

データ駆動型学習用コーパス SCoRE を構成する英文の評価：

英文難易度，教育用語句のカバー率，

および意味分野別カバー率の分布から

<http://www.score-corpus.org/>

中條 清美 ・ 濱田 彰 ・ 小林雄一郎（日本大学）

平成29-32年度 科学研究費補助金基盤研究（B）（17H02366）

流れ

1. はじめに

- データ駆動型学習 Data-Driven Learning (DDL)
- 教育用例文コーパス SCoRE (第4次開発版)

2. SCoREコーパスの客観的評価

- 評価1 : 英文難易度評定
- 評価2 : 教育用語句のカバー率
- 評価3 : 語彙の意味分野別分布

データ駆動型学習 Data-Driven Learning (DDL)

- ▶ **豊富な言語使用の例**に触れることで言語の規則性やパターンを**発見して学ぶ**学習方法
- ▶ コーパスを利用して**授業や自習**において語彙や文法を学習する
- ▶ **適切なDDL**は語彙や文法学習において**高い効果**を発揮するという**メタ分析**の結果
(Mizumoto & Chujo, 2015; Boulton & Cobb, 2017)

たくさんの例文に触れて言葉の規則性に気づく

where		サンプリング	なし	5	10	20	ソート	出現順	左	キーワード	右	表示	KWIC	センテンス
1	Is there a place where we can talk ?	私たちが話のできる場所がありますか？												
2	Is there a place where I can rest ?	私が休むことのできる場所がありますか？												
3	Is there a place where she can sit ?	彼女が座れる場所がありますか？												
4	I know a place where we can hide .	私は私たちが隠れることのできる場所を知っています。												
5	She knows a place where we can go .	彼女は私たちが行くことのできる場所を知っています。												
6	Here is the place where the accident happ...	ここが事故が起こった場所です。												
7	Do you mean the place where we met ?	私たちが会った場所という意味ですか？												
8	Do you remember the place where we met ?	あなたは私たちが会った場所を覚えていますか？												
9	It 's a place where we can relax .	そこは私たちがくつろぐことのできる場所です。												
10	Is there a place where we can eat ?	私たちが食事のできる場所がありますか？												
11	Here is the room where you will sleep .	ここはあなたが眠る予定の部屋です。												
12	Is this the room where you slept ?	これはあなたが眠った部屋ですか？												
13	Here is the room where we paint .	これは私たちがペンキを塗る部屋です。												
14	Here is the room where we dance .	これは私たちがダンスをする部屋です。												
15	There is the room where he reads .	彼が読書をする部屋があります。												
16	That is the room where she works .	あれは彼女が仕事する部屋です。												
17	This is the room where it all began .	これはすべてが始まった部屋です。												
18	Here is the room where she died .	ここは彼女が死んだ部屋です。												
19	Here is a room where you can play .	ここはあなたが遊ぶことのできる部屋です。												
20	This is the room where the children sleep .	これは子供たちが眠る部屋です。												

教育用例文コーパスSCoRE

(Sentence Corpus of Remedial English)

<http://www.score-corpus.org/> (無料・登録不要)

1. N S が作例した簡潔で自然な英語例文 10,460文
2. 英語例文が高校英語教科書レベルの範囲内
3. 学習者の不得意な文法項目 (22項目) を含む
4. 初級・中級・上級レベル (文長と語彙習得学年)

初：8語以内, 1～2年, 中：5-11語, 1～3年, 上：9語以上, 4年以上

5. 日本語対訳付き

ツールを開く▶

パターンブラウザ

コンコーダンス

適語補充問題

ダウンロード

SCoREとは

SCoREの開発と特徴

プロジェクトメンバー

成果

プロジェクトで開発した
コーパス検索ツール

ユーザガイド

DDL教材バンク

動作環境

更新履歴

用例の投稿

SCoREとは

教育用例文コーパスSCoRE (Sentence Corpus of Remedial English) は、簡潔で自然な英文とその日本語対訳をウェブ上で自由に（無料・登録不要）閲覧・検索・コピー・ダウンロード・Webテストに利用できる「データ駆動型英語学習支援プログラム」です。

コンコーダンス | SCoRE

ツールの切り替え ▼

検索

検索 * to

▼ to不定詞 1175

- 動詞 + to不定詞 312
- 動詞 + 目的語 + to不定詞 209
- 名詞 + to不定詞 290
- 不定詞 (目的) 90
- 形容詞 + to不定詞 284

▼ 受動態 354

- be + 過去分詞 354

▼ 現在完了 548

- have + 過去分詞 380
- got 23
- been 49
- seen 39
- gone 35
- done 46
- come 35
- made 41
- given 37
- lost 36
- changed 39

サンプリング なし 5 10 20 ソート 出現 左 キーワード 右 表示 KWIC センテンス

1	Her parents wanted her to be a doctor .	彼女の両親は彼女に医者になってもらいたいと思いまし...
2	ed up in front of the store , wanting to be the first to ...	何手もの客たちは、新しいiPhoneを買うのに一番になり...
3	It began to rain .	雨が降り始めました。
4	The flowers began to open .	花が咲き始めました。
5	n the ink on the documents began to fade , she knew ...	書類のインクがうすくなり始めたので、彼女はプリンタ...
6	He tries to spread a message...	彼は親切と思いやりのメッセージを広めようとしていま...
7	I tried to finish the book be...	私は期限前にその本を読み終えようとしたましたが、全く...
8	It 's starting to snow , so may...	雪が降り始めているので、おそらく明日授業はキャンセ...
9	rted to rise and the animals began to move around .	太陽が昇り始め、動物たちは動きまわり始めました。
10	, the traffic started to slow down to avoid accidents o...	いったん雨が降り始めると、交通はぬれた道路での事故...
11	I never learned to ski but I did lea...	私はスキーは習ったことがありませんがスケートは習い...
12	ve shown that children who learn to read and write in ...	研究は、就学前に読み書きを学ぶ子供たちは人生をより...
13	He forgot to lock the door .	彼はドアにかぎをかけるのを忘れました。
14	Did you forget that I need to use the car today ?	あなたは今日私が車を使う必要があるのを忘れましたか...
15	me that he used the credit card to pay for the tickets .	彼はチケットを買うのにクレジットカードを使ったこと...
16	ne of your debts before you try to buy a house .	家を購入しようとする前に、負債の一部を清算すること...
17	Dan loves to travel internation...	ダン是世界各国を旅行するのが大好きです。
18	I love to read all kinds of bo...	私はあらゆる種類の本を読むのが大好きです。
19	I do n't know what you hope to accomplish by doi...	あなたがそれをして何を達成したいのかわたしにはわか...
20	She hoped to get some help c...	彼女はこれらすべてのバッグを階段の上に乗るのを手伝...

2. SCoREコーパスの 客観的評価

評價①

英文難易度評定

英文難易度評価の例文抽出に使用したコーパス

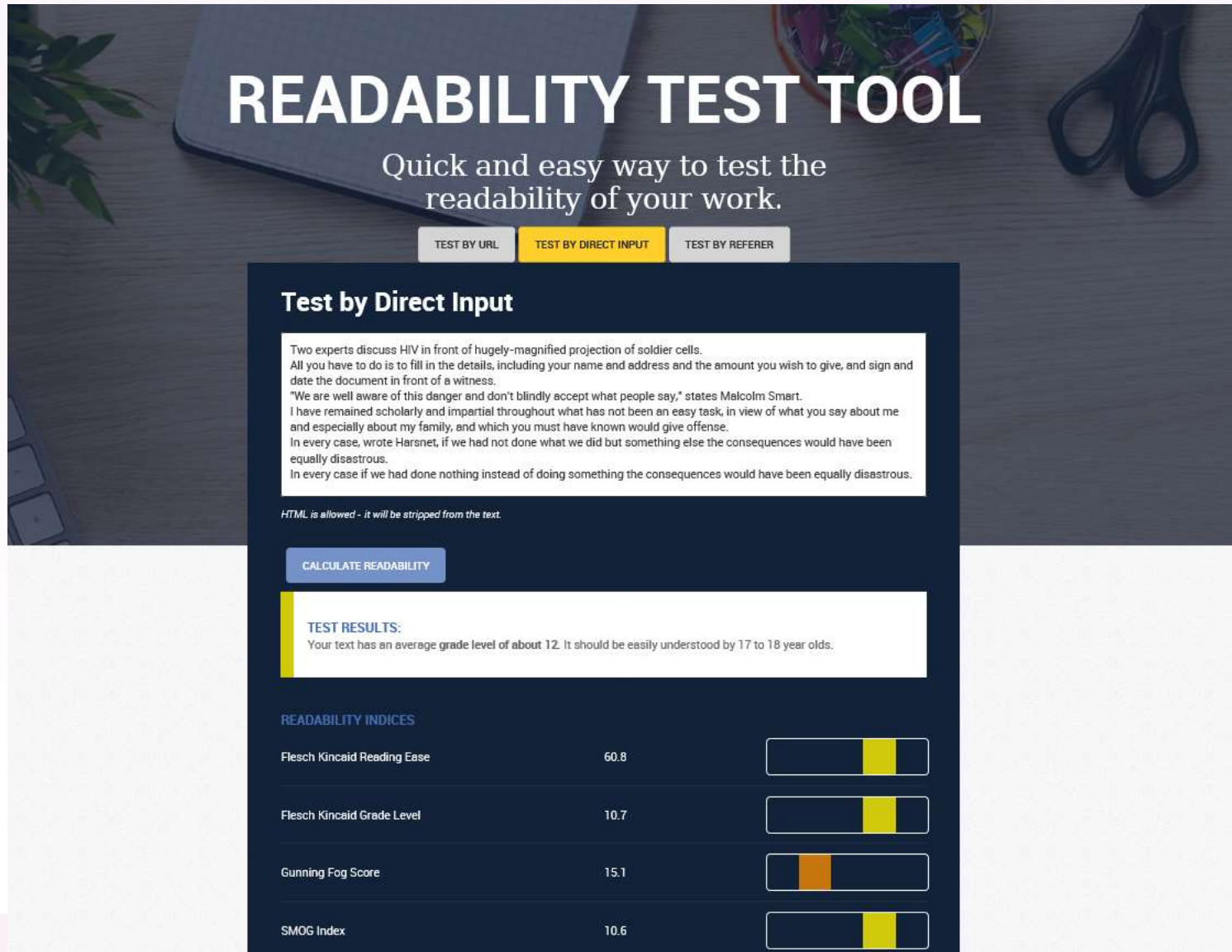
- 1) SCoRE 初級・中級・上級 2) CoBLE (米国教科書Corpus of Beginner Level English)
3) WPN (新聞 WebParaNews) 4) BNC (British National Corpus)

	だれでも 利用可能	不安にな らない	日本語訳 がある	上級レベ ルあり	世界的に 通用する	教育用に 配慮
SCoRE	✓	✓	✓	✓	✓	✓
CoBLE		✓	✓		✓	✓
WPN	✓		✓	✓	✓	
BNC	✓			✓	✓	

評価用英語例文の抽出方法

- 英語例文の文法項目
 仮定法, 関係詞, 前置詞
- **3種**の検索語句
 [If * had *], [what * say/sa*], [in front of]
- SCoRE 3レベルを含む**6種**コーパス
- 各コーパスの検索結果の先頭から**2文**を選択
- 合計例文数 : **3 種語句 × 12文**

Readability による評価 <https://www.webpagefx.com/tools/read-able/>



The screenshot shows the 'Readability Test Tool' interface. At the top, there's a header with the title 'READABILITY TEST TOOL' and a subtitle 'Quick and easy way to test the readability of your work.' Below this are three buttons: 'TEST BY URL', 'TEST BY DIRECT INPUT' (which is highlighted in yellow), and 'TEST BY REFERER'. The main section is titled 'Test by Direct Input' and contains a text area with sample text. Below the text area is a 'CALCULATE READABILITY' button. The results section, titled 'TEST RESULTS:', shows that the text has an average grade level of about 12, which should be easily understood by 17 to 18 year olds. At the bottom, there's a table of 'READABILITY INDICES' with four rows: Flesch Kincaid Reading Ease (60.8), Flesch Kincaid Grade Level (10.7), Gunning Fog Score (15.1), and SMOG Index (10.6). Each row has a corresponding progress bar indicating the score's position relative to a scale.

READABILITY TEST TOOL

Quick and easy way to test the readability of your work.

TEST BY URL TEST BY DIRECT INPUT TEST BY REFERER

Test by Direct Input

Two experts discuss HIV in front of hugely-magnified projection of soldier cells.
All you have to do is to fill in the details, including your name and address and the amount you wish to give, and sign and date the document in front of a witness.
"We are well aware of this danger and don't blindly accept what people say," states Malcolm Smart.
I have remained scholarly and impartial throughout what has not been an easy task, in view of what you say about me and especially about my family, and which you must have known would give offense.
In every case, wrote Harsnet, if we had not done what we did but something else the consequences would have been equally disastrous.
In every case if we had done nothing instead of doing something the consequences would have been equally disastrous.

HTML is allowed - it will be stripped from the text.

CALCULATE READABILITY

TEST RESULTS:

Your text has an average grade level of about 12. It should be easily understood by 17 to 18 year olds.

READABILITY INDICES

Flesch Kincaid Reading Ease	60.8	<div></div>
Flesch Kincaid Grade Level	10.7	<div></div>
Gunning Fog Score	15.1	<div></div>
SMOG Index	10.6	<div></div>

(1) 次の 12 文は, “**in front of**” の検索結果です。

授業で検索して出てきた英文としてよいと思われる文かどうか, 各文について右側の評価項目について,

5. 強くそう思う, 4. そう思う, 3. どちらとも言えない, 2. そう思わない, 1. 全くそう思わない, の中から当てはまるものを 1 つ選んで「数字」を書いてください。

		文の難し さはちよう どよい	文の長さ はちようど よい	単語の難 しさはちよ うどよい	英語として 自然な文 である	文の構造 はわかりや すい	文の意味 はわかりや すい	目標の文 法を学習 しやすい	日本語訳 がなくても 理解できる
1	He stepped in front of her.								
2	She stood in front of the door.								
3	Liz crouched in front of the puppy to pet it.								
4	There is a large box in front of the door.								
5	Erica saw the lion in front of the pet store.								
6	Two experts discuss HIV in front of hugely-magnified projection of soldier cells.								
7	After a while, Tom's large van finally arrived in front of our house.								
8	All the bicycles in front of the station had been locked so they were removed using tools.								
9	They sat right in front of the large screen to get a nice view of the abstract artwork.								
10	Because it was built in front of Kyongbokkung Palace, the symbol of the South Korean people, South Korean administrations repeatedly planned to demolish the building.								
11	The pinnacle will be lowered to the ground by a crane during an Aug. 15 ceremony in front of the museum to commemorate the 50th anniversary of South Korea's liberation from Japan.								
12	All you have to do is to fill in the details, including your name and address and the amount you wish to give, and sign and date the document in front of a witness.								

データ分析の方法

- ランダムフォレストを用いて，文法学習のしやすさに影響を与える要因を特定した
- 文法学習のしやすさ評定の回答パターンに対する個々の要因の影響力（変数重要度）を推定した。
- 質問紙8項目のうち「目標の文法を学習しやすい」を従属変数，残り7項目を独立変数とした。

学習者による英文難易度評価からわかったこと

- SCoRE, CoBLE, WebParaNews, BNCのコーパスの英文のうち, SCoREは, 文構造や文意のわかりやすさの点で, 他の3種のコーパスよりも評価が高かった。
- ランダムフォレストを用いて, 文法学習のしやすさの評価に影響を与える質問項目の重要度を推定した。その結果, 文法学習のしやすさに大きく影響を与えていたのは, 文構造のわかりやすさ, 文意のわかりやすさ, および日本語訳の有無であった。

評価②

教育用語句のカバー率

教育用語句を抽出するための教育用書籍

- 1) マーフィーのケンブリッジ英文法（初級編）

Basic Grammar in Use (Murphy, 2011)



- 2) マーフィーのケンブリッジ英文法（中級編）

Grammar in Use Intermediate (Murphy, 2009)

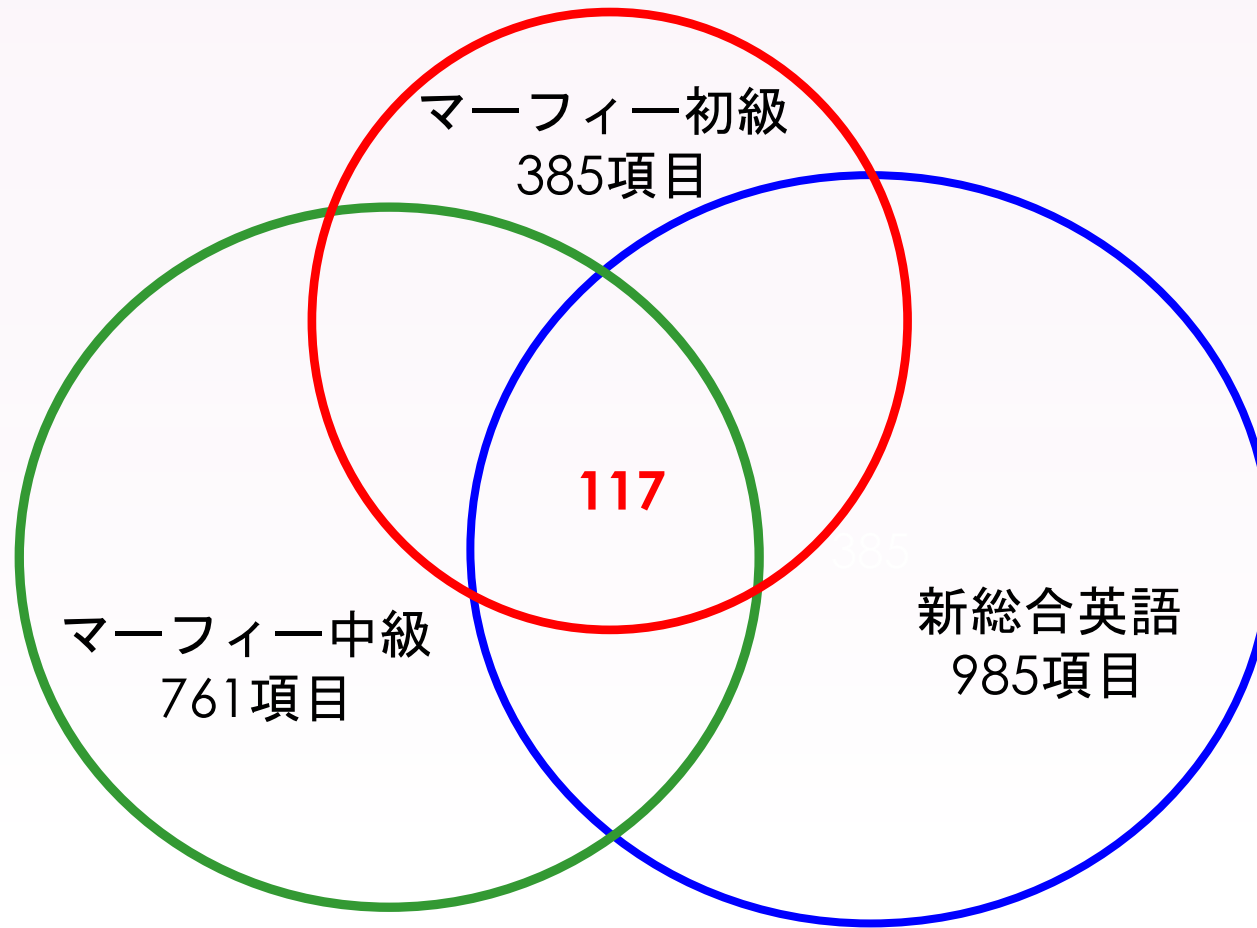


- 3) 基礎からの新総合英語(チャート式シリーズ)

(高橋・根岸, 2003)



和集合 1,610項目, 積集合 117項目
117項目のうちSCoREにないのは, have got, ought to, till



評価③

語彙の意味分野別分布

SCoRE英文構成語彙の意味領域の分布を調査

- 「どんな検索語句を入力しても何らかの検索結果が出てほしい」
- tea や rice だけでなく subway や hockey や firefighter も出てきてほしい
- このような要素を測るひとつの目安として、英文を構成する語彙の意味領域を調査
- 調査の対象：
 - (1) SCoREの**初級**例文3,756文
 - (2) SCoREの**中級**例文3,506文
 - (3) SCoREの**上級**例文3,198文
 - (4) **COCAの Thematic Words** 例文5,217文 (次回の第5次開発の追加分)
 - (5) SCoRE の 全例文10,467文 + COCAの例文5,217文 (合計15,684文)
- 調査方法： WMatrix3 (Paul Rayson)の **USAS Semantic Tagging System** を利用
 - (1) から (5) の SCoRE の英文例文の構成語彙を 21 の意味領域に分類し、その割合を調査

Wmatrix の semantic tagging

Wmatrix tag wizard

[Tag Wizard...](#) | [My Tag Wizard...](#) | [My Dictionaries...](#) | [Domain Tag Wizard...](#) | [Load file without tagging...](#)]
[[Folders > My folders](#) | [Details](#) | [Create...](#) | [Delete...](#) | [Archive...](#) | [Extract...](#) | [Join...](#) | [Share...](#) | [NEW! CrossTab...](#) | [Empty TRASH](#)]
[[Options > Switch to Simple Interface](#) | [Edit user options...](#)]
[[Help > Contents](#) | [Availability](#) | [Tagsets: POS & Semantic](#) | [USAS: Lexicon & MWEs & Context rules](#) | [Updates](#) | [Feedback](#)]

You are logged in to Wmatrix3 as: nihonchujo

[You are here > [My folders](#)]

Upload file

→

Part-of-speech tagging

→

Semantic tagging


→

Frequency lists

1. Enter new folder name:

2. Click the browse button to select a file:
 参照...

3. Upload now Reset form



The Wmatrix tag wizard takes you automatically through the POS (CLAWS tagger) and Semantic analysis (USAS tagger) stages, and produces frequency lists from your text file. Please do not run large texts (e.g. with more than 1,000,000 words) through the tag wizard. These are better run off-line and loaded into Wmatrix afterwards. Please get in touch with Paul to do this.

File types:
The **raw text type** is plain ASCII, as created by Windows notepad for example. If you have a MS Word file, you should convert it to a 'txt' file before uploading to Wmatrix. Open the file in MS Word, select 'save as' and then the file type should be set to 'text only with line breaks' with MS-Word 2000 and earlier. If using MS-Word 2003 select 'plain text', then in the dialog box click 'insert line breaks' and 'allow character substitution' and then save. This last option is required for replacing Windows apostrophes for ASCII equivalents. Raw files will have their contents enclosed in a <wmatrix> tag to enable CLAWS tagging. Wmatrix is also capable of dealing with text in HTML, SGML or XML format. The taggers do not require parsable encoding, it is necessary only that left and right angled brackets are well-balanced.

Further [input format guidelines](#) are available including for example how to avoid problems with less-than and greater-than symbols in the input text.

One text per folder:
If you do not specify a folder, one will be created with a unique name. It is recommended that you use a **new folder for each file**. If your corpus consists of more than one file, then we recommend concatenating the files together first.

Large texts:
Please do not run large texts (e.g. with more than 1,000,000 words) through the tag wizard. These are better run off-line and loaded into Wmatrix afterwards. Please get in touch with Paul to do this. Since Wmatrix does not pre-index the texts, speed of concordancing is proportional to the text size. Producing a concordance from larger texts may take up to 1 minute and may cause the web browser to timeout. A good rule of thumb is that texts should not exceed 1 million words (even when run through the tag wizard offline).

Wmatrix

©2000-17 UCREL, Lancaster University.
For technical queries please contact Paul Rayson : p.rayson@lancaster.ac.uk

UCREL Sematic Analysis System (USAS) によるSemantic Tagging

英文の構成語彙を「生物・身体・感情・食物・物質・科学・言語」**21の意味領域に分類**

A general and abstract terms	B the body and the individual	C arts and crafts	E emotion
F food and farming	G government and public	H architecture, housing and the home	I money and commerce in industry
K entertainment, sports and games	L life and living things	M movement, location, travel and transport	N numbers and measurement
O substances, materials, objects and equipment	P education	Q language and communication	S social actions, states and processes
T Time	W world and environment	X psychological actions, states and processes	Y science and technology
Z names and grammar			

まとめ

- SCoREに収録されている英文について，（１）英文難易度評定，（２）教育用書籍で扱われている教育用語句のカバー率，（３）SCoREの語彙の意味領域の分布の観点から，SCoREコーパスがDDLを行う上で教育用コーパスとしてどの程度適切なのかを評価した。
- これらの結果からSCoREの現状分析と次回開発への改良点をさぐった。
- 本研究はSCoREのような教育用コーパスを評価する「初めの一歩」
- 今後：評価方法の精査，他の観点からの評価方法の考案，他のコーパスとの比較
- 謝辞：本研究は平成29-32年度 科学研究費補助金基盤研究（B）（17H02366）（中條・アントニ・水本・西垣・松下・濱田・小林）を受けて行われました。