

# MeThAL

## Vers une macroanalyse du théâtre alsacien

*Séminaire LiLPa*

Pablo Ruiz Fabo



# Plan

- Macroanalyse
- Projet MeThAL
- Démonstration

# Plan

- **Macroanalyse**
- Projet MeThAL
- Démonstration

# Macroanalyse

- Grands corpus textuels
- Caractérisation de sous-corpus
  - Période, groupe d'auteurs
- Traits linguistiques concrets
  - Traitement automatique
  - Analyse quantitative
- Complémente la lecture « proche »

# MACROANALYSIS

(2013)

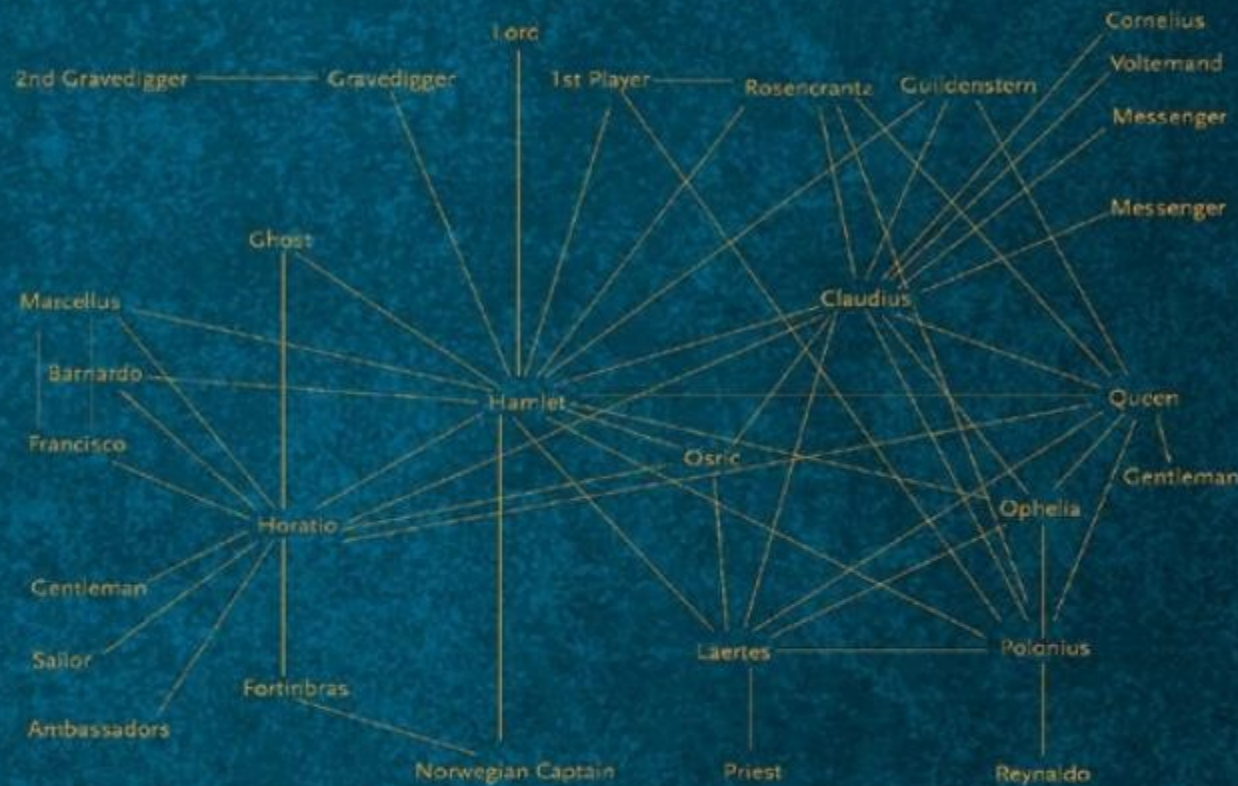
*Digital Methods & Literary History*



MATTHEW L. JOCKERS

# DISTANT READING

(2013)



Franco Moretti

# Macroanalyse des textes de théâtre

- Théâtre classique français (Schöch, 2017)
  - Caractérisation des sous-genres via topic models

## **DHQ: Digital Humanities Quarterly**

2017

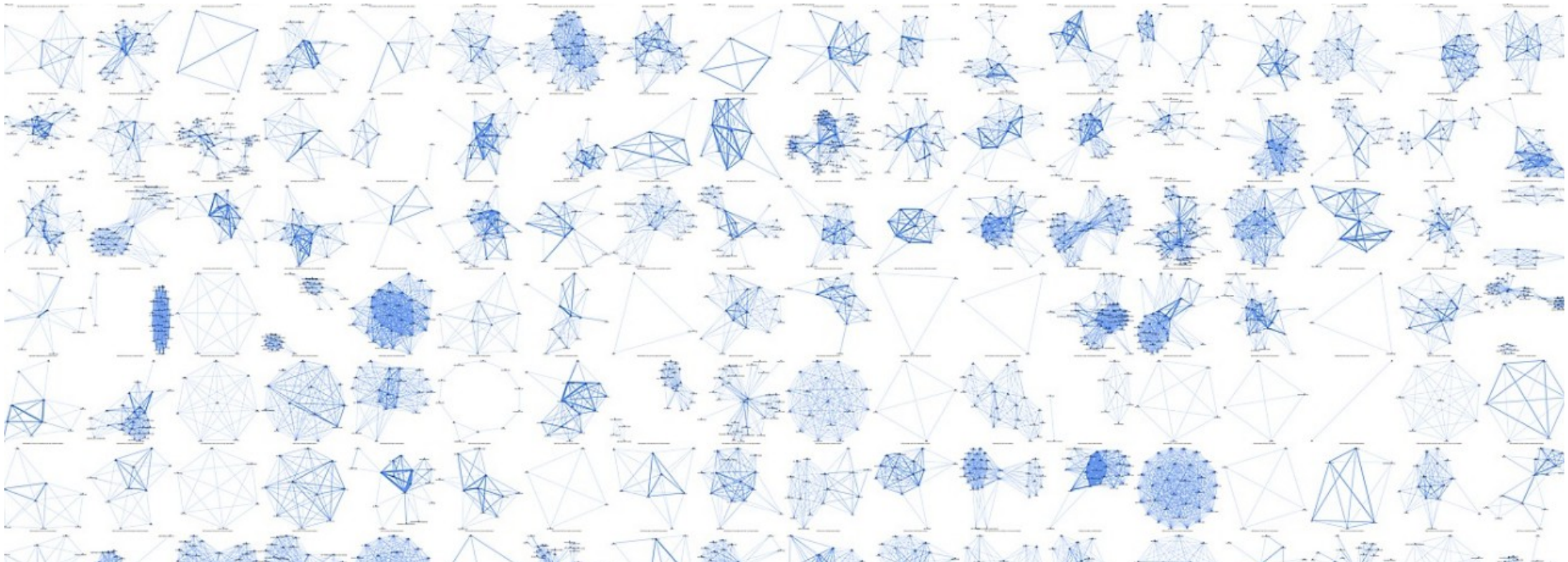
Volume 11 Number 2

### **Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama**



# Macroanalyse des textes de théâtre

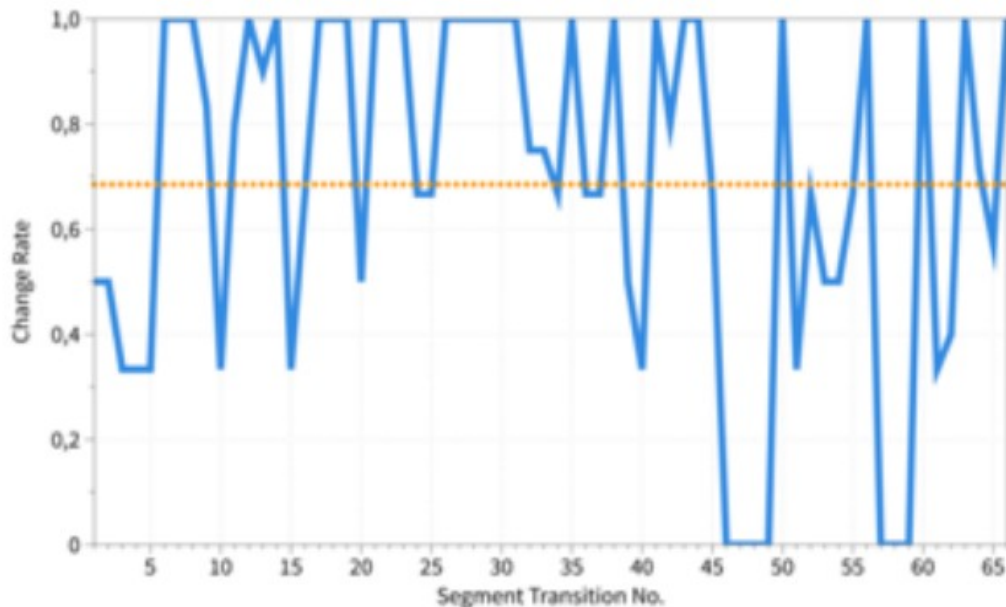
- Théâtre allemand
  - DLINA (Fischer et al.) : réseaux de personnages



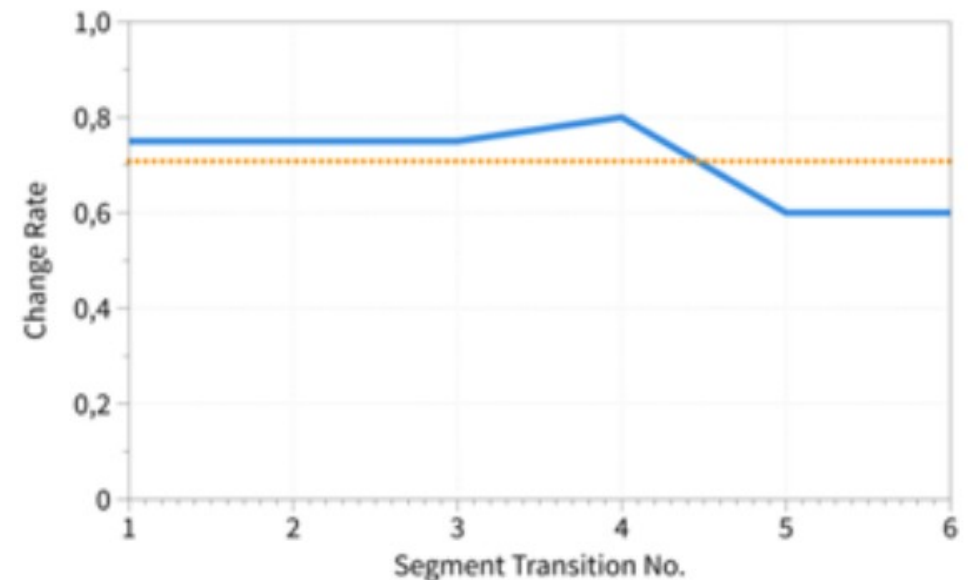


# Macroanalyse des textes de théâtre

- Théâtre allemand
  - *Dynamisme* : changement du nombre de personnages sur scène (Fischer et al., 2017)



« Die Jubelfeier der Hölle » (Benkowitz, 1801)



« Anatol » (Schnitzler, 1893)

# Macroanalyse des textes de théâtre

- Grandes traditions européennes
  - Théâtre allemand
    - *Quantitative Drama Analysis* (QuaDramA)  
Reiter et al., Stuttgart
    - Identification automatique des protagonistes
    - Classification automatique en types de personnages



# Macroanalyse des textes de théâtre :

## Corpus numériques en format standard

Drama Corpora Project (DraCor)

### German Drama Corpus

Number of plays	474
person + personGrp	10,734 (male: 7365, female: 2301)
text	8,342,891 tokens
sp	330,206 (7,966,041 tokens)
stage	160,681 (1,016,145 tokens)
Last update	10/12/2019, 18:23:17

```
<sp who="#damon">
  <speaker>DAMON.</speaker>
  <p>Warum lacht Ihr?</p>
</sp>
<sp who="#lisette">
  <speaker>LISETTE.</speaker>
  <p>Ja das ist gewiß. Auf seine Treue ki
    Ihnen in Ihrer Not redlich beistehen
</sp>
<sp who="#damon">
  <speaker>DAMON.</speaker>
  <p>Das hoffe ich auch gewiß.</p>
</sp>
<sp who="#lisette">
  <speaker>LISETTE.</speaker>
```

# Théâtre en alsacien ?

- Impossibilité de ces types d'analyse
- Corpus requis non disponibles
- Projet **MeThAL** un premier pas vers  
una macroanalyse de la tradition

# Plan

- Macroanalyse
- **Projet MeThAL**
- Démonstration

# Projet MeThAL

- Équipe
- Objectifs et défis
  - Encodage TEI d'un corpus de théâtre
  - Interface d'exploration du corpus
  - Analyses (socio)linguistiques
  - Analyses littéraires
- Collaborations



# MeThAL : Équipe

- Porteur : Pablo Ruiz
- Delphine Bernhard
- Pascale Erhart
- Dominique Huck
- Carole Werner

# MeThAL : Objectifs & défis

- Encodage TEI d'un corpus de théâtre en alsacien
- Interface d'exploration du corpus
- Analyses (socio)linguistiques
- Analyses littéraires « distantes »

# Encodage TEI



## Text Encoding Initiative (TEI) Guidelines

- + 1 [The TEI Infrastructure](#)
- + 2 [The TEI Header](#)
- + 3 [Elements Available in All TEI Documents](#)
- + 4 [Default Text Structure](#)
- + 5 [Characters, Glyphs, and Writing Modes](#)
- + 6 [Verse](#)
- + 7 [Performance Texts](#)
- + 8 [Transcriptions of Speech](#)
- + 9 [Dictionaries](#)
- + 10 [Manuscript Description](#)
- + 11 [Representation of Primary Sources](#)
- + 12 [Critical Apparatus](#)
- + 13 [Names, Dates, People, and Places](#)
- + 14 [Tables, Formulæ, Graphics and Notated Music](#)
- + 15 [Language Corpora](#)
- + 16 [Linking, Segmentation, and Alignment](#)
- + 17 [Simple Analytic Mechanisms](#)
- + 18 [Feature Structures](#)
- + 19 [Graphs, Networks, and Trees](#)
- + 20 [Non-hierarchical Structures](#)
- + 21 [Certainty, Precision, and Responsibility](#)
- + 22 [Documentation Elements](#)
- + 23 [Using the TEI](#)

# Encodage TEI



## Text Encoding Initiative (TEI) Guidelines

- + 1 [The TEI Infrastructure](#)
- + 2 [The TEI Header](#)
- + 3 [Elements Available in All TEI Documents](#)
- + 4 [Default Text Structure](#)
- + 5 [Characters, Glyphs, and Writing Modes](#)
- + 6 [Verse](#)
- + 7 [Performance Texts](#)
- + 8 [Transcriptions of Speech](#)
- + 9 [Dictionaries](#)
- + 10 [Manuscript Description](#)
- + 11 [Representation of Primary Sources](#)
- + 12 [Critical Apparatus](#)
- + 13 [Names, Dates, People, and Places](#)
- + 14 [Tables, Formulæ, Graphics and Notated Music](#)
- + 15 [Language Corpora](#)
- + 16 [Linking, Segmentation, and Alignment](#)
- + 17 [Simple Analytic Mechanisms](#)
- + 18 [Feature Structures](#)
- + 19 [Graphs, Networks, and Trees](#)
- + 20 [Non-hierarchical Structures](#)
- + 21 [Certainty, Precision, and Responsibility](#)
- + 22 [Documentation Elements](#)
- + 23 [Using the TEI](#)

# Encodage TEI

```
<listPerson>
  <person xml:id="lisette" sex="FEMALE">
    <persName>Lisette</persName>
  </person>
  <person xml:id="die_witwe" sex="FEMALE">
    <persName>Die Witwe</persName>
  </person>
  <person xml:id="damon" sex="MALE">
    <persName>Damon</persName>
  </person>
  <person xml:id="leander" sex="MALE">
    <persName>Leander</persName>
  </person>
  <person xml:id="oronte" sex="MALE">
    <persName>Oronte</persName>
  </person>
</listPerson>
```

```
<div type="scene">
  <head>Neunter Auftritt</head>
  <stage>Lisette. Damon.</stage>
  <sp who="#lisette">
    <speaker>LISETTE.</speaker>
    <p>So alleine? und so betrübt?</p>
  </sp>
  <sp who="#damon">
    <speaker>DAMON.</speaker>
    <p>Ach Lisette, meinen Kummer zu erleichtern, muß ich  
Ich bin unglücklich gewesen. Mein Schiff ist in eine  
gewisseste Nachricht. Himmel! und ich verliere zugl  
-</p>
  </sp>
  <sp who="#lisette">
    <speaker>LISETTE.</speaker>
    <p>Was? So ist es an Leanders Unglücke nicht genug ge  

```

# Encodage TEI :

## Sources numériques (BNU)

**NUMiSTRAL**

TOUT

Rechercher...

RECHERCHE  
AVANCÉE

Accueil > 868 résultats page 1 sur 58 > Consultation

 D'r poetisch Oscar : Einakter / vun Marie Hart ; nej bearweit vun E. Deck,... H...

  
Aide



**D'r poetisch  
Oscar**

Einakter  
vun  
**MARIE HART**

Nej bearweit vun E. DECK, Regisseur vom Els. Theater  
in Hawenau.



# Défis encodage TEI : OCR

- Reconnaissance optique des caractères
  - Modèle alsacien du projet RESTAURE
  - Défi : typographie gothique

August 28, 2018 (1)

Dataset

Open Access

## Tesseract OCR models for the Alsatian dialects

Bernhard, Delphine;

This dataset provides trained Tesseract (<https://github.com/tesseract-ocr/tesseract>) OCR models for the Alsatian dialects. These models were developed in the context of the RESTAURE project, funded by the French ANR. Two models are provided : The first model, ISKO\_2015, has been presented in

Uploaded on August 28, 2018



# Défis encodage TEI : Normalisation

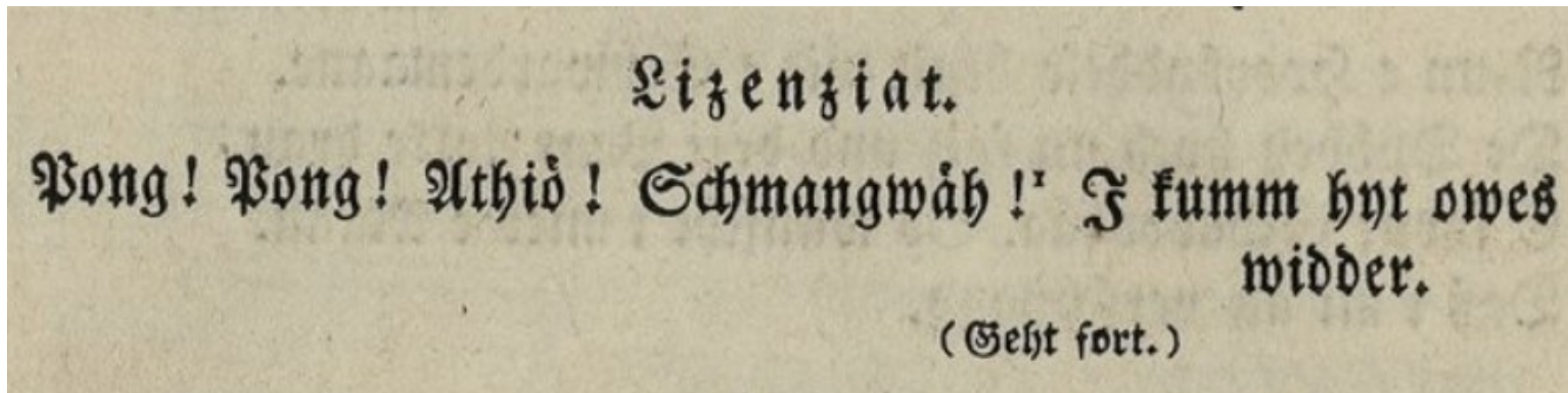
- Grande variété orthographique en alsacien
- Recherche en TAL sur la normalisation automatique

# Défis encodage TEI :

## Mélange de langues dans les pièces

- Difficulté pour le TAL

Pfingstmontag  
(Arnold, 1816)



**Lizenziat.**

Pong! Pong! Athiö! Schmangwäh! I kumm hyt owes widder.

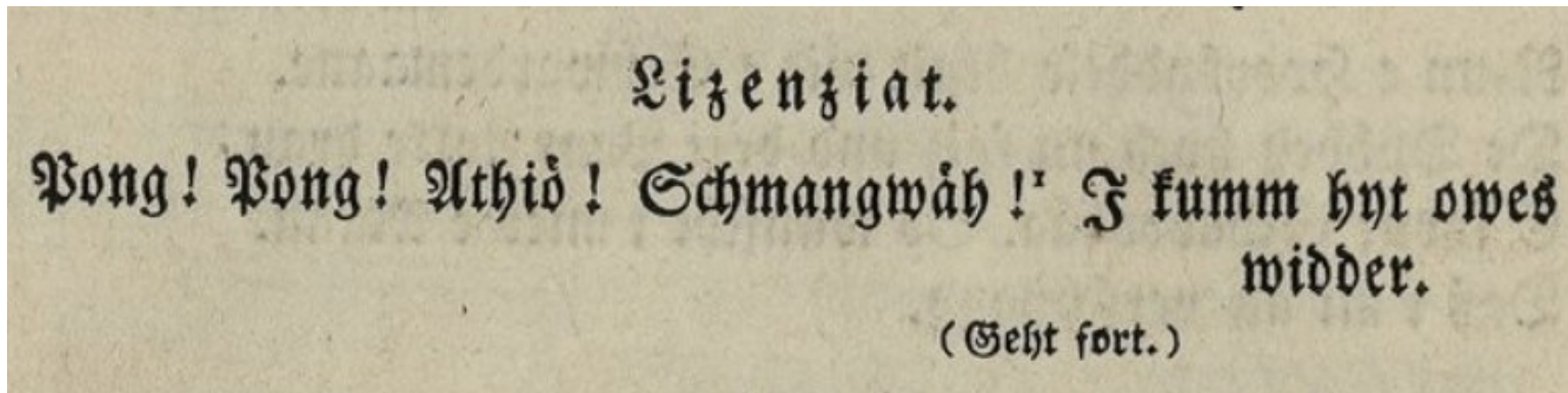
*Geht fort.*

# Défis encodage TEI :

## Mélange de langues dans les pièces

- Difficulté pour le TAL

Pfingstmontag  
(Arnold, 1816)



**Lizenziat.**

Pong! Pong! Athiö! Schmangwäh! I kumm hyt owes widder.

Geht fort.

# Applications du corpus :

## Sociolinguistique historique

- Pratiques orthographiques et linguistiques
- Comment les situer par rapport aux pratiques en français et allemand ?
- Évolution temporelle

# Applications du corpus :

## Analyses littéraires

- Analyse quantitative des interactions
  - Qui parle à qui ?
  - De quels sujets ?
  - Quelles variables jouent un rôle ?
    - Position sociale, sexe du personnage
- Aspects formels de la technique dramatique
- Sous-genres



# Plan

- Macroanalyse
- Projet MeThAL
- **Démonstration**

# *Pfingstmontag* sur DraCor

- *Der Pfingsmontag* (Arnold, 1816) :  
Pièce fondatrice du théâtre alsacien

## TRAVAIL EFFECTUÉ

- Encodage en TEI
- Intégration dans la plate-forme DraCor  
pour l'analyse quantitative



```
<listPerson>
  <person xml:id="dorthee" sex="FEMALE">
    <persName>Frau Dorothee (Dorthee)</persName>
  </person>
  <person xml:id="starkhans" sex="MALE">
    <persName>Hr. Starkhans</persName>
  </person>
  <person xml:id="lissel" sex="FEMALE">
    <persName>Lieschen (Lissel)</persName>
  </person>
  <person xml:id="christinel" sex="FEMALE">
    <persName>Christinchen (Christinel)</persName>
  </person>
  <person xml:id="reinhold" sex="MALE">
    <persName>Reinhold</persName>
  </person>
  <person xml:id="lizentiat" sex="MALE">
    <persName>Lizenziat Mehlbrüh</persName>
  </person>
  <person xml:id="rosine" sex="FEMALE">
    <persName>Frau Rosine</persName>
  </person>
  <person xml:id="mehlbrüej" sex="MALE">
    <persName>Herr Mehlbrüh (Mehlbrüej)</persName>
  </person>
```



## Alsatian Drama Corpus

Number of plays	1
person + personGrp	16 (male: 8, female: 8)
text	36,362 tokens
sp	1,334 (35,959 tokens)
stage	211 (719 tokens)
Last update	07/12/2019, 18:27:03

## Greek Drama Corpus

Number of plays	39
person + personGrp	437 (male: 275, female: 110)
text	321,599 tokens

## Calderón Drama Corpus

Number of plays	54
person + personGrp	839 (male: 550, female: 284)
text	568,412 tokens
sp	23,966 (544,039 tokens)
stage	4,758 (27,898 tokens)
Last update	28/09/2019, 11:31:40

## Roman Drama Corpus

Number of plays	36
person + personGrp	385 (male: 253, female: 104)
text	306,144 tokens

## German Drama Corpus

Number of plays	474
person + personGrp	10,734 (male: 7365, female: 2301)
text	8,342,891 tokens
sp	330,206 (7,966,041 tokens)
stage	160,681 (1,016,145 tokens)
Last update	10/12/2019, 18:23:17

## Russian Drama Corpus

Number of plays	210
person + personGrp	3,636 (male: 2550, female: 858)



## Alsatian Drama Corpus

Number of plays	1
person + personGrp	16 (male: 8, female: 8)
text	36,362 tokens
sp	1,334 (35,959 tokens)
stage	211 (719 tokens)
Last update	07/12/2019, 18:27:03

## Calderón Drama Corpus

Number of plays	54
person + personGrp	839 (male: 550, female: 284)
text	568,412 tokens
sp	23,966 (544,039 tokens)
stage	4,758 (27,898 tokens)
Last update	28/09/2019, 11:31:40

## German Drama Corpus

Number of plays	474
person + personGrp	10,734 (male: 7365, female: 2301)
text	8,342,891 tokens
sp	330,206 (7,966,041 tokens)
stage	160,681 (1,016,145 tokens)
Last update	10/12/2019, 18:23:17

## Greek Drama Corpus

Number of plays	39
person + personGrp	437 (male: 275, female: 110)
text	321,599 tokens

## Roman Drama Corpus


Number of plays	36
person + personGrp	385 (male: 253, female: 104)
text	306,144 tokens

## Russian Drama Corpus

Number of plays	210
person + personGrp	3,636 (male: 2550, female: 858)

# dracor.org/als

Drama Corpora Project (DraCor) Corpora ▾ SPARQL API

Easy Linavis 

## Alsatian Drama Corpus

Authors ↑↓	Title ↑↓	Network Size ↑↓	Year (normalized) ↑↓	Source ↑↓	ID ↑↓
Arnold, Johann Georg Daniel PND: <a href="#">118927698</a>	<a href="#">Der Pfingstmontag</a> Lustspiel in Straßburger Mundart Wikidata: <a href="#">Q77593548</a>	16	1816 📅 1835 📅 1816	<a href="#">MeThAL</a> (TEI version)	als000001



Arnold, Johann Georg Daniel ([118927698](#)) ▴  
**Der Pfingstmontag** ([Q77593548](#))  
*Lustspiel in Straßburger Mundart*  
📅 1835 📖 1816

Downloads

Network Metrics

Cast list (in order of appearance)

1. Frau Dorothee (Dorthee)

2. Hr. Starkhans

3. Lieschen (Lissel)

4. Christinchen (Christinel)

5. Reinhold

6. Lizenziat Mehlbrüh

7. Frau Rosine

8. Herr Mehlbrüh (Mehlbrüej)

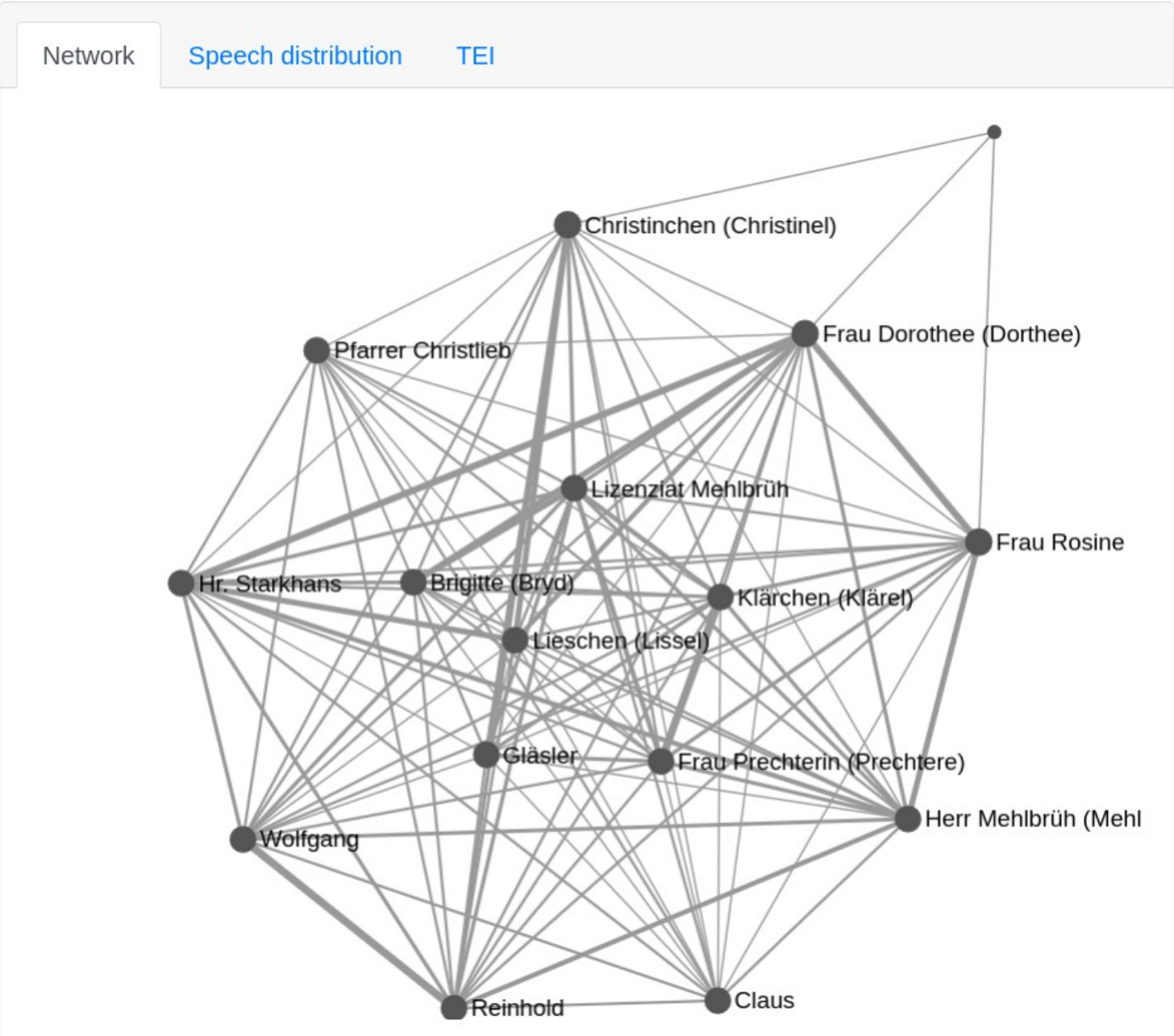
9. Wolfgang

10. Brigitte (Bryd)

11. Frau Prechterin (Prechtere)

12. Klärchen (Klärel)

...



# Interaction entre les personnages

source	target	weight
All	All	All
reinhold	wolfgang	7
prechtere	klärel	7
dorthee	starkhans	6
dorthee	rosine	6
dorthee	bryd	5
starkhans	lissel	5
lissel	christinel	5
lizentiat	bryd	5
rosine	mehlbrüej	5
dorthee	lissel	4
starkhans	mehlbrüej	4
christinel	gläser	4
reinhold	mehlbrüej	4
lizentiat	prechtere	4
lizentiat	klärel	4
dorthee	lizentiat	3
dorthee	mehlbrüej	3
starkhans	bryd	3

Pfingstmontag (Arnold, 1816)

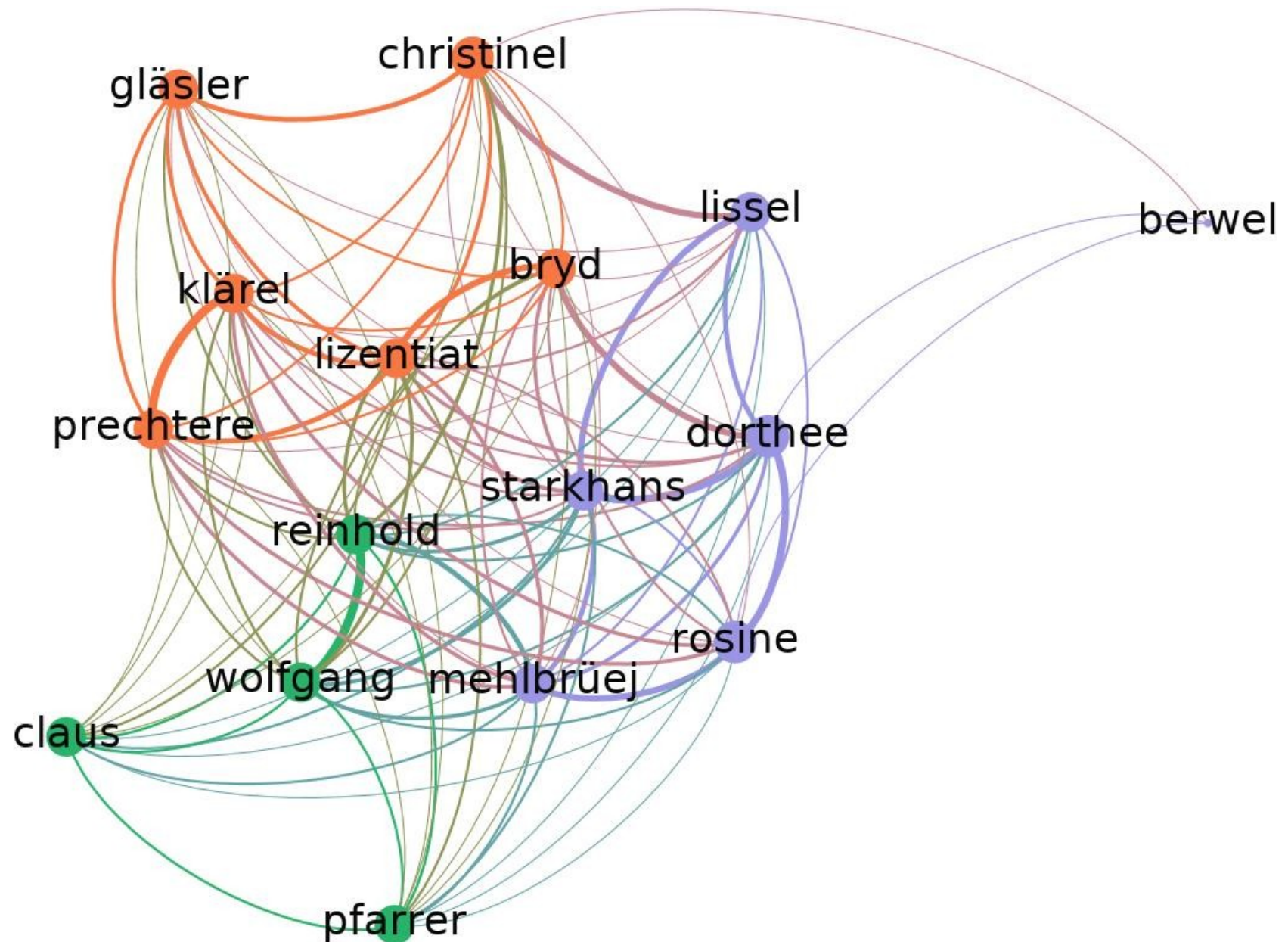
# Interaction entre les personnages

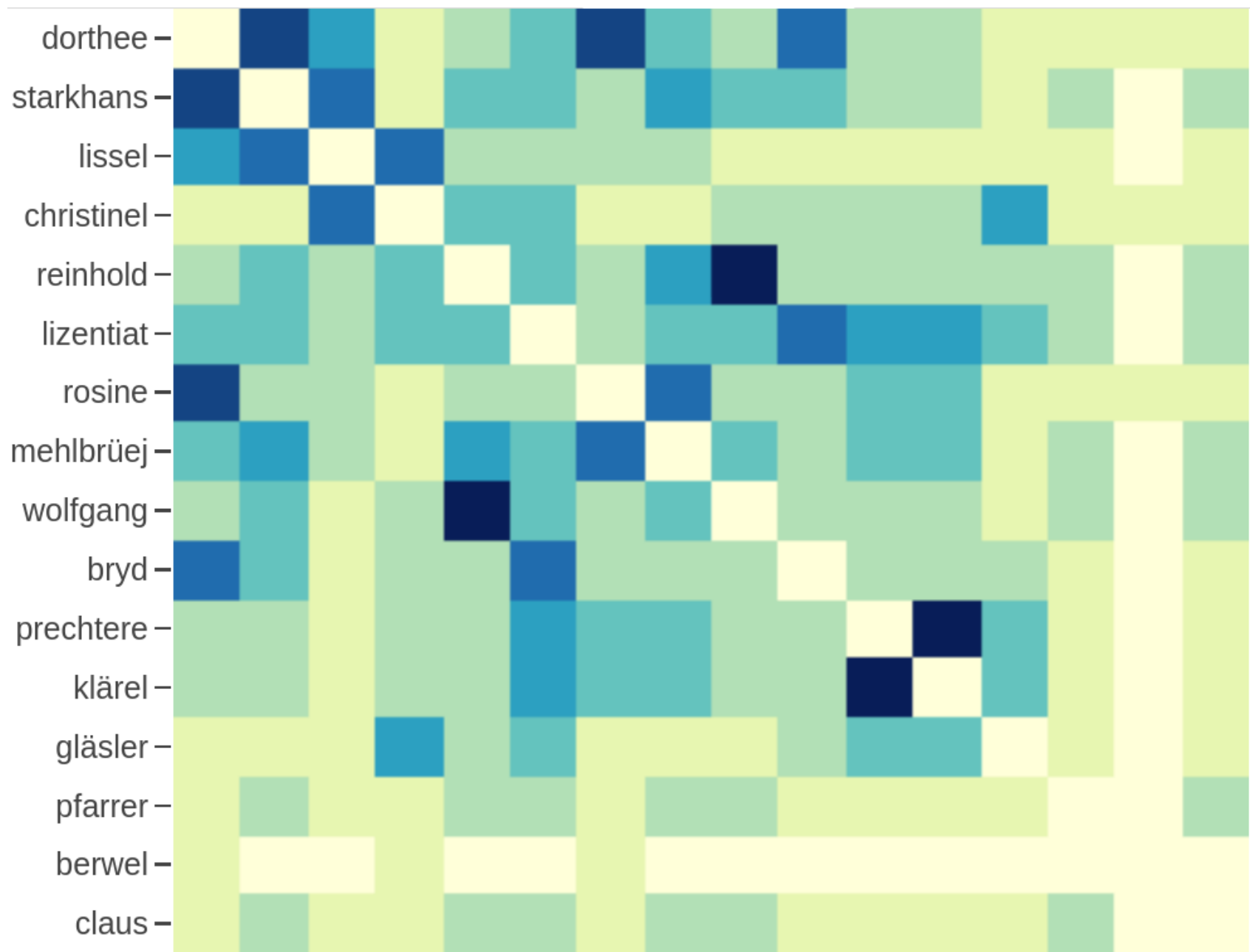
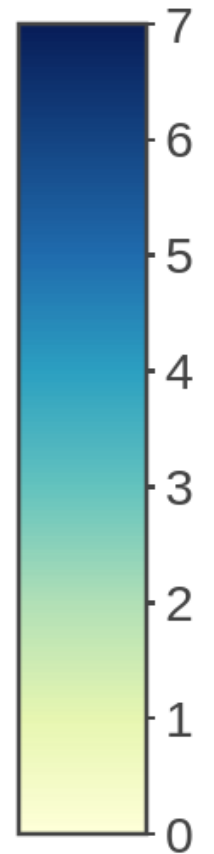
Pfingstmontag (Arnold, 1816)

name	betweenness	closeness	strength	degree	average_distance
Al	All	All	All	All	All
dorthee	4.00	1.0000	40	15	1
christinel	4.00	1.0000	30	15	1
rosine	4.00	1.0000	34	15	1
starkhans	0.00	0.9375	39	14	1.06666666666667
lissel	0.00	0.9375	29	14	1.06666666666667
reinhold	0.00	0.9375	38	14	1.06666666666667
lizentiat	0.00	0.9375	42	14	1.06666666666667
mehlbrüej	0.00	0.9375	38	14	1.06666666666667
wolfgang	0.00	0.9375	34	14	1.06666666666667
bryd	0.00	0.9375	32	14	1.06666666666667
prechtere	0.00	0.9375	35	14	1.06666666666667
klärel	0.00	0.9375	35	14	1.06666666666667
gläsler	0.00	0.9375	25	14	1.06666666666667
pfarrer	0.00	0.9375	20	14	1.06666666666667
claus	0.00	0.9375	20	14	1.06666666666667
berwel	0.00	0.5556	3	3	1.8

# Interaction entre les personnages

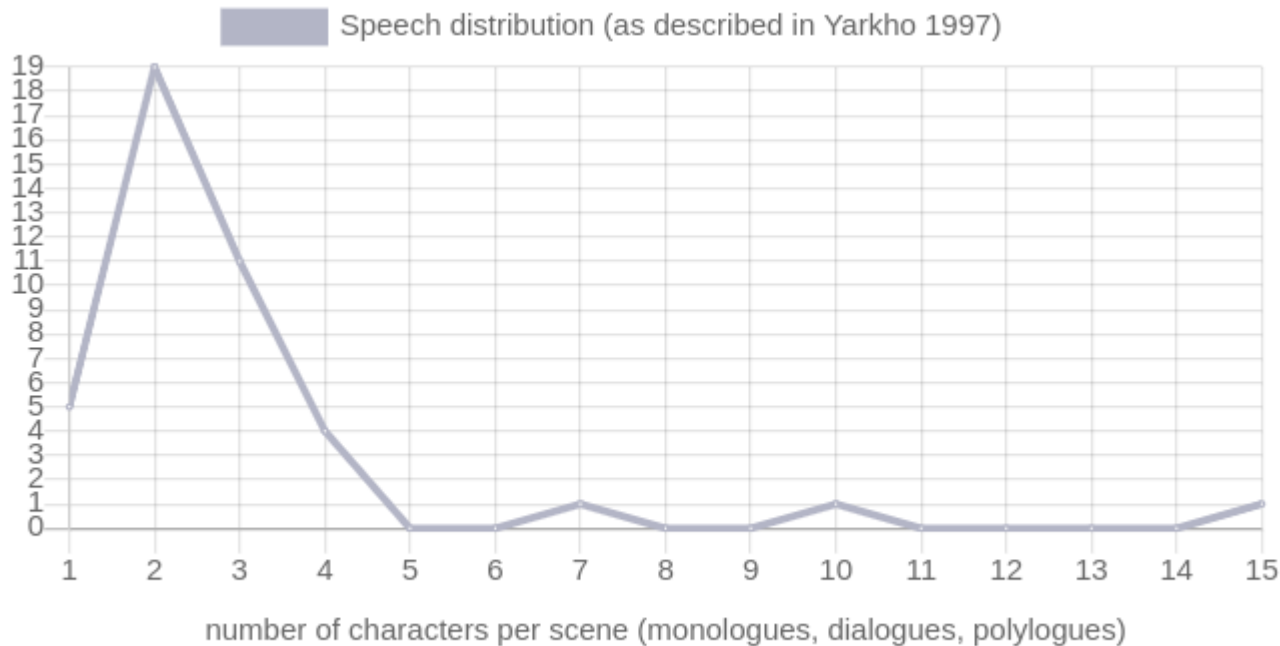
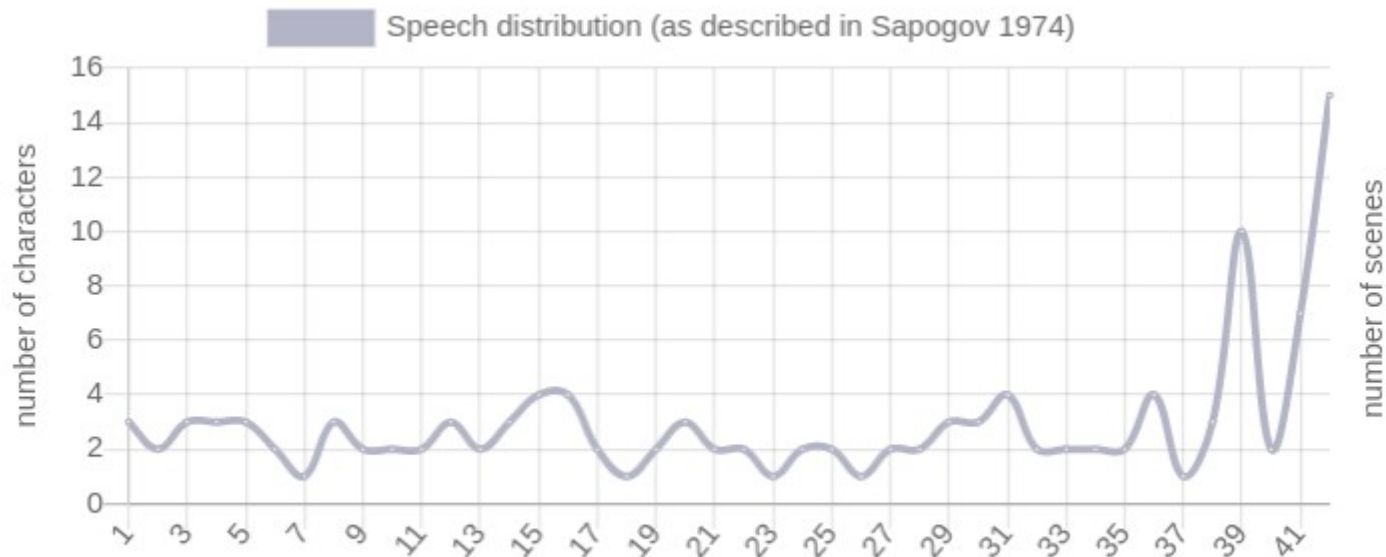
Pfingstmontag (Arnold, 1816)





# Nombre de personnages sur scène

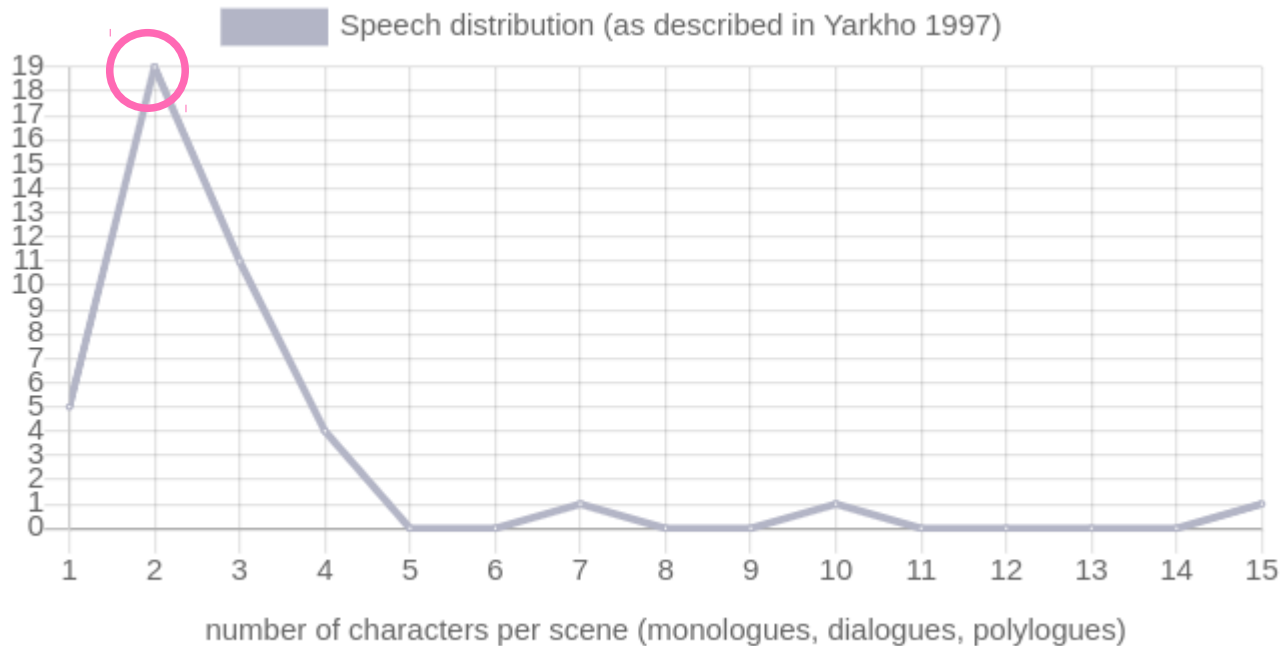
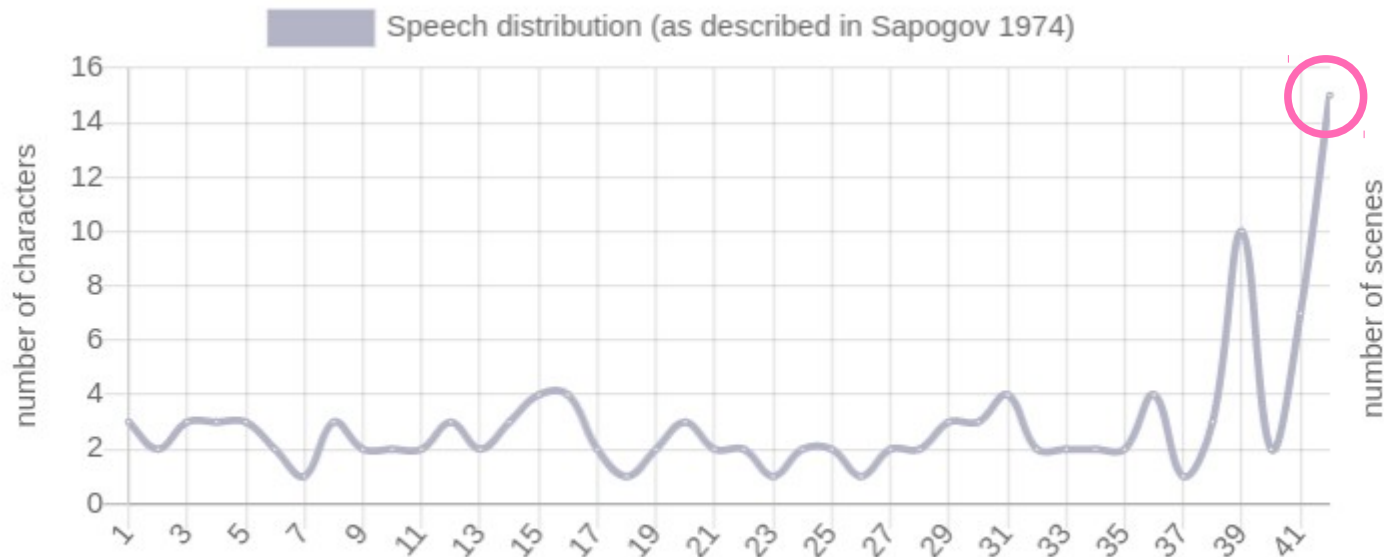
*Pfingstmontag* (Arnold, 1816)





# Nombre de personnages sur scène

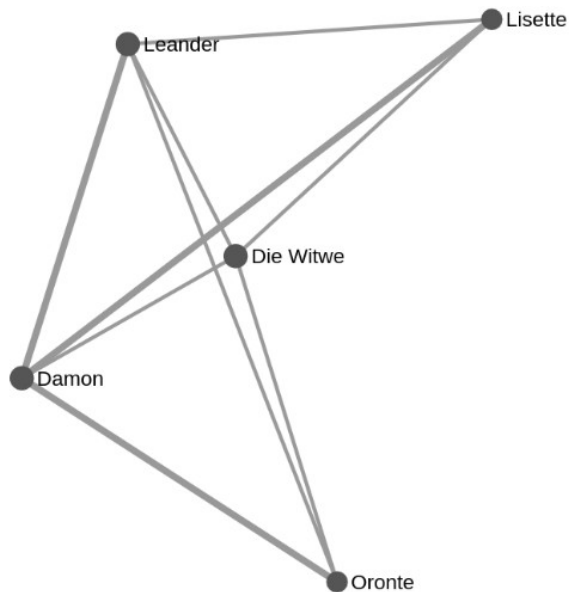
*Pfingstmontag (Arnold, 1816)*



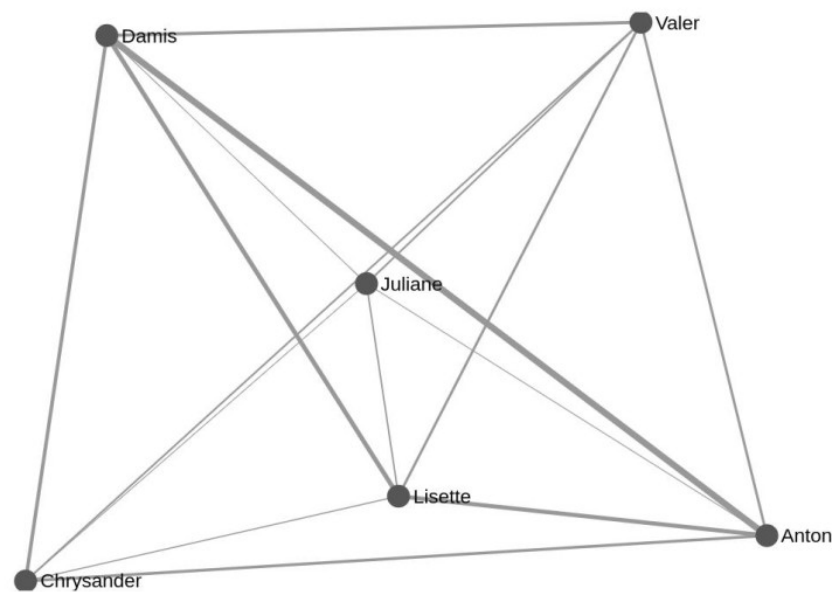
# Tendances :

## Comparaison de pièces dans GerDraCor

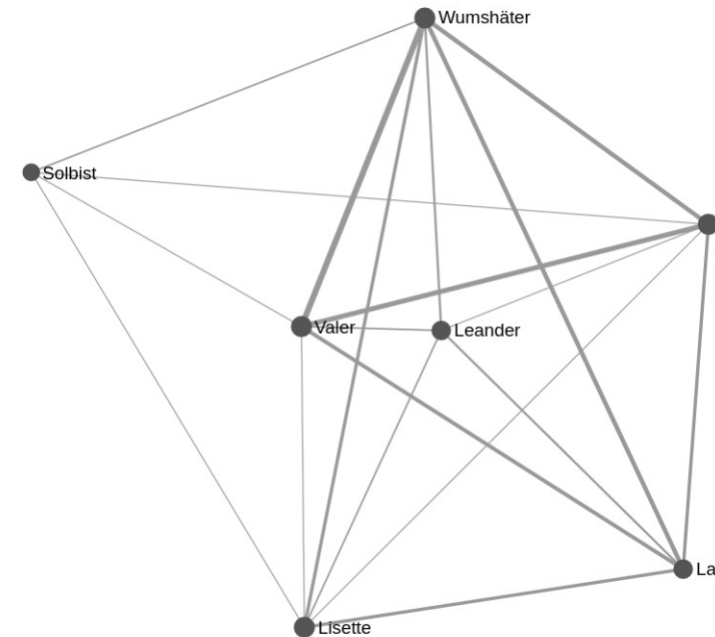
- 3 Comédies de Lessing (1747 à 1755)



*Damon*



*junge Gelehrte*



*Mysogin*

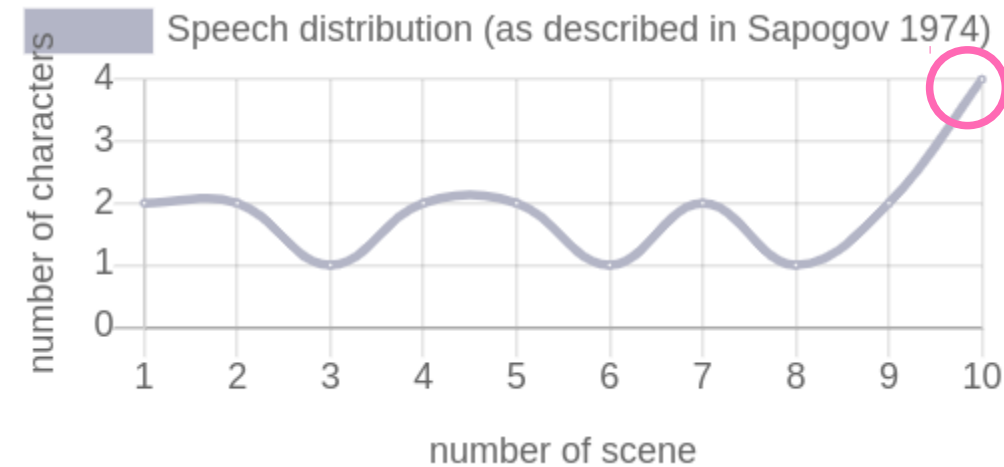
Complexité comparable dans les interactions



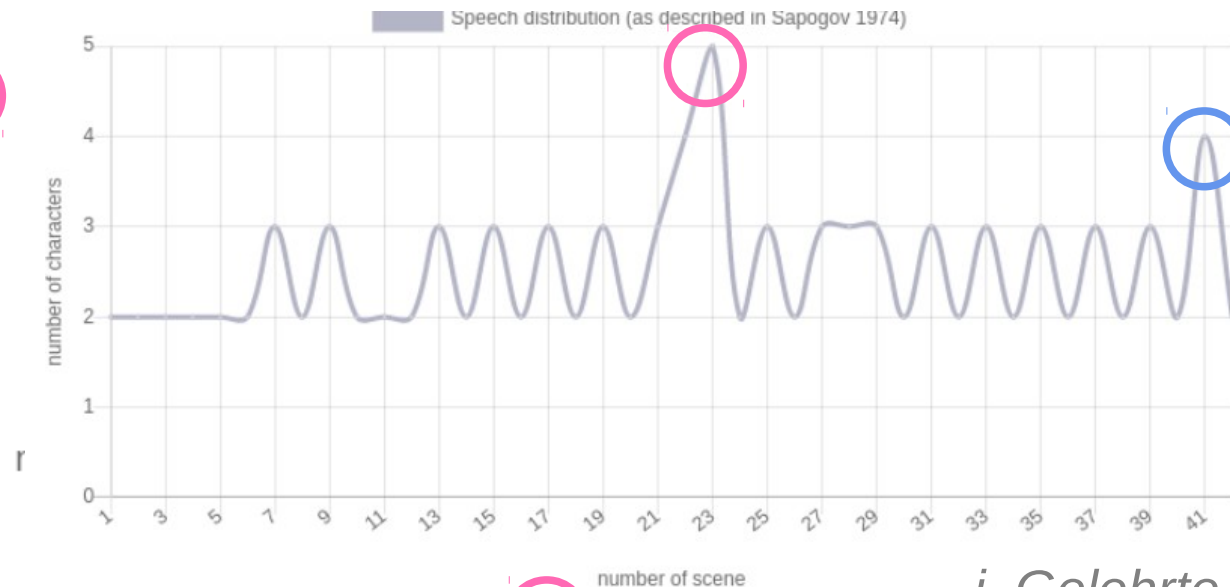
# Tendances :

## Comparaison de pièces dans GerDraCor

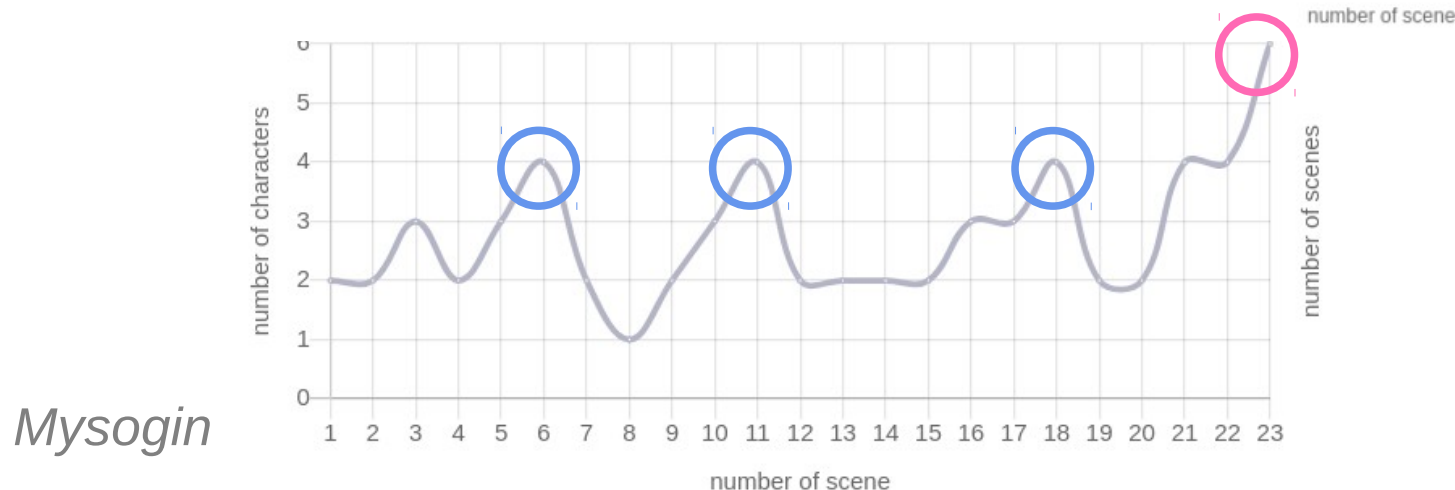
- 3 Comédies de Lessing (1747 à 1755)



*Damon*



*j. Gelehrte*



*Mysogin*

Répartition des personnages sur scène diffère

# MeThAL : Collaborations

- BNU
- Bibliothèque alsatique du Crédit Mutuel
- Comité scientifique
  - F. Fischer (DraCor)
  - N. Reiter (QuaDramA)
  - C. Schöch (Univ Trier)

(Der Vorhang fällt.)

ENDE

# Merci!

ruizfabo@unistra.fr