Cleaning the names of meeting attendees

Why do we need to do it manually?

It turns out that many MEPs have very inconsistently reported the attendees of their meetings. For example, a firm may have been reported in Hungarian, Italian, English or with an abbreviation. This leads to differences in the spellings of each name, even making some names unrecognisable. This makes it difficult for us to uncover patterns behind MEP meetings. Therefore, we have to standardise the names of each attendee if we want to properly analyse their lobbying activity. Unfortunately automated tools are not very good at finding and fixing all these different spellings and translated versions (although you're welcome to experiment with AI tools as long as you carefully check their output).

Going through the list of attendees manually also allows us to classify them into groups. This is not easy to automate as it requires an understanding of what the attendee does and how it operates.

Steps for the classification of attendees in the files

1. Download the attendee list for your chosen dossier from this folder.

2. Create 4 columns next to the attendees column:
    - fixed_names: standardised/fixed spelling of attendees
    - class: "NGO", "business", "government", "academia" or "other"
    - structure: "single actor" or "umbrella group"
    - comment: remarks for groupmates or future notes

3. Copy the contents of the attendees column into the fixed_names column.
   Before you can begin correcting the spelling of the entries, read the following notes:

    - **DO NOT** change entries in the attendees column. We need this column to stay intact if we want to successfully merge the correct spellings into the larger data later on. You can standardise the spelling by changing the entrees in the fixed_name column.
    - Some entries are very similar. Standardising their spelling works best if you replace the varying names in the fixed_name column by copy+pasting the most suitable name into the other rows.
    - It is preferable to have attendee names that include a transparency register number, if listed.
    - Be aware that some attendees have different names in different languages. Thus, you will have to do some research to discover which organisation they are part of. The same goes for attendee entries that solely consist of a number.

- The process works the fastest if you also fill in the other columns (class, structure, comment) as you fix the spelling of each attendee.

4. In order to overcome language differences and discover the class/structure of attendees, you will need to use a variety of tools. Here's a list with some suggestions:
   - Google translate.
   - Transparency Register: Zoeken in het register - Europese Unie (europa.eu)
   - Lobbyfacts: Search | lobbyfacts
   - Websites of the attendees.
   - ChatGPT or AI tools can be useful if you are really unsure about the class or structure of a given attendee. However, make sure it identifies the same attendee as the one in the data.

   Below is an example of what a finished table may look like:

	A	B	C	D	E
1	**attendees**	**fixed_names**	**class**	**structure**	**comment**
2	Amnesty International Limited	Amnesty International Limited	NGO	Single_actor	
3	APEAL	APEAL - Transparency Register Number: 700369	Business	Umbrella_group	
4	Apeal	APEAL - Transparency Register Number: 700369	Business	Umbrella_group	
5	APEAL - Transparency Register Number: 70036906501-10	APEAL - Transparency Register Number: 700369	Business	Umbrella_group	
6	Human Rights Watch	Human Rights Watch	NGO	Single_actor	
7	International Commission of Jurists	International Commission of Jurists	NGO	Single_actor	
8	McDonald's	McDonald's	Business	Single_actor	
9	US Ambassador	US Ambassador	Government	Single_actor	
10	Vinces Consulting	Vinces Consulting	Other	Single_actor	Consultant
11					

5. **Do not forget** to repeatedly save the file. Use a clear name, like 2018_0902_NLE_clean.xslx (replace with your own dossier's reference code).

6. After the table has been filled in, we can then import it into R and join it with our meeting data file.