

音変化に着目した英語ディクテーション 学習支援システムのための音声合成機能の実現

川本達也[†] 國近秀信^{††} 竹内 章^{††} Robert Chartrand^{†††}

初学者が英語を聞き取る際に難しく感じる原因の一つに、音変化が挙げられる。音変化とは、英単語が連なって発声される際に、音が変化する現象のことである。我々は、英語の音変化を学習するためのディクテーション学習支援システムの実現を目指している。本研究では、そのために必要な音変化に対応した音声合成機能の実現を行った。本機能は、英文を発音記号へ変換し、音変化規則に従って音変化が生じる発音記号を置換し、音声合成エンジンを用いて音声出力する。本機能は、消失、弱化、連結、脱落、同化の他、不規則な音変化を対象とする。

Implementation of a Text-to-Speech Function for an English Dictation System for Learning Relaxed Pronunciation

Tatsuya Kawamoto[†] Hidenobu Kunichika^{††}
Akira Takeuchi^{††} and Robert Chartrand^{†††}

It is difficult for novice learners of English to listen English speech. One of the main reasons is that relaxed pronunciation almost always happens in normal speech. Relaxed pronunciation is a phenomenon that happens when words are slurred together. The aim of this project is to implement an English dictation system for learning relaxed pronunciation. In order to achieve the aim, we have implemented a text-to-speech function which generates speech including relaxed pronunciation. The function translates English sentences into phonetic symbols, replaces some phonetic symbols by referring to the rules of relaxed pronunciation, and generates speech by using a text-to-speech engine.

1. はじめに

英語リスニングを困難にする原因の一つとして、単語が文中で発音される際に生じる音変化に不慣れなことが挙げられる。ディクテーションは、リスニングスキル習得のための有効な学習法の一つであり、これまで多くのディクテーション学習システムが開発されている。しかし、それらの多くはあらかじめ用意された音声ファイルを再生し、音声出力するため、多様な英文テキストを利用した学習は困難である。

そこで我々は、多様な英文テキストを利用し、リスニング初学者が音変化した英文を聞いて書き取るディクテーション学習支援システムの実現を目指している。本研究は、そのために必要な音変化に対応した音声合成機能を実現することが目的である。以下、第2章では関連研究について述べ、第3章では音変化について述べる。その後、第4章で本システムの概要を説明し、第5章で音声合成機能の実現について述べる。最後に、第6章にて今後の展望について述べる。

2. 関連研究

2.1 音声合成

音声合成とは、人の声を人工的に作り出す技術のことであり、録音編集方式および規則合成方式へ大別される[3]。録音編集方式は、人間の音声を、単語やフレーズ単位で録音しデータベースに蓄積、その音声単位をつなぎあわせて合成音を生成する技術である。実際に人間の音声を録音するため高品質の音声を得られる反面、言語情報の種類が比較的少なく、合成できるメッセージが限定されるという制約がある。また新たに単語やフレーズを追加する場合、同じ発話者が録音しなければならないため、データベースの構築には大きな手間がかかるといえる。一方、規則合成方式は、入力された文字から、音声学的な規則によって音声を合成する。この方式では、合成のために必要となる記憶容量を大幅に削減することができる。任意のテキストを合成対象とするため、新たに単語やフレーズを追加する場合にも音声収録は不要だが、肉声感に乏しいという欠点もある。本研究では、教材作成者が多様なテキストを用意し、それを用いて学習できるようにするため、規則合成方式を利用する。

[†]九州工業大学大学院情報工学府

Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology

^{††}九州工業大学大学院情報工学府

Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology

^{†††}久留米大学外国語教育研究所

Institute of Foreign Language Education, Kurume University

2.2 ディクテーション学習支援システム

これまでに多くのディクテーション学習支援システムが開発されている（例えば、[2][5]）。しかし、既存のディクテーション学習支援システムは、正誤判定の正確さや学習者のレベルに合わせたヒントの提示などに焦点を当てており、英語の音変化に焦点をあてた物はほとんどないのが現状である。

音声の再生方式には、あらかじめ用意した音声ファイルを再生するものが多い。しかし、音変化の規則は複雑で数も多く、また学習者の理解状態を考慮して出題することを考えると、音声ファイルを用意することは大変困難である。さらに、問題作成者が新たな問題を追加するたびにその音声ファイルを用意しなければならないという問題がある。

3. 英語の音変化

英語の音変化には、規則的に変化するものとして、消失 (elision)、弱化 (reduction)、短縮 (contraction)、連結 (linking)、脱落 (deletion)、同化 (assimilation) がある[9]。また、不規則な音変化も存在する。以下、各音変化について説明する。

3.1 消失

強勢がなく弱く発音される音節において、音韻の一部が省略される。

- (1) 強勢のある音節の後に強勢のない音節が連なると、強勢のある音節直後の母音 [ŋ] が省略される。
- (2) 調音点や調音方法が同じか近い子音が連続すると、先頭の子音が省略される。
 - (a) 同じ音が連続する場合
 - (b) 破裂音 [p][t][k][b][d][g] と破裂音が連続する場合
 - (c) 破裂音に鼻音 [m][n][ŋ] が続く場合

3.2 弱化

強勢がなく弱く発音される機能語において、母音があいまい母音となる。

- (1) 機能語の母音は曖昧母音となる。機能語とは人称代名詞（主格、目的格、所有格）、助動詞、前置詞、冠詞、接続詞、関係代名詞のことを指す。
- (2) ストレスがなく語末の文字が y, ay, ey のとき、[i:] [ei] [ai] の音素が弱い音 [i] となる。

3.3 短縮

特定の単語と単語が組み合わされるとき、一方の単語の一部が省略され、他方と結合して短くなる。例えば、”I am” は、”I’m” と短縮される。

- (1) 主語と be 動詞の結合では、be 動詞が短縮される。
- (2) 主語と have 動詞の結合では、have 動詞が短縮される。
- (3) 主語と助動詞の結合では、助動詞が短縮される。
- (4) 疑問詞と be 動詞、have 動詞あるいは助動詞の結合では、be 動詞、have 動詞あるいは助動詞が短縮される。
- (5) be 動詞、have 動詞あるいは助動詞と “not” の結合では、“not” が短縮される。

3.4 連結

先行する単語の語尾の子音と、後続する単語の語頭の母音が 1 つのまとまりとして発音される。

- (1) 語尾の [n] と語頭の母音が連続する場合、連結が生じる。
- (2) 語尾の [r] と語頭の母音が連続する場合、連結が生じる。
- (3) 後続する単語が、不定冠詞 (a, an)、代名詞 (it)、前置詞 (in, of, on, up, at など)、接続詞 (and) である場合、連結が生じる。

3.5 脱落

消失が語中における省略であったのに対し、脱落は連続する単語が結合する際に音が省略される現象である。

- (1) 語尾と語頭で同じ音韻が連続する場合、脱落が生じる。
- (2) 語尾が破裂音 [p][t][k][b][d][g] で、後続する語の語頭が破裂音、破擦音 [tʃ][dʒ][dʒ]、摩擦音 [f][v][θ][ð][s][z][ʃ][ʒ]、鼻音 [m][n][ŋ] または側音 [l] の場合に脱落が生じる。
- (3) 有声/無声の違いはあるが発音方法の同じ音韻が連続する場合、脱落が生じる。

3.6 同化

連続する前後の音韻の影響を受けて別の音韻に変化することである。

- (1) 進行同化
進行同化には、無声化、有声化、鼻音化がある。
 - (a) 無声化
“-s” は先行する [k] の影響を受けて無声化し、[s] となる。
“-ed” は先行する [k] の影響を受けて無声化し、[t] となる。
 - (b) 有声化
“-s” は先行する [v] の影響を受けて有声化し、[z] となる。
“-ed” は先行する [v] の影響を受けて有声化し、[d] となる。
 - (c) 鼻音化
“-t-” は脱落で [t] となり、さらに [n] の影響を受けて鼻音化する。
- (2) 逆行同化
逆行同化には、進行同化の無声化、有声化、鼻音化に加え、両唇音化、硬口蓋音化がある。

- (a) 無声化
“-s”は後続する[t]の影響を受けて無声化し、[s̥]となる。
 - (b) 有声化
形容詞の“-th”は、後続する母音の影響を受けて有声化し、[ð]となる。
 - (c) 両唇音化
“-n”は後続する両唇音[b]の影響を受けて両唇音化し、[m]となる。
 - (d) 硬口蓋音化
“-s”は後続する硬口蓋音[j]の影響を受けて硬口蓋音化し、[ʃ]となる。
- (3) 相互同化
相互同化が生じる条件は、以下の4つである。
- (a) [t]と[j]が同化して[tʃ]となる。
 - (b) [d]と[j]が同化して[dʒ]となる。
 - (c) [s]と[j]が同化して[ʃ]となる。
 - (d) [z]と[j]が同化して[ʒ]となる。

3.7 不規則変化

代表的な音変化の中には、前述の音変化規則では対応することができない物が存在する。例えば、“have to”は“hafta”と発音され、“is it”は“zit”と発音される。

4. 英語ディクテーション学習支援システム

我々はこれまでに英語の音変化に着目した英語ディクテーションシステムを試作した[1]。本章では、本システムの概要を述べる。

4.1 システム構成

図1に、本システムの構成図を示す。本システムは、発音記号変換部、音声処理部、正誤判定部、弱点発見部、オーサリングツールおよび問題選定部から成る。まず、教材作成者がオーサリングツールを利用して問題データを作成する。その後、本システムは学習者モデルを参照しながら問題を選定し、音声出力による出題、正誤判定および弱点発見を繰り返す。以下、オーサリングツールおよびディクテーション機能について説明する。

4.2 オーサリングツール

教材作成者が準備した英文テキストを読み込み、各英文について次の処理を行う。まず、発音辞書を参照して、英文を発音記号へ変換する。次に、音変化規則が適用可能か否かを検査し、もし適用可能な場合は、全ての適用可能な音変化規則識別子と単語の位置及び当該英文テキスト問題データへ記録する。

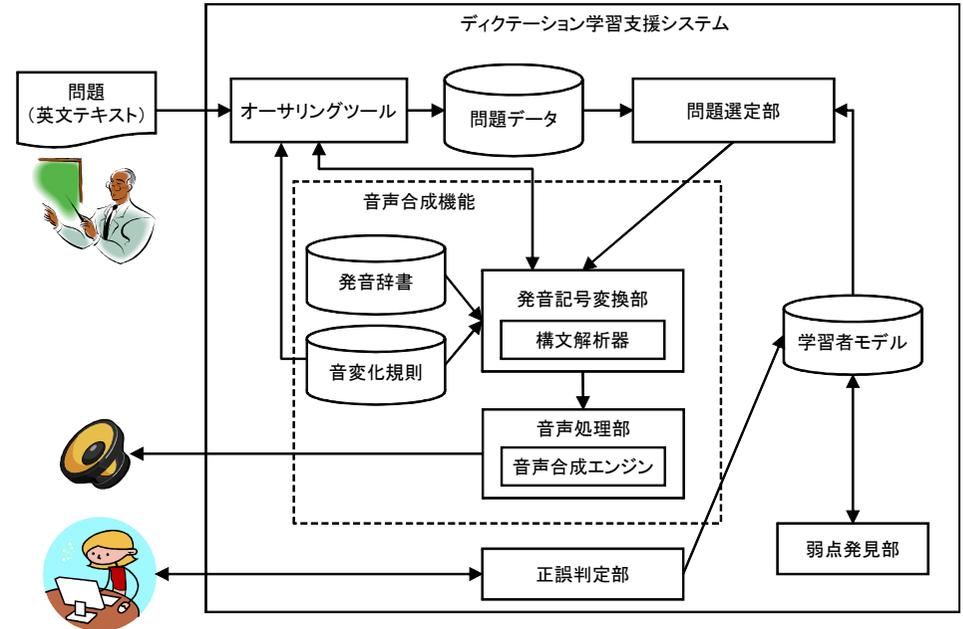


図1 システムの構成

4.3 ディクテーション機能

ディクテーションは、まず音声を聞き、次に聞き取った文を書き取り、最後に正確に書き取れたか否かを確認するという流れで行われる。

(1) 問題選定

問題は10問を1セットとして出題される。本機能は、後述の弱点発見部で弱点と判断された音変化を含む問題をランダムに5問選択し、その他の音変化が起こる問題をランダムに5問選択する。

(2) 音声出力

まず、出題する英文を単語毎に分割し、発音辞書を参照して発音記号へ変換する。その後、音変化規則に従って音変化が生じる部分を置換する。この発音記号を音声

合成エンジンに渡すことにより、音変化規則に従った音声を出力する。詳しくは、第5章で述べる。

(3) 正誤判定

2段階 DP マッチング法[5]を用いて、学習者が入力した英文と正解英文とを比較し、文字レベルおよび単語レベルで、正解（一致）または不正解（欠落、置換、過多）を判定する。その結果を学習者へ提示するとともに、学習者モデルへ記録する。もし不正解と判定された場合は、誤りの場所、種類および音変化規則を学習者へ提示し、学習者モデルへ記録する。例えば、正解が”I want to beat that team”で学習者が入力した英文が、”I wana be that team”であった場合、図2に示すメッセージを出力する。

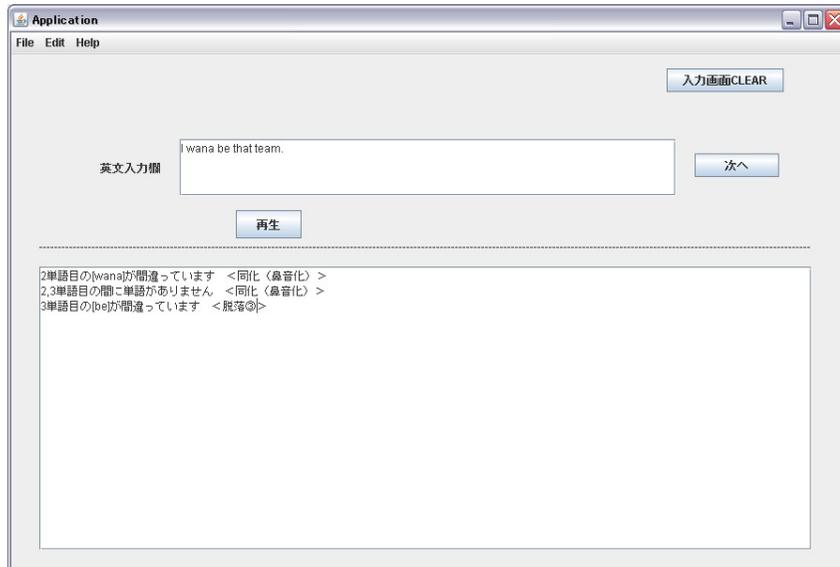


図2 学習インターフェースの例

(4) 弱点発見

学習者モデルを参照し、各音変化規則の誤り頻度を求める。続いて、最も誤り頻

度が高い音変化規則を弱点と同定する。最後に、その結果を学習者モデルへ記録し、(1) 問題選定へ戻る。なお、本ステップで同定する弱点は、音素の種類により分類した音変化規則のいずれかである。

5. 音声合成機能

問題の英文を入力とし、音変化した音声を出力する機能である。本機能は、まず出題する英文を単語毎に分割し、辞書データを参照して発音記号に変換する。また、音変化規則の中には品詞を必要とする物も含まれるため、Stanford parser[7]を用いて英文を解析する。その後、音変化規則に従って音変化が生じる部分を置換する。この発音記号を音声合成エンジンへ渡すことにより、音変化規則に従った音声を出力する。

5.1 発音辞書

発音記号へ変換のために発音辞書として The CMU Pronouncing Dictionary[6]を使用した。辞書データは、英単語とその発音表記が対となっているテキストファイルである。The CMU Pronouncing Dictionary の発音記号の対応表を表1に示す。本辞書においてアルファベットの後の数字はストレスを意味し、0, 1, 2 はそれぞれ No stress, Primary stress, Secondary stress を表す。

5.2 発音記号変換部

問題の英文を音変化させた音声として出力するために、英文から音変化後の発音記号を生成するモジュールである。手順としては、まず、問題の英文から発音記号への変換を行う。次に、音変化規則に従い発音記号を置換する。

5.2.1 発音記号への変換

問題の英文を読み込み、ピリオド、コンマ、クエスチョンマーク、エクスクラメーションマークなどを英文から取り除く。The CMU Pronouncing Dictionary では単語が全て大文字で登録されているため、英文を大文字へ変換する。その後、英文を単語ごとに分割し、発音辞書と比較し同じ単語が見つかった場合、発音記号へ置き換える。これにより、英文データを発音記号へ変換することが出来る。本システムでは、音素と音素の区切りを半角スペース1つで表し、単語と単語の区切りを半角スペース2つで表す。以下に例を示す。

英文: I want to beat that team.
発音記号: AY1 WAA1 NT TAH1 BIY1 T DHAE1 T TIY1 M

表 1 The CMU Pronouncing Dictionary の発音表記

Phoneme	Example	Translation	Phoneme	Example	Translation
AA	Odd	AA D	L	lee	L IY
AE	At	AE T	M	me	M IY
AH	Hut	HH AH T	N	knee	N IY
AO	Ought	AO T	NG	ping	P IH NG
AW	Cow	K AW	OW	oat	OW T
AY	Hide	HH AY D	OY	toy	T OY
B	Be	B IY	P	pee	P IY
CH	Cheese	CH IY Z	R	read	R IY D
D	Dee	D IY	S	sea	S IY
DH	Thee	DH IY	SH	she	SH IY
EH	Ed	EH D	T	tea	T IY
ER	Hurt	HH ER T	TH	theta	THEY T
EY	Ate	EY T	UH	hood	HH UH D
F	Fee	F IY	UW	two	T UW
G	Green	G R IY N	V	vee	V IY
HH	He	HH IY	W	we	W IY
IH	It	IH T	Y	yield	Y IY L D
IY	Eat	IY T	Z	zee	Z IY
JH	Gee	JH IY	ZH	seizure	S IY ZH ER
K	Key	K IY			

5.2.2 音変化

本節では、音変化の法則に従った発音記号へと変換する処理について述べる。本研究では、第3章で述べた音変化の規則を計算機で処理できるよう実現した。ただし、短縮については、発音辞書にあらかじめ短縮形で登録されているため、本機能の処理対象から除外した。短縮については、教材作成者が短縮形を用いた英文を入力することで対応する。本機能では、不自然な音変化を避けるため、まず不規則な音変化を適用し、その後、弱化、消失、脱落、連結、同化の順で規則を適用する。以下、各音変化規則について説明する。

(1) 消失

消失(1)「強勢のある音節直後の母音[a]が省略される」の処理の流れを述べる。まず、生成された発音記号を音素ごとに分割する。強勢のある音素には最後に1が付与されているため、1が付与された音素が見つかった後、母音が見つかるか否かを判別する。母音が見つかった場合、その音素が[a]を表す“AH0”ならば省略する。

消失のもう一つの規則である「調音点や調音方法が同じか近い子音が連続すると、先頭の子音が省略される」について説明する。消失(1)と同様に、まず単語を音素ごとに分割する。消失(2)-(a)は、子音とその次の子音が同じ音素であった場合に一つ目の音素を削除する。消失(2)-(b)は、破裂音と破裂音が連続した場合に一つ目の音素を削除する。消失(2)-(c)は破裂音の次に鼻音が続いた場合に破裂音を消去する。以下に消失(1)の例を示す。

英文： station
 変化前の発音記号： S T EY1 SH AH0 N
 変化後の発音記号： S T EY1 SH N

(2) 弱化

まず、構文解析の出力結果から品詞情報を取り出して、単語が機能語か否かを判別する。もし、機能語であるならば弱化を適用し、母音を曖昧母音に変化させる。この処理が弱化(1)にあたる。また、弱化(2)はある英単語が、語末かつ y, ay, ey で終わるとき、y, ay, ey の部分の音素を弱い[i]の音へ置換する。以下に弱化(1)の例を示す。

英文： of
 変化前の発音記号： AH V
 変化後の発音記号： AH0 V

(3) 連結

連結が起こった際は、発音時に1つのプレスで発話されるように2単語を1単語へまとめる。具体的には、単語間にある2つのスペースを1つのスペースへと置換する。連結(1)は、語尾の[n]と語頭の母音が連続する場合に連結する。連結(2)は語尾の[r]と語頭の母音が連続する場合に連結を行う。連結(3)は、後続する単語が不定冠詞、代名詞、前置詞、接続詞である場合に連結する。以下に連結(1)の例を示す。

英文： in an hour
 変化前の発音記号： IH0 N AE1 N AW1 ER0
 変化後の発音記号： IH0 N AE1 N AW1 ER0

(4) 脱落

脱落(1)は、語尾と語頭で同じ音韻が連続する時に、語尾の方の音素を削除する。脱

落(2)は、語尾が破裂音で、後続する語の語頭が破裂音、破擦音、摩擦音、鼻音、側音の場合に語尾の破裂音を削除する。脱落(3)は、語尾と語頭が、発音方法の同じ音韻のであるか確かめる。その場合、語尾の音素を削除する。以下に脱落(1)の例を示す。

英文： good day
変化前の発音記号： G UH1 D D EY1
変化後の発音記号： G UH1 D EY1

(5) 同化

現在処理可能な同化規則は、同化(1)-(c)、同化(2)-(a)、同化(2)-(c)、同化(2)-(d)および同化(3)の全てである。本節では、一部のどうか規則について説明する。同化(1)-(c)は、[鼻音 T]の場合に[鼻音]へ置換する。同化(2)-(a)は、英単語が”s”で終わり、次の単語が[T]から始まるならば[S]へ置換する。同化(3)-(a)は[T Y]の場合に[CH]へ置換する。以下に同化(1)-(c)の例を示す。

英文： want to
変化前の発音記号： W AA1 N T T AH0
変化後の発音記号： W AA1 N AH0

(6) 不規則変化

本研究では、Wikipedia の Relaxed pronunciation[8]に記載されている 37 種類の不規則変化を参照し、不規則変化テーブルを作成した。本システムは、英文中に特定の単語の並びが存在するか否かを調べ、存在した場合は不規則変化テーブルを参照し、当該単語に相当する発音記号を変化後の発音記号へ置換する。例えば、英文中に should have が存在する場合、should have の発音記号を shoulda の発音記号へ置換する。以下に例を示す。

英文： He should have arrived.
変化前の発音記号： HH IY1 SH UH1 D HH AE1 V ER0 AY1 V D
変化後の発音記号： HH IY1 SH UH1 D AH0 ER0 AY1 V D

5.3 音声処理部

音声処理部は、発音記号変換部で生成された発音記号を受け取り、音声合成エンジン用の発音記号へ置換し、音声合成エンジンを用いて音声を出力するモジュールである。現在は、音声合成エンジンとして Microsoft Windows 7 の音声合成エンジン[4]を利用している。

Microsoft Windows 7 の音声合成エンジン用の Speech API と The CMU Pronouncing Dictionary とでは使用されている発音記号が一部異なるため、置換が必要である。具体的には、以下の処理を行う。

- ”AH0”, ”HH”をそれぞれ”AX”, ”H”へ置換する。
- ストレスを表す”1”および”2”については、音素との間に半角スペースを入れる。
- No Stress を表す”0”は使用できないため削除する。
- 単語と単語の区切りは”&”へ置換する。
- <PRON SYM = “”>のダブルクォーテーションの間へ発音記号を挿入する。

6. おわりに

本研究では、英語の音変化を学習するためのディクテーションシステムに必要な音声合成機能を実現した。今後は、より自然な音声となるよう改良するとともに、学習支援機能を充実させる予定である。

参考文献

- (1) 有松友幸, 國近秀信, 竹内章: 音変化に着目した英語ディクテーション学習支援システムの実現, 2010 年春 JSiSE 学生研究発表会, pp.259-260 (2010).
- (2) 東原義訓, 橋本直子, 大木吉功, 中山和彦: 音変化して発音される英語の聞き取りのためのコースウェアの開発と効果, 日本教育工学会大会講演論文集, Vol.11, pp. 579-580 (1995).
- (3) 渡辺聡, 岩木健, 兼安勉, 三木敬: コーパスベース音声合成とその応用, 沖テクニカルレビュー, Vol.73, No.2, pp.62-65 (2006).
- (4) Microsoft Speech API (SAPI) 5.4, [http://msdn.microsoft.com/en-us/library/ee125663\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/ee125663(VS.85).aspx)
- (5) 望月源, 神谷泰弘, 奥村学, 島津明: 英語ディクテーション学習支援システムの構築, 教育システム情報学会誌, Vol.19, No.3, pp. 145-153 (2002).
- (6) The CMU Pronouncing Dictionary, <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
- (7) The Stanford Natural Language Processing Group, <http://nlp.stanford.edu/software/lex-parser.shtml>
- (8) Wikipedia Relaxed pronunciation, http://en.wikipedia.org/wiki/Relaxed_pronunciation
- (9) 山田恒夫, 足立隆弘, 英語リスニング科学の上達法, 講談社 (1998).