

英国地方議会会議録コーパスの構築に向けた分析

木村 泰知^{†1} 渋木 英潔^{†2} 森 辰則^{†2} 神門 典子^{†3†4}

^{†1}小樽商科大学 ^{†2}横浜国立大学 ^{†3}国立情報学研究所 ^{†3}総合研究大学院大学
E-mail: kimura@res.otaru-uc.ac.jp, {shib,mori}@forest.eis.ynu.ac.jp, kando@nii.ac.jp

1 はじめに

議会会議録は、首長や議員の議論が書き起こされた話し言葉のデータであり、長い年月の議論が記録された通時的データである。議会会議録には、「議員本人の発言が記録された一次情報である」、「マニフェストなどと比較して、内容が現在までに実際に行ってきた/行っていることが中心である」といった特徴がある。また、地方議会を対象としたデータは地域性の比較にも有用であることから、政治学、経済学、言語学、情報工学といった分野における研究対象として利用されている。しかしながら、地方議会会議録は一般に地方自治体ごとに公開されており、統一的な枠組みで研究者が利用することが困難な状況であった。それゆえ、我々は文献 [1, 2] などにおいて国内の地方自治体を対象とした地方議会会議録コーパスの構築を行った。しかしながら、日本国内のみが対象であり、海外のデータおよび研究者のことは考慮されていなかった。

海外の政治情報を扱った研究には、文献 [3, 4, 5, 6] などが存在する。これらの研究で用いられているデータは、ブログ、Twitter、新聞記事、大統領選のディベートであり、議会会議録と比較すると以下のような違いがある。ブログや Twitter といったデータは、一部に議員本人の発言も含まれているかもしれないが、基本的には一般の人々の発言である。新聞記事は、議員の発言が引用となり、一次情報ではない。大統領選のディベートは、議員など候補者本人の発言であるが、選挙というイベントにおける発言である。以上から、海外の議会会議録を対象としたコーパスを構築するという本研究には価値がある。

海外の地方議会会議録コーパスを構築する第一歩として、本稿ではイギリスの地方議会を対象とし、イギリス地方議会会議録コーパスの構築を行う。イギリスを対象としたのは、英語であることや国の知名度に加え、日本と同じ島国で立憲君主制を敷き同程度の国土面積であることから、既に構築した日本国内の地方議会会議録コーパスとの比較対象として適切と考えたか

表 1: 発言テーブル

	フィールド	説明
1	識別子	発言 ID
2	都道府県名	例:「北海道」「青森県」
3	回	会議の回数 (第〇回)
4	号	会議の号 (〇日目)
5	年	開催年 (和暦) 23, 24, 25...
6	月	開催月 1~12
7	日	開催日 1~31
8	会議種別	定例会・臨時会
9	表題	例:「第 21 回岩手県議会臨時会」
10	役割	例:「議長」「質問者」「答弁者」
11	発言者 ID	議員テーブルを参照する外部キー
12	発言者名	例:「工藤孝男」「佐々木博」
13	役職	例:「知事」「教育長」
14	発言文	例:「次に、冬の節電対策について...」
15	発言以外	例:「(拍手)」 「[...]君登壇」
16	URL	ダウンロード元の URL
17	保存場所	HTML ファイルの保存場所

らである。このような比較を容易にするためには統一的なデータ構造が望ましく、本稿ではその点も合わせて検討する。

2 地方議会会議録コーパス

我々がこれまで取り組んできた、地方議会会議録コーパスについて簡単に説明する。地方議会会議録コーパスの収集対象は、全国 47 都道府県議会の本会議であり、収集期間は平成 23 年 4 月の統一地方選挙から平成 27 年 4 月の統一地方選挙の前 (平成 27 年 3 月) までの 4 年間である。その結果、地方議会会議録コーパスに収録されたデータのサイズは、約 400 万文、1.5GB となっている。収集された会議録データは、表 1 に示す 17 項目からなる「発言テーブル」により、統一されたデータ構造で管理されている。一般に日本の会議録では「第 2 回定例会」のように「年」「月」「日」による日付よりも「回」や「号」で識別するため、冗長ではあるが両方のフィールドで管理している。また、議員のプロフィール情報は名前やふりがななどの 8 項目からなる「議員テーブル」で管理されている。

日本の地方議会会議録の例として、図 1 に東京都議会会議録 (本会議) の一部を示す。日本の会議録の特

001 - 010 行	議長	議事進行
011 - 020 行	議長	議事進行
021 - 030 行	議長	議事進行
031 - 040 行	議長	議事進行
041 - 050 行	議長	議事進行
051 - 060 行	議長	議事進行
061 - 070 行	議長	議事進行
071 - 080 行	議長	議事進行
081 - 090 行	所信表明	
091 - 100 行	石原知事	所信表明
101 - 110 行	石原知事	所信表明
111 - 120 行	石原知事	所信表明
121 - 130 行	石原知事	所信表明
131 - 140 行	石原知事	所信表明
141 - 150 行	石原知事	所信表明
151 - 160 行	石原知事	所信表明
161 - 170 行	石原知事	所信表明
171 - 180 行	石原知事	所信表明
181 - 190 行	石原知事	所信表明
191 - 200 行	石原知事	所信表明
201 - 210 行	石原知事	所信表明
211 - 220 行	石原知事	所信表明
221 - 230 行	石原知事	所信表明
231 - 240 行	議長	議事進行
241 - 250 行	議長	議事進行
251 - 260 行	山下議員	質問概要
261 - 270 行	山下議員	質問1
271 - 280 行	山下議員	質問1
281 - 290 行	山下議員	質問2
291 - 300 行	山下議員	質問3
301 - 310 行	山下議員	質問4
311 - 320 行	山下議員	質問5・6
321 - 330 行	山下議員	質問7
331 - 340 行	山下議員	質問8
341 - 350 行	山下議員	質問9
351 - 360 行	山下議員	質問10

東京都議会会議録 平成23年度第2回定例会

256行目 257行目 (中略) 265行目	私は、都議会民主党を代表して、都政の主要課題について知事並びに関係局長に伺います。 東日本大震災より3カ月余りが過ぎました。 … (中略) … しかし、私たちは、この結果を尊重するとともに、もう1方の公選によって私たちに付託された都民の期待を踏まえ、今後も都民の生活を第1とする都政の実現に取り組むことを表明させていただきます。
質問1 山下議員 266行目 267行目	まず、東日本大震災における被災地支援と東京の防災対策について伺います。 三月十一日、マグニチュード九・〇、最高震度七の強い揺れが東日本一帯を襲うとともに、大津波、海砂を巻き込んだ黒く重い海水の塊が太平洋沿岸の防波堤を軒並み破壊し、海水や瓦れきが市街地に流れ込み、最大な被害を引き起こしました。 福島第一原子力発電所にも大津波が押し寄せ、冷却電源を失った原子炉建屋は爆発、格納容器が損傷して、放射性物質が広範囲に拡散しました。 原発周辺の住民の皆さんは、自宅があるのに怖れない深刻な状況が続いています。 私たちは、この未曾有の複合災害に対していち早く被災地支援と都内の震災対策を充実させること、そして補正予算の編成を知事に申し入れました。 また、各議員は、党の被災地支援活動やNPOと連携した取り組みを行うなど、被災地支援に取り組んでまいりました。 そこで伺います。
268行目 269行目 270行目 271行目 272行目 273行目	都は、児童生徒への心のケアや、災害時要援護者の救護など、医療人材の継続的な派遣や、地元雇用を推進する自治体事業、キャッシュ・フォー・ワークといった取り組みへの支援をするなど、被災者の皆さんが希望を見出し、一歩踏み出すことのできるよう、生活再建をともにリポートしていくことが重要です。 また、各県が創造的復興、もしくは再生を目指し、独自復興計画を策定、実現させていくことを都が後押しし、安全な地域社会の再建に寄与していく必要があります。
274行目 275行目 276行目	このように被災地が取り組むべき課題は山積し、日々刻々地域ごとに状況が変化しております。 被災地のニーズを的確に把握し、被災地、被災者が真に必要とする支援に今後とも継続して取り組むべきと考えますが、知事の見解を伺います。
277行目 278行目 279行目	現在、部内には福島県などから自主避難してきた約五千名の避難者の皆さんが都営住宅などに仮住まいをさせていただいております。 故郷から遠く離れ、いつ帰れるのかという思いを持って生活している皆さんに、都は寄り添う形でその生活を支えていくべきと考えます。 避難者は、見知らぬ東京での生活が不安であり、特に高齢者の方々については、引きこもりがちになるなど、孤立化も懸念されます。
280行目 281行目 282行目	先日、特別区の都営住宅で、自治会の皆さんが避難者と懇談会を開き、福島での共通の話題で盛り上がりしました。 こうしたかわり合いをふやす場でもあるミニ懇談会を開催し、避難者同士や地域との交流機会を創造することを求めています。 また、福祉も含めた総合的な相談を区市町村や災害復興まちづくり支援機構、NPOなどと連携して開催するなど、広い協働の形で避難者の暮らしを支えることも重要と考えます。
283行目	都は、コミュニティにも配慮した避難者に対する支援の取り組みを行っていくべきと考えますが、都の見解を伺います。
質問2 山下議員 284行目	東日本大震災を教訓に、東京においても震災時における社会対応力の強化や防災リーダーなど、地域人材の育成などに一層取り組み、東京を災害に強い持続可能な都市としていかねなければなりません。
285行目 286行目 287行目 288行目	現在、各道府県や市町村で地域防災計画などを見直す動きが出ています。 今回の震災による大津波は、近年研究が進みつつあった平安期の貞観地震に類似したものといわれています。 高知県や茨城県では、既に江戸期の地震の実例を盛り込み、地域防災計画の策定や浸水想定を行っています。 東京においても、江戸期に三連動地震による大津波、これに続く暴風雨や富士山噴火による複合災害が起きており、過去の災害分析からも改めて被害想定を研究すべきと考えます。
289行目 290行目	実践的訓練やサイロインの耐震化、震災化のさらなる推進も必要です。 福島原発事故を踏まえるのであれば、近い将来必ず起きるといわれている東海地震による静岡県浜岡原発リスクをも想定した放射能対策も万行なければなりません。
291行目	地震、津波の被害想定を検討や防災対策の総点検、そして東京の総合防災力をさらに高める取り組みが必要だと考えますが、知事の見解を伺います。

図 1: 東京都議会会議録 (本会議) の平成 23 年第 2 回定例会の例

徴として、実際の発言の書き起こしに近いことが挙げられる。厳密には、フィラーの削除や助詞の訂正など、読みやすくなるよう整文を行っているため書き起こしではない [7] が、いわゆる発言の要旨というレベルまで要約されてはいない¹。そのため、図に示すように一つの発言が非常に長い記述となっている。会議録中で「質問」や「答弁」など、どのような発言がどれだけなされているかを調査したものが図の左側で、10行単位で「議事進行」、「所信表明」、「質問概要」、「質問」、「答弁」といったラベルを振ったものである²。日本の会議録が書き起こしに近いものであるのは、47都道府県議会全てにおいて同様である。

3 イギリス地方議会会議録

本稿で対象としたのは、ロンドン議会³、ウェールズ議会⁴、スコットランド議会⁵、北アイルランド議会⁶、ジブラルタル議会⁷の5議会である。日本の地方議会

¹ 要旨のレベルで要約されたものは『議会だより』のような形で会議録とは別にまとめられていることが多い。

² ラベルは調査のために振った情報で、地方議会会議録コーパスには含まれていない。

³ <https://www.london.gov.uk/about-us/london-assembly>

⁴ <http://www.assembly.wales/en/Pages/Home.aspx>

⁵ <http://www.parliament.scot/>

⁶ <http://www.niassembly.gov.uk/>

⁷ <http://www.parliament.gi/>

会議録コーパスでは本会議を中心に構築したため、イギリス地方議会会議録コーパスでも plenary (総会) を中心に構築することとする。しかしながら、同じ総会の会議録であっても、日本の47都道府県議会の間の違いと比較すると、イギリスの5議会の違いは大きい。以下で各議会の特徴を述べる。

3.1 ロンドン議会会議録

ロンドン議会の会議録は、ASP ファイルで公開されている。内容は基本的に agenda and minutes として議題ごとの要旨がまとめられており、日本の会議録とは大きく異なっている。会議中の発言を書き起こしたものの (transcript) は、議題ごとの appendix として、それぞれ PDF および RTF ファイルで別に公開されている。また、plenary と別に市長への質問 (mayor's question time) がまとめられているのも特徴である。

3.2 ウェールズ議会会議録

ウェールズ議会の会議録もロンドン議会と同様に、agenda and minutes として議題と要旨が ASP および PDF ファイルで公開されている。会議中の発言は transcript として HTML ファイルで別に公開されている。ウェールズ議会の transcript の例を図2に示す。各

Cyfarfu'r Cynulliad am 13:30 gyda'r Llywydd (Elin Jones) yn y Gadair.		The Assembly met at 13:30 with the Llywydd (Elin Jones) in the Chair.	
 Y Llywydd / The Llywydd 13:30:03			
1 Galw'r Aelodau i drefn.		I call Members to order.	
1. Cwestiynau i'r Prif Weinidog		1. Questions to the First Minister	
 Y Llywydd / The Llywydd 13:30:05			
2 Yr eitem gyntaf ar ein agenda'r prynhawn yma yw'r cwestiynau i'r Prif Weinidog, a'r cwestiwn cyntaf, Dawn Bowden.		The first item on our agenda this afternoon is questions to the First Minister, and the first question is from Dawn Bowden.	
Buddsoddi mewn Twristiaeth ym Merthyr Tudful a Rhymny		Tourism Investment in Merthyr Tydfil and Rhymney	
 Dawn Bowden AM 13:30:12			
1. A wnaiff y Prif Weinidog ddatganiad am fuddsoddi mewn twristiaeth ym Merthyr Tudful a Rhymny? OAQ52176		1. Will the First Minister make a statement on tourism investment in Merthyr Tydfil and Rhymney? OAQ52176	

図 2: ウェールズ議会 Plenary(2018/05/15) の transcript (左側がウェールズ語, 右側が英語)

発言はウェールズ語 (図の左側) と英語 (図の右側) の両方で書かれており, 日本の会議録にはない特徴である。HTML ファイルでは以下のように両方の言語がペアで記述されており, これらの情報をデータ化することでパラレルコーパスとしての役割も期待できる。

```
<div class="contributionText" id="A42654">
  <div class="verbatim">
    1. Cwestiynau i'r Prif Weinidog
  </div>
  <div class="translation">
    1. Questions to the First Minister
  </div>
</div>
```

3.3 スコットランド議会会議録

スコットランド議会の会議録は, official reports として ASP および PDF ファイルで公開されている。official reports は要旨ではなく書き起こしに近く, 日本の会議録と近い形式である。「中断 (*Interruption*)」や「笑い声 (*Laughter*)」のような記述もあり, 議会の様子を忠実に記録しようとしている点も日本の会議録と似ている。

3.4 北アイルランド議会会議録

北アイルランド議会の会議録は, official report (hansard) として ASP および PDF ファイルで公開されている。内容は, スコットランド議会と同じで議会の様子を記録した書き起こしに近いものとなっている。北アイルランド議会の取り組みとして, オープンデータとして Web API による提供を行っている。

3.5 ジブラルタル議会会議録

ジブラルタル議会の会議録は, hansard として PDF ファイルで公開されている。内容は, スコットランド議会および北アイルランド議会と同じく議会の様子を記録した書き起こしである。ジブラルタル議会に限らずイギリス地方議会では基本的に日付により会議を識別しており, その単位で会議録 (ファイル) を作成している。しかしながら, ジブラルタル議会では同じ日付であっても morning session と afternoon session で別のファイルを作成しており特徴的である。

以上の 5 議会の特徴を表 2 にまとめる。

表 2: イギリス地方議会会議録の特徴

	内容	公開
ロンドン	要旨+書き起こし	ASP, PDF, RTF
ウェールズ	要旨+書き起こし	ASP, PDF, HTML
スコットランド	書き起こし	ASP, PDF
北アイルランド	書き起こし	ASP, PDF, Web API
ジブラルタル	書き起こし	PDF

4 データ構造の検討

前節で述べた特徴を基に、表 1 のフィールドが適用できるか考察する。まず、「都道府県名」だけでは不十分で「国名」のフィールドが必要である。また、ジブラルタル議会の事例から、「年」「月」「日」に加えて「時間」のフィールドが必要である。「発言文」に関しても、ウェールズ議会のように、異なる言語で同じ発言をしている場合があり、その情報を保持する必要がある。一方で、データベースとして管理することを考えると、日本語や英語といった「言語」フィールドを用意して、言語ごとに記録することが効率的である。そこで、「翻訳文」フィールドを用意し、他言語の同じ発言の「識別子」を記録することを考えている。

イギリス地方議会会議録を分析することで、我々が地方議会会議録コーパスでこれまで用いてきたデータ構造を修正する必要があることが分かった。今回の修正は、日本とイギリスを比較した上で行われたものであり、他の国の会議録にも対象できるかについては未検討である。今後も各国の会議録を実際に扱う際に段階的に修正を行い、統一的なデータ構造を作成していきたいと考えている。

5 まとめ

本稿では、イギリス地方議会会議録コーパスとして、ロンドン議会、ウェールズ議会、スコットランド議会、北アイルランド議会、ジブラルタル議会の 5 議会を対象としてデータを収集し、日英両国の地方議会会議録を統一的に扱うためのデータ構造を検討した。現在、2016 年 1 月からの 1 年分のデータを収集しており、今後コーパスの公開方法とともに、さらにデータを拡充していきたいと考えている。また、我々は評価型ワークショップ NTCIR-14 において、地方議会会議録コーパスを用いた「政治情報を対象とした質問応答タスク」QA Lab-PoliInfo を開催しており、イギリス地方議会会議録においても同様の取り組みができないか検討していきたい。

謝辞

この研究は平成 30 年度国立情報学研究所公募型共同研究の助成を受けています。

参考文献

- [1] 筒井貴士, 我満拓弥, 大城卓, 菅原晃平, 永井隆広, 渋谷英潔, 木村泰知, 森辰則. 地方議会会議録コーパスの構築および政治情報システム構築を目標としたアノテーションの一提案. 自然言語処理, Vol. 21, No. 2, pp. 125–156, 2014.
- [2] Yasutomo Kimura, Keiichi Takamaru, Takuma Tanaka, Akio Kobayashi, Hiroki Sakaji, Yuzu Uchida, Hokuto Ototake, Shigeru Masuyama. Creating Japanese Political Corpus from Local Assembly Minutes of 47 Prefectures. Coling 2016 workshop, The 12th Workshop on Asian Language Resources, pp.78–85, 2016.
- [3] Hassanali, Khairun-nisa, and Vasileios Hatzivassiloglou. Automatic detection of tags for political blogs. Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media, Association for Computational Linguistics, pp. 21-22, 2010.
- [4] Chambers, N., Bowen, V., Genco, E., Tian, X., Young, E., Harihara, G., and Yang, E. Identifying political sentiment between nation states with social media. Proceedings of EMNLP, pp. 65-75, 2015.
- [5] Lerman, Kevin and Gilder, Ari and Dredze, Mark and Pereira, Fernando. Reading the markets: Forecasting public opinion of political candidates by news analysis. Proceedings of the 22nd International Conference on Computational Linguistics Vol. 1, pp. 473-480, 2008.
- [6] Cano-Basave, Amparo Elizabeth and He, Yulan. A Study of the Impact of Persuasive Argumentation in Political Debates. Proceedings of NAACL-HLT, pp. 1405-1413, 2016.
- [7] 高丸圭一, 木村泰知. 栃木県の地方議会会議録における整文についての基礎分析-本会議のウェブ配信と会議録の比較-. 宇都宮共和大学 都市経済研究年報, Vol. 10, pp. 74–86, 2010.