コーパスを教育利用するための要件としての適応学年指標
― 文法コーパスと TED コーパス構築

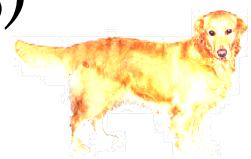
Chrome
(recommended)

文法コーパス　SCoRE on Talkies
http://www.mintap.com/talkies/?score

TED コーパス　selected360 on Talkies
http://www.mintap.com/talkies/?ted360

ミント音声教育研究所　田淵　龍二

言語処理学会第 24 回年年次大会(NLP2018)
2018 年 3 月 13 日（火）15：50～



研究の動機と目的

ICT 環境の進展と社会のニーズに応じて細分化と大規模化が進むコーパスは、語学教育に豊かな多様性を与えてくれる。

コーパスには，データドリブン（データ駆動型）学習としての利用もあるが，現状ではまだ研究と産業用が主だと見受けられる。

そこで，コーパスを教育，特に学校での語学教材とする要件について研究し、コーパスを構築した。

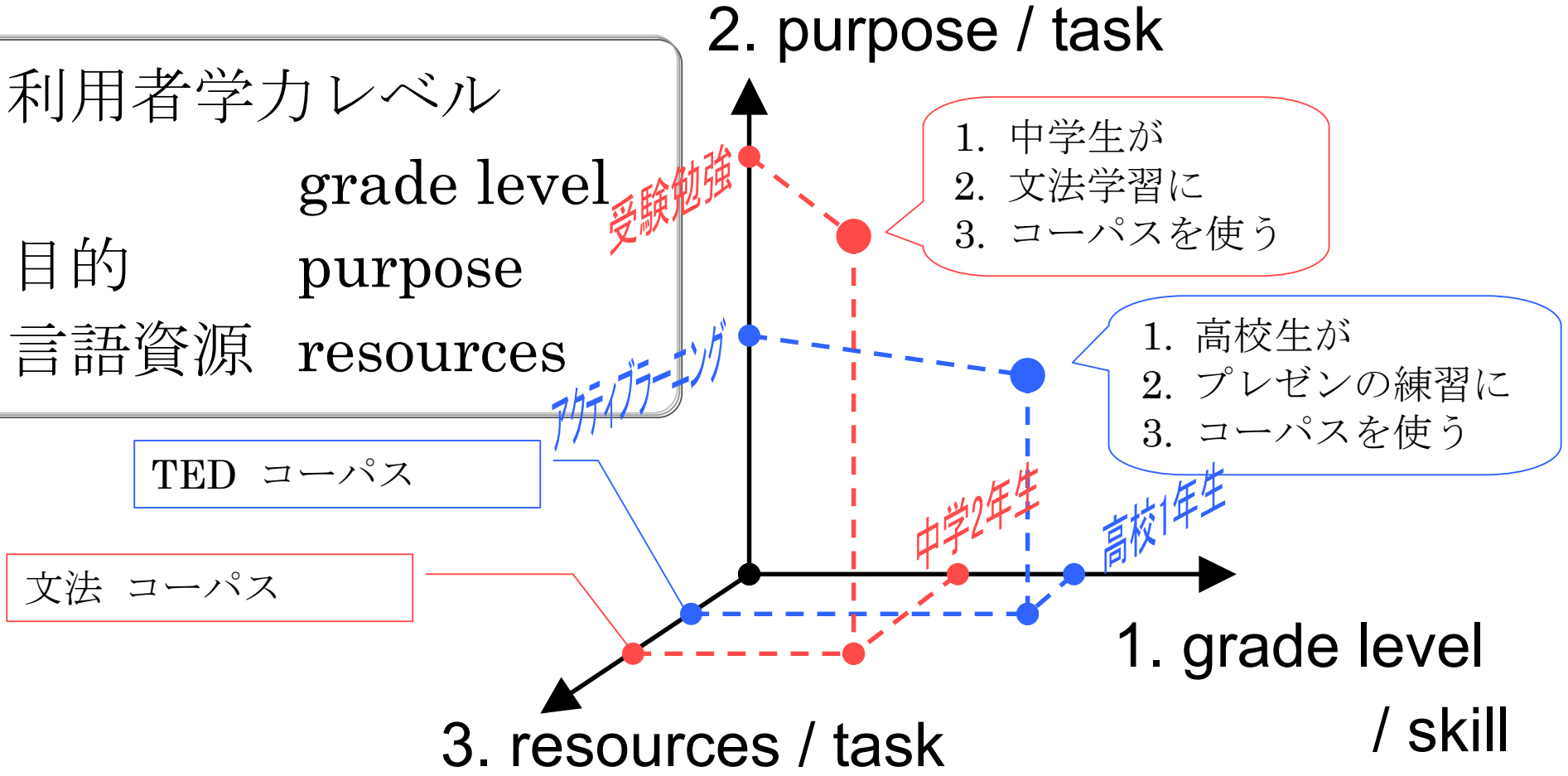
学習用コーパスの３項関係

研究・産業用

学習用

誰（何年生）が、何のために、何を

1. 利用者学力レベル　grade level
2. 目的　purpose
3. 言語資源　resources

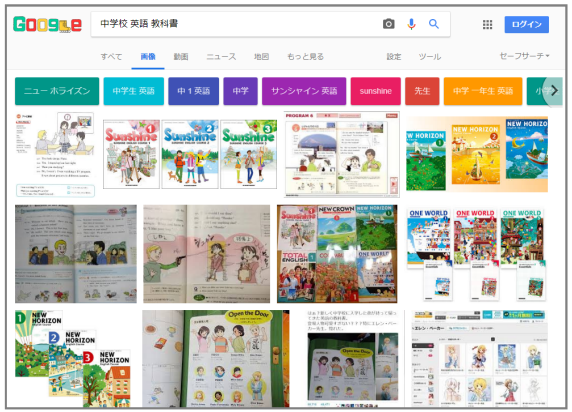


コーパスの適応学年

おもちゃには対象年齢

https://www.kosodate-ryouhin.co.jp/SHOP/134684/list.html

教科書には適応学年がある



https://www.google.co.jp/

学習用コーパスにも適応学年が欲しい

では、コーパスの適応学年は どうやって測るか？
この問いに答えるにはヒトの読解プロセスの知見が必要

モデル・ヒューマン・プロセッサ

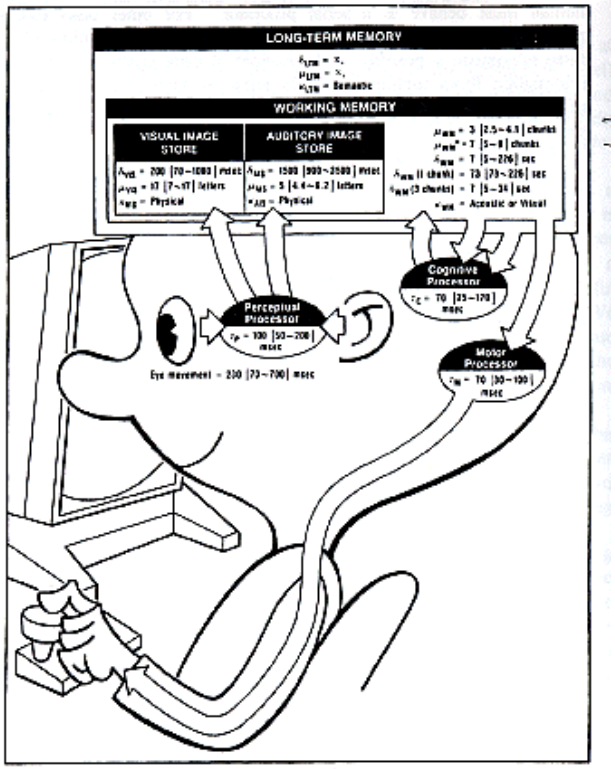
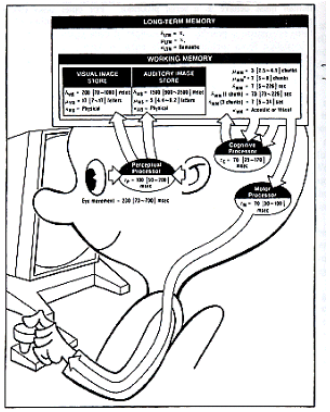


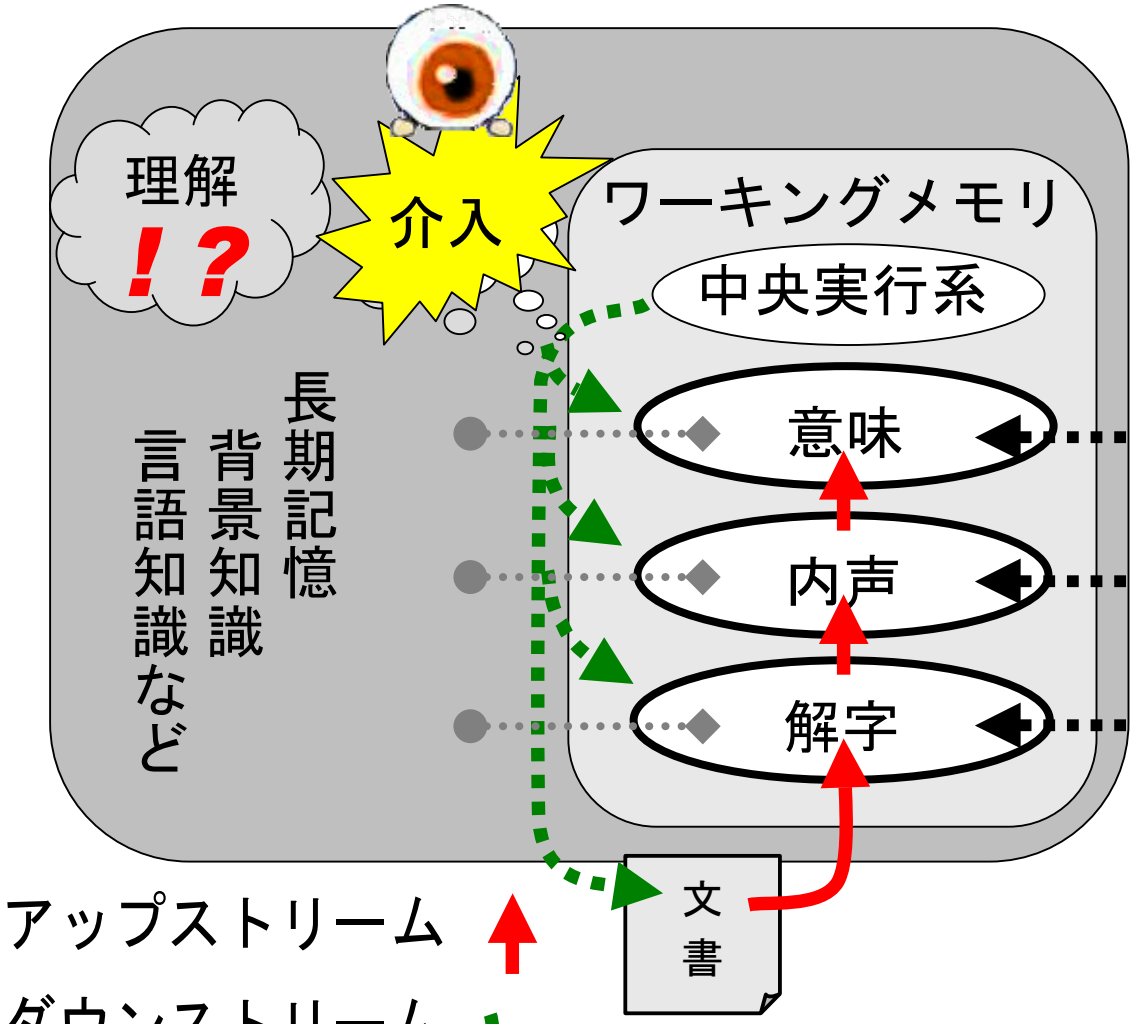
Figure 2.1. The Model Human Processor—memories and processors.
Sensory information flows into Working Memory through the Perceptual Processor. Working Memory consists of activated chunks in Long-Term Memory. The basic principle of operation of the Model Human Processor is the *Recognize-Act Cycle of the Cognitive Processor* (P0 in Figure 2.2). The Motor Processor is set in motion through activation of chunks in Working Memory.

Card, S., Moran, T. P., & Newell, A. (1983). The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

ヒトの視聴覚機能をコンピュータに見立てて、反応速度や記憶容量、記憶時間などの計測結果を統合した。
（Card 1983）

読解プロセス Reading Process (1)





semantic
4±1 chunks
acoustic
2±1 sec
physical
230 [70-700] msec

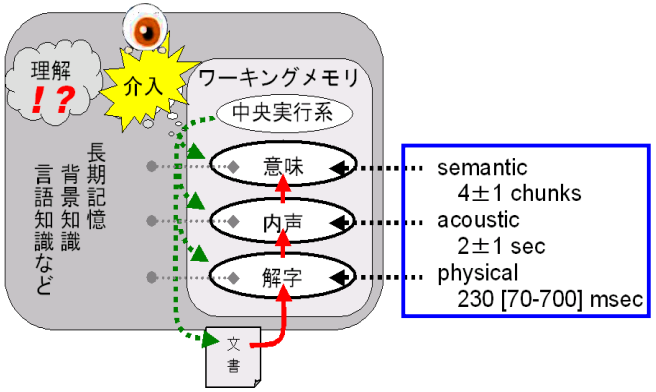
田淵・湯舟（2014）『外国語教育メディア学会（LET）第 54 回全国研究大会発表要項』pp.116–117 を改変（田淵）

読解プロセス Reading Process (2)

読解プロセスと読解指標

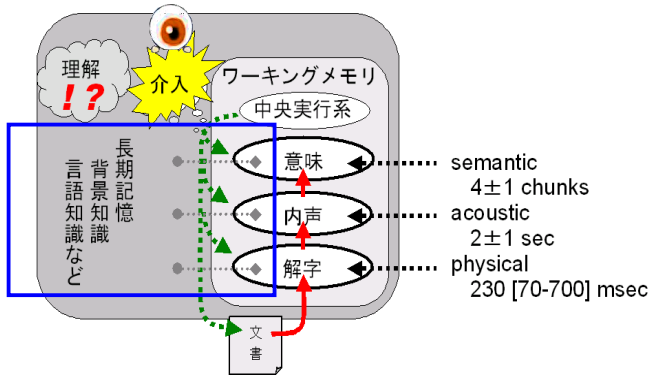
リーダビリティ

- 主に アップストリーム
- テキストがワーキングメモリの時間的容量的特性の許容範囲内か？



語彙レベル

- 主に 長期記憶とのアクセス
- アップストリームが円滑か？
- 中央実行系の介入が頻繁か？



… 文法、文型、文構造、段落構成などは扱っていない

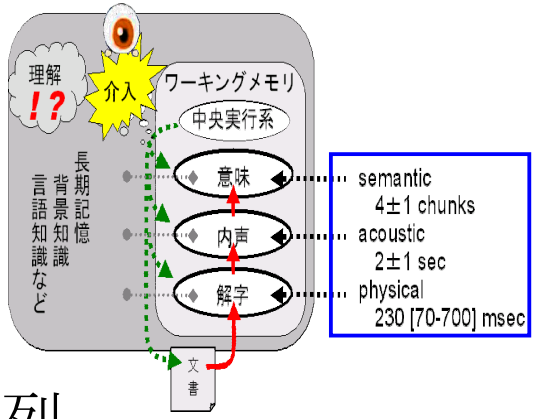
リーダビリティ公式（1）

Mint Grade Level of English Text for Readers in Japan

MGJP = 0.07496*(3*S+2*C)+7.926*LOG(P)+4.618

where

S: １フレーズに含まれる平均音節数
C: １フレーズに含まれる平均子音数
P: １節に含まれる平均フレーズ数

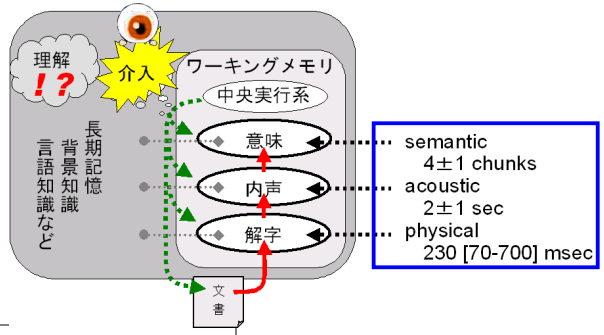


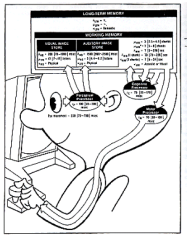
LOG: 常用対数
フレーズ: 英数字以外で区切られた文字列
節: . ! ? ; などの記号で区切られた文字列

田淵龍二・湯舟英一（2016）.「音韻符号化の予測時間に基づく日本人英語学習者向けリーダビリティ公式の開発」. Language Education & Technology, 52, 359-388.
http://mintap.kir.jp/public/news/pic/2016_3bg.pdf

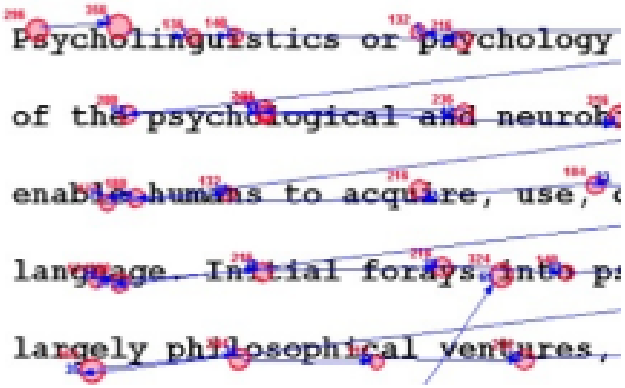
読解プロセス Reading Process (2)

アップストリーム

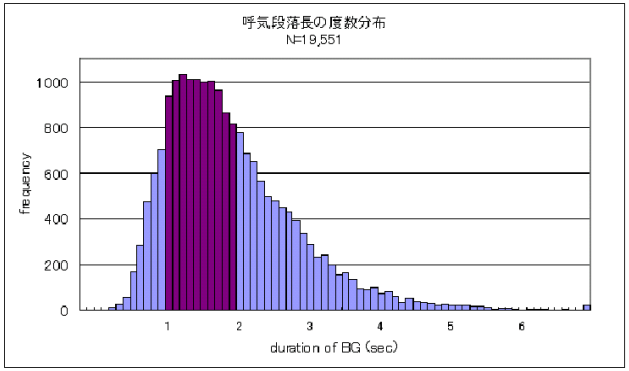




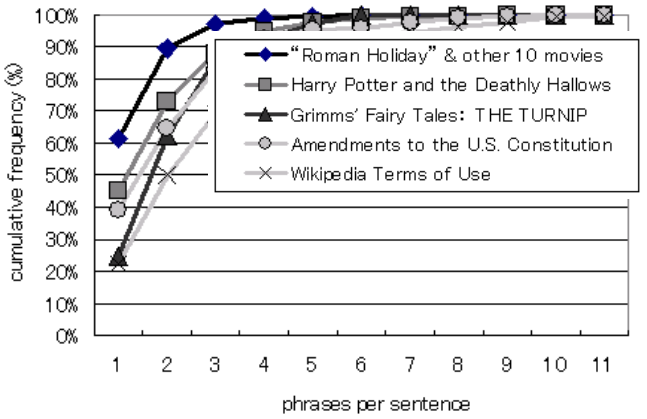
解字 physical	内声 acoustic	意味 semantic
眼球運動解析	呼気段落長解析	句数による文長解析
eye movement	breath group	sentence size
230 [70-700] msec	2±1 sec	4±1 chunks



http://ameblo.jp/psycholinguistics2003/entry-11746454176.html

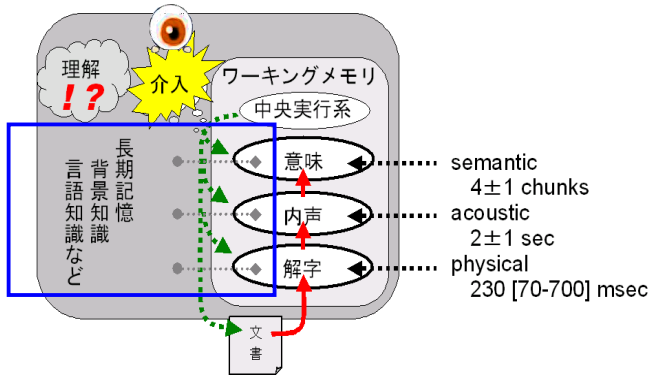


田淵・湯舟（2017）「TED Talks 字幕の表示時間の特徴とその教育的利用に向けた考察」. 外国語教育メディア学会 第 54 号 (2017) 167–192 など



田淵・湯舟（2015）「音韻符号化の予測時間に基づく日本人英語学習者向けリーダビリティ公式の開発」. 外国語教育メディア学会 第 52 号 (2015)

語彙レベル公式（1）



考え方： 長期記憶に入っている語彙を推測

Q1 対象読者は？

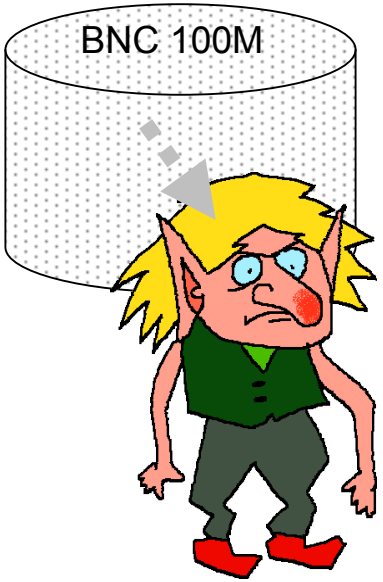
A1 日本の英語学習者（中高大生）

Q2 彼らが接してきた英文は何か？

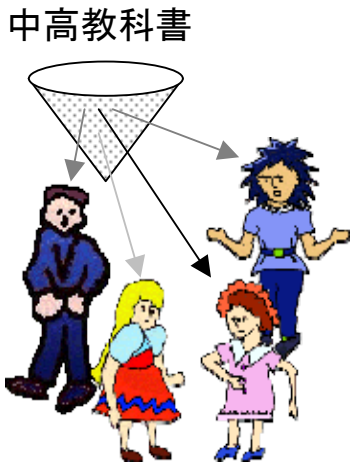
A2 学校の教科書＋入試過去問

BNC 頻度表を使わなかった理由

- 対象が日本の非母語話者
- 高親密度日常単語の低位傾向

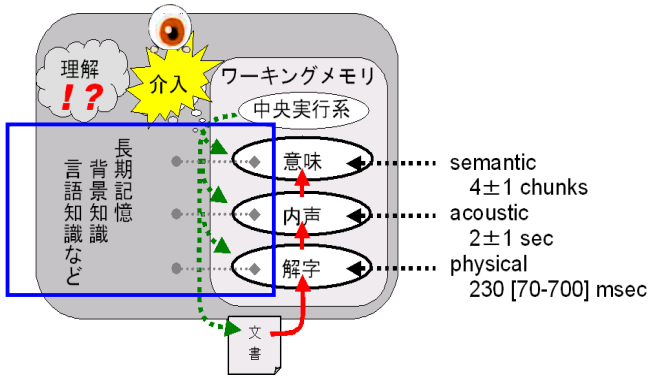


BNC 100M を読破した幻の巨人



教科書を 6 年間学んだ大学生

語彙レベル公式（2）



考え方

長期記憶に入っている語彙を推測する

日本の学制区分に即した学年指標にする

推測方法

1. 頻度　　より多くの者が読むテキスト
2. 親密度　　ある学年が共通して学ぶ語彙

語彙レベルを推測するための言語資源

- 市販検定教科書（中学、高校）
- 入試（おもに大学入試）など試験問題

語彙レベル公式（3）

Vocabulary Grade Level of English Text for Readers in Japan

VGL = 30.371*H+8.7914

where

H: 大学受験レベル単語数 ÷ 総単語数

田淵龍二（2017）.「日本人英語学習者向け語彙レベル適応学年算出公式の試験的開発」. LET Kanto Journal, 1, 25-35.
http://mintap.kir.jp/public/news/pic/let_k1_2_201703.pdf

MGJP, VGL と 学制との対照表

MGJP, VGL	学制
9	中 3 / junior high school
10	高 1 / senior high school
11	高 2
12	高 3
13	大 1 / university, college

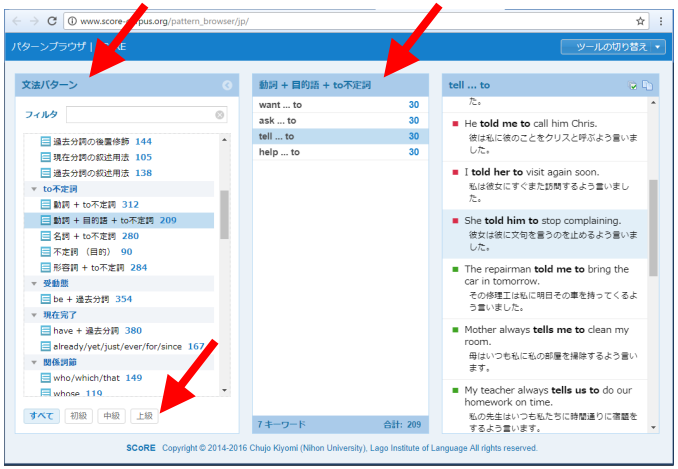
小数点以下は四捨五入して学年とした。

適応学年指標を持つコーパスを構築（1）

言語資源

文法コーパス　SCoRE

http://www.score-corpus.org/pattern_browser/jp/

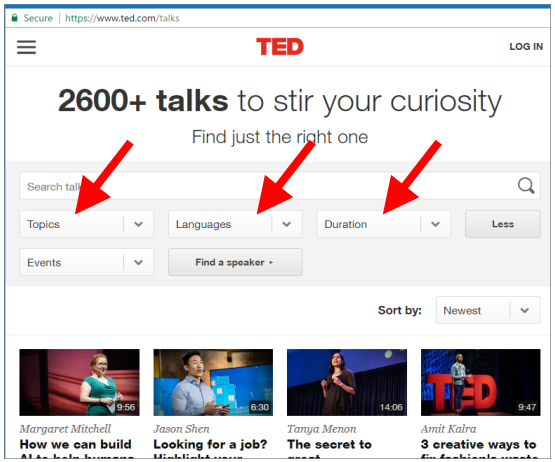


フィルタ　文法項目、級分け

構築　SCoRE on Talkies

講演動画　TED

https://www.ted.com/talks



トピックス、長さなど

selected360

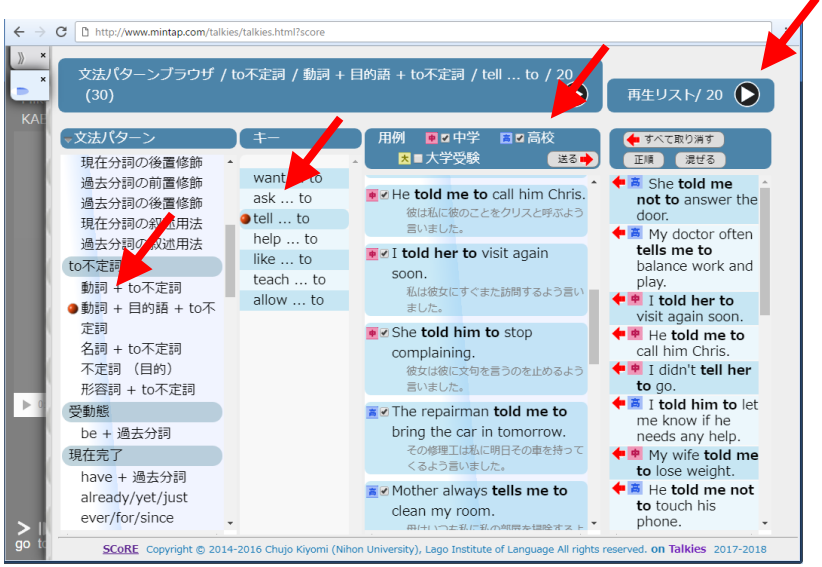
適応学年指標を持つコーパスを構築（2）

1. SCoRE on Talkies （オープンサイト）

http://www.mintap.com/talkies/?score

Chrome
(recommended)

Talkies 対応にリメイク
級分けを適応学年（中高大区分）に



フィルターチェック

再生 / PC

スマホ

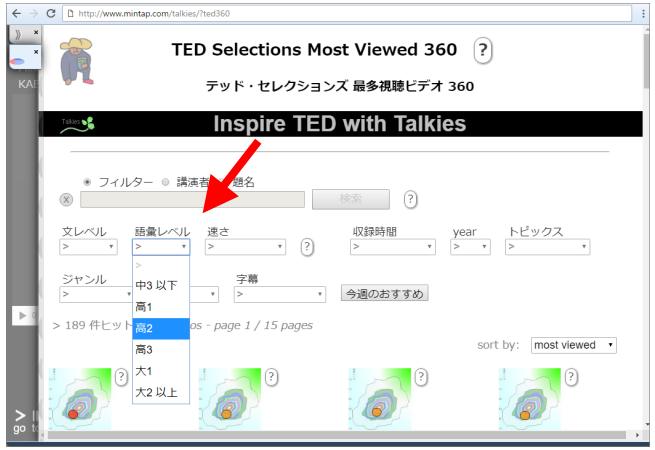
適応学年指標を持つコーパスを構築（2）

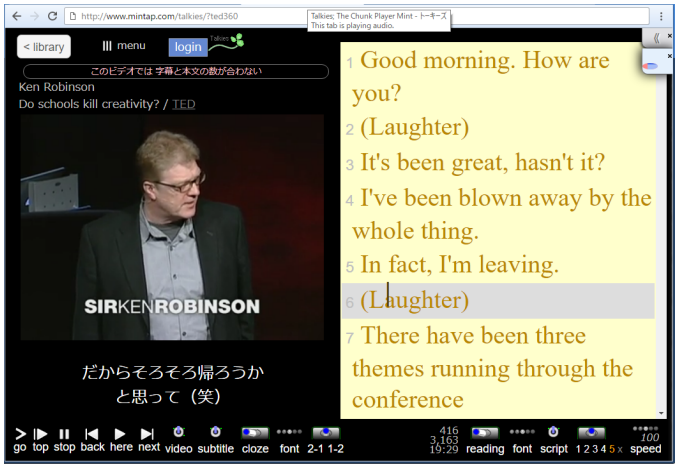
2. selected360　（オープンサイト）

http://www.mintap.com/talkies/?ted360

Chrome
(recommended)

最多視聴ビデオ上位 360 本
文・語彙適応学年と速さレベル





フィルターチェック　選択　再生 /　PC　　スマホ

ありがとうございました

アンケートは回収箱へ

ハンドアウト（PDF 版）や TED コーパスなどの記事を配信するメルマガをご希望の方はメールでお知らせください

tabuchiryuji@nifty.ne.jp