情 報 学 基 礎 61－5
自 然 言 語 処 理 142－5
（2001．3．5）．

マルチモーダル対話コーパス検索/再生ツールの実装

伊藤 一成　斎藤 博昭

慶應義塾大学大学院 理工学研究科 計算機科学専攻
〒223-8522 神奈川県横浜市港北区日吉 3-14-1
Phone: 045(563)1141
Email: {k_ito,hxs}@nak.ics.keio.ac.jp

あらまし

コンピュータと人間の自然なマルチモーダル対話を実現するためには，コンピュータが人間同士のマルチモーダル対話過程と同様な形で処理出来る事が望ましい．そのためには，発話文の意味解析，また表情，感情を表現出来る，動画，音声による人間への自然な応答が不可欠である．意味的・語用論的な構造をテキストに明示する方法としてXMLのタグセットであるGDA(大域文書修飾)が提案されている．本稿では，動画コーパスとその転記テキストをGDA化したデータを用い，意味や構文情報に基づく検索を行い，該当する文節に対応する動画像をユーザに対し提示することが可能なツールについて報告する．

キーワード：マルチモーダル，対話コーパス，GDA，XQL，情報検索

A Search/Playback Tool of a Multimodal Dialogue Corpus

Kazunari ITO　Hiroaki SAITO

Department of Computer Science
Keio University
3-14-1, Hiyoshi, Kouhoku-ku, Yokohama 223-8522, Japan
Phone: +81-45-563-1141
Email: {k_ito,hxs}@nak.ics.keio.ac.jp

Abstract

To express a semantic structure and pragmatic information in a text, GDA(Global Document Annotation), a tagset of XML, has been proposed. Tagging a transcription of a dialogue movie corpus has also been tried using GDA.

This paper reports a tool which enables the user to search a phrase and/or a semantic structure in a tagged corpus and shows the user the matched part of the movie corpus. An XQL format is allowed for search patterns as well as a plain phrase. As for playback functions, various modes are equipped including variable playback speed.

KeyWord: multimodal, dialogue corpus, GDA, XQL, information retrieval

1 はじめに

近年，人とコンピュータの間により自然なコミュニケーション環境を実現するため，声，身ぶり，表情等のマルチモーダル情報を使ってコンピュータと対話できるマルチモーダルインタラクティブシステムの開発研究が多く行われている．しかしながらコンピュータがあたかも本当の人間の様に振舞うためには，機械が人間も理解可能な知識ベースを有し，さらに感情を有する音声や表情を表すことが出来る動画の形式で応答する機構が必要である．

ところでテレビドラマ，ニュース，語学教材などの動画データはエンターテイメントや教育等にまつわる用途をそれ自体として持っているが，これに意味内容に基づくインタラクティブな検索や提示が出来る様になれば，実用的な価値を更に高めることが出来る．そのような対話的なアクセスが可能な知的コンテンツは機械翻訳，情報検索，質問応答，知識発見システムなどを実用化する上で，今後益々必要になってくると考えられる．

現在，テキストに構文・意味等に関する情報を付加するXMLタグセットであるGDA(Global Document Annotation; 大域文書修飾)が提案されている[1]．GDAで記述されたテキストを，音声や動画像と有機的に結び付けることは，これらの技術の基礎研究と応用開発の推進に寄与するであろう．

本稿では，はじめに，GDAテキストと，動画，音声などのマルチモーダルコーパスファイルとの結び付けをするために拡張タグを定義する．その上で，動画コーパスとそれを書き起こしたGDAタグ付きコーパスを用いて，意味や構文情報に基づく検索を行い，該当する文節に対応する動画像をユーザに対し提示するツールについて報告する．

2 GDAタグ付きコーパス

2.1 GDA

GDAは，電子化文書の意味的・語用論的な構造を明示するXMLのタグ集合を策定，公開し，これに基づく多用途の知的コンテンツをサポートする応用技術の開発と普及を推進することにより，インターネット網でこのタグ集合を広めることを目指すプロジェクトである[2]．概要としては，基本的に文章の意味(意味論的および語用論)解析を記述するために使用する．意味と言っても様々あるが，主に主題役割，修辞関係，及び照応に関する情報を記述するためのタグセットである．

```
<q who="A">
<su syn="fc" id="kakure">
 <ad>で</ad>
 <n arg="X">みみ</n>
 <ad sem="obj">は</ad>
 <adp>けっこう</adp>
 <v>かくれ</v>
 <ad>て</ad>
 <v>る</v>
 <v>か</v>
 <v>な</v>
</su>
</q>
```

図 1: GDA タグ付きコーパス例

GDAタグを施したテキストの例を図1に示す．<q>はその部分が直接誰かによって発せられた語であることを意味し，who属性の値はその発話者を表している．<su>は文，つまり，発話の他の部分と統語的な関係を持たない部分を指す．synは文の統語的構造で，fcは前向き連鎖依存関係であることを意味する．<v>エレメントは動詞または動詞句，<n>エレメントは名詞または名詞句，<ad>,<adp>エレメントは副詞や後置詞句や連体詞である．GDAのタグはこのように単語を単位とする細かい統語構造を表示できるように作られている．

2.2 GDAコーパスに対する時刻情報の付与

ある動画ファイルとその書き起こしテキストの関連付けをする場合，テキスト中の発話文が，その動画中で発せられた時刻をテキストに埋め込む方法が主に用いられている．同様の方式でGDAファイルと動画ファイルを関連付けるために新しいタグ，

tst タグ (タイムスタンプタグ) を定義した．tst タグはそれ自体要素を持たない，空要素タグであり，以下の様に記述される．

<tst val=”発話開始時刻 (秒)”/>

tst タグは各文節毎に付けられる．先述の GDA タグ付きコーパス (図 1) に tst タグを追加した場合の例を図 2 に示す．

```
<q who="A">
<su syn="fc" id="kakure">
 <ad>で<tst val="60.453870"/></ad>
 <adp rel="obj">
 <n arg="X">みみ<tst val="60.815086"/></n>
 <ad>は<tst val="61.100256"/></ad>
 </adp>
 <adp>けっこう<tst val="61.604057"/></adp>
 <v>かくれ<tst val="61.907619"/></v>
 <ad>て<tst val="62.192223"/></ad>
 <v>る<tst val="62.383523"/></v>
 <v>か<tst val="62.615927"/></v>
 <v>な<tst val="62.849301"/></v>
</su>
</q>
```

図 2: 時刻情報付き GDA コーパスの例

3 マルチモーダル対話コーパス検索/再生ツール

本章では実装したツールについて，機能および操作方法について述べる．

3.1 動作環境

本ツールは Java で実装されている．また Java プログラム上で動画を再生するために JMF(Java Media Framework)[3] を使用している．よって Java2 をサポートする JRE(Java 実行環境) 及び，JMF がインストールされているマシン上であれば，OS を問わず実行可能である．

3.2 使用したデータ

本ツールの実装にあたり社団法人 電子情報技術産業協会の対話理解技術専門委員会 [4] が提供している，対話コーパスを使用した [5][6]．この対話コーパスは 1999 年 6 月に，協会より CDROM 媒体にて配布されている．動画コーパスの内容は，2 人が特定のタスクに対して対話を行っているもので，1 分から十数分程度の 9 対話からなる．今回は，GDA ファイルには手作業にて tst タグを追加した．

3.3 画面構成

はじめに GDA ファイルと動画ファイルを指定すると，図 3 に示す実行画面が表示される．本ツールはメインウィンド及び複数の内部ウィンド (図 3-[1] ～ [4]) から構成されている．メインウィンドは内部ウィンドを表示する部分，及び検索を行うための検索式を入力する部分（図 3-[5]）から構成される．

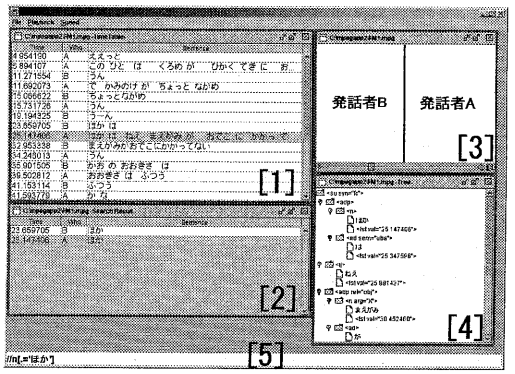

図 3: 実行画面 (動画部分は肖像権の都合上カットした)

3.4 各ウィンドの機能

タイムテーブルウィンド （図 3-[1]）
読み込まれた GDA ファイルを基に，各文に対して，発話開始時間，発話者，テキスト部分を抽出したものをまとめて表形式にて表示する (図 4 参照)．また，動画の再生されている部分の行の背景色が変化して表示される．再生したい行をクリックすると，その文に対応する動画が再生される．

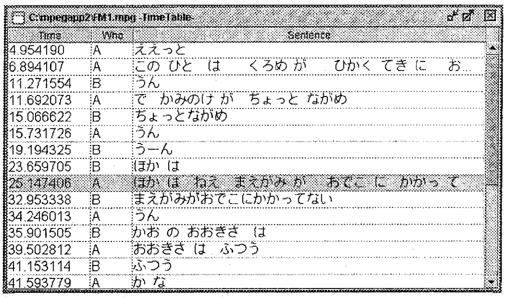

図 4: タイムテーブル表示ウィンド

検索結果表示ウィンド (図 3-[2])
検索式に適合した文節または文に関する情報を，タイムテーブル表示ウィンドと同一形式で表示する(図 5 参照)．また再生したい行をクリックするとその文節を含む文に対応する動画が再生される．

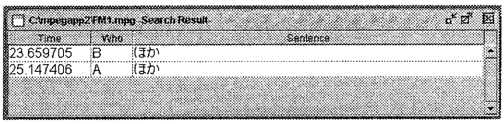

図 5: 検索結果表示ウィンド

動画再生用ウィンド (図 3-[3])
読み込まれた動画を表示，再生する．またウィンド下部のコントローラにて，任意場所での再生，停止が可能である．
GDA のツリー構造表示用ウィンド (図 3-[4])
現在再生されている文に対応する GDA ファイルの木構造を表示する (図 6 参照)．これにより文の意味情報，構文情報がリアルタイムに確認出来る．

3.5 検索および再生方法

メインウィンドのテキスト入力フィールド (図 1-[5]) に検索式を入力することで，読み込まれた GDA ファイルに対して検索を行う．入力する検索式は XQL (XML Query Language)[7] で定義されている入力を受け付ける．XQL は W3C にて勧告化が進んでいる，XML をデータモデルとして使用する照会言語である．XQL は，すでに Web ブラウザー，文書保管システム，XML ミドルウェア，Perl ライブラリ，コマンド行ユーティリティといった多方面にわたるソフトウェアに実装されている．

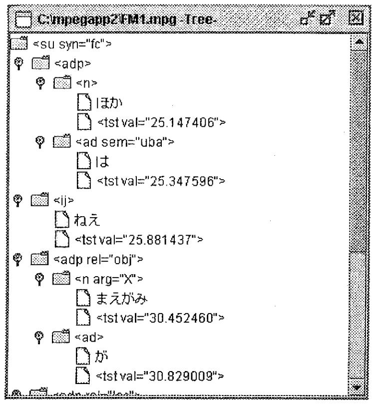

図 6: ツリー構造表示ウィンド

これにより構造に基づく検索が可能となる．

GDA ファイルに適用出来る XQL 検索式の例を次に示す．

- //su　すべての文
- //n[.=' はい']　「はい」という名詞句
- //q[@who='A']　発話者が A である文

図 7 に実際に検索を行った場合の表示例を示す．図 7 では，発話者が'B' かつ名詞で' かいとう' という条件で検索し，その結果，検索結果表示ウィンドに 1 件の該当したものが表示されている．そして，テーブルの各行をクリックすることにより，動画表示ウィンドにて，該当部分の動画が再生される．

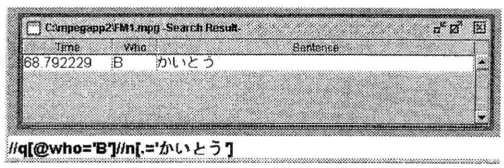

図 7: XQL 検索式による結果結果例

また，単語そのものを，テキストフィールドに入力することにより，単純なマッチングによる検索を行うことが可能である．/で始まるフレーズを検索するには一重引用符で囲めばよい．図 8 に例を示す．図 8 では，' おおきさ' で検索し，該当する 2 件が表示されている．この場合も同様の方法で再生が可能である．

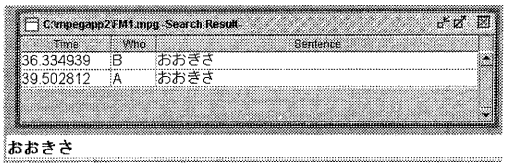

図 8: 単語入力による検索結果例

3.6 その他の機能

ツール上部のメニューバーから，動画再生方法に関する設定が可能である．

- **検索結果に対する再生区間指定**
  デフォルトでは，再生する文を指定した場合，その文のみを再生する．しかしながら，その文を含む対話の部分的な内容を把握するためには，前後数文の情報が必要な場合も多い．また実際の発話開始時刻より 1，2 秒前から再生し，また発話終了時刻後も暫く再生し続けた方が，利用者側からみてもその文全体が聞き取り易い．本ツールでは，(1) 指定文のみ，(2) 前後 1 文を含む，(3) 前後 2 文を含む，(4) 文の前後 1 秒間を含む，(5) 文の前後 2 秒間を含む，(6) 最終まで停止せず再生，の 6 通りの再生方法をメニュー項目から指定出来る (図 9 参照)．

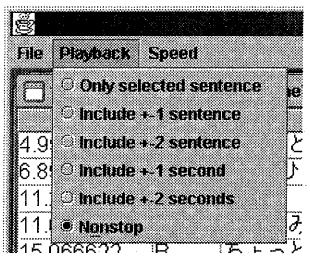

図 9: 再生区間の設定画面

- **再生速度倍率指定**
  動画の音声，及び映像情報を転記する場合は，本来よりも低い速度倍率で再生しながら行う必要がある．また対話の全体内容を把握する場合，逆に速度倍率を上げて再生することにより，それに必要な時間を大幅に短縮する事が出来る．本ツールでは，メニュー項目から指定することにより，任意の速度倍率で再生する事が可能である (図 10 参照)．

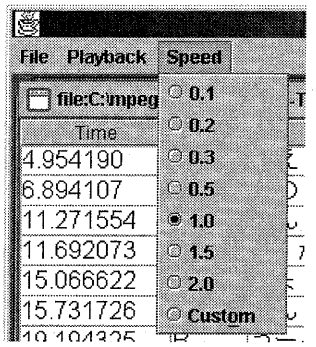

図 10: 再生速度倍率の設定画面

4 課題及び展望

今後，様々な機能を追加していく予定である．具体例を以下に列挙する．

- **検索を行うための GUI の開発**
  現在の方法では，条件に合致する文節や文が存在することを前提とした検索であるので，結果があるとは限らない．候補として存在しうる文，文節，品詞を順に提示しながら候補を絞って行くトップダウン方式による検索方法が必要と考える．また係受け関係まで検索の条件に含める場合は，現在の XQL のみによる方式では非常に難しい．そのため，W3C で勧告され，XML 木構造のノードの親子及び兄弟関係を記述するのに適した Xpath[8] のパーザを組込む予定である．さらにユーザの利便性を考えれば，XQL あるいは Xpath に準じた検索式を直接入力するのではなく，それら両者のフロントエンドとして働く GUI が必要である．

- **複数ファイルに対する検索**
  現状ではローカルに保存された単一のファイルに対してのみ検索が可能である．今後はネットワーク上に存在するコーパスファイルを多数蓄積したデータベースに対して検索要求を

行い，必要な情報のみをローカルにダウンロードし再生するといった，クライアント・サーバモデルへの対応を行う．

- **他のマルチモーダルデータとの融合による検索方法**
  音声の韻律情報を記述したJ-ToBIファイル，またFACSに基づいて人の表情について表記する表情タグを付与したテキストと関連づけることにより，疑問形の'ほんとう'，ためらいの'そうですか'といった検索が可能となる．また音声だけではなく，映像情報を転記したテキストとも関連付けを行う予定である．これにより身ぶり，動作，視線，場所といった情報も検索条件に含むことが出来るようになる．また動画の情報から該当するテキスト部分を抽出する逆検索等も可能となる．

- **自動タグ付け，関連付け機能の追加**
  今回使用したデータは人手によって，GDAタグ付け及びタイムスタンプ付与を行ったものである．このようなデータは，基礎研究のために必須だが，大量に作成し普及させるためには，テキストからタイムスタンプ付きGDAファイルを手間をかけることなく自動生成できる機構が必要不可欠である．書き起こしテキストに形態素解析，構文解析ツールを使用し，さらにフィルタを施すことによりGDAファイルを自動的に生成するシステムが提案されている[9]．しかし，現状の音声認識技術の精度を考えれば，生成されたGDAファイルと動画の同期を取って，自動的にかつ，正確にタイムスタンプ付与することは難しい．しかしながら人手であっても，動画を再生しながら各発話の開始をクリックして指定する事によりGDAファイルに対するtstタグの付与は可能であり，この機能を追加する予定である．

- **モジュール化**
  質問応答システム，知的エージェントシステム等に対して本機能を組込むために，本ツールのモジュール化を進めている．また本ツールでは機能を追加する場合，新しく内部ウィンドを追加するだけでよく，容易に機能拡張ができる．さらに用途に応じて必要な内部ウィンドだけ利用するといったことも可能である．

5 まとめ

本稿では，マルチモーダル対話コーパス検索/再生ツールについて述べた．はじめにGDAコーパスファイルに時刻情報を付与するための拡張タグを定義した．次にツールの機能説明を行った．今後は課題及び展望で述べた様々な拡張機能を実装していく予定である．

謝辞

本ツールを開発する機会を与えて頂き，また作成過程において多大なる助言をいただきました橋田浩一さんをはじめとする社団法人電子情報技術産業協会の対話理解技術専門委員会の方々に感謝します．

参考文献

[1] The GDA Tag Set ホームページ
http://www.etl.go.jp/etl/nl/GDA/tagset.html

[2] 橋田 浩一, "GDA 意味的修飾に基づく多用途の知的コンテンツ", 人工知能学会論文誌, Vol. 13, No4, pp.528–535, 1998.

[3] JMF ホームページ
http://java.sun.com/products/java-media/jmf/index.html

[4] 社団法人 電子情報技術産業協会 対話理解技術専門委員会ホームページ
http://it.jeita.or.jp/jhistory/committee/mmc/mmc.htm

[5] "自然言語処理システムに関する調査報告書" 社団法人 日本電子工業振興協会, 2000-3.

[6] 金子 拓也, 石崎 俊, "マルチモーダル対話コーパスの構築—マルチモーダルデータのタギングについて—" 電子情報通信学会 思考と言語研究会, TL99-3, pp.17–23, 1999.

[7] XQL ホームページ
http://www.w3.org/TandS/QL/QL98/pp/xql.html

[8] Xpath ホームページ
http://www.w3c.org/TR/xpath

[9] 鈴木 潤, 橋田 浩一, "GDA タグを利用した回答抽出システムの提案", 言語処理学会 第 7 回年次大会, 2001.(掲載予定)