De la protection des données à la protection de la personne :

Réflexions sur l'impact des nouvelles réglementations sur la collecte des corpus

Muriel Lalain, Gilles Pouchoulin, Béatrice Priego-Valverde et Serge Pinto

Édition électronique

URL : <https://journals.openedition.org/corpus/5895>

DOI : 10.4000/corpus.5895

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Référence électronique

Muriel Lalain, Gilles Pouchoulin, Béatrice Priego-Valverde et Serge Pinto, « De la protection des données à la protection de la personne :

Réflexions sur l'impact des nouvelles réglementations sur la collecte des corpus », *Corpus* [En ligne],

22 | 2021, mis en ligne le 15 février 2021, consulté le 04 mars 2022. URL : <http://journals.openedition.org/corpus/5895> ; DOI : <https://doi.org/10.4000/corpus.5895>

Ce document a été généré automatiquement le 4 mars 2022.

© Tous droits réservés

De la protection des données à la protection de la personne : Réflexions sur l'impact des nouvelles réglementations sur la collecte des corpus

Muriel Lalain, Gilles Pouchoulin, Béatrice Priego-Valverde et Serge Pinto

Ce document de synthèse émane des réflexions et discussions menées dans le cadre du groupe éthique du Laboratoire Parole et Langage ; nous remercions nos collègues pour les riches échanges et débats autour de ces nouvelles réglementations dont les traces écrites sont téléchargeables sur le site du LPL : <http://www.lpl-aix.fr/recherche/ethique/>

1. Introduction

- 1 Les travaux de recherche en linguistique nécessitent le recueil de données langagières, sélectionnées et rassemblées pour former ce que l'on appelle un corpus. La diversité des approches, des pratiques et des objectifs a complexifié cette notion de corpus. Cependant, qu'il s'agisse de données textuelles, orales, ou multimodales, que le recueil permette la collecte de données authentiques, écologiques, ou de laboratoire, que l'étude suive une méthodologie empirico-déductive ou inductive, les données langagières émanent nécessairement de locuteurs. Ainsi en linguistique, mais aussi dans les différents champs de son interdisciplinarité (sociolinguistique, psycholinguistique, etc.), nos travaux nous conduisent à impliquer la personne humaine et à utiliser les données de cette dernière, à des fins de recherche et/ou de formation. La constitution d'un corpus implique alors nécessairement la contribution de locuteurs dont nous recueillons la parole (donnée primaire) qui fera l'objet d'analyse(s) linguistique(s) (données secondaires) mais aussi d'éléments d'informations

complémentaires qui permettront d'interpréter les résultats des analyses (métadonnées).

- 2 La récente mise en application des deux cadres réglementaires du Règlement Général sur la Protection des Données Personnelles (RGPD 2018) et de la Loi Jardé (Jardé 2012) est venue bousculer les habitudes de la communauté scientifique en SHS de manière générale, puisque ces réglementations s'appliquent finalement à tous les secteurs faisant intervenir la personne humaine. En linguistique, l'impact de ces nouvelles réglementations est appréciable dès la collecte des données langagières. Cependant, il n'est que difficilement mesuré tant les expériences et témoignages de terrain, ou tout simplement les informations claires en matière de protection des données et des personnes sont rares, imprécises et parcellaires. Pour pallier ce manque, les auteurs présentent à travers cet article l'ensemble des questions et réflexions soulevées autour de ces deux nouveaux cadres réglementaires par un groupe de travail mis en place au sein du Laboratoire Parole et Langage (LPL 2020).
- 3 Le RGPD définit les principes juridiques et l'obligation de sécurité qui garantissent les droits et les libertés des personnes. L'application de ce règlement européen dans les établissements de recherche nécessite ainsi la mise en place d'actions de sensibilisation, d'informations permettant d'en comprendre les enjeux et d'en clarifier le cadre juridique applicable depuis le 25 mai 2018. Sa mise en œuvre exige de répondre à de nombreuses questions très diverses telles que : Ai-je le droit de collecter ces données ? Dans quel but ? Sur qui ? Pour qui ? Quelles données ? Où et comment les stocker ? Pour quelle durée ? S'agit-il d'un traitement dit « à risque » ? Quels sont les acteurs traitant les données ? Cette approche réflexive vise à amener la communauté scientifique à s'interroger et à réfléchir sur ses pratiques et usages en matière de collecte et de traitement des données personnelles. Les réponses apportées à ces interrogations décriront la documentation du traitement des données à caractère personnel qui devra être enregistrée auprès du service du Délégué à la Protection de Données (DPD¹) dont dépend le responsable du traitement.
- 4 La loi Jardé (Jardé 2012) est née d'une volonté de simplification de la loi de protection des personnes, jusque-là régie par la loi Huriot-Serusclet (Huriot 1988), spécifiquement applicable aux « personnes se prêtant à la recherche biomédicale ». Entrée en vigueur en novembre 2016, la loi Jardé fixe les dispositions réglementaires relatives aux recherches impliquant plus largement la personne humaine. Dès lors, la première interrogation concerne le champ d'application de la loi Jardé. La diversité des recherches en SHS ainsi que l'interdisciplinarité des laboratoires, rendent ce questionnement à la fois crucial et complexe. Viennent ensuite les questions relatives aux « catégories de recherche », aux « risques de la recherche », aux démarches, aux acteurs, aux modalités de consentement, etc.

2. Le règlement général sur la protection des données

- 5 En préambule de cette partie dédiée au RGPD, nous tenons à distinguer deux grandes natures de dossier à instruire dans une structure de recherche :
 1. Les traitements dits de « *fonctionnement* », *i. e.* tout ce qui concerne l'administration de la recherche : gestion du personnel, achats, missions...
 2. Les traitements dits de « *recherche* », *i. e.* mis en œuvre pour des raisons de recherche, avec une finalité de recherche.

- 6 Dans cet article, nous nous intéresserons uniquement aux traitements de recherche concernant les données personnelles. Le texte Européen (RGPD 2018) évoque à plusieurs endroits des spécificités autorisées pour les traitements à finalité de recherche, à travers l'expression « à des fins de recherche scientifique ». Notre objectif initial était de cerner la portée de ces références afin de proposer en premier lieu aux chercheurs de notre laboratoire des pratiques sécurisées et conformes à la protection des droits fondamentaux des personnes contribuant à la fiabilité du travail scientifique.
- 7 En outre, les auteurs clarifient ici la portée du terme « donnée » utilisé dans cette partie. En effet, il est d'usage courant en linguistique de corpus de distinguer différentes catégories de données selon leur degré d'élaboration :
- primaire *i. e.* expérimentale, observable, mesurable... qui est issue directement du terrain
 - secondaire *i. e.* dérivée de la donnée primaire... qui enrichit, annote, interprète...
 - métadonnée *i. e.* information technique, descriptive, administrative... qui caractérise une autre donnée.
- 8 Qu'elle soit qualifiée de primaire ou de secondaire, l'information véhiculée par une donnée peut concerner une personne physique et est donc susceptible de porter atteinte à sa vie privée. Or il est important de souligner qu'une métadonnée² peut représenter une sensibilité en matière de protection de la vie privée aussi forte que celle de la donnée qu'elle caractérise. Le texte Européen du RGPD ne fait aucune distinction entre ces différentes catégories de données, si ce n'est de redéfinir le caractère personnel de la donnée. Pour cette raison et afin de lever toute ambiguïté et confusion pour le lecteur, le terme « donnée » utilisé dans cette partie englobera toute donnée soumise à la protection du RGPD et ce, quelle que soit son appartenance à l'une des 3 catégories décrites ci-dessus. Elle sera alors qualifiée de « personnelle » ou « à caractère personnel » afin de la différencier de la donnée sans caractère personnel.

2.1. Applicable dès le 25 mai 2018

- 9 Depuis l'entrée en vigueur du RGPD en mai 2018, de nombreux chercheurs s'inquiètent des répercussions induites par la mise en conformité du RGPD sur leurs activités de recherche utilisant des données à caractère personnel. D'autres perçoivent son application complexe et contraignante pour leur travail de chercheur. Ces craintes et appréhensions sont légitimes et compréhensibles car en SHS, il est d'usage courant dans la collecte de corpus oraux et multimodaux d'utiliser des données personnelles à des fins de recherche et comme il est stipulé dans le considérant 159 du RGPD :
- Lorsque des données à caractère personnel sont traitées à des fins de recherche scientifique, le présent règlement devrait également s'appliquer à ce traitement.
- 10 Pourtant, la France a su être précurseur en matière de droit des données personnelles en adoptant le 6 janvier 1978 la loi relative à « l'informatique, aux fichiers et aux libertés », plus connue sous l'appellation de loi « informatique et libertés » (LIL 1978). Toujours en vigueur en France dans une nouvelle rédaction simplifiée, la LIL (LIL 2019) :
- [...] comporte notamment les dispositions relatives aux « marges de manœuvre nationales » autorisées par le Règlement général sur la protection des données (RGPD) que le législateur a choisi d'exercer [...]
- 11 L'entrée en vigueur du RGPD ne peut pas être vue comme une nouveauté majeure si l'on considère que 70 à 80% des exigences posées par la LIL ont été reprises, voire

renforcées pour certaines, et ce même si de nouvelles règles y ont été introduites au niveau européen.

- 12 En revanche, la nouveauté est que d'une part, le RGPD amplifie le pouvoir coercitif de la CNIL³, instituée en 1978 pour veiller au respect de la LIL, et que d'autre part, le RGPD oblige au principe de responsabilité⁴ (« *accountability* »), à l'exigence de documentation et de renversement de la charge de la preuve : c'est au responsable de traitement, y compris au chercheur, de prouver que le traitement de données personnelles est conforme au règlement et documenté. De plus, même si les personnes bénéficiaient déjà de droits et de protections inhérents à la LIL, ces engagements sont renforcés par le RGPD.
- 13 Conscient de la difficulté pour la communauté scientifique de proposer des projets impliquant des traitements de données personnelles, le législateur a prévu un régime dérogatoire pour les activités de recherche scientifique. Abrogées au 1^{er} juin 2019, ces dérogations apparaissent dans le règlement (Dérogations 2019) comme des alternatives activables dans la législation nationale de chacun des États-membres de l'Union Européenne. Il faudra donc se référer à la nouvelle rédaction (LIL 2019) pour s'informer sur les dispositions applicables en France en matière de recherche scientifique.

2.2. Champ d'application de la réglementation

- 14 Le RGPD encadre le traitement et la circulation des données à caractère personnel sur le territoire de l'Union Européenne. C'est-à-dire toute opération, informatisée ou non, effectuée sur des données à caractère personnel, quel que soit le procédé ou le support. L'article 4.2 du RGPD définit la notion de traitement par
[...] la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction
- 15 d'une donnée à caractère personnel. Il apparaît clairement que le traitement ne se limite pas à la simple analyse ou exploitation de la donnée personnelle. Bien au contraire, son champ d'intervention s'applique à l'ensemble du cycle de vie de la donnée.
- 16 À titre informatif concernant les métadonnées, il faut préciser que leur cycle de vie s'étend très fréquemment au-delà de celui des données qu'elles décrivent. En effet, il faut savoir que :
 - elles doivent être créées avant le traitement de collecte/production des données qu'elles décrivent afin d'informer (qui, quand, comment, où...)
 - elles peuvent aussi être conservées même si les données qu'elles décrivent, sont détruites.
- 17 Il apparaît clairement ici que la qualité et la pertinence des métadonnées jouent un rôle primordial sur la visibilité et la réutilisation des données qu'elles caractérisent.
- 18 À propos des données à caractère personnel, le RGPD les définit comme tout élément qui, de manière directe ou indirecte, permet d'identifier un individu : nom, prénom, NIR⁵, date de naissance, courriel nominatif, téléphone, adresse IP, photo, voix, vidéo, empreinte digitale... Il faudra aussi considérer avec la même attention et vigilance, l'identification d'individus par recoupement d'informations issues de diverses sources comme par exemple reconnaître précisément un individu en corrélant les différentes

informations « artisan boulanger de tel âge », « propriétaire de son local commercial situé dans telle commune » et « membre de tel club sportif ».

- 19 Dans la phase initiale d'élaboration d'un projet de recherche, la première question à se poser est de savoir si le RGPD s'applique et si le projet implique un traitement de données personnelles ? Est-ce que des opérations telles que collecter, enregistrer, analyser, consulter, conserver, archiver... sont opérées sur des données à caractère personnel ? Si tel est le cas alors le traitement est soumis au RGPD. Par exemple, le traitement effectué au LPL dans le cadre de « l'étude des sourires lors des transitions thématiques de la conversation » a nécessité, pour atteindre la finalité déclarée, la collecte directe de données sur l'état civil de la personne (nom, prénom, âge, lieu de naissance, adresse mail) et de données audio/vidéo sur lesquelles porteront les analyses linguistiques. De ce fait, le traitement de cette étude entre bien dans le cadre de la réglementation sur la protection des données à caractère personnel et doit faire l'objet d'une inscription au registre du DPD.
- 20 Ensuite, il faut s'assurer que parmi les données personnelles, il n'y a pas de données « sensibles » telles que définies dans la réglementation, c'est-à-dire des données pouvant relever directement ou indirectement de « l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits. » (Article 9 du RGPD).
- 21 Par principe, ces données sont interdites à la collecte et à l'exploitation. Cependant des exceptions existent au regard de la loi et du règlement. L'article 9, paragraphe 2, du RGPD énonce les différentes conditions pour lesquelles cette interdiction ne s'applique pas, et plus particulièrement :
- (a) la personne concernée a donné son consentement explicite au traitement de ces données à caractère personnel pour une ou plusieurs finalités spécifiques, sauf lorsque le droit de l'Union ou le droit de l'État membre prévoit que l'interdiction visée au paragraphe 1 ne peut pas être levée par la personne concernée ;
 - (j) le traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques, conformément à l'article 89, paragraphe 1, sur la base du droit de l'Union ou du droit d'un État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée.
- 22 Lorsque le traitement de données sensibles est autorisé, sa mise en conformité est soumise à un certain nombre de dispositions complémentaires, bien plus dures et protectrices. Ce qui justifie ce renforcement de la loi est que les traitements contenant des données sensibles sont susceptibles de porter atteinte aux droits et libertés des personnes concernées. Leur mise en œuvre peut aboutir à une demande d'autorisation à la CNIL.
- 23 Des dispositions juridiques spécifiques sont prévues par le législateur pour les traitements de données de santé à caractère personnel ayant pour finalité l'évaluation ou l'analyse des pratiques ou des activités de soins et de prévention. Ces traitements sont autorisés s'il y a [1] information et non-opposition de la personne concernée [2],

avis du CEREES⁶ [3], et anonymisation des données à caractère personnel ou garanties suffisantes pour sécuriser et assurer la confidentialité des données.

2.3. Les nouveaux acteurs de contrôle

- 24 Parmi les acteurs majeurs du RGPD, le responsable de traitement (RT) joue un rôle central puisqu'il est en charge du respect de l'application du règlement dans une entreprise, un organisme, une association, etc. Le RGPD le définit dans l'article 4 comme « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement ».
- 25 Un autre acteur important est le sous-traitant (ST). L'article du RGPD le définit comme « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui traite des données à caractère personnel pour le compte du responsable du traitement ». Lié par un contrat ou un acte juridique écrit au RT, un sous-traitant est un prestataire qui intervient pour conserver la donnée, la stocker, la transmettre, la diffuser, etc. Il collabore avec le RT pour garantir le respect de ses diverses obligations (droits des personnes, violations des données, analyses d'impact...) et engage sa responsabilité à l'égal de celle du RT.

2.4. L'inscription au registre des traitements de données personnelles

- 26 Dans chacune de nos tutelles, un Délégué à la Protection des Données (DPD) est désigné pour s'assurer du respect des obligations du RGPD et pour gérer le registre dédié aux traitements des données à caractère personnel. Sur les conseils et avec l'assistance du DPD, il est d'usage courant que ce soit le responsable scientifique du projet qui réalise, à la place du RT, la procédure formalisée⁷ d'inscription de son traitement au registre.
- 27 Il est important de préciser que les traitements de données personnelles doivent être inscrits, avant leur mise en œuvre, au registre du DPD dont dépend le responsable de traitement.
- 28 À la réception d'un formulaire d'inscription d'un traitement, le service à la protection des données (SPD) l'analyse en portant une attention toute particulière à la finalité définie et la nature des données traitées. Le SPD peut soit autoriser la mise en œuvre du traitement et l'inscrire au registre, soit informer le RT de la nécessité de réaliser une étude d'impact sur la vie privée (EIVP) des personnes concernées.
- 29 Hormis l'obligation d'inscrire au registre les traitements mis en œuvre dans son unité, le RT doit s'assurer de leur mise à jour : leur suppression si le traitement est terminé et leur modification si d'éventuels changements sont apportés au traitement.

2.5. Les principes clés du RGPD

- 30 Afin de légitimer la collecte, le traitement et la conservation des données à caractère personnel, le RGPD s'appuie sur 6 principes juridiques⁸ que le RT doit obligatoirement respecter et dont il doit être capable de prouver leur respect.

2.5.1. PRINCIPE DE LIMITATION DES FINALITÉS DU TRAITEMENT DE DONNÉES PERSONNELLES

- 31 Avant toute chose, le chercheur doit déterminer « à quoi » va servir son traitement de données personnelles. Quels sont les objectifs à atteindre ? Quelle est la finalité du traitement ? Tel que mentionné au considérant 33 du RGPD, le législateur accepte le fait qu'il soit parfois difficile pour le scientifique « de cerner entièrement la finalité du traitement » avant sa mise en œuvre :

Souvent, il n'est pas possible de cerner entièrement la finalité du traitement des données à caractère personnel à des fins de recherche scientifique au moment de la collecte des données. [...] Les personnes concernées devraient pouvoir donner leur consentement uniquement pour ce qui est de certains domaines de la recherche ou de certaines parties de projets de recherche, dans la mesure où la finalité visée le permet.

- 32 Ainsi le chercheur doit décrire la finalité de son traitement à travers un bref résumé⁹ de 2 à 3 lignes expliquant l'objectif scientifique visé. Parfois, il peut être nécessaire de définir des sous-finalités pour des projets organisés en plusieurs axes d'investigation. Le cas échéant, le chercheur devra les détailler directement sous la finalité principale du traitement.
- 33 Quoi qu'il en soit, le règlement qualifie la finalité de « déterminée, explicite et légitime ». Cela oblige à ce que la donnée personnelle collectée soit en adéquation directe avec la finalité préalablement définie, clairement comprise et licite. À savoir que si les conditions d'utilisation de la donnée étaient amenées à changer pour d'autres fins incompatibles avec la finalité initiale, le chercheur devrait en avertir son DPD, réinformer les personnes concernées et, le cas échéant, leur refaire signer un consentement. On apprécie ici l'importance de déterminer clairement la finalité d'un traitement avant sa mise en œuvre.
- 34 Aborder ce principe en préalable est particulièrement pertinent lorsque l'on inscrit son traitement au registre du DPD car toutes les questions inhérentes aux autres principes seront interprétées au regard de la finalité, des objectifs visés. Ce principe pose les règles d'utilisation et de réutilisation des données personnelles, intra et post-traitement.

2.5.2. PRINCIPE DE MINIMISATION DES DONNÉES À CARACTÈRE PERSONNEL

- 35 Le principe de minimisation oblige le RT à ne collecter que les données personnelles utiles et pertinentes pour atteindre les objectifs fixés et déclarés. L'idée ici, c'est la proportionnalité, la minimisation des données personnelles : je ne collecte que ce qui est strictement nécessaire au regard de la finalité de mon traitement. Ai-je vraiment besoin de ces données à caractère personnel dans le cadre de mon projet ? Est-ce que je collecte ces données personnelles car je pourrais en avoir besoin un jour ?
- 36 Une fois que le chercheur a déterminé l'ensemble des données personnelles nécessaires à ses besoins, il doit s'assurer que son traitement obéit à des garanties conformes aux droits et libertés des personnes concernées. Ai-je vraiment le droit de collecter ces données à caractère personnel ? Est-ce que les personnes concernées seraient d'accord ? Est-ce que les données sont sensibles ?
- 37 Par principe, la collecte des données sensibles est interdite. Néanmoins si celles-ci sont indispensables pour atteindre les objectifs du traitement à finalités de recherche,

l'article 9 du RGPD mentionne des exceptions autorisant le traitement des données sensibles : [9-a] le consentement explicite [9-e] les données manifestement rendues publiques [9-j] le traitement est nécessaire à des fins de recherche scientifique¹⁰. Pour cette dernière, l'État Français a introduit une condition assujettie à un avis préalable de la CNIL.

- 38 En conséquence, le chercheur doit documenter et justifier par tout moyen de l'intérêt de la donnée personnelle collectée. On est vraiment dans un processus de renversement de la charge de la preuve, c'est au chercheur de prouver qu'il est en conformité avec la loi, qu'il documente et justifie par tout moyen qu'il a besoin de cette information-là pour atteindre tel objectif scientifique.

2.5.3. PRINCIPE DE LIMITATION DE LA CONSERVATION DES DONNÉES À CARACTÈRE PERSONNEL

- 39 Un autre point discuté dans la réglementation, et pourtant déjà prévu dans la LIL, concerne le principe de la durée de conservation des données à caractère personnel qui doit être définie par le RT au préalable de leur traitement.
- 40 Le RGPD¹¹ précise que les données à caractère personnel doivent être
[...] conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées ; [...]
- 41 Ce principe de conservation soulève trois aspects importants qui interrogent sur la durée, les moyens à employer et le devenir des données personnelles à l'issue de cette étape.

2.5.3.1. La détermination de la durée de conservation des données à caractère personnel

- 42 Premièrement, la durée de conservation détermine la période d'utilisation des données à caractère personnel c.-à-d. la période nécessaire pour atteindre la finalité du traitement. Généralement, l'accès aux données personnelles est restreint aux seuls membres du projet scientifique durant cette période. Les durées peuvent être très variables selon les projets. Par exemple, un projet de recherche qui propose le suivi longitudinal d'enfants de leur naissance jusqu'à leur majorité, aura une durée de conservation très longue. À la différence d'un corpus collecté dans le cadre d'une thèse, celle-ci ne sera que de quelques années au-delà desquelles les données personnelles permettant l'identification (in)directe des personnes devront être « anonymisées »¹² ; le corpus ainsi « anonymisé » pourra être préservé au-delà de cette période pour une réutilisation ou être archivé. L'idée est de conserver les données à caractère personnel tant qu'elles sont nécessaires pour atteindre l'objectif fixé.
- 43 Actuellement, la durée peut être déterminée ou déterminable. L'intérêt de la rendre déterminable est que si le chercheur est amené à prolonger la durée de conservation initialement prévue, il n'est pas tenu d'en informer les personnes si celle-ci respecte la valeur initiale et la règle de gestion¹³ associée, mentionnées dans la notice d'informations. En effet, la durée de conservation des données personnelle doit être mentionnée dans la notice d'informations à remettre aux personnes concernées.

- 44 Quoi qu'il en soit, l'important est de définir cette durée en considérant d'éventuelles obligations concernant certaines données, de se l'approprier et de la justifier dans la documentation.

2.5.3.2. Les moyens à mettre en œuvre sur la période de conservation

- 45 Le second point concerne les moyens, matériels et techniques, à mettre en œuvre pour assurer l'effectivité de la période de conservation. Qui reçoit une alerte prévenant du dépassement de la période de conservation ? Y a-t-il un outil automatique qui gère la datation des données à caractère personnel, leur effacement... ? Comment gérer la particularité des durées déterminables ? Ces questions importantes doivent être abordées en concertation entre le chercheur et le RT.
- 46 Comme rappelé dans (Ghio *et al.* 2020), et tel que proposé par le RGPD et les guides de bonnes pratiques comme le guide pour la recherche en SHS du CNRS (CNRS-InSHS 2019), nous préconisons fortement aux chercheurs de notre laboratoire de pseudonymiser¹⁴ les données personnelles de leurs traitements dès la collecte. Rappelons ici que le RGPD ne s'applique pas sur des données personnelles anonymes ou anonymisées c.-à-d. sur des données personnelles à partir desquelles la personne concernée n'est pas ou plus identifiable.
- 47 Parmi les différentes techniques de pseudonymisation (chiffrement, généralisation, agrégation, bruitage...), la substitution consiste à remplacer dans un corpus chaque ensemble de données permettant d'identifier (in)directement un individu par une clé d'identification arbitraire. En pratique, les données telles que nom, prénom, mail, téléphone... sont retirées du jeu initial de données personnelles pour être isolées dans une table de correspondance. Pour chacune des entrées de la table de correspondance, une clé d'identification est ajoutée sous une forme normalisée¹⁵, respectant des règles de nommage et de codage établies par le chercheur. Cette clé doit aussi être reportée dans le jeu initial de données personnelles (nettoyé des données identifiantes) afin de garantir le lien entre les données personnelles pseudonymisées et la personne concernée. Ici l'utilisation du terme de pseudonymisation est en parfaite adéquation avec la définition faite dans l'article 4.5 du RGPD où la table de correspondance constitue « ces informations supplémentaires » permettant d'attribuer à « une personne concernée précise » des données personnelles.
- 48 En pratique, la table de correspondance permet de ré-identifier les personnes des données pseudonymisées afin de faire valoir leurs droits, de les informer de modifications majeures dans le traitement les concernant, de les avertir en cas de violation/perde de leurs données personnelles... Pour garantir un niveau de protection optimal, elle devra être conservée séparément des autres données (personnelles, sonores, vidéos...) et faire l'objet de mesures de sécurités et de confidentialités, appropriées à la sensibilité des données. À la fin de la période de conservation, il suffira alors de détruire la table de correspondance (le cas échéant, les autorisations ou consentements obtenus auprès des personnes concernées) pour permettre la réutilisation et préservation du corpus « anonymisé ».
- 49 Cette recommandation de bonnes pratiques s'inscrit dans une démarche visant d'une part à renforcer la sécurité des données personnelles et d'autre part, à gérer les traitements arrivant au terme de leur durée de conservation.

2.5.3.3. Le devenir des données à caractère personnel à l'issue de la période de conservation

- 50 À l'issue du traitement, le RGPD oblige à ce que les données personnelles soient détruites, anonymisées ou archivées. Néanmoins l'article 5.1 du RGPD précise que « [...] les données à caractère personnel peuvent être conservées pour des durées plus longues dans la mesure où elles seront traitées exclusivement à des fins de recherche scientifique [...] ». De surcroît, en tant qu'unité publique de recherche, les données arrivant à échéance de leur période de conservation légitime, doivent être archivées « conformément aux règles applicables en matière d'archives publiques et d'archives privées ».
- 51 Toujours est-il qu'il incombe au responsable du traitement de notifier au service dédié à la protection des données dont il dépend, l'ensemble des traitements arrivant au terme de leur période de conservation.
- 52 Au LPL, nous préconisons à nos chercheurs de déposer leurs données « anonymisées » dans des entrepôts nationaux tels que Nakala¹⁶ pour les SHS ou Ortolang¹⁷ pour les sciences du langage. Garantissant un haut niveau de sécurité et de disponibilité, ces plateformes d'archivage proposent à l'ensemble de la communauté scientifique des services permettant le dépôt, la documentation, la diffusion et la valorisation de leurs ressources numériques.
- 53 Néanmoins, du fait de la diversité des recherches en SHS et de l'interdisciplinarité des laboratoires, il existe des données personnelles qui ne sont ni partageables, ni anonymisables en raison de leur nature « sensible » ou même, des risques de perdre leur utilité scientifique¹⁸. Comment gérer cet état de fait lorsque les chercheurs reçoivent l'injonction de travailler en Open Data et de publier en Open Science ? Pour répondre à cela, il revient au chercheur de définir les modalités de partage, de conservation et d'archivage des données collectées/produites en définissant leurs conditions d'accès et de réutilisation, tout en appliquant le paradigme de l'Open Data « Aussi ouvert que possible, aussi fermé que nécessaire ».

2.5.4. PRINCIPES D'INTÉGRITÉ ET DE CONFIDENTIALITÉ

- 54 Le responsable de traitement doit préserver la sécurité et la confidentialité des données à caractère personnel traitées. À ce titre, il doit prendre toutes les mesures nécessaires, adaptées à la nature des données et aux risques liés au traitement, pour interdire toutes intrusions hostiles et empêcher l'altération ou la perte des données. Il doit s'assurer que la confidentialité des données personnelles est respectée et que seules les personnes habilitées accèdent à celles-ci.
- 55 L'analyse des risques, réalisée sur la base (i) de l'identification de leurs sources qu'elles soient d'origine humaine ou non (ii) et de l'évaluation de leurs impacts sur les droits et libertés des personnes concernées, facilite grandement la mise en œuvre des moyens techniques et organisationnels garantissant l'intégrité et la confidentialité des données personnelles traitées.
- 56 Sans négliger l'importance du risque lié au facteur humain, les postes de travail informatique constituent une des principales failles pour les intrusions illégales et les tentatives de piratage. Il est donc essentiel de les sécuriser pour empêcher les tentatives d'accès malveillants, la perte ou altération de données, les virus, etc.

57 Voici quelques recommandations de sécurité :

- verrouillage automatique de session
- arrêt complet de l'ordinateur en cas d'absence de surveillance
- antivirus régulièrement mis à jour
- pare-feu pour le contrôle des entrées/sorties du réseau
- moyens de chiffrement (clé USB, disque amovible, portable)
- protocoles garantissant la confidentialité et l'authentification (SFTP, HTTPS)
- fonction de hachage pour assurer l'intégrité des données
- signatures numériques pour vérifier l'origine et l'authenticité de l'information
- chiffrement pour garantir la confidentialité des données et des messages
- ...

2.5.5. PRINCIPES DE LICÉITÉ, DE LOYAUTÉ ET DE TRANSPARENCE

58 Il est essentiel de définir la base juridique sur laquelle repose un traitement de données à caractère personnel. L'identification du fondement juridique assure que le traitement est conforme à la loi et oblige le RT à un devoir de documentation attestant de la conformité des traitements dont il est responsable.

59 Le RGPD précise qu'il faut qu'au moins une des conditions parmi les 6 proposées soit remplie, pour que le traitement soit licite : (extrait de l'Article 6.1 du RGPD)

- a) la personne concernée a consenti au traitement de ses données à caractère personnel pour une ou plusieurs finalités spécifiques ;
- b) le traitement est nécessaire à l'exécution d'un contrat auquel la personne concernée est partie ou à l'exécution de mesures précontractuelles prises à la demande de celle-ci ;
- c) le traitement est nécessaire au respect d'une obligation légale à laquelle le responsable du traitement est soumis ;
- d) le traitement est nécessaire à la sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique ;
- e) le traitement est nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement ;
- f) le traitement est nécessaire aux fins des intérêts légitimes poursuivis par le responsable du traitement ou par un tiers, à moins que ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée qui exigent une protection des données à caractère personnel, notamment lorsque la personne concernée est un enfant.

60 Dans le domaine des sciences humaines et sociales, le fondement d'un traitement de recherche repose le plus souvent sur (6.1.a) le consentement, (6.1.e) l'exécution d'une mission d'intérêt public ou (6.1.f) les intérêts légitimes.

61 Il faut noter que le consentement n'est pas uniquement un fondement juridique. Il apparaît aussi comme une des exceptions pour collecter des données sensibles. À titre informatif, au LPL si la loi n'exige pas le recueil du consentement pour un traitement de données personnelles, nous demandons à nos chercheurs d'obtenir l'autorisation d'enregistrement de la voix et/ou de l'image auprès des personnes susceptibles d'être filmées, photographiées ou enregistrées vocalement.

62 Une fois la base juridique d'un traitement identifiée, elle doit être obligatoirement communiquée à la personne concernée, complétée par les informations mentionnées dans le chapitre III du RGPD¹⁹ « Droits de la personne concernée ». La loi impose que

l'information à la personne concernée doit être faite d'une façon « concise, transparente, compréhensible et aisément accessible ». Il est fortement recommandé qu'elle soit fournie par écrit, voire par voie électronique lorsque la situation est appropriée. De plus, elle doit être faite préalablement à la collecte des données (voir les articles 13 et 14 du RGPD) et devra être adaptée en fonction du fondement juridique du traitement :

- sur les intérêts légitimes du RT, l'information doit spécifier ces intérêts légitimes ;
 - sur le consentement, l'information doit mentionner à la personne concernée le droit de retirer son consentement à tout moment, ainsi que le droit à la portabilité de ses données ;
 - sur l'exécution d'une mission d'intérêt public ou sur l'intérêt légitime du RT, l'information doit mentionner à la personne concernée le droit d'opposition.
- 63 Un point important à souligner concerne la manière dont l'information doit être divulguée. Informer ne veut pas dire obtenir le consentement. Il n'est donc pas nécessaire de conserver la manifestation expresse de la volonté de la personne informée. Ce qu'il faut, c'est conserver la manière dont l'information a été faite auprès des personnes, être capable de prouver que les personnes ont bien été informées. À une information orale, il faudra privilégier l'affichage, le mailing ou la remise directe d'un document d'informations.
- 64 Lorsque les données personnelles sont collectées directement auprès de la personne, la réglementation RGPD précise que les personnes doivent être informées au moment où les données sont obtenues. Par contre, lorsque celles-ci ne sont pas collectées directement auprès de la personne, le RT doit fournir les informations légales à la personne dans un délai d'un mois à compter du début du traitement.
- 65 De plus, le RGPD impose que l'information soit claire et compréhensible, et comporte plusieurs mentions d'information (coordonnées du RT/DPD, finalités du traitement, destinataires des données personnelles, durée de conservation, droits^{20 21}applicables...).

2.5.6. PRINCIPE D'EXACTITUDE

- 66 Ce dernier principe impose que les données personnelles soient exactes et, le cas échéant, tenues à jour (actualisées, corrigées, effacées...). Cela signifie qu'il faut expliquer, clairement et formellement, à la CNIL que les données personnelles utilisées dans le traitement, sont exactes et mises à jour. Par exemple, que toutes les adresses électroniques des personnes utilisées dans un traitement sont valides et correctes.
- 67 Ce principe est adapté au traitement de fonctionnement comme la gestion du personnel ou un annuaire, mais pas pour les traitements de recherche où les données personnelles collectées n'ont pas vocation à être mises à jour après leur collecte ou pour une utilisation ultérieure.

2.6. Pour conclure cette partie...

- 68 Hormis le cadre juridique, fondamental et essentiel, l'articulation de RGPD avec la recherche scientifique conduit le chercheur à s'interroger sur son travail en matière de gouvernance, de protection et d'usages des données traitées. Cette réflexion épistémologique, propre à chaque discipline, devrait contribuer à l'amélioration des pratiques dans le respect des normes éthiques et scientifiques. Concernant le cadre

législatif, ces réflexions pourront être complétées par l'attention portée à la réglementation de la recherche impliquant la personne humaine.

3. La loi relative aux recherches impliquant la personne humaine dite loi Jardé

- 69 Dans cette partie consacrée à la loi Jardé, nous présenterons les éléments de sa genèse qui selon nous, sont à l'origine des difficultés rencontrées dans sa lecture et son interprétation au regard des recherches menées en sciences humaines et sociales et en particulier pour la collecte de données en linguistique. Nous exposerons ensuite le travail de réflexion mené par le groupe éthique du LPL et l'évolution de notre compréhension de la loi au fil de la parution des différents arrêtés et décrets relatifs à celle-ci. Enfin, nous en donnerons une grille de lecture issue de nos réflexions fondées sur les différents textes de loi et documents de nos tutelles.

3.1. Quid Act Jardé ?

- 70 Pour bien comprendre pourquoi la publication et la mise en application de la loi Jardé ont tellement pu émouvoir les différentes communautés de chercheurs, tant en SHS que dans d'autres disciplines, il convient de se pencher sur sa genèse.
- 71 La loi Jardé, relative aux recherches impliquant la personne humaine (RIPH) a été votée en 2012 et est entrée en vigueur en 2016, suite à la parution de son décret d'application. Cette loi, comme toutes finalement, est née dans un contexte socio-politique national et européen animé de divers événements et changements – l'affaire du Mediator® ou la transposition de la directive Européenne aux essais cliniques de médicaments à usage humains, par exemple, constituent deux événements notables dans son histoire.
- 72 Au-delà de ce contexte, cette loi est présentée par son auteur lui-même, Olivier Jardé, comme une nécessaire révision de la loi Huriot-Sérusclat, qui avait déjà été modifiée par la loi de Santé Publique de 2004, et qui depuis 1988 constituait le cadre réglementaire pour la protection des personnes qui se prêtaient à des recherches biomédicales. La loi Jardé s'inscrit donc, et ce n'est pas anodin, dans la lignée de lois précédentes, qui posaient le cadre des recherches interventionnelles liées aux essais cliniques médicamenteux, autorisant et encadrant la recherche sur l'homme. Au sujet de la loi Huriot-Sérusclat, Lemaire et Matei (2012) précisent :
- ... [qu'elle] avait été conçue pour encadrer la recherche sur le médicament.
- 73 La loi Huriot-Sérusclat de 1988, modifiée par la loi de Santé Publique de 2004 est donc la première loi de protection de la personne humaine se prêtant à des recherches biomédicales, dont l'encadrement est assuré par des responsabilités réparties entre promoteur, investigateur et CCPPRB (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale). Rédigée par des pharmacologues (Jaillon & Demarez 2008), elle est clairement inscrite dans le domaine médical. Elle présentait 2 limites essentielles qui concernaient (i) la distinction entre les recherches avec et sans bénéfice direct et (ii) les recherches avec les personnes « hors d'état de consentir » ; Ces deux limites rendaient de fait certaines recherches impossibles à conduire.
- 74 La loi de Santé Publique de 2004 a constitué une première évolution de la loi Huriot en proposant des modifications sur les notions de « bénéfice », en modifiant les missions

des CCPPRB devenus les CPP (Comités de Protection des Personnes) et en retravaillant les modalités de consentement. Cette première révision est, au même titre que la loi Huriet, inscrite dans le domaine de la recherche médicale avec notamment un cadre spécifique pour les « recherches portant sur les soins courants ».

- 75 C'est alors également dans le domaine de la santé que s'inscrit la deuxième révision de la loi Huriet, proposée par O.Jardé. Ses propositions émanent d'une volonté de renforcement de la protection de la personne mais souhaitent surtout répondre à trois écueils principaux qui ont résisté aux précédentes révisions :

- La restriction de la réglementation à la recherche sur les médicaments, c'est-à-dire aux recherches interventionnelles
- Les dispositions de consentement inadaptées (pourtant révisées en 2004) qui rendaient impossibles certaines recherches
- L'absence de validation éthique pour les recherches hors champs qui rendait la publication des résultats de ces études impossibles dans les revues internationales

- 76 La loi Jardé de 2012 propose ainsi un cadre réglementaire pour toutes les recherches impliquant la personne humaine avec des modifications majeures :

- L'intégration des recherches non interventionnelles dans le champ d'application de la loi (de laquelle découle la classification des RIPH en trois catégories, définies dans la section suivante)
- Une réflexion basée sur la notion de risque (et non plus sur celle de bénéfice)
- Une obligation de demande d'autorisation au CPP (pour toutes les catégories de recherche)
- La création de la Commission Nationale des Recherches Impliquant la Personne Humaine (CNRIPIH)

- 77 On se rendra aisément compte en examinant précisément les articles de cette loi de 2012, et en particulier le décret d'application du 16 novembre 2016 ainsi que les arrêtés de décembre 2016 et mai 2017, que tout comme les lois Huriet-Sérusclat et de Santé Publique, la loi Jardé est également une loi qui s'adresse prioritairement aux acteurs la recherche clinique médicale.

- 78 En témoigne le lexique utilisé, non seulement dans les textes réglementaires suscités (diagnostic, traitement, surveillance Art. 1121-1), mais également dans les articles de revue dont la loi a pu faire l'objet ; par exemple, au sujet des recherches impliquant la personne humaine, Rat et collègues (2017) précisent :

Dans le cadre de ces recherches, les actes pratiqués sur les personnes participantes (traitement, décision de traiter ou non, moyen de surveillance et de suivi médical) sont définis par un protocole de recherche qui doit être respecté par les médecins investigateurs.

- 79 Ou encore concernant les catégories de recherches :

Une notion clé, véritable pivot de la loi et critère majeur de qualification des recherches, est l'intervention que la recherche ajoute au soin ou à la prise en charge habituelle. (Matei & Lemaire 2013)

- 80 Le champ lexical utilisé n'a cependant pas permis de cantonner la loi Jardé au domaine de la santé, puisque portant dans son titre la mention « personne humaine », son champ d'application s'est vu étendu à l'ensemble des recherches impliquant l'humain.

3.2. Quae Act Jardé ?

- 81 La loi Jardé n° 2012-300 du 5 mars 2012 modifie l'article L. 1121-1 du code de la santé publique en précisant que les recherches organisées et pratiquées sur l'être humain en vue du développement des connaissances biologiques ou médicales sont autorisées dans les conditions prévues par la loi et sont désignées par les termes « Recherches Impliquant la Personne Humaine » (RIPH).
- 82 Comme précisé plus haut, la loi Jardé est organisée à partir de la notion de « risque » dont découlent 3 catégories de recherches :
- 1° Les Recherches Interventionnelles (RI) qui comportent une intervention sur la personne non justifiée par sa prise en charge habituelle
 - 2° Les Recherches Interventionnelles qui ne comportent que des Risques et des Contraintes Minimales (RICRM), dont la liste est fixée par arrêté du ministère de la santé, après avis du directeur de l'Agence Nationale de Sécurité du Médicament et des produits de Santé (ANSM)
 - 3° Les recherches Non Interventionnelles (RNI) qui ne comportent aucun risque ni contrainte et dans lesquelles tous les actes sont pratiqués et les produits utilisés de manière habituelle.
- 83 La liste des recherches mentionnées au 2° et au 3° de l'article L. 1121-1 sont par ailleurs fixées par l'arrêté du 12 avril 2018.
- 84 Ainsi, la loi Jardé propose un cadre réglementaire qui préserve la personne se prêtant à la recherche et catégorise les recherches en fonction du risque encouru par la personne. Ceci constitue la première grande modification de la loi de Santé Publique puisqu'à présent, les recherches qui ne présentent aucun risque, *i. e.* les recherches non interventionnelles, entrent dans le champ d'application de la loi. De ce fait, et c'est aussi la deuxième modification majeure, les recherches des 3 catégories sont soumises à l'autorisation d'un CPP (Art. L. 1123-6 et L. 1123-7), la constitution du dossier de demande et la procédure d'évaluation étant d'autant plus allégée que le risque encouru est faible.
- 85 Concernant les CPP, des modifications sont également apportées (Lemaire 2019) ; elles concernent l'attribution des dossiers à évaluer qui se font à présent par tirage au sort (voir Rat *et al.* 2017 pour le parcours de constitution et de dépôt des dossiers au CPP). Les missions des CPP sont également augmentées puisqu'à l'évaluation éthique et scientifique des dossiers s'ajoute la mission de vérification de la protection des données personnelles des participants. Enfin, la coordination des CPP est confiée à la Commission Nationale des Recherches Impliquant la Personne Humaine (CNRI), dernière grande modification apportée par la loi de 2012.
- 86 Les modifications sont donc nombreuses, initialement nées d'une volonté de simplification du cadre réglementaire relatif aux recherches impliquant la personne humaine. De nombreuses modifications qui ont fait l'effet d'un raz de marée sur la communauté des chercheurs en sciences humaines et sociales, puisque de fait, nos recherches impliquent la personne humaine.

3.3. Regards sur la loi Jardé et les SHS

- 87 Dès lors, comment situer nos recherches en sciences humaines et sociales ? Comment mener nos travaux dans le respect de la réglementation en vigueur qui a priori nous

concerne puisque notre objet d'étude est l'humain. Et tout comme les sciences du vivant, les sciences humaines recouvrent plusieurs disciplines, la question se pose alors de savoir si toutes sont concernées au même titre.

- 88 Depuis 2017, la direction du LPL a souhaité mettre en place une cellule de réflexion autour de cette nouvelle réglementation. Ce groupe s'est constitué de plusieurs membres du laboratoire, représentants de la diversité des champs de recherche et des méthodes utilisées dans cette unité réunissant des membres de l'INSHS, de l'INS2I et de l'INSB. Nous avons entrepris d'examiner et comprendre les termes de la loi Jardé, ce dans le but de pouvoir l'appliquer, tout en tenant compte de la diversité des recherches menées au LPL. Notre groupe s'est donc penché sur les textes officiels (loi, décrets, arrêtés) ainsi que sur les documents de support préparés par nos institutions²².
- 89 À partir de ces documents, nous avons, pas à pas, questionné les termes de la loi pour répondre à la première interrogation : nos recherches, menées au LPL, sont-elles concernées par ce nouveau cadre réglementaire ? Une première interrogation à laquelle se sont ajoutées les nombreuses suivantes : Si oui, quel sera l'impact sur nos pratiques habituelles ? Comment nos travaux s'inscrivent-ils dans ce champ législatif ? Comment déterminer si telle ou telle étude rentre dans telle ou telle catégorie ?...
- 90 Pour répondre à ces questions, nous avons dans un premier temps tâché de définir les RIPH : « recherches organisées et pratiquées sur l'être humain en vue d'améliorer les connaissances biologiques ou médicales ».
- 91 Nous avons ainsi pu déterminer, à l'appui des documents que :
- L'expression « recherches organisées » implique le recrutement de sujets et l'existence d'un protocole.
 - L'expression « pratiquées sur l'être humain » implique la réalisation d'actes pouvant porter atteinte à l'intégrité physique ou psychique de la personne.
 - L'expression « en vue du développement des connaissances biologiques ou médicales » implique la connaissance de la vie, au sens de l'étude du développement de la physiologie, du comportement... sans qu'il y ait de frontière entre les disciplines.
- 92 À ce stade, les études menées en linguistique, en psycholinguistique et en neurolinguistique dans notre laboratoire semblaient répondre à la définition des RIPH.
- 93 Cette dernière expression « en vue du développement des connaissances biologiques ou médicales » soulève la question de la finalité de la recherche qui semble alors déterminante pour situer nos travaux dans ou hors du champ d'application de la loi Jardé. Par ailleurs, la notice explicative de l'INSB du CNRS insiste sur les données de la recherche, en particulier le type de données et les modalités de recueil. Nous avons longuement débattu afin de déterminer laquelle de ces deux entrées « finalité » ou « données » nous permettrait d'avancer sur notre première interrogation, alors encore sans réponse. En filigrane, le sentiment du groupe était que cette loi concernait surtout les recherches cliniques médicales.
- 94 Nous avons pris connaissance des arrêtés du 12 avril 2018, fixant la liste des recherches mentionnées au 2° et au 3° de l'article L. 1121-1 (c'est-à-dire les recherches de catégorie 2 et 3, respectivement les RIRCM et les RNI). Encore une fois, le lexique utilisé est ancré dans la médecine. Pour autant, le 7° point de l'arrêté fixant la liste des recherches de catégorie 2, portant sur les
- Techniques de recueil et de collecte de données au moyen de capteurs ou de méthodes d'imagerie

- 95 nous interpelle puisque sont couramment utilisés au LPL les dispositifs de recueil de données biodermales, électro-encéphalographiques, électro-magnéto-articulographiques.
- 96 Dans ce même Arrêté, la liste des actes et procédures pouvant être réalisés dans le cadre d'une recherche de catégorie 3 (RNI) comprend au 4^e point le
Recueil par capteurs extra-corporels non invasifs, notamment [...] enregistrement par [...] électroencéphalogramme [...] capteurs de mouvement [...].
- 97 Puis au 5^e point les
Enregistrements audio, vidéos, photographiques hors imagerie médicale.
- 98 Et enfin en 8^e point les
Entretiens, observations, tests et questionnaires [...].
- 99 Après l'examen de cet Arrêté, nous admettons que toutes les recherches menées au LPL relèvent de la loi Jardé et que toutes les recherches devront faire l'objet d'une demande d'autorisation au CPP. Nous mesurons l'ampleur du changement à venir dans nos pratiques (constitution de dossier, soumission et délai de réponse), puisque jusque-là, un petit nombre seulement de chercheurs avait eu à demander un avis au CPP.
- 100 Forts de cette conclusion, nous abordons la question des catégories de recherche dans le but de proposer un guide qui nous permettrait (i) de situer nos recherches dans l'une des trois catégories de la loi et (ii) d'identifier le parcours des demandes d'autorisations à suivre. Ce travail nous a conduits à répartir nos recherches essentiellement entre les catégories 2 (RICRM) et 3 (RNI), à l'appui de l'Arrêté d'avril 2018 qui précisait la nature des données et les conditions de recueils. Notre répartition était donc basée sur les méthodologies de la recherche.
- 101 Ce guide a été restitué aux membres du laboratoire et nous avons poursuivi notre travail de « veille législative » au cours de nos réunions mensuelles.

3.4. Parce que nous le valons bien...

- 102 Soucieux de suivre l'évolution de la réglementation en vigueur, nous assistons à une conférence donnée par O. Jardé et F. Lemaire invités par l'université Paris Nanterre²³ ; l'occasion pour nous de questionner directement les auteurs de la loi et des décrets relatifs à son application. C'est à cette occasion que nous apprenons qu'au-delà de notre groupe de réflexion, une réaction s'est faite plus musclée : une grande industrie du cosmétique, inquiète de voir son activité de recherche paralysée, a saisi le conseil d'État pour que soient définies comme n'étant pas des recherches impliquant la personne humaine les recherches
qui bien qu'organisées et pratiquées sur des personnes saines ou malades, n'ont pas pour finalités celles mentionnées au I, et qui visent :
a) Pour les produits cosmétiques, conformément à leur définition mentionnée à l'article L. 5131-1, à évaluer leur capacité à nettoyer, parfumer, modifier l'aspect, protéger, maintenir en bon état le corps humain ou corriger les odeurs corporelles
- 103 Ceci sera en effet mentionné dans le décret de mai 2017 modifiant certaines dispositions réglementaires relatives aux recherches impliquant la personne humaine.
- 104 Nous pourrions également lire, dans ce même décret :
Ne sont pas des recherches impliquant la personne humaine au sens du présent titre les recherches qui, bien qu'organisées et pratiquées sur des personnes saines

ou malades, n'ont pas pour finalités celles mentionnées au I et qui visent : [...]

d) À réaliser des expérimentations en sciences humaines et sociales dans le domaine de la santé (Art. R. 1121-1.-II-1°d)

- 105 La parution du décret de 2017 a clairement modifié notre lecture de la loi Jardé. En reprenant notre réflexion initiale, nous avons pu proposer une nouvelle ventilation des recherches menées au laboratoire qui fait apparaître les recherches en fonction des méthodes de recueil des données et selon qu'elles entrent ou pas dans le champ d'application de la loi. Conformément au décret de 2017, toutes nos recherches en Sciences humaines et sociales, dont la finalité concerne cette discipline n'entrent pas dans le champ d'application de la loi, qu'elles impliquent des personnes saines ou malades.
- 106 Lors de cette conférence à Nanterre, les auteurs O. Jardé et F. Lemaire ont clairement expliqué à l'assemblée que ce qui déterminait l'appartenance de la recherche au champ d'application de la loi était sa finalité.
- 107 Cependant, les recherches couramment menées au LPL, et en particulier les recherches impliquant des questions cliniques et/ou des populations de patients se situent, selon les membres du groupe, à la frontière du champ d'application de la loi. Nous en avons conclu, qu'en plus de la finalité, il conviendra le cas échéant de s'interroger sur la méthode de recueil des données et la population visée. En effet, au-delà de la finalité, ce seront ces deux points qui pourront être déterminants pour conclure. Prenons trois exemples de travaux menés au LPL :
- 108 [1] Caractéristiques de l'humour réussi (Priego-Valverde 2018) : cette étude est menée afin de clarifier le concept d'« humour réussi ». Elle propose une analyse de séquences humoristiques dans des interactions en face-à-face. Les données sont constituées de l'enregistrement audio-visuel de 3 interactions. Cette étude, dont la finalité concerne une meilleure connaissance de l'humour dans les interactions au travers d'une analyse linguistique est sans conteste positionnée dans le champ des SHS. Les participants et les méthodes de recueil des données le sont également. Le décret de 2017 situe donc cette étude en dehors du champ d'application de la loi Jardé.
- 109 [2] Contribution des paramètres intonatifs et temporels à la réalisation de certaines variations mélodiques (Petrone, Lalain & Mattei) L'objectif de cette étude est de mieux comprendre la contribution relative des paramètres intonatifs et temporels dans la réalisation de certains patrons prosodiques. L'étude est menée auprès de sujets sains et de patients porteurs d'une paralysie laryngée. Les données sont constituées d'enregistrements audio de productions de phrases. On pourrait être plus hésitants que dans le cas précédent pour qualifier cette étude, notamment parce qu'elle implique des patients ; pourtant, parce que sa finalité concerne l'amélioration des connaissances concernant la réalisation de contrastes prosodiques, et que l'enregistrement audio est une méthode de recueil des données classiquement utilisée, cette étude s'inscrit bien dans le champ des SHS. Conformément au décret de 2017, cette étude se situe en dehors du champ d'application de la loi Jardé.
- 110 [3] Production de la parole et dysarthrie parkinsonienne (Sarr *et al.* 2009) Cette étude est conduite auprès de patients parkinsoniens afin de déterminer si la production de la parole de ces patients est améliorée par un traitement neurochirurgical particulier, la stimulation cérébrale profonde. Des analyses des valeurs de pression intra-orale permettent de caractériser cet effet potentiel sur le contrôle pneumo-phonatoire. L'étude est menée par des chercheurs de SHS, auprès de patients ; la méthode de recueil

de données est non invasive (enregistrements aérodynamiques). Les données sont recueillies dans deux conditions : lorsque les patients ont la stimulation cérébrale en marche, ou à l'arrêt. Pour cette étude, la finalité concerne la production de la parole ; elle entre donc bien dans le champ des SHS. Cependant, les conditions de recueil de données impliquent une modification du soin courant. Dans cet exemple donc, bien qu'elle consistait à « réaliser des expérimentations en sciences humaines et sociales dans le domaine de la santé », cette recherche serait qualifiée aujourd'hui de « recherche[s] Interventionnelle[s] (RI) qui comporte[nt] une intervention sur la personne non justifiée par sa prise en charge habituelle ».

- 111 Les trois exemples ci-dessus, illustrent 3 catégories de cas pour lesquels nous nous sommes interrogés au LPL. Un premier cas, pour lequel le décret de 2017 a levé tout doute concernant la qualification de ce type de recherche.
- 112 Un second cas, pour lequel, tout en nous appuyant sur le décret de 2017, nous avons porté notre vigilance sur la population et le mode de recueil de données.
- 113 Enfin le dernier cas montre bien que la question de la finalité ne peut être en réalité le seul et unique argument pour qualifier la recherche et qu'il convient de considérer, le cas échéant, précisément le risque que la personne encourt.
- 114 Au-delà de ces cas de figure plutôt simples, il peut arriver que le doute subsiste dans la qualification de la recherche, en particulier, comme nous l'avons illustré plus haut, lorsque la recherche présente des questions en lien avec la santé. Lorsque c'est le cas, nous l'avons précisé, c'est l'analyse de la finalité, de la population et de la méthodologie de recueil des données qui permettra de lever le doute et de conclure sur l'appartenance ou non de la recherche au champ d'application de la loi. Lorsqu'une recherche en SHS relève de la loi Jardé, en général, elle entre essentiellement dans la catégorie 3 de la loi (les RNI) et la liste des actes et procédures pouvant être réalisés dans ce cadre est fixée par l'arrêté de 2018. Cette liste comprend notamment le recueil de données par capteurs extra-corporels non invasifs par exemple : EEG, capteurs de mouvement, enregistrements audio, vidéo, entretiens et questionnaires. Que la recherche relève de la catégorie 3 est un moindre mal puisque dans ce cas, l'avis au CPP sera nécessaire, mais la procédure pour les RNI est une procédure allégée qui permet d'obtenir un n° IRB et ne nécessite pas d'assurance.
- 115 Finalement, dans la grande majorité des cas, nos recherches en SHS sont des recherches n'impliquant pas la personne humaine au sens de la loi Jardé et sont donc hors champs, ou bien sont bien des RIPH et entrent généralement dans la catégorie 3 des RNI pour lesquelles le parcours d'autorisation est considérablement allégé.
- 116 Dans les deux cas, les responsables des recherches peuvent demander l'avis d'un comité éthique et de recherche scientifique ; pour les recherches n'impliquant pas la personne humaine, cette démarche permettra de valider la recherche d'un point de vue éthique et scientifique et permettra l'obtention d'un numéro qui pourra être fourni lors de la soumission des résultats à une revue.
- 117 Pour les recherches dont les responsables sont dans l'incertitude, le comité éthique pourra donner un avis consultatif et préciser si oui ou non la recherche doit faire l'objet d'une demande de CPP.
- 118 Pour les recherches qui entrent dans le champ d'application de la loi Jardé, la demande d'un CPP permettra d'assurer la validité de la recherche d'un point de vue légal, éthique et scientifique.

- 119 Nous soulignons ici la différence entre le cadre éthique et le cadre légal : demander une validation éthique et scientifique de sa recherche devient indispensable pour répondre aux exigences éditoriales des revues. Mais une recherche éthique n'entre pas nécessairement dans le champ d'application de la loi Jardé et ne nécessite pas forcément un avis du CPP. C'est précisément aux comités éthique et de recherche scientifique des universités qu'est dévolue cette mission.

3.5. L'impact de la loi Jardé sur les recherches en SHS, en bref

- 120 Le contexte et les raisons qui ont vu naître la loi Jardé, ainsi que les qualifications et professions de ses auteurs constituent une part d'explication au désarroi que l'on a vu naître chez les acteurs de la recherche en sciences humaines et sociales : la loi Jardé est en définitive une loi proposée par des médecins, pour encadrer la recherche clinique (Lemaire 2019) ; le champ lexical de la loi elle-même est, nous l'avons vu, très spécifique au domaine médical et ce sont en particulier les termes cliniques qui sont utilisés dans les publications traitant de la loi Jardé, des recherches interventionnelles, etc. Toute l'ambiguïté à laquelle chacun a pu être confronté à la lecture des textes réglementaires a disparu à la publication du décret de mai 2017.
- 121 En SHS, pour le recueil des données linguistiques, il s'agira dans un premier temps d'identifier clairement la finalité de la recherche afin de pouvoir la situer hors ou dans le champ d'application de la loi Jardé, en se référant aux articles R. 1121-1.I et R. 1121-1.II-1°-d du décret n° 2017-884 du 9 mai 2017. Nous pensons également nécessaire l'examen, outre la finalité, de la population impliquée et de la méthode de recueil des données. Puis, le cas échéant, il conviendra de situer sa recherche dans l'une des trois catégories RI, RIRCM ou RNI et d'accomplir les démarches de rigueur.
- 122 La loi Jardé, nous l'avons vu, propose un cadre réglementaire et inclut dans les attributions des CPP la prise en charge les dispositions d'information et de consentement des personnes volontaires se prêtant à la recherche. Cependant, ce point concerne toutes les recherches impliquant la personne humaine ; ainsi, les questions d'information et de consentement restent une préoccupation obligatoire et réglementée du responsable du projet de recherche. Elle est définie cette fois par le cadre réglementaire de la protection des données à caractère personnel codifié par le RGPD dont les nouvelles dispositions ont fait l'objet de la première partie de cet article.

4. Conclusion

- 123 Les nouveaux cadres réglementaires pour ce qui concerne la protection des données et la protection des personnes ont généré beaucoup d'inquiétudes chez les différentes communautés de chercheurs qui ont senti leur pratique sur le point d'être bouleversées.
- 124 Ces inquiétudes n'ont pas épargné les membres du groupe éthique du LPL. Mais après deux années à prendre connaissance des textes et débattre de leur contenu, nous avons progressé dans la compréhension de ces nouveaux cadres réglementaires et avons pu poser un regard plus positif sur les changements, finalement moindres, qu'ils entraînent dans notre pratique courante.

- 125 En réalité, en SHS, la nouvelle réglementation sur la protection de la personne n'a pas d'impact majeur, puisque dans la majorité des cas, les recherches en SHS n'entrent pas dans le champ de la loi Jardé. En définitive, le dossier que l'on peut avoir à déposer de manière plus certaine, constitue la demande d'approbation d'un comité éthique et de recherche des universités qui nous permettra de valider nos questionnements éthiques et satisfaire les exigences éditoriales des revues. Le nouveau règlement de protection des données n'est finalement pas non plus une nouveauté en ce sens que les contraintes qu'il impose existaient pour la plupart déjà et étaient seulement méconnues des chercheurs, ces derniers n'ayant alors pas conscience que leurs données pouvaient être concernées.
- 126 Plus qu'un frein à nos pratiques, ces nouvelles réglementations peuvent être vues de manière positive puisque par les réflexions qu'elles nous conduisent à avoir sur nos projets, elles nous permettent d'élaborer, de manière sans doute plus précise, les notions relatives à la protection des personnes impliquées dans nos études, et à prendre en considération de façon plus systématique les possibles écueils et biais méthodologiques relatifs au traitement et à la protection de données. Les formulaires que nous sommes amenés à remplir peuvent ainsi être vus comme des outils d'amélioration de nos pratiques professionnelles. Peu à peu, nos tutelles ont mis en place des supports d'aide (documents en ligne, personnel dédié) qui rendent la tâche plus facile. Enfin, l'énergie que nous mettons pour être en conformité avec la loi Jardé, le RGPD et l'éthique de la recherche n'est pas vaine puisqu'elle est maintenant amenée à être réinvestie pour répondre aux nouvelles demandes de Data Management Plan (DMP) qui émanent des organismes financeurs tels que l'ANR.

BIBLIOGRAPHIE

Arrêté (2018). Arrêté du 12 avril 2018 fixant la liste des recherches mentionnées au 3° de l'article L. 1121-1 du code de la santé publique. NOR : SSAP1810240A. <https://www.legifrance.gouv.fr/eli/arrete/2018/4/12/SSAP1810240A/jo/texte>.

CNRS-InSHS (2019). André-Poyaud I., Astor S., Baude O., Boudjaaba F., Bujan G., Collignon B., Dubois F., Kessous E., Maurel L. & Roger M. *Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte*. InSHS – Guide pour la recherche, juin 2019. Récupéré sur https://inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf.

Décret (2017). Décret n° 2017-884 du 9 mai 2017 modifiant certaines dispositions réglementaires relatives aux recherches impliquant la personne humaine NOR : AFSP1706303D <https://www.legifrance.gouv.fr/eli/decret/2017/5/9/AFSP1706303D/jo/texte>.

Dérogations (2019). Article 100-1, abrogé au 1^{er} juin 2019. Section 5 : Garanties et dérogations applicables au traitement à des fins de recherche scientifique ou historique ou à des fins statistiques. Récupéré sur <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006052581&dateTexte=20180812#LEGISCTA000037282494>.

Ghio A., Pouchoulin G., Viallet F., Giovanni A., Woisard V., Crevier-Buchman L., Hirsch F., Fauth C. & Fredouille C. (2020). « Du recueil à l'exploitation des corpus de parole "pathologique" : comment accéder à la variation physiopathologique ? », *Corpus* 22.

Huriet (1988) loi n° 88-1138 du 20 décembre 1988 dite Huriet relative à la protection des personnes qui se prêtent à des recherches biomédicales NOR : SPSX8810045L.

Jaillon P. & Demarez J.-P. (2008). « L'histoire de la genèse de la loi Huriet-Sérusclat de décembre 1988 », *Médecine/Sciences* 24(3) : 323-327 ; DOI : 10.1051/medsci/2008243323.

Jardé (2012). Loi n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine. Version consolidée au 26 janvier 2020. Récupéré sur <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000025441587>.

Lemaire F. (2019). « La loi Jardé : ce qui change », *La Presse Médicale* 48(3) Part. 1 : 238-242.

Lemaire F. & Matei M. (2012). « De la loi Huriet à la loi Jardé », *Réanimation* 21 : 373-374.

Matei M. & Lemaire F. (2013). « La loi Jardé facilitera-t-elle la recherche clinique ? », *La lettre du Cardiologue*, 464-465.

LIL (1978). Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Récupéré sur <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000886460&categorieLien=id>.

LIL (2019). La loi « Informatique et Libertés », nouvelle édition du 17 juin 2019. Récupéré sur <https://www.cnil.fr/fr/la-loi-informatique-et-libertes>.

Loi de Santé Publique (2004). Loi n° 2004-806 du 9 août 2004 relative à la politique de santé publique, NOR : SANX0300055L <https://www.legifrance.gouv.fr/eli/loi/2004/8/9/SANX0300055L/jo/texte>.

LPL (2020). GT éthique. Récupéré sur <http://www.lpl-aix.fr/recherche/ethique/>.

Priego-Valverde B. (2018). « Sharing a laugh at others : Humorous convergence in French conversation ». *European Journal of Humour Research*, International Society for Humor Studies, 6(3), DOI : 10.7592/ejhr2018.6.3.priego, HAL : hal-01923354.

Rat C., Tudrej B., Kinouani S., Guineberteau C., Bertrand P., Renard V., Saint-Lary O. & le Comité d'éthique du Collège national des généralistes enseignants (2017). « Encadrement réglementaire des recherches en médecine générale », *Exercer* 135 : 327-334.

RGPD (2018). Le règlement général sur la protection des données - RGPD, 23 mai 2018. Récupéré sur <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>.

Sarr M.M., Pinto S., Jankowski L., Teston B., Purson A., Ghio A., Régis J., Peragut J.C. & Viallet F. (2009). « Contribution de la mesure de la pression intra-orale pour la compréhension des troubles de la coordination pneumophonique dans la dysarthrie parkinsonienne », *Revue Neurologique* 165 : 1055-1061, [IF: 1.762] DOI : 10.1016/j.neurol.2009.03.012.

NOTES

1. Appelé aussi DPO pour *Data Protection Officer*.
2. Citons par exemple, la métadonnée qualifiant le type de dysarthrie d'un corpus constitué d'enregistrements sonores comme des extraits du texte de « La chèvre de Mr Seguin » lus par des personnes souffrantes de maladie neurologique.
3. Commission Nationale de l'Informatique et des Libertés.

4. Passage d'une logique déclarative ou d'autorisation préalable à une logique de responsabilisation et de contrôle du responsable de traitement.
5. Numéro d'Inscription au Répertoire (INSEE) ou encore appelé Numéro de Sécurité Sociale.
6. Comité d'Expertise pour les Recherches, les études et les évaluations dans le domaine de la santé.
7. Par exemple, le formulaire d'inscription au registre du CNRS accessible depuis l'intranet <https://intranet.cnrs.fr/>.
8. Article 5 du RGPD, paragraphe 1 (a-f).
9. En général, il correspond à une description plus explicite que le titre du projet.
10. Au sens de l'article L. 112-1 du code de la recherche [En ligne].
11. Article 5 du RGPD, paragraphe 1 (e).
12. Voir « L'anonymisation des données, un traitement clé pour l'Open Data », CNIL, 17 octobre 2019, <https://www.cnil.fr/fr/lanonymisation-des-donnees-un-traitement-cle-pour-lopen-data>.
13. Par exemple, une durée de 7 années (+ l'année en cours) à laquelle se cumule 2 années supplémentaires après chaque publication de résultats scientifiques utilisant les données.
14. « La pseudonymisation permet de traiter les données d'individus sans pouvoir identifier ceux-ci de façon directe », CNIL.
15. Par exemple, une clé codée sur 7 caractères correspondant à la concaténation de l'initiale en majuscule du nom et du prénom, du caractère tiret et d'un nombre incrémental sur 4 chiffres.
16. <https://www.nakala.fr/>
17. <https://www.ortolang.fr/>
18. Par exemple, le bruitage ou la déformation d'enregistrements sonores rendrait impossible toute recherche en linguistique ou en phonétique clinique.
19. <https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre3>
20. Droit à l'information, d'accès, de rectification, à l'effacement, à la limitation, d'opposition, à la portabilité, au retrait du consentement...
21. Droit de définir des directives relatives au sort de ses données à caractère personnel après sa mort (Chap. II Art. 48 LIL 17 juin 2019).
22. <http://www.lpl-aix.fr/recherche/ethique/>
23. <https://www.parisnanterre.fr/agenda/olivier-jarde-francois-lemaire-les-recherches-impliquant-la-personne-humaine-ce-que-change-la-loi-jarde--907338.kjsp>

RÉSUMÉS

Dans le cadre de la collecte des corpus oraux et multimodaux en Sciences Humaines et Sociales (SHS), le respect de l'application du nouveau Règlement Général sur la Protection des Données à caractère personnel provoque un profond bouleversement dans l'organisation et les méthodes de travail des chercheurs. C'est aussi le cas de la Loi Jardé, relative aux recherches impliquant la personne humaine, qui a étendu la notion de protection des personnes au-delà du champ de la recherche biomédicale. Ces deux importants changements de cadre législatif obligent aujourd'hui la communauté scientifique à s'interroger à la fois sur la protection de la personne et sur la protection des données à caractère personnel.

En définitive, avant d'initier un travail de recherche, un important travail de réflexion et d'actions est à mettre en place pour collecter des données dans le respect de la réglementation de

protection des données et du cadre législatif de protection de la personne humaine. Identifier les difficultés pour envisager les solutions les plus adaptées est la mission que s'est fixée le groupe de réflexion constitué autour de la question « éthique » au Laboratoire Parole et Langage (LPL 2020). Nous proposerons dans cet article les réflexions argumentées de notre groupe sur l'impact de ces nouvelles réglementations sur les recherches en linguistique.

The new legal regulations for personal data protection (in France, Règlement Général sur la Protection des Données - RGPD) have had a profound impact on the organization and methodology concerning the collection of oral and multimodal corpora for human science research. This was also the case with the Jardé law for research involving human beings, which extended the concept of individual protection beyond the field of biomedical research to include research in the humanities. These two important changes to the French legislative framework bring the scientific community to question the protection of participants and personal data.

In compliance with the legislative framework, important reflection and actions are necessary before initiating any research or data collection. To deal with these ethical questions, the think tank at the Laboratoire Parole et Langage (LPL 2020) set out to identify the difficulties in determining the most suitable solutions. In this article, we will share the documented reflections of our group on the impact of these new legal regulations on linguistic research.

INDEX

Mots-clés : RGPD, données personnelles, Loi Jardé, recherche impliquant la personne humaine

Keywords : RGPD, personal data, Jardé law, research involving human beings

AUTEURS

MURIEL LALAIN

Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France

GILLES POUCHOULIN

Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France

BÉATRICE PRIEGO-VALVERDE

Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France

SERGE PINTO

Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France