



一般向け医学文章における 難解語特定に関する研究

東京大学大学院 医学系研究科
社会医学専攻 山田恵美子



一般向けに作成された医学文章はそれほど易しくない

例：動脈硬化（gooヘルスケアより）

臨床的に最も重要である粥状硬化は、大動脈、脳動脈、冠動脈
一番問題となる

(括弧で括って移動)

脈など比較的太い動脈に起こる硬化で、動脈の内膜にコレステロールなどの脂肪からなる粥腫(じゅくしゅ)(アテローム)ができ、

(リンク)

次第に肥厚することで動脈の内腔が狭くなります。粥腫が破れ

ると血栓がつくられ、動脈は...

血の塊 できて



易しくない要因の分類

本研究の対象

	text1	text2	合計(%)
難解な語	9	13	37.9
説明の欠如	3	7	17.2
不要な語	1	5	10.3
難解な語列	3	3	10.3
助詞の欠如	2	2	6.8
語の不統一	1	1	3.4
接続詞の欠如	2	0	3.4
難解な係り受け	1	1	3.4
その他	3	2	8.6



方法

メルクマニュアル家庭版より11ページサンプルを収集
人手(一人)で難しい語をアノテート(152箇所)

CRF(IOBフォーマット)で学習

素性

POS

医学文章における出現頻度

一般文章における出現頻度 (=Google Hit)

その文章における出現頻度

文字の種類(平仮名・カタカナ・漢字)

文字列の長さ

(一般・医学)辞典にのっているかのっていないか

予備実験結果

11次交差検定

精度	98.12%
適合率	74.88%
再現率	51.01%



問題点

アノテートが難しい

- 作業するうちに作業者の知識が増加する
作業途中で評価基準が変わる
- 個々の語は理解できても複数組み合わせると分からない場合がある
どうアノテートするか
「語は理解できるのに、読んでいくうちにいつのまにかよくわからなくなっていた」
- 得られる情報は「難しい」か否かではなく「難しいと思う」かどうか
理解しているつもりで理解していない場合もある

大量にデータを作るのが大変

メルクマニュアルに依存

具体例

明らかな意識障害がなくても、先に述べた肝機能検査で著しい異常を認めた場合には、劇症肝炎へ移行する危険性があることを考え、肝臓専門医と相談しながら治療を行い、できるだけ早期に総合的な治療が可能な医療機関へ移送することが大切です。

ビリルビン分子の複数個が互いに結合して...

原因がわかれば、それをすみやかに除去します。

潰瘍部位にバリウムがたまるため、ニッシュと呼ばれる特有の像を示します。

上腹部痛などの症状を伴わない例が多いので、治療を受けないまま悪化して出血を起こしたり、難治性の潰瘍に移行する例が多いといわれています。

線維腺腫は小さい充実性のしこりで、癌ではなく良性の腫瘍です。

最近、E型肝炎ウイルスによる劇症肝炎例も報告され、いずれも海外渡航歴のない国内に在住する人からの発病として話題になっており、非A非B非C型の場合には検索が必要です。