ネットワークアクセス可能な機能的音声データベースシステムの概念設計

川下 太郎，柳田 益造

同志社大学

〒 610-0321 京都府京田辺市多々羅都谷 1-3
E-mail: dta0715@mail4.doshisha.ac.jp, myanagid@mail.doshisha.ac.jp

あらまし　音声データベースは，音声に関する全ての局面において極めて重要な研究基盤である．音声データベースでは，単に音声波形，テキスト，音素表記，発話者に関する情報などの他に韻律タグ，形態素情報，統語情報などの入力，音声資料の追加・補充が可能でなければならず，また種々の条件による検索，例えば音素列，韻律，テキスト，文法カテゴリ，発話者条件，発話状況などによる検索ができなければならない．さらに音声データベースは音声認識システムの評価に使えるという点も重要である．

しかしながら，音声データベースの問題点として，欲しい情報をなかなか検索できないことや，配布に時間とコストがかかる，あるいはクライアント側で必要な機能を持たないと何もできないということがある．ここではそのような問題点を解決する機能的音声データベースシステムの概念設計が提案されている．

キーワード　音声データベース，データベース管理システム，インターネット，関係データベース

Conceptual Design of A Network Accessible Functional Speech Database

Taro KAWASHITA and Masuzo YANAGIDA
Doshisha University

1-3, Tatara-Miyakodani, Kyo-Tanabe, Kyoto, 610-0321 Japan
E-mail:dta0715@mail4.doshisha.ac.jp, myanagid@mail.doshisha.ac.jp

Abstract　Speech database is an important base for research concerning speech. It should have facilities of storing not only speech signals, text data, phonemic description and speaker description, but also tags for prosody, morpheme and syntactic structure and facilities of adding new speech data to existing database and retrieving by phoneme symbols, prosody, texts, grammatical categories, speaker specification, utterance circumstance and so forth. In addition to that, it is important that it can be used as materials for evaluating speech recognition systems. Conventional speech database, however, generally has difficulties in retrieving what users want and in consuming time for distribution, moreover it is of no importance if a client has no necessary facilities. This paper proposes conceptual design of a network accessible functional speech database that might solve the problems mentioned above.

Keyword　Speech Database, Database Management System, Internet, Relational Database

1　はじめに

音声は，人間が用いるコミュニケーション手段の中で最も重要なものである．話し手は，聞き手に伝えたい内容を文の形式に構成し，発声器官を動かして相手側が知覚可能な音声信号を生成する．音声認識の主目的は連続的に発声された音声信号を観測して，それを音素や音節あるいは文章を表す離散的な言語記号に変換することである．更に発展させて，その意味を抽出することを目指す場合には，音声理解と呼ばれる．現在の音声研究において，この音声理解システムの構築は大きな目標であり，実用的な場面で使えるものが求められている．

音声研究を行うにあたり，音声は発話者の個人性や心理状態や発声状況によって大きく変動し，また発話者の出身地や年齢や社会的地位などによっても表現が異なることがあるため，大量の音声データが必要不可欠である．大量の音声データを扱うことによって個別の要因に影響されない普遍的な特徴を発見したり，あるいは体系的な変動モデルを構築することが可能になると考えられる．また音声分析や音声認識システムの動作確認のためには，多様な，かつ多数の話者の発声した大規模な音声資料から成るデータベースが必要である．

音声処理技術の研究開発における音声データベースの重要性は古くから認識されており，音声データベース構築の試みがこれまで多くの研究機関でなされてきた[1, 2, 3, 4]．しかしながら，構築されたデータベースが複数の研究目的に効率的に共同利用されている例は，音声認識システムの構築あるいは評価以外には少ない．音声データベースを構築するためには必ず目的が存在する．音声データベースは，その構築目的に沿うように構築されている．よって個別の研究目的のための大規模音声データが収集・蓄積されていても，その中から本来の目的とは異なった研究目的に合ったサブセットあるいは特定音声区間を検索・抽出することが困難であると考えられる．

音声データを多様な目的の音声研究で利用しやすい形にするためには，音声のどの部分がどの音素に対応しているかを示す音素ラベリングの作業や，場合によっては韻律の抽出等の作業が必要である．しかし，フォルマント周波数や基本周波数を含めて，これらの音声情報を自動的かつ高信頼度に得る方法がまだ確立されていなかったり，不十分であったため，人間が視察によってこれらの情報を確認したり抽出したりしていた．そのため，これらの情報を備えた音声データベースの構築には膨大な時間と労力が必要であった．このため大規模音声データベースを構築するには，これらの音声情報を可能な限り自動的に獲得するシステムを装備することが必要である．ラベリング作業に関してもいくつかの研究がある[5, 6, 7, 8]が，まだ大部分が人手で行うもので十分なものではない．そこで本研究では，これらを支援する機能を備えた音声データベースシステムの検討を示す．

2　データベースシステム

コンピュータ技術やネットワーク技術の急速な進展に伴い，各種コンピュータアプリケーションが対象とするデータは，いっそう多様化，複雑化，大規模化している．このような状況下では，各種アプリケーションが取り扱うべきデータ資源を有機的に統合して蓄積管理し，効率的な共有と，より高度な利用を図ることが必要となる．この要求を満たすものがデータベースシステムである．20 年ほど前までのデータベースは CD-ROM で配布される形態であったが，近年はインターネットの普及に伴い，Web 環境で利用されるデータベースが多くなってきた．

この Web 環境で利用されるデータベースは一般的に「Web データベース」と呼ばれ，クライアントのブラウザから Web サーバにアクセスし，Web サーバを介してデータベースに蓄積されているデータを得るという形態を持つ．インターネットを利用するため，データの格納場所を問わず，世界中の様々な場所からアクセスすることが可能である．クライアントではブラウザのみを使用するため OS やソフトウェアを問わないといった利点も挙げられる．

したがって，データベースを Web で公開することは，蓄積された資料を幅広く利用でき

るようにするために非常に重要となる．ただし，この場合，クライアント側に必ずしも利用のためのツールが備わっているとは限らないという状況にも対処できるようにしておくことが要求される．

3　音声データベース

音声データベースは単に音声を記録・保存するだけでなく，どういう人が発声したどのような音声がどこに保存されているかについての情報を持っている．これによって，指定した語や文字を即座に音声として聴取することはもちろん，指定した条件を満たす音声データを取り出したり，指定したアクセント型を持つ語を聞いてみたり，指定した音声データ群を音声認識の学習用や評価用に取り出すことができる．音声データベースは，従来から様々な機関で構築されているが，その機関に所属する研究者の研究のために構築されたものが多い．そのためそれを使うにはその機関と同じような能力を持った計算機と大容量外部記憶装置を必要とし，データベース自体はCD-ROMの形で配布されることが多かった．

コンピュータ技術が進歩するにつれて処理可能なデータ量が増大し，そのためデータベースのデータ量が大幅に増してきた．最近，特に音声の研究では統計的手法の発達により大量の音声データがシステムの学習のために必要とされるようになった．

音声情報処理システムの研究・開発を行うためには，分析・合成・認識の各種の手法を適切に比較・評価することが必要とされる．これを行う方法としては現在のところ，共通の音声データを用いてこれらの処理を行い，その動作を比較するという方法が採られる．このようなことから，共通利用可能な各種・大量の音声データを収録し，保管・公開することは研究・開発過程での利用および認識システムの性能評価の両面から強く求められている．この目的のために用いられる音声データに対しては，音声学的な条件による検索よりも音声データそのものの質と量が重要であると考えられる．

一方，音声学あるいは音韻・韻律関係の研究のためには，音声学的な条件検索機能が必須である．たとえば，関西方言話者で話中の「が」を/ŋa/で発声しているケースを抽出するとか，単独発声では“LH”となる2モーラ語が，後に格助詞「が」が付くと“LL”になるケースを抽出する，あるいはもっと複雑な条件，例えば，単独では“LH”であるが，後に「が＋述語」が続くときに,「が」の高さが後続の述語が高起式か低起式かによって“L”になるべき場合（例：関西方言の「肩が痛い」），“H”になるべき場合（例：関西方言の「肩が凝った」）があるが，それを誤って発声している音声資料をデータ中から抽出するというようなことができることが望まれる．

4　機能的音声データベースシステムの概念設計

大規模な汎用音声データベースを効率的に種々の研究目的に利用していくためには，蓄積された音声データの検索・提供を行う音声データベース管理システムが不可欠である．このような機能をもったデータベースを機能的データベースと呼ぶことにする．ここでは音声を主なコンテンツとした「機能的音声データベースシステム」についての検討を示す．

4.1　要求仕様

これまでの音声データベースの問題点である欲しい情報をなかなか検索できないことや，配布に時間とコストがかかる，あるいはクライアント側で必要な機能を持たないと何もできないということを考慮した結果，音声データベースシステムを利用して多様な音声研究を進めるためには，音声データベースシステムは以下の機能を備えることが望まれる．

1. ユーザは自分の計算機に音声処理用のツールやプログラムを持っていなくてもネットワーク接続さえできれば音声データベースを利用できること．
2. ユーザが提供の意思表示をした音声データやラベル情報ならびにツール，あるいは既登録のそれらについて，管理者がその追加，削除を統一的に行えること．
3. 音声データ検索のための種々の条件をユーザが定義できること．また，その検索を実行するために必要な情報がデータベー

スに欠けていればそれを分析等によって動的に生成することができること．

4. 音声データベース作成のための音声の切り出し，音素の手動や種々のラベリング支援プログラム，音声分析ツールなどのユーティリティが整備されていること．
5. クライアント側から C 言語や JAVA 言語で書かれたプログラムによってアクセスできること．
6. ユーザが希望すれば，その資格に応じて，データをダウンロードできること．

4.2 システムの構成の検討

4.2.1 システムの基本設計

上記の要求仕様を満たすものとして Windows を OS とするシステムと UNIX と OS とするシステムの二種類のデータベースシステムを実験的に構築し，どちらが有用か検討する．

Windows を用いたシステムのための OS としては，インストールが容易であり，高速なネットワークファイルアクセスを提供できることを考慮して，サーバ用 OS である NT Server 4.0 を採用する．

UNIX を用いたシステムのための OS としては，互換性が高く，安価で，ソースコードが容易に参照でき，自由に改変することができる点を考慮して，Linux を採用する．本研究ではその中で日本語環境に優れている VineLinux を採用する．

4.2.2 Web サーバ

Apache は現在 Web 上で最も多く採用されている Web サーバソフトであり [18]，フリーソフトでありながら，多彩なプラットフォームで動作し，また，これに関する書籍や Web 上での情報も豊富であるので，機能的音声データベース構築のための Web サーバとして最適であると考え，これを両システムに採用する．この Apache のサーバ側アプリケーションとして Servlet を導入する．Servlet は Web サーバで動作し，パフォーマンス，データベース接続性，安定性，およびセキュリティの面で高い機能を備えている．

4.2.3 データベース管理システム

データベース管理システムには複数ユーザからのアクセスを処理する機能や，入力されているデータを保護するセキュリティ機能などが求められる．また，サーバ上での安定した動作や Web サーバとの連携性が必要とされる．

ここではデータベース管理システムとして NT Server，VineLinux の両 OS で動作し，Web サーバとデータベースの接続が比較的容易な MySQL を採用する．

4.2.4 Web サーバとデータベースの接続

両システムともに JDBC[1] というデータベースアプリケーションとデータベースの間の Java インタフェースを用いて接続を行う．MySQL の JDBC ドライバを用いることにより，プラットフォームに依存せずにアプリケーションの開発が可能となる．

4.3 二つのシステムの比較

これまで述べてきたように，本研究では二種類の OS を用いてサーバを構築してきた．その比較を以下に述べる．

(a) プログラム開発に要する労力
Servlet は，Web サーバにカスタム機能を追加でき，Java 言語で書かれたサーバ側コンポーネントである．Java 言語は，オブジェクト指向言語であることが特徴で，ネットワーク対応であり，セキュリティに優れており，様々なコンピュータ上で動作し，一度作成すればどのプラットフォームでも動くことなど，利点が多い．これにより，両システムでプログラム開発に要する労力に差はないと考える．

(b) 応答時間
両システムで検索の応答時間を測定した．2 万件のデータに対して SQL をクライアントのブラウザからサーバに 100 回または 500 回発行し，結果が出力するまでの時間を測定した．結果を Table 1 に示す．Linux のシステムの方が NT Server に比べ応答時間が若干短かった．

(c) セキュリティ
Web サーバを構築し公開する際に，最も注意しなければならないのはセキュリティである．ネットワークアクセス可能にすると不特定多数のユーザがアクセスする

[1] Java Database Connectivity

Table 1: 両システムの応答時間

発行回数	NT	Linux
100 回	28 秒	23 秒
500 回	128 秒	105 秒

ため，クライアントはサーバを構築する際に指定されたファイルやデータベース以外にはアクセスできないようにする必要がある．特にデータベースをインターネットに公開する場合，データベースのデータが悪意あるユーザによって変更されることがないようにしなければならない．しかし，Web サーバのセキュリティホールは日々発見される一方，セキュリティホールを埋めるプログラムも開発者から提供されている．その際，セキュリティホールの発見からプログラムの配布までの時間が問題となるが，オープンソースの Linux ならばその時間が非常に短いと言える．このため，Linux で構築する Web サーバの方が NT によるよりも安全性は高いと言える．

4.3.1 総合比較

これまで述べてきた結果より，NT Server の長所は容易なインストール，Linux の長所は安全性と安定性である．これらのことより，Linux がここで考えているデータベースシステムの構築に適していると考える．

Table 2: 両システムの比較

	NT	Linux
プログラム開発に関する労力	○	○
応答時間	△	○
セキュリティ	△	○

4.4 システム構成

本研究で提案するシステム構成を Fig. 1 に示す．本システムは，サーバ側が Web サーバに Apache，サーバ側アプリケーションとして Servlet の Tomcat，データベース管理システムとして MySQL で構成される．サーバ側アプリケーションとデータベース管理システムは JDBC を用いて接続を行う．クライアントからはブラウザでアクセスを行い，サーバはそれに備えて常時待機している．Fig. 1 にそってシステム全体の動作の流れを説明する．

4.5 システムの動作

(a) ユーザ登録
新規ユーザの登録は提供される各種の音声資料のデータ収集の目的，収録データの内容，被験者構成，データ数 , 収録概要，データベースの利用条件，配布条件などを確認の上，申請を行う．個人情報を送信する場合は，安全性を考慮して SSL[2]を用いて暗号化を行い，音声データベース利用のための誓約書の提出にはデジタル署名を用いる．

(b) アクセス
クライアント◯が研究に用いる音声データや分析データを入手するために Web サーバにブラウザでアクセスしする．サーバは音声データへのアクセスを認証したユーザにのみ検索を許可しているため，このアクセス制限機能によりクライアント側にユーザ認証画面を表示させる．サーバは入力された ID, パスワードをデータベースより認証するユーザリストと照合し，ユーザ認証に失敗した場合は認証失敗画面を出し，成功した場合は音声データベース利用目次を表示する．

(c) 検索
研究に用いる音声データを取得したいユーザは，ユーザごとに許可された音声データベースの中から目的にあう音声データを様々な音素表記，カナ漢字表記，形態素情報，構文情報，統語情報などの条件によって検索することが可能である．

(d) データのダウンロード
サーバ側で準備されている分析ツール以外によってクライアント側が分析を行いたい場合は，Fig. 1 のクライアント◯のように研究に用いる音声データ群をサーバから取得し，目的を達成することができる．

[2] Secure Socket Layer

(e) 分析ツールの利用
サーバ側が提供する分析ツールの分析結果のみの取得も可能である．サーバ側はデータベース上にユーザが定義した検索に要する情報を持たない音声データ関しては分析ツールを用いて動的に音声データの分析を行い，それに基づいて検索を行う，分析結果は，一旦テンポラリファイルに格納される．分析結果をデータベース本体に格納するべきかどうかは管理者が判断する．分析において，自動的に行える処理と人手をかけてやるべき処理を区別し，自動化可能な処理は自動化し，完全自動化不能な処理はどこまでを自動でやらせるかをユーザに指定させて対処する．

(f) ユーザからのデータ等の提供
クライアントは，当然，内容の変更，追加，削除はできないようにすべきであるが，それらの提案はできるようにする．

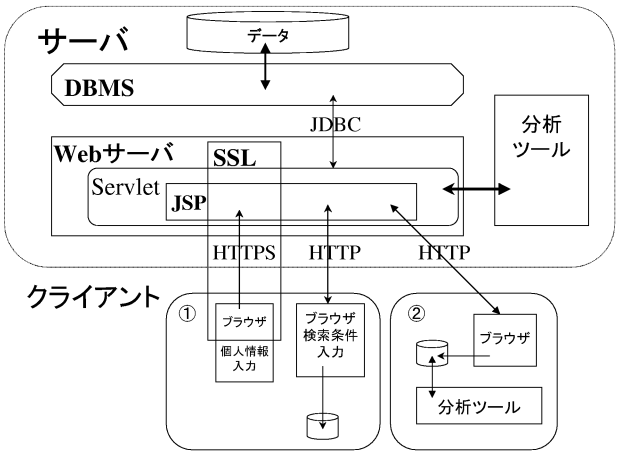

Fig. 1: Configuration of a Network Accessible Functional Speech Database

4.6 音声データの格納方法

音素表記，カナ漢字表記，形態素情報，構文情報，基本周波数，フォルマント周波数などの格納方法を解説する．関係データベースは，複数の 2 次元のテーブルを提供するのに対して，音声データベースの情報では単純な単語の並びといった線形関係や構文情報や係り受け関係といった階層的な関係も記述する必要がある．それについては個々の階層を一つのテーブルに対応させることで格納することができる．各テーブルとして文，形態素が与えられるときには，各テーブルに一意な通し番号の ID を付与する．次に存在する線形順序関係はシーケンシャルな ID で表現し，係り受けの順序関係は各 ID へのポインタとして表現する．各テーブルの属性は，一意な ID と非線形な順序関係が必要な場合は階層構造を表現するためのポインタと，そのテーブルに付与された属性から構成される．Table 3〜8 に各テーブルの例を示す．Table 9 は，文を形態素解析にかけてその結果を格納したものである．なお，外来語辞書テーブルは単語辞書テーブルから外来語だけを抜き出して作成したテーブルであるので，同じ ID が存在することはない．

4.7 検索事例

関西方言者の音声データを次の条件で検索したい場合について考える．例えば，関西方言話者で，外来語を日本語の韻律則から逸脱して，原語の韻律に近い韻律で発声している音声データを探す場合について考える．関西方言話者とは出生地，育成地が関西であり，父，母が関西出身者の話者と定義する．

原語アクセント位置が関西アクセントと異なっている単語が入っている文を検索し，話者が関西方言者である文を抽出することができる．以下の SQL の問い合わせにより実現可能である．

関西方言話者が日本語アクセントでなく原語のアクセントで発話している文を抽出する SQL

```
SELECT DISTINCT 話者.氏名, 外来1.単語, 文.文
FROM 外来語辞書 as 外来1, 外来語辞書 as 外来2,
        形態素解析, 話者, 文,
        方言 as 方言1, 方言 as 方言2,
        方言 as 方言3, 方言 as 方言4, 発話
WHERE 外来1.原語≒外来2.関西 and
        外来1.ID=形態素解析.単語 ID and
        形態素解析.文 ID = 文.ID and
        文.ID = 発話.発話 ID and
        発話.話者 ID = 話者.ID and
        方言1.方言='関西' and
        方言2.方言='関西' and
        方言3.方言='関西' and
        方言4.方言='関西' and
        話者.育成地=方言1.都市 and
        話者.出生地=方言2.都市 and
        話者.父=方言3.都市 and
        話者.母=方言4.都市
```

Table 3: 文テーブル

ID	文
1023	キャリアは学校で決まるものではない.
⋮	⋮
2051	サグラダファミリアはバルセロナにある.
⋮	⋮

Table 6: 方言テーブル

都市	方言
大阪	関西
京都	関西
神戸	関西
姫路	関西
東京	関東
横浜	関東
⋮	⋮

Table 5: 発話テーブル

話者 ID	発話 ID
8	1023
8	2051
⋮	⋮

Table 7: 単語辞書テーブル

ID	単語	品詞	補足
152	は	係助詞	
160	で	格助詞	
161	に	格助詞	
⋮	⋮	⋮	⋮
1094	学校	名詞	
3562	もの	名詞	
⋮	⋮	⋮	⋮
12965	決まる	自動詞	五段・ラ行
13094	ある	自動詞	五段・ラ行
14477	ない	特殊・ナイ	
14478	だ	特殊・ダ	
⋮	⋮	⋮	⋮

5 結論

ネットワークアクセス可能な機能的音声データベースシステムの概念設計を示した．今後，本システムの実装を行い，性能評価やユーザインタフェースの評価を行うことが必要である．

謝辞

本研究の一部は，文部科学省科研費（特定領域研究 (B)「韻律」）および同志社大学学術フロンティア事業「知能情報科学とその応用」の援助を受けた．

参考文献

1) 板橋秀一：「騒音データベースと日本語共通音声データ」，日本音響学会誌，47 巻，12 号，pp.951-953，(1991).
2) 田中和世，速水悟，山下洋一，鹿野清宏，板橋秀一，岡隆一：「RWC 計画における音声対話データベースの構築」，情報処理学会研究報告，96-SLP-11-7，pp.37-42，(1996).
3) 匂坂芳典，浦谷則好：「ATR 音声・言語データベース」，日本音響学会誌 Vol.48，No.12，pp.878-882，(1992).
4) 小林哲則，板橋秀一，速水悟，竹沢寿幸：「日本音響学会研究用連続音声データベース」，日本音響学会誌 Vol.48，No.12，pp.888-893，(1992).
5) 武田一哉：「音声データベース構築のための視察に基づく音韻ラベリング」，ATR technical report，(1987).
6) 壇辻正剛：「音声データベースの音声表記法」，人文学と情報処理 Vol.12，pp.12-20，(1996).
7) 樋口宣男：「韻律的特徴の記述法」，人文学と情報処理 Vol.12，pp.27-32，(1996).
8) 斉藤 隆，阪本正治：「テキスト音声合成を利用した音素・韻律統合ラベリングシステム」，電子情報通信学会技術研究報告，SP99-88，(1999).
9) 板橋秀一：「音声データベース／コーパスとは」，人文学と情報処理 Vol.12，pp.6-11，(1996).
10) 竹沢寿幸，末松博：「音声・テキストコーパスとその構築技術，標準動向」，人工知能学会誌 Vol.10，No.2，pp.168-180，(1995).
11) 山本幹雄：「音声対話データベース構築の現状」，日本音響学会誌 54 巻，第 11 号，pp.797-802，(1998).
12) 鹿野清宏，伊藤克亘，河原達也，武田一哉，山本幹雄：「音声認識システム」，オーム社，(2001).
13) 武田一哉：「簡易検索言語をもつ音声データベース管理システム」，ATR technical report，(1987).
14) 匂坂芳典：「多層音韻ラベルをもつ日本語音声データベース」，ATR technical report，(1987).
15) 武田一哉：「研究用日本語音声データベース 利用解説書」，ATR technical report，(1988).
16) 阿部匡伸：「研究用日本語音声データベース利用解説書　連続音声データ編」，ATR technical report，(1990).
17) 桑原尚夫：「研究用 ATR 日本語音声データベースの作成」，ATR technical report，(1989).
18) Netcraft Web Server Survey. http://www.netcraft.com/survey/

Table 4: 話者テーブル

ID	氏名	氏名カナ	性別	生年月日	年齢	出生地	育成地	父	母
8	MY	○○○○	男	19460204	55	神戸	神戸	姫路	神戸
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
23	TK	□□□□	男	19771008	24	東京	横浜	大阪	京都
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 8: 外来語辞書テーブル

ID	単語	関東	関西	原語	品詞	意味	言語	綴り
1065	キャリア	1	1	2	名詞	経歴	Eng	career
1066	キャリア	2	2	1	名詞	運ぶ人	Eng	carrier
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1478	サグラダファミリア	2	2	3	固有名詞	聖家族教会	Sp	Sagrada Familia
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1894	バルセロナ	2	3	4	名詞	地名（スペイン）	Sp	Barcelona
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 9: 形態素解析テーブル

ID	文 ID	単語 ID	表層語	基本形	活用
10351	1023	1065	キャリア	キャリア	一般
10352	1023	152	は	は	提題
10353	1023	1094	学校	学校	一般
10354	1023	160	で	で	道具
10355	1023	12965	決まる	決まる	連体
10356	1023	3562	もの	もの	一般
10357	1023	14478	で	だ	連用
10358	1023	152	は	は	提題
10359	1023	14477	ない	ない	終止
10360	1023	0	.	.	句点
⋮	⋮	⋮	⋮	⋮	⋮
21584	2051	1478	サグラダファミリア	サグラダファミリア	建物名
21585	2051	152	は	は	提題
21586	2051	1894	バルセロナ	バルセロナ	都市名
21587	2051	161	に	に	場所
21588	2051	13094	ある	ある	終止
21589	2051	0	.	.	句点
⋮	⋮	⋮	⋮	⋮	⋮