

Analysing NE: a literary perspective

Virtual Training School on NER
COST Action CA 16204 « Distant
reading »

23rd of March 2021

The Character: a Complex Narratological Object

- Open text (fiction > text), or closed text (fiction = text)?
- ‘parcellar’ character vs. ‘complete’ character
- Russian formalists: character defined by its actantial role

Literary theory about character

- Hamon: the semiotical status of a character; definition by contrast with other characters
 - qualification (predicates)
 - Distribution (appearances)
 - autonomy
 - function
 - Conventional designation
 - Portrait (explicit comments)

From NE to characters

- Characters
 - « bags of words »
 - (words >>> mental or physical representations)
 - 1° what the character says
 - 2° what is said about it
 - 3° what it does
 - 4° what it thinks
- NE: capture only partially no. 2
- Co-reference chains (min. 3 nominal and pronominal expressions pointing towards a discursive object) are more complete, but also important differences

NE and co-reference chains

Tall and athletic, straight in his chair, speaking in a deep voice, **Muyassar** displays a certain form of pride. **The young man** is born in Tadjikistan 25 years ago, in the mountain region of Pamir. (*Langages*, 195/ 2014)

NE, co-reference chains, characterisation

Tall and athletic, straight in his chair, speaking in a deep voice, **Muyassar** displays a certain form of pride. **The young man** is born in Tadjikistan 25 years ago, in the mountain region of Pamir.
(*Langages*, 195/ 2014)

Extended co-reference chains

- + Nominal phrases

- + Epithets

- + Appositions

 - "Timocrate, chief of the conjurers."

- + Attributes

- + anaphorical personnel pronouns

- + complements

- Grammatical agreement


- demonstratives et possessives?

Identifying the markables

- Boundaires

Tall and athletic

Tall and athletic, straight in his chair, speaking in a deep voice, **Muyassar** displays a certain pride. **The young man** is born in Tadjikistan 25 years ago, in the mountain region of Pamir. (*Langages*, 195/ 2014)

A red line with two downward-pointing arrows connects the phrase 'The young man' to the phrase 'born in Tadjikistan', indicating a grammatical or semantic relationship between the subject and the location of birth.

Identifying the markables

Ex.

THE COUNTESS

Is he rich?

LE BARON

Yes.

Ex. Her aunt > The Countess

Or

Her > Lucile

Aunt > The Countess

Identifying the markables

Should we use an attribute (which one?) to indicate who is characterising the character?

- "narrator"

"Timocrate, chief of the conjurers."

- The character itself

I am the best decorator.

- Another character

O God! He's so skilful!

Identifying the markables

- Maybe for theatrical texts, where speeches are annotated, we do not need an indication about the « authorship » of the characterisation?
- BUT

DORINE

She [Mme Pernelle] will tell you [...] that she's not that aged as to deserve this name.

(Molière *Tartuffe*)

➤ We need at least to annotate PPA 3rd person

Corpus

- 35 plays by Louis de Boissy (1694-1758)
- 15 character annotated
- Initial works on Cortazar *Continuidad de los Parques*; Borges *Le Jardin aux sentiers qui bifurquent* (WG « R » of CAHIER)

Characterising the characterisations

@ana values

- Mimetic elements (create the illusion of a real-life person)
- Synthetic elements (the character is linked to a class of literary objects)

Characterising the characterisations

Mimetic elements

- Ontological status
- Age
- Sex
- Nationality
- Social status
- Occupation
- Family position
- Physical aspect
- Temperament
- Faith
- Political adhesion

Annotation sample

```
<sp xml:id="I_1_1" who="#Comtesse">
    <speaker>LA COMTESSE<pc type="affirmative">.</pc></speaker>
<l>As-tu fait tes efforts pour dévoiler <rs ref="#Lucile" ana="#FPN">ma
    nièce</rs> ?</l>
</sp>
<sp xml:id="I_1_2" who="#Lisette">
    <speaker>LISETTE<pc type="affirmative">.</pc></speaker>
    <l><rs ref="#Comtesse" ana="#SS">Madame</rs>, j'ai perdu près
    d'elle mon adresse.</l>
    <l><rs ref="#Lucile" ana="#TEMP">Son air paraît ouvert, son cœur ne
        l'est jamais</rs>,</l>
    <l>On ne peut pénétrer dans ses replis secrets ;</l>
    <l>À le développer vainement on s'attache,</l>
    <l><rs ref="#Lucile" ana="#TEMP">C'est par timidité, peut-être, qu'il se
        cache</rs>.</l>
</sp>
```

Some results

Paméla en France (1745, a parody)

- 50 characters
- 29 extended co-référence chains
- 20 chains have one referential object, 7 have several referential objects, 2 have collective objects

Types of entities referred to

1 entity		Several entities	Collective entities
Paméla Mère Père Marquis Comtesse Maître Mathurin Oncle Chevalier Nérine Dame	Tante Le Plaisir La Sagesse L'Amour Lutin 1 La Décence La Gaïeté La Joie L'Hymen	Père, Mère Marquis, Comtesse Paméla, Chevalier Oncle, Tante Paméla, Mathurin Sagesse, Paméla Plaisir, Marquis	Danseurs Les Français

Lenght of co-reference chains

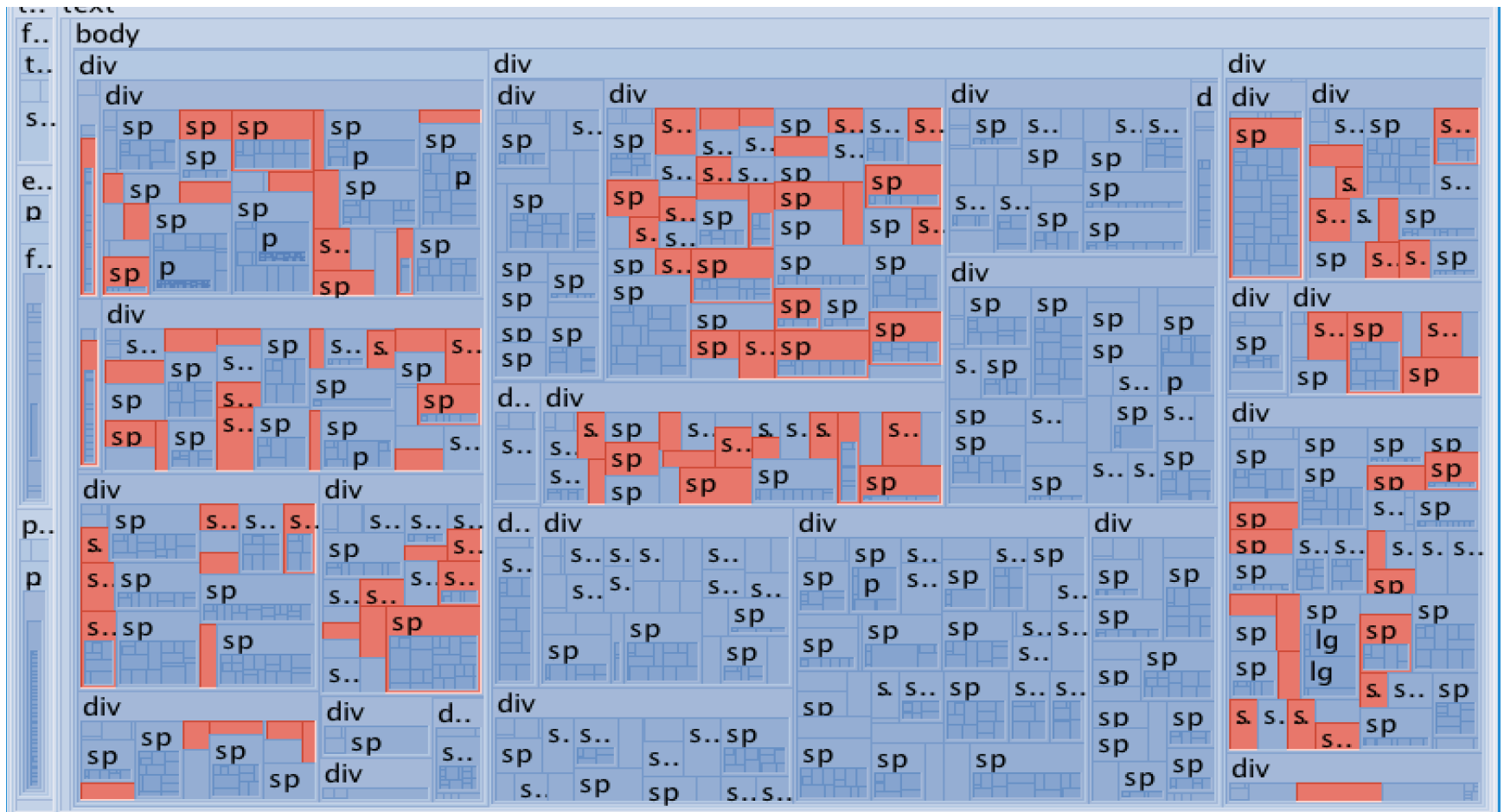
Chaînes	Longueur	Longueur +
Paméla	503	534
Marquis	242	307
Mathurin	212	217
Chevalier	157	160
Nérine	78	80
Comtesse	33	71
Le Plaisir	19	41
Marquis, Comtesse		37
Maître	27	29
La Sagesse	7	28
Dame	24	25
Plaisir, Marquis		20
Sagesse, Paméla		19

Lenght of co-reference chains

Père	4	15
Mère	3	14
La Décence	14	14
Les Français		12
L'Amour	11	11
La Gaieté	11	11
Père, Mère		11
Oncle	3	6
Danseurs		6
Lutin 1	5	5
Tante	1	4
La Joie	4	4
Paméla, Mathurin		4
L'Hymen	3	3
Oncle, Tante		3

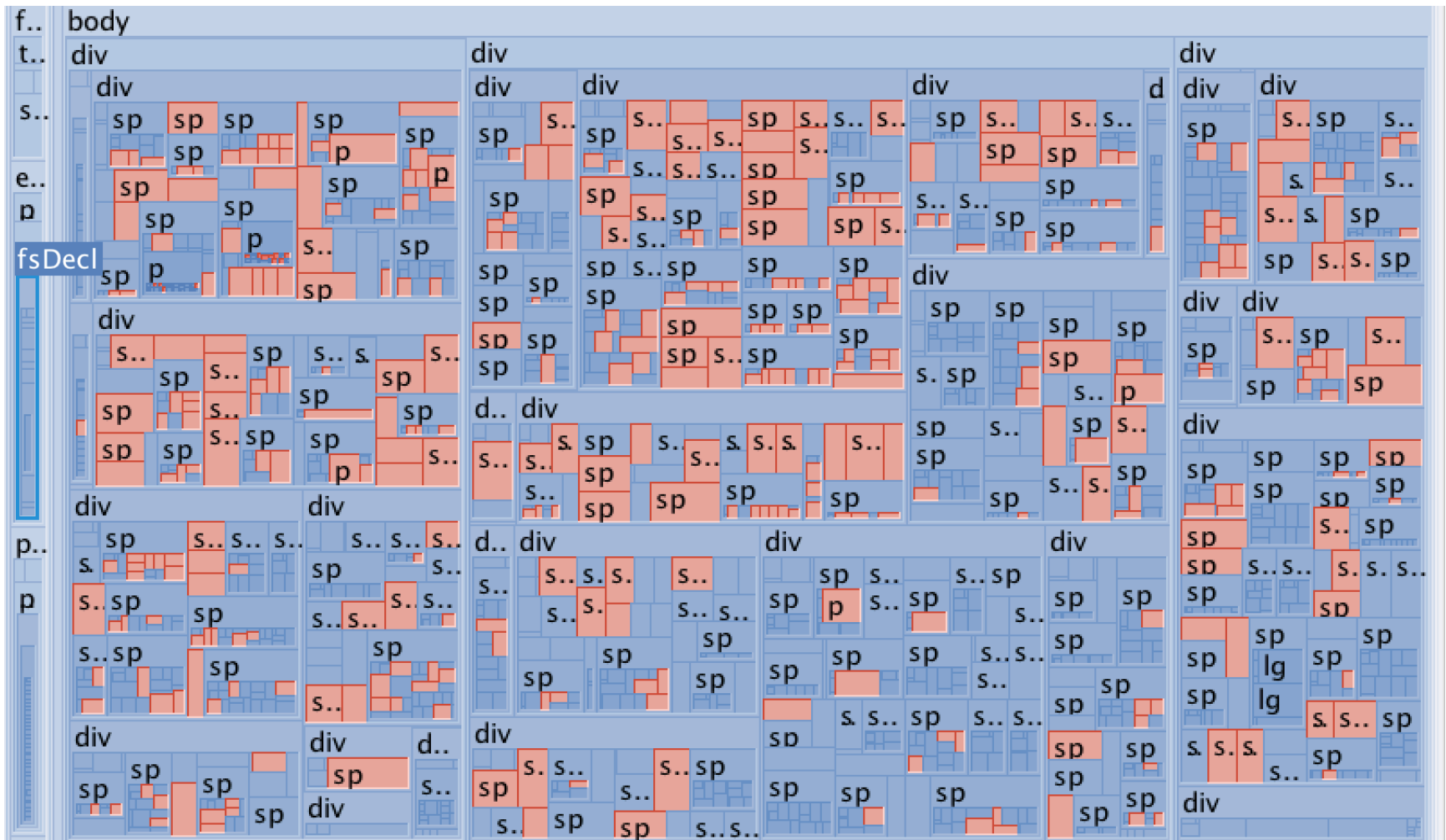
Character's distribution

Pamela as an acting character



Character's distribution

Pamela in the co-reference chains



What's in a chain

The co-reference chain for Pamela

je
la
la
elle
Vous
vous
vous
un talent
je
pour moi
Ma chère
vous
je
Je
me
je m'exprime assez mal en François
une jeune étrangère
la naïveté
J
je me

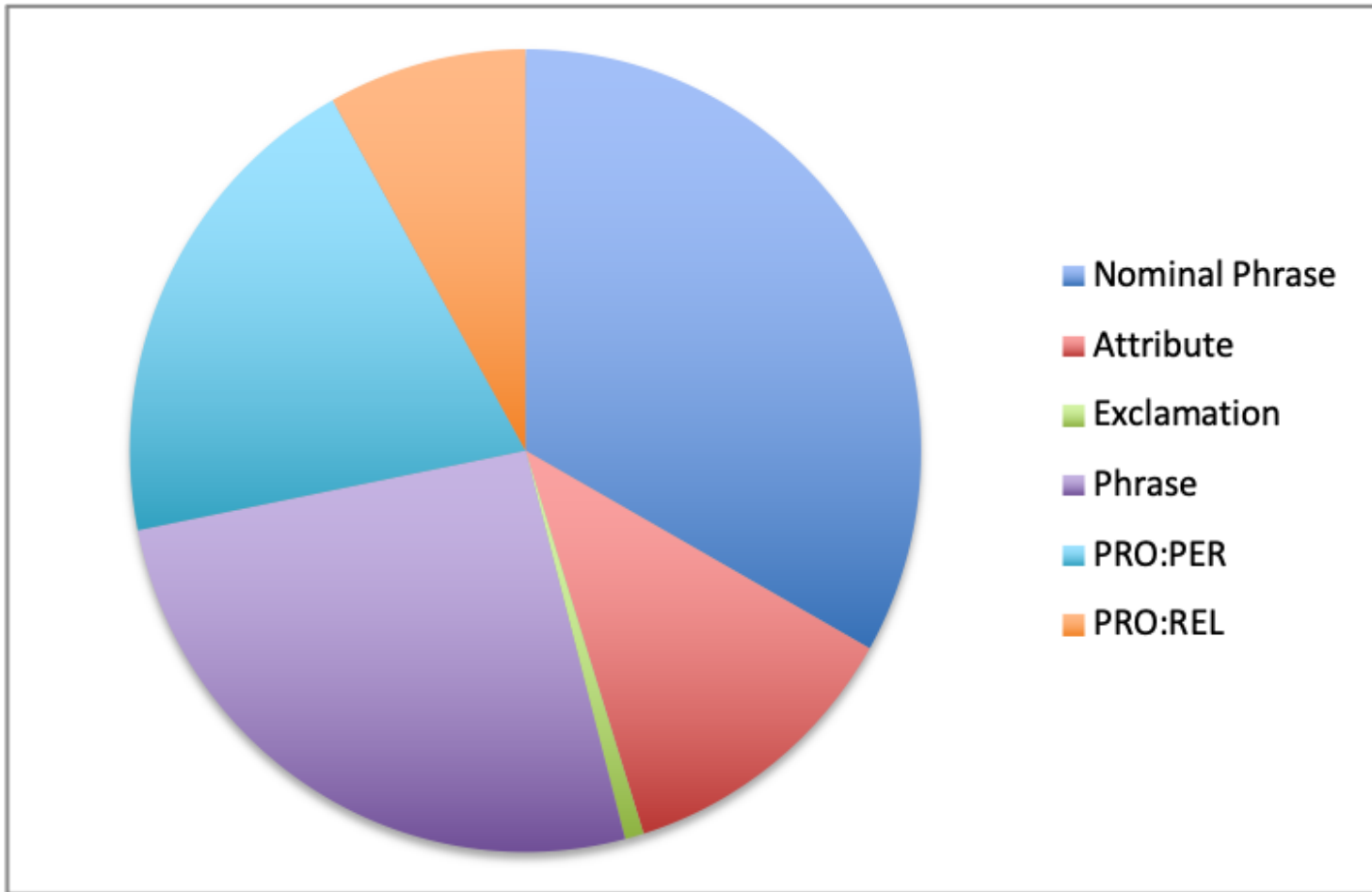
j
je
vous
moi
m
être sage, & modeste &
fée,
Je
m
ma sagesse
elle
je
vous
moi
la femme de chambre
la plus tendre de ses amies
je
j
j
me

m
vous
Tout fied àPaméla
Sa grace est partout naturelle
Elle pare l'habit
elle
rien n'est étranger pour elle
Je
vous en êtes plus belle
m
me
j'aime la Musique & la danse
je
Je
J'écris ce que
je
pense un peu trop librement

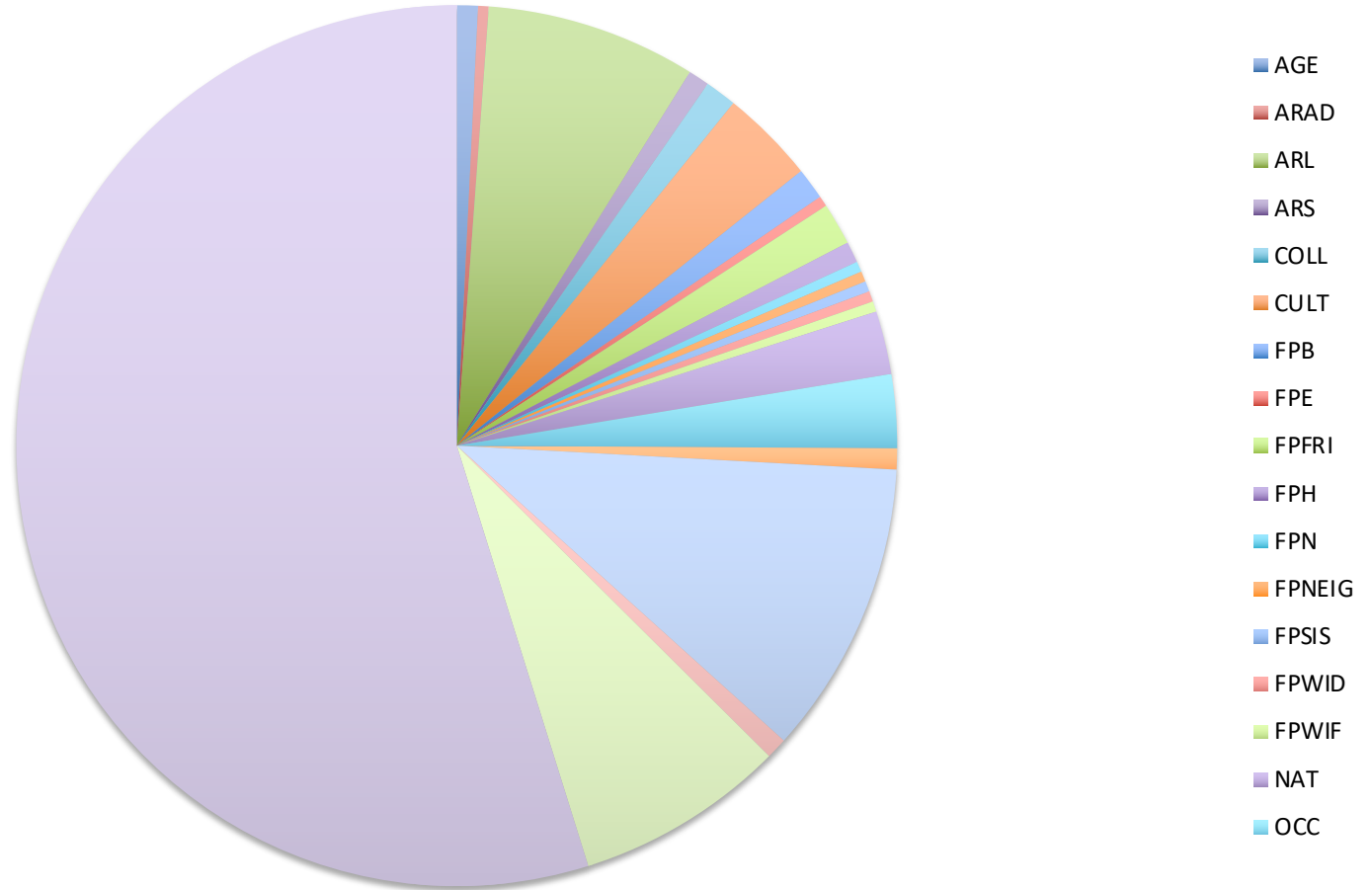
Simple units/ characterisation units

Character	Ratio characterisation/ units	Character	Ratio characterisation/ units
Paméla	102/503 (20%)	Sagesse, Paméla	3/19
Marquis	30/242 (12%)	Père	0/4
Mathurin	36/212 (16%)	Mère	1/3
Chevalier	26/157 (16%)	La Décence	0/14
Nérine	9/78 (11%)	Les Français	8/12
Comtesse	9/33 (27%)	L'Amour	0/11
Le Plaisir	2/19 (10%)	La Gaïeté	1/11
Marquis, Comtesse	8/37 (21%)	Père, Mère	2/11
Maître	12/27 (44%)	Oncle	1/3
La Sagesse	0/7	Danseurs	3/6
Dame	8/24 (33%)	Lutin1	1/5
Plaisir, Marquis	0/20	Tante	0/1

POS repartition



What kind of characterisations?



TEMP characterisation

	TEMP	Total ana	
Plagiaire	69	215	32,09%
Vie	245	574	42,68%
Triomphe intéri	113	317	35,65%
Talents	115	311	36,98%
Sage étourdi	113	246	45,93%
Rival favorable	156	282	55,32%
Retour paix	19	52	36,54%
Prix silence	323	452	71,46%
Pouvoir sympa	217	436	49,77%

TEMP: constant distribution over each
play

