

なぜ「女医」はあるのに「*男医」はないのか
—情報理論的分析—

藪下 克彦

1. はじめに

「女医」という言葉（単語）はあるが、「*男医」¹というのはない。これはなぜであろうか。この問いが本論文のテーマである。「女医」に関しては、しばしば、（性）差別語である、という主張を聞く。そのように主張する人たち（例えば、上野ほか（1996）、田中・諸橋（1996）、中村（1995））は、「女医」という表現は、次のような男性中心的な考え方を反映しているとする。医師は伝統的に男性が就くのが「標準」であって、それから外れた女性医師には女性であることを強調するための「有標」(marked)な表現である「女医」をあて、「標準」である男性医師には有標な表現を与える必要がない。これが「女医」は存在するが「*男医」は存在しない理由である。この説明には問題がある。一つ目は、「女医」が「差別語」であることを前提としている。さらに、後で見るが、「A-」と「B-」という反意接頭辞と語基「X」がある時、「AX」という単語は実在するが「*BX」という単語は実在しないというのは、「女医」と「*男医」のペアに限らず、他にも多くの例がある。例えば、「短パン」と「*長パン」が例の1つである。この場合は、もちろん（男女）差別は関係ない。

本論文は、まず、「女医」が、コーパスなどの言語使用例を見る限り、差別性をもって使用されているケースはないが、逆に中立的、好意的に使われているケースは見つけることができることから、「女医」は「差別語」でないことを立証する。その後、「AX/*BX」現象に対して、その他の例にも当てはまる確率に基づく情報理論的分析を提案する。

2. 「女医」は差別語か

まず、「「女医」は差別語か」という命題、厳密には、疑問に答えるためには、「差別語」とは何かに関する共通理解を明らかにしておく必要がある。例えば、佐藤（2005: 206）は、差別語の判断基準として次の3点を挙げている：①被差別者を傷つける表現であること、②悪意に基づく表

現であること、③偏見を助長する表現であること。小林(2016:20)は「...、他者の人格を個人的にも集団的にも傷つけ、蔑み社会的に排除し、侮蔑・採殺する暴力性をもつ言葉のことです。しかも、もっぱら自己選択できない自然的・社会的属性を差別の対象とされた人や集団を卑しめていう蔑称語です」と定義している、さらに、差別語には必ず差別の実態を伴うと付け加えている。言い換えれば、差別の実態が伴わなければ、差別語ではないということになる。

本論文では、ある言葉・表現が「差別語」であることの必要条件として、その言葉・表現の使用の前提に、ある個人または集団をある属性に基づき差別する差別の実態があることとする。

それでは、早速「女医」という表現が差別語の必要条件を満たすかどうか検討する。「女医」に差別の実態が伴っているであろうか。医師である、さらに、女性であるという属性に基づいて、その属性をもつ個人または集団、つまり女性医師が誰かまたはいかなる集団によって差別されている実態があるであろうか。少なくとも私には思いつかない。「女医」が「差別語」であるのではないかと考えていたある人が挙げた根拠は、ある病院で遭遇した次のような光景だそうである。その人によると、ある(男性)老人がある女性医師に関して、「女医は、やっぱり...」という非難がましい発言をしていたというものである。これは、差別というより、「学者は、頭でっかちで、...」、「公務員は、...」に見られるような、「学者」や「公務員」と同様な使われ方で「憎まれ口」、「悪態」の例であると考えられる。天皇でさえも憎まれ口の対象になるので、上記のような根拠で「女医」を「差別語」と認定すると、「学者」や「公務員」を含め人や集団を指す言葉がすべて「差別語」となってしまうので受け入れがたい。

「女医」を差別の対象として誰かを蔑めるために使われている例を見つけるのは難しいのに対し、中立的、または、好意的に使われている例を見つけるのは易しい。中立的な例として、医療機関を探すためのウェブサイト【病院ナビ】(<https://byoinnavi.jp>)で「女医」が検索語のボタンの1つとして設定してあることを挙げるができる。このような不特定多数の人が利用するウェブサイトの管理者が差別語をそれと認識して使うことは考えにくく、差別性があるとは考えていないことは明らかである。女性医師に好意を持って「女医」が使われている例-特に、「女医」に「-さん」を付けて言う例-を見つけるのは非常に容易である。例えば、(鳴門市民劇場機関誌 2016: 6)のインタビュー記事における女優の田岡美也子さんの次のような発言を見ていただきたい。

Q:「趣味があれば聞かせて下さい」

A(田岡):「3年くらい前から俳句をやりだしたの。句会がすごく楽しくて、色んな職種の人と知り合っただけね。女の人ばかりの句会にも入っているの。皆すごいユニークで、女医さんがいたり編集者がいたりして面白いの。…」(傍点は筆者)

上の記事での「女医」の使用に差別意識がないのは明らかであろう。さらには、「-さん」という接尾辞とともに用いられていることからすると、敬意とまではいなくても親しみをもって、その個人の女性医師なり女性医師一般を捉えているということができる。

以上、「女医」には、女性医師に対する差別を意図した使用例を見つけることは難しいことを述べたが、私個人の主観や偏った言語経験に起因している可能性があるため、実際に使用された言語データを集めたデータベースであるコーパスを参照することにした。参照したコーパスは、大学共同利用機関法人人間文化研究機構国立国語研究所と文部科学省科学研究費特定領域研究「日本語コーパス」プロジェクトが共同で開発した『現代日本語書き言葉均衡コーパス』(BCCWJ: Balanced Corpus of Contemporary Written Japanese)の簡易検索版、通称『少納言』(<http://www.kotonoha.gr.jp/shonagon/>)である。「女医」を入力して検索した結果、196件の用例が出力された(検索日:2019年10月19日)。そこには、読者自身で検証していただくには無いが、女医に対する差別的意図を伴ったただ1つの用例も見つけることができない。従って、「女医」が差別語ではないことが、コーパスによっても裏付けられる、少なくとも、差別的に使用されている例を見つけることができないということができる。²

コーパスの用例を調べていて、面白い発見があった。検索結果の1つとして出てきた次の用例を見ていただきたい。

「...大阪府女医会 副会長 長野 禮子・お問い合わせ 大阪府女医会「市民公開講座事務局」...」(2008年 Yahoo! ブログ) [検索結果表示番号:138]

お気づきいただけたであろうか。「女医会」とある。「公益社団法人 日本女医会」(<http://www.jmwa.or.jp>)という組織があり、大阪府女医会は地方支部の1つである。ここで注目していただきたいのは、女性医師が会員である組織を自ら「女医会」と称していることである。もし「女医」が差別語であるならば、被差別者である女性医師が差別語である「女医」と自

ら称するであろうか。具体的には、どの明らかな差別語でも良いので、その差別語で自らのことを称している被差別者を想像していただきたい。想像できないであろう。その意味では、「女医」と自ら称することを女性医師自身がよしとすることは、「女医」が差別語でないことを如実に表している。

以上、「女医」が差別語でないことを、日常でもコーパスでも「女医」が差別的に使われている例を見つけることができないこと、さらに、女性医師自身が自らを「女医」と称することをよしとしている事実を根拠に、立証した（と思っている）。以下では、本論文の本題である「女医」という単語はあるのに、なぜ「*男医」という単語はないのか」という疑問に差別語とは別の角度からアプローチする。

3. 「女医」型単語の言語学的特徴

この節では、「女医」という表現の言語学的特徴を検討する。

3.1. 派生語としての「女医」

「女医」は、形態的には、接頭辞「女-」と語基「医」（医者の意味）からなっている派生語である。それに対し、同じ意味を表す「女性医師」は、「女性」と「医師」の二つの語根からなっている複合語である。

3.2. 「女医」はあるのに「*男医」はない

上記で「女医」が派生語であることを確認したが、「男性医師」に対応する「*男医」という派生語はない。この特徴が、「女」と「男」の違いに対応しているので、第1節で述べたように、男女差別に起因しているという単純な憶測が生まれるのも無理はない。しかし、第2節で考察したように、「女医」という表現の使用に女性差別が伴っているという事実が見当たらないことを確認した。従って、「女医」という派生語はあるが、「*男医」という派生語はないという事実の説明を男女差別以外に求めなければならないが、その手がかりとして、「女医」と「*男医」のようなペアとして他にどのようなものがあるのか見ることにする。筆者は以下のような例に思い当たった。読者の皆様も例を考えていただきたい。

- (1) 「男娼」³-「*女娼」, 「女王」-「*男王」, 「外車」-「*内車」,
「外人」-「*内人」, 「短パン」-「*長パン」, 「冷や汁」-「*熱汁」,
「軽自動車」-「*重自動車」, 「朝風呂（朝湯）」-「*夜風呂（夜湯）」

従って、「女医」と「*男医」のような例は、次のような現象の一例であると見ることができる。

- (2) 反意の対である接頭辞「A-」と「B-」、そして語基 X が存在する時、「AX」という派生語は実在するが、「BX」という派生語は実在しない場合がある。

(2)に適合する「AX」と「*BX」のペアに共通する特徴は何であろうか。試行的近似として直感的な言葉で言うと、「*BX」で意図される人、動物、モノ、出来事など、つまり指示対象は、「AX」で表されるそれに較べて「普通」である。逆に言えば、「AX」の指示対象は「*BX」のそれにくらべて「普通ではない」と特徴付けることができると思われる。例えば、夜、風呂に入るとは「普通」だが、朝、風呂に入るのは（前者に較べて）「普通ではない」と一般に理解されているのではないだろうか。しかし、「普通（ではない）」という表現は価値・規範的判断を想起させる言葉であるので、人によって「AX」の指示対象が「普通ではない」、そして「*BX」のそれは「普通」であるという判断が一致しない恐れがある。判断が一致しないばかりか、例えば、「男娼」の指示対象は「普通ではない」のに対し、「*女娼」のそれは「普通」である」というような命題には、人によっては生理的・道徳的嫌悪感を抱いてしまい判断をするとかしないとか以前の問題にもなりうる。従って、(1)に見られるような「AX」と「*BX」のペアの総てに共通する特徴を確率の概念を使って以下のように一般化する。

- (3) 「AX」と「*BX」に関する確率的特質：

反意の対である接頭辞「A-」と「B-」、そして語基「X」が存在する時、「AX」という派生語は実在するが、「*BX」という派生語は実在しない場合、「X」で表される個人、個体、事象などの指示物が A である確率のほうが、X が B である確率よりも小さい。ここで、X が A である確率を $P(A|X)$ 、X が B である確率を $P(B|X)$ と表記する。従って、上で述べたことは、 $P(A|X) < P(B|X)$ と表すことができる。

また、「女医」はあるが「*男医」はない」という事実は、次のような原理を示唆している。

(4) 語彙の経済性：

反意の対である接頭辞「A-」と「B-」,そして語基「X」が存在する時、「AX」と「BX」の両方を語彙項目として使用するのではなく、何らかの有用性に基づき、どちらか1つを単語として採用する。そうすることにより、語彙サイズのむやみな増大を抑制することができる。

次節において、「AX」と「BX」が(3)の特徴を満たす時、「AX」の方が、「BX」よりも、情報の増加という意味においてより有用であることを情報理論による分析で明らかにする。

4. 情報理論的分析

「女医」と「*男医」の様に、AとBが反意対である時、「AX」という派生語は存在するが「*BX」という派生語は存在しない場合がある。その場合の特徴として、前節の(3)にまとめたように、一般的に、あるXがAかBかが問題になっている状態でXがAである確率、 $P(A|X)$ のほうが、XがBである確率、 $P(B|X)$ よりも小さいという特徴があることを見た。この確率的特徴を聞いただけで、「AX」と「*BX」の対比の「説明」になっていると思う人はほとんどいないと思われるので、本論文では、確率を基礎概念に据え、そこから「情報(量)」を数学的に定量化する学問である Shannon (1948) によって創始された情報理論の道具立てを使い、さらなる分析を試みる。

4.1. 情報理論入門

本節では、情報理論に馴染みのない読者のために、情報理論の基本的な考え方と道具を紹介する。⁴

4.1.1 情報量の定義

一般的に、ある事柄について何かを知ると情報を得る、情報を得ることによって情報が増える。その時に、どんな内容かによって、たくさん知ったという気になる場合とそうでもないという場合がある。このことは、情報に量の多寡があることを示している。例で考えてみよう。ジョーカーを抜いたトランプのデッキから一枚の札が引かれて、その札を引いた人は、その札が何か知っているが、あなたはその札が何か知りたしとする。札を引いた人から次の2つのヒントのどちらかをもらえらしたらどちらを選ぶであろうか：(i) その札のスーツ、つまり、スペード、クラブ、ハート、ダイヤモンド

のいずれか、または、(ii) その札の番号、つまり、1 から 13 までのいずれかの番号。(ii)のほうのヒントを選ぶのではないだろうか。(i)を聞いた場合、52 枚の中 12 枚に絞られるのに対し、(ii)の場合、52 枚の中 4 枚に (まで) 絞られるのであるから。このことから、(ii)の情報のほうが(i)のそれに較べて量が多いとすることができる。別の見方をすると、情報の量は、不確実性を減らす量と捉えることもできる。⁵

上の状況、つまり引かれた札がなんであるのかが問題になっている状況で、なぜ(ii)の情報のほうが(i)のそれより多いかに関する定式化に際し情報理論が着目するのは確率である。つまり、(ii)は引かれた札の番号を伝え、その事象の確率は 1/13 であるのに対し、(i)は引かれた札のスーツを伝え、その事象の確率は 1/4 である。言い換えれば、(ii)は 1/13 の確率で起こる事象が起こったことを伝える情報で、(i)は 1/4 で起こる事象が起こったことを伝える情報である。一般に生起する確率が低ければ低いほど、その事象が起こったという情報の量は高くなり、「驚き」「意外さ」は大きくなる。従って、情報量の定義として次のものが広く使われている。

(5) 定義：(自己) 情報量 (選択情報量, 自己エントロピー)

事象 E が起こる確率を $P(E)$ とするとき、事象 E が起こったことを知らされたとき受け取る情報量、 $I(E)$ を

$$I(E) = -\log P(E) = \log \frac{1}{P(E)}$$

と定義する。

なお、対数 ($\log(\text{arithm})$) の底 (base) としてなにを選んでも以下の論旨に影響はないが、一般的に 2 を使うことが多いので、本論文でも特に断りがない限り底を 2 とする。さらに底を 2 とした時の情報量の単位をビット (bit) と呼ぶ。

(5) の定義を上記のトランプの場合に当てはめてみる。52 枚の札の中から一枚引かれ、その札が何であるか聞いた時の情報量は、それぞれのカードが引かれる確率は 1/52 であるので、(5) によると $\log_2 52 (\approx 5.70)$ になる。(i) の情報、つまり引かれた札のスーツが何であるかを聞いた時の情報量は、それぞれのスーツである確率が 1/4 であるので、 $\log_2 4 (= 2)$ であり、(ii) の情報、つまり引かれた札の番号を聞いた時の情報量は、それぞれの番号である確率が 1/13 なので、 $\log_2 13 (\approx 3.70)$ になる。この結果は、上で確認した(ii)の情報のほうが(i)のそれよりも多いという直感的判断と一致する。さらに、引かれた札が何であるかという情報は、(i) と(ii) を合わせた情報に一致すると考えられる。なぜ

なら、札のスーツと番号が分かれば、その札を特定できるからである。ここで注目して頂きたいのが、引かれた札が何であるかの情報量、 $\log_2 52$ と(i)の情報量、 $\log_2 4$ と(ii)の情報量、 $\log_2 13$ の間に次の関係があることである。 $\log_2 52 = \log_2 4 + \log_2 13$ 。これは、対数の次の一般的性質による。

$$(6) \quad -\log p \cdot q = (-\log p) + (-\log q) \quad (\log \frac{1}{p \cdot q} = \log \frac{1}{p} + \log \frac{1}{q})$$

(6)は次のことに対応している。ある事象が2つのお互いに独立した事象の組み合わせと見なせる場合（例えば、上のトランプの場合、引かれた札が何であるかという事象は、その札のスーツが何かという事象と番号が何かという事象の組み合わせと見ることができる。）、2つの事象の起きる確率を p と q とすると、元の事象の起きる確率は、 $p \cdot q$ となる。(6)は、元の事象の情報量は、その事象を構成している2つの事象の情報量の和に等しいと言っていると解釈することができる。これは元の事象の情報量は構成する事象の情報量の総和に一致するという直感、「情報量の加法性」にうまく対応している。このことは、事象の起きる情報量を対数を使って(5)のように定義することの妥当性の証左とみることができる。

それでは、「女医」に関わる次の事象の情報量を考えてみたい。ある人が医師であること（ただそれだけ）が分かった時、さらに、その人が女性（医師）であると分かった時の情報量と男性（医師）であると分かった時の情報量を考える。医師が女性である確率を $P(W|D)$ 、医師が男性である確率を $P(M|D)$ と表すと、医師が女性であることを聞いた時の情報量と男性の場合のそれは、それぞれ、 $I(P(W|D))$ と $I(P(M|D))$ と表すことができる。⁶ ここで、注意しておきたいことは、 $P(W|D)$ と $P(M|D)$ は、それぞれ、(3)の $P(A|X)$ と $P(B|X)$ に相当し、 $P(W|D) < P(M|D)$ という関係が成り立つ。少なくとも現在までの日本において、一般的に言って、医師が女性である確率のほうが男性であるそれよりも小さいということができる。⁷ このことは、情報量という意味では、 $I(P(W|D)) > I(P(M|D))$ 、つまり、医師が女性であると聞いた時の情報量のほうが、男性であると聞いた時のそれより大きいことを意味している。⁸

4.1.2 エントロピー（平均情報量、シャノン情報量）

前小節では、ある事象が起こったと聞いた時の情報量の定義とその具体的例を見たが、今度は、いくつかの事象が起こりうるがどの事象が起きたのかわからない状態での情報量を考える。上の医師の性別の例で言うと、その医師が女性である事象と男性である事象の2つの中、どちらなのかわからない状

態の情報量である。一般的に言って、起こる可能性のある事象が A_1, A_2, \dots, A_n あり、それぞれが $P(A_1), P(A_2), \dots, P(A_n)$ ($\sum_{i=1}^n P(A_i) = 1$) の確率で起きるとすると、その中の 1 つの事象が成り立っているがどの事象が成り立っているのかわからない状態の情報量は、それぞれの事象の情報量の期待値の和だと考えることができる。この情報量は「エントロピー（平均情報量，シャノン情報量）」と呼ばれ、以下のように定義される。

(7) 定義：(情報) エントロピー（平均情報量，シャノン情報量）

事象の集まり A_1, A_2, \dots, A_n があり、それぞれの生起確立、 $P(A_1), P(A_2), \dots, P(A_n)$ が $\sum_{i=1}^n P(A_i) = 1$ を満たす時、確率分布 P のエントロピー $H(P)$ は、次のように定義される。

$$H(P) = -\sum_{i=1}^n P(A_i) \log P(A_i) = \sum_{i=1}^n P(A_i) \log \frac{1}{P(A_i)}$$

エントロピーは、定義からも分かるように、各事象の情報量の期待値の和（事象の情報量をそれぞれの確率で平均した値の和）である。ある事柄（試行）に関して、起こりうる事象が複数ある場合、エントロピーは、その中のどの事象が起こったのか聞く時、どのくらいの情報量を得ることが期待できるのかを表す。逆に言うと、どの事象が起こったのか明らかにするのに必要になる平均的信息量を表している。例を使ってエントロピーの意義を確認する。さいころを 1 回振って出た目を知りたい 2 つの状況を考える。1 つ目の状況 A は、どの目が出たか全く分からない状況で、2 つ目の状況は出た目が偶数であることが分かっている状況である。A の状況における事象は、1 から 6 までの目のそれぞれが出た A_1, A_2, \dots, A_6 の 6 つの事象であるのに対し、B の状況における事象は、2, 4, 6 の目が出たそれぞれの事象 B_1, B_2, B_3 の 3 つの事象である。A と B の状況における確率分布は次の通りである。 $P(A_i) = 1/6$ ($1 \leq i \leq 6$)、 $Q(B_i) = 1/3$ ($1 \leq i \leq 3$)。この 2 つの状況のエントロピーは、それぞれ、 $H(P) = -\sum_{i=1}^6 P(A_i) \log P(A_i) = 6 \cdot 1/6 \cdot \log 6 \approx 2.58$ 、 $H(Q) = -\sum_{i=1}^3 Q(B_i) \log Q(B_i) = 3 \cdot 1/3 \cdot \log 3 \approx 1.58$ である。直感的に、出た目を聞いて得る情報量は、A の状況の方が B の状況でよりも多い、また、出た目の不確かさも、A の状況の方が B の状況よりも大きい。エントロピーは、その直感的理解にもうまく対応している。我々の関心のある状況、つまり、ある人が医者である時、その人が女性であるか男性であるかを知りたい時のエントロピーを考えてみたい。起こりうる事象は、女性である事象 W と男性である事象 M である 2 つの事象で、その生起確率は、それぞれ $P(W)$ と $P(M)$ ($P(W) + P(M) = 1$) とする。一般的に、

起こりうる事象が2つの場合、2つの生起確立が等しい時、つまり、1/2の時、エントロピーが最大（対数の底が2の時、1ビット）になる。今のケースは、上で確認したように、（過去から）現在の女性医師と男性医師の数の割合を考えると、 $P(W) < P(M)$ ということができる。因みに、仮に $P(W) = 1/4$ と $P(M) = 3/4$ とすると、この時のエントロピー $H(P) = \frac{1}{4} \cdot \log 4 + \frac{3}{4} \cdot \log \frac{4}{3} \approx 0.81$ になる。

4.1.3 カルバック・ライブラー情報量（相対エントロピー、情報利得）(Kullback and Leibler, 1951)

前節で複数のお互いに独立した事象がそれぞれある確率で起きると想定されている状況の期待情報量、不確定度を測る尺度としてエントロピーがあることを見た。本節では、同じ事象に関して、それぞれの事象の確率が更新されると情報量に変化するが、その変化の度合いの尺度、言い換えれば、確率の更新がもたらす情報量の貢献度を表す指標を導入する。

(8) 定義：カルバック・ライブラー情報量（相対エントロピー、情報利得）

P と Q を同じ確率空間における（離散）確率分布とする。 P の Q に対する（ Q が P に変更されることによる）カルバック・ライブラー情報量は以下のように定義される。

$$D_{KL}(P||Q) = \sum_{i=1}^n P(A_i) \log \frac{P(A_i)}{Q(A_i)}$$

例えば、あるコインを投げて表が出る事象 H(ead)と裏が出る事象 T(ail)の確率が同じ、つまり 1/2 である（と思っていた）確率分布を Q とする。そして、（実は）HとTの確率がそれぞれ 1/4 と 3/4 であると分かった時の確率分布を P とする。この場合のカルバック・ライブラー情報量は、 $D_{KL}(P||Q) \approx 0.19$ となる。

4.2. 「女医」の情報理論的的意義

「女医」と「男医」のように、「A」と「B」が反意関係にあり、「AX」という単語は存在するが「*BX」という単語は存在しない場合、Xの指示物がAである確率がBであるそれよりも小さいと一般的に言えるという観察から、上記の言語学的特質は確率に関係しているという仮説に基づき、今まで、基盤を確立（論）に置く情報理論の基礎概念を紹介してきた。ここでは、「女医」

と「*男医」との間でどのような情報理論的な違いがあるのか、カルバック・ライブラー情報量を使って見ていく。

「女医」と「*男医」は、ある医師が女性か男性か問題になっている状況で、その医師がそれぞれ女性、または男性であると確定する (disambiguate) 働きがある。ある医師が女性であるか男性であるかわからない状況は、次の確率分布で表すことができる。その医師が女性である事象 W と男性である事象 M があり、その発現確率は、それぞれ $Q(W)$ と $Q(M)$ で、さらに $Q(W) + Q(M) = 1$, $Q(W) < Q(M)$ という関係が成り立つ。この状況で、(この場合、「女医」という表現によって) その医師が女性であると確定することは、確率分布が Q から $P: P(W) = 1, P(M) = 0$ に変化することを意味している。(「女医」の使用により) その医師が女性であると確定することにより得られる情報量の増加は、 $D_{KL}(P||Q) = P(W) \log \frac{P(W)}{Q(W)} + P(M) \log \frac{P(M)}{Q(M)} = \log \frac{P(W)}{Q(W)} + P(M) \log \frac{P(M)}{Q(M)} = \log \frac{P(W)}{Q(W)} [P(W) = 1, P(M) = 0 \text{ のため}] = \log P(W) - \log Q(W) = \log 1 - \log Q(W) = 0 - \log Q(W) = -\log Q(W)$ 。従って、ある医師の性別が女性か男性かはっきりしない (かつ、女性である確率、 $Q(W)$ が男性である確率、 $Q(M)$ より低い) 状況において、「女医」の使用により、医師の性別が女性に確定された時の情報量の増加は、カルバック・ライブラー情報量を使うと $-\log Q(W)$ と表すことができる。ここで、お気づきになった読者もいらっしゃると思うが、医師の性別が確定していない状況から女性であると確定した状況のカルバック・ライブラー情報量は、以前の状況において、医師が女性であると聞いた時の (自己) 情報量と同じである。例えば、「女医」の使用以前における医師が女性である確率を $1/4$ で、男性である確率が $3/4$ の場合、「女医」の使用により増える (カルバック・ライブラー) 情報量は、 $-\log 1/4 = 2$ になる。一般的に次の命題が成り立つ。

- (9) 命題: 相反する事象 A と B があり、その確率分布 Q は次のようなものであるとする: $Q(A) + Q(B) = 1, Q(A) < Q(B)$ 。確率分布が Q から次の P に更新されたとする: $P(A) = 1, P(B) = 0$ 。その時のカルバック・ライブラー情報量、 $D_{KL}(P||Q)$ は、確率分布 Q における A の (自己) 情報量と一致する。つまり、 $D_{KL}(P||Q) = -\log Q(A)$ である。⁹

証明: (8) のカルバック・ライブラー情報量の定義によると、

$$D_{KL}(P||Q) = P(A) \log \frac{P(A)}{Q(A)} + P(B) \log \frac{P(B)}{Q(B)} = P(A) \log \frac{P(A)}{Q(A)} [P(B) = 0 \text{ のため。}] = \log \frac{1}{Q(A)} [P(A) = 1 \text{ のため。}] = \log 1 - \log Q(A) = 0 - \log Q(A) = -\log Q(A)。$$

QED

命題(9)から次の命題が導き出される。

- (10) 命題：(9)の状況における確率分布 Q が P ではなく、次のような R に更新された場合を考える。 $R(A) = 0, R(B) = 1$ 。その時の情報増加量、 $D_{KL}(R||Q)$ は、 $D_{KL}(P||Q)$ よりも少ない、つまり $D_{KL}(P||Q) > D_{KL}(R||Q)$ 。

証明：(9)により $D_{KL}(P||Q) = -\log Q(A)$ 。また、(9)と同様に、 $D_{KL}(R||Q) = -\log Q(B)$ 。 $Q(A) < Q(B)$ 。従って、 $D_{KL}(P||Q) > D_{KL}(R||Q)$ 。QED

(10)の命題は、2つの相反する事象 A と B があり、さらに、 A が起きる（と想定される）確率の方が、 B が起きるそれよりも小さい時、事象 A が起きたと聞いた（ A の確率が1になった）時の方が、事象 B が起きたと聞いた（ B の確率が1になった）時よりも、情報の増大が大きいことを表している。このことは、「女医」と「*男医」に関して次のことを意味している。ある医師が女性と男性のどちらか確定していない（しかし、女性である確率が男性である確率よりも低い）状況で、「女医」の使用によって女性であると確定する場合の方が、「*男医」の使用により男性であると確定する場合よりも、情報量の増大が大きいことを表している。言い換えれば、「女医」のほうが「*男医」よりも情報の増加という意味では、より有用であると言える。「女医」と「*男医」だけでなく、「 AX 」と「 $*BX$ 」のパターンのペア全般に次のような一般化が成り立つ。

- (11) ある事象 X が相反する属性 A と B によって相反する2つの事象、 AX と BX に二分割され、 A が起きる確率の方が B のそれよりも低い時、単語「 AX 」によって、 X が A であると確定する方が、単語「 $*BX$ 」によって、 X が B であると確定するよりも、情報量の増大が大きい。従って、単語「 AX 」のほうが単語「 $*BX$ 」よりも情報量（の増大）という意味において有用であると考えられる。

(11)に示された情報理論的事実と(4)の「語彙の経済性」から、「女医」と「*男医」を一例とする「 AX 」という単語は存在するが「 $*BX$ 」は存在しないという現象を導き出すことができる。

5. 結論と今後の課題

日本語の単語に「女医」はあるが「*男医」はないという現象を受け、先ず、表面的には、「男女」（「女男」）に関係があるので、その現象を男女差別に帰する見方、具体的には、「女医」が差別語であるからとする考え方を支持する実態がないことを、コーパスなどの実際の使用例における「女医」の使われ方を検討することによって確認した。その中でも、注目すべきは、「女医会」という組織が全国都道府県に存在していることである。どこに自分たちの被差別者としての蔑称を自分たちの組織の名称とする人たちがいるであろうか。

本論文では、「女医」と「*男医」を「AX」と「*BX」の一般的なパターンの単なる一例に過ぎないと位置づけ、先ず、このパターンのペアに共通して、「AX」の指示対象の出現率、または、遭遇率が「*BX」のそれより低いという特徴があることを確認した。さらに、「単語の数をむやみに増やさない」という「語彙の経済性」を想定し、その原理から AX と *BX の中どちらを取るかという視点で 2 つを見た時、「AX」のほうが「*BX」よりも情報量の増大という意味では、より有用であることが情報理論的分析によって明らかになった。今後の課題としては、「語彙の経済性」仮説の妥当性の検証、さらには、「語彙の経済性」仮説を基にしない「AX/*BX」現象の分析の可能性を探る必要がある。

注

¹ 「*BX」のアスタリスクは、「BX」が実在しない単語であることを表している。

² 『少納言』は、各種出版データや東京都下の公共図書館の蔵書等にみられる 11 種データ、具体的には、書籍、雑誌、新聞、白書、教科書、広報誌、Yahoo!知恵袋、Yahoo!ブログ、韻文、法律、国会会議録を母集団として、そこから無作為に用例を抽出した「均衡」コーパスであるので、そこに「女医」が差別的に使われている用例がないということは、一般にそういう用例は無いと結論づけてもよいと思われる。

³ 「女医」を女性のことを差別する（性）差別語であると主張する人たちは、「男娼」を（性）差別語だと判断するのであろうか。また、そう判断する場合、どのような論理によって、男性と女性にどちらに対する（性）差別語とするのであろうか。興味があるところである。

⁴ より厳密な情報理論の入門には、甘利 (2011)や Cover and Thomas (2006)を参照のこと。

- ⁵ (ii) を聞いた後は、引かれた札の候補の数が 52 から 4 に減るのに対し、(i) の場合は 52 から 13 にしか減らない。「不確定性/不確実性」という意味では、情報量の多寡は不確定性減少の貢献度の高低に相当する。
- ⁶ $P(X|Y)$ のような表記の確率は Y が成り立つ時の X の確率を表し、条件確率 (conditional probability) と呼ばれる。また、 W, M, D はそれぞれ、woman, man, doctor を表している。
- ⁷ 「平成 28 年 (2016 年) 医師・歯科医師・薬剤師調査」(厚生労働省, 2017) の結果によると、平成 28 年 12 月 31 日現在における全国の届出「医師数」は 319,480 人で「男」251,987 人 (総数の 78.9%), 「女」67,493 人 (同 21.1%) となっている。
- ⁸ 医師が男性であると聞くのは、女性であると聞くよりも「普通」であること、また「驚き」が少ないことに対応している。さらに、【病院なび】のように医師や医療機関を検索する場合、「女医」でヒットする検索数のほうが「*男医」でのそれより少なく、検索の絞り込み度が高いことにも対応している。
- ⁹ 起こりうる相反する事象の中のどの事象が起きたのか確定しない状況 (確率分布) から 1 つの事象が起きたと確定した状況への変化が引き起こす情報の増加 (カルバック・ライブラー情報量) が、その事象が起こったと聞いた時の (自己) 情報量に一致するのは至極当然のことであり、情報量の増大の値としてカルバック・ライブラー情報量の定義が妥当であることを示す結果であると考えられる。

参考文献

- 甘利俊一. (2011). 『情報理論』東京: 筑摩書房
- 上野千鶴子+メディアの中の性差別を考える会. (1996). 『きつと変えられる性差別語 - 私たちのガイドライン』東京: 三省堂.
- 厚生労働省. (2017). 「平成 28 年 (2016 年) 医師・歯科医師・薬剤師調査の概況」. 更新日 2017 年 12 月 14 日. 最終アクセス 2019 年 11 月 12 日.
- 小林健治. (2016). 『最新 差別語・不快語』東京: 人間出版.
- 佐藤裕. (2005). 『差別論-偏見理論批判-』東京: 明石書店.
- 田中和子, 諸橋泰樹. (1996). 「新聞は女性をどう表現しているか」. 田中和子, 諸橋泰樹 (編著). 『ジェンダーからみた新聞のうら・おもて - 新聞女性学入門-』第 I 部第 1 章. 東京: 現代書館.
- 中村桃子. (1995). 『ことばとフェミニズム』. 東京: 勁草書房.
- 鳴門市民劇場機関誌. (2016). 『観劇ひろば 渦潮』第 111 号 (2016 年 11 月

27 日発行) p.6 のインタビュー記事「楽屋訪問 73」。鳴門市: 鳴門市民劇場。

Cover, T. M. and Thomas, J. A. (2006) *Elements of Information Theory* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Kullback, S., and Leibler, R. A. (1951) On information and sufficiency. *Annals of Mathematical Statistics* **22**: 79-86.

Shannon, C. E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* **22**: 379-423 and 623-656.