Constitution d'une base de données issue de greffes de tribunaux de grande instance

Thomas Soubiran [1]

[1] Ceraps - UMR 8026 du CNRS

Demi-journée plateforme DATA

« Approches quantitatives des données textuelles »

Lille, 18 octobre 2016

Introduction :

Présentation d'un exemple de traitement numérique de données textuelles :

- Analyse **syntaxique** et non sémantique
- **Extraction d'informations** contenues dans des documents papiers pour constituer une **base de données**

L'enquête Comparutions immédiates

La collecte réalisée dans le cadre de l'enquête Comparutions immédiates

L'enquête Comparutions immédiates

La collecte réalisée dans le cadre de l'enquête Comparutions immédiates :

- Enquête réalisée en réponse à un appel à projet de la Mission Droit et Justice publié en 2012 portant sur le recours aux comparutions immédiates (CI) par AC Douillet, T. Léonard, T. Soubiran et H. Yazdanpanah
- Visait notamment à analyser les déterminants du recours aux CI et des décisions qui y sont prononcées

L'enquête Comparutions immédiates

La collecte réalisée dans le cadre de l'enquête Comparutions immédiates :

- Enquête réalisée en réponse à un appel à projet de la Mission Droit et Justice publié en 2012 portant sur le recours aux comparutions immédiates (CI) par AC Douillet, T. Léonard, T. Soubiran et H. Yazdanpanah
- Visait notamment à analyser les déterminants du recours aux CI et des décisions qui y sont prononcées
- Collecte de données :
  - **Volet qualitatif :** entretiens auprès de magistrats du siège et du parquet ainsi qu'auprès d'avocats
  - **Volet statistique :** constitution d'une base de données renseignant notamment le type de procédure, le nombre de prévenus par jugement, différentes caractéristiques démographiques et sociales des prévenus, leur(s) chef(s) d'inculpation ainsi que la décision du tribunal les concernant

L'enquête Comparutions immédiates

La collecte réalisée dans le cadre de l'enquête Comparutions immédiates :

- Enquête réalisée en réponse à un appel à projet de la Mission Droit et Justice publié en 2012 portant sur le recours aux comparutions immédiates (CI) par AC Douillet, T. Léonard, T. Soubiran et H. Yazdanpanah
- Visait notamment à analyser les déterminants du recours aux CI et des décisions qui y sont prononcées
- Collecte de données
- Le volet statistique portant notamment sur l'étude de l'orientation en CI et ses évolutions dans le temps, une attention toute particulière a été prêtée à la bonne **représentation** des procédures pénales (les tests d'hypothèse nécessitaient la modélisation des effets d'interaction avec le temps)

Collecter des données dans des greffes de tribunaux :

Le ministère de la Justice agrège et diffuse des données remontant de l'activité des tribunaux.

Malheureusement, les informations disponibles étaient trop limitées pour les analyses prévues. Les informations pertinentes ne se trouvant que dans les minutes de jugement, il a fallu retourner à la source dans les archives des tribunaux

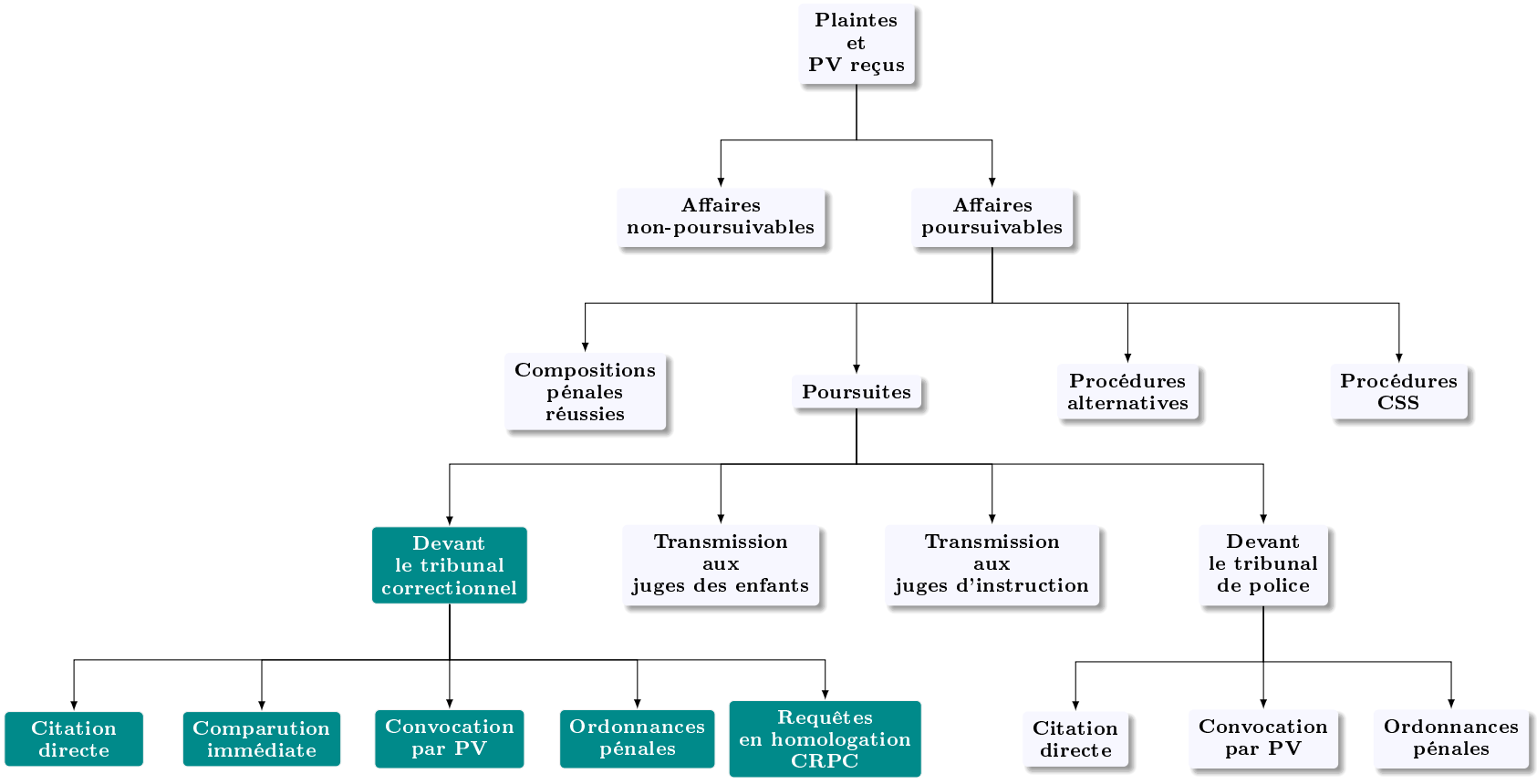
La constitution de la base a consisté en **trois étapes** :

- Sélection des minutes de jugement
- Numérisation des minutes retenues dans les greffes des tribunaux
- Reconnaissance optique des textes et extraction des informations par des scripts Perl

Plan :

Plan

Les procédures pénales :



Note : le détail des procédures alternatives et CSS ont été omis pour plus de lisibilité

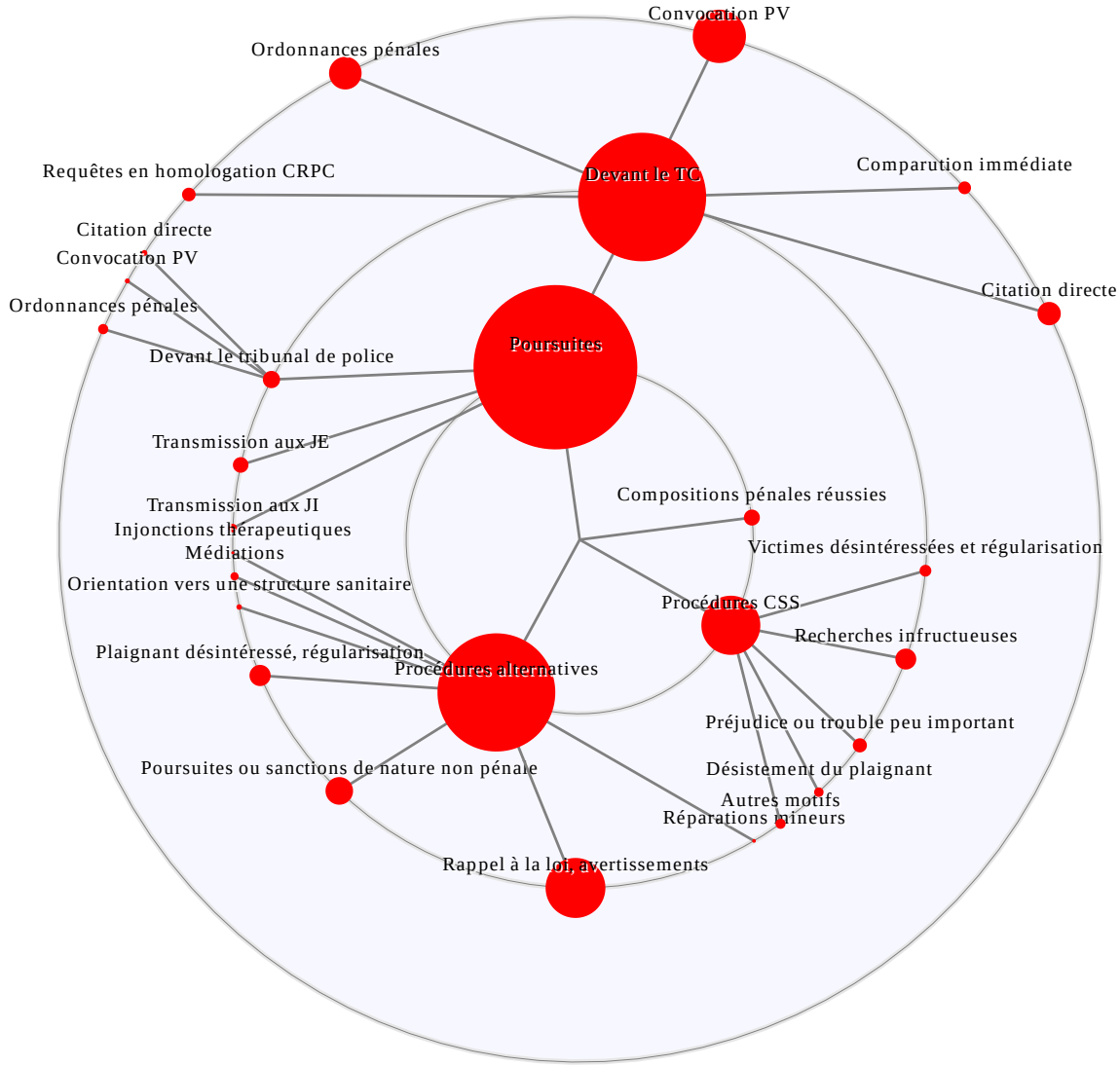
Les procédures devant les TC :

Les comparutions devant les tribunaux correctionnels font suite à **cinq types de procédures** différentes :

- les comparution immédiates (CI), anciennement, les flagrants délits
- les convocations par PV (CPPV)
- les citations directes
- les ordonnances pénales
- les requêtes en homologation comparution sur reconnaissance préalable de culpabilité (CRPC), le « plaider-coupable à la française »

Note : Les CRPC ont été exclues du champ de l'enquête (cf. *infra*)

Répartition des plaintes et PV reçus (2000-2009) :



Période d'observation :

Les minutes de jugement ont été recueillies sur une période de 10 ans (2000-2009) :

- La limite supérieure est celle de la date d'application de la **réforme de la carte judiciaire** initiée par la Garde des sceaux Rachida Dati en 2007 qui a notamment conduit à la suppression de 23 tribunaux de grande instance dont le tribunal d'Hazebrouck
- La durée a été choisie de façon à pouvoir observer les effets des lois dites **Perben I (2002) & II (2004)** [1] ainsi que les effets de l'évolution concomitante des politiques pénales menées
- La loi Perben II crée notamment les **CRPC** qui n'ont été mises en œuvre qu'à partir de 2005, soit en milieu de période d'observation. De plus, il s'agit d'une procédure assez différente des autres (ne serait-ce qu'au regard des informations contenues dans les minutes de jugement). C'est pourquoi elles ont été exclues (même si elles concurrencent d'autres procédures)

1. loi n° 2002-1138 du 9 septembre 2002 d'orientation et de programmation pour la justice et loi n° 2004-204 du 9 mars 2004 portant sur l'adaptation de la justice aux évolutions de la criminalité

Sélection des minutes de jugement :

Devant le nombre d'affaires poursuivies devant les tribunaux correctionnels (4 518 017 au total sur la période 2000-2009), leur numérisation dans leur intégralité était impossible au regard des moyens dont nous disposions.

Un **plan de sondage** a donc été conçu pour sélectionner les minutes à numériser :

- **Objectif du plan de sondage :** obtenir une bonne représentation des proportions de chaque procédure et de leurs évolutions sur la période 2000-2009 avec une attention particulière aux CI
- **Difficultés :**
  - Pas de données auxiliaires autres que les chiffres diffusés par le ministère agrégés par année et par juridiction
  - Difficultés pratiques variables pour chacune des juridictions (déserte en train, configuration des locaux, accès aux minutes, ...) et budget limité
  - De plus, l'enquête exploratoire a montré que la distribution des procédures suivait des patrons particuliers qui devaient être pris en compte dans la sélection

Sélection des juridictions :

Restriction aux juridictions de la **Cour d'appel de Douai** (région Nord Pas-de-Calais).

La première étape a d'abord consisté à sélectionner **cinq juridictions** parmi les **onze** que comptait la Cour d'appel de Douai jusqu'en 2009.

Critère de sélection, la **« taille » latente** des juridictions

Sélection des juridictions :

Restriction aux juridictions de la **Cour d'appel de Douai** (région Nord Pas-de-Calais).

La première étape a d'abord consisté à sélectionner **cinq juridictions** parmi les **onze** que comptait la Cour d'appel de Douai jusqu'en 2009.

Critère de sélection, la « **taille** » **latente** des juridictions :

- Les juridictions varient grandement quant aux **volumes d'affaires** qu'elles ont à traiter.

  Par exemple, en 2009, le tribunal d'Hazebrouck a traité au total **5 113** affaires pénales contre **128 709** pour le tribunal de Lille.
  Le croisement avec des données issues du recensement a de plus montré que ces différences sont **corrélées** aux fortes disparités des territoires des juridictions.
  La taille latente des juridictions permet donc d'appréhender d'autres différences que le volume (différences d'infractions commises et donc de procédures, politiques pénales différentes, ...) par ***proxy***

Sélection des juridictions :

Restriction aux juridictions de la **Cour d'appel de Douai** (région Nord Pas-de-Calais).

La première étape a d'abord consisté à sélectionner **cinq juridictions** parmi les **onze** que comptait la Cour d'appel de Douai jusqu'en 2009.

Critère de sélection, la « **taille** » **latente** des juridictions :

- Les juridictions varient grandement quant aux **volumes d'affaires** qu'elles ont à traiter.
- Les juridictions ont été réparties en **sept classes** au moyen d'un modèle de mélange fini (McLachlan and Peel (2000)) :

$$f(y|x,\Theta) = \sum_{k=1}^{K} \pi_k f_k(y|x,\theta_k) \quad \text{avec} \quad \sum_{k=1}^{K} \pi_k = 1 \quad \text{et} \ \pi_k > 0 \quad \forall k$$

où y désigne la variable dépendante univariée ou multivariée suivant une densité conditionnelle f, f_k désigne la distribution de y pour la classe k avec $y \sim f_k(y,\theta_k)$, x est un vecteur de variables indépendantes, π_k la probabilité (inconnue) d'observer la classe k, θ_k le vecteur de paramètres spécifiques à la distribution k et $\Theta = (\pi_1,\ldots,\pi_k,\theta_k^\top,\ldots,\theta_k^\top)$ est le vecteur regroupant tous les paramètres.

Sélection des juridictions (suite) :

- Sélection « **raisonnée** » des juridictions de façon à couvrir au mieux le spectre des tailles latentes

  Il ne manque que la classe IV et la classe VII qui correspond aux juridictions les plus importantes et qui sont spécifiques à la région parisienne (Bobigny, Nanterre, ...)
- Un tirage aléatoire semblait en effet difficile à ce stade du fait du nombre limité de strates combiné aux contraintes organisationnelles de l'enquête
- Les juridictions retenues sont : **Arras, Avesnes-sur-Helpe, Béthune, Hazebrouck et Lille**

  Elles représentent un total de 117 751 poursuites devant le TC

Enquête préparatoire et sélection des minutes de jugement :

L'enquête exploratoire a notamment consisté à examiner les audiences de trois semaines choisies de façon arbitraire dans les cinq juridictions retenues.

Il en ressort en particulier que :

- Les procédures tendent à être **regroupées** lors des audiences, particulièrement les comparutions immédiates (effet de grappe)
- Pour les juridictions les plus importantes comme Lille, certaines audiences sont même **exclusivement** consacrées aux comparutions immédiates
- Les audiences de CI ne sont **pas uniformément réparties** au fil de la semaine (souvent plus nombreuses en début de semaine pour juger les infractions commises pendant le week-end)
- **Saisonnalité :** les volumes des différentes procédures varient en fonction du mois de l'année (infractions différentes, calendrier scolaire, ...)
- Enfin, l'organisation du calendrier des audiences se fait à l'échelle de la semaine

Sélection des minutes de jugement (suite) :

La distribution des CI a donc conduit à un changement d'échelle : sélection de **semaines** et non plus de minutes :

- La semaine est une strate « naturelle » puisqu'elle est utilisée par les greffiers pour organiser les audiences
- Elle permet de « lisser » les variations quotidiennes évoquées précédemment
- Les probabilités d'inclusion des minutes peuvent de plus être calculées à partir de celles de la semaine

Sélection des semaines :

Là encore, l'absence de données auxiliaires est problématique (saisonnalité).

- Il a donc semblé préférable de procéder à une **répartition uniforme** des semaines sélectionnées de façon à ce que chacune d'entre elle (indexée par son numéro d'ordre dans l'année) ne soit retenue qu'**une seule fois**
- Tirage simple ou systématique ne permettaient pas de satisfaire à ces critères
- Alternative : techniques de **coordination des échantillons**

Techniques de coordination des échantillons :

- Techniques développées à partir du début des années 1970 de façon indépendante par différents organismes de statistique publique
- Visent à *maximiser* (coordination **positive** : Eq. 1) ou *minimiser* (coordination **négative** : Eq. 2) la probabilité qu'une unité soit sélectionnée lors de tirages successifs t et u (Nedyalkova et al. (2006)) :

$$\pi_k^t \pi_k^u \leq \pi_k^{tu} \leq \min(\pi_k^t, \pi_k^u) \quad (1)$$

$$\max(0, \pi_k^t + \pi_k^u - 1) \leq \pi_k^{tu} \leq \pi_k^t \pi_k^u \quad (2)$$

où $\pi_k^t = E(S_k^t),\ t = 1, \ldots, T$ et $\pi_k^u = E(S_k^u),\ u = 1, \ldots, T$ désignent les probabilités d'inclusion de premier ordre en t et u , $\pi_k^{tu} = E(S_k^t S_l^u),\ k \in U^t \cap U^u,\ t = 1, \ldots, T$ la probabilité d'inclusion longitudinale et S_k désigne la variable indicatrice marquant la présence ou l'absence de l'unité k dans l'échantillon

- La coordination positive correspond aux panels et la coordination négative aux enquêtes ordinaires auprès des ménages ou des entreprises

Tirage de l'échantillon :

Plusieurs techniques de coordination ont été proposées.

La technique retenue repose sur les **PRN** (*Permanent Random Numbers*) (Ohlsson (1995)).

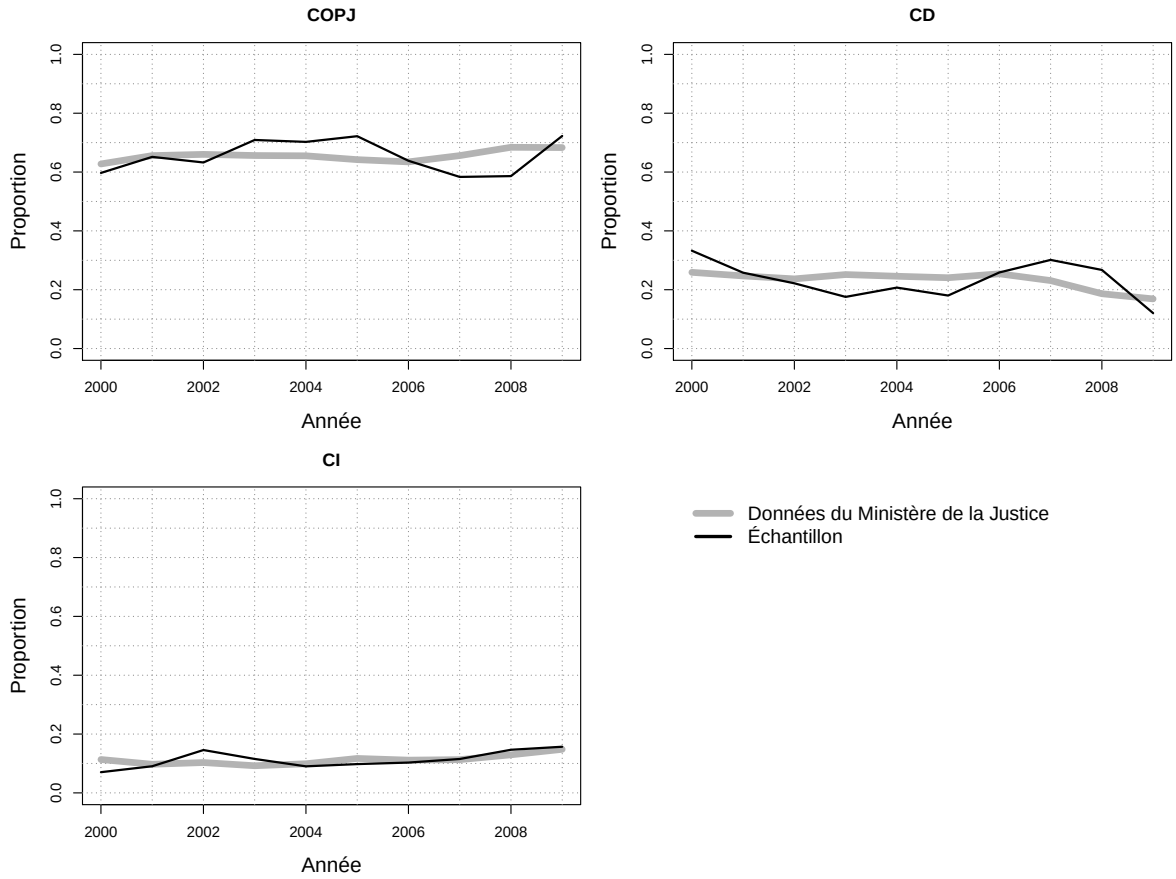
Elle consiste à :

1. Stratifier des semaines en trois groupes de quatre mois
2. Dans chaque strate, attribuer un PRN u_i à chacune des semaines avec $u_i \sim U(0,1)$
3. Trier des semaines dans chaque strate suivant leur PRN dans un ordre croissant ou décroissant
4. Sélectionner les dix premières semaines de chaque strate (ou les dix dernières dans le cas d'un tri décroissant)

Résultats :

- Au final, **trois semaines** sont sélectionnées pour chaque année (une par quadrimestre)
- Cette approche a, en plus de sa simplicité, l'avantage d'être équivalent à un **tirage aléatoire simple stratifié** selon la méthode de Sunter (Tillé (2001))
- Les logiciels standards d'analyse de données d'enquête peuvent donc être utilisés pour l'estimation des paramètres
- Le nombre de minutes ne pouvant pas être connu à l'avance, des simulations ont été réalisées à partir des saisies opérées lors de l'enquête exploratoire
- Les distributions marginales ne correspondant pas exactement aux chiffres du ministère, la pondération a été corrigée *ex post* par un algorithme de **calage sur les marges**

Évolution des proportions des procédures (données du ministère et échantillon) :



Note : Les COPJ ont dû être regroupées aux ordonnances pénales pour correspondre aux données du ministère

Plan

Extraction des informations :

L'extraction d'informations s'est appuyée sur le caractère très **normé** de la rédaction des minutes :

- Structuration identique des minutes
- Utilisation de formules récurrentes (« Attendu qu'il est prévenu d'avoir..., le `<date/>` à `<hh/>` heures `<mm/>`, en tout cas sur le territoire national et depuis temps n'emportant pas prescription... »)
- Cependant,
  - Variations d'une juridiction à l'autre et parfois même en fonction des greffiers
  - Formatage parfois difficile à analyser
  - Fautes de frappes, ...
  - Annotations ou corrections manuelles qui perturbent l'OCR
  - Certaines minutes se trouvaient aussi dans un très mauvais état (particulièrement à Arras pour les années 2000-2002)

Expressions régulières :

L'extraction a principalement reposé sur l'usage d'**expressions régulières** (ou « rationnelles » pour les puristes) :

- Automate fini non déterministe (sic)
- Chaînes de caractères qui vont permettre de générer un programme pour trouver, et éventuellement extraire, des portions de texte correspondant à un **motif**
- Généralisation des *wild cards*.

  Par exemple, le caractère `"*"` pour tronquer les mots : `"ls *.tex"` retourne tous les fichiers se terminant par `".tex"` dans le répertoire courant
- Combinaison de deux types de caractères :
  - les caractères dits **littéraux** (`"a"`, `"b"`, `"c"`,..., `"1"`, `"2"`, `"3"`,..., `"\u20AC"`,...)
  - les caractères **spéciaux** de substitution (`"|"`, `"["`, `"]"`), de groupement (`"("`, `")"`,...) et de quantification (`"?"`, `"*"`, `"+"`,...) complétés par d'autres **métacaractères** (`"\d"`, `"\w"`, `"\s"`, `"\W"`,...)

Exemple :

Récupérer la date de naissance d'un prévenu au format suivant :

DATE DE NAISSANCE : <jour/>/<mois/>/<année/>

Expression régulière Perl :

```
qr/
   DATE\W+DE\W+NAISSANCE\W*:?(?<match>.+) ## on récupère tout
        après ":"
/xi
```

Si on a trouvé quelque chose :

```
/(
          (?<jour>
             [I\d]+                    ## chiffres et "I"
          )\/
          (?<mois>
             [I\d]+
          )\/
          (?<an>
             [I\d]+
          )
)/x
```

Puis validation de la date en ayant au préalable transformé les `"I"` en `"1"` le cas échéant pour corriger les ratés de l'OCR

Exemple (suite) :

L'exemple précédant correspond au cas le plus favorable où les caractéristique des prévenus étaient listées en colonne :

NOM : ...
DATE DE NAISSANCE : ...
LIEU DE NAISSANCE : ...
...

D'autres cas étaient moins favorables comme le format « en ligne » : `<civilité/>` `<prénom/>` `<nom/>`, né le `<date-de-naissance/>` à `<lieu-de-naissance/>`, ...

En effet,

- Pas de mots-clefs
- Le nombre d'items n'était pas toujours le même
- Les séparateurs pouvaient varier (`","`, `";"`)
- Nécessité de découper récursivement ces portions de texte en « ancrant » avec des mots clefs les expressions régulières en commençant par les portions de texte les plus régulières
- Plus généralement, c'est le principe qui a été appliqué pour l'analyse syntaxique des minutes

Divide-and-conquer :

Les expressions régulières ont été utilisées pour mettre en œuvre une stratégie de type *divide-and-conquer*

Divide-and-conquer :

Les expressions régulières ont été utilisées pour mettre en œuvre une stratégie de type *divide-and-conquer* :

- Les minutes ont été découpées en blocs de plus en plus petits jusqu'à isoler l'information recherchée

Divide-and-conquer :

Les expressions régulières ont été utilisées pour mettre en œuvre une stratégie de type *divide-and-conquer* :

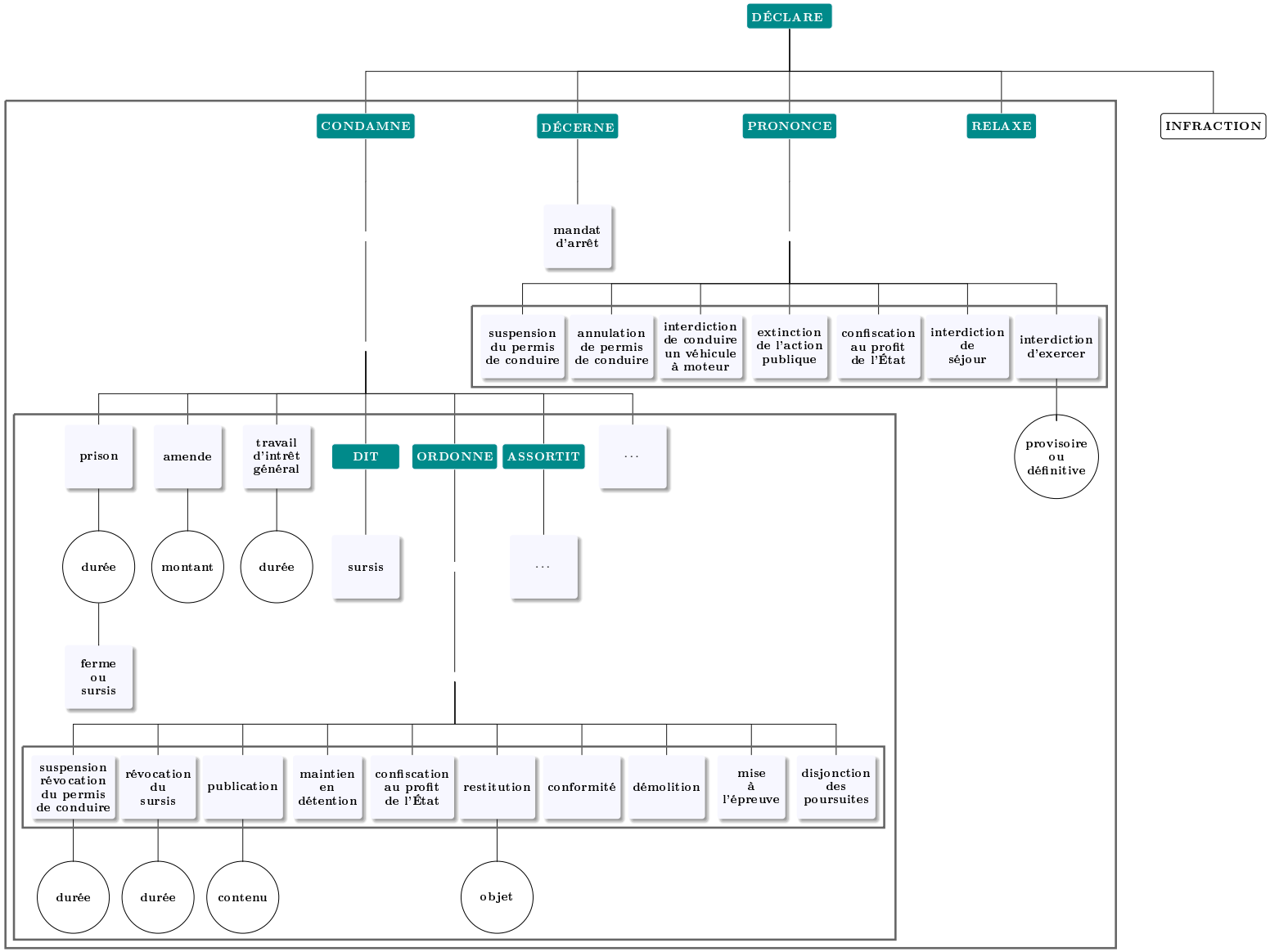
- Les minutes ont été découpées en blocs de plus en plus petits jusqu'à isoler l'information recherchée
- Premier découpage : les minutes et les sections des minutes
  - Les minutes ont été décomposées en cinq sections : la composition du tribunal, les parties en présence (partie(s) civile(s) et prévenu(s)), le(s) chef(s) d'inculpation, les débats, le verdict et, le cas échéant, la ou les peines prononcées
  - Le découpage a été réalisé à partir d'une fonction spécifique appliquant des critères passés en arguments spécifiques à chaque paquet de minutes
  - Nul besoin alors de tout réécrire en fonction des minutes, il suffit de modifier les règles
  - Coder cette fonction a permis une plus grande flexibilité notamment pour gérer les problèmes posés par les variations dans la présentation des minutes ainsi que les scories qui ont perturbé l'OCR

Divide-and-conquer :

Les expressions régulières ont été utilisées pour mettre en œuvre une stratégie de type *divide-and-conquer* :

- Les minutes ont été découpées en blocs de plus en plus petits jusqu'à isoler l'information recherchée
- Premier découpage : les minutes et les sections des minutes
- Puis découpage des sections en sous-parties puis en sous-sous-parties...
  - L'analyse syntaxique des décisions s'est révélée la plus complexe
  - Plusieurs prévenus, plusieurs chefs d'inculpation et donc autant de décisions avec possibilité de relaxes partielles, distinction des peines fermes du sursis...
  - Nombreuses variantes dans la rédaction
  - L'analyse a nécessité un long travail d'énumération qui a résulté dans un code en deux passes et de multiples branchements
  - Mais, là encore, le caractère normé de la prose juridique permet de limiter la combinatoire

Analyse lexicale des décisions :



Fazit :

Avantages :

- Automatisation
- Pas d'interprétation : application uniforme des mêmes règles
- Adaptativité : apprentissage
- Si les règles sont suffisamment strictes, les erreurs se soldent généralement par des `NA` (cf. les dates de naissance)

Inconvénients :

- Pas d'interprétation
- Nécessité de savoir précisément ce qu'on cherche et d'énumérer les alternatives (long processus d'essais-erreurs)
- Supporte mal le flou et les ambiguïtés
- Fonctionne bien dans ce cas car la rédaction du texte répond à des standards assez strictes qui en font (presque) l'expression d'un *langage régulier*

  Application plus difficile dans le cas de discours moins normés
- Ne peut remplacer la saisie manuelle lorsque les originaux sont en mauvais état

Les valeurs manquantes :

- L'automatisation conduit sans doute à une plus grande **perte d'informations**
- Les `NA` constituent à la fois une source de **biais** et représenter une **perte de précision** des estimateurs
- Néanmoins, les données manquantes apparaissent indépendantes des variables d'intérêts
- D'un point de vue statistique, elles peuvent être considérées comme **MAR** (*Missing At Random*) :

$$P(R|y^S) = P(R|y^{NA})$$

- Les valeurs manquantes représentent toutefois une perte d'informations et donc de précision (**inflation de la variance** et donc des erreurs types) mais elles n'introduisent pas de **biais de sélection**
- Enfin, sauf peut-être pour Arras...

Plan

La base de données :

Au final, la base de données se compose de trois tables :

- Table audiences :

  `date, heure, composition du tribunal`

- Table prévenus :

  `sexe, année de naissance, domicile fixe, situation familiale, statut d'activité professionnelle, casier, procédure, comparant`

- Table décisions :

  `infractions, décisions, condamnations, quanta de peines, sursis`

Au total, la base renseigne 7 882 minutes de jugement de 9 361 prévenus (mais plus de 500 minutes sont manquantes)

Notes : : Différentes sous-parties des minutes comme celles relatives aux parties civiles n'ont pas été exploitées.

De plus, certains renseignements extraits n'ont pas été intégrés.

Estimation sur des données d'enquêtes :

Les données sont issues d'un échantillonnage, ce faisant le plan de sondage doit être **pris en compte** dans l'estimation des paramètres :

- Le plan de sondage rompt en effet avec les **postulats habituels** des tests (population infinie, $VA \sim iid$, ...)
- Ignorer les effets du plan de sondage peut, par exemple, conduire à augmenter le risque de **rejeter l'hypothèse nulle** alors qu'elle est vraie (erreur de type I), notamment lorsque les grappes sont homogènes
- Les variables d'intérêt étant pour la plupart **catégoriques**, les hypothèses ont été testées au moyen de test du χ^2 de Pearson et de modèles log-linéaires.
  Les paramètres ont été corrigés au moyen de la méthode proposée par **Rao et Scott** (Chaudhuri and Stenger (2005)) avec le package R survey (Lumley (2010)). Dans le cas d'un tableau croisé, cette correction a pour forme :

$$X^2_{RS} = \frac{(R-1)(C-1)X^2_P}{\bar{\hat{\delta}}}$$

avec R le nombre de lignes, C le nombre de colonnes, X^2_P la statistique du χ^2 et $\bar{\hat{\delta}} = E[X^2] = \sum_{k=1}^{K-1} \hat{\delta}_k$. $\hat{\delta}_k$ désigne ici la k-ième valeur-propre de la matrice d'effet du plan $\boldsymbol{D} = \hat{\boldsymbol{V}}_{H_0}^{-1} \hat{\boldsymbol{V}}_{\mathcal{S}}$.

Quelques résultats :

- Alignement des « petites » et « moyennes » juridictions sur les « grandes » dans le recours aux comparutions immédiates
- Les CI tendent à déboucher sur des peines plus sévères (prison, ferme et sursis)
- L'orientation en CI est fortement déterminée par les antécédents judiciaires des prévenus
- L'absence de « garanties de représentation » augmente aussi les chances de passer en CI
- Les précaires et les étrangers y sont aussi sur-représentés

Plan

Conclusion :

- Pertinence de l'utilisation des techniques de sondage et, plus particulièrement, des **plans complexes** pour sélectionner des documents issus de fonds d'archives

  L'approche présentée montre comment pallier l'absence de données auxiliaires dans certaines circonstances (sans avoir à en collecter de nouvelles)
- Les **expressions régulières** constituent un outil puissant d'extraction textuelle

  Néanmoins, elles nécessitent des textes suffisamment réguliers pour fonctionner pleinement
- Question (ouverte) : recours à des approches alternatives fondées sur des algorithmes d'apprentissage, IA, la logique floue. . . ?

Merci pour votre attention

Plan

Références bibliographiques

Chaudhuri, A. and H. Stenger (2005). *Survey Sampling : Theory and Methods* (2 ed.). Boca Ranton : CRC Press.

Douillet, A.-C., T. Léonard, T. Soubiran, and H. Yazdanpanah (2015). Logiques, contraintes et effets du recours aux comparutions immédiates. Étude de cinq juridictions de la Cour d'appel de Douai. rapport de recherche pour la mission de recherche droit et justice, Centre d'études et de recherches administratives.

Lumley, T. (2010). *Survey Sampling : Theory and Methods*. New York : Wiley.

McLachlan, G. and D. Peel (2000). *Finite Mixture Models*. New York : Wiley.

Nedyalkova, D., J. Pea, and Y. Tillé (2006). A Review of Some Current Methods of Coordination of Stratified Samples. Introduction and Comparison of New Methods Based on Microstrata. rapport de l'université de neuchâtel, Université de Neuchâtel.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In A. C. Brenda G. Cox and M. J. Colledge (Eds.), *Business Survey Methods*, Chapter 9, pp. 153–183. New York : Wiley.

Tillé, Y. (2001). *Théorie des sondages : échantillonnage et estimation en populations finies*. Paris : Dunod.