くすりの適正使用協議会」の会員と賛同企業が任意で作成している、A4用紙1枚の基本フォーマットに従い必要最小限の情報にしぼった一般向け文書である。一方「添付文書」には薬物動態、臨床成績、薬効薬理、理化学的知見など医療専門家向けの情報が含まれる。

「くすりのしおり®」を対象としたのは、「中学生レベルを目標として書いた」と明記されていることと¹⁴⁾、一般の患者によるトライアル調査で「わかりやすさ」に一定の評価を得ている¹⁵⁾資料だからである。

対象とする医薬品を選ぶためには、効能別に分類して医療薬を掲載している「今日の治療薬2006」¹⁶⁾を用いた。68ある分類から風邪の諸症状に関連する5分類(分類

10.非ステロイド抗炎症薬、鎮痛・解熱薬、総合感冒薬;分類13.抗ヒスタミン薬;分類14.抗アレルギー薬;分類38.鎮咳薬;分類39.去痰薬)を選び、各分類から内服薬2種類ずつを乱数表の順で抽出し、計10の医薬品を選んだ。「風邪」をもととしたのは幅広い年齢層で自らが服薬する一般的な医薬品を選ぶためである。また、内服薬に限定したのは服薬方法によって異なる表記の傾向が出ることを避けるためである。

選択した10の医薬品について「くすりのしおり®」を商品名で検索し、主成分量の違う商品などは便宜的に最初のものを選択し、印刷用画面から文書を印刷した。次に同じ医薬品の「添付文書」を検索しPDFファイルを印刷した。

10種類の医薬品について2種類ずつ計20の文書の印刷物をもとにマニュアルで、文の数と、出現した漢字を「連続する漢字数」別にセットとして数えた。数え方の例は図1のとおりである。文の数は見出しもひとつとして数えた。また、連続する漢文字列が意味的に2語以上に分割できても、漢字が連続していれば1として数えた。対象とした箇所は、「くすりのしおり®」の定形文章で占められる「保管方法その他」を除く全項目と、それに相当する「添付文書」の項目とした。

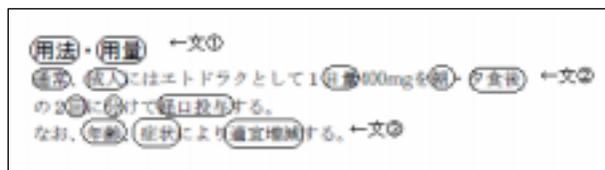


図1. 文と漢字数の数え方例

(文は3、漢字セットは単漢字3、2連続漢字7、3連続漢字1、4以上連続漢字2と数える)

5. 調査結果

表1は、2種類の文書別に、文の数と連続漢字数別の漢字セット数を数えた10の医薬品文書の集計結果である。文の数は1文書平均「くすりのしおり®」は35で、「添

付文書」の 79 の半分以下で同じ項目の内容を表現していることがわかった。また、1 文中の漢字セット数合計は 4.9 と 4.8 とほぼ同じである。そのうち「くすりのしおり®」は 1 漢字(2.0)と 2 漢字(2.1)が、「添付文書」は 2 漢字(2.3)と 4 以上連続する漢字(1.1)が平均 1 回以上出現している。

表 1 . 集計概要

| | 文の数 | 1漢字 | 2漢字 | 3漢字 | 4漢字 | 漢字セット合計 |
|---------|-----|-----|-------|-----|-----|---------|
| 添付文書 | 789 | 564 | 1,795 | 533 | 835 | 3,727 |
| 1文書平均 | 79 | 56 | 180 | 53 | 84 | 373 |
| 1文中平均 | - | 0.7 | 2.3 | 0.7 | 1.1 | 4.8 |
| くすりのしおり | 351 | 699 | 742 | 204 | 78 | 1,723 |
| 1文書平均 | 35 | 70 | 74 | 20 | 8 | 172 |
| 1文中平均 | - | 2.0 | 2.1 | 0.6 | 0.2 | 4.9 |

連続漢字数別の傾向の違いをより明確に示したのが、漢字セット中の連続漢字数別割合(図 2)である。「くすりのしおり®」は全漢字セット中 2 文字連続と単漢字(43.1%; 40.6%)が多く、「添付文書」では 2 文字について 4 文字以上(48.2%; 22.4%)が多い。



図 2 . 漢字セットの連続数別の割合

「添付文書」に特徴的な 4 文字以上連続の漢字セットは最も可読性を阻むと推定される。そこで「くすりのしおり®」に依然として使われている 4 文字以上の漢字文字列を個々に見ることとした。のべ 79 個 56 種類のうち 2 回以上出現した 14 種類を表 2 に示した。この例にも見られるとおり医学特有の疾病、症状、服薬に関する語が多い。4 文字以上連続する漢字セットの全体 56 種類のうち 51 種類がこれらの語であった。

最後に SMOG に模して可読性の学年レベルを表す数値を得ることができないか試

みた。2 種類の文書を同じ文の数に調整して、可読性を阻むと推定される 4 文字以上連続の漢字セット数の平方根をとってみたのが表 3 である。200 から 400 の文において「くすりのしおり®」では中学生の学年レベルに相当する 7~9、「添付文書」では既存研究で示された科学雑誌論文のレベル相当の 15~21 の数値が導き出されている。

表 2 . 「くすりのしおり®」4 以上漢字*

| 分類 | 4以上漢字 | 出現回数 |
|--------|-------|------|
| 初期症状 | | 6 |
| 適宜増減 | | 5 |
| 食欲不振 | | 3 |
| 急性上気道炎 | | 3 |
| 呼吸器疾患 | | 2 |
| 変形性関節症 | | 2 |
| 医学 | | |
| 副鼻腔炎 | | 2 |
| 気管支喘息 | | 2 |
| 分娩後疼痛 | | 2 |
| 貧血症状 | | 2 |
| 前立腺肥大 | | 2 |
| 低出生体重児 | | 2 |
| 服用方法 | | 2 |
| 一般 | 効果以外 | 2 |

* 2 回以上出現したもの

表 3 . 「4 以上漢字の数」と平方根

| 文の数 | 添付文書 | | 「くすりのしおり®」 | |
|-----|-------|-----|------------|-----|
| | 4漢字の数 | 平方根 | 4漢字の数 | 平方根 |
| 200 | 212 | 15 | 44 | 7 |
| 300 | 317 | 18 | 67 | 8 |
| 400 | 423 | 21 | 89 | 9 |

6 . 考察

「くすりのしおり®」は、項目数を減らすだけでなく、同じ項目でも情報量を減らし、全体として分量を少なくしている。これは「くすりのしおり®」が「最小限の情報」提供を方針としているため、同じ項目でも、専門家向けの研究データなどを省略していることによる。

また、「くすりのしおり®」は読みやすくなった結果、4 以上連続する漢字のセットが減っている。残された 4 文字以上の漢字セットには医学用語が多く、読みにくく理解しにくいことが推定され、可読性判定の目安として利用できると思われる。

可読性判定として、200 から 400 の文中の 4 漢字以上の数の平方根から「くすりのしおり®」と「添付文書」のそれぞれ推定される可読性に相当する学年レベルを算出することができた。しかし、本調査はサンプルが小さく、また「くすりのしおり®」と「添付文書」の学年レベルが厳密に設定されたものではない。また、SMOG のように人に対する標準テストによる検証はしていないので、そのまま公式として提案することはできない。

7. 結論

日本語文書の可読性判定について SMOG の応用可能性を探るため「くすりのしおり®」と「医薬品添付文書」の漢字の連続数別セット数を数え、比較分析した。その結果、連続する 4 文字以上の漢字セット、結果として医学用語を減らすなどの手法で「くすりのしおり®」が読みやすさを確保していることがわかった。また、可読性の判定には連続する 4 文字以上の漢字数を用いる可能性が認められた。

今後は、漢字セットの特徴、形態素分析ツールなどを用いた大きなサンプルでの分析に加え、漢字用語や医学用語に関する言語学研究による考察、さらには人に対する検証などを通じて、妥当性のある可読性判定ツールの検討を進めたい。

本研究は平成 18 年度厚生労働科学研究費補助金「患者 / 家族のための良質な保健医療情報の評価・統合・提供方法に関する調査研究」の補助を得ている。

引用・参考文献

- 1) Roter, D. L.; Rudd, R. E., Comings, J. Patient literacy. A barrier to quality of care. *Journal of General Internal Medicine*. Vol.13, No.12, 1998,p. 850-851
- 2) Estey, A.; Musseau, A., Keehn, L. "Patient's understanding of health information: a multihospital comparison." *Patient Education and*

Counseling. Vol.24, No.1, 1994,p. 73-78.

- 3) Graber, M. A.; Roller, C. M., Kaeble, B. "Readability levels of patient education material on the World Wide Web." *Journal of Family Practice*. Vol.48, No.1, 1999,p. 58-61.

- 4) Meadows, A. J. "Quantitative study of factors affecting the selection and presentation of scientific material to the general public". *Scientometrics*. Vol.20, No.1, 1991,p. 113-120

- 5) McLaughlin, G.H. "SMOG grading: a new readability formula." *Journal of Reading*. Vol.12, No.8, 1969, p.639-646.

- 6) Gunning, R. *The Technique of Clear Writing*. New York, McGraw-Hill, 1952, p.329.

- 7) Fry, E. "A readability formula that saves time". *Journal of Reading*. Vol.11,1968, p.513-516,575-578.

- 8) Kincaid, J.P.; Fishburne, R.P.; Rogers, R.L.; Chissom, B.S. *Derivation of New Readability Formula for Navy Enlisted Personnel*. Millington, TN: Navy Research ranch, 1975, 40p.

- 9) SMOG Reading Level Calculator

http://www.linda-andrews.com/readability_tool.htm [cited 2006-10-18]

- 10) 久保鈴子.厚生労働科学研究研究費補助金 医薬品・医療機器等レギュラトリーサイエンス総合研究事業 患者及び国民に理解される副作用等医薬品情報内容の構築と医薬品適正使用への患者参加推進に関する研究 平成 16 年度総括・分担研究報告書. 2005. 160p.

- 11) 浮田潤ほか.日本語の表記形態に関する心理学的研究.日本心理学会, 1996,114p.

- 12) くすりのしおり® <http://www.rad-ar.or.jp/siori/> [cited 2006-10-18]

- 13) 医療用医薬品の添付文書情報

http://www.info.pmda.go.jp/psearch/html/menu_tenpu_base.html [cited 2006-10-18]

- 14) 海老原格; 藤原充雄, 和田有一. "「くすりのしおり」を良く理解してもらうために". *臨床医薬*. Vol.16, No.1, 2000,p. 27-38

- 15) 日本 RAD-AR 協議会. 「対話のある医療」をめざして. 東京, 日本 RAD-AR 協議会, 1997.68p.

- 16) 今日の治療薬 2006(28 版). 東京, 南江堂, 1,271p.