T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	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
Title(English)	A Scene Retrieval System for Language Education Videos utilizing Telop and Web Searching Information
著者(和文)	周 清楠, 渡辺陽介, 勝山 裕, 直井 聡, 横田治夫
Authors(English)	Qingnan ZHOU, Yousuke WATANABE, Yutaka KATSUYAMA, Satoshi NAOI, Haruo YOKOTA
出典(和文)	, , , D4-2
Citation(English)	, , , D4-2
発行日 / Pub. date	2010, 3

Powered by T2R2 (Science Tokyo Research Repository)

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（O）

周　清楠† 渡辺　陽介†† 勝山　裕††† 直井　聡††† 横田　治夫††,††††

† 東京工業大学　情報工学科
†† 東京工業大学　学術国際情報センター
††† 株式会社　富士通研究所
†††† 東京工業大学大学院　情報理工学研究科計算工学専攻
E-mail: †{seinan,watanabe}@de.cs.titech.ac.jp, ††{katsuyama,naoi.satoshi}@jp.fujitsu.com,
†††yokota@cs.titech.ac.jp

あらまし　近年，語学学習サイトが数多く提供されるようになったが，会話フレーズそのものの文字列や音声しか提供していないものがほとんどである．テレビの語学番組を利用することで，雰囲気も含めてフレーズの使い方を学習することが可能となるが，大量の語学番組の中から学習したいフレーズに関連する会話シーンを探し出すことは，多くの時間と労力を要する．本稿では，テロップの情報を使い，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を提案する．提案システムでは，Web 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，シーン区切りの検出及び会話シーンの判定を行う．また，テロップ情報に基づく転置インデックスを用意し，検索結果をテロップの文字とキーワードとの合致度によりランキングする．

キーワード　テロップ，シーン検出，シーン検索

A Scene Retrieval System for Language Education Videos utilizing Telop and Web Searching Information（O）

Qingnan ZHOU†, Yousuke WATANABE††, Yutaka KATSUYAMA†††, Satoshi NAOI†††, and Haruo YOKOTA††,††††

† Department of Computer Science, Tokyo Institute of Technology
†† Global Scientific Information and Computing Center, Tokyo Institute of Technology
††† Fujitsu Laboratories Ltd.
†††† Department of Computer Science, Graduate School of Information Science and Engineering Tokyo Institute of Technology
E-mail: †{seinan,watanabe}@de.cs.titech.ac.jp, ††{katsuyama,naoi.satoshi}@jp.fujitsu.com,
†††yokota@cs.titech.ac.jp

Abstract　In recent years, a lot of language study sites have been offered. Most of the language study sites only provide the strings or the voices of phrases which users want to learn. However, it is impossible for users to learn phrases while gasping the atmosphere of actual conversation. For increasing linguistic ability and learning the real conversation, it is useful to study from language study program. However, it is difficult to find the necessary scene from lots of videos. In this study, using the telop, we aim to propose a system to retrieve the scene that includes the conversations related to keywords given by users. We use web searching information to correct telop recognition results, and then detect duaration of scenes based on appearing time of the Telop.

Key words　Telop, Scene Detection, Scene Retrieval

1. はじめに

近年，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た．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，多くの学習サイトは，会話フレーズそのものの文字列や音声しか提供していないものがほとんどである．語学学習においては，前後関係を把握しながら会話を修得することが重要である．テレビの語学番組を利用することで，雰囲気も含めてフレーズの使い方の学習することが可能となるが，大量の語学番組を見て，その中から学習したいフレーズに関連する会話シーンを探し出すことは，多くの時間と労力を要する．

適切な会話シーンを検索するためのアプローチとして，画面中に表示されるテロップを利用することができる．テロップとは，テレビなどの画面に表示される文字情報のことで，例えば，ニュースの重要事項，語学番組の字幕などが挙げられる．また，テロップはクローズドキャプションと違い，重要な場面や強調したい場面に出現するため，シーンの検索や区切りに有効であると考えられる．しかしこれまでのテロップを対象とした動画検索技術を適用しただけでは，検索対象のフレーズに対応するテロップが表示されている区間が特定できるだけで，会話の雰囲気を知るための前後関係を含んだシーンを抽出することはできない．本稿では，テロップの情報を使い，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を提案する．

提案システムでは，まず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からテロップを抽出する．テロップの認識結果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，そのシーンが会話であるかどうか判定を行う．抽出した会話シーンに対して，その中に出現するテロップの文字情報に基づく転置インデックスを用意し，与えられたキーワードに対して検索を行う．検索結果シーンはテロップの文字とキーワードとの合致度によってランキングして提示する．

本稿の構成は以下のようになっている．まず，2. 節で関連研究について述べる．3. 節において本稿の提案手法について説明を行い，4. 節でプロトタイプシステムについて述べる．そして，5. 節で本システムに関する評価実験の結果を示し，6. 節においてまとめと今後の課題について述べる．

2. 関連研究

2.1 テロップ認識度の向上

我々の研究グループは，Web データを活用した TV テロップ認識率向上手法 [3] を提案した．この手法はニュース番組のテロップの認識結果に対し，Web 上のニュース記事を用いて，誤認識などを検出し，自動に修正する手法である．

本稿もこの手法を用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のテロップと違い，短時間で変化し，なおかつ出現する位置が固定ではない．その結果，認識結果に多くの誤認識や意味不明な内容がある．そこで，本稿は，誤認識や意味不明な内容を除去した後に，フゥンらが提案した手法でテロップの修正を行う．

2.2 テロップを用いたニュース検索

H.Kuwano らが提案した Telop-on-demand system [4] は，ニュース番組のテロップ情報を用いて，入力キーワードを含むテロップが表示されている区間を検出する．しかし，この手法は，検索対象のフレーズに対応するテロップが表示されている区間が特定できるだけで，会話の雰囲気を知るための前後関係を含んだシーンを抽出することはできない．

本稿では，テロップの情報を使い，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を提案する．

3. 語学番組シーン検索システム

3.1 語学番組検索における問題点

本研究の目的は，テロップの情報を使い，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実現であるが，技術的には以下のような問題点が存在する．

- テロップ認識結果に多くの意味不明な文字列や誤認識が含まれており，どれが必要なテロップかを判断する必要がある．本研究では，テロップの修正を行い，誤認識を考慮したシーンの検索手法を提案する．
- 複数のテロップ認識結果が得られた時，どこからどこまでが一つの論理的に繋がっているシーンかを検出する必要がある．本研究では，テロップの出現時間間隔を利用し，シーンの検出を行う．
- 会話シーンのみを検索したい利用者のために，テロップ認識結果のうち，どれが会話シーンかをシステムが区別できるようにする必要がある．本研究では，シーン検出を行った後，シーン中のテロップの出現時間長及びテロップの個数により，会話シーンの判定を行う．

3.2 システム構成

本システムの構成図を図 1 で示す．本システムは二つのサブシステムから構成される．メタデータ作成サブシステムと検索サブシステムである．

メタデータ作成サブシステムは，検索サブシステムに使われるデータを作成する．以下のステップで処理を行う．

（1） 認識ツールを用いて動画からテロップを認識する．

（2） 認識結果に意味不明な文字列などが含まれているため，ノイズフィルタを用いて除去する．

（3） Web から取得した正しいフレーズデータを用いて，テロップ修正を行う．

（4） テロップの出現時間間隔を利用して，シーン検出する．

（5） シーン中のテロップの出現時間長及びテロップの個数を用いて，会話シーンの判定を行う．

（6） 検索エンジン用の転置インデックス及び Web 上に埋め込むストリーム動画配信のメタファイルを作成し，検出した

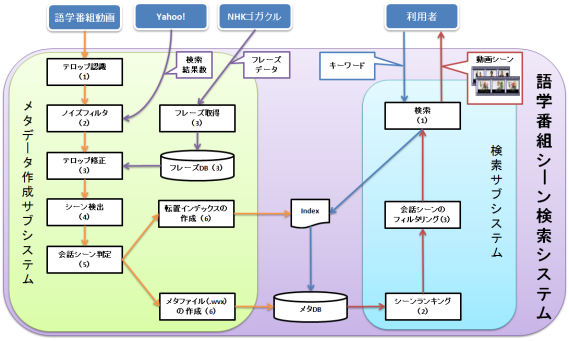

図 1　システムの構成図

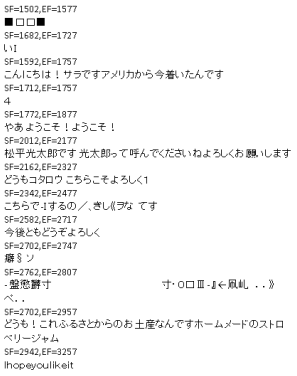

図 2　認 識 結 果

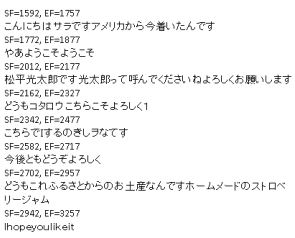

図 3　ノイズ除去後の結果

シーンの情報と共にメタ DB に格納する．

検索サブシステムは，利用者が与えたキーワードを受取り，シーンを検索し，ランキングをする．以下のステップで処理を行う．

（1） 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 DB から探る．

（2） 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．

（3） 利用者が選択したオプションに従い，シーンの提供を行う．

本稿では NHK の英語番組を対象とし，(株）富士通研究所により開発したイメージ文字認識システム [5] を利用しテロップ認識を行い，NHK ゴガクル [1] のフレーズデータでテロップの修正を行う．

本節の以降ではシステムの各ステップの詳細を述べる．まず，メタデータ作成サブシステムのステップ（2）,（3）に必要となるテロップ文字列の類似度について 3.3 で述べた後，ステップ（2）,（3）の詳細を 3.4 で説明する．次に，ステップ（4）,（5）を 3.5，3.6 で述べる．検索サブシステムのステップ（1）を 3.7 で説明を行い，ステップ（2）,（3）は 3.7.3 で述べる．

3.3　テロップ文字列同士の類似度

テロップ認識ツールを用いてテロップ情報を認識する際，すべて正しく認識するとは限らない．認識結果には誤認識や認識漏れなどが存在し，フレーズ DB に蓄えた正しいフレーズで修正する際，どの認識結果がどの正しいフレーズに対応しているかの判断が必要である．

また，認識結果には類似なテロップが多く存在する．これらはテロップを認識する際，動画の背景などが変化したとき，出続けているテロップが新しいテロップと判定され，再認識された結果である．これらの類似テロップを放置しておくと，一つのテロップの正確な出現時間長を把握することができず，会話シーンか否かの判定に影響を及ぼすため，類似のテロップをまとめる必要がある．

そこで，本研究では N-gram [6] を用いて，テロップ同士の類似度を算出する．テロップ a，b の長さを l_a，l_b とし，テロップ a，b における 2-gram の共通キーワード数を C とし，類似度 S を以下のように定義する．

$$S = \frac{C}{max\ (l_a, l_b) - 1}$$

3.4　テロップ修正

3.4.1　ノイズの除去

次にノイズの除去について述べる．本稿におけるノイズとは，動画中にテロップが出現していないにもかかわらず，背景の画像などを誤ってテロップとして認識し，認識結果に意味不明な文字列として出力されたものとする．

テロップ認識ツールを用いた出力結果を図 2 で示す．図 2 の中の“ SF ”はテロップの出現開始時刻フレーム，“ EF ”はテロップの終了時刻フレーム．「SF=***， EF=***」の下の行は表示されたテロップの文字列である．

図 2 の「■口口■」,「い I 」,「癖§ ソ」のようなノイズがテロップ認識出力結果の約 6 割を占めており，ノイズの除去をしないと，後ほど述べるテロップ修正及びシーン検出などに影響を及ぼすため，以下の 3 ステップでノイズの除去を行う．

（1） 記号の除去

語学番組中，一つのテロップに複数個の記号が含まれることは極めて少ない．そこで，「テロップ中の記号の割合 C_k が T_k 以上（$C_k \geqq T_k$）」の場合，そのテロップを除去する．それ以外の場合はテロップに含まれる記号を除去する．

（2） 短いテロップの除去

語学番組は言語の正しい使い方を教えることを目的としているため，不完全センテンスや省略語などは少ない．そこで，「テロップの長さ L がしきい値 T_l 以下（$L \leqq T_l$）」の場合，そのテロップを除去する．

（3） 意味不明な文字列の除去

意味不明かどうかの推測は機械にとって困難であるため，本研究では，YahooAPI [7] で取得したサーチエンジンでのヒット数を用いて判断する．まずすべてのテロップに N-gram を適用し，分割した文字列を空白で繋いで一つの問合せとする．そして，OR 条件で検索し，ヒット数を得る．「ヒット数 R がしきい値 T_m 未満（$R < T_m$）」の場合，テロップを除去する．

図 3 は図 2 の認識結果に $T_k = 0.3$，$T_l = 2$，$T_m = 1000000$，N-gram（$N = 3$）のパラメータでノイズ除去を適用した場合の出力例である．図 3 から分かるように，「■口口■」,「い I 」,「癖§ ソ」など意味不明な文字列が除去されている．

3.4.2　Web 情報を用いたテロップ修正

3.4.1 でノイズを大幅に除去したが，ノイズではない実際に

表 1　一組のフレーズの例

英　語：Nice to meet you.
日本語：よろしくお願いします.
番組名：英語が伝わる！100 のツボ
放送日：2009/09/28

図 4　オーバーラップ

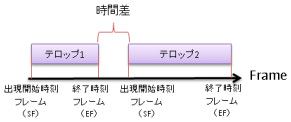

図 5　非オーバーラップ

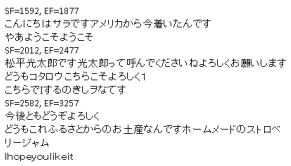

図 6　シーン検出後の結果

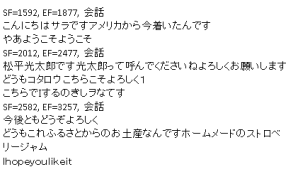

図 7　会話シーン判定後の結果

表示されたテロップに対しては，正しい文字を別の文字に誤認識した場合の修正作業が必要である．そこで，フゥンらが提案したテロップ認識率向上手法 [3] のアイディアに基づき，テロップ修正を行う．

フゥンらは Web 上のニュース記事を大量に蓄えて，ニュース番組のテロップを修正した．本稿の対象は語学番組であるので，語学学習サイト NHK ゴガクルの提供するフレーズデータを用いて，NHK の英語番組のテロップを修正する．NHK ゴガクルには各番組のフレーズデータが計 6000 個以上蓄えられており，1 フレーズは表 1 のような一組で表されている．

ここでは，定期的（週一回程度）に NHK ゴガクルから上記の情報を取得し，フレーズ DB に蓄える．

フレーズ DB の情報とテロップ認識結果を照合してテロップ修正を行う．以下にその手順を示す．

（1）　認識結果と番組名が一致するフレーズデータの類似度 S を測る

（2）「類似度 S がしきい値 T_n 以上（$S \geqq T_n$）」の場合，フレーズデータの内容で置き換える．

3.5　シーン区切り検出

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には，論理的に繋がっているシーンを検出しなければならない．本稿ではテロップの出現時間間隔を利用して，シーンの区切りを検出する．

3.5.1　テロップ間の時間関係

実際の動画中，一つの画面に一つだけのテロップが出現するとは限らない，また同じ画面に出現するテロップの出現開始時刻と終了時刻が一緒とは限らない．テロップ間の時間関係はオーバーラップ（図 4）と非オーバーラップ（図 5）に分けることができる．

オーバーラップになる原因は二つ挙げられる．

- 実際に複数個のテロップが同時に表示された．
- 一つのテロップを出現時間の重なる別々のテロップとして認識した．これは認識ツールの仕様として，長時間にわたり出続けるテロップは，背景などに変化があると，別のテロップとして再認識され，類似したテロップがオーバーラップして出る場合がある．

非オーバーラップになる原因も二つ挙げられる．

- シーンとシーンの切れ目．語学番組中のテロップは主に会話中の字幕やフレーズ解説中の例文などとして，テロップ情報が必要な場面に出現する場合が多い．反対に，シーンからシーンに切り替わる際，テロップは出現しないことが多い．
- 論理的に繋がったシーンの一部のテロップの認識漏れ．

3.5.2　シーン区切り検出処理の流れ

ノイズ除去後の結果に含まれる出現開始時刻フレーム（SF），終了時刻フレーム（EF）を用いてシーン区切りを検出する．

シーン区切り検出は以下のステップで行う．

（1）　テロップ間の時間関係に基づき，オーバーラップか，それとも非オーバーラップかを判定する．

（2）　オーバーラップ区間の場合は必ず一つのシーンの一部であるため，連続したテロップをシーンとしてまとめる．

（3）　非オーバーラップ区間の場合は常にシーンの区切りとは限らない．そこで,「テロップの出現時間間隔 D がしきい値 T_d 以上（$D \geqq T_d$）」であった場合のみシーン区切りとみなす．それ以外の場合シーンとしてまとめる．

シーンとしてまとめる際，3.5.1 で述べた類似なテロップが存在する場合がある．これらの類似した連続テロップを一つのテロップにまとめないと，一つのテロップの出現した時間の長さを正確に得ることができないため，3.6 で述べる会話シーン判定に影響を及ぼす可能性がある．そこで,「テロップ同士の類似度 S がしきい値 T_b 以上（$S \geqq T_b$）」であった場合のみ同一シーンの同一テロップとみなす．それ以外の場合同一シーンに属する別のテロップとして扱う．また，一つのテロップにまとめる際，3.4.1 で得たヒット数 R を用いて，類似テロップ中最も R が大きいテロップを選択する．

図 6 は図 3 の結果に $T_d = 25$，$T_b = 0.5$ のパラメータでシーン区切り検出手法を適用した場合の出力例である．

3.6　会話シーンの判定

語学番組には解説のシーンや会話をしているシーンなどが含まれる．3.5 でシーンの検出を行ったが，利用者が求めている会話が行われているシーンの検出はまだ実現されていない．そこで，本研究ではテロップの出現時間長及びシーン中のテロップの個数を用いて，会話シーンの判定を行う．

会話シーンの特徴は，各テロップの出現時間長が短い，なおかつ一つのシーンに複数個のテロップがある．そこで，本研究では，各シーン中,「すべてのテロップの出現時間長 J がしきい値 T_j 以下（$J \leqq T_j$）」，なおかつ「シーン中のテロップ数が 2 以上」の場合，会話シーンと判定する．それ以外の場合は解説シーンと判定する．

図 7 は，図 6 に対し，$T_j = 565$ のパラメータで会話シーン判定を行った場合の出力例である．会話シーンに対しては,“ EF ”の後に「会話」ラベルが付与される．

3.7 シーン検索

3.7.1 転置インデックス

本システムは，検索の効率性を上げるために，転置インデックスを用いて検索を行う．また，本システムでノイズ除去及びテロップの修正を行ったとしても，すべてのテロップが正しい文字列になるとは限らない．そこで，本システムは N-gram を用いて転置インデックスを作成することにより，検索キーワードに完全一致しないシーンであっても候補として取得することが可能になる．

3.7.2 検索手法

本システムでは利用者から与えられるキーワードは，単語または複数の単語を繋いだフレーズであることを想定している．3.7.1 で述べたが，ノイズ除去及びテロップの修正を行ったとしても，すべてのノイズや誤認識などを完全に除去，修正することは不可能である．そこで誤認識や認識漏れを考慮した検索手法を用いる．

シーン検索は以下のステップで行う．

（1） 入力として複数単語が与えられた場合には連続したフレーズとみなして単語間の空白を除去する．例えば「Nice to meet you」を「Nicetomeetyou」にする．

（2） 得た文字列に対し N-gram を適用する．

（3） 分割した各文字列に対し，転置インデックスから対応するテロップ ID を取得する．

（4） 取得した全てのテロップ ID の和集合を取る．

（5） 各テロップ ID が含まれるシーン情報を取得する．

3.7.3 シーンのランキング

3.7.2 で述べた検索手法は，部分一致検索が可能であるが，関連がないテロップ情報にも多くヒットしてしまうという欠点がある．よりよいシーンを検索結果の上位に出現させるため，シーンのランキングが必要と考えられる．

そこで，利用者が入力したキーワードと 3.7.2 で得た各テロップ ID の N-gram との共通キーワード数 C_h をカウントし，その結果で降順ソートする．従って，入力キーワードとより合致度が高いテロップが含まれるシーンのランクを上げることが可能になる．

本システムでは，利用者に会話シーンのみを検索対象とするか，それともすべてのシーンを検索対象とするかを検索オプションで選択可能である．もし，利用者が会話シーンのみを指定した場合，3.6 で特定された会話シーンのみに対し，前で述べたランキング付けを行う．

4. プロトタイプシステム

前で述べた語学番組シーン検索システムを実装した．メタデータ作成システムは Java で実装し，データベースは Postgres を使用した．検索サブシステムのインターフェースは JSP で実装し，検索画面を図 8 で示す．

入力キーワードが「気に入った」で，会話シーンのみを指定した場合，検索結果は図 8 のようになる．再生ボタンをクリックすると，再生画面（図 9）に移り，再生を自動に開始する．再生画面中の関連動画の欄には，キーワード「気に入った」の全シーンを指定した場合の検索結果が表示される．

図 8 検索画面

図 9 再生画面

5. 評価実験

5.1 実験の目的

本実験の目的は，テロップ情報のみを用いた場合，利用者が与え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どれだけ正しく検索できるかの検証である．この目的を達成するために，以下の三つの評価実験を行った．

- シーン区切り検出手法に関する実験

目的は，テロップ情報のみを用いてどれほどシーンを正しく区切れるかの検証である．また，ノイズ除去を適用した場合としない場合とで，シーン区切りに対する効果も検証する．

- 会話シーン判定手法に関する実験

テロップ情報のみを用いてどの程度会話シーンの判定が正しくできるかを検証でする．

- シーン検索に関する実験

利用者が入力したキーワードに関連ある会話シーンをどれほど取得できるかを検証する．また，会話シーンのみを指定した場合としない場合とで，検索性能を比較する．

本節の以降では，各実験の詳細を述べる．シーン区切り検出手法に関する実験は 5.2 で述べ，会話シーン判定手法に関する実験は 5.3 で説明する．シーン検索に関する実験は 5.4 で述べる．

5.2 シーン区切り検出手法に関する実験

本実験は，テロップ情報のみを用いてどれほどシーンを区切

れるかの検証である．また，ノイズ除去を適用した場合としない場合とで，シーン区切りに対する効果も検証する．

5.2.1 実験データ

今回は 2009 年 9 月〜2010 年 1 月に放送された NHK の英語番組を使用する．

評価のために人間が区切りを指定したものを正解シーンとした．実験データから正解シーンを作成する際,「人間からみて論理的繋がっている」と「長い会話シーンや解説シーン中，内容が変わったら，新たなシーンとする」の二つの条件に従い作成した．

英語番組の内容，動画本数及び正解シーン総数を表 2 で示す．

表 2　実験データ

番組名	動画本数	正解シーン総数
英語が伝わる！100 のツボ	5 個	92 個
ハートで話そう！マジカル英語塾	5 個	141 個
リトル・チャロ　カラダにしみこむ英会話	5 個	150 個
ニュースで英会話	5 個	114 個
トラッド　ジャパン	5 個	133 個

5.2.2 評 価 方 法

正解シーンは人間が作成するため，必ずシーンの開始や終了にタイムラグが生じる．そこで今回は許容範囲 T_c を用いて，システムが検出したシーンの開始時刻フレーム (SF) と正解シーンの開始時刻フレーム (SF') の差の絶対値が T_c 以下，なおかつシステムが検出したシーンの終了時刻フレーム (EF) と正解シーンの終了時刻フレーム (EF') の差の絶対値が T_c 以下の場合，検出したシーンは正しいと判定する ($|SF - SF'| \leqq T_c \cap |EF - EF'| \leqq T_c$)．

今回は F-尺度 (F-$measure$：式 1) を用いて評価する．

$$F\text{-}measure = \frac{2 * precision * recall}{precision + recall} \tag{1}$$

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

$$recall = \frac{|Result \cap Scene|}{|Scene|} \tag{3}$$

ただし，$Result$ はシステムが検出したシーンの集合，$Scene$ は作成した正解シーンの集合を表している．

5.2.3 実 験 結 果

本実験ではテロップ間の時間差でどれほどシーンを区切れるかを調査するため，本実験で使われるテロップの出現時間間隔しきい値 T_d 以外のパラメータは，事前に何回か実験を行い，良い結果を得たパラメータを使用する．ノイズ除去に必要とするパラメータについては $T_k = 0.3$，$T_l = 2$，$T_m = 1000000$，$N = 3$，類似テロップを一つにまとめる類似度しきい値 $T_b = 0.4$，正解シーン判定に必要する許容範囲 $T_c = 240$ に固定した．全動画に対し，テロップの出現時間間隔しきい値 T_d を 25，55，85，115 と変化させた場合に，ノイズ除去を適用する前と適用した場合の F-$measure$ を算出し，平均を取った結果を図 10 で示す．

縦軸は F-$measure$，横軸はテロップの出現時間間隔しきい値 T_d,「N_F」は 3.4.1 で述べたノイズ除去後の F-$measure$ であり,「F」はノイズ除去前の F-$measure$ である．

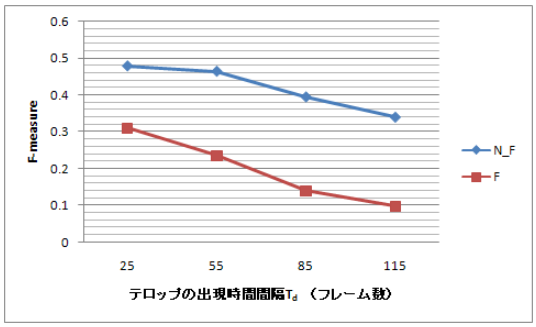

図 10　シーン検出手法適用の結果

本実験の考察は以下である．

- 「N_F」の方が良く,「F」との差は平均で約 2 倍である．その理由として，ノイズが本来のシーンとシーンの切れ目を繋いでしまったことが原因である．ノイズ除去することで，本来の切れ目の部分も明らかになり，適切にシーンを区切ることに成功した．
- テロップの出現時間間隔しきい値 T_d を 25 から 115 まで調整をした結果，ノイズ除去に関係なく，F-$measure$ が下がる傾向が明らかになった．その理由として，今回の正解シーンが全体的に細かく区切って作成されていることが挙げられる．T_d が小さいほど，本システムはシーンを細かく区切る傾向があるため，良い結果を得た．
- この手法はテロップ情報のみ用いてシーン区切りを行うため，テロップが出現する区間しか検出できない．しかし，実際の動画では，テロップが消えた時，確実にシーンが終わるとは限らない．テロップが消えても，まだそれに関して説明している場合がある．検出できなかったシーン区切りのうち，テロップ情報だけでは検出が不可能と思われるシーンは 4 割ほど見られた．そのようなシーンを正しく検出するためには，テロップ情報だけでなく，画像や音声などとの併用が必要であると考えられる．それについては今後の課題である．

5.3　会話シーン判定手法に関する実験

続いて，テロップ情報のみを用いてどれほど正しく会話シーンの判定ができるかを検証する．

5.3.1 実験データ

本実験で用いる実験データは以下のものを用いた．

- 会話シーンを含む動画

本実験は，会話シーンの判定手法の評価をするため，会話シーンが含まれる「英語が伝わる！100 のツボ」と「リトル・チャロ　カラダにしみこむ英会話」を使用する．

- シーン区切り情報

会話シーン判定は，シーン検出後に行うため，シーン検出結果が必要である．そこで 5.2.3 で得た最適なテロップの出現時間間隔しきい値 $T_d = 25$ でシーン検出し，そのうちの正しく検出したシーン情報のみを実験データとして使用する．

- 正解会話シーン情報

上記シーン検出結果中の正しいシーンを用いて，人手により正

表 3　実験データ

番組名	動画本数	検出した正しいシーン総数	正解会話シーン総数
100 のツボ	5 個	88 個	21 個
リトル・チャロ	5 個	72 個	23 個

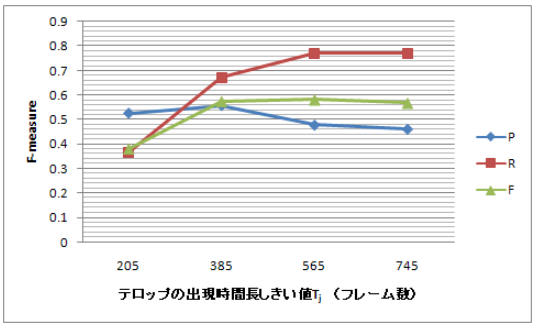

図 11　会話シーン検出手法適用の結果

解会話シーンのラベル付けを行った．

英語番組の内容，動画本数，検出した正しいシーン総数及び正解会話シーン総数を表 3 で示す．

5.3.2　実 験 結 果

前で述べた二つの番組に対し，3.6 で述べたテロップの出現時間長のしきい値 T_j を 205，385，565，745 と変化させた場合に，提案手法を適用した会話シーン判定結果の *Precision*，*Recall*，*F-measure* を算出し，平均値を図 11 に示す．

縦軸は *F-measure*，横軸はテロップ出現時間長しきい値 T_j，「P」,「R」,「F」はそれぞれ *Precision*，*Recall*，*F-measure* である．

本実験の考察は以下である．

- テロップ出現時間長しきい値 $T_j = 565$ あたりで最大の *F-measure* $= 0.581$ を得た．また，T_j を 565 からさらに大きくしても，*F-measure* はそれほど変化がなかった．その理由として，会話シーンと解説シーンに含まれるテロップの出現時間長の差が調査した範囲より大きいことが考えられる．T_j をより大きくすれば，ほぼすべての解説シーンが会話シーンと判定され，*F-measure* は下がることが予想される．
- 本来，すべてのシーンが正しく会話シーンと判定された場合，*Recall* $= 1$ になるはずだが，今回は *Recall* $= 0.77$ であった．本システムが誤って会話シーンを解説シーンに判定した原因は，主に「シーン中テロップ数は 2 個以上」の条件を満たしていなかったためである．実際の会話シーン中，キャラクタの独り言やナレーションが一行のテロップとして表わされているシーンが含まれているため，本手法では，この問題に対応できない．キャラクタの独り言やナレーションが一行のテロップとして表わされているシーンにも対応できる会話シーン判定手法の提案は今後の課題である．

5.4　シーン検索に関する実験

本実験では，利用者が入力したキーワードに関連ある会話シーンをどれほど取得できるかを検証する．また，検索時のオプションとして，会話シーンのみを指定した場合としない場合とでの検索性能についても比較する．

表 4　実験データ

番組名	動画本数	検出したシーン総数	検出した会話シーン総数
100 のツボ	27 個	687 個	338 個
ハートで話そう！	11 個	436 個	147 個
リトル・チャロ	5 個	229 個	108 個
ニュースで英会話	8 個	172 個	28 個
トラッド　ジャパン	6 個	212 個	56 個

表 5　検索キーワードの結果

	会話シーン指定前			会話シーン指定後		
検索キーワード	検出結果	正解	平均適合率	検出結果	正解	平均適合率
こっちがいい	15	2	0.266	9	2	0.583
無理だ	8	4	0.457	4	4	1.000
よろしく	13	6	0.484	7	6	0.734
お金をくずして	8	2	0.320	3	2	0.638
なにが入ってるの？	31	2	0.226	19	2	0.583
気に入った	8	3	0.633	4	3	1.000
大丈夫	15	9	0.567	11	9	0.792
どうしたの？	17	7	1.000	12	6	1.000
それで思い出した	18	4	0.433	9	3	1.000
よかった	38	7	0.883	20	7	0.910
Thanks	53	3	0.600	18	2	0.833
mean	59	2	0.250	22	2	1.000
should	99	10	0.470	40	9	0.989
need	28	2	0.236	13	2	0.450
better	89	3	0.588	31	3	1.000
know	44	5	0.342	25	5	0.624
right	42	3	0.300	20	3	0.532
just	40	7	0.221	17	6	0.392
hope	21	2	0.700	6	2	1.000
like	36	6	0.484	17	6	0.857

5.4.1　実験データ

本実験で用いる実験データは以下の通りである．

- 対象動画

2009 年 9 月～2010 年 1 月に放送された NHK の英語番組のうち，テロップ認識ツールが最後まで正しく認識した動画，計 57 個を使用する．

- 検索サブシステムの索引データ

57 個の動画に対し，5.2.3，5.3.2 で得た最適なパラメータに基づいて，検索用のデータを作成する．作成手順は以下で示し，作成結果を表 4 で示す．

（1） テロップの出現時間間隔しきい値 $T_d = 25$ に基づいて，全動画のシーン区切りを行う．

（2） テロップ出現時間長しきい値 $T_j = 565$ に基づき，検出した全シーンに対し，会話シーン判定を行う．

（3） 前の 2 ステップで得たデータに基づき，転置インデックス及びメタファイルを作成する．

- 検索キーワード

本実験では,「会話シーンテロップに含まれる」と「日常会話に

使われる」という二つの基準に従い，日本語 10 個，英語 10 個のキーワードを選出した（表 5）．

5.4.2 評価方法

本実験では，検索結果がランキングされているため，平均適合率 [8] を用いて評価する．平均適合率は，検索システムの評価に用いることが多く，各正解が表示された順位までの適合率を求め，それらを全正解にわたって平均することで求められる．平均適合率（AP）は，L を検索結果数，N を検索結果中の正解数とすると，式 4 で求めることができる．

$$AP = \frac{1}{N}\sum_{i=1}^{L} P(i)I(i) \tag{4}$$

$$P(i) = \frac{\text{第 } i \text{ 位までの正解数}}{i} \tag{5}$$

$$I(i) = \begin{cases} 1 & (\text{第 } i \text{ 位が正解}) \\ 0 & (\text{上記以外}) \end{cases} \tag{6}$$

5.4.3 実験結果

5.4.1 で説明した検索キーワードを用いて実験を行い，会話シーン指定した場合としない場合とで，各キーワードに対する検索結果の平均適合率を算出した結果を表 5 で示す．また，日本語キーワード，英語キーワード及び全キーワードの平均適合率の平均を取った結果を図 12 に示す．

本実験の考察は以下である．

- 今回の全キーワードの実験結果から見ると，利用者が会話シーンを指定した場合，より多くの求めている会話シーンが上位に現れることが明らかになった．その理由として，会話シーン指定後，多くの解説シーンを除外することによって会話シーンの順位が上がったことが考えられる．
- 日本語キーワードの場合と英語キーワードの場合の平均適合率の結果から見ると，日本語キーワードの方がよりよい結果を得られることが明らかになった．英語キーワードに本検索手法を適用した場合，検索結果が非常に多く，なおかつ検索結果中に多くの同一ヒット数の検索結果が得られたためである．
- 会話シーンを指定した場合としなかった場合，正解会話シーンの数があまり変化しなかったことが明らかになった．その理由としては，5.3.2 で述べた会話シーン判定評価実験の *Recall* が高いことが考えられる．会話シーンの約 8 割が，本システムでも正しく会話シーンと判定できたため，会話シーンを指定した場合としなかった場合とで，取得できる正解会話シーンの数の減少が小さかった．
- 会話シーンを指定した場合でも，数多くの関連が薄いシーンがランキング下位に含まれる．その理由として，今回の検索手法が N-gram であることが考えられる．3.7.3 で述べたように，この検索手法は，部分一致検索を実現できるが，その反面関連が薄い結果も得ることになる．関連シーンが薄いランキング下位の結果を除外する検索手法の改良については今後の課題である．

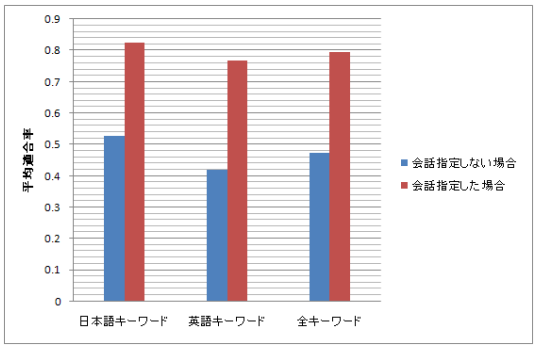

図 12 キーワードの平均適合率の平均値

6. まとめと今後の課題

本稿では，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を提案した．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．これにより利用者は検索時に会話シーンのみを指定できる．本研究では，シーン区切り，会話シーン判定，キーワード検索結果の評価実験を行い，提案手法が利用者の与えたキーワードに関連する会話シーンを検索できることを明らかにした．

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．例えば，ノイズ除去におけるパラメータの最適値の調整やより多様なキーワードを用いたシーン検索などである．次に提案手法の改良を行う．例えば，画像や音声などとの併用でシーンを区切ることが考えられる．また，キーワードの類義語やキーワードを他言語に変換したものを用いた検索なども考えられる．

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．

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] 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.
[4] H. Kuwano, Y. Taniguchi, H. Arai, M. Mori, S. Kurakake and H. Kojima,“ Telop-on-demand: Video structuring and retrieval based on text recognition ”, Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference, vol.2, pp.759-762 (2000)
[5] 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.
[6] 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.
[7] YahooAPI, http://developer.yahoo.co.jp/
[8] 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.