

字幕・副音声付きテレビニュース放送を利用可能な 語学学習教材作成システムとリスニング教材プレイヤー[†]

田中敬志・小林聡・中川聖一*
豊橋技術科学大学*

これまでに多くの語学学習教材が開発されてきているが、その製作においては多大な時間と労力がかけている。また、これまでの語学学習教材では、そのコンテンツが限られているため、長時間飽きずに使用を継続することは困難であり、さらに学習者が興味のあるコンテンツを選択する余地はない。そこで本研究では、学習者に興味のある TV ニュース放送をパソコンに取り込み、そのニュース放送を素材とした学習教材を教師または学習者自身で手軽に作成することができるシステムを開発した。勿論、本システムは教師や学習者が自前で収録したビデオを語学学習教材化することもできる。本論文では特に、副音声と字幕の同期手法を取り入れた教材作成システムと教材プレイヤーによって実現される各機能とシステム全体の評価について述べる。教材プレイヤーを用いた被験者実験では学習効果が得られ、それらの評価アンケートでも教材について肯定的な意見が得られた。

キーワード：CALL, 語学学習教材, 日本語学習, 英語学習, リスニング, TV ニュース, 字幕, 音声処理, 評価

1. ま え が き

語学学習には繰り返し学習が必要であるが(DARWING 1989), 既存の教材では話題が限定されている。そのため、学習者が興味を持ち続けられるか否かという問題がある。それに対し、より広い話題や時期に即した話題に触れられるならば、学習者が外国語学習に興味を持ち続けることが可能になると考えられる。しかし、適当な話題のニュース放送などを用いて、学習者個人や語学教師が語学学習教材を作成するには限界があり、また、幅広い話題の内容を教材として利用するためには手間がかかるので、教材を半自動的に手軽に作成できるツールが必要となる。

本論文においては、字幕付き TV ニュース放送を元

に、自然な発話による豊富な話題の提供、時期に即した話題の提供、音声と字幕との同期、字幕への品詞付与、読みの付与という機能を持つ語学学習教材作成システム(小林・田中・森・中川 2002)を土台として、今回、学習者にとって有用な翻訳音声(副音声)の利用機能、問題作成支援機能、学習者管理機能を追加した。また、本教材作成システムによって得られる教材用データを利用し、音声と字幕の同期表示、教材の繰り返し視聴、単語の語義表示、聞き取りテストの作成など、市販の CALL ソフトウェアの機能を強化したリスニング教材プレイヤーについて述べる。調査した CALL ソフトウェアについては(豊橋技術科学大学 中川研究室)に特徴等を挙げてある。教材プレイヤーについては、Web 技術を利用して学習者の操作記録を保存できる学習者管理機能を追加した。本教材作成システムおよび教材プレイヤーでは、日本語学習用教材、英語学習用教材ともに作成・利用することができる。本システムのように音声情報処理技術と自然言語処理技術を利用した語学学習教材の自動的な作成を目指した研究はこれまでにない。このようなシステム、あるいは手法を開発したところに本研究の意義がある。また、要素技

2003年2月12日受理

[†] Takashi TANAKA, Satoshi KOBAYASHI and Seichi NAKAGAWA* : CALL Contents Creator and CALL Player Using TV News with Closed Caption and Subchannel

* Toyohashi University of Technology, 1-1, Hibarigaoka, Tempaku, Toyohashi, Aichi, 441-8580

術の発達により、本システムを構成するコンポーネントを入れ替えることにより、提案システムの性能の向上が見込まれる。

最後に、教材プレイヤーを2ヶ月間、16名の被験者に使用してもらい、学習期間の前後にテストを行なうことで学習教材としての評価実験を行ない、実験後にアンケートを取った。また、英語の1クラスの受講者全員に本システムの使用に関するアンケート調査を行なった。

2. 教材作成システム

2.1. 教材作成手順の概要

(小林・田中・森・中川 2002)

現在、NHKの「ニュース7」では音声認識技術を用いてアナウンサーの発話内容の字幕が数秒遅れて字幕放送として送られてくる(安藤ほか 2001)。また、他の番組においても字幕放送やWebを通して字幕を取得できる番組が増えてきている。さらに字幕を用いた教材として、映画の字幕を用いた研究もされている(五十里・佐野・緒方・有木 2001)。本研究で構築しているシステムは、このTVニュース放送の音声と字幕から(半)自動的に、語学のリスニング教材を作成しようというものである。教材作成システムの処理の流れを以下に示す。

- (1) 2枚のTVチューナーボードを使用し、MPEG1形式のビデオと字幕のテキストをPCに取り込む。英語版の場合はWebから取得する¹⁾。
- (2) 字幕の形態素解析を「茶筌」(奈良先端科学技術大学院大学 自然言語処理学講座)を用いて行なう。また、形態素解析の結果の確認・修正を行なう。英語版では「Brill's Tagger」(BRILL 1995)を使用して品詞解析を、読み付与は「CMU Dictionary」(CMU)を使用して行なう。
- (3) 手順(2)の形態素解析の結果として得られる読みを利用し、音節(音素)系列を作成する。
- (4) 録画から得られる音声进行分析し特徴パラメータ系列に変換する。
- (5) 音声認識技術(甲斐・中川 1997)を利用して、手順(3)で得た字幕の音節(音素)系列と音声の特徴パラメータ時系列との照合により、字幕と音声との同期情報を作成する。
- (6) 手順(2)で得た品詞、発音情報、および手順(5)で得た同期情報を、XMLのデータ形式を用いて字幕テキスト中に埋め込む。
- (7) 副音声を音声認識技術を利用して単語列に変換

し、これと字幕中の各単語を英語に変換したものを照合し、同期情報を作成する。

- (8) 最終的に、MPEG1形式の映像+音声ファイルを「QuickTime」形式に変換し、それと手順(6)で得られる様々な情報を含む字幕により、Javaで開発した専用のプレイヤーを用いて再生する語学学習教材となる。この語学学習教材の画面は図6(後述)のようになる。

字幕取得においてはプログラムの制約により字幕が正確に収録されていない、字幕中に発話されていない部分が存在する、などの問題点があるため、現在の教材作成システムは半自動となっているが、自動的な教材用データの作成の可能性を示せていると考えている。この部分の問題を解決すれば、技術的にはTVニュース放送から多数の教材が簡単に利用可能となり、他のCALLシステムの欠点といえる話題の制限を克服できると考えている。

本システムを用いて教材を作成する作業にかかる時間は、コンピュータの習熟度合いにもよるが、1、2分のニュース放送のデータであれば、ファイル形式変換や同期計算時間を含めて全体で15~30分程度、このうち教材作成者の作業時間は10~20分程度である。

本教材作成システムによって、教材用データとして、音声と字幕との同期情報、各単語への分割(日本語版)、各単語の品詞情報、各単語の読みの情報が得られる。これらの情報を用い、教材プレイヤーにおいては4節に述べるような機能が実現されている。このように、教材作成システムと教材用データ、教材プレイヤーは不可分のものとして教材を提供することになる。

2.2. 音声と字幕の同期

今回、語学学習教材作成システムとして、副音声と字幕との同期処理機能を追加した。ここでは従来からの主音声と字幕の同期処理と、今回新たに開発した副音声と字幕との同期処理について述べる。

2.2.1. 主音声と字幕の同期

(小林・田中・森・中川 2002)

現在、NHKによる字幕放送は、音声に対して数秒から十数秒の遅れを伴っている。英語ニュースの場合、発話と字幕はほぼ同時に提示されているが、字幕は画像情報である。Web上のニューススクリプトも利用可能であるが、発話との同期情報までは含まれていない。また、いずれの場合も、単語単位での同期情報は放送からは得られない。

字幕のかな文字列(あるいは音素記号列)からそれに

対応する音節(音素)単位 HMM 系列を構成し、HMM 系列と音声の特徴ベクトルを Viterbi アルゴリズムを用いて最適な時系列同士のマッチングを行ない、音声と字幕の同期情報を得ている。ただし、この際、無音区間をパワーの閾値により除く処理を行なっている。

特徴ベクトルの分析条件はサンプリングが12kHz、フレーム周期が8ミリ秒、LPC14次による分析であり、日本語版は LPC メルケプストラム10次元を4フレームまとめてKL展開で20次元に圧縮したものと Δ ケプストラムと $\Delta\Delta$ ケプストラム、 Δ パワー、 $\Delta\Delta$ パワーの計42次元、英語版は LPC メルケプストラム10次元とその Δ ケプストラムの20次元を使用している。

この方法により得られる同期の精度を正確に測定するため、各名詞の終端点での同期のずれを集計したところ、日本語の音声データで平均14ミリ秒(サンプル数39)、英語の音声データで平均92ミリ秒(サンプル数25)となった。英語の音声に対しては、背景音などの比較的大きなノイズが含まれるため(英語音声の S/N は16dB~20dB、日本語音声の S/N は30dB~33dB²⁾)、また音響モデルの精度が不十分(音響モデル作成のためのトレーニングデータ量が少ない)であることが、日本語音声の精度よりも悪くなる原因となっている。

2.2.2. 副音声と字幕の同期

2.2.2.1. TV ニュース放送の副音声

現在の TV ニュース放送では、2カ国語放送を行っているものがある。例えばNHKの「ニュース7」、「BS23」では、主音声日本語で、その対訳の英語音声副音声で放送され、また、BS等で放送されているPBS、CNN、ABC等では、主音声英語で、その対訳の日本語音声副音声で放送されている。

これら TV ニュース放送を数本ずつ録画し、日本語ニュース、英語ニュースそれぞれ10分程度について、主音声と副音声の関係を調べた結果、以下のようなことがわかった。

- ・主音声よりも早く、それに対応する副音声が発声されることがある
- ・主音声と副音声の文毎のずれは、平均で3秒程度である
- ・主音声と副音声は文単位で対応しており、英語1文に対して対訳日本語2文、また、英語2文に対して対訳日本語1文など、最小の対訳が複数文対複数文であることも多い(1対1の対応は全体の6割程度)
- ・主音声のうち、まれに副音声の対訳が付かないものがある

2.2.2.2. 字幕と副音声の対応付け

字幕と副音声の文単位での対応付けは図1のように行った。以下に、その手順を示す。以下では、英語ニュース(主音声：英語、副音声：日本語、字幕：英語)の場合を考える。

- (1) 字幕中の各単語を英和辞書で検索し、その単語を日本語の語彙(辞書の内容をそのまま利用)で置き換える。その際、前置詞は除外し、単語の原型でも検索を行っている。
- (2) 副音声を音声認識し、単語列に変換する。英語ニュース放送の副音声中には主音声も重畳されているため、副音声から主音声の成分をスペクトルサブトラクション(SS)を用いて除去した音声で音声認識を行った。

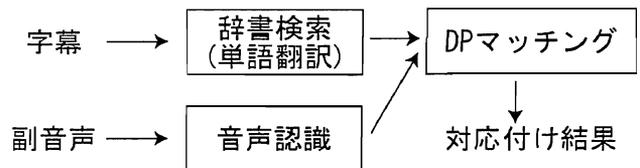


図1 字幕と副音声の対応付け

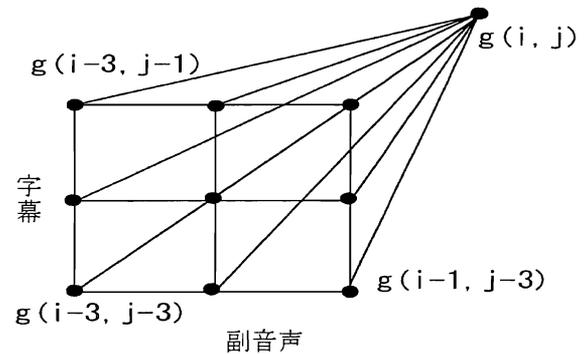


図2 DPパス

t を時間とし、主音声のスペクトラム系列を $X_{main}(t)$ 、副音声のスペクトラム系列を $X_{sub}(t)$ 、副音声の中の日本語のクリーンな音声を $X_{clean}(t)$ としたとき、英語ニュース放送の副音声のデータ系列は以下のように表すことができる。

$$X_{sub}(t) = X_{clean}(t) + m(t) \times X_{main}(t)$$

$m(t)$ は主音声の重みで、日本語の発話がない時には $m(t)$ は1.0程度であり、日本語の発話中は $m(t)$ が0.1程度に変化する。そのため、各フレーム毎に $m(t)$ を計算($X_{clean}(t)$ の各パワースペクトル成分が0以上となる $m(t)$ の値を利用して、スペクトルサブトラクションを行い、 $X_{clean}(t)$ を求めている。

(3) (1)と(2)の DP マッチング(中川 1999)を文単位で行い, 対応付け結果を求める. DP パスは図2のものを用い, 最大3文対3文の対応付けまで考慮する. $g(i, j)$ は以下の式を用いた.

$$g(i, j) = \min_{0 \leq x, y \leq N} \left\{ \begin{array}{l} g(i-x, j-y) \\ + (x+y)^\alpha \times d(i-x, \dots, i, j-y, \dots, j) \end{array} \right\}$$

N : DP パスのサイズ. 図では $N=3$

α : 文の長さに対する重み. $\alpha \geq 1$

また, 距離 $d(i-x, \dots, i, j-y, \dots, j)$ は副音声の $i-x$ から i 番目までの文と字幕の $j-y$ から j 番目までの文の距離で, 次のように定めた.

①副音声を音声認識した結果の $i-x$ 番目の文から i 番目の文までを取り出し, 単語毎にまとめ, $i-x$ 番目から i 番目の文までの単語のリスト W を作る.

②字幕の $j-y$ 番目の文から j 番目の文までの中で, ①で作成した単語リスト W が含まれているかどうかを調べ, 含まれていた単語のリスト W_{inc} を作る.

③距離 d を以下の式で求める.

$$d(i-x, \dots, i, j-y, \dots, j) = \exp\left(-\frac{|W_{inc}|}{|W|}\right)$$

$|W|$: W のサイズ, $|W_{inc}|$: W_{inc} のサイズ,

2.2.2.3. 字幕と副音声の対応付け結果

実験条件として表1のものを用いた.

テストセットとして, 英語ニュース放送の4分弱の4つのニュースを収録し, 主音声と副音声の対訳をそれぞれ25対ずつ抜き出し, それをテストセットとした. それぞれのニュースのトピックと, 英文と和文の文数を表2に示す.

表1 実験条件

形態素解析	ChaSen ver. 2.1
辞書	英辞郎 ver. 0.50
音声認識	
デコーダ	Julius
音響モデル	話者依存, 対角音節モデル, ASJ-JNAS, ASJ-DB を使用
言語モデル	bi-gram(1pass), tri-gram(2pass) 毎日新聞データベースを使用 2万語彙
特徴ベクトル	MFCC+ Δ MFCC+ Δ POW (計25次元)

表2 テストセットの各トピック毎の文数
(括弧内は対応する和文がない英文の数)

トピック	英文	和文
アメリカの軍事	25	36
航空問題	27(2)	28
薬事問題	27(1)	33
西ナイルウイルス	30	28

表3 副音声の単語認識率 [%]

(SS:スペクトルサブトラクション)

SS	正解率	正解精度
なし	63.5	45.9
あり	72.8	67.6

表4 対応付け結果

単語列への変換方法	単語正解率	ずれの平均
音声認識 (SSなし)	63.5%	0.41文
音声認識 (SSあり)	72.8%	0.23文
手動書き起し	100%	0.14文

上記テストセットの副音声(日本語)の認識結果を表3に示す. SSなしでは, 前節(2)で示したように副音声中に主音声の成分が含まれているために, 挿入エラーが増え, 正解精度が顕著に低くなる結果となっている.

表4に字幕と副音声の対応付け結果を示す. 音声認識については表3と同じもので, 手動書き起しは, 副音声を人手で書き起したものである. また, ずれの平均は, 対応付けの結果が表2のテストセットの対応付けからどの程度ずれているかを, 文毎にずれた文数を計算し, 平均をとったものである. DP 計算時のパスのサイズ N , 文の長さに対する重み α については, それぞれ, 2~5, 1.0~2.0の中から, 最も対応付けの結果が良かったものを使用した.

手動で書き起したものを用いて対応付けを行った結果, ずれの平均は0.14文となっており, 対応付けがうまく行われていることがわかる. 音声認識を用いた結果では, SSなしで0.41文, SSありで0.23文となっており, 音声認識率が大きく影響していることがわかる.

具体的には, 平均のずれが0.1文程度であれば, かなり正確に対応付けがとれており, 学習での使用も問題ないと考えられる. 今回最も良かった結果は0.2文程度のずれであり, このままでも学習に利用することは可能であるが, 認識率の向上, 未知語への対応(固有名詞が多いため)など, 改善の余地がある.

また、図3にはNと α を変化させたときの対応のずれの結果を示している。図に示すように、 α が1.4前後ではNにかかわらず安定した結果となっている。図4は自動対応付けの結果例で、この場合の平均のずれは0.17文で、「Chace」が未知語のため「チェイス」(実際は音声)と対応が取れていない。

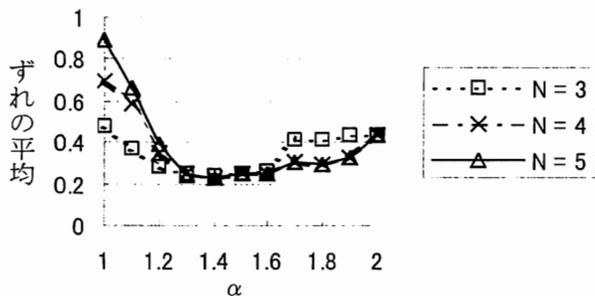


図3 対応付け結果とDP時のパラメータの関係
音声認識(SSあり)

All of that began to change, intervention in fact was redefined with the advent of the Cold War, the creation of the CIA, instead of the Marines storming the beaches we had covert operations Guatemala in 1953 in Iran we put the shah back on the Peacock throne.

海兵隊が海岸に上陸する代わりに、隠密作戦が取られました。たとえば1953年のグアテマラです。またイランでは、国王を再び即位させました。

(解体が海岸に上陸する代わりにホーム作戦がとられました。例えば千百五十三年残ってまだです。またイランでは歩行再び即位させました)

So I think when we're talking about intervention, we want to be very careful not to limit it to traditional military operations.

そこで、干渉について語る時、伝統的な軍事行動に限定すべきではありません。チェイス先生。

(そこで感情について語る時伝統的な軍事行動に限定すべきではありません)

Professor Chace, given that we have a long and rich history of intervention, but what about really military invasion as opposed to covert operations, would you say there's a long history, even if you define it by the military, as a military assault.

確かにアメリカの外国に対する介入の歴史は長いかもしれませんが、軍事介入という点についてはどうでしょうか。それに絞った場合、つまり軍事介入に絞った場合もやはりこれまでたびたびあったと言えるのでしょうか。

(十年前昔からアメリカの外国に対する介入の歴史は長い間も幸せだ軍事介入と友系についてはどうでしょうか。でそれに絞った場合五枚軍事介入に絞った話もありこれまで二人がやったと言えるでしょうか)

図4 対応付け結果の例

(日本文はわかりやすいように副音声を書き起こしたもので、実際には日本語音声である。括弧内は音声認識結果である。)

2.3. 問題作成支援

今回、語学教師が問題を作成するのを支援するツールを作成した。このツールは、問題のリスト表示、字幕、問題作成、問題表示の各部分がある(図5)。問題のリスト表示部分では、これまでに作成した問題が並んで表示される。字幕表示部分では、現在対象としている教材の字幕が表示される。問題作成部分では内容一致問題や穴埋め問題、単語の並び替え問題が作成できる。この際、字幕表示部分の任意の個所をクリックすることにより、その文が問題作成部に取り込まれる。内容一致問題では取り込んだ文を修正し、また他の候補を入力することで問題が作成できる。穴埋め問題では、取り込んだ文の任意の個所をドラッグで指定することにより、そこを穴埋め個所とする問題が作成できる。単語の並び替え問題では、取り込んだ文の単語に対して表示順序を指定することで問題が作成できる。最後の問題表示部分では、現在作成中の問題の画面イメージが表示される。全体的にGUIで構成されており、問題作成を容易にしている。

また、作成した問題は、XML形式で保存され、教材プレイヤーによって利用可能となる。

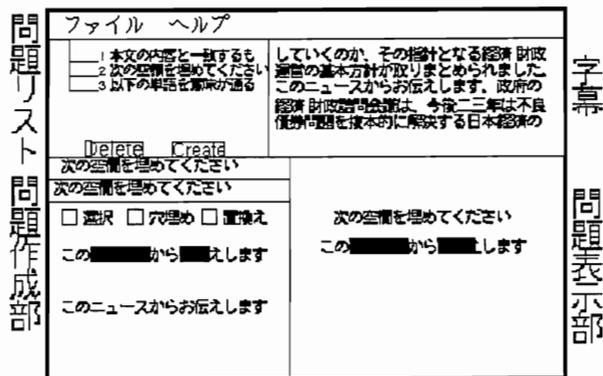


図5 問題作成支援ツール

3. 教材プレイヤーの各機能について

教材プレイヤーの実行例を図6に示す。教材プレイヤーでは、教材作成システムによって得られる教材用データを使用し、市販の語学CALL教材に見られるような以下のような機能を実現している。また、本研究室で別に開発している発音の自動評価システム(中村・中川 2002, 三輪・中川 2002)を組み込める利点がある。

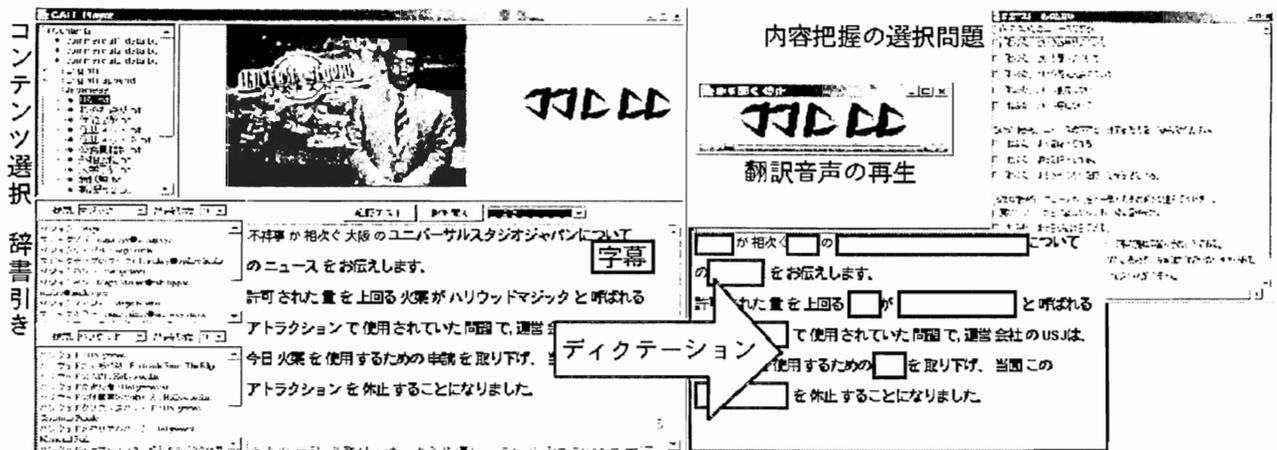


図6 教材プレイヤーの実行例

3.1. ビデオの再生

ビデオの再生には「QuickTime for Java」(Apple Computer)を用い、Java アプリケーション上で「QuickTime」の機能呼び出すことでビデオのスムーズな再生をJava上で実現している。字幕中の任意の単語をクリックすることで、その単語の時間までビデオの再生位置を戻すことができ、それにより聞きたい箇所を繰り返し視聴することが容易となっている。

3.2. 字幕

字幕の提示は、漢字かな混じりでの提示(標準用)と、ルビ(日本語)や発音記号(英語)を付加した提示(初心者用)、特定の品詞のみの提示(中級者用)、および提示なし(上級者用)などの提示方法が選択可能である。

また各提示方法において、再生中のビデオの再生位置に対応する単語がハイライトされ、カラオケ様に表示される。

3.3. 辞書引き

通常の電子辞書と同じように任意の語句をキーボードからタイプする以外にも、字幕中の単語にマウスカーソルをあわせることで、その単語(と単語の原型)が自動的に辞書引きされ表示される。

また、日本語版の日英辞書には「和英辞郎」を、英語版の英日辞書には「英辞郎」を用いている。「英辞郎」「英辞郎」(英辞郎)はともに、翻訳家の有志によって作成されているフリーな辞書で、収録語彙が多い(100万語以上)ことからニュース学習用の辞書としては適していると考えられる。

また、日中・中日辞書にも対応している。

3.4. 翻訳音声

教材のコンテンツとして利用しているニュース放送は、主音声は自国語で副音声はその対訳音声となって

おり、日本語のニュースでは英語の、英語のニュースでは日本語の副音声が発送されている。

そこで、日本語版では英語の、英語版では日本語の翻訳音声をコンテンツ毎に聞けるようにしたもの、および、字幕に副音声を同期させるようにしたものが本機能である。

3.5. ディクテーション問題

字幕中の単語をランダムで入力フィールドとすることにより、自動的にディクテーション問題を作成することができる。機能的には品詞毎にディクテーションを行う単語を選択する機能もあるが、品詞解析の精度の悪さ、学習者の操作の煩雑さなどから後述の被験者実験には使用していない。

入力された文字列が単語と一致した場合に入力フィールドがもとの字幕表示に戻るようになっている。

3.6. 内容把握問題

各コンテンツの理解度を確認するためのテストとして、学習中任意に内容把握を行なうための選択問題等を実行できる。これは他の機能とは異なり、自動的にTV番組と字幕から作成することが現時点で困難なため、現時点では語学教師等、指導者や教材開発者が問題を作成することになる。

このような、教材作成時における教師の介入やそれを支援する機能として、2.3.に述べた問題作成支援ツールを構築した。このツールで作成された問題はCALL Player上で回答することで、回答の正誤が自動採点される。

3.7. 学習状況の管理

本システムにおける学習はノートパソコンなどに教材用データと教材プレイヤーをインストールし、ネットワークに接続されている場所で使用すれば、Web技術

(CGI)を利用し研究室のサーバにおいてユーザー認証を行なう。また、同様に学習者の学習状況をログに保存し管理するようになっている。ログファイルは学習者ごとに時刻とともに動作を保存するようになっている。これにより、以下のような事象の記録が可能となっている。

- ・使用回数
- ・使用時間
- ・単語検索回数
- ・ビデオ再生回数
- ・語彙検索回数
- ・字幕の切り替え回数
- ・副音声の再生回数
- ・ディクテーション問題の回答回数
- ・選択問題の回答回数と回答

4. 被験者実験

4.1. 評価方法

教材作成システムによって得られる教材用データ、および教材プレイヤーを用いることによる学習効果を検討するために、被験者実験を行なった。被験者実験では、教材プレイヤー使用前にプリテストを行なった。その後、教材プレイヤーを8週間使用してもらった。8週間の学習の後に、ポストテストを行なった。このプリテストとポストテストとの得点の変化により教材としての効果を評価する。最後に教材についてのアンケートを行なった。今回の評価実験では、特に統制条件は用意せず、プリテストとポストテストの得点の変化のみから評価を行なっている。なお、今回の実験は被験者数が充分集められず、また評価期間も2ヶ月間と短かったため、予備実験と位置付けられる。

被験者は授業などで募集を募り、8名の留学生が日本語の、8名の日本人学生が英語の学習を行った。留学生は中国語を母語とする者が5名(うち台湾が1名)、韓国語を母語とする者が2名、ベトナム語を母語とする者が1名で、日本語のレベルは日本語検定1級からほとんど話せないレベルまで分散していた。日本人の英語のレベルは英検3級程度であった。

学習に使用したコンテンツは、英語版、日本語版ともにニュース放送を用い、90秒程度づつに区切ったものを1コンテンツとしてそれぞれ20コンテンツ作成し、そのうち10コンテンツには事前に作成した内容把握問題(コンテンツ毎の理解度テスト、選択問題)を利用できるようにした。また日本語版ではニュースのコンテンツの他に、学部学生のための研究室紹介を日本語で

行っているビデオ(6研究室、各7分程度)を用いて、研究室紹介のコンテンツも学習に使用してもらった。

学習は休憩10分を途中に入れた30分を1セットとして、1週間で3セット、8週間で24セット、計12時間を基準として設定した。学習の方法としては、1週間で2コンテンツを学習していくことを被験者に勧めたのみで、具体的な学習方法は全て被験者に任せた。学習は本人所有のPCや、研究室、大学内のWeb教育教室のPCに教材用データと教材プレイヤーをインストールし、進めた。

4.2. プリ・ポストテスト

4.2.1. 内容

プリテストとポストテストの間が8週間あり、かつその解答を被験者に教えていないため、プリテストを行ったことによる学習効果は微々たるものと仮定し、プリテスト、ポストテストともに同じ問題を使用した。テストの内容は日本語、英語ともに、検定試験の問題とニュース音声を用いた問題をそれぞれ20問、合計40問用意した。

日本語の検定試験の問題は、平成8年度の日本語能力検定試験の聴解の問題(日本国際教育協会1996)から1, 2, 3, 4級それぞれ5問ずつ選び合計20問とした。英語の検定試験の問題は、「TOEIC Friends 5」(国際コミュニケーションズ2002)というTOEIC公式の問題集のListeningから20問を選択した。

ニュース音声を用いた問題は、内容把握問題と書き取り(5秒弱の音声の書取。1回聴取)をそれぞれ10問ずつ用意し、合計20問とした。

日本語のニュース音声を用いた問題は「ニュースで学ぶ日本語パート2」(三井・堀・森松1998)という教材を用いた。内容把握問題はその教材から10問抜き出し、書き取り問題はそこから内容把握問題と重ならない音声を10問用意した。

英語のニュース音声を用いた問題は、教材で使用していない素材を用いた。内容把握問題は独自に作成し、英語教師に添削してもらったものを10問用意し、書き取り問題はそのうち内容把握問題と重ならない音声を10問用意した。

4.2.2. 結果

検定試験、書き取り問題、内容把握問題、それぞれ100点満点で採点し、合計点を出す際のそれぞれの重みは問題数に応じてそれぞれ0.5:0.25:0.25とした。

書き取り問題はDPマッチングを用いて正解文と回答文のDP距離より点数を算出した。正解文と回答文を音節(音素)系列に分解し、挿入、脱落をコスト1、

置換をコスト2としてDP距離を計算し、1問10点として1問あたりの点数を以下のように算出した。

$$\text{点数} = (1 - \text{DP 距離}) \times 10$$

また、回答文には単語のスペルミスなどもあるため、そのようなミスを修正した回答文でも点数を算出し、修正したものとしていないものとの平均を取り、それを書き取りの点数とした。

表5に、プリテストとポストテストの平均点と標準偏差を示す。留学生、日本人学生ともに成績が向上しているのがわかる。実際、ポストテストの得点がプリテストの得点よりも高いかどうかt検定を行なったところ、留学生では危険率 $p < 0.001 (t = 3.896)$ で、また日本人学生では危険率 $p < 0.02 (t = 2.372)$ と、有意差が見られた(共に自由度7)。

表5 プリテスト、ポストテストの平均点(標準偏差)

被験者	プリテスト	ポストテスト
留学生平均	71.0(12.0)	85.3(8.3)
日本人平均	36.7(16.0)	48.4(11.4)
全平均	53.8(22.4)	66.8(21.3)

4.3. アンケート

4.3.1. 内容

この教材に対するアンケートは5段階評価(1:悪い~5:良い)で行い、教材プレイヤーの各機能の使い易さ、学習に役立つかなどについて、またコンテンツや他の学習方法との比較など約50項目について、被験者16名にポストテストの終了後に記入してもらった。主な結果を表6に示す。

4.3.2. 結果

教材プレイヤーの各機能が学習に役立っているかという評価(表6(a))の、字幕のカラオケ表示では、2.2節で示したように、英語の字幕と音声の同期精度が不十分なために、留学生と日本人学生との差が開いているものと考えられる。また、辞書引き、ディクテーション問題でそれぞれ留学生の点数が低くなっているのは、辞書引きでは日本語版での辞書に和英辞書を利用しているため、ディクテーション問題では日本語の入力が面倒なため(ひらがな・漢字をタイプ入力しなければならない)であると考えられる。しかし、どの項目についても3点以上で、留学生、日本人学生のどちらかは4点以上となっており、それぞれの機能が語学の学習に際して有用であり、必要とされていることが示された。

表6 アンケート結果

(a) 教材プレイヤーの各機能が学習の役に立つか

質問項目	留学生	日本人学生	平均
字幕表示	4.5	3.6	4.1
辞書引き	3.9	4.4	4.1
ディクテーション	3.9	4.3	4.1
選択問題	4.6	4.1	4.4
翻訳音声の再生	3.9	4.0	3.9

(b) 教材自体についての評価項目

質問項目	留学生	日本人学生	平均
教材が学習の役に立つか	4.5	3.8	4.1
教材を授業で使用してもよいか	4.4	3.6	4.0
コンテンツが自動更新されるのであれば、教材を使い続けたいか	3.9	3.3	3.6

点数が低い項目には各機能の使い勝手がよく挙げられ、特に辞書引きの使い勝手では日本人学生の平均で2.4点(留学生3.5点)となっていた。辞書引きについては4.3節で説明したように、字幕の任意の単語にマウスカーソルをあわせただけで辞書引きされてしまうので、その点で使いにくいという意見があった。また、辞書の内容について、「解説内容が多すぎて語義を選択するのに迷う」という意見も多く、辞書引きの機能についてはまださらに改良していかなければならない。

教材自体についての質問項目の評価(表6(b))では、「この教材が学習の役に立つか」という質問に対しては、4点以上と答えた被験者の割合は88%(留学生100%、日本人学生63%)となり、かなりの割合の被験者がこの学習教材が有効であると考えているのがわかった。また、「この教材を語学の授業で使用してもよいか」という質問でも、81%(留学生88%、日本人学生75%)の被験者が4点以上の点数を付けている。「コンテンツが自動更新されるのであれば、この学習教材を使い続けたいか」という質問に対しては44%(留学生63%、日本人学生25%)の被験者が4点以上をつけた。約半数の被験者が8週間の教材使用後も使い続けることに対して肯定的であり、本論文で示したような語学学習教材の自動的な作成機能についての要望があることを示している。

4.3.3. 講義における使用

本教材プレイヤーを実際の講義で使用し、75分の講義中に15分、計6回の使用の後、教材についてのアンケートを取った。アンケート項目は(京都大学 総合情報メディアセンター2002)に従った。評価は5段階

(1:悪い~5:良い)で行なった。また、アンケート項目を大きく「操作性」に関するものと「学習意欲」に関するものに分け、集計した。受講者34名中、5名は上記実験の被験者であるため、このアンケートの集計からは除いている。結果として表7に見られるような回答傾向となった。操作性および学習意欲の向上ともに4点弱の結果となっており、教材プレイヤーの操作性はだいたい良く、また本教材作成システムによって作成される教材用データと教材プレイヤーによって学習意欲も増すことが示された。

表7 講義における使用に対するアンケート結果

	平均点
操作性	3.6
学習意欲	3.7

5. ま と め

本論文では、まずTVニュース放送とその字幕から語学学習教材を半自動的に作成するシステムについて、特に音声(主、副)と字幕との同期機能について述べた。このような音声情報処理技術と自然言語処理技術を利用した語学学習教材の自動的な作成を目指した研究はこれまでにない。また、教材作成システムによって作成される教材用データを用いる教材プレイヤーの機能について述べた。次に、教材プレイヤーを用いての被験者実験とアンケートを行なった。被験者実験の前後に行なったテストの結果、成績の向上が見られた。本研究では被験者について全体的に学習効果が見られたが、教材としては、すべての学習者に有効なものが望ましいものの、一部の学習者にとって有効なものも有意義である。被験者実験の最後に行なったアンケートでも、教材プレイヤーの機能および教材自体の評価で肯定的な結果が得られた。最後に、英語1クラスで実際に教材プレイヤーを使用し、アンケートを取った。その結果、操作性および学習意欲の向上について肯定的な結果が得られた。本教材作成システムおよび教材プレイヤーの有効性については、今後の詳細な評価が必要である。

なお、本研究のように著作物を加工するような場合、著作権処理が必要となる。作成した教材を組織内で共用するなどの場合は、TV放送会社と契約する必要がある。本研究では、教師以外に、学習者自身でも教材を作成できるツールを目指しており、この場合は著

作権の問題は発生しない。また、評価実験でも一部使用しているように、TVニュース放送以外のコンテンツも本システムにより教材化できるため、著作権処理を回避することが可能である。本システムは、音声情報処理技術と自然言語処理技術を用いた教材作成システムの根幹を開発したものであり、TVニュースに特化したシステムではない。

謝 辞

本研究は文科省科学研究費特別領域研究「メディア教育利用」の助成によって行なわれた。

ニュース放送の研究利用を快諾いただいた、NHKに感謝いたします。

注

- 1) 英語ニュースの場合、字幕が画像として音声と同期して表示されている。そのため、パソコンにテキストとして読み込むためにはOCR等の処理が必要である。現在は、Webに公開されているニュースクリプト(字幕と同等)を利用している。
- 2) 音声自動認識では、S/Nが20dBより悪くなると急激に認識率が低下する。研究段階でよく用いられているクリーンな音声のS/Nは40~50dB程度である。

参 考 文 献

- 安藤彰男, 今井亨, 小林彰夫, 本間真一, 後藤淳, 清山信正, 三島剛, 小早川健, 佐藤庄衛, 尾上和穂, 世木寛之, 今井篤, 松井淳, 中村章, 田中英輝, 都木徹, 宮坂栄一, 磯野春雄(2001) 音声認識を利用した放送用ニュース字幕製作システム, 電子情報通信学会論文誌, J84-D-II(6):877-887
- Apple Computer, QuickTime for Java, <http://developer.apple.com/quicktime/qtjava/>
- BRILL, E. (1995) Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging, *Computation Linguistics*, 21(4):543-565, <http://www.cs.jhu.edu/~brill/>
- The CMU Pronouncing Dictionary, <http://www.speechcs.cmu.edu/cgi-bin/cmudict>.
- DARWING, T. M.(1989) Information type and its relation to non-native speaker comprehension, *Language Learning*, 39:157-172
- 英辞郎, <http://www.alc.co.jp/eijiro/index.html>

五十里慎吾, 佐野輝希, 緒方淳, 有木康雄(2001) ユーザー発話のセグメンテーションと発話評価機能をもつ英語学習支援システム, 情報処理学会 音声言語情報処理 SLP 2001(40):7-12

甲斐充彦, 中川聖一(1997) 冗長語・言い直し等を含む発話のための未知語処理を用いた音声認識システムの比較評価, 信学論, J80-D-II(10):2615-2625

小林聡, 田中敬志, 森一将, 中川聖一(2002) 字幕付きテレビニュース放送を素材とした語学学習教材作成システム, 人工知能学会論文誌, 17(4):500-509

国際コミュニケーションズ(2002), 「TOEIC Friends 5」, 東京

京都大学 総合情報メディアセンター年報(2002), 第2章 語学教育部門 : 51-63

三井豊子, 堀歌子, 森松映子(1998) ニュースで学ぶ日本語パート2, 日本語の凡人社, 東京

三輪多恵子, 中川聖一(2002) 日本人英語音声とネイティブ英語音声のリズム検出と比較, 電子情報通信学会 信学技報, SP2001-142 : 51-58

中川聖一(1999) パターン情報処理, 丸善

中村直生, 中川聖一(2002) 日本人の英語発音の評価法, 電子情報通信学会 信学技報, SP2002-20 : 13-18

奈良先端科学技術大学院大学自然言語処理学講座. 茶筌, <http://cl.aist-nara.ac.jp/lab/nlt/chasen>

日本国際教育協会 国際交流基金 平成8年度 日本語能力検定試験 試験問題と正解, 凡人社, 東京

豊橋技術科学大学中川研究室 英会話 CAI ソフトアンケート結果, <http://www.slp.ics.tut.ac.jp/CALLsoft/>

Summary

There are many kinds of Computer Assisted Language Learning (CALL) materials that have been developed. However making those materials takes a lot of time and efforts. And many times, learners lose interests if they use these materials repeatedly, because the contents are fixed (they are usually distributed by CDs). If dynamic materials can be presented continuously and easily, then learners can pick materials that they are interested in.

For that purpose, we developed a CALL contents creating system (CALL Content Creator) that can construct CALL materials from TV news programs or video materials and captions. The Creator can be used either by learners or teachers in an easy procedure. The Creator can synchronize speech and captions, and we developed a method to synchronize a subchannel and captions. We also made a CALL listening materials player (CALL Player) as a tool to carry out materials that are made by the Creator. The Player can display speech synchronously with captions. We described functions of the Creator and the Player in this paper. In our experiment, we proved that by using the Player, subjects could increase their language test results. In questionnaires for subjects, we got positive answers for the CALL materials.

KEY WORDS: CALL, LANGUAGE LEARNING MATERIAL, JAPANESE LEARNING, ENGLISH LEARNING, LISTENING, TV NEWS, CAPTION, SPEECH PROCESSING, EVALUATION