

la procédure judiciaire criminelle au corpus textuel

Après avoir présenté le contexte de recherche, il est nécessaire de procéder à la description des données exploitées en [analyse criminelle](#) et qui serviront de base à notre étude. Pour cela, nous avons étudié plusieurs dossiers de [procédures](#) d'atteintes aux personnes (homicides, disparitions) comme ceux traités par le [DSAC](#).

Avant d'entamer leur description, précisons que nous avons pu les consulter au sein des bâtiments du [PJGN](#), sous le contrôle des officiers et sous-officiers responsables du Département, et dans le cadre d'une convention de partenariat entre le [PJGN](#) et l'université de Cergy-Pontoise comportant une clause de confidentialité. La présentation des données sera donc d'ordre général, sans description précise des faits investigués, sans mention d'information portant atteinte à la vie privée des personnes impliquées ou d'éléments permettant d'identifier l'affaire. Les extraits cités pour des besoins d'illustration le seront après masquage ou pseudonymisation ¹.

L'objectif de ce chapitre est de décrire le contenu d'une [procédure](#) judiciaire criminelle telle qu'elles sont confiées aux [analystes criminels](#), puis d'en dégager les aspects textuels afin de déterminer d'une part leur potentiel en tant que ressource d'étude et de développement, et d'autre part les modalités d'automatisation de leur traitement.

1. La pseudonymisation consiste à remplacer une information par une autre équivalente afin de préserver l'anonymat (une date par une autre date, un prénom par un autre prénom, etc.). La pseudonymisation n'est réalisée que pour les besoins de ce manuscrit, les travaux ont été réalisés sur le texte original.

1 Qu'est-ce qu'une procédure judiciaire ?

Les chroniqueurs judiciaires font parfois mention lorsqu'ils relatent un procès de la masse des dossiers d'archives qui compile le travail d'investigation ayant été mené jusque-là. Si l'on peut supposer que la numérisation des services de la justice permettra à l'avenir de se passer du papier, les **procédures** sont encore parfois confiées aux analystes criminels sous la forme de plusieurs boîtes d'archives, comme à la figure 4. Néanmoins, ce cas de figure se raréfie, et même les **cold cases** circulent aujourd'hui sous forme électronique². Au-delà de sa forme matérielle, commençons par définir la **procédure** judiciaire.



FIGURE 4 – Un dossier de **procédure** relatif à une disparition au format papier

La « **procédure** judiciaire » est l'ensemble des démarches effectuées pour mener une action en justice. Son déroulement, selon les configurations, est encadré par les articles 53 et suivants du **Code de procédure pénale** pour l'**enquête de flagrance**³, 75 et suivants pour l'**enquête préliminaire**⁴, 74 et suivants pour les découvertes de cadavres et les disparitions inquiétantes⁵ et 151 pour la commission rogatoire⁶.

Le dossier de **procédure** judiciaire rassemble les documents compilés par les enquêteurs dans l'objectif de la manifestation de la vérité.

2. Pour les **cold cases**, la numérisation des documents devient d'ailleurs une étape préalable à la reprise des investigations.

3. Légifrance : https://www.legifrance.gouv.fr/affichCode.do;jsessionid=OFB4EA802845B5D4FF4651709A000D1C.tplgfr42s_2?idSectionTA=LEGISCTA000006151876&cidTexte=LEGITEXT000006071154&dateTexte=20040310#LEGIARTI000006575016 [consulté le 11 décembre 2019]

4. Légifrance : https://www.legifrance.gouv.fr/affichCode.do;jsessionid=OFB4EA802845B5D4FF4651709A000D1C.tplgfr42s_2?idSectionTA=LEGISCTA000006151877&cidTexte=LEGITEXT000006071154&dateTexte=20040310 [consulté le 11 décembre 2019]

5. Légifrance : <https://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000020632053&cidTexte=LEGITEXT000006071154&dateTexte=20090514> [consulté le 11 décembre 2019]

6. Légifrance : https://www.legifrance.gouv.fr/affichCodeArticle.do;jsessionid=62F4026FB69ECD7B0C8350B6FB04D828.tplgfr42s_2?cidTexte=LEGITEXT000006071154&idArticle=LEGIARTI000006575360&dateTexte=20191211&categorieLien=cid#LEGIARTI000006575360 [consulté le 11 décembre 2019]

Dans le cas de la [commission rogatoire](#), le cadre juridique de production de ce dossier est précisé par l'article 81 du [Code de procédure pénale](#)⁷ (extrait) :

Le [juge d'instruction](#) procède, conformément à la loi, à tous les **actes d'information** qu'il juge utiles à la manifestation de la vérité. Il instruit à charge et à décharge.

Il est établi une **copie de ces actes** ainsi que de toutes les **pièces de la procédure** ; chaque copie est certifiée conforme par le greffier ou l'officier de police judiciaire commis [...]. Toutes les pièces du dossier sont **cotées** par le greffier au fur et à mesure de leur rédaction ou de leur réception par le juge d'instruction.

Toutefois, si les copies peuvent être établies à l'aide de procédés photographiques ou similaires, elles sont exécutées à l'occasion de la **transmission du dossier**. Il en est alors établi autant d'exemplaires qu'il est nécessaire à l'administration de la justice. Le greffier certifie la conformité du dossier reproduit avec le dossier original [...].

Dans cet article sont introduites les notions d'acte d'information, de pièce de [procédure](#), et il est précisé quelques modalités de leur circulation. On comprend que les actes d'information, qui sont réalisés au cours de l'[enquête](#), débouchent sur la production de documents appelés pièces de [procédure](#) qui relatent ces actes, et que l'ensemble des pièces de [procédure](#) constitue le dossier. Dans la suite de ce chapitre, nous nous efforçons de donner un aperçu de la nature de ces documents, qui constituent donc la base du travail des [analystes criminels](#).

2 Les pièces de procédure

Nous avons répertorié une quinzaine de types de documents constituant des pièces de [procédure](#), que nous avons rassemblés en deux catégories (tableau 1). D'un côté, la catégorie des documents réglementaires et de [procédure](#) constituée par les pièces liées au déroulement légal et opérationnel de l'[enquête](#), et de l'autre les documents d'information, qui convoient des éléments d'information relatifs au faits.

7. Légifrance : [Article 81 du Code de procédure pénale](#) [consulté le 09 mai 2019].

Documents réglementaires et de procédure	Documents d'information
<ul style="list-style-type: none"> - Réquisitions, - Bordereau d'envoi, - Inventaire des pièces à conviction, - PV de notification de garde à vue, - PV de saisie, - Certificats médicaux préalables à la garde à vue 	<ul style="list-style-type: none"> - PV transport constatations mesures prises, - PV d'audition de témoin et de garde à vue, - PV d'actes d'enquête divers (investigations, renseignements, perquisition...), - PV de synthèse, - Rapports d'expertises, - Retours des réquisitions : téléphoniques, bancaires, péages, compagnies d'assurance, etc., - Documents graphiques : photos et vidéos, - Autres documents.

TABLE 1 – Récapitulatif des documents rencontrés en **procédure** traitées en **analyse criminelle**.

Les sous-parties suivantes sont consacrées à la description des pièces, en détaillant notamment leur format, leur rôle pour l'**enquête**, et leur émetteur.

L'inventaire de la matière documentaire que nous allons réaliser dans cette sous-partie n'est pas exhaustif et ne constitue en aucun cas un référentiel rigide de la constitution de la **procédure**. Il s'agit d'un aperçu livré sur la base de nos observations qui de plus ne concerne que les pratiques de la **Gendarmerie nationale**. Dans la Police les pratiques et les formats de documents diffèrent, ce qui d'ailleurs dans notre objectif est à prendre en compte lorsqu'une affaire « circule » entre Police et Gendarmerie, puisque le dossier mélange alors des pièces issues des deux institutions.

2.1 Documents réglementaires et de procédure

Les documents réglementaires et de **procédure** sont des documents ne convoyant pas en tant que tel des informations à propos aux faits. Leur fonction est de garantir la légalité et la conformité des opérations ainsi que le respect des droits des personnes impliquées.

Nous classons dans cette catégorie :

- Les bordereaux d'envoi : il s'agit de documents faisant office de sommaire et de liste des pièces lors de la transmission du dossier,
- Les inventaires de pièces à conviction, faisant la liste des **scellés**,

- Les **réquisitions** de tiers et d'experts (figure 5) afin d'obtenir des informations, une assistance ou des moyens matériels, elles sont adressées aux opérateurs téléphoniques, établissements bancaires, médecins légistes, maires, pompes funèbres, laboratoires photographiques, interprètes, etc.
- Les certificats médicaux préalables à la **garde à vue**, qui attestent que l'état de santé de la personne est conforme avec la mesure de **garde à vue**,
- Les **procès-verbaux** de notification de **garde à vue**, qui en consignent les modalités : heure de début et de fin, notification des droits de la personne gardée à vue, consultation d'un médecin, assistance d'un avocat, temps de repos, etc.,
- Les **procès-verbaux** de saisie, qui documentent la collecte d'un élément matériel placé sous **scellé**.

Au-delà de leur fonction au sein de la **procédure**, nous avons regroupé ces documents car ils sont semi-structurés : soit comme des formulaires, soit sous la forme d'une langue très normée comme la figure 5 l'illustre, soit parce qu'ils constituent des inventaires d'éléments de la **procédure**. Leur rôle au sein de la **procédure** ne concerne pas directement les faits, mais ils fournissent des informations sur le déroulé des investigations. C'est pourquoi nous avons décidé de les considérer à part du reste de la **procédure**.

2.2 Documents d'information

Nous regroupons comme documents d'information des documents convoyant des informations directement liées aux faits ou des comptes-rendus d'opérations menées par les enquêteurs. Ces documents consignent une information dense et non-structurée qui est le matériau informationnel au cœur de l'**enquête**. Bon nombre d'entre eux sont des **procès-verbaux**.

Le dictionnaire Larousse en ligne propose la définition suivante pour le terme **procès-verbal** :

Acte écrit rédigé par un magistrat, un officier ou agent de police judiciaire, un officier public, qui rend compte de ce qu'il a fait, entendu ou

République Française – Ministère de la Justice
TRIBUNAL DE GRANDE INSTANCE DE [REDACTED]
Parquet du Procureur de la République

D 606
C

REQUISITIONS

Le Procureur de la République près le Tribunal de Grande instance de [REDACTED]

Vu la découverte le [REDACTED] à [REDACTED]
 du cadavre de : [REDACTED]

Vu l'enquête suivie :

<input type="checkbox"/> Vu l'article 74 du Code de Procédure Pénale Attendu que les causes de la mort apparaissent inconnues ou suspectes	<input checked="" type="checkbox"/> Vu l'article 60 - 77-1 du Code de Procédure Pénale Attendu qu'au vu des premiers renseignements apportés par l'enquête il y a lieu de faire procéder à des constatations, examens techniques ou scientifiques
---	--

REQUIERT

■ Monsieur le Professeur [REDACTED] Expert inscrit sur la liste de la Cour d'Appel de [REDACTED]

AUX FINS DE

- ✕ • Procéder à l'analyse des prélèvements qui font l'objet des scellés n° 1A et 16A, 17 et 20 A.
- ✕ • Procéder sur ces prélèvements à la recherche de sperme - de sang d'origine humaine ;
- ✕ • Procéder sur ces prélèvements à toutes analyses et recherches d'empreintes génétiques ;
- Procéder à tous prélèvements de comparaison utiles sur la personne de [REDACTED] et sur ces derniers prélèvements à toutes analyses et recherches d'empreintes génétiques aux fins de les comparer aux résultats obtenus par les analyses exécutées sur les scellés ;
- ✕ • dresser un rapport d'ensemble détaillé des opérations et des conclusions.

Les ouvertures nécessaires de scellés seront faites conformément aux dispositions de l'article 60 alinéa 3 du Code de Procédure Pénale.

Fait au Parquet, le [REDACTED]
 Le Procureur de la République,

[REDACTED]

[REDACTED] - [REDACTED] Cedex
 [REDACTED] - Fax : [REDACTED]

FIGURE 5 – Exemple de réquisitions d'expert

constaté dans l'exercice de ses fonctions. (Abréviation familière : P.-V.)⁸

Cette définition est en accord avec l'article 429 du [Code de procédure pénale](#)⁹ :

Tout procès-verbal ou rapport n'a de valeur probante que s'il est régulier en la forme, si son auteur a agi dans l'exercice de ses fonctions et a rapporté sur une matière de sa compétence ce qu'il a vu, entendu ou constaté personnellement. [...]

Le [procès-verbal](#) est donc un document officiel rédigé par une personne qualifiée et rapportant des faits qu'elle a perçus personnellement.

En [Gendarmerie nationale](#), les [procès-verbaux](#) s'organisent selon une structure verticale¹⁰ (figures 6, 7) : un en-tête mentionne les références de l'unité et la nature du document, suivie des date, heure, lieu de la rédaction, suivies des nom, grade et affectation du ou des auteurs du document, et enfin du cadre légal des opérations rapportées (cadre général et renseignements sur la commission rogatoire). Ensuite vient le corps du texte à proprement parler, et enfin le document s'achève par la ou les signatures et visas nécessaires.

Néanmoins, cette structure n'est pas rigide et s'adapte selon les besoins et la finalité du [procès-verbal](#). Par exemple, les [auditions de témoins](#) comportent une partie de renseignements d'état-civil à propos de la personne auditionnée.

2.2.1 Procès-verbaux de transport constatations et mesures prises

Il s'agit d'un document, qui, comme son nom l'indique, relate le déplacement et les premières mesures adoptées par un ou des gendarmes suite au signalement ou à la découverte de faits criminels. Lors de la découverte d'une scène de crime par exemple, les techniciens en identification criminelle ([TIC](#)) se rendent sur les lieux et procèdent aux actes de police technique et scientifique comme le prélèvement et la protection des éléments matériels ou la prise de mesures et de photographies de la scène de crime. Ce document peut se structurer en plusieurs rubriques : la saisine, la

8. https://www.larousse.fr/dictionnaires/francais/procès-verbal_procès-verbaux/64067 [consulté le 11 juin 2019]

9. Légifrance : article 429 du [Code de procédure pénale](#) <https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006071154&idArticle=LEGIARTI000006576551> [consulté le 11 juin 2019]

10. En comparaison, les [procès-verbaux](#) de Police comportent une marge à gauche où figurent les références et l'objet des investigations.

description de la situation à l'arrivée des enquêteurs, les mesures prises, la liste des personnes présentes, la situation et l'environnement des lieux (topographie régionale, d'ensemble et rapprochée), les conditions climatiques, le corps du délit (dans le cas d'un homicide, position du cadavre, habillement, éventuellement éléments matériels le camouflant, blessures visibles), les prélèvements effectués, les constatations en présence du [médecin légiste](#), les saisies d'éléments matériels et enfin les conclusions de cette première étape de l'[enquête](#). Deux paragraphes d'exemple ci-dessous sont tirés d'un [procès-verbal](#) de transport, constatations et mesures prises suite à la découverte d'un cadavre.

1- SAISINE

Le 03 juillet 2012 à 15 heures 15, nous recevons un appel téléphonique de la Brigade de Recherches d'ETAMPES nous informant de la découverte d'un corps humain enterré dans la forêt au Nord de l'agglomération de DOURDAN (91). Cette unité demande notre concours aux fins de procéder aux constatations et opérations de police technique.

Nous nous transportons immédiatement sur les lieux en compagnie du Major GAGNAC commandant notre unité.

[...]

4) - position générale du corps :

Le cadavre est en position décubitus dorsal légèrement désaxé côté droit. L'ensemble tête-tronc, membre inférieur gauche se trouvent sensiblement dans le prolongement. La tête est légèrement inclinée sur la gauche. Le corps présente une légère cambrure au niveau des reins ainsi qu'une faible torsion au niveau du thorax. Le membre inférieur droit est relevé, et fait un angle droit à l'articulation du genou. Le talon droit se trouve au niveau de la partie médiane de la face interne du mollet gauche.

Les [procès-verbaux](#) de transport, constatation et mesures prises sont donc des documents rapportant un certains nombres d'opérations en s'appuyant sur une pratique formalisée : l'enquêteur est saisi, se déplace, décrit la situation et prend des mesures. Le texte du document est à la fois narratif et descriptif (situation et opérations), organisé en paragraphes, exprimé à l'aide de phrases complètes rédigées et

comporte plusieurs pages.

2.2.2 Procès-verbaux d'auditions et d'interrogatoires

Au cours d'une **enquête**, deux sortes d'**auditions** peuvent être menées : les **auditions** de **témoins**, où la personne entendue, libre, se présente sur convocation ou de sa propre initiative, et les **auditions** de personnes mises en cause, où la personne entendue est gardée à vue ou **mise en examen**.

Les conditions d'**audition** des premières sont précisées par les articles 101 à 113¹¹ et 113-1 à 113-8¹² du **Code de procédure pénale**. Quant aux personnes mises en cause, il s'agit des articles 114 à 121 du **Code de procédure pénale**¹³ dans le cadre de la **commission rogatoire**, 77 et suivants dans le cadre de l'**enquête préliminaire**¹⁴, et 63 et suivants dans le cadre de l'**enquête de flagrance**¹⁵.

En pratique, les **auditions** sont réalisées par deux ou trois personnels ayant la qualification d'**officiers de police judiciaire**, ou par le **juge d'instruction** pour les personnes **mises en examen**. L'**audition** est une conversation entre les enquêteurs et le **témoin**. La personne entendue s'exprime oralement, les enquêteurs posent des questions, et les propos sont pris en note par l'un des enquêteurs au fil de la discussion. Selon les rapports de la personne entendue avec les faits investigués, l'**audition** peut concerner toutes sortes de sujets. On interrogera les proches sur les éléments psychologiques, la description du caractère, des habitudes, des connaissances, du train de vie de la victime, pour les **témoins** oculaires, on s'intéressera aux personnes, aux lieux, aux horaires et à tout élément ayant pu attirer leur attention.

11. Légifrance : Articles 101 à 113 du **Code de procédure pénale** https://www.legifrance.gouv.fr/affichCode.do;jsessionid=24358CA45B3E6900D6BE9EC1D3A3D9D9.tplgfr42s_2?idSectionTA=LEGISCTA000006182923&cidTexte=LEGITEXT000006071154&dateTexte=20190510 [consulté le 10 mai 2019].

12. Légifrance : Articles 113-1 à 113-8 du **Code de procédure pénale** https://www.legifrance.gouv.fr/affichCode.do;jsessionid=24358CA45B3E6900D6BE9EC1D3A3D9D9.tplgfr42s_2?idSectionTA=LEGISCTA000006182888&cidTexte=LEGITEXT000006071154&dateTexte=20190510 [consulté le 10 mai 2019].

13. Légifrance : Articles 114 à 121 du **Code de procédure pénale** https://www.legifrance.gouv.fr/affichCode.do;jsessionid=24358CA45B3E6900D6BE9EC1D3A3D9D9.tplgfr42s_2?idSectionTA=LEGISCTA000006167425&cidTexte=LEGITEXT000006071154&dateTexte=20190510 [consulté le 10 mai 2019].

14. Légifrance : <https://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000006575132&cidTexte=LEGITEXT000006071154&dateTexte=19930301> [consulté le 11 décembre 2019]

15. Légifrance : https://www.legifrance.gouv.fr/affichCodeArticle.do;jsessionid=62F4026FB69ECD7B0C8350B6FB04D828.tplgfr42s_2?idArticle=LEGIARTI000006575065&cidTexte=LEGITEXT000006071154&dateTexte=20001231 [consulté le 11 décembre 2019]

Des dispositions spécifiques sont prévues pour certains types de publics : les **auditions** de mineurs doivent être filmées, tout comme les **auditions** de personnes mises en cause (voir section 2.2.9). Ces enregistrements sont versés au dossier de **procédure**. Si la personne n'est pas capable de s'exprimer en français, il peut être fait appel à un interprète en langue des signes ou en langue étrangère ou régionale. Enfin, la personne entendue peut être assistée par un avocat.

La figure 6 présente la première page d'une **audition** de **témoin**. Les propos tenus lors de l'**audition** sont consignés après les différents éléments de contexte que nous avons décrits ci-dessus. Au bas de chaque page, la personne auditionnée et les **officiers de police judiciaire** apposent leurs signatures. La longueur des **procès-verbaux** d'**audition** varie de une à plusieurs dizaines de pages (notamment pour les **auditions** de **garde à vue**).

Dans l'exemple 6, les propos sont consignés sous la forme d'un dialogue entre les gendarmes et le **témoin** entendu. Bien que se calquant sur l'échange oral de l'**audition**, cette forme dialogique du texte du **procès-verbal** n'est pas systématique : nous avons rencontré des **auditions** consignées entièrement sous la forme de récit à la première personne, comme cela est préconisé notamment par la méthode d'interrogatoire ProGREAI (BLANCHET et al., 2013).

Contrairement aux autres documents d'information de l'**enquête**, la pratique de l'**audition** a déjà fait l'objet de recherches en sciences humaines. Plusieurs domaines l'ont étudiée : le droit en étudie la constitutionnalité et les rapports entre **audition** libre et **audition** sous contrainte (LAMY, 2012), la psychologie a produit des recherches sur les entretiens cognitifs (BRUNEL et al., 2013) ou les **auditions** d'enfants (VERKAMPT et al., 2010; VILAMOT et al., 2010; VERKAMPT, 2013), et dans une certaine mesure, la linguistique s'est également penchée sur le sujet par le biais de l'interprétation (DRIESEN, 2016). Hormis ces travaux, nous n'avons pas rencontré de recherches qui se seraient intéressées à la matière textuelle produite par les **auditions** d'**enquête**. Or, l'**audition** d'**enquête**, par la quantité d'information qu'elle introduit dans le dossier, les conditions linguistiques de sa réalisation, et la singularité des pièces qu'elle produit, mérite une attention toute particulière. Nous avons tâché de l'étudier et d'en proposer une définition en tant que genre textuel au sein du chapitre V afin de cerner les enjeux à la fois pratiques et théoriques de cette technique

d'enquête.

2.2.3 Procès-verbaux des actes d'enquête

Sous la catégorie des **procès-verbaux** d'actes d'enquête, nous rassemblons différents types de documents : les comptes-rendus de **perquisition**, de porte-à-porte, de ratissage (fouille d'une zone à la recherche d'éléments matériels), de **tapissage**, les filatures, les écoutes téléphoniques, les contrôles d'identité, les relevés de plaques de véhicules... Dans ce type de documents, les enquêteurs indiquent toute opération menée dans l'objectif de la recherche d'éléments ou de témoignages, ou consignent des faits qu'ils ont constatés pouvant intéresser les investigations. Ces **procès-verbaux** peuvent aller de une à plusieurs pages selon les opérations rapportées.

L'exemple de la figure 7 présente les éléments typiques du **procès-verbal** avec les en-têtes suivies du récit des opérations menées par l'auteur, qui est seul à signer le document. Les propos éventuellement rapportés par des **témoins** le sont au discours indirect, contrairement aux **auditions**. Le contenu est textuel et varié, allant de phrases complètes rédigées à des listes d'éléments d'intérêt (identités, immatriculations de véhicules, etc.).

2.2.4 Procès-verbaux de synthèse

Nous avons déjà évoqué les **procès-verbaux** de synthèse dans le chapitre I section 5.4.2. Leur rédaction n'est pas réservée au stade de l'**analyse criminelle**, mais peut intervenir à d'autres moments de l'**enquête**, par exemple à la clôture de l'**enquête de flagrance**. Le **procès-verbal** de synthèse fait le bilan des investigations réalisées et des moyens mis en œuvre au moment de la rédaction et produit des hypothèses de travail appuyées sur les éléments référencés en **procédure**, comme l'illustre l'extrait page 38.

GENDARMERIE NATIONALE		PROCEDURE SUR COMMISSION ROGATOIRE PROCES - VERBAL D'AUDITION	N° Pièce <i>05/AL</i>	N° Feuillet 1
Compagnie ou escadron SECTION RECHERCHE				
Unité				
C.U.	Procès Verbal			
	907/07 BT			
(ANALYSE ET REFERENCES)				
Audition _____ – Père de _____ et _____ dont _____ avait la charge				
Le _____ à dix heures et trente minutes,				
Nous soussigné Mdl-Chef _____ de la SR _____ et Gendarme _____ de la BR _____, Officiers de police judiciaire,				
Vu les articles 16 à 19 et 151 à 155 du Code de Procédure Pénale.				
Nous trouvant à _____, 10 rue _____, au _____,				
rapportons les opérations suivantes :				
RENSEIGNEMENTS SUR LA DELEGATION				
Date et numéro - NOM et fonction du magistrat				
N° _____, en date du _____ délivrée par Madame _____, Vice-présidente chargée de l'Instruction au TGI de _____				
Information ouverte contre :				
X				
Pour (qualification des infractions) :				
Homicide volontaire				
Transmission :				
Mission :				
ENQUETE				
Nous faisons comparaître devant nous, le témoin ci-après nommé et lui donnons connaissance des faits pour lesquels sa déposition est requise. ----				
Après avoir prêté serment de dire toute la vérité, rien que la vérité, le témoin, entendu séparément dépose ainsi qu'il suit :				
Je me nomme: _____				
née le _____ à _____ (FRANCE). Nationalité française. Célibataire				
Filiation: née de _____ et _____				
Je demeure: 10 rue _____ à _____ (FRANCE) et 18 rue de _____ à _____, chez sa soeur _____, tél: 06. _____				
Profession : Peintre en bâtiment				
---Je suis la père de _____ née le _____, d' _____ née le _____ et de _____ né le _____				
Concernant ses filles _____ et _____ :				
Question: Pouvez-vous nous donner les raisons du placement de vos deux filles, _____ et _____ ?				
Réponse: Comme ma femme a eu un accident de scooter le _____ et qu'elle est hospitalisée et que moi je n'ai pas encore reconnu les enfants, j'ai téléphoné à une assistante sociale de _____ pour trouver une solution. D'autant plus qu'avec mon travail je ne peux pas m'en occuper. Je suis passé avec les filles au centre médico-social de _____ le _____, il me semble. J'ai vu M. _____ et une assistante sociale. Ils m'ont proposé le placement dans une famille d'accueil. J'ai dit que ma femme en aurait pour 1 an de convalescence, et on m'a dit que les filles seraient placés pour une durée minimum de 1 an. ----				
Pour moi c'était une bonne solution, au moins les enfants étaient placés. Nous avons un fils, _____, qui est également placés. C'est _____ qui avait demandé que _____ soit placé, car elle avait des problèmes pour s'en occuper, car il était très actif et ne faisait que des conneries. ----				
J'ai assisté à la décision de placement du Juge au Tribunal à _____ Il me semble que c'était salle 122 ou 126. M. _____ était également présent ainsi que l'assistante sociale de _____, Mme _____, ----				
La personne entendue			Les OPJ	
_____			_____	

FIGURE 6 – Exemple d'audition de témoin

GENDARMERIE NATIONALE <small>GROUPEMENT</small> [REDACTED] <small>UNITE</small> [REDACTED] <small>BT</small> [REDACTED] <small>C.U.</small> : [REDACTED] <small>PV N°</small> 905/2000	PROCEDURE SUR COMMISSION ROGATOIRE PROCES-VERBAL DE RENSEIGNEMENT	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;"><small>Pièce</small></td> <td style="width: 20%; text-align: center;">21/1</td> <td style="width: 20%;"><small>Feuille</small></td> </tr> </table>	<small>Pièce</small>	21/1	<small>Feuille</small>
<small>Pièce</small>	21/1	<small>Feuille</small>			

Nous soussignés, [REDACTED], Gendarme, Officier de police judiciaire,
 en résidence à la brigade de [REDACTED]

Vu les articles 16, 17 à 19, 151 à 155 du Code de Procédure Pénale,
 Rapportons les opérations suivantes que nous avons effectuées en exécution de la commission rogatoire jointe
 délivrée en application de l'article 18 alinéa 4 du Code de procédure pénale.-----
 Nous trouvant à [REDACTED]. Après en avoir informé le magistrat mandant à [REDACTED] ainsi que l'officier de police
 judiciaire de territorialement compétent :-----

●**RENSEIGNEMENTS SUR LA DELEGATION** Date [REDACTED] N° 200/0068
Nom et fonction du magistrat :
 Mme [REDACTED], Vice-président chargée de l'instruction au tribunal de
 grande instance de [REDACTED]

Vu l'information ouverte pour :
 Homicide volontaire

A l'encontre de :
X.....

●**TRANSMISSIONS :** Date N°
Grade, Nom, Prénom :
Fonction :

●**MISSION : VOIR COMMISSION ROGATOIRE JOINTE**

Le mercredi [REDACTED]

---Dans le cadre de la commission rogatoire de référence, nous nous portons à [REDACTED] à la
 brigade de Gendarmerie afin de contrôler un individu ayant une ressemblance avec le portrait robot
 n° 5.-----

---L'intéressé M. [REDACTED], commerçant à [REDACTED] se présente à notre unité à
 [REDACTED]. Nous procéderons à son audition ce même jour. Il sera de même effectué avec son accord
 à un tapissage puis sur prescription du juge d'instruction à un prélèvement sanguin.
 (CF pièce 21/1, 21/2, 21/4).-----

---Le témoin, [REDACTED], ne reconnaît pas [REDACTED] qui par ailleurs elle
 connaît en temps que personne demeurant sur la commune de [REDACTED].-----
 (CF pièce 21/3)

---Le cousin de [REDACTED], [REDACTED] sera également entendu dans
 le cadre de la présente vérification.
 (CF pièce 21/6).

---Le [REDACTED] nous sommes destinataires des résultats d'analyse de l'ADN de
 [REDACTED]. La comparaison est négative.

---Fait et clos à [REDACTED] le [REDACTED]

L'Officier de police judiciaire

FIGURE 7 – Procès-verbal de renseignement

- Les hypothèses de travail :

Le suicide ayant été écarté d'emblée, les premières hypothèses de travail ont pu être échaffaudées à l'aide des témoignages puis sur les axes habituels d'investigations d'enquêteurs.

[...]

HYPOTHESE 2 :

Ces derniers ont quitté les lieux à 13H00 en emportant des sacs et en utilisant leur voiture. Ce départ pouvait paraître suspect d'autant qu'ils savaient que la jeune fille était recherchée. Ils sont en outre des témoins importants quant à l'emploi du temps de la victime. - Cf pièces n°15, 16, 26, 47, 48, 49, 50 -

2.2.5 Rapports d'expertise

Les rapports d'expertise sont obtenus sur **réquisition**. Différents types de rapports d'expertises peuvent figurer dans la **procédure** avec pour objectif de fournir des éléments de preuve scientifique : dans les cas d'homicides, les examens médico-légaux comme l'**autopsie** permettent de dater le décès, d'en identifier la cause, ainsi que de récupérer des **traces** biologiques du ou des auteurs. Les expertises psychiatriques dressent le profil psychologique d'un individu, permettant notamment de décrire sa personnalité ou de déterminer sa responsabilité pénale.

Les expertises complémentaires portent sur les prélèvements biologiques lors de l'autopsie, comme le prélèvement des organes pour procéder à des analyses biologiques (**traces** de sperme, de sang, d'**ADN**), toxicologiques (alcool, produits stupéfiants). Les **scellés** collectés sont aussi analysés à la recherche de ce type de **traces**. Les analyses sont documentées par des rapports versés au dossier, éventuellement agrémentés de planches photographiques (notamment pour les **autopsies**), des documents qui se distinguent de ceux que nous avons décrit jusqu'à présent. Comme le montrent les extraits aux figures 8 et 9, ils relèvent d'une pratique professionnelle spécifique et convoient une information technique dont la compréhension nécessite des compétences médicales ou des connaissances en chimie et biologie. Ces rapports présentent souvent des résultats sous forme de tableaux et de listes de données chiffrées.

I - TECHNIQUES

- **La recherche de l'alcool éthylique** a été réalisée sur l'échantillon de liquide séro-hématique par chromatographie en phase gazeuse avec un système d'injection par espace de tête (Headspace) de l'alcool volatilisé dans un flacon scellé, séparation sur colonne capillaire HP Wax et détection par ionisation de flamme en utilisant le propanol 1 comme standard interne, selon la méthode officielle (arrêté du 6 mars 1986). La limite de détection est de 0,01 g/l. Cette technique permet l'identification et la quantification de l'éthanol, du méthanol et de l'isopropanol.

FIGURE 8 – Extrait de rapport d'analyse présentant une technique

POUMONS

Les différents fragments pulmonaires ont été débités en coupes semi-sériées
Les poumons ont perdu leur caractère aérien.
Les alvéoles et cloisons interalvéolaires ont à peu près complètement disparu car elles sont noyées dans une sérosité hémorragique.
Quand les septa interalvéolaires sont visibles, ils sont souvent rompus.
Les bronches et les bronchioles ont conservé leur revêtement de type cylindrique respiratoire et elles sont libres de tout obstacle.
Les vaisseaux sont tous dilatés et gorgés de sang.

FIGURE 9 – Extrait de rapport d'analyse d'organe

Autopsies. La Société Française de Médecine Légale a publié en 2009 des lignes directrices pour la rédaction de rapports d'autopsie (SOCIÉTÉ FRANÇAISE DE MÉDECINE LÉGALE, 2009), dans lesquelles est soulignée la difficulté d'harmonisation face à la diversité des formations et des pratiques. On y préconise une structure en plusieurs rubriques rappelant la situation de découverte, décrivant l'examen externe (figure 10) puis l'examen interne (l'autopsie médico-légale), les prélèvements effectués et enfin la conclusion. Le rapport d'autopsie est rédigé dans une langue faisant appel au vocabulaire médical, tout en apportant en conclusion des interprétations accessibles à un lecteur non-expert (figure 11 et extraits suivants).

- les multiples lésions de la nuque indiquent que la tête de la victime devait alors être fléchie en avant, c'est-à-dire que la victime était à ce moment-là soit décédée, soit sans connaissance ; là encore, cette topographie lésionnelle indique une victime non pas debout, mais assise.

Suite à la levée de corps et aux opérations thanatologiques les Experts

A/ EXAMEN EXTERNE :

SEXE : Féminin
 HABITUS : corpulence mince.
 Corps plein de sable.

PHÉNOMÈNES CADAVÉRIQUES :

- RIGIDITE : marquée et généralisée.
- LIVIDITES : majoritairement de siège postérieur et très violacées, avec respect des zones d'appui et minoritairement sur la face antérieure du tronc.
- TACHE VERTE ABDOMINALE : non développée.

TÊTE :

- Sable dans la cavité buccale. Présence d'un cerclage des arcades dentaires indiquant un suivi orthodontique.
- Léger écoulement rosé par les narines, lors de la mobilisation du corps.
- Etat congestif des vaisseaux sous-conjonctivaux avec pétéchies à la partie supérieure des globes oculaires
- Une excoriation superficielle, punctiforme, entre les sourcils..
- Intégrité apparente des os propres du nez.
- Une excoriation punctiforme de 0,3 cm sur la partie gauche du milieu de l'arrête nasale.

FIGURE 10 – Extrait de rapport d'autopsie

évoquent une mort par asphyxie mécanique, engendré par le double mécanisme d'une strangulation et d'une suffocation. Actuellement il leur est impossible de déterminer s'il y a eu agression sexuelle.

2.2.6 Factures téléphoniques

Les factures téléphoniques détaillées (couramment appelées « fadet » par les enquêteurs) listent les communications d'un client et d'autres informations intéressantes : pour un client donné, on y voit les numéros utilisés, les numéros contactés avec horodatage, le numéro IMEI¹⁶ du boîtier téléphonique depuis lequel la communication a été passée ou reçue (ce qui permet d'identifier l'utilisation d'un même téléphone par plusieurs numéros), le relais téléphonique par lequel la communication est passée, le sens de la communication et la nature de l'échange. Ces documents étaient encore récemment adressés sur [réquisitions](#) par les opérateurs téléphoniques. En France, depuis avril 2017 une agence du ministère de la justice spécialisée dans

16. Le numéro IMEI est un numéro d'identification des appareils de téléphonie mobile.

CONCLUSION

1°/ Il s'agit du corps de ██████████, née le ██████████, découvert sans vie, le ██████████, sur ██████████ 2 kilomètres au nord de ██████████ et porté disparue depuis la fin d'après-midi ██████████.

2°/ D'après les Enquêteurs, le corps été recouvert de sable.

3°/ La levée de corps, effectuée le jour même vers 19 heures, par le docteur P. ██████████ permettait de noter : un corps en décubitus dorsal, froid, rigide et portant des lividités très cyanotiques à la face postérieure du corps et paradoxalement sur la face antérieure du tronc, laissant évoquer une mobilisation secondaire du corps. La mort pouvait être estimée comme remontant 10 à 24 heures avant le ██████████ 19 heures.

FIGURE 11 – Extrait de conclusion de rapport d'autopsie

IMSI	Date appel	Heure début	Durée	Numéro appelant	Numéro appelé	N° de série du téléphone	Numéro Cellule	Sens d'appel	Nature d'appel
208 ██████████	██████████	11:27:26	00:00:19	3369 ██████████	3365 ██████████	3543 ██████████	208 ██████████	Sortant	Voix
208 ██████████	██████████	13:27:12	00:00:01	3369 ██████████	3361 ██████████	3543 ██████████	208 ██████████	Sortant	Voix
208 ██████████	██████████	13:27:35	00:00:31	3361 ██████████	3369 ██████████	3543 ██████████	208 ██████████	Entrant	Voix
208 ██████████	██████████	14:11:31	00:00:00	3369 ██████████	3362 ██████████	3543 ██████████	208 ██████████	Sortant	SMS
208 ██████████	██████████	14:11:32	00:00:00	3362 ██████████	3369 ██████████	3543 ██████████	208 ██████████	Accusé de réception	SMS
208 ██████████	██████████	14:12:55	00:00:00	3362 ██████████	3369 ██████████	3543 ██████████	208 ██████████	Entrant	SMS

FIGURE 12 – Extrait de facture téléphonique détaillée

l'interception des communications numériques a été mise en place, l'Antenj. Cette plateforme pilote la plate-forme nationale des interceptions judiciaires (PNIJ), un portail qui permet aux enquêteurs d'obtenir des informations liées aux télécommunications via une demande en ligne, dont entre autres les factures détaillées. En plus de faciliter l'accès aux données de télécommunication, la mise en place de cette plateforme a permis l'uniformisation des formats de documents communiqués par les opérateurs, et ainsi le développement d'utilitaires de traitement des données issues des factures téléphoniques, ce qui n'était pas le cas auparavant.

La figure 12 est un extrait de facture téléphonique, où l'on peut voir que les factures sont structurées sous la forme de tableaux de données où chaque colonne correspond à un type d'information et chaque ligne à une communication.

2.2.7 Relevés bancaires

Les relevés bancaires sont fournis sur [réquisition](#) par les établissements bancaires afin de constater les mouvements de fonds : virements, salaires, prestations sociales,

20.01	Carte	Air France Roissy Cdg Cede	16/01	-245,52	<input type="checkbox"/>
20.01	Carte	Amazon Premium PARIS	18/01	-49,00	<input type="checkbox"/>
20.01	Carte	SNCF Paris	19/01	-3,40	<input type="checkbox"/>
22.01	22.01	Avoir	Carte Amazon Premium	21/01	49,00 <input type="checkbox"/>
22.01	Carte	Match Dac	21/01	-55,00	<input type="checkbox"/>
23.01	23.01	Prlv	Free Mobile	-19,99	<input type="checkbox"/>
27.01	27.01	Carte	Apple Itunes Store-eu	24/01	-17,99 <input type="checkbox"/>
27.01				25/01 11H42	-20,00 <input type="checkbox"/>

FIGURE 13 – Extrait de relevé bancaire. La mention « RET DAB » indiquant un retrait au distributeur automatique a été surlignée par l’enquêteur (dernière ligne), suivie de la localité du retrait (camouflée).

retraits d’espèces, émissions de chèques, pensions, etc. Les retraits en espèces permettent par exemple de vérifier la présence géographique d’une personne à un moment donné, et les transferts d’argent attestent de contacts entre personnes. On peut également identifier des dépenses suspectes, comme des achats d’objets potentiellement impliqués dans les faits (figure 13).

Les relevés bancaires présentent sur une ligne la date, la nature de l’opération, son objet et le montant entrant ou sortant. Tout comme les données téléphoniques, ce sont des documents structurés sous forme de tableaux, mais à l’inverse, les formats ne sont pas uniformisés entre établissements bancaires.

2.2.8 Autres documents versé au dossier de procédure

Toutes sortes d’autres documents peuvent être versés au dossier de [procédure](#) par les enquêteurs. Nous avons relevé une très grande variété de documents, souvent ajoutés en annexe de [procès-verbaux](#) :

- Avis de recherche,
- Lettres anonymes, de revendication,
- Documents administratifs : actes d’état-civil, permis d’inhumation, copies de livrets de famille, dossiers pénitentiaires,
- Coupures de presse,
- Correspondance avec les représentants des parties : notamment les demandes d’actes,
- Documents fournis par les services de police municipale ou des sociétés de sécurité : compte-rendus de ronde par exemple,
- Documents commerciaux : fichiers clients,

- Documentation diverse, catégorie au sein de laquelle nous avons rencontré de la documentation sur des sectes et des éléments bibliographiques sur des techniques d'enquête.

Cette liste, bien entendu, est loin d'être exhaustive. N'importe quel document peut être ajouté au dossier pourvu qu'il puisse intéresser l'enquête.

2.2.9 Documents photographiques, vidéos ou graphiques divers

De nombreuses étapes des investigations sont documentées par des photographies et des vidéos : prises de vue de la scène de crime et de ses environs (y compris des vues aériennes), photos du cadavre, photos de l'autopsie, photos de véhicules, de scellés, d'éléments matériels, portraits de personnes suspectes, photographies de tapissage, photographies prises lors de reconstitutions, etc., portant parfois des annotations ou des mentions. En plus de ces photos visant à documenter l'enquête, des photos issues d'appareils appartenant à des personnes impliquées peuvent être collectées comme pièces à conviction. Elles sont soit intégrées à des documents sous forme de planches photographiques, soit rassemblées dans des répertoires informatiques.

Les portraits-robots sont des représentations graphiques du visage de personnes recherchées, établies d'après une description fournie par un témoin. De nos jours, ils sont réalisés à l'aide de logiciels spécialisés qui permettent d'assembler des traits physiques.

Pour les vidéos, on trouve les enregistrements des gardes à vue¹⁷ et de certaines auditions, ou encore des extraits de caméras de vidéo-surveillance.

Enfin, des cartes type IGN ou état-major peuvent être ajoutées au dossier, tout comme des plans manuscrits ou des schémas produits par des témoins.

Les documents graphiques, de natures variées, apportent des informations complémentaires aux autres documents de la procédure. Leur fonction peut être illustrative (plans, cartes, photos des lieux et des faits), légale (captation des auditions),

17. Les conditions d'enregistrement sont précisées aux articles 64-1 <https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006071154&idArticle=LEGIARTI000006575093&dateTexte=&categorieLien=cid> et D15-6 <https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006071154&idArticle=LEGIARTI000006514842&dateTexte=&categorieLien=cid> du Code de procédure pénale. [consultés le 25 juin 2019]

d'investigation ([portraits-robots](#)), ou être des éléments de preuve en tant que tels (vidéo-surveillance).

2.3 Synthèse

Le volume de chaque type de document au sein de la [procédure](#) est dépendant de l'affaire traitée. Les pratiques d'[enquête](#) ont recours à l'analyse des télécommunications et des mouvements de fonds en ce qui concerne le trafic de stupéfiants. Dans ces cas, les données téléphoniques et bancaires représenteront donc un volume de documents important au sein du dossier de [procédure](#), alors que dans d'autres types d'affaires, les [auditions](#) auront la place centrale. Les autres types de documents, notamment les rapports d'expertise, sont moins nombreux et ne sont pas la source principale de l'information exploitée par l'[analyse criminelle](#).

Dans cette section, nous avons présenté un panorama non-exhaustif des documents qui constituent le dossier de [procédure](#) judiciaire traité en [analyse criminelle](#). Nous avons pu constater une grande variété dans la nature des documents, aussi bien en ce qui concerne leurs conditions de production, leurs fonctions, leurs formes et leurs émetteurs. À propos des émetteurs, il faut noter pour la poursuite de notre réflexion la multiplicité des acteurs impliqués dans la production des documents, membres de l'institution judiciaire ou civils réquisitionnés : enquêteurs, techniciens, magistrats, médecins, experts... Cet aspect fait de l'élaboration du dossier de [procédure](#) un processus collaboratif orchestré par la direction de l'[enquête](#).

Malgré ces aspects, tous les documents de la [procédure](#) visent à clarifier les faits investigués. Ils partagent donc au moins, dans les grandes lignes, une finalité commune. Dans la suite de ce chapitre, nous essaierons de déterminer les aspects qui rapprochent le dossier de [procédure](#) d'un [corpus](#) textuel, ce qui permettra par la suite d'envisager l'adaptation des méthodes de la linguistique de [corpus](#) à la pratique de l'[analyse criminelle](#).

3 Le texte de la procédure judiciaire

La description de la matière rassemblée dans le dossier de **procédure** judiciaire, remise dans la perspective des besoins de l'**analyse criminelle** présentée au chapitre I, conduit à envisager les directions suivantes :

- Le développement d'approches automatiques d'**extraction d'information**,
- À terme, la conception d'une méthodologie d'accès au texte inspirée de la linguistique de **corpus** et de l'analyse des données textuelles (**ADT**).

Nous concevons ces deux directions comme complémentaires : il s'agirait d'élaborer un outil d'exploration textuelle *ad hoc* pour les **analystes criminels**, impliquant, entre autres possibilités d'application de technologies de traitement automatique des langues (**TAL**), un repérage automatique de certaines informations et permettant une navigation facilitée au sein des textes.

Afin de situer nos données de travail au regard de ces deux dernières disciplines, il convient de s'attarder sur un concept qui leur est fondamental : celui du **corpus**. Toute archive documentaire ne constituant pas nécessairement un **corpus**, nous souhaitons déterminer si et selon quelles modalités le dossier de **procédure** judiciaire se prête à l'application des méthodologies en question. Pour cela, nous considérerons la (les) définition(s) du **corpus** selon la linguistique de **corpus** ainsi que deux types de **corpus**.

3.1 Le concept de corpus selon la linguistique de corpus

La linguistique de **corpus** est une branche de la linguistique qui prend comme objet d'étude des **corpus**, c'est-à-dire des ensembles de textes authentiques, au contraire d'autres branches de la linguistique qui travaillent à des niveaux différents, par exemple celui de la phrase (l'énoncé), et sur des exemples construits pour la démonstration (artificiels).

Avec le développement de méthodes informatiques appliquées au texte, les pratiques de la linguistique de **corpus** comme le relevé d'occurrences, les concordances, ou l'application de méthodes statistiques comme l'analyse factorielle des correspondances (**AFC**) (BENZÉCRI, 1973; SALEM, 1986; LEBART et SALEM, 1994; HABERT et al., 1997) ont été décuplées. Aujourd'hui, de nombreux logiciels et outils s'adressent

aux linguistes plus ou moins familiers de l'informatique pour leur offrir des voies d'entrées dans les **corpus**. Le développement technique s'est accompagné d'une réflexion théorique sur la constitution des **corpus**, les modes d'approche (*corpus-based* ou *corpus-driven*), la validité des pré-traitements (annotations, lemmatisation), la représentativité des résultats, etc. Les pratiques de la linguistique de **corpus** et leur potentiel dans notre cas d'étude seront plus amplement présentés dans le chapitre III.

L'un des grands questionnements de la linguistique de **corpus** et qui nous intéresse tout particulièrement réside dans la définition même du **corpus** : SINCLAIR (1991) le définit comme

A collection of naturally occurring language text, chosen to characterize a state or variety of a language,¹⁸

MAYAFFRE (2002) renvoie à la littérature spécialisée et à une définition encyclopédique selon laquelle un **corpus** est

un rassemblement de textes ou une collection de textes regroupés sur la base d'hypothèses de travail en vue de les interroger

À propos de la structure du **corpus**, RASTIER (2004) oppose conception documentaire à conception philologique-herméneutique, et propose la définition suivante :

Un **corpus** est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications.

GARRIC et LONGHI (2012) indiquent qu'il n'existe pas de consensus sur la notion de **corpus** tout en proposant la définition suivante :

Un **corpus** est défini comme un ensemble raisonné de textes, structuré par une cohérence interne.

De ces définitions, on retient l'idée d'un **rassemblement de textes** qui n'est pas le fruit du hasard puisque le **corpus sert une intention** et répond à une **structure interne**.

18. « Une collection de textes linguistiques naturels, choisis pour caractériser un état ou une variété d'une langue » (traduction libre)

Ces définitions presque dictionnairiques sont complétées dans la littérature par des critères de bonne formation parmi lesquels on rencontre homogénéité, contrastivité, diachronicité selon MAYAFFRE (2002), signifiante, acceptabilité, exploitabilité selon PINCEMIN (2012a). L'objectif de ces critères est de légitimer les observations du chercheur en garantissant la qualité de la production du *corpus* exploité : le critère d'acceptabilité par exemple évalue entre autres la représentativité des données considérées afin d'objectiver au maximum les observations.

3.2 Corpus réflexifs, corpus hétérogènes

Parmi les différents *corpus* impliqués en linguistique de *corpus*, deux types en particulier attirent notre attention : les *corpus* réflexifs et les *corpus* hétérogènes.

Le concept de *corpus* réflexif a été établi par D. MAYAFFRE dans son article précédemment cité. La réflexivité du *corpus* tiendrait dans « le fait que ses constituants [...] renvoient les uns aux autres pour former un *réseau sémantique* performant dans un tout (le *corpus*) cohérent et auto-suffisant ». Connectée à l'hypertextualité et à l'intertexte, cette notion mêle la question de la clôture du *corpus* et celle des ressources nécessaires à son interprétation. L'exemple employé par l'auteur est celui du champ des discours politiques, dans lequel les productions verbales s'inscrivent dans un réseau de discours déjà énoncés.

En ce qui concerne l'homogénéité, PINCEMIN (2012a) déclare qu'elle ne doit pas être une condition de prise en charge d'un *corpus*, pourvu que l'on dispose d'une bonne connaissance de ses défauts et que l'on en fasse un usage éclairé, ce que l'auteur appelle le critère d'*interprétabilité*. L'intérêt de la proposition est de dépasser les critères de bonne formation d'un *corpus* en posant un cadre pour l'étude de *corpus* imparfaits. Deux formes d'hétérogénéité sont rapportées : des aspects techniques tels que l'encodage des fichiers ou les questions d'échantillonnage, et des aspects de variation linguistique liés, notamment, au genre textuel.

3.3 Et la procédure ?

L'hétérogénéité des composants du dossier de *procédure* est flagrante à la lecture de la première section de ce chapitre : les contenus, genres, fonctions, volumes

et émetteurs variés des documents génèrent une hétérogénéité sur les plans aussi bien techniques que linguistiques. Que ce soit à propos de la méthode de l'[analyse criminelle](#) ou de notre question de recherche (pour rappel, l'adaptation d'approches automatiques à la pratique de l'[analyse criminelle](#)), cette forte hétérogénéité des documents et du texte est un aspect crucial à prendre en compte dans notre perspective de recherche, car c'est aux outils et aux méthodes de s'adapter aux données et pas l'inverse (PINCEMIN, 2012a).

L'aspect réflexif est moins évident mais néanmoins intéressant à souligner car il se retrouve sur deux plans.

Tout d'abord, comme nous l'avons décrit dans le cas des [procès-verbaux](#) de synthèse par exemple (en 2.2.4), les pièces de [procédure](#) peuvent référer explicitement les unes aux autres en mentionnant leurs cotes. On retrouve ainsi le critère de ressource d'interprétation interne : une information est avancée sur la base d'un autre passage du [corpus](#), les informations peuvent être interprétées en regard les unes des autres. L'aspect d'auto-suffisance est même prévu par la loi puisque l'[analyse criminelle](#) et plus largement l'action judiciaire reposent sur l'exploitation des éléments actés dans le dossier de [procédure](#) exclusivement.

D'autre part, les [procès-verbaux](#), dans leurs en-têtes légaux, font mention des textes de loi cadrant la production du document (exemple à la figure 7 : « Vu les articles 16, 17 à 19, 151 à 155 du [Code de procédure pénale](#) »). Le texte de loi est la ressource interprétative des aspects légaux du document, que chaque partie impliquée dans sa constitution est, théoriquement, censée connaître puisque nul n'est censé ignorer la loi. Les textes de loi ne sont pas inclus à proprement parler dans le dossier de [procédure](#). Mais est-il nécessaire et pertinent d'intégrer la loi, texte public par excellence, référencé et structuré, que nul n'est censé ignorer, et surtout pas les professionnels de la justice, au dossier de [procédure](#) ?

Réflexif et hétérogène, il reste néanmoins à caractériser le caractère de [corpus](#) du dossier de [procédure](#). Si l'on reprend les éléments définitoires énoncés à la sous-partie 3.1 :

- Rassemblement : comme archive ou comme [corpus](#), le dossier de [procédure](#) compile des documents,

- Textes : tous les documents ne sont pas des textes, mais une partie non-négligeable en est,
- Finalité : tous concourent à un même objectif, celui de faire la lumière sur des faits criminels,
- Structure : si elle manque de normalisation, la structure du dossier de **procédure** n'est pas inexistante pour autant. Au minimum, elle correspond au découpage en pièces de **procédure**.

Globalement, le dossier de **procédure** remplit les critères définitoires du **corpus** précédemment dégagés. La finalité est un aspect qui diverge des définitions du **corpus** produites par la linguistique textuelle, puisqu'elle n'est pas d'étudier des phénomènes linguistiques. On peut la réajuster dans la perspective de notre recherche, dans ce cas, la finalité du dossier de **procédure** devient d'être un support de développement d'approches de **TAL** et de description des phénomènes linguistiques de l'**enquête**. Ceci posé, nous pouvons nous intéresser plus précisément aux manipulations et pré-traitements nécessaires avant son utilisation pour notre recherche.

3.4 Constitution du corpus de recherche

Est dit *textuel* ce qui procède d'une langue, se transcrit, s'articule en lettres et en mots dans une écriture, par opposition à ce qui relève d'autres médias : les images, les sons. (PINCEMIN, 1999, p. 135)

Dans le jargon des bases de données, les champs textuels s'opposent aux champs factuels et numériques. Ce qui est factuel, c'est ce qui prend sa valeur parmi un ensemble donné d'alternatives (vrai / faux, codes départementaux, date, répertoire de noms d'auteurs, etc.). Pour ce qui est textuel, il n'y a pas de liste de possibilités prévues, la seule contrainte est en général une longueur maximale. (*ibid.* p. 136)

Resituons notre recherche : les **procédures** que nous avons à notre disposition sont donc, de notre point de vue, à envisager comme un **corpus** que nous allons employer comme « bac à sable » de test et d'adaptation de méthodes de **TAL**, en

orientant nos travaux par la réalisation d'une étude des textes de la [procédure](#) judiciaire, notamment des réalisations linguistiques des entités qui intéressent l'[analyse criminelle](#).

Nous sélectionnerons une partie de la [procédure](#) pour remplir ce rôle, afin de retenir les documents qui s'y prêtent le mieux, aussi bien en termes d'exploitabilité immédiate que de l'intérêt des informations qu'ils comportent.

Il ressort de la présentation des pièces de [procédure](#) réalisée lors de la première partie de ce chapitre que :

- Tous les documents ne sont pas des textes (documents graphiques),
- Certains documents sont des extractions de bases de données (documents téléphoniques et bancaires),
- Certains documents sont des textes ou comportent du texte, mais leur statut de documents réglementaires réduit leur apport en information et leur confère une forme rigide et procédurale (bordereaux d'envoi, inventaires des pièces à conviction, [procès-verbaux](#) de notification de [garde à vue](#), [réquisitions](#), certificats médicaux de [garde à vue](#), [procès-verbaux](#) de saisie),
- Certains consignent des pans de texte libre ([procès-verbaux](#) transport constatations mesures prises, synthèse, [actes d'enquête](#), [auditions](#), rapports d'expertise).

La dernière catégorie, celle des documents comportant des pans de texte libre, est donc celle qui paraît le mieux convenir à notre objectif, néanmoins, le volume de chaque type de document en son sein est variable. Pour une affaire donnée, on trouvera un [procès-verbal](#) transport constatations mesures prises, une ou deux synthèses selon le stade de l'[enquête](#), un nombre indéterminé d'[actes d'enquête](#) et d'[auditions](#). En ce qui concerne les expertises, leur nature et leur nombre sont imprévisibles, et bien qu'elles comportent des pans de texte, une partie de leur contenu peut aussi prendre la forme de tableaux de données et d'images, ce qui ne correspond pas à nos contraintes. Le document d'expertise le plus exploitable serait le rapport d'autopsie, mais quel que soit le nombre de victimes de l'affaire étudiée, il sera toujours vraisemblablement inférieur au nombre de [témoins](#) auditionnés. Dans les [procédures](#) dont nous disposons, chaque affaire comporte une autopsie, ce qui ne représente pas un volume suffisant pour être traité selon les méthodes que nous envisageons.

L'étude des rapports d'expertise criminalistique, qui n'est pas dénuée d'intérêt, demanderait la constitution d'un **corpus** spécifique de documents collectés à travers différentes affaires.

Dernier critère décisif : les **analystes criminels** rapportent se baser en grande partie sur le contenu des **auditions** pour produire leurs analyses. Ce type de document représente en effet un volume conséquent de la **procédure** (dans nos exemples, de plusieurs centaines à plusieurs milliers de documents), convoie une information dense, non-structurée, et porte des enjeux linguistiques très spécifiques qui méritent une étude approfondie.

En conséquence, nous retenons la catégorie des **auditions** pour la poursuite des travaux. Cette catégorie catalyse nos différents intérêts de recherche. D'une part, les « aspects **TAL** » sont satisfaits en constituant une quantité de données pertinente et en rapportant les informations qui intéressent l'**analyse criminelle**, et d'autre part, les « aspects linguistiques » sont remplis car les **auditions** constituent un type de texte singulier dont la description et la compréhension pourront alimenter le développement des méthodes informatiques.

Parmi les **procédures** disponibles pour nos recherches, une seule présente une qualité de reconnaissance de caractères exploitable, c'est-à-dire ne nécessitant pas de correction et permettant son traitement informatique. Cette **procédure** contient 370 **auditions**, qui feront désormais office de **corpus** de travail.

4 Conclusion

Dans ce chapitre, nous avons détaillé la composition du dossier de **procédure** judiciaire exploité par l'**analyse criminelle** d'après des exemples d'affaires réelles.

L'enjeu était de structurer l'organisation de la **procédure** en types de documents, puis d'en cerner les aspects textuels afin de délimiter à la fois le futur champ d'application des outils et les ressources nécessaires à leur développement. Cette phase de description a démontré la singularité du dossier de **procédure** judiciaire, rassemblant des documents de formes et de fonctions très variées, émanant d'un grand nombre d'acteurs différents, mais qui sont tout de même tous produits et consignés dans un seul et même but. La variété des documents, les volumes et natures de texte

qu'ils contiennent imposent un tri à la fois pour notre étude et pour l'intégration des méthodes automatiques de traitement, pour proposer dans un premier temps des observations et des pistes de développement. Guidée par l'ajustement entre l'intérêt, la disponibilité et l'exploitabilité des données, nous faisons le choix d'utiliser un type particulier de documents, les [auditions](#) de [témoins](#).

Maintenant que contexte et matière ont été présentés en détails, nous consacrons le chapitre suivant à constituer un état de l'art adapté à nos préoccupations, c'est-à-dire aux questions de linguistique outillée et d'extraction automatique d'information que nous considérerons dans la perspective de l'[analyse criminelle](#).