The Automatic Transcription System for the Parliamentary debates at the House of Representatives

Mr. Najib EL KHADI
Secretary General at the House of Representatives

Kingdom of Morocco
Parliament
House of Representatives

Geneva, October 15th 2018
Synopsis

• Introduction

• Evolution of the transcription of parliamentary debates in the House of Representatives

• Description of the automatic transcription system

• Conclusion
Importance of reporting of parliamentary debates

The sessions taking place at the House of Representatives are public, and the full minutes of the debates are published at the Parliament’s Official Bulletin

Article 68 of the 2011 Constitution (2011)

The list of the newly elected deputies is indicated in the full minutes of the debates and published on the Official Bulletin.

Article 21 of the Internal Rules of the House of Representatives
The Importance of the parliamentary debate minutes

A detailed report is drafted after each parliamentary committee meeting ... Members of parliament (MP’s) have the right to consult on the spot the original minutes of parliamentary committee meetings

**Article 110 of the Internal Regulations**
The list of deputies absent without justification is indicated in the minutes of parliamentary committee meetings

**Article 111 of the Internal Regulations**
Deductions will be made from the salaries of deputies absent without justification during the plenary sessions or during the meetings of the committees ... The list of the sanctioned deputies is published in the Official Bulletin of the Parliament

Article 106 of the Internal Regulations

Written questions and answers by the Government are published in the Official Bulletin of Parliament

Article 277 of the Rules of Procedure

Are also published: the composition of the House, the motions on the legislation, the reports of the constitutional bodies of governance ...
The Parliament’s Official Bulletin

Since 2014 the Official Bulletin of Parliament is available only in electronic version on the portal of the Chamber. Previously, it was published in paper form by the General Secretariat of the Government.
The Evolution of the Parliamentary Debate transcription at the House of Representatives

**Prior to 2006**

- **Plenary Session**
  - The session is recorded as an audio tape
  - A 2 hours 30 minutes questioning session requires approximately 3 tapes of 45 minutes each.

- **The Transcriber**
  - Listening to the tape and transcription manuscript
  - On average it requires 1 day per transcriptionist
  - About 3 days to transcribe the entire session.

- **The Secretary**
  - Enter text on computer and print out
  - 3 additional days needed

- **The Corrector**
  - Correction of misspellings
  - Replaying the whole session for correction of other types of errors or omissions
  - 3 additional days

**Prior to 2006**

- 10 days
The Evolution of the Parliamentary Debate transcription at the House of Representatives

• **Prior to 2006**

  • An additional day for the layout format compatible with the Official Bulletin

  • In this period, the General Secretariat of Government who printed the Official Bulletin

  • At the beginning of 2006 a considerable delay was registered and corresponded to almost three years of cumulative unpublished debates
The Evolution of the Parliamentary Debate transcription at the House of Representatives

• Between 2006 and 2017

• Setting up a manual transcription system
• A transcript room was made available to the service sessions with Transcriptionists (8) and computers
• Each transcriber is equipped with a headset and a pedal to control the audio flow
• The room was equipped with a television which transmits the plenary session live.
The Evolution of the Parliamentary Debate transcription at the House of Representatives

• **Between 2006 and 2017**
  
  • A server at the control room records the audio of the session
  • After 15 minutes of the start of the session, the 5 minute audio pieces are transferred to the transcribers
  • Another server assigns the piece to the transcriber that is available
  • Each transcriptionist listens to the piece and transcribed into text format
  • The text is transferred to the corrector, then the publisher that controls and checks the entire text
  • A final step is the layout of the report before transmission for publication in the Official Bulletin
The Evolution of the Parliamentary Debate transcription at the House of Representatives

- **Between 2006 and 2017**
  - On average, the transcript of a session of oral questions requires 3 days
  - The verification and correction by the controller requires two days
  - The layout requires an extra day
  - In total, it will be necessary to wait 6 days before being able to publish the report
  - Nevertheless, it was possible to make up for the delay of the 3 years not transcribed
  - As soon as they were produced, the minutes of the plenary meetings were published immediately on the portal of the House of Representatives
The Evolution of the Parliamentary Debate transcription at the House of Representatives

- **E-Parliament project**

  - Effectiveness, efficiency and good governance
  - Valuation of Parliamentary Work
  - Consolidating the capacities of the House of Representatives
  - Openness and Transparency

  **Strategic Orientations**
  - Zero paper
  - Virtual Office
  - Archive Valuation
  - Communication with the citizens and with the partners
• E-Parliament Project

- Citizens
- Constitutional Institutions
- Government
- House of Councillors

---

- IT Infrastructure
  - IT Security
    - Process Management
    - GOL
    - Automatic Transcription
    - Portal and Social Networks
    - Archive management System
    - IT applications
The Evolution of the Parliamentary Debate transcription at the House of Representatives

• **E-Parlement Project**

  • A reflection was made internally to examine the possibility of setting up an automatic transcription system
  • This is the first time in Morocco that such a system has been designed to support the Arabic dialectal language as it is used by parliamentarians
  • The Chamber made the decision to explore this path, and a call for tenders was issued
  • Only one Moroccan company responded and was selected
The Evolution of the Parliamentary Debate transcription at the House of Representatives

- **E-Parlement Project**

- The proposed system is based on artificial intelligence technology
- Based on a corpus that has been prepared and contains hundreds of hours of parliamentary debate, the system learning the vocabulary and pronunciation associated
- Data formatting took more than 6 months
- The system became operational after almost a year
- The recognition rate is about 80%, it can be increased or decreased depending on the circumstances
The Evolution of the Parliamentary Debate transcription at the House of Representatives

• **Since 2017**

• The system of automatic transcription of parliamentary debates is actually in operation

• After 10 minutes from the beginning of the session, the automatically transcribed pieces are transferred to the correctors (former transcribers)

• The report is about two days before publication

• The year 2018 marked the beginning of the use of the system for the transcription of committee meetings
Description of automatic Transcription system

- A transcription team receives the audio recording, broken down into pieces

- The transcription team develops the transcription P.V.s for each piece received. The group is provided with a web word processing application, a pedal, and an audio headset.

- Correctors and editors re-read and correct the debates

- Audio files and texts are stocked in application servers and database for subsequent research
Description of automatic Transcription system

- Creation of parliamentary sessions at the beginning of each year.
- The system administrator shall be tasked to configure the session-related information: the legislature, the year, the session, the starting and ending dates.
- These information shall have to be entered once at the beginning of each parliamentary session.
At the beginning of each parliamentary session the administrator creates the day session.

- S/he enters the information geared to the session: agenda, chairperson, date ... etc.

- Audio files are automatically broken down into a parameterable duration.

- Files may be manually or automatically attributed to the session.
• The transcription module is made of: a media player,

• an information recap on the session being transcribed,

• a text editor with the main features and functionalities.

• The transcriber, provided with an audio headset and a pedal, works out a P.V. corresponding to the audio piece received, then forwards it to the corrector.
Description of automatic Transcription system

• The corrector receives the P.V. of each transcriber on a different tab. Piece 1 goes to tab 1, and piece 2 to tab 2 … etc.
• S/he corrects the text, deletes the redundancies and transfers the pieces towards the global content.

• The global content corresponds to the total transcription of the session.

• The audio file relating to the session can be listened to, and the session Word file may be generated.
Once the session P.Vs published, multi-criteria research can be made on the data base.

The research makes possible the retrieval the text and audio files by using criteria such as the date, the year, key words and the session subject.
Description of automatic Transcription system

- **Fin 2017: Extension aux commissions**

<table>
<thead>
<tr>
<th>Supported languages</th>
<th>Transcription is made for both Classical and dialectal Arabic.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPEECH RECOGNITION RATE</strong></td>
<td>The Speech Recognition Rate averages 80% on classical and dialectal Arabic. The said rate may increase or decrease when it comes to other languages like English or French.</td>
</tr>
<tr>
<td><strong>Deadlines for the preparation of verbatim reports (PVs)</strong></td>
<td>Verbatims are made available for the proof-readers by the server <strong>in real time for the plenary sessions, and 24 hours after for the commissions</strong></td>
</tr>
<tr>
<td><strong>Proposed System Capacity</strong></td>
<td>The system may cater for plenary sessions and the 9 parliamentary commissions with a preparation deadline for the aforementioned P.V.s</td>
</tr>
</tbody>
</table>
Description of automatic Transcription system

The automatic transcription makes use of the latest evolution in terms of artificial intelligence.

The language learning by the transcription core is made thanks to the already available hundreds of hours and texts.

A unique dictionary of words, a grammar, a list of words seconded by the different possible pronunciations for each word are used for learning purposes.

Linguistic data, like the language model, statistic models are developed, and acoustic models have been developed for each context.

All thes elements are subsequently used like basic data for the creation of an automatic transcription server which uses a neuron network.
A comparison between manual and automatic transcriptions

**Manual Transcription**

- Debate transcription is thoroughly manually undertaken by the transcribers
- The transcribers prepare the text which is corrected by proofreaders
- Transcription is made only for plenary sessions
- Application accessible only on the PC where it has been installed

**Automatic Transcription**

- Transcription undertaken by a transcription server without the intervention of a transcriber
- Proofreaders proceed to a direct text correction. Transcribers are assigned other tasks
- Transcription can be undertaken both for plenary sessions and for commissions
- Web application available on the whole local network of the House of Representatives
Conclusion

• The system can improve further by sensitizing parliamentarians to respect certain rules: only one speaker at a time, talking near the microphone, not mixing several languages ...

• The system will be used in committees to help board officers easily produce meeting minutes and comply with the rules of procedure

• The House of Representatives is ready to share its experience in this area with other parliaments