Vol.2010-DBS-151 No.17
2010/11/13

テロップ情報による語学番組シーン検索手法の評価

渡辺陽介[†1]　勝山裕[†2]
直井聡[†2]　横田治夫[†3]

動画データの録画・蓄積環境の進歩により，大量のビデオデータを扱うことが容易になってきた．我々の研究グループでは語学番組を対象に，録画した番組の中からキーワードに関連したシーンを検索するシステムを開発している．本システムの特徴として，番組中に出現するテロップ情報の活用が挙げられる．テロップの出現間隔からシーン間の区切りを検出し，テロップ文字列をシーンの索引語として用い，テロップの出現頻度・出現時間からシーンの役割を推定している．テレビ番組データにおいて利用可能なテロップ以外の情報として，話者の発言内容を格納したクローズドキャプションがあるが，ニュース番組検索ではこちらが利用されることも多い．しかし，テロップとクローズドキャプションのどちらがシーン検索に適しているか，特に語学番組検索というコンテキストにおいての比較・評価はされていなかった．本稿では，語学番組シーン検索におけるテロップとクローズドキャプションの利用の比較・評価を行う．

Evaluation of Scene Retrieval Method Using Telop Information in Language Education Video

Yousuke Watanabe,[†1] Yutaka Katsuyama,[†2]
Satoshi Naoi[†2] and Haruo Yokota[†3]

Due to the development of video management technologies, we can easily store and watch video data. Our research group has been developing a video retrieval system to obtain scenes that include conversations related to the specified keywords. The system uses telop information in the video to generate metadata for keyword search. It detects scenes and classifies them based on durations of telops and frequency of telops. We investigated retrieval performance of the system in the previous paper, but did not clarify advantages of using telops over other TV metadata such as closed captions. In this paper, we compare use of telops and closed captions in the context of scene retrieval for language education video archives.

1. はじめに

近年の国際化による語学学習の重要性の増加に伴って，デジタル教材による語学学習支援の需要が高まっている．多数の Web サイトで語学学習のための情報が提供されるようになっており，例えば，NHK ゴガクル[1)]，スペースアルク[2)] などが存在する．しかし，多くの学習サイトで提供されている教材の中心は重要なフレーズの文字列データや音声データであり，動画データの教材は多くない．一方，テレビの語学番組は，フレーズの使用例が実際の会話場面を用いて示されるため，会話の雰囲気や前後関係を把握しながら学習する用途に適している．そこで，語学番組のデータを大量に録画・保存して，語学学習のための教材として利用することが考えられるが，その実現には膨大な動画データの中から学習したいフレーズが使われている場面だけを探し出すといった，動画検索の機能が必要となってくる．

本稿では語学番組の動画内における「連続的な一繋がりの会話の開始から終了まで」をシーンと呼ぶものとし，本稿における動画検索の目的を「学習したい特定のフレーズを含むシーンを探すこと」とする．動画検索においては，動画の内容を説明するメタデータを利用することが一般的であるが，語学番組の放送元から発話内容やシーン区切りのような情報が常に提供されているわけではないため，自動生成によりこれらのメタデータを付与する必要がある．本研究では，動画に適切なメタデータを付与するためのアプローチとして，動画中に出現するテロップの情報を利用する．テロップとは，画面に表示される文字情報のことで，語学番組では重要事項の解説や，話者の発話内容またはその翻訳文を提示するためなどに使用される．番組によっては日本語と外国語の対訳が続けてテロップとして表示されることもあるため，テロップ情報をシーン検索の索引語にできれば，日本語と外国語の両方のキーワードで同じシーンの検索を実現できる可能性がある．また，話者の発話内容を単に記述したクローズドキャプションと違い，テロップは重要な場面や強調したい場面に出現することが多く，またテロップの出現時間や前後関係の情報はシーン検索のための重要な手掛かりとなる．例えば，会話シーンの場合はテロップが頻繁に入れ替わるが，解説が行われて

†1 東京工業大学 学術国際情報センター
Global Scientific Information and Computing Center, Tokyo Institute of Technology
†2 株式会社富士通研究所
Fujitsu R&D Inc.
†3 東京工業大学 大学院情報理工学研究科
Graduate School of Information Scienece and Engineering, Tokyo Institute of Technology

© 2010 Information Processing Society of Japan

いるシーンの場合は重要事項の解説テロップが長時間表示される，といった語学番組特有のテロップの使用方法に着目したシーンの分類も可能である．

我々の研究グループではこれまでに，テロップの情報を用いて利用者が入力したキーワードを含んだ一連の会話が行われているシーンを検索するシステムを開発してきた[3),4)]．本システムでは，既存のテロップ認識ツールを用いて語学番組中に出現するテロップ文字列および出現時間の情報を抽出する．このとき，テロップ文字列には誤認識や認識漏れが含まれているため，Web 情報を用いて認識結果の修正を行う．次に，テロップの出現時間間隔に着目して，論理的に繋がっているシーンの区切りを検出する．さらに，テロップの出現時間長及びシーン中のテロップの個数により，そのシーンの内容が会話中心か，解説中心であるかのシーン種別の判定を行う．このように抽出したメタデータ対して，利用者から与えられたキーワードによるシーン検索を行い，候補となるシーンをランキングして提示する．

これまでにシステム単体についての性能評価[3)] については行ってきたが，テロップ以外のメタデータを用いた場合との比較はされていなかった．特に，テロップとクローズドキャプションのどちらがシーン検索に適しているか，特に語学番組検索というコンテキストにおいて明らかにはされていない．そこで，本稿では語学番組シーン検索におけるテロップとクローズドキャプションの利用の比較・評価を行う．

本稿の構成は以下のとおりである．2 節では，クローズドキャプションとテロップについて，それぞれの特徴を述べる．3 節では，我々が研究開発を行っている語学番組シーン検索システムの概要を説明する．4 節では，テロップとクローズドキャプションを用いたシーン検索の比較について述べる．5 節で関連研究について紹介し，6 でまとめと今後の課題を述べる．

2. クローズドキャプションとテロップ

まず，本稿で用いるクローズドキャプションとテロップについて解説する．

2.1 クローズドキャプション

クローズドキャプション (closed caption. 以下 CC とも表記する) は狭義には北米で行われている字幕放送を指すが，ここではテレビ番組中の話者の発言内容や効果音等を文字で表した字幕データ全般を指す語として用いる．クローズドキャプションの文字情報は発話内容が主であるため，話者のいないシーンがほとんどない語学番組では，大半のシーンにおいて対応する文字情報を取得することが期待できる．話者ごとに字幕の色を変えることができる点も特徴である．ただし，クローズドキャプションはテロップほどに表現の自由があるわけではなく，テロップのように内容の重要部分だけを一部色や大きさを変えて強調したり，画面内の任意の位置に表示したりといったことはできない．

字幕放送の普及度合いに関しては，対応した番組は徐々に増えてきているが，2010 年 8 月時点の NHK に限っても，放送されている全ての語学番組が字幕放送対応になっているわけではない．また，過去に放送された語学番組データの蓄積があっても，これらに対して放送元から後付けで字幕データが提供される見込みは薄い．そのため，字幕情報が利用できない語学番組は相当数存在する．字幕が利用できない番組には，音声認識による認識結果でクローズドキャプションの情報の代用とするというアプローチもあるが，本研究では考慮しない．

2.2 テ ロップ

テロップは画面内に表示される文字情報である．クローズドキャプションとほぼ同様な目的で字幕として使用されることもあるが，それ以外の多くの用途にも用いられる．テロップには，話者の話す内容を表示する以外にも，重要事項の強調や，そのシーンの主題を表示するといった用途に使われることがあるため，単なる文字情報だけではなく，テロップの出現時間や出現位置，前後関係，文字色や大きさといった要素も情報として利用できる．

特に語学番組では，例えば，解説シーンにおいて説明対象になっている重要フレーズを解説の開始から終了までテロップとして出し続けるといった使い方がされる．この場合には，テロップの表示内容とクローズドキャプションの内容は異なり，さらにテロップの表示期間から解説シーンの開始・終了を特定することも可能である．また，番組内における特定のコーナーであることを表すために，コーナーが続く期間中，コーナー名をテロップとして画面端に出し続けるという使い方もされる．本研究では，シーン間の切れ目の検出や，シーンが会話中心か解説中心かの種別判定において，テロップの出現時間の情報を利用しているが，同様のことをクローズドキャプションや音声認識の結果を用いて実現することは難しいと考えられる．

テロップ情報は，画像認識によって映像中から文字情報を抽出することができるため，放送元から文字データが直接提供されていなくても使えるという利点がある．ただし，テロップの認識精度の問題は存在するため，ノイズ除去などの処理が必要となる．

3. テロップ情報を用いた語学番組シーン検索システム

本節では，我々の研究グループが開発しているシーン検索システム[3),4)] の概要について述べる．4 節の比較実験でも，本システムをベースに行っている．本システムは，利用者か

© 2010 Information Processing Society of Japan

Vol.2010-DBS-151 No.17
2010/11/13

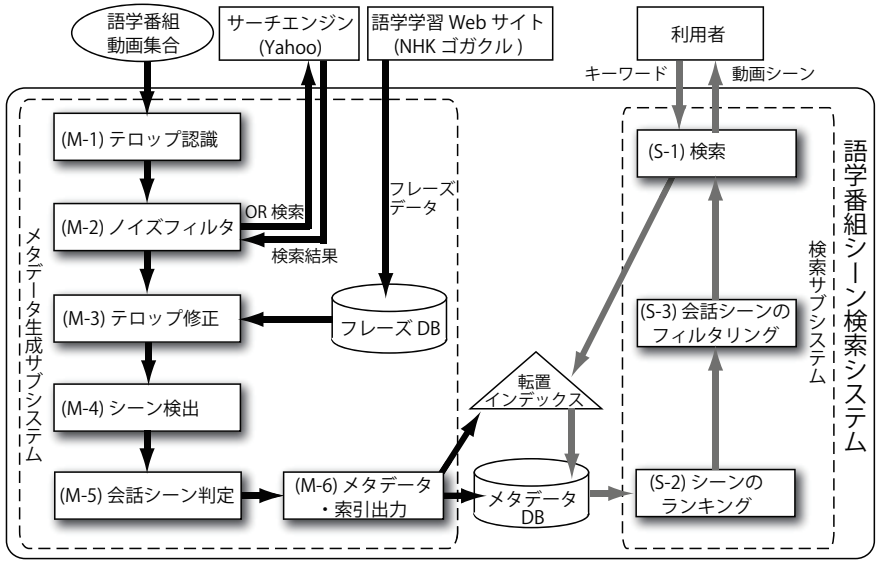

図 1 シーン検索システムアーキテクチャ

ら与えられたキーワードと，探したいシーンの種別 (会話シーンか解説シーンか) を入力として受け取り，出力としてキーワードに関連したシーンのリストを返す．システム構成図を図 1 で示す．本システムはメタデータ生成サブシステムと検索サブシステムの二つから構成され，索引の作成処理とシーン検索処理をそれぞれが担当する．

3.1 メタデータ生成サブシステム

メタデータ生成サブシステムは，検索サブシステムが使用するためのシーンに対する索引データを作成する．以下の 6 ステップによる処理を行う．

(M-1)**テロップ認識**: テロップ認識ツールを用いて，動画からテロップ文字列・開始時刻 (フレーム)・終了時刻 (フレーム) の情報を取得する．図 2 は認識結果の例で，1 つの認識結果が 2 行ごとに出力されている．1 行目の SF はテロップの開始フレーム，EF はテロップの終了フレームを表し，次の行の文字列が認識されたテロップ文字列である．認識結果のうち,「■口口■」,「いⅠ」,「癖§ ソ」などは誤認識によるノイズである．

(M-2)**ノイズフィルタ**: 認識結果のテロップ文字列には意味不明な文字列が含まれている可能性があるため，事前に定義した 3 つのルールに基づいてそれらを除去する．(1) テロップ文字列に対しての記号文字の割合が T_k 以上の場合は，通常の文章では起こりにくいためノイズとみなす．(2) テロップ文字列の長さが T_l 以下の短すぎる文章はノイズとみなす．(3) テロップ文字列を文字単位で N-gram に分解して，それぞれの断片を検索エンジン[9] に検索語として与え，処理結果の合計数が T_m 以下しか戻ってこない場合は，意味不明な文字列の可能性が高いのでノイズとみなす．

(M-3)**テロップ修正**: テロップ文字列の誤認識が部分的であれば，類似する正しい文章と照合することで修正可能である．本システムでは，Web 上の語学学習サイト「NHK ゴガクル」[1] から取得したフレーズデータとテロップ文字列を比較する．文字列間の類似度が T_n 以上の文章が見つかった場合には，それを本来の文章とみなして誤認識の修正を行う．類似度の計算法について，本研究は特定の手法に依存するものではないが，実装上は各文字列を N-gram に分解したときの共通要素の割合で計算している．4 節の実験では，外部サイトの情報が変化することによる結果への影響を除くためにこのステップは実行していない．

(M-4)**シーン検出**: テロップの出現時間間隔を利用して，シーンとシーンの切れ目を検出する．語学番組では，シーンとシーンの切れ目において，テロップが表示されない空の区間が発生するため，テロップの出現間隔からシーン検出が可能である．出現間隔が T_d 以上であった場合をシーンの切れ目とみなす．図 4 のようなテロップ出現の場合は，連続した 1 つのシーンに出現したテロップであるとみなし，図 5 のように空白期間があるものは別のシーンに分割する．

(M-5)**会話シーン判定**: そのシーンが会話主体であるか，解説主体であるかを，シーン内のテロップの出現時間長及び出現テロップ個数により判定する．語学番組では，会話シーンにおいては会話に合わせてテロップが切り替わるため，出現時間が短く，連続出現するテロップ数も多い．逆に，解説シーンにおいては，重要フレーズについてのテロップを表示したまま解説が進行することが多いため，一つ当たりのテロップの出現時間が長い．この特性を考慮して，シーン内のテロップ数が 2 以上，またはシーン内の全テロップの出現時間が T_j 以下の場合は，会話シーンであると判定する．

(M-6)**メタデータ・索引出力**: 検索サブシステム用の転置インデックス及びストリーム動画配信用のメタファイルを作成し，検出したシーンごとの情報をメタデータ DB に格納する．インデックスの作成においては，シーン内のテロップ文字列を文字単位で N-gram に分解し，各断片をそのシーンへの索引語とする．N-gram を用いる理由は，

© 2010 Information Processing Society of Japan

```
SF=1502,EF=1577
■□□■
SF=1682,EF=1727
いI
SF=1592,EF=1757
こんにちは！サラですアメリカから今着いたんです
SF=1712,EF=1757
4
SF=1772,EF=1877
やあ ようこそ！ようこそ！
SF=2012,EF=2177
松平光太郎です 光太郎って呼んでくださいねよろしくお 願いします
SF=2162,EF=2327
どうもコタロウ こちらこそよろしく1
SF=2342,EF=2477
こちらで-Iするの／、きし《ヲな てす
SF=2582,EF=2717
今後ともどうぞよろしく
SF=2702,EF=2747
癖§ソ
SF=2762,EF=2807
-盤懲爵寸                    寸･○□Ⅲ-』←凧屾 ..》
ぺ..
SF=2702,EF=2957
どうも！これふるさとからのお 土産なんですホームメードのストロ
ベリージャム
SF=2942,EF=3257
Ihopeyoulikeit
```

図 2 認識結果

```
SF=1592, EF=1757
こんにちはサラですアメリカから今着いたんです
SF=1772, EF=1877
やあようこそようこそ
SF=2012, EF=2177
松平光太郎です光太郎って呼んでくださいねよろしくお願いします
SF=2162, EF=2327
どうもコタロウこちらこそよろしく1
SF=2342, EF=2477
こちらでIするのきしヲなてす
SF=2582, EF=2717
今後ともどうぞよろしく
SF=2702, EF=2957
どうもこれふるさとからのお土産なんですホームメードのストロベ
リージャム
SF=2942, EF=3257
Ihopeyoulikeit
```

図 3 ノイズ除去後の結果

図 4 オーバーラップ

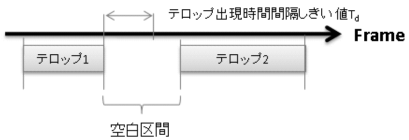

図 5 非オーバーラップ

テロップ文字列中に除去しきれなかった誤認識文字や，認識漏れした文字があった場合にも，利用者キーワードに部分マッチしやすくなり，検索結果の再現率を改善できるためである．

3.2 検索サブシステム

検索サブシステムは，利用者が与えたキーワードとシーン種別の検索条件を受取り，関連するシーンを取得し，ランキング結果を返す処理を行う．以下の 3 ステップにより実現されている．

(S-1) **検索**: 転置インデックスを用いて，利用者が入力したキーワードに関連する結果をメタデータ DB から探る．利用者キーワードを N-gram に分解し，各断片とのマッチングを行う．N-gram を用いる理由は，前述したとおり，誤認識を含むテロップ文字列とキーワードが部分マッチしやすくなるようにするためである．

(S-2) **シーンのランキング**: 検索結果シーンをテロップの文字とキーワードの合致度によってランキングする．

(S-3) **会話シーンのフィルタリング**: 利用者からシーン種別が検索条件として与えられた場合は，指定された種別に対応するシーンだけを残す．会話シーンのみ，解説シーンのみ，両方を含む，の 3 パターンが可能である．

4. 評 価 実 験

本節では，本研究において行った評価実験について述べる．実験の目的は，テロップ情報をシーン検索に用いた場合と，クローズドキャプションを用いた場合を比較検討することである．まず実験データについて述べ，次に比較対象であるクローズドキャプションを用いたシーン検索の概要を述べる．なお，今回の評価では，ステップ M-3 のテロップ修正の処理は，外部の語学学習サイトの影響を強く受けるために行わないものとした．

4.1 実験データ

本研究では，評価用の語学番組データとして，NHK が地上波アナログ放送で 2010 年 1 月に放送した語学番組を用いる (表 1)．動画 1 の放送時間は約 10 分で，日本語キーフレーズとして「やっぱりこっちがいいよ」，それに対応する英語フレーズとして「I still like this one better」が紹介される．動画 2 の放送時間は約 20 分で，日本語キーフレーズとして「想像もしてなかった」，それに対応する英語フレーズとして「I never imagined」が紹介される．

テロップ情報の取得には，富士通研究所が開発したテロップ認識ツール[7] を使用する．このツールは，MPEG 形式の動画データを与えると，動画中に出現するテロップの開始時刻，終了時刻，認識文字列をセットにして返す．対象番組の解析を行ったところ，認識できたテロップ数は動画 1 では 88 個，動画 2 では 157 個であった．図 6 は動画 1 について認識されたテロップの情報を時系列順に表したものである．横軸が時間軸を表している (一番上は開始から 2 分 30 秒まで，2 番目は 2 分 30 秒から 5 分 00 秒まで，3 番目は 5 分 00 秒から 7 分 30 秒まで，一番下は 7 分 30 秒から最後まで)．画面には表示されたものの認識結果としては出力されなかったテロップもあった．それらについては，テロップ認識ツール側の改善が必要となる．

クローズドキャプション情報は，本来ならば配信されている情報を使うべきであるが，今

© 2010 Information Processing Society of Japan

表 1 実験動画

	動画 1	動画 2
動画詳細	英語がわかる 100 のツボ！ (2010/1/4)	リトルチャロ (2010/1/4)
動画時間	約 10 分	約 20 分
認識されたテロップ数	88	157
クローズドキャプション数	142	292
人手によるシーン作成数	27	31

表 2 人手によるシーン区切り (動画 1)

開始時刻	終了時刻	説明	開始時刻	終了時刻	説明
00:00	00:22	タイトル	04:46	05:03	会話シーン 9
00:23	00:56	人物紹介	05:04	05:17	会話シーン 10
00:57	01:08	会話シーン前置き	05:18	05:20	会話シーン 11
01:09	01:23	会話シーン 1	05:21	05:26	会話シーン 12
01:23	01:28	会話シーン 2	05:27	05:34	会話シーン 13
01:29	01:55	会話シーン 3	05:35	05:42	解説シーン 4
01:56	02:00	会話シーン 4	05:43	06:40	解説シーン 5
02:01	02:06	会話シーン 5	06:41	07:38	解説シーン 6
02:07	02:14	会話シーン 6	07:39	07:59	解説シーン 7
02:15	02:18	会話シーン 7	08:00	09:07	会話シーン 14
02:19	02:35	会話シーン 8	09:08	09:14	会話シーン 15
02:36	02:57	解説シーン 1	09:15	09:35	会話シーン 16
02:58	03:10	解説シーン 2	09:36	09:45	エンドクレジット
03:11	04:45	解説シーン 3			

回用いた番組はアナログおよびデジタルでも字幕放送を行っていないため，人手によって話者の発話内容と発言時刻をテキストに落とした上で評価を行う．今回対象とした動画 1 では 142 発言，動画 2 では 292 発言が含まれていた．テロップに比べると，話者の発言は移り変わりがより頻繁であることがわかる．

シーンの区切りの正解データとして，人手によりシーンの作成を行った．動画 1 では 27 シーン，動画 2 では 31 シーンを作成した．スペースの関係上，ここでは動画 1 ついて作成した正解シーンのみ，表 2 に示す．

4.2 比較対象：クローズドキャプションを用いたシーン検索

ここでは，テロップと同様にシーンの索引付けに用いられることの多い，クローズドキャプションの情報を用いる手法を比較対象とする．テロップを用いる提案手法と比較するために，発言開始時刻と発言終了時刻の情報から 3.1 節の各ステップと同様の手法を用いて，シーン検出・シーン種別判定を行う．すなわち，ある発言から次の発言までの間に空白期間が一定時間あった場合に，そこをシーンの切れ目とする．シーン種別判定についてもテロップ情報の場合と同様の判断基準を適用する．

表 3 シーン検索パラメータ

変数名	値	役割
T_k	0.3	ノイズフィルタ 1 で，ノイズとみなす記号文字の割合の閾値
T_l	2	ノイズフィルタ 2 で，ノイズとみなす短い文字列の閾値
T_m	100,000	ノイズフィルタ 3 で，ノイズとみなす検索エンジンの処理結果数の閾値
T_d	4sec	シーン検出で，シーンの切れ目を決める空白区間の閾値 クローズドキャプションに適用する場合のみ，2sec と 3sec を使用.
T_j	19sec	会話シーン判定で，会話テロップと判断するための出現時間の閾値
N	3	文字列を N-Gram に分解するときの単位

表 4 テロップを用いたシーン検出結果

調査項目	動画 1 テロップ	動画 1 CC	動画 2 テロップ	動画 2 CC
入力テロップ (CC) 数	88	142	157	292
ノイズフィルタによる除去数	19	5	34	20
ノイズフィルター後のテロップ数	69	137	123	272
システムの検出したシーン数	17	14	25	29
発見した正解シーン数	8	5	7	4

4.3 評価方法

シーン区切りの精度評価実験では，人間が区切ったシーンと比較するものとする．ただし，人手により作成したシーンの切れ目であっても，区切り時刻に関しては必ずしも秒単位で正確とは言えない．そこで，人間の区切りの時刻とシステムによる区切り時刻の間に起きるずれを，前後 8 秒までは許容するものとした．

評価尺度は，Precision(式 1), Recall(式 2), F 値 ($F\text{-}measure$：式 3) を用いる．

$$precision = \frac{|Result \cap Scene|}{|Result|} \tag{1}$$

$$recall = \frac{|Result \cap Scene|}{|Scene|} \tag{2}$$

$$F\text{-}measure = \frac{2 * precision * recall}{precision + recall} \tag{3}$$

4.4 実験結果

まず実験に用いた各パラメータは，表 3 の通りである．パラメータ T_d については，テロッ

© 2010 Information Processing Society of Japan

プを用いる手法については 4sec で統一し，クローズドキャプションの方は動画 1 で 2sec，動画 2 で 3sec を用いた．クローズドキャプションの閾値 T_d を動画ごとに変える必要があった理由は，発言の間隔がテロップ出現間隔よりも短く，また番組ごとに話す速度が異なっているためである．

ノイズ除去の結果とシーン検出の結果を表 4 に示す．クローズドキャプションのデータも一部ノイズとして除去されているが，これらは「あ」や「おー」や「ねー」など，短すぎる発言である．シーン検出における両者の比較結果を図 7，図 8 に示す．図より，どちらも精度に改善の余地はあるが，テロップを用いた方がクローズドキャプションよりもシーン検出精度がすぐれていることが分かる．クローズドキャプションが悪かった理由として，語学番組では話者の発言がほぼ絶え間なくあり，明確な無言シーン以外では論理的な切れ目として検出することは難しかったためである．

次に，シーン種別の判定精度についての比較結果を図 9 に示す．図 9 は，動画 1 と動画 2 の両方において，システムが会話シーン/解説シーンとそれぞれ判定したシーンのうち，人間が確認して実際に種別が正しかったものの割合を表している．テロップを用いて発見した会話シーン 9 件は全て正解で，解説シーンの方は 6 件中 3 件が正解だった．クローズドキャプションの場合は，会話シーンで 8 件中 7 件正解したが，解説シーンは 1 件しか発見できず，正解が含まれなかった．この結果から，シーン種別の判定制度についても，テロップ情報を用いる手法の有効性が確認された．クローズドキャプションは，会話であっても解説であっても発言頻度に差がほとんど見られないため，発話内容の意味解析まで踏み込まない限りはシーン種別の判定が難しいと思われる．

5. 関 連 研 究

動画を扱うシステムにおいて，コンテンツの内容を説明するメタデータをどのようにして得るのかという点は非常に重要である．本研究では，テロップ情報からメタデータを生成しているが，それ以外のメタデータ作成アプローチとして，人手による付与，クローズドキャプションの利用，音声認識，電子番組表の利用などがあげられる．橋本ら[11] は，野球動画のダイジェスト生成のために，1 球ごとの投球結果や攻守交代などの論理的なシーン区切りを人手で入力してメタデータを付与している．人手によるメタデータ付与は最も確実で信頼できる方法ではあるが，コストが高いという問題点がある．河合ら[12] は，番組紹介映像の自動生成のためにクローズドキャプションを用いており，さらにクローズドキャプション内の単語に重みづけをするために電子番組表の文字情報も用いている．本研究では，シーン間の切れ目の検出や，シーンが会話中心か解説中心かの判定において，テロップの出現時間の情報を利用しているが，同様のことをクローズドキャプションや音声認識の結果を用いて実現することは難しい．

次にメタデータとしてテロップ情報を用いた動画検索の関連研究について述べる．Kuwano らが提案した Telop-on-demand system[6] は，ニュース番組を対象として入力キーワードを含むテロップが表示されている動画を検索する．このシステムではテロップの文字情報が利用されている．それに対して本研究では，テロップの文字情報を索引語として利用するだけでなく，出現時間の情報も用いて，シーン間の切れ目の検出やシーンが会話中心か解説中心かの判定に利用している点が特徴である．また，本研究では対象がニュース番組でなくて語学番組であるという点も異なっている．

本研究では，テロップ認識には既存のテロップ認識技術[7] を用いているが，ツールの認識精度を向上させるために認識結果の文字列の修正処理を行っている．文字列の誤り検出及び訂正については，これまで文脈を考慮した手法や統計的言語モデルをベースとした手法などが提案されている[13)–16)]．これらの手法では辞書データを用いているが，語学番組では最近のニュース記事を教材として扱う場合などがあり，新しい用語が使われることも多いため，辞書が新しい情報に対応していない場合がある．我々の研究グループは過去の論文において，Web データを活用した TV テロップ認識率向上手法[5] を提案した．この手法では，テレビのニュース番組のテロップ認識結果に対し，Web 上のニュース記事を用いて誤認識などを検出し，自動修正を行った．本研究でもこの手法に基づいてテロップの修正を行うが，語学番組のテロップはニュース番組よりも短時間で変化し，なおかつ出現する位置がニュース番組ほど固定ではないため，認識結果により多くの誤認識や意味不明な内容が含まれる．そこで，本研究ではサーチエンジンのヒット数等を用いて意味不明な文字列を検出・除去した後に，Web 上の語学学習サイトから取得したフレーズの文字データを用いて，上記手法でテロップの誤字修正を行っている．

6. おわりに

本稿では，複数の語学番組ビデオデータから，利用者が入力したキーワードに関連する一連の会話が行われているシーンを検索するシステムの評価を行った．本システムは，Web 上の情報を用いてテロップ認識結果の修正を行い，テロップの出現時間間隔，出現時間長及び個数を利用し，シーンの区切りを検出し，会話シーンの判定を行う．本稿で述べた実験では，クローズドキャプションを用いたシーン検出手法との比較を行い，クローズドキャプ

© 2010 Information Processing Society of Japan

ションではシーン検出が難しく，テロップ情報が有効であることを示した．

今後の課題として，まずより詳細な評価実験が挙げられる．今回はごく少ない数の番組でしかシーン検出の評価が行えなかったが，より多くの番組を用いた評価は当然必要と考える．キーワードに対する検索精度についての評価も行う必要がある．また，提案手法の拡張も課題の一つである．特にクローズドキャプションとテロップを組み合わせることは検討すべきと考えている．

謝　　辞

本研究の一部は文部科学省科学研究費補助金特定領域研究（#21013017）の助成により行われた．

参 考 文 献

1) NHK ゴガクル，http://gogakuru.com/index.html
2) スペースアルク, http://www.alc.co.jp/
3) 周 清楠, 渡辺陽介, 勝山 裕, 直井 聡, 横田治夫, テロップと Web 情報を用いた語学番組シーン検索システム, DEIM Forum 2010.
4) 周 清楠, 渡辺 陽介, 勝山 裕, 直井 聡, 横田治夫, 語学番組検索システムにおけるシーン区切り検出手法, 情報処理学会全国大会, 2010.
5) ドウンゴフゥン，勝山裕，直井聡，横田治夫，Web サーチを活用した TV テロップ認識率向上手法，信学技報, vol.108, no.93, DE2008-29, pp.163-168, Jun.2008.
6) H. Kuwano, Y. Taniguchi, H. Arai, M. Mori, S. Kurakake and H. Kojima, Telop-on-demand: Video structuring and retrieval based on text recognition, IEEE International Conference on Multimedia and Expo., vol.2, pp.759-762, 2000.
7) Y. Katsuyama, H. Bai, H. Takebe and K. Fujimoto, A study for caption character pattern extraction, IEICE Tech. Rep., vol. 107, no. 491, PRMU2007-239, pp. 143-148, Feb. 2008.
8) 田淵浩章，坂本廣，北村泰彦，N-gram に基づく用例対訳検索手法，信学技報, vol.108, no.441, AI2008-52, pp.43-48, Feb.2009.
9) YahooAPI, http://developer.yahoo.co.jp/
10) 酒井哲也，よりよい検索システム実現のために，情報処理，vol.47, no.2, pp.147-158, Feb.2006.
11) 橋本隆子，加登岡隆，飯沢篤志，モバイル環境におけるプロ野球パーソナルダイジェスト配信システム，日本データベース学会 Letters (DBSJ Letters), Vol.1 No.2,pp.24-27, 2003.
12) 河合吉彦，住吉英樹，八木伸行，電子番組表における紹介文を利用した番組紹介映像の自動生成手法，電子情報通信学会論文誌 D, Vol.J91-D No.8 pp.2157-2165, 2008.
13) Karen Kukich, “ Techniques for automatically correcting words in text ”, ACM Computing Surveys, vol.24, no.4, pp.377-439 (1992)
14) X. Tong and D. Evans, “ A statistical approach to automatic ocr error correction in context ”, In Proceeding of the Fourth Workshop on Very Large Corpora. Copenhagen, Denmark, pp.88-100 (1998)
15) Masaaki Nagata,“ Context-based spelling correction for japanese ocr ”, In Proceeding of the 16th conference on Computational linguistics, vol.2, pp.806-811 (1991)
16) Masaaki Nagata, “ Japanese ocr error correction using character shape similarity and sttistical language model ”, In Proceeding of the 17th conference on Computational linguistics, vol.2, pp.922-928 (1998)

© 2010 Information Processing Society of Japan

Vol.2010-DBS-151 No.17
2010/11/13

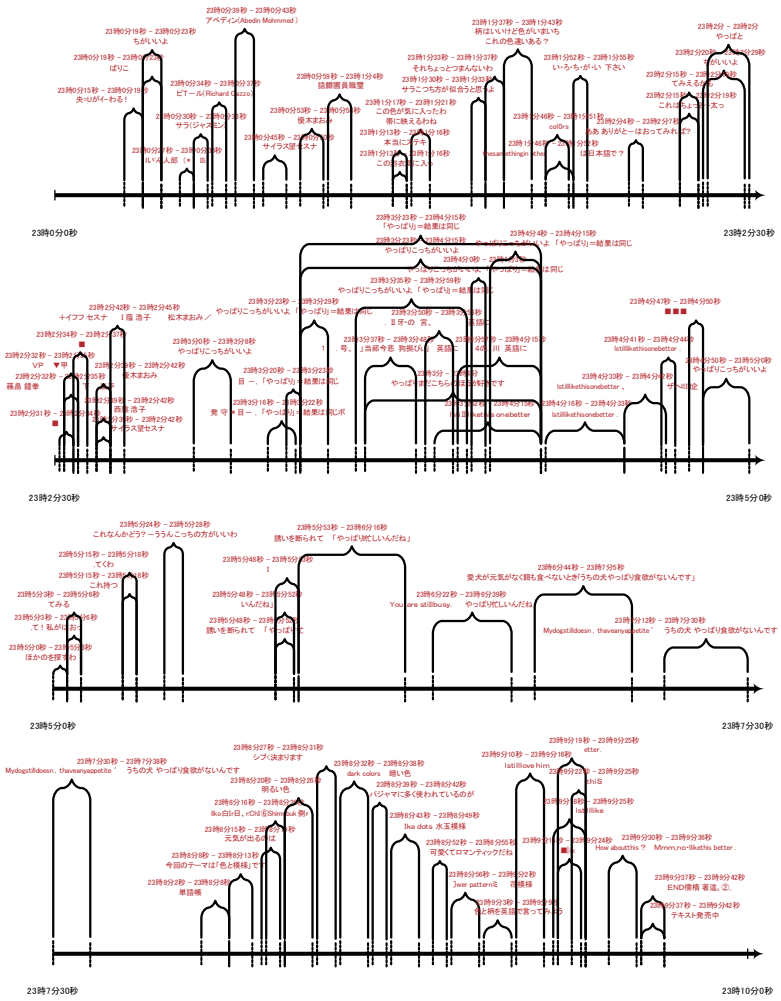

図 6 テロップ認識結果 (動画 1)

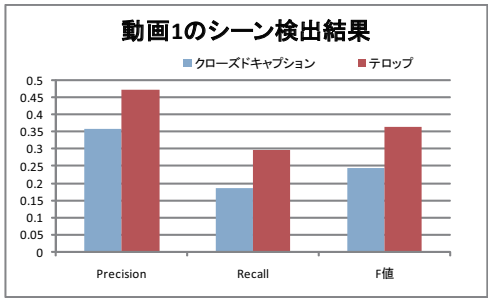

図 7 シーン検出精度 (動画 1)

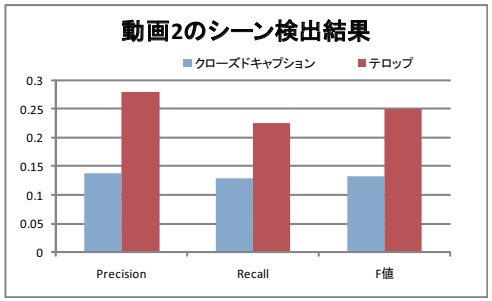

図 8 シーン検出精度 (動画 2)

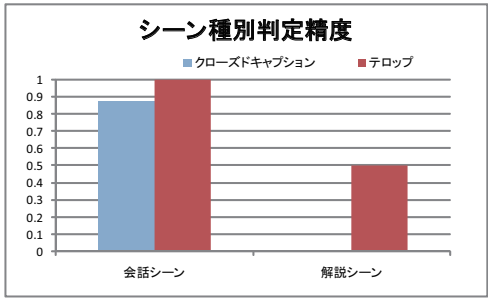

図 9 シーン種別判定精度

© 2010 Information Processing Society of Japan