

APPENDIX II

Construction of the dataset

This appendix presents the origin of the raw data used in the master thesis. It also describes extensively how the dataset used for the empirical analysis was built.

The final dataset contains information about the French *député*'s (members of the lower house) participation over the XIIth, XIVth and XVth (up to 21 June 2018) legislatures. More precisely, all MPs who were reelected at least once between 2007 and 2018 is in this dataset. The dataset contains 1,356,119 observations, organized in a panel data structure with 423 unique MPs (i dimension) and 14,489 meetings (t dimension). The dataset contains 134 variables.

Contents

1	License and origin of the raw data	101
1.1	License	101
1.2	Origin of the data	101
1.3	<i>Regards Citoyens</i>	101
1.4	Links	102
I	Extracting information from the main database	103
2	Overall presentation of the dataset	103
2.1	Map of the dataset	103
2.2	Dataset construction	104
3	Construction of the parlementaires tables	106
3.1	The <i>baseline</i> file	106
3.1.1	Step one: cleaning the file and adding some basic information	106
3.1.2	More editing with STATA	107
3.2	The <i>career</i> files	108

3.2.1	Past MP offices	108
3.2.2	Other offices: looking for multiple office holders	109
3.2.3	Past other offices: capturing the political career	115
3.2.4	District information	118
4	Construction of the meetings tables	120
4.1	Floor (public) meetings	120
4.1.1	Extracting the relevant information from the database	120
4.1.2	Distinguishing legislative and control meetings	120
4.2	Committee meetings	121
4.2.1	A word on the National Assembly's agencies	121
4.2.2	Extracting and editing the tables	123
5	Construction of the complementary tables	123
5.1	Amendments	123
5.2	Rapporteurs	124
5.3	Questions	125
6	Additional information	125
6.1	Seats in the hemicycle	125
6.1.1	Motivation and variables	125
6.1.2	Construction of the variables	126
6.2	More data	127
II	Conversion into a panel dataset	128
7	Reconstructing the timeline	128
8	Incorporating attendance and participation data	129
9	Further modifications	130

List of Tables

1	Political group acronyms	107
2	Generated variables based on the previous offices	109
3	Summary of the jargon	110
4	Error coding explanations	114
5	Generated variables based on the other offices	115
6	Error coding explanations	116
7	Summary of the variables for the past offices. i indicates the tier.	117
8	Generated variables for the district	118
9	Map of the parlementaires files	119
10	Ids of the permanent committees	122

List of Figures

1	Map of the main database	103
2	Visual representation of the versions	105
3	Sketch of the administrative organization of Paris, Lyon and Marseille	112
4	Seats in the hemicycle	127

1 License and origin of the raw data

1.1 License

All the data we use comes from SQL dumps of the website *NosDéputés.fr*, an initiative of *Regards Citoyens* (hereafter RC) based on the data provided by the French National Assembly and the Official Journal of the French Republic. The data is published under ODbL (Open Database License) license. This license allows anybody to publicly exploit databases, for commercial purpose or not, to the extent that the database and its modifications remain under this license. The SQL dumps are available at the address <https://data.regardscitoyens.org/nosdeputes.fr/>

1.2 Origin of the data

All data comes from the National Assembly website and the *Journal officiel*. These are public information, updated every 8 hours. The National Assembly website states that *les débats et les documents parlementaires [sont publics et] peuvent donc être reproduits [sauf] à des fins commerciales ou publicitaires*¹

Attendance in committee is recorded and publicly available, contrary to the interventions in the hemicycle. As such, the presence of an MP during a floor meeting is recorded through its oral interventions as transcribed in the reports of the meetings. However, it is likely that an MP attended a given floor meeting without actually speaking during this meeting. As such, the data recollection underestimates by construction the real attendance of the MPs to floor meetings. Also, in the case of floor meetings, a difference is made between interventions and participation, the latter being the fact that the MP spoke more than 20 words. Short interventions are indeed seen as unable to discuss in depth a given topic.

1.3 *Regards Citoyens*

Regards Citoyens is an association of volunteer citizens who want to offer a simplified access to the functioning of the French institutions based on the publicly available information. RC was founded in July 2009 and now gathers around 40 members including 7 administrators. Since June 2010, RC is an association as defined by the law of 1901. It is not affiliated to any political party and the website is entirely run by volunteers. We are not affiliated to them.

¹Which could be translated as follows: "debates and parliamentary documents [are public] and can be reproduced, [except] for commercial or advertising ends."

The initiative of Regards Citoyens is inspired by similar international initiatives such as TheyWorkForYou in Great-Britain, Parlorama.eu and VoteWatch.eu in Europe or the work of the Sunlight Foundation in the United States. In the case of France, similar work is proposed by Olivier de Solan with the website Mon-Depute.fr or initiatives such as *Mémoire Politique* or Candidats.fr.

1.4 Links

NosDéputés.fr: <http://www.nosdeputes.fr>

Regards Citoyens: <http://www.regardscitoyens.org>

French National Assembly: <http://www.assemblee-nationale.fr>

ODbL License: https://en.wikipedia.org/wiki/Open_Database_License

of all the MPs interventions, the attached meeting and the number of words for each intervention. Amendement collects all amendments tabled by the MPs and the Government. For the XIIIth legislature, only hemicycle amendments were collected. From the XIVth legislature onwards, all amendments (both committee and hemicycle) are collected. Presence indicates if there exists a proof of presence for a given MP at a given meeting. Various additional tables are combined with these "major" tables in the process of recovering information (e.g. for the amendments tabled or the rapporteur status).

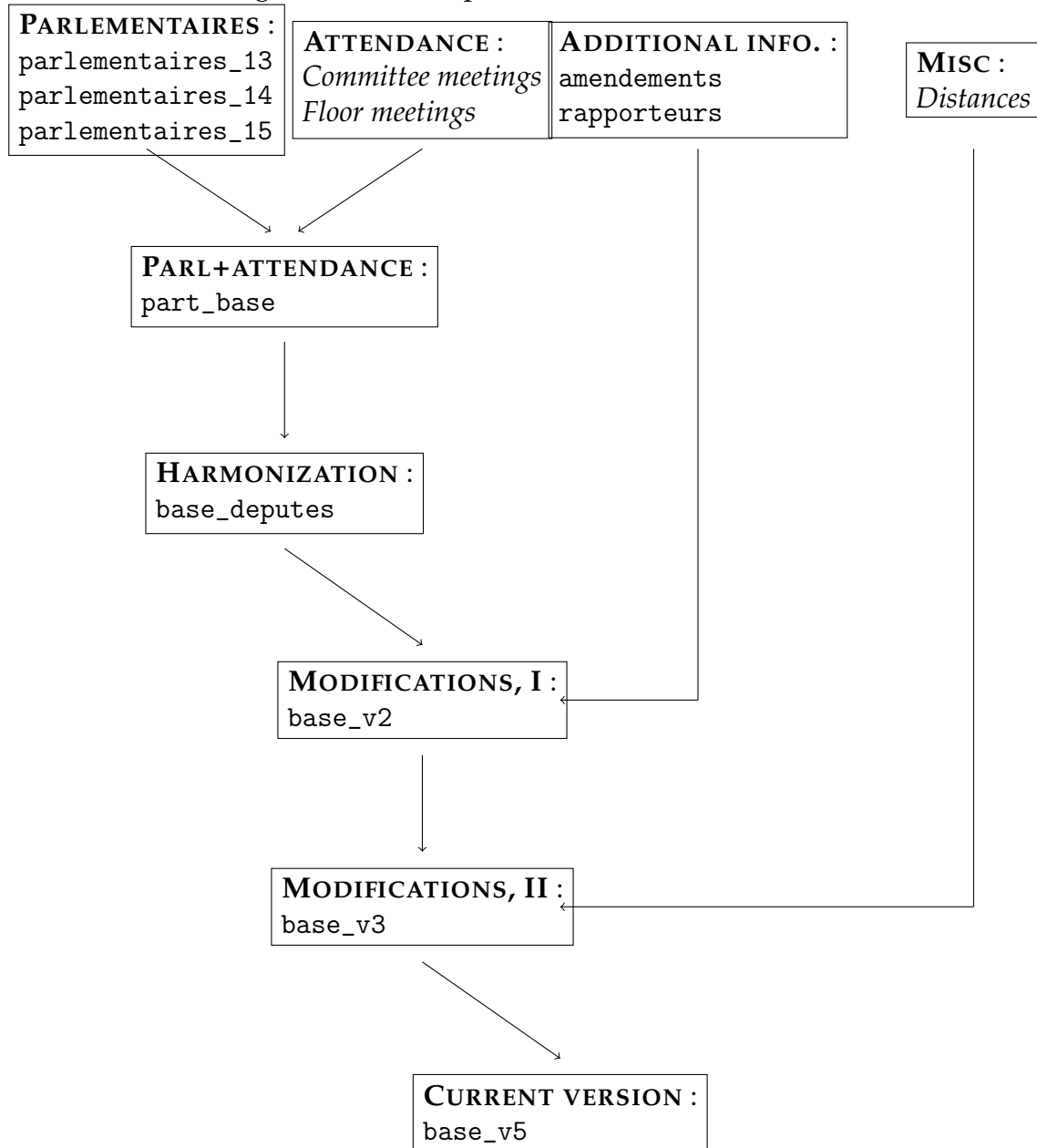
The dump for the XVth legislature includes information up to June, 21st, 2018.

2.2 Dataset construction

In practice, we generated several tables from the main database and rearrange them so as to get the final data set. Figure 2 presents how the files were combined. Further sections will be devoted to the description of this construction process.

The first line corresponds to the files that are extracted from the database. These files are ready or almost ready to use. The only exception is the tables parlementaires which require some intermediary steps before being used. Subsequent sections present the construction of the datasets of the first line. Part III presents how we went from these files to the final dataset.

Figure 2: Visual representation of the versions



3 Construction of the parlementaires tables

The parlementaires tables are by far the most complicated to construct since any information regarding the political career of the MPs was encapsulated in a code that needed to be translated into usable variables. This process involved the use of additional files and scripts. In table 9 we distinguish two parts in the construction of the database (*baseline* and *career*).

Overall, there are 3 tables: parlementaires_13, parlementaires_14, parlementaires_15. The construction of the parlementaires databases follows the following steps:

1. Extract the relevant table from the SQL database
2. Generate 5 .csv files
 - 1 from which basic biographic information will be derived [named hereafter *baseline*]
 - 4 that from which information regarding the political career, past and present of the MP, will be derived [named hereafter *political career*]
3. Once 4 readable files are generated, merge them together into one .dta file
4. Merge the latter *career* file with the *baseline*
5. Edit a little more the merged file so as to get the final dataset.

3.1 The *baseline* file

3.1.1 Step one: cleaning the file and adding some basic information

This file contains the basic biographic information about the MPs. In Excel we first add the following variables :

- mandat_complet
- suppleant

mandat_complet takes the value 1 if the MP completed his term, 0 otherwise. This is to account for different durations later in the analysis. On the other hand, suppleant takes the value 1 if the MP is the substitute to another, 0 otherwise.

We also provide information regarding the job of the MP. To do it, we rely on information gathered on the *Assemblée nationale* website³.

³Sources: http://www.assemblee-nationale.fr/qui/xml/cat_soc_prof.asp?legislature=13
http://www.assemblee-nationale.fr/qui/xml/cat_soc_prof.asp?legislature=14
<http://www2.assemblee-nationale.fr/deputes/liste/cat-sociopro>

We generate an external file so as to match each MP with its occupation and then bring it back into the baseline file. This is all done manually. A feature at this stage is that some MPs do not have any jobs. They are marked as "*Néant*" and we manually check their socio-professional status manually.

3.1.2 More editing with STATA

More serious editing is done with STATA. we turned *sexe* into a dummy variable *sex* that takes 1 if the MP is a woman. Regarding the political groups, we generated some new variables:

- *indep* if the MP is an independent
- *left* if the MP is left-leaning
- *right* if the MP is right-leaning
- *centrist* if the MP is centrist
- *maj* which will be the variable of interest and takes 1 if the MP belongs to the majority group

The political groups are defined by the *Assemblée* as a group of 15 or more MPs. They do not have to belong to the same political party to be in the same group. Actually, the notion of political party does not exist within the *Assemblée*, yet the political groups overlap the different parties that exist outside the House.

At the beginning of each legislature, groups are formed. For the three legislatures, they are the following. Bold highlights the groups that belong to the *majorité présidentielle* (as defined by Wikipedia)⁴.

Table 1: Political group acronyms

XIIIth Legislature		XIVth Legislature ⁵		XVth Legislature	
Group name	Orientation	Group name	Orientation	Group name	Orientation
UMP	right	SER	left	LREM	center
NC	right	RRDP	left	MODEM	center
SRC	left	GDR	left	LR	right
GDR	left	LR	right	UAI	right
NI	–	UDI	center	NG	left
		NI	–	LFI	left
			–	GDR	left
			–	NI	–

⁴In the eyes of the *Assemblée*, the definition of the majority group is not the same: the majority is defined as the biggest non-opposing group.

MPs who do not belong to a group are marked as "NI" or *non inscrits*. As such, it sort of defines a group. we did not have values for 26 MPs (for the XIIIth legislative term), so we completed their group belonging manually. For our definitions, we rely on Wikipedia. What is more, some MPs had no affiliation, so we checked their pages and see what was their political group or political orientation.

Last but not least, we created an age variable corresponding to the age of the MP when he began his term.

3.2 The career files

The career part falls into 4 main parts: [3.2.1](#), the past MP mandates, [3.2.2](#), the current political career, which keeps track of potential multiple offices holders. Part [3.2.3](#) gathers information about the past other offices of the MP. Finally the last "career" part, [3.2.4](#), is devoted to the introduction of additional information about the district of the MP.

3.2.1 Past MP offices

See the Python script `prev_mp_xp.py` for more technical and detailed information. The general idea is to slice an initial sequence of characters so as to get 2 objects: a integer that counts the overall number of mandates (counting the current one) and a $k \times 3$ matrix (where k is the number of mandates) gathers the start year, end year and rationale for the en of the mandates of the MPs, i.e, it kind of looks like this:

$$\begin{pmatrix} 2007 & 2012 & \text{fin de la legislature} \\ 2002 & 2007 & \text{fin de la legislature} \\ \vdots & & \\ 1988 & 1988 & \text{annulation de l'élection} \end{pmatrix}$$

The variables we generated using these objects are summarized in table [2](#)

⁵For this legislature, we manually attributed MPs to their groups (mostly to SER). It is based on Wikipedia and the *Assemblée nationale* website <http://www.assemblee-nationale.fr/14/qui/declarations-groupes.asp>

⁶For MPs who entered the *Assemblée* after 2007, we marked it as 0 since the ongoing span does not reflect any accumulated experience

Table 2: Generated variables based on the previous offices

Variable	Support	Description
nb_mdt	\mathbb{N}	Total number of mandates of the MP
mdt_cons	\mathbb{N}	Total number of consecutive mandates (current spell)
first_elec	Year	Date of the 1st election of the MP
seat_since	Year	Date of the last entry in the <i>Assemblée</i>
consec_years	\mathbb{N}	Number of consecutive years in the <i>Assemblée</i> prior to 2007
total_exp	\mathbb{N}	Total experience as of 2007 ⁶
annulation	$\{0,1\}$	Dummy taking 1 if one of the MPs mandate has ever been canceled
deces	$\{0,1\}$	Takes 1 if the MP died during his term
demission	$\{0,1\}$	Takes 1 if the MP resigned
nomination	$\{0,1\}$	Takes 1 if the MP was nominated in the government
in_out	$\{0,1\}$	Takes 1 if the MP completed 2 or more distinct spells during the legislature

3.2.2 Other offices: looking for multiple office holders

The idea was again to extract interesting information regarding the other offices held (or not) by the MP. An MP, if he holds multiple offices, can belong either to the executive or legislative branch at the local level. we wanted to take this into account. Also, we added some additional information once the *département* or the *région* or the MP was identified.

The general idea is to attribute a set of dummy variables so as to characterize the MP local mandates (if he holds multiple offices) and to add some information about the *département*, the *région* or the city (e.g. distance and travel time to Paris) when it is relevant.

We remain agnostic on the moment where the MP held multiple offices. If the MP was, for instance, mayor and MP, it could be between 2007 and 2012 for the latter and either mayor between 2007-2008, 2008-2012 or 2007-2012 (the MP was reelected).

Table 5 summarizes the variables we created. Subsequent sections are devoted to a deeper presentation of the construction of these variables.

A quick overview of the French local administration The French local administration is often referred as the *mille-feuille territorial* due to its complexity and its entanglement of structures and prerogatives. The purpose of this part is to clarify the great principles and the biggest changes this administration undergo for the past 10 years so as to be able to understand how the MPs who are multiple officeholders integrate into this *mille-feuille*⁷.

The three tiers of local government There are three tiers of local government in France: the *commune*, the *département* and the *région*. There are 36,000 *communes*, 101 *départements* and 27/18 *regions*.

⁷Sources : <http://www.vie-publique.fr>,
<http://www.collectivités-locales.gouv.fr>

The lowest tier is the *commune*. Every 6 years, lists compete so as to be represented in the *Conseil municipal*. Depending on the size of the village, the rules governing the election slightly change. Nevertheless, the main idea is that *conseillers municipaux* are elected through these lists and then seat in the *Conseil municipal*.

Once the *Conseil municipal* is elected, it itself elects the mayor and his deputies. The total amount of deputies cannot exceed 30% of the number of people seating in the *Conseil municipal*. It is said that the *Conseil municipal* is the *organe délibératif* of the city, while the mayor and its deputies form the *organe exécutif*. This can roughly be assimilated to the legislative and executive branches at the city level.

The second tier of government are the *départements*. The *département* is split into *cantons* and in each *canton*, a candidate is elected (through a two-round system). All the elected candidates form the *Conseil général*, which is elected for 6 years. More precisely, half of the *Conseil général* is renewed every three years⁸. At each renewal, the *Président* and *vice-présidents* of the *Conseil général* are elected within the *Conseil général*. They form the executive branch.

The rules evolved in 2014. The *Conseil* is now fully renewed every six years, voters now vote for a pair and the number of *canton* was reduced. Also, the elections are renamed *élections départementales* and the *Conseil général* becomes the *Conseil départemental*. As for the election of the presidents and vice-presidents, the procedure remains roughly the same.

The last tier corresponds to the *régions*. The *Conseil régional* is composed of *Conseillers régionaux* elected through lists and for six years. The *Conseil* is fully renewed every six years. The president and vice-presidents are elected within the *Conseil*. The following table summarizes the terminology⁹.

Table 3: Summary of the jargon

Tier	Executive branch	Legislative branch
City	<i>Maire</i> <i>Adjointes au maire</i>	<i>Conseil municipal</i>
<i>Département</i>	<i>Président</i> <i>Vice-présidents</i>	<i>Conseil général (until 2014)</i> <i>Conseil départemental</i>
<i>Régions</i>	<i>Président</i> <i>Vice-présidents</i>	<i>Conseil régional</i>

⁸*Cantoniales* elections were held in 2008 instead of 2007 so as not to overload the electoral agenda of 2007.

⁹Note that this is not very precise. The goal is to fix ideas so as to later define relevant variables.

The *intercommunalités* Between the cities and counties, an intermediate tier exists. It corresponds to the *intercommunalités*. we will not dig into the technicalities, but the overall idea is that several cities mutualize some of their attributions with the surrounding cities. The juridical definition of this mutualization constantly evolves and there exist many different types of *établissements publics de coopération intercommunale* (which are the administrative structure for this mutualization). The most important of them for our purpose falls into 4 categories:

- *Communautés de communes*
- *Communautés d'agglomération*
- *Communautés urbaines*
- (Since 2015 : *Métropoles*)

For each of these *EPCI*, there exists a "legislative" and an "executive" branch. It happens very often that a mayor is also the president of the *EPCI* the city belongs to. The 4 types of *EPCI* correspond to 4 levels of integrations between cities and to 4 thresholds in terms of population density. Indeed, a *communauté de communes* will typically be found in rural areas whereas *métropoles* correspond to the quasi-merger of cities in densely populated areas. As such, the president of a *communauté urbaine* will have responsabilités in bigger cities than the president of a *communauté de communes*. Also, *métropoles* appeared in 2014¹⁰. Within the *métropoles*, the *Métropole de Lyon* has a particular status it replaces the *département* and its attributions on its territory.

The particular cases of Paris, Lyon, and Marseille The *Loi relative à l'organisation administrative de Paris, Marseille, Lyon* (or "*loi PLM*") voted in 1982 organizes the administrative status of the cities of Paris, Lyon and Marseille. Each city is divided into *arrondissements* and each *arrondissement* as its town hall. On top of that, the city is governed by a city hall that is sort of "superimposed" over those *mairies d'arrondissements*. Also, if the *mairies d'arrondissements* are elected just as regular town halls, they rely on the city hall for their credits. As for the city hall, it is composed of the third of the *conseillers municipaux d'arrondissement*. Then, at the city level, a mayor and its deputies are elected. These three cities can roughly be seen as two-layer cities with the lower-layer corresponding to a regular municipality and the upper-layer duplicating the lower-layer. See the figure hereafter for a more visual explanation.

Still there are some exceptions. The *arrondissements* and the *secteurs*, which are the elective district, do not necessarily coincide. In Marseille, there are 16 *arrondissements* but 8 *sec-*

¹⁰*Loi de modernisation de l'action publique et d'affirmation des métropoles* or "*loi MAPTAM*"

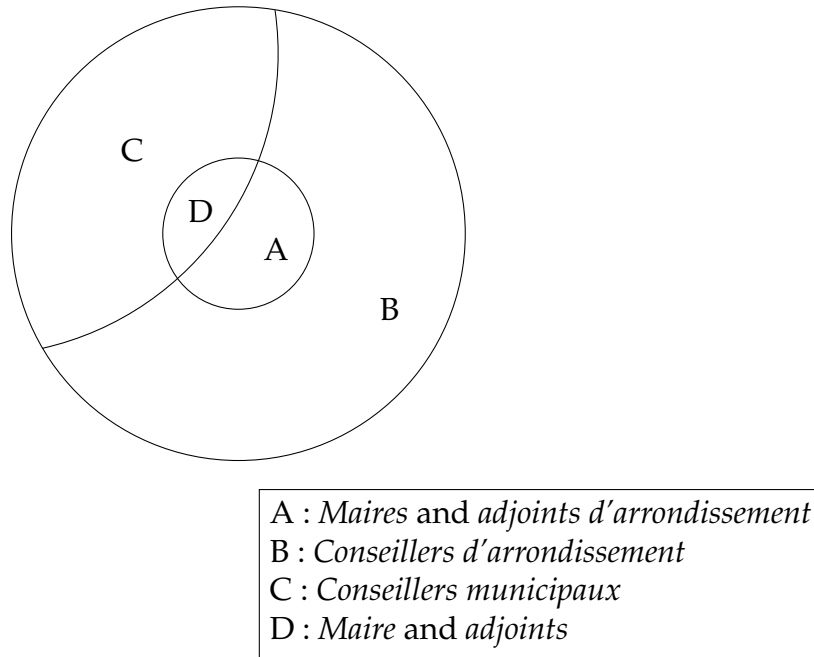


Figure 3: Sketch of the administrative organization of Paris, Lyon and Marseille

teurs, hence 8 *mairies de secteur*. Also, in Paris, from 2020 onwards, there will only be 17 *secteurs* (the 1st, 2nd, 3rd and 4th *arrondissements* will merge). Last but not least, The upper-layer of Paris, the *Conseil de Paris* has the same prerogatives as a *département*. Indeed, since 1964, Paris is both a city and a county. Yet for historical reasons, the powers of the *Mairie de Paris* are limited to the benefit of the *Prefecture de Police* (*arrêté des Consuls du 12 messidor an XII*).

A word on overseas territories The French territorial administration also features *territoires d'outre-mer*, i.e. overseas territories. Depending on their degree of independence, they range from a *sui generis* status (Nouvelle Calédonie) to a regular county. Also, France's latest county is Mayotte (since 2011). It often comes that an overseas territory is both a *région* and a county (e.g. Guadeloupe, Martinique). Corsica has since 2018 the status of *collectivité territoriale unique*, that is, a *région* with a specific status.

For the sake of simplicity and because it is not crucial for our purpose, we ignore these specificities. Oversea territories are considered as both *départements* and *régions* in our datasets, irrespective of what their actual status is.

Generating new variables and adding information An MP, when holding multiple offices, can (at least before 2017) either belong to the legislative or executive branch of one of the local levels. That is why we defined *delib* and *exec* as dummies taking one if the

MP belongs to either of these groups. Also, we saw that the executive branch was elected within the local assembly, meaning that $\text{exec} = 1 \Rightarrow \text{delib} = 1$. These dummy variables are defined for the three tiers of government and the *intercommunalités* in the same way. For the latter, we also define an ordinal variable ranging from 1 to 4 depending on the type of EPCI one belongs to.

These different variables are built so as to compute future interaction effects depending on the local level of responsibility.

We also add information on the number of inhabitants of the municipality (*habitants*). Indeed, the data often features such an information at the city level. For the counties and *régions*, we also decided to add information from an external file compiled from the Insee databases. More precisely, the county is always indicated, whenever the MP holds an office at the city and/or county level. The information provided at the county level is the following:

- County ID
- Distance to Paris (km)
- Travel time by train (hours, decimal)

Also, when the MP is a member of the county assembly, we add information on the population of the county. we proceed similarly at the *région* level¹¹; information on the distance and travel time to Paris and the population is provided based on another external dataset. Finally, we provide dummy variables whenever the MP hold an office in Paris, Lyon or Marseille (*plm*) or overseas (*oultre_mer*)

We chose to include both the travel time and the distance because the salient parameter for an MP traveling from his local office to the *Assemblée* is more the travel time than the distance. Also, we chose to include the train because MPs can travel by train for free and the train offers downtown to downtown connection.

Yet, we had to consider some exceptions. Ardèche is the only county that does not have a train station. As such, we considered the travel distance by road. The starting point in

¹¹For the *région*, additional information is only provided when the MP holds an office at this tier of government, whereas for the *département* part of the information is common to the two lowest tier of government (distance and travel time)

counties we considered are the prefectures. This is because the prefecture is where the *conseil général* usually meets (or nearby) and when a connection to Paris exists, it often goes through this city.

As for overseas territories and Corsica, we chose not to include any travel time, only distances. For the regions, we considered the region capital, which is one of the counties prefectures (e.g. Rennes is the prefecture of *Ille-et-Villaine* and the capital of the *Région Bretagne*). Depending on the dataset, the populations are given for the years 2007, 2012 or 2017.

Some remarks about the construction of the variables The codings for the counties and *régions* are arbitrary and are only used for further identification (e.g. with the birthplace of the MP). For the mainland counties, it turns out that my coding matches the official numbers for the counties ranging from *Côte d’or* (21) to *Val-d’Oise* (95). This is purely a coincidence.

Finally, when a variable is empty, we chose specific codes depending on the reason. These codings and their meanings are summarized in table 4.

Table 4: Error coding explanations

Code	Meaning
-99	Empty tier, but the MP is a multiple office holder
-98	The MP does not hold multiple offices. All variables for this MP are coded this way
-97	No information regarding the number of inhabitants of the city available
-96	Omitted travel time (Corsica, overseas)

Table 5: Generated variables based on the other offices

Variable	Support	Description
plm	$\{0, 1\}$	Dummy taking one if the city in question is Paris, Lyon or Marseille
delib_m	$\{0, 1\}$	Takes 1 if the MP is member of a City Council
exec_m	$\{0, 1\}$	Takes 1 if the MP is member of the Board of the City Council <i>If equal to 1, it implies delib_m = 1</i>
delib_v	$\{0, 1\}$	Same as delib_m but for Paris, Lyon or Marseille
exec_v	$\{0, 1\}$	Same as exec_v but for Paris, Lyon or Marseille <i>If equal to 1, it implies delib_v = 1</i>
habitants	\mathbb{N}	Number of inhabitants in the City (if indicated)
delib_epci	$\{0, 1\}$	Dummy taking 1 if one if the MP is member of the Council of an <i>intercommunalité</i>
exec_epci	$\{0, 1\}$	Takes 1 if the MP is member of the board of an <i>intercommunalité</i> <i>If equal to 1, it implies delib_epci = 1</i>
epci_type	$\{1, 2, 3, 4\}$	Gives the type of <i>intercommunalité</i> between: <i>Communauté urbaine</i> (1) <i>Communauté d'agglomération</i> (2) <i>Communauté de communes</i> (3) <i>Métropoles</i> (4)
delib_d	$\{0, 1\}$	Takes 1 if the MP is member of the <i>Conseil général/départemental</i>
exec_d	$\{0, 1\}$	Takes 1 if the MP is member of the board of the <i>Conseil général</i> <i>If equal to 1, it implies delib_d = 1</i>
pop_d	\mathbb{N}	Indicates the number of inhabitants in the <i>département</i> (when relevant)
distance	\mathbb{N}	Indicates the distance (km) relative to Paris
temps_trajet	\mathbb{R}	Indicates the travel time between the prefecture to Paris by train* (*Except for Ardèche (car), Corsica and overseas)
dept_no	\mathbb{N}	Indicates <i>département</i> number* (* It matches the real number up to some exceptions)
delib_r	$\{0, 1\}$	Takes 1 if the MP is member of the <i>Conseil régional</i>
exec_r	$\{0, 1\}$	Takes 1 if the MP is member of the board of the <i>Conseil régional</i> <i>If equal to 1, it implies delib_r = 1</i>
pop_r	\mathbb{N}	Indicates the number of inhabitants in the <i>région</i> (when relevant)
distance_r	\mathbb{N}	Indicates the distance (km) relative to Paris
tps_trajet_r	\mathbb{R}	Indicates the travel time between the local capital to Paris by train*
reg_code	\mathbb{N}	Indicates <i>region</i> code (arbitrary)
oultre_mer	$\{0, 1\}$	Takes 1 if the <i>département/région</i> is located overseas.

3.2.3 Past other offices: capturing the political career

General idea This part consists of merging the time dimension developed for the past offices and the spatial dimension developed for the other offices. The general idea is to generate 4 variables for each branch at each level:

- **primo**: the year the MP was elected at this office for the 1st time
- **exp**: the total experience of the MP at this office
- **consec**: if, when the MP term began, he was holding another office, this tells us for how long

- depuis: the year the MP took office

For the two last variables, consider the case of somebody who was elected mayor in the 1990s and then again in 2001. Also, this guy was elected MP in 2007. Then for the mayor layer, consec will be 7 (we consider the ongoing year as a full year) and depuis will be 2001, while primo will be any time in the 1990s and exp will take into account both his current and past terms.

When we speak about layers, we consider two things: the "legislative" and "executive" branches at the local level, as defined in the previous section and the layer in question (city, *intercommunalité*, county, region). As such, for each tier, my variables fall into two parts, exec and delib, which correspond to the executive and legislative branches respectively. Finally, the tiers are identified by m, v¹², epci and r.

On top of that, dummy variables are included so as to assess whether the MP has an experience for any given (tier, branch), if the MP has an experience in several different counties or regions (pluri) and additional variables regarding geographical features are included (distance and travel time to Paris and population for the regions). For these latter variables, the description and limitations are exactly the same as those described in the previous section.

Some technical precisions The error codes of the previous part are augmented with the following:

Table 6: Error coding explanations

Code	Meaning
-95	Missing information due to offices in different counties/regions
-94	No data for this (tier,branch)
-93	No active mandate at this (tier,branch) when elected MP
-91	Not concerned by this city tier
-90	Holds multiple offices

Also, this third part requires information that is extracted in the first part (namely the year the MP took office). As such, this part must be the last to be computed. Also, as mentioned in the foreword of the script, the original dataset needs some refinements on the counties names, otherwise there could be some mistakes (the counties' names overlap so the information gathered based on these names can be wrong).

¹²v will only be applied to the cases of Paris, Lyon, and Marseille as described in the previous section.

Finally, the procedure is not fully automated, although it saves quite a lot of time, it is still required in the end to check that there are no mistakes. What is more, further updates regarding the population are required at each stage.

Table 7: Summary of the variables for the past offices. i indicates the tier.

Variable	Support	Description
Common variables ($i \in \{m, v, epci, d, r\}$)		
oultre_mer	$\{0, 1\}$	Takes 1 if the local tier is overseas
Executive branch		
exec_i	$\{0, 1\}$	Dummy taking 1 if the MP ever had executive responsibilities
exec_i_primo	Year	Year the first election to this office
exec_i_exp	\mathbb{N}	Total experience for this office
exec_i_depuis	Year	Year of the last "entry" into this office
exec_i_consec	\mathbb{N}	Consecutive years at this office
Legislative branch		
delib_i	$\{0, 1\}$	Dummy taking 1 if the MP ever had executive responsibilities
delib_i_primo	Year	Year the first election to this office
delib_i_exp	\mathbb{N}	Total experience for this office
delib_i_depuis	Year	Year of the last "entry" into this office
delib_i_consec	\mathbb{N}	Consecutive years at this office
City/County/Region variables ($i \in \{m, v, d, r\}$)		
pluri_i	$\{0, 1\}$	Takes 1 if the MP had responsibilities in different counties/regions
City/County Variables		
dist	\mathbb{R}	Distance between the prefecture and Paris
tps_trajet	\mathbb{R}	Travel time by train to Paris
dept_no	\mathbb{N}	County identification number
Municipality only		
plm	$\{0, 1\}$	Dummy taking 1 if one is in Paris, Lyon or Marseille
Region only		
distance_r	\mathbb{R}	Distance between the capital and Paris
tps_trajet_r	\mathbb{R}	Travel time
pop_r	\mathbb{R}	Population
reg_code	\mathbb{N}	Regional id

3.2.4 District information

Using the file containing information about the counties, we generated new variables using the same method. These variables are the following:

Table 8: Generated variables for the district

Variable	Description
dept_n_c	Number of the birth county
circ_n	Number of the district's county
circ_pop	Population of the district's county
circ_dist	Distance from Paris of the district's county
circ_tps	Travel time to Paris of the district's county
circ_om	Dummy indicating if the county is located overseas
parachute	Dummy indicating whether the birth and district county are different

There are also 3 new error codes: 95 (no information regarding the birth district), -92 (no information regarding the district) and -89 (no information regarding the county).

Table 9: Map of the parlementaires files

Baseline			Career			
			PRELIMINARY FILES			
Files	Auxiliaries	Remarks	Input files	Scripts	Output files	Remarks
parlementaires_baseline.csv	profession.csv ¹³		parl_modif_noheader.csv	circ_info.py	complements.csv	
	parlementaire.csv		parlementaire_input.csv	past_mp_xp.py	sortie_finale.csv	District Information
	scrap_csp.html		parlementaire_input_2.csv	other_offices.py	sortie_finale4.csv	Other offices Information
	parlementaire_csp.xlsx		dept_stats.csv			
			reg_stats.csv			
			parlementaire_input_3.csv	past_officies.py	sortie_finale5.csv	Past others Information
			dept_stats.csv			
			reg_stats.csv			
			past_mp_short.csv			
STATA FILES ¹⁴						
File name			File names			
parlementaires_baseline.dta			circo.dta			
			past_mp.dta			
			others.dta			
			past_others.dta			
			political_career.dta			
			Merged file			
			parlementaires_xx.dta			
			Associated .do file: parlementaires_construction.do			

¹³profession.csv is a $n \times 2$ vector that gathers the CSP as defined by the *Assemblée nationale* website for the current legislature and the MPs names and sorted in scrap_csp.html. parlementaire.csv is a file extracted from the database with the name and the ID of the MP. The socio-professional category of each MP is manually merged on parlementaire_csp.xlsx, then ordered and pasted in parlementaire_baseline manually.

¹⁴These files correspond to the simple conversion into a .dta format of the generated preliminary files.

4 Construction of the meetings tables

There is two kinds of meetings: the public meetings (or the floor meetings), presented in subsection [4.1](#) and the committee meetings (see [4.2](#)). By committee meetings, we mean the *commissions*, where the bills are discussed, but also more broadly all the agencies within and outside the National Assembly to which MPs participate. Of the 6,000 meetings recorded, around 1,300 are floor meetings. The rest corresponds to committee meetings.

4.1 Floor (public) meetings

4.1.1 Extracting the relevant information from the database

As always, we start from the database and extract the initial raw material we need. More precisely, we consider all the floor meetings, so we use the command `SELECT id date annee session FROM seance WHERE "type"="hemicycle" ORDER BY date ASC`. In words, we needed the session id, its date, and session.

This leaves me with an Excel sheet containing n lines, each line corresponding to a recorded meeting. we have:

- 1306 recorded meetings for the XIIIth legislature
- 1359 recorded meetings for the XIVth legislature
- 316 recorded meetings for the XVth legislature^{[15](#)}

This corresponds to the file `liste_seance.xlsx`.

4.1.2 Distinguishing legislative and control meetings

The initial problem The problem is that the only information we have about the floor meetings is whether they are floor meetings... which is not very informative. Indeed, depending on the type of the meeting, the repartition of the speaking time will be different between the various MPs. For instance, when examining a bill, the majority is endowed with half the overall speaking time, while for other types of meeting this might not be the case.

This, of course, has an influence on the prior probability of each MP to speak. Also, it does not make sense to penalize an MP who did not speak while he couldn't. As such, the

¹⁵The data from the XVth legislative term only contains information between June, 21st, 2017 and June, 21st, 2018.

rationale for recovering the meetings type is to be able to control for the type of meeting and therefore being able to disentangle the effect of the speaking rules and the pure effect of our parameters of interest.

Identification of the type of public meeting between *control* and *legislative* meetings is not straightforward. To identify the type of meeting, we collapsed the interventions recorded for the meeting and indicate whether at least one of these interventions was labeled as a question. Manual verification using the transcript of the meeting highlighted the fact that the control meetings were mostly devoted to questioning, but also that the end or the beginning of the meeting was devoted to legislative matters.

4.2 Committee meetings

4.2.1 A word on the National Assembly's agencies

The French National Assembly is made of numerous agencies whose role are to organize the Assembly (*Bureau, Questure*), the parliamentary debates (*groupes parlementaires*), the work of the MPs (*commissions, groupes d'études, ...*).

We will present here quickly the agencies related to the missions of the MPs, i.e. the *commissions*, the *missions extraparlimentaires* and the *groupes* (excluding the parliamentary groups).

The *commissions* and *missions* There are different types of committees:

- *permanentes* whose goal is to prepare the legislative work and control the government. They are a sort of small replication of the hemicycle but with a focus on a specific topic. This is here that most of the legislative work is done
- *spéciales* who are in charge of investigating a specific topic. Since 2008 (but before in practice), the permanent committees are the first concerned when a bill comes to the Assembly.
- *d'enquête* are in charge of the contingent investigation on a given topic. It is one of the tools the MPs have to control the government.

Missions are another way for the parliament to control the government. Note that those inquiry commissions and missions can be created jointly with the upper chamber. As such, the *missions* aim at gathering information on a given topic.

The extra-parliamentary groups ¹⁶ are agencies MPs or Senators can be member of. This can be institutions such as the CNIL (in charge of the data privacy law enforcement) or the supervisory board of the *Caisse des dépôts*, a public national institution in charge for long-term public investments.

The study and friendship groups The latter corresponds to groups that gather MPs sharing a common interest regarding a foreign country. Their goal is to develop links between the national and foreign parliaments. They are also in charge of the reception of foreign officials in the Assembly. As for the study groups, they are instances designed so that MPs can enrich their knowledge regarding a specific topic. Study groups can have access to facilities within the assembly but they have no financial autonomy.

Identification in the data We restrict our attention to committee meetings. These meetings are recorded as *organismes* in the dataset. Table ¹⁰ summarizes the different values attached to the permanent committees¹⁷. The reason why we focus on the permanent committees is that they make up most of the committee meetings and are the relevant committees when it comes to first examine a bill at the National Assembly.

Table 10: Ids of the permanent committees

	XIIIth leg.	XIVth leg.	XVth leg.
Committee	id	id	id
<i>Affaires culturelles et éducation</i>	237	16	11
<i>Affaires économiques</i>	204	14	13
<i>Affaires étrangères</i>	13	10	18
<i>Affaires sociales</i>	22	9	14
<i>Défense nationale et forces armées</i>	2	12	16
<i>Développement durable et aménagement du territoire</i>	11	11	17
<i>Finances</i>	211	15	12
<i>Lois</i>	212	13	15
<i>Affaires culturelles, familiales et sociales</i>	203	–	–
<i>Affaires économiques, environnement et territoire</i>	220	–	–

¹⁶Source : <http://www.senat.fr/oep/index.html>

¹⁷Note that during the XIIIth legislature, 2 permanent committees disappeared the 1st of July, 2009. More precisely, the *commission des affaires culturelles et de l'éducation* and the *commission des affaires sociales* took over the *commission des affaires culturelles, familiales et sociales* while the *commission des affaires économiques* and *commission du développement durable et de l'aménagement du territoire* took over the *commission des affaires économiques, de l'environnement et du territoire*. Source: <http://www.assemblee-nationale.fr/13/commissions/commissions-index.asp>

4.2.2 Extracting and editing the tables

The code used to generate the .csv file is the following:

```
SELECT seance.id, seance.date, seance.annee, seance.type, seance.organisme_id,  
organisme.id, organisme.nom, organisme.type  
FROM seance  
INNER JOIN organisme ON  
seance.organisme_id = organisme.id
```

We then recover the Excel file and import it into Stata. The precise instructions are written in the associated file `committees-edit.do`. The first edition consists of converting the date into the right format, keeping only relevant variables and rearranging them.

5 Construction of the complementary tables

Three important variables were not included in the first versions of the dataset: amendments, rapporteur, and questions. These variables were important since they are important predictors of the probability to speak for an MP during floor meetings. Definitions of these variables are provided in the main section of the master thesis. Here, we focus on the construction of the tables.

5.1 Amendments

Knowing whether an MP tabled an amendment is important because it allows him to speak in order to present his amendment. Amendments are attached to a bill. They can be tabled either during committee or floor meetings. The only information we have for the three legislative terms deal with hemicycle amendments. These hemicycle amendments are interesting because they used to be used as a way to filibuster by the opposition.

The problem with the database is that we do not have information on the agenda of hemicycle meetings. What is more, amendments must be tabled prior to the meeting in order to be evaluated by the Bureau of the Assembly. Indeed, amendments can be rejected if they do not meet some requirements or withdraw by the MP himself. Also, during the meeting, amendments can "fall" if similar or opposite amendments are adopted. In this case, the MP will not be granted time to explain his amendment, even if it was neither rejected nor withdrawn.

For these reasons, we restrict our attention to amendments who were either accepted or rejected and drop all of those who were refused, withdrawn or those who fell.

We dealt with the agenda identification issue by combining several datasets in order to associate amendments tabled by MPs with the relevant floor meeting. By relevant floor meeting, we mean the meeting during which the bill the amendments relates to was discussed. In practice, we combined the `amendements` table with the `texteloi` table, which contains all bills. These were combined with various identifiers of the moment of the intervention in order to finally get the list of meetings during which the bill was discussed. We do not distinguish among the *discussion générale* or *phase d'examen détaillé*. The amendment variable takes value 1 if the MP tabled an amendment (either accepted or rejected) for *this* bill, irrespective of whether the amendment itself was discussed at the beginning or the end of the subsample of meetings devoted to this bill.

For technical reasons, we extract information from the database in two parts. Both parts are then combined and finally encapsulated into a STATA file which contains for each MP the list of the floor meetings during which he tabled an amendment.

The requests for the first (up) and the second (down) part are:

```
select parlementaire_amendement.parlementaire_id,  
parlementaire_amendement.numero_signataire, amendement.id, amendement.numero,  
amendement.sort, amendement.texteloi_id, texteloi.id_dossier_an  
from parlementaire_amendement  
left join amendement on parlementaire_amendement.amendement_id=amendement.id  
left join texteloi on amendement.texteloi_id=texteloi.id
```

```
select section.id_dossier_an, section.section_id, intervention.seance_id  
from section  
right join intervention on section.section_id=intervention.section_id
```

5.2 Rapporteurs

Information on the rapporteurs is gathered in the same way as for the amendments. The same problem of non-identification of the meeting's agenda was dealt with in the same way. Also, we restrict on the amendment status only for floor meetings and not committee meetings. The request to recover the initial data is as follows:

```
select parlementaire_texteloi.parlementaire_id, parlementaire_texteloi.fonction,
texteloi.id, texteloi.id_dossier_an
from parlementaire_texteloi
left join texteloi on parlementaire_texteloi.texteloi_id=texteloi.id
```

5.3 Questions

Questions are a precise type of meeting intervention. They are directly identified from the intervention table. They were used to identify the type of meetings, here interventions are used for themselves and the variable question indicates if, for a given meeting, the MP asked at least one question.

6 Additional information

The construction of the database needed some additional information about the seats of the MPs and various information regarding their interventions.

6.1 Seats in the hemicycle

6.1.1 Motivation and variables

We use the `place_hemicycle` variable in order to extrapolate some information about the MPs. We know that placement is endogenous: each group is attributed a quota in the hemicycle and then MP chose to seat where they want. However, not all seats are equal. More experienced MPs tend to seat in the front of the hemicycle, while the least experienced seat in the back. Microphones are located in the spans of the Assembly, so more talkative MPs might choose to sit next to microphones.

What is more, the higher the average distance between the MPs and the group president, the bigger the group. Therefore, this average distance can be used as a proxy measure of the size of the group. In practice, this average measure is combined with the raw size of the group.

Also, we know that the group president chooses the speakers. We might think that there is a correlation between the proximity of the MP and his group president and their degree of affinity, therefore the closer the MP, the higher the chances that he will be picked as a speaker.

This leads us to define the following variables:

- hauteur: height (in m) of the seat of the MP

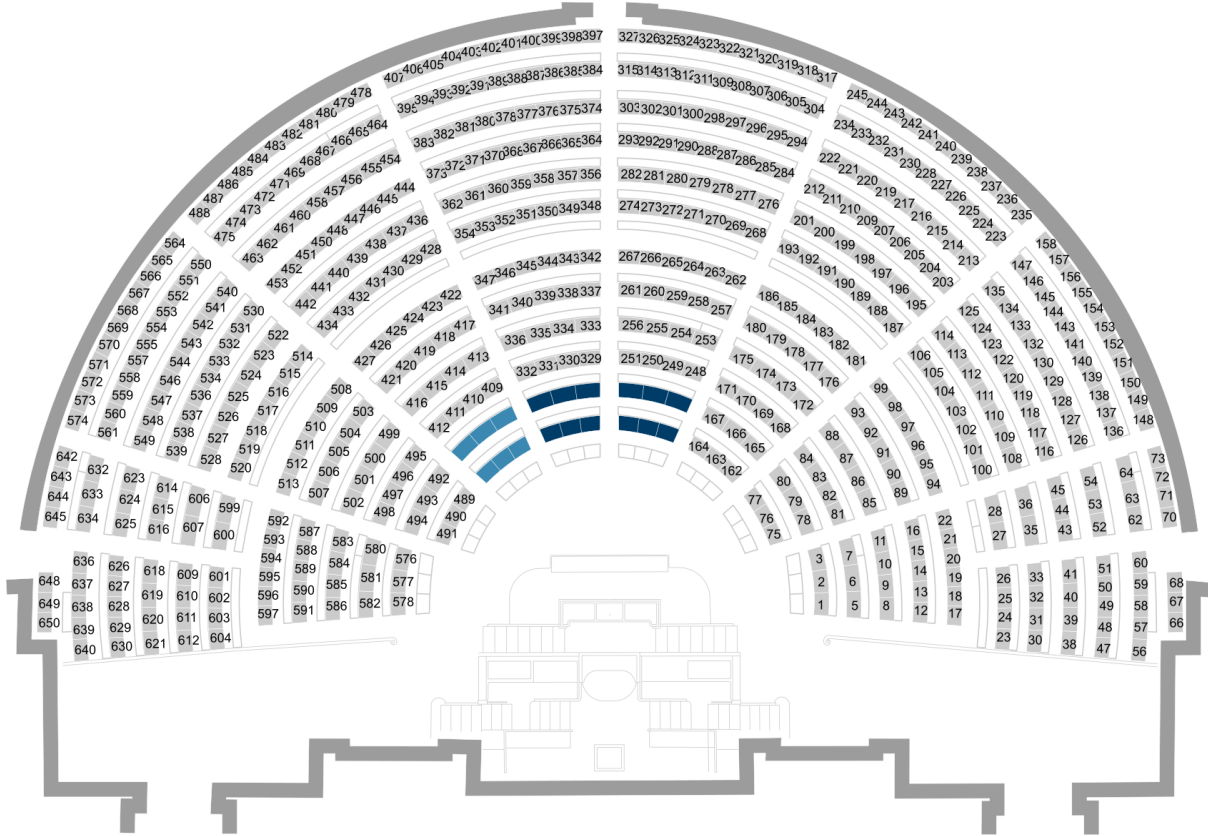
- `travee`: indicates if the seat is located near a span
- `distance_grp`: distance (in m) between the MP and his group president
- `dist_ave`: average distance (in meters) between the group president and the MPs attached to the group

6.1.2 Construction of the variables

Based on a generic map of the seats of the Assembly, we attribute coordinates to each seat. Figure 4 serves as the coordinates systems. Each seat number is associated with a unique set of coordinates in the (x, y) plane. To get the z dimension, we use the fact that the approximate height of the size of the desk of the Speaker of the Assembly is 3 meters high and is as high as the last row. Therefore, height is then attributed and is comprised between 0.8m and 3m. Each seat being associated with (x, y, z) coordinates, we then compute the Euclidean distance between the seat of the group president and the seat of the MP. As for `hauteur` and `travee`, they are intrinsic properties of the seats and do not change over time. Conversion is based on the fact that the approximate diameter of the hemicycle is 18,57m across.

Regarding the group presidents, a particular case is the green group of the XIVth legislature. This group wanted to enforce a co-presidency, which is forbidden by the internal regulation of the Assembly. Therefore, the Group president officially changed every six months. However, in order to account for the practice, we considered the president-pairs over time and attributed a distance based on the nearest president. For all presidents, the distance for them is set to 0.

Figure 4: Seats in the hemicycle



6.2 More data

We extracted information about the interventions, the attendance, but we also include information regarding the number of words said, the number of long interventions (i.e. more than 20 words in the hemicycle) as well as the total number of intervention. All these information that serve as outcome variables are extracted using the `intervention` table of the original database.

In all cases, the quantity of interest is extracted and associated to the MP's id in the database. As such, the intermediate file (whatever the quantity of interest is) contains for each MP the list of meetings and the associated value of interest (e.g. number of words said, number of long intervention, number of intervention, etc).

Part II

Conversion into a panel dataset

After extracting and converting as much information as possible from the initial database, we shaped our data so that in the end we recover a panel dataset. The idea is that each MP i will be observed several times, the time dimension being captured by the meetings.

The difficulty that arises with this method is that all MPs are not required to attend all meetings. Indeed, they register to a committee meeting and are only supposed to attend these meetings. The only meetings all MPs are supposed to attend are the floor meetings.

We take into account the fact that the MPs are registered in one given permanent committee and will only assign to each MP the list of committee meetings that he was supposed to attend.

Building the final dataset requires three major steps: reshaping the data in a panel-form, creating the attendance and participation variables and finally harmonizing all labels across the three legislatures.

All .do files and intermediary files are saved.

7 Reconstructing the timeline

We begin with the `parlementaire` file and duplicate each MP by the number of days he remains in office. We then remove the dates before the first meeting and after the last meeting.

Using the information provided by the National Assembly website, we then assign each MP to his corresponding group and keep track of the modifications that occurred during the legislature¹⁸. During this step, group presidents (as well as the length of their "term") are identified.

The longest step is to assign each MP to a permanent committee. All MPs must register to a committee but they can change whenever it is necessary. Therefore, some MPs will always attend the same committee while other will change from time to time. The reason behind such changes might be that it is necessary for each political group to have a proportional representation across all committees. We then drop all MPs who do not belong to the list of those who were reelected at least once.

Before digging into the modifications manually, use the attendance records of each MP to see what committee he attended the most. If 90% of his records come from the same

¹⁸For the XIVth legislative term, the changes are recorded here: <http://www.assemblee-nationale.fr/14/qui/modifications.asp>

committee, this committee is assigned as the MP as his permanent committee for the whole legislature. MPs below the 90% threshold are identified and their committee is manually assigned using the information available on the website of the National Assembly¹⁹. Once we are done with the MPs below the 90% threshold, we control manually for the permanent committee of those above the threshold and if necessary modify manually the committee he belongs to.

During this step, we also identify the members of the boards of each permanent committee and, as for the group presidents, the length of their "terms". The members of the Bureau of the National Assembly are also identified. Members of the different governments of the legislative term are also dropped when nominated since they cannot take part in the meetings while in office (see *Ordonnance n° 58-1099 du 17 novembre 1958 portant loi organique pour l'application de l'article 23 de la Constitution*²⁰). Finally, MPs elected or reelected during a by-election are identified. This process is repeated three times exactly the same way. All commands are saved in a .do file.

8 Incorporating attendance and participation data

Once we have the panel baseline of the MPs, we attribute to each MP the list of committee meetings and floor meetings he was supposed to attend given his spell in office and the permanent committee he is registered in at a given time.

Using the information about the interventions of the MPs, we create seven variables: three for the committee meetings and four for the floor meetings:

- *Committee meetings*
 - `pres_com`, which indicates if the MP attended the committee meeting
 - `c_inter`, which counts the number of interventions
 - `c_mots`, which counts the number of words
- *Floor meetings*
 - `pres_h`, which indicates if the MP intervened at least once during the meeting
 - `t_inter`, which counts the total number of interventions
 - `t_mots`, which counts the number of words pronounced during the meeting

¹⁹See for instance: http://www2.assemblee-nationale.fr/deputes/fiche/OMC_PA267306

²⁰Source: <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006069202&dateTexte=20101217>

- *inter*, which counts the number of long interventions interventions at a given meeting

We also indicate if the MP attended a committee meeting of his committee or of another committee (which in this case was not compulsory). For the floor meetings, we indicate the type of meeting (control or legislative).

We then check with the data displayed on the website *NosDéputés*²¹ to see if we did not make any mistake.

It appears that their website sometimes indicates interventions for MPs they did while they were in office as government members (e.g. Xavier Bertrand). In our dataset, we corrected this small error.

Again, this process is repeated three times.

9 Further modifications

Once we have the panel dataset for all legislature, we need to homogenize the labels and create a panel that keeps track of the activity of the MPs across all the legislatures they seat in. We also replace the negative values by missing values in the multiple offices holding variable in order to avoid spurious computations. We also homogenize the socioeconomic categories across all three legislatures. The labels for the sessions and the political groups are also the same across all three legislatures.

Modifications that occurred after the first version of the dataset was finished were always conducted the same way: the dataset was decomposed into the three legislatures and the modifications were done the same way (e.g. amendments, seats) three times. The dataset was then recomposed.

²¹More precisely, information for the current legislature is available on <http://www.nosdeputes.fr>, <https://2012-2017.nosdeputes.fr> for the XIVth legislative term and <https://2007-2012.nosdeputes.fr> for the XIIIth legislature.