一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

テキストアナリティクスと音声解析と認知科学と検索エンジン

田淵 龍二

ミント音声教育研究所 〒370-0013 群馬県高崎市萩原町 950-31
E-mail: tabuchiryuji@nifty.ne.jp

あらまし フレッシュ・キンケイド公式などアメリカのリーダビリティ公式を音声解析による認知科学的視点から再評価したところ、作動記憶（ワーキングメモリ）との関連が極めて濃いことがわかってきた。基礎研究としてのテキスト読解過程の認知科学的基盤、およびテキストから音声情報を抽出する方法を説明する。こうした知見を語学教育における教材選びに応用し、適応学年ごとに学習に最適なテキストや動画を選ぶ検索エンジンを開発した。こうした基礎研究から応用までの流れを紹介をすることで、単語や形態素以外の要素に注目したテキストアナリティクスの多様な一面を明らかにする。

キーワード テキストアナリティクス，音声解析，認知科学，検索エンジン，リーダビリティ公式

Text analytics, sound analytics, cognitive science and search engine

Ryuji TABUCHI

Mint Phonetics Education Institute 950-31 Hagiwara, Takasaki-shi, Gunmma, 370-0013 Japan
E-mail: tabuchiryuji@nifty.ne.jp

Abstract We reevaluated readability formula in American such as Flesch-Kincaid from a cognitive scientific point of view by speech analysis and found that the relationship between readability and working memory is extremely strong. We developed an application that applied this knowledge to teaching materials for language education. We explain cognitive science base of text reading process, introduce audio information extraction method from text and demonstrate the use of search engines that can select texts and videos that are optimal for learning for each grade. In this way, we will clarify various aspects of text analytics focusing on elements other than words and morphemes.

Keywords text analytics, sound analytics, cognitive science, search engine, readability formula

1. はじめに

自動音声認識技術の進捗が著しいとはいえ、音声解析の精度はまだ十分とは言えない。それに比べて文解析（テキストアナリティクス）は、良質な資料を大量かつ手軽に入手でき、緻密かつ安定的に解析できることが特徴である。文解析を俯瞰すると、単語（token, lemma）や形態素（morpheme）・構文（structure）などを手段としつつ共起や相関を調べたり、目的に応じて分野や語句を絞り込む手法が多く見受けられる。

そうした中で今回は、単語や形態素・構文以外の要素に注目したテキストアナリティクスの多様な一面として、テキストから音声情報・時間情報を抽出して活用する方法を紹介する。

以下の議論の対象言語は英語（アメリカ）である。

2. 呼気段落と認知科学

2.1. 発声と呼気段落

音声は連綿とつながっているように聞こえるが、よく観察すると 1 秒以下の無音区間（pause）によって数秒程度に区切られている。これを呼気段落（breath group）と呼ぶ。音が区切れる主な要因は吸気である。図 1 に十数秒の英語音声の波形とスペクトログラムを示す。A, B, C, D が呼気段落であり、その段落が始まる直前のスペクトログラムには、a, b, c,d の位置に吸気にともなう雑音がかすかに観察される。

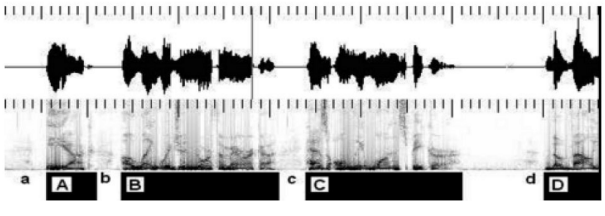

図 1. 連続音声の視覚化と呼気段落 [1]

かつて服部[2]は呼気段落について「一つの単音あるいは音節よりなるものから，極めて長いものまで存在し，その長さを規定すべき法則がなく，その成立原因は極めて複雑で全部を科学的に研究することはほとんど不可能」とし、

その後､呼気段落の研究はほとんど進んでいなかった。

2.2. 呼気段落長

音声を使った英語教材を作成していた筆者は意味の塊りとしてのチャンクに注目し、チャンクごとに自在に再生可能なプレーヤーを開発した。その過程で、チャンクが吸気による無音区間で区切られていることに気が付いた。チャンクのほとんどは呼気段落であった。そこで呼気段落を解析した[1]ところ、その継続時間は対数正規分布となり（図 2)、呼気段落長の平均は約 2 秒で、最頻区間は 1 秒から 2 秒（棒グラフの濃い部分）であった。右上がりの曲線は累積割合を示している。3 秒までで 9 割を超えているのがわかる。呼気段落の単語数と発話速度と度数の立体散布図（度数を高さとした等高線図）は牡蠣殻を伏せたような独特の形状（図 3）となった。単語数で測った発話速度の単語数ごと平均値（中黒点のある矩形）は対数関数で近似された。この近似曲線を話速曲線と呼ぶ。単語数が多い呼気段落ほど話速が速いことが見て取れる。

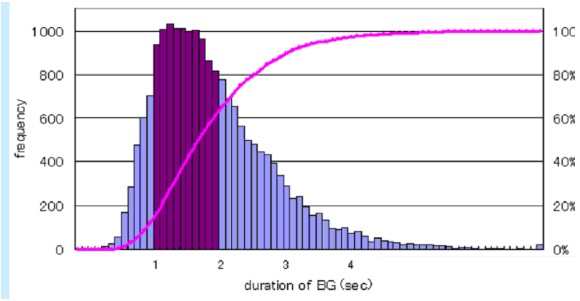

図 2. 呼気段落長の度数分布（n=19,551）[3] を改変

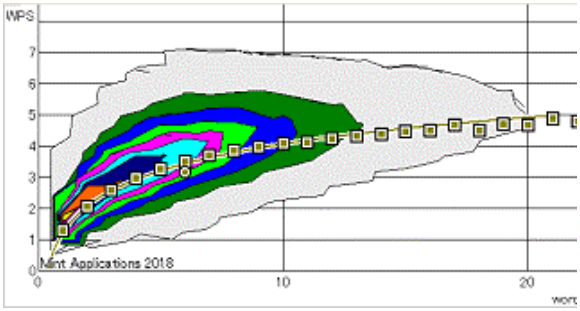

図 3. 単語数ごと発話速度の分布と平均値 [3] を改変

2.3. 呼気段落と作動記憶（working memory）

呼気段落長の対数正規分布特性と話速曲線の対数特性から、強い規則性が示唆されたことから、認知科学との関連を調べたところ、Card の Model Human Processor（MHP）[4]と通じることが判明した。つまり、作動記憶の聴覚イメージ貯蔵庫時間特性は平均 1.5 秒で分布範囲は 0.9 秒から 3.5 秒との説である。

こうした呼気段落と聴覚作動記憶（音韻ループ; phonological loop）の時間幅の一致から、発声においては相手に聴き取りやすいように作動記憶の時間特性に合わせて 2 秒程度の音の固まりを形成するようになったのではないだろうか。逆に、人の音声が 2 秒程度のまとまりで発信されるのでそれに合わせて認知機構が進化したのかもしれない。あるいは、両者の時間幅がたまたま一致したので人の音声言語が発達できたのではないかと筆者は考えた。

2.4. 読解プロセス（reading process）

文字を読むときには頭の中で声を出している場合が多い。これを内声（inner voice）などと呼ぶ。経験的には、耳から入る外声とよく似ている。この現象を音韻符号化（phonological coding）と呼ぶ。そこで筆者は、MHP（Card）と呼気段落長解析などを総合して、読解プロセスの概要図（図 4）を作成した。なおここでは詳述しないが、意味処理（semantic）の部分では、句数で数えた文長解析結果（図 5）も勘案している。一般文書では 4 句以下で構成された文が全体の 9 割以上を占めていることが見て取れる。

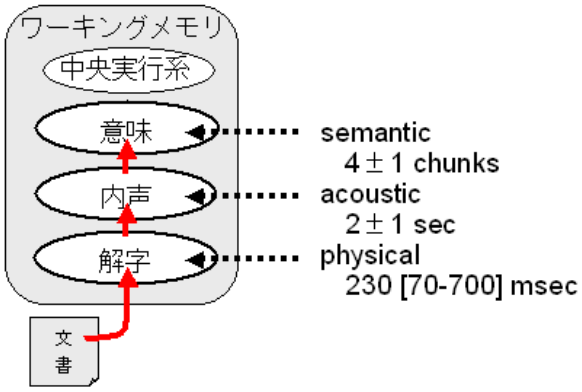

図 4. 読解プロセスの概要（長期記憶は省略）[5]を改変

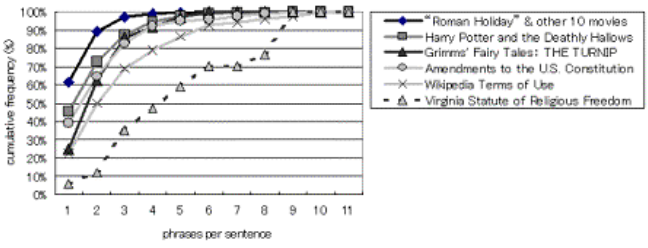

図 5. 句数で数えた文長ごとの累積（句点区切りを文とし、読点区切りを句とした） [6]

2.5. 音韻符号化時間予測

テキストを黙読（音韻符号化）するときにどのくらいの速さで読んでいるかを直接測る技術はないので、

音声（外声）から類推して式①を作成した[6]。

D = 120× Sy + 80× Cn ・・・ ①

where

D: estimated duration in milliseconds of a BG

Sy: number of syllables in a BG

Cn: number of consonants in a BG

BG: breath group

スピーチ動画（TED）の字幕表示継続時間と、式①で予測した字幕黙読時間を比較した（図 6）ところほぼ同等であり、よく近似できていることがわかった。

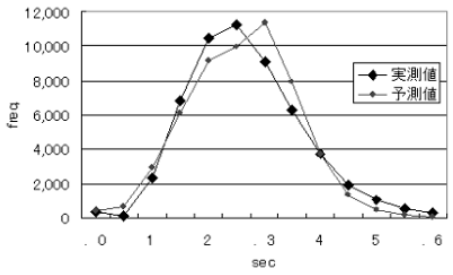

図 6. 字幕表示実測時間と字幕黙読予測時間の度数分布（7 秒以上は省略, n=54,253） [7]

3. 音響解析結果と文解析結果の応用

3.1. 認知科学に基づいたリーダビリティ公式の設計

テキストの読みやすさを数値化するリーダビリティ公式はアメリカで 100 年以上の歴史を持ち、100 以上の公式が開発されてきた[6]。しかし、それらは実用本位であり、原理についての研究はほとんどなかった。

そこで筆者は音響解析（呼気段落長）と文解析（音韻符号化時間予測と、句で数えた文長解析）、およびそれらを認知科学の観点から統合した読解プロセス概要（図 4）を基盤にした新しいリーダビリティ公式の設計をおこなった。主な点を表 1 に示す。

表 1. 新しいリーダビリティ公式の基本設計

1.	音韻符号化予測時間が短いほど読みやすい
2.	句数の少ない文ほど読みやすい
3.	読者の母語に配慮して公式を使い分ける
4.	公式の返り値を学年とする（適応学年）
5.	各国の語学教科書の指定学年を基準とする
6.	小学中学高校を中心とする（日本は中高のみ）

3.2. 新しいリーダビリティ公式の構築

公式を作成するにあたっては、式②を標準形として定めた。標準形とは言語に拠らないという意味である。母語と対象言語ごとに定数の組（a, b, c）が決定される。定数 a が乗ぜられている式が音韻符号化（句長）にかかわり、定数 b が乗ぜられている式が句数（文長）にかかわる。平均音節数と平均子音数の項にそれぞれ 3 と 2 を掛けているのは、式①の係数 120 と 80 の比である。文長の項で対数を取っているのは、図 5 の一般文書のグラフが対数で近似できるとみなしたからである。実際この対数処理のおかげで、文長が大きなテキストでの算出値の爆発を抑制する効果が得られた。

MG=a*(3*SyPP+2*CPP)+b*LOG(PPC)+c ・・・ ②

where

MG： Glade Level of the text

SyPP： Syllables Per Phrase

CPP： Consonants Per Phrase

PPC： Phrases Per Clause

Phrase： a sequence of words devided by non-letters

Clause： a sequence of words devided by ". ! ? ;" and is almost equal to a sentence

LOG： Common Logarithm

a, b, c： constant

アメリカの場合は、語学（language）の教科書と達成度テストを、日本の場合は英語検定教科書を基準文書とした。基準文書から SyPP, CPP, PPC を算出し、その文書の適応学年を MG として②に代入した結果を回帰分析して a, b, c を求めた。結果を式③④に示す。

MGEN=0.07662*(3*SyPP+2*CPP)+19.554*LOG(PPC)-3.141 ・・・ ③

MGJP=0.07496*(3*SyPP+2*CPP)+7.926*LOG(PPC)+4.618 ・・・ ④

where

MGEN： Glade Level for readers in USA

MGJP： Glade Level for readers in Japan

3.3. 新しいリーダビリティ公式の検証

公式③④のうち比較可能な公式が多数存在するアメリカの式③（MGEN）についてニューヨーク州学年別語学達成度テストの指定学年との誤差を計算した（表 2）。指定学年との誤差は平均で 0.2 学年と小さく、他の公式と比べて最小であった。ばらつきを示す標準偏差は 0.7 で、算出値に±1 するのが妥当だとわかった。ちなみに、日本で有名なフレッシュ・キンケイド公式（FKGL）の誤差は 1 学年以上であった。

表 2. MGEN と，達成度テスト学年との誤差

GL-ABP	MGEN	FKGL	CLI	FORC	FryG	GFog	SMOG	LWF	ARI
3	1.3	-0.8	0.8	5.1	-0.7	1.5	3.1	0.1	-1.8
4	-0.6	-1.3	0.5	4.6	-0.8	1.1	2.6	-0.7	-2.2
5	0.5	-0.8	1.2	3.6	0.8	1.5	2.5	0.2	-1.2
6	0.4	-0.5	1.6	3.3	0.3	1.6	2.3	0.1	-0.5
7	-0.3	-1.4	-0.1	2.3	-0.9	1.0	1.6	-0.6	-2.0
8	-0.4	-1.6	-0.1	1.3	-0.6	0.7	1.0	-0.7	-1.6
mean	0.2	-1.1	0.7	3.4	-0.3	1.2	2.2	-0.3	-1.6
SD	0.7	0.4	0.7	1.4	0.7	0.4	0.8	0.4	0.6

[6]を改変

4. 語学学習用コーパス・検索エンジンの要件

近年のウェブ環境の進展とともに、かつては考えられなかったほどの情報に手軽にアクセスできるようになったことはよいことである。しかし、人の処理能力には限度があり、特に習熟途中の学習者にとっては、過大な情報量はかえって意欲を低下させる恐れもある。

そこで「学習利用する言語資源の適応学年を素材選択肢に加えたコーパスを構築し，授業や個別学習を現場で支援」[8]する工夫が求められる。ここでは、日本で教育を受けている中高生や大学生、あるいはそれらを終えた成人が英語学習を目的として、映像音声を利用する場合を想定する。

4.1. 語学学習に活用するコーパスの要件

語学学習に活用するコーパスに要求される条件を表 3 に示す。

表 3. 語学学習コーパスの要件

1.	ストーリーがあること
2.	音声があること
3.	書き起こしと、翻訳の 2 つの字幕があること
4.	翻訳字幕は、フレーズ訳が望ましい
5.	映像があることが望ましい
6.	収録作品ごとの適応学年があること
7.	字幕ごとの適応学年があること
8.	適応学年にはリーダビリティと語彙レベルの 2 本立てが望ましい

以下に表 3 の要件について順に解題を列挙する。

1. 収録作品は、映画のように 1 時間を越えるもの、スピーチのように 15 分前後のものから、1 センテンス、あるいはフラッシュカードのような 1 単語のものまである。映画やスピーチであれば当然ストーリーが備わっているが、1 センテンスや 1 単語であってもストーリーを備えたい。これは、運用可能な言語を習得するには意味、特に場面の意味が不可欠であることによる。例えば、文法コーパス[9]は、短文であっても意味理解が容易であるように工夫されている。
2. 日常言語の基本は音声であり、音声とテキストが具備されていることは記憶の定着に効果が期待できる。また黙読時の内声を補助する役割も期待できる。
3. 習得途上の学習者でもテキストの意味を理解しつつ学び続けるには母語による補助が必須である。また聞こえた音声を正しく言語として認知するためには、音声の書き起こしが必須となる。
4. 聞こえている連続音声（一般には呼気段落で 2 秒前後）の意味処理を同時に行えるようにすることは学習効果を高め定着を促進すると考えられる。そのためには、フレーズ訳が最適である。英語と日本語のように語順が逆転しやすい言語にあっては困難が伴いやすいフレーズ訳の作成ではあるが、語学学習用のコーパス構築を目的とするのならば、フレーズ訳に努めたい。これは聞こえた順に意味理解を可能にする手法でもあり、リスニングや同時通訳の訓練にもなる。
5. 音声をテキストにすると抜け落ちる情報を、短時間（ほぼ一瞬）に把握するには映像が最適である。映像には、言語化しにくい人間関係や感情が態度や表情から把握できるし、背景や時間、年齢や社会関係も服装や小道具や舞台装置から推察できる。これらは聴解の助けになる。
6. ビデオ鑑賞のような場合には、作品ごとの適応学年が有効である。
7. 文法や語句などの表現学習には短文が適しているので、フレーズごとの適応学年が必要となる。字幕ごとの適応学年があれば、作品全体としては難易度が高くても、部分としての活用機会が増大する。
8. リーダビリティは作動記憶の生理的制約に関わる評価値であり、語彙レベルは作動記憶の意味処理に関わる評価値であるので、使い分ける必要がある。例えば、文法や語彙表現を学ぶには語彙レベルの低いものを選び、語彙学習をするには簡単な表現のものを選ぶのが望ましい。

4.2. 語学学習用検索エンジンの要件

次に、コーパスの検索エンジンに要求される条件を表 4 に示す。

表 4. 語学学習用検索エンジンの要件

1.	音声再生に合わせて字幕が表示されること
2.	字幕は音声の呼気段落と同期していること
3.	書き起こし字幕は一覧表示が望ましい
4.	書き起こし字幕は音声に同期して強調表示されること
5.	字幕単位で音映像の反復再生ができること
6.	任意の連続字幕をまとめて段落にできること
7.	段落単位で音映像の反復再生ができること
8.	クローズテストなど簡単な自動生成ドリルがあることが望ましい

以下に表 4 の要件について順に解題を列挙する。

1. 聞こえている音声と字幕がずれていると生理的

な違和感が生じやすい。

2. 一般に字幕には字数制限があり、一度に読み取れる分量になっている。また、1 本の字幕は意味のまとまりでもある。この性質は呼気段落の性質（2 秒程度の意味のまとまり）と同じであることから、字幕と音声が同期していると学習者の負担が少ない。
3. 音声も字幕も消えては現れる一過性を特徴としているが、学習的観点からはテキストの前後を閲覧できる方が「今の音」の理解を促進させる。これは、黙読の速度は音声の速度より速いからである。
4. 視覚的な注意資源を移動させたときでも、一覧の中からすぐに「今の音」の位置に戻せるからである。
5. 記憶の定着には反復学習が必要であり、作動記憶の時間制限に合致している 2 秒程度の分量ごとに作業することが適しているからである。多くの場合、初見の表現でも無理なく復唱できる。
6. 意味が完結している最小単位が文であり、文は通常 2 ないし 4 個の句（字幕）で構成されているからである。
7. 1 文、あるいは数個の文（段落）での学習が必要となるからである。例えば映画のスキットや、プレゼンでの少しまとまった言い回しでは十数秒から数十秒の段落となることが多い。
8. 音映像の視聴だけでは学習が単調になる場合もあるので、一部の情報を隠したドリルがあると、気分が変わる。また習熟度の確認にも使える。

5. コーパス・検索エンジンへの応用と実装

最後に語学学習用検索エンジンへの実装について述べる。実例としては筆者が開発した 2 つのコーパスである日英対訳コーパス CORPORA と TED ビデオコーパス selected360 を取り上げる。ここで紹介するコーパスはオープンサイト（無料・無登録・無制限）でスマホでもすぐにアクセスできるように QR コードを付けておく。

5.1. 日英対訳コーパス CORPORA

日英対訳コーパス CORPORA は会話とスピーチの 2 つのコーパスを実装している。会話は 24 本の映画、スピーチは 2,439 本の TED Talks から構成される。英語の語句で検索し、語句を含む字幕のシーンがヒットする。ヒットしたシーンはヒット字幕の前後 20 秒程度のスキットとして音映像の視聴が可能となる。

ヒットした字幕それぞれについて、文レベル（リーダビリティ）と語彙レベルで中学・高校・大学に分類されている。学習者のレベルに応じて難易度を選べる仕組みになっている。

get away で検索した様子を図 7 に示す

図 7. get away で検索した様子と QR コード
URL： http://www.mintap.com/talkies/pac/corpora.html

5.2. TED ビデオコーパス selected360

TED ビデオコーパス selected360 は TED Talk のなかから人気の高い上位 360 本を検索対象とした特定分野コーパスである。フィルターとして、文レベル・語彙レベル・しゃべる速さ・ビデオ長・ジャンルなどがある。文レベルは中学 2 年から大学 3 年までの 8 段階、語彙レベルは中学 3 年から大学 2 年までの 6 段階に分かれている。

文レベルを高校 2 年としたときの様子を図 8 に示す。

図 8. 文レベルを高 2 としたときの様子と QR コード
URL： http://www.mintap.com/talkies/?selected360

selected360 でヒットしたビデオには、等高線図が付いている。これは TED Talks の中でのそのビデオの相対的難易度を視覚化したものである。横軸が文レベル、縦軸が語彙レベル、背景の山（等高線）が Talks の度数分布を示している。丸印は対象としているビデオの相対的難易度位置を示す。一例を図 9 に示す。丸印が山頂の左上にあることから、文レベルはやや低く語彙レベルがやや高いことがわかる。

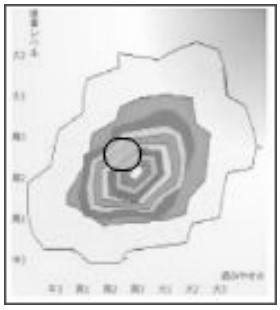

図 9. あるビデオの相対的難易度を示す等高線図とコメント（右）

6. まとめ

今回の報告は、自然言語の音声言語としての原点に立ち戻り、文字言語化により失われたかに見える時間情報を再現する理論と技術と応用に焦点を当てたものである。

自然言語を解析するにあたり、入れ物（container）と内容物（content）に概念区分した場合、入れ物は解析単位であり、内容物は解析対象と考えられる。英語を例に取ると、テキストを空白で区切り、英字列を抽出する作業は、処理単位（入れ物）を単語（word）としたことを意味する。この場合の内容物は英字列の特殊な並び方（token）であり、その性質や近隣内容物との関係を解析する作業（形態素・構文解析）である。この方法は長年研究され成果を挙げてきた。

今回の報告では、空白の代わりに句読点でテキストを区切った。句読点でテキストを区切り、単語列を抽出する作業は、処理単位を句（phrase）とし、内容物（words）から時間情報（音韻符号化処理時間）を計算するとともにいくつかの句を結合して文（sentence）とする構造化作業を通して、テキストの読みやすさ指標（readability grade level）を算出して活用したことになる。

こうした作業を認知科学的視点から見ると、形態素・構文解析と単語処理は、長期記憶とアクセスしながらの作動記憶（中央実行系）の意味処理（の緻密化）の過程（図 10 の B 破線）に相当し、読みやすさ指標は、入力刺激（文字）を音声変換（音韻符号化）して中央実行系での意味処理に渡していく過程（図 10 の A 実線）に相当すると考えられる。この過程の生理的制約（時間［2±1 秒］と容量［4±1 個］の制約）とテキストとの関係性の指標化とも言えるだろう。近年研究が盛んになったアノテーションは長期記憶（背景知識、言語知識）を総動員して意味理解を正確にする過程（図 10 の C 破線）に相当すると思量される。

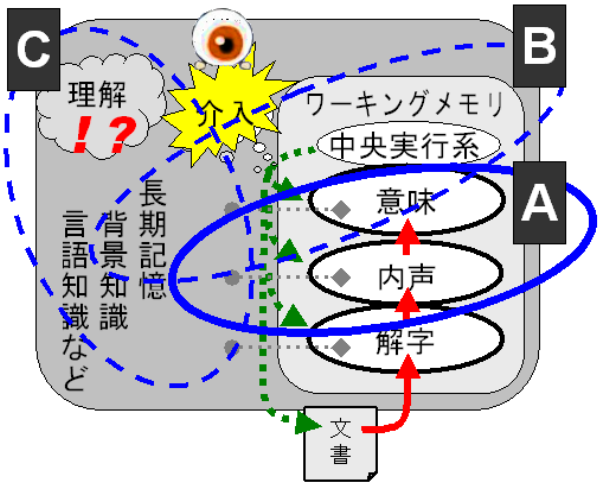

図 10. 読解プロセスとテキストアナリティクス

今回の研究報告を契機に、テキストアナリティクスの幅が広がり、AI 開発の一助となることを期待している。

文　献

[1] 湯舟英一，田淵龍二，“映画音声コーパスを利用した Breath Group 長の分析，”Language Education & Technology, 50, 23–41，2013.

[2] 服部四郎，音声学，岩波書店，1950

[3] 田淵龍二，“ワーキングメモリ理論と一息の連続音声継続時間に基づいた Direct Method としてのチャンク音声提示法の提案，”言語教育エキスポ 2015, 87-88. 2015.

[4] Card, S. K., Moran, T. P., & Newell, A.: The psychology of human-computerinteraction.Lawrence Erlbaum Associates, Inc.，1983.

[5] 田淵龍二，湯舟英一，“1 次的読解速度予測に基づく日本人英語学習者向けリーダビリティ公式とその教育的示唆，”外国語教育メディア学会 LET 関東支部第 132 回（2014 年度）研究大会，発表要項, 12-13. 2014.

[6] 田淵龍二，湯舟英一，“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, 359–388，2015.

[7] 田淵龍二，“TED 字幕表示時間と音韻符号化予測時間の比較研究，”Language Education & Technology, 54. 41−54. 2017.

[8] 田淵龍二，“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，”言語処理学会第 24 回年次大会発表論文集 2018_E2-1, 360-363. 2018.

[9] 中條清美，内山将夫，赤瀬川史朗，西垣知佳子，“データ駆動型英語学習における教育用例文コーパス SCoRE の活用，”言語処理学会第 22 回年次大会発表論文集 2016_P19-3, 1081-1084. 2016.