

# U-LeNet. Files with MPs characteristics

Cleaning and merging data for seven convocations 1998-2019

*Oleksa Stepaniuk*

*25 December, 2019*

U-LeNet dataset includes a separate file with the MPs characteristics for all convocations of Ukrainian Parliament starting from the third one till the most recent (1998-2019). Each file includes Ukrainian and English versions of MPs names, their faction, political affiliation, and the method of their election (party system or constituency). Depending on convocation, there is also additional information, for instance, education, previous education, or the number of assistants. Importantly if the MP changed several factions during the convocation, only the last one is included in the dataset.

In the raw form, the usage of these files is complicated by several factors:

1. Each MP has an ID, but the same person that was MP in several convocations has different IDs;
2. The spelling of parties, factions, and names differs across files so that the same party can have three variants of spelling. Such confusion complicates comparison;
3. A large share of observations has missing or inaccurate information on faction/party affiliation. At the same time these variables are among the most important for understanding MPs behavior;

As a result, we decided to create a unified database of all MPs across seven convocations included in the U-LeNet.

## 1 Cleaning data for each convocation before merging

### 1.1 Operations applied to all files

**Non-affiliated MPs.** When a person prematurely stops being an MP (for example, by joining the government), he or she automatically stops being a member of the parliamentary faction. Such a person is still included in the list of MPs on the Parliament website (source of U-LeNet data) but is designated as ‘Non-affiliated.’ As a result, some of the most prominent MPs such as Yulia Tymoshenko (leader of Tymoshenko Bloc) or Volodymyr Lytvyn (leader of Lytvyn Bloc) are designated as Non-affiliated. We thought that this could distort the analysis. Consequently, we checked all the Non-affiliated MPs across seven convocations. If MP resigned before the end of convocation, we added to the database his faction before the resignation.

**Rada Management.** When a person joins the Parliament management (Chairman, First deputy, Deputy Chairman of the Verkhovna Rada), he or she stops being a member of a faction. However, such individuals are representatives of some faction and wield considerable power over Parliamentary proceedings. So we decided to indicate the faction affiliation for each member of Rada management for each convocation.

**Faction position.** The file for the ninth convocation has detailed information on faction position - chairmen of factions and MPs groups, their deputies, and membership status of other MPs. The title of individual in a rada management is also indicated in this field. We added similar information to files for convocations from third to eight. Though information for these convocations is less detailed - we did not specifically indicate the deputy heads of factions/groups. Also, we unified the spelling of leader titles - instead of using both ‘head’ and ‘chairman,’ we decided to use only ‘chairman.’

**Switching to executive or judicial position.** For convocations from fifth to ninth, we checked if MP prematurely stopped his service to join the government or other position in state service. To separate such individuals, we added the variable ‘to\_exec\_judicial.’ It is more complicated to conduct analogous analyses for more remote convocations (4,5) due to the lower quality of available data. However, we plan to do this at the beginning of 2020.

**Institutions.** Files related to convocations from third to eight have records for ‘President’ and ‘Government.’ For the ease of filtering, we indicated the faction of such records as ‘Institution.’

## 1.2 Rada 9

**Numerical codes.** All categorical variables in the raw file with MPs characteristics stored using the numerical codes instead of strings (variables party, faction, faction\_position, committee, committee\_position). The meaning of those codes is stored in the text file ‘mps\_info\_9\_README.rtf.’ Such an arrangement is not convenient and also not acceptable if we want to merge/compare files across convocations. Also, codes belonging to different variables are mixed/confused in the mps\_info\_9\_README. We moved codes to the table format (file ‘mps\_9\_codes.csv’) and then substituted the codes in the main file with text values.

**Non-affiliated and Dovira MPs group.** Ninth convocation only began its work in the summer of 2019. Naturally, it is still undergoing changes in the structure of factions. We updated the MPs affiliations by adding a new MPs group (‘Dovira,’ Ukr. Trust) and correcting when necessary other affiliations. As a result, we have correct data on MPs affiliations as of December 2019.

### Comments/adjustments to variables.

- Committee and International organizations: we provided a more concise name for the committees and international organizations by removing ‘Verkhovna Rada of Ukraine’ from the name and using abbreviations, i.e., OSCE instead of Organization for Security and Co-operation in Europe. It should be noted that MPs are often members of several international organizations. The dataset includes information only about one organization per MP (the first one on the Parliament website).
- Assistants: file includes several columns with information of MPs assistants: total number of assistants, number of full-time, part-time, or volunteer assistants. For some observations, these data are unreliable - the total number of assistants is not equal to the sum of specific types of assistants. As a result, we decided to keep only a total number of assistants.
- Committee position: similarly to faction position variable, we used ‘chairman’ in all titles instead of using both ‘chairman’ and ‘head.’

## 1.3 Rada 8

Data on eighth Rada were collected before the convocation termination. As a result there were some inconsistencies in the faction affiliations. We checked all the MPs and made changes were necessary.

## 1.4 Radas 7, 6, 3

Files of sixth and seventh Radas had completely missing faction affiliation. Moreover party affiliation was indicated instead of faction affiliation. File for the third Rada had faction affiliation for only 11% of observation. We added faction affiliations to these files and saved them as ‘mps\_3\_revised.csv’, ‘mps\_6\_revised.csv’ and ‘mps\_7\_revised.csv.’

## 1.5 Rada 4

File of the fourth Rada had completely missing party affiliation. We added it and saved the new file as ‘mps\_4\_revised.csv.’

# 2 Cleaning data after merging all files

## 2.1 Unifying spelling of Parties and Factions across convocations

We substituted the names of all parties and factions so that the party/faction has the same name across all convocations. For example, ‘Block of Yulia Tymoshenko’ and ‘Yuliia Tymoshenko Block’ was changed to

‘Bloc Tymoshenko.’ The original and the new names for parties/factions are stored in files ‘values\_party.csv’ and ‘values\_faction.csv.’

## 2.2 Unify transliterations of Ukrainian names

During the period of 1998-2019 there were several changes in the official methodology for transliterating the Ukrainian names to English. Also some people change the transliteration of their name between Ukrainian/Russian style of spelling. Consequently, 79 MPs had several versions of their English names and 7 MPs had several versions of their Ukrainian names. We substituted the duplicates with the most recent transliteration version. Original and substituted names are stored in the file ‘values\_name\_eng.csv’ and ‘values\_name\_ukr.csv.’

## 2.3 Other adjustments

- We added a unique ID one for each MP. We also kept the old, non-unique IDs.
- All date variables were harmonized to have format ‘year-month-day.’
- The case of father and son with identical names (Omelchenko Oleksandr Oleksandrovych) was resolved by adding ‘(son)’ to the sons name.
- The “ ” symbol was removed from English names to avoid possible processing problems.

Merged database was saved in both csv and rds formats. All the code related to cleaning and merging is stored in the file ‘mps\_merging\_code.Rmd.’