A French Corpus of Québec’s Parliamentary Debates

Pierre André Ménard, Centre de recherche informatique de Montréal
Desislava Aleksandrova, Université de Montréal
Québec’s Parliament Composition

- One of ten Canadian provinces and three territories
- Only province with French as only official language (one other bilingual province)
- Unicameral since 1963 (bicameral “Conseil Législatif” from 1867)

 Territory is divided into
  - 17 administratives regions
  - 125 provincial circonscriptions
  - 78 federal circonscriptions

 Workings
  - First-past-the-post voting for each provincial circonscriptions
  - 125 elected members (from 71 in 1867)
  - Prime minister is the leader of the majority party
  - Chair (Président de l’Assemblée nationale) is elected by a secret vote of the members
  - Legislature with each governing party change, session when priorities change
Source

Parliamentary proceedings available on the ANQ website (http://assnat.qc.ca)
- Daily debates of the Parliament
- Committees (economy, education, public finance, ...)

Transcriptions
- 1908 - 1963: reconstituted from notes, news, summaries.
- 1964 - 1989: curated to normalize language, style, ...
- 1989 - today: verbatim statements (minor spelling corrections)
  (Free translation)
  "That is the concrete reality, M. President. And, about the member, maybe, when she gets up... She had time to think. The discussions she had with M. Arsenault for her to continue to vote against the holding of a commission of inquiry there, where... What was the deal with him there? What was it, the agreement you had to refuse to Quebeckers to have a commission of inquiry..."
- French only (not translated)
- Preliminary version available same day
- Final version available within 2-3 days

Excerpt of a proceeding's web page from March 12, 1996 with annotated elements.
Structure and processing

Data in HTML format
- Daily transcriptions
- List of elected members

Manual retrieval
- Historical list of political parties (Wikipedia)

Processing steps
- Extract political party timeline for each member profile
- Speaker detection and disambiguation in transcriptions
- Transcription text extraction and enrichment
- Linguistic annotation with Trankit in CONLL-U
- Conversion to TEI Parlamint format

La Verge Noire
La Secrétaire adjointe
La présidente Mme Houda-Pepin
Le Président suppléant (M. Cousineau)
L'Orateur suppléant, M. Vautrin
Le Président (M. Ouimet, Marquette)
Son Honneur le lieutenant-gouverneur
M. LE PRESIDENT (M. Gauthier, Berthier)
M. L'Heureux (Gilbert)
M. L'Heureux (président du comité plénière)
Corpus statistics

| Legislature | 28 |
|-------------|----|
| Sessions    | 78 |
| Total period| 88 years |
| Spanning period | 114 years |
| Debates periods | 5,948 days |
| Members     | 1,310 |
Corpus statistics
# Data readiness overview

Data quality related to its use in a ML/NLP project

| Dimension               | Source format                                      | Parlamint format                                           |
|-------------------------|----------------------------------------------------|------------------------------------------------------------|
| Completeness            | (~) Some information is missing                    | (+) Improved, enriched with external data.                |
| Consistency             | (-) Time dependent consistency                     | (+) Improved, but not perfect.                             |
| Integrity/Completeness  | (-) No explicit integrity between data types       | (+) Linked data source.                                   |
| Integrity/Consistency   | (-) No explicit integrity between data types       | (~) Some unresolved links (parties, speakers)              |
| Timeliness              | (+) Same-day availability as offered               | (-) Conversion process not integrated                      |

* Evaluation from an external viewpoint. Results would be different if applied to internal systems.
Conclusion

- Conversion complexity linked with source evolution and segmentation
- Improved data readiness with conversion to Parlamint standard
- Accessibility : work in progress
- Future work
  - Improve information integrity
  - Language model comparison with various flavors of French
  - Sentiment analysis (non parliamentary expressions)
Questions?
Présidence de l'honorable J.-M.-P. Sauvé
La séance est ouverte à 8 h 30.
Prière.
M. l'Orateur:
À l'ordre, Messieurs! Que les portes soient ouvertes!
Rapports de comités:
M. Taché (Hull):
M. l'Orateur, j'ai l'honneur de présenter à la Chambre le deuxième rapport du comité permanent des comptes publics. Voici le rapport :
Le comité permanent des comptes publics fait rapport que M. Valentine qui avait été assigné à comparaître audit comité, a fait défaut de ce faire.
Le comité vous prie respectueusement, en conséquence, d'adopter toutes les mesures nécessaires pour que l'ordre donné à M. Valentine soit respecté.
(Applaudissements à droite)
M. Pelletier (Lotbinière):
Relativement à cet incident, je tiens, à titre de membre du comité des comptes publics, à protester contre l'outrecuidance de certains témoins.
À l’ordre des Messieurs, que les portes soient ouvertes!