

Exercice pratique

April 23, 2020

1 Sujet 5: Analyse des dialogues dans l'Avare de Molière

J'ai choisi de partir du format markdown. Je commence par télécharger le fichier.

```
[1]: import os
import urllib.request
data_url = "http://dramacode.github.io/markdown/moliere_avare.txt"
data_file = "moliere_avare.txt"
if not os.path.exists(data_file):
    urllib.request.urlretrieve(data_url, data_file)
    print("Download from http://dramacode.github.io/markdown/moliere_avare.txt")
```

Je voudrais faire un dataframe contenant pour chaque scène les interventions de chaque personnage.

```
[2]: import pandas as pd
```

Je commence par lire le fichier markdown.

```
[3]: text=open(data_file,'r')
Lines = text.readlines()
```

Je parse le texte de façon à récupérer le nom des personnages.

```
[4]: Perso = []
for l in Lines:
    if l[0:4]==' ':
        name=l[4:].split('.')[0].split('*')[0].split(',')[0]
        if name not in Perso and name[-1]!=' ':
            Perso.append(name)
Perso
```

```
[4]: ['VALÈRE',
'ÉLISE',
'CLÉANTE',
'HARPAGON',
'LA FLÈCHE',
'MAÎTRE SIMON',
'FROSINE',
```

```
'MAÎTRE JACQUES',  
'LA MERLUCHE',  
'BRINDAVOINE',  
'MARIANE',  
'LE COMMISSAIRE',  
'ANSELME']
```

Je récupère de même les actes et les scènes:

```
[5]: Scenes=[]  
for l in Lines:  
    if l[0:2]=='##':  
        if l[2]!='#':  
            Acte=l.split(' ')[2].split('.')[0]  
        else :  
            Scene=Acte+'_'+l.split(' ')[2].split('.')[0]  
            Scenes.append(Scene)  
  
Scenes
```

```
[5]: ['Premier_Première',  
      'Premier_II',  
      'Premier_III',  
      'Premier_IV',  
      'Premier_V',  
      'II_Première',  
      'II_II',  
      'II_III',  
      'II_IV',  
      'II_V',  
      'III_Première',  
      'III_II',  
      'III_III',  
      'III_IV',  
      'III_V',  
      'III_VI',  
      'III_VII',  
      'III_VIII',  
      'III_IX',  
      'IV_Première',  
      'IV_II',  
      'IV_III',  
      'IV_IV',  
      'IV_V',  
      'IV_VI',  
      'IV_VII',  
      'V_Première',  
      'V_II',
```

```
'V_III',
'V_IV',
'V_V',
'V_VI']
```

Après avoir vérifié que je récupère bien tous les personnages et toutes les scènes, je construis un dataframe qui contient les répliques de chaque personnage dans chaque scène.

```
[6]: df=pd.DataFrame()
Scene=''
for l in Lines:
    if l[0:2]=='##':
        if l[2]!='#':
            Acte=l.split(' ')[2].split('.')[0]
        else :
            Scene=Acte+'_'+l.split(' ')[2].split('.')[0]
            new_row = pd.Series(data={}, name=Scene)
            #append the row to the dataframe
            df = df.append(new_row, ignore_index=False)
            Name=''
    elif l[0:4]==' ':
        Name=l[4:].split('.')[0].split('*')[0].split(',') [0]
        if Name[-1]==' ':
            Name=Name[:-1]
        if Name not in df.columns:
            df[Name]=None
        if isinstance(df[Name][Scene],list)==False: df[Name][Scene]=[]

    elif Scene!='' and Name!='' and l[0] != '\n':
        df[Name][Scene].append(l.split('\n')[0])
```

Je vérifie une entrée:

```
[7]: df['ÉLISE']['Premier_II']
```

```
[7]: ["Me voilà prête à vous ouïr, mon Frère. Qu'avez-vous à me dire ?",
'Vous aimez ?',
'Vous êtes-vous engagé, mon Frère, avec celle que vous aimez ?',
'Suis-je, mon Frère, une si étrange personne ?',
"Hélas ! mon Frère, ne parlons point de ma sagesse. Il n'est personne qui n'en
manque du moins une fois en sa vie ; et si je vous ouvre mon cœur, peut-être
serai-je à vos yeux bien moins sage que vous.",
'Finissons auparavant votre affaire, et me dites qui est celle que vous
aimez.',
"J'en vois beaucoup, mon Frère, dans les choses que vous me dites ; et pour
comprendre ce qu'elle est, il me suffit que vous l'aimiez.",
```

'Oui, je conçois assez, mon Frère, quel doit être votre chagrin.',
 'Il est bien vrai que tous les jours il nous donne, de plus en plus, sujet de
 regretter la mort de notre Mère, et que...']

```
[8]: print(len(df['ÉLISE']['Premier_II']))
```

9

Il y a 9 répliques d'Élise dans l'acte I scène 2.

```
[9]: len(df['ÉLISE']['Premier_II'][4].split(' '))
```

[9]: 41

Il y a 41 mots dans la 5ème réplique (en coupant simplement aux espaces).
 J'ai à présent un objet que je peux utiliser pour répondre aux questions posées.

2 Question 1

Classez les personnages selon la quantité de parole grâce à une analyse syntaxique du texte (scènes / répliques / mots). En particulier, quel est celui qui parle le plus ? Quel est celui qui ne parle pas du tout ? Attention, les noms des personnages ne sont pas forcément homogènes (casse et accents par exemple).

```
[10]: df
```

```
[10]:
```

		VALÈRE	\
Premier_Première	[Hé quoi, charmante Élise, vous devenez mélanc...		
Premier_II		NaN	
Premier_III		NaN	
Premier_IV		NaN	
Premier_V	[C'est vous, Monsieur, sans contredit., Non. M...		
II_Première		NaN	
II_II		NaN	
II_III		NaN	
II_IV		NaN	
II_V		NaN	
III_Première	[Je n'ai jamais vu de réponse plus impertinent...		
III_II	[À ce que je puis voir, Maître Jacques, on pay...		
III_III		NaN	
III_IV		NaN	
III_V		NaN	
III_VI		NaN	
III_VII	[Il a perdu le sens.]		
III_VIII		NaN	
III_IX	[Cela ne sera rien., C'est assez.]		
IV_Première		NaN	

IV_II		NaN
IV_III		NaN
IV_IV		NaN
IV_V		NaN
IV_VI		NaN
IV_VII		NaN
V_Première		NaN
V_II		NaN
V_III	[Que voulez-vous, Monsieur ?, De quel crime vo...	
V_IV	[Ce ne sera point votre passion qui jugera l'a...	
V_V	[Qui songe à votre argent, dont vous me faites...	
V_VI		NaN

ÉLISE \

Premier_Première	[Non, Valère, je ne puis pas me repentir de to...	
Premier_II	[Me voilà prête à vous ouïr, mon Frère. Qu'ave...	
Premier_III		NaN
Premier_IV	[Nous ne venons que d'arriver., Quoi ?, Pardon...	
Premier_V	[Vous moquez-vous, Valère, de lui parler comme...	
II_Première		NaN
II_II		NaN
II_III		NaN
II_IV		NaN
II_V		NaN
III_Première	[Oui, mon Père.]	
III_II		NaN
III_III		NaN
III_IV		NaN
III_V		NaN
III_VI	[Vous avez fait, Madame, ce que je devais fair...	
III_VII		NaN
III_VIII		NaN
III_IX		NaN
IV_Première	[Oui, Madame, mon Frère m'a fait confidence de...	
IV_II	[Voilà mon Père.]	
IV_III		NaN
IV_IV		NaN
IV_V		NaN
IV_VI		NaN
IV_VII		NaN
V_Première		NaN
V_II		NaN
V_III		NaN
V_IV	[Ah ! mon père, prenez des sentiments un peu p...	
V_V		NaN
V_VI		NaN

	CLÉANTE \
Premier_Première	None
Premier_II	[Je suis bien aise de vous trouver seule, ma S...
Premier_III	NaN
Premier_IV	[Rien, mon Père., Quoi ? mon Père., Non., Nous...
Premier_V	NaN
II_Première	[Ah ! traître que tu es, où t'es-tu donc allé ...
II_II	[Lui aurait-on appris qui je suis ? et serais-...
II_III	NaN
II_IV	NaN
II_V	NaN
III_Première	[Moi, mon Père, mauvais visage ; et par quelle...
III_II	NaN
III_III	NaN
III_IV	NaN
III_V	NaN
III_VI	NaN
III_VII	[Madame, à vous dire le vrai, c'est ici une av...
III_VIII	NaN
III_IX	[Qu'est-ce, mon Père ? Vous êtes-vous fait mal...
IV_Première	[Rentrons ici, nous serons beaucoup mieux. Il n...
IV_II	[Puisque vous n'y allez pas, mon Père, je m'en...
IV_III	[Ce qui m'en semble ?, Là, là., À vous en parl...
IV_IV	[Je me moque de cela., Je n'en démordrai point...
IV_V	[Je vous demande pardon, mon Père, de l'emport...
IV_VI	[Qu'y a-t-il ?, Comment ?, Quoi ?, Qu'est-ce q...
IV_VII	NaN
V_Première	NaN
V_II	NaN
V_III	NaN
V_IV	NaN
V_V	NaN
V_VI	[Ne vous tourmentez point, mon Père, et n'accu...

	HARPAGON \
Premier_Première	None
Premier_II	None
Premier_III	[Hors d'ici tout à l'heure, et qu'on ne répliq...
Premier_IV	[Certes, ce n'est pas une petite peine que de ...
Premier_V	[Ici, Valère. Nous t'avons élu pour nous dire ...
II_Première	NaN
II_II	[Mais croyez-vous, Maître Simon, qu'il n'y ait...
II_III	[Attendez un moment. Je vais revenir vous parl...
II_IV	NaN
II_V	[Tout va comme il faut. Hé bien, qu'est-ce, Fr...
III_Première	[Allons.Venez çà tous, que je vous distribue m...
III_II	NaN

III_III		NaN
III_IV		NaN
III_V	[Ne vous offensez pas, ma Belle, si je viens à...	
III_VI	[Vous voyez qu'elle est grande ; mais mauvaise...	
III_VII	[Voilà un compliment bien impertinent. Quelle ...	
III_VIII	[Dis-lui que je suis empêché, et qu'il revienne...	
III_IX	[Ah, je suis mort !, Le traître assurément a r...	
IV_Première		NaN
IV_II	[Ouais ! mon Fils baise la main de sa prétendu...	
IV_III	[Ô ça, intérêt de Belle-Mère à part, que te se...	
IV_IV	[Me parler avec cette impudence !, Laisse-moi ...	
IV_V	[Cela n'est rien., Et moi, j'ai toutes les joi...	
IV_VI		NaN
IV_VII	[Il crie au Voleur dès le jardin, et vient san...	
V_Première	[Tous les Magistrats sont intéressés à prendre...	
V_II	[Qui ? celui qui m'a dérobé ?, Il n'est pas qu...	
V_III	[Approche. Viens confesser l'action la plus no...	
V_IV	[Ah ! fille scélérate ! Fille indigne d'un Père...	
V_V	[Ah ! Seigneur Anselme, vous me voyez le plus ...	
V_VI	[Où est-il ?, N'en a-t-on rien ôté ?, Il faut,...	

LA FLÈCHE \

Premier_Première		None
Premier_II		None
Premier_III	[Je n'ai jamais rien vu de si méchant que ce m...	
Premier_IV		NaN
Premier_V		NaN
II_Première	[Oui, Monsieur, et je m'étais rendu ici pour v...	
II_II	[Que veut dire ceci ? Notre maître Simon qui p...	
II_III		NaN
II_IV	[L'Aventure est tout à fait drôle Il faut bien...	
II_V		NaN
III_Première		NaN
III_II		NaN
III_III		NaN
III_IV		NaN
III_V		NaN
III_VI		NaN
III_VII		NaN
III_VIII		NaN
III_IX		NaN
IV_Première		NaN
IV_II		NaN
IV_III		NaN
IV_IV		NaN
IV_V		NaN
IV_VI	[Ah, Monsieur, que je vous trouve à propos !Su...	

IV_VII	NaN
V_Première	NaN
V_II	NaN
V_III	NaN
V_IV	NaN
V_V	NaN
V_VI	NaN

	MAÎTRE SIMON \
Premier_Première	None
Premier_II	None
Premier_III	None
Premier_IV	None
Premier_V	None
II_Première	None
II_II	[Oui, Monsieur, c'est un jeune Homme qui a bes...
II_III	NaN
II_IV	NaN
II_V	NaN
III_Première	NaN
III_II	NaN
III_III	NaN
III_IV	NaN
III_V	NaN
III_VI	NaN
III_VII	NaN
III_VIII	NaN
III_IX	NaN
IV_Première	NaN
IV_II	NaN
IV_III	NaN
IV_IV	NaN
IV_V	NaN
IV_VI	NaN
IV_VII	NaN
V_Première	NaN
V_II	NaN
V_III	NaN
V_IV	NaN
V_V	NaN
V_VI	NaN

	FROSINE \
Premier_Première	None
Premier_II	None
Premier_III	None
Premier_IV	None

Premier_V		None
II_Première		None
II_II		None
II_III		[Monsieur...]
II_IV	[Hé c'est toi, mon pauvre la Flèche ! D'où vie...	
II_V	[Ah, mon Dieu ! que vous vous portez bien ! et...	
III_Première		NaN
III_II		NaN
III_III	[Savez-vous, Maître Jacques, si votre maître e...	
III_IV	[Mais pourquoi, et quelle est votre inquiétude...	
III_V	[C'est qu'elle est encore toute surprise ; et ...	
III_VI	[Qu'elle vous trouve admirable., L'aventure es...	
III_VII	[Non, il vaut mieux que de ce pas nous allions...	
III_VIII		NaN
III_IX		NaN
IV_Première	[Vous êtes, par ma foi, de malheureuses gens l...	
IV_II		NaN
IV_III		NaN
IV_IV		NaN
IV_V		NaN
IV_VI		NaN
IV_VII		NaN
V_Première		NaN
V_II		NaN
V_III		NaN
V_IV	[Voici un étrange embarras.]	
V_V		NaN
V_VI		NaN

MAÎTRE JACQUES \

Premier_Première		None
Premier_II		None
Premier_III		None
Premier_IV		None
Premier_V		None
II_Première		None
II_II		None
II_III		None
II_IV		None
II_V		None
III_Première	[Châtiment politique., Oui ; le vin pur monte ...	
III_II	[Morbleu, Monsieur le nouveau venu, qui faites...	
III_III	[Oui vraiment il y est, je ne le sais que trop.]	
III_IV		NaN
III_V		NaN
III_VI		NaN
III_VII		NaN

III_VIII		NaN
III_IX		NaN
IV_Première		NaN
IV_II		NaN
IV_III		NaN
IV_IV	[Eh, eh, eh, Messieurs, qu'est-ce ci ? à quoi ...	
IV_V		NaN
IV_VI		NaN
IV_VII		NaN
V_Première		NaN
V_II	[Je m'en vais revenir. Qu'on me l'égorge tout ...	
V_III	[Oh, oh. Aurais-je deviné sans y penser ?, Écr...	
V_IV	[Tu me payeras mes coups de bâton.]	
V_V	[Vous voyez que je ne dis rien.]	
V_VI	[Hélas ! comment faut-il donc faire ? On me do...	

	LA MERLUCHE \	
Premier_Première		None
Premier_II		None
Premier_III		None
Premier_IV		None
Premier_V		None
II_Première		None
II_II		None
II_III		None
II_IV		None
II_V		None
III_Première	[Quitterons-nous nos siquenilles, Monsieur ?, ...	
III_II		NaN
III_III		NaN
III_IV		NaN
III_V		NaN
III_VI		NaN
III_VII		NaN
III_VIII		NaN
III_IX	[Il vient en courant, et fait tomber Harpagon...	
IV_Première		NaN
IV_II		NaN
IV_III		NaN
IV_IV		NaN
IV_V		NaN
IV_VI		NaN
IV_VII		NaN
V_Première		NaN
V_II		NaN
V_III		NaN
V_IV		NaN

V_V	NaN
V_VI	NaN

	BRINDAVOINE \
Premier_Première	None
Premier_II	None
Premier_III	None
Premier_IV	None
Premier_V	None
II_Première	None
II_II	None
II_III	None
II_IV	None
II_V	None
III_Première	[Vous savez bien, Monsieur, qu'un des devants ...
III_II	NaN
III_III	NaN
III_IV	NaN
III_V	NaN
III_VI	NaN
III_VII	NaN
III_VIII	[Monsieur, il y a là un Homme qui veut vous pa...
III_IX	NaN
IV_Première	NaN
IV_II	NaN
IV_III	NaN
IV_IV	NaN
IV_V	NaN
IV_VI	NaN
IV_VII	NaN
V_Première	NaN
V_II	NaN
V_III	NaN
V_IV	NaN
V_V	NaN
V_VI	NaN

	MARIANE \
Premier_Première	None
Premier_II	None
Premier_III	None
Premier_IV	None
Premier_V	None
II_Première	None
II_II	None
II_III	None
II_IV	None

II_V		None
III_Première		None
III_II		None
III_III		None
III_IV	[Ah ! que je suis, Frosine, dans un étrange ét...	
III_V		NaN
III_VI	[Je m'acquitte bien tard, Madame, d'une telle ...	
III_VII	[Je puis dire la même chose. C'est une rencont...	
III_VIII		NaN
III_IX		NaN
IV_Première	[C'est une douce consolation, que de voir dans...	
IV_II		NaN
IV_III		NaN
IV_IV		NaN
IV_V		NaN
IV_VI		NaN
IV_VII		NaN
V_Première		NaN
V_II		NaN
V_III		NaN
V_IV		NaN
V_V	[Hélas ! à vos paroles, je puis ici répondre, ...	
V_VI	[Mais vous ne savez pas, que ce n'est pas asse...	

	LE COMMISSAIRE \
Premier_Première	None
Premier_II	None
Premier_III	None
Premier_IV	None
Premier_V	None
II_Première	None
II_II	None
II_III	None
II_IV	None
II_V	None
III_Première	None
III_II	None
III_III	None
III_IV	None
III_V	None
III_VI	None
III_VII	None
III_VIII	None
III_IX	None
IV_Première	None
IV_II	None
IV_III	None

IV_IV		None
IV_V		None
IV_VI		None
IV_VII		None
V_Première	[Laissez-moi faire. Je sais mon métier, Dieu m...	
V_II	[Ne vous épouvantez point. Je suis Homme à ne ...	
V_III		NaN
V_IV		NaN
V_V		NaN
V_VI	[Holà, Messieurs, holà.Tout doucement, s'il vo...	

		ANSELME
Premier_Première		None
Premier_II		None
Premier_III		None
Premier_IV		None
Premier_V		None
II_Première		None
II_II		None
II_III		None
II_IV		None
II_V		None
III_Première		None
III_II		None
III_III		None
III_IV		None
III_V		None
III_VI		None
III_VII		None
III_VIII		None
III_IX		None
IV_Première		None
IV_II		None
IV_III		None
IV_IV		None
IV_V		None
IV_VI		None
IV_VII		None
V_Première		None
V_II		None
V_III		None
V_IV		None
V_V	[Qu'est-ce, Seigneur Harpagon, je vous vois to...	
V_VI	[Le Ciel, mes Enfants, ne me redonne point à v...	

Je génère un dataframe NW qui contient le nombre de mots de chaque personnage dans chaque scène.

```
[11]: def NWords(Name, Scene):
        N=0
        if isinstance(df[Name][Scene],list):
            for i in range(0,len(df[Name][Scene])):
                N+=len(df[Name][Scene][i].split(' '))
        return N

        NWords('VALÈRE','Premier_Première')
```

[11]: 593

```
[12]: Perso = df.columns
        NW=pd.DataFrame(index=df.index)
        for Name in Perso:
            NW[Name]=[NWords(Name,Scene) for Scene in df.index]
        NW
```

```
[12]:
```

	VALÈRE	ÉLISE	CLÉANTE	HARPAGON	LA FLÈCHE	MAÎTRE SIMON	\
Premier_Première	593	463	0	0	0	0	
Premier_II	0	150	725	0	0	0	
Premier_III	0	0	0	459	249	0	
Premier_IV	0	158	211	1094	0	0	
Premier_V	666	36	0	268	0	0	
II_Première	0	0	350	0	852	0	
II_II	0	0	126	159	13	186	
II_III	0	0	0	21	0	0	
II_IV	0	0	0	0	281	0	
II_V	0	0	0	520	0	0	
III_Première	245	3	74	729	0	0	
III_II	107	0	0	0	0	0	
III_III	0	0	0	0	0	0	
III_IV	0	0	0	0	0	0	
III_V	0	0	0	100	0	0	
III_VI	0	16	0	66	0	0	
III_VII	5	0	565	172	0	0	
III_VIII	0	0	0	20	0	0	
III_IX	6	0	38	75	0	0	
IV_Première	0	54	229	0	0	0	
IV_II	0	3	12	49	0	0	
IV_III	0	0	383	364	0	0	
IV_IV	0	0	158	148	0	0	
IV_V	0	0	157	120	0	0	
IV_VI	0	0	15	0	40	0	
IV_VII	0	0	0	392	0	0	
V_Première	0	0	0	84	0	0	
V_II	0	0	0	168	0	0	
V_III	600	0	0	429	0	0	

V_IV	19	139	0	120	0	0
V_V	358	0	0	240	0	0
V_VI	0	0	125	87	0	0

	FROSINE	MAÎTRE JACQUES	LA MERLUCHE	BRINDAVOINE	MARIANE \
Premier_Première	0	0	0	0	0
Premier_II	0	0	0	0	0
Premier_III	0	0	0	0	0
Premier_IV	0	0	0	0	0
Premier_V	0	0	0	0	0
II_Première	0	0	0	0	0
II_II	0	0	0	0	0
II_III	1	0	0	0	0
II_IV	116	0	0	0	0
II_V	1448	0	0	0	0
III_Première	0	734	22	20	0
III_II	0	184	0	0	0
III_III	18	11	0	0	0
III_IV	182	0	0	0	173
III_V	21	0	0	0	0
III_VI	7	0	0	0	32
III_VII	40	0	0	0	215
III_VIII	0	0	0	18	0
III_IX	0	0	28	0	0
IV_Première	395	0	0	0	218
IV_II	0	0	0	0	0
IV_III	0	0	0	0	0
IV_IV	0	292	0	0	0
IV_V	0	0	0	0	0
IV_VI	0	0	0	0	0
IV_VII	0	0	0	0	0
V_Première	0	0	0	0	0
V_II	0	317	0	0	0
V_III	0	11	0	0	0
V_IV	4	7	0	0	0
V_V	0	7	0	0	185
V_VI	0	24	0	0	34

	LE COMMISSAIRE	ANSELME
Premier_Première	0	0
Premier_II	0	0
Premier_III	0	0
Premier_IV	0	0
Premier_V	0	0
II_Première	0	0
II_II	0	0
II_III	0	0

II_IV	0	0
II_V	0	0
III_Première	0	0
III_II	0	0
III_III	0	0
III_IV	0	0
III_V	0	0
III_VI	0	0
III_VII	0	0
III_VIII	0	0
III_IX	0	0
IV_Première	0	0
IV_II	0	0
IV_III	0	0
IV_IV	0	0
IV_V	0	0
IV_VI	0	0
IV_VII	0	0
V_Première	103	0
V_II	151	0
V_III	0	0
V_IV	0	0
V_V	0	378
V_VI	23	105

Je génère un dataframe NS qui contient le nombre de répliques de chaque personnage dans chaque scène.

```
[13]: def Nsentences(Name, Scene):
      N=0
      if isinstance(df[Name][Scene],list):
          N=len(df[Name][Scene])
      return N
```

```
[14]: NS=pd.DataFrame(index=df.index)
      for Name in Perso:
          NS[Name]=[Nsentences(Name,Scene) for Scene in df.index]
      NS
```

```
[14]:
```

	VALÈRE	ÉLISE	CLÉANTE	HARPAGON	LA FLÈCHE	MAÎTRE SIMON	\
Premier_Première	8	8	0	0	0	0	
Premier_II	0	9	10	0	0	0	
Premier_III	0	0	0	42	33	0	
Premier_IV	0	24	29	56	0	0	
Premier_V	24	4	0	22	0	0	
II_Première	0	0	21	0	20	0	
II_II	0	0	6	9	1	5	

II_III	0	0	0	1	0	0
II_IV	0	0	0	0	6	0
II_V	0	0	0	35	0	0
III_Première	11	1	3	40	0	0
III_II	14	0	0	0	0	0
III_III	0	0	0	0	0	0
III_IV	0	0	0	0	0	0
III_V	0	0	0	2	0	0
III_VI	0	1	0	6	0	0
III_VII	1	0	23	17	0	0
III_VIII	0	0	0	2	0	0
III_IX	2	0	2	6	0	0
IV_Première	0	2	10	0	0	0
IV_II	0	1	1	3	0	0
IV_III	0	0	22	23	0	0
IV_IV	0	0	8	9	0	0
IV_V	0	0	19	18	0	0
IV_VI	0	0	5	0	6	0
IV_VII	0	0	0	5	0	0
V_Première	0	0	0	6	0	0
V_II	0	0	0	19	0	0
V_III	30	0	0	30	0	0
V_IV	1	2	0	4	0	0
V_V	15	0	0	11	0	0
V_VI	0	0	4	10	0	0

	FROSINE	MAÎTRE JACQUES	LA MERLUCHE	BRINDAVOINE	MARIANE \
Premier_Première	0	0	0	0	0
Premier_II	0	0	0	0	0
Premier_III	0	0	0	0	0
Premier_IV	0	0	0	0	0
Premier_V	0	0	0	0	0
II_Première	0	0	0	0	0
II_II	0	0	0	0	0
II_III	1	0	0	0	0
II_IV	5	0	0	0	0
II_V	44	0	0	0	0
III_Première	0	30	2	1	0
III_II	0	12	0	0	0
III_III	2	1	0	0	0
III_IV	5	0	0	0	6
III_V	1	0	0	0	0
III_VI	2	0	0	0	5
III_VII	2	0	0	0	10
III_VIII	0	0	0	2	0
III_IX	0	0	4	0	0
IV_Première	6	0	0	0	6

IV_II	0	0	0	0	0
IV_III	0	0	0	0	0
IV_IV	0	21	0	0	0
IV_V	0	0	0	0	0
IV_VI	0	0	0	0	0
IV_VII	0	0	0	0	0
V_Première	0	0	0	0	0
V_II	0	24	0	0	0
V_III	0	2	0	0	0
V_IV	1	1	0	0	0
V_V	0	1	0	0	3
V_VI	0	1	0	0	1

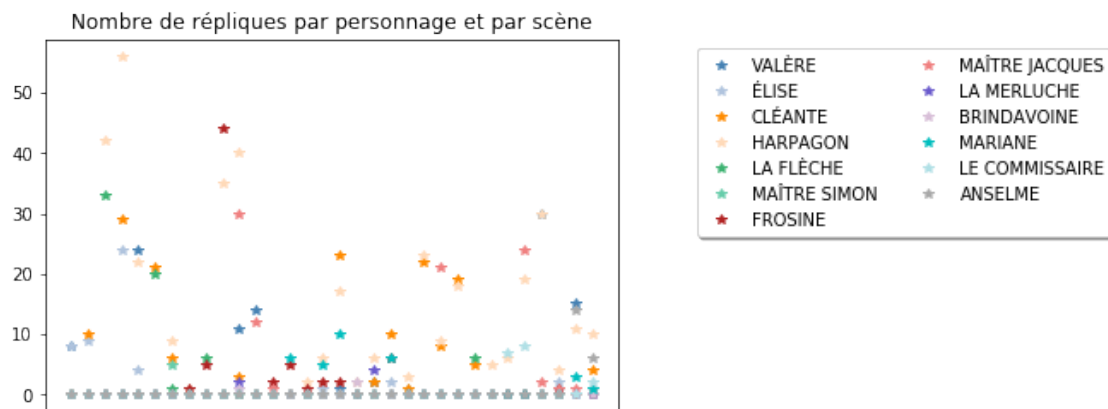
	LE COMMISSAIRE	ANSELME
Premier_Première	0	0
Premier_II	0	0
Premier_III	0	0
Premier_IV	0	0
Premier_V	0	0
II_Première	0	0
II_II	0	0
II_III	0	0
II_IV	0	0
II_V	0	0
III_Première	0	0
III_II	0	0
III_III	0	0
III_IV	0	0
III_V	0	0
III_VI	0	0
III_VII	0	0
III_VIII	0	0
III_IX	0	0
IV_Première	0	0
IV_II	0	0
IV_III	0	0
IV_IV	0	0
IV_V	0	0
IV_VI	0	0
IV_VII	0	0
V_Première	7	0
V_II	8	0
V_III	0	0
V_IV	0	0
V_V	0	14
V_VI	2	6

Je fais des graphs simples pour visualiser rapidement les données:

```
[15]: %matplotlib inline
import matplotlib.pyplot as plt
fig = plt.figure()
colors=['steelblue','lightsteelblue','darkorange','peachpuff','mediumseagreen','mediumaquamarine']
NS.plot(style='*', color=colors)
plt.legend(loc='upper center', bbox_to_anchor=(1.5, 1.0), shadow=True, ncol=2)
plt.title('Nombre de répliques par personnage et par scène')
```

```
[15]: Text(0.5,1,'Nombre de répliques par personnage et par scène')
```

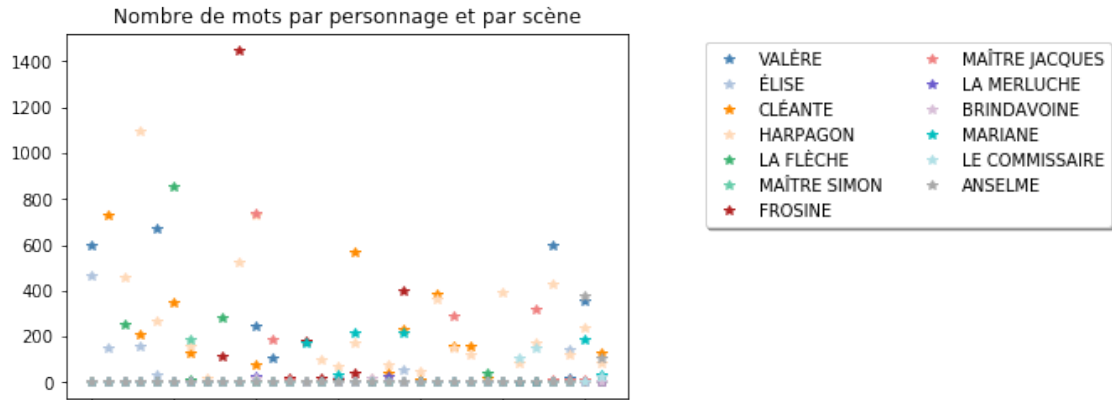
<Figure size 432x288 with 0 Axes>



```
[16]: fig = plt.figure()
NW.plot(style='*', color=colors)
plt.legend(loc='upper center', bbox_to_anchor=(1.5, 1.0), shadow=True, ncol=2)
plt.title('Nombre de mots par personnage et par scène')
```

```
[16]: Text(0.5,1,'Nombre de mots par personnage et par scène')
```

<Figure size 432x288 with 0 Axes>



Je calcule le nombre total de répliques et de mots prononcés par chaque personnage.

```
[17]: NW.sum(axis=0)
```

```
[17]: VALÈRE          2599
      ÉLISE          1022
      CLÉANTE        3168
      HARPAGON        5884
      LA FLÈCHE       1435
      MAÎTRE SIMON     186
      FROSINE         2232
      MAÎTRE JACQUES  1587
      LA MERLUCHE       50
      BRINDAVOINE       38
      MARIANE          857
      LE COMMISSAIRE    277
      ANSELME          483
      dtype: int64
```

Du plus silencieux au plus bavard:

```
[18]: NW.sum(axis=0).sort_values()
```

```
[18]: BRINDAVOINE       38
      LA MERLUCHE       50
      MAÎTRE SIMON     186
      LE COMMISSAIRE    277
      ANSELME          483
      MARIANE          857
      ÉLISE          1022
      LA FLÈCHE       1435
      MAÎTRE JACQUES  1587
      FROSINE         2232
```

VALÈRE	2599
CLÉANTE	3168
HARPAGON	5884

dtype: int64

C'est donc Harpagon qui prononce le plus de mots, et Brindavoine qui en prononce le moins. Je fais pareil avec les répliques:

```
[19]: NS.sum(axis=0).sort_values()
```

```
[19]: BRINDAVOINE      3
      MAÎTRE SIMON    5
      LA MERLUCHE     6
      LE COMMISSAIRE  17
      ANSELME         20
      MARIANE         31
      ÉLISE          52
      LA FLÈCHE       66
      FROSINE        69
      MAÎTRE JACQUES  93
      VALÈRE         106
      CLÉANTE        163
      HARPAGON       376
      dtype: int64
```

De façon rassurante, en ne comptant que les répliques, on tombe sur un ordre très similaire. J'ai vérifié manuellement que Brindavoine n'a bien que 3 répliques, et Maître Simon 5, c'est bon.

3 Question 2

Réalisez un graphique qui montrera le nombre de mots que chaque acteur prononce dans chaque scène. Pour cela, vous pouvez vous inspirer de l'étude de l'Avare de Molière réalisée par l'OBVIL (graphe de gauche). Dans ce graphique, les lignes sont de longueur égale et la hauteur représente le nombre de mots prononcés au total dans la scène. La largeur de chaque rectangle indique le pourcentage de la scène qu'un acteur occupe.

Pour calculer les hauteurs de chaque rectangle:

```
[20]: Height = NW.sum(axis=1)
      Height
```

```
[20]: Premier_Première  1056
      Premier_II      875
      Premier_III     708
      Premier_IV     1463
      Premier_V       970
      II_Première    1202
```

```

II_II          484
II_III         22
II_IV         397
II_V         1968
III_Première  1827
III_II        291
III_III        29
III_IV        355
III_V         121
III_VI        121
III_VII       997
III_VIII       38
III_IX        147
IV_Première   896
IV_II         64
IV_III        747
IV_IV         598
IV_V          277
IV_VI         55
IV_VII        392
V_Première    187
V_II          636
V_III        1040
V_IV          289
V_V          1168
V_VI          398
dtype: int64

```

Je transforme ensuite le nombre de mots en pourcentage:

```

[21]: NW_percent=NW.div(Height, axis=0)
      NW_percent

```

```

[21]:
Premier_Première  VALÈRE    ÉLISE    CLÉANTE  HARPAGON  LA FLÈCHE  \
Premier_II        0.000000  0.171429  0.828571  0.000000  0.000000
Premier_III       0.000000  0.000000  0.000000  0.648305  0.351695
Premier_IV        0.000000  0.107997  0.144224  0.747779  0.000000
Premier_V         0.686598  0.037113  0.000000  0.276289  0.000000
II_Première       0.000000  0.000000  0.291181  0.000000  0.708819
II_II            0.000000  0.000000  0.260331  0.328512  0.026860
II_III           0.000000  0.000000  0.000000  0.954545  0.000000
II_IV            0.000000  0.000000  0.000000  0.000000  0.707809
II_V             0.000000  0.000000  0.000000  0.264228  0.000000
III_Première      0.134100  0.001642  0.040504  0.399015  0.000000
III_II           0.367698  0.000000  0.000000  0.000000  0.000000
III_III          0.000000  0.000000  0.000000  0.000000  0.000000

```

III_IV	0.000000	0.000000	0.000000	0.000000	0.000000
III_V	0.000000	0.000000	0.000000	0.826446	0.000000
III_VI	0.000000	0.132231	0.000000	0.545455	0.000000
III_VII	0.005015	0.000000	0.566700	0.172518	0.000000
III_VIII	0.000000	0.000000	0.000000	0.526316	0.000000
III_IX	0.040816	0.000000	0.258503	0.510204	0.000000
IV_Première	0.000000	0.060268	0.255580	0.000000	0.000000
IV_II	0.000000	0.046875	0.187500	0.765625	0.000000
IV_III	0.000000	0.000000	0.512718	0.487282	0.000000
IV_IV	0.000000	0.000000	0.264214	0.247492	0.000000
IV_V	0.000000	0.000000	0.566787	0.433213	0.000000
IV_VI	0.000000	0.000000	0.272727	0.000000	0.727273
IV_VII	0.000000	0.000000	0.000000	1.000000	0.000000
V_Première	0.000000	0.000000	0.000000	0.449198	0.000000
V_II	0.000000	0.000000	0.000000	0.264151	0.000000
V_III	0.576923	0.000000	0.000000	0.412500	0.000000
V_IV	0.065744	0.480969	0.000000	0.415225	0.000000
V_V	0.306507	0.000000	0.000000	0.205479	0.000000
V_VI	0.000000	0.000000	0.314070	0.218593	0.000000

	MAÎTRE SIMON	FROSINE	MAÎTRE JACQUES	LA MERLUCHE \
Premier_Première	0.000000	0.000000	0.000000	0.000000
Premier_II	0.000000	0.000000	0.000000	0.000000
Premier_III	0.000000	0.000000	0.000000	0.000000
Premier_IV	0.000000	0.000000	0.000000	0.000000
Premier_V	0.000000	0.000000	0.000000	0.000000
II_Première	0.000000	0.000000	0.000000	0.000000
II_II	0.384298	0.000000	0.000000	0.000000
II_III	0.000000	0.045455	0.000000	0.000000
II_IV	0.000000	0.292191	0.000000	0.000000
II_V	0.000000	0.735772	0.000000	0.000000
III_Première	0.000000	0.000000	0.401752	0.012042
III_II	0.000000	0.000000	0.632302	0.000000
III_III	0.000000	0.620690	0.379310	0.000000
III_IV	0.000000	0.512676	0.000000	0.000000
III_V	0.000000	0.173554	0.000000	0.000000
III_VI	0.000000	0.057851	0.000000	0.000000
III_VII	0.000000	0.040120	0.000000	0.000000
III_VIII	0.000000	0.000000	0.000000	0.000000
III_IX	0.000000	0.000000	0.000000	0.190476
IV_Première	0.000000	0.440848	0.000000	0.000000
IV_II	0.000000	0.000000	0.000000	0.000000
IV_III	0.000000	0.000000	0.000000	0.000000
IV_IV	0.000000	0.000000	0.488294	0.000000
IV_V	0.000000	0.000000	0.000000	0.000000
IV_VI	0.000000	0.000000	0.000000	0.000000
IV_VII	0.000000	0.000000	0.000000	0.000000

V_Première	0.000000	0.000000	0.000000	0.000000
V_II	0.000000	0.000000	0.498428	0.000000
V_III	0.000000	0.000000	0.010577	0.000000
V_IV	0.000000	0.013841	0.024221	0.000000
V_V	0.000000	0.000000	0.005993	0.000000
V_VI	0.000000	0.000000	0.060302	0.000000

	BRINDAVOINE	MARIANE	LE COMMISSAIRE	ANSELME
Premier_Première	0.000000	0.000000	0.000000	0.000000
Premier_II	0.000000	0.000000	0.000000	0.000000
Premier_III	0.000000	0.000000	0.000000	0.000000
Premier_IV	0.000000	0.000000	0.000000	0.000000
Premier_V	0.000000	0.000000	0.000000	0.000000
II_Première	0.000000	0.000000	0.000000	0.000000
II_II	0.000000	0.000000	0.000000	0.000000
II_III	0.000000	0.000000	0.000000	0.000000
II_IV	0.000000	0.000000	0.000000	0.000000
II_V	0.000000	0.000000	0.000000	0.000000
III_Première	0.010947	0.000000	0.000000	0.000000
III_II	0.000000	0.000000	0.000000	0.000000
III_III	0.000000	0.000000	0.000000	0.000000
III_IV	0.000000	0.487324	0.000000	0.000000
III_V	0.000000	0.000000	0.000000	0.000000
III_VI	0.000000	0.264463	0.000000	0.000000
III_VII	0.000000	0.215647	0.000000	0.000000
III_VIII	0.473684	0.000000	0.000000	0.000000
III_IX	0.000000	0.000000	0.000000	0.000000
IV_Première	0.000000	0.243304	0.000000	0.000000
IV_II	0.000000	0.000000	0.000000	0.000000
IV_III	0.000000	0.000000	0.000000	0.000000
IV_IV	0.000000	0.000000	0.000000	0.000000
IV_V	0.000000	0.000000	0.000000	0.000000
IV_VI	0.000000	0.000000	0.000000	0.000000
IV_VII	0.000000	0.000000	0.000000	0.000000
V_Première	0.000000	0.000000	0.550802	0.000000
V_II	0.000000	0.000000	0.237421	0.000000
V_III	0.000000	0.000000	0.000000	0.000000
V_IV	0.000000	0.000000	0.000000	0.000000
V_V	0.000000	0.158390	0.000000	0.323630
V_VI	0.000000	0.085427	0.057789	0.263819

Je vais avoir besoin du nombre total de mots:

```
[22]: NWtot=Height.sum()
      NWtot
```

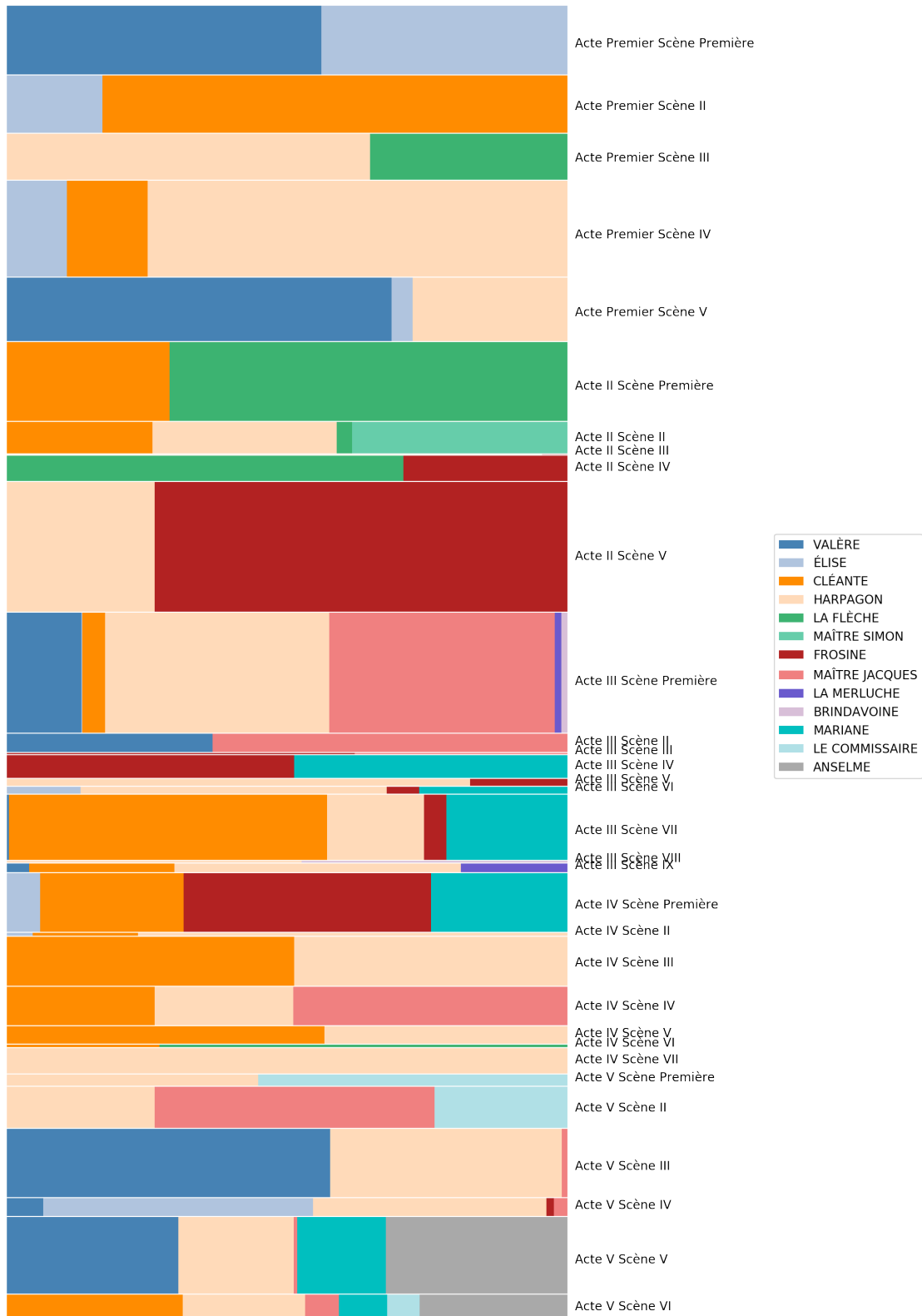
[22]: 19818

Il reste à faire une jolie figure avec tout ça...

```
[24]: previous=1
fig=plt.figure()
plt.rc('axes',edgecolor='w')
plt.rcParams['figure.figsize'] = [8, 15]
plt.rcParams['figure.dpi'] = 200
for scene in NW.index:
    h=Height[scene]/NWtot
    acte=scene.split('_')[0]
    sc=scene.split('_')[1]
    ax = fig.add_axes([0.05, previous-h, .8, h], xticks=[],yticks=[]) # je fais
    ↳un subplot / scène et j'enlève les ticks
    fig.text(0.86, previous-h*0.6, 'Acte '+acte+' Scène '+sc,fontsize=10) # je
    ↳mets le nom de la scène à droite
    start=0
    i=0
    for perso in NW.columns: # j'empile des barplots pour chaque personnage en
    ↳utilisant 'left='
        width = NW_percent[perso][scene] # la longueur de la barre est la
        ↳portion de la scène où le personnage parle
        ax.barh([acte],[width] , left=[start], height=1, color=[colors[i]],
        ↳label=perso)
        start+=width
        i+=1
    plt.ylim(-.5,.5)
    plt.xlim(0,1)
    previous-=h
plt.legend(loc='upper center', bbox_to_anchor=(1.5, 30.0)) # je mets une légende
fig.text(0.3, 1.02, 'L\'Avare de Molière',fontsize=20) # j'ajoute un titre
```

```
[24]: Text(0.3,1.02,"L'Avare de Molière")
```

L'Avare de Molière



Et voilà, c'est pas hyper joli, mais ça répond à la question!