

医学用語の選択に見られる特徴

**Characteristics of the Choice of Japanese Medical Words
in the Corpora of Scientific and Clinical Documents**

金子 周司 Shuji Kaneko

京都大学大学院薬学研究科

Kyoto University Graduate School of Pharmaceutical Sciences

発表者について

京都大学大学院薬学研究科・生体機能解析学分野
www.pharm.kyoto-u.ac.jp/channel/ja/index.html

京都大学大学院薬学研究科 生体機能解析学分野
Department of Molecular Pharmacology, Graduate School of Pharmaceutical Sciences, Kyoto University

アクセス
メンバー
研究概要
募集
ラボツアー
業績
実験プロトコル
教育
辞書
内部連絡

薬学研究科トップ
京都大学トップ

◆ 中川准教授らの研究成果が「The Journal of Neuroscience」誌に掲載され、国際的に注目されています。

◆ 平成24年度より大学院の入試制度（修士、博士）が大幅に変更になりました。

◆ 薬学部6年制（薬学科）学生に対する[研究への励ましメッセージ](#)を掲示しています。

◆ イオンチャンネル創薬とは？ → [研究概要での解説をご覧ください](#) →

◆ [ライフサイエンス辞書](#)がシソーラスにバージョンアップ。

©2004-2012 Department of Molecular Pharmacology

WebLSD検索結果: apoptosis
http://lsd.pharm.kyoto-u.ac.jp/cgi-bin/lsdproj/ejlook Google

ONLINE LIFE SCIENCE DICTIONARY
ライフサイエンス辞書オンラインサービス

プロジェクト WebLSD O/D英語教材 EtoJ Vocab EtoJ WebSpell
ミラーサイト [京都1 | 京都2 | 東京] フォントサイズ [小 | 中 | 大 | 特大] ENGLISH

新規対訳の受け付け | 検索結果の読み方 |

英和・和英 シソーラス 英語共起表現 オプション表示/非表示

英和・和英検索: apoptosis search

▶ 英和検索結果

▶ **apoptosis** ***** 音声 音声 シソーラス 共起表現 Scholar, Entrez, Google, Wikipedia
(遺伝子にプログラムされた能動的な細胞死) **アポトーシス**, **アポプトーシス**, **プログラム細胞死**, **予定死**
【関連語】 apoptotic, PCD, programmed cell death, programmed death
【用法】 induce apoptosis [アポトーシスを誘発する] / T-cell receptor-induced apoptosis [T細胞受容体が誘導するアポトーシス] / undergo apoptosis [アポトーシスを起こす] / UV-induced apoptosis [紫外線が誘発するアポトーシス] 用例

▶ **apoptosis-inducing factor** ** シソーラス 共起表現 Scholar, Entrez, Google, Wikipedia
アポトーシス誘発因子

▶ **apoptosis induction** *** 共起表現 Scholar, Entrez, Google, Wikipedia
アポトーシス誘導
【関連語】 induction of apoptosis

▶ **apoptosis regulatory protein** * シソーラス Scholar, Entrez, Google, Wikipedia
アポトーシス調節タンパク質
【関連語】 anti-apoptotic protein, pro-apoptotic protein

136543812th transaction of WebLSD@Pharm node1 since Nov 11, 1996
WebLSDに未収録の専門用語（用法）は「新規対訳の受け付け」から投稿できます。

Page Top
Copyright (C) 2009 ライフサイエンス辞書プロジェクト

- 中枢神経薬理学の研究をしつつ、20年来、ライフサイエンス辞書を構築しています。

はじめに

- 医療や生命科学の急激な進歩は、莫大な数の専門用語を新たに生み出している。
- 発表者は医学系学生や研究者が英語の専門用語を学習・活用するための電子辞書の開発に20年来取り組んできたが、日本語については訳語として位置づけ、あまりその特性について深く考察してこなかった(金子2006「ライフサイエンス辞書とは」[情報管理](#), 49:1, pp.24-35)。
- しかし今後、医療や教育の電子化がますます進展し、自然言語処理が医療サポートや知識発見に応用されていくことを考えると、医学用語の日本語表記について理解を深めることが必要と思われる(金子、大武 2010「ライフサイエンス辞書からクリニカルインフォマティクスへ」[情報管理](#), 53:9, pp.473-479)。
- 本研究ではどのようにして和文で医学用語が選択されているかを少しでも知るために、医学文献や医薬品解説書を元にしたコーパスを構築し、専門用語を抽出した上で異表記を収集、解析した。
- コーパスの解析結果からも、日本語では漢字、カタカナ、ひらがな、英語綴りなどを混在して用いることができるため、異表記が非常に多いという特徴があり、編集者や許認可者による修正を経た後の文書においても、医学用語の多様性は維持されていた。いくつかの例を紹介して考察してみたい。

コーパスの概要

- 株式会社羊土社の協力を得て、1996年から2005年にわたって『実験医学』誌に発表された医学研究総説の全文テキストを**実験医学コーパス (37.3Mbyte)**とした。
- 財団法人日本医薬情報センター(通称JAPIC)が有料で販売している医療用医薬品全13,000種の添付文書情報(2008年版)について、解析目的での使用許諾を得てテキスト化し、**JAPICコーパス (49.6Mbyte)**とした。
- 解析は、ライフサイエンス辞書に収録している157,347語の日本語をクエリーとして、コーパス中で一致する文字列の頻度をPerlスクリプトにより求めた。
- これら2種類のコーパスの概要を表1に示す。以下においては同規模のコーパスとして頻度(語数)の比較を行う。

表1 本研究で用いたコーパスの概要

	実験医学	JAPIC
文字数	20, 235, 504	25, 247, 795
読点数	271, 158	418, 839
句点数	504, 847	687, 669
「など」頻度	22, 897	26, 997
「血管」頻度	13, 146	12, 345
「高い」頻度	4, 265	4, 087

各コーパスの特徴

- 表2はそれぞれのコーパスで求めた頻度のうち、一方での値が他方の100倍以上であった特徴語を示している。
- 実験医学コーパスにおいては、最先端の成果を研究者自らが執筆していることもあり、「遺伝子」「タンパク質」「配列」といった生体分子の名称や物性を表す語が多く、「シグナル」や「ドメイン」のように専門家の間でのみ通用するjargonとも考えられるカタカナ語が多用されている点が特徴的である。
- 一方、JAPICコーパスで最も頻度が高いのは「本剤」であるが、これは医薬品添付文書における主語として多用されるためである。その他にも「経口投与」「血中濃度」など添付文書における解説として専門家に注目されるべき特徴語が見られる。医薬品は同一作用機序をもつ類似薬が多いこともあり、それらの添付文書間では記述も似ている傾向がある。このことは実験医学コーパスからは50,257種類の語が抽出されたのに対して、JAPICコーパスからは36,449語しか抽出されなかった解析結果に反映されている。
- 図1には各コーパスを構成している文字種の割合を示した。いずれのコーパスにおいても英数字とカタカナが3～4割、漢字の割合も3～4割を占めており、きわめて専門用語に満ちた文書であることがわかる。

表2 コーパスの特徴語(頻度データ)

語	実験医学	JAPIC
遺伝子	45,676	294
タンパク質	31,544	53
シグナル	18,755	14
ドメイン	14,474	7
配列	10,640	16
本剤	35	99,945
経口投与	90	21,550
血中濃度	135	19,973
既往症	2	14,879
妊婦	11	14,007

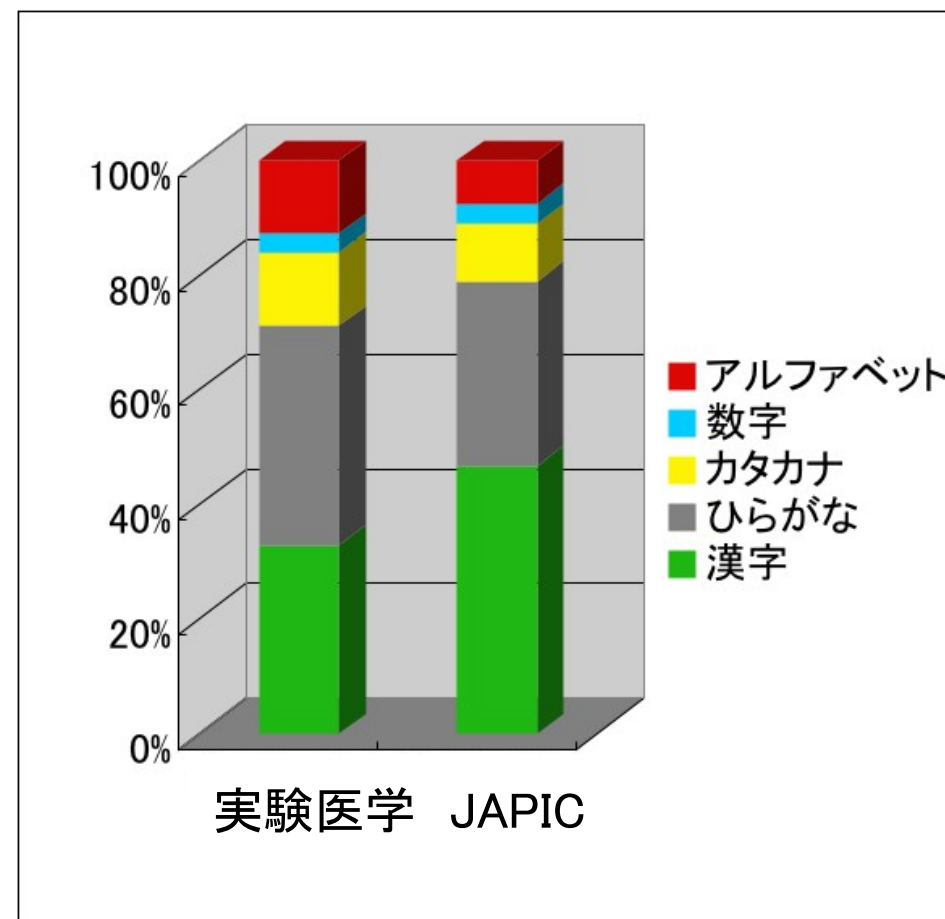


図1 コーパスを構成する文字種

結果1:「protein」使い分けの実情

- Proteinとはアミノ酸がペプチド結合によって連なる生体構成分子種である。英語においてはprotein以外の異表記は存在しない。
- 日本語においてはproteinが卵白に多く含まれることに起源をもつ「蛋白」から「蛋白質」という語を生み出し、日本医学会は用語集で「蛋白質」を推奨している。しかし文部科学省は学術用語として「タンパク質」という表記を標準としており、新聞や報道等においては「たんぱく質」という表記が多く採用されている。それぞれから「質」を除去した表記も多く用いられ、さらには「プロテイン」と表記すれば一般社会においてはサプリメントとして用いられる補助栄養食品を指すかのように微妙に使い分けられている。
- 今回、構築したコーパスにおいて調べてみた結果を表3に示すが、それぞれ編集者や許認可者の手が入った文書であるにもかかわらず、多様な表記が検出された。

表3 「protein」の日本語表記における選

語	実験医学	JAPIC
たんぱく質	0	14
タンパク質	31,544	53
タンパク	4,330	174
蛋白質	2,693	489
蛋白	1,067	4,337
プロテイン	801	77

実験医学コーパスで「タンパク質」が多く、JAPICコーパスで「蛋白」が多いのは基礎医学と医療という分野間の差異であると考えられる。

- 実験医学コーパスにおいて、前に「結合」や後に「分解」が接続する場合はいずれも「タンパク質」が多く用いられていた(表4)。「プリオン」との接続においては「プリオン蛋白」という表記が特異的に高い傾向が見られた。このことはカタカナ同士が接続した場合に元の語の境界が分かりづらくなることを避けている表現なのかもしれない。しかし、タンパク質をリン酸化する酵素であるprotein kinaseを表す際には、そのままカタカナ語として「プロテインキナーゼ」が最も頻出した。
- JAPICコーパスにおいては、「蛋白結合」のように「蛋白」という表記が全般的に好んで用いられていたが、この用語はいずれの省庁や団体も推奨している表記ではない。一部においては「糖蛋白質」のように「質」をつけた表記が集中しているケースも見られたが、これは他の類似薬で用いられた文書をそのまま流用して使っているために複製増幅効果が現れたものと推察される。

表4 「protein」の接続語に応じた選択

コーパス	接続語	タンパク質	タンパク	蛋白質	蛋白	プロテイン
実験医学	結合～	1,327	199	139	29	0
	プリオン～	53	9	2	54	0
	～分解	676	128	18	11	0
	～キナーゼ	75	42	3	0	489
JAPIC	～結合	0	43	0	1,693	0
	糖～	2	20	114	76	0

結果2：薬物のカテゴリーを表す名称

- 表5は腫瘍の増殖に対して抑制的に作用するカテゴリーの薬物に与えられる一般的な名称を調査した結果である。
- この結果から、いずれのコーパスにおいても多様な表記が混在していることがわかる。専門的には「癌≠悪性腫瘍」であり「癌＝上皮細胞の(つまり一部の)悪性腫瘍」であることを加味すると、このように階層の異なる概念を同一視している現状は好ましいとは言えない。

表5 腫瘍増殖を抑制する薬物の名称

語	実験医学	JAPIC
抗癌薬	19	0
抗癌剤	763	32
抗がん薬	0	9
抗がん剤	3	6
制癌剤	37	3
抗腫瘍薬	13	0
抗腫瘍剤	12	52
抗悪性腫瘍薬	10	13
抗悪性腫瘍剤	1	741
悪性腫瘍治療薬	1	0

まとめ

- 医学用語は長らく標準化の方向性で議論されていた。
- しかしながら、本研究で編集者や許認可者の修正を経た文書コーパスを解析した結果、コントロールされた状況においても医学用語の多様性は失われていないことが明らかになった。
- 実際に現場で作成される文書(例えば電子カルテや学会抄録など)はさらに多様で混沌としているであろうことは容易に想像できる。
- 今後、医療文書の電子化などによって情報の利活用を目指す場合、このように多様な異表記に耐えうる(かつ英語表記や略記にも対応した)頑強なシソーラスを早急に整備することが必要と思われる。

(参考)テキスト自動タグづけ



file:///Users/skaneko/Desktop/MedNLP/tagged.html

file:///Users/s...NLP/tagged.html

【家族歴】
気管支鏡検査 気管支内腔に明らかな異常なし BAL マクロファージ71.6% リンパ球10.6% 好中球14.6% 好酸球3.2%

【剖検所見】 肝S4,5,8に最大割面10×8cm大の腫瘍。
心雑音なし
腹部:軟:腸蠕動音低下。
母糖尿病
海外旅歴 東南アジア

(2)感染性腸炎

【生活社会歴】
腹部症状強いため、消化管穿孔、虚血性腸炎を考え消化器外科オンコールコンサルト、腸管内に便が貯留し、便秘の可能性高いが、虚血性腸炎は否定できず、かりのそうであった場合でも年齢、基礎疾患を考えると、手術の適応は難しいとの判断であった。
プロブレムリスト
腸蠕動音:正常。
SpO2 (自発呼吸O2:1L/分)98%

【主訴】
胸部X線写真:特に異常所見なし。
【内分泌】成長ホルモン(GH)0.91 ng/ml,LH 31.5mIU/ml,FSH 52.3mIU/ml,プロラクチン 11.2ng/ml,コルチゾール 15.9 μg/dl,ACTH 60.1 pg/ml,ADH 5.4pg/ml

【現病歴】
当院初診。
【生活社会歴】
12月12日にERCPを施行したところ乳頭部と思われた部位に腫瘍をほぼ全周性に認めた。
対座法にて右側半盲あり。
fiber type groupingは認められないAcid-pd浸潤細胞の一部は陽性となり大食細胞であると考えられた。
その先に正常の十二指腸開口部確認。
今回冠動脈造影時にあわせて両上肢血管造影を行った。
免疫学所見:CRP 0.09 mg/dl,HbA1c 7.0%,C-peptide R 1.8 ng/ml
【血算】WBC 4200/μl(Lym 1428/μl),RBC 288×10⁴/μl,Hb 7.7g/dl,Ht 23.2%,Plt 11.6×10⁴/μl

(模擬電子カルテの解析例)

異表記をすべて検出して、
統制語への紐づけをする

★PCでお見せできます