# An Attempt to Measure the Quality of Questions in Question Time of the Australian Federal Parliament

Andrew Turpin
University of Melbourne
aturpin@unimelb.edu.au

## ABSTRACT

This paper uses standard information retrieval techniques to measure the quality of information exchange during Question Time in the Australian Federal Parliament's House of Representatives from 1998 to 2012. A search engine is used to index all answers to questions, and then runs each question as a query, recording the rank of the actual answer in the returned list of documents. Using this rank as a measure of quality, Question Time has deteriorated over the last decade. The main deterioration has been in information exchange in "Dorothy Dixer" questions. The corpus used for this study is available from the author's web page for further investigations.

## Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Miscellaneous

## General Terms

Application

## Keywords

Information retrieval, Question Time, Parliament, Search Engine

## 1. INTRODUCTION

Question Time in the House of Representatives in the Australian Federal Parliament begins at 2pm on every sitting day, and is a period of about one hour where members of the parliament can ask Ministers questions and receive an answer typically limited to several minutes in length. While a seemingly essential part of a functioning democracy, Question Time has come under attack in recent days [10], and after the last election [1], for lacking content and relevance. Mark Rodrigues' publication on parliamentary reform from 2010 sums it up as follows.

Question Time in the House of Representatives is often criticised for declining parliamentary standards and accountability. Oppositions are inclined to use partisan attacks disguised as questions to embarrass the government to which Ministers respond with lengthy answers of marginal relevance. Ministers often use Question Time to attack the opposition with pre-prepared statements in response to "Dorothy Dix" questions from their own side. Much of the theatre of Question Time is characterised by disorder contrived to make the evening news [9].

In this paper we use an off-the-shelf, open source information retrieval (IR) system to quantify whether questions and answers in Question Time contain less information and relevance in recent years than in past years. IR systems index a collection of documents, and then for any provided query, rank the documents in the collection against the query, with the top ranked document being deemed the most relevant to the query. In this study, all answers to all questions from Question Time in the period 1998 through to 2012 were indexed as a single document collection, then each question was run against the collection as a query. In theory, the answer provided by a parliamentarian to a question should be the top ranked document returned by the system; assuming, of course, that the question was answerable, and that they actually provided the answer. The hypothesis tested by this paper, therefore, is that the rank of answer documents to questions posed in 2012 are lower than the rank of answer documents in 1998. That is, questions and answers in 1998 contained more content that was decipherable by a search engine than 2012.

Hansard is a written daily record of everything that occurs in the parliament. The next section discusses the creation of the Question Time corpus from downloaded Hansard XML files: a non-trivial task, as it turned out. Then we present some results to address the hypothesis, and the final section of the paper discusses the results and limitations of this study.

## 2. CORPUS CREATION

All of the Hansard from 1998 to the current day can be downloaded from the Australian Parliament House Hansard Web site [4] in XML format. Unfortunately, from a computational point of view, the XML format changed mid-2011, and so different parsers are required to extract information for Hansards after March 2011 and before. This section describes how we extracted questions and answers from the

XML. The final collection is available for download in TREC format from the author's home page.

The extraction begins with an XSL script to extract the "Questions Without Notice" `<debate>` from each Hansard file – one per day of sitting – and then the first `<question>` and `<answer>` from each `<subdebate.1>` of that debate. Sometimes each subdebate contains multiple questions and answers (called supplementary questions, in parliamentary language), but these were excluded from extraction as they are difficult to automatically validate. Often the supplementary questions are ruled "out of order" and so no answer is supplied, or the XML is incorrectly formed so that supplementary questions and answers are difficult to align. Furthermore, supplementary questions often contain references back to the original question (for example, pronouns), and so do not stand alone as suitable queries to an automated retrieval system.

Once each `<question>` and `<answer>` are identified, extracting the actual text is complicated by *interjections*. These are interruptions by people other than the nominated talker, and can occur from anywhere in the parliamentary chamber, including from The Speaker (chairperson of the debate). Post March 2011, these interjections are marked as separate paragraph tags that contain a `<span>` tag with an attribute that contains the string "Interjecting". For example,

```
<p class="HPS-Normal">
    <span class="HPS-Normal">
        <a href="EOH" type="MemberInterjecting">
        <span class="HPS-MemberInterjecting">
            Mr Laming:
        </span>
        </a>
            Superclinics?
    </span>
</p>
```

Thus all paragraphs are extracted within a question or answer that do not contain a `<span>` with an attribute that contains "Interjecting".

A second form of interruption to the text can come form The Speaker, either to call for order, or to address specific questions about the procedure of the debate ("Points of Order"). In these cases, the `<span>` tag contains an attribute that contains the substring "Office". For example,

```
<p class="HPS-Normal">
  <span class="HPS-Normal">
    <span class="HPS-OfficeContinuation">
        The DEPUTY SPEAKER:
    </span>
    The member for Wannon has got one last chance.
  </span>
</p>
```

Again, these paragraphs are excluded from the question and answer. The exact XSL used is given in Appendix A.

Ideally all sections of the debate would be tagged correctly, but often many people are speaking at once, and the debate moves at a furious pace. In these circumstances it is inevitable that mistakes are made, with interjections falling into parts of speeches, interruptions by The Speaker being tagged as part of a speech, and so on. Any errors that we came across during development of the XSL were corrected in the original XML by hand. In total we made 24 edits to

files in 2012 and 1999. If those years are representative, then there are most likely 5 to 10 errors in the other years that remained uncorrected, and so may have filtered through to the final corpus.

In an effort to capture and correct errors in the final extracted question-answer pairs, and to tidy up the text for indexing, some post-processing of the XSL output was performed as described in Appendix A.

The raw XML contains other information that might be useful for analysis of Question Time, such as time stamps on speeches, names of speakers and interjectors, supplementary questions, and so on. For this study, the other piece of information extracted was the `<party>` of the questioner and responder. This allowed identification of "Dorothy Dixers": questions directed to members of the same political party to give them a platform for politicking. Occasionally a backbencher might ask a genuine question of their own front bench, but this was not distinguished in my analysis. We categorised the `<party>` fields into six parties as follows.

1. `AG`
2. `ALP`
3. `AUS`
4. `IND Ind Ind.`
5. `LP NATS NP Nats NaysWA NPActing CLP`
6. `UNKNOWN N/A`

Thus if the party of the questioner and responder fell into one of these categories, the question was deemed a Dorothy Dixer.

The final corpus contained 16310 question-answer pairs as outlined in Table 1 from 53 periods of parliament covering 941 sitting days.

The mean length of answers has remained about the same over all years: approximately 430 words or 190 words with stop words removed (linear regression, $R^2$ $p = 0.29$). The mean length of questions has decreased over time ($R^2 = 0.69, p = 0.0001$).

## 3. RANK OF ANSWERS

The most obvious experiment is simply to run each question as query and record the rank of the answer. For this purpose, we used the Zettair search engine (version 0.9.3) [12] with default Okapi BM25 [7, 8] parameters, stemming and stop list. The index created over all answers contained 16,310 documents, 54,971 distinct index terms, and 6,929,465 terms.

Figure 1 shows the median and inter-quartile ranges (IQR) of the ranks of the answer documents separated by year. The line indicates a simple linear regression. As can be seen, the median rank of the answer documents has significantly increased over the period of the corpus (t-test on slope, $p = 0.001$; $R^2 = 0.38$, $F = 8.0$, $p = 0.014$), and the IQR has not increased significantly (t-test on slope, $p = 0.19$; $R^2 = 0.16$, $F = 2.5$, $p = 0.14$).

Figure 2 shows the median rank and score for each answer document separated by year and with Dorothy Dixers separated out from the other questions. The score is the similarity score computed by the Zettair implementation of the Okapi BM25, and has no absolute meaning, but can be

**Table 1: Number of question-answer pairs for each year in the corpus (DD signifies Dorothy Dixer), and length of question and answer documents in words (after stopping).**

| Year | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. non-DD | 493 | 679 | 674 | 463 | 611 | 600 | 473 | 624 | 640 | 451 | 641 | 587 | 474 | 472 | 340 |
| No. DD | 494 | 674 | 668 | 461 | 595 | 567 | 476 | 626 | 634 | 443 | 641 | 576 | 466 | 403 | 281 |
| Total | 994 | 1363 | 1348 | 926 | 1209 | 1169 | 957 | 1265 | 1287 | 900 | 1284 | 1165 | 941 | 880 | 622 |
| **Answers** | | | | | | | | | | | | | | | |
| Mean Len. | 168 | 157 | 181 | 198 | 190 | 178 | 182 | 138 | 144 | 181 | 217 | 245 | 220 | 192 | 154 |
| Max Len. | 503 | 718 | 563 | 763 | 771 | 635 | 533 | 375 | 442 | 498 | 777 | 839 | 620 | 326 | 306 |
| **Questions** | | | | | | | | | | | | | | | |
| Mean Len. | 32 | 30 | 31 | 31 | 31 | 32 | 30 | 29 | 31 | 26 | 21 | 25 | 25 | 25 | 25 |
| Max Len. | 191 | 114 | 157 | 91 | 126 | 146 | 141 | 132 | 159 | 133 | 130 | 137 | 158 | 81 | 100 |
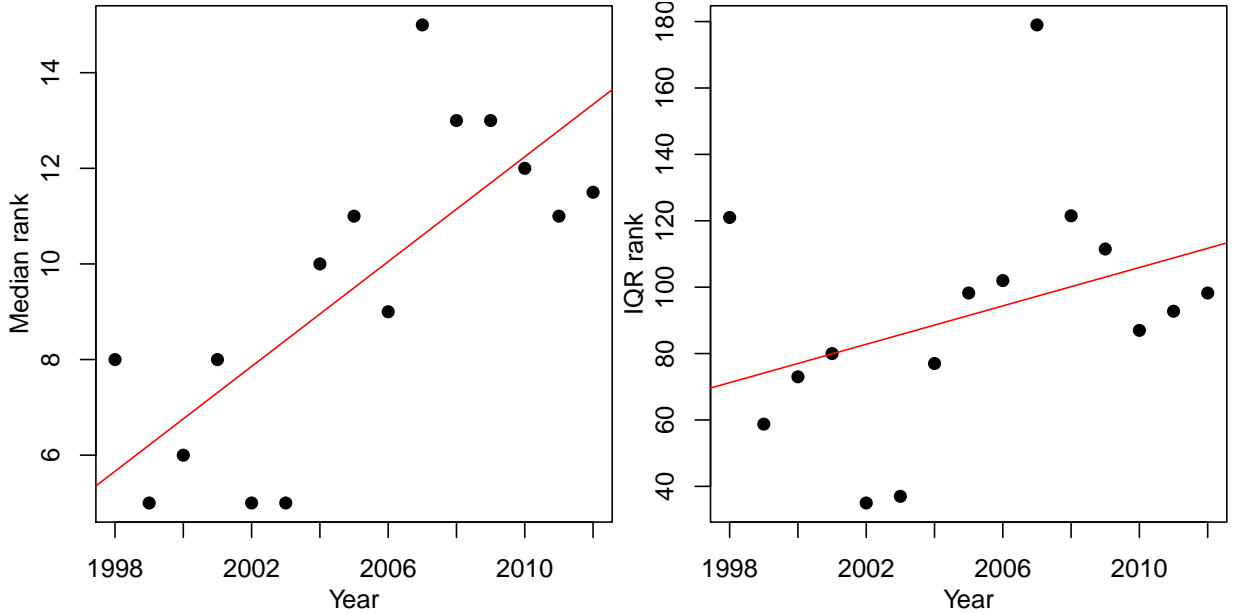


**Figure 1: Median and inter-quartile ranges (IQR) of the rank of the answer document for each question.**

compared relatively within a corpus. Again, lines shown are simple linear regression lines. Surprisingly, Dorothy Dixers show a significant increase in rank ($p = 0.002$) and decrease in score ($p < 0.001$) over time. Questions that were not Dorothy Dixers did not show a significant change in rank ($p = 0.454$), nor score ($p = 0.030$). This provides some evidence for Rodrigues' claim that Dorothy Dixers are increasingly used as a platform for making party political statements that have little relevance to the question.

### Questions in Writing

As an alternate experiment, the XSL scripts were modified to extract "QUESTIONS IN WRITING" from Hansard. The post processing scripts were also altered as the Questions in Writing had no interjections, and a slightly different introductory sentence. 7032 question-answer pairs were extracted from the corpus, covering years 2004 onwards as described in the first row of Table 2. As can be seen in the final three rows of the table, the IR engine ranks the answer to a question in position one generally over 50% of the time: far more success than on questions from Question Time, where the

IR engine only gets about 20% of matching answers in rank position one. This seems to match public opinion that questions and answers in Question Time have less informative content than general parliamentary debate.

## 4. NEAR DUPLICATES

It is conceivable that some answers are similar to another, and so it may be unfair to expect an IR system to find exactly the right answer to a specific question from a pool of documents that are very similar. In fact, 59 answer documents are one word, "No.", and 8 are "Yes". Browsing documents of 15 words or fewer reveals that very few of them would contain terms that would match a specific question using any TFxIDF style IR metric: many are simply wordy ways of saying "no". For example, "The answer to the honourable gentleman's question is no" or "In precise answer to the tail end of the question: no". There are 251 documents with 15 words or less.

Taking this one step further, if an answer is run as a query against the collection of answers, where will itself rank? If the answer is unambiguous according to the IR engine, then
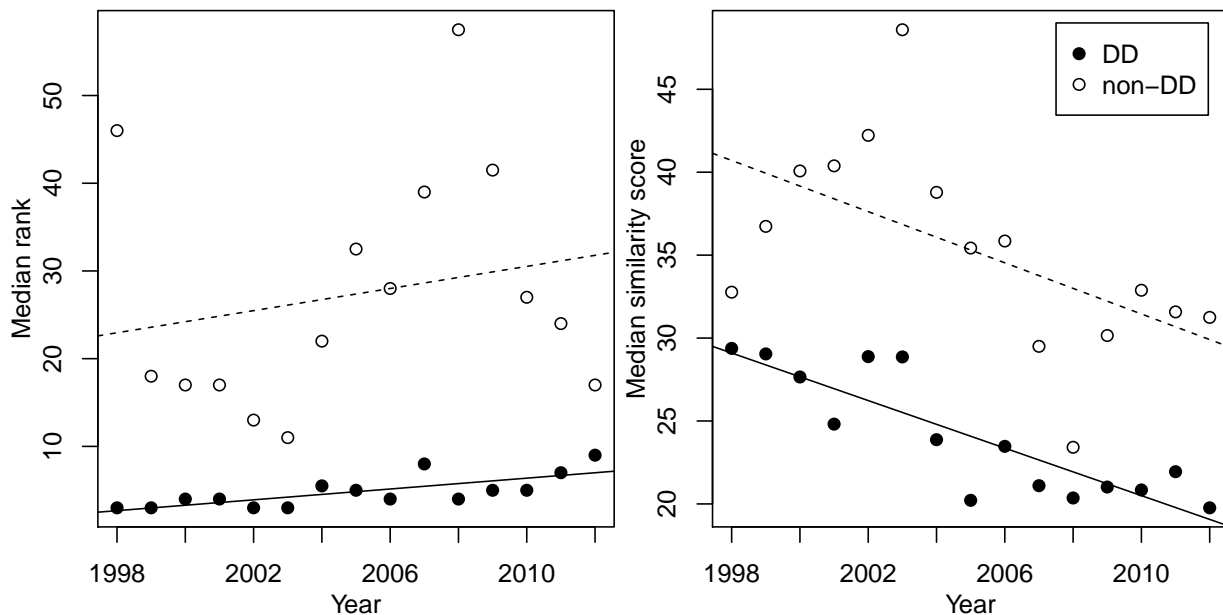
Figure 2: Median and inter-quartile ranges (IQR) of the rank of the answer document for each question separated by Dorothy Dixer (filled) and non-Dorothy Dixer (open).

Table 2: Statistics on the Questions in Writing corpus. The last three rows show a summary of ranks of the answer documents as ranked by Okapi BM25 when the question is posed as a query.

| Year | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|
| Number | 5 | 1841 | 1560 | 1205 | 303 | 588 | 279 | 496 | 330 |
| Med. Rank | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| IQR Rank | 1 | 12 | 16 | 10 | 3 | 4 | 3 | 5 | 9 |
| % at Rank 1 | 40% | 50% | 47% | 44% | 58% | 55% | 64% | 58% | 53% |

it should appear at rank position one. There are 91 documents where this is not the case. These documents have an average length of about 8 words, with the longest being 22 words.

A similar exercise can be undertaken for questions. Building a document collection of just the questions as documents, and then running each question as a query, 44 do not come up with themselves as the first ranked document, 39 of which are Dorothy Dixers. The majority of these questions are of the form: "Would the minister please update the House about...". Perhaps some more sophisticated phrase stopping technique might distill these questions to their essence, but this was not tried in this work.

If we remove all question-answer pairs where one or the other does not return itself as the top ranked document, or the answer is 15 words or less, this excludes 386 pairs of questions and answers. Not surprisingly, this made no difference to the trend in median ranks or IQRs shown in Figure 1.

## 5. PERSONNEL

Figure 3 shows the median and IQR of ranks for each sitting period (typically about twenty days). One could conjecture that the 39th and 40th parliaments (prior to November 2004) contained Question Times that were more informative than that of the 41st, 42nd and current 43rd parliament post November 2004. One could also argue that the Prime Minis-

ter may not be responsible, as Howard was PM both before and after November 2004. Perhaps the quality of Question Time as an information exchange is driven by the Opposition Leader, who is the chief questioner, or the Speaker of the House, who chairs the debate and monitors content.

Figure 5 shows the ranks of returned answers grouped by the two people over the lifetime of the corpus. It is not that illuminating: no one obviously stands out as a culprit. Kim Beazley is the only person in the corpus to be opposition leader twice, and there is a significant increase in ranks (Wilcoxon $p < 0.001$) between his two terms, perhaps indicating that he alone did not control the content of Question Time in that period.

Perhaps the only observation that can be drawn from this part of the analysis is that the 6 sessions from February 2002 to December 2003 (Prime Minister Howard, Opposition Leader Crean, Speaker Andrew) had low ranks relative to other periods where personnel remained unchanged.

## 6. ON MESSAGE

It is currently common practice for parliamentarians from the major parties to be issued with "talking points" for a day, and there is a strong emphasis to "stay on message". Perhaps, therefore, the answers given to questions, particularly Dorothy Dixers, are very similar, making it difficult for an IR engine to pick the correct answer to any given question. To explore this possibility, we compared the scores of
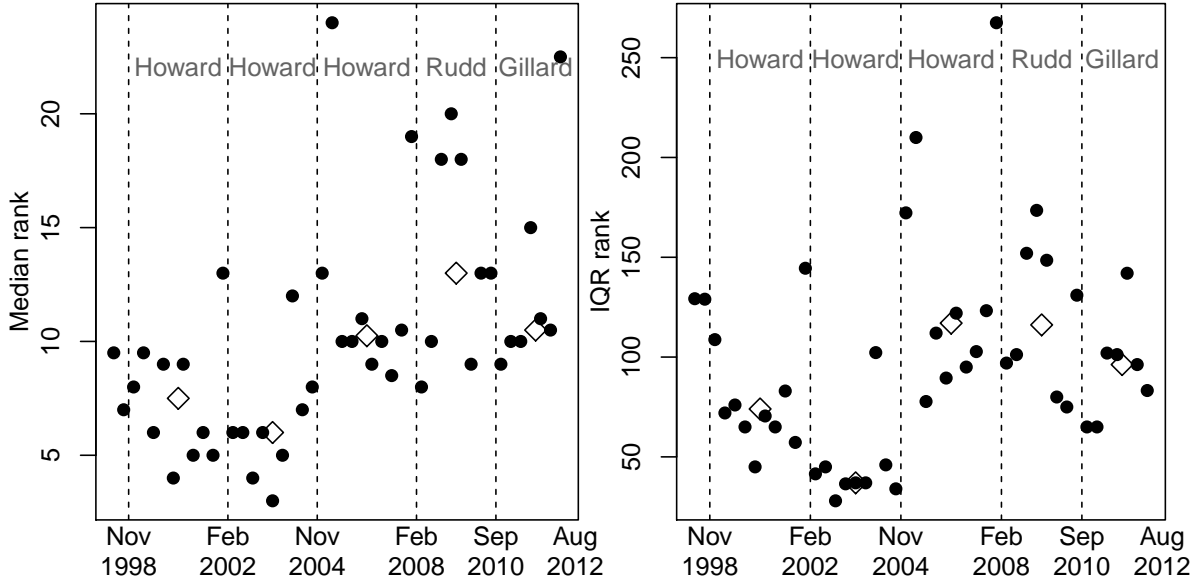
**Figure 3: Median and inter-quartile ranges (IQR) of the rank of the answer document for each question with one circle for each sitting period. Dotted lines show a change in parliament, with the names of the Prime Minister at the top of each parliament. The diamond gives the median value over the whole parliament.**
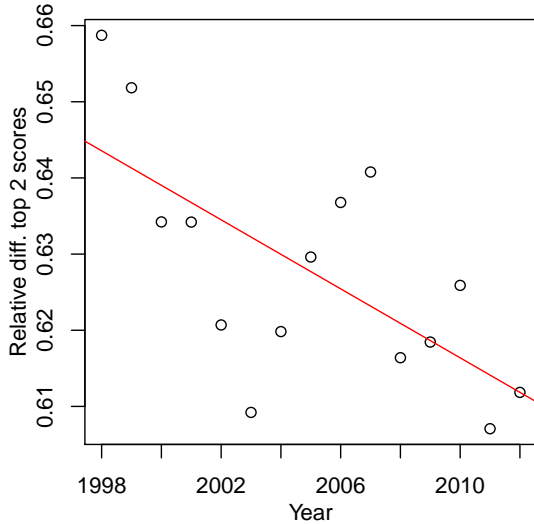


**Figure 4: Median relative difference in the Okapi scores of the top two answer documents returned when each answer is run as a query.**

the top two ranked documents when using each answer as a query.

For each answer document $a$, define $\Delta_a$ to be the difference between the top two scores of documents returned when $a$ is issued as query, divided by the top score. Thus if $\Delta_a$ is high, the second document is scored as very different from the first, while a low $\Delta_a$ indicates that the top two documents are similar, according to the IR engine's similarity measure. There is no correlation, between the median rank of answers as returned by the IR engine in a parliamentary

period and median $\Delta_a$ (Pearson, $p = 0.4$). But, there is a linear trend for median $\Delta_a$ decreasing from about 66% in 1998 to 61% in 2012 ($p = 0.007$). Figure 4 shows the data.

Splitting the data into Dorothy Dixers and non Dorothy Dixers reveals little. Both follow a similar pattern as the aggregated data, decreasing from bout 66% to 61%.

## 7. DISCUSSION

This paper attempts to objectively measure the quality of Question Time in the House of Representatives of the Australian Federal Parliament using state-of-the-art document ranking schemes that are the underlying technology of modern search engines. All answers given in Question Time in the period 1998 through to August 2012 are collected and indexed by the IR system, then each question is posed as a query to the system and the rank of its known answer recorded. The IR engine has more trouble finding answers to questions in recent years than in the past, leading to the conclusion that Question Time has reduced in quality over the last decade.

Throughout the experiments the Okapi BM25 metric [7, 8] with default settings was used as the ranking metric. We also generated Figure 1 using a Language Model scheme [6] with Dirichlet smoothing [13] using $\mu = 2000$ and $\mu = 500$, but there was no appreciable difference from the Okapi results, hence they are not reported in this paper. Perhaps there are other similarity metrics that are more suited to this retrieval task. Going beyond the standard approaches to include the meta-data identifying speakers, parties, and so on, or even exploiting the audio recordings of question time to determine speech patterns would be interesting areas for future study.

While the length of questions in words has decreased in recent years (see Table 1), the marked increase in median rank of answers occurs for 2003 where question length was still on a par with previous years. It isn't until 2007 that
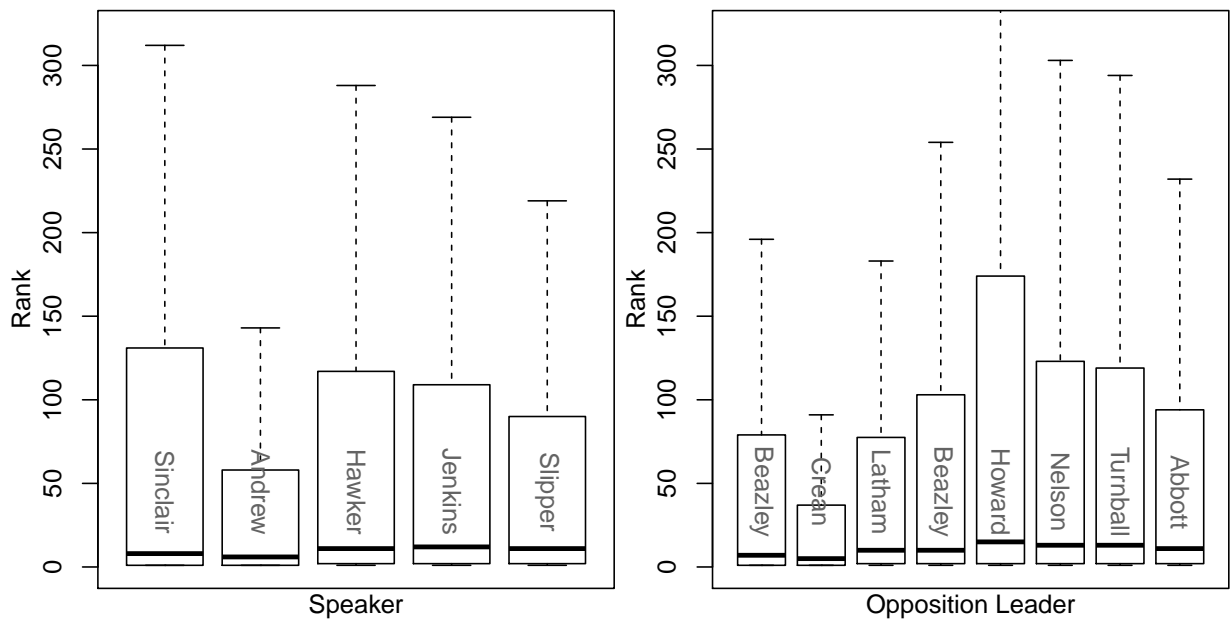
**Figure 5: Ranks of answer documents for each question grouped by Speaker (left) and Opposition Leader (right).**

questions become noticeably shorter. It is unlikely, therefore, that the number of words in questions or answers can account for the results.

Obviously using an IR system to measure quality of Question Time has its limitations. All of the IR approaches tested use stemming, stopping, treat questions and answers as bags-of-words, and ranks documents using a TFxIDF scheme. Such a scheme increases the similarity score between a document and a query if query words that occur frequently in a document (TF) and do not occur frequently in other documents (IDF). This approach is currently the best available technology for efficiently matching queries to documents [2]. It does not, however, incorporate the myriad of nuances that are present in Question Time question and answers. Politicians include doublespeak, metaphor, half-truths, red herrings, retorts to interjections, and any other manner of rhetorical and Vaudeville devices in their questions and answers. Also the "documents" used here are transcripts of the spoken word, so include social pleasantries (or ugliness) that it not usually found in written documents. The TFxIDF ranking used here captures none of this variation. As an aside, note that research into Spoken Word Retrieval [5] has focused predominately on the speech recognition side of the process, and not on the IR component. This seems like a fertile area for future research using this corpus.

In order to validate the use of TFxIDF rankings as a measure of quality of Question Time data, human judgements of question-answer pairs are required. Given all of the complications of the information exchange in Question Time, annotating this data set requires careful planning and instruction to annotators. We will explore this avenue in future work.

Any judgement of answer quality depends on what one sees as the aim of Question Time. If you subscribe to Mary Crawford's observation [3] that "For better or worse, it represents the translation of the political process into the everyday", then reducing the content of Question Time purely to an information matching problem is a futile exercise. If one would rather that an hour of an Australian politician's life was used in information sharing, rather than theatre, then the results in this paper directly apply. There is most likely a middle ground, where theatre does not obstruct information sharing, the text of which presents a major challenge to current automatic language analysis and information retrieval systems, and also a significant challenge to any manual annotation of the data.

Regardless of the theatrical content of Question Time, reduction of the repetition of information may be welcome. The results in Section 6 show that the information content of the answers to Dorothy Dixers are becoming more similar to each other as time goes on, and this could be the primary cause of the increased median rank of answer documents. Further investigation is required.

## 8. CONCLUSION

The quality of information exchanged in Question Time, as measured by an IR system, has decreased from 1998 to 2012. There has been a sharp deterioration since the end of 2003 that continues through to 2012. Surprisingly, the information contained in Dorothy Dixers has decreased over this period, while the information in other question-answer pairs has remained consistently poor when compared with questions in writing.

## 9. ACKNOWLEDGMENTS

# APPENDIX

## A. TECHNICAL DETAILS OF CORPUS CONSTRUCTION

Figure 6 shows the final XSL used. Prior to March 2011, interjections were tagged separately as `<interjection>` tags, simplifying the XSL considerably. Figure 7 shows the relevant XSL.

Post processing of the extracted data proceeded as follows.

1. TREC-style [11] tags were inserted to demarcate questions and answers.

2. The "name (electorate) (time):" prefix of all questions and answers were removed.

3. The first sentence was removed from all questions that contained more than one sentence. Typically the first sentence is of the form "My question is to ...". While informative, this project is interested in the content of the question, and not the intended recipient.

4. The phrase "Members interjecting" and its many variants was removed from all questions and answers using the `sed` script shown in Figure 8.

5. During a question, if a member is interrupted, their name appears in the question after their continuation in upper case and followed by a colon. This was removed from all questions with the following `sed` script.

```
s/Mr [A-Z]*://g
s/Mrs [A-Z]*://g
s/Dr [A-Z]*://g
```

6. The phrase "(Time expired)", with possible punctuation between the 'd' and '(', was removed from all questions and answers.

7. Extended ASCII characters were mapped to standard ASCII characters in both questions and answers. In particular: the Euro and Pound symbols were mapped to `EURO` and `GBP` respectively; various length hyphens were all mapped to a simple minus sign; and various quotation marks were mapped to simple single quotes. Double quotes `"` were also substituted with a single `'` to prevent phrase processing by the IR engine.

## B. REFERENCES

[1] AAP. Squabble breaks out over parliamentary reform. *Sydney Morning Herald*, Sept. 2010.

[2] T. G. Armstrong, A. Moffat, W. Webber, and J. Zobel. Improvements that don't add up: ad-hoc retrieval results since 1998. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM '09, pages 601–610, New York, NY, USA, 2009. ACM.

[3] M. Crawford. Question Time: don't change the contest we want to watch. `http://theconversation.edu.au`, Sept. 2012.

[4] `http://www.aph.gov.au/Parliamentary_Business/Hansard`.

[5] M. Larson and G. J. F. Jones. Spoken content retrieval: A survey of techniques and technologies. *Foundations and Trends in Information Retrieval*, 5(4-5):235–422, 2011.

[6] J. M. Ponte and W. B. Croft. A language modelling approach to information retrieval. In W. B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson, and J. Zobel, editors, *Proc. ACM SIGIR*, pages 275–281, Melbourne, Australia, 1998.

[7] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society For Information Science*, 27:129–146, 1976.

[8] S. E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In W. B. Croft and C. J. van Rijsbergen, editors, *Proc. ACM SIGIR*, pages 232–241, Dublin, Ireland, 1994.

[9] M. Rodrigues. Parliamentary reform. `http://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook43p/parliamentaryreform`, 2010.

[10] M. Turnball. Republican virtues: truth, leadership and responsibility. *The George Winterton Lecture 2012 delivered at The University of Western Australia*, Sept. 2012.

[11] E. M. Voorhees and D. K. Harman. *TREC: experiment and evaluation in information retrieval.* MIT Press, 2005.

[12] `http://www.seg.rmit.edu.au/zettair`, 2012.

[13] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '01, pages 334–342, New York, NY, USA, 2001. ACM.

```
<xsl:for-each select="//debate[debateinfo/title='QUESTIONS WITHOUT NOTICE']/subdebate.1[question and answer]">
    <xsl:for-each select="question">
        <xsl:if test="count(preceding-sibling::question) = 0">
            <xsl:for-each select="node()//span[@class='HPS-MemberQuestion']">
                <xsl:for-each select="ancestor::body/p">
                    <xsl:if test="not(.//span[contains(@class, 'Interjecting') or contains(@class, 'Office')])">
                        <xsl:value-of select="."/>
                    </xsl:if>
                </xsl:for-each>
            </xsl:for-each>
        </xsl:if>
    </xsl:for-each>
</xsl:for-each>
```

**Figure 6: The XSL used for extracting the first question in each subdebate from Hansard XML published after March 2011. Answers are extracted similarly, replacing "question" with "answer" in line 3.**

```
<xsl:for-each select="//debate[debateinfo/title='QUESTIONS WITHOUT NOTICE']/subdebate.1[question and answer]">
    <xsl:for-each select="question">
        <xsl:if test="count(preceding-sibling::question) = 0">
            <xsl:for-each select=".//para">
                <xsl:if test="not(./ancestor::interjection)">
                    <xsl:value-of select="."/>
                </xsl:if>
            </xsl:for-each>
        </xsl:if>
    </xsl:for-each>
</xsl:for-each>
```

**Figure 7: The XSL used for extracting the first question in each subdebate from Hansard XML published before March 2011. Answers are extracted similarly, replacing "question" with "answer" in lines 2 and 3.**

```
s/Opposition members interjecting//g
s/Government members interjecting//g
s/Honourable members interjecting//g
s/Opposition member interjecting//g
s/Government member interjecting//g
s/Honourable member interjecting//g
s/A government member interjecting//g
s/An opposition member interjecting//g
s/Mr [a-zA-Z'"'"']* interjecting//g
s/Mrs [a-zA-Z'"'"']* interjecting//g
s/Ms [a-zA-Z'"'"']* interjecting//g
s/Dr [a-zA-Z'"'"']* interjecting//g
s/Mr [a-zA-Z]* [a-zA-Z'"'"']* interjecting//g
s/Mrs [a-zA-Z]* [a-zA-Z'"'"']* interjecting//g
s/Ms [a-zA-Z]* [a-zA-Z'"'"']* interjecting//g
s/Dr [a-zA-Z]* [a-zA-Z'"'"']* interjecting//g
s/The member for [A-Za-z]* is interjecting//g
s/The member for [A-Za-z]* [A-Za-z]* is interjecting//g
s/The member for [A-Za-z]*-[A-Za-z]*is interjecting//g
```

**Figure 8: Sed script used to remove interjections that were not tagged in the XML.**