

# 国会審議映像検索システムと同形異音語の分析

——金大中と金日成——

増 山 幹 高

1. はじめに
2. 国会審議映像検索システム
3. 同形異音語の分析例：金大中と金日成
4. おわりに

## 1. はじめに

比較議会情報プロジェクトでは<sup>1)</sup>、文字情報に偏ってきた国会審議や立法過程に関する従来の分析から脱し、音声や画像、映像を活用することによって、これまで捨象されてきた立法情報の異次元を解明することに取り組んでいる。

国会の会議録が重要な立法情報であることに疑いはないが、国会で起きていることが全て会議録に残るわけではない。国会の記録とはなにか。憲法第 57 条は国会に会議の記録を保存し、公表・頒布することを求めている。記録技術の限られた時代において、紙媒体に記述された文字情報は唯一の記録であり、会議録は帝国議会の最初から 1 世紀余に連綿と残されている。国会は比較的に発言を忠実に記録しているが、会議録は発言を 100% 再現しているわけではない。発言中に「、」や「。」と口にするわけでもなく、「えー」「そのー」は会議録には残っていない。ケバという無駄な言葉を取り除き、

---

1) <https://gclip1.grips.ac.jp/~clip/>

話し言葉を書き言葉に直す整文が行われている。つまり、会議録は公的な記録として貴重な情報源であるが、何らかの加工を施された文字情報であり、様々な文字にならない情報が捨象されていることに留意する必要がある。

国会の会議録は、国会図書館の「国会会議録検索システム」から、第1回国会以降の全てにインターネットでアクセスすることが可能であり、キーワード検索もすることができる。また、帝国議会についても会議録にインターネットでアクセスすることが可能であり、戦後の会議録についてはキーワード検索も可能である。

<https://kokkai.ndl.go.jp/#/>

<https://teikokugikai.ndl.go.jp/#/>

このように会議録がテキスト・データとして容易に入手できることで、会議録は言語社会学的な分析にも活用されており、日本国中から集まった議員が長期間にわたって言語コーパスを提供するものとして、方言だけでなく、年齢差や性差などのバリエーション、また発話の長期的な変化が分析されている<sup>2)</sup>。こうした言語学的な研究が会議録を単なる文字情報ではなく、例えば、「朝刊」を「長官」と混同しないように「あさかん」と発言されても「朝刊」と整文されることに留意し、そうした整文がいかに行われるのかを分析する。これに対して、政治学的な研究は、実際の発言と会議録の相違に目を向けることはなく、分析ツールの普及もあって会議録を単にテキスト・データとして、政治的な対立構造の抽出といったことに主眼を置く<sup>3)</sup>。

しかし、首相の所信表明や施政方針といった演説についても、概ね原稿を読んでおり、実際の発言と会議録には大きな差はないが、会議録は話し言葉から書き言葉への整文を経たものであり、そうした整文の方法も長期的には変化してきている。テレビ・インターネットのおかげで、国会審議が中継され、動画の視聴が可能となることにより、会議録は実際の発言により忠実なものになってきている。話し方自体、世代的な変化は著しく、国会での発言

---

2) 松田 (2008)。

3) Curini et al. (2020).

もより日常言語に近づいてきている。したがって、会議録のテキスト分析で政治的な対立構造に長期的な変化が見出されたとしても、情報通信技術や言語社会環境の変化にも留意すべきである。

さらに、注意すべきことは、文字情報として発言に忠実であろうとするものの、会議録が会議で起きたことを全て記録するものではないということである。無論、全てを記録し、把握することは不可能であろう。近未来において通信容量が許すようであれば、国会審議の全てを保存し、疑似体験をすることも可能になるかもしれない。しかし、国会で行われていることの全てを追体験しないと国会審議が把握できないとなると、情報としての包括性は高まるが、効率的な情報収集とは言い難い。何らかの情報の集約やダイジェストは必要であるが、文字情報としての会議録にだけ依存するのではなく、音声や映像を活用して国会審議を分析し、多元的な議会空間を解明することが課題である。

国会審議の音声や映像を活用しようとする「国会審議映像検索システム」では、発言の音声認識によるテキスト・データと国会図書館の「国会会議録検索システム」で公開される会議録を同期させることで審議映像の時間情報と文字情報を同刻させており、その利点を生かせば、同形異音語がどのように発音されたのかを効率的に確認することができる。本稿では、「国会審議映像検索システム」を概観したうえで、国会議員が金大中と金日成を現地読みするか、日本語読みするかかの分析を紹介する。

## 2. 国会審議映像検索システム

国会では、1990年代の終わりから衆参両院事務局が審議映像をインターネットで配信し、ライブストリーミングでの視聴やライブラリに保存された審議映像がいつでもどこでも視聴できるようになっている<sup>4)</sup>。

<https://www.shugiintv.go.jp/index.php>

<https://www.webtv.sangiin.go.jp/webtv/index.php>

比較議会情報プロジェクトは、こうした時空間的制約のさらなる解消を目

指す試みとして、国会図書館の提供する会議録と衆参両院の事務局が配信する審議映像をリンクさせ、発言のキーワード検索から審議映像をピンポイントで再生することを目指してきている。具体的には、音声認識によって映像情報と文字情報を同刻することにより、審議映像を発言単位でキーワード検索し、審議映像を配信元である衆参両院事務局のサーバー上で再生しつつ、字幕をタイムラグなしに付すことを可能にしている。こうした試みにより、審議映像の利用方法が革新的に改善され、国会審議に関する文字情報と映像情報を体系的に検証することが期待され、審議映像が視覚や聴覚に支障のある場合でも活用されるようになるとともに、音声認識による映像検索を地方議会や審議会などの会議全般、ニュースの映像配信にも応用することが可能となる。

国会審議について議員の発言内容に対応する審議映像を検索し、該当する審議映像の部分的視聴を可能にする「国会審議映像検索システム」は、2012年11月26日から一般公開され、2014年4月18日からは自動翻訳機能を活用した英文入力サイトも運用されている<sup>5)</sup>。

<https://gclip1.grips.ac.jp/video/>

国会審議映像検索システムは、国会図書館の提供する国会会議録検索シス

- 
- 4) 衆議院は2010年以降の映像を公開しているが、参議院の運用では公開を1年に限っており、参議院事務局サーバーでの審議映像の公開終了後には映像は再生できない。
  - 5) 国会審議映像検索システムについては、これまでも国内外の学会や専門誌で紹介している。Masuyama (2012, 2016, 2018a, 2018b)、Masuyama and Kawahara (2019)、Masuyama and Takeda (2014)、増山 (2016, 2017a, 2017b, 2018, 2020, 2021, 2022)、増山・竹田 (2012, 2015)、鈴木他 (2014)。また、列国議会同盟 (Inter-Parliamentary Union) の Innovation Tracker Issue 4 (12 Feb. 2020) においても人口知能を活用した審議動画検索として紹介され、その他にも科研費ニュースや Intersteno (情報処理・情報伝達の世界連盟) においても紹介されている。

<https://www.ipu.org/innovation-tracker/story/artificial-intelligence-innovation-in-parliaments>

[https://www.jsps.go.jp/j-grantsinaid/22\\_letter/data/news\\_2017\\_vol4/p05.pdf](https://www.jsps.go.jp/j-grantsinaid/22_letter/data/news_2017_vol4/p05.pdf)

<https://www.intersteno.org/2016/04/e-news-73-april-2016-was-published/>

テムで発言内容を検索できるように、会議録の文字情報から、発言に対応する映像にピンポイントで到達することを目指している。これにより、例えば、ニュースや新聞記事で「〇〇議員が△△と国会審議で発言」という情報に接した際、「〇〇 △△」でキーワード検索し、その瞬間の審議映像をピンポイントで再生できるようにし、会議録からは分からない発言者の表情や臨場感、会議の流れなどが把握できるようになる。また、審議映像に字幕を付すことにより、聴覚や視覚に支障がある場合でも審議映像を利用することが可能になり、発言の瞬間を URL として表現することにより、SNS を通じて審議映像をインターネットで共有することも容易になる<sup>6)</sup>。

具体的には、審議映像の発言内容を検索するため、検索インデックスとして、動画中の発言のテキスト・データと、そのテキスト・データと動画中の映像シーンの同期情報が必要となる (図 1)。国会審議の場合、従来は速記により、衆議院では現在は音声認識により、会議録が作成されており、国会審議映像検索システムでは、会議録のテキスト・データと動画の音声データの同期データを作成することにより、審議映像中の発言内容を検索することが可能になっている。

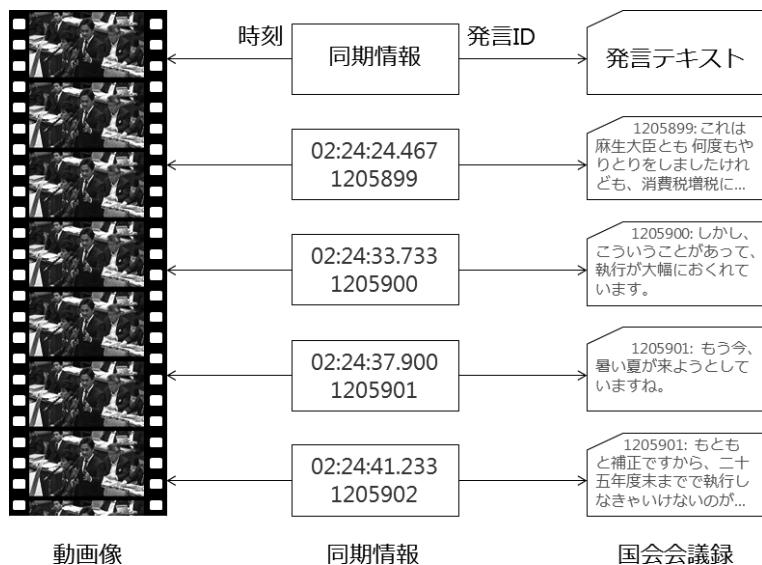
審議映像と会議録テキストの同期データの作成には、京都大学の開発した「字幕自動生成システム」を利用している<sup>7)</sup>。字幕自動生成システムでは、音声ストリーム・データとテキスト・データを付与すると、それらの同期情報が生成され、数時間にわたる審議映像とテキスト・データの同期情報の作成が可能となり、審議映像の検索インデックスを作成することができる。

図 2 に示すように、国会審議映像検索システムは、インターフェースとデータベースの二つから構成される。データベースには、会議録のテキスト・データと会議録中の発言と審議映像の同期データが含まれる。会議録のテキスト・データから全文検索用のインデックスが生成され、それらが利用

6) 参議院の審議映像は、発言内容で検索可能であり、映像の部分再生も可能であるが、字幕が付されず、映像の再生と同時に対応する会議録の確認はできない。

7) 河原 (2012)、秋田・河原 (2013)。

図1 審議動画と国会会議録の同期情報



者の入力したキーワードに対応する発言を検索する際に利用される。検索インデックスには、会議名や発言者名もメタデータとして付与されており、これらの組み合わせによる検索を可能にする。会議録のテキスト・データは、審議映像の再生中の字幕表示や会議録の表示にも利用される。

国会審議における発言と映像の同期データは、検索結果に対応するシーンの視聴ページに遷移する際に利用される。このデータを生成・保持することにより、発言に対応する映像の冒頭からピンポイントで視聴することが可能となる。また、この同期データは、審議映像の再生中に表示する字幕の切替えや、会議録の該当発言を表示する切り替えにも利用されている。なお国会審議の動画自体は、データベースでは保持せず、配信元である衆参両院事務局のウェブサイトで再生している。つまり、国会審議映像検索システムは、そうした審議映像データのプロキシとして機能している。

また、図3は審議映像と会議録の同期データ作成の流れを示している。

図 2 国会審議映像検索システムの構成

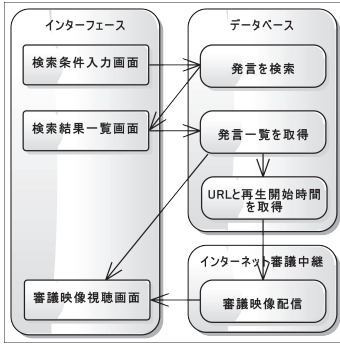
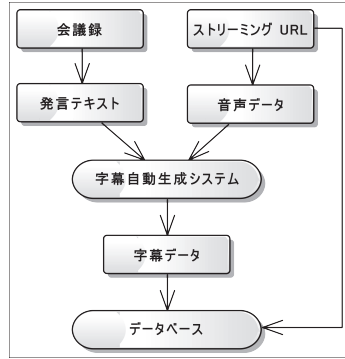


図 3 発言と映像の同期データ作成の流れ



同期データを作成するには、会議録のテキスト・データと審議映像の音声データが必要となる。まず国会図書館の国会会議録検索システムから、会議録のテキスト・データを取得する。国会審議映像検索システムでは、会議録取得の一連の作業をシステム上で行う機能を開発し、取得したデータは字幕生成の処理ができる書式に変更される。

審議映像については、衆参両院事務局のウェブサイトから会議録に対応する審議映像のストリーミング・データを取得し、音声データを抽出して音声ファイルとして保存する。国会審議映像検索システムでは、動画の URL を特定すると、動画のストリーミング・データを取得し、音声ファイルとして保存する処理が自動的に行われる。このように会議録と音声ファイルを抽出し、字幕自動生成システムを用いて同期データの作成を行い、会議録のテキスト・データおよびストリーミング・データの URL と合わせて、発言と映像の同期データが国会審議映像検索システムに登録される。

国会審議映像検索システムのトップページはリニューアルを重ね、現在では上部の検索キーワード入力欄に検索条件を入力するだけでなく、検索オプションから発言、会議、議員、議案、フリップのいずれかを選択し、検索条件を入力することで国会審議映像を検索することもできる (図 4)。例えば、検索オプションを選択しない場合、「気候変動 安全保障」と入力すると、

会議録が単純に検索されるが、検索オプションで「議員」を選んだ場合、議員情報と併せて検索結果が議員毎に表示される。

図5は検索条件に「気候変動 安全保障」と入力した単純な検索結果を示している。入力したキーワードを含む発言の一覧が右側の検索結果エリアに最近のものから表示され、個々の検索結果には、タイトル部分に、会議名、開会日が表示され、発言者名に続いて、検索キーワードが含まれる発言部分の会議録が表示される。タイトル部分は審議映像視聴ページへのリンクとなっており、ここをクリックすることで映像を視聴することができる。また、左側のフィルタエリアには、最初に入力した検索条件に加えて、開会日<sup>8)</sup>、衆参のいずれか、年齢層、性別、選挙区、会派、発言の出現回数の多い発言者5名および10会議、回次による絞り込みメニューが表示される。

検索結果のいずれかをクリックすると、国会審議映像の視聴ページに遷移する(図6)。この視聴ページに遷移すると、選択した発言の部分再生が始まる。左側上部の動画再生エリアには審議映像が再生され、その下の字幕エリアに発言者の氏名と発言内容の字幕、再生操作のコントロールが表示される。右側の発言リストには、再生している会議の会議録全体が発言毎に表示され、部分再生しているシーンに該当する発言の箇所が反転表示される。検索した審議映像の部分再生が終了すると、自動的に続く発言の審議映像を再生し、映像の再生の進行に合わせて、会議録の反転表示している箇所も変化する(1分間再生し、1分を超えた場合は3発言までで停止する初期設定になっているが、右上部にある「設定」で変更できる)。また、右側の発言リストのいずれかをダブルクリックすると、自動的に部分再生する審議映像を切り替えることができ、検索結果の発言は黄色くハイライトされて残る。さらに、検索結果以外の発言における検索語も黄色くハイライトし、関連発言を目視できるとともに、発言リスト上部に検索入力欄があり、当該会議録を別のキーワードでも検索できる。

動画再生エリアの下部に部分再生している発言のシーンに対応する審議映

---

8) 検索期間の初期設定は過去1年間としている。



図 4 国会審議映像検索システムのトップページ

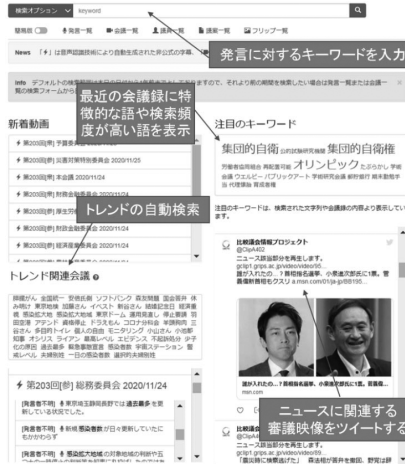


図 5 検索結果の一例 (2020 年 11 月 25 日)

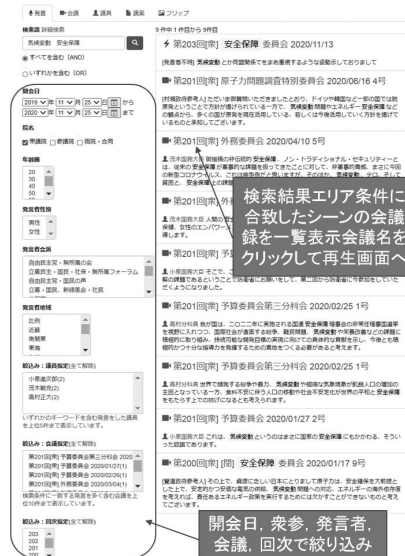


図6 審議映像の視聴ページ



像の URL を表示しており、その URL をツイートできるようにしてある。再生している動画部分をツイートしたい場合、再生中にツイートするためのボタンをクリックすると、発言と URL をツイートすることができ、注目する発言のシーンをインターネットで共有することができる。URL を時分秒で表示することにより、審議映像の開始だけでなく、終了も設定できるようになっている。比較議会情報プロジェクトでは、国会関連のニュースと併せて、会議録を読むだけでは把握し難く、審議映像を確認することによって、その実態がより詳しく理解されると思われる審議映像 URL をツイートし<sup>9)</sup>、国会審議映像検索システムのトップページでも確認できるようにしている(図4右下)。再生エリア上部の会議名にリンクを追加し、国会図書館の国会会議録検索システムにおける当該会議の会議録が確認できるとともに、視聴ページ下部には、発言者のプロフィールを表示し、会議で審議されている議案と参加している議員名がリストが表示される(図6には表示されていない)。

国会審議映像検索システムは、国会議員の発言内容をピンポイントで再現しようとするものであり、様々な活用法を考えることができる。図6のように、小泉環境大臣(当時)の発言を部分的に再生することによって、会議録の文字情報としてだけでなく、実際の会議における肉声として、顔の表情

9) <https://twitter.com/ClipA402>

や声の調子、会場の雰囲気や会議の流れを併せて理解することができる。

国会審議には会議録に含まれない様々な情報がある。例えば、委員会審議で頻繁に用いられる参考資料は、質疑に立っている議員が特に求めない限り、会議録には含められず、国会事務局や国会図書館で保管されているだけで、存在さえ国会関係者でなければ知る由もない。しかし、そうした参考資料は、国会審議の要点を示す視覚的情報であることが多く、カメラの寄った審議映像が残っていれば、内容を確認することもできる。例えば、衆議院の会議録には、(パネルを示す) というト書きがあり、キーワード検索によって委員会審議で用いられたパネルを審議映像から判別することができる<sup>10)</sup>。

国会審議映像検索システムでは、キーワード検索とパターン認識を組み合わせることによって、こうしたパネルを自動的に抽出する機能を開発しており、トップページの検索オプションにフリップ一覧を追加している(図4)。これをクリックすると、抽出された画像のサムネイルと会議の一覧が検索結果として示され、いずれかの会議リンクをクリックすると、当該会議において抽出されたパネルが一覧表示され、サムネイル画像とともにパネルの文字認識結果が示される(図7・8)。文字認識結果はデータベース化され、それらはキーワード検索にも利用される。

また、視聴ページにおける会議録の分析機能も開発してきており、視聴ページ右の上段のタグをクリックすると、当該会議で自動抽出されたパネルの一覧を確認でき、パネルの画像をクリックすると、審議映像がパネルを写している瞬間に切り替わる(図9)。また、一つ下のタグをクリックすると、10分間隔で抽出したキーワード一覧を示し、審議の流れもキーワードで確認できる(図10)。

さらに、一つ下のタグをクリックすると、会議録を1割に集約した自動要約が示され、一つ下のタグをクリックすると、ワードクラウドによる会議録

10) <https://gclip1.grips.ac.jp/~clip/panel/>

参議院の審議映像はパネルにカメラを寄せることがないため、パネルは特定できるものの、抽出したパネルの判別が難しい。

図7 フリップ検索結果



図8 フリップ文字認識



図9 パネルと審議映像



図 10 キーワードによる審議経過



図 11 自動要約



のテキスト分析が表示される (図 11・12)。最下段のタグをクリックすると、議長や委員長の定型的な発言を自動的に検索した結果が表示され、それらをクリックすると、審議映像が切り替わり、議事進行に応じて審議の状態を確認することができる (図 13)。

国会審議映像検索システムでは、審議映像に付す字幕を国会図書館の国会会議録検索システムから取得した会議録のテキスト・データに拠っている。会議録に基づいて審議映像が検索できるようになり、字幕を付した形の審議映像の再生が可能となるには、会議録が公開されるまでの時間が必要となる (会議名の前にカメラマークがあり、国会図書館の国会会議録検索システムで該当する会議録へのリンクが付されている)。現在の運用では、審議映像が衆参

図 12 ワードクラウド



図 13 議事進行



両院事務局のサーバー上のライブラリにおいて視聴可能となると、まず会議録情報のないまま審議映像の視聴が自動的に可能となり、トップページの新着動画に追加される。一両日中に音声認識版の会議録が自動的に作成され、それによって審議映像が検索可能となり、顔識別による発言者の特定も行い(会議名にカメラマークのあるもの)、上記のようなツイートやニュース配信で引用できるよう速報性の向上に努めている<sup>11)</sup>。

11) 音声認識・顔認識版による審議映像検索までは、視聴ページの動画再生エリア上部に示す会議名に音声認識・顔認識版によることを明記し、字幕付与機能は停止させている。これら一連の処理を全て自動的に実施する機能開発は2017年度に完成し、2018年から実装している。また、顔識別は2020年から実装を開始しているが、新型コロナウイルスの感染拡大防止のため議員がマスクを着用するようになっており、本来の顔識別の精度が実現できない状況にある。

また、トップページの「注目のキーワード」欄には、新たに集中的に発言され、検索される単語の上位 15 位が表示されている。こうしたキーワードは、検索可能な審議映像に対応する会議録から、会議録に頻出する単語として、最新のものであるほどウェイトがかけられ、また特定の会議のなかで頻度の高いものによりウェイトがかけられるように抽出されるとともに、検索頻度の高さも加味されている<sup>12)</sup>。また、利用者側の注目度を反映する仕組みとして、実際に検索入力されるものほど大きく表示している。これらのキーワードのいずれかをクリックすることで、そのキーワードを入力して検索することとなり、続けて検索条件を追加指定することも可能である。さらに、トップページのトレンド会議関連は SNS でトレンドになっているキーワードで会議録を自動的に検索し、社会的に関心のある話題と国会審議の関連性を可視化しようとしている (図 4 左下)<sup>13)</sup>。

### 3. 同形異音語の分析例：金大中と金日成

この節では、文字、音声、映像が複合的、組織的に集積されるという観点から議会を捉え直し、書き言葉として加工される会議録では捨象され、文字にならない立法情報の分析を試みる。「国会審議映像検索システム」では、発言の音声認識によるテキスト・データと国会図書館の「国会会議録検索システム」で公開される会議録を同期させている。審議映像の時間情報と文字情報を同刻させる「国会審議映像検索システム」を活用すれば、複数の読みがある同形異音語がどのように発音されたのかを効率的に確認することが可

---

12) 具体的には、 $s(word) = \frac{f(word)}{\max(f(word))} + (1-w) \frac{q(word)}{\max(q(word))}$  による。 $f(word)$

は単語の会議録における頻度であり、時間的に最新であるほど、また特定の会議中における頻度の高さでウェイトをかけている。 $q(word)$  は単語の検索頻度である。 $w$  はこれら両頻度のウェイトであり、現時点では 0.5 を想定している。

13) 具体的には、毎時 Twitter の日本のトレンドを自動的に取得・保存し、直近 7 日間のキーワードで直近 100 日間の会議録を検索している。

能である。

言語学においては、例えば、外国の地名を英語でどう発音するのかということについて、アメリカの議員が Iraq の a を /æ/ か /ɑ:/ と発音するかで政治的姿勢を識別することが試みられている<sup>14)</sup>。日本では、例えば、韓国の人名を現地読みする場合と日本語読みする場合がある。この節では、そうした現地読みか日本語読みかと、発言者の党派性に関連があるのかを検証してみる。

具体的には、同形異音語の分析例として、国会議員が金大中と金日成を現地読みするか、日本語読みするかを分析する。「国会審議映像検索システム」では、衆議院に関しては2010年1月18日に始まる第174回国会以降、参議院に関しては2012年12月26日に始まる第182回国会以降の審議映像が検索可能である。日本では、1990年代以降、韓国の人名について現地読みすることが一般的になってきている。例えば、韓国の大統領名をキーワードとして審議映像を検索した場合、最近の大統領名は現地読みであり、古くは李承晩だと日本語読みで李承晩ラインとして言及されることが多かった。一方、北朝鮮については、金正恩を日本語読みすることはなく、金正日も日本語読みする場合は少なかった。こうした予備的調査から、ある程度の審議映像が検索され、両方の読みが混在する金大中と金日成を分析対象とした。

前節で概説したように、国会審議映像検索システムによるキーワード検索から、審議映像視聴ページの URL をクリックし、現地読みか日本語読みかを確認したところ、金大中 42 件のうち 3 件、金日成 48 件のうち 18 件で日本語読みをしていることがわかった<sup>15)</sup>。金大中か金日成かの個人の問題な

---

14) Hall-Lew et al. (2010).

15) 審議映像の検索結果は増山 (2022) 付録にまとめる通りである。参考人、公述人の発言は除いた。脚注 4 でも述べているように、参議院の 1 年以上前の審議映像は配信されなくなるため、音声認識の際に抽出した音声などによって確認している。会議録では金大中と記載され、現地読みと日本語読みがされている場合があり (2017 年 4 月 19 日第 193 回国会衆議院外務委員会笠井亮)、ここではそれぞれ 1 件としている。  
<https://gclip1.grips.ac.jp/video/video/6205?t=2h38m33s>



のか、韓国か北朝鮮かという国の問題なのかは判別できないが、金日成のほうが日本語読みされるということは  $\chi$  二乗検定からも確認される (11.540 [0.001])。また、衆議院か参議院かで比べると、衆議院 63 件中 18 件が日本語読みであり、参議院 27 件中 3 件が日本語読みであり (3.221 [0.073])、答弁か質問かで比べると、答弁 20 件中 2 件が日本語読み、質問 70 件中 19 件が日本語読みであった (2.556 [0.110])。

発言者の所属政党が自民党議員か否かで比べると、自民党議員 42 件中 9 件が日本語読みであり、非自民党議員 48 件中 12 件で日本語読みであり (0.160 [0.689])、民主党系か共産党の左派議員か否かで比べると、左派議員 35 件中 5 件が日本語読みであり、非左派議員 55 件中 16 件が日本語読みであった (2.621 [0.105])。こうした集計的分析では発言者の党派性と現地読みか日本語読みの関係は明瞭ではないが、拉致問題に関わる発言が多く検索されており、拉致議連の議員では松原仁 (民主党・答弁)、渡辺周 (民主党・質問)、古屋圭司 (自民党・質問) の発言が含まれるが、こうした党派横断的な拉致議連の議員か否かで比べると、拉致議連の議員 11 件中 6 件が日本語読みであり、拉致議連の議員でない 79 件中 15 件が日本語読みであることから (6.825 [0.009])、現地読みか日本語読みかは党派性と単純な関係になく、拉致問題などの外交問題に関する議員個々の姿勢と複合的に関連していることが推測される。

#### 4. おわりに

本稿では、文字情報に偏ってきた国会審議や立法過程に対する従来からの分析から脱し、音声や画像、映像を活用する試みとして、国会審議について議員の発言内容に対応する審議映像を検索し、該当する審議映像の部分的視聴を可能にする「国会審議映像検索システム」を概説した。「国会審議映像検索システム」は、審議映像の時間情報と会議録の文字情報を同刻させており、複数の読みがある同形異音語が実際にどのように発音されたのかを効率的に確認することができる。本稿では、同形異音語の分析例として、韓国の人名、

具体的には、金大中と金日成を国会議員が現地読みするか日本語読みするかを確認し、その議員の党派性や拉致問題への取り組みとの関連を分析した。

国会審議映像検索システムは、国会会議録の文字情報から、発言に対応する映像にピンポイントで到達することを目指すものであり、その瞬間の審議映像の部分再生から、会議録では分からない発言者の表情や臨場感、会議の流れなどを把握できるようにする。また、審議映像に字幕を付すことにより、視聴覚に支障がある場合でも国会審議映像を利用することが可能になり、発言の瞬間を URL として表現することにより、SNS を通じて審議映像をインターネットで共有することも容易になる。さらに、国会審議には会議録に含まれない様々な情報があり、例えば、委員会審議で用いられる参考資料についても、国会審議映像検索システムを活用すれば、部分再生した画像から瞬時に確認することができる。

こうした国会審議映像に関する映像情報と文字情報の音声認識による同刻といった技術は、地方議会や審議会などの会議全般、またニュース報道などへの対象の拡大という意味においても<sup>16)</sup>、あるいは会議録といった「正解」のない映像の検索・視聴という同刻方法や検索・視聴する利用方法という意味においても多様な応用可能性を検討することができる。また、発言の瞬間を URL として表現し、映像の部分再生を容易にする国会審議映像検索システムの利点を生かし、SNS を通じて映像を部分的に共有することも容易であり（増山 2017b）、文字情報と映像情報とで国会審議に対する理解が異なるのかという実験的分析も容易になる（増山 2017a、2018、Masuyama 2018b）。

国会の会議録が重要な立法情報であることに疑いはないが、国会で起きていることが全て会議録に残るわけではない。会議録は公的な記録として貴重な情報源であるが、何らかの加工を施された文字情報であり、様々な文字にならない情報が捨象されていることに留意する必要がある。例えば、日常的

---

16) 増山 (2016)。また「地方議会審議映像検索システム」も試行的に運用してきている。

<https://gclip1.grips.ac.jp/local-assembly/>

な会話ではアイコンタクトやうなずきによって共感や同意が確認される。実際のコミュニケーションでは声だけでなく、視線や仕草、声音によって伝わる内容が大きく異なるのであれば、文字か映像かで国会審議の理解が左右されても驚くことではない。しかし、議会において文字、音声、映像が集積されながら、それらが複合的、組織的に活用されているとは言い難く、議会研究における会議録偏重は根強い。

比較議会情報プロジェクトの最終的な目標の一つは審議過程のハイライト化にある。国会審議映像検索システムのパネル抽出やキーワード抽出、自動要約、テキスト分析、議事進行分析を改良するとともに、感情分析、音圧分析、話速分析、フィラー分析などを統合する機能を開発し<sup>17)</sup>、画像の類似度を機会学習によって識別し、審議の盛り上がりや白熱しているかといったことから、議論の展開や成り行きを左右する場面を絞り込むことにより、審議のハイライトを自動的に選別・可視化し、会議録を読むのでもなく、審議映像を頭から視聴せずとも、議会で繰り返されていることを効率的に理解し、時空間的に再現することを目指している。これは、野球に喩えれば、従来の会議録偏重によるテキスト分析がスコアブックで試合を理解しようとするのであれば、得点シーンや試合の結果を左右するプレーをニュースのハイライトを見るように試合の流れとして理解する試みと言える。比較議会情報プロジェクトは議会を文字、音声、映像が複合的、組織的に集積されるものと捉え直し、国会審議をニュースのハイライトのようにダイジェスト化・可視化し、会議録から削ぎ落とされる国会の時空間の解明を目指している。本稿においては、国会審議映像検索システムを同形異音語の分析に活用する例として、金大中と金日成が現地読みされるか日本語読みされるかの分析を紹介し、同形異音語の読みという会議録から捨象されてきた立法の異次元を解明する可能性を提示した<sup>18)</sup>。

17) 増山 (2020)。

18) 審議映像検索はすでに言語研究において活用されており、例えば、松田 (2016) に拠れば、「場合」を「ばあい」「ばわい」「ばやい」と読むかを判別する研究に貢献している。

## 付記

本稿は「政策情報のユニバーサル化・国際化に関する実証と実践」(2015年度～2019年度・課題番号15H05727)、「コミュニケーション空間分析としての国会研究」(2020年度～2022年度・課題番号20H00062)による部分的成果である。

## 参考文献

- 秋田佑哉・河原達也. 2013. 「音声認識を用いたオンライン自動字幕作成・編集システム」『日本音響学会秋季研究発表会講演論文集』2-8-4.
- 河原達也. 2012. 「議会の会議録作成のための音声認識—衆議院のシステムの概要」情報処理学会研究報告 SLP-93-5.
- Curini, Luigi, Airo Hino, and Atsushi Osaka. 2020. "The Intensity of Government-  
Opposition Divide as Measured through Legislative Speeches and What We  
Can Learn from It: Analyses of Japanese Parliamentary Debates, 1953-2013."  
*Government and Opposition* 55, no. 2: 184-201.
- Hall-Lew, Lauren. 2010. "Indexing Political Persuasion: Variation in the Iraq  
Vowels" *American Speech* 85: 91-102.
- Masuyama, Mikitaka. 2012. "Text-based Search on Diet Deliberation Video Clips"  
*Presented at the 2012 Annual Meeting of the Association for Asian Studies*,  
Toronto, Canada.
- Masuyama, Mikitaka. 2016a. "Reducing the Space and Time between Citizens  
and Parliament: Video Retrieval System for Diet Deliberations" *Prepared for  
delivery at the 2016 General Conference of the European Consortium for  
Political Research*, Charles University, Prague, Czech Republic.
- Masuyama, Mikitaka. 2018a. "Heat of Discussion: A New Approach to  
Understanding Parliamentary Discussion" *Prepared for delivery at the 25th  
IPSA World Congress of Political Science*, Brisbane, Australia, July 21-26,  
2018.
- Masuyama, Mikitaka. 2018b. "Reading and Viewing Parliamentary Debates: The  
Case of the Japanese Diet" *Prepared for the 2018 Annual Meeting of the  
American Political Science Association*, Boston, August 30-September 2, 2018.
- Masuyama, Mikitaka, and Kaori Takeda. 2014. "Instant Parliamentary  
Deliberations Are in Our Reach" *Presented at the 2014 Annual Meeting of*

- the American Political Science Association*, Washington, D.C., USA.
- Masuyama, Mikitaka, and Kawahara Tatsuya. 2019. "Automatic Speech Recognition and Video Retrieval System for the Japanese Diet" *Prepared for delivery at the ECPR Standing Group on Parliaments Conference*, Leiden University, Netherlands, June 27-29, 2019.
- 増山幹高. 2016. 「国会審議映像検索システムとそのニュース動画への応用・検証」 *GRIPS Discussion Paper*, 16-23.
- 増山幹高. 2017a. 「国会審議映像検索システムとその実験的応用の可能性」 *GRIPS Discussion Paper*, 17-11.
- 増山幹高. 2017b. 「国会審議映像検索システムとその SNS 的活用の可能性：ニュースと審議映像の Tweet」 2017. *GRIPS Discussion Paper*, 17-12.
- 増山幹高. 2018. 「国会審議映像検索システムとその実験的応用の試み」 *GRIPS Discussion Paper*, 18-07.
- 増山幹高. 2020. 「国会審議映像検索システムの機能開発状況」 *GRIPS Discussion Paper*, 20-11.
- 増山幹高. 2021. 「国会審議の非会議録研究の試み：安倍首相の国会発言における変化」 *GRIPS Discussion Paper*, 21-05.
- 増山幹高. 2022. 「国会審議映像検索システムと同形異音語の分析：金大中と金日成」 *GRIPS Discussion Paper*, 22-05.
- 増山幹高・竹田香織. 2012. 「いかに見たい国会審議映像に到達するか？」 2012 年度日本選挙学会（於筑波大学）.
- 増山幹高・竹田香織. 2015. 「いかに見たい国会審議映像に到達するか？—国会審議映像検索システムの概要—」 『レヴァイアサン』 56 号、54-79 頁.
- 松田謙次郎. 2008. 『国会会議録を使った日本語研究』 ひつじ書房.
- 松田謙次郎. 2016. 「大正～昭和戦前期の SP 盤演説レコードにおける「場合」の読みについて」 『国立国語研究所論集』 11 号、63-81 頁.
- 鈴木泰山・内山雄司・青木保一・相良毅・秋田佑哉・河原達也・竹田香織・増山幹高. 2014. 「音声認識技術の活用による国会審議映像検索システムの実現」 情報処理学会（於 NHK 放送研究所）.