

Sémème au microscope : genèse et variation sémiques d'une unité lexicale

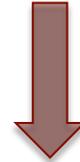
Reutenauer C.¹, Lecolle M.², Jacquy E.¹, Valette M.¹

¹ ATILF, Nancy ; ² CELTED, Metz

JADT 2010, Rome

1. Introduction

Même signifiant, signifié inconnu ou en évolution (néologie, glissement de sens)



Allocation d'un
nouveau signifié

1. Introduction

Même signifiant, signifié inconnu ou en évolution (néologie, glissement de sens)

Cadre théorique : sémantique interprétative (Rastier, 2001)

- Rôle majeur de la textualité
- Description en **sèmes**
de mécanismes sémantiques
 - ✓ Regroupements de sèmes
 - ✓ Récurrences locales de sèmes



Allocation d'un
nouveau signifié

1. Introduction

Même signifiant, signifié inconnu ou en évolution (néologie, glissement de sens)

Cadre théorique : sémantique interprétative (Rastier, 2001)

- Rôle majeur de la textualité
- Description en **sèmes** de mécanismes sémantiques
 - ✓ Regroupements de sèmes
 - ✓ Récurrences locales de sèmes

Allocation d'un nouveau signifié

Vers une automatisation
(Reutenauer *et al.*, 2009)

- Annotation en traits sémantiques
- Repérage de candidats-sèmes d'une unité lexicale, observation de leur activation

1. Introduction

Même signifiant, signifié inconnu ou en évolution (néologie, glissement de sens)

Cadre théorique : sémantique interprétative (Rastier, 2001)

- Rôle majeur de la textualité
- Description en **sèmes** de mécanismes sémantiques
 - ✓ Regroupements de sèmes
 - ✓ Récurrences locales de sèmes

Allocation d'un nouveau signifié

Vers une automatisation
(Reutenauer *et al.*, 2009)

- Annotation en traits sémantiques
- Repérage de candidats-sèmes d'une unité lexicale, observation de leur activation

Cas du nom propre *Outreau*
(Lecolle, 2007)

- Évolution de sens entre 2001 et 2006
- Emergence et disparition de facettes sémantiques

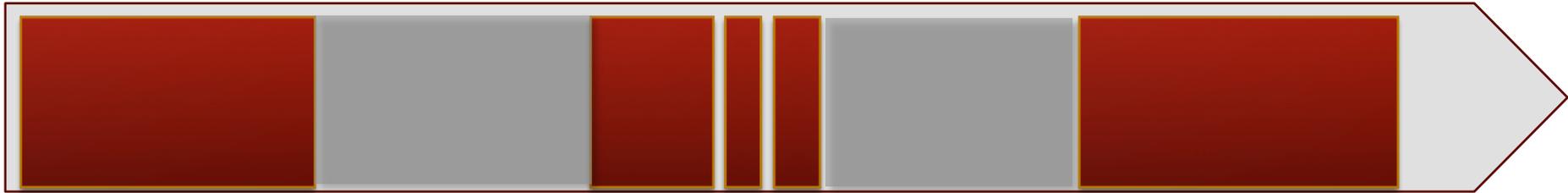
2. Ressources et outils

2.1 Ressource textuelle : corpus *Outreau*, (Lecolle 2007)

- Thème : affaire d'Outreau
- Articles de presse généraliste francophone
- ~ 400 000 occurrences de formes
- De 2001 à 2006 (5 périodes)
- (Lecolle, 2007) : constitution du corpus et analyse manuelle

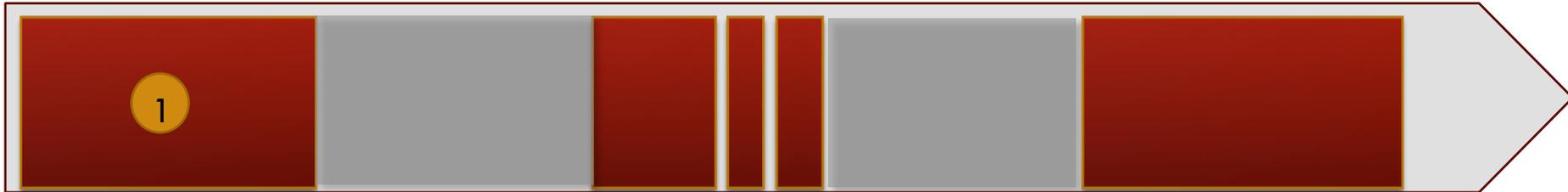
2. Ressources et outils - corpus

2.1 Evolution du sens d'*Outreau* dans les discours (Lecolle, 2007)



2. Ressources et outils - corpus

2.1 Evolution du sens d'*Outreau* dans les discours (Lecolle, 2007)



1

2001 – 2002 : « découverte » d'un réseau pédophile, arrestations

2. Ressources et outils - corpus

2.1 Evolution du sens d'*Outreau* dans les discours (Lecolle, 2007)



1

2001 – 2002 : « découverte » d'un réseau pédophile, arrestations

2

mai-juin 2004 : procès de Saint-Omer

2. Ressources et outils - corpus

2.1 Evolution du sens d'*Outreau* dans les discours (Lecolle, 2007)



1

2001 – 2002 : « découverte » d'un réseau pédophile, arrestations

2

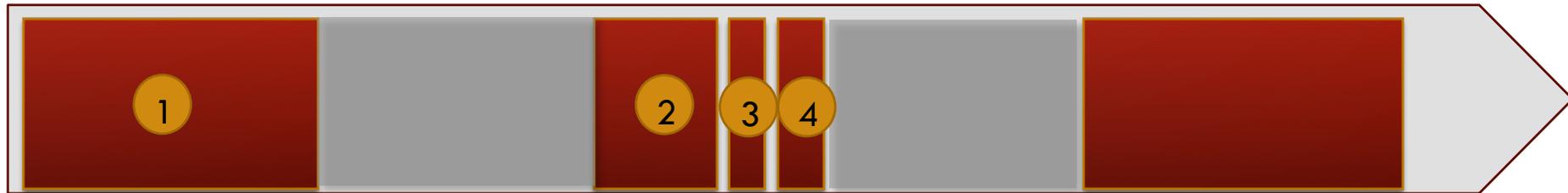
mai-juin 2004 : procès de Saint-Omer

3

1-2 juil. 2004 : attente du verdict

2. Ressources et outils - corpus

2.1 Evolution du sens d'*Outreau* dans les discours (Lecolle, 2007)



1

2001 – 2002 : « découverte » d'un réseau pédophile, arrestations

2

mai-juin 2004 : procès de Saint-Omer

3

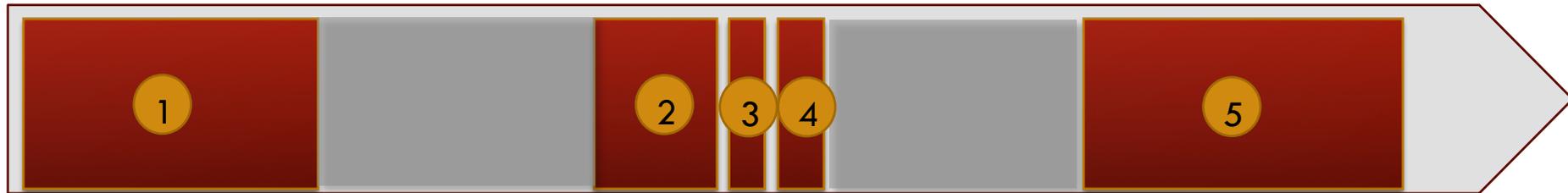
1-2 juil. 2004 : attente du verdict

4

3-8 juil. 2004 : verdict

2. Ressources et outils - corpus

2.1 Evolution du sens d'*Outreau* dans les discours (Lecolle, 2007)



1

2001 – 2002 : « découverte » d'un réseau pédophile, arrestations

2

mai-juin 2004 : procès de Saint-Omer

3

1-2 juil. 2004 : attente du verdict

4

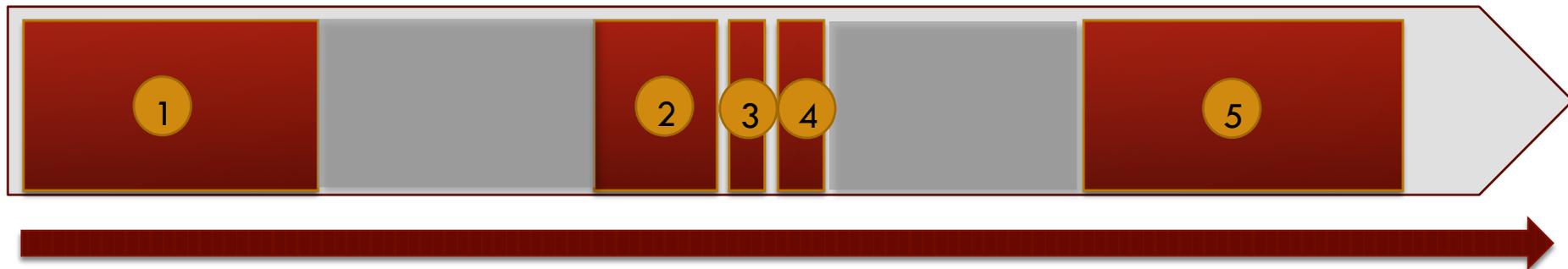
3-8 juil. 2004 : verdict

5

déc. 2005 – avr. 2006 : procès en appel à Paris ; commission d'enquête parlementaire

2. Ressources et outils - corpus

2.1 Evolution du sens d'*Outreau* dans les discours (Lecolle, 2007)



lieu habité

l'erreur judiciaire par excellence

1

2001 – 2002 : « découverte » d'un réseau pédophile, arrestations

2

mai-juin 2004 : procès de Saint-Omer

3

1-2 juil. 2004 : attente du verdict

4

3-8 juil. 2004 : verdict

5

déc. 2005 – avr. 2006 : procès en appel à Paris ; commission d'enquête parlementaire

2. Ressources et outils

2.2 Constitution d'une image sémique du corpus

(Valette, Estacio-Moreno, Petitjean, Jacquey, 2006), SEMY (Grzesitchak, 2008)

Version lexicale



~ 400 000 occurrences
de formes

... la cour se résout à libérer sept
accusés. Ces hommes...

2. Ressources et outils

2.2 Constitution d'une image sémique du corpus

(Valette, Estacio-Moreno, Petitjean, Jacquey, 2006), SEMY (Grzesitchak, 2008)

Version lexicale



~ 400 000 occurrences
de formes

... la cour se résout à libérer sept
accusés. Ces hommes...

*élimination des mots-outils
étiquetage, lemmatisation*

... ■ cour se résoudre ■ libérer sept
accusé. ■ homme...

2. Ressources et outils

2.2 Constitution d'une image sémique du corpus

(Valette, Estacio-Moreno, Petitjean, Jacquey, 2006), SEMY (Grzesitchak, 2008)

Version lexicale



~ 400 000 occurrences
de formes

... la cour se résout à libérer sept
accusés. Ces hommes...

*élimination des mots-outils
étiquetage, lemmatisation*

... ■ cour se résoudre ■ libérer sept
accusé. ■ homme...

définition lexicographique (TLFi)

→ **ACCUSÉ**¹, ÉE, part. passé et subst.

I.— *Part. passé* du verbe *accuser**.

II.— *Emploi subst. masc. et fém., DR. et lang. commune.* Celui, celle qui est déféré(e) à la justice comme auteur d'une infraction, d'un délit ou d'un crime :

- 1. ... la détention pour dettes est un fait judiciaire si rare en province que, dans la plupart des villes de France, il n'existe pas de maison d'arrêt. Dans ce cas, le débiteur est écroué à la prison où l'on incarcère les *inculpés*, les

2. Ressources et outils

2.2 Constitution d'une image sémique du corpus

(Valette, Estacio-Moreno, Petitjean, Jacquey, 2006), SEMY (Grzesitchak, 2008)

Version lexicale



~ 400 000 occurrences
de formes

... la cour se résout à libérer sept
accusés. Ces hommes...

élimination des mots-outils
étiquetage, lemmatisation

... ■ cour se résoudre ■ libérer sept
accusé. ■ homme...

définition lexicographique (TLFi)

→ **ACCUSÉ**¹, ÉE, part. passé et subst.

I.— Part. passé du verbe *accuser**.

II.— Emploi subst. masc. et fém., DR. et lang. commune. **Celui, celle qui est déféré(e) à la justice comme auteur d'une infraction, d'un délit ou d'un crime :**

- 1. ... la détention pour dettes est un fait judiciaire si rare en province que, dans la plupart des villes de France, il n'existe pas de maison d'arrêt. Dans ce cas, le débiteur est écroué à la prison où l'on incarcère les *inculpés*, les

/accusé/ /déférer/
/justice/ /auteur/
/infraction/ /délit/
/crime/

candidats-sèmes
= lemmes des mots
composant la définition

2. Ressources et outils

2.2 Constitution d'une image sémique du corpus

(Valette, Estacio-Moreno, Petitjean, Jacquy, 2006), SEMY (Grzesitchak, 2008)

Version lexicale



~ 400 000 occurrences
de formes

... la cour se résout à libérer sept
accusés. Ces hommes...

*élimination des mots-outils
étiquetage, lemmatisation*

... ■ cour se résoudre ■ libérer sept
accusé. ■ homme...

définition lexicographique (TLFi)

→ **ACCUSÉ**¹, ÉE, part. passé et subst.

I.— Part. passé du verbe *accuser**.

II.— Emploi subst. masc. et fém., DR. et lang. commune. **Celui, celle qui est déféré(e) à la justice comme auteur d'une infraction, d'un délit ou d'un crime :**

- 1. ... la détention pour dettes est un fait judiciaire si rare en province que, dans la plupart des villes de France, il n'existe pas de maison d'arrêt. Dans ce cas, le débiteur est écroué à la prison où l'on incarcère les *inculpés*, les

Version sémique



~ 10 millions occurrences
de candidats-sèmes

/accusé/ /déférer/
/justice/ /auteur/
/infraction/ /délit/
/crime/

candidats-sèmes

= lemmes des mots

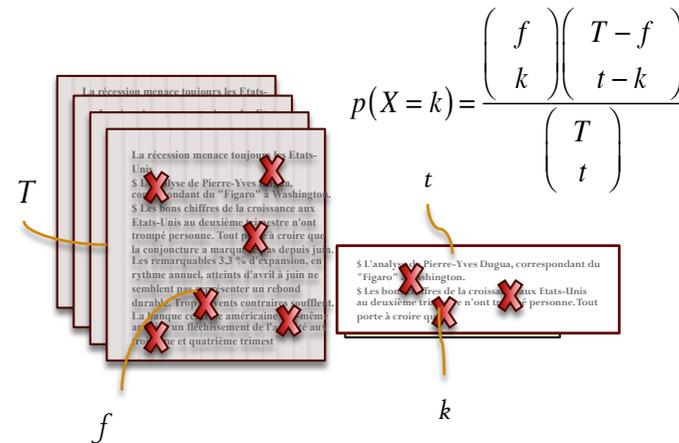
composant la définition

2. Ressources et outils

2.3 Outil lexicométrique

Lafon (1984), Salem *et al.* (2003)

- Calcul de spécificités
 - comparaisons partie - tout
 - logiciel Lexico 3

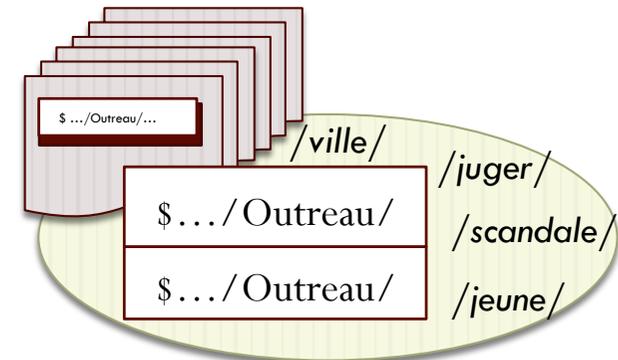
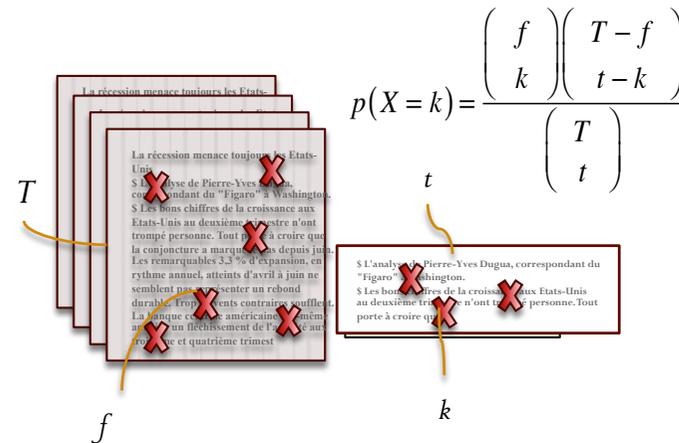


2. Ressources et outils

2.3 Outil lexicométrique

Lafon (1984), Salem *et al.* (2003)

- Calcul de spécificités
 - comparaisons partie - tout
 - logiciel Lexico 3
- Sous-corpus des voisinages d'*Outreau*
 - paragraphes contenant *Outreau*
 - jeu sur
 - des partitions en temps (périodes)
 - différents niveaux (lexical / sémique)



3. Sélection de candidats-sèmes

3.1 Trois critères à vérifier

3. Sélection de candidats-sèmes

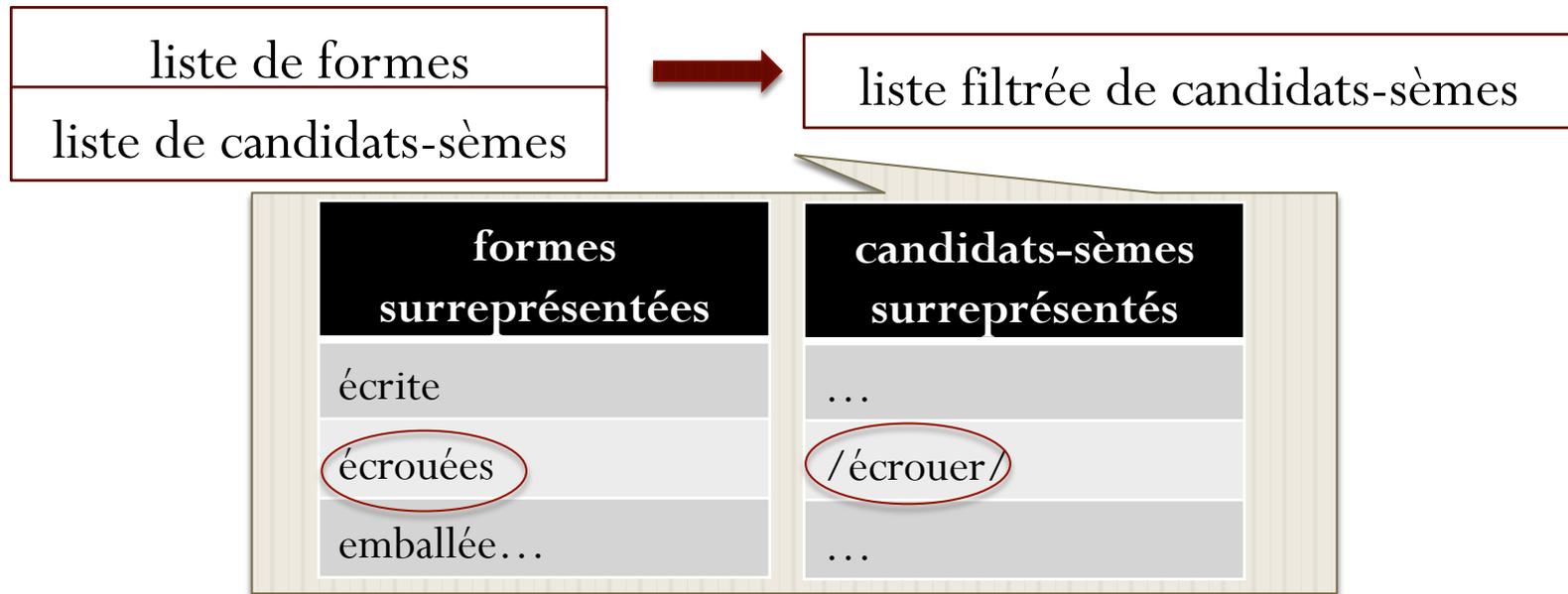
3.1 Trois critères à vérifier

- Critère d'activation : spécificité supérieure à un seuil positif

3. Sélection de candidats-sèmes

3.1 Trois critères à vérifier

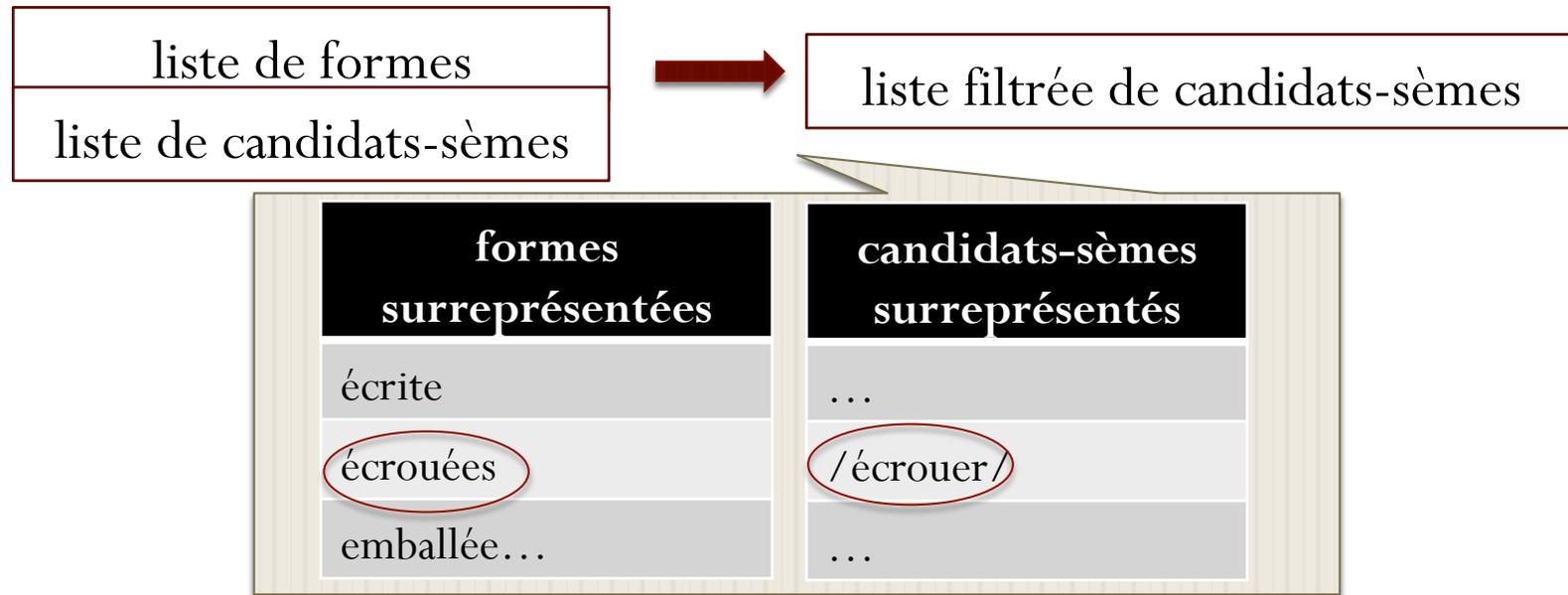
- Critère d'activation : spécificité supérieure à un seuil positif
- Critère de co-actualisation



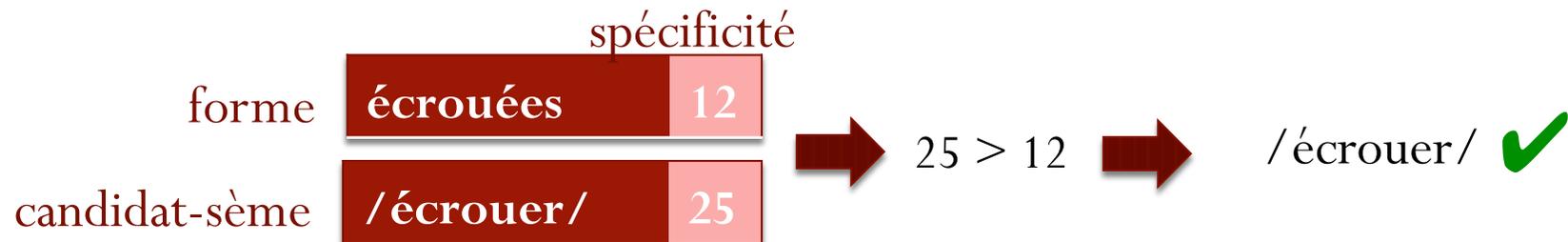
3. Sélection de candidats-sèmes

3.1 Trois critères à vérifier

- Critère d'activation : spécificité supérieure à un seuil positif
- Critère de co-actualisation

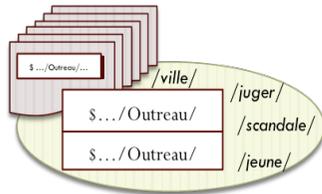


- Critère de renforcement

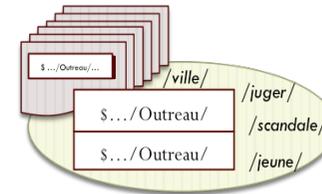


3. Sélection de candidats-sèmes

3.2 Variation diachronique des ensembles retenus



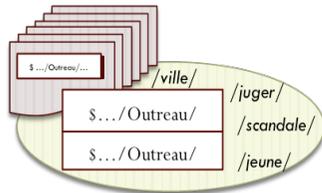
Deux méthodes
de construction
de sémème



3. Sélection de candidats-sèmes

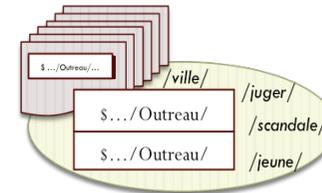
3.2 Variation diachronique des ensembles retenus

Par période (enrichissement)



Deux méthodes
de construction
de sémème

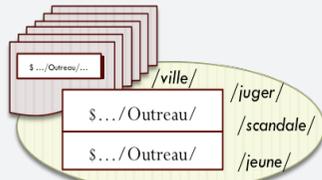
Globale (activation - inhibition)



3. Sélection de candidats-sèmes

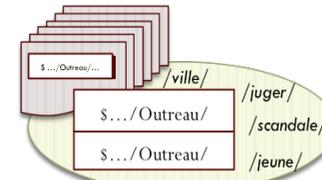
3.2 Variation diachronique des ensembles retenus

Par période (enrichissement)

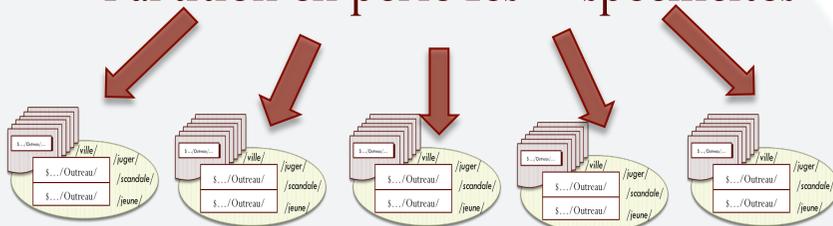


Deux méthodes
de construction
de sémème

Globale (activation - inhibition)



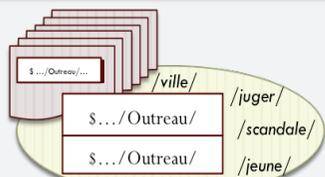
Partition en périodes + spécificités



3. Sélection de candidats-sèmes

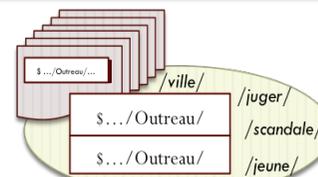
3.2 Variation diachronique des ensembles retenus

Par période (enrichissement)

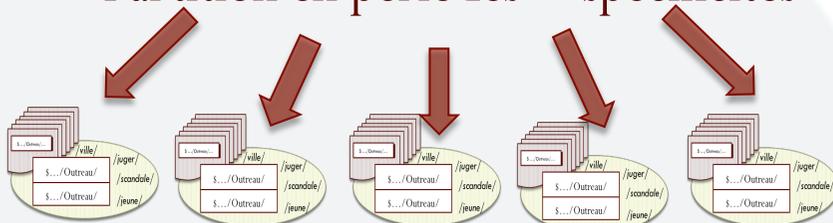


Deux méthodes
de construction
de sémème

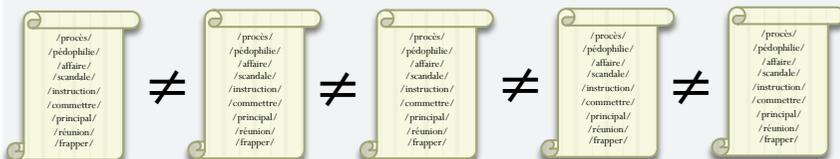
Globale (activation - inhibition)



Partition en périodes + spécificités



Sélection de candidats-sèmes

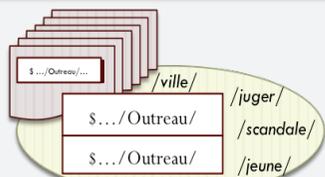


cinq listes de candidats-sèmes

3. Sélection de candidats-sèmes

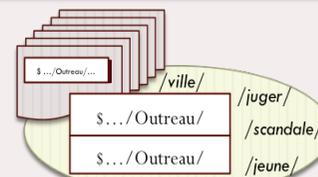
3.2 Variation diachronique des ensembles retenus

Par période (enrichissement)

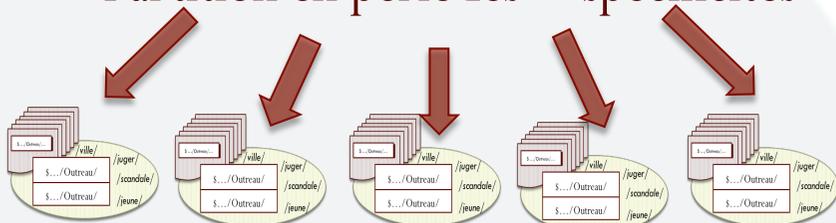


Deux méthodes
de construction
de sémème

Globale (activation - inhibition)



Partition en périodes + spécificités

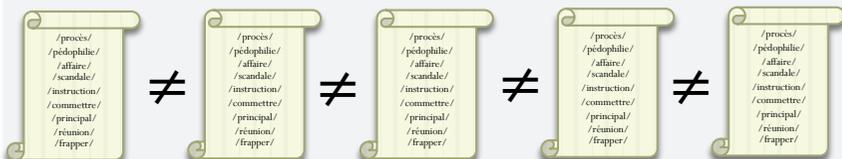


Sélection de candidats-sèmes



une liste de
candidats-sèmes

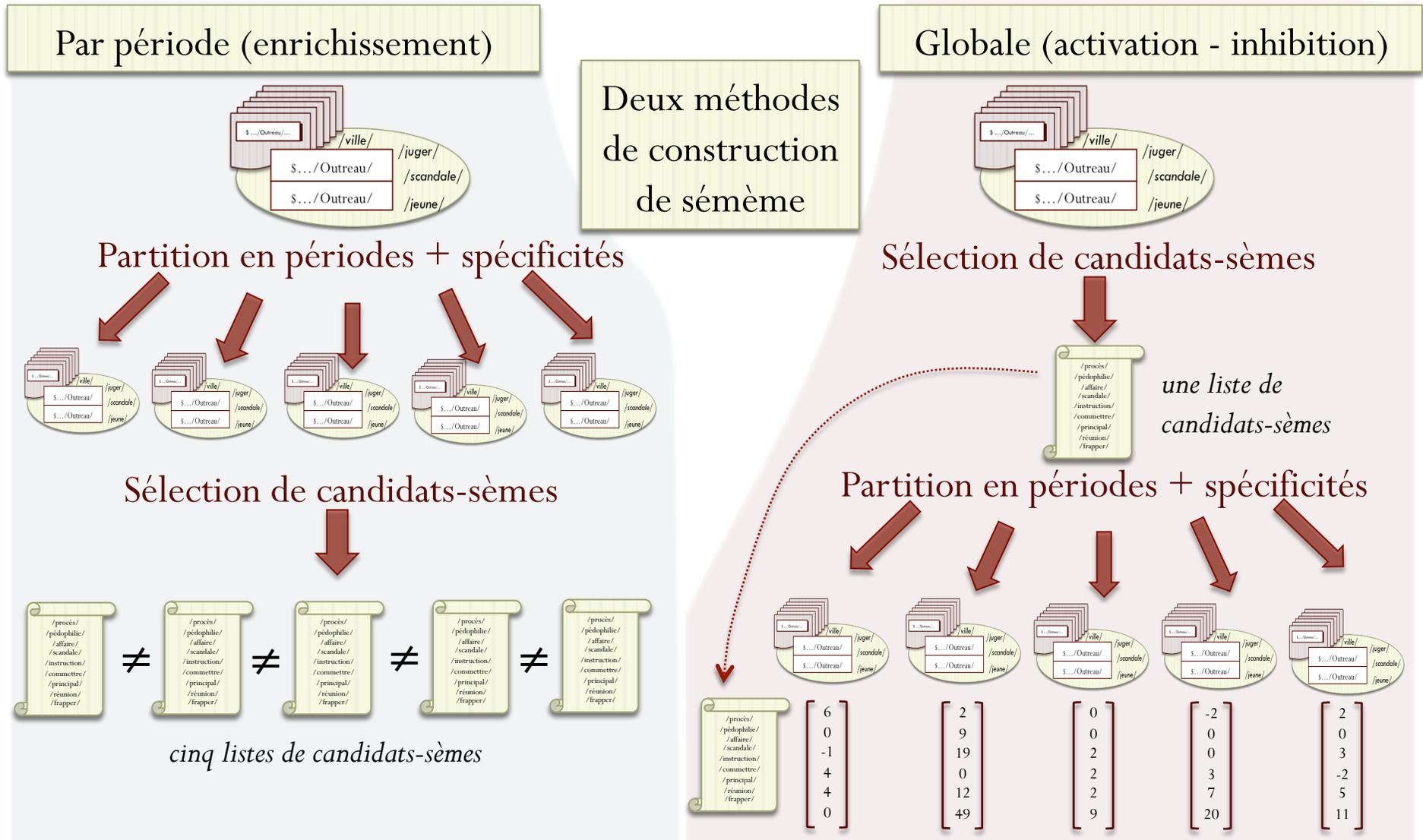
Sélection de candidats-sèmes



cinq listes de candidats-sèmes

3. Sélection de candidats-sèmes

3.2 Variation diachronique des ensembles retenus



4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation



Comment ?

- Constitution manuelle de classes
 - ✓ Regroupements de candidats-sèmes
 - ✓ Dénomination des classes

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation



Comment ?

- Constitution manuelle de classes
 - ✓ Regroupements de candidats-sèmes
 - ✓ Dénomination des classes
- Intervention de deux linguistes

✓ Sémème par période

✓ Sémème global

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation



Comment ?

- Constitution manuelle de classes
 - ✓ Regroupements de candidats-sèmes
 - ✓ Dénomination des classes
- Intervention de deux linguistes

✓ Sémème par période

✓ Sémème global

Exemples de classes

LIEU – VILLE ET HABITANTS (période 1)

lieu immeuble enfant ouvrier
misère cité jardin couple taxi
grève nord habitant fille
chauffeur habiter adulte
infirmier appartement membre
maire banlieue frère

JUDICIAIRE

acquitter affaire assise audience
cour débat débattre délibérer
écrouer emprisonner huis
institution instruction judiciaire
justice magistrat pédophilie
plaidoirie procès saint-omer
verdict

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation



Comment ?

- Constitution manuelle de classes
 - ✓ Regroupements de candidats-sèmes
 - ✓ Dénomination des classes
- Intervention de deux linguistes

✓ Sémème par période

✓ Sémème global

Exemples de classes

LIEU – VILLE ET HABITANTS (période 1)

lieu immeuble enfant ouvrier
misère cité jardin couple taxi
grève nord habitant fille
chauffeur habiter adulte
infirmier appartement membre
maire banlieue frère

JUDICIAIRE

acquitter affaire assise **audience**
cour débat débattre délibérer
écrouer emprisonner huis
institution instruction judiciaire
justice magistrat pédophilie
plaidoirie **procès** saint-omer
verdict

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation



Comment ?

- Constitution manuelle de classes
 - ✓ Regroupements de candidats-sèmes
 - ✓ Dénomination des classes
- Intervention de deux linguistes

✓ Sémème par période

✓ Sémème global

Exemples de classes

LIEU – VILLE ET HABITANTS (période 1)

lieu immeuble enfant ouvrier
misère cité jardin couple taxi
grève nord habitant fille
chauffeur habiter adulte
infirmier appartement membre
maire banlieue frère

JUDICIAIRE

acquitter affaire assise audience
cour **débat débattre** délibérer
écrouer emprisonner huis
institution instruction judiciaire
justice magistrat pédophilie
plaidoirie procès saint-omer
verdict

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation



Comment ?

- Constitution manuelle de classes
 - ✓ Regroupements de candidats-sèmes
 - ✓ Dénomination des classes
- Intervention de deux linguistes

✓ Sémème par période

✓ Sémème global

4. Analyse des résultats

4.1 Classification des candidats-sèmes

Répartition des candidats-sèmes en classes



Pourquoi ?

- Confrontation aux résultats de (Lecolle, 2007) pour validation



Comment ?

- Constitution manuelle de classes
 - ✓ Regroupements de candidats-sèmes
 - ✓ Dénomination des classes
- Intervention de deux linguistes

✓ Sémème par période

✓ Sémème global



*Observation
des classes*

- Apparition, disparition, récurrence de classes
- Poids de la classe
 - ✓ Proportion de candidats-sèmes affectés
 - ✓ Valeurs de spécificités (moyenne)

4. Analyse des résultats

4.2 Sémème par période : classe **LIEU**

4. Analyse des résultats

4.2 Sémème par période : classe LIEU

Répartition des candidats par classe

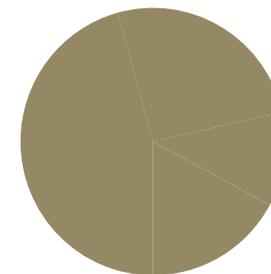
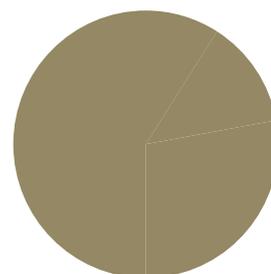
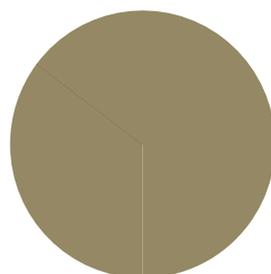
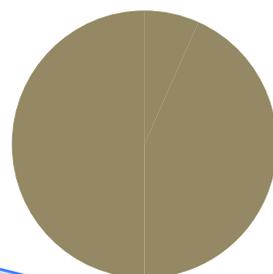
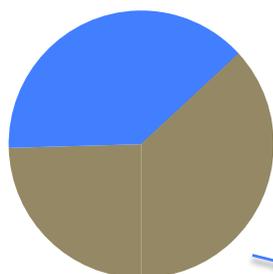
Période 1

Période 2

Période 3

Période 4

Période 5



LIEU (VILLE ET HABITANTS)

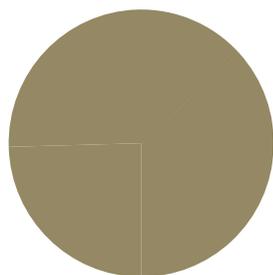
lieu immeuble enfant ouvrier
misère cité jardin couple taxi
grève nord habitant fille
chauffeur habiter adulte infirmier
appartement membre maire
banlieue frère

4. Analyse des résultats

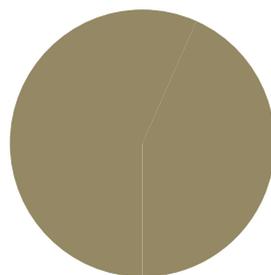
4.2 Sémème par période : classe POLITIQUE

Répartition des candidats par classe

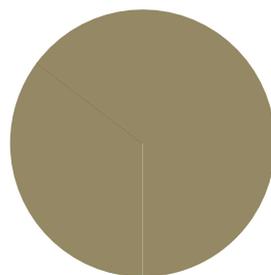
Période 1



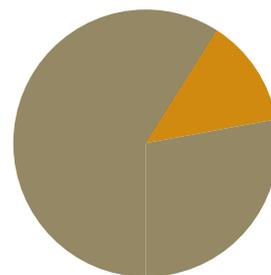
Période 2



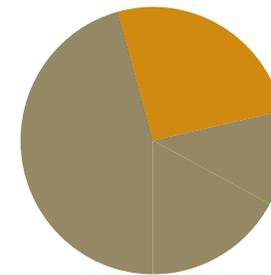
Période 3



Période 4



Période 5



POLITIQUE

réforme président
national présidence
présidentiel polémique
traiter préventif

POLITIQUE

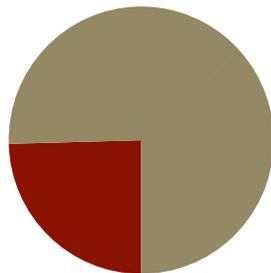
public parlementaire
pouvoir politique société
député enquête réforme
texte

4. Analyse des résultats

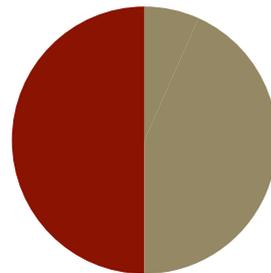
4.2 Sémème par période : classe **JUDICIAIRE**

Répartition des candidats par classe

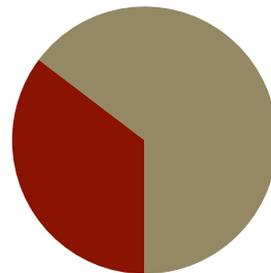
Période 1



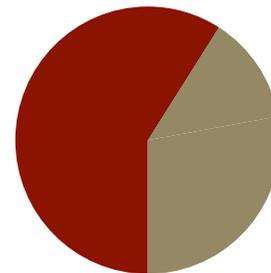
Période 2



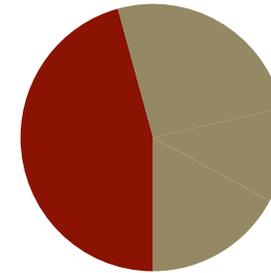
Période 3



Période 4



Période 5



JUDICIAIRE ET POLICIER

écrouer police
arrestation
démanteler
procureur
incarcération
suspicion soupçon
incarcérer dénoncer
huissier suspect
identité matériel

JUDICIAIRE ET POLICIER

procès instruction séance
débatte assise justice
désigner cour affaire
accusation examiner
audience suspendu
déposition entendre ouvert
débat disculper clamer
innocence justifier
plaidoirie dénonciation
témoigner huis innocenter
initial principal fond
psychologie

JUDICIAIRE ET POLICIER

affaire
plaidoirie
avocat
plaider juré
acquittement

JUDICIAIRE ET POLICIER

jugement peine judiciaire
délibération prison détention
coupable affaire commettre
remise juge tribunal faute
justice loi provisoire instruire
condamnation droit criminel
procès prononcer innocent
assise chancelier coupable
instruction juger magistrat
substitut conclusion institution
juré magistrature réclamation
verdict

JUDICIAIRE

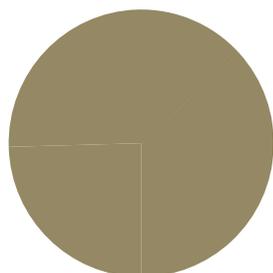
enquête judiciaire
jugement coupable
innocent libérer
provisoire recueillir
juge libéré libre
procureur sanction
délibération
délibérer
responsabilité

4. Analyse des résultats

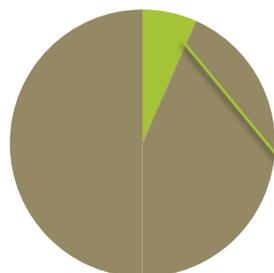
4.2 Sémème par période : classes DÉSASTRE et FIASCO

Répartition des candidats par classe

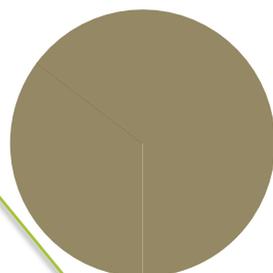
Période 1



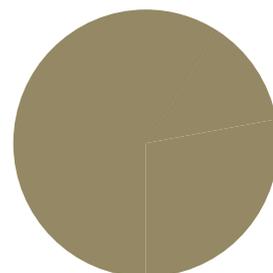
Période 2



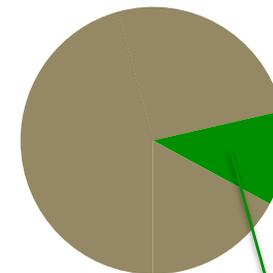
Période 3



Période 4



Période 5



DÉSASTRE

FIASCO

DÉSASTRE

scandale
désastre
incohérence
naufnage

FIASCO

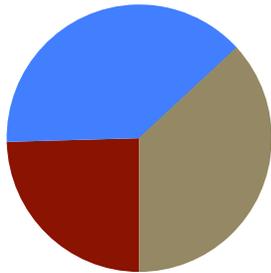
faute
naufnage
drame
faillite

4. Analyse des résultats

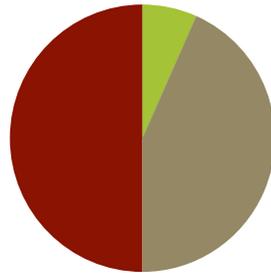
4.2 Sémème par période : évolution des classes

Répartition des candidats par classe

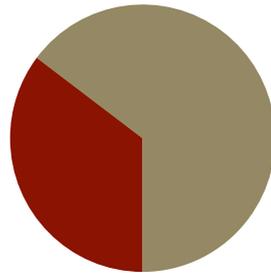
Période 1



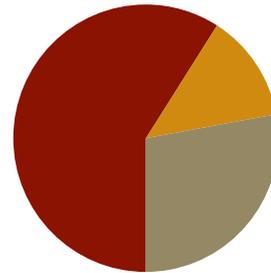
Période 2



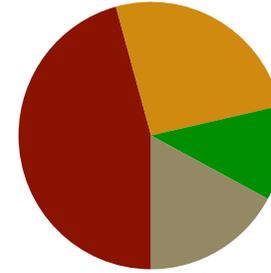
Période 3



Période 4



Période 5

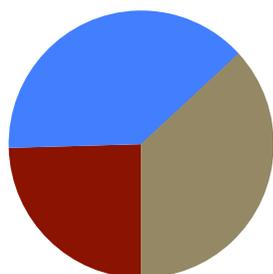


4. Analyse des résultats

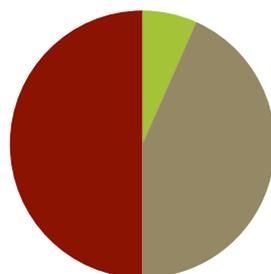
4.2 Sémème par période : évolution des classes

Répartition des candidats par classe

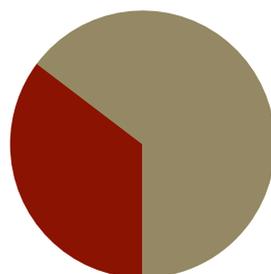
Période 1



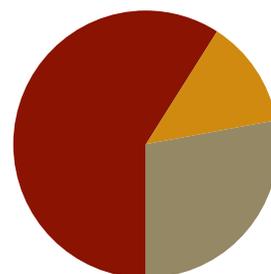
Période 2



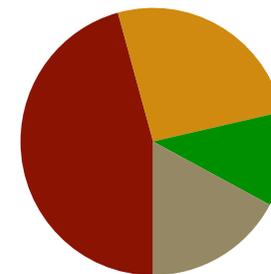
Période 3



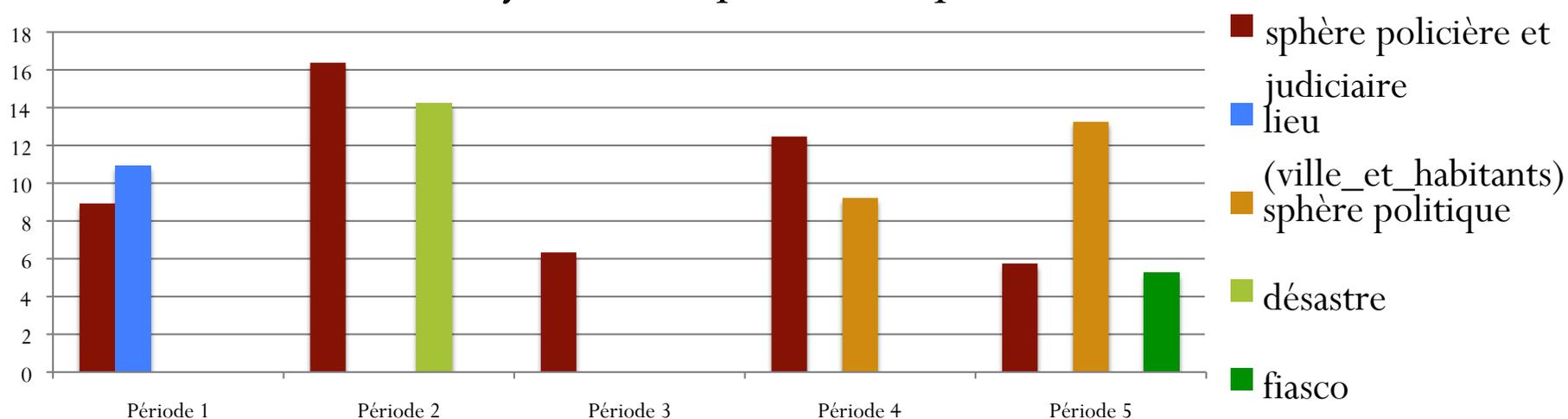
Période 4



Période 5



Moyenne de spécificités par classe

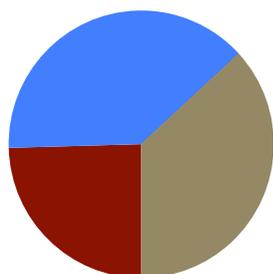


4. Analyse des résultats

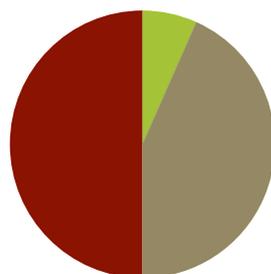
4.2 Sémème par période : évolution des classes

Répartition des candidats par classe

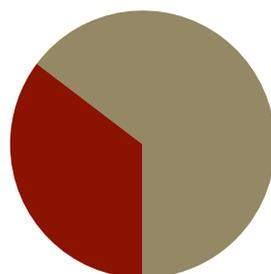
Période 1



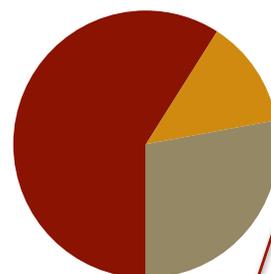
Période 2



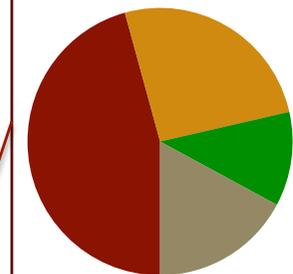
Période 3



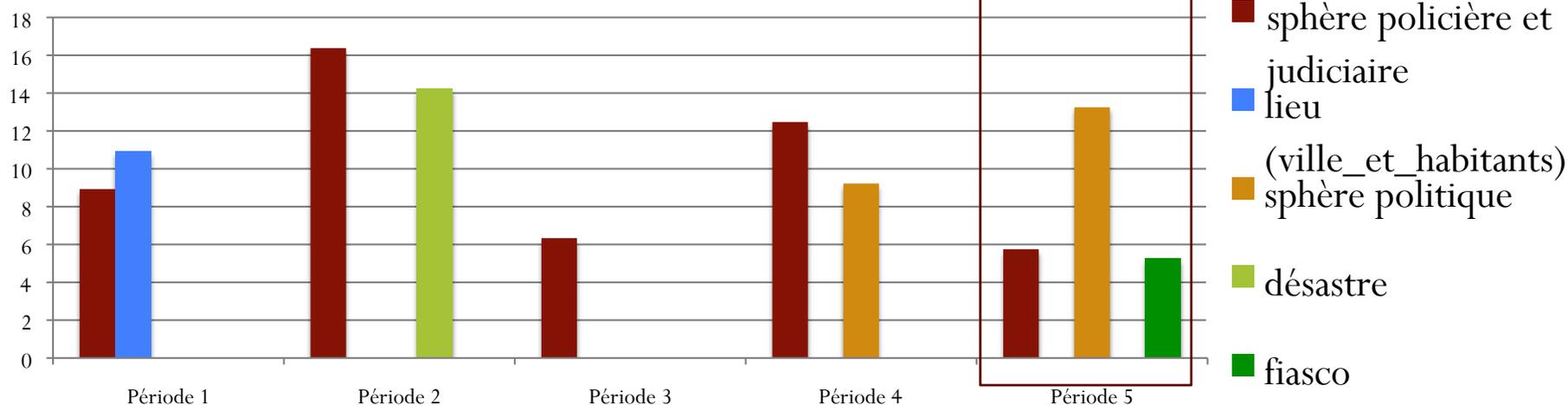
Période 4



Période 5



Moyenne de spécificités par classe



4. Analyse des résultats

4.2 Sémème global

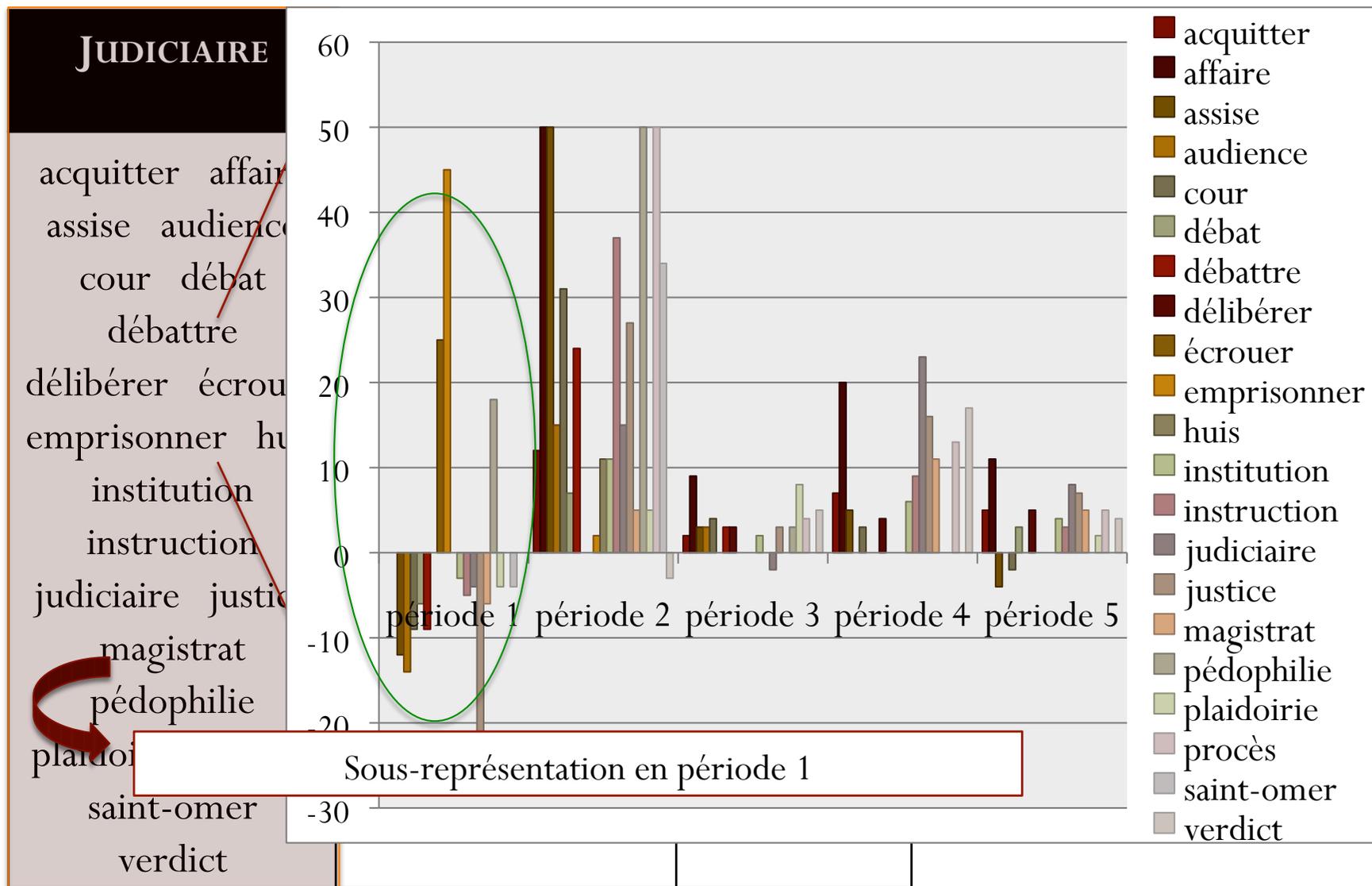
JUDICIAIRE	LIEU (VILLE ET HABITANTS)	DRAME - DÉSASTRE
acquitter affaire assise audience cour débat débattre délibérer écrouer emprisonner huis institution instruction judiciaire justice magistrat pédophilie plaidoirie procès saint-omer verdict	cit� habitant ouvrier personne	d�sastre dramatique drame fiasco horreur nauffrage scandale tragique

1 s m me pour
les 5 p riodes

Pas de classe
POLITIQUE

4. Analyse des résultats

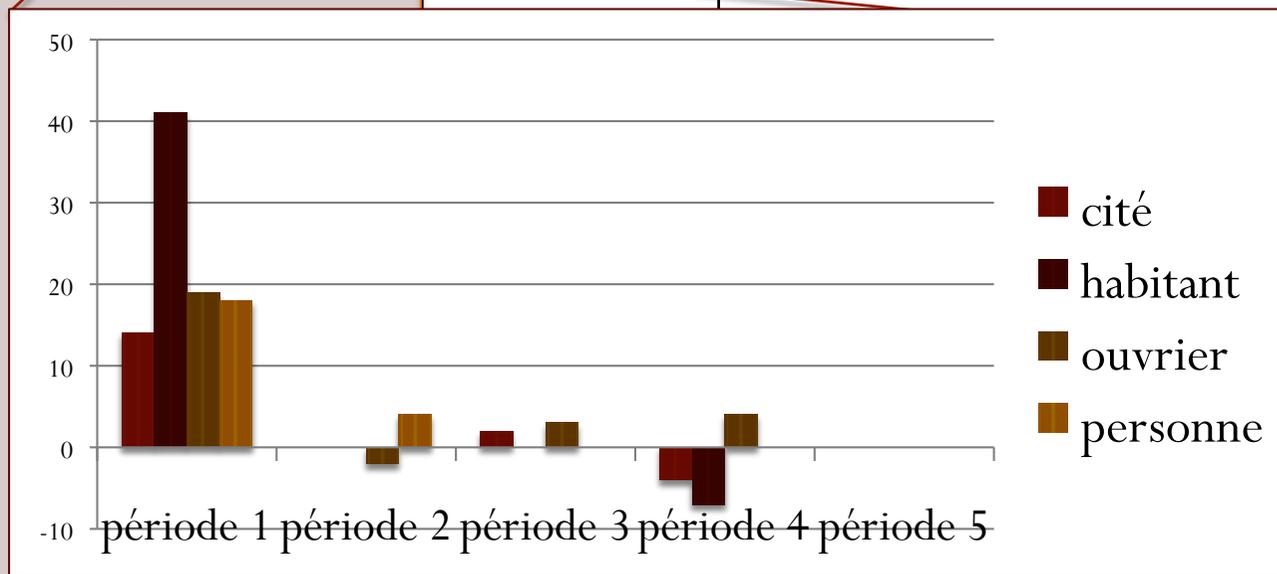
4.2 Sémème global : classe JUDICIAIRE



4. Analyse des résultats

4.2 Sémème global : classe LIEU

JUDICIAIRE	LIEU (VILLE ET HABITANTS)	DRAME - DÉSASTRE
acquitter affaire	cité	désastre
assise audience	habitant	dramatique
cour débat	ouvrier	drame
débattre	personne	fiasco
délibérer écrouer		
emprisonner huis		
institution		
instruction		
judiciaire justice		
magistrat		
pédophilie		
plaidoirie procès		
saint-omer		



Evolution des spécificités de la classe LIEU conforme à l'analyse manuelle

4. Analyse des résultats

4.3 Bilan : confrontation des deux approches

Sémème par période	<i>Outreau</i>	Sémème global
<u>Convergence avec l'analyse manuelle</u>		

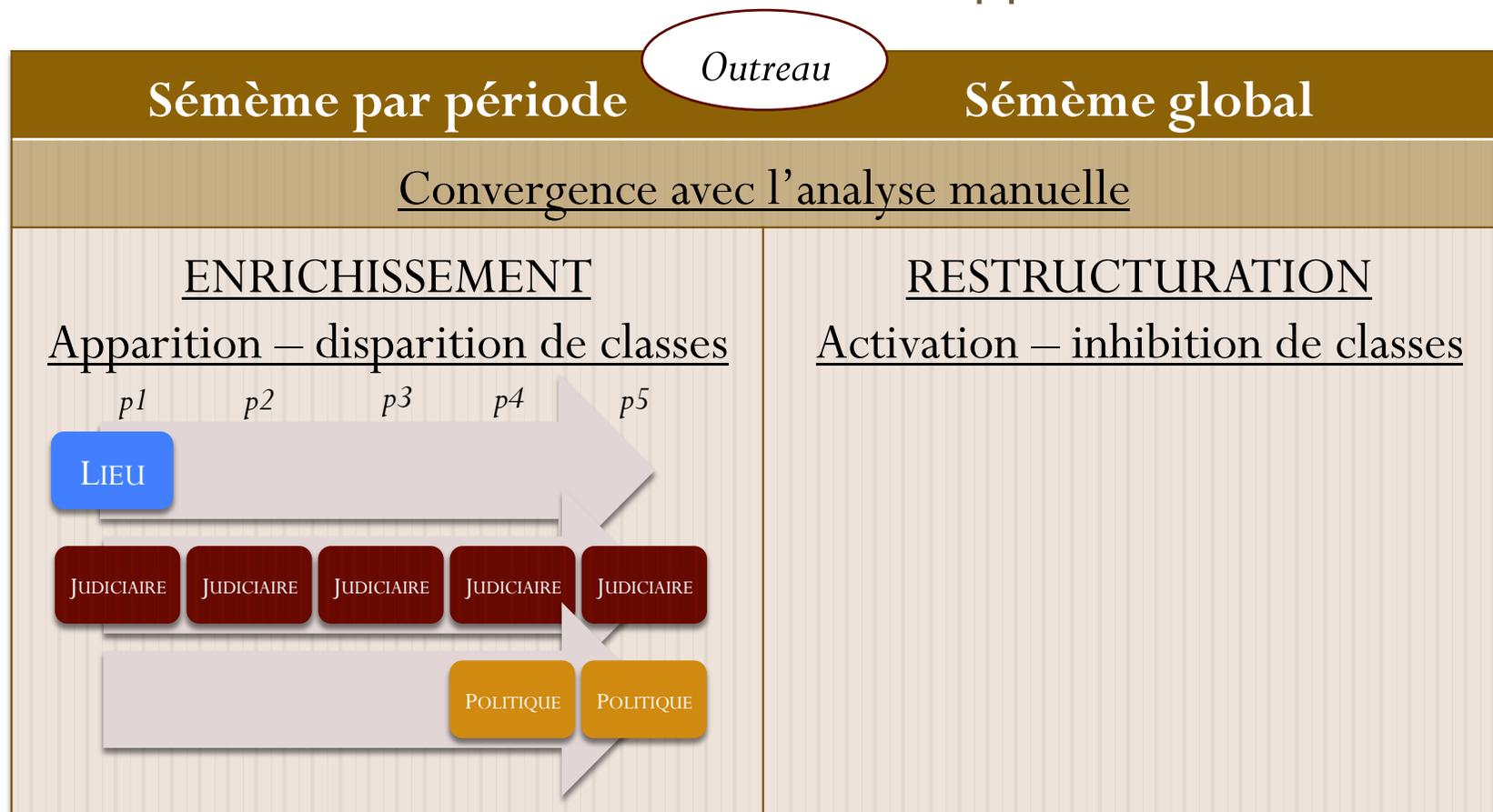
4. Analyse des résultats

4.3 Bilan : confrontation des deux approches

Sémème par période	Sémème global
<u>Convergence avec l'analyse manuelle</u>	
<u>ENRICHISSEMENT</u> <u>Apparition – disparition de classes</u>	<u>RESTRUCTURATION</u> <u>Activation – inhibition de classes</u>

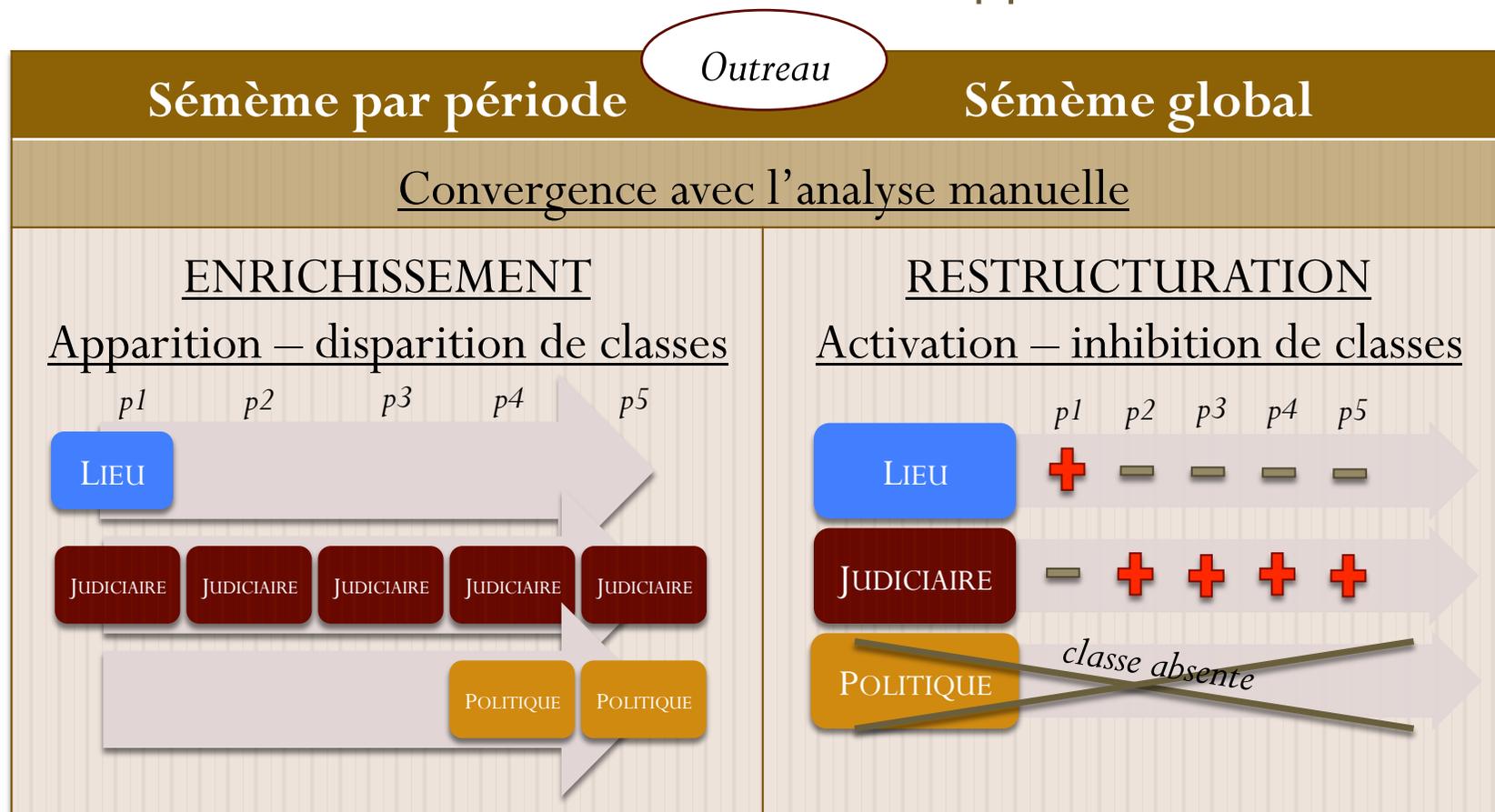
4. Analyse des résultats

4.3 Bilan : confrontation des deux approches



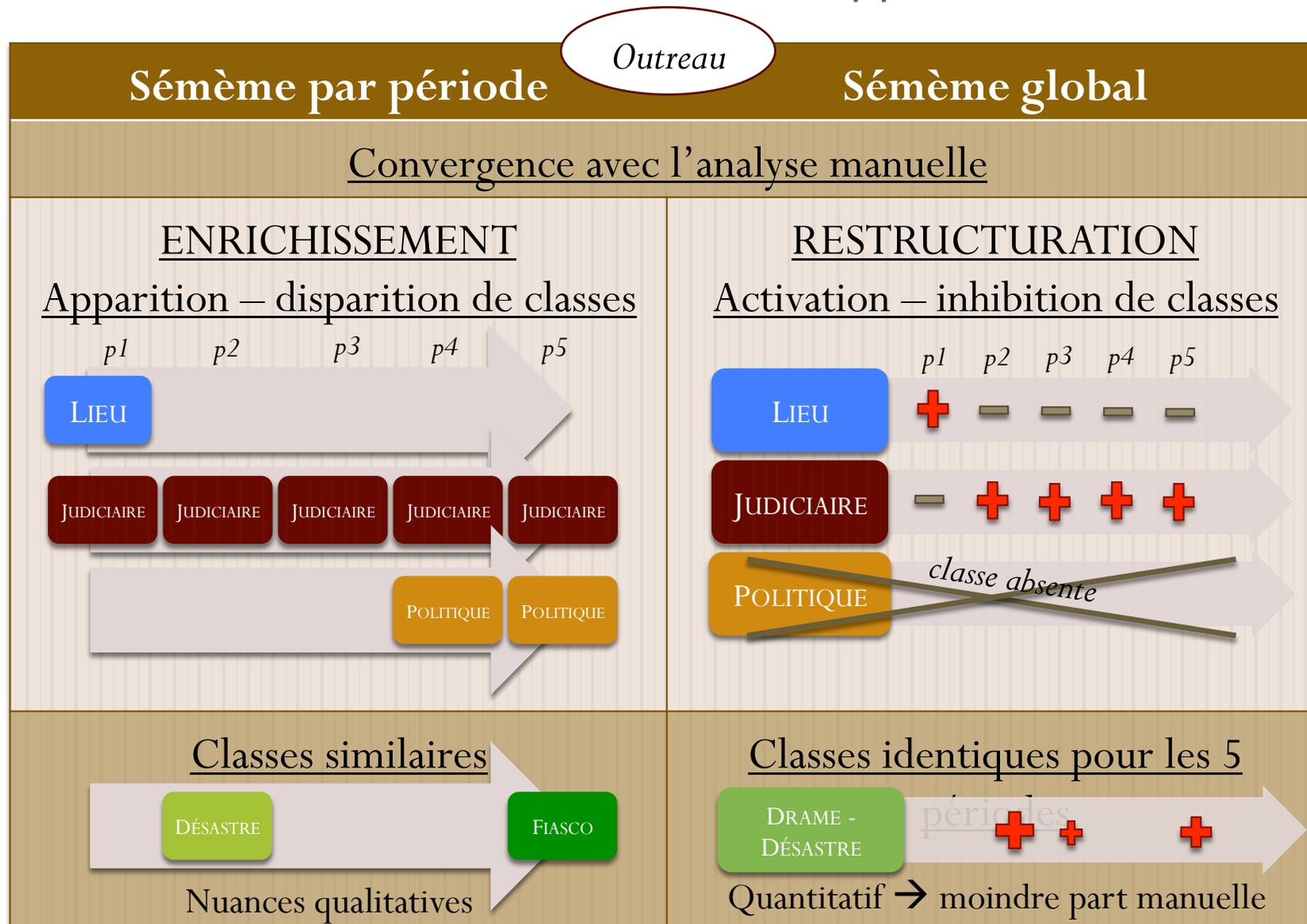
4. Analyse des résultats

4.3 Bilan : confrontation des deux approches



4. Analyse des résultats

4.3 Bilan : confrontation des deux approches



5. Bilan et perspectives

- Résultats valides
 - Convergence des deux méthodes de construction de sémèmes
 - Apport propre à chacune
- Limites
 - pas d'émergence de la dimension épistémique
 - filtrage : élimination de candidats pertinents

5. Bilan et perspectives

- Résultats valides
 - Convergence des deux méthodes de construction de sémèmes
 - Apport propre à chacune
- Limites
 - pas d'émergence de la dimension épistémique
 - filtrage : élimination de candidats pertinents

- Structuration de sémème
 - recours à d'autres niveaux (syntaxe, morphologie)
 - jeu sur la granularité (de générique à spécifique)
- Modélisation de processus sémiques de diffusion et sommation