



Kobe Shoin Women's University Repository

Title	言語資料としての国会会議録検索システム The on-line full-text database of the Minutes of the Diet: Its potentials and limitations
Author(s)	松田 謙次郎 Kenjiro Matsuda
<i>Citation</i>	Theoretical and applied linguistics at Kobe Shoin, No.7 : 55-82
Issue Date	2004
Resource Type	Bulletin Paper / 紀要論文
Resource Version	
URL	
Right	
Additional Information	

言語資料としての国会会議録検索システム*

松田謙次郎

The on-line full-text database of the Minutes of the Diet: Its potentials and limitations

Kenjiro Matsuda

Abstract

The on-line full-text database of the Minutes of the Diet offers linguists a unique resource for corpora study of the modern Japanese language; with all the debates and information of the session searchable by keywords, and the name of the speaker, date, House, etc., all laid out in an easy-to-use interface. The database is accessible from ordinary internet browsers, and the search results are easily downloadable to the user's PC. This article explores this resource's possibilities for various linguistic research (lexicon, syntax, dialectology and discourse analysis), demonstrates actual searches and their results, and carefully examines the limitations that necessarily arise from several sources (e.g. editorial practices of the Diet Office transcribers).

1. はじめに

国会会議録は、衆参両議院の国会会議録と、その委員会記録を納めたものである。日本は1890年の国会開設以来、この記録が断続することなく連続と続けられているという世界的にも稀有な記録を持つ国の1つであるが、この記録は従来印刷物としてのみ頒布されて来ている。衆参両院事務局と国立国会図書館は国会会議録フルテキスト・データベース・システムの構想を1992年より持ち始め、共同で構築作業を開始し、1999年より一

*本稿の執筆に当たっては、衆議院速記者養成所と参議院速記者養成所において、資料を閲覧させて頂いた。参議院記録部と広報課には、いくつか質問に答えて頂いている。また、朝日祥之、太田一郎、助川泰彦、フィリップ・シュベルティ、マーク・スコット、高野照二、橋本力、二階堂整の各氏から有益なコメントを賜った。感謝申し上げたい。本研究の一部は、日本学術振興会科学研究費補助金（基盤研究（B）「言語における制約間のインターフェースに関する総合的研究」（平成12年度～15年度）、研究代表者：西垣内泰介、課題番号12410129）を受けている。

部運用開始、2001年よりインターネット上での公開が本格的に開始された(鈴木(1994, 77)、参議院50年のあゆみ編集委員会(1998, 296)、奥村(1997, 41-42)、総務庁(1999)、国立国会図書館(2001))。これで新憲法下で開催された国会のほぼすべての会議録が、¹国会会議録検索システムウェブサイト(<http://kokkai.ndl.go.jp/>)においてオンラインで検索できるようになったわけである。国会会議録は、戦後政治史研究者にとっての第一級史料であるが、別な観点からこれを考えると、日本各地出身である成人の、戦後60年近くにわたる改まった話し言葉での膨大な発話記録とも捉えることができる。話者(議員)の生年で考えると、19世紀後半から20世紀後半までとなり、理論上は100年にわたる範囲の話者の発話を納めていることになる。しかもそれが簡単に検索できる状態に置かれているわけである。言語研究者であれば、当然国会会議録をコーパスとして使うことを考えても不思議はない。

ここでは、このオンライン版国会会議録(正式名称は「国会会議録検索システム」、以後「検索システム」と略称)を言語資料として使用する場合の注意点・問題点を指摘し、併せてこうした議会会議録をデータとした研究の現状を概観し、その可能性を探ることとしたい。なお、国会会議録は、「国会議事録」という名称も使われることがあるが、ここでは以後「会議録」という名称を使うことにする。

2. 国会会議録

まず、国会会議録について、従来のシステムを簡単に確認しておこう。国会がその会議録を速記によって逐語的記録として作成し(衆議院規定第15章、参議院規定第10章)、それを公表頒布すること(憲法第57条)はいずれも法の定めるところである。国会会議録はこうした法律的根拠によって取られる、国会の公式記録である。

会議録の作成には、速記現業(速記、反訳)、調査、校閲編集といった行程(これらについては§5.3を参照)を経るが、これらの作業は両院の記録部によって行われている。こうした作業のアウトプットは、本会議、予算委員会などの翌日に発行される速報版、官報号外として発行される本会議会議録、そして委員会会議録²がある。速報版は議員対象であるが、後の2つについては一般でも入手可能である。

会議録は長年にわたり紙媒体に記録されてきたわけだが、こうした記録の電子化に当たっては、1998年の第144回国会までの記録については過去の会議録をOCRで読みとり、読みとりエラーを修正して電子テキストとしている。第145回国会からは、電子的に作成された会議録を、そのまま公開に回している。縦書きが横書きに変更された以外は、すべてのフォーマットがそのまま継承されている。このことは、検索結果の本分表示画面から呼び出せる、会議録の紙媒体での画像からも確認できる(§3.2、図11参照)。

会議録検索システムの登場によって、第1回国会(1947年5月開会)以降の本会議、全ての委員会が電子的に公開されることになったわけであり、検索システムウェブサイト

¹ 閲覧・検索できない例外については§4.で述べる。

² 参議院での正式名称は「委員会会議録」。両院はそれぞれ独自の速記者養成所を持つなど、速記録作成についても微妙に異なるところがある。

の「FAQ よく寄せられる質問」(http://kokkai.ndl.go.jp/KENSAKU/www_faq_top.html)によれば、最新の国会にしても2~3週間で公開に至るという。なお、これはまだ実現していないが、このデータは将来はCD-ROMに収められ公開される予定とのことである(参議院50年のあゆみ編集委員会, 1998, 297)。そうなれば、研究者各自のパソコン上で、今より格段に自在にかつ効率的に検索が行えるようになるだろう。

3. 国会会議録検索システムの使い方

3.1 検索条件指定

では早速検索システムの実際の使い方を見てみよう。³ サイト (<http://kokkai.ndl.go.jp/>) にアクセスすると、まず図1の画面が出る。⁴

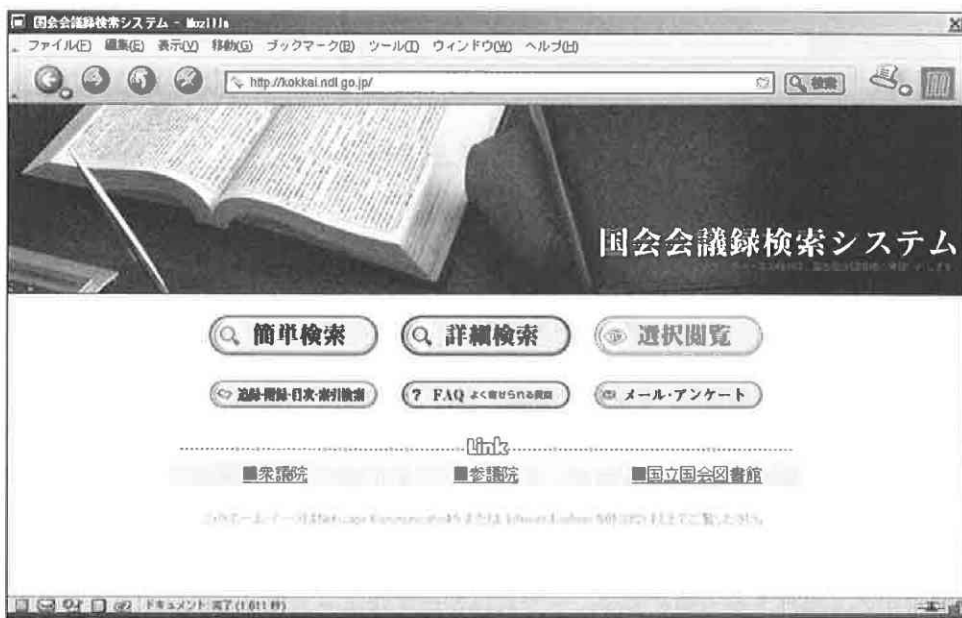


図1: 国会会議録検索システムトップページ

検索に最も関わる箇所は、画面中央の「簡単検索」「詳細検索」の2箇所である。検索に2つのオプションがあるわけだが、「簡単検索」(図2)と「詳細検索」(図4)の違いは、前者がシンプルな指定のみなのに対して、後者ではさらに検索のオプションが提供されている点である。いずれのページでも、画面が縦に2分割されている。

³検索プログラムはJavaScriptを使用しているので、ブラウザの設定で、JavaScriptを許可にしないと作動しない。なお、この稿を書くに当たって、Windows98SE/Internet Explorer 5、WindowsXP/Internet Explorer 6、MacOSX/Internet Explorer 6、Vine Linux (Linux version 2.4.18-0v13)/Mozilla/5.0の4つの組み合わせで検索サイトにアクセスしているが、いずれの場合も問題はなかった。以後解説で表示してある画面は、LinuxでMozillaを使用した際のものである。

⁴このアドレスからも分かる通り、このサイトは物理的には国立国会図書館に属しているが、データベース全体の著作権は国会図書館に、個々の発言の著作権は各発言者にある(検索システム「FAQページ」による)。

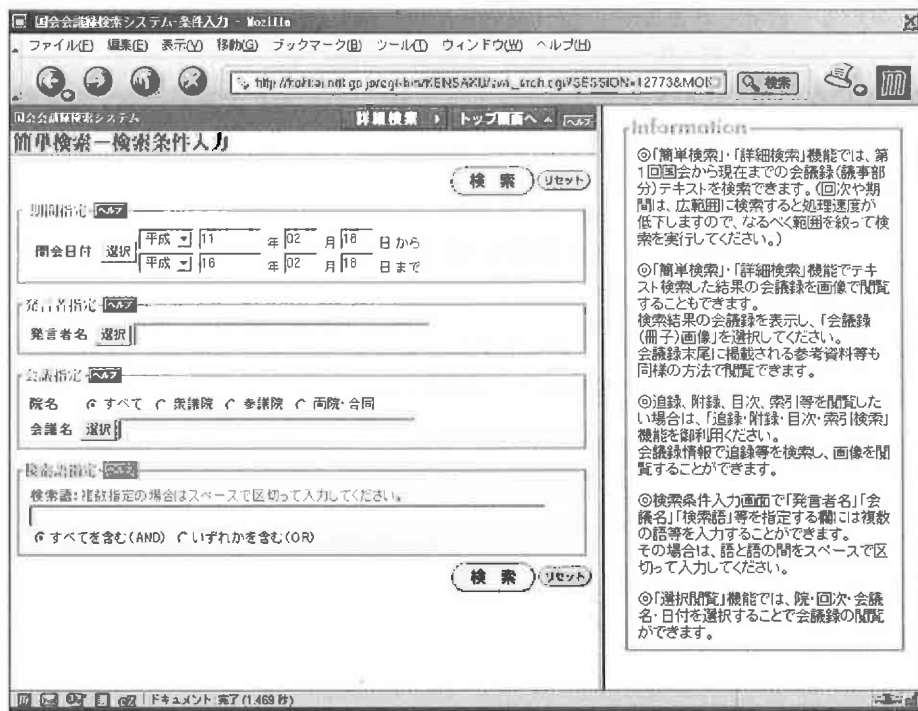


図 2: 簡単検索トップページ

簡単検索では、開会日付、発言者、会議、そして検索語から検索できる仕組みになっている。このうち、最初の開会日付のみが必須項目である。ページを開けた段階では、下段に現時点で最新の会議録の日付が、上段にその 5 年前の日付が表示されている。日付指定には、年号、年、月、日の各ボックスに自分で日付をタイプしていく方法⁵と、「選択」ボタンを押して、右ウィンドウに会期や種別（臨時・特別・通常）と共に表示される国会回次を選択する方法がある。この日付のみを指定して検索を実行すると、その期間に開催された国会の会議録すべてがヒットするわけである。例えば、「昭和 22 年 5 月 20 日」から「昭和 22 年 5 月 20 日」と指定し、「会議指定」が「すべて」（デフォルト値）であると、第 1 回国会が 2 件（参議院本会議と衆議院本会議）がヒットする。

この「選択」ボタンによる指定は、次の「発言者指定」と「会議指定」でも有効である。発言者指定は、検索対象とする議員名を入力するわけだが、窓に直接議員名を入力することも、また「選択」ボタンによって候補を探すことも可能である。「選択」ボタンをクリックすると、やはり右ウィンドウの議員名検索のための窓の下に、議員一覧も表示される。この議員一覧から選ぶことも可能だが、この方法だと 209 ページもあるページの選択をしなければならない。むしろ、「部分一致」と「先頭一致」というオプションを持つ議員名検索を使う方が多くの場合賢明だろう。いずれかの方法で議員名を検索し

⁵ 予め入力してある日付は 1 桁であっても「01」とゼロを足した 2 桁形式になっているが、1 桁で入力しても何ら支障はない。また、半角でも全角でも構わない

たら、「セット」ボタンをクリックすれば、左のウィンドウにその議員名が入力されるようになっている。議員によっては、作家などの場合に、別名で議員として登録されている場合がある。その場合、検索をすると「別名選択画面」が現れ、改めて議員を選択する仕組みになっている。図3は、「金子洋文」という名で議員登録をしていた「金子吉太郎」を検索した場合の画面である。ここで「金子洋文」にチェックを入れてから、検索実行すればよい。

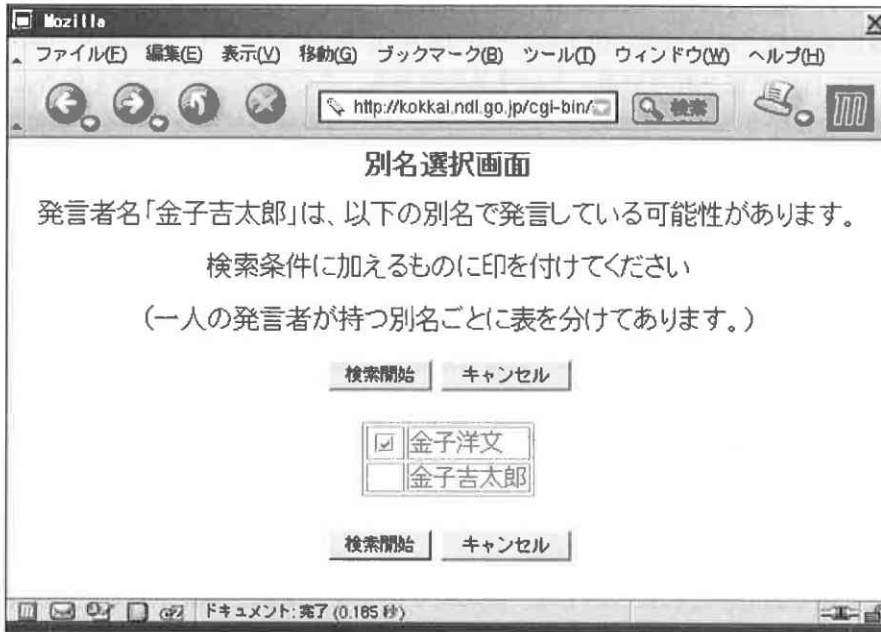


図3: 「金子吉太郎」で引いた別名選択画面

次に「会議指定」の項目を見てみよう。「すべて」「衆議院」「参議院」「両院・合同」の4つがラジオボタンで選択できるようになっているほか、会議名が選択となっている。「両院・合同」というのは、〇〇合同審査会とか、〇〇両院協議会という名称で開催されている会合を指す。これに対して「すべて」というのは、「衆参両院、また合同もあわせたすべての会議」という意味である。「会議指定」も「期間指定」や「発言者指定」と同じく、「選択」ボタンをクリック、出てくる右画面で検索・選択ができる。「期間指定」も同様だが、「院名」で選択範囲を指定する形になっているので、例えば「衆議院」を選択しておく、「選択」ボタンを押して出てくるのは衆議院の本会議・委員会のみである。もちろん、すでに正確な会議名が分かっているならば最初から「会議名」の欄に入力すれば良い。

そして4つ目の指定が「検索語指定」である。単語間をスペースで区切って入力すれば、複数単語の指定も可能で、それらの関係を AND (すべての語を含む) か OR (どれかの語を含む) で指定できる。ただしワイルドカードや、正規表現による検索式は使えない。またヘルプファイルに書いてないことだが、英数字の半角・全角、また大文字・

小文字の区別はしない。よって a A A a のいずれを入力しても PTA や AP がヒットする（ただし会議録のテキストでは全角を採用している）。「簡単検索」の画面は、ほぼ以上に尽きる。

では「詳細検索」はどこが違うのだろうか。（図4）は「詳細検索」の画面である。これを見ると、「期間指定」には「開会回次」が、「発言者指定」には「肩書き」「所属会派」「役割」といった項目が、そして「会議指定」には「号数」の指定が可能になっている。「号」とは、同一会議の（例えば第159回衆議院予算委員会）の会合順を表す。例えば、第159回衆議院予算委員会の第1号は平成16年01月23日であり、次の会合である2号は同月26日に開催されている、という具合である。この号数指定の右端を見ると、「閉会中審査」というボックスがあり、チェックが入れられるようになっている。実は国会が閉会中であっても、常任委員会および特別委員会は、各議院の決議さえあれば議案の審査や調査が可能である（国会法第47条）。「閉会中審査」ボタンは、こうした閉会中審査を検索する場合に使われる。

簡単・詳細検索いずれの検索指定項目でも、入力窓のそばにヘルプボタンが配置されており、別ウィンドウで詳しい説明が見られる配慮がなされている。このため、ほとんどの検索は全くの初心者でも難なくこなせるはずである。なお、検索ヒット数の上限は1,000件に設定されており、1,000件を超えると検索結果の一覧ができなくなってしまう（図5）。この場合、「簡単検索—検索結果表示」画面の上にある「検索条件変更画面」を選んで、入力した検索条件を訂正するか、右の「検索条件入力」を選択して、最初から条件入力を行うことになる。

検索結果は画面でも確認可能だが、ファイルとして使用中のPCにダウンロードすることも可能である。これについては、次のセクションで事例と共に示すことにする。

3.2 検索実行・結果の表示・ダウンロード

では実際にこうして検索条件指定をした上で、検索を実行してからの流れを、実際に検索を実行しながら解説してみよう。

単純な検索語1語を指定した場合

まず試しに、「考えれる」という語を使ってみよう。これはいわゆるら抜き言葉で、しかも語幹が長いことから、「考えられる」という形で共通語では使われることが多い。そこで、国会会議録での出現状況を見るために、以下の条件で検索をしてみよう。

【開会日付】 昭和22年5月20日から平成16年2月4日まで

【発言者指定】 なし（発言者名＝空白）

【会議指定】 すべて（会議名＝空白）

【検索語指定】 考えれる

開会日付は、第1回国会開会日から最新の会議録までのものであり、検索語指定では、AND にしてあるが、デフォルト値がANDなので、そのままにしてあるだけである。こ

図 4: 詳細検索トップページ

The screenshot shows a web browser window titled "国会会議録検索システム-条件入力 - Mozilla". The main content area is titled "詳細検索-検索条件入力". It contains several sections for inputting search criteria:

- 期間指定:** Includes fields for "開会日付" (Start Date) and "国会回次" (Session Number). The start date is set to Heisei 11, year 02, month 16. The session number is set to Heisei 16, year 02, month 16.
- 発言者指定:** Includes fields for "発言者名" (Speaker Name), "肩書き" (Title), and "所属党派" (Political Party). There are also checkboxes for "証人" (Witness), "参考人" (Reference Person), and "公述人" (Public Speaker).
- 会議指定:** Includes fields for "院名" (House Name), "会議名" (Meeting Name), and "号数" (Number). The "院名" is set to "すべて" (All). There are checkboxes for "衆議院" (House of Representatives), "参議院" (House of Councillors), and "両院・合同" (Both Houses/Joint). There are also fields for "号数" (Number) with "号から" (From Number) and "号まで" (To Number) options, and a checkbox for "開会中審査" (Review During Session).
- 検索語指定:** Includes a text input field for "検索語" (Search Term) and a note: "検索語:複数指定の場合はスペースで区切って入力してください。" (Search Term: If multiple terms are specified, separate them with spaces). There are checkboxes for "すべてを含む(AND)" (Include all) and "いずれかを含む(OR)" (Include any).

Buttons for "検索" (Search) and "リセット" (Reset) are located at the top right and bottom right of the form area.

Information

- ◎「簡単検索」・「詳細検索」機能では、第1回国会から現在までの会議録(議事部分)テキストを検索できます。(回次や期間は、広範囲に検索すると処理速度が低下しますので、なるべく範囲を絞って検索を実行してください。)
- ◎「簡単検索」・「詳細検索」機能でテキスト検索した結果の会議録を画像で閲覧することもできます。検索結果の会議録を表示し、「会議録(冊子)画像」を選択してください。会議録末尾に掲載される参考資料等も同様の方法で閲覧できます。
- ◎追録、附録、目次、索引等を閲覧したい場合は、「追録・附録・目次・索引検索」機能を御利用ください。会議録情報で追録等を検索し、画像を閲覧することができます。
- ◎検索条件入力画面で「発言者名」「会議名」「検索語」等を指定する欄には複数の語等を入力することができます。その場合は、語と語の間をスペースで区切って入力してください。
- ◎「選択閲覧」機能では、院・回次・会議名・日付を選択することで会議録の閲覧ができます。

The status bar at the bottom indicates "ドキュメント: 完了 (1.587 秒)".



図 5: 検索結果が制限値を越えた場合の表示画面

これらの条件を入力すると、画面は図 6 のようになる。そのまま検索を実行した結果の画面が図 7 である。

青字で検索ヒット数 (26 件) が表示され、その横に「検索結果一覧表示」と書かれたボタンが置かれている。下の段は検索条件の再表示である。「検索結果一覧表示」のクリックで、画面は図 8 に変わる。

デフォルトでは、上から新しい順に「考える」が出現した会議が並べられている。これは「日付による並べ替え」ボタンで逆順での並べ替えが可能であり、また複数の同じ会議がある場合は、「会議名によるグルーピング」が便利だろう。今回の検索の場合、一画面では収まりきれないので、2 ページ目にも続いているが、続きを見るには画面下の「次ページ」をクリックする。古い順に並べ替えて最初の項目を見ると、国会会議録ではすでに昭和 22 年 8 月 5 日の第 1 回国会（参議院 財政及び金融委員会 第 8 号）で使用されていることが分かる。実際の発言を見るには、それぞれの会議名をクリックすればよい。「参議院財政及び金融委員会 第 8 号」をクリックして現れた画面が、図 9 である。

画面は、4 フレームで構成されている。左上が発言者フレーム、右上が会議録情報フレーム、左下がダウンロード用フレーム、そして画面の大部分を占めている右下が本文フレームである。発言者フレームは、この会議での発言を発言順に並べたものである。冒頭（ファイル番号 000）の「会議録情報」とは、その会議の開催日時、審査議題などを記したものである。⁶ 実際にこの文字部分をクリックして中身を見ると、この会議が昭和 22 年 8 月 5 日（火曜日）午前 10 時 38 分に開会され、「國民貯蓄組合法の一部を改正す

⁶ 第 1 号の会議録には、すべての委員名が国会の慣例に従って「君」という統一の敬称を付けられて記されている。

期間指定

開会日付
 昭和 年 月 日から
 平成 年 月 日まで

国会回次
 _____ 回から _____ 回まで

発言者指定

発言者名 _____
 肩書き _____
 所属会派 _____
 役割 証人 参考人 公述人

会議指定

院名 すべて 衆議院 参議院 両院・合同
 会議名 _____
 号数 _____ 号から _____ 号まで 開会中審査

検索語指定

検索語: 複数指定の場合はスペースで区切って入力してください。
 | 考えれる |
 すべてを含む(AND) いずれかを含む(OR)

図 6: 「考えれる」の検索条件入力画面

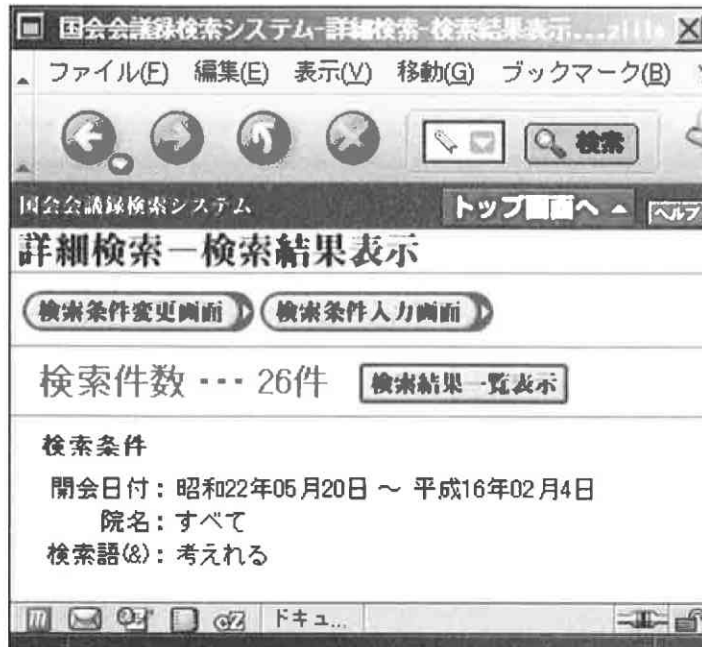


図7: 「考えられる」の検索結果画面



図8: 「考えられる」の検索結果一覧表示画面

図 9: 参議院財政及び金融委員会第 8 号 表示画面

国会会議録検索システム-本文表示(MAIN) - Mozilla

ファイル(E) 編集(E) 表示(V) 移動(G) ブックマーク(B) ツール(I) ウィンドウ(W) ヘルプ(H)

http://kokkai.ndl.go.jp/cgi-bin/KENSAKU/swk_dispdoc.cgi?SESSION=9686&SA

国会会議録検索システム

本文表示

検索結果一覧画面 国会会議 次会議 検索条件入力画面

[026/026] 1 - 参 - 財政及び金融委員会 - 8号
 昭和22年08月05日

発言者: 前次 2 / 30 検索語: 前次

発言者名 対応頁

[000]会議録情報	1
[001]黒田英雄	1
[002]福田赳夫	1
[003]黒田英雄	2
[004]深川タマエ	2
[005]福田赳夫	2
[006]森下政一	2
[007]福田赳夫	3
[008]玉屋喜章	3
[009]福田赳夫	4
[010]玉屋喜章	4
[011]福田赳夫	4
[012]玉屋喜章	4
[013]福田赳夫	4
[014]黒田英雄	4
[015]赤澤與仁	4
[016]福田赳夫	5

全選択 全削除 発言のみ選択

番号指定 番～ 番 選択

ダウンロード

本文表示

見ますと、非常に重大な問題と相成りまして、その後預金をするという傾向が極めて貧弱なものとなつて参つたのであります。従いまして、三月以降におきまして、毎月六、七十億という金額が通貨増強として出まして、そうしてこの勢は決して底止するところないような状況でありますと、結局いつの日にか通貨全体の量というものが非常に多額な量に上る、非常に通貨の増強ということになりますれば、そのこと自體だけでも、通貨の信頼感という見地から見ますと、これは破滅的な問題となつて来るというふうに考えられるのであります。尚、終戦後の経済全体の建直しの見地から見ても、國民全體がその日その日の生活から幾らかの蓄積を残して行く、産業にいたしまして、その日その日の企業經理の結果というものが、何らかの形でプラスになつて行くことで、これは企業のいわゆる生産擴大ということにはならないのであります。さような見地から見ても、その裏をなすものは、貯蓄ができる、資金の蓄積ができるという問題であるのであります。これは貯蓄という問題を進行して行くことが必要であらうということを感じて至つたのであります。幸いに昨年の夏の議會が開かれております間におきまして、衆議院を中心といたしまして、議會方面におきまして、これは通貨の信任、再建という問題で、非常に關心を拂はなければいかんということに相成りました。共産黨以外の當時の五黨の決議を以ちまして、通貨安定決議案というものが衆議院に提出され、その決議を見たわけでありまして、たまたま政府におきまして、只今早上げた通り、さような趣旨のことを考えておりますので、議會、政府一體となりまして、この問題を推進するということに相成りまして、今日に至つてあるわけでありまして、

さような氣運の起ります時までにおける預金の趨勢というものは極めて貧弱であつたのであります。衆議院におきまして通貨安定の議が起りました頃から前後いたしまして、預金は相當増加の趨勢を示しております。即ち十月には七十二億圓の預金の増加を見て、十一月には七十二億圓、十二月が百五億圓、更に今年に入りまして、一月に六十八億圓、二月に七十八億圓、三月に百三十四億圓、四月に六十三億圓、五月に九十五億圓、六月に百十億圓、かような数字が出て多つておるのであります。昨年の十一月からの貯蓄増強ということの本當に計畫といたしてやつたのであります。十一月から三月までの貯蓄増加の目標は五百億であります。それに對しまして、達成した貯蓄の額というものは四百五十四億、即ち九〇%の成績を収めておるわけでありまして、尚、本年度におきましては、自由預金の増加目標額を月百億圓と決めておるわけでありまして、百億圓と決めましたのであります。これをいかに使うかという問題であります。これは百億圓の増強という目標を達成するに必要とするものは、

ドキュメント: 完了 (6 832 件)

る法律案」他9件がこの委員会に付託されており、そのうちこの法律案だけを審査したことがわかる。⁷

「会議録情報」の下からは、発言順にチェックボックス（ダウンロード時に使用）＋[3桁番号]＋議員名のフォーマットで発言者が並ぶ。検索語を含む発言をした発言者の名前は、太字で表示され、そこだけ青い帯が掛けられている（これが見あたらない場合は、下に向かって探すと良い）。青い帯の掛けられている議員名をクリックすれば、その発言を右のウィンドウで読むことができる（ただし「財政及び金融委員会」をクリックした段階で、すでにその発言部分は表示されているはずである）。

会議録情報フレームでは、「検索結果一覧画面」など4つの大きめのボタンが一目に並ぶ。「検索結果一覧画面」とは、この前の画面であり、要は「戻る」のような機能を果たす。「前会議録」「次会議録」で、検索語を含んだ会議録間を移動できる（この2つのボタンの「前」「次」は、あくまで検索結果一覧の並びについての順番である）。右端の「検索入力画面」は、「簡単検索」の最初の検索条件入力画面に戻る。この場合、前に入力した情報がすべてクリアされた状態になる。以前に入力した検索条件を一部手直ししたい場合は、一旦「検索結果一覧画面」に戻り、そこから「検索条件変更画面」に戻った方がよい。

4つのボタンの下には、現在閲覧している会議情報と日付が書かれている。会議情報のフォーマットは、[検索結果一覧での順番／検索ヒット総数][国会会期][衆参別]-[会議名]-[号数]である。さらにその下にあるのが、発言者と検索語の移動ボタンである。発言者の「前」「次」ボタンを使えば、左ウィンドウで発言者を指定しなくても、次々と発言者をたどっていくことができる。「検索語」は、隣接する検索語出現箇所へ移動する。

検索語出現箇所を示す本文フレームには、発言者のその発言全体が表示される。財政及び金融委員会を選択した段階で、このフレームには、福田赳夫議員の発言が表示されているはずである。「考えれる」は図10のような文脈で出現することがわかる（下線筆

従いまして、三月以降におきまして、毎月六、七十億という金額が通貨増発として出まして、そうしてこの勢は決して底止するところなしというような状況でありますと、結局いつの日にか通貨全体の量というものが非常に多額な量に上る、非常に通貨の増発ということになりますれば、そのこと自體だけでも、通貨の信頼感という見地から見ますと、これは破壊的な問題となつて来るというふうに考えれるのであります。

図10: 「考えれる」の文脈：昭和22年8月5日第1回国会 参議院財政及び金融委員会 第8号 福田赳夫議員の発言

者)。これで見ると、確かに抜き言葉であるが、昭和22年の段階で出現していることは、全くあり得ないことではないにしろ、やはり驚きだという感想を持つだろう。そ

⁷国会に提出された法案は、原則として、まず提出された院の議長が所管委員会に付託し、そこでの審査・可決後に本会議で審議に至る（国会法56条2項）。

ここで、この発言が本当にら抜きであったのかを確認することが必要になる。

会議録作成が現在のように電子化される以前の会議録については、すでに発行されている会議録をスキャンし、OCRで読みとっているため、その際の読みとりエラーという可能性がある。これを検証するには検索システムに付属している、紙媒体で発行された会議録の原本画像閲覧機能を使えばよい。これは、「本文表示」右上フレーム右端にある「会議録(冊子)画像」というボタンで行える。実際にクリックすると、紙媒体バージョンでの当該箇所の画像(TIFF形式、200Kb程度)が見られる。この方法で福田議員の当該発言を確認すると、やはりそれが「考えられる」であったことが確認できる(図11)。⁸

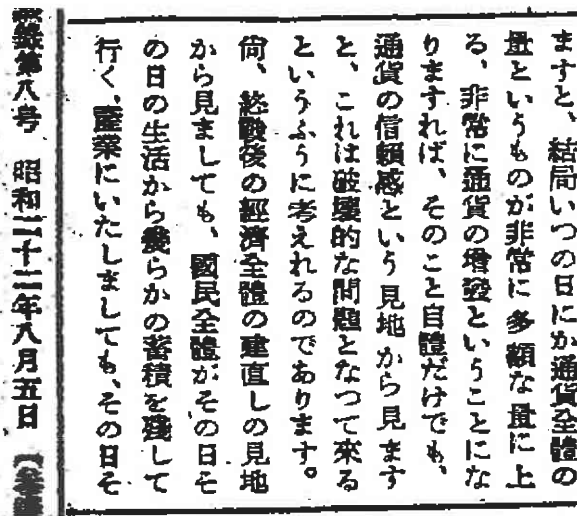


図 11: 参議院財政及び金融委員会第 8 号 福田発言部分 [拡大表示]

検索結果のダウンロード

ところで、こうした結果を自分の PC にダウンロードするにはどうすればよいだろう。検索システムは、この点非常によいインターフェースを提供している。上の福田議員の例で見てみよう。発言者情報フレームを見ると、発言者名の前にチェックボックスが見えるはずである。ダウンロードしたい発言者にチェックを入れて、その下のダウンロード用フレームで「ダウンロード」をクリックすると、発言者名確認画面を経て、「download.txt」というファイル名でダウンロードが開始される。もちろん、ここで青帯の付いていない発言者を選択しても構わない。拡張子からもわかる通り、ファイルのフォーマットはテキストファイルで、冒頭に会議録情報が付されていて、その発言すべてが含まれている

⁸実はこうして確認でミスが見つかることも実際にある。例えば検索システムで探すと、「考えられない」の国会会議録における最初の出現は、1947年10月15日の衆議院農林委員会第33号における大島義晴議員の発言(「もちろん水害の原因に山林の伐採ということが関係なしとは申されませんが、これだけが水害の原因であつたとは考えられないのであります」)となるはずだが、同箇所を画像ファイルで確認すると、「考えられる」の誤りであることがわかる(2004年2月22日時点での確認)。こうしたミスは、検索システムの「メール・アンケート」ページを使うなどして、利用者がこまめに報告してデータの修正を促し、データの質を高めて行くべきだろう。なお、この点については §5.1 でも触れる。

(図 12)。

1-参-財政及び金融委員会-8号 昭和 22 年 08 月 05 日
 ○政府委員（福田赳夫君） それでは貯蓄運動を開始してからの状況について御説明申し上げます。御承知の通り、昨年三月三日に金融緊急措置という非常の措置をとりました。あの措置は通貨の信任という問題から論じますと、極めて重大な問題であつたわけでありまして、大藏當局といたしましては、最後までああいう措置をとるということにつきまして、非常な逡巡をしておつたわけでありまして、諸般の情勢上ああいう措置をとつたわけでありまして、果せるかな、この措置をとりました後におきましては、通貨に対する信任という見地から見ますと、非常に重大な問題と相成りまして、その後預金をするという傾向が極めて貧弱なものとなつて参つたのであります。従いまして、三月以降におきまして、毎月六、七十億という金額が通貨増發として出まして、そうしてこの勢は決して底止するところなしというような状況でありますと、結局いつの日にか通貨全體の量というものが非常に多額な量に上る、非常に通貨の増發ということになりますれば、そのこと自體だけでも、通貨の信頼感という見地から見ますと、これは破壊的な問題となつて來るといふふうに考えられるのであります。

図 12: download.txt の内容（冒頭部分）

複数検索語指定—AND 指定の場合

簡単な検索手順を見た所で、次にもう少し複雑な同じく可能形を検索例にして見てみよう。「見る」の可能形で、保守的な「～られ」形の可能形と「～れ」形の革新的可能形で、肯定と否定のペアを考えるとする。すると、見られる—見られない、見られる—見れない、見れる—見られない、見れる—見れない、の4つの組み合わせが考えられる。ここで同じ発言者がこれらを使う場合を考えると、おそらく等位接続（「見られてもみられなくても」、「見れるか見れないか」など）で2つが繋がれているか、少なくとも近接して出現している可能性が高いと考えられよう。その場合、革新形と保守形を組み合わせるよりは、どちらも革新、ないしは保守形に揃える傾向があることが予想できるだろう。つまり、保守—保守、革新—革新という組み合わせの方が保守—革新、革新—保守という組み合わせよりも一般的だという仮説である。⁹ この仮説を簡単に確かめるには、上のそれぞれのペアを検索語指定に入力し、「すべてを含む (AND)」をチェックした上で検索することを繰り返し、それぞれのヒット数を比較すればよい。¹⁰ 実際に行ってみると上の仮説がほぼ正しいことが確かめられる（表 1）。

⁹ 実際に自然発話データではではこうした傾向が Matsuda (1993, 31) に報告されている。

¹⁰ 無論これはきわめて粗い比較である。正確には、少なくとも 1. テキスト画像の検討、2. 保守形が本当に可能として出現しているかの確認、3. 同一会議録での複数回出現の場合の出現頻度チェック、4. 近接の度合いを何らかの形で測定した上での革新形の出現しやすさとの相関分析、そして 5. 他の動詞での同様な比較、の5つはすべきであろう。ここでは、あくまで AND 機能の使用例を示すための便宜上この例を引いている点に注意されたい。

ペアの種類	ヒット数
見られる & 見られない	708
見れる & 見れない	60
見られる & 見れない	24
見れる & 見られない	13

表1: 「見る」の可能形 — 肯定・否定、保守・革新形の組み合わせによる出現頻度数

複数検索語指定 — OR 指定の場合

OR 指定が威力を発揮するのは、表記のゆれの場合であろう。日本語では漢字、カタカナ、ひらがな、ローマ字の4文字種が使われるが、単語によってはその表記が一定でないものがある。例えば、「空き缶」については、「あき」か「空き」か「空」か、そして「缶」についても「カン」「罐」などの異表記が見られる。よって「空き缶」の表記にはいくつもの可能性があることになるわけで、こうした表記のゆれが日本語情報処理の大きな問題となっていることは、周知の事実である。国会会議録では、『国会会議録用字例』（衆議院記録部・参議院記録部, 1995）に表記基準を定めてはいるが、必ずしもこれですべてが解決するわけではない。実際、国会会議録でも、「空き缶」については、「あき缶」、「空きカン」、「空缶」、「空罐」等の表記が見られる（ただしこの中には「あき缶処理対策協会」といった固有名詞も含まれる）。表記に関して付け加えると、旧かな、旧字体の問題もある。旧字体から新字体への切り替えは、ほぼ昭和29年頃には終了しているようだが、単語により大きな差があるようである。例えば「総理大臣」と「総理大臣」という表記が第1回国会から混在している。こうした単語を検索システムで対象にする場合、使われている表記すべてをOR指定で並べることで解決がつけられる。

ただし、OR指定で陥りやすいのは、検索ヒット数の上限を超えてしまうという事態である。この場合には、条件変更画面で期間を限定して検索し直すことがよいだろう。

4. 会議録に含まれないもの

さて、ここまで検索の実例を通して検索システムの有用性を見てきたが、ここからこのシステムの限界に触れてみたい。まず、この会議録には国会で実際に発話されたものであっても、様々な事情で含まれないものがあるという事実がある。つまり、検索システムで必ずしも国会での記録すべてがアクセスできるわけではない。こうした、会議録に現れないケースを、主なものに絞って、前田(1997)に従って簡単に見ていこう。

4.1 議長の許可を得ない発言（「不規則発言」やじ）

会議録の原則は、議長の許可を得た正規の発言を記載するというものである。よって、議長の許可のない発言は、議事進行に関係があるか、議長か委員長が特に取り上げない限り、原則的に会議録に記載されない。こうした発言のことを「不規則発言」という。簡単に言えば、やじである。議場が騒然とした場合について発言者限認不能、発言聴取不

可能な場合には、(議場騒然、聴取不能)といった決まった記述がなされる決まりになっている。例えば、第150回国会参議員議院運営委員会4号での、次の場合などがそうである：

○委員長(西田吉宏君) ただいまから議院運営委員会を開会をいたします。(議場騒然) 本会議における議案の趣旨説明聴取に関する件を……(議場騒然、聴取不能)といたします。

図13: 議場騒然、聴取不能の例：第150回国会参議員議院運営委員会4号における西田吉宏委員長の発言

4.2 秘密会記録

国会には「秘密会」と呼ばれるものが存在する。憲法57条第2項に、国会は特に秘密を要すると議決された部分は公表の必要がないと定められており、こうした会議の記録は存在しても、一般からのアクセスは不可能である。これは同じく憲法の定める国会の会議公開の原則(憲法57条1項)の例外になるわけで、¹¹そのために秘密会開催には出席議員の3分の2以上の多数での議決というハードルが設けられている(大山, 2003, 244-5)。本会議ではまだ秘密会は開催されていないが、委員会レベルでは、かなり古い記録になるが、1994年1月30日に閉会した第128回国会までに92回の開催が確認されている(前田, 1997, 3-4)。よく知られた秘密会としては、内容はすべて公表されているが、「ロッキード問題に関する調査特別委員会」がそうである。¹²

4.3 不穏当な発言

国会会議中に、不穏当と認められる発言があった場合、議長が取り消し権を行使して取り消すことが可能である。言論の府であり、議員がその発言を巡って民事および刑事上の責任を問われないことが憲法第51条で保証されている以上、ここで当然何が「不穏当」に相当するかということが問題となる。国会法と議院規則によって規定されている不穏当発言の例は、無礼な発言または他人の私生活に関する発言、敬称の不使用、書籍等の朗読、私語、妨害的発言、議題外の発言であるが、不穏当かどうかを実際に決定するのは、議長の判断である(前田, 1997, 4)。議長が取り消すわけであるから、議事録に記載されないはずであるが、これについては衆参で扱いが異なる。衆議院では禁止後も発言が継続されたことのみを示し、発言内容を記さないが、参議院では発言禁止後の発言を〔〕でくくって記載しているので、検索可能である。

不穏当な発言の例として、「無礼」とされて削除された吉田茂首相のいわゆるバカヤロウ発言、¹³青島幸男議員の男めかけ発言などがあるが、いずれも発端となった各発言その

¹¹ただしここで憲法が公開すべしとしているのは、本会議であり、委員会ではない(大山, 2003, 244)。

¹²帝国議会期の秘密会については、近年になって秘密会議速記録(貴族院、衆議院)が公開されている(鈴木(1995)、大山(2003, 247))。

¹³この事件の経緯については(若宮, 1994)が詳しい。

ものは会議録から削除されている。ただし、それに続く国会論戦の中で、その発言自体を取り上げる場合は、その限りではない。よって、会議録自体には、「ばかやろう」（「バカヤロウ」もあり）も「男めかけ」も登場する。

4.4 プレス・コード

かなり特殊な検索不能なケースとして、占領下のプレス・コードによる削除がある。プレス・コードは、正確には、連合軍最高司令官名で 1945 年に出された「日本に与える新聞準則」であるが、これが、日本におけるあらゆる刊行物に適用されるということで、国会会議録も含まれ、プレス・コードに抵触した文言が削除されているわけである。ここで削除された発言は、現在でもそのままの状態では会議録に残されており、元の発言を検索することはできない。この例として、1951 年 1 月 27 日衆議院本会議における川上貫一議員の発言を会議録から探ってみよう。「——」部分が削除部分である。

そもそもこの飛行基地は、すべて国民の税金でつくったものです。終戦処理費によつてつくられた飛行場であります。——。政府がいかに強弁されようとも、国民は——という事実をおおい隠すことはできません。（拍手）しかもこの終戦処理費は、十六年度において一千二十七億が計上されております。この莫大な金が、次から次へと飛行基地や軍事道路や潜水艦基地のために使われたらどうなるか。国民の税金はますます高くなり、至るところで田畑は取上げられ、農民は立ちのきを命ぜられ、——ことは明白である。

図 14: プレス・コードによる削除例：1951 年 1 月 27 日衆議院本会議における川上貫一議員の発言

こうした会議録の削除が極端になると、議員の演説すべてが削除、ということも実際に起きてしまう。¹⁴政治的妥当性はさておき、このことは国会会議録が、文字通り「国会発言のすべて、そのまま」ではなく、中には実際に発言されているながら何らかの理由で一般にはアクセスできない記録があるという事実を示している。このこと自体は、コーパスとしての利用価値を大きく損なうものではないが、ある種の語彙が「不穏当」なものとしてシステムティックに削除されている場合、語彙研究にとっては障害になるだろう。

5. 言語資料としての問題点

さて、ではこうした国会会議録を言語資料として使用する場合に問題となるのはどのような点だろうか。ここでは OCR による誤字・脱字、外字処理、そして整文の問題の 3 点に絞って解説してみたい。

¹⁴この例となるのが、1956 年 11 月 27 日の第 25 回国会衆議院本会議 7 号における、中曽根康弘議員の演説である（この事件の経緯については（若宮, 1994）を参照のこと。

5.1 OCRによる誤字・脱字

§2. や注8でも触れたように、第145回国会以降のデータは、入力した原稿がそのまま電子的に加工されて、インターネット上での公開に回されているが、それ以前のものについては、紙媒体（官報号外）バージョンをスキャナで読みとり、OCRで文字に変換している。しかし、OCRには当然誤認識が付き物であり、たとえ99%の認識率であっても、裏を返せばほぼ100字中1字は誤認識をする計算になる。これは国会会議録ほどの量のデータになると、膨大な誤字があることがあることになってしまう。¹⁵ もちろん、両院記録部でテキストデータには何重にも人の目を通してチェックをしているはずだが、それでもこうした誤字・脱字を修正し切れていない。当然、これは言語資料として使用する際の大きな障害となるのであり、実際注8で触れたような、言語調査にとって致命的なエラーも起きることになる。

こうした事態に、利用者である言語研究者はどう対処するべきだろう。まず言えることは、データとして使用する場合に、検索結果を鵜呑みにせずに、画像データに戻って確認する作業を行うことが重要だということである。画像データであれば、少なくとも原本である官報号外との異同は確認可能である。録音データが残っていない場合、官報号外が遡及しうる最古の資料となるのであるから、これと一致していれば一応確認しうる所までは確認したと言えるだろう。ただし、この方法でチェックできるのは、あくまでヒットしたデータについてのみである。もともと入力されているテキストデータ自体が誤っていたためにヒットしなかった場合については、どうすることもできない。また、上の検索例でもわかる通り、国会会議録ほどのデータになると、検索ヒット数が数百というのはざらであり、そうした場合には、この作業はそれだけでも大変な労力と時間を必要とすることになる。研究者の言語データ検索の労力と時間を節約するのがコーパス利用の大きな利点の一つだとすれば、これは検索システムをコーパスとして使用する際の大きな欠点となる点は否めない。

しかし、ここでは同時にまた、このシステムが誕生して日が浅い事実も考慮されるべきだろう。むしろ、注8でも述べたように、入力ミスを発見した場合には、利用者がまめにそれを報告することで、管理者側での修正を促し、結果的にデータの信頼性を高めるという方向で取り組むことを考えるべきだろう。これが、利用者ができることの2点目である。

利用者の貢献の3点目として挙げたいのは、入力ミスのサンプリング調査である。膨大なデータのすべてを個人でチェックするのは不可能に近いが、サンプリング法を用いれば、データに存在するミスの量のある程度の精度で推定することは可能である。データのミスの程度がわかれば、検索調査の精度の推定にも繋がるはずである。言語データとして検索システムが使われるのであれば、どこかでこうした作業は必要になるであろう。

¹⁵ ちなみに（河合, 2001）によれば、紙ベースの国会会議録は年間25,000ページ前後で1ページに約3,400字の文字情報が納められていると言う。

5.2 外字処理

問題点の2つ目は、外字処理の問題である。検索システムで公開されている国会会議録データの制作過程では、まず校閲 (§5.3 参照) を経て確定された会議録最終稿に基づいて、2つのファイルが作られている。一つは院内閲覧用のデータベースファイルであり、このファイルにさらに手が増えられ、財務省印刷局へ送稿され、官報号外として印刷されるに至る。このバージョンでは、議員名にも外字が使用されている。もう一つのバージョンは、ホームページ用のものであるが、ここで外字が使えないために、JIS 第1・2水準の字体に置き換える作業を行っている(浅水, 1997, 29-30)。つまり紙バージョンとインターネットバージョンとで、会議録の表記に差が出るわけである。国会会議録を使って表記調査をする場合、この点は十分留意する必要があるだろう。

5.3 整文の問題

最後の問題点は、整文(字句の整理)に関するものである。まず、会議録制作の流れをおおまかに把握しておこう。¹⁶ 会議録制作は、最初に会議場で発言を2人組で速記にとどめることから始まり、次にそれを速記記号から漢字かな交じり文に直す「反訳」と呼ばれる作業に入る。反訳では組になった2人速記者の記録が互いに、さらに録音と照合されて原稿に仕上がる。反訳された原稿は、次に校閲・編集部門に送られる。ここでは、速記・反訳段階での誤りのチェック、字句の整理(整文)、表記のチェック、発言者・時間(開議、休憩、散会)などの確認、そして会議録冒頭につけられる会議録情報、末尾につける議案・報告書などを校閲し、厳重なチェックを経た上で、印刷およびホームページ用に回されることになる。¹⁷ このうち言語研究者として関心を持たざるを得ないのは、字句の整理(整文)および表記のチェックであり、実際国会会議録を言語資料として用いる場合に、これが一番厄介な問題を作り出していると言っても過言ではない。

この整文作業についてまず述べなければいけないのは、その実情を伝える資料が、非常に少ないと言うことである。それには、字句の整理が会議録の改竄ではないかという誤解を招くおそれがあるという事情があるが(青山(1989, 44))、同時にこれは、整文化作業の詳細がわからないということで、会議録の使用者に一定の不安を与えていることも否定できないだろう。ともあれ、ここでは字句の整理についてその詳細を伝える数少ない文献である青山(1989)に従って、その実際を検討することにする。

青山(1989, 44)によると、1972年に参議院記録部に設置された整文委員会で、検討の結果字句の整理の規準として、以下の4点が設けられたという(元の文書には簡単な具体例が付されていたとのことだが、青山の論文ではこれらは紹介されていない)：

1. 言い誤り、脱落、不整などのため発言の趣旨を明確に文字に表現しがたいと判断さ

¹⁶ 会議録制作過程に関する記述は、青山(1989)、石堂(1990)、鈴木(1994)、浅水(1997)を参考にしている。ただし、これらの文献はいずれも国会情報システムが整備されつつある過程の中で執筆されている点は、注意を要する。現在はこれらの文献執筆時点より、はるかに院内の情報ネットワーク化が進行していると考えられるため、若干の変化があるかもしれないが、現在の状況をこれらと同レベルで詳説した文献は見あたらなかった。

¹⁷ ちなみに鈴木(1994, 74)によると、速記者1組が担当する10分間の速記の反訳作業に最低2時間、それらをつなげた6時間分程度の原稿を校閲し終わるのに最低3日要するという。

れる場合は、軽微なものに限り、社会通念上認められる表記の方法に従って当該部分の整理を行う。

2. 字句の整理は、一步誤ると改竄につながることを常に念頭に置き、必要最小限度において慎重に行い、軽微かどうか判断しがたい場合は発言者等に確認した上で行う。
3. 発言そのものが問題となるおそれがあると判断される場合は、字句の整理を行わない。
4. 会議録主任が発言の訂正の請求を受けた場合は、その訂正が軽微なものである時は会議録主任において処理し、訂正の内容が問題になるおそれがあると認められるときは、必要に応じ委員長の許可を求めものとする。

この4つの原則では、もちろん1が最も興味を引く部分である。速記の世界で「けば」と呼ばれるフィラー、ある種の終助詞・接続詞などが含まれるのは、理解できる。さらに字句の整理には、語順、口語的発音、さらに文脈の整理も含まれる。これは、実際の会議を聞いたままに書き取ったものと、会議録に実際に収録されたものを比較すると一目瞭然である。この貴重な例を、青山(1989, 43)から引いてみよう(図15・16)。図15が録音テープから漢字仮名混じりに起こされたもの、そして図16が会議録に収録された記録である。青山では明記されていないが、調べてみるとこれは、1972年10月9日開催の、第69回国会参議院建設委員会閉3号における田中一議員の発言であることがわかるので、会議録バージョンは検索システムから引用した(そのためかどうか「角柴」の表記が異なっている)。少々長いが、整文過程を窺える貴重な例であるので、すべて引用しておこう。

2つを比較してわかるのは、図16が「読み物」として、はるかに読みやすいということである。「発言そのまま(verbatim)」の前者では、可読性に大きな問題が生じてしまう。「速記は逐音で書き反訳は逐語でしろ」という速記者のスローガンは、会議録の目的を考えた場合、大いに納得できるものである。

また青山(1989)には挙げられていないが、会議録を見た限りでは、いわゆる助詞の脱落が極端に少ないことに気が付く。関西方言では特にそうだが、東京方言でも助詞の脱落は自然談話では頻繁に起きる現象である(松田, 2000)。これなどもおそらく字句の整理と言うことで、助詞が補われていることは想像に難くない。

いずれにしても、上のような形で反訳が行われているということから、少なくとも会議録が自然談話そのものではないことは動かし難いことだとわかるだろう。会議場発言に何らかの加工が加えられており、しかもその詳細があまり明らかにされていないわけである。これはコーパスとして利用するものにとっては、大きな不安材料となってしまうことは否めない。¹⁸

¹⁸それではこうした整文に関する内部文書を、情報公開の手続きを取って請求すればよいという考えがあるかもしれない。しかし、衆参両議院は立法府に属するのであり、2001年に施行された情報公開法(正式名称は「行政機関の保有する情報の公開に関する法律」)ではカバーされないのである。

今ここへ来るまでにですねあのお見えになるまでにですねどうもその—角栄君のあの著述がね—改造論という奴が相当山間僻地の土地までですね値段が上がっておると　そうして私はずーっとこの—休会中には歩くんですよよく　そうすこの—何がどうしてこんなことま—してたんだと言うと用地の問題もう値上がりになってことしの予算じゃとてもですね今までの公団から与えられた予算じゃとても買えませんこう言って嘆いてま—おるんです上がってないという道路局長言っているから実態を調べて下さいと言って調べる約束を調査をするしてその当委員会で報告する約束しましたからこれはいいんですがあなたは私はその—日本列島改造論についての論議はこれはいづれするんですがただ気になるのですいづれの機会にしますみんなのみんなの聞いているところでなるべく宣伝力あるような質問をしたいと思うんですよ　トリアエズワデスネゼーキンワーホージンイワユルショーヒントシテモッテオラナイデネアガリヲマッテイルトーキテキナエーカネモチガザイバツガザイバツテナーフドーサンガイシャノカウノウオゼンブコレワショーヒンデスガソーユーモノガトチヲカイシメテソーシテネアガリヲマッテイルンダトユーコレニタイシテワデスヨキョージッパーセントノカゼーラシヨートユーマーコトヲダイジンエーアナタワホーゲンコレモホーゲンノヒトツダナワシニイワセリヤ　ホンキデヤルキナンテナイトオモウンデスヨ

図 15: 国会発言を漢字仮名混じりで起こしたもの：1972 年 10 月 9 日第 69 回国会参議院建設委員会閉 3 号における田中一議員の発言

いまここへお見えになるまでに、どうも角榮君のあの著述、改造論という、相当山間僻地の土地まで値段が上がっている。そうして、私はずっとよく休会中に歩くんですよ。そうすると、どうしてこんなことしてんだと言うと、用地の問題、もう値上がりになって、ことしの予算じゃ、いままでの公団から与えられた予算じゃとても買えません、こう言って嘆いておるんです。ところが、御承知のように、どんどん山間僻地が値上がりしているんです。まあ道路局長は、市街化区域のほうが上がっているんだ、一般道路のほうが上がっているんだと言いますが、山間僻地を日本の大手の不動産業者がどのくらい買っているか。これは、手をつければそれがすぐ時価になる、時価、時価と言うけれども。先行投資として、商品として買うんですから、これはどうにもなりません、買って売るんですから。それで上がっているんです。上がってないと高橋道路局長言っているから、実態を調べてくださいと言って、調査をして当委員会で報告する約束をしましたから、これはいいんです。私は、日本列島改造論についての論議は、これはいずれかの機会にします。みんなの聞いているところで、なるべく宣伝力のあるような質問をしたいと思うものですから。ただ気になるのは、とりあえずは、税金は——いわゆる商品として持っておらないで値上がりを待っている投機的な法人、金持ち、財閥が——不動産会社の買うのは全部これは商品ですが、そういうものが土地を買い占めて、そうして値上がりを待っているんだという、これに対しては九〇%の課税をしようということを大臣、あなたは放言——これも放言の一つだな、私に言わせれば。本気でやる気なんてないと思うんですよ。

図 16: 実際に収録されたバージョン：1972年10月9日第69回国会参議院建設委員会閉3号における田中一議員の発言

6. 議会資料と言語研究

ここで、焦点を国会会議録自体から、会議録と言語研究の接点を考えてみよう。これまで見てきたような会議録を言語研究に使う場合、どのような資料として位置づけるべきだろう。まず§5.で見た通り、字句の整理を経た会議録は、言い誤り、冗長性、繰り返し、語順の乱れなどといった自然談話の持つ大きな特徴をいくつか失っている。また、国会中継を見る限り、原稿を質問・答弁で原稿や資料を読むことも多い。当然これを自然談話とは見なせない。結局のところ、口語的特徴も多分に残しているが、書き言葉と話し言葉の中間的性格を持つものと位置づけるのが正しいものと思われる。同様な意見は、Slembrouck (1992) によってイギリス国会の会議録について述べられているところである。

しかしながら同時に、国会会議録が戦後60年近くの日本語の記録であるという事実にも、公平な目を向ける必要がある。これほどまでに多量な、口語的性格を多分に残す、時間的幅を持った電子化現代語資料は、日本語コーパスの増えてきた昨今の事情を鑑みても、容易に見つかるものではない。¹⁹ 整文化作業を経ているという、言語研究者には欠点になる部分はあるとしても、むしろこのような会議録の性質を生かした活用法を考えるのが、正しいあり方であろう。

例えば、ケバ取り作業が行われている以上、フィラーの研究を会議録を使って行うのは適当ではない。同様に、助詞の脱落や、言い誤りの分析も不向きである。言うまでもなく、音声などは絶望的である。しかし、言語研究者以外にはあまり知られていない変異現象であれば、整文化の目を逃れて、変異が生き残っている可能性は高い。例えば井上(2003)は、「終わらせてください」、「しゃべらせてください」など、五段動詞の使役に「サセル」をつける「サ入れ言葉」に関する実態調査を国会会議録で行った卒論に触れている。ここでは、サ入れ言葉に「戦後まもなくの使用例が見つかった」井上(2003, 53)とされ、それ以上の詳細は明らかにされていないが、これはサ入れ言葉が整文化作業の目を逃れた（またはもともと対象になっていなかった）ということを示す例と言えるだろう。

助詞についても同じことが言える。脱落はチェックされるとしても、助詞の変異は脱落ばかりではない。南部(2003)は、吉田茂と鳩山由紀夫の会議録中の発話に基づき、「が・の交替」(僕が読んだ本～僕の読んだ本)の使用率の変化を分析し、Harada(1971)の仮説通りに、が・の交替規則が衰退していくことを示している(表2参照)。データは、まず鳩山について2003年5月16日から遡って2000年4月19日までの200回発言分約16万字を、そして吉田は1947年7月3日から1949年4月8日までの同量程度のデータを用いている。ちなみに、吉田は1878年生まれ、鳩山は1947年生まれで、いずれも東京出身者である。この研究は、現代日本語における現在進行中の文法変化を見事に捉えた点で、特筆に値するものだが、同時に国会会議録の持つ資料的価値も示している点を見逃してはならない。

¹⁹ もちろん文学作品であれば、『新潮文庫の百冊』やネット上ならば「青空文庫」(<http://www.aozora.gr.jp/>)など、複数のコーパスを挙げることができるが、これらは完全な書き言葉である。

	の	が ^s	合計
吉田茂	149 (69.6%)	65 (30.4%)	214 (100.0%)
鳩山由紀夫	57 (34.8%)	107 (65.2%)	164 (100.0%)
合計	206	172	378

$$X^2 = 45.53 (p < 0.001)$$

表 2: 国会会議録に見る「が・の交替規則」の変遷：吉田茂と鳩山由紀夫の発言より（南部(2003)に基づくデータ）

文法ということでは、帝国議会議事速記録であるが、当為表現の分析を行った諸星(1986)は、国会議事録を使用する上でも大いに参考になるだろう。また明治期における大分、和歌山、山形3県の県議会記録に見られる助動詞、助詞を調査し、地域差を検討した平沢(2000)の研究も同様である。いずれも議会議録の持つ資料的可能性を示唆している。こうして見てくると、文法項目だけでもかなりバラエティに富んだ分析が可能なのである。

語彙論的研究に目を移してみよう。国会会議録が、戦後60年近い日本語の移り変わりを反映している点に注目すれば、例えば橋本(2003)が新聞社説で発見した外来語増加パターンを検証することもできる。外来語に限らず、新語・流行語の発生と消失も、細かに捉えることができるだろう。

方言研究はどうだろう。一般に国会会議録には、方言的特徴はきわめて薄いことから、整文化作業の影響はもちろん否定できない。しかし、これは逆に「気づかれにくい方言」(沖, 1991)の分析の余地を残しているとも言える。全国から選出されてきている、国会議員の発話データであるから、この手の研究には、むしろお誂え向きとも言えるだろう。もちろん、南部(2003)のような研究を議員の属性と絡めて分析するのも可能である。議員の出身地や生年については、1990年までの国会であれば衆議院・参議院(編)(1990b, 1990a)で一括して、それ以降であれば「衆議院要覧」「参議院要覧」で調べが付けはらずである。最近では、議員自身のウェブサイトが開設されているので、より詳細な情報も入手できる。

また、議会議録ということでは、検索機能を付属させた会議録を公開しているのは、何も国会に限ったことではない。2004年2月16日現在、各都道府県議会レベルでは、栃木県と福井県を除いた全都道府県でそのようなウェブサイトを開設している。この両県にしても、検索機能が始動するのは、もはや時間の問題と見て良い。²⁰さらに、政令都市でも川崎市を除いた12都市(札幌、仙台、さいたま、千葉、横浜、名古屋、京都、大阪、神戸、広島、北九州、福岡)で同様なウェブサイトが稼働している。気づかれにくい方言の分析には、こうした地方議会の会議録を含めると、より効果的なことは論を待たない所である。²¹

²⁰両県とも、議会自体のウェブサイトは開設されている。地方議会の会議録については、「都道府県別地方議会議録検索サイト」(<http://www.asahi-net.or.jp/~gb4k-ktr/localgov.htm>)や「地方議会 Explorer」(<http://www2s.biglobe.ne.jp/~L-Fairly/chihouex.html>)などのサイトに詳しい情報がある。

²¹ただし、上で国会会議録について述べた事柄以上に、地方議会にはさらなる問題を指摘しなければならぬ

社会言語学と言うことでは、談話分析を考えてみよう。ファイラーの研究は無理であっても、議場のやりとりは、談話分析には興味深いデータを提供してくれることに変わりはない。管見では国内には研究例がまだ見あたらないが、イギリスでは1990年代から国会の会議録(Hansard)を用いた談話研究関係の論文が発表されてきている(Slembrouck, 1992; Shaw, 2000; Harris, 2001; Pérez de Ayala, 2001; Christie, 2004)。Slembrouck(1992)の先駆的研究は、まさにイギリス版整文化過程の分析と言え、議事録が基本的に書き言葉であるという事実を、録音と会議録を対照させたデータを示した上で例証している。また2000年以降の研究は、ポライトネスとジェンダーの関わりに集中して多様な分析を展開してきており、ビデオ中継資料や録音資料も使用されている。²²一定の限界を持ちつつも、この分野では少なくとも議会会議録が談話分析資料として定着しつつあることを伺わせるものと言えよう。²³こうした海外における研究からは、会議録を使った日本との対照研究といった見通しも開けて来るのである。

最後になるが、Slembrouck(1992)のように、整文化が加えられている国会会議録を逆手に取り、詳細な実体の不明な整文化そのもののありさま・移り変わりを検討する研究も、また面白い知見を提供するであろうし、それ以上に言語資料として会議録を使用するに当たり必要不可欠な作業である。例えば、参議院記録部によれば、ら抜き言葉については、一昔前までは修正がなされていたが、近年ら抜き言葉が一般化するにつれ、そのまま表記するようになったと言う。²⁴だとすれば、可能形の変異について年を追って会議録を調べ、いつから整文化方針が変化したのかが気になるところであるし、他の同様な変異現象についても同種の調査が必要となろう。同時期の可能形の実態調査と比較できるのであれば、国会会議録のより正確な社会言語学的性格も浮き彫りになる。これはまさに整文化と60年近いデータの蓄積を持つ、国会会議録ならではの研究となるだろう。²⁵

い。それは、会議録の品質の問題である。地方議会では、速記や反訳作業を外注しているケースも少なくない。国会は、衆参それぞれ記録部で速記者養成所を持ち、専門速記者を自前で養成しているのだから、その速記者の教育体制や会議録の品質管理には信頼が置けるものと思つて良い。これが、外注となるとプロとは言え、その品質にはばらつきが出てくることも考えられる(こうした問題については、鈴木(1997)が参考になる)。地方議会会議録を利用する場合には、この点を十分にふまえておくべきだろう。

²²国会においても、近年まだまだ数は少ないとは言え、その審議中継がインターネットを通じて視聴可能である。衆議院は「衆議院TV」(<http://www.shugiintv.go.jp/top.cfm>)、参議院は「参議院インターネット審議中継」(<http://www.webtv.sangiin.go.jp/webtv/index.php>)という別サイトだが、それぞれ衆議院、参議院からリンクが張られている。ただし過去のの中継録画が見られるビデオライブラリについては、衆議院が平成12(2000)年からの会議が視聴可能であるのに対して、参議院は過去1国会分のみと差がある。

²³築山(1997)、国立国会図書館議会官庁資料室ウェブサイト・関連リンク集(http://www.ndl.go.jp/horei_jp/links/link.htm)などの資料からもわかる通り、もはや検索機能付き議会会議録をインターネットで公開している国々は、少しも珍しくはない。それにしても談話分析研究が非常に盛んで、しかもThomasという検索システム(<http://thomas.loc.gov/>)を持つアメリカよりも、イギリスに議会会議録を資料とした談話分析研究が集中しているかのように見えるのは、興味深い。

²⁴筆者の質問に対する、参議院広報部を経由した筆者宛の返信(電子メール、2003年10月27日)。

²⁵議会資料の言語研究利用の特殊例としては、カナダ議会のHansard Corpusがある。これは、Linguistic Data Consortium(LDC)から2000年より入手可能なコーパスであるが、カナダ議会の1970年代半ばから1980年代終わりくらいまでの英/仏両語による議事録である。IBMとBellcore(ベル電話会社の研究機関)がデータを提供しているものであるが、英語とフランス語の対照になっている、parallel corpus(対訳コーパス)の代表例であることから、機械翻訳研究での使用が非常に盛んである。ただし、これは議会資料としての性格よりは、対訳コーパスとしての性格によるものである。

7. おわりに

繰り返しになるが、国会会議録は、立法府の最高機関における戦後 60 年近い期間にわたる、議員の発言を記録した稀有な資料である。言語研究を目的として公開されたものではないとは言え、無料でこれほどのデータにアクセスできる現実には、もっと日本語研究者に広く享受されるべきだろう。公開されてまだ日の浅い検索システムは、もっと言語研究者に知られる必要があるが、ほとんどの言語学者にはふだん余りなじみのない国会会議録には、あまりまとまった解説書もないようである。筆者の本稿の執筆動機も、まさにここにある。

ただしすべての言語資料がそうであるように、国会会議録も万能なものではない。整文を始めとする問題点があることは縷々述べてきた所である。それでも捨てるにはあまりに惜しいこのコーパスを使いこなすには、それらの問題点と長所を知り、自分の研究目的とつきあわせるしかないだろう。つまるところ、コーパスなどすべての資料同様、結局研究者が調査しようとする事象と、うまくマッチしたものであるかどうかの問題なのである。

今後国会会議録が信頼しうる言語資料として日本語研究者に使われていくためには、どうしても整文化作業の影響を明らかにする必要がある。同時に、異分野の研究者が検索システムを通じて様々な分析を施すことにより、その資料としての性質を浮かび上がらせることが必要である。こうした作業を通じて、OCR の入力ミスが修正されて行き、またインターフェースに対する利用者の要望（例えば現在の単純な検索法ばかりでなく、ワイルドカード指定や、さらには正規表現による検索の導入など）もサービスを提供する側である国会図書館なり、両院記録部に伝わることになる。こう考えてくると、将来の国会会議録検索システムを育てていくのは、利用者である我々自身でもあると言っても過言ではない。国会会議録検索システムを、真のインタラクティブなシステムにするのは、利用者である我々の責務でもあるだろう。

参考文献

- 青山學司 (1989). 会議録作成に携わって—字句の整理を中心として—. 『立法と調査』, No. 152, 42–47.
- 浅水信昭 (1997). 会議録作成の現況と近未来—衆議院記録部の場合—. 『議会政治研究』, No. 43, 22–33.
- Christie, Chris (2004). Politeness and the linguistic construction of gender in Parliament: An analysis of transgressions and apology behaviour. In *Sheffield Hallam Working Papers: Linguistic Politeness and Context*. <http://www.shu.ac.uk/wpw/politeness/christie.htm>.
- Harada, Shin'ichi (1971). Ga-No conversion and idiolectal variations in Japanese. *Annual Bulletin RILP*, 5, 99–113.

- Harris, Sandra (2001). Being politically impolite: Extending politeness theory to adversarial political discourse. *Discourse & Society*, **12** (4), 451–472.
- 橋本和佳 (2003). 戦後の新聞社説に見る外来語の増加. 『国語学会 2003 年度秋季大会予稿集』, pp. 95–102.
- 平沢啓 (2000). 明治期の県議会の言語—県議会議事録速記録を資料として—. 『きのくに国文』, No. 6, 11–32.
- 井上史雄 (2003). 『日本語は年速一キロで動く』. 講談社.
- 石堂尚 (1990). 会議録・「作成と保存」から「情報の発信」へ. 『立法と調査』, No. 161, 40–43.
- 河合美穂 (2001). 国際協力—21 世紀における国立国会図書館の役割. 日本情報の国際共有に関する研究 文部省科学研究費補助金基盤研究 A2 (課題番号 / 10044018、研究代表者 / 内藤衛亮) 研究成果報告書 平成 12 年度報告. <http://www.nii.ac.jp/publications/kaken/HTML%93%FA%96%7B%8F%EE%95%F12000/2000Kawai-J.htm>.
- 国立国会図書館 (2001). *National Diet Library Newsletter No. 119*. http://www.ndl.go.jp/en/publication/ndl_newsletter/119/191.html.
- 前田英昭 (1997). 国会の不穏当な発言と会議録の削除. 『議会政治研究』, No. 43, 1–13.
- Matsuda, Kenjiro (1993). Dissecting Analogical Leveling Quantitatively: The Case of the Innovative Potential Suffix in Tokyo Japanese. *Language Variation and Change*, **5**, 1–34.
- 松田謙次郎 (2000). 東京方言格助詞「を」の使用に関わる言語的諸要因の数量的検証. 『国語学』, **51** (1), 61–76.
- 諸星美智直 (1986). 国語資料としての帝国議会議事録速記録—当為表現の場合—. 『国学院大学大学院紀要 (文学研究科)』, **17**, 217–251.
- 南部智史 (2003). 「が」と「の」の交替の史的変化について. 2003 年度神戸松蔭女子学院大学大学院 英語学研究 XII レポート.
- 沖裕子 (1991). 気付かれにくい方言—アスペクト形式「～かける」の意味とその東西差. 『日本方言研究会第 53 回研究発表会発表原稿集』, pp. 21–33.
- 奥村卓石 (1997). 衆議院の情報化と今後の展開—立法情報・議事運営にシステム導入—. 『議会政治研究』, No. 41, 39–46.
- 大山礼子 (2003). 『国会学入門 (第 2 版)』. 東京:三省堂.

- Pérez de Ayala, Soledad (2001). FTAs and Erskine May: Conflicting needs? Politeness in Question Time. *Journal of Pragmatics*, **33**, 143–169.
- 参議院 50 年のあゆみ編集委員会 (編) (1998). 『参議院 50 年のあゆみ』. 財団法人参友会, 東京.
- Shaw, Silvia (2000). Language, gender and floor appointment in political debates. *Discourse & Society*, **11** (3), 401–418.
- Slembrouck, Stef (1992). The parliamentary Hansard ‘verbatim’ report: The written construction of spoken discourse. *Language and Literature*, **1**, 101–119.
- 総務庁 (1999). 『平成 11 年版 通信白書』. http://www.soumu.go.jp/joho_tsusin/policyreports/japanese/papers/99wp/99wp-0-index.html.
- 鈴木威男 (1997). 会議録外部委託の現状と課題. 『議会政治研究』, No. 43, 14–21.
- 鈴木威男 (1995). 『貴族院秘密会議事速記録』の公開. 『立法と調査』, No. 189, 79–81.
- 鈴木悌夫 (1994). 国会会議録作成の実務と課題. 『議会政治研究』, No. 30, 70–78.
- 衆議院記録部・参議院記録部 (1995). 『国会会議録用字例』. 大蔵省印刷局, 東京.
- 衆議院・参議院 (編) (1990a). 『議会制度百年史 参議院議員名鑑』. 東京: 大蔵省印刷局.
- 衆議院・参議院 (編) (1990b). 『議会制度百年史 衆議院議員名鑑』. 東京: 大蔵省印刷局.
- 築山信彦 (1997). 米議会における「議事録」及び「会議録」. 『議会政治研究』, No. 43, 34–47.
- 若宮啓文 (1994). 『忘れられない国会論戦—再軍備から公害問題まで』. 中央公論社.

Author's E-mail Address: kenjiro@shoin.ac.jp