Production de ressources multilingues pour l'aide à la traduction du droit pénal en hindi, ourdou et français

Bénédicte Parvaz Ahmad

▶ **To cite this version:**

Bénédicte Parvaz Ahmad. Production de ressources multilingues pour l'aide à la traduction du droit pénal en hindi, ourdou et français. Linguistique. Institut National des Langues et Civilisations Orientales- INALCO PARIS - LANGUES O', 2019. Français. NNT : 2019INAL0017 . tel-02916259v2

HAL Id: tel-02916259
https://inalco.hal.science/tel-02916259v2

Submitted on 16 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Institut National des Langues et Civilisations Orientales

École doctorale n° 265

Langues, Littératures et Sociétés du monde

Équipe de Recherche Textes, Informatique, Multilinguisme

Thèse

présentée par :

Bénédicte Diot-Parvaz Ahmad

soutenue le 30 novembre 2019

pour obtenir le grade de docteur de l'Inalco
en Sciences du Langage : linguistique et didactique des langues

Production de ressources multilingues pour l'aide à la traduction du droit pénal en hindi, ourdou et français

Thèse dirigée par :

M. Mathieu Valette	Professeur des universités, Inalco
Mme Annie Montaut	Professeure des universités émérite, Inalco

Rapporteuses:

Mme Sylvie Monjean-Decaudin	Professeure des universités, Université Paris-Panthéon Sorbonne
Mme Natalie Kübler	Professeure des universités, Université Paris-Diderot

---

Membres du jury :

Mme Fayza El Qasem	Professeure des universités, Université Paris-Sorbonne Nouvelle
Mme Natalie Kübler	Professeure des universités, Université Paris-Diderot
Mme Sylvie Monjean-Decaudin	Professeure des universités, Université Paris-Panthéon Sorbonne
Mme Annie Montaut	Professeure des universités émérite, Inalco
Mme Marie-Caroline Saglio-Yatzimirsky	Professeure des universités, Inalco
M. Mathieu Valette	Professeur des universités, Inalco

Remerciements

Cette thèse a représenté un long marathon de trois ans qui a demandé un investissement total, physique et moral. Le franchissement de la ligne d'arrivée n'aurait pas été possible sans les personnes suivantes que je tiens à remercier chaleureusement.

Mes remerciements vont à Annie Montaut sans qui mon amour pour les langues indiennes se serait cantonné à une excentricité de jeunesse. Merci de m'avoir permis de m'épanouir dans ce domaine et de m'avoir fait – et de continuer à me faire – découvrir les inépuisables richesses et curiosités de cette aire linguistique. Merci de m'avoir encouragée à proposer ce projet de thèse, d'y avoir cru, et de m'avoir accompagnée pendant ces trois années.

De chaleureux remerciements également à Mathieu Valette, qui m'a accueillie dans son équipe de recherche, malgré mon parcours de traductrice-interprète néophyte en linguistique de corpus. Merci de m'avoir intégrée à cette formidable équipe de recherche, merci pour son soutien et sa grande disponibilité, dans le temps de la recherche comme en dehors.

Merci à tous les formidables collègues de l'équipe ERTIM, dont les qualités humaines ont rendu cette période passionnante. Parmi l'équipe, un merci tout spécialement à François Stuck pour son aide en ingénierie infromatique. Sans lui, les applications informatiques (concordancier, occurrencier, tests d'évaluation des termes) n'auraient pas vu le jour. Je tiens aussi à le remercier pour le temps qu'il a consacré à cette collaboration, et pour sa relecture scrupuleuse de mon manuscrit. Merci aussi à Jean-Michel Daube qui m'a assistée dans mes débuts hésitants en programmation. Enfin, merci à Cyril Grouin d'avoir pris le temps de me donner des conseils méthodologiques toujours pertinents pour le traitement de mes corpus multilingues.

Merci à Océane, Amélie, Lucie, Jennifer et Liyun pour leur aide précieuse, notamment en informatique, aux débats constructifs et aux échanges de toutes sortes que nous avons eus pendant cette période riche.

Merci aux informateurs et évaluateurs/trices qui ont bien voulu participer à ce projet : Yasser Qureshi, présenté par la regrettée Maryam Abou Zahab, pour les précieux documents pakistanais ; M. Khan, avocat à Lahore, M. Haider Taj, parlementaire au Pakistan, qui ont apporté leurs lumières sur le système pakistanais ; M. Rajeev Surana, avocat à la Haute Cour de Jaipur en Inde, pour ses renseignements sur la procédure indienne ; Mme Bernadette Anton-Bensoussan, magistrate qui a longtemps œuvré à la Cour d'Appel de Paris, pour son engagement envers les interprètes et traducteurs judiciaires, sa compréhension de leurs problèmes et son ouverture transdisciplinaire, pour sa participation à l'évaluation terminologique en français et pour ses talents de pédagogue lors de ses formations à la langue judiciaire.

Merci également à Yves et Babette pour le soutien familial qu'ils m'ont apporté.

Enfin, merci à tous mes ami-e-s qui, pendant cette période passionnante mais compliquée, m'ont soutenue pendant les moments de doute. Leur bienveillance a contribué à ma santé tant physique que mentale. Parmi eux, un merci tout particulier à Matthieu Y.

Table des matières

Liste des abréviations pour la glose ix

Introduction 1

1 Traduire le droit : contextes juridiques et contextes linguistiques 5

1.1 **Des traditions judiciaires culturellement très distantes** . . . 7
1.1.1 Présentation du système pénal indien et pakistanais . . . 8
1.1.1.1 La tradition juridique dans le sous-continent indien . . . 9
1.1.1.1.1 *Période pré-britannique* . . . 10
1.1.1.1.2 *Période coloniale* . . . 19
1.1.1.1.3 *Les langues de la justice* . . . 21
1.1.1.2 Organisation contemporaine du judiciaire . . . 22
1.1.1.2.1 *L'organisation de l'État indien* . . . 23
1.1.1.2.2 *Le système judiciaire indien* . . . 25
1.1.1.2.3 *L'organisation de l'État pakistanais* . . . 29
1.1.1.2.4 *Le système judiciaire pakistanais* . . . 30
1.1.2 Présentation de la justice en France . . . 36
1.1.2.1 Tradition judiciaire en France . . . 36
1.1.2.1.1 *Sources du droit français* . . . 36
1.1.2.1.2 *Notions de droit, loi, justice* . . . 38
1.1.2.1.3 *Droit pénal* . . . 40
1.1.2.2 Organisation de la justice pénale française . . . 41
1.2 **Les politiques linguistiques des États** . . . 43
1.2.1 Le contexte multilingue en Asie du Sud . . . 44
1.2.2 Politique linguistique en Inde . . . 47
1.2.2.1 Le statut du hindi . . . 47
1.2.2.1.1 *Hindi et ourdou, une origine commune* . . . 48
1.2.2.1.2 *L'écriture devanagari* . . . 52
1.2.2.1.3 *D'après la Constitution* . . . 52
1.2.2.1.4 *Autres textes de loi* . . . 55
1.2.2.2 Les outils officiels de promotion du hindi . . . 57
1.2.2.2.1 *Le Department of official language* . . . 57
1.2.2.2.2 *Le BSTT et la Committee for Scientific and Technical Terminology ou vaigyānik evam taknīkī āyog (CSTT)* 57
1.2.2.2.3 *La « three-language formula »* . . . 59
1.2.2.3 L'ourdou en Inde . . . 60
1.2.2.3.1 *Particularités graphiques de l'ourdou* . . . 60
1.2.2.3.2 *Une langue minoritaire* . . . 61
1.2.3 Politique linguistique au Pakistan . . . 63
1.2.3.1 L'ourdou, langue nationale . . . 63
1.2.3.2 Textes législatifs entérinant le statut de langue officielle . 65
1.2.3.3 Perception de l'ourdou dans la société pakistanaise . . . 66
1.2.3.4 Promotion de l'ourdou au Pakistan . . . 67

1.2.4 Politique linguistique en France . . . 68
1.2.4.1 Un pays, des patois . . . 68
1.2.4.2 Une République, une langue . . . 69
1.2.4.3 Promotion du français . . . 69
1.3 **Élaboration des terminologies** . . . 71
1.3.1 En hindi : vers la sanskritisation . . . 71
1.3.1.1 Cadre du développement terminologique . . . 71
1.3.1.2 Fonds de création terminologique . . . 76
1.3.1.2.1 *Le fonds sanskrit* . . . 76
1.3.1.2.2 *Le fonds arabo-persan* . . . 76
1.3.1.2.3 *Le fonds anglais* . . . 77
1.3.2 Développement terminologique en ourdou . . . 78
1.3.3 Élaboration de la terminologie en français . . . 80
1.3.3.1 Les institutions de contrôle terminologique . . . 81
1.3.3.2 Mécanismes de formation terminologique . . . 82
1.3.3.3 Utilisation de la terminologie . . . 83
1.4 **Place de la traduction judiciaire dans les trois États** . . . 84
1.4.1 France : une place indispensable, reconnue mais mal comprise . . . 84
1.4.2 Inde et Pakistan : la traduction, une phase invisibilisée . . . 85
1.4.3 Conclusion . . . 86

2 La traduction judiciaire 89
2.1 **Spécificités de la traduction judiciaire** . . . 89
2.1.1 Un champ d'étude récent . . . 89
2.1.2 La traduction judiciaire selon la Justice en France . . . 91
2.1.2.1 L'interprétation du droit : une prérogative du juriste . . . 91
2.1.2.2 Traducteurs et interprètes, techniciens transcodeurs . . . 92
2.1.3 Sélection et rôle du traducteur-interprète judiciaire . . . 94
2.1.3.1 Critères de sélection et de contrôle . . . 94
2.1.3.2 Rôle de l'interprète judiciaire . . . 95
2.1.3.3 Rôle du traducteur judiciaire . . . 96
2.1.3.4 Les destinataires de la traduction . . . 97
2.1.4 Les équivalents fonctionnels de traduction . . . 98
2.2 **Déroulement de la procédure pénale** . . . 100
2.2.1 En France . . . 100
2.2.1.1 Signalement d'une infraction . . . 100
2.2.1.2 L'instruction . . . 102
2.2.1.2.1 *Éléments spécifiques à la procédure correctionnelle* . . . 104
2.2.1.2.2 *Éléments spécifiques à l'instruction criminelle* . . . 105
2.2.1.3 Le procès . . . 106
2.2.1.3.1 *Au tribunal correctionnel* . . . 106
2.2.1.3.2 *À la cour d'assises* . . . 107
2.2.1.3.3 *En appel* . . . 108
2.2.1.3.4 *Détention* . . . 108
2.2.1.4 Les gens et les fonctions de la justice . . . 108
2.2.1.4.1 *Les magistrats* . . . 109
2.2.1.4.2 *Magistrats du parquet* . . . 109
2.2.1.4.3 *La police judiciaire* . . . 110
2.2.1.4.4 *Greffe et greffier* . . . 110
2.2.1.4.5 *Les auxiliaires de justice* . . . 111

2.2.2 En Inde et au Pakistan . . . 112
2.2.2.1 Ouverture d'une information judiciaire . . . 113
2.2.2.1.1 *À partir d'un rapport de police* . . . 113
2.2.2.1.2 *À partir de la plainte d'un individu* . . . 113
2.2.2.2 Trois procédures principales . . . 113
2.2.2.2.1 *Warrant case* . . . 114
2.2.2.2.2 *Summons case* . . . 115
2.2.2.2.3 *Summary trials* . . . 115
2.2.2.3 Précisions sur la procédure pakistanaise . . . 115
2.2.2.4 Déroulement d'une audience . . . 116
2.3 **Genres textuels du judiciaire** . . . 119
2.3.1 Le genre textuel . . . 119
2.3.2 En France . . . 121
2.3.2.1 Textes entraînant un effet juridique . . . 121
2.3.2.1.1 *Ordonnances de clôture d'instruction et jugements* . . . 121
2.3.2.1.2 *Ordonnances du juge des libertés et de la détention* 127
2.3.2.1.3 *La Commission rogatoire internationale* . . . 128
2.3.2.1.4 *Les convocations, citations et avis à audience* . . . 129
2.3.2.1.5 *L'inscription à un fichier national* . . . 129
2.3.2.2 Les procès-verbaux . . . 129
2.3.3 En Inde et au Pakistan . . . 131
2.3.3.1 Énonciation, phraséologie et lexique . . . 132
2.3.3.1.1 *Le cadre énonciatif* . . . 132
2.3.3.1.2 *Phraséologie* . . . 132
2.3.3.1.3 *Lexique* . . . 132
2.3.3.2 Exemples en hindi . . . 133
2.3.3.2.1 *Textes entraînant un effet juridique* . . . 133
2.3.3.2.2 *Procès-verbaux* . . . 139
2.3.3.2.1 Synthèse pour le hindi . . . 141
2.3.3.3 Exemples en ourdou (Pakistan) . . . 142
2.3.3.3.1 *Textes entraînant des effets juridiques* . . . 142
2.3.3.3.2 *Procès verbaux* . . . 149
2.3.4 Conclusion . . . 154

3 Constitution et traitement informatique du corpus multilingue 157
3.1 **Définition et objectifs de la constitution du corpus** . . . 158
3.1.1 Qu'est-ce qu'un corpus ? . . . 158
3.1.1.1 Exhaustivité et représentativité . . . 159
3.1.1.2 Disponibilité des données . . . 160
3.1.1.3 Corpus parallèle, corpus comparable . . . 162
3.1.2 Objectifs de l'analyse de corpus . . . 163
3.1.2.1 Valider et formuler des hypothèses à partir du corpus . . . 163
3.1.2.2 Extraction lexicale et alignement multilingue des corpus . 164
3.1.2.3 Observation des usages théoriques et des usages réels . . . 165
3.2 **Présentation du corpus** . . . 166
3.2.1 Textes parajuridiques : un sous-genre entre pédagogie et vulgarisation 166
3.2.2 Texte juridique ou judiciaire ? . . . 168
3.2.3 Corpus de référence . . . 170
3.2.4 Langue de spécialité et langue commune . . . 170
3.3 **Traitement informatique d'un corpus multilingue** . . . 171
3.3.1 Le projet de la chaîne de traitement . . . 172

3.3.2 Ressources et outils disponibles en hindi et ourdou pour le traitement des corpus . . . 173
3.3.2.1 Les acteurs de la recherche en linguistique de corpus hindi-ourdou . . . 173
3.3.2.2 Les corpus disponibles . . . 175
3.3.3 Les outils de traitement à disposition et leur insertion dans la chaîne de traitement . . . 176
3.3.3.1 D'un encodage hétérogène des documents à leur normalisation en Unicode . . . 176
3.3.3.1.1 *Le problème de la représentation numérique des caractères* . . . 177
3.3.3.1.2 *Qu'est-ce qu'Unicode ?* . . . 178
3.3.3.1.3 *Et UTF-8 ?* . . . 179
3.3.3.2 Conversion des fichiers en formats interopérables . . . 180
3.3.3.3 Reconnaissance optique des caractères . . . 181
3.3.3.3.1 *OCR pour l'ourdou* . . . 182
3.3.3.3.2 *OCR en hindi* . . . 184
3.3.3.4 Aspiration de pages internet . . . 185
3.3.3.5 Nettoyage du contenu numérisé . . . 186
3.3.3.6 Segmentation, tokenisation, étiquetage . . . 187
3.3.3.7 Étiquetage morphosyntaxique en hindi et ourdou . . . 189
3.3.3.7.1 *Phase d'apprentissage* . . . 194
3.3.3.7.2 *Évaluation des POS tagger* . . . 194
3.3.3.8 Ressources lexicales et ontologies WordNet . . . 201
3.3.3.9 Anonymisation des sous-corpus juridictionnels . . . 203
3.3.4 Conclusion . . . 203

4 Résultats des analyses textométriques et du phraséolexique judiciaire 205
4.1 **Analyse textométrique** . . . 205
4.1.1 La textométrie, une approche à travers le texte et le contexte . . . 206
4.1.2 Extraction terminologique . . . 206
4.1.2.1 La recherche des spécificités . . . 206
4.1.2.2 Limites de la recherche des spécificités . . . 207
4.1.3 Recherche des segments répétés . . . 208
4.1.4 Contraste entre genres textuels . . . 209
4.1.4.1 Contraste entre les sous-genres parajuridique et normatif en hindi . . . 210
4.1.4.2 Contraste entre les forums et les textes de loi en hindi . 210
4.1.4.2.1 *Distribution exclusive des lexèmes* . . . 211
4.1.4.2.2 *Recherche des segments répétés dans les Forums* . . . 212
4.1.4.2.3 *Recherche des segments répétés dans les Lois* . . . 212
4.1.4.3 Contraste entre les articles de presse et les textes de loi en hindi . . . 213
4.1.4.3.1 *Recherche des termes exclusifs à une partition* . . . 214
4.1.4.3.2 *Recherche des termes spécifiques à une partition* . . . 215
4.1.4.3.3 *Recherche des segments répétés dans la langue de presse* 215
4.1.4.3.4 *Conclusion* . . . 217
4.1.5 Recherche des cooccurrents . . . 217
4.2 **Alignement de corpus parallèles multilingues et équivalences traductionnelles** 218
4.2.1 Indices de surface et cognats . . . 218
4.2.2 Exemple de recherche d'équivalence traductionnelle . . . 218

4.3 **Alignement de corpus comparables et recherche des équivalences traductionnelles** . . . 219
4.3.1 Méthodes d'alignement de corpus comparables . . . 220
4.3.2 Alignement multilingue des codes pénaux indien et français . . . 221
4.3.2.1 Alignement intermédiaire avec l'anglais pour langue pivot 221
4.3.2.1.1 *Extraction des unités lexicales les plus fréquentes* . . . 221
4.3.2.1.2 *Organisation de la structure textuelle* . . . 222
4.3.2.1.3 *Répartition en réseaux notionnels* . . . 223
4.3.2.1.4 *Alignement des versions française et hindi* . . . 226
4.3.2.1.5 *Extraction de définitions grâce à l'alignement multilingue* . . . 229
4.4 **Étude du lexique et de la phraséologie** . . . 231
4.4.1 Classement notionnel des lexèmes et unités phraséologiques (phraséolexique) . . . 233
4.4.1.1 Méthodologie . . . 234
4.4.1.1.1 *Les réseaux notionnels* . . . 235
4.4.1.1.2 *Étiquettes notionnelles et unités lexicales* . . . 236
4.4.1.1.3 *Collocations et unités phraséologiques* . . . 237
4.4.1.1.4 *Présentation des résultats* . . . 237
4.4.1.1.5 *Mise à disposition des résultats en ligne* . . . 245
4.4.2 Analyse des résultats et équivalences traductionnelles . . . 247
4.4.2.1 Recherche d'équivalences traductionnelles . . . 248
4.4.2.1.1 *Correspondances parallèles* . . . 248
4.4.2.1.2 *Correspondances partielles* . . . 249
4.4.2.1.3 *Absence de correspondance* . . . 252
4.4.2.2 Prise en compte du destinataire, variation entre langue de spécialité et parajuridique . . . 253
4.4.2.2.1 *Opacité dans les textes normatifs* . . . 254
4.4.2.2.2 *Spécialisation lexicale judiciaire dans le parajuridique* 255
4.4.2.2.3 *Conclusion pour l'analyse des résultats et équivalences traductionnelles* . . . 257
4.4.3 Conclusion . . . 258

A Abréviations et glossaire 265

B Ressources utilisées pour la constitution des corpus 273
B.1 Ressources en hindi . . . 273
B.2 Ressources en ourdou . . . 274
B.3 Ressources en français . . . 275
B.4 Programmes d'extraction des articles de presse judiciaire . . . 277

C Index thématique des termes judiciaires 279
C.1 Index des termes en français . . . 279
C.2 Index des termes en hindi . . . 280
C.3 Index des termes en ourdou . . . 281

D Tableaux monolingues des cooccurrences et unités phraséologiques par réseau notionnel 283
D.1 Infractions . . . 283
D.2 Mesures et sanctions . . . 295
D.3 Instruction . . . 300
D.4 Mécanique du procès . . . 308

D.5 Lieux de justice 313
D.6 Protagonistes 317

Index 323

Bibliographie 329

Liste des abréviations pour la glose

première personne (1)
deuxième personne (2)
troisième personne (3)
accompli (ACC)
adjectif (ADJ)
arabo-persan (A.P.)
adverb(ial) (ADV)
arabe ou d'origine arabe (AR.)
auxilliaire (AUX)
causatif (CAUS)
converbe (COV)
démonstratif (DEM)
déterminant (DET)
distant (DIST)
duratif (DUR)
emphatique (EMPH)
ergatif (ERG)
féminin (F)
futur (FUT)
génitif (GEN)
hindi (HI.)
honorifique (HON)
impératif (IMP)
inaccompli (INACC)
indicatif (IND)
infinitif (INF)
instrumental (INST)
intransitif (INTR)
locatif (LOC)
masculin (M)
négation, négatif (NEG)
oblique (OBL)
ourdou (OU.)
pluriel (P)
participe (PART)
passif (PASS)
passé (PASSÉ)
possessif (POSS)
présent (PRES)
proximal (PROX)
persan ou d'origine persane (PS.)
réflexif (REFL)
relatif (REL)
singulier (S)
sanskrit ou d'origine sanskrite (SKT.)
subjonctif (SUBJ)
suffixe (SUF)
transitive (TR)
vocative (VOC)

Transcription et translittération des langues sud-asiatiques

La translittération des mots en langues sus-asiatiques (hindi, ourdou, sanskrit) adoptée ici suit les règles prescrites par la norme ISO-15919[1], à deux exceptions près. Tous les mots translittérés sont en caractères italiques. Lorsque des mots sont employés dans un cadre plus large que celui d'un énoncé ou d'une désignation terminologique, ils suivent l'orthographe la plus couramment utilisée , en version romaine[2]. Ex : « Nyaya Panchayat » ou « Huddood » dans la description du système pénal, « devanagari », noms d'auteurs et de personnages politiques, titres d'œuvres non confidentielles (« Dharmashastra », « Véda »)

Translittération des caractères arabo-persans

Le premier tableau (fig. 1) indique les normes pour la translittération des caractères arabo-persans pour l'ourdou. Elles diffèrent des règles de translittération de l'arabe et du persan afin de conserver une cohérence avec les phonèmes communs entre l'ourdou et les autres langues d'Asie du Sud. Une seule modification y a été apportée, la translittération du caractère ourdou ؤ est v en position de consonne, *ū* en position vocalique, au lieu de *w*. Trois caractères, dont les transcriptions manquaient au tableau, ont été ajoutés : la nasalisation vocalique ں translittérée ̃ le hamza ء, translitteré par ’, et la voyelle *alef* longue suscrite ٰ, translittérée par *ā*. Les tables reproduites ci-après sont tirées de la documentation officielle (op. cit. ci-dessus).

caractère arabo-persan	translitération	caractère arabo-persan	translittération
ث	s̲	ط	ṯ
ح	ḥ	ظ	ẓ
خ	k̲h	ع	ʿ
ذ	z̲	غ	ġ
ز	z	ف	f
ژ	ž	ق	q
ص	ṣ	و	v, ū
ض	ż	ں	̃
ء	’	ٰ	ā

Tab. 1: Translittération des caractères arabo-persans

Translittération des caractères en devanagari

Le second tableau (tab. 1) indique les normes suivies pour les langues utilisant la devanagari (hindi et sanskrit). Pour les voyelles nasalisées du hindi, le *candra-bindu* (ँ) a été translitteré par un tilde couvrant (अँ *ã*, इँ *ĩ*, उँ *ũ*, एँ *ẽ*, ओं *õ*). Dans le cas des diphtongues, le tilde recouvre la dernière voyelle (औं *aũ*, ऐं *aĩ*).

[1]ISO-15919, édition de 2001, information and documentation - Transliteration of Devanagari and related Indic scripts into Latin characters, `https://www.chatranjali.fr/Scripts/Standards/ISO15919.pdf`

[2]Sauf pour les titres d'œuvres, comme le veulent les conventions typographiques du français.

Ref.	devanagari	translittération	Ref.	dev.	translittération	Ref.	dev.	translittération
005	अ	a Rule 2 [a]	027	छ	cha	053	व	va
006	आ ा	ā	028	ज	ja	054	श	śa
007	इ ि	i	029	झ	jha	055	ष	ṣa
008	ई ी	ī	030	ञ	ña	056	स	sa
009	उ ु	u	031	ट	ṭa	057	ह	ha
010	ऊ ू	ū	032	ठ	ṭha	088	क़	qa
011	ऋ ृ	ṛ	033	ड	ḍa	089	ख़	kha
096	ॠ ॄ	ṝ	034	ढ	ḍha	090	ग़	ġa
012	ऌ ॢ	ḷ	035	ण	ṇa	091	ज़	za
097	ॡ ॣ	ḹ	036	त	ta	092	ड़	ṛa
015	ए े	e	037	थ	tha	093	ढ़	ṛha
016	ऐ ै	ai	038	द	da	094	फ़	fa
019	ओ ो	o	039	ध	dha	048a	ऱ	ŕ [c]
020	औ ौ	au	040	न	na	051	ळ	ḷa
013	ऍ ॅ	ê [b]	042	प	pa	002	ं	ṁ Rules 3, 5, 8 [a]
017	ऑ ॉ	ô [b]	043	फ	pha	001	ँ	◌̃

FIG. 1: Translittération de la devanagari

Note sur la correspondance entre la graphie et la prononciation

Le hindi et l'ourdou ont une origine commune mais s'écrivent avec des systèmes d'écriture différents. Dans les exemples cette étude, on pourra observer qu'un même mot est translitteré différemment en fonction du système d'écriture. Par exemple le commissariat sera translitteré *thānā* à partir de la devanagari et *thānah* à partir du système arabo-persan, alors que la prononciation est la même.

De plus, le système phonologique de l'ourdou ne différencie pas de nombreuses lettres de l'alphabet arabe et reste proche de celui du hindi :

- ز *z*, ض *ż*, ذ *ẕ* et ظ *z* sont prononcées *z*,
- ت *t* et ط *ṭ* sont prononcées *t*,
- س *s*, ص *ṣ* et ث *s* se prononcent toutes *s*,
- ک *k* et ق *q* sont rarement différenciées, sauf dans un registre soutenu, et prononcées *k*,
- de même que گ *g* et غ *ġ*, souvent prononcées *g*,

– ع ‘ ne constitue pas un phonème en ourdou mais sa présence modifie les voyelles autour,
– enfin, le ء ’ ne se prononce pas.

Introduction

La distance entre les langues et les cultures est une difficulté constamment présente dans le travail du traducteur ou de l'interprète, chaque langue ayant sa représentation et son découpage conceptuel du monde. Deux langues proches par leur parenté peuvent également présenter des divergences sur des concepts, alors que leurs termes se ressemblent (ce sont les fameux « faux amis »). De par la proximité historique, géographique ou culturelle, les difficultés terminologiques et traductionnelles de ces langues sont documentées et diverses solutions ont été proposées, même si elles restent parfois débattues, grâce à des ressources linguistiques comme les dictionnaires bilingues, les concordanciers et les bitextes. Des instances officielles, régionales, nationales ou internationale sont même à l'origine de l'élaboration de telles ressources et fixent les terminologies officielles (cas de gouvernements multilingues, des instances européennes ou de l'ONU). Des langues plus éloignées génétiquement ou culturellement, comme le français et l'arabe littéraire, peuvent aussi bénéficier de ce type de ressources, du fait de leur histoire et des échanges internationaux qui provoquent depuis longtemps le besoin de traduction de l'un à l'autre. Il existe, pour d'autres langues « éloignées », des ressources terminologiques dans des domaines précis (la philosophie, la religion, les arts). Dans les autres cas se pose le problème du manque de ressources multilingues dans les combinaisons linguistiques de travail du traducteur-interprète[3].

C'est notamment le cas, dans le domaine judiciaire, pour la traduction ou l'interprétation du français vers les langues du sous-continent indien et inversement. Ces langues, bien que parlées par des communautés de locuteurs parmi les plus importantes de la planète (442 millions pour le hindi, plus de 100 millions pour l'ourdou, le bengali ou le pendjabi), ne possèdent pas suffisamment de liens historiques culturels, économiques ou migratoires avec la France pour initier un travail de traduction systématisé, comme cela a pu être le cas pour des langues européennes ou l'arabe. Pourtant, la diaspora originaire

[3] « Traducteur-interprète » désigne dans cette étude les traducteurs (qui pratiquent leur spécialité à l'écrit) et les interprètes (qui la pratiquent à l'oral), ou les traducteurs-interprètes (qui pratiquent les deux).

d'Asie du Sud (en majorité d'Inde, du Pakistan, du Bangladesh ou du Sri Lanka) compte en France plusieurs centaines de milliers de personnes, et les locuteurs de hindi-ourdou y sont près de 100 000. Autant de femmes et d'hommes susceptibles d'avoir affaire à la justice pénale – en qualité de plaignant·e, de victime, de mis·e en cause ou de témoin – lors d'une procédure dans une langue qu'ils ne maîtrisent pas, ou mal. La France, en tant que membre de l'Union européenne, se doit de prévoir pour eux comme pour toute personne ne parlant pas français, une assistance linguistique en matière judiciaire, afin de garantir leur bonne compréhension de la procédure et leur droit de se défendre.

Le traducteur-interprète judiciaire, qui doit non seulement assurer la bonne communication des acteurs de la justice avec les justiciables non francophones, traduit ou interprète également à destination de la justice des pièces en langue étrangère utiles à la *manifestation de la vérité*. Pour mener à bien ces missions, Le traducteur ou l'interprète de hindi-ourdou qui travaille en France se trouve démuni. Ces langues, parlée en Inde et au Pakistan par des centaines de millions d'individus, manquent cruellement de ressources numériques lexicales monolingues ou bilingues dans le champ du droit, plus encore en matière judiciaire. Il doit s'appuyer sur quelques dictionnaires papier difficiles à obtenir de la langue source vers l'anglais (parfois dans l'autre sens, s'il est chanceux), et de quelques glossaires juridiques élaborés dans les pays d'accueil d'une forte diaspora indo-pakistanaise (Grande-Bretagne, États-Unis, Canada). Il recourt également à l'observation de bitextes. Dans tous ces cas, l'anglais est présent comme langue source ou cible de la traduction.

Fort de ce constat, comme tout traducteur technique, le traducteur-interprète judiciaire constitue ses lexiques bilingues ou multilingues, puis des mémoires de traduction pour consigner les tournures et les segments récurrents rencontrés au fil des missions. Les limites de cette démarche, non systématique, apparaissent vite. En effet, non seulement le hindi et l'ourdou ne partagent que très peu de référents linguistiques avec le français, mais en plus les systèmes juridiques des pays qui les emploient diffèrent grandement du système français.

En dehors de l'Inde et du Pakistan, les pays d'accueil de fortes communautés indo-pakistanaises possèdent un système juridique fondé sur la *common law*, tandis que le système français est civiliste[4]. L'essentiel du lexique et de la phraséologie du système

[4]Même si le Canada, dans sa partie francophone, applique un droit civiliste, de nombreuses traductions de concepts du droit anglo-saxon, qui ont été calquées de l'anglais vers le français, ne peuvent être utilisées telles qu'elles pour le droit français. De plus, la diaspora indo-pakistanaise n'est pas très nombreuse dans

anglo-saxon, que l'on retrouve en Asie du Sud, désignent des concepts de justice qui ne peuvent être transposés tels quels en français. Le simple usage d'un pivot linguistique, en l'occurrence l'anglais, s'en trouve disqualifié comme point de passage entre des langues et des systèmes.

Le traducteur-interprète se retrouve par conséquent face à deux obstacles : celui d'une langue pivot, l'anglais, et celui de la différence des systèmes judiciaires. Ces difficultés favorisent la multiplicité des choix de traductions en l'absence de standardisation. On peut opter pour la traduction normative d'un concept, au risque d'utiliser un terme incompréhensible par le justiciable. On peut aussi choisir un calque notionnel, avec une périphrase explicative dans la langue cible. On peut encore décider de créer des équivalents fonctionnels, ou même encore opter pour l'une de ces solutions en fonction du contexte. Quelle que soit la solution retenue, le traducteur-interprète doit pouvoir l'assumer et la justifier.

La terminologie textuelle est un moyen de surmonter ces obstacles tant linguistiques que culturels. À partir de la collecte d'un corpus de textes judiciaires en trois langues, français, hindi et ourdou, nous procéderons à l'extraction terminologique et à l'alignement de corpus parallèles ou comparables afin d'élaborer des ressources linguistiques multilingues, dans le but d'aider le traducteur-interprète à comparer termes et concepts, et à valider ses choix de traduction. Dans la base de données numérique et accessible en ligne, qui se veut la partie production lexicale multilingue de ce travail, l'utilisateur trouvera des définitions de concepts et des termes dans les trois langues. Le traducteur-interprète retrouvera les occurrences en contexte de ces termes et expressions recherchés, afin de choisir la solution la plus appropriée pour le cas qui l'intéresse. Il pourra aussi trouver des équivalents de traduction dans les trois langues[5].

Cette étude propose, par la création de ressources multilingues, d'aider à la recherche d'équivalents traductionnels dans des textes judiciaires issus de pays aux traditions juridiques aussi éloignées que la France, l'Inde et le Pakistan, et dans les langues de la justice de ces pays.

Dans la première partie, nous présenterons le contexte juridique et linguistique des

---

les zones francophones, ce qui s'explique par l'obstacle de l'apprentissage d'une nouvelle langue.

[5]Ici, il sera question avant-tout de traduction, par conséquent de documents écrits, mais les résultats présentés ici pourront trouver aussi bien des applications dans le domaine de l'interprétation, et même dans la didactique de l'enseignement de la langue de spécialité.

pays concernés : nous expliquerons dans un premier temps sur quel substrat historique et culturel s'appuient les systèmes judiciaires de la France, de l'Inde et du Pakistan, puis la manière dont les politiques linguistiques y sont menées, afin de comprendre la place qui y est accordée à la traduction et à l'interprétation judiciaire.

La deuxième partie sera consacrée à la pratique proprement dite de la traduction judiciaire. Nous exposerons les caractéristiques de cette discipline et en retracerons le développement qui a abouti à sa singularité. Nous soulèverons la divergence de vue entre la justice et le traducteur sur les objectifs de la traduction, puis nous aborderons les textes judiciaires susceptibles d'être traduits afin d'en comparer les caractéristiques génériques et terminologiques dans les trois langues d'étude.

La troisième partie sera consacrée à la production des ressources pour la traduction et l'interprétation. En s'appuyant sur les objectifs de la traduction judiciaire exposés dans les deux parties précédentes, ainsi que sur les genres textuels mis en jeu dans la pratique de terrain, on y traitera de la collecte du corpus multilingue et de la chaîne de traitements informatiques utilisée pour extraire les ressources multilingues. Les ressources terminologiques alignées et la comparaison des phraséologies produiront des équivalents traductionnels qui tiendront compte non pas de la terminologie normative de chacun de pays telle qu'elle est exposée dans les lexiques spécialisés, mais de ce qui est constaté sur le terrain dans la pratique langagière. Les équivalents traductionnels sont ainsi élaborés en tenant compte du destinataire du message. Nous présenterons à la fin de cette partie l'élaboration et le fonctionnement de l'outil de ressources multilingues qui se présente sous forme d'un moteur de recherche explorant différents corpus et alimenté par des lexiques résultant de l'extraction terminologique ou de l'alignement présentés précédemment.

Nous espérons que, pour le justiciable lambda ou le travailleur associatif désireux d'apporter des conseils juridiques au migrant hindiphone ou ourdouphone, la plateforme de ressources multilingue fournira l'accès à une information claire et fiable. La fiabilité est d'autant plus requise qu'il s'agit de droit pénal, impliquant une responsabilité importante quant aux choix des termes et à leurs conséquences.

Chapitre 1

Traduire le droit : contextes juridiques et contextes linguistiques

Si traduire le droit doit faire l'objet d'une étude si particulière, c'est que l'activité touche à la fois un objet culturellement très marqué et que cette traduction est sujette à une double articulation heuristique (Gémar, 2002; Harvey, 2002) : une articulation juridique, d'interprétation intralinguistique, selon laquelle on doit d'abord définir le sens juridique du texte dans la langue où il a été écrit, rôle qui revient au juriste. Une fois ce sens compris, la deuxième articulation consiste à réussir la transposition de ce sens dans un autre système juridique et une autre langue.

En effet, le droit est une matière culturellement marquée car il s'est élaboré en corrélation avec les particularités de la société auquel il est attaché, cette dernière organisant le monde selon une vision conceptuelle qui lui est propre. Même dans des pays aussi proches que l'Espagne et la France, de même tradition latine et civiliste, les différences juridiques existent. De plus, le droit est également sujet à interprétation intralinguistique à un double niveau. Tout d'abord, en tant que domaine de spécialité, il possède une terminologie, une phraséologie et des genres textuels qui lui sont propres et qu'il faudra adapter au destinataire, plus ou moins versé dans la connaissance juridique. Mais ce n'est pas tout : le droit dit aussi la légalité ou l'illégalité des actes, il reconnaît des infractions et définit des peines. À ce titre, il doit être interprété par un magistrat pour s'adapter au cas légal pratique pour lequel on le sollicite. La langue utilisée pour dire le droit peut donc elle-même être sujette à interprétation. Le traducteur juridique, davantage qu'un traducteur de tout autre domaine technique, est ainsi soumis à diverses forces contraignantes tout au long de son travail.

L'Inde et le Pakistan d'un côté, la France de l'autre, sont culturellement et linguistiquement distant. Avant de se pencher sur la manière dont on traduit les textes judiciaires dans ces pays, une mise en contexte s'impose. C'est l'objet de ce premier chapitre, qui pose en premier lieu les bases du contexte juridique (1.1) dans chacun des

pays étudiés, puis la politique linguistique de chaque État (1.2), afin de comprendre les mécanismes de genèse de la terminologie juridique pour le hindi, l'ourdou et le français (1.3).

Les systèmes juridiques indien et pakistanais ont été élaborés à partir du système de common law et d'equity britannique, mais celui-ci n'en constitue pas l'unique source (1.1.1). Quant au système juridique français, il est de tradition civiliste (1.1.2). La difficulté du traducteur ou de la traductrice-interprète qui travaille avec le hindi, l'ourdou et le français réside dans la distance culturelle et linguistique qui imprègne les textes d'Asie du Sud et les textes français. Cette distance implique qu'il n'existe que très peu de référents communs pour servir de points d'ancrage à des concepts ou des phraséologies. Le traducteur-interprète devra donc s'attacher à connaître et comprendre les différences de nature et de fonctionnement des systèmes afin de mener à bien sa mission puisqu'il lui faudra, dans la mesure du possible, transposer le droit non seulement d'une langue à l'autre mais également d'un système à un autre.

Les langues qui disent le droit ont leur importance. Elles font l'objet de politiques linguistiques (1.2) pour hiérarchiser l'utilisation des langues ou garantir des droits linguistiques dans un contexte multilingue comme en Inde et au Pakistan (1.2.1). Dans un tel contexte, les locuteurs non-natifs de la langue officielle la manipulent comme langue seconde avec plus ou moins d'aisance. Il est alors possible à l'État d'apporter des aménagements linguistiques pour la rendre plus facile à comprendre et faciliter l'accès au droit des citoyens. Les politiques linguistiques, en définissant les usages officiels des langues, encadrent aussi leur développement terminologique (1.3). Ainsi, voulant s'affranchir à la fois du passé colonial et d'une histoire commune, l'Inde et le Pakistan ont opté pour une stratégie de différenciation (1.3.1 et 1.3.2). Parallèlement, la France a développé une politique monolingue (1.3.3) et développe sa terminologie dans le cadre de l'Europe.

À la lumière des contextes juridiques exposés (1.1), de la compréhension du statut linguistique du hindi, de l'ourdou et du français (1.2) et de leurs développements terminologiques (1.3), nous aborderons pour chaque pays la place faite à la traduction judiciaire (1.4).

Avant d'étudier l'attitude des gouvernements envers les langues utilisées en justice, penchons-nous sur les différents systèmes judiciaires qui nous intéressent et sur les langues qui leur sont associées. Cette présentation en contexte permettra de mieux situer la place du traducteur-interprète judiciaire entre la justice et les langues.

1.1 Des traditions judiciaires culturellement très distantes

La connaissance des systèmes juridiques est déterminante pour une bonne traduction judiciaire. Elle permet, en les comparant, de montrer quelles en sont les affinités et les divergences. On pourra ainsi constater la distance culturelle qu'il faudra franchir dans l'opération de traduction.

L'Inde et le Pakistan héritent d'un système judiciaire commun dans les grandes lignes (1.1.1) fondé sur la tradition britannique de common law et d'equity. Tout au long de leur histoire commune, des traditions juridiques variées se sont développées (1.1.1.1). Leurs systèmes judiciaires contemporains (1.1.1.2) présentent donc des structures communes, mais 70 ans d'histoire indépendante ont aussi engendré des variations (1.1.1.2.1 et 1.1.1.2.3). De son côté, la France possède un système civiliste (1.1.2.2). Ces différences de traditions et de systèmes impliquent une organisation des procédures et un découpage conceptuel différents.

Les systèmes juridiques de common law et civiliste sont parmi les plus répandus à travers le monde (cf. 1.1[1]). La common law tire son nom de l'unification du droit coutumier de Grande-Bretagne[2]. La famille du droit « civiliste »[3] ou « romano civiliste », ou encore « romano-germanique », prévaut en Europe.

Une des grandes distinctions entre ces deux systèmes est la manière dont est réalisée l'instruction. Alors que le droit pénal de common law est qualifié d' « accusatoire » en la matière, la procédure civiliste est qualifiée d' « inquisitoire ». Dans le système civiliste, le magistrat (en France, il s'agit du juge d'instruction) dirige la procédure : il ordonne les actes d'enquête et dirige les débats. Dans la procédure accusatoire, la charge de la preuve revient à l'accusation, menée par le Procureur. Chaque partie présente ses pièces et ses témoins, qui seront validés ou rejetés par le tribunal.[4].

---

[1] source : University of Ottawa, Civil Law department Domaine public,https://commons.wikimedia.org/w/index.php?curid=1179721.

[2] La common law a ultérieurement été complétée par l'equity qui en tempérait les jugements. Les deux fonctionnent aujourd'hui à l'unisson, sans avoir fusionné.

[3] Attirons l'attention du lecteur non spécialiste du droit sur la distinction entre « civiliste », qui qualifie un système juridique, et « civil », qui désigne une branche du droit privé, par opposition au droit pénal.

[4] On dit traditionnellement que la common law se construit davantage sur la jurisprudence tandis que le système civiliste, plus longtemps lié à la hiérarchie religieuse, est plus vertical et s'appuie sur un appareil de textes législatifs plus fourni pour appliquer le droit. Dans la réalité, les limites de l'influence des textes législatifs ou de la jurisprudence sont beaucoup plus souple. Dans le système civiliste, la

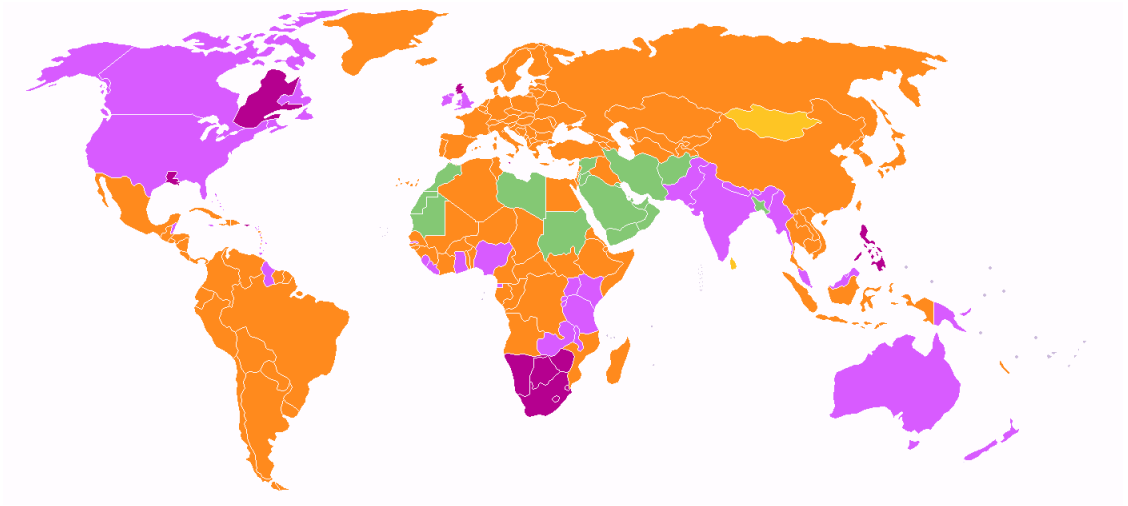

FIG. 1.1: Répartition des systèmes juridiques dans le monde : *violet*: droit de common law ; *orange*: droit civiliste ; *en vert*: droit musulman ; *pourpre*: système mixte common law / civiliste ; *jaune*: droit coutumier (source : Wikipedia)

Cette principale distinction entraîne, comme on le verra dans la description de la procédure judiciaire (2.2), des termes différents pour les acteurs et les pièces de la procédure. Cependant, ce n'est pas la seule distinction qui sépare les systèmes indiens, pakistanais et français.

1.1.1 *Présentation du système pénal indien et pakistanais*

Avant leur indépendance en 1947 au terme de la partition de l'Empire britannique, l'Inde et le Pakistan ont partagé une histoire multimillénaire, parcourue de très nombreuses cultures (1.1.1.1). S'ils partagent une organisation et des procédures judiciaires rattachées à la common law et à l'equity, ce système n'est pas leur unique source de droit. Les deux États se réfèrent aux mêmes textes pour le droit pénal (1860 pour le Code pénal et 1898 pour le Code de procédure pénale) rédigés lors de l'époque du Raj (1.1.1.1.2), période où la Couronne britannique gouvernait officiellement la majeure partie du sous-continent[5]. Cette période coloniale, qui débute à la fin du XVIII[e] siècle, fut précédée par une histoire juridique multi-millénaire (1.1.1.1.1), dans laquelle se distinguent une période avant l'arrivée des souverains musulmans (1.1.1.1.1) et une autre qui s'achèvera à la

jurisprudence influence notablement les jugements ultérieurs, tandis qu'un texte législatif aura préséance sur une jurisprudence dans le système de common law.

[5] À côté de ce gouvernement britannique subsistaient plus de 500 États princiers de tailles très diverses, avec des degrés d'indépendance variables vis-à-vis du pouvoir britannique. Nombreux étaient ceux qui avaient intégré la présence d'un *resident* britannique dans leur cour.

colonisation britannique (1.1.1.1.1).

Aujourd'hui, l'Inde et le Pakistan n'ont toutefois pas le même fonctionnement judiciaire car les sept dernières décennies ont apporté à chacun des évolutions politiques qui ont creusé des divergences (1.1.1.2). Par exemple, le Pakistan, créé initialement pour rassembler les populations musulmanes du sous-continent indien, a entamé dans les années 1970 une islamisation de la société qui s'est traduite également dans le champ judiciaire (1.1.1.2.4). C'est ainsi qu'au début des années 1980, une transposition partielle de la Charia dans le droit pénal fut opérée (1.1.1.2.4). De son côté, si le système judiciaire indien (1.1.1.2.2) est resté laïque, certaines procédures et peines ont été modifiées en ce qui concerne la minorité, les agressions sexuelles ou le trafic et l'exploitation d'êtres humains. Le code pénal [6] ainsi que le code de procédure pénale (appelé à partir de ce point CrPC) ont suivi les évolutions des sociétés de chaque pays, accusant la distanciation juridique[7].

1.1.1.1 La tradition juridique dans le sous-continent indien

FIG. 1.2: Carte de l'Asie du Sud. Source : http://www.russellmarlowe.com

Traiter la totalité des formes juridiques qui ont existé dans les territoires de l'Inde et du Pakistan actuels dépasserait le cadre de ce projet, tant les variations locales ont pu être nombreuses et la période longue. Une esquisse de l'histoire de la tradition

[6] Appelé dorénavant IPC, pour Indian Penal Code.

[7] On peut ajouter aujourd'hui pour l'Inde l'abrogation de l'article 377 de l'IPC (depuis le 6 septembre 2018) qui criminalisait l'homosexualité.

juridique de ces pays est cependant nécessaire pour comprendre la complexité actuelle de la représentation et de l'expression linguistique actuelle du droit. On distingue trois grandes périodes dans cette tradition : celle qui précède la venue des colons britanniques (1.1.1.1.1), allant de l'Inde antique à la période des souverains moghols (1.1.1.1.1), la période du Raj (gouvernement britannique, 1.1.1.1.2), et la période postérieure à l'Indépendance de 1947. On précisera dans la description de la période contemporaine (1.1.1.2) l'organisation juridique dans les deux pays.

1.1.1.1.1 Période pré-britannique

De l'Antiquité jusqu'à l'Indépendance en 1947, le sous-continent indien se caractérise par la diversité de ses peuples et de ses cultures. Le hindi s'apparente au sanskrit, langue indo-européenne antique, dont les terminologues revendiquent la prédisposition à enrichir les terminologies modernes du hindi. Il est utile, pour cette raison, d'étudier le système judiciaire dépeint à travers le sanskrit, afin de comprendre ce qu'il reste de cette époque et ce qui a changé ou disparu depuis. Le sanskrit n'était toutefois pas la seule langue pratiquée à l'époque (cf. partie 1.2.2 sur la politique linguistique en Inde).

Avant d'aborder les détails du système judiciaire, on doit se familiariser avec l'idée que se fait le monde indien antique de la justice et de l'ordre.

Dans la période indienne antérieure à la prise de contrôle officielle du pouvoir britannique, qui commence officiellement 1857, aucun concept ne se définit comme « droit » au sens de la tradition occidentale lui donne, pas plus que l'on ne trouve de concept semblable à la religion (cf. sous-section 1.1.2.1)[8]. Les notions que recouvrent ces concepts, loin d'être absentes de la pensée indienne, se trouvent englobées dans une vision du monde structurée différemment dans laquelle l'atteinte à l'ordre social, réprimée ou réparée par la peine en Occident, se traduit par une perturbation du *dharma*, l'ordre du cosmos qui assigne un rôle et des devoirs à chaque créature afin de perpétuer l'harmonie de l'Univers (Montaut, 2014). Le mot même de *dharma* vient de la racine sanskrite DHAR, qui signifie « tenir, maintenir ». Le *dharma* est donc un cadre structurant qui tient l'univers

[8]Le débat sur la définition du droit n'est pas clos : pour Robert Lingat, (Lingat, 1968), il s'agit de l'« ensemble des règles qui gouvernent actuellement et impérativement les hommes dans un milieu et à une époque donnés.», tandis que Norbert Rouland (Rouland, 1988), rappelle que pour les anthropologues juridiques, tout ce qui sert à la médiation et à la résolution de conflits au sein d'une communauté représente du droit, les systèmes alternant de la médiation à la répression selon les lieux et les époques.

et qu'il convient de préserver au moyen de modèles de comportement individuels[9]. À l'échelle de l'individu, le *dharma* permet à l'être de se réaliser et d'assurer son bien-être après la mort. Il s'agit donc d'un devoir moral, envers soi et l'univers. Vidé de ce sens premier, le mot a pris en hindi moderne le sens de « religion », inconnu de l'Inde ancienne[10].

Gardons cependant à l'esprit que cette vision de l'univers est celle d'un groupe social dominant, celui des brahmanes, dans le cadre de la communauté aryenne, une parmi d'autres dans le sous-continent[11].

a. L'Inde antique et pré-musulmane

La tradition brahmanique rapporte que le *dharma* a été révélé aux Rishis (les « sages ») à l'époque antique, au travers de la littérature de la shruti (*śruti*), dont font partie les Védas (*veda*), ces quatre recueils qui constituent le premier monument littéraire du sous-continent[12]. Le concept de dharma a été abondamment commenté dans les *dharmasūtra* (littéralement « aphorismes du dharma ») qui s'appuient sur l'autorité du Véda, de la Tradition et de la Bonne coutume pour définir à l'aide de nombreuses règles les rapports sociaux et les activités[13], constituant ainsi une forme de législation. La première source, le Véda (« connaissance »), a été collecté par différentes écoles d'interprétations, les *caraṇa*[14], qui en possédaient leurs propres recensions appelées *saṁhitā*. Dans cette littérature d'exégèse, la *vidhi* constituait une injonction qui s'appuyait sur un passage des

[9]Voir (Lingat, 1968): « Conformité avec ce que les Hindous regardent comme étant l'ordre naturel des choses, et c'est ce qui explique sa connexité avec le droit ».

[10]Le lien entre dharma et justice se retrouve encore dans la devise de cour suprême indienne : *yatō dharmastatō jayaḥ*, en sanskrit, qui signifie « Là où se trouve la vertu / la justice (le dharma) se trouve la victoire ».

[11]Les Aryens, dont on situe le berceau en Asie centrale, sont arrivés par périodes successives dans le sous-continent indien. Ils représentent l'un des quatre grands groupes de peuplement de la péninsule, avec les Austro-asiatiques, les Dravidiens et les Tibéto-birmans.

[12]Les textes les plus anciens du premier des Védas, le Rigveda (*ṛgveda*), remontent à -1500 ans. La période de rédaction s'étale jusqu'à -150.

[13]On trouve ainsi exposée dans le texte connu comme Lois de Manu la théorie des quatre Varṇa qui explique la division de la société en quatre grandes catégories d'humains : *brahmanes* (prêtres et intellectuels) en tête, suivis par les *kṣatriya*-s (guerriers et souverains), les *vaiśya*-s (commerçants) et *śudra*-s (serviteurs), auxquels incombent des devoirs spécifiques.

[14]Parmi ces école, la Mimansa (*mimāṁsā*), fondée par le sage Jaimini, est considéré comme l'inventrice du mode indien du raisonnement juridique (Lingat, 1968).

Védas. En hindi moderne, c'est ce terme qui désigne la « loi ».

La littérature de la smriti (*smṛti*), représente la tradition mémorisée par les brahmanes qui conseillent les souverains et sont amenés à juger des litiges. Cette tradition est consignée sous forme d'aphorismes (*sūtra*). La littérature de la smriti se divise en deux séries de textes. Dans la première, on trouve les *dharmasūtra*-s, consacrés aux devoirs et privilèges de la communauté aryenne, dont fait partie le Code de Manu (*manu saṁhita*), une partie des Purāṇa-s (III^{e} - X^{e} s.) et des textes composés aux IX^{e} et X^{e} siècles par Yājñavalkya, Viṣṇu, Nārada, Bṛhaspati, Kātyāyaṇa, Devala, Atri, Parāśra et Yama. Les textes du 2^{e} groupe sont des commentaires de ceux du premier.

La Bonne coutume, *sadācāra* ou *śiṣṭācāra*, recueille la pratique de la vie religieuse conseillée pour acquérir des mérites spirituels.

D'autres textes hors de la smriti, comme l'Arthashastra (*arthaśātra*), traité de gouvernement écrit par Kautilya sous l'Empire Maurya (du -IV^{e} au -II^{e} s.), ou des œuvres profanes comme les pièces de Kalidasa (IV^{e} - V^{e} s.), ou la *Rājatarangiṇī* de Kalhana (XII^{e} s.), nous renseignent sur les pratiques judiciaires de l'Inde de l'Antiquité jusqu'au XII^{e} siècle (Das, 1990).

Dans les Upanishads (*upaniṣad*, du -VI^{e}s. au XV^{e} s.), qui font également partie de la shruti, on trouve mention de *pariṣad* ou *samiti*, assemblées amenées à juger de la conformité des pratiques avec le *sadācāra*. Si, d'après (Lingat, 1968), ces formations de jugement disparaissent rapidement, les termes qui les désignent sont parvenus jusqu'en hindi moderne où ils désignent respectivement le conseil et le comité.

La loi universelle théorisée dans le *dharma* n'est cependant pas unique pour toutes les créatures : elle diffère en fonction de la place occupée par l'individu dans une société très hiérarchisée. Un membre d'une basse caste n'aura ainsi pas les mêmes obligations qu'un prince ou un guerrier de haut rang, de même qu'une femme ou un vieillard n'auront pas à remplir les mêmes devoirs qu'un commerçant chef de famille. L'acte dit *adharma*, contraire au *dharma*, peut entraîner sanction et déchéance.

Dans une moindre mesure, les buts de l'Homme (*puruṣārtha*) que sont l'*artha*(l'Intérêt) et le *kāma* (le Plaisir)[15] constituent d'autres règles de conduite de l'activité humaine. Elles restent toutefois subordonnées au *dharma*.

À côté de ce principe moral individuel très théorique qui dicte à chacun sa conduite, il existe une réalité coutumière et locale beaucoup plus pragmatique, fonctionnant sur le

[15]Traductions extraites de Malamoud (2016).

consensus, qui permet de remédier aux disputes ou de poser une médiation entre les individus, et qui garde toute sa puissance décisionnelle dans l'Inde d'aujourd'hui[16]. Cette coutume, désignée par les termes *rīt* puis *riti-rivāj* ou *rasm-rivāj*[17] (Montaut, 2014), repose sur les décisions prises par une formation de jugement au sein de la communauté (Derret et Duncan, 1999). Généralement au nombre de cinq (*pañc* en sanskrit, *pāñc* en hindi et ourdou), ces « sages » forment le conseil du panchayat (*pañcāyat*).

À plus vaste échelle, l'Inde antique a connu quelques périodes de relative unification avec un État fort lors desquelles existait une forme de législation portée par le *śāsan*. Dérivé de la racine *śās* « punir, gouverner », il désigne aussi bien à l'époque antique un traité, une charte (dans le Rigveda), qu'un décret (dans le Mahābhārata) (Monier-Williams, 1960)[18]. En hindi moderne, il désigne l'action de gouverner et peut également prendre le sens de « discipline », « règle ». Incarnant la législation, le souverain, *rājā*, avait pour devoir moral (*rājādharma*) d'assurer l'administration de la justice. Le Dharmaśāstra indique qu'il avait ainsi la possibilité de punir ou d'acquitter, qu'il protégeait les intérêts des parties en personne ou par l'intermédiaire d'un juge en chef (Rocher, nov.1968-fév.1969). Il était conseillé dans cette tâche par les brahmanes qui, versés dans la connaissance des textes, pouvaient vendre leurs services aux parties[19]. Le roi n'est toutefois pas au dessus des lois et peut faillir.

Outre les textes mentionnés plus hauts, des témoignages de visiteurs étrangers autour des premiers siècles avant et après l'an zéro, comme le Grec Mégasthène (-IV^{e}, -III^{e} s.) sous les Maurya, le pèlerin chinois Fa-Hien (+400-411) et plus tard de son compatriote Hiuen-Tsang (VII^{e} siècle), racontent dans leurs descriptions de la vie quotidienne des aspects de la justice contemporaine. Ces indices ont été étudiés par (Das, 1990), et (Moudgil, 2015) les reprend dans sa thèse.

On ne trouve pas de trace d'administration judiciaire à l'époque védique. Celle-ci est en revanche très élaborée dans le Dharmaśāstra et l'Arthaśāstra. La procédure judiciaire, *vyavahāra* (Kātyāyaṇa), y est expliquée comme une procédure dans laquelle un plaignant

[16]Nous y reviendrons dans la description des institutions modernes du droit.

[17]L'étymologie même de ces termes montre à quel point la coutume dépasse les communautés religieuses : *rīt* ou *riti* « rite, observance », du skt. ṛt « ordre », est associé aux termes*rasm* et *rivāj* d'origine arabe (Montaut, 2014).

[18]Parmi les exemplaires les plus spectaculaires de ces traités, citons les édits d'Ashoka, III^{e} s. av.J.-C., gravés sur des piliers disséminés dans tout son empire pour diffuser au peuple des principes moraux et son programme de gouvernement.

[19]U.C. Sarkar, Epochs in Hindu Legal History (1958), p.37.

et un défendant, qui se disputent, recherchent une décision[20]. La Nāradasmṛti (tradition découlant du sage Nārada) évoque pour la première fois une représentation devant un tribunal. Le tableau 1.1 reprend des termes extraits de l'ensemble de ces textes qui esquissent déjà les rôles des acteurs de la justice (Rocher, nov.1968-fév.1969; Das, 1990).

vyavahāra	procédure légale
vyavahārapada	sujet de différend
arthi	plaignant
pratyarthi	représentant du plaignant
niyoga	représentation légale
niyogin / niyukt / niyogakṛt	représentant d'une partie non professionnel et non rémunéré.
prativādin	représentant de l'une ou l'autre partie. Les représentants, qui sont exclus des poursuites graves (meurtre d'un brahmane, meurtre, vol, agression sexuelle d'une femme mariée, d'une servante, consommation d'aliments interdits, contrefaçon de monnaie…), sont des intimes du plaignant ou du défendant.
vivād	affaire au tribunal (Yājñvalkya (YāJÑ.))
kāyastha	scribe d'un tribunal inférieur
vivāka	litige
prāḍvivāka	juge présidant un tribunal (délégué représentant le roi)
prāḍ	celui qui pose des questions (aux parties et aux témoins)
dharmādhikārin	juge
sabhya	conseillers juristes (assesseurs)
lekhaka	scribe qui enregistre les décisions (Bṛhaspati (BṚHAS.))
sadhyapāla, sapuruṣa	huissier (BṚHAS.), convoque et garde les témoins et les parties

TAB. 1.1: Termes juridiques utilisés en sanskrit entre la période antique et le XII[e] s.

La plupart des juristes de l'époque distinguent de 18 à 20 catégories de litiges. Les litiges civils, d'*arthavivāda*, et les affaires pénales, dites *sāhasa*, sont jugés par les mêmes tribunaux.

Les juristes ne s'accordent pas sur la hiérarchie des tribunaux (*sabhā*, littéralement « assemblée »), qui a dû varier localement et au cours d'une période si longue[21].

Pour BṚHAS., il existe quatre types de tribunaux :

- *pratiṣṭhita*, tribunal fixe, dans une ville ou un village
- *apratiṣṭhita*, tribunal ambulant (forain)

[20] *vi* = « variés » ; *ava* = « doute » ; *hāra* = « qui ôte ».

[21] Das (1990), *op. cit.*

– *mudrita*, cour présidée par un juge qui représente le roi (et possède le sceau royal)
– *sasita*, cour présidée par le roi en personne.

Les auteurs Kātyāyaṇa et Yājñavalkya mentionnent 3 tribunaux populaires[22] :

1. *kula*, 1re instance pour les affaires mineures (pour Mitākṣarā, le tribunal est formé des membres aînés de la famille)
2. *śreni*, tribunal de corporation, de guilde, dont les membres peuvent être de castes différentes (cf. Mitā.)[23]
3. *puga*, assemblées d'hommes de castes et de professions différentes, mais habitant le même village ou la même ville. Aussi désignée comme *pañcamaṇḍalī*, cette assemblée fait penser au panchayat moderne.

D'autres auteurs citent le tribunal *gaṇa*, parfois comme synonyme de *puga* (ans le vyavahāra prakāśa), parfois de *kula* (cf. Kātyā.)[24]. Chez Kauṭilya, le panchayat de village ne juge pas les affaires graves. Il peut tout de même prendre des décisions concernant des affaires d'accident, d'homicide ou de suicide.

Différents fonctionnaires occupent des fonctions policières[25]. Kālidāsa mentionne le *nāgaraka* comme chef de la police et les rakṣinah (Moudgil, 2015, p.84). Pendant la période kuṣāṇa (du Ier au IIIe s.), on trouve la trace d'un officier de haut rang dont les fonctions varient, selon les textes, de celles d'un préfet à celles d'un commissaire ou d'un magistrat : le *daṇḍanāyaka*[26]. Un autre officier de police, qui exécute la peine, est appelé *daṇḍapāśika* ou *daṇḍabhogika*. L'officier de police habilité à traiter dix catégories d'infraction (*aparādha*) est nommé *daśāparādhika*[27]. Le chef d'un service de police est qualifié de *daṇḍādhikārin*[28]. De manière intéressante, on observe que la notion de peine (*daṇḍa*) prévaut dans le vocabulaire des forces policières, chargées de maintenir l'ordre et d'appliquer les sanctions, alors que c'est en hindi moderne la notion qui qualifie ce qui touche au pénal, que l'on soit au niveau de la police ou du tribunal. Dans le lexique des juges, les notions d'arbitrage entre des parties et de dharma sont présentes, mais nulle trace encore de *nyāya* qui marque en hindi moderne ce qui touche au judiciaire et aux

[22] Das (1990), *op. cit.*

[23] Mitākṣarā.

[24] Kātyāyaṇa.

[25] Das (1990) et Moudgil (2015), *op. cit.*

[26] Étymologiquement : le chef des peines.

[27] de *daśa*, « dix » + *aparādha* + suffixe d'adjectivation.

[28] *daṇḍa* « peine » + *adhikārin* « celui qui a des droits sur ».

tribunaux.

Lorsque nous aborderons la question du développement terminologique du hindi moderne, qui revendique le sanskrit comme langue de culture (cf. 1.3.1), nous remarquerons que certains des termes évoqués ici ont été retenus, tandis que la plupart ont été remplacés. La période musulmane marquera une stabilisation de l'administration et de ses termes.

b. L'Inde des souverains musulmans

Les souverains musulmans, turcs puis moghols, s'installent durablement dès le XII[e] siècle. Ils administrent des parties plus ou moins étendues du sous-continent jusqu'à l'instauration du gouvernement britannique en 1857, sauf dans les royaumes restés indépendants jusqu'en 1947. Leur gouvernement ne modifie pas fondamentalement l'état de la justice villageoise locale. La structure administrative instaurée par les nouveaux venus s'attache avant-tout à tirer des revenus produits dans les subah (*ṣubah* « provinces ») dont la gestion est déléguée à des officiers loyaux envers le régime. Les souverains musulmans ne cherchent pas à convertir en masse leurs sujets et la pleine application du fiqh (*fiqḥ*), droit islamique, se limite à la communauté musulmane. Les autres communautés placées sous leur administration, qualifiées de *ẕimī*, « protégées » moyennant une taxe[29], pouvaient continuer à pratiquer leurs coutumes sauf en matière fiscale et criminelle (Markovits, 1994, p.35).

Il va de soi que l'administration de la justice a évolué pendant cette période, et d'un souverain à l'autre. Cependant, les termes d'origine arabe, turcique ou persane qui sont restés dans la terminologie judiciaire[30] jusqu'à aujourd'hui témoignent d'institutions et de fonctions suffisamment stables au cours de cette période. En voici un certain nombre, rassemblés dans une courte synthèse, largement inspirée de l'œuvre très documentée de (Ahmad, 1941).

Sous l'Empire moghol, droit et religion sont indissociables. Il n'existe toutefois, pas plus que pour le droit dit hindou, de forme écrite uniforme pour l'interprétation de la chose juridique : le droit musulman se fonde sur l'interprétation de la jurisprudence en fonction d'écoles juridiques qui reconnaissent une collection plus ou moins grande d'ouvrages de référence et de méthodes (Dupret, 2012). Le droit musulman n'est pas figé. En plus des prescriptions tirées de l'interprétation des textes sacrés, on recourt à la loi

[29] Abolie sous le règne d'Akbar.

[30] Il en existe encore de très nombreux dans la procédure pénale et l'administration en général.

urfi (*ʿurfī*). Cette loi est une autre approche du droit musulman où le Coran représente une source éthique, non prescriptive, laissant au juge une grande marge de manœuvre : on y trouve des règles d'éthique concernant le commerce, la propriété, les impôts, la guerre…

Les fonctions judiciaires et exécutives étaient séparées, sauf s'agissant de la personne du souverain qui était le dernier recours judiciaire. On considérait alors qu'il existait trois catégories d'infractions mettant en péril l'ordre social : celles contre les prescriptions religieuses (hadd, *ḥadd*), celles relatives à un homicide (*qisas*), relevant du droit privé, et toutes celles n'appartenant pas à ces deux catégories (tazir, *taʿzīr*). Ces distinctions existent toujours dans le Code pénal pakistanais. De fait, les personnages et les institutions chargées de rendre la justice pouvaient également avoir un rôle religieux (cas des qazis) et être garants de la morale (sadar, *ṣadar*).

Le ministère de la justice était connu sous la dénomination de mahakma-e-adalat (*maḥakma-e-ʿadālat*) sous les Moghols, distinct du mahakma-e-shariat (*maḥakma-e-śariat*) qui s'occupait de la nomination des représentants administratifs religieux. Précédemment, sous la période du sultanat de Delhi (XIII^{e} - XVI^{e} s.), la justice était désignée par le terme *qaza*, qui a donné qāzī. Le ministère était dirigé par le *Qazi-ul Quzat* (ou Aqzal Quzat), chef de tous les autres qāzî-s. On trouvait ensuite un qāzī à la tête de chaque province (*qāzī-e-ṣubah*), de chaque district (*qāzī-e-sarkār*) et de chaque subdivision de district (*qāzī-e-parganah*), juges magistrats qui présidaient le système judiciaire à chacun de ces échelons.

Des fonctionnaires[31], dont certains ont subsisté dans le système britannique, étaient présents pour assister les qāzīs dans leur mission. À l'échelon du tribunal de district (*sarkār*), plusieurs fonctions s'occupent d'enregistrer les témoignages et les preuves (*katib*), d'administrer les archives (*peśkar*), de s'occuper des reliures et du matériel d'écriture (*daftarī*), tâches qui ressemblent à celles du greffier français. Un autre fonctionnaire est chargé de délivrer les notifications et les assignations (*mirdaha*), comme l'huissier en France. L'*amin* est chargé des affaires judiciaires et le *mucalkā navīs*, tel l'huissier-audiencier en France, vérifie la présence des parties et des témoins. Daftari, amin, mirdaha et muchalka navis existent toujours dans la procédure sous le règne britannique.

Au niveau du village, les panchayats, avec à leur tête le chef du village (*muqaddam*, *paṭel* ou *caudhrī*), réglaient les litiges civils ou pénaux mineurs (Ahmad, 1941).

[31] Ahmad (1941), *op. cit.*

Les bâtiments qui hébergeaient les magistrats et leurs cours étaient appelés adalat khana (*adālat k̲h̲ānā*) ou, comme aujourd'hui, kachehri (*kacehrī*).

Pour maintenir l'ordre, il existait un système policier au niveau du district (sarkar, *sarkār*) et en dessous[32]. Le responsable au niveau du sarkar ou, en dessous, du pargana était le faujdar (*faujdār*). Aujourd'hui encore, la procédure pénale en ourdou est qualifiée de *faujdārī*, en référence à cet aspect répressif de la justice. Vers la fin de l'Empire britannique, le faujdar remplit des fonctions identiques à celles du Magistrat habilité au titre de la section 30 (Section 30 magistrate). Il est responsable de l'ensemble des thanas (*thānā*, postes de police), qui ont chacun à leur tête un *thānedār*. Il existe en plus des avant-postes (chauki, *caukī*), termes toujours d'actualité. Dans les milieux urbains, le faujdar a sous ses ordres un kotwal (*kotwāl*) qui peut juger des affaires pénales mineures. Il dispose aussi d'une force montée (sarwar, *sarvar*), à pied (ahadi [*ahadī*], ou sipahi [*sipāhī*]), d'une force armée (barqandazi [*barqandazī*]). Au niveau du village, le chaukidar (*caukīdār*) assure la sécurité tandis que le chef de village représente le faujdar.

On retrouve des éléments de la procédure actuelle dans la procédure de l'époque moghole : arrestation (*ḥirāsat*), libération sous caution (*zamānat*), enquête et perquisition. L'*ʿadil*, ou témoin réalisant une déclaration sous serment, existe déjà.

La fonction d'avocat (*vakīl*) est mentionnée dans le Fiqh-e-Firoz Shahi (compilé entre 1351 et 1388) et le Fatawa-e-Alamgiri (compilé vers 1670). Sous Aurangzeb et Shahjahan est mentionné le *vakīl-e-sharaʿī*, sorte d'avocat commis d'office dans les affaires civiles pour assurer les intérêts de la partie s'opposant à l'État. Le principe du *vakālat nāmah* (procuration qui permet à l'avocat de représenter son client) existait déjà dans la même fonction qu'aujourd'hui.

On ne trouve pas de terme désignant spécifiquement l'appel, mais les juridictions qui les traitent sont les *ʿadālat-ul aliah*, ou *tajwīz sānī* (2e proposition), ou *ʿadālat murafiah* (à rapprocher du terme *ʿurfī*), ou simplement *istigasah*. Le terme utilisé dans la langue moderne est emrpunté à l'anglais « appeal ».

Tous ces éléments montrent que les Britanniques n'ont pas instauré ex nihilo des règles venues de Grande-Bretagne, ils se sont appuyés sur des structures déjà existantes qu'ils ont découvertes lors de la traduction des recueils de loi « musulmanes » (Fiqh-e-Firoz Shahi ou Fatawa-i-Alamgiri entre autres) au cours des XVIIIe et XIXe siècles. Cela est visible notamment dans certaines dénominations du Code pénal qui sont empruntées

[32] Ahmad (1941).

au système pré-colonial. Un grand nombre de ces termes se sont maintenus dans la terminologie judiciaire pakistanaise et, dans une moindre mesure, dans la terminologie indienne ainsi que dans la langue populaire (2).

1.1.1.1.2 Période coloniale

À partir du XVIII[e] siècle, les Britanniques s'implantent progressivement dans le sous-continent au gré de leurs transactions commerciales. À travers la Compagnie des Indes (East India Company), ils prennent le contrôle de territoires que leur louaient les souverains moghols et deviennent des zamindars (*zamīndār*, personne qui exploite la terre et en tire un revenu) comme les autres. Ils obtiennent également de ces mêmes souverains la prérogative d'y appliquer le droit (*dīvānī*, « droit civil »)[33](Markovits 1994, Jain 1972, p.68).

D'après Raj Kapil (Kapil, 2008), à l'arrivée des Britanniques à Calcutta (courant XVIII[e] siècle), les communautés de marchands (*banyan*, dubash [*dubaś*]) se spécialisent dans la rédaction des contrats commerciaux et représentent l'East India Company devant les instances mogholes régionales. La première Cour suprême britannique en territoire sud-asiatique est fondée à Calcutta en 1773 afin de régler les litiges selon la loi britannique[34]. Parallèlement s'ouvrent deux autres cours de justice à usage des indigènes, la Suddar Diwani Adawlat (*ṣaddar dīvānī ʿadālat*) pour les affaires civiles et la Sudder Nizamat Adawlat (*ṣaddar niẓāmī ʿadālat*) pour les affaires pénales (Kapil, 2008, p.88).

L'administration nécessite l'implication de la population indigène car les Britanniques sont très peu nombreux (quelques milliers tout au plus sur l'ensemble du territoire). Les spécialistes de la justice sous le gouvernement moghol ou dans les cours princières (Awadh, Hyderabad), versés dans les langues classiques arabe, sanskrite ou persane, se mettent au service du nouveau pouvoir. (Kapil, 2008, p.79) en cite huit catégories : le *qāzī* (magistrat, aujourd'hui personne versée dans le droit civil musulman, chargé notamment de célébrer les mariages), le kotwal (*kotvāl*, « policier », désigne aujourd'hui un officier de police), le pargana sarrishtdar / qanungo (*parganah sarriśtdār* / *qānūngo*, responsable des archives régionales, le premier terme est tombé en désuétude, le second, d'origine persane, signifie « juriste » en hindi-ourdou moderne), le patwari (*paṭvārī*, responsable des

[33] *dīvānī*, du persan, désigne le bureau et les prérogatives du diwan [*dīvān*] dans sa juridiction. Le diwan désigne à l'origine un hall d'audience, et par extension à la fois un tribunal, une cour et un ministre.

[34] Le Regulating Act de 1773 qui instaure la juridiction précise que la Cour est compétente pour juger l'ensemble des sujets britanniques, sans toutefois définir qui entre dans cette catégorie.

cadastres et des arpenteurs, celui qui détient les « *paṭṭī* », titres de propriété conservés sur de longs feuillets), l'amil (percepteur), le katib (*kātib*, scribe), le maulvi ou pandit (*maulvī* ou *paṇḍit*, « instituteur », à l'origine lettrés religieux respectivement musulman et hindou), auxquels on peut ajouter le munshi (*munśī*), scribe maîtrisant le persan, alors langue officielle de la cour des Moghols.

En 1781, une madrasa[35] ouvre sous l'égide de la Company afin de former du personnel juriste de tradition musulmane et d'expression persane, tandis que le droit coutumier local pénètre à la Cour Suprême. John Williams, juge à la Cour Suprême de Calcutta, réunit une cinquantaine d'érudits tant hindous que musulmans pour rédiger des écrits juridiques dans les langues locales d'administration (persan, arabe et sanskrit) et traduire des textes juridiques locaux en anglais. Les Britanniques s'intéressent en effet aux manuels d'administration légués par les Moghols, tel que l'Ain-i-Akbari (*ʿain-i-akbarī*), datant de la seconde moitié du XVI[e] siècle, pour y puiser des modèles pour leur nouvelle administration locale.

Dans cette même perspective, comprenant que le droit de leur métropole ne sera pas transposable tel quel dans la nouvelle colonie, les Britanniques cherchent à compiler les pratiques juridiques en vigueur dans leurs juridictions afin de créer des codes juridiques adaptés, au risque de figer un droit jusque là largement consensuel[36]. Des codes de loi ont ainsi été créés pour les différentes confessions : l'Anglo-Muhammadan law (loi anglo-musulmane) à partir de ce qui était perçu comme le droit musulman, et l' Anglo-Hindu law (loi anglo-hindoue) issue d'une interprétation similaire de textes hindous supposés être des canons juridiques. C'est à cet effet que les Britanniques ont constitué le Sanskrit College de Bénarès dès 1772, puis celui de Calcutta, pour rassembler des pandits spécialistes des Dharmashastras *dharmaśāstra* et codifier le droit hindou à partir des pratiques d'une minorité dominante de la population, les brahmanes (Duncan et Derett, 1999, p.321).

Les Britanniques introduisent également leur pratique de la common law et de l'equity qui se réfèrent respectivement à la jurisprudence et aux lois édictées par le Gouvernement (Setalvad, 1966). En matière pénale, ils rédigent un code dès 1860 qui

[35]École d'études religieuses islamiques pouvant s'ouvrir à d'autres sujets que l'exégèse des textes religieux.

[36]En 1772, Warren Hastings, administrateur du Bengale puis premier Gouverneur général, fut le premier à entamer la compilation d'un droit hindou (Derrett et Duncan, 1999, p. 225).

constitue toujours la base des codes pénaux indien et pakistanais actuels[37].

Dans les autres colonies européennes, les codes juridiques appliquent le plus souvent le droit de la métropole (Portugal, France) en laissant une place au droit coutumier local (Annoussamy, 2001)[38].

Le système pénal actuel en Inde et au Pakistan tire son origine du code pénal rédigé sous la colonisation britannique, lors de l'époque du Raj. La première version du code pénal s'appliquait aux territoires sous administration britannique (ce qui n'était pas le cas de tout le territoire du sous-continent, constitué alors de centaines de royaumes plus ou moins indépendants du régime britannique). Auparavant, les affaires pénales étaient traitées en fonction du droit local et coutumier. La mainmise de la Grande-Bretagne sur l'économie puis l'administration des territoires du sous-continent amèneront les colons à réfléchir à leurs interactions avec leurs administrés. Ils devront s'interroger sur la langue à adopter pour administrer la loi et rendre la justice dans un empire composite.

1.1.1.1.3 Les langues de la justice

Dans toute l'histoire du sous-continent indien, avant la période contemporaine, la justice et, plus généralement, l'administration, ont été exercées dans la langue de culture de la cour. Ceux qui occupaient la fonction de conseillers juridiques étaient alors les pandits ou les qazis. Avant l'arrivée des souverains musulmans, c'était en sanskrit que l'on rendait la justice. Du temps des Moghols, on utilisait principalement le persan. Les Britanniques ont utilisé, parallèlement à l'anglais, le persan puis l'ourdou dans les différentes juridictions qu'ils contrôlaient, suscitant la réprobation de ceux qui, comme Malaviya au Bihar, rejetaient ce jargon aussi inintelligible que l'anglais (Montaut, 1997a). Seules certaines cours princières ont utilisé les langues locales (hindi, tamoul, gujarati, etc.) pour noter les procédures. En 1901, dans les provinces de l'Oudh (actuel Uttar Pradesh) et du nord, le hindi et la graphie devanagari (devanāgarī), utilisée pour noter le hindi[39] font leur entrée dans les cours de justice[40].

Globalement, le sujet ordinaire, qui ne connaissait ni la langue du colonisateur, ni celles des cultures classiques, n'avait pas un accès direct à la justice et devait s'en remettre

[37] Auxquels on se référera désormais comme IPC pour Indian Penal Code et PPC pour Pakistan Penal Code.

[38] A Pondichéry, la renonciation permettait au XIXe siècle aux habitants d'être jugés par le seul Code civil français. Ils renonçaient ainsi aux pratiques coutumières. À Goa, on appliquait le code civil portugais.

[39] La devanagari note d'autres langues, comme le sanskrit, le marathi ou le népali.

[40] *Op. cit.*

à des intermédiaires sur lesquels nous reviendrons. Cette opacité du monde juridique pour l'écrasante majorité de la population est une constante qui se retrouve encore aujourd'hui dans la pratique judiciaire.

De part la durée du règne moghol, la langue persane a fortement et durablement influencé la terminologie administrative et judiciaire du sous-continent. Cette tendance a été renforcée par l'emploi continu d'une caste particulière de fonctionnaires-scribes, les munshis, de l'administration moghole à l'administration britannique. Cette imprégnation par la langue persane a tant et si bien réussi que de nombreux termes modernes dans les langues vernaculaires n'ont pas réussi à supplanter les termes persans de l'époque coloniale, voire antérieure. Nous étudierons plus loin (cf. chapitre 2) les termes qui subsistent dans l'usage de la langue judiciaire.

Dans cette sous-section, nous avons vu qu'il n'existait pas dans l'Inde pré-islamique de concepts de loi et de droit tels que se les représente l'Occident ou le droit moderne. Dans l'Inde pré-britannique coexistaient des traditions juridiques variées, fondées sur des religions (brahmanisme et islam) ou sur des coutumes. Les colons britanniques, en apportant leur vision du droit, n'en ont pas moins rejeté les droits locaux. Ils ont recensé et traduits les textes hindous et musulmans, fixant ainsi une vision orthodoxe du droit dans ces traditions. Ils se sont appuyés sur les structures administratives existantes, en particulier celles laissées par les Moghols. En conséquence de la reconnaissance de l'ourdou comme langue d'administration, de nombreux termes judiciaires arabo-persans se sont maintenus pendant la période du Raj et au-delà.

1.1.1.2 Organisation contemporaine du judiciaire

Les sociétés indienne et pakistanaise ont hérité des structures mises en place sous l'Empire britannique. Elles ont également conservé la plupart des textes législatifs, dont les deux codes fondateurs du droit pénal : le Penal Code (1860) et le Criminal Procedure Code (1898). A partir de l'Indépendance en 1947, la partition du sous-continent donne naissance à une autre entité : le Pakistan. Les deux pays entament alors leur histoire en empruntant des voies divergentes. Afin de comprendre comment sont organisés les systèmes judiciaires indien (1.1.1.2.2) et pakistanais (1.1.1.2.4), présentons rapidement l'organisation de chacun des États (1.1.1.2.1 et 1.1.1.2.3), fédéraliste dans les deux cas.

1.1.1.2.1 L'organisation de l'État indien

Aujourd'hui, l'Inde est une fédération démocratique d'États et de territoires où règne un multilinguisme à plusieurs niveaux (1.1.1.2.1). Le système judiciaire n'échappe pas à cette structuration hiérarchique (1.1.1.2.2).

Avant l'Indépendance, le sous-continent indien était morcelé entre des territoires sous différents contrôles. La Couronne britannique (voir carte 1.3) possédait trois Presidencies depuis 1793 (Madras, Bengal et Bombay) et d'autres Provinces. D'autres puissances européennes possédaient quelques comptoirs, comme Yanaon, Chandernagor, Mahé, Karikal et Pondichéry pour la France, ou Daman, Diu, Dadra et Haveli pour le Portugal. En outre, plus de 500 États princiers dont la superficie pouvait varier de plusieurs milliers de km^2 (Jammu-et-Cachemire, Nizam d'Hyderabad dans le Deccan) à quelques dizaines seulement (Junagadh, Kathiawar, etc., (Markovits, 1994, pp.478-486)) se partageaient le territoire restant. Dans chacune de ces entités étatiques était appliqué un droit particulier.

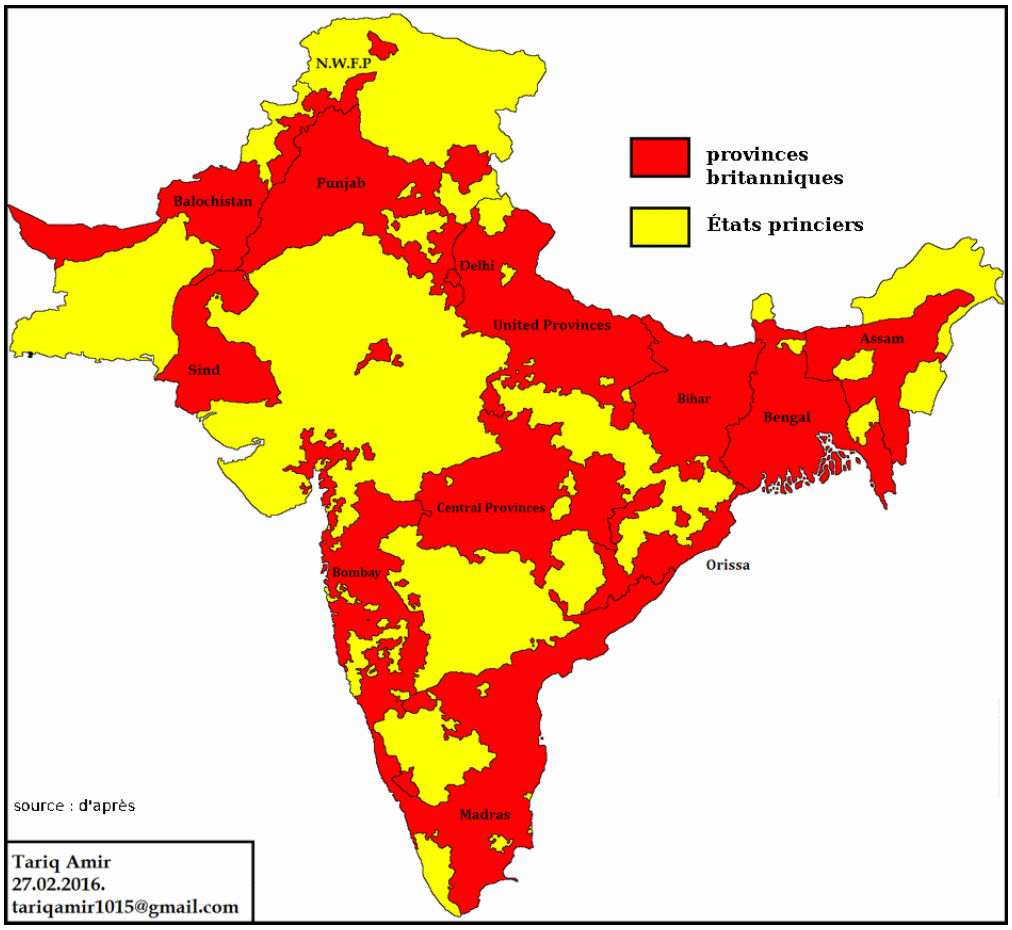

FIG. 1.3: L'Asie du Sud avant 1947. Source : https://pakgeotagging.blogspot.com/2016/03/princely-states-of-british-indian.html

Aujourd'hui, l'Inde indépendante, intitulée officiellement Union indienne, est une

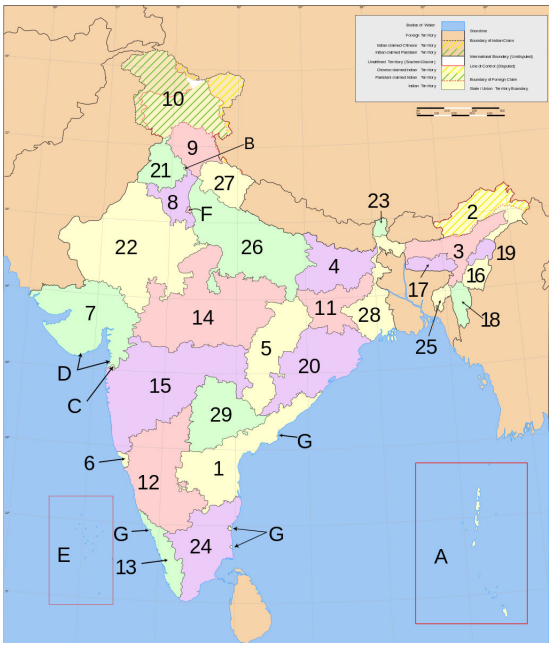

FIG. 1.4: Carte des États et territoires indiens. Voir les notes précédentes pour les noms des États et territoires de l'Union. Source : Wikipedia, *States of India numbered*, 2016.

république fédérale démocratique, dotée d'une constitution depuis 1950 et formée de 29 États[41] créés sur des critères linguistiques et ethniques [42] [43], ainsi que de sept territoires fédéraux administrés par le Gouvernement central qui siège à New Delhi[44]. Ces territoires sont soit d'anciens comptoirs coloniaux cités plus haut, soit des lieux à l'importance historique ou politique bien particulière (Delhi, capitale fédérale, Chandigarh, capitale à la fois du Pendjab et du Haryana, les îles Andaman et Nicobar, à l'importance stratégique). Au sein de cette fédération, l'État de Jammu-et-Cachemire conserve un statut juridique particulier assuré par l'article 370 de la Constitution [45], régulièrement contesté.

[41] Liste officielle des États : Andhra Pradesh (1), Arunachal Pradesh (2), Assam (3), Bihar (4), Chhatisgarh 5), Goa (6), Gujarat (7), Haryana (8), Himachal Pradesh (9), Jammu et Cachemire (10), Jharkhand (11), Karnataka (12), Kerala (13), Madhya Pradesh (14), Maharashtra (15), Manipur (16), Meghalaya (17), Mizoram (18), Nagaland (19), Odisha (ex-Orissa) (20), Pendjab (21), Rajasthan (22), Sikkim (23), Tamil Nadu (24), Telangana (25), Tripura (26), Uttarakhand (27), Uttar Pradesh (28) et Bengale occidental (29), d'après le portail internet du gouvernement indien :`http://goidirectory.nic.in/state.php`..

[42] (Markovits, 1994, p.332).

[43] Bien qu'aucun État ne soit aujourd'hui monolingue ou mono-ethnique, ni même la moitié des districts du pays (Montaut, 2012).

[44] Liste des territoires de l'Union : Andaman et Nicobar (A), Chandigarh (B), Dadra et Nagar Haveli (C), Daman et Diu (D), îles Laquedives (E), Delhi (F), Puducherry (ex-Pondichéry) (G), site cf. note précédente.

[45] Jusqu'en 1965, le Jammu-et-Cachemire a bénéficié de sa propre constitution et le chef de l'exécutif, le Sadr-e-Riyasat, y avait un statut de Premier ministre (Ganguly, 1997).

Sur les plans législatif et exécutif, le gouvernement indien possède deux chambres au niveau central, la *lok sabhā* (chambre du Peuple) et la *rājya sabhā* (chambre des états), et une assemblée législative au niveau de chaque État, la *vidhān sabhā* . En fonction des domaines, il revient au pouvoir central ou aux États d'élaborer les lois. La langue première de rédaction des lois est l'anglais. Les textes sont ensuite traduits vers le hindi, puis vers les autres langues.

1.1.1.2.2 Le système judiciaire indien

Le code pénal indien (*daṇḍ saṁhitā*, littéralement « peine-code »), comme précisé plus haut, a gardé les bases de la version de 1860 et a été amendé plusieurs fois. La version sur laquelle s'appuie cette étude date de 2013[46]. Le code de procédure pénale (Criminal Procedure Code, *prakriyā daṇḍ samhitā*), lui, date de 1973 et s'applique à tout le territoire indien à l'exception du Jammu-et-Cachemire (voir le préambule du CrPC). Il a été révisé en 1998[47].

Les juridictions indiennes, civiles (civil juridictions / *sivil adhikāritā*) ou pénales (criminal juridictions / *dāṇḍik adhikāritā*), s'organisent de manière pyramidale. Au premier niveau, les tribunaux inférieurs (1.1.1.2.2) regroupent un système judiciaire classique et un système traditionnel. Les affaires pénales peuvent être traitées par le tribunal du magistrat (Magistrate's Court / *majisṭreṭ nyāyālaya*), le Nyaya Panchayat rural (*nyāya pañcāyat*), (1.1.1.2.2) ou la cour d'assise (Sessions Court /*seśan korṭ*, 1.1.1.2.2), selon leur gravité. Au niveau de l'État se trouve la Haute Cour (High Court - *ucca nyāyālaya*, 1.1.1.2.2). Enfin, au niveau de l'Union indienne se trouve la Cour suprême (*sarvocca nyāyālaya* ou *uccatam nyāyālaya*). Les cours pénales fonctionnent selon un modèle institué par le Criminal Procedure Code (CrPC) de 1898.

a. Tribunaux inférieurs

Chaque État est divisé en districts qui comportent un certain nombre de magistrats judiciaires répartis en différentes catégories (art. 11 du CrPC). Ces magistrats ont chacun leur tribunal (magistrate court / *majisṭreṭ nyāyālaya*).

[46]Pour le hindi et l'anglais seulement. La mise à jour en ourdou est plus ancienne.

[47]La version utilisée dans cette étude a été mise à jour au 31 décembre 2018, cf. `https://indiacode.nic.in/handle/123456789/1611?view_type=search&sam_handle=123456789/1362`.

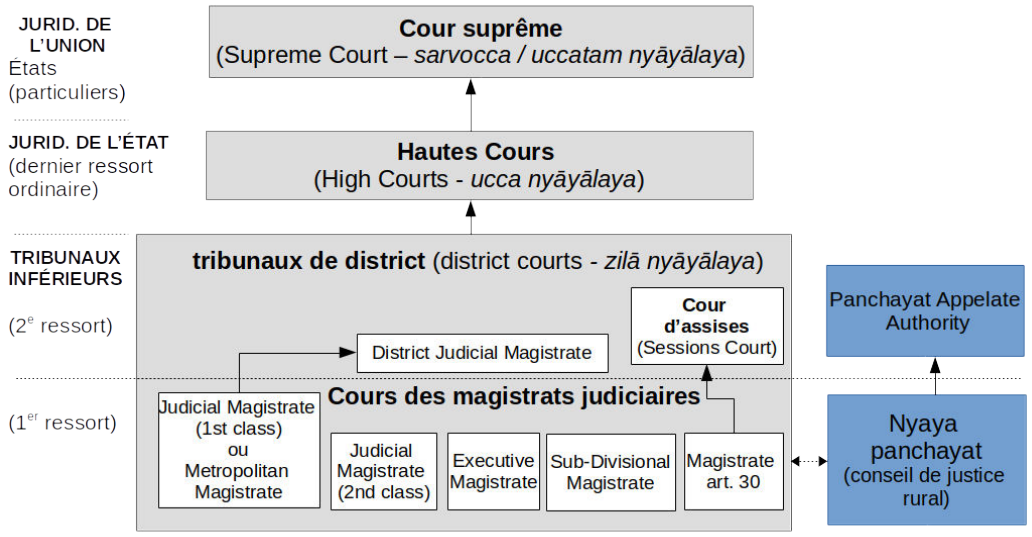

FIG. 1.5: Organisation de la justice pénale en Inde

a.1. Magistrats judiciaires de 3e classe (third class judicial magistrates / *ṭṛtīya varg nyāyik majisṭreṭ*)

Ils connaissent des affaires pénales les plus légères. A ce titre, les peines qu'ils prononcent ne peuvent excéder un mois d'emprisonnement.

a.2. Magistrats judiciaires de 2e classe (second class judicial magistrates/ *dvitīya varg nyāyik majisṭreṭ*)

Ils jugent les affaires passibles d'une peine d'emprisonnement allant jusqu'à six mois et d'une amende jusqu'à 200 roupies.

a.3. Magistrats judiciaires de 1re classe (first class judicial magistrates / *pratham varg nyāyik majisṭreṭ*)

Ils peuvent prononcer des peines allant jusqu'à trois ans d'emprisonnement et 10 000 roupies d'amende. Dans les métropoles, ces magistrats sont remplacés par les Metropolitan Magistrates (*mahānagar majisṭreṭ*).

a.4. « Section 30 magistrate » (*dhārā tīs majisṭreṭ*)

Le gouvernement d'un État peut conférer au magistrat de première classe de son choix le pouvoir de prononcer des peines allant au-delà de sept années de prison (à l'exception de la peine capitale) par le truchement de l'article 30 du CrPC.

a.5. Magistrat judiciaire de district (district judicial magistrate / *nyāyik zilā majisṭreṭ*)

Il est responsable de toute la magistrature à l'échelle du district. Les magistrats de district adjoints sont, comme lui, des magistrats de première classe. Tous deux ont pouvoir d'interjeter appel (de formuler un appel) et d'entendre en appel des litiges traités par des magistrats de 2e ou 3e classe.

a.6. Magistrat de sous-division (sub-divisional magistrate / *upkhaṇḍ majisṭreṭ*)

Peuvent être nommés à ce poste des magistrats de première ou seconde classe.

a.7. Magistrat exécutif (executive magistrate / *kāryapālak majisṭreṭ*)

Il est chargé de faire appliquer la législation et possède des fonctions judiciaires mineures décrites dans les articles 107 à 110 du CrPC. Entre autres, il entend les individus arrêtés sur ordre d'une autre juridiction que la sienne et peut empêcher l'entrée et la circulation de certains individus sur son territoire. Il n'a pas le pouvoir de juger.

Le système des cours de magistrats judiciaires correspond approximativement aux tribunaux de police et correctionnels français.

a.8. La cour d'assises (Session Court / *seśan korṭ*)

On en trouve une dans chaque district. Cette cour traite les affaires criminelles les plus graves : vol à main armée ou en réunion, homicide, vol aggravé en récidive, etc. Le juge des assises a le pouvoir de prononcer toute peine prévue au code pénale. Cependant une décision d'appliquer la peine capitale doit toujours être confirmée par la Haute Cour.

Le juge des assises peut aussi entendre en appel des affaires traitées en première instance par un magistrat de district ou investi de pouvoirs supplémentaires au titre de l'article 30 du CrPC. La cour d'assises se trouve dans l'enceinte du tribunal de district.

a.9. Les Nyaya Panchayats (*nyāya pañcāyat*)

Au niveau des tribunaux inférieurs, il existe un système de justice alternatif, issu d'une institution traditionnelle dans le sous-continent : le *panchayat* (pañcāyat).

A l'origine, un panchayat est constitué de cinq personnes chargées de juger les conflits et les litiges administratifs ou judiciaires (cf. 1.1.1.1.1). Il existait des panchayats

pour chaque caste. Par la suite, il en a été établi au niveau des villages. L'institution du panchayat a survécu à la période du Raj, où elle était promue par les indépendantistes, à la tête desquels Gandhi, pour son mode d'administration indigène.

À l'indépendance, le système des panchayats a été ravivé. En 2009, une loi appelée *Nyaya Panchayat Bill* définit les compétences des Nyaya Panchayats (ou panchayat de justice), qui offrent en milieu rural une alternative au système judiciaire classique engorgé.

Les Nyaya Panchayats sont des instances élues au niveau du village ou du groupement de villages. Ils peuvent traiter des affaires civiles et pénales qui se présentent dans leur zone géographique. Bien que les affaires criminelles les plus graves ne soient pas de leur ressort, ils peuvent traiter des infractions concernant plus d'une quarantaine d'articles du Code pénal. Une affaire instruite par une juridiction classique inférieure (tribunaux des magistrats pour les affaires pénales) peut leur être transférée à toute étape de la procédure.

En cas de contestation de la décision, les justiciables peuvent interjeter appel au niveau du district auprès du *Panchayat Appelate Authority* ou de l'*Ombudsman* (médiateur).

Le Nyaya Panchayat privilégie la résolution des conflits par la conciliation. Il peut prononcer des sanctions sous forme d'amendes ou enjoindre au paiement de compensations, mais il ne peut ordonner de peines de prison. L'article 26 de la loi indique que la procédure doit se dérouler dans la langue habituellement utilisée dans la localité et en particulier par les parties en présence.

b. La Haute Cour (High Court / *ucca nyāyālaya)*

Cette juridiction traite les appels, les révisions et a un rôle consultatif. Elle reçoit les appels des décisions des cour d'assises ou des magistrats (section 30) dans certains cas. La juridiction a en outre le pouvoir de confirmer une peine capitale. C'est également le dernier ressort en appel pour les affaires criminelles, sauf dans de rares cas pouvant aller jusqu'à la Cour suprême.

Actuellement, 25 hautes cours traitent les affaires des 29 États et des sept territoires avec à leur tête un *Chief Justice* désigné par le président après consultation du président de la Cour suprême (Chief Justice of India).

c. La Cour suprême (Supreme Court, *sarvocca nyāyālaya* ou *uccatam nyāyālaya*)

C'est la juridiction la plus élevée du pays. Elle traite les litiges entre les États ou

entre les États et le Centre, et peut à de rares occasions constituer une juridiction d'appel pour des affaires criminelles graves. Récemment, sa compétence a été étendue aux litiges d'intérêt public (Public Interest Litigation), pour lesquels tout individu ou groupe peut introduire une requête. La Cour suprême joue également le rôle de cour de cassation et vérifie la conformité des lois à la Constitution. Composée de 25 juges, elle est présidée par le Chief Justice of India[48].

Le système judiciaire indien comprend une voie classique, héritée du système instauré sous le Raj, et une vois alternative, sous la forme de conseils pour la résolution des litiges pénaux les moins graves. Les tribunaux inférieurs, dirigés par des magistrats, traitent les affaires en première instance. Ils sont placés sous l'autorité d'un magistrat de district. Des voies de recours existent dans ces tribunaux ou auprès de juridictions d'appel (High Court, Supreme Court). Dans la procédure classique, les juges sont tous professionnels. Il n'existe pas de jury populaire comme en France. Le système indien est resté entièrement laïque, contrairement au système pakistanais dont la description va suivre.

1.1.1.2.3 L'organisation de l'État pakistanais

Le Pakistan s'est constitué en une fédération de provinces et de territoires. Le rapport du Centre envers ces entités est plus vertical que dans l'union indienne. Un des motifs de création du Pakistan étant une revendication religieuse, l'islam joue un rôle, quoique relatif, dans le système judiciaire du pays.

Le Pakistan (fig. 1.6) est une république depuis sa création en 1947[49]. Le pays connaît une histoire politique tourmentée, avec divers épisodes de dictature militaire, et sa Constitution n'est achevée qu'en 1973. Jusqu'à cette date, c'est l'India Government Act adopté en 1935 qui fait office de constitution provisoire. L'Objective Resolution, adoptée en 1949, devait préparer la future Constitution et sert encore de socle aux lois actuelles pakistanaises.

Sur le plan exécutif, l'État pakistanais est doté aujourd'hui d'un Parlement fédéral (*majlis-e-śūrā*) situé à Islamabad, qui comprend une Assemblée nationale (*qaumī asamblī*) et un Sénat (*aiwān-e-bālā*). Chacune des quatre provinces (Pendjab, Sindh, Balouchistan et Khyber Pakhtunkhwa) est représentée par un Gouverneur et une Assemblée provinciale

[48]Source : `http://supremecourtofindia.nic.in`.

[49]Nous ne traiterons pas dans ce paragraphe du Pakistan occidental, futur Bangladesh.

FIG. 1.6: Provinces et territoires du Pakistan

(*subāī asamblī*). Les autres territoires administratifs (Zones tribales, Azad Jammu-Cachemire, Gilgit-Baltistan) sont administrés selon différents systèmes[50]. Au niveau fédéral, l'ourdou et l'anglais sont les langues officielles, ainsi que dans les parlements provinciaux, alors que dans chaque province, une autre langue que l'anglais ou l'ourdou domine numériquement.

1.1.1.2.4 Le système judiciaire pakistanais

Le système judiciaire pakistanais (fig. 1.7) partage avec son homologue indien une organisation héritée de codes juridiques rédigés sous la période coloniale. Dans chaque pays, ces codes ont été amendés afin de suivre les évolutions des jeunes républiques. Ainsi, au Pakistan, une dimension religieuse du droit a été introduite, qui concerne également une partie de la branche pénale. Comme son homologue indien, le système judiciaire pakistanais se caractérise par un engorgement (on ne compte que 4200 juges pour 180 millions d'habitants en 2013) et des délais de procédure très longs[51].

[50] L'Azad Jammu-Cachemire, considéré comme un État de la fédération, possède une Assemblée législative, il a à sa tête un président et un premier ministre. Le territoire de Gilgit-Baltistan, lui aussi, possède sa propre assemblée législative depuis 2009 et bénéficie d'un statut d'autonomie avancée. En revanche, les Zones tribales (FATA, ou Federally Administrated Tribal Areas), sont représentées au Parlement fédéral mais ne possèdent pas d'assemblée législative.

[51] cf. Hussain, 2015 et Chemin, 2009, p.23.

a. Évolution post-indépendance et islamisation du droit

A l'Indépendance, la Constitution provisoire se base sur le Government of India Act de 1935. Les trois Hautes Cours (High Courts) du Pakistan déjà présentes avant 1947 continuent de fonctionner dans le prolongement de leurs prérogatives et une autre cour est instaurée à Dacca au Pakistan oriental. Différents amendements dans le cadre des multiples versions de la Constitution viendront modifier légèrement les pouvoirs des tribunaux supérieurs, de manière non notable. Ainsi, la Federal Court (Cour fédérale) devient la Supreme Court (Cour suprême) en vertu de la Constitution de 1973.

Un changement notoire intervient en 1979-1980, avec l'introduction du droit islamique, notamment sous la forme des Ordonnances de Hudood (Hudood Ordinances), également appelées Lois de Hudood[52]. Sous le gouvernement du général Zia ul Haq, dont la volonté est d'islamiser davantage le pays, un Tribunal Fédéral de la Charia (Federal Shariat Court) est institué. Ce tribunal religieux fondé sur le droit islamique est instauré en vertu de l'article 203-B de la Constitution afin de vérifier si les lois en vigueur sont en accord avec les principes de l'islam (art. 203-D).

La loi pénale islamique avait disparu pendant la présence du législateur colonial et les avocats du projet de création d'un État pakistanais, Ali Jinnah en tête, ne souhaitaient pas y voir l'instauration d'une république religieuse. Cependant, la question de l'islamisation du jeune État est posée dès l'introduction de la Résolution des Objectifs (Objectives Resolution) en 1949, après la mort de Jinnah :

> Sovereignty over the entire universe belongs to Allah Almighty alone and the authority which He has delegated to the state of Pakistan, through its people for being exercised within the limits prescribed by Him is a sacred trust[53].

Texte qui sera inséré dans les trois versions de la Constitution (1956, 1962 et 1973).

a.1. Les lois de Hudood (hudūd)

Le courant d'islamisation du pays prend de l'ampleur à la fin des années 1970. Un programme d'islamisation de la loi prévoit officiellement de remplacer les lois « d'origine

[52] Au cours des dernières évolutions politiques, le gouvernement est partiellement revenu sur l'application de ces lois, jugée discriminatoires à plusieurs titres, cf. suite.

[53] Traduction : « *La souveraineté sur l'univers tout entier appartient à Allah Tout Puissant et l'autorité qu'il a déléguée à l'État du Pakistan, par le truchement de son peuple, afin qu'elle soit exercée dans les limites qu'Il a prescrites, est un bien sacré.* » Source : texte de l'Objectives Resolution (1949), `http://www.pakistani.org/pakistan/constitution/annex.html`.

occidentale » par des lois fondées sur des sources islamiques (Wasti, 2009, p.3) et le Président de la République déclare pour la première fois que l'islam est le fondement de la raison d'être du Pakistan.

À son arrivée au pouvoir, le général Zia ul-Haq fait de l'islamisation de la société pakistanaise sa priorité. Les cinq Ordonnances de Hudood sont adoptées en 1979, elles concernent alors cinq types d'infraction : l'adultère (*zinā'*), la fausse accusation d'adultère (*qazf)*, le vol (*sariqah*), le vol à la tire (*harabah*)[54] et la consommation d'alcool (*khamr*). Le terme *hudūd* est le pluriel arabe de *hadd*, il désigne les catégories d'infraction pour lesquelles une une peine est prévue par le Coran ou la Sunna comme un droit de Dieu. Aucun individu ou instance humaine ne peut alors annuler ces peines qui peuvent aller de l'administration de coups de fouet à la peine de mort[55].

a.2. Introduction des *qisas* et *diyat*

Dans les années 1980, des religieux, emmenés par Gul Hasan, souhaitent voir appliquer des règles plus proches du canon juridique islamique. Ils contestent les dispositions du Code Pénal de 1860 et du Code de procédure pénale de 1898, les jugeant contraires à l'islam. Il n'était pas possible, dans le cadre de ces textes, de demander par exemple le prix du sang (*diyat*) ou le pardon pour les héritiers ou les descendants d'une victime d'homicide (qisas).

Le *qiṣāṣ* et la *diyat* constituent deux autres concepts du droit islamique (cf. 1.1.1.1.1). Ils concernent les homicides[56], les atteintes aux personnes ou aux biens.

- le *qiṣāṣ* désigne les représailles, ou l'égalité dans les représailles,
- la *diyat* représente une compensation ou « prix du sang ».

Ces principes ont été adoptés en droit pénal en 1997 (Criminal Law Act). Dans les deux cas, la conciliation ou *ṣulah* est recherchée et le pardon constitue une alternative au châtiment.

b. Juridictions pénales au Pakistan

Pour le reste, la justice reste organisée en trois niveaux, comme dans le système indien : les tribunaux spécialisés, de district et les cours d'assises (1.1.1.2.4), les Hautes Cours (High Courts, 1.1.1.2.4) et la Cour suprême (Supreme Court,1.1.1.2.4) (cf. fig. 1.7).

[54]Les peines pour vol n'ont jamais été appliquées.

[55](Wasti, 2009, p. 33).

[56]Certains homicides sont exclus du *qiṣāṣ*, cf. art. 306 du code pénal pakistanais.

À cela s'ajoute un réseau de juridictions islamiques chargé de contrôler la conformité du droit aux principes de l'islam (1.1.1.2.4).

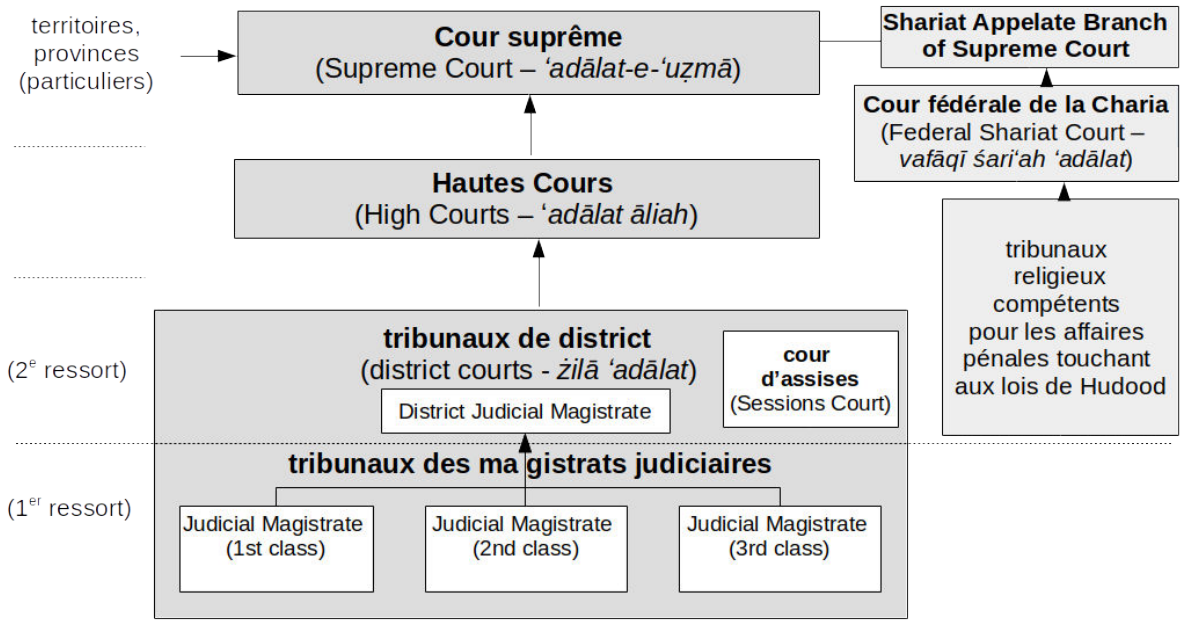

FIG. 1.7: La justice pénale pakistanaise. Adapté de Hussain (2015).

b.1. Les tribunaux inférieurs

Ils se subdivisent en juridictions civiles, fondées dans le cadre de la Civil Courts Ordinance de 1962, et en juridictions pénales (criminal courts) se référant au Code de procédure pénale de 1898. En marge de ce système se trouvent aussi des tribunaux spéciaux établis en vertu de lois spéciales.

Comme en Inde, les cours pénales se composent de Judicial Magistrates Class I, II, III (magistrats judiciaires de 1re, 2e ou 3e catégorie) traitant d'infractions plus ou moins graves, d'un Sessions Judge (juge des assises), et d'Additional Sessions Judges (juges adjoints des assises).

b.1.1. Magistrat de 3e classe (Magistrate 3rd Class / *soam darjah majiṣ-reṭ*)

Il peut prononcer des peines allant jusqu'à un mois d'emprisonnement et 3 000 roupies d'amende.

b.1.2. Magistrat de 2e classe (Magistrate 2nd Class / *doam majisṭreṭ*)

Peut prononcer des peines allant jusqu'à un an d'emprisonnement et 15 000 roupies d'amende.

b.1.3. Magistrat de 1re classe (Magistrate 1st Class / *avval darjah majisṭ-reṭ*)

Juge des infractions passibles de trois ans d'emprisonnement au maximum et prononce des amendes allant jusqu'à 45 000 roupies.

b.1.4. Le juge de district, d'assises et son assesseur (District and Sessions Judge / Additional District and Sessions Judge)

En matière pénale, ils sont chargés des affaires criminelles les plus graves (meurtre, terrorisme, trafic de drogue lorsque la peine de mort est encourue, toutes les affaires pénales du ressort des lois de Huddood, cf. 1.1.1.2.4). Ils reçoivent également les appels et les demandes en révision émanant des tribunaux des magistrats inférieurs.

Il existe encore d'autres tribunaux spécialisés dans certaines infractions : corruption, affaires de narcotiques, délinquance financière, droits des consommateurs, etc.

b.2. La Haute Cour (High Court / *adālat āʿliah*)

Chaque province possède une Haute Cour. Il en existe une supplémentaire à Islamabad, la capitale fédérale. Chaque Haute Cour est présidée par un juge (le Chief Justice) et compte un nombre variable de juges en fonction de la population de la province.

Les prérogatives des Hautes Cours sont les questions touchant aux Droits Fondamentaux et les appels des affaires civiles ou pénales suivant les jugements des juridictions inférieures. Chaque Haute Cour possède aussi un Tribunal de la Charia (Shariat Bench), formation de jugement islamique traitant certaines affaires relevant des ordonnances de Hudood (cf. 1.1.1.2.4).

b.3. La Cour suprême (Supreme Court / *adālat-e-ʿuẓmā*)

Comme en Inde, elle représente le plus haut degré de juridiction du pays. Elle possède également une fonction d'appel et d'arbitrage des questions de droit (articles 184, 185, 186 de la Constitution). Elle règle les différends entre gouvernements provinciaux, entre les provinces et le gouvernement fédéral, connaît des infractions aux Droits Fondamentaux

et des questions « d'intérêt public » (article 184(3)). C'est une juridiction d'appel pour les affaires privées comme pénales qui traite aussi les appels des décisions de la Cour de la Charia (Shariat Court, cf. 1.1.1.2.4). La Cour est composée d'un Chief Justice (président) et de 16 autres juges nommés par le Président selon la procédure indiquée dans la Constitution (art 175).

b.4. La Cour fédérale de la Charia (Federal Shariat Court / vafāqī śariʿah ʿadālat)

Cette juridiction, instituée dans le cadre de l'article 203 de la Constitution, est composée de huit juges musulmans dont le Président (Chief Justice), désignés par des membres anciens ou actuels de la Cour suprême, en consultation avec le Ministre fédéral de la Loi et de la Justice. Trois de ces juges doivent être des oulémas, érudits en droit islamique.

Tout citoyen ou gouvernement (fédéral comme provincial) peut saisir la Cour et lui demander d'examiner un point de loi pour juger de sa conformité ou non aux injonctions islamiques (article 203-D). Tout appel des décisions de cette Cour se fait à la Cour suprême auprès du Shariat Appelate Bench où siègent trois juges musulmans dont deux oulémas au plus. Dans le cas où un point de loi ne serait pas jugé conforme à l'islam, le Gouvernement doit amender la loi. Autre point important, la Cour traite en appel les affaires criminelles visées par les ordonnances de Hudood (article 203-D).

Cette cour fédérale religieuse, ainsi que les tribunaux subalternes qui en dépendent, ne sont pas des juridictions originelles. Elles ne peuvent être saisies en premier lieu dans une affaire pénale. La Cour fédérale de la charia est controversée depuis sa création en 1980. D'une part, elle vient doubler les prérogatives des juridictions supérieures non religieuses. D'autre part, son indépendance vis-à-vis de l'exécutif n'est pas garantie. Mais par dessus tout, c'est sa validation des lois de Hudood (châtiments islamiques) et son interférence avec les droits des femmes qui fait débat. Certains auteurs, dont Wasti (2009), tendent même à démontrer que l'institution de ce système religieux a contribué à la dégradation du système judiciaire du pays. Depuis 2006 et la promulgation du Protection of Women (Criminal Laws Amendment), ses prérogatives ont été restreintes. On lui a retiré notamment les affaires touchant au *zinā*.

c. Systèmes judiciaires alternatifs

Devant l'ampleur des délais (parfois plusieurs dizaines d'années), de nombreux

Pakistanais se tournent vers des solutions alternatives pour résoudre leurs litiges judiciaires. Dans les zones tribales (FATA), l'institution des « *jirgah*-s » est très puissante et fleurit là où les institutions étatiques font défaut. Les membres qui constituent ces assemblées sont les aînés (les « barbes blanches ») et ne sont pas élus. Les femmes n'y sont pas représentées.

Pour les infractions répertoriées au Criminal Procedure Code (CrPC) comme « compoundable »[57], il est possible de recourir, en dehors de tribunaux, au *ṣulah* (« compromis »), forme de consensus entre personnes privées pour régler les litiges, et de faire reconnaître cette décision par des tribunaux classiques.

Plus de 70 ans d'indépendance n'ont pas foncièrement modifié les systèmes judiciaires indiens et pakistanais, issus d'une origine commune, malgré des parcours politiques assez différents. À l'exception des juridictions religieuses au Pakistan, dont les prérogatives restent relativement limitées dans le pénal (aucune affaire n'y est traitée en première instance), les deux systèmes restent organisés de manière similaire, avec trois degrés de juridiction et une cour spécialisée dans les litiges entre les provinces ou les États. Tous les deux connaissent des modes de justice alternative plus ou moins officiels, qu'il s'agisse du *ṣulah* ou de la *jirgah* au Pakistan, ou des Nyaya Panchayats en Inde. Ces modes fondés sur le consensus et la proximité des citoyens pallient les déficiences d'un système classique jugé trop lent, corrompu et distant.

1.1.2 *Présentation de la justice en France*

La tradition juridique française (1.1.2.1) s'appuie avant-tout sur des racines grecques et latines, tout comme les concepts de droit, loi ou justice (1.1.2.1.2). Le droit coutumier forme également une source juridique. Ces sources ont fusionné et se sont affranchies de la tutelle religieuse pour donner naissance au droit pénal contemporain (1.1.2.1.3).

1.1.2.1 Tradition judiciaire en France

1.1.2.1.1 Sources du droit français

Les racines du droit français plongent dans la culture grecque qui a lui transmis les notions de « droit », « justice » et de « loi ». Droit et absolu (religieux ou surnaturel) sont alors intimement liés. Ils ne se sépareront qu'à une date très récente.

[57]Livre Major Acts, CrPC.

À l'époque grecque, il existait à côté d'un droit divin (thémis) une justice (díkē) faite de formules régissant les usages. Ces désignations trouvent leur origine dans la mythologie et font coexister un droit surnaturel et un droit dont l'humain est le co-créateur et l'interprète. Les philosophes grecs inscriront le droit davantage dans le monde humain, avec la notion de « nomos » : d'abord « ordre universel s'imposant aux astres comme aux cités », puis « lois propres aux hommes envisagées dans le cadre politique de la cité » (Rouland, 1988, p.76).

Ces notions continuent de s'éloigner de l'absolu religieux sous le droit romain, sans jamais s'en affranchir totalement. Aux fondements de Rome, un collège de prêtres détient le monopole de la connaissance du droit. Les formules qui codifient les procédures restent pendant longtemps cachées au public (Rouland, 1988, p.81). Peu à peu, le droit passe aux mains des aristocrates. Suite à la pression de la plèbe qui demande des lois écrites accessibles à tous, douze tables de loi sont rédigées en -450. Il s'agit de la première édification de droit privé depuis la fondation de Rome. Elle pose les bases des règles de procédure, du droit de la famille, du droit des biens et du droit pénal. La société romaine connaît les concepts de *juge*, de *juriste* et les avis de ces derniers sont consignés dans des compilations de *jurisprudence*. Ils peuvent être retenus ou écartés par un juge (Rouland, 1988, p.89). Comme dans le droit islamique, les juristes fondent leurs décisions sur la déduction, la comparaison ou l'analogie. On interprète aussi le droit déjà existant.

Le *Corpus juris civilis* de Justinien, élaboré à Byzance au VI^{e} siècle, constituera une autre œuvre juridique fondatrice, dans laquelle l'influence chrétienne est prépondérante. Ces fondements tomberont dans l'oubli lors des invasions de l'Europe occidentale qui marqueront la chute de Rome.

À côté du droit officiel, la coutume, plus pragmatique, est reconnue comme source juridique. Les droits et coutumes locaux perdureront dans tout l'Empire romain et parfois l'influenceront.

Au Moyen-Âge, on redécouvre le droit justinien au XI^{e} siècle à Bologne. Cet épisode marque un tournant de l'évolution du droit et la constitution d'une tradition dite romano-canonique, rencontre du droit romain détaché du religieux et d'une Église en quête de règles d'organisation. Un système institutionnel pyramidal est alors mis en place avec à son sommet un Pontife. C'est également à cette période que naît progressivement la terminologie de base du droit en français.

Sous l'Ancien Régime, le territoire français actuel était culturellement divisé entre Nord et Sud. Au Sud, le droit était écrit, sous l'influence du droit romain. Au Nord, on

pratiquait un droit coutumier, oral, d'influence germanique. Les coutumes furent rédigées progressivement entre les XV^e^ et XIX^e^ siècles et unifiées grâce au droit romain qui sert de base pour un droit commun. A l'aube du XIX^e^ siècle, l'Assemblée constituante née de la Révolution avait pour projet de poursuivre cette unification. Celle-ci sera réalisée par Napoléon Bonaparte qui fit rédiger en 1804 le Code civil, connu aussi sous le titre de Code Napoléon. Le XX^e^ siècle voit s'opérer une accélération de la production de lois avec l'élaboration de différents codes (du travail, des affaires sociales, etc) au fur et à mesure que l'État prend de la place dans de nouveaux domaines.

1.1.2.1.2 Notions de droit, loi, justice

Au fil de l'histoire française du droit se sont développés les concepts de droit, loi et justice, dont les traits notionnels sont assez différents de ceux des concepts indiens, avant leur rencontre avec les droits européens.

a. Droit

Deux couples de notions liées sont à analyser ici. En grec, on distingue deux types de droit : *thémis* et *díkē* (qui ont pour équivalents latins *fas* et *jus*). *Thémis* vient de la racine indo-européenne **dhe* qui signifie « poser, établir », tout comme *dharma* et *vidhi* en sanskrit. Il désigne le droit d'origine divine qui s'applique à la famille, au *genos*, un droit non écrit qui fixe chez le roi la conduite à tenir chaque fois que l'ordre du *genos* est menacé (Benveniste, 1969, p.99). Il correspond au neutre indéclinable *fas* en latin qui désigne « l'énonciation solennelle, d'origine divine ».

Díkē, « l'usage, la manière d'être », vient de la racine indo-européenne **dikei*, signifiant « montrer ce que l'on doit faire, prescrire la norme » (*jus*, *juris* en latin : « le droit prescrit par la coutume[58] »). Elle désigne la justice entre les familles, matérialisée par des formules appliquées par un juge. Progressivement, ces deux notions se sont rapprochées.

Le bas latin *directum*, à rapprocher du latin *rex* « celui qui a autorité pour tracer les emplacements des villes et déterminer les règles de droit[59] », qui a fait souche dans les langues romanes pour donner en français « droit », évoque la rectitude et par conséquent la règle de droit. On retrouve cette notion en italien et en allemand dans les termes

[58]Benveniste (1969), p.110.

[59]Cf. Benveniste (1969), p.9.

diritto et *Recht*, bien différenciés du terme « loi » (legge, Gesetz) qui, comme en français, désigne « ce qui est posé »[60].

En français, le lexème « droit » fait référence à différentes notions influencées par les cultures grecque et romaine antiques puis judéo-chrétienne. D'un point de vue juridique, le droit est en premier lieu une norme venant régir le comportement des hommes en société. Il existe d'autres systèmes normatifs : la morale, la religion, la déontologie, l'éthique. Cependant de nos jours, le droit est devenu le seul mécanisme venant corriger officiellement les dysfonctionnements de la société.

Dans la culture du droit français, le droit représente un idéal à atteindre et entérine parfois des états de fait (mariage pour tous, PACS), attachant des conséquences juridiques à des situations données, comme dans le cas de la réparation d'un préjudice. Le droit ne peut cependant pas tout dire ni prévoir et c'est pour pallier ce défaut qu'ont été instaurés certains principes généraux du droit[61] :

- **généralité de la règle de droit** : la règle s'adresse à tous ou du moins à des catégories générales. Même les décisions prises à titre individuel découlent de l'application de ces règles générales. Le droit peut toutefois faire preuve de souplesse : si son application stricte conduit à une injustice, alors l'équité adapte au cas particulier.
- **caractère obligatoire du droit** : contrairement à la morale, on peut contraindre à l'application du droit via un tribunal ayant le pouvoir de sanctionner.
- **contenu normatif de la règle de droit** : selon le Conseil constitutionnel, « la loi a pour vocation d'énoncer des règles et doit par suite être revêtue d'une portée normative (2005) ».

b. Loi

Dans son acception large, elle désigne la règle de droit qui s'impose à tous. Elle peut être orale ou écrite. Dans un sens plus restreint, la loi est une règle de droit émanant du Parlement, comme le rappelle l'article 34 de la Constitution : « La loi est votée par le Parlement. »

[60] cf. Fabre-Magnan (2010).

[61] Ne pas juger un fait est d'ailleurs un délit selon le Code civil (art. 4) : « Le juge qui refusera de juger, sous prétexte du silence, de l'obscurité ou de l'insuffisance de la loi, pourra être poursuivi comme coupable de déni de justice. »

c. Justice et judiciaire

Autres notions importantes à définir pour comprendre le concept de droit pénal en France, celles de « justice » et de « judiciaire », dérivés du *jus* latin.

Pour les philosophes grecs antiques, la justice est un ordre naturel des choses visant à l'harmonie sociale (Platon, *la République*, IV), pour le bien commun (Aristote, *la Politique*, livre III, 17). Avec l'avènement du christianisme, l'ordre naturel des philosophes antiques est assimilé à la volonté de Dieu qui surpasse les pouvoirs temporels institués. C'est la religion qui dicte la norme et la morale.

Dans son sens moderne, la justice se rattache aux concepts d'égalité et de respect de l'autre. Dans le dictionnaire Larousse, les sens 1) et 2) de l'entrée justice indiquent : « principe moral qui exige le respect du droit et de l'équité » puis « qualité morale qui invite à respecter les droits d'autrui ». Le mot justice vient du latin *justitia*, où il a le sens de « conformité avec le droit, sentiment d'équité ; esprit d'équité » ; en bas latin, il désigne au singulier la notion abstraite : « le droit, les lois », tandis qu'il prend un sens concret au pluriel : « jugements, préceptes »[62].

L'adjectif « judiciaire », utilisé depuis la fin du XIV^{e} siècle, est un emprunt au latin *judiciarius*, « relatif aux tribunaux ». En français moderne, il désigne ce « qui se rapporte à la Justice, à l'administration de la Justice. »[63]

Ces quelques notions définies, le paragraphe suivant permet de cerner plus précisément ce que représente le droit pénal dans le champ du judiciaire, qui constitue le contexte de notre étude.

1.1.2.1.3 Droit pénal

Tel qu'il sera défini ci-après, le droit pénal est un concept européen qui sera appliqué aux droits indien et pakistanais au terme d'un voyage par l'intermédiaire du droit britannique.

Le droit se subdivise en deux grandes catégories : droit privé et droit public. Le droit privé règlemente les relations entre personnes privées, physiques ou morales, tandis que le droit public régit les relations impliquant une personne publique. Bien que le droit pénal s'attache à la protection des intérêts de la société, il fait partie du droit privé. Il

[62]Source : Centre National de Ressources Textuelles et Lexicales, `https://www.cnrtl.fr/etymologie/justice`.

[63]Source de la définition : `https://www.cnrtl.fr/definition/judiciaire`.

se définit comme un droit répressif qui sanctionne la faute commise par un membre de la société, rompant ainsi le principe de solidarité (Teillot et Ubransky, 2008). On retrouve ici la connotation morale, presque religieuse, de la rupture d'un ensemble de valeurs et de croyances.

L'adjectif pénal trouve son étymologie dans le latin *poena*, « peine, sanction ». Le droit pénal est donc celui qui détermine les peines, les sanctions en cas d'infraction. Sous l'Ancien Régime, où l'incrimination était fondée sur la coutume, la notion de responsabilité pénale tenait déjà compte de l'intention de l'auteur, de son âge et de sa santé mentale. La peine était déterminée en fonction de la qualité de l'auteur de l'infraction et de la victime ainsi que des circonstances de l'acte. Parallèlement à cela, le roi pouvait, en vertu du pouvoir de la « justice retenue », engager directement des poursuites et en arrêter d'autres définitivement, conférant au système une dimension arbitraire.

Parmi les philosophes des Lumières, Montesquieu en France (De l'esprit des Lois, 1748), puis Beccaria (1764) énoncent les premiers le principe de légalité de la peine établissant la proportionnalité de l'acte et de sa sanction incarnée dans la maxime *nullum crimen, nulla poena sine lege* (« il n'existe ni crime, ni peine sans loi »). Autrefois, le droit pénal était appelé « droit criminel », à ne pas interpréter en français moderne comme le droit qui traite les infractions les plus graves. On retrouve cette dénomination dans l'anglais *criminal law*.

En France, le droit pénal prend sa source dans les textes nationaux que sont la Constitution et le Code Pénal, et dans des textes internationaux comme la Déclaration Universelle des Droits de l'Homme, et divers traités ou conventions internationaux, notamment européens.

1.1.2.2 Organisation de la justice pénale française

En France comme dans les deux États précédents, le judiciaire est séparé du législatif et de l'exécutif. Comme nous l'avons mentionné en introduction, le système judiciaire français est, à la différence des deux autres, civiliste, ce qui implique une organisation différente de la procédure. La France est en outre un État à l'administration beaucoup plus centralisée. Il n'existe que deux degrés de juridiction pénale, sans Cour suprême en dernier recours.

Les juridictions pénales françaises se répartissent en deux degrés, ou ressorts.

Au premier degré, les affaires pénales sont réparties en trois catégories en fonction de la gravité de l'infraction.

- les contraventions sont jugées par le tribunal de police ;

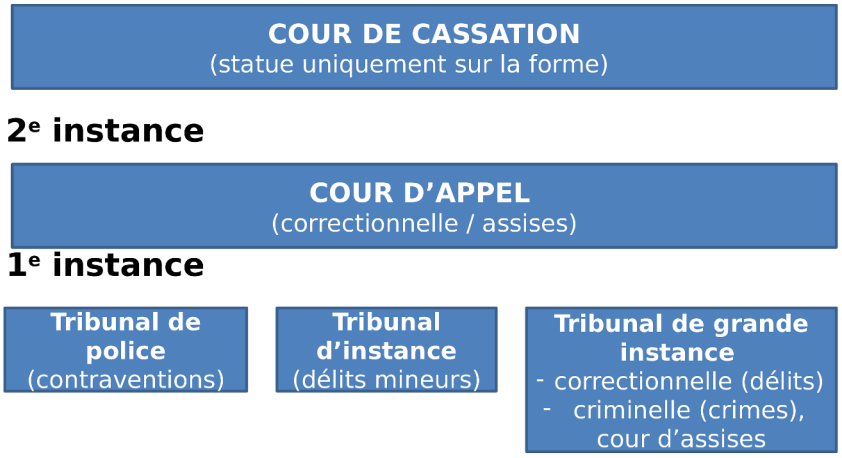

FIG. 1.8: Les juridictions pénales en France

- les délits sont jugés par le tribunal d'instance ou de grande instance, en fonction de la gravité et du montant de l'amende encourue.[64] On parle alors d'une *procédure correctionnelle* ou délictuelle ;
- les crimes, qui sont les infractions les plus graves, sont jugés en cour d'assises, par un jury populaire tiré au sort sur les listes électorales. En ce cas, la procédure déclenchée est dite *criminelle*.

Le français distingue ainsi sur le plan terminologique les infractions selon leur gravité, ce qui aura des répercussions sur toute la terminologie de la procédure (cf. chapitre 2). Cette distinction est absente du système anglais et, par conséquent, des terminologies indienne et pakistanaise.

Le second degré est réservé aux appels. Un justiciable, ou le Ministère public, peut interjeter appel d'une décision concernant un délit ou un crime. En revanche, le jugement du tribunal de police est définitif.

La France dispose en outre d'une Cour de cassation qui, loin d'être l'équivalent d'une Cour suprême, ne constitue pas un degré supérieur. Elle peut « casser » les jugements délivrés par les juridictions d'appel, ce qui signifie qu'elle ne juge pas sur le fond de l'affaire, mais sur la forme. Si elle estime que la règle de droit n'a pas été respectée, elle peut désigner une nouvelle juridiction d'appel pour qu'un nouveau procès ait lieu. Souvent, la Cour de cassation crée du droit de par la jurisprudence qu'elle produit

[64] Le tribunal d'instance traite les affaires dont les montants en jeu sont inférieurs à 10 000 euros.

en interprétant les textes.

Voici posé le cadre légal et judiciaire des trois pays de notre étude, l'Inde, le Pakistan et la France, dont les juridictions sont récapitulées dans le tableau 1.9. Les concepts qui président à la fondation des institutions judiciaires n'étaient pas similaires à l'origine, mais il existait toutefois des systèmes de jugement et de répression des infractions suivant la même logique : une plainte contre un individu, une institution ou des biens, une confrontation des versions des parties et la délivrance d'une décision par un juge en sont les éléments centraux. La police est également présente à travers les différentes époques et les différents pays pour maintenir l'ordre ou réprimer les délinquants. On note aussi dans les trois cas un lien fort entre la justice et le divin, ou du moins le supranaturel. Ce lien se retrouve aussi dans l'approche de la traduction du droit en Occident, liée aux textes sacrés bibliques.

En ce qui concerne les types de tribunaux, on observe dans la péninsule indienne un recours plus important aux tribunaux des « pairs », qu'ils soient de la famille, de la caste ou de la profession, et à la recherche de conciliation. Le droit coutumier, important en France avant la Révolution, est rendu par des institutions, de manière systématiquement verticale.

Toutes ces institutions ne peuvent fonctionner sans l'usage de la langue, puisqu'appliquer la loi consiste à interpréter les règles fixées dans des textes reconnus pas l'ensemble de la communauté. Nous venons de voir comment les mots de la justice et du droit sont nés, ont évolués et dans quelles langues ils étaient formulés. Dans la section suivante, nous nous intéresserons au contrôle de l'État sur la langue du droit, afin de montrer son rôle actif dans la standardisation de la langue et son développement terminologique. Nous verrons notamment que les décisions officielles en matière linguistique contraignent les usages langagiers réels au point de freiner la participation active des justiciables au processus judiciaire.

1.2 Les politiques linguistiques des États

La politique linguistique d'un État définit le soutien du gouvernement envers une ou plusieurs langues dans les différents usages possibles : administratif, éducatif, médiatique, etc. En somme, la politique linguistique définit ce que peuvent ou doivent

<table>
<tr><th></th><th>France</th><th colspan="2">Inde</th><th colspan="3">Pakistan</th></tr>
<tr><td>autre recours</td><td>cour de cassation</td><td colspan="2"></td><td colspan="3"></td></tr>
<tr><td>2e /3e instance</td><td></td><td colspan="2">Cour suprême</td><td colspan="2">Cour suprême</td><td>division d'appel du tribunal de la Charia à la Cour suprême</td></tr>
<tr><td rowspan="2">2e instance</td><td rowspan="2">cour d'appel</td><td>Haute Cour</td><td rowspan="2">autorité d'appel du Panchayat</td><td>Haute Cour</td><td colspan="2" rowspan="2">Cour fédérale de la Charia</td></tr>
<tr><td>tribunal de district, cour d'assises</td><td>tribunal de district, cour d'assises</td></tr>
<tr><td rowspan="2">1e instance</td><td rowspan="2">tribunal de police, tribunal d'instance, TGI (tribunal correctionnel / cour d'assises)</td><td colspan="2">Haute Cour (litiges État / État ou État / Centre)</td><td colspan="3">Haute Cour (litiges province / province ou province / État)</td></tr>
<tr><td>Cour de Magistrat, tribunal de district, cour d'assises</td><td>Nyaya Panchayat</td><td>Cour de Magistrat, tribunal de district, cour d'assises</td><td>jirga (KPK)</td><td>tribunal de la Charia</td></tr>
</table>

FIG. 1.9: Tableau comparatif des juridictions pénales (France, Inde et Pakistan)

parler les individus dans leurs interactions avec l'État en fonction des contextes. En outre, chaque État a mis en place des organismes pour soutenir sa politique linguistique et contrôler les usages langagiers sur le terrain.

La particularité des pays de l'Asie du Sud est leur situation de multilinguisme (1.2.1). C'est dans ce contexte que nous verrons comment les langues sud-asiatiques étudiées ici, hindi et ourdou, ont en réalité une origine commune (1.2.2.1.1) et de quelle manière des décisions politiques les ont différenciées.

En Inde, le multilinguisme est géré par une politique linguistique (1.2.2) soutenue par la Constitution (1.2.2.1.3) et des lois linguistiques fédérales (1.2.2.1.4). Des organismes sont chargés de promouvoir le développement et la diffusion du hindi (1.2.2.2) tandis que l'ourdou y a un statut de langue minoritaire (1.2.2.3).

Au Pakistan, selon la Constitution, les langues reconnues comme officielles sont l'ourdou et l'anglais (1.2.3).

En France, l'usage du français est de rigueur dans l'administration (1.2.4). Cette langue n'a pas de concurrent régionaux sur le territoire français. Le pays développe désormais sa politique linguistique dans un cadre européen (1.2.4.3).

1.2.1 *Le contexte multilingue en Asie du Sud*

L'Inde et le Pakistan, à la différence de la France, sont des pays fortement multilingues. Les locuteurs monolingues y sont relativement rares. Il arrive fréquemment qu'un individu parle trois ou quatre langues : la première au sein de sa famille, une autre

dans ses relations quotidiennes de travail ou sociales, encore une autre à l'école ou dans les bureaux de l'administration, et qu'il prie dans une autre langue encore (Khubchandani, 1991). Toutes ces langues ne sont en général pas maîtrisées au même niveau. Certaines sont seulement orales, d'autres écrites. Elles correspondent à des registres et des fonctions sociales définis[65].

Le sous-continent indien, ou Asie du Sud, est une vaste aire linguistique où cohabitent quatre grandes familles.

La famille indo-européenne, qui compte comme anciens représentants le sanskrit, le perse, le grec ou le latin, y est aujourd'hui représentée par la famille indo-aryenne, qui domine en nombre de locuteurs (cf. carte 1.2.1) avec le hindi-ourdou, le bengali, le pendjabi, le gujarati, le cachemiri, le marathi et le sindhi (pour n'en citer que quelques unes), et la famille iranienne, dont font partie le pashto, le balouchi et le persan. Elle s'introduit dans la péninsule à partir de -1500, avec la langue des Védas (le sanskrit dit « védique »).

La famille dravidienne arrive numériquement au deuxième rang et occupe aujourd'hui le sud de la péninsule, avec une exception au Balouchistan pakistanais. On ignore si elle est rattachée à une famille plus vaste, les tentatives de classification génétique n'ayant pas abouti pour l'instant[66]. Les langues dravidiennes étaient sans doute présentes dans la péninsule sur une aire bien plus grande avant l'arrivée des Indo-Aryens. Ses représentants actuels les plus connus sont le télougou, le tamoul, le kannada et le malayalam, sans oublier le brahui au Pakistan.

La famille austro-asiatique représente un petit nombre de locuteurs en comparaison des précédentes. Les langues munda-s, parlées dans le centre-est de l'Inde (Jharkhand, Bihar), comptent toutefois plusieurs millions de locuteurs. Rarement écrites, ce sont les langues associées aux *ādivāsī*-s, littéralement les habitants originaux du sous-continent. Le môn-khmer leur est apparenté.

Pour finir, la famille tibéto-birmane, rattachée à la plus vaste famille sino-tibétaine, regroupe des langues parlées aux marges, dans les contreforts montagneux de l'Himalaya. Les langues tibéto-birmanes sont parlée en Inde par 12 257 382 locuteurs[67]. Ses représen-

[65]Ce qu'avait déjà traduit le théâtre sanskrit antique en assignant une langue particulière à un statut ou une fonction sociale du personnage.

[66]Certaines hypothèses rattachent le dravidien à la civilisation de l'Indus, présente entre -4000 et -1500 de l'Afghanistan jusqu'au Haryana actuel.

[67]*Census of India*, 2011, consulté sur le site `http://www.censusindia.gov.in/2011Census/C-16_2506201`

tants les plus importants en nombre sont le manipuri (plus d'1,7 million de locuteurs) et le bodo (près d'1,5 millions de locuteurs)[68].

Il existe encore d'autres langues non classées, dont les parlers des îles Nicobar (dits negrito) et le burushaski dans les territoires du Nord au Pakistan.

La plupart de l'ensemble des langues du sous-continent sont orales. Il existe cependant de nombreux alphasyllabaires, tous issus de l'écriture *brāhmī* à l'exception des langues utilisant l'écriture arabe (pashto, ourdou, sindhi, balochi, kashmiri…)[69] (Bühler, 1898). Utilisée entre le IV^e^ siècle av. J.C. et le V^e^ siècle ap. JC., la brahmi a servi à noter le sanskrit et différents prakrits. Ultérieurement, les alphabets indiens ont évolué selon deux grands ensembles pour donner naissance au nord aux alphasyllabaires sharada (*śāradā*, dont dérivent la gurmukhi qui sert à noter le pendjabi, le dogri), *nāgarī* puis *devanāgarī* (pour le sanskrit, le hindi, le népali ou le marathi, le tibétain, le bengali….). La variante sud de la brahmi a donné naissance aux alphabets des langues dravidiennes et, avec la diffusion du bouddhisme theravada, à la plupart des alphabets utilisés en Asie du Sud-Est.

L'écriture devanagari apparaît vers le VII^e[70]^ siècle et sa forme actuelle est stabilisée vers le XV^e^ siècle. La devanagari n'est donc pas à l'origine spécifique à l'écriture du sanskrit. D'ailleurs, le hindi s'est écrit dans d'autres scripts[71], comme l'ont été d'autres langues de la région, telles le cachemiri, le sindhi, le dogri ou le pendjabi, qui se sont écrites tantôt en sharada, en alphabet arabe, en devanagari ou en gurmukhi.

On recensait plus de 1369 langues déclarées comme « maternelles » [72] en Inde dans le recensement de 2011[73] et environ 80 au Pakistan au recensement de 2017. Les millénaires de cohabitation et de contact, qui ont créé des convergences linguistiques (lexicales, morphologiques ou syntaxiques), n'ont pas réduit la complexité linguistique de

---

8_NEW.pdf.

[68] *Op. cit.*

[69] Si l'on considère que la brahmi dérive d'un alphabet sémitique, tous les scripts ont bien une origine commune.

[70] cf. https://www.britannica.com/topic/Devanagari.

[71] Au Pendjab pakistanais, on peut encore voir des ex-voto dans le sol des anciens temples gravés en écriture arabe et transcrivant du hindi.

[72] La notion même de langue est difficile à définir, tant elle contient des éléments politiques et identitaires variés. Les enquêteurs ont en réalité recueilli 19 569 dénominations différentes, ramenées à 1369 après vérifications.

[73] *Census of India*, 2011, *op. cit.*, partie « Language », p. 4.

l'aire, bien au contraire.

En Inde et au Pakistan, les enjeux politiques autour des langues et des identités linguistiques sont toujours vivaces.

1.2.2 *Politique linguistique en Inde*

La place du hindi comme langue de gouvernance et plus généralement la politique linguistique en Inde ont toujours été des sujets épineux, sans cesse remis en cause et prêtant à controverse.

Après la Partition et l'Indépendance, le pays fut découpé en États suivant des critères linguistiques. Ces critères furent contestés à de multiples reprises et le sont encore actuellement[74]. Depuis son Indépendance, l'Inde a œuvré à renforcer, ou du moins à maintenir, l'usage du hindi comme langue officielle du pouvoir fédéral. L'opposition des territoires non-hindiphones, craignant de perdre leur identité linguistique et culturelle, d'une part, et l'influence d'une élite éduquée au moyen de l'anglais, langue de communication internationale et ouverture vers l'Occident, poussèrent le gouvernement indien à conserver l'anglais à côté du hindi. La décision fut prise au terme de longs et houleux débats.

1.2.2.1 Le statut du hindi

Langue comptant le plus de locuteur dans le pays, le hindi possède un statut privilégié dans l'administration aux côtés de l'anglais. Nous verrons tout d'abord les conditions qui ont mené le hindi à cette place, avec pour système d'écriture la devanagari, qui a longtemps cohabité avec l'alphabet arabo-persan. Par la suite, nous nous intéresserons aux textes de loi et aux institutions qui assurent au hindi son rayonnement national et son dynamisme terminologique.

Le hindi est officiellement langue maternelle de 43,63% de la population indienne en 2011[75], tandis que l'anglais se situe au 40^{e} rang en nombre de locuteurs natifs (Montaut, 1997b), et que moins de 10% de la population indienne parle aujourd'hui l'anglais. Le statut du hindi est déterminé officiellement en tout premier lieu par la Constitution, qui entre en vigueur en 1950, puis par un ensemble de lois qui en définissent les usages et

[74]Cf. les revendications du Haryana pour être séparé du Pendjab en 1956, grève de la faim tragique de Potti Sreeramulu pour la constitution d'un État télougouphone en 1952 ou encore les demandes de création d'États comme le Bodoland pour la communauté Bodo dans le Nord-Est.

[75]www.censusindia.gov.in/2011Census/Language-2011/Statement-4.pdf.

les moyens de promotion. Ce statut a été entériné après des débats mouvementés entre les tenants de l'anglais et ceux du hindi comme langue d'État.

En effet, les partisans du hindi souhaitaient promouvoir par son usage une identité indienne affranchie du passé colonial, dans la lignée des mouvements nationalistes qui avaient pris leur essor dans la lutte pour l'Indépendance.

De l'autre, les tenants de l'anglais avançaient plusieurs arguments. D'une part, l'élite menée par Franck Antony considérait que l'anglais, langue de communication internationale, apporterait une ouverture sur le monde[76]. D'autre part, les États dravidiens du sud de la péninsule, non hindiphones, rejetaient l'usage d'une langue indo-aryenne, perçue comme une tentative d'hégémonie culturelle du Nord sur le Sud. L'anglais permettait de garantir un médium de communication plus neutre culturellement. Enfin, Ambedkar, avocat de la cause des intouchables, considérait l'adoption de toute langue indienne standardisée comme une imposition de la culture des castes dominantes sur les castes dominées dont les parlers étaient considérés comme non standard. Là encore, l'anglais offrait une médiation neutre. L'empreinte de ces débats se retrouve dans la rédaction de la Constitution qui donne une place particulière au hindi et à l'anglais dans l'Union, tout en reconnaissant des droits linguistiques aux autres locuteurs.

Cependant, la reconnaissance du hindi comme langue officielle de l'Union ne signe pas la clôture des débats sur la langue et ses usages. Les parlementaires de l'aube de l'indépendance devaient encore s'entendre sur la nature du hindi à adopter. Avant de poursuivre, la contextualisation du paysage linguistique historique de la zone hindiphone permettra de comprendre la complexité des débats qui ont mené aux décisions de politiques linguistiques.

1.2.2.1.1 Hindi et ourdou, une origine commune

Au paragraphe 1.2.1, nous avons vu que le hindi appartenait au groupe indo-aryen, et que ce groupe était représenté à l'époque antique par le sanskrit, langue classique et très documentée, utilisée pour les rites religieux et l'administration. Le sanskrit n'est lui-même pas homogène, il connaît plusieurs phases de développement depuis l'époque védique et a été « mis en forme » par les grammairiens classique dont Pāṇini (environ -V^{e} siècle) était un éminent représentant.

Quoi qu'il en soit, le sanskrit n'était pas la seule langue indo-aryenne à l'époque,

[76]Sridhar (1987), p.309.

et il n'était pas la langue du peuple. La tradition recense un certain nombre de prakrits (parlers « vulgaires », par opposition au sanskrit, langue « complète », « raffinée ») en usage à la même époque dans la population[77]. Dans le théâtre classique antique, la notion de langue associée à une classe sociale ou à des usages précis est déjà bien ancrée. Dans le traité de dramaturgie Nātyaśāstra[78], écrit entre le -V^{e} et le +V^{e} siècle, le chapitre 18 précise quels personnages peuvent s'exprimer en sanskrit (les rois, princes et brahmanes, dans certaines conditions les reines et les femmes de haut rang), dans un prakrit (les femmes de toutes conditions, les enfants ainsi que les personnages de faible extraction) ou en d'autres langues[79], en descendant au fur et à mesure dans la hiérarchie sociale. Si certains prakrits sont parvenus à atteindre le statut de langue littéraire ou religieuse (comme le pali sous Ashoka), la diglossie prévaut et nul système formalisé ne semble mis en place pour passer d'une langue à l'autre.

Le passage des anciennes langues indo-aryennes (Old Indo-Aryan, indo-aryen ancien (OIA))aux New Indo-Aryan, langues actuelles de la famille indo-aryenne (NIA) est mal établi. Les langues de la période charnière, autour de l'an 1000, s'appellent d'ailleurs *apabhraṃśa* (« corrompue »). Elles ont évoluées en une multitude de parlers dans la plaine du nord l'Inde. Le hindī moderne descend de la *khaṛī bolī*, parler de la région de Delhi et plus au nord, sans tradition littéraire prestigieuse. On recense les premières occurrences de l'appellation hindi, ou plutôt en *hindvī* / *hinduī* ou *hindavī* chez le poète Amir Khusrau (1253-1325) (Montaut, 2012), puis dans les cours de la région du Deccan (centre de la péninsule), où il est parlé par les soldats des camps, véritables creusets culturels, puis par la noblesse centre-asiatique indianisée. Une autre appellation, l'ourdou, qui tire son nom du turcique signifiant « camp militaire », langue qui métisse des mots persans, arabes, turciques et de langues locales. Le terme *urdū*, qui semble apparaître au XVIII^{e} siècle seulement (Montaut, 2012), ne s'impose réellement qu'à partir du XIX^{e} siècle. D'autres appellations sont également en usage : *bhākhā*, *bhāṣā*, qui signifient simplement « langue », ou vers le XVII^{e} siècle *hindustānī*, terme donné par les souverains en persan (Montaut, 2012). Quel que soit son nom, cette langue s'écrit à l'aide de différentes écritures : devanāgarī ou arabo-persane.

Cette situation syncrétique perdure jusqu'au XIX^{e} siècle, lorsque les identités

[77]Le Nātyaśāstra en cite sept : māgadhī, āvantī, prācyā, śaurasenī, ardhamāgadhī, bāhlikā et dākṣiṇāṭyā.

[78]Muni (1951), pp. 323-333.

[79]Le Nātyaśāstra cite : le śākarī, le cāṇḍalī, l'ābhirī, le śābarī, le drāviḍī et l'oḍrī.

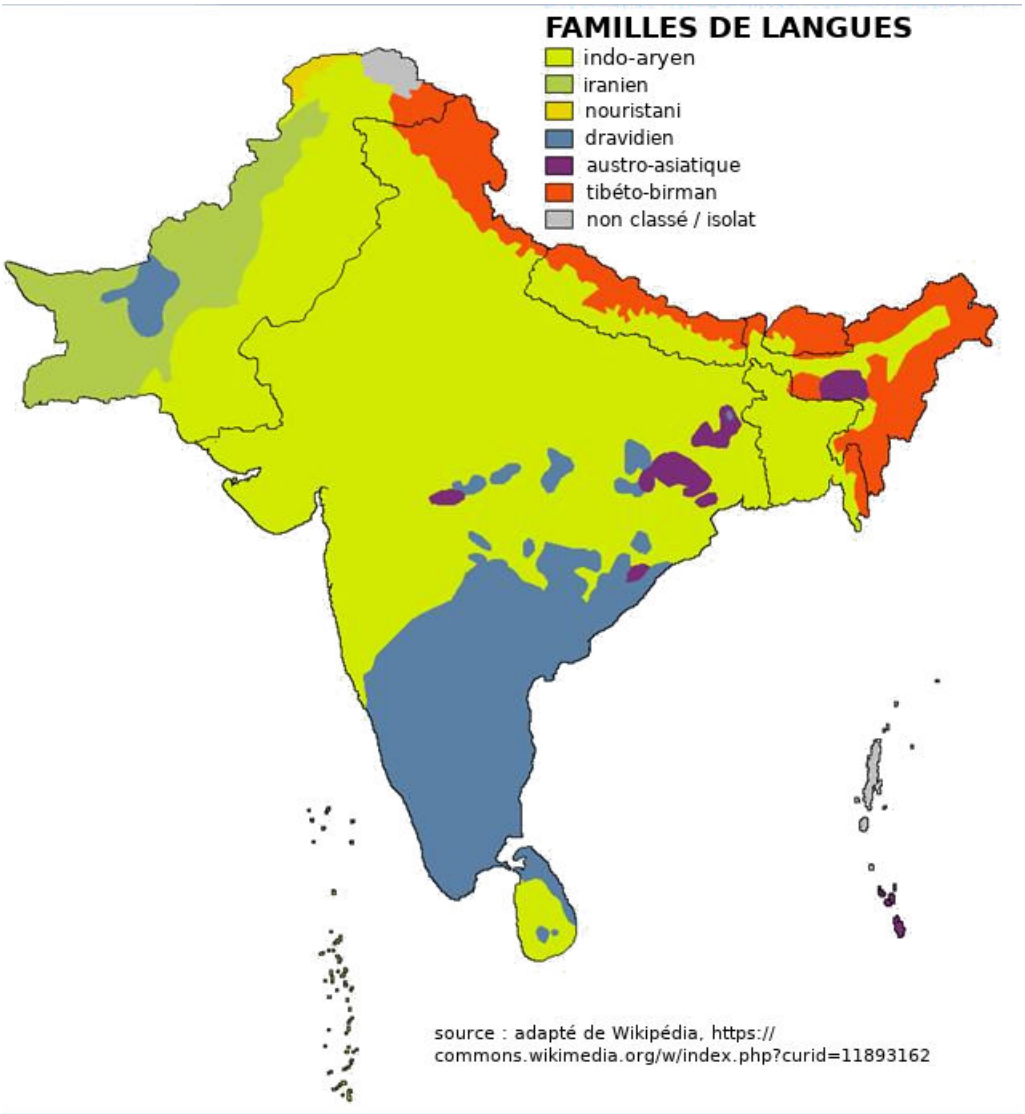

FIG. 1.10: Répartition des langues en Asie du Sud

communautaires et linguistiques commencent à se figer, pendant la période coloniale britannique. À cette époque, les Britanniques s'interrogent sur la langue à utiliser pour l'administration du territoire colonisé. Deux camps s'affrontent : l'orientaliste et juge William Jones ainsi que les missionnaires sont favorables à l'adoption des langues vernaculaires, tandis que McCaulay et les anglicistes jugent que les langues régionales n'ont pas suffisamment de génie pour transmettre un quelconque savoir moderne et scientifique (Hock et Bashir, 2016, p. 647). Parallèlement, les langues vernaculaires se développent dans les royaumes souverains.

Dans la population, on parle lors des échanges quotidiens ou dans l'espace public l'hindoustani, mâtiné de lexique arabo-persan, sanskrit ou d'autres mots dits *deśī*, « indigènes », dont l'étymologie est indéfinie.

Cependant, au niveau administratif, les Britanniques choisissent de s'appuyer sur la

minorité musulmane, qui conservait encore un pouvoir symbolique à travers les derniers souverains moghols. Ils s'aident des munshis, ces scribes qui tenaient les registres de l'administration, du commerce et des terres, et qui étaient versés dans la culture persane. L'ourdou, qui désigne alors un état de la langue persianisée et écrite avec le script arabo-persan, est à partir de 1836 reconnue comme langue administrative (Montaut, 2012) et employée dans les tribunaux au côté de l'anglais.

En réaction à ce qui est perçu comme une mesure de soutien à la communauté musulmane, un mouvement revivaliste prône la purge du hindi de ses mots d'origine arabo-persane ou étrangère, c'est-à-dire non sanskrite. On commence à associer le hindi à une forme de la langue plus sanskritisée, écrite avec les lettres devanagari et associée à l'hindouisme. Le clivage grandit au fil du siècle précédant l'indépendance, en même temps que les dissensions entre hindous, sikhs et musulmans.

Toutes ces sensibilités linguistiques se retrouvèrent dans les débats à l'aube de l'Indépendance : des enjeux d'identité culturelle nationale, de hiérarchie de classes et de castes continuaient de sous-tendre les discussions. Certains, comme Gandhi et ses héritiers, prônaient l'adoption de l'hindoustani, langue du peuple, plus facile à adopter par une population largement analphabète qui en était déjà familière. D'autres jugeaient que l'hindoustani n'avait pas suffisamment de génie pour exprimer des concepts scientifiques et techniques car elle n'était qu'une langue de communication entre différents groupes (Sridhar, 1987). Selon eux, le hindi devait revenir à sa base indo-aryenne classique sanskrite, seule capable de lui fournir le vocabulaire nécessaire pour former les langues des spécialités scientifiques, juridiques et autres. Ils arguaient que le fonds lexical arabo-persan se prêtait mal à la dérivation lexicale, du fait de la morphologie même des langues persanes et arabes. Autre argument des avocats d'un hindi plus sanskritisé : le sanskrit, bien qu'indo-aryen, est une langue culturelle qui s'est diffusée dans la péninsule pendant des millénaires, comme en atteste le vocabulaire des langues dravidiennes, très largement sanskritisé. Le recours à des emprunts au sanskrit devait donc permettre une compréhension plus facile des mots par les locuteurs non indo-aryens et contribuer à la formation d'une terminologie panindienne. Or, comme l'ont montré plusieurs linguistes[80], si le sanskrit a bien produit de nombreuses lexies dans les diverses langues indiennes, au fil des siècles, des dérivations et des emprunts, ces mots ont évolué et des morphèmes similaires peuvent souvent avoir pris des sens très différents. Ce socle classique ne garantit

[80]Sridhar (1987), p. 307.

donc pas l'intercompréhension entre les langues indiennes. En 1949, le hindi est adopté comme « langue officielle ». La Constitution qui paraît l'année suivante prévoit qu'il s'enrichira des apports des autres langues indiennes, afin de faciliter son apprentissage et sa diffusion auprès des non hindiphones. Le développement du hindi dans les décennies suivantes comme langue administrative en fera une variété de plus en plus coupée de la langue populaire (Montaut, 1997a).

1.2.2.1.2 L'écriture devanagari

Nous l'avons vu précédemment (cf. paragraphe 1.2.1), l'écriture devanagari provient de l'alphasyllabaire *brāhmī*. Son organisation est la même que celle de toutes les écritures qui proviennent de la brāhmī : des caractères notent les voyelles pleines (en position initiale ou isolée dans un mot) et les consonnes, chaque son consonantique est suivi d'une voyelle inhérente /a/ (a bref). L'ajout d'une autre voyelle que le /a/ à une consonne s'effectue au moyen d'un symbole vocalique différent de la forme pleine qui s'écrit au-dessus, en dessous, avant, après ou autour du caractère consonantique (cf. fig 1.11). Plusieurs consonnes peuvent se lier et former une ligature (cf. fig. 1.2).

FIG. 1.11: Caractère क combiné aux voyelles /ā/, /i/, /ī/, /u/, /ū/, /r̥/, /R̥/, /e/, /o/, /au/ [81]

caractères isolés	ligature	transcription (IAST)
क् + छ	क्छ	kcha
क् + श	क्श	kṣa
क् + र	क्र	kra
र् + क	र्क	rka
र् + क् + त् + इ	र्क्ति	rkti

TAB. 1.2: Exemple de ligatures le caractère क /ka/ [82]

1.2.2.1.3 D'après la Constitution

Les articles traitant des questions linguistiques, de 343 à 351, se trouvent dans la partie XVII intitulée "Official language". L'article 343 stipule :

(1) The official language of the Union shall be Hindi in Devanagari script. […][83]

Le hindi y est donc reconnu comme langue officielle (*rāj bhāshā*) au niveau fédéral, ce qui ne signifie pas qu'elle est l'unique langue officielle du pays, comme l'indiquent les articles suivants ainsi que l'annexe VIII des langues reconnues à usage officiel. Le paragraphe suivant de l'article prolonge pour quinze ans l'usage de l'anglais dans les fonctions officielles de la fédération, lui conférant ainsi le statut de *sah-rāj bhāshā* (co-langue officielle) :

> (2)Notwithstanding anything in clause (1), for a period of fifteen years from the commencement of this Constitution, the English language shall continue to be used for all the official purposes of the Union for which it was being used immediately before such commencement :
>
> Provided that the President may, during the said period, by order authorise the use of the Hindi language in addition to the English language and of the Devanagari form of numerals in addition to the international form of Indian numerals for any of the official purposes of the Union.[84]

Cette période devait permettre aux employés des administrations de maîtriser suffisamment le hindi pour alors s'affranchir totalement de l'anglais.

La Constitution indienne dispose en annexe VIII (8th schedule) de la liste des langues officiellement reconnues dans l'ensemble de l'Union[85]. Chaque État possède ainsi une langue pour ses usages officiels.

Les articles 348 et 349 définissent le statut prééminent de l'anglais comme langue législative. L'article 348 précise que tous les textes législatifs doivent être proposés, promulgués et votés en anglais. Cependant, le Gouverneur d'un État peut autoriser l'usage du hindi ou de toute autre langue à usage officiel pour la procédure judiciaire en Haute Cour, à l'exception des jugements. En ce cas, les textes doivent être accompagnés d'une traduction en anglais. L'article 349 protège ce statut pour une période de 15 ans à partir de la mise en œuvre de la Constitution, soit jusqu'en 1965. La limite d'adaptation de

---

[83](1) *"La langue officielle de l'Union sera le hindi, en écriture devanagari."*.

[84] *"(2) Nonobstant la disposition du paragraphe 1, pendant une période de quinze ans à compter de la date de promulgation de la présente Constitution, l'anglais doit continuer d'être utilisé aux fins officielles de l'Union pour lesquelles il était déjà utilisé immédiatement avant la promulgation / Pendant ladite période, le président peut autoriser par décret, aux fins officielles de l'Union, l'usage de la langue hindi, en plus de la langue anglaise, et des chiffres selon la forme devanagari, en plus des chiffres indiens selon la forme internationale."* Traduction de J. Leclerc, *L'aménagement linguistique dans le monde*, `http://www.axl.cefan.ulaval.ca/asie/inde_const1.htm`.

[85]La 8e annexe de la Constitution, se rapportant aux articles 344(1) et 351, reconnaît 23 langues : le hindi et l'anglais, l'assamais, le bengali, le bodo, le dogri, le gujarati, le kannada, le cachemiri, le konkani, le maithili, le malayalam, le manipuri, le marathi, le népali, l'oriya, , l'ourdou, le pendjabi, le sanskrit, le sindhi, le tamoul, le télougou.

quinze fut finalement levée en raison de l'opposition des États du Sud, et l'anglais fut conservé comme co-langue officielle.

L'article 350 définit l'utilisation des langues dans les cours de justice :

> *350. Language to be used in representations for redress of grievances.-*
> Every person shall be entitled to submit a representation for the redress of any grievance to any officer or authority of the Union or a State in any of the languages used in the Union or in the State, as the case may be.[86]

Cela signifie que tout justiciable peut s'adresser à la justice dans l'une des 23 langues de l'annexe VIII.

Les articles 350A et B insistent sur la protection des droits linguistiques des minorités en instaurant un Officier spécial pour les minorités linguistiques et en reconnaissant le droit à recevoir une instruction primaire dans sa langue maternelle. Cependant, la définition d'une langue maternelle n'est pas établie. Si les droits linguistiques sont reconnus, l'assistance linguistique, elle, n'est pas clairement définie.

On constate au travers des articles évoqués qu'un nombre important de langues est évoqué et que ces langues n'ont pas toutes le même statut dans les différents organes administratifs. Une activité importante de traduction est donc à prévoir, qui n'est pas évoquée à ce stade.

L'article 351 consacre le statut du hindi comme langue d'envergure nationale. Il indique la volonté du pays de promouvoir son usage à travers l'Union et d'en facilité l'adoption en empruntant dans les diverses langues indiennes une proportion de son vocabulaire :

> *351. Directive for development of the Hindi language.*
> It shall be the duty of the Union to promote the spread of the Hindi language, to develop it so that it may serve as a medium of expression for all the elements of the composite culture of India and to secure its enrichment by assimilating without interfering with its genius, the forms, style and expressions used in Hindustani and in the other languages of India specified in the Eighth Schedule, and by drawing, wherever necessary or desirable, for its vocabulary, primarily on Sanskrit and secondarily on other languages.[87]

---

[86] En français : *« Langue des représentations visant à réparer un tort : Quiconque a le droit de faire une représentation visant à réparer un tort auprès d'un représentant de l'Union ou d'un État dans n'importe laquelle des langues utilisées dans l'Union ou dans l'État, selon le cas. »* Traduction de J. Leclerc, *L'aménagement linguistique dans le monde*, `http://www.axl.cefan.ulaval.ca/asie/inde_const1.htm`.

[87] *« Directive concernant la promotion du hindi : L'Union a pour devoir de promouvoir l'usage du hindi, de développer cette langue de manière à ce qu'elle serve de moyen d'expression à tous les éléments de la culture composite de l'Inde et d'assurer son enrichissement par l'assimilation, sans nuire à son génie, des formes, du style et de l'expression propres à l'hindoustani et aux autres langues de l'Inde énumérées à l'Annexe VIII. Chaque fois que ce sera nécessaire ou désirable, le hindi puisera, pour enrichir son vocabulaire, dans celui du sanskrit, puis dans celui d'autres langues. »*, Traduction adaptée de J. Leclerc, *L'aménagement linguistique dans le monde*, `http://www.axl.cefan.ulaval.ca/asie/inde_const1.htm`.

L'article fait référence à l'hindoustani, variété du hindi intégrant de nombreux mots d'origine persane ou *deśī* dont Gandhi se faisait le champion. Cette variété populaire s'oppose au hindi très sanskritisé dont l'usage se développe dans les administrations à partir de l'Indépendance, tendance encouragée par ce même article. L'article 351, de par sa formulation ambiguë, autorise ainsi deux interprétations antithétiques : l'une prône la continuité des usages linguistiques populaires d'une variété de langue qui revendique son métissage, l'autre privilégie le recours au fonds lexical sanskrit pour promouvoir la langue.

Tout un appareillage est ainsi créé pour accompagner le développement et la diffusion du hindi tant dans les administrations qu'auprès du grand public. Le Central Hindi Committee, le Central Hindi Directorate (Montaut, 2003), ainsi qu'un organe de création terminologique pour pallier le manque de termes techniques dans tous les domaines grâce à la production de néologismes, le Board for Scientific and Technical Terminology (1950), qui deviendra en 1961 le CSTT, voient le jour.

1.2.2.1.4 Autres textes de loi

Deux autres textes législatifs garantissent le statut et la promotion des langues officielles à travers l'Union : l'Official Language Act de 1963 et l'Official Language Rule de 1976.

a. L'Official Language Act (1963) - loi sur les langues officielles

La loi de 1963 sur les langues officielles, amendée en 1967, indique que les lois et jugements prononcés dans les Hautes cours de justice (High Courts) doivent être traduits en hindi lorsqu'ils ont été rédigés initialement en anglais. Si un jugement a été rédigé ds une autre langue, il doit être traduit vers l'anglais, mais pas nécessairement vers le hindi. Cette loi permet de prolonger l'usage de l'anglais comme langue officielle aux côtés du hindi. De fait, l'anglais conserve une force législative plus grande puisque la législation centrale continue d'être produite initialement en anglais. La loi de 1963 distingue les États de langue officielle hindi des autres. Dans l'article 7, il est stipulé que :

> *7. Optional use of Hindi or other official language in judgments, etc., of High Courts.*— As from the appointed day or any day thereafter the Governor of a State may, with the previous consent of the President, authorise the use of Hindi or the official language of the State, in addition to the English language, for the purposes of any judgment, decree or order passed or made by the High Court for that State and where any judgment, decree or order is passed or made in any such language (other

than the English language), it shall be accompanied by a translation of the same in the English language issued under the authority of the High Court.[88]

Il en ressort que tout ce qui n'est pas en anglais à l'origine doit être traduit dans cette langue au niveau de la Haute Cour et que l'usage du hindi, ou d'une autre langue, y est optionnel.

b. L'Official Language Rule (1976) - règlement sur les langues officielles

Cette loi, qui découle de l'article 8 de la loi précédente, définit les usages linguistiques entre les États et les organes de l'Union (niveau fédéral). À la différence de l'Act de 1963, il distingue trois catégories d'États en fonction de leur utilisation du hindi[89].

- Régions A : États ou territoires dont le hindi est langue officielle ;
- Régions B : États ou territoires dont la langue officielle est une langue indo-aryenne, et où la maîtrise ou la compréhension du hindi reste répandue ;
- Régions C : États ou territoires autres. Ce sont principalement les États du Sud, dravidiens, dont l'identité culturelle forte est plus éloignée de la zone d'influence du hindi.

Le texte définit deux niveaux de maîtrise de la langue : « proficiency in Hindi » (maîtrise du hindi) indique que le locuteur a suivi au moins dix ans de scolarité en hindi, ou qu'il a opté pour une étude renforcée du hindi pendant sa scolarité, ou qu'il est titulaire d'un diplôme du supérieur en hindi. « Working knowledge of Hindi » indique un niveau de connaissance courant. Le locuteur a passé le hindi comme matière lors de l'examen de secondaire Matriculation ou un examen de niveau équivalent en hindi.

Selon le texte, un bureau administratif possède une connaissance courante du hindi si au moins 80% de ses employés ont atteint ce niveau.

Les deux textes législatifs définissent la création de comités chargés d'évaluer périodiquement la progression de l'usage du hindi dans les administrations, toujours dans la ligne fixée par la Constitution.

---

[88] *« Utilisation facultative du hindi ou d'autres langues officielles dans les jugements des Hautes Cours : Conformément au jour désigné ou de tout autre jour, le gouverneur d'un État peut, avec le consentement préalable du président, autoriser l'emploi du hindi ou de la langue officielle de l'État, en plus de l'anglais, pour l'usage de tout jugement, décret ou arrêté passé ou édicté par la Haute Cour de cet État et pour tout jugement, décret ou arrêté passé ou édicté dans une telle langue (autre que l'anglais), il sera accompagné d'une traduction équivalente en anglais émise sous l'autorité de la Haute Cour. »*, Traduction adaptée de J. Leclerc, *L'aménagement linguistique dans le monde*, `http://www.axl.cefan.ulaval.ca/asie/inde_const1.htm`.

[89] Ni le Tamil Nadu, ni le Jammu et Cachemire ne sont concernés par cette loi.

1.2.2.2 Les outils officiels de promotion du hindi

Poursuivant les objectifs posés par la Constitution, l'État a mis en place divers organes pour promouvoir l'usage du hindi : le Département pour la langue officielle (Department of official language), la Commission pour la terminologie scientifique et technique (Commission for Scientific and Technical Terminology) et des organes semi-publics pour former à la traduction administrative entre l'anglais et le hindi comme le Bharatiya Anuvad Parishad (Conseil indien de traduction). Il promeut également au niveau de l'éducation la formule trilingue dite *Three Language Formula*.

1.2.2.2.1 Le Department of official language (rājbhāṣā vibhāg)

Sous la tutelle du ministère de l'Intérieur (Ministry of Home Affairs), il est chargé de la bonne application des textes législatifs concernant l'usage du hindi. Il favorise la diffusion du hindi parmi les employés du gouvernement central au travers du réseau du Kendriya Hindī Samiti et supervise les questions de traduction officielle (production de terminologie et de manuels). D'après les rapports publiés par ce département, le hindi a gagné en visibilité mais n'est cependant pas encore majoritairement adopté comme langue de travail dans les administrations, où l'anglais reste dominant[90].

1.2.2.2.2 Le BSTT et la CSTT

Deux organes de production terminologique voient successivement le jour : tout d'abord le Board for Sicentific and Technical Terminology (BSTT) en 1950, puis la CSTT en octobre 1961 sous l'égide du Ministère de l'Éducation, dans le cadre de l'article 344 de la Constitution[91].

La CSTT a pour mission de décrire la terminologie scientifique et technique du hindi puis des autres langues régionales. Elle doit aussi définir un cadre pour la standardisation de ces terminologies et réviser les termes déjà existants. Elle développe des néologismes dans les domaines de spécialité si besoin et cherche à en diffuser l'usage dans les administrations, les milieux académiques, scientifiques et éducatifs. Dans ce dernier cas, la terminologie fixée entre alors dans la rédaction des manuels scolaires. La CSTT en évalue également la diffusion dans ces milieux. À terme, les terminologies des langues indiennes seront alignées en vue d'une terminologie pan-indienne.

---

[90] `http://www.axl.cefan.ulaval.ca/asie/inde-3pol-union.htm`.

[91] Comme le Department of official language.

Au sein de cette Commission sont élaborés des dictionnaires et des glossaires de spécialité bilingues (anglais-hindi, hindi-anglais ou anglais en combinaison avec une autre langue) ainsi que des glossaires multilingues. Jusqu'ici, l'effort a été avant-tout porté sur la traduction anglais-hindi.

Le travail des terminologues part d'une sélection de termes techniques anglais à traduire. Cette sélection est soumise à des experts du monde académique pour la spécialité concernée, professeurs, écrivains ou journalistes, qui envoient leurs proposition à un panel d'évaluation. Les propositions sont soit acceptées, soit modifiées ou encore remplacées par un néologisme. Il existe pour les langues régionales une sélection similaire au niveau des États.

Le BSTT, puis la CSTT, ne parviennent pas à absorber les besoins en création terminologique. En 1955, seuls 5000 néologismes sont créés[92], et 34 000 termes juridiques en 1979. Les traducteurs sont peu nombreux, les moyens dérisoires (Montaut, 1997a). En plus des termes très sanskritisés, la syntaxe du hindi officiel est celle de la phrase complexe anglaise, rendant cette variété de la langue artificielle et opaque[93]. Le hindi officiel devient l'objet de critiques tous azimuts. L'objectif de standardisation de la terminologie n'est pas non plus atteint, on trouve des variantes terminologiques régionales (en Uttar Pradesh, au Bihar, Montaut (1997a).

La diffusion et l'usage des termes produits par cette commission restent médiocres. Cela s'explique par la réticence des hautes autorités à appliquer les consignes d'utilisation des termes[94], entre autres pour des raisons matérielles, et en raison de l'artificialité de la langue, très éloignée du parler populaire. Le hindi « officiel » possède un statut plus symbolique que communicationnel[95], la langue de la réussite professionnelle et sociale demeurant l'anglais. Le prestige de l'anglais étant plus fort que le hindi[96], l'élite trouve peut d'intérêt à investir des moyens dans son développement réel.

Dans la pratique, de très nombreux termes scientifiques et techniques anglais sont utilisés dans la langue hindi et leurs équivalents en hindi sont méconnus du lectorat. Leur

[92] Le besoin était alors de 300 000 termes pour le seul Ministère de l'éducation (Montaut, 1997a)

[93] *Op. cit.*

[94] *Op. cit.*

[95] *Op. cit.*

[96] Les classes moyennes investissent énormément dans l'éducation en anglais de leurs enfants, cette langue étant perçue comme la porte vers la réussite professionnelle et sociale. Dans les concours aux grandes écoles scientifiques (IIT) ou de management, l'anglais est incontournable (Montaut, 1997a).

introduction dans les manuels techniques et scolaires rend ces derniers incompréhensibles. La commission a d'ailleurs assoupli sa position sur la question depuis quelques années et l'on considère qu'environ un quart du vocabulaire validé en hindi est simplement translittéré de l'anglais[97]: « computer » *kampyūṭar*, « hard disk » *hārḍ ḍisk*, ou les éléments chimiques en sont des exemples. D'autre part, la demande de la part des domaines de spécialité reste faible : les glossaires sont peu achetés et peu utilisés, les termes forgés peinent à sortir du milieu académique. En 2004, un arrêt de la Cour suprême a même dû forcer le NCERT (National Council of Education Research, Conseil National pour la Recherche Pédagogique) à les utiliser, mentionnant dans sa décision que la mise en place de la CSTT « et les dépenses engagées (…) ne rimeraient à rien si les terminologies (…) n'était pas dans les faits utilisées par le Gouvernement et les organes sous son contrôle.»

1.2.2.2.3 La « three-language formula »

Proposée par le gouvernement central, la « three-language formula » consiste à former des citoyens trilingues. Chaque Indien doit maîtriser sa langue maternelle, le hindi et l'anglais. Dans le cas des locuteurs natifs hindiphones, l'apprentissage d'une autre langue indienne constituerait la troisième langue. Ce dispositif se voulait équitable vis-à-vis des ressortissants du sud du pays : en échange de leurs efforts d'apprentissage du hindi comme langue de communication, leurs compatriotes du nord feraient l'effort d'apprendre des langues dravidiennes (Sridhar, 1987, p.309). Le projet a échoué car si les Indiens du sud apprenaient le hindi, les Indiens hindiphones, eux, préféraient apprendre une langue proche comme l'ourdou ou le sanskrit, plutôt qu'une langue dravidienne. Le Tamil Nadu, se fondant sur le fait que la politique éducative n'est pas centralisée, a décidé d'opter à la place pour une politique bilingue tamoul-anglais[98].

[97] Article de Samanth Subramaniam paru le 28 avril 2011 dans The National, consulté sur `https://www.thenational.ae/world/asia/india-struggles-to-find-hindi-words-for-english-science-terms-1.453891`.

[98] À ce problème, il faut ajouter les difficultés des communautés parlant des langues non répertoriées à recevoir une éducation primaire dans leur langue avant d'apprendre la langue régionale. Cette situation provoque de graves taux d'échec scolaire chez des populations déjà économiquement défavorisées (Montaut, 1997a).

1.2.2.3 L'ourdou en Inde

En 2011, l'ourdou est déclaré langue maternelle par plus de 50 millions de locuteurs indiens, soit 4,19% de la population totale en 2011[99], ce qui en fait la sixième langue la plus parlée à l'échelle nationale. Comme nous l'avons vu, le hindi et l'ourdou sont intimement liés depuis des siècles (cf. 1.2.2.1.1). On parle d'ailleurs couramment de hindi-ourdou dans les descriptions linguistiques. Hindi et ourdou partagent une grammaire commune, ainsi qu'un vaste lexique fondamental qui permet l'intercompréhension des locuteurs de l'une ou l'autre langue. Les méandres historiques ont progressivement investi l'une et l'autre variété de charges culturelle, religieuse et politique, ouvrant la voie à la divergence à une date récente. L'ourdou possède au sein de l'Union indienne deux singularités : son écriture est la seule à ne pas descendre de la *brāhmī* et elle fait partie des langues minoritaires.

1.2.2.3.1 Particularités graphiques de l'ourdou

Une des distinctions majeures de l'ourdou est d'ordre graphique : le hindi s'écrit officiellement en caractères devanagari, tandis que l'ourdou s'écrit dans la variante nastaliq *nastaʿalīq* de l'écriture arabo-persane (du persan *naskh* نسخ et *taʿalīq* تعلیق). Dans cette variante, qui n'est pratiquement plus utilisée de manière usuelle qu'en Asie du Sud, les lettres suivent une ligne descendante dans l'ordre d'écriture. Elle peuvent former des ligatures et se déformer plus ou moins pour s'écrire les unes sous les autres, ce qui n'est pas possible dans la variante dominante d'origine arabe appelée naskh (*naskh*).

Fig. 1.12: Variantes nastaliq et naskh en ourdou

Cette écriture est actuellement la seule du sous-continent – en dehors de l'anglais– qui n'émane pas de la brahmi et dont la direction va de la droite vers la gauche. En outre, l'alphabet arabo-persan n'est pas adapté à la structure vocalique des langues indiennes,

[99]Source : www.censusindia.gov.in/2011Census/Language-2011/Statement-4.pdf.

mais à la structure dérivationnelle trilitère des langues sémitiques : les voyelles brèves ne sont pas notées, les semi-consonnes /v/ et /y/ sont représentées par les mêmes symboles que les voyelles longues /ū/, /o/ et /ī/ et /e/ (respectivement les lettres و et ی).

En comparaison de l'alphabet persan, l'alphabet ourdou s'est enrichi de symboles pour représenter les sons rétroflexes : /ṭ/ ٹ, /ḍ/ ڈ, /ṛ/ ڑ, sauf pour /ṇ/ qui n'est pas distingué du /n/ dental ن.

1.2.2.3.2 Une langue minoritaire

L'ourdou est reconnu comme l'une des 23 langues à usage officiel de l'Union indienne. Elle est langue officielle au Jammu-et-Cachemire où, comme nous le verrons pour le Pakistan, elle n'est pas la langue première de la majorité de la population[100]. Cet État ayant cependant une situation très particulière dans la fédération indienne (la plupart des lois promulguées au Centre ne s'y appliquent pas, il possède ses propres lois et son code pénal[101]), il n'est pas représentatif de la situation de l'ourdou dans la péninsule. L'ourdou est co-langue officielle avec le télougou dans le jeune État du Telangana et possède un statut de « seconde langue officielle » (« second official language ») au Bihar, au Jharkhand, en Uttar Pradesh, au Bengale occidental et dans le territoire de Delhi. Ce statut est octroyé aux langues parlées par au moins 60% de la population d'un district.[102]

Bien qu'étant une autre variante du hindi-ourdou, l'ourdou est considéré comme langue minoritaire et, à ce titre, les droits linguistiques de ses locuteurs sont protégés par l'article 29 de la Constitution. La Commission Nationale pour les Minorités (National Commission for Minorities), créée en 1992, veille à l'application de ces garanties. La National Commission for Minority Educational Institutions (NCMEI), créée en 2004, protège les établissements qui enseignent dans des langues minoritaires.

Malgré toutes ces garanties légales, de nombreux médias et chercheurs constatent le déclin de l'ourdou depuis l'indépendance (l'ourdou perd encore 1% de locuteurs entre le recensement de 2001 et celui de 2011). Différents facteurs peuvent expliquer ce déclin. Dans les régions hindiphones, apprendre l'ourdou comme langue principale signifie souvent

[100] On parle majoritairement dogri dans la région de Jammu, gujari vers Poonch, cachemiri dans la Vallée, balti vers Kargil, ladakhi à Leh… De plus en plus, le hindi y est enseigné dans les écoles au détriment de l'ourdou, surtout dans les zones à majorité hindoue.

[101] L'adoption par le Parlement indien le 5 août 2019 de la révocation de l'article 370 A, de la Constitution, qui garantissait ce statut particulier, risque de remettre en cause cette situation.

[102] Source : `http://www.axl.cefan.ulaval.ca/asie`.

aller dans une école religieuse, avec une éducation dans les autres matières de moindre qualité. Le système scolaire indien favorise l'anglais comme langue d'enseignement, car elle permet ensuite d'accéder aux enseignements supérieurs ou d'envisager une poursuite d'étude à l'étranger. Pour la recherche d'emploi, l'anglais est un plus incomparable : un journaliste ou un écrivain s'exprimant en anglais pourra prétendre à un salaire bien plus élevé que son homologue ne parlant qu'une langue vernaculaire. L'individu scolarisé en anglais ne privilégiera pas l'ourdou comme langue seconde à l'école, mais plutôt le hindi ou une autre langue non indienne, d'usage plus courant ou ouvrant davantage d'opportunités professionnelles (Rahman, 2008).

La diffusion de l'ourdou a sans doute également été freinée par la technologie. Du fait de sa graphie particulière, l'ourdou est plus complexe à traiter et n'a pas pu bénéficier des avancées informatiques des autres langues à script arabo-persan. Jusqu'à très récemment (milieu des années 2000), il était courant de trouver des ouvrages en ourdou calligraphiés à la main puis imprimés. Dans les années 1990, on formait encore et on utilisait les services de calligraphes pour la rédaction de journaux locaux. L'ourdou était peu pratique à écrire avec des traitements de texte informatiques (voir chapitre 3). Trop souvent encore, la qualité des documents numérisés est très inférieure à celle des documents en hindi ou en anglais, et leurs contenus sont beaucoup plus difficiles à manipuler.

L'ourdou possède un Conseil national pour sa promotion, le NCPUL (National Council for Promotion of Urdu Language), depuis 1996. Créé sous l'égide du Ministère du développement des ressources humaines (Ministry of Human Resource Development), il a pour objet de « promouvoir, développer et propager la langue ourdou » [103]. Le site internet du NCPUL[104] recense les formations pour apprendre l'ourdou et les manifestations promouvant cette langue. Il propose quelques ressources avec, en particulier, des textes de loi traduits en ourdou, mais pas de ressources pour le droit pénal. Les documents, tous au format pdf images, ne sont pas transformables et leur qualité de numérisation est mauvaise. Il est regrettable que le site ne fournisse pas de dictionnaire en ligne par exemple.

D'autres initiatives promeuvent l'ourdou, comme la National Translation Mission (cf. partie 3).

[103]Blog du NCPUL : `https://ncpul.blogspot.com/2012/12/urdu-in-india.html`.

[104]`http://www.urducouncil.nic.in/`.

Au cours des trois années de recherche sur le sujet, de nombreux changements ont été remarqués concernant les ressources numériques juridiques accessibles dans les langues indiennes autres que l'anglais : de nombreux sites ainsi que des textes de lois ont été traduits et mis en ligne entre 2016 et 2019.

Quel que soit le statut des langues vernaculaires ou du hindi, la langue ayant le statut linguistique le plus élevé demeure l'anglais. C'est dans cette langue que s'écrivent les lois et qu'est délivrée une éducation supérieure de qualité. C'est elle qui permet d'accéder aux positions économiques et politiques dominantes. Dans les administrations et la justice, l'anglais reste le moyen de communication privilégié. On retrouve chez le voisin pakistanais une discrimination linguistique similaire.

1.2.3 *Politique linguistique au Pakistan*

Comme en Inde, le Pakistan s'est doté d'une langue à vocation fédérale en plus de l'anglais. Or l'ourdou, contrairement au hindi, compte peu de locuteurs natifs. Nous verrons dans un premier temps comment l'ourdou a été choisi, puis comment il est perçu dans la société pakistanaise contemporaine. Enfin, nous verrons que l'État a mis en place des outils pour promouvoir cette langue, ciment national, mais qu'ils semblent insuffisants au regard de l'importance symbolique de l'anglais.

1.2.3.1 L'ourdou, langue nationale

L'histoire linguistique du Pakistan est liée à celle de l'Inde jusqu'en 1947. Le pays nouvellement indépendant, créé pour des raisons politico-religieuses, doit se doter d'éléments favorisant l'intégration des différentes identités qui le composent. L'ourdou sera l'un d'eux.

Le Pakistan voit le jour en 1947, suite à la douloureuse partition de l'Inde britannique en deux entités indépendantes. Les décennies qui précèdent l'accès à l'Indépendance sont marquées par une recrudescence des émeutes interreligieuses à travers l'Inde. Les musulmans, minoritaires, s'inquiètent de leur avenir dans une Inde indépendante peuplée très majoritairement d'hindous. En mars 1940, la Conférence de Lahore propose la création d'un foyer indépendant pour tous les musulmans du sous-continent indien, afin que leurs intérêts ne se retrouvent pas absorbés dans la future Inde.

À l'Indépendance, le Pakistan est constitué de deux entités territoriales, le Pakistan occidental et le Pakistan oriental, distantes de près de 1800 km de territoire indien. Ces

unités distinctes n'ont qu'un seul point commun, la religion majoritaire de leurs habitants : l'islam. En dehors de ce critère règne une grande diversité.

Le Pakistan oriental, qui deviendra indépendant en 1971 sous le nom de Bangladesh, possède une forte identité culturelle et linguistique bengalie.

Le Pakistan occidental se divise en quatre provinces : le Pendjab, où l'on parle pendjabi, la North-West Frontier Province[105], où l'on parle majoritairement le pashto, le Balouchistan avec pour langue dominante le balouchi, et le Sindh, de langue sindhi. Comme en Inde, aucune province n'est homogène sur le plan linguistique et les variations dialectales au sein d'une même langue sont très importantes (voir le potohari dans la région d'Islamabad, considéré comme une variété de pendjabi, avec lequel le taux d'intercompréhensibilité est pourtant faible, ou le shinwari, variante du pashto très isolée).

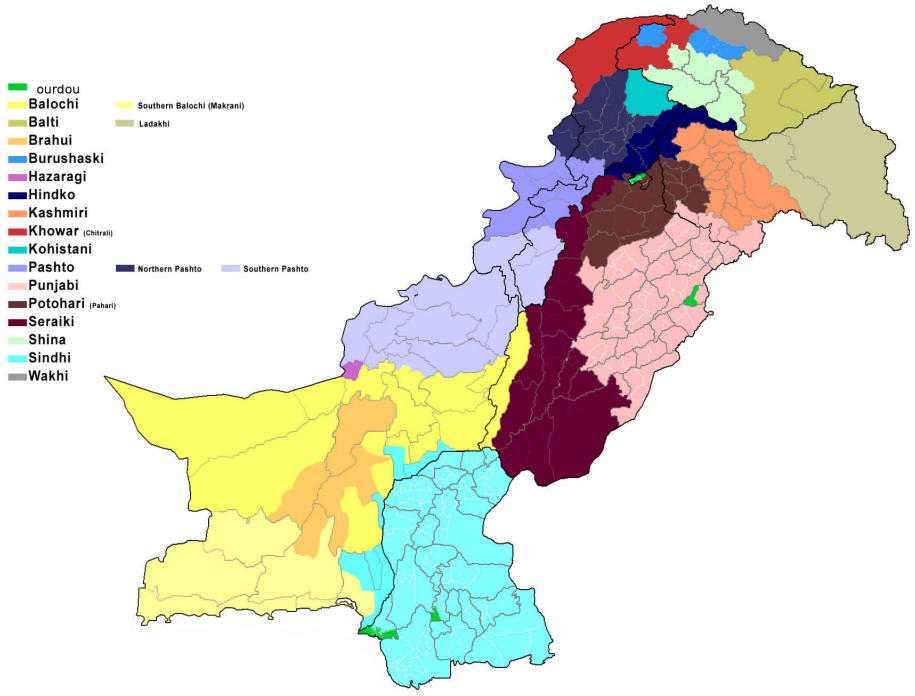

FIG. 1.13: Carte linguistique du Pakistan. Source : `https://commons.wikimedia.org/wiki/File:Pakistani_languages_map.jpg`

Sur le plan du nombre de locuteurs, le pendjabi est la langue la plus parlée (entre 44% et 54% de la population, selon le mode de comptage[106]), suivi par le pashto

[105] Aujourd'hui Khyber Pakhtunkhwa (KPK).

[106] Le pendjabi possède de nombreuses variantes et, n'étant pas langue d'enseignement, n'est pas standardisée. Le dernier recensement national porte à 44,15% le nombre de locuteurs de pendjabi sans le siraiki (ou saraiki), considéré par de nombreux linguistes comme une simple variante dialectale qui couvre le sud de la province. En incluant cette variante, le chiffre monte à 54,68% de la population.

(plus de 15%) et le sindhi (14%)[107]. L'ourdou, quant à lui, arrive en quatrième position et n'est majoritaire dans aucune province. C'est à l'origine la langue d'une élite urbaine, aujourd'hui encore langue première de moins de 8% de la population du pays. Elle est identifiée aux Mohajir (s *mohājir*, les immigrants musulmans venus des territoires aujourd'hui situés en Inde à partir de 1947. Comme l'islam, l'ourdou se voulait ciment de la population globale du Pakistan, au-delà des différences ethniques, de caste ou de langue, selon la formule du premier dirigeant du pays Ali Jinnah : « One nation, one language » [paragraphe 6.2.2.2][](Hock et Bashir, 2016), car s'il est peu parlé en tant que langue première, il est largement utilisé dans les échanges en dehors du cercle des intimes et du village. L'ourdou est donc choisi comme langue officielle du nouveau pays avec l'anglais[108], à l'exclusion de toutes les autres langues régionales.

1.2.3.2 Textes législatifs entérinant le statut de langue officielle

Ce statut privilégié de l'ourdou est entériné dans la Constitution dont le pays se dote en 1973, à l'article 251:

> *Section 251– National language.*
> 1. The National language of Pakistan is Urdu, and arrangements shall be made for its being used for official and other purposes within fifteen years from the commencing day.
> 2. Subject to clause (1), the English language may be used for official purposes until arrangements are made for its replacement by Urdu.
> 3. Without prejudice to the status of the National Language, a Provincial Assembly may by law prescribe measure for the teaching, promotion and use of a provincial language in addition to the national language.

En français :

> 1. La langue nationale est l'ourdou et des aménagements seront réalisés afin de lui permettre d'être utilisé à des fins officielles ou autres dans un délai de 15 ans à compte de ce jour.
> 2. En vertu du paragraphe (1), la langue anglaise peut être utilisée à des fins officielles le temps que les aménagements nécessaires soient réalisés pour son remplacement par l'ourdou.
> 3. Sans porter atteinte au statut de la langue nationale, une Assemblée Provinciale peut, en vertu d'une loi, prescrire des mesures pour l'enseignement, la promotion et l'usage d'une langue provinciale en sus de la langue nationale (*art. 251*).

Comme pour le hindi en Inde, il était prévu initialement que l'ourdou soit, à moyen terme, la seule langue officielle de l'État. D'autre part, l'article 251 autorise une place aux langues régionales en marge de la langue nationale.

---

[107]Source : `http://www.pbs.gov.pk/sites/default/files//tables/POPULATION%20BY%20MOTHER%20TONGUE.pdf`, consulté le 20-06-2019, sur le site du Pakistan Bureau of Statistics..

[108]Rahman (2004) indique qu'en 2000, 105 millions de locuteurs peuvent prétendre à un degré de maîtrise de l'ourdou au Pakistan, contre seulement 26 millions pour l'anglais.

Plus loin dans la Constitution, l'article 255, qui concerne les prestations de serment, autorise en théorie l'intéressé(e) à prêter serment « de manière préférable en ourdou » mais aussi « dans une langue comprise par cet individu »[109].

Dans les premières décennies du Pakistan indépendant, les Mohajirs, dont la langue maternelle est l'ourdou, sont nombreux aux postes de pouvoir. Leur attitude vis-à-vis des langues « provinciales » est perçue comme arrogante et provoque du ressentiment envers l'ourdou (Rahman, 2004), surtout chez les Bengalis du Pakistan oriental qui perçoivent l'imposition de l'ourdou comme une forme d'impérialisme de l'élite pendjabie du Pakistan occidental.

Dans la pratique, on décourage l'usage des langues provinciales à l'école, sauf dans le Sindh, et depuis la création de la province du Khyber Pakhtunkhwa en 2012, où cinq langues régionales ont été introduites dans les cursus scolaires.

Au niveau administratif, le Sindh reconnaît depuis près de 30 ans le sindhi dans les usages civils et judiciaires : les actes de naissance délivrés localement portent des mentions en cette langue, en plus de l'ourdou ; les greffiers des tribunaux qui ne sont pas locuteurs natifs sindhis doivent passer un examen de compétence en cette langue dans les trois ans suivant leur nomination, et les greffiers de langue première sindhi doivent passer un examen similaire prouvant leur maîtrise de l'ourdou[110].

1.2.3.3 Perception de l'ourdou dans la société pakistanaise

(Rahman, 1999) présente l'ourdou comme situé à une place intermédiaire dans la hiérarchie sociale pakistanaise. D'un côté, il est l'un des éléments symboliques de l'identité nationale pakistanaise, tout en étouffant l'épanouissement des langues régionales pourtant numériquement plus importantes. Il est vécu par les différents mouvements de revendication régionalistes comme hégémoniste. La maîtrise de l'ourdou est un élément indispensable constitutif du *capital linguistique*, et ne pas le maîtriser cantonne le locuteur d'une langue provinciale à une position sociale inférieure. L'auteur reprend ici la notion développée par (Bourdieu, 1977) selon laquelle le capital linguistique est constitué par la capacité de parler la langue utilisée dans les domaines de pouvoir pour parvenir à devenir membre d'un « club » d'avantages. Entre l'apprentissage de l'ourdou et celui de sa langue locale, le locuteur non-ourdouphone tendra vers la première option qui lui

[109]Constitution du Pakistan, version en ligne consultée sur le site de l'Assemblée Nationale pakistanaise, `http://www.na.gov.pk/uploads/documents/1333523681_951.pdf`.

[110]Judicial Staff Service Rules, 1992, article 10.

permettra de progresser plus facilement dans les positions de pouvoir et de domination économique. Il est courant d'observer des familles dans lesquelles les parents, locuteurs natifs non-ourdouphones, parlent à leurs enfants en ourdou afin de favoriser leur réussite future.

D'un autre côté, l'ourdou se situe, dans la course au « capital linguistique », bien en dessous de l'anglais, co-langue officielle qui ouvre vers le monde. Si l'anglais représente la langue des anciens colonisateurs, c'est aussi la langue de formation de la génération des administrateurs du pays, politique et militaires, au moment de l'Indépendance. (Rahman, 1999) explique que son capital ne cesse d'augmenter au Pakistan comme en Inde car il ouvre la porte à de meilleures perspectives de carrière. Il reste réservé, hier comme aujourd'hui, à une élite. Par opposition, l'ourdou apparaît alors comme une langue aux valeurs conservatrices, liée aux partis religieux qui s'opposent à la modernisation de la société et s'accrochent à des valeurs identitaires.

1.2.3.4 Promotion de l'ourdou au Pakistan

(Rahman, 1999) souligne l'attitude ambivalente de l'élite pakistanaise sur la question linguistique. En effet, le gouvernement semble promouvoir l'ourdou et d'autres langues locales dans les textes, à travers des initiatives telles que la proposition de création d'une National Language Commission[111] (qui n'a toujours pas vu le jour), ou la promulgation entre 2001 et 2014 de lois locales concernant l'usage des langues régionales.

Des institutions favorisant la diffusion est la promotion de l'ourdou ont même été créées. La National Language Authority a ainsi vu le jour en 1979, avec des missions similaires à son homologue indienne la CSTT : standardiser la langue, créer des ressources linguistiques en ourdou mais également dans les autres langues du pays, et traduire des documents scientifiques ou techniques de l'anglais vers l'ourdou. Le nom de cet organe a récemment changé en National Language Promotion Department [112] et est rattaché au Ministry of National Heritage and Integration.

Dans la réalité, le gouvernement octroie peu de moyens à l'ourdou, et encore moins aux langues provinciales, sur le terrain : l'école publique, en langue ourdou, manque cruellement de fonds tandis que l'élite politique et militaire scolarise ses propres enfants

[111] Voir l'article du journal Chitral Today : `http://www.chitraltoday.net/commission-to-set-criteria-for-national-status-to-local-languages/`.

[112] Voir le site institutionnel`http://www.nlpd.gov.pk/`.

en anglais[113]. L'élite conserve ainsi sa position en ne partageant pas l'accès au capital linguistique représenté par l'anglais avec les masses.

Ce constat est identique en Inde. Dans les deux pays s'est développé un accès inégal à l'éducation fondé sur la langue. Si en Inde l'éducation peut être délivrée dans les langues régionales ou en anglais, seul l'anglais permet véritablement la poursuite d'études au niveau supérieur et, de ce fait, est considéré comme la langue la plus prestigieuse. Cette attitude vis-à-vis des langues locales explique le manque de moyens mis en œuvre au développement des termes de spécialité et à leur promotion auprès d'une population beaucoup plus large.

1.2.4 *Politique linguistique en France*

À la différence de l'Inde et du Pakistan, la France est aujourd'hui un pays monolingue à l'administration centralisée. Cette situation est relativement récente au regard de l'histoire du pays. Bien qu'il existe des langues régionales qui peuvent bénéficier d'un soutien en matière éducative ou culturelle, aucune n'est reconnue pour des usages officiels.

1.2.4.1 Un pays, des patois

L'ordonnance de Villers-Cotterêts en 1539, sous François 1er, est connue comme texte fondateur faisant du français la langue d'administration. En réalité, ce texte contient avant tout des dispositions sur l'organisation de la justice. Il limite le rôle de l'Église dans ce domaine et renforce celui de la monarchie. Deux articles seulement, sur 192, concernent le français :

> *Article 110*
>
> Afin qu'il n'y ait cause de douter sur l'intelligence des arrêts de nos cours souveraines, nous voulons et ordonnons qu'ils soient faits et écrits si clairement, qu'il n'y ait ni puisse avoir ambiguïté ou incertitude, ni lieu à demander interprétation.
>
> *Article 111*
>
> Nous voulons donc que tous arrêts, et toutes autres procédures, soient prononcés, enregistrés et délivrés aux parties en langage maternel français et non autrement.

L'ordonnance ne sera pas appliquée dans les provinces qui seront annexées ultérieurement au royaume de France (Leclerc, 2019). À cette époque, 99% des Français parlent une autre langue que le français, qualifiée du terme général de « patois ». On distingue deux grandes zones linguistiques et culturelles : le pays des langue d'oïl, au nord, où

[113]L'armée pakistanaise, qui représente un lobby considérable, possède son propre réseau d'écoles en langue anglaise, construites sur des terrains publics et réservées aux enfants de militaires.

l'influence germanique est plus forte, et l'aire des langues d'oc, restées plus proches du fonds roman. À cela s'ajoutent d'autres langues d'origine distincte, comme le breton ou le basque.

Sous Louis XIV, la langue française se consolide. On la façonne en langue élitiste, prestigieuse, sans qu'il y ait besoin de légiférer. Le français reste cependant largement ignoré des habitants de la France jusqu'à la Révolution française de 1789.

1.2.4.2 Une République, une langue

À partir de la Révolution de 1789, les concepts de langue et de nation deviennent liés. Les patois sont considérés comme des idiomes féodaux de l'Ancien régime qui font obstacle à la vision universaliste exprimée dans les idéaux révolutionnaires d'égalité et transmise en français. Désignés comme promoteurs de l'obscurantisme social, ils doivent être anéantis. Une douzaine de lois linguistiques y veillera.

Le décret du 2 thermidor (20 juillet 1794) marque le début d'une politique d'hégémonisation linguistique, pendant la Terreur. Tous les actes administratifs doivent être rédigés en français et les agents qui enregistreraient ou délivreraient des actes en patois seront punis. Le 25 octobre 1795, le français est décrété seule langue d'instruction. Vers 1830, l'éradication des patois s'accélère au travers de consignes plus strictes et de punitions dans l'éducation. Tout au long du XX^{e} siècle, les langues régionales sont largement ignorées et il n'y a pas ou peu de textes législatifs concernant les droits linguistiques des minorités.

De manière assez insolite, la reconnaissance de la langue française comme langue de la République ne se fait qu'en 1992, par amendement de la Constitution de 1958. L'objectif de ce changement n'était pas dirigé contre les langues régionales mais contre l'influence grandissante de l'anglais dans le domaine des médias et des entreprises.

1.2.4.3 Promotion du français

La Délégation générale à la langue française et aux langues de France (DGLFLF), créée en 1989, promeut l'usage et le bon usage du français. Elle contribue aussi à son enrichissement. Elle travaille en collaboration avec des départements ministérielles, les secteurs professionnels et d'autres partenaires.

Dans le contexte européen, la France a signé la Charte européenne des langues régionales ou minoritaires. À travers ce texte, les pays européens s'engagent à financer des enseignements, des contenus pour les médias et des traductions des textes officiels en langues régionales. La France ne l'a toujours pas ratifiée car le Conseil constitutionnel a

jugé que le texte enfreignait l'article 2 de la Constitution, qui indique que la langue de la République est le français. L'adoption du texte aurait contrevenu au principe d'indivisibilité et d'unicité du peuple français. Bien que l'État français ne soit ainsi tenu par aucun engagement de promotion ou de sauvegarde des langues régionales, la DGLFLF a pour mission de les valoriser et de les protéger (par exemple en numérisant des corpus oraux ou écrits, en éditant des ressources numériques ou des applications en ces langues).

Paradoxalement, la France a fait du plurilinguisme un de ses objectifs prioritaires, avec le développement de ressources plurilingues et une attention soutenue au respect des langues nationales dans les réunions publiques ou les droits des consommateurs.

On constate donc que cette volonté du plurilinguisme s'attache surtout à ce que le français soit représenté, entre autres langues, sur le plan international. Il n'est pas dirigé vers l'acquisition d'une meilleure connaissance ou prise en charge des langues étrangères dans le pays.

La France contemporaine n'est pas confrontée à des enjeux identitaires semblables à ceux qui ont lieu dans le sous-continent indien à travers l'emploi des langues. Elle a pris le parti, au cours de sa récente histoire, d'éradiquer le plurilinguisme qui régnait sur son territoire au nom de l'égalité des citoyens. Elle lutte cependant pour le plurilinguisme à l'échelle internationale afin d'y permettre le maintien du français. Intégrée à l'Union européenne, elle est soumise à l'obligation d'assurer une assistance linguistique dans le champ juridique à ceux qui ne parlent ni ne lisent le français (article 6 de la Convention européenne de sauvegarde des droits de l'Homme (CESDH)). D'autres chartes d'assistance linguistique se développent également dans les services publics, sur la base du volontariat.

Trois types de politiques linguistiques se dégagent de cette étude. L'Inde, tout d'abord, s'attache à l'expression de sa diversité linguistique et met en place des institutions, des lois et des moyens pour lui permettre d'exister. On constate toutefois que les différentes langues ne sont pas traitées à égalité, et que l'anglais, suivi du hindi, tiennent une place privilégiée. L'expression linguistique officielle du hindi, qui fut définie au terme d'âpres débats, se veut pan-indienne mais reflète en réalité un repli sur la culture et la langue sanskrite, perçue comme indigène. Le Pakistan, quant à lui, a choisi une formule linguistique plus centralisée et néglige les langues régionales, au motif d'assurer l'unité et la cohésion nationales. Dans ces deux premiers pays, l'anglais reste une langue symbole de prestige et de pouvoir. La France, enfin, a adopté une politique monolingue musclée qui a fini par faire pratiquement disparaître ses langues régionales. Elle ne s'ouvre véritablement au plurilinguisme que dans le cadre de la législation européenne.

Une fois le cadre des politiques linguistiques respectives mis en place, il est possible de comprendre l'élaboration des terminologies en général et de la terminologie juridique en particulier, objet de la section suivante.

1.3 Élaboration des terminologies

La dynamique terminologique d'une langue dépend de la politique linguistique décidée par l'État. Ce sont les choix de celui-ci qui influenceront la façon dont les termes seront créés, développés, standardisés et diffusés. En Inde, la tendance après l'Indépendance est allée vers une sanskritisation à outrance (1.3.1), créant une langue artificielle et opaque. Nous verrons qu'il n'en a pas toujours été ainsi. Un mouvement similaire a opéré au Pakistan (1.3.2), dans le sens cette fois-ci d'une hyper-arabisation des termes, l'arabe était considéré comme langue de l'identité musulmane, ciment de ce pays nouvellement créé. Ces choix politiques ont des conséquences sur la langue juridique et judiciaire car ils permettent d'intégrer le justiciable au processus judiciaire en rendant la langue de la justice plus accessible. De telles questions identitaires ne se posent pas dans la terminologie du français (1.3.3), qui dérive du latin, car la langue est en France en situation de monopole.

1.3.1 *En hindi : vers la sanskritisation*

Dans un premier temps, nous étudierons le cadre du développement terminologique du hindi, qui débute vers la fin du XIX^{e} siècle, dans un contexte colonial, de compétition avec l'ourdou et d'effervescence scientifique. Nous verrons ensuite dans quelle mesure le hindi a puisé dans trois fonds linguistiques (sanskrit, arabo-persan et anglais) pour assurer son renouvellement terminologique.

1.3.1.1 Cadre du développement terminologique

Sous la colonisation britannique, les Indiens sont exposés à la diffusion des connaissances scientifiques et technologiques venues d'Occident. Des initiatives sont prises dès le XIX^{e} siècle pour les traduire dans les langues vernaculaires. Les concepts véhiculés étant très innovants par rapport à la culture indienne, on réfléchit déjà à la façon de former des termes nouveaux et autres équivalents de traduction.

Raja Rajendra Lal Mitra, considéré comme le premier indologiste indien et originaire du Bengale, publie en 1877 un fascicule intitulé *A Scheme for Rendering European Scientific*

Terms into Vernaculars of India[114]. Il prône la traduction de tout ce qui n'est pas nom d'objet technique ou binôme de la classification latine et propose de forger des termes à partir de racines sanskrites[115].

Missionné par le maharadja de Baroda, le professeur Tribhuvan Kelyandas Gajjar est chargé de traduire la littérature scientifique dans des langues indiennes.

À Bénarès, la *Kashi Nagari Pracharani Sabha*, fondée en 1893 pour promouvoir la diffusion du script devanagari, publie en 1898 des fascicules de terminologie scientifique en hindi. Elle publiera aussi des dictionnaires et une encyclopédie en hindi.

Enfin, à partir de la première moitié du XX^{e} siècle, Raghuvira s'engage dans un monumental projet lexicographique. Il entame ses travaux en 1931 pour publier son premier fascicule en 1943. Dans le premier volume de son dictionnaire, il affirme répertorier 150 000 termes et expressions. Son objectif est de passer en revue les deux millions de termes anglais de l'époque. Pour mener son projet à bien, il s'entoure de 200 spécialistes et effectue un travail de comparaison des dictionnaires terminologiques existants en anglais et dans diverses langues européennes, afin de les prendre pour modèle dans son élaboration des termes hindi. Du côté indien, il utilise des sources sanskrites[116], palies, prakrites et les langues indo-aryennes modernes. Les sources arabo-persanes, en dehors du lexique des mots des langues modernes indo-aryenne, semblent être exclues.

Dans son dictionnaire, (Vira, 1962, p. 49) expose sa méthode d'élaboration terminologique. Il revendique le sanskrit comme source primaire, duquel il titre des racines verbales, des suffixes et des préfixes[117] qui peuvent se combiner afin de créer tous les néologismes nécessaires, en comparaison avec la formation des néologismes des langues occidentales, fondée sur des racines et des formants gréco-latin[118]. Il instaure quelques principes généraux pour limiter la dérivation et les combinaisons :

- univocité du terme : il ne doit y avoir qu'un seul sens primaire ;
- un terme simple anglais doit correspondre autant que possible à un terme simple

[114]Ray (1983).

[115]Pradip Kumar Bose, *Health and Society in Bengal*, 2006, SAGE publishing, Inde.

[116]Ses sources en sanskrit sont : Kośhas, XIX^{e} et XX^{e} siècle : *śabda-kalpa-druma* de Rādhākātadeva, et des dictionnaires bilingues : sanskrit / allemand avec *Sanskrit-Wörterbuch*, d'Otto Böhtlingk et Rudolph Roth, le dictionnaire sanskrit / anglais de Monier-Willians et de Franklin Edgerton.

[117]Suivant les descriptions de Pānini, Raghuvira recense près de 500 racines usitées.

[118]Raghuvira recense 33 suffixes primaires, que l'on peut ajouter aux racines verbales, 47 suffixes secondaires ajoutés aux adjectifs et aux noms et 20 préfixes combinables entre eux pour former un total de 340 combinaisons.

en hindi ;

– les préfixes et suffixes sanskrits sont traduits par les préfixes et suffixes d'origine gréco-latine correspondants, dans les mêmes domaines de connaissance. Par exemple :

préfixe sanskrit	**préfixe gréco-latin**
pari	peri
anu	sub
san	con / syn
ap	ab
prati	anti
su	eu
dus	dys

Tab. 1.3: Équivalents des préfixes en sanskrit et en grec / latin

– les termes doivent avoir le même nombre de dérivatifs dans les deux langues. Exemple avec le terme vidhi (« law », loi) :
– un même terme peut avoir plusieurs formes abrégées (exemple en anglais : oxy-, ox-, oxo- pour « oxygen »)
– les termes créés doivent être faciles à prononcer pour la phonologie indienne ;
– la terminologie ainsi créée est panindienne ;
– pour les nouvelles idées, de nouveaux suffixes sont choisis. Ils dérivent de mots existants dont on ôte en général la première consonne ou syllabe pour la suffixer à une base portant le sens principal. Il n'existe pas plus de 12 nouveaux suffixes.

Cette méthode, centrée sur le point de vue occidental, sous-entend que la façon d'appréhender le monde et de se représenter les concepts est exactement la même dans les deux cultures, et transposable à l'identique dans les deux langues.

dérivé anglais	**dérivé sanskrit**	**traduction en français**
law	vidhi	loi
law-ful	vidhi-vat, vidhi-sangat	légal
legal	vaidh	légal
il-legal	a-vaidh	illégal
legisl-ation	vidh-ān	législation
legisl-ative	vidh-āyī	législatif (adj.)

Tab. 1.4: Dérivation à partir de *vidhi / law*

Les différentes tentatives de diffuser les savoirs vers les langues vernaculaires restent isolées. Elles ne sont pas appuyées par une volonté politique du gouvernement britannique qui, de son côté entreprend depuis le XIX[e]s. l'élaboration de dictionnaires et lexiques anglais-hindoustani pour ses besoins de gouvernance (lexiques de la marine, militaires)[119]. Les résultats sont assez variables, en fonction des informateurs. Dans les troupes militaires, de nombreux informateurs ne sont pas de langue maternelle hindoustani et les termes consignés peuvent être des variantes dialectales ou même issus d'autres langues (Saleem, 2013). Certains travaux sont remarquables, comme le dictionnaire de Platts[120], en trois graphies (nastaliq, devanagari et latine), qui servira très longtemps de référence à l'élaboration de dictionnaires bilingues anglais-hindi. Cependant, ces ouvrages écrits d'un point de vue britannique ne contribuent pas au développement des connaissances dans les langues indiennes.

Avant même l'adoption de la nouvelle constitution du pays, Raghuvira est officiellement mandaté en 1948 pour développer la terminologie dans le but de traduire des textes scientifiques et techniques anglais en hindi et dans les autres langues indiennes. Il publiera en octobre 1949, alors que les débats sur la tournure que doit prendre le hindi viennent de s'achever à l'Assemblée constituante, un premier dictionnaires administratif anglais-hindi. De nombreux dictionnaires techniques paraîtront dans les années suivantes, jusqu'en 1965.

En 1950, dotée d'institutions solides, l'Inde nouvelle met en place le Board of Scientific Terminology pour accélérer l'entreprise de traduction technique et terminologique. L'opinion du Conseil se divise entre les *puristes*, qui ne jurent que par le sanskrit pour l'enrichissement de la langue technique, les *expérimentalistes*, favorables à l'hybridation d'éléments de différentes origines pour fabriquer les termes, les *rétentionnistes*, adeptes de la conservation des mots anglais en l'état, et les *synthétistes* (Sharma, 1978). Comme vu dans la partie précédente, le point de vue des puristes finit pas l'emporter et la formation à partir du fonds sanskrit devient la règle. La CSTT, qui prend la suite du

[119]Le collège de Fort William, à Calcutta, est le centre de formation des officiers britanniques pour la langue hindoustani. On commence à y publier des ouvrages en deux graphies, nastaliq et devanagari, dès le début du XIX[e] siècle.

[120]Platts, John Thompson, *A dictionary of Urdu, classical Hindi and English*, Londres : W. H. Allen & Co., 1884. Ce dictionnaire sert notamment de base au dictionnaire anglais-hindi / anglais-ourdou en ligne de l'université de Chicago : `https://dsal.uchicago.edu/dictionaries/platts/`. McGregor s'en est aussi servi pour son dictionnaire bilingue hindi-anglais publié en 1994 (Oxford University Press).

Board of Scientific Terminology en 1961, pose le cadre de la formation lexicographique, parmi lesquels ces quelques principes[121]:

- les mots anglais ou étrangers dont l'usage est déjà très répandu (radio, télévision, téléphone) sont **translittérés** en nāgarī au plus près de la prononciation anglaise standard ;
- il est possible de former des **termes hybrides**, comportant une base lexicale provenant d'un fonds différent de celui de ses affixes : « *ionīkaran* » pour ionisation, « *oxīkaran* » pour oxydation, ou encore « *regiṣṭrīkaran* » pour enregistrement.
- Les **termes désignant des concepts** sont prioritairement traduits par des racines sanskrites : *urjā* (énergie), *śikṣā* (éducation), *kriṣī* (agriculture)…
- Dans le pensum de 2012, l'attention est portée à la *simplicité de la langue*, pour éviter « obscurantisme et purisme » ;
- l'identité maximale avec les autres langues indiennes, perçue comme exprimée dans la langue culturelle commune sanskrite, doit être recherchée.

Plusieurs problèmes demeurent, notamment celui des synonymes renvoyant à un même concept : la Commission est chargée d'assigner autant que possible un seul sens à chaque terme, de manière à désambiguïser la langue de spécialité. Le problème des calques de l'anglais, qui consistent à traduire littéralement à l'aide d'éléments sanskrits des éléments anglais, est omniprésent. On imite également la formation des mots savants construits à l'aide de formants grecs ou latins. Selon (Ray, 1983, p. 95), 70% de la terminologie hindi proviendrait de la translittération de termes internationaux ou de la traduction à partir de racines sanskrites, laissant seulement 30% à l'apport local du hindi ou des autres langues indiennes. (Sharma, 1978) souligne également ce mouvement d'emprunts ou de calques massif, qui ne se limite pas au lexique mais influence toute la phraséologie et la dérivation lexicale du hindi :

> During the last three decades, major Indian langages and specially Hindi has developped a number of devices to encode modern politico-economic, as well as scientific and technological concepts : but in the course of this development it has succumbed to « imitative transfer » in various degrees. Not only the meanings of English words but also their micro-structures at the morpheme level have been indiscriminately translated (e.g. « re + action + *ary* » , *prati* + *kriyā*+ *vādī*, far + east [extrême orient], *sudūr* + *pūrv*, etc.). Sentence and clause construction in the Hindi of the legal

[121]Ces principes sont rappelés dans des sortes de digestes, les rābhaṣā sahāyikā, qui, dans les administrations et les entreprises publiques, rappellent les lois sur l'usage du hindi, la grammaire de base de la langue ainsi que des glossaires de spécialité pour l'administration ou les branches techniques concernées. Les principes mentionnés ici sont tirés de (Gupta, 2008), chapitre 6, et de (Anonyme, 2012).

field has uninhibitedly moved in close step with the traditional legislative drafting in English.[122]

1.3.1.2 Fonds de création terminologique

Nous l'avons vu précédemment, le sanskrit représente le fonds lexical le plus important dans la création de termes en hindi. Les fonds arabo-persan et anglais constituent également des fonds de ressources non négligeables et, dans une moindre proportion, on trouve des termes créés à partir d'autres langues ou de langues indiennes non indo-aryennes.

1.3.1.2.1 Le fonds sanskrit

Langue de culture classique, le sanskrit se divise traditionnellement en trois types de lexies qui serviront de racines ou de formants aux termes hindi. Ces types se différencient par leur degré d'altération par rapport au mot sanskrit d'origine. Le premier type, appelé *tatsam* (littéralement « semblable à lui ») est un emprunt direct au sanskrit, sans modification. L'*ardhatatsam* (« semi-tatsam ») dérive d'un emprunt tardif au sanskrit, et le *tadbhava* est une lexie qui a suivi des modifications phonétiques au cours du développement de l'indo-aryen, dans les parlers populaires. Ce sont principalement les *tatsam*-s qui servent de base à la création terminologique en hindi. Quoi qu'il en soit, l'indo-aryen et le hindi ont subi un long développement linguistique depuis la période sanskrite. Quel que soit leur degré d'altération phonétique, tatsam, ardhatatsam ou tadbhava ont pu subir des évolutions sémantiques importantes, comme nous l'avons vu avec les exemples de lexies resémantisées *vidhi* et *dharma* (cf. 1.1.1.1.1).

1.3.1.2.2 Le fonds arabo-persan

Héritier de plusieurs siècles de culture moghole, il fournit sans doute plus de termes dans les domaines administratif et juridique que dans d'autres domaines de spécialité. Dans le sous-domaine judiciaire, on trouve par exemple *zamānat* (la caution) ou *ḥāżir(ī)* (fait d'être présent lors d'une procédure). Si la première lexie n'a pas d'équivalent d'origine sanskrite, la seconde est employée dans les textes normatifs en concurrence avec son

[122] Traduction : *« Pendant les trois dernières décennies, les langues indiennes majeures, et en particulier le hindi, ont développé de nombreux procédés pour encoder les concepts tant politico-économiques que scientifiques ou technologiques modernes : mais au fil de ce développement, [le hindi] a succombé au « transfert par imitation » à des degrés divers. Non seulement le sens des mots anglais a-t-il été traduit, mais également leurs micro-structures au niveau du morphème, sans discrimination. » .*

synonyme sanskritisé *upasthit(i)*. Les termes composés à partir de lexies de ce fonds sont encore plus nombreuses dans les textes de procédure pénale (non normatifs, tels que des jugements, rapports ou discussion sur des forums).

1.3.1.2.3 Le fonds anglais

Les lexies anglaises interviennent directement peu dans la création néologique. On trouve quelques cas d'hybridation avec des formants d'origine sanskrite comme *rajiṣṭrī-karaṇ* (« registry » + suffixe indiquant un processus), « enregistrement », ou *apīl-ārthī* (« appeal » + suffixe indiquant l'agent). Plus fréquemment, les termes anglais sont translittérés en hindi au plus près de la prononciation de l'anglais standard (près de 25% de la terminologie totale, comme *vāraṇṭ*, *rejiṣṭrār* ou *majiṣṭreṭ*) selon les principes énoncés plus haut par la CSTT. Plus rarement, lorsque les termes sont plus anciens, ou mieux ancrés dans la population, la translittération est modifiée pour suivre les usages phonétiques locaux, érosion similaire à celle des mots *tadbhava* issus du sanskrit : *rapaṭ* (« report », le rapport), *pulis* (« police »).

L'influence indirecte de l'anglais au niveau lexical est en revanche très importante sur le mode de création terminologique. De très nombreux termes sont des calques provenant de la traduction littérale des unités anglaises qui les composent (cf. chapitre 3). Ce processus est très visible dans les noms des postes et des entités administratives. On peut aussi le constater dans les noms de nombreux concepts. Dans un mot dérivé, la racine et les affixes anglais, souvent d'origine latine, peuvent être littéralement traduits par des racine et affixes d'origine sanskrite. Le terme obtenu devient relativement opaque pour l'individu qui ne connaîtrait pas le terme dans la langue source. Par exemple *adhivaktā*, l'avocat, est un calque de l'anglais *advocate* (d'origine latine), celui qui « parle pour » quelqu'un. *adhi* traduit le préfixe « ad- », et *vaktā* celui qui parle. Ou encore *homicide*, décomposé en « homme » et « meurtre », donnant *mānav vadh*, deux composants sanskrits ayant des sens identiques.

Une autre influence indirecte du fonds lexical anglais se retrouve dans la formation de termes dont le sens ou la phonologie converge vers le sens ou la phonologie du mot-source. Par exemple, dans le néologisme hindi *antarim*, on retrouve phonétiquement le mot-source anglais « interim » et le formant sanskrit *antar* dont un des sens signifie « qui sépare, qui se trouve entre ».

1.3.2 *Développement terminologique en ourdou*

Au cours du XIX^e^siècle, les identités culturelles et religieuses se cristallisent autour du hindi et de l'ourdou, variantes d'origine commune[123]. L'ourdou devint la variante de *khaṛī bolī* écrite en caractères nastaliq et persianisée, avec le soutien des publications britanniques du College de Fort William, à Calcutta, destinées à l'apprentissage de la langue de communication locale par les officiers militaires et administrateurs (cf. 1.3.1.1. Ces derniers devaient pouvoir communiquer avec leurs subordonnés dans des cadres spécifiques mettant en jeu des techniques et des sciences occidentales. La nécessité de faire évoluer la terminologie locale pour y intégrer ces nouveaux concepts était impérieuse. Tandis que les Britanniques ont élaboré des lexiques et dictionnaires mêlant les graphies devanagari, nastaliq et latines[124], les travaux des lexicographes indiens en ourdou et en hindi ont été menés séparément.

Le Fort William College fut la première instance officielle à formaliser le cadre pour la néologie et la création terminologique (Saleem, 2013). Il fut suivi en 1840 par le Delhi College, qui publia un recueil de règles de terminologie (Suleman et Dr. Sibtain, 2016, p.209).

Dans l'effort de modernisation qui occupe les progressistes du XIX^e^s. et de la première partie du XX^e^s., des lexicographes indiens se préoccupent d'absorber les nouvelles connaissances techniques venues d'Occident en ourdou. Au XIX^e^siècle déjà, plusieurs universités, comme Jamia Usmania à Hyderabad, proposaient des enseignements scientifiques en ourdou.

Le lexicographe Rai Sohan Lal, au XIX^e^ s., propose de définir une terminologie simple, dans les langues locales, en forgeant des termes à partir de leur définition et de leurs traits notionnels (Durrani, 1993)[125].

Le Nawab Amadul Malk Syed Hussain Bilgram, au début des années 1920, critique la démarche de Lal et propose de puiser davantage dans le fonds lexical arabe, qui possède des terminologies riches. Il prône de puiser uniquement dans le persan, l'arabe, le hindi et l'anglais pour les emprunts lexicaux, sans toutefois recourir à des mots arabes ou hindi trop complexes. Il encourage la création de locutions verbales à partir de bases

[123]Montaut (1997a) souligne à ce propos : « Paradoxalement, il y a plus de différence linguistique entre un dialecte occidental (comme l'haryanvi, ou le bangaru) et un dialecte oriental (comme le bhojpuri, le maithili) du vaste groupe des parlers hindi qu'entre hindi et ourdou parlés. »

[124]Cf. le célèbre dictionnaire de Platts, de la fin du XIX^e^siècle, qui sert encore de référence.

[125]Durrani, A. (1993), *Urdu Istalahat Nigari*, Islamabad : *Muqtadra Qaumi Zaban*. Cité par (Saleem, 2013).

nominales.

Cette ligne de conduite, allant vers une persianisation du vocabulaire sans toutefois rejeter totalement les apports du hindi – perçu ici comme variante sanskritisée – sera celle suivie par le département de traduction de l'université Jamia Usmania de Hyderabad, qui sera créé en 1917 pour produire des traductions à destination des masses. Un de ses éminents protagonistes fut Waheeduddin Saleem, dont l'ouvrage *Waz-i-Istelaahaat*, que fit paraître l'Anjuman Taraqqi-i-Urdu (mouvement de promotion de l'ourdou) dans sa première édition en 1921, expose les règles de formation des termes, donne la liste des affixes avec des exemples. Ce département travaillera jusqu'en 1950 sur le développement terminologique.

À cheval sur deux époques, celles pré- et post-indépendance, Molvi Abdul Haq avait publié un dictionnaire anglais-ourdou en 1937. Après l'Indépendance, il se met à réviser les anciennes terminologies. Haq (1949) puis Hassan (1969) posent les règles de la formation terminologique pour l'ourdou :

- si les termes scientifiques n'ont pas d'équivalent en ourdou, il faut les translittérer (emprunts écrits en caractères nastaliq) ;
- si des équivalents conceptuels existent en ourdou, ils doivent être utilisés. Si cette approche fonctionne pour les sciences « dures », les équivalents conceptuels sont en revanche en nombre beaucoup plus réduits dans le champ social ou dans un champ aussi marqué que le juridique ;
- si les termes sont des compositions de morphèmes non libres, on les conserve tels quels ;
- si le terme est une synapsie dont un des membres seulement possède un équivalent en ourdou, on peut utiliser une combinaison comprenant un élément traduit et l'autre translittéré.

Au Pakistan, le cadre du développement terminologique est assuré par le NLPD (National Language Promotion Development) dont les missions demeurent de fournir une terminologie, collecter les tentatives privées ou institutionnelles de création terminologique et dicter les règles de formation terminologique

Hindi et ourdou ont connu une phase de développement terminologique similaire à l'arrivée des colons britanniques. Deux mouvements allaient alors converger vers la création de dictionnaires et la traduction technique et scientifique de l'anglais vers les langues indiennes, après une phase de collection des textes religieux et juridiques des langues indiennes classiques vers l'anglais. D'une part, les Britanniques ont créé des glossaires à des

fins administratives, dans un but de se faire comprendre de leurs subordonnés et non dans une démarche encyclopédique. D'autre part, les Indiens comprenaient l'intérêt de diffuser les connaissances apportés par les Occidentaux dans des langues compréhensibles par les masses populaires. Ces deux mouvements ont divergé avec la cristallisation des identités culturelles et religieuses. Les principes de développement terminologique se caractérisent autour de l'Indépendance et ultérieurement par un tropisme pour les fonds lexicaux censés représenter les identités culturelles et religieuses associées à chacune des langues. Les premiers travaux de terminologie bilingue se sont appuyés sur des ouvrages écrits par des Occidentaux à l'usage d'Occidentaux, impliquant une conceptualisation qui était le fruit d'une tradition gréco-latine, non transposable telle quelle au moyen d'un simple calque de traduction dans les langues indiennes. C'est malheureusement ce qui a souvent sous-tendu la création terminologique en Inde, et qui explique l'opacité ressentie face aux terminologies actuelles. (Saleem, 2013) donne l'exemple de la traduction de l'anglais « atom », dont la définition « particule insécable » a été traduite en ourdou, mais n'est plus valide. Cette définition n'est pas transparente dans le mot anglais « atom », ce qui fait qu'il peut continuer d'être utilisé même si son sens n'est plus en accord avec son étymologie. Dans sa pratique, le traducteur-interprète qui cherche à utiliser les ressources issues de ces développements ressent profondément ce décalage entre une terminologie prescriptive, imposée par des instances officielles, et une terminologie par assez descriptive. (Saleem, 2013), lui-même traducteur, propose par exemple de cesser de s'acharner à utiliser des néologismes incompréhensibles pour remplacer des mots d'anglais déjà bien implantés dans les habitudes langagières de la population.

1.3.3 *Élaboration de la terminologie en français*

De manière peu surprenante, les termes du droit français viennent pour la plupart du latin, par évolution de la langue. Si la langue latine a longtemps été utilisée pour les usages administratifs jusqu'au XVI^{e} siècle, cela n'empêcha pas le développement et l'usage parallèles de termes juridiques et judiciaires en français, souvent issus du latin. Certains de ces mots reprennent des termes juridiques latins, comme avocat, déjà attesté au XII^{e} siècle, d' « advocatus »(celui que l'on appelle comme conseiller dans un procès, du verbe « advocare », appeler). Ce n'est cependant pas toujours le cas : les mots du droit peuvent provenir de termes agricoles ou ruraux (ex : cour, attesté depuis le XI^{e} siècle dans le sens de « siège de justice », provient du latin « cohors »et du latin médiéval « curia »(coin de ferme) ; « parquet », qui désigne aujourd'hui le service du

Procureur de la République, vient du bas latin « parricus », désignant un petit enclos). Un certain nombre de termes utilisés dans le domaine juridique le sont aussi ou l'ont été dans le domaine religieux, ce qui est assez cohérent avec les sources du droit français et le rôle qu'a longtemps joué l'Église dans l'administration de la justice.

S'il serait trop ambitieux de décrire ici la formation de la terminologie juridique et en particulier pénale en français, on dira quelques mots sur les institutions qui l'encadrent (1.3.3.1), les procédés de création terminologique actuels (1.3.3.2), et l'inscription de la France dans l'espace multilingue plus vaste que constitue l'Europe (1.3.3.3).

1.3.3.1 Les institutions de contrôle terminologique

La création terminologique juridique est contrôlée par des instances officielles relevant de l'État. D'une part, le Ministère de la Justice possède, comme la plupart des ministères, sa Commission spécialisée pour la terminologie : la bien nommée Commission de terminologie et de néologie en matière juridique, créée en 2003, qui établit des rapports quadriennaux de son activité. Cette Commission est liée, comme les autres commissions ministérielles, à la Commission d'enrichissement de la langue française [126]. La Commission générale valide les propositions de groupes d'experts, qui incluent des chercheurs du CNRS.

L'AFNOR (Association française de Normalisation), né en 1926, et placée sous la tutelle du Ministère de l'Industrie est chargée d'entériner ces termes qui deviendront des normes d'application obligatoire. Bien que cet organisme s'occupe avant-tout de la terminologie des domaines techniques, il liste également les terminologies relatives aux services publics et travaille donc en lien avec la Commission.

Cette dernière travaille principalement selon deux axes d'enrichissement terminologique : la traduction de mots ou expressions issus de l'anglais – principalement –, et le renouvellement du lexique à l'occasion d'une refonte de textes législatifs, pour l'adapter aux changements sociétaux, ou dans un souci de simplification du langage administratif. Par exemple, à l'occasion de la refonte du Code pénal, les termes d' « inculpation »et de « détention préventive »ont été remplacés par « mise en examen »et « détention provisoire ». Le premier terme, « inculpation », était jugé trop partial par rapport au principe de présomption d'innocence.

Les nouveaux termes sont ensuite publiés au Journal Officiel, à partir de quoi leur emploi devient obligatoire dans les administrations. Les néologismes peuvent aussi être

[126]Commission générale de terminologie et de néologie jusqu'en 2015.

définis dans les nouveaux textes de loi, comme dans le système indien ou pakistanais (cf les premiers articles du IPC, définitoires).

Malgré ce caractère contraignant et obligatoire, certains termes ne connaissent pas le succès escompté et les juristes peuvent continuer d'utiliser les anciens termes (par exemple, refus d'utiliser « majeur protégé »au lieu de « majeur incapable ». Dans la langue de la justice utilisée par les non-spécialistes, on retrouve souvent ces anciens termes qui continuent d'être diffusé via les œuvres de fiction.

(Lino et Deschamps, 2016) indiquent que le droit pénal est un sous-domaine relativement peu productif, comparé, par exemple, au droit de la famille. En effet, assez peu de listes ont été publiées par l'AFNOR ces dix dernières années [127].

1.3.3.2 Mécanismes de formation terminologique

(Lino et Deschamps, 2016) citent deux grands types de formation terminologique : par matrice interne ou externe.

La formation par matrice interne indique une dérivation ou une composition à partir d'une ou plusieurs bases lexicales existantes.

- l'affixation permet, de manière classique, d'obtenir des dérivés (féminicide, se re-pacser). Il n'est pas certain que ces termes soient nés dans le domaine juridique, mais ils y ont fait souche.
- la composition par réunion de morphèmes non libres (textopornographie, pédopornographie, sexto) ou libres (synapsies telles qu' action de groupe).
- la mutation grammaticale, très utilisée autrefois : formation de substantifs à partir de participes (requérant, plaignant, appelant, délégué, inculpé, réfugié).
- la réduction : on forme des sigles qui deviennent des termes (JEX, JAP, CESEDA, FIJAIS…)

La formation par matrice externe consiste à emprunter des termes à une langue étrangère (en général l'anglais), puis à les adapter progressivement en français (ce fut le cas d'action de groupe, emprunté à l'anglais « class action »).

Outre la formation de termes nouveaux, l'évolution du droit oblige aussi parfois à redéfinir le sens de termes existants. Par exemple le mariage, qui peut désormais être contracté entre deux personnes de même sexe.

[127] La liste NOR : CTNX1328341K de 2013 contient une dizaine de néologismes concernant le droit pénal ; celle de 2011, NOR : CTNX1107922X, en compte quatre. On compte un néologisme en 2006 (NOR : CTNX0609246X) at en 2014 (NOR : CTNX1419591X).

1.3.3.3 Utilisation de la terminologie : rédaction des textes normatifs et projets multilingues européens

Une fois les termes créés, ils sont utilisés dans la rédaction des textes normatifs. Depuis 2005, un guide de légistique présente les normes de rédaction des textes normatifs (lois, ordonnances, décrets, arrêtés)[128].

Les termes créés contribuent aussi à l'enrichissement de projets juridiques européens. Parmis eux, citons IATE, une base de données des traducteurs de l'Union européenne. Cette base ne propose cependant pas d'équivalents de traduction, au contraire de Legivoc, qui se propose de comparer les droits nationaux afin d'en aligner les concepts dans les diverses langues européennes[129].

Enfin, BabelLex, autre projet européen de traduction multilingue, s'inscrit dans le cadre de l'article 6 de la CESDH qui prévoit l'assistance linguistique à toute personne ne comprenant pas la langue de la procédure judiciaire. L'espace, qui n'est pas encore opérationnel, doit permettre la mise en relation de tous types de traductions judiciaires dans toutes les langues, pas seulement les langues officielles des États européens, puisqu'il s'agit de permettre l'accès au droit à tous les ressortissants étrangers, y compris hors Union européenne. La plateforme permettra aussi de trouver un interprète ou un traducteur dans la langue souhaitée[130].

La base de la terminologie judiciaire française, issue majoritairement de mots latins, existait déjà au moyen-âge. Cette terminologie est loin d'être figée. Elle continue de se développer avec la création terminologique qui suit à la fois des mécanismes internes au lexique de la langue, et emprunte aux langues avec lesquelles elle est en contact. Ces développements, qui lui permettent d'évoluer avec le droit, s'adaptent au contexte multilingue européen à travers des projets de traduction et de droit comparé.

Dans la section 1.1, nous avons analysé les différences notionnelles et systémiques de la justice en Inde, au Pakistan et en France. Il a été montré dans la section 1.2 que chacun des pays avait mis en place une politique linguistique particulière, soit pour

---

[128]Ce guide est téléchargeable depuis le lien `https://www.legifrance.gouv.fr/Droit-francais/Guide-de-legistique2`.

[129]Le projet n'est pas encore abouti, voir la fiche de présentation du projet `http://www.justice.gouv.fr/art_pix/LEGIVOC_fiche_20130506.pdf` ou (Vibert *et al.*, 2013).

[130]Pour une présentation du projet sur le site du Ministère de la Justice : `http://www.justice.gouv.fr/europe-et-international-10045/babellex-lacces-aux-langues-pour-le-droit-dans-lue-28953.html`.

gérer le multilinguisme en son sein, soit pour promouvoir une ou plusieurs langues dans le contexte international. Ces politiques linguistiques produisent des terminologies parfois opaques (1.3) et influencent la manière dont les langues sont considérées et traduites dans le contexte judiciaire. Le statut du traducteur ou de la traductrice-interprète s'en trouve affecté, ainsi que la possibilité pour le justiciable ne parlant pas la langue de la justice de bénéficier d'une assistance linguistique.

1.4 Place de la traduction judiciaire dans les trois États

Dans le droit français, la traduction en contexte judiciaire existe et est reconnue (1.4.1) depuis l'édit de Villers-Côteret qui impose l'usage du français dans les institutions de l'État. En Inde et au Pakistan, pays multilingues, la place de la traduction en justice reste étonnamment invisible (1.4.2) alors qu'elle est omniprésente. Elle semble être prise en charge par le greffe ou les avocats et il n'existe pas d'interprète ou de traducteur dont le rôle soit déterminé par les textes. Ceci n'empêche pas qu'il existe, au moins en Inde, des conditions imposant la traduction des textes de loi et des jugements en anglais vers le hindi puis les langues vernaculaires (cf. 1.2.2 « Politique linguistique en Inde »).

1.4.1 *France : une place indispensable, reconnue mais mal comprise*

En France, l'interprétation et la traduction des divers documents de la procédure pénale permet d'une part d'éclairer le juge d'instruction, et d'autre part d'offrir une source d'information et de débats éventuels pour le parquet et les avocats (Monjean-Decaudin, 2012). Elle représente une garantie procédurale : toutes les parties ont accès aux mêmes informations et en prennent connaissance dans les mêmes conditions, principe du débat contradictoire.

Ce souci de garantie à travers la traduction orale existait déjà sous l'Ancien Régime, où dès 1670 une ordonnance obligeait le juge à nommer un interprète pour les personnes ne comprenant pas le français (Ordonnance de 1670, titre XIC, art 1, Monjean-Decaudin (2012)). Cette disposition a été reprise sous la Révolution dans le Code Brumaire de l'An IV (Monjean-Decaudin, 2012). Dans le Code d'instruction criminelle, qui a précédé l'actuel Code de procédure pénale, les articles 332 et 333 reconnaissent l'assistance d'un interprète. La reconnaissance de la traduction s'est faite plus progressivement.

Dès l'origine, le rôle de l'interprète et celui du traducteur étaient perçus comme d'importance différente. Cette particularité tient sans doute à la conception du droit comme

un domaine quasi-sacré et au fait que l'histoire de la traduction soit intimement liée à l'étude des textes religieux (Monjean-Decaudin, 2012, pp. 231-237).

L'assistance à un interprète et la traduction des actes de la procédure est également garantie au niveau européen qui a imposé des standards minimaux (Monjean-Decaudin, 2012). La CESDH, et notamment ses articles 5 et 6, définissent ainsi qui peut être bénéficiaire de cette assistance linguistique et quels sont les actes concernés, le tout dans un souci d'équité du procès.

Les bénéficiaires de ces garanties sont en tout premier lieu les personnes arrêtées : l'article 5, alinéa 2 stipule :

> « toute personne arrêtée doit être informée, dans le plus court délai et dans une langue qu'elle comprend, des raisons de son arrestation et de toute accusation portée contre elle. »

La personne accusée a également droit à cette même garantie, comme le réaffirme l'article 6 :

> « tout accusé a droit notamment à être informé, dans le plus court délai, dans une langue qu'il comprend et d'une manière détaillée, de la nature et de la cause de l'accusation portée contre lui. »

Dans les deux cas, l'assistance est offerte par l'État, quelles que soient les conditions de ressources. Ces dispositions ont été enregistrées dans la directive européenne 2010/64/UE, puis transposées en droit français dans la circulaire n° 2013-711 du 5 août 2013. On remarque qu'il n'est nullement fait ici état d'une quelconque assistance pour les parties civiles ou pour les témoins. Dans les faits, en tout cas en France, ceux-ci peuvent bénéficier des mêmes conditions d'assistance linguistique dans la procédure pénale.

Contrairement à l'Inde ou au Pakistan, non seulement l'interprétation et la traduction judiciaire sont-elles reconnues dans la procédure judiciaire française, mais les interprètes et les traducteurs aussi. Ils ont, pour les experts, le statut d'auxiliaire de justice (cf. chapitre 2). Toutefois, nous le verrons au chapitre suivant, cette reconnaissance varie selon les phases de la procédure et il reste des malentendus quant aux méthodes à employer pour travailler en bonne entente avec les acteurs judiciaires.

1.4.2 *Inde et Pakistan : la traduction, une phase invisibilisée*

Dans les droits indiens et pakistanais, le traducteur judiciaire n'a pas de statut officiel. Les actes de la procédure sont traduits au sein des tribunaux. En Inde, on confie la traduction judiciaire à la personne reconnue comme la plus compétente dans la combinaison linguistique recherchée. Cette personne est souvent un ancien employé de

l'administration ou du greffe du tribunal[131]. Quoi qu'il en soit, il n'est nulle part fait mention de l'acte de traduction et le traducteur n'est cité nulle part dans les traces de la procédure.

Au niveau des dépositions dans les commissariats ou des rapports de la police locale, tout se déroule en hindi. Le déposant est entendu par un officier qui consigne ses déclarations après les avoir reformulées afin de correspondre aux critères du genre textuel. Si l'affaire doit aller plus loin, jusqu'à la juridiction de la High Court, alors la procédure requiert l'anglais. En effet, si le hindi est habituellement utilisé à ce niveau, les mutations des magistrats d'un État à un autre impliquent que des magistrats non-hindiphones peuvent se retrouver à siéger dans des cours d'États hindiphones. L'anglais est ainsi devenu la langue de la consignation des débats et des jugements à ce niveau. Or rappelons que le justiciable indien, dans 90% des cas, ne comprend pas l'anglais (cf. 1.2.2.1). Il peut bénéficier d'une assistance linguistique, mais rien n'est obligatoire : il en va du pouvoir discrétionnaire du magistrat. Bien souvent, le justiciable s'en remet à son avocat. Lorsqu'une décision est rendue et que le condamné ne comprend pas la langue du tribunal, cette décision ne sera pas traduite ni à l'écrit, ni oralement par un professionnel. C'est l'avocat qui se chargera de faire comprendre à son client le sens du jugement.

1.4.3 *Conclusion*

Pour résumer, nous avons vu que les systèmes judiciaires pakistanais et indiens étaient très similaires, puisqu'ils sont originaires d'un fonds commun, mais pas identiques, car le Pakistan a évolué vers l'islamisation d'une partie de son droit pénal. Dans ces deux pays existe en outre une procédure alternative à celle des tribunaux classiques : le Nyaya Panchayat pour l'Inde, et les jirgas au nord-ouest du Pakistan. Les jugements émanant de ces juridictions traditionnelles peuvent être reconnus par les juridictions classiques. On peut en interjeter appel dans un tribunal de premier degré. Le passage en revue de l'histoire des institutions judiciaires dans ces deux pays a montré que si les concepts de droit et de justice n'étaient pas nommés comme tels, dans la réalité, justice et maintien de l'ordre étaient assurés par des institutions étatiques ou locales, religieuses ou coutumières, et faisaient souvent intervenir le consensus de la communauté.

Des anciennes dénominations sanskrites, il est resté assez peu de termes en usage actuellement. En revanche, la période musulmane a été fructueuse en production de termes

[131]Communication personnelle de M. Ranjeev Surana, avocat à la High Court de Jaipur, Rajasthan.

judiciaires qui se sont maintenus à l'époque coloniale et ensuite. En effet, les Britanniques n'ont pas plaqué leur législation dans leurs territoire coloniaux, ils se sont documentés et appuyés sur les systèmes du passé pour construire un système de droit pénal qui demeure en place actuellement.

L'Indépendance en 1947 marque une rupture linguistique et culturelle entre l'Inde et le Pakistan, chacun cherchant à s'identifier en opposition à son voisin. Le développement terminologique marque un repli vers les fonds lexicaux créant une forte identité culturelle (le sanskrit pour le hindi, l'arabe et le persan pour l'ourdou. À ce stade, les politiques linguistiques de ces deux pays ont favorisé ces divergences.

De son côté, la justice pénale en France provient d'un héritage gréco-romain et coutumier, avec une influence de l'Église catholique. Le système contemporain est hiérarchisé, sans possibilité de système de règlement alternatif au tribunal pour les affaires pénales. Les décisions sont rarement prises par des pairs, sauf dans le cas des jurys d'assises, qui n'existent ni en Inde, ni au Pakistan.

Le vocabulaire de la procédure judiciaire trouve son origine principalement dans le latin et le pays n'a pas eu besoin de recourir massivement à la production terminologique dans le domaine. Il ne fait donc pas face au problème d'artificialisation de la langue de spécialité que connaissent le hindi et l'ourdou.

Enfin, la France est des trois pays celui qui offre l'assistance linguistique la plus complète, avec une définition de l'interprète / traducteur de justice. Dans les autres pays, cette fonction n'est pas officielle.

Cette double présentation, à la fois juridique et linguistique, était nécessaire pour prendre la mesure de l'écart entre les notions, les institutions et les langues à traduire dans cette étude. Dans la partie suivante, nous définirons plus concrètement l'intervention des interprètes et traducteurs et analyserons la langue à laquelle ils ont affaire. Nous verrons en particulier que le point de vue des traducteurs-interprètes et celui de la justice sur la mission des premiers sont parfois en conflit.

Chapitre 2

La traduction judiciaire

L'appellation « traduction juridique » évoque en premier lieu la traduction du droit en général. Celui-ci se divise en de nombreuses branches : droit commercial, droit humanitaire, droit international, droit de la propriété intellectuelle… Le domaine judiciaire ne couvre qu'une petite partie de ce champ juridique, celle qui occupe la justice et ses institutions. Il se subdivise lui-même en plusieurs branches, l'une civile et l'autre pénale, cette dernière constituant l'objet de notre étude[1].

La traduction judiciaire est un champ d'étude récent (2.1.1) étroitement lié à la réflexion sur la traductologie des textes juridiques. Cette réflexion a mis en lumière les spécificités de la traduction judiciaire (2.1), qui doit composer avec les exigences de la justice (2.1.2). Nous avons vu au chapitre 1 les bases des systèmes juridiques français, indien et pakistanais. Dans la pratique, le traducteur-interprète judiciaire doit également connaître le déroulement de la procédure (2.2) dans chaque pays, afin d'identifier les phases, les acteurs et les concepts qu'il sera amené à traduire ou interpréter. En France, le traducteur-interprète peut intervenir à n'importe quelle étape de la procédure. On peut lui demander de traduire des documents appartenant à des genres textuels variés (2.3), caractérisés par des phraséologies diverses. Cette étape permettra de relever les notions et unités phraséologiques à retenir pour l'étape suivante d'extraction terminologique et d'alignement en vue de rechercher des équivalents de traduction (voir chapitre 3).

2.1 Spécificités de la traduction judiciaire

2.1.1 *Un champ d'étude récent*

La réflexion sur la traduction judiciaire est très récente et découle de la réflexion plus vaste sur les stratégies de traduction qui animent la traductologie, discipline qui

[1]Pour la notion de droit pénal, le lecteur est invité à se reporter au 1.1.2.1.3.

s'émancipe de la linguistique dans les années 1970 outre-Atlantique.

Au départ, la traduction juridique émane de la linguistique juridique qui a été initiée par des juristes plutôt que par des linguistes. Il s'agissait alors, pour les juristes, de clarifier les éventuels problèmes d'interprétation des textes juridiques. Au sein même de la linguistique juridique, ce sont les champs de la lexicologie et de la lexicographie qui ont été en premier lieu les plus fertiles. Puis la réflexion s'est structurée dans le domaine de la traductologie, où l'on s'interrogeait sur les critères assurant une bonne traduction juridique. Le droit comme objet spécifique de la traduction a commencé à constituer une branche spécifique de la traductologie au Canada dès la fin des années 1970, où elle prendra le nom de jurilinguistique en 1982. En France, la traduction juridique devient un champ d'étude dans les années 1990. Elle connaît des développements orientés plutôt vers le droit ou vers la traduction, notamment avec la juritraductologie, au début des années 2000.

Nombreux sont les auteurs outre-Atlantique à s'être penchés sur la question, en particulier au Canada (Horguelin, 1977 et Gouin, 1977, Jean-Claude Gémar 1982, L.-P. Pigeon ,1982[2]. Des auteurs européens ont également animé la recherche, tels Gérard Cornu (1990), Claude Bocquet (1994) ou Jean-René Ladmiral[3].

Au fil de l'élaboration de la discipline, les spécificités de la traduction judiciaire sont apparues plus prégnantes à l'intérieur de la traduction juridique : la justice impose des contraintes qui lui sont propres (littéralité, fidélité) et qui semblent incompatibles avec une traduction réussie du point de vue du traducteur. À cela s'ajoutent la nature des textes, le caractère performatif, les objectifs visés de communication et les publics destinataires du travail traduit comme autant de défis à relever pour le traducteur.

L'adjectif judiciaire qualifie ce « [q]ui se rapporte à la Justice, à l'administration de la Justice. »[4]. Les textes judiciaires font partie d'une procédure pour laquelle la justice est sollicitée. Ces textes sont certes intégrés dans une procédure de droit mais ne sont pas obligatoirement des textes juridiques (Monjean-Decaudin, 2012), ce qui est déjà une particularité de ce domaine.

La traduction judiciaire, tout comme la traduction juridique dans son ensemble, est inscrite dans une double articulation heuristique. La première articulation consiste en

[2]cf. Gémar (2011), Elmer A. Driedger (1982) et Wagner et Gémar (2015).

[3]C'est à Jean-René Ladmiral que l'on doit l'introduction de la traductologie en France.

[4]source : `http://www.cnrtl.fr/definition/judiciaire`.

la compréhension puis en l'interprétation du droit dans sa langue d'énonciation. En effet, le droit a la particularité d'utiliser des mots de la langue commune dans des acceptions très spécifiques. Par exemple « renvoyer » une affaire ne signifie pas la rejeter mais, au contraire, la présenter lors d'une audience devant un tribunal. La seconde articulation consiste en l'effort de traduction vers une autre langue, un autre système de justice. Cet effort est d'autant plus difficile que le droit est extrêmement marqué culturellement, ce qui signifier que les référents utiles à sa compréhension ne sont pas universels : ils dépendent de l'histoire et du développement tout entier de la communauté qui le pratique (Gémar, 2002). Avant de traduire le droit, il faut donc connaître son propre système de droit puis le comparer à l'autre et le comprendre.

La traduction judiciaire est le sous-domaine de la traduction juridique attaché à l'institution de la Justice. Cette institution possède sa culture, ses référents et son mode de fonctionnement. Elle les impose au traducteur-interprète dont elle délimite le statut et les fonctions.

2.1.2 *La traduction judiciaire selon la Justice en France*

La justice française fait quotidiennement appel aux traducteurs et interprètes pour toutes sortes de missions. Le recours à ces auxiliaires de justice est une garantie procédurale. Il permet de faire avancer la recherche de la vérité, principe fondamental de l'instruction judiciaire, d'informer les parties et de nourrir le débat contradictoire. Toutefois, la justice réserve aux magistrats et autres juristes la complexité de l'opération d'interprétation (2.1.2.1). Elle impose au traducteur ou à l'interprète des exigences de fidélité et de littéralité qui contraignent l'exercice du traducteur-interprète et découlent d'une réflexion sans concertation avec les traducteurs (2.1.2.2). Les contraintes techniques imposées par l'institution de la Justice, qui impose un cadre souvent limitatif, ne tiennent pas compte de la réalité de l'effort de traduction ou d'interprétation..

2.1.2.1 L'interprétation du droit : une prérogative du juriste

Les textes sacrés sont à l'origine de la traduction et du droit dans le monde occidental, ce qui explique l'attachement à la littéralité comme gage de fidélité à l'original. On connaît la légende de la traduction des Septantes qui raconte que soixante-dix traducteurs de la Bible, animés par leur foi dans les saintes écritures, ont produit de manière isolée des traductions rigoureusement identiques. Cette légende postule qu'il n'y aurait qu'une seule bonne traduction possible d'une œuvre originale, et que la traduction ne laisse pas de place à la créativité ou à la subjectivité du traducteur. Une bonne

traduction serait donc une traduction littérale, dans laquelle n'apparaîtrait pas le rôle du traducteur, qui n'est qu'un outil de transposition d'un discours d'une langue à une autre. L'histoire de la traduction a été traversée par le débat entre traduction littérale, fidèle au texte original, ou traduction libre, fidèle au sens (Guidère, 2008, chapitre 3.8). Plus récemment, la traductologie, dans sa réflexion sur les stratégies de traduction, a montré que l'affaire n'était pas si simple et que le contexte historique, politique, culturel qui entourent une œuvre ou la culture réceptrice de sa traduction concourent grandement au processus traductionnel (Guidère, 2008). On reconnaît désormais diverses formes de fidélité, en fonction des objectifs remplis par la traduction et de ses destinataires. En somme, une traduction, tout comme une analyse de texte, est indissociable d'un contexte. Pour la traductologie, qui étudie la traduction en tant que processus et le produit qui en découle, la question clé des modes de traduction n'est désormais plus la fidélité mais la recherche d'équivalence de traduction, qu'elles soient formelles, sémantiques ou fonctionnelles[5].

2.1.2.2 Traducteurs et interprètes, techniciens transcodeurs

L'institution judiciaire française qui, comme nous l'avons vu, s'est posé la question du rôle de la traduction en justice depuis plusieurs siècles, a conservé le principe idéalisé et sacralisé de la fidélité en traduction [6] comme critère d'évaluation. Selon le législateur, la fidélité, gage d'exactitude, de conformité et de sincérité, ne s'obtient que par une traduction littérale[7]. Cette absence de reconnaissance de la complexité du processus de traduction a d'ailleurs animé de longs débats sur le statut légitime d'expert accordé au traducteur, l'expert n'étant pas seulement un technicien mais un spécialiste de son domaine.

Monjean-Decaudin (2012) explique que deux types de mission peuvent être assignées à un traducteur : pour certaines, il sera requis, tandis que pour d'autres, il sera commis.

Le traducteur qui travail sur réquisition est sollicité en tant que technicien : il doit effectuer une « traduction simple », c'est-à-dire au sens du juriste transposer un code linguistique dans un autre code. Il n'a pas d'avis à donner sur l'interprétation du contenu. On préférera qu'il fournisse une traduction littérale, même si celle-ci sonne étrangement ou si son sens est difficile à saisir.

---

[5]Guidère (2008), *op. cit.*

[6]Cette approche est un peu plus souple vis-à-vis de l'interprétation qui n'a pas le même caractère formel du fait même qu'elle est orale.

[7]Dans d'autres pays, le critère de fidélité n'est pas placé au premier plan. Au Canada, c'est le principe de continuité qui prime, et qui impose à l'interprète de ne pas résumer le contenu des échanges (Bergeron, 2002).

En tant qu'expert, le traducteur qui travaille dans le cadre d'une commission d'expertise n'est pas astreint à la littéralité : il peut donner son avis de spécialiste sur des faits de langues[8]. L'écrasante majorité des demandes de traduction sont faites dans le cadre de réquisitions[9], replaçant le traducteur dans un rôle d'exécutant, la place de l'interprétation devant revenir au juriste.

De leur côté, les juristes ont leur propre définition de l'interprétation, qu'ils considèrent comme une activité « consubstantielle à l'activité juridictionnelle » (Monjean-Decaudin, 2012, p.244). Il s'agit pour eux d'une recherche de la connaissance. À partir d'une interprétation sémiotique, le juge doit interpréter l'intention de l'auteur, soit au moment où il a écrit, soit au moment où ses mots sont interprétés. Il peut à cette fin s'éclairer d'autres textes. C'est au juge d'authentifier le sens des mots du droit et de leur attribuer une signification qui leur permettra de servir une fonction.[10] L'activité interprétative du traducteur est alors perçue par lui comme une interférence, une étape intermédiaire susceptible d'accumuler des biais de perception nuisible à la recherche de cette connaissance. De nombreux juristes ne perçoivent par la complexité du processus de traduction qu'ils considèrent comme une activité simple de transposition de code (Judith Lavoie)[11]. Certains magistrats préfèrent d'ailleurs donner le moins d'informations possible au traducteur-interprète afin de ne pas l'influencer dans sa mission, pensant se prémunir ainsi de tout biais interprétatif et garantir une totale neutralité du résultat. Or le traducteur-interprète a besoin de bien connaître le contexte du discours pour bien le comprendre et l'interpréter. (Lederer, 1994) rappelle qu' « on ne peut traduire sans interpréter »[12].

On perçoit ici le décalage entre les théories de la traduction développées par les traductologues, qui ont constaté l'échec de la traduction littérale, hors contexte, pour s'attacher à la recherche du sens et de la fonction, et le point de vue des juristes qui considèrent la langue comme un étiquetage de la réalité. Pour ces derniers, l'opération de traduction écrite ou orale consiste en un simple transcodage, l'opération d'interprétation

[8]G. le Poittevein, Code d'instruction criminelle annoté, t2, Paris, 1926, cité par Monjean-Decaudin (2012).

[9]La réquisition est une procédure plus simple, elle permet au juge de solliciter le traducteur qu'il souhaite sans justification (Monjean-Decaudin, 2012).

[10]J.-J. Le Sueur, Une introduction à la théorie du droit, p.147.

[11]J. Lavoie, « Droit et traductologie : convergence et divergence », in J.-C. Gémar et N. Kasirer (dir.), *Juriliguistique : entre langues et droits*, p.534.

[12]M. Lederer, La traduction aujourd'hui, Paris, Hachette, 1994, pp.9-15.

étant une exégèse visant à rechercher la connaissance juste d'un texte.

2.1.3 *Sélection et rôle du traducteur-interprète judiciaire*

Dans la partie 1.4.1, nous avons vu que l'existence du traducteur et de l'interprète judiciaire était actée depuis longtemps dans la justice française. Nous avons également vu que l'assistance à un interprète et la traduction des actes de la procédure étaient des garanties procédurales inscrites dans le droit européen. Pour rappel, le traducteur s'occupe de traduire des textes écrits, tandis que l'interprète s'occupe de traduire, ou d'interpréter, des propos oraux.

La justice expose certains critères de sélection et de contrôle des traducteurs et interprètes (2.1.3.1). Dans la procédure, l'interprète a un rôle direct et actif (2.1.3.2) à toutes les étapes. Les occasions d'intervention du traducteur (2.1.3.3) sont indirectes mais néanmoins cruciales.

2.1.3.1 Critères de sélection et de contrôle

(Monjean-Decaudin, 2012) expose les qualités et les garanties d'exercice du point de vue de la justice :

- l'interprète-traducteur doit être majeur (18 ans minimum),
- ce peut être tout personne à l'exception d'un membre de la cour, du greffe, des parties ou d'un témoin à l'affaire (article 344 du Code de procédure pénale). Il n'est donc pas nécessaire que l'interprète-traducteur possède des diplômes ou soit inscrit professionnellement comme tel,
- la justice doit s'assurer de ses compétences. Dans la pratique, ce contrôle n'est assuré par aucun organe particulier. La formation professionnelle n'est pas un critère de sélection, il n'y a pas non plus de test de compétence en langue.
- il doit être sincère dans sa mission et s'engage dans la prestation de serment « d'apporter son concours à la justice en son honneur et sa conscience » (art 102, 344, 407 CPP). On retrouve ici le caractère sacré du droit (Monjean-Decaudin, 2012).

Les critères de sélection de l'interprète ou du traducteur sont peu nombreux et le contrôle des compétences linguistiques absent. N'importe qui, parlant la langue recherché, peut devenir interprète après une prestation de serment. En complément, chaque cour d'appel possède une liste des experts traducteurs et interprètes, qui ont postulé en remplissant un dossier et fourni des justificatifs de leur savoir-faire. Les services judiciaires

peuvent puiser dans cette liste le spécialiste de la langue recherchée[13]. Là non plus, aucun test linguistique n'est imposé. La compétence en langue peut être valorisée par un diplôme acquis dans un pays étranger, même si le domaine étudié n'est pas en lien avec la traduction, l'interprétation ou le droit. Les experts, qu'ils soient traducteurs, interprètes ou autres, sont soumis une obligation de formation continue en matière juridique : ils suivent des conférences qui leur permettent d'acquérir des connaissances de base sur les principes directeurs d'un procès ou la langue juridique. Un traducteur-interprète expert présente donc plus de chances d'être familiarisé avec la langue du droit.

2.1.3.2 Rôle de l'interprète judiciaire

L'interprète intervient à toutes les étapes de la procédure, le plus souvent dans un rôle de liaison. Cela suppose qu'il interprète dans les deux sens, de la langue A vers B puis de B vers A. Avant l'audience du procès (si elle a lieu), il participe à l'information préliminaire (voir 2.2) : il retranscrit les interceptions de communications, fait la liaison entre interpellé(s) et force de police lors de l'interpellation, de la perquisition, des interrogatoires lors de la garde à vue. Il est également présent pour les personnes qui souhaitent déposer plainte. L'interprète est présent lors de l'instruction : lors des auditions par le juge d'instruction des différents protagonistes (mis en examen, témoins ou parties civiles), des confrontations, lors des expertises (notamment psychologique et psychiatrique dans le cas d'une procédure criminelle), et lors des entretiens entre client(s) et avocat(s).

Pendant l'audience, il traduit les propos de la personne pour laquelle il est convoqué, les questions de la formation de jugement qui lui sont adressées ainsi que les échanges de tous les protagonistes afin de permettre à l'intéressé de suivre correctement le déroulement des débats : la tâche qui incombe à l'interprète en audience est ardue et demande un effort soutenu, car les séances peuvent durer plusieurs heures, voire plusieurs jours, au cours desquels sa concentration est mise à rude épreuve.[14]. Il doit être partout à la fois et s'adapter à divers modes de travail : en simultanée lorsqu'il répète à l'intéressé les échanges qui se déroulent sous ses yeux (exposé des faits, questions et réponses des autres protagonistes, témoignages, rapports des experts, etc.), en consécutive lorsqu'il traduit les échanges directs entre l'intéressé et le juge. Lors des audiences où

[13]Il existe aussi des listes d'interprètes dits CESEDA auprès des tribunal de grande instance (TGI). Ceux-ci ne sont pas experts et n'interviennent que dans les affaires concernant le droit des étrangers.

[14]En comparaison, l'interprète de conférence, remplacé toutes les 20 minutes, ne traduit que de la langue A vers la langue B, ou de B vers A.

de nombreux intervenants parlent la même langue, il est d'usage de recourir à plusieurs interprètes pour alléger la tâche, surtout si les protagonistes relèvent pour les uns de l'accusation et pour les autres de la défense, mais ce n'est pas une obligation[15].

Après l'audience et la condamnation (2.2), l'interprète peut être amené à intervenir lors d'entretiens avec le juge d'application des peines ou en commission lorsque le détenu demande une autorisation de sortie ou sa libération avec une peine aménagée.

2.1.3.3 Rôle du traducteur judiciaire

La traduction sert deux catégories de public : le justiciable, non initié au droit, et le magistrat, qui le maîtrise.

L'assistance linguistique vaut aussi pour la traduction afin de respecter la règle d'équité du procès et le caractère contradictoire des débats. Il est important que la personne non francophone puisse accéder aux pièces de la procédure et les comprenne. Implicitement, cela signifie que l'on doit les lui mettre à disposition dans une langue et un niveau de langue qu'elle comprend. D'après les disposition de la CESDH[16], et après des débats qui ont duré sur des décennies (Monjean-Decaudin, 2012), ce droit à la traduction ne porte que sur des actes limités. Par exemple, la règle de droit stipule, comme vu précédemment (1.4.1) que l'acte d'accusation soit traduit à la personne accusée. En revanche, il n'y a pas d'obligation à lui traduire l'acte de jugement (Monjean-Decaudin, 2012). En général, la priorité dans la traduction est donnée aux acte de privation de liberté, afin que la défense puisse les contester. Cette sélection s'explique dans la mesure où la traduction écrite demande du temps et des ressources humaines tandis que le temps de la procédure impose des délais à respecter et des ressources limitées[17].

La traduction sert également au juge instructeur. Elle lui permet de prendre connaissance du contenu de pièces en langue étrangère versées au dossier. Ces pièces peuvent être de nature très diverses, allant de l'agenda retrouvé lors d'une perquisition au contrat de travail en passant par les courriers échangés entre un détenu et ses proches, des actes d'état civil ou des pièces de procédure judiciaire envoyées dans le cadre d'une

[15]Il m'est ainsi arrivé de me retrouver lors d'une reconstitution avec trois autres collègues interprétant la même langue, chacun de nous représentant une personne impliquée ou témoin, et d'avoir lors d'une audience aux assises à interpréter en continu pour des parties civiles et ponctuellement pour des témoins de la défense.

[16]Convention européenne de sauvegarde des droits de l'homme.

[17]Les traducteurs et les interprètes judiciaires sont rémunérés par le Ministère de la justice en leur qualité de COSP (collaborateurs occasionnels du service public).

coopération pénale internationale (cas des procédures antiterroristes, par exemple).

L'interprétation et la traduction judiciaire sont exercées par des interprètes et des traducteurs qui n'ont pas obligatoirement étudié le droit ou la traduction-interprétation. Cette situation est en cohérence avec le point de vue de la justice, pour laquelle un traducteur ou un interprète n'exerce pas de réflexion sur la langue ou sur le message (voir 2.1.2.1). La justice réserve l'opération d'interprétation, synonyme de quête de connaissance, au seul juriste, et relègue le traducteur-interprète au rôle de simple transcodeur. Ainsi, n'importe quelle personne parlant correctement une langue B sera potentiellement un bon traducteur-interprète, indépendamment de ses capacités de compréhension de l'affaire, de déverbalisation puis de reformulation du message.

Cette conception des opérations de traduction et d'interprétation par les traducteurs-interprètes va à l'encontre de la théorie interprétative de Lederer et Séleskovitch[18]. Ces interprètes de conférence ont théorisé le processus de transformation sous-jacent à la traduction-interprétation, travail sur le sens du message et non sur les mots. L'interprète ou le traducteur, après une phase de compréhension du message, l'assimile lors d'une phase de transfert appelée déverbalisation, puis le reformule afin de le ré-exprimer. Pour mener à bien cette opération traduisante, il doit donc bien connaître la langue de départ, le sujet du message, son contexte, et la langue cible. C'est pourquoi le traducteur doit comprendre comment fonctionne l'objet de son travail (dans notre cas, la procédure pénale), sans en être un spécialiste. Lors de la phase de restitution, ou reformulation, l'interprète-traducteur doit également prendre en compte le contexte d'énonciation : il doit adapter son expression au destinataire.

2.1.3.4 Les destinataires de la traduction

Pour Lederer et Séleskovitch, le travail de traduction est nécessairement une opération d'interprétation sur le sens du message (cf. 2.1.3.3). Il doit prendre en compte, en plus du contexte et de la maîtrise des langues source et cible, le destinataire. Dans la traduction et l'interprétation judiciaire, ces destinataires se divisent en deux catégories (cf. 2.1.3.3) : les initiés au droit et les non-initiés.

Parmi les initiés se trouvent les professionnels du droits, magistrats et avocats. Les non-initiés représentent les justiciables (auteurs d'infraction, victimes, témoins, experts non juridiques). Pour chacune de ces catégories, les objectifs de la traduction diffèrent. Le

[18]Pour une présentation plus approfondie, voir Herbulot (2004).

magistrat sollicitera une traduction dans le but de prendre connaissance de pièces versées au dossier (objectif informatif) ou pour transmettre un document entraînant des effets de droit (objectif juridique). Par exemple, dans une enquête antiterroriste, le magistrat envoie à un gouvernement étranger une commission rogatoire internationale afin que l'enquête puisse se poursuivre hors du territoire français.

Pour le justiciable, l'enjeu de la traduction et de l'interprétation est informatif. Une traduction lui permet de prendre connaissance du dossier, de ce qui lui est reproché, des mesures prises à son encontre ou de ce que les parties adverses ont versé au débat. Grâce à la traduction ou à l'interprétation, le justiciable est en mesure d'exercer son droit au silence, à contester ou à demander des actes procéduraux supplémentaires. Le justiciable n'étant pas un spécialiste de la langue juridique, le traducteur-interprète doit tenir compte de son idiolecte et de son bagage cognitif (par. 288 Monjean-Decaudin, 2012) afin de rendre le sens du message compréhensible. Non seulement le traducteur-interprète doit tenter de saisir l'intention de l'auteur, mais il doit aussi évaluer le niveau de compréhension du destinataire pour ajuster sa reformulation du message. Cependant, comme le rappelle (Monjean-Decaudin, 2012), il reste impératif de conserver la juridicité du contenu dans ce transfert de registre.

2.1.4 *Les équivalents fonctionnels de traduction*

Le débat sur la fidélité en traduction a fait rage pendant des siècles, pour ne pas dire des millénaires, renouvelant à chaque génération les débats entre traduction littérale, fidèle au texte original, ou traduction libre, fidèle au sens (Monjean-Decaudin, 2012) ; ou encore entre stratégie de traduction *sourciste*, privilégiant le texte source quitte à sonner étrangement, afin de transmettre l'atmosphère de la culture d'origine (Guidère, Venuti) opposée à la stratégie *cibliste*, qui tente d'adapter l'original au goût et à la culture du public qui lira l'œuvre, afin de rendre celle-ci compréhensible. L'essor de la traductologie dans les années 1970 déplace le débat en reconnaissant qu'il existe diverses formes de fidélité, en fonction des objectifs remplis par la traduction et de ses destinataires. Le problème actuel est plutôt dorénavant de trouver des équivalences de traduction, qu'elles soient formelles, sémantiques ou fonctionnelles.

Dans le cas de la traduction judiciaire, la recherche d'équivalents de traduction est donc là aussi nécessaire, puisqu'une traduction littérale, décontextualisée, serait incomplète et qu'il n'est pas possible de traduire des concepts juridiques sans comprendre ou expliquer au destinataire leur contexte juridique et culturel de référence.

Dans le cas de cette étude, d'une part, les langues source et cible possèdent peu de référents linguistiques communs. D'autre part, les systèmes juridiques des trois pays sont culturellement différents les uns des autres. L'apport de la culture juridique britannique, bien que ne relevant pas du même système que la justice française, a donné l'occasion de créer et de développer dans des langues d'Asie du sud des concepts inédits plus proches des nôtres (1.3) : les concepts de droit, la loi, l'appareil législatif en sont des exemples. Ils sont aujourd'hui en Inde et au Pakistan très éloignés du concept de *dharma* de l'Antiquité.

Nous avons vu en 1.3 comment les terminologues du hindi et de l'ourdou s'étaient attaqués au problème des équivalences terminologiques. Ils ont utilisés des termes anciens auxquels ils ont donné un sens nouveau (exemple de *vidhi*, la loi), dans le cas du hindi ils ont transposé des calques de l'anglais en sanskrit, en traduisant littéralement chaque formant d'origine latine ou grecque (exemple de *mānav vadh*, l'homi-cide), stratégies qui rendent la langue opaque pour le non initié. Ils ont encore eu recours aux emprunts depuis l'anglais (tels quels : *korṭ*, *juḍiśal*, ou déformés : *rapaṭ* pour « report », qui peuvent apporter un peu de transparence au non initié s'ils sont déjà implantés et utilisés dans la langue non spécialisée.

L'équivalent fonctionnel est un processus par lequel le traducteur vise au transfert du sens du texte dans la langue et dans la culture d'arrivée.(Monjean-Decaudin, 2012, p. 479). Il permet d'éviter les calques en mettant en parallèle des éléments de système différents. Par exemple, le procureur de la République n'a pas exactement le même rôle qu'un *prosecutor* dans le système de common law, mais il s'avère un bon candidat à l'équivalence fonctionnelle pour informer le justiciable de la personne qui le poursuit. Pigeon prône ce principe :

> « Le principe même de l'équivalence fonctionnelle signifie que l'on traduit en utilisant un mot qui ne correspond pas rigoureusement au même concept juridique mais à un concept analogue. »[19]

(Gémar, 2002) souligne que des équivalences de systèmes juridiques qui ont déjà été entérinées par la jurisprudence peuvent servir d'équivalents fonctionnels.

Dans le chapitre suivant, nous nous appuierons sur ces principes d'équivalence fonctionnelle pour établir des alignements entre des textes comparables en français, hindi et ourdou.

Traduction et interprétation font l'objet de champs de recherche récents. D'une part

[19]Pigeon, *La traduction juridique*. L'équivalence fonctionnelle, p. 280, cité dans (Monjean-Decaudin, 2012).

la jurilinguistique, qui associe le droit et la traduction, et d'autre part la juritraductologie, qui part du point de vue de la traduction pour mener une réflexion sur la spécialité juridique. Les avancées de ces recherches ont affirmé les spécificités de la traduction du droit, qui ne sont pas reconnues partout et restent en Inde et au Pakistan des disciplines techniques, malgré la polysémie bien plus importante de la terminologie juridique. Au sein de la traduction juridique, la traduction judiciaire possède des singularités qui tiennent à sa mission auprès de la justice. Cette dernière, qui dénie au traducteur-interprète sa compétence réflexive, s'oppose aux avancées des traductologues qui ont reconnu depuis des décennies que l'opération de traduction s'effectuait sur le sens du message et non sur les mots qui codent celui-ci. Pour comprendre le message et le reformuler en utilisant des équivalents fonctionnels si besoin, le traducteur-interprète doit maîtriser les termes de la procédure pénale selon les différents systèmes. Cette maîtrise passe par une connaissance des étapes de la procédure judiciaire et des différents genres textuels du domaine judiciaire.

2.2 Déroulement de la procédure pénale

La procédure pénale regroupe les étapes qui mènent du signalement d'une infraction à la décision finale d'un magistrat. Nous verrons successivement son déroulement en France (2.2.1), en Inde et au Pakistan (2.2.2).

2.2.1 *En France*

La procédure pénale est déclenchée par le signalement d'une infraction (2.2.1.1). L'instruction (2.2.1.2) qui s'ensuit peut aboutir à un procès (2.2.1.3) ou s'achever avant. Lors de cette procédure, diverses personnes exerçant des rôles précis (2.2.1.4) interviennent. Une des particularités de la procédure pénale, à retenir pour le traducteur-interprète, est son hétéronymie selon les degrés et les types de procédure.

2.2.1.1 Signalement d'une infraction

En France, lors d'une procédure pénale, le dépôt d'une *plainte* déclenche l'ouverture d'une *enquête* (art. 40 et 85 et suivants du CPP[20]) qui aboutit ou non à la tenue d'un procès au terme duquel est prononcé un jugement. Tout commence par l'ouverture d'une

[20]Code de procédure pénale.

information judiciaire[21]. Celle-ci est ouverte lorsque le procureur de la République est *saisi* d'une plainte (art. 40-1 du CPP). Cette plainte, appelée alors *saisine*, émane soit d'un citoyen ordinaire, soit d'un officier de police judiciaire ayant constaté une infraction, soit du procureur de la République qui a le pouvoir de s'autosaisir. Le déroulement d'une procédure judiciaire est tenu secret (art. 11 du CPP).

A travers son *réquisitoire*, le procureur décide ensuite de confier l'*information judiciaire*, c'est-à-dire l'instruction de la plainte, à un *juge d'instruction* (art. 80 du CPP).

L'instruction est obligatoire en cas de crime, mais pas pour les délits. La terminologie française diverge ici de l'anglais et du hindi-ourdou pour lesquels il n'existe pas d'hétéronymes distinguant les diverses catégories d'infractions. En effet, Le droit français distingue trois niveaux de gravité : la *contravention*, le *délit* et le *crime*.

Les contraventions sont les infractions les moins graves : stationnement interdit, légers coups et blessures, outrage, etc. Elles se règlent au tribunal de police et ne feront pas l'objet de la présente étude car leur procédure est directe : la *formation de jugement* est composée d'un juge unique et il n'y a pas d'appel possible. La traduction des contraventions ne fait pas partie des missions du traducteur judiciaire[22].

Les délits font l'objet d'une *procédure correctionnelle* ou délictuelle. Ils sont passibles d'une peine de prison inférieure ou égale à 10 ans et de dommages et intérêts égaux ou supérieurs à 3 750 euros. On trouve dans cette catégorie les vols, abus de biens sociaux, le harcèlement moral, les agressions sexuelles (hors viol[23]), les homicides involontaires et la fraude financière. Les délits sont jugés par un tribunal correctionnel formé de trois magistrats.

Type d'infraction la plus grave, le crime est traité selon la *procédure criminelle*. Cette catégorie regroupe les faits graves de fraude financière, meurtre et tentative de meurtre, assassinat et tentative d'assassinat, terrorisme, séquestration, violence volontaire grave, viol…). Contrairement aux délits, les crimes sont jugés en cour d'assise par un jury populaire dont le nombre de jurés peut varier, assisté du président de la cour et de ses assesseurs. Si les dommages-intérêts démarrent, comme en correctionnelle, à 3 750 €, la peine de prison prévue peut aller de 15 ans à la perpétuité.

---

[21] Le site service-public.fr propose une description claire et détaillée de la procédure : `https://www.service-public.fr/particuliers/vosdroits/N19807`.

[22] Communication de Mme Anton-Bensoussan, formation aux ETI de la cour d'appel de Paris du 17 juin 2017.

[23] Il existe une procédure de correctionnalisation de la procédure dans ce cas-ci.

La qualification d'une infraction en délit ou crime est fondamentale pour la terminologie de la procédure judiciaire : les actes et les acteurs y sont désignés par des termes différents selon que la procédure soit correctionnelle ou criminelle, bien qu'ils remplissent des fonctions similaires.

2.2.1.2 L'instruction

Le Parquet désigne un juge d'instruction qui sera chargé d'instruire le dossier à charge et à décharge. Il ne recherche donc pas uniquement des éléments défavorables au mis en examen mais doit déterminer s'il y a réellement eu infraction, quels en sont les auteurs, la nature et les circonstances (art. 49 du CPP).

Le juge d'instruction dispose de moyens d'action à travers les services de police judiciaire (policiers, gendarmes ou douaniers) qu'il mandate en fonction des besoins par des *commissions rogatoires* (art. 81 du CPP). Ces commissions permettent d'interpeller des personnes, placer des mis en cause sur écoute, perquisitionner, demander des expertises, auditionner des témoins, etc.

C'est le plus souvent dans le cadre de ces commissions rogatoires que les interprètes-traducteurs sont sollicités à ce niveau de la procédure : ils peuvent traduire des conversations téléphoniques ou des messages, interpréter lors de la perquisition ou de l'audition de témoins, ou encore traduire des scellés récupérés lors de l'intervention policière.

Si les éléments probants sont suffisants, une interpellation peut aboutir à un placement en garde à vue décidée par le Procureur (art. 62-3 du CPP), mesure de sûreté qui peut durer jusqu'à 48 heures ou 96 heures en fonction du type d'affaire (art. 63 du CPP). Pendant la garde à vue, le mis en cause[24] est interrogé plusieurs fois sur les faits, voire confronté à d'autres acteurs de l'affaire. Toutes les déclarations formulées lors des auditions sont scrupuleusement actées sous forme de procès-verbal. A défaut, l'intégralité de la procédure peut être remise en cause pour vice de forme.

Ici encore, l'interprète intervient et assiste aux interrogatoires de garde à vue. À l'issue de ces derniers, il relit les déclarations du procès-verbal à l'intéressé(e) qui le signe s'il en approuve le contenu.

Si les preuves de l'infraction sont suffisamment étayées, le mis en cause devient *mis en examen* (art. 80-1). On peut entendre ou lire dans la presse le mot « inculpé », terme hors d'usage dans la langue judiciaire, car il enfreignait le principe de présomption

[24]On ne parle pas de suspect dans la terminologie pénale actuelle, encore moins d'inculpé.

d'innocence. Le terme prévenu, lui, ne désigne que les justiciables traduits devant un tribunal de police ou un tribunal correctionnel.

Le mis en examen doit répondre d'un ou plusieurs *chefs de prévention* (en procédure criminelle on parle de *chef d'accusation*) pour des faits qu'on lui reproche. La mise en examen est décidée et notifiée à l'intéressé(e) à l'issue de l'*audition de première comparution* dans le cabinet du juge d'instruction, lors de laquelle le juge vérifie l'identité du prévenu et lui présente la *qualification* juridique retenue (art. 80-1) : il lui explique d'après quels textes de loi sont identifiés les faits qui lui sont reprochés.

D'autres auditions ont ensuite lieu, au cours desquelles le mis en examen répond aux questions du juge. Tout l'entretien est soigneusement retranscrit par le greffier et, dans le cas d'une procédure criminelle, filmé (art. 64-1 du CPP). Pendant l'interrogatoire, le mis en examen est assisté de son avocat. La retranscription du greffier est rédigée en procès-verbal qui est relu et signé par le juge, la personne interrogée et l'interprète s'il y en a un. Comme dans le cas des auditions ou des interrogatoires de police, les procès-verbaux du cabinet d'instruction ont des codes et formulations bien précis. Pendant la mise en examen, des commissions rogatoires permettent de poursuivre les investigations (enquête de la police judiciaire, expertises, etc…)

Dans le cas où il y a un risque de fuite du mis en examen, on dit alors qu'il ne présente pas de *garanties* suffisantes *de représentation*. Le juge d'instruction a alors le pouvoir de prononcer à son encontre une mesure de *contrôle judiciaire* le temps de l'instruction (art. 144 du CPP) ou de le placer en *détention provisoire*[25] en demandant au Juge des Libertés et de la Détention un *mandat de dépôt* (art. 141-2). La personne mise en examen est alors envoyée en *maison d'arrêt*. Ce cas de figure est relativement courant pour des étrangers, surtout s'ils sont en situation irrégulière, puisqu'ils ne disposent ni de domiciliation ni d'attaches suffisamment solides pour garantir qu'ils se présenteront aux convocations de la justice. L'interprète a par conséquent également pour mission de se rendre dans les maisons d'arrêt pour accompagner les avocats ou les experts rendant visite au détenu.

Tous les actes de l'instruction sont *contradictoires*, signifiant par là que toutes les parties à l'affaire sont informées de leur déroulement et des résultats (article préliminaire du CPP).

Lorsque le juge d'instruction estime qu'il a réuni suffisamment d'éléments, il délivre

[25]Le terme « détention préventive » n'est plus en usage en France mais l'est en Belgique.

un *avis à partie* qui implique que tous les actes ont été accomplis et que l'affaire peut être renvoyée devant une *formation de jugement* (art. 175 du CPP). Dans un délai de un à trois mois selon que le mis en examen est détenu ou libre, les parties et le procureur de la République doivent faire part de leurs observations et de leurs réquisitions. Les parties disposeront d'un délai supplémentaire de dix jours à un mois pour formuler des observations complémentaires. Le procès se déroulera donc au moins quatre mois après la fin de l'instruction.

2.2.1.2.1 Éléments spécifiques à la procédure correctionnelle

a. Comparution immédiate ou différée

Dans une procédure correctionnelle, le mis en cause peut être, en fonction des faits, présenté immédiatement devant le juge (comparution immédiate, art. 395 du CPP) ou être convoqué au tribunal à une date ultérieure (comparution différée, art. 388). En réalité, la comparution immédiate a lieu dans un délai allant de quelques jours à un mois et le mis en cause a toujours la possibilité de la refuser afin de prendre le temps de préparer sa défense.

Lors d'une audience de comparution immédiate, les affaires se succèdent toute la journée et sont jugées en quelques minutes. Le justiciable dispose d'un temps relativement court pour s'entretenir avec son avocat dans des box à proximité de la salle d'audience. Il passe ensuite un rapide entretien avec un enquêteur de personnalité chargé de retracer brièvement son parcours de vie et les éléments de sa personnalité qui semblent importants pour l'affaire jugée : charge familiale, situation financière, état de santé, etc. Ces données permettent de donner de la « chair » au dossier, de juger une personne et non simplement des faits. Pendant ce temps, le greffe recherche le casier judiciaire du mis en cause.

b. Comparution sur reconnaissance préalable de culpabilité (comparution sur reconnaissance préalable de culpabilité (CRPC) ou plaider-coupable)

Dans ce cas précis, le mis en cause reconnaît les faits qui lui sont reprochés et le procureur de la République peut directement lui proposer une peine, sans procès. Cette procédure n'est pas possible pour tous les types de délits, en particulier les plus graves (art. 495-7 du CPP).

2.2.1.2.2 Éléments spécifiques à l'instruction criminelle

Dans la procédure criminelle, à la différence des interrogatoires correctionnels, les interrogatoires devant le juge d'instruction sont filmés. Les expertises psychiatriques et psychologiques de l'accusé, ainsi que des parties civiles, sont obligatoires, ainsi que des enquêtes de personnalité plus approfondies, en vue d'un procès aux assises.

a. Expertises psychologique et psychiatrique

Pendant la phase d'instruction, obligatoire dans la procédure criminelle, la prise en compte de la personnalité de chacune des parties est plus déterminante. Avant tout jugement au fond, le mis en examen passe systématiquement deux entretiens avec des experts judiciaires rattachés à une cour d'appel (art. 706-115 du CPP). L'une de ces expertises est réalisée par un psychiatre qui doit déterminer si le mis en examen était affecté d'une pathologie mentale particulière qui aurait pu affecter son jugement au moment de la commission des faits (art. D47-21), tâche ardue s'il en est car la procédure s'étale parfois sur plusieurs années et le psychiatre est appelé à évaluer l'état mental dans lequel se trouvait une personne des années plus tôt. Cette mission se complique lorsque les individus mis en examen développent des pathologies dues à leur incarcération.

La seconde expertise est menée par un psychologue chargé de déterminer le niveau intellectuel du mis en examen et les principaux traits de sa personnalité. Ces experts ne disposent pas des pièces versées au dossier pendant l'instruction et n'ont qu'une connaissance sommaire de ce dernier. Par leurs questions, ils amènent le justiciable à parler de lui, de ses relations avec ses proches, de sa perception des faits qui lui sont reprochés.

b. Enquête de personnalité

Un enquêteur de personnalité peut aussi intervenir à ce stade (art. R121-1). Il prendra plus de temps qu'à l'occasion d'une comparution immédiate pour retracer le parcours de vie du justiciable et peut interroger des proches afin de corroborer certains points. C'est une étape importante qui permettra au mis en cause de se présenter sous un jour plus humain.

Ces trois entretiens existent également pour les parties civiles. Ils permettent de déterminer le retentissement psychologique et psychiatrique que les faits ont pu avoir sur elles. Cette évaluation sera prise en compte lors du jugement pénal et de l'indemnisation des parties civiles. La remise du rapport des experts marque généralement la fin de

l'instruction.

À l'issue de l'instruction correctionnelle, lorsque les éléments sont suffisants pour maintenir les charges, le juge d'instruction délivre une *ordonnance de renvoi devant le tribunal correctionnel* (ORTC). On parle d'une *ordonnance de mise en accusation* (OMA) dans le cas d'une procédure criminelle, en vue d'un procès à la cour d'assises.

2.2.1.3 Le procès

2.2.1.3.1 Au tribunal correctionnel

Une affaire peut être jugée en correctionnelle lors d'une audience qui regroupe d'autres affaires, c'est notamment le cas dans les procédures de comparution immédiate, ou occuper tout le temps d'une audience, voire se dérouler sur plusieurs jours. Le prévenu et la victime doivent être présents ou représentés par leur avocat.

Les affaires correctionnelles sont généralement jugées par un collège de trois juges : un président et deux assesseurs. L'un des assesseurs rappelle quels sont les faits reprochés et, dans un exposé concis, relate les étapes de la procédure. La formation de jugement interroge le prévenu, qui a le droit de garder le silence. Il peut ensuite être interrogé par le procureur de la République, qui représente les intérêts de la société, et par l'accusation, incarnée par l'avocat de la victime. Témoins et experts sont appelés à la barre pour apporter leurs témoignages et être tout à tour interrogés par le tribunal, l'accusation et la défense. Ils déposent sous serment, sauf dans certains cas[26]. Vient ensuite l'interrogatoire de la victime. Chaque prise de parole du tribunal ou de l'accusation doit donner à la défense l'occasion de répondre.

Les plaidoiries des avocats et réquisitoire du Ministère public suivent les débats. Elles exposent l'argumentaire de chaque partie visant à demander une condamnation ou l'acquittement du prévenu. Le dernier mot est à la défense.

La décision peut intervenir immédiatement après, pour des affaires simples, à la fin de l'audience lorsque plusieurs affaires sont jugées à la suite, ou être rendu à une autre date. On dit alors qu'il est *mis en délibéré*. Si le tribunal prononce une condamnation, on dit qu'il *entre en voie de condamnation*. Celle-ci est généralement accompagnée d'une peine, bien qu'une dispense soit possible dans certains cas. Lorsque la culpabilité du prévenu n'est pas établie, celui-ci est *relaxé*.

La peine peut être ferme, il faut alors l'exécuter, ou avec sursis, ce qui signifie

[26]Ne déposent pas sous serment les témoins qui ont un lien familial ou de subordination avec le prévenu.

que la peine n'est pas à exécuter moyennant certaines conditions. Dans le cas contraire, lors d'une condamnation ultérieure, la peine qui était en sursis viendra s'ajouter à la nouvelle. Les parties ont un délai de dix jours pour *interjeter appel*.

Les audiences sont généralement publiques. L'interprète, s'il est expert, n'est pas tenu de prêter serment avant le début de l'audience dans les tribunaux du ressort de sa cour d'appel. Dans le cas contraire, il doit jurer de prêter son concours à l'exercice de la justice en son honneur et conscience.

2.2.1.3.2 À la cour d'assises

Comme expliqué au chapitre 1 (voir 1.1.2.2), cette juridiction juge les crimes les plus graves[27]. Elle est saisie par une décision de *mise en accusation*, équivalent du renvoi en correctionnelle de la procédure délictueuse. À la différence du tribunal correctionnel, la cour d'assise n'est pas permanente et se réunit en sessions au cours desquelles plusieurs affaires sont traitées sur plusieurs semaines[28]. Au début de chaque session, des citoyens sont tirés au sort à partir des listes électorales pour former un jury populaire qui participe au jugement aux côtés des magistrats professionnels (art. 259 à 267 du CPP). À chaque affaire de la session, plusieurs jurés sont tirés au sort pour participer au jury. Ceux qui ne sont par appelés ou ont été récusés par les parties peuvent partir. Au début de la session, les jurés ignorent tout de l'affaire qu'ils s'apprêtent à juger[29].

À la différence d'un procès en correctionnelle, un procès aux assises est une procédure entièrement orale, et ne comptent que les déclarations faites au cours de l'audience. Les jurés n'ont pas connaissance des pièces du dossier, ils ne savent de l'affaire que ce qui leur est lu par le président sur le déroulé des faits et les phases de l'instruction, et peuvent prendre des notes.

La *partie civile* (s'il y en a une), les témoins et les experts viennent déposer à la barre et sont interrogés successivement par l'avocat général, qui représente le Ministère public[30], l'avocat de la (des) partie(s) civile(s) et celui de la défense. L'accusé (on ne parle plus de prévenu dans une cour d'assises) est interrogé à son tour par le tribunal et les

[27]Une nouvelle cour, la cour criminelle, a été créée par la loi du 23 mars 2019. Elle juge des majeurs ayant commis des crimes punis de peines allant de 10 à 20 ans de réclusion.

[28]Il est rare qu'une affaire aux assises dure moins de deux jours.

[29]Dans le cas d'une cour d'assises spéciale, composée pour juger des actes de terrorisme, il n'y a pas de jury et les juges sont tous des magistrats professionnels.

[30]Il est l'équivalent du procureur de la République à la cour d'assises.

différents avocats. La présence d'un avocat est pour lui obligatoire à l'audience. Comme dans la procédure correctionnelle, les interrogatoires et témoignages sont suivis par les plaidoiries et réquisitoire, et la parole est laissée en dernier à la défense. Le jury se retire alors pour délibérer, assisté par le président de la cour.

La décision d'un jury d'assises qui reconnaît la culpabilité ou l'innocence d'un accusé est un *verdict*. Si l'accusé est reconnu non coupable, il est *acquitté*. Sinon, il peut être condamné à une peine de prison, une amende, et à verser des dommages et intérêts à la victime (la partie civile n'est reconnue victime que dans le cas de la condamnation de l'accusé). Dans le cas où la peine de prison est supérieure à dix ans, on parle de *réclusion*. Comme en correctionnelle, les parties peuvent faire appel sous dix jours.

2.2.1.3.3 En appel

Lors d'un procès en appel, qu'il s'agisse d'un procédure correctionnelle ou criminelle, le déroulement reste identique. Seuls quelques termes changent alors que les fonctions demeurent.

- Le procureur de la République devient *avocat général*.
- La décision n'est plus un jugement, elle devient *arrêt*.

2.2.1.3.4 Détention

Le mis en cause placé en détention provisoire peut exercer son droit à demander une *remise en liberté* que le juge des libertés et de la détention (JLD) acceptera ou rejettera.

Après le jugement, on parle de *peine de détention* jusqu'à 10 années, et de *peine de réclusion* au-delà.

Un détenu peut bénéficier de *parloirs* pour recevoir des visiteurs ou s'entretenir avec son avocat. Il peut aussi être placé à l'*isolement*, sans contact avec les autres détenus ou des visiteurs. Les personnes condamnées sont suivies par le SPIP (service pénitentiaire d'insertion et de probation) pendant leur détention et après leur sortie.

2.2.1.4 Les gens et les fonctions de la justice

Tout au long de l'instruction, du procès et après la condamnation, le traducteur-interprète intervient auprès de professionnels de la justice : magistrats, police judiciaire, greffiers, huissiers, médiateurs, dont les fonctions sont expliquées ci-après.

2.2.1.4.1 Les magistrats

À l'exception des jurés qui ne sont pas professionnels, en matière pénale, tous les juges sont des magistrats. On distingue les *magistrats du siège*, qui rendent la justice, des *magistrats debout* ou *du parquet*.

a. Magistrats du siège

Outre le juge d'instruction, qui prend la décision d'envoyer ou non un mis en examen devant une formation de jugement, le traducteur-interprète peut avoir à assister un devant le juge de la liberté et de la détention et le juge d'application des peines (le JAP), qui interviennent également dans la procédure pénale[31].

a.1. Juge des Libertés et de la Détention (JLD)

La privation de liberté ne devant pas excéder une certaine durée, le mis en examen doit passer régulièrement devant un Juge des Libertés et de la Détention tout le temps de l'instruction. Les délais de présentation varient selon la nature de la procédure. Le juge décide de la pertinence ou non de garder le prévenu en détention provisoire. Ce dernier peut alors formuler une demande de remise en liberté dans la limite maximale d'une par jour, à laquelle le JLD a l'obligation de répondre en motivant sa décision (art. 137-2 du CPP). Si le JLD décide de prolonger la détention, il délivre une ordonnance de maintien en détention provisoire (art. 146). Il en va de même pour les mesures de placement sous surveillance électronique ou sous contrôle judiciaire.

a.2. Juge d'application des peines (JAP)

Le JAP suit l'évolution du condamné à l'intérieur et à l'extérieur des prisons. Il se prononce sur la libération conditionnelle et l'aménagement de peine. Ce magistrat n'intervient donc qu'après la décision au procès.

2.2.1.4.2 Magistrats du parquet

Historiquement, les magistrats qui plaidaient debout se tenaient dans un petit parc délimité par des barrières, le parquet. Aujourd'hui, le Parquet représente les intérêts de la société devant les juridictions, au-delà de ceux de la victime. On parle aussi de *Ministère*

[31] Il existe de nombreux autres magistrats du siège : juge des enfants, juge délégué aux victimes, et en matière civile : juge aux affaires familiales (JAF), juge de l'exécution (JEX), juge de la mise en état…

public. Par exemple si un individu retire sa plainte et que le parquet estime que les intérêts de la société ont été lésés (dans le cas d'une agression particulièrement violente, par exemple), il peut poursuivre l'individu contre qui la plainte était initialement dirigée. Le *procureur de la République*, au sein de chaque TGI, est en charge du Parquet et poursuit le ou les auteurs d'une infraction. Il supervise l'action de la Police judiciaire pendant l'instruction et peut s'opposer aux conclusions d'un juge d'instruction dans son *réquisitoire*. Ce terme désigne le moyen d'expression du Ministère public. À l'instruction ou lors d'un procès, le Procureur *prend ses réquisitions*. Lors d'un procès, le réquisitoire désigne le discours rhétorique du Procureur destiné à protéger les intérêts de la société, au terme duquel ce dernier proposera la peine qu'il souhaite voir appliquée.

Des actions ou des agents similaires sont désignés par des hétéronymes différents en fonction des procédures ou des degrés de juridiction, ou par différenciation d'avec la branche civile du judiciaire. Dans un procès aux assises, aboutissement de la procédure criminelle, le procureur de la République devient ainsi l'avocat général.

2.2.1.4.3 La police judiciaire

Il s'agit ici d'une mission et non d'un service particulier. Les officiers de police judiciaire (OPJ) et agents de police judiciaire (APJ) peuvent être policiers, gendarmes ou douaniers. La police judiciaire, comme son nom l'indique, est liée au tribunal, et plus précisément au procureur de la République qui en supervise les activités. Elle est sollicitée par le juge d'instruction pour enquêter. En dehors des tribunaux, ce sont les OPJ qui réquisitionnent les traducteurs-interprètes pour les assister dans leurs activités d'investigation.

2.2.1.4.4 Greffe et greffier

Le greffe est le service qui assiste les magistrats dans leur mission. On distingue les greffes correctionnels de ceux des cours d'assises ou d'appel.

Il existe plusieurs sortes de greffiers[32], tous assermentés. Dans le cabinet d'instruction, le greffier assiste le juge d'instruction en prenant les notes des procès-verbaux d'audition ou d'interrogatoire, ou en mettant en forme la procédure. Le greffier (ou la greffière) est l'interface initiale entre le juge et le traducteur-interprète. C'est lui qui le contacte et peut lui donner des précisions sur la mission que le juge souhaite lui confier.

[32]Profession très majoritairement féminine.

Lors d'une audience de jugement, le greffier prend des notes et vérifie la disponibilité au *rôle*[33] pour les affaires à juger. Dans tous les cas, c'est lui qui délivre au traducteur-interprète les pièces nécessaires à sa rémunération.

2.2.1.4.5 Les auxiliaires de justice

Tous les acteurs judiciaires suivants contribuent de manière pérenne ou temporaire au bon fonctionnement de la justice pénale, en assistant soit les magistrats, soit les justiciables.

a. L'avocat

Appelé également *conseil*, il représente son client et en défend les intérêts lors de la procédure. Il peut être choisi ou désigné dans certaines circonstances. Selon les conditions de ressources, l'avocat peut être commis d'office et ses frais pris en charge par l'administration du tribunal. Les avocats sont inscrits à un *barreau* dans chaque département. Un barreau est présidé par un *bâtonnier*. L'interprète judiciaire peut être amené à assister un avocat lors de ses entretiens avec son client, à son cabinet, au tribunal, en garde à vue ou en détention.

b. Les huissiers

On distingue les *huissiers de justice* et les *huissiers audienciers*. Les premiers achètent leur charge et assument la délivrance et la notification des actes (citation en procédure pénale ou assignation en justice au civil). On dit qu'ils *régularisent un exploit*. Présent au tribunal correctionnel, l'huissier audiencier tient le rôle de l'audience. Il appelle chaque affaire et vérifie si les intéressés sont présents. L'huissier audiencier n'a pas fait d'études de droit.

c. Les experts

Les affaires pénales peuvent nécessiter l'avis d'un expert pour des questions techniques, afin d'apporter un éclairage sur le dossier et faire émerger la vérité. Il existe toutes sortes d'experts, dans tous les corps de métiers. Chaque cour d'appel en possède une liste. L'expert est choisi sur dossier par le service des experts de sa cour d'appel. Pour

[33]Liste des affaires prévues à l'audience.

les interprètes et les traducteurs, aucun test de compétence en langue n'est exigé[34]. Un expert est nommé une première fois pour une période probatoire de trois ans, puis pour cinq ans. Au terme de ces périodes, il doit justifier d'un nombre d'heures de formation à la procédure judiciaire.

L'expert est commis par les autorités judiciaires qui lui délivrent une *commission d'expertise* afin de recueillir son avis sur une question technique. Seuls les traducteurs-interprètes font exception pour les raisons vues en 2.1.3, puisqu'ils sont le plus souvent *requis*[35]. N'importe quel spécialiste d'un domaine peut être désigné expert, même s'il n'est inscrit sur aucune liste. Il lui suffit de prêter serment (par écrit ou oralement).

2.2.2 *En Inde et au Pakistan*

Dans le chapitre 1 (voir 1.1.1.1), nous avons vu que le système judiciaire du sous-continent était un héritage de la période coloniale britannique. Il a été rappelé toutefois que l'héritage administratif antérieur des Moghols, ainsi que la recension des textes légaux hindous et musulmans, avaient influencé l'écriture des codes. Par la suite, ces systèmes ont su s'adapter aux sociétés locales et des notions issues du droit britannique ont pu évoluer par rapport aux notions originales, raison pour laquelle il ne suffit pas de traduire les termes utilisés en anglais indien vers le français pour obtenir une traduction claire des concepts. Ce constat, lié à l'imbrication étroite entre le droit et la culture locale, est le même dans d'autres langues : droits belge et français sont proches sans être identique, le droit allemand d'Allemagne ne sera pas exactement similaire au droit de la communauté germanophone belge, etc.

La procédure en Inde et au Pakistan reste relativement similaire dans le cas des affaires pénales, et est indiquée dans le CrPC[36] (de 1973 pour l'Inde et de 1898 pour le Pakistan). Si la procédure d'ouverture de l'information judiciaire (2.2.2.2) est commune aux deux pays, la procédure qui suit est exposée en (2.2.2.2) pour l'Inde, et les variantes de la procédure pakistaniaise se trouvent en (2.2.2.3).

---

[34]Dans de nombreux autres pays, des tests linguistiques permettent d'apprécier leur compétence et de leur donner en fonction un degré d'habilitation à traduire ou à interpréter.

[35]De fait, les expert traducteur-interprète (ETI) ne jouissent pas exactement du même statut que les autres experts, ce qui se constate physiquement dans les autorisations d'accès et de déplacement à l'intérieur du tribunal.

[36]Criminal Procedure Code, cf. annexe A.

2.2.2.1 Ouverture d'une information judiciaire

Comme en France, il est possible d'ouvrir une information judiciaire à partir d'un rapport de police ou du signalement d'un plaignant auprès d'un magistrat.

2.2.2.1.1 À partir d'un rapport de police

Le first information report (FIR) est rédigé à partir de la déposition de toute personne ayant des renseignements à donner au sujet d'une infraction commise (art. 154 du CrPC en Inde, art. 164 du CrPC au Pakistan). Ce rapport d'informations préliminaires constitue le point de départ à l'ouverture d'une information judiciaire. Il est rédigé dans la langue du déposant en Inde, en ourdou au Pakistan. Le déposant répond aux questions d'un officier de police, puis les échanges sont reformulés de façon à ce que le contenu apparaisse comme un récit des faits à la première personne. Dans une seconde partie, le document présente l'action de la police.

Un officier enquêteur (*investigation officer* / *anveṣaṇ adhikārī* (HI.) / *tafteśī afsar* (OU.)) mène l'enquête. Il collecte les preuves, auditionne les témoins, etc. Au terme de son enquête, il adresse un rapport (*police report* ou *chalan* /cālān (HI., OU.)) au magistrat. C'est lui qui indique s'il y a lieu de poursuivre (rédaction d'une *charge-sheet*, *ārop patr* (HI.), en ourdou *fard-e-jurm* (OU.)) ou de clore la procédure.

Dans la langue populaire, cette phase d'ouverture d'une procédure judiciaire est appelée *parcā darj karānā* (faire enregistrer une « page »).

2.2.2.1.2 À partir de la plainte d'un individu

Une plainte est adressée à un tribunal. Celui-ci entend le plaignant (*parivādī* (HI.), *muddaʿī* (OU.)) et ses témoins (*sākṣī* (HI.), ourdou : *gavāh* (HI.)). S'il décide d'ouvrir une enquête, ses conclusions l'amènent soit à reconnaître qu'il existe suffisamment de charges contre l'accusé pour poursuivre, soit à clore l'information En ce dernier cas, il doit motiver sa décision. Un mandat (*vāraṇṭ* (HI.OU.) / *parvānah* (OU.)) d'arrêt ou une convocation (*samman* (HI.OU.)) sont délivrés à l'encontre du mis en cause en fonction des faits.

2.2.2.2 Trois procédures principales

Le CrPC indien prévoit trois types de procédures criminelles : avec délivrance d'un mandat d'arrêt (warrant cas), d'une convocation à une audience (summons case) et

des procédures directes (summary trials)[37].

2.2.2.2.1 Warrant case (affaire pénale avec délivrance d'un mandat d'arrêt)

L'expression se trouve généralement en anglais. Cette procédure concerne des crimes punis de mort, de réclusion à perpétuité ou d'une détention supérieure à 2 ans (art. 2(x) du CrPC). Les affaires avec mandat d'arrêt (*vāraṇṭ giraftārī* (HI.OU.)/ *parvānah giraftārī* (OU.)) sont instruites différemment selon qu'elles sont ouvertes à partir d'un rapport de police ou d'une plainte. Dans ce type d'affaire, l'accusé a le droit d'accéder à toutes les pièces versées au dossier le concernant. Il est placé sous mandat d'arrêt mais peut bénéficier d'une libération sous condition dans l'attente du procès (*zamānat* (HI.) / *żamānat* (OU.)) en fonction des infractions et des circonstances.

Lorsque l'affaire est enregistrée à partir d'un rapport de police, si les charges retenues sont suffisantes, alors les faits doivent être qualifiés en référence à des articles du Code pénal (framing of charges). Le mis en cause doit en avoir connaissance, les comprendre, et indiquer s'il reconnaît les faits qui lui sont reprochés ou non (plea of guilty).

S'il plaide coupable, le juge peut choisir de le condamner directement. Le tribunal demande au ministère public (*prosecution*, *abhiyojan* (HI.), *isteġāṣah* (OU.)) de produire des preuves de sa culpabilité et des témoignages. Ce seront les témoins de l'accusation (*prosecution witnesses*, *abhiyojan kā sākṣī* (HI.), *gavāh isteġāṣah* (OU.)), qui peuvent subir un contre-interrogatoire (cross-examination) de la part de l'avocat de la défense. En Inde, les déclarations des témoins sont filmées depuis peu afin d'éviter que le témoin ne se rétracte lors de l'audience, cas extrêmement courant (Berti, 2012). Dans la procédure indienne ou pakistanaise, c'est à l'accusation que revient la charge de la preuve.

L'accusé peut s'expliquer sur les faits ou garder le silence, il ne dépose par sous serment. La défense peut produire des preuves orales ou écrites, mais ce n'est pas obligatoire puisqu'elle n'a pas la charge de la preuve.

Le jugement est une décision du tribunal. Dans le seul cas où l'accusé est condamné, toutes les parties débattent de la peine à appliquer. À l'issue du procès, les parties peuvent faire appel de la décision.

[37]Source : Garima Singh, `https://www.lawfarm.in/blogs/process-of-criminal-trial-and-proceedings-in-india`.

2.2.2.2.2 Summons case (affaire pénale avec convocation)

Cette procédure s'applique pour tous les faits qui ne relèvent pas d'un warrant case et pour lesquels la peine encourue n'excède pas deux ans (art. 2(w) du CrPC indien). La méthode est moins stricte, les charges sont présentées à l'oral et le mis en cause n'est pas obligé d'être présent. Il n'a pas accès à toutes les pièces du dossier (la procédure est non contradictoire) et ne peut que contre-interroger les témoins.

Le mis en cause ne prend connaissance des faits et de leur qualification que devant le magistrat, à l'audience. Comme précédemment, l'accusé choisit de plaider coupable ou non. S'il plaide coupable, le juge peut le condamner à une peine. Dans le cas contraire, le procès commence. L'accusation et la défense y présentent leurs preuves et témoignages. Le mise en cause peut déposer à la demande du magistrat.

Lorsque la culpabilité est établie, la peine à appliquer n'est pas débattue. Elle est décidée par le juge seul. Les parties peuvent faire appel.

2.2.2.2.3 Summary trials (procès rapides)

Les affaires de cette catégorie concernent un nombre limité d'infractions mineures (passibles au plus de trois mois de prison) qui sont alors traitées en procédure accélérée (fast track proceeding)(art. 260-265 du CrPC) et jugées en une seule audience. Elles sont du ressort des Chief judicial magistrate, Metropolitan Magistrate ou du Judicial Magistrate de première classe. La procédure d'instruction est la même que pour les affaires relevant de la catégorie précédente. Si la décision du juge est moins détaillée que pour un procès des catégories précédentes, elle doit toutefois être motivée.

2.2.2.3 Précisions sur la procédure pakistanaise

Le CrPC pakistanais reconnaît les deux types de procédure avec mandat (article 90) ou convocation à l'audience. Il existe en outre une procédure visant les justiciables en fuite (art. 87) et la possibilité pour le tribunal de saisir des biens du mis en cause (art. 88). Ce même code prévoit, après l'enregistrement d'un FIR, qu'une enquête de police soit immédiatement diligentée et achevée sous 14 jours (art. 173 du CrPC), durée maximale de détention en garde à vue. Les charges rédigées contre le prévenu doivent l'être en anglais.

Dans la réalité, les policiers ne peuvent conclure une enquête aussi rapidement

et remettre le rapport (chalan) au magistrat[38]. Le magistrat délivre généralement un délai supplémentaire pour remettre le rapport et commence à instruire l'affaire malgré tout.

Si l'affaire est trop importante pour être traitée directement par le tribunal du magistrat (*qābil-e-samā'at*, « audiençable »), elle est transférée à la Session court.

D'autre part, lorsqu'un individu est interpellé par la police, il doit être présenté au magistrat local sous 24 heures. En réalité le délai de déferrement est plus long. À ce stade, le détenu a le droit d'accéder gratuitement au dossier d'investigation.

2.2.2.4 Déroulement d'une audience

Des précisions sur les échanges linguistiques lors d'audiences pénales en Inde nous sont apportées par (Berti, 2012, 2015; Berti et Tarabout, 2015), ethnologue qui a étudié les procédures pénales dans une région hindiphone du nord de l'Inde. Elles m'ont été corroborées pour le Pakistan par Me Yasser Qureshi, avocat franco-pakistanais.

Pendant l'audience, les paroles des témoins qui s'expriment en hindi ou dans une autre langue locale sont directement enregistrées par écrit en anglais, et reformulées avec des tournures plus juridiques. Ici, c'est le juge qui traduit en anglais et dicte au greffier chargé de l'enregistrement. Le procureur ou l'avocat de la défense peuvent aussi intervenir dans la traduction (note 6 Berti, 2012). La personne dont on a traduit les propos n'est pas avertie des échanges qui interviennent en anglais entre professionnels du droit et qu'elle ne comprend pas.

Contrairement à la procédure française, il n'y a pas ici de recherche du contexte social ou psychologique qui permettrait d'expliquer les faits. Une autre différence est la tendance des témoins de l'accusation à se dédire. On parle alors de « hostile witness ». Cette situation est problématique dans une procédure accusatoire où la charge de la preuve revient entièrement à l'accusation.

Après l'audition des témoins, le juge, le procureur et l'avocat de la défense débattent. L'avocat de la défense expose ensuite les décisions de jurisprudence sur lesquelles il bâtit son argumentation.

Le tableau synoptique 2.1 récapitule les phases de la procédure ainsi que ses acteurs. La terminologie de la procédure française se caractérise par une abondance de termes hétéronymes pour des actes ou des acteurs similaires appartenant à des

[38] Communication vidéo de Shauket Rafique Bajwa, avocat, vue et traduite de l'ourdou sur `https://www.youtube.com/watch?v=MoyngoudBhU` (traduction personnelle).

procédures différentes (correctionnelle ou criminelle, de première instance ou d'appel). Cette hétéronymie impose un apprentissage supplémentaire au traducteur-interprète qui doit savoir les discriminer, en retenir le sens, et le restituer par le bon terme (ou la bonne tournure en hindi-ourdou). Dans la partie suivante, nous explorerons les genres textuels du domaine judiciaire afin d'en repérer les caractéristiques terminologiques et phraséologiques.

	français		hindi (Inde)	ourdou (Pakistan)
déclenchement	plainte			
information préliminaire	enquête préliminaire, réquisitoire		enquête préliminaire, enquête de police, rapport de fin d'enquête (*chalan*)	
instruction	enquête dirigée par le juge d'instruction		enquête dirigée par le magistrat judiciaire	
selon gravité de l'infraction...	délit (≤ 10 ans de prison)	crime (≥ 10 ans de prison)	infraction (≤ 2 ans)	infraction (≥ 2 ans)
... type de procédure	procédure correctionnelle	procédure criminelle	convocation	mandat d'arrêt
lieu du procès en 1ère instance	tribunal correctionnel	cour d'assises	tribunal du magistrat judiciaire	tribunal de district ou cour d'assises
lieu du procès en appel	C.A. correctionnelle	C.A. d'assises	tribunal de district ou cour d'assises	Haute Cour
dernier recours national	cour de cassation (sur la forme)		Cour suprême	
décision	décision / arrêt (appel)	arrêt, verdict (jury)	décision	
parties au procès	plaignant, partie civile (accusation) ministère public mis en cause (défense) témoins, experts		plaignant (accusation) procureur (accusation) défense témoins, experts (de la défense, de l'accusation)	

FIG. 2.1: Tableau synoptique des grandes lignes de la procédure pénale

2.3 Genres textuels du judiciaire

Aux actes des procédures pénales détaillées en 2.2 sont associés des textes à traduire à l'écrit ou à l'oral. Ces textes s'inscrivent dans des genres textuels et se structurent en différentes dominantes séquentielles. La connaissance des caractéristiques du genre textuel permet au traducteur de comprendre le raisonnement et le but recherché par le rédacteur. Après une définition du genre textuel (2.3.1), nous verrons les caractéristiques associés aux genres des textes de la procédure (2.3).

2.3.1 *Le genre textuel*

Pour les théoriciens de l'analyse textuelle comme Rastier (1989) ; Adam (1997) ; Bronckart (2004) ou Maingueneau (2004), texte, discours et genre sont inextricablement liés. Dans son sens linguistique général, le discours est défini comme une « actualisation de la langue par un sujet parlant » ou le résultat de ce processus[39]. Il ne se limite cependant pas à la forme orale, il faut l'entendre comme production langagière. Un discours est un acte de langue produit dans le cadre d'une situation d'interaction sociale. Il s'articule sous forme de texte dans une « sphère d'usage » (Bakhtine, 1984) caractérisée par un domaine thématique. Les textes sont des objets complexes:

> configurés par toutes sortes de contraintes qui déterminent l'énonciation représentée, le mode de lecture qu'ils appellent, la représentation de leurs destinataires, les mondes qu'ils semblent ouvrir (…). Rastier (1989)

Un texte est donc une actualisation du discours dans un cadre énonciatif donné. Pour (Adam, 1997), le texte est réglé par divers plans d'organisation : sémantique, énonciative et illocutoire (suite d'actes de langage).

L'interaction sociale qui définit le cadre d'énonciation du texte est son genre. Pour (Rastier, 1989), tout texte se place dans un genre, celui-ci étant caractérisé par un plan d'ensemble du texte, une structure séquentielle au niveau micro-linguistique et une dominante séquentielle (narrative, explicative, argumentative, etc.).

Dans le cas du champ judiciaire, les documents à traduire sont de nature très hétéroclite (Monjean-Decaudin, 2012). En réalité, tout document rattaché à une procédure judiciaire est considéré comme judiciaire. Or ces documents ne sont pas tous de nature juridique. Ce n'est pas le cas, par exemple, des retranscriptions d'écoutes téléphoniques dont les conversations se produisent dans une langue familière, voire codée. Ces échanges

[39] http://www.cnrtl.fr/lexicographie/discours.

peuvent concerner tout autre sujet que le juridique (par exemple la remise d'une somme d'argent ou la prise d'un rendez-vous). Ce n'est pas non plus le cas d'un agenda ou de documents comptables saisis lors d'une perquisition. Certains documents sont de nature juridique ou judiciaire tandis que d'autres intègrent ces deux aspects.

Un texte juridique est empreint de la langue de spécialité du droit, ou « langage du droit »)[40], et entraîne des effets de droit. Ce langage des énoncés du droit, normes, décisions, conventions, déclarations etc, possède un caractère performatif. Le langage juridique n'est cependant pas intégralement constitué d'éléments spécialisés. Il entretient une relation de force variable entre la langue générale (langue commune, ou naturelle) et le langage normatif (celui des textes de loi) (Wroblewski, 1988).

Les spécialistes qui se sont penchés sur le discours juridique et la traduction du droit ont proposé différentes classifications des textes traduits : sémantique, pragmatique, normative entre autres. Claude Bocquet, pour sa part, classe les textes juridiques en fonction des types de forme logique et de discours selon la typologie suivante (cf. tableau 2.1, repris de (Monjean-Decaudin, 2012)) :

texte juridique	**discours**
textes normatifs	discours du législateur
textes des décisions qui appliquent ces normes	discours du juge
textes qui exposent le contenu des règles de droit	discours de la doctrine

Tab. 2.1: Typologie des types de textes juridiques par C. Bocquet

Dans notre perspective de recherche terminologique, nous ne détaillerons pas ici les caractéristiques de tous les genres de textes susceptibles d'entrer dans le cadre d'une procédure. Nous nous intéresserons aux textes produits par des juridictions (tribunaux, officiers de police judiciaire) à destination des justiciables, car ce sont eux qui contiennent la terminologie et les phraséologies judiciaires pouvant poser difficulté au traducteur-interprète. Ils appartiennent, selon la classification de Bocquet, à la deuxième catégorie (décisions appliquant des normes) et contiennent un fort degré de technicité juridique. Connaître les spécificités de ces textes permet au traducteur de comprendre le raisonnement du rédacteur et le but qu'il recherche (Damette, 2013).

Selon (Wroblewski, 1988), le langage juridique jurisprudentiel – actualisé dans les textes juridictionnels – est très hétérogène. Il repose à la fois sur des éléments de la

[40]Cornu (2010).

langue de spécialité et sur des énonciations en langue générale (non spécialisée). Ces textes juridictionnels sont émis en contexte judiciaire, et rédigés par des « initiés », professionnels de la justice (Monjean-Decaudin, 2012). Leur plan général est structuré en séquences aux styles hétérogènes. On distingue deux grandes sous-catégories : les textes entraînant un effet juridique (décisions, convocations, demande d'actes, inscription à un fichier, (2.3.2.1)) et les procès-verbaux (2.3.2.2), qui transcrivent des échanges oraux. Ces sous-catégories se retrouvent dans les trois pays.

2.3.2 *En France*

La présentation suivante expose la structuration caractéristique des textes juridictionnels en fonction de leur finalité (performative ou informative). Cette structuration est identifiable par la phraséologie du texte et ses éléments énonciatifs.

2.3.2.1 Textes entraînant un effet juridique

2.3.2.1.1 *Ordonnances de clôture d'instruction et jugements*

Les magistrats du parquet rendent seuls des décisions appelées *ordonnances* tandis que les tribunaux correctionnels, les cours d'assises et leurs cours d'appel rendent respectivement des *décisions*, *verdicts* ou *arrêts*.

Le juge d'instruction rédige à l'issue de son instruction une ordonnance de renvoi devant le tribunal correctionnel (ORTC), s'il s'agit d'une procédure correctionnelle, une ordonnance de mise en accusation (OMA), dans le cas d'une affaire criminelle, ou une ordonnance de non-lieu (ONL) s'il estime qu'il n'y a pas lieu à renvoyer l'affaire devant une formation de jugement[41]. Le juge a la possibilité de modifier la qualification ou d'en abandonner une partie. La rédaction de ces documents est structurée autour de l'argumentation, appelée « motifs », qui conduit le magistrat à sa prise de décision[42]. Ce syllogisme formel, que l'on retrouve dans tous les textes d'application de la norme, se construit en trois temps (Damette, 2013).

L'en-tête du document rappelle les références de l'affaire et les parties en présence. Elle s'organise selon le modèle suivant :

[41] Un non-lieu est souvent prononcé pour insuffisance d'éléments, mais ce n'est pas toujours le cas. Lorsque les éléments sont nombreux et établis, le mis en cause peut ne pas être renvoyé devant une juridiction en raison de son irresponsabilité pénale, de la prescription des faits, de son décès ou autre.

[42] Cf. Frison-Roche (1995), « (…) l'argumentation est la technique tant de l'établissement des faits que de l'application du droit. »

– références de l'affaire (tribunal, cour d'appel, numéros de dossier),
– mention du magistrat (fonction et lieu) introduit par « Nous », à laquelle fera écho la formule de décision dans la conclusion,
– mention de la personne mise en cause introduite par la formule « vu l'information suivie contre »,
– récapitulatif des échanges entre les membres du parquet et de l'instruction (ou entre le requérant et le tribunal dans le cas d'un appel), chaque échange de pièce est introduit par « vu »,
– qualification introduite par « mis en examen des chefs de », puis par « faits prévus et réprimés » pour les références aux textes de loi. Les limites dans le temps et l'espace sont exprimées par l'expression figée « en tout cas sur le territoire national et depuis temps non prescrit ».

Les unités phraséologiques de cette introduction suivent l'ordre d'une phrase qui ne se conclut qu'à l'énoncé de la décision. Elles sont nombreuses et posent la solennité du cadre.

La première partie, les « attendus », est introduite par l'expression figée « attendu que », purement stylistique[43]. Elle commence par un résumé non polémique des faits[44]. C'est l'exposé des faits, que l'on retrouve dans sa version orale en début d'audience, en correctionnelle comme aux assises. On constate des éléments de l'enquête, sans émettre de jugement. On y trouve les résultats des rapports d'expertise, puis les déclarations des différents protagonistes rapportées pendant l'instruction. Cette partie se termine par une rubrique « Renseignements et personnalité » qui fournit des éléments de contexte personnel et judiciaire sur le mis en examen. Y sont indiquées les conclusions des expertises psychiatrique et psychologique, ainsi que d'éventuelles mentions au casier judiciaire. Comme nous l'avons vu dans le déroulement de la procédure pénale (2.2), la procédure en France s'attache à juger les faits commis par un individu dans un contexte.

La partie « discussion » est celle dans laquelle se manifeste l'argumentation. On y présente sous forme de raisonnement syllogique les arguments à partir des éléments factuels évoqués dans l'exposé des faits. La prémisse majeure du raisonnement syllogique

[43]Communication de Mme Anton-Bensoussan, magistrate, lors de la formation du 30 janvier 2019 aux experts traducteurs et interprètes près la cour d'appel de Paris.

[44]Damette (2013), *op. cit.*

est représentée par les textes normatifs. La prémisse mineure, par le cas d'espèce. Enfin, la conclusion constitue la décision[45].

Les arguments, à charge ou à décharge de la personne mise en examen, servent de fondement à la qualification des faits. S'il y a suffisamment de raisons (« charge suffisante »), ils motivent le renvoi du mis en examen devant un tribunal. Sinon, la décision sera un non-lieu. Cette partie est également appelée « motivation » de la décision. Elle se conclut par une tournure telle que « au vu de l'ensemble de ces éléments ».

Enfin, le « dispositif », qui comprend la décision, constitue la dernière étape. Cette partie est toujours introduite par la formule « par ces motifs », puis par la reprise de la première personne du pluriel utilisée dans l'en-tête dans la phrase : « Ordonnons le renvoi / la mise en accusation de [mis en cause] », suivi d'un rappel du chef de prévention et des articles de loi visés, ainsi que des formules de procédure.

a. Cadre énonciatif et structure du document

L'énonciateur, en l'occurrence le magistrat, se manifeste au tout début et à la fin du document par la 1e p. du pluriel. Il est présent à travers le pronom personnel « nous » (« Nous, X, juge d'instruction au tribunal de grande instance de Y, ») et des verbes au présent de l'indicatif (« ordonnons », « informons »). Dans tout le reste du document, il reste en retrait, contrairement à son homologue indien ou pakistanais.

La structure du texte commence par un en-tête rappelant les références de l'affaire et le cadre légal. S'en suit l'exposé des faits, puis la partie argumentative appelée « discussion ». On y présente sous forme de raisonnement syllogique en trois temps (Damette, 2013) les arguments à partir des éléments factuels évoqués dans l'exposé des faits. La prémisse majeure du raisonnement syllogique est représentée par les textes normatifs. La prémisse mineure, par le cas d'espèce. Enfin, la conclusion constitue la décision[46]. Ce type de raisonnement se retrouve dans tous les textes d'application du droit.

Les arguments, à charge ou à décharge de la personne mise en examen, servent de fondement à la qualification des faits. S'il y a suffisamment de raisons (« charge suffisante »), ils motivent le renvoi du mis en examen devant un tribunal. Sinon, la

[45]Communication informelle de M. Pierre Beaudoin.

[46]Communication informelle de M. Pierre Beaudoin.

décision sera un non-lieu. Cette partie est également appelée « motivation » de la décision. La conclusion énonce la décision et le rappel des éléments du cadre référentiel.

Chacune de ces parties est introduite, structurée ou conclue par des unités phraséologiques.

b. Phraséologie

La phraséologie de cette catégorie de document, à caractère performatif, se doit de poser un cadre solennel, celui de la décision de justice. On y trouve de nombreuses séquences polylexicales parfois longues et à fort degré de figement. Certaines de ces unités phraséologiques sont typiques du seul genre judiciaire et marquent la structuration de chaque partie du document (introduction, exposé des faits, argumentation et décision). Ce sont dans l'introduction et dans la conclusion que les expressions figées et les séquences les plus stéréotypées sont les plus nombreuses. Dans le corps du document, formé par l'argumentation, le degré de technicité du lexique est plus faible.

Dans l'en-tête, chaque étape de la présentation est introduite par une expression figée suivant sa fonction :

- la présentation de la ou des personnes mise(s) en cause est introduite par le syntagme figé « vu l'information suivie contre »
- chaque échange de pièce entre les membres du parquet et de l'instruction (ou entre le requérant et le tribunal dans le cas d'un appel) est introduit par « vu »,
- les chefs de prévention ainsi que les textes de loi qui les sanctionnent sont introduits par les expressions figées « mis en examen des chefs de » et se concluent par « faits prévus et réprimés » suivi des références légales. La mention est indispensable puisqu'il n'est pas possible, selon le code pénal, d'être jugé pour des infractions qui ne sont pas répertoriées dans la loi.
- les limites spatio-temporelles de l'infraction sont énoncées par l'expression figée « en tout cas sur le territoire national et depuis temps non prescrit », qui indique que le droit français peut s'appliquer, puisque les faits ont été commis sur le territoire où il s'applique et dans une limite de temps valable.

Ces expressions sont autant d'étiquettes qui posent le référentiel de l'énoncé.

La conclusion, appelée « dispositif », est également richement structurée en expressions figées qui maintiennent le cadre solennel de l'énoncé. Y sont rappelés les chefs de prévention, motif du renvoi, et les références aux textes légaux, éléments déjà présentés dans l'en-tête.

Le corps du texte est composé de deux parties. La première, à dominante factuelle, est l'exposé des faits. La seconde, appelée « discussion », est à dominante argumentative. La première partie, celle des attendus, est introduite par l'expression figée « attendu que », purement stylistique[47]. Elle commence par un résumé non polémique des faits[48].

Certains éléments, comme les attendus, ne sont pas réservés aux textes judiciaires. On les retrouve dans des textes de loi. D'autres, comme les expressions figées destinées à poser le cadre référentiel, sont caractéristiques du genre.

c. Lexique

Les mots qui entrent dans la composition des unités phraséologiques de ces textes sont rarement limités au domaine de spécialité. Ils prennent leur sens particulier du fait de leur association déterminée. Par exemple « territoire » ou « national » appartiennent à la langue générale, tandis que leur association dans le syntagme « en tout cas sur le territoire national » est limitée à l'énoncé judiciaire. Dans le corps du texte, l'exposé des faits sera composé d'énonciations en langue générale, émaillées de lexèmes techniques pour structurer les phases chronologiques de la relation ou rapporter des actions judiciaires (« Une information a été ouverte », « bornage », « exploitation de la ligne téléphonique », « commission rogatoire »), ou pour les rapports d'expertise (termes médicaux, fiscaux, psychologiques…). Dans l'exposé des faits, que l'on retrouve dans sa version orale en début d'audience, en correctionnelle comme aux assises, on constate des éléments de l'enquête, sans émettre de jugement. On y trouve les résultats des rapports d'expertise, puis les déclarations des différents protagonistes rapportées pendant l'instruction. Cette partie se termine par une rubrique « Renseignements et personnalité » qui fournit des éléments de contexte personnel et judiciaire sur le mis en examen. Y sont indiquées les conclusions des expertises psychiatrique et psychologique, ainsi que d'éventuelles mentions au casier judiciaire. Comme nous l'avons vu dans le déroulement de la procédure pénale (2.2), la procédure en France s'attache à juger les faits commis par un individu dans un contexte.

Les exemples qui suivent illustrent ces caractéristiques :

[47]Communication de Mme Anton-Bensoussan, magistrate, lors de la formation du 30 janvier 2019 aux experts traducteurs et interprètes près la cour d'appel de Paris.

[48](Damette, 2013), *op. cit.*

Nous, Mme A, Vice-président chargé de l'instruction au tribunal de grande instance de [lieu],
Vu l'information concernant :
-M. X Détenu à [lieu] : [date] Mandat de dépôt : [date], 2e prolongation (…)
Personnes mises en examen
du(des) chef(s) de :
Pour avoir à [lieu], en Île de France, sur le territoire national, entre le [date] et le [date], en tout cas depuis temps non prescrit, commis des faits de
aide à l'entrée, à la circulation ou au séjour irrégulier d'un étranger en France, faits prévus et réprimés par les articles l 622-1 alinéas I et 2, l 622-3 du Code des étrangers ;(…)
Attendu que l'information a permis d'établir les faits suivants :

L'exemple ci-dessous, extrait d'une ordonnance de mise en accusation, montre comment se construit le raisonnement déductif qui conduit à la qualification d'assassinat. Le magistrat doit argumenter en faveur de l'existence de la préméditation. Dans le cas contraire, les faits seraient qualifiés de meurtre.

Dans l'exposé des faits, on trouve la description suivante, sans effet modal :

« Les analyses effectuées dans les prélèvements effectués sur et sous la literie, et sur et sous la tête de la victime, mettaient en évidence la présence d'essence pour automobiles. Le laboratoire central de la préfecture de police concluait à une mise à feu perpétrée sur un lit dont le matelas s'était entièrement consumé (…). »

Dans la partie discussion, le juriste revient sur ce point pour souligner la préméditation :

« Le crime apparaît personnalisé, notamment dans son mode opératoire, ainsi [la victime] a été brûlée par le déversement d'essence automobile sur son visage et le haut de son corps, dimension mettant en évidence une symbolique forte dans la volonté de son auteur de personnifier son crime. »

Toujours dans le même texte, la décision est ainsi rédigée :

Au final, au vu de l'ensemble des ces éléments, ceux à charge l'emportant sur ceux à décharge, X devra être mis en accusation du crime d'assassinat.
Par ces motifs
Ordonnons la mise en accusation de M. X devant la cour d'assises de [lieu] pour avoir à [lieu] le [date], en tout cas sur le territoire national et depuis temps non prescrit, avec préméditation, volontairement donné la mort à Y.
Crime prévu et réprimé par les articles 221-3, 221-8, 221-9, 221-9-1 et 221-1 1 du code pénal (natinf 5014)

Le caractère performatif du document est ici actualisé au niveau lexical par l'auxiliaire modal « devoir », le verbe « ordonner » et l'usage du présent de l'indicatif (1re p. de pluriel) exprimant l'injonction et la présence d'un énonciateur solennel (le nous

de majesté). La dernière phrase, qui reprend la partie introductive du document, appartient au registre normatif, avec les références au texte légal. Cette partie finale est structurée par des marqueurs logiques (« au final, au vu de l'ensemble », « par ces motifs »).

Dans le cas où le magistrat instructeur décide ne pas renvoyer l'affaire devant un tribunal, l'ordonnance de non-lieu se conclut ainsi :

> Et attendu qu'il n'existe dès lors pas de charges suffisantes contre : M. X, d'avoir commis les infractions susvisées ;
> Déclarons n'y avoir lieu à suivre en l'état et ordonnons le dépôt du dossier au greffe pour y être repris s'il survenait des charges nouvelles.

Dans cet exemple, le caractère performatif est exprimé par le verbe « déclarer », toujours à la 1re personne du pluriel. Le marqueur logique qui le précède, « et attendu que », structure la fin du raisonnement qui abouti à la décision conclusive.

Le jugement (ou arrêt pour une cour d'appel ou de cassation) suit la même trame argumentative que les documents précédents (présentation des faits, discussion et décision). Il diffère de l'ordonnance de clôture d'instruction dans son introduction et sa conclusion.

Dans l'introduction, la date figure en toutes lettres ainsi que la composition de la formation de jugement. L'arrêt commence par une formule solennelle qui marque le caractère performatif, « A statué ainsi qu'il suit », suivie de l'exposé des faits.

2.3.2.1.2 Ordonnances du juge des libertés et de la détention (JLD)

Le JLD délivre des ordonnance relatives à la privation de liberté. L'ordonnance de maintien en détention (OMD), de mise en liberté (OML, assortie ou non de mesures de restriction), de maintien du contrôle judiciaire ou de rejet de la demande de mise en liberté (ORDML) en sont des exemples. Le texte ne présente pas de partie contenant la relation de l'« exposé des faits » comme dans les ordonnances qui concluent une instruction (l'ORTC, l'ONL, l'OMA) ou le jugement. Le magistrat énonciateur rappelle les articles de loi qui motivent la détention dans le registre normatif. Il construit ensuite ses arguments en faveur du maintien du mis en cause. Ces arguments sont à chaque fois introduits pas l'expression figée « attendu que ». Dans l'exemple ci-dessous, extrait d'une OMD, ses arguments sont la fuite et le renouvellement potentiels de l'infraction :

Attendu que les obligations d'une mesure de contrôle judiciaire ou d'assignation à résidence avec surveillance électronique auxquelles la personne peut être astreinte se révèlent insuffisantes au regard des fonctions définies à l'article 137 du code de procédure pénale ;
Attendu que la détention provisoire de X est l'unique moyen, au regard des éléments précis et circonstanciés résultant de la procédure, de parvenir à l'un ou plusieurs des objectifs suivants :

- de garantir le maintien à la disposition de la justice ; en ce que le mis en examen d'origine indo-pakistanaise, en situation irrégulière sur le territoire français et interpellé en possession de faux documents administratifs ne présente pas de garantie suffisante de représentation.
- de prévenir le renouvellement de l'infraction ; en ce le mis en examen vient d'être condamné le [date] par le tribunal correctionnel de [lieu] pour des faits identiques.

PAR CES MOTIFS,
ORDONNONS le maintien en détention provisoire de X jusqu'à sa comparution devant le tribunal.

À chaque fois, un argument est étayé par un passage du texte légal (moyens de parvenir à la réalisation de certains buts).

2.3.2.1.3 La Commission rogatoire internationale (commission rogatoire internationale (CRI))

La CRI n'est pas une décision de justice, c'est une procédure par laquelle la justice française demande l'assistance d'un pays étranger dans les actes de l'instruction. Ce genre de texte ne comporte pas d'argumentation juridique. D'une portée internationale et destiné à de hautes instances, son style se veut diplomatique, marqué par des formules de politesse.

Le texte commence par la mention de l'État français (« République française ») et de sa devise. La présentation du mis en cause et des chefs de prévention est identique à celle d'une ORTC ou d'une ordonnance de mise en accusation (OMA).

Une formule d'adresse polie précède l'exposé des faits :

« Nous avons l'honneur de prier les Autorités judiciaires compétentes de [pays] de bien vouloir procéder aux opérations ci-dessous précisées et de nous renvoyer dans les meilleurs délais la présente commission rogatoire accompagnée des pièces relatives à son exécution. »

L'exposé des faits est suivi d'une partie intitulée « mission » dans laquelle le magistrat indique les actes judiciaires que l'État français souhaite voir menés par le pays destinataire.

Après la formule de remerciement et de salutation,

« Nous prions les hautes autorités judiciaires de [pays] de bien vouloir agréer l'expression de notre haute considération. »

la CRI se conclut par une annexe citant les textes répressifs applicables. Ce genre textuel contient des traits caractéristiques de la correspondance (formules d'adresse, de politesse et de salutation), un exposé neutre des faits, sans élément polémique, et une séquence appartenant au genre normatif (texte de loi) selon le classement de Bocquet (cf. 2.1).

2.3.2.1.4 Les convocations, citations et avis à audience

Ces documents sont délivrés par le tribunal, transmis par huissier ou remis en main propre par un officier de police judiciaire (cas de la COPJ). Convocation, citation et avis, sont des termes hétéronymes qui désignent des documents ayant la même fonction : informer ou sommer le destinataire à se rendre à une audience. Dans les affaires pénales, on dit qu'ils sont notifiés. Le lexique actualise le degré d'injonction plus ou moins fort du document : les prévenus ou parties civiles sont « cités à comparaître », tandis que la victime est « invitée à comparaître ». Le détenu, quant à lui, sera « convoqué ».

Outre l'introduction déjà évoquée pour les autres documents, convocations comme citations reprennent les chefs et les articles des textes répressifs (citation de textes normatifs). Le magistrat énonciateur s'adresse à la première personne (« je » ou « nous ») au justiciable (représenté par la 3^{e} personne). Le document est complété par une annexe informative indiquant les modalités de représentation, les frais et les recours de la procédure.

2.3.2.1.5 L'inscription à un fichier national

Une personne condamnée peut se voir inscrite à un fichier national. C'est le cas des personnes condamnées dans des procédures d'infractions sexuelles ou violentes, qui sont inscrites au FIJAIS. Elles reçoivent une notification de cette inscription, à caractère informatid, et des obligations qui y sont liées. Le texte destiné au condamné est à la troisième personne et comminatoire, avec de nombreuses occurrences de lexies exprimant l'obligation (« obligation », « justifier »,« devoir ») et un rappel des risques encourus en cas de non respect de ces injonctions.

2.3.2.2 Les procès-verbaux d'échanges oraux

Dans les cas où le justiciable fait des déclarations au magistrat instructeur ou à l'officier de police judiciaire (OPJ), celles-ci sont retranscrites dans un procès-verbal de déposition (pour le plaignant), d'audition (pour les victimes, parties civiles ou témoins) ou

d'interrogatoire (pour le mis en cause, en garde à vue ou devant le juge d'instruction)[49]. Le document est relu et signé par le déclarant, le magistrat et l'interprète.

Dans ce genre textuel, le déroulement des séquences est encadré par une trame de questions. Devant les OPJ, la séance commence par un interrogatoire d'identité. Les questions, fixes, ne seront pas retranscrites sur le procès-verbal. Seules figureront les réponses de cette première partie, retranscrites à la première personne du singulier.

La séquence suivante, commune au cabinet d'instruction et au service de police judiciaire, concerne les faits. Plusieurs éléments viennent interférer dans les déclarations des protagonistes. D'une part, il est impossible de retranscrire la totalité des échanges linguistiques : les personnes hésitent, se reprennent, reformulent leurs déclarations. L'officier enquêteur peut également reformuler des propos dans un registre qu'il juge plus adapté au style écrit. Parfois, une question imprévue est posée ou une réponse suscite une remarque.

Chez le juge d'instruction, les échanges sont plus codifiés. Le binôme magistrat-greffier dispose de différentes stratégies de travail. Le juge pose les questions, prend en note la réponse et la propose au greffier (ou à la greffière) qui les écrit s'il est d'accord. Le greffier peut aussi saisir à la volée les réponses. En cas de conflit sur la compréhension d'une réponse, la décision du greffier prévaut[50]. Lorsqu'un échange imprévu ou non consigné dans une question intervient, celui-ci est intégré dans la réponse donnée. L'exemple ci-dessous en est une illustration.

Le magistrat interroge un mis en examen :

juge : - Monsieur X, connaissez-vous monsieur Y ?
X : - Non, je ne le connais pas.

Le juge montre alors une photo de Y, X répond à nouveau « Non, je ne le connais pas ». La présentation de la photo et la réponse de X sont consignées comme si elles étaient rapportées par X :

« Vous me présentez une photo de Y, mais je confirme ne pas le connaître ».

La reformulation des déclarations est une caractéristique de ce genre textuel. Elle permet de ne pas interrompre un échange et de synthétiser plusieurs renseignements

[49]Les confrontations relèvent aussi de ce genre textuel.

[50]Communication informelle de M. Pierre Beaudoin, magistrat à Cayenne, qui précise qu' « avant l'avocat, le greffier est le véritable contre-pouvoir du juge : c'est lui qui atteste, qui certifie ce qui s'est dit / passé, [...] cela permet notamment le contrôle de l'action du juge. »

répondant à un question posée. Les magistrats, comme les officiers de police judiciaire, ont la possibilité de consigner des remarques sur des éléments de contexte non verbaux (une attitude, une pause).[51]. L'exercice de reformulation peut troubler l'interprète car lors de la relecture du procès-verbal, ses propos auront subi une modification.

On distingue dans les textes juridictionnels un genre écrit entraînant des effets juridiques et un genre oral retranscrit, informatif. Le genre écrit se caractérise par la formalité des tournures articulant les différentes parties du texte (introduction, corps et conclusion). Parmi les textes écrits se trouvent des textes présentant une argumentation (jugements, ordonnances, arrêts) présentant une succession de séquences à dominantes variables : on passe d'un exposé factuel, neutre, à un argumentaire polémique. D'autres textes comprennent des séquences relevant de la catégorie des textes normatifs. Ces textes présentent des passages techniques inégalement répartis.

Le genre oral retranscrit des procédures verbales qui serviront de preuves dans la procédure. Il produit un figement à l'écrit d'échanges oraux plus ou moins fluides et ne peut être intégral. Il se caractérise par une reformulation des déclarations des protagonistes. Le risque est de voir le message initial altéré.

2.3.3 *En Inde et au Pakistan*

La même typologie de genres textuels juridictionnels se retrouve dans ces deux pays, divisée entre textes entraînant des effets juridiques et retranscriptions d'échanges oraux à valeur informative pour la procédure. Les séquences de ces textes juridictionnels sont empreints d'un degré plus ou moins fort de technicité juridique. Dans certaines séquences, c'est la langue courante[52] qui domine (exposé des faits). Dans d'autres, on retrouve des caractéristiques propres au discours administratif (verbes et locutions verbales passifs, mise à distance de l'acteur et mise en avant du processus, emploi des déictiques) tandis que certaines caractéristiques enfin ne s'observent que dans le genre judiciaire. Sur le plan lexical, certaines locutions ou termes prennent un sens particulier dans le contexte judiciaire.

[51]Conversation avec une magistrate de l'instruction, au TGI de Paris, courant 2018.

[52]Pour une discussion de l'usage de cette expression, cf. 3.2.4

2.3.3.1 Énonciation, phraséologie et lexique

2.3.3.1.1 Le cadre énonciatif

Sur le plan énonciatif, l'énonciateur-magistrat est davantage présent dans les textes entraînant des effets juridiques indiens et pakistanais. À la différence du texte français, il transparaît tout au long des textes par l'usage de la première personne (généralement au singulier) et l'emploi de verbes désignant son implication dans le processus (j'ai observé, j'ai vu).

Dans les transcriptions des procès verbaux, l'énonciateur alterne, entre l'officier qui rédige le compte-rendu, et le plaignant. Parfois, dans le cours de la déposition faite à la 1re personne du singulier, le plaignant est désigné à la 3e personne, sans que la transition entre le discours direct et le discours indirect ne soit marquée.

L'emploi du passif, très répandu dans le style administratif et dans la langue de la presse, se retrouve ici fréquemment. Ce n'est pas un usage qui appartient à la langue commune, mais il dépasse le cadre du judiciaire.

D'autre part, le caractère performatif des documents (décision du magistrat, souhait du plaignant de voir initié une action judiciaire) est actualisé par l'emploi du mode subjonctif, à la différence du français qui utilisait l'indicatif.

2.3.3.1.2 Phraséologie

De nombreuses unités phraséologiques ne relèvent pas exclusivement de la langue judiciaire, mais de la langue administrative plus généralement, avec des constituants très sanskritisés ou très persianisés, selon la langue source. Dans les procès verbaux (dépôts de plainte, FIR), des unités polylexicales opaques structurent l'ouverture de la déposition et la conclusion qui décrit l'action de la police. Ces retranscriptions sont des reformulations du discours d'un non-initié, le justiciable, qui s'exprime en langue commune. Ses compétences linguistiques en matière juridique ne lui permettent pas de s'exprimer avec le lexique et les unités phraséologiques à haut degré de spécificité – pour ne pas dire totalement opaques – qui figurent sur les documents, qu'il doit pourtant signer en indiquant qu'il en a compris le sens et que les mentions sont correctes.

2.3.3.1.3 Lexique

Les textes juridictionnels indiens et pakistanais se caractérisent au niveau lexical par l'abondance de marqueurs honorifiques qui s'adaptent à la fonction de la personne désignée ou interpellée. Autre particularité, liée à la morphologie de la langue, des

morphèmes de pluriel qui n'apparaissent pas dans la langue commune sont employés pour désambiguïser le nombre. Dans la continuité, il est fait usage de mots grammaticaux (conjonctions de coordination, postpositions) qui n'appartiennent pas à la langue commune, mais à cette langue administrative et, dans le cas du hindi, artificielle dont nous avons parlé dans le chapitre consacré aux politiques linguistiques (cf 1.2.2).

Le fonds lexical persan caractérise le lexique du genre, même lorsqu'il n'est plus majoritaire, comme dans les textes en hindi. En effet, dans cette langue subsistent des termes incontournables d'origine arabo-persane, qui n'ont pas été remplacés par des néologismes sanskritisés. Dans d'autres séquences, termes sanskritisés et leur équivalent persianisés sont en concurrence.

Les exemples qui suivent, d'abord en hindi puis en ourdou, illustrent les observations formulées ci-dessus.

2.3.3.2 Exemples en hindi

2.3.3.2.1 Textes entraînant un effet juridique

Les documents qui suivent sont des décisions (*ādeś*) en hindi. L'une est extraite du site de la Haute Cour de Jaipur. Le requérant a fait appel d'une précédente décision. Sa requête, n'ayant pas présenté suffisamment d'éléments probants pour justifier une procédure, fait l'objet d'un rejet par la cour. Dans l'ensemble, la langue utilisée est hautement sanskritisée, avec quelques unités phraséologiques persianisées. L'introduction présente les parties et les membres de la formation de jugement. Le développement de l'argumentation est exposé en plusieurs points, jusqu'à la décision de rejet au dernier paragraphe. L'organisation du texte, découpé en paragraphes simplement numérotés[53], suit le raisonnement du magistrat.

La seconde décision émane d'un tribunal de magistrat de l'Uttar Pradesh. Dans une affaire de harcèlement pour obtenir une dot supplémentaire[54], le juge accepte la plainte de la plaignante et lui ordonne de produire la liste de ses témoins afin de convoquer (*samman jārī karnā*) la belle-famille mise en cause.

---

[53] À la différence du français qui découpe en exposé des faits, discussion, renseignements et personnalité, PCM…

[54] Lors des mariages, et bien que la pratique en soit interdite par la loi depuis 1961, la mariée doit apporter des biens ou une somme d'argent conséquente à sa belle-famille. Les tractations entre les deux familles, souvent âpres, peuvent se prolonger après le mariage par le harcèlement de l'épouse, et parfois son assassinat.

L'introduction des deux décisions commence par la référence de l'affaire et les parties en présence. *banām*, mot d'origine persane, marque lexicalement la séparation des parties adverses. Elle signifie « envers » ou « contre » quelqu'un, et n'est utilisée que dans ce contexte. Dans la décision de Jaipur, on utilise également le lexème d'origine arabe, *ʿurf*, pour « alias », qui n'a pas d'équivalent d'origine sanskrite dans ce contexte[55].

(1) *manoj ʿurf guḍḍū evaṁ anya banām rājasthān rājya*
Manoj alias Guddu et autres contre Rajasthan État
« Manoj alias Guddu et autres contre l'État du Rajasthan »

Les membres de la formation de jugement sont précédés d'un adjectif honorifique *mānnīya* (« honorable », « digne de respect »), dimension absente des textes français, où seul est présent le titre de civilité (« Madame », « Monsieur ») pour tous les protagonistes, suivi, s'il y a lieu, de la fonction (« juge d'instruction », « vice-président », « avocat », « expert »…) :

(2) *mānnīya nyāyādhipati śrī es es koṭhārī*
honorable président du tribunal M. S. S. Kothari
« honorable président du tribunal Monsieur S. S. Kothari »

Ces termes de désignation honorifiques renvoient à des catégories d'individus. Pour l'avocat, le procureur, l'ensemble de la formation de jugement ou le juge de la cour d'assises, l'adjectif consacré est *vidhyamān* ou *vidvān* (« érudit »).

(3) *vidhyamān / vidvān adhivaktā, vidhyamān seśan nyāyādhīś, vidvān lok abhiyojak*
érudit / érudit avocat, érudit assises juge, érudit procureur
« Maître, Monsieur le juge, Monsieur le procureur de la République. »

Dans les deux décisions, à plusieurs reprises, le morphème *-gaṇ*, marqueur de pluriel collectif, est suffixé à des substantifs humains pour différencier des lexèmes dont le nombre au nominatif n'est pas marqué autrement. Le hindi-ourdou est une langue flexionnelle à deux cas (direct et oblique), deux genres et deux nombres. Au cas direct, de nombreux noms masculins ont la même forme au singulier et au pluriel. Dans une décision, ce suffixe de pluriel permet de préciser le nombre de personnes concernées.

[55]Pour la signification des abréviations de la glose, cf. p. iii .

(4) *prārthī-gaṇ*
requérant-P
« Les requérants »

(5) *abhiyukta-gaṇ*
mis en cause-P
« Les mis en cause »

Ce morphème de pluriel n'est jamais utilisé dans la langue commune, il appartient à la langue administrative.

Il en va de même pour des mots grammaticaux comme la postposition[56] *hetu*, qui indique le but (*jāñc hetu*, litt. enquête-pour, « pour l'enquête »), et qui a pour synonyme *ke lie* dans la langue commune. Autre exemple avec la très présente conjonction de coordination *evaṁ* « et », tatsam[57] sanskrit, utilisée à la place du *aur* de la langue commune, ou encore la locution exprimant le lien de cause à conséquence est formulé *ke anusaraṇ mẽ*, que l'on retrouve surreprésentée dans tout le corpus judiciaire (cf. chapitre 3) :

(6) *abhiyukta-gaṇ* *kī or se kī gaī prāthanā* ke anusaraṇ mẽ *saimpal*
mis en cause-P requête faite de la part de suite à échantillon
jāñc hetu *bhejne par koī* *fīs* *deya nahĩ hogī.*
pour examen à l'envoi quelque frais ne devra pas être donné
« *Suite à* la requête émanant des mis en cause, l'envoi de l'échantillon pour examen ne sera pas facturé. »

Nous reviendrons sur ce point plus en détails dans le chapitre 3.

Autre trait lexical partagé avec la langue administrative, l'emploi d'adjectifs tirés de participes passés sanskrits : *pārit* (« passé »), 10 occurrences dans le jugement (1267 tokens), *ākṣepit* (« rejeté »), 6 occurrences, ainsi que 16 autres hapax de même formation en *-it*[58], *ukta* (« mentionné »), 7 occurrences.

[56]Une postposition est un relateur qui se place après le mot dont il détermine la fonction. Le hindi et l'ourdou sont des langues à ordre SOV (sujet, objet, verbe) dont presque tous les relateurs sont postposés. En ourdou, sous l'influence du persan, quelques prépositions se sont imposées.

[57]cf. 1.3.1.2

[58]Dérivé de la terminaison *-(i)ta* de participes passés sanskrits.

Les déictiques d'origine sanskrite, communs avec la langue administrative, sont également surreprésentés : se pūrva (« avant »), ke paścāt (« après »), nimnalikhit (« mentionné ci-dessous »), ūparukta (« mentionné ci-dessus »), ukta (« mentionné »). Les synonymes de la langue commune, moins présents, ne sont toutefois pas exclus des textes.

Au niveau des unités phraséologiques, le document de Jaipur se caractérise par des unités polylexicales introductives ou conclusives déterminées, qui structurent le discours argumentatif :

– L'exposé des faits est introduit par la formule :

(7) *pr̥ṣṭhbhūmi is prakār hai ki*
contexte DET.PROX manière est que
« Le contexte est le suivant »

– Des expressions figées impliquant l'énonciateur à la première personne se retrouvent après l'exposé des faits et des procédures de première instance. Elles expriment un jugement ou une prise de décision.

(8) *mujhe vidvān nigarānī nyāyālaya ke is ākṣepit ādeś mẽ*
je.OBL érudit supervision tribunal de.POSS ce rejeté ordre-dans
koī bhī avaidhatā yā truṭi dr̥ṣṭigat nahĩ hotī hai.
quelque.EMPH illégalité ou erreur apparent NEG est
« Aucune illégalité ou erreur ne m'apparaît dans la présente ordonnance de rejet du tribunal d'inspection. »

– Dans le document de la Haute Cour, la partie conclusion et décision est introduite par l'expression suivante, qui n'est pas typique du langage judiciaire et est partagé par la langue commune comme marqueur argumentatif :

(9) *is prakār spaṣṭ hai ki …*
DET.PROX manière clair est que …
« De cette manière, il est clair que … »

Dans la décision du tribunal inférieur, la formulation est beaucoup plus directe :

(10) *ādeś / abhiyukta-gaṇ X, Y, Z ko dhārā 498e, 504, 506, 406 bhā. daṇ. sam.*
ordre / mis en cause-P X, Y, Z-OBL article 498A, 504, 506, 406 I.P.C
tathā 3/4 ḍī pī ekṭ (…) ke antargat zarie samman dināṅk …/…/2012 ke lie
et 3/4 D.P. Act en vertu de par convocation date pour

talab kiyā jātā hai
demandé être.PASS.PRES.3MS
« ordonnance / en vertu des articles 498A, 504, 506, 406 du code pénal indien et des articles 3 et 4 de la loi sur la prévention de la dot, les mis en cause X, Y et Z sont cités à comparaître par convocation en date du .../.../2012. »

Les éléments énonciatifs typiques de cette décision sont l'usage du mode subjonctif et de la voix passive :

Le subjonctif est ici le mode de l'injonction :

(11) *yah nirṇay kiyā jāe ki*
DET.prox. décision faire.PASS.SUBJ. 3.S que
« qu'il soit décidé que » (glose)

(12) *parivādinī gavāhān sūcī dākhil kare*
plaignante témoins liste enregistré faire.SUBJ. 3.S
« Que la plaignante enregistre [sa] liste des témoins. »

Dans cette même séquence, le français utilise une première personne au mode indicatif.

Les phrases passives sont abondantes et, en l'absence d'agent, mettent en avant le processus :

(13) *abhiyukt se do bār fīs lī gayī thī*
mis en cause de deux fois frais prendre.PASS.PST.3fs
« Des frais avaient été demandés deux fois au prévenu »

(14) *aparādhik kāryavāhī samāpt kī gayī*
criminelle procédure terminé faire.PASS.PRES
« Il a été mis fin à la procédure criminelle »

Ce trait est partagé par la langue administrative et la presse. Dans la langue commune, les structures passives sont spontanément peu utilisées.

L'énonciateur se manifeste à plusieurs reprises, à la première personne du singulier :

– au paragraphe 6:

(15) *maĩne donõ pakṣõ ke tarkõ par vicār kar-ke …*
1s-ERG deux partie GEN argument-sur pensée faire- COV B
« J'ai réfléchi aux arguments des deux parties… »

– au paragraphe 8:

(16) *is prakār se maĩ niścit rūp se is*
DET.prox manière-avec 1s apparence déterminée-avec DET.prox
mat mẽ hū̃ ki…
opinion-dans être.1s que
« Ainsi, je suis formellement d'avis que… »

– lors du passage en revue des pièces présentées par les parties (mémoires, ordonnances, registres), correspondant à la formulation « attendu que » dans les documents comparables en français :

(17) *maĩne […] kā avalokan kiyā… , maĩne […] bhī dekhā*
1s.ERG observation de … faire.ACC 1s.ERG aussi voir.ACC
…j'ai observé […], j'ai également vu […]

Au niveau lexical, le texte performatif se caractérise par l'emploi de termes d'origine arabe ou persane pour lesquels il n'existe pas d'équivalent en hindi (*ʿurf*, *banām*), trace de l'héritage de l'administration en langue persane puis en ourdou (cf. 1.1.1.1.1 et 1.1.1.1.2). Les postpositions du hindi courant sont remplacées par des synonymes hautement sanskritisés, marquant un registre écrit et soutenu (*hetu* à la place de *ke lie* pour le but, *ke paścāt* au lieu de *ke bād* pour « après »).

Les unités phraséologiques qui marquent l'articulation argumentative (7, 8 et 9) ne sont pas restreintes aux textes judiciaires. Elles apparaissent dans les textes administratifs en général.

Le niveau énonciatif est marqué par l'usage du passif sans agent, voire sans sujet, permettant ainsi une distanciation par rapport à l'acteur présumé des faits ou une emphase sur le processus. L'aspect polémique de l'énoncé s'en trouve neutralisé. Le subjonctif, parfois passivé, est utilisé pour formuler les ordres. Une autre caractéristique, en comparaison du français, est l'implication de l'énonciateur au fil de l'argumentation (« j'ai entendu les deux parties », « j'ai observé », « je suis d'avis que ». Cet usage de la première personne du singulier, qui va à l'encontre de l'opacité du lexique administratif-judiciaire

et du caractère artificiel de l'emploi du passif, rend le discours plus compréhensible[59]. Il est également présent dans les versions anglaises des décisions, qui sont les originales.

2.3.3.2.2 Procès-verbaux

Les documents suivants sont deux plaintes qui ont fait l'objet d'un FIR ainsi qu'une plainte rapportée dans une décision d'un tribunal inférieur. Comme nous l'avons vu dans la partie (2.2.2.4), les procès-verbaux de dépositions sont des reformulations de question-réponses écrites dans un style plus judiciaire et à la première personne. Ces procès-verbaux sont divisés en trois parties : une synthèse des principaux éléments (références, personnes impliquées, reprise des éléments normatifs que sont les articles du code pénal visés) sous forme de tableau, la déposition sous forme de narration dans une énonciation reformulée, puis les démarches entreprises par le commissariat d'enregistrement avec des unités phraséologiques spécifiques.

Le choix du lexique varie de manière notable. Les lexèmes utilisés dans les phraséologies, sémantiquement stables, sont, selon les documents et les séquences, majoritairement arabo-persans, sanskritisés, ou encore mêlent les deux fonds lexicaux.

On retrouve les éléments décrits à la section précédente concernant la langue artificielle administrative (déictiques, relateurs et conjonctions de coordinations très sanskritisés, emplois du passif absents de la langue commune, en particulier orale). Ces éléments, non seulement présents dans les séquences introductives et conclusives où l'énonciateur est l'officier de police, se retrouvent aussi dans le discours rapporté du plaignant, visiblement reformulé.

Dans la plainte relatant un accident de la circulation, les unités phraséologiques du genre judiciaire sont très persianisées. Elles sont incompréhensibles pour un hindiphone moyen et difficiles à saisir même pour un locuteur d'ourdou plus familier du lexique d'origine arabe ou persane.

Ci-dessous sont donnés quelques exemples de formules très arabisées ou persianisées (AR. pour l'origine arabe, PS. pour l'origine persane) insérées dans des formulations plus sanskritisées (signalées par *skt*) :

(18) *vākiyāt* *muqadmā* *hāz̲ā* *is* *prakār* *hai* *ki*
faits AR. affaire AR. présente AR. cette façon est que

[59]C'est d'ailleurs un trait qu'adoptent de plus en plus les magistrats français dans la rédaction des décisions, dan un but de lisibilité pour un public non spécialiste.

Les faits de la présente affaire se présentent ainsi…

riporṭ par muqadmā uparokta dhārā X IPC mẽ darj kar
rapport-sur affaire susmentionné article X IPC-dans enregistré faire.COV
taftīś kī gaī
enquête faire.PASS.ACC
Une affaire AR. a été enregistrée AR. sur la base de ce rapport au titre des articles X du code pénal indien et l'enquête PS. a été menée.

Ces emplois très persianisés, voire arabisés, ne se retrouvent pas dans les textes normatifs, au lexique très sanskritisé. Ils ne sont pas non plus répandus dans la langue commune. Ils reflètent en réalité la tradition de la procédure judiciaire dont ils sont les caractéristiques lexicales.

La déposition est écrite comme s'il s'agissait d'une narration à la première personne du singulier alors que le déclarant a en réalité été auditionné par un officier de police qui lui a posé les questions. Comme dans les procès-verbaux en français, la reformulation est la règle. On la constate dans l'enchaînement sans fin de propositions relatives tout au long de l'énoncé, en une seule et longue phrase, peu probable dans un discours réel. Cet enchaînement marque le lien de cause à conséquence qui relie les faits. La fin de la déposition est marquée par une autre formule figée supposée prononcée par le déclarant :

(19) *riporṭ kartā hū̃. ucit kāryavāhī karāvẽ*
rapport faire.PRES.1S adéquate procédure faire.SUBJ3P
« Je rapporte [les faits]. Menez la procédure appropriée. »

La partie décrivant les actions entreprises par les policiers est introduite par l'expression figée *kāryavāhī pulis* (« procédure policière ») suivie d'une reprise des éléments du genre normatifs déjà mentionnés en introduction. Les faits sont qualifiés en référence aux articles du code pénal :

(20) *jurm dhārā 279, 337, 304-e bhā-da-sa, 112/183, 184, 134/18 em-vī ekṭ kā bannā*
crime article 279, 337, 304A IPC, 112/183, 184, 134/18 MV Act de devenir
pāyā jāne par …
obtenir.PASS.COV sur …
ayant été qualifié de crime prévu aux art. 279 , 337 , 304A du code pénal indien et aux art. 112/183 , 184 , 134/18 de la Loi sur les véhicules motorisés …

La qualification des infractions est systématiquement exprimée par un syntagme comprenant le verbe *pānā* (« obtenir », « réussir ») à la voix passive, spécificité du genre.

La suite décrit la transmission de la déposition et du FIR (rapport d'information préliminaire) à l'officier chargé de l'enquête. Là encore, les structures passives dominent :

(21) *mazmūn riporṭ se māmalā aparādh dhārā 279-337 IPC kā ghaṭit honā*
sujet rapport-origine affaire infraction article 279-337 IPC de survenu être
pāyā jātā hai. ataḥ mulazim X dhārā uparokta mẽ
obtenir.PASS.PRES. par conséquent prévenu X article susmentionné-dans
kāyam kar taftīś śrī Y e-es-āī ke kī gaī jo pūrva se
maintenu faire.COV enquête M. Y ASI de faire.PASS.ACC qui auparavant
taftīś mẽ maśrūf hai asal FIR bād kāyamī ASI ke pās
enquête-dans occupé est original FIR après constance à l'ASI
badast āmadā LIEU ke ASI bhejī jātī hai pratiyā̃ FIR
en main propre préparée ASI de LIEU envoyer.PASS.PST exemplaires FIR
niyamānusār jārī kī gaī̃.
selon la règle délivrées faire.PASS.ACC
« D'après l'objet du rapport, l'infraction aux articles 279-337 de l'IPC est avérée. Par conséquent l'enquête par le sous-inspecteur adjoint sur le prévenu X au titre des articles susmentionnés, déjà engagée, est poursuivie. L'exemplaire du rapport d'information préliminaire détenu par le sous-inspecteur adjoint est envoyé au sous-inspecteur adjoint de LIEU. Des copies du rapport d'information préliminaire sont délivrées conformément au règlement. »

2.3.3.2.1 Synthèse pour le hindi Il est ici fait usage d'un très grand nombre de termes d'origine arabo-persane que nous retrouverons à l'identique dans les documents de la procédure pakistanais. Pour une étude lexicale plus détaillée, on se reportera au chapitre 3.

Les exemples de textes de la procédure indienne étudiés ci-dessus montrent un recours massif aux tournures passives, ce qui est caractéristique de la langue administrative, avec une emphase sur les processus par rapport aux autres arguments (agent, patient). Les termes d'origine arabo-persane ou sanskrite sont employés en cooccurrence en proportions variables selon les documents et les séquences. Bien que l'usage de doublets provenant des deux fonds lexicaux soit courant en hindi moderne dans la langue commune, il s'agit dans ce cas-ci de termes et d'unités phraséologiques spécialisés, opaques pour la plupart des individus. Le corpus ne présente pas assez d'exemples pour dégager une tendance quant aux proportions moyennes de ce mélange. Une étude sur un échantillon plus vaste pourrait établir si ces choix lexicaux varient en fonction d'usages régionaux

ou individuels. Par ailleurs, dans le cas des procès verbaux, le discours des plaignants ou témoins est systématiquement reformulé et artificialisé à l'aide de lexèmes et de phrasèmes qui n'appartiennent pas à la langue commune.

Dans tous les cas et types de documents, certains des termes d'origine arabo-persane utilisés sont incontournables car ils n'ont pas d'équivalent fonctionnel d'origine sanskrite. Ils reflètent d'une part l'influence de la langue de l'administration judiciaire précoloniale et sa persistance dans la langue moderne, malgré la création officielle d'une terminologie sanskritisée[60]. Le tableau 2.2 récapitule les termes d'origine arabo-persane observés avec leurs équivalents hindi, d'origine sanskrite, et leur traduction en français.

terme arabe / persan observé	**terminologieofficielle**	**traduction**
asal	*mūl*	original
āmādā	*taiyār*	prêt
tahrīrī	*likhit*	écrit
jurm	*aparādh*	infraction
badast	*hāth* se	en main propre
bayān	*kathan, bayān*	déclaration
mukām	*sthān*	lieu
mazmūn	*viṣaya*	sujet, objet
may	*sahit*	y compris
mazīd	*adhik*	supplémentaire
mustagīs	*abhiyogī*	plaignant
hāz̲ā	*yah, is*	ceci
vākiyāt	*ghaṭnāẽ*	évènement
daurane X	X *kī avadhi mẽ*	pendant X
murtib karnā	*sampādit karnā*	rédiger
taftīś	*jā̃c, anveṣaṇ, anusandhān*	enquête

TAB. 2.2: Variations entre la terminologie officielle et les usages dans les textes juridictionnels en hindi. La translittération est réalisée à partir des caractères devanagari, suivant les normes du tableau de l'annexe1.

2.3.3.3 Exemples en ourdou (Pakistan)

2.3.3.3.1 Textes entraînant des effets juridiques

Les textes en ourdou entraînant des effets juridiques étudiés ci-après sont des décisions de la Cour suprême du Pakistan. Ces décisions, rendues publiques sur le site

[60]La partie IV du glossaire juridique officiel publié par le gouvernement indien en 2001 contient près de 2400 termes d'origine arabo-persane avec leurs équivalents en hindi. cf. bibliographie : Minsitry of Law, Justice and Company Affairs (2001).

de la Cour suprême, sont disponibles pour le public[61]. Une mention sur le document signale que la version en ourdou est une traduction d'un original en anglais réalisé par le département de la traduction interne à la cour. En réalité, les versions en ourdou contiennent en moyenne moins de pages que les versions anglaises. Toutes les parties de la décision ne sont donc pas traduites.

Parmi les deux décisions collectées, l'une fait suite à une affaire de corruption, l'autre à des accusations de violence envers un chef de parti politique. On y retrouve la plupart des points soulevés dans les textes en hindi : l'ordre de la présentation des parties du document, avec l'en-tête présentant le référentiel de l'affaire, le corps du document comprenant la présentation des faits, des pièces du dossier, l'argumentation de la formation de jugement, et la décision en conclusion. On y retrouve des éléments de phraséologie observés en hindi : emploi de formules honorifiques, mécanisme de marquage du pluriel pour désambiguïser le nombre, implication de l'énonciateur à la première personne, tournures passives et usage du subjonctif pour formuler les ordres. En revanche, le registre lexical puise exclusivement dans les fonds arabo-persan et anglais, sans mélange avec le fonds sanskrit.

a. Lexique

Le lexique puise dans le fonds lexical arabo-persan et, dans une moindre mesure, emprunte davantage que le hindi à l'anglais. Cette présence de l'héritage linguistique arabo-persan affecte jusqu'au niveau morphologique des unités lexicales, puisque la formation arabe ou persane des pluriels, jugée plus soutenue, est préférée à la formation régulière du hindi-ourdou commun.

La locution *banām* (ici « contre »), déjà vue dans les textes juridictionnels hindi, se retrouve dans l'en-tête pour opposer les deux parties.

Le lexème honorifique de désignation ou d'adresse *janāb* est utilisée à l'égard des juges, du procureur (*prāsekyūṭar*) et devant un titre administratif :

(22) *janāb* *jasṭis* *X*
HON.3.M.S Juge (Haute Cour) X
« Monsieur le Juge X »

[61] Adresse du site de la Cour suprême du Pakistan : `http://www.supremecourt.gov.pk/web/`.

(23) *janāb* *prāsekyūṭar*
HON.3.M procureur
« Monsieur le procureur »

(24) *janāb* *Y ḍepṭī* *ḍāyrekṭar qaumī ahtasāb byūro*
HON.3.M.S Y adjoint directeur Bureau National des Comptes
« Monsieur Y, directeur adjoint du Bureau National des Comptes »

En effet, comme signalé pour le hindi, l'ourdou juridique utilise des relateurs (de but, spaciaux ou temporels) plus soutenus que dans la langue ordinaire, issus du persan ou de l'arabe : *barāe* « pour » et *az* « à partir de » (du persan, antéposés et non postposés à leur complément), *bālā* « au dessus, plus haut » et *z̲ail* « en dessous, ci-après », de l'arabe, etc.

Sur le plan de la deixis, on note l'utilisation du démonstratif de proximité arabe *hāz̲ā* postposé au mot qu'il détermine :

(25) *muqadmah* *hāz̲ā*
affaire DEM
« *cette* affaire /*la présente* affaire »

(26) *ʿadālat* *hāz̲ā*
tribunal DEM
« *ce* tribunal / *le présent* tribunal »

Ce démonstratif n'est utilisé que dans des contextes administratifs. Dans les décisions étudiées, il n'apparaît qu'en cooccurrence avec *muqadmah*, « affaire », et *ʿadālat*, « tribunal ». Ce démonstratif est absent de la langue commune et n'est pas employé à l'oral. La langue ourdou commune, comme en hindi, utilise des démonstratifs de proximité ou d'éloignement d'origine indo-aryenne : *yah* / *is* pour la proximité, *vah* / *us* pour l'éloignement au singulier[62]. Ils sont toujours antéposés au déterminé. Le démonstratif arabe *hāz̲ā* est donc un élément spécifique que la phraséologie judiciaire partage avec le langage administratif. On remarque qu'il n'existe pas d'unité phraséologique équivalente en hindi judiciaire ou administratif.

[62]La variation correspond au cas direct ou oblique. Au pluriel, le démonstratif de proximité est *ye* / *in* (HI.), *yah* / *in* OU., et pour le démonstratif lointain : *ve* / *un* (HI.), *vah* / *un* (OU.).

Comme en hindi, des formes de pluriel appartenant à un registre soutenu sont préférées aux formes régulières communes au hindi et à l'ourdou. La marque du cas oblique pluriel (terminaison *-õ*) est neutralisé par ces formes d'origine persane ou arabe. Dans les cas où le pluriel n'est pas marqué (cas des pluriels masculins au cas direct),ces formes permettent de les distinguer des formes au singulier, ou simplement de donner de la solennité au texte. Ce phénomène est partagé par tous les textes au registre soutenu, qu'ils soient administratifs ou relevant d'autres genres. Le tableau 2.2 récapitule pour quelques termes les formes plurielles de la langue commune et les pluriels irréguliers constatés dans les décisions :

singulier	**pluriel courant**	**pl. soutenu**	**traduction**
gavāh	*gavāh*	*gavāh-ān*	témoin(s)
huk(u)m	*huk(u)m*	*ahkām / ahkāmāt*	ordre(s), commandement(s)
vajah	*vajahẽ*	*vajūhāt*	raison(s), motif(s)
mohājir	*mohājir*	*mohājirīn*	exilé(s)

FIG. 2.2: Pluriels communs ou soutenus en ourdou

Le premier terme, *gavāh*, d'origine persane, prend le suffixe de pluriel persan *ān*. Les autres termes sont d'origine arabe. Dans cette langue, une manière régulière de former des pluriels masculins est de suffixer le morphème *īn* (cf. *mohājir-īn*, « exilé-s »). Un autre mécanisme consiste à modifier les voyelles autour du cadre consonantique du mot (pluriel interne ou brisé), ce que l'on observe avec *huk(u)m* et *vajah*. Ces procédés de formation des pluriels suffixés ou internes arabes font partie des traits plus soutenus de la langue ourdou. La situation est analogue à celle où le hindi ajoute un morphème de pluriel aux classes d'humains.

Des formants provenant de participes persans entrent dans la composition de collocations : *kardah*, participe passé du verbe *kunan* (faire), *śudah*, participe passé du verbe *śodan* (être) :

(27) *X ko jārī kardah do pāsporṭ*
X à *délivré faire.PART.PASSÉ* deux passeports
« Les deux passeport [qui ont été] délivrés à X … »

terme observé	terme en anglais	terme ourdou (non utilisé)
prāsekyūṭar	prosecutor	*sarkārī vakīl*
saprīm korṭ	Supreme Court	*ʿadālat-e-ʿuẓmā*
cearmen	chairman	*ṣadar*
ḍāyrekṭar	director	*nāẓim*

TAB. 2.3: Emprunts anglais et terminologie ourdou (Pakistan)

(28) *manẓūr śudah* *parvānah giraftārī*
accepté être.PART.PASSÉ mandat arrestation
« mandat d'arrêt validé. »

Les deux décisions totalisent 6 occurrences de *kardah* et 4 de *śudah*.

À la différence des décisions en hindi, le recours aux termes anglais translittérés est fréquent (11 occurrences de *cearmen* pour « chairman », 11 de *ḍāyrekṭar* pour « director », 8 occurrences de *saprīm korṭ* pour « Supreme Court » sur les deux décisions)[63]. Ces termes ne se retrouvent pas en concurrence avec des équivalents terminologiques ourdou :

Cet usage de l'anglais montre une adaptation de l'énonciateur à utiliser une désignation plus connue et plus utilisée à l'oral, moins opaque que les équivalents officiels en ourdou.

b. Phraséologie

L'énonciateur est présent à la première personne du pluriel tout au long de la décision[64]

(29) *hamāre maẕkurah bālā faiṣle*
nos-1P mentionné dessus décisions
« Nos décisions mentionnées précédemment » (paragraphe 2)

(30) hamāre *ḥukm par*
notre-1P ordre sur

[63]Sur un corpus totalisant 4 810 tokens.

[64]Le tribunal est formé de deux juges, d'où le pluriel.

« Sur notre ordre » (paragraphe 2)

(31) hamne *muddat dī*
nous-ERG. délai donner.ACC.3.S
« Nous avons accordé un délai. »

(32) hamẽ *[…] naẓar nahĩ āī*
à-nous […] vision n'est-pas-venue
« Nous n'avons pas vu […] »

L'énonciateur exprime un jugement à travers des question rhétoriques, contrastant avec l'abondance des tournures passives qui mettent en avant le processus :

(33) *lihāz̲ā qaumī aḥtasāb byūro ko hī vażāḥat karnī cāhie*
par conséquent national comptabilité bureau-à EMPH explication faire falloir
kih X mẽ aisī kaun-sī ġair-maʿmūlī qābliyat thī jis kī muʿabid se
que dans X ainsi quelle extraordinaire compétence était REL aide INST
maz̲kurah maʿāmle kā maʿāinah un-se karāyā jānā nā-guzer
mentionnée affaire-deGEN examen 3.P.HON.INST faire.CAUS.PASS indispensable
thā ?
était ?
« Il reviendra par conséquent au Bureau National des Comptes d'expliquer en quoi les compétences hors du commun de Monsieur X rendaient indispensable l'examen de ladite affaire par celui-ci. » (paragraphe 8)

Les phrases passives sont nombreuses :

(34) *saprīm korṭ se muʿāvinat kī dar̲kh̲vāst* kī jātī hai
de la Cour suprême aide-deGEN requête faire.PASS.PRES.3.F.S
« Une requête *est adressée* à la Cour suprême. »

(35) *tamām aqdāmāt pahle hī kie jā cuke haĩ*
toutes mesures avant EMPH *faire.PASS.PRES.3.M.P*
« Toutes les mesures ont auparavant déjà été prises. »

On recense 93 occurrences de syntagmes verbaux à la forme passive pour les deux décisions. Comme en hindi, l'utilisation fréquente de la voix passive n'est pas un trait de la langue commune. Elle caractérise en revanche la langue administrative et de presse.

Le subjonctif conserve en ourdou le rôle déjà décrit en hindi : il sert à formuler les ordres (25 occurrences recensées)[65].

(36) *T. S. ko ḥirāsat mẽ le liyā jāe*
T.S. OBL détention dans emmener.PASS.SUBJ.3S
« Que T.S. soit conduit en détention. »

(37) *vah un afrād kī niśāndehī karẽ*
3P DET.3.P.DIST hommes de indication faire.SUBJ.3.P
« Qu'ils localisent ces hommes. »

Les occurrences de collocations formées d'une base nominale ou adjectivale suivie du formant *kardah* (participe passé du v. persan « faire ») ou *śudah* (part. passé du verbe persan « être ») sont plus fréquentes que la moyenne (cf. 2.3.3.3.1). En persan comme en hindi-ourdou, les verbes simples, composés d'une seule unité, sont moins nombreux que les prédicats complexes composés d'une base nominale ou adjectivale suivie d'un verbe support. Les plus fréquents de ces verbes sont *karnā* « faire » ou *honā* « être ». Les collocations mentionnées plus haut, formées en composition avec les participes passés persans *kardah* et *śudah*, sont des variantes du registre soutenu des collocations équivalentes formées avec *karnā* ou *honā* en verbes supports :

1.3.7 **registre soutenu**	**registre courant**	**traduction**
jārī kardah	*jārī kiyā gayā (jārī + karnā.PASS.ACC)*	[qui a été] délivré
jārī śudah	*jārī huā (jārī + honāACC)*	délivré
manẓūr śudah	*manẓūr huā (manẓūr + honāACC)*	accepté

Ici, le formant adjectival *kardah* a pour équivalent la forme participiale accomplie et passive de *karnā* (*kiyā gayā*), utilisée en fonction adjectivale.

Concernant l'articulation du texte, on retrouve des formulations équivalentes, tant dans leur forme que dans leur fonction, à celles relevées dans les décisions en hindi. Dans la phase d'introduction de l'exposé des faits :

(38) *muqadmah ke pas-e-munaẓar mẽ vahī faiṣlah hai kih …*
affaire de contexte dans 3.S.EMPH décision est que

[65]Il sert également à formuler des demandes polies à la forme plurielle de 1re personne, mais cette fonction n'est pas une caractéristique du genre : elle fait partie de la langue commune.

« Dans le contexte de cette affaire, la décision est la suivante … »

Dans le corps du texte, des unités phraséologiques faisant référence à la procédure sont récurrentes :

– *bayān kiyā jānā kih* (« être déclaré que »), ou encore

(39) *unke bayānāt se yah ẓāhir hotā hai kih …*
leur déclarations-de ceci clair est que
« Il ressort de leurs déclarations que … » (paragraphe 10, décision 2)

– *maʿāmlah kisī ke rū bah rū peś kiyā jānā* se dit d'une affaire qui est présentée à un tribunal,

– *ʿadālat hāz̲ā ke savāl kā javāb dete hue* (« répondant à la question de ce tribunal »)

Elles contiennent, ici encore, des éléments lexicaux arabo-persans soutenus : terminaison du pluriel arabe en *-āt* dans *bayān-āt*, locution adverbiale persane *rū-ba-rū*, emploi du démonstratif arabe *hāz̲ā* postposé.

Dans les deux affaires, les décisions renvoient les parties à une date d'audience et précisent sous forme d'ordres les mesures à prendre d'ici-là. Il n'y a pas de formule particulière introduisant cette conclusion comme le « PCM »[66] français.

Dans les textes analysés ressortent des points comparables aux textes de même genre en hindi : les emplois verbaux (tournures passives, subjonctifs), les termes honorifiques, les morphèmes de pluriel ou le choix d'un lexique soutenu difficilement compréhensible par un citoyen ordinaire. L'ourdou se distingue du hindi par un registre soutenu empruntant massivement des traits lexicaux et morphologiques à l'arabe et au persan (prépositions, formation du pluriel, utilisation de participes persans adjectivaux en remplacement des participes adjectivaux courants) et en empruntant davantage de termes à l'anglais.

2.3.3.3.2 Procès verbaux

Les procès-verbaux pakistanais en ourdou reprennent les caractéristiques du sous-genre vues en hindi.

Le lexique et la phraséologie sont ici aussi largement influencés par la tradition administrative arabo-persane. Alors que les dépositions se font dans la langue générale, le

[66]Par ces motifs, cf. 2.3.2.1.1.

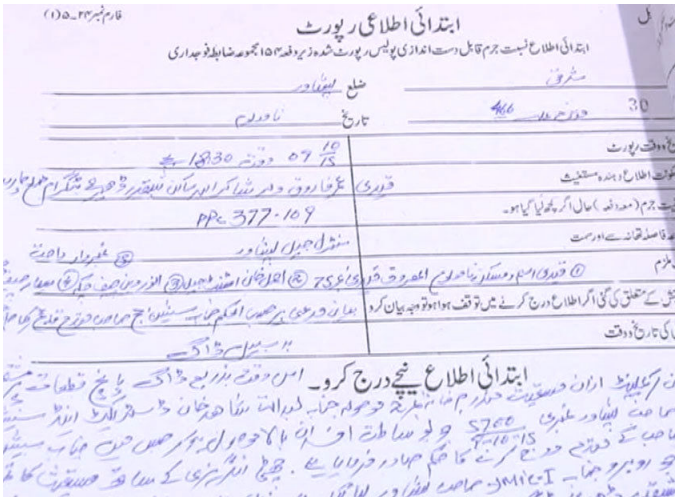

FIG. 2.3: Exemple de FIR : en-tête. Source : site du journal pakistanais Dawn `www.dawn.com`, crédit : Ali Akbar, 13/10/2015

compte-rendu fourni par les documents montre des expressions trop spécialisées pour être compréhensibles d'un large public. Comme en hindi, les procès verbaux, dont la forme n'a pas varié depuis l'indépendance, conservent un aspect opaque, souvent critiqué.

a. Structure générale

Le corpus de cette recherche compte six textes juridictionnels informatifs dont la structure séquentielle oscille entre le genre normatif exprimé dans un langage légal (rappel des articles de loi, des étapes la procédure) et une relation des faits à la première personne dans une langue plus générale émaillée de termes administratifs. Ici encore, la structure est mixte et le niveau de technicité des passages varie entre la langue générale et la langue judiciaire.

Dans les FIR[67] (au nombre de cinq dans notre corpus), qui sont à la base d'une procédure pénale, un en-tête rappelle l'article du code de procédure pénale fondant le cadre juridique du rapport (article 154). Suit un tableau reprenant les circonstances du rapport : lieu des faits, nom du déposant, articles de loi visés, état des dommages, état de la procédure judiciaire. Les propos du déposant, reformulés comme vu précédemment, sont ensuite consignés par écrit. Enfin, une rubrique fait état de l'action judiciaire qui a suivi la déposition. Chaque section est clairement délimitée par des titres.

[67]Pour rappel : rapport d'information préliminaire.

TERME A.P.	GLOSE	SENS	LANGUE COURANTE
is isnā mẽ	DEM.PROX période dans	au même moment	*is vaqt*
masmī	dénommé	dénommé	–
kis-ān	homme.P	homme.P	*ādmī*
sakin	qui demeure (adj)	demeurant	*kā rahnevālā*
savārī moṭar kār	passager voiture-à-moteur	à bord d'un voiture	*gāṛī mẽ savār*
musalaʿ aslaʿ-e-ātiś	armé arme-à-feu	munis d'armes à feu	[nom de l'arme]
barkhilāf	envers-contre	contre	*ke khilāf*

TAB. 2.4: Équivalences lexicales langue spécialisée / langue commune

b. Lexique

Le lexique très arabo-persianisé (A.P.) indique une reformulation de l'énoncé du déposant pour correspondre au sous-genre du rapport :

Ces éléments ne font pas partie du lexique de la langue courante mais de la langue soutenue et administrative. On y retrouve au niveau des morphèmes la marque de pluriel persane *ān* (*kis-ān*), le relateur persan appelé izafah (*-e-*), ainsi que la particule persane *bar* (« avec, contre ») dans *barkhilāf*.

Toute la procédure qui suit est décrite dans des termes empruntant au registre arabo-persan spécialisé (en bleu) :

(40) *is vaqt ek tahrīrī darkhvāst mustaġīs khānah nambar 2 mere rū ba-rū*
en ce moment une écrite requête plaignant case numéro 2 face à moi
barkhilāf masmī X vaġairah peś kī gaī jo maẓmūn darkhvāst se
contre dénommé X etc. présenter.PASS.ACC REL sujet requête.INST
baṣūrat jurm 324 C.P.P. pāī jātī hai. jis par
à l'apparence infraction 324 C.P.P. obtenir.PASS.PRES sur lequel
riporṭ ibtadāī hāzā murtib karke is darkhvāst se naqal FIR
rapport préliminaire DEM rédiger. COV DEM requête.INST copie FIR
baġraẓ taftīś afsarān ko arsāl kar dī hai.
avec intention enquête officiers-à envoyer.PASSÉ.COMP.
« Une plainte écrite du plaignant [signalé à la] rubrique numéro 2 m'est présentée contre le dénommé X et autres. L'objet [de la plainte] est une infraction prévue à l'article 324 du CPP. Le présent rapport préliminaire est rédigé en conséquence. La présente requête ainsi qu'une copie du rapport d'information préliminaire sont envoyées aux officiers enquêteurs. »

On retrouve dans la formulation ci-dessus des lexèmes d'origine arabo-persane et

des tournures présents dans le FIR indien du Rajasthan (cf. 2.3.3.2.2) : *muqadmā hāzā* (HI.) / *muqadmah hāz̲ā* (OU.) « présente affaire », *taftīś* « enquête), *jurm* (infraction), *pāyā jānā* (être obtenu), *badast* (en main propre), *asal* (HI.) / *aṣ̲al* (OU.) « original »), *mazmūn* (HI.) /*mażmūn* (OU.) « sujet ».

c. Phraséologie

La déposition est généralement introduite par l'expression figée :

(41) *bak̲hidmat janāb SHO ṣ̲āḥab thānah [lieu]*
à l'attention de Monsieur-HON SHO Monsieur-HON commissariat [lieu]
janāb-e-ʿālī ! guzāriś hai kih …
Monsieur-HON ! requête est que
« À l'attention de Monsieur le Responsable du commissariat de (lieu). Monsieur, voici ma requête : »

La locution *bak̲hidmat* est composée de la particule persane *ba* indiquant le moyen ou l'association, et du substantif *k̲hidmat*, « l'aide, le service ». Les termes *janāb* et *ṣ̲āḥab* sont des éléments honorifiques dénominatifs. L'unité polylexicale suivante, *janāb-e-ʿālī*, sert de moyen d'adresse honorifique. L'expression figée suivante *guzāriś hai ki* introduit la déposition, qui passe ensuite à la première personne. Cette expression sert habituellement à formuler un souhait ou une requête, or celle-ci (le déclenchement des poursuites) ne sera formulée avec une autre expression figée qu'à la fin de la narration des évènements :

(42) *mulzamān ke k̲hilāf muqadmah darj kiyā jāe aur*
mis-en-cause.P deGEN contre affaire enregistrée faire.PASS.SUBJ et
qarār vāqaʿī sazā dī jāe
définitivement peine donner.PASS.SUBJ
« Qu'une affaire soit enregistrée contre les mis en cause et qu'une peine définitive soit prononcée. »

ou une variante :

(43) *mulzamān ke k̲hilāf qānūnī kārravāī kī jāe,*
mis-en-cause.P deGEN contre légale procédure faire.PASS.SUBJ,
sak̲ht se sak̲ht sazā dī jāe
peine la plus stricte donner.PASS.SUBJ
« Qu'une procédure soit menée à l'encontre des mis en cause, que la peine la plus stricte soit donnée. »

Pour un rapport effectué par un policier à un supérieur, l'adresse introductive est plus directe, sans élément honorifique :

(44) *bak̲hidmat afsar in cārj thānah [lieu] …*
à l'attention de officier en charge commissariat [lieu]
« À l'attention de l'officier en charge, commissariat de (lieu) … »

La suite de la déposition est une narration à la première personne, comme vu précédemment. Parfois, le texte repasse à la 3e personne, comme dans cet exemple (FIR-2), alors que l'énonciateur est toujours le déposant :

(45) *maĩ masmī X vald Y sākin Z kā rihāiśī hū̃ […] merā*
1S dénommé X fils de Y qui demeure Z de résident être.1S mon
kārobār pārṭnar nahĩ hai […] jab sāil apne daftar thā …
associé d'affaires n'est pas quand plaignant son bureau était
« Je soussigné X fils de Y, demeure à Z. Je n'ai pas d'associé d'affaires. Quand le plaignant était dans son bureau … »

La partie conclusive, qui concerne la procédure entreprise par la police, est celle qui présente le lexique et les unités phraséologiques les plus techniques et les plus arabo-persanes. Elle est introduite par les expressions figées *kārravāī pūlīs* (procédure-police, « action de la police ») ou moins fréquemment *az thānah* (« depuis le commissariat », avec la préposition persane *az*).

d. Synthèse pour l'ourdou

Dans les textes des procès-verbaux se retrouve l'emploi de structures et de lexique empruntés à l'arabe et au persan. Ces éléments sont majoritairement communs à la langue administrative, voire correspondent plus largement au registre soutenu de l'ourdou, qui prend pour modèle les sources arabe et persane. Ils caractérisent un niveau de langue spécialisé et sont difficilement compréhensibles par un public non initié au droit. L'opacité de ces tournures et du lexique est critiquée par les avocats. [68] qui soulignent que la formulation des FIR n'a pas évolué depuis l'Indépendance et que, même lorsque le déposant est illettré, le document écrit comporte des tournures très persianisées, signifiant

[68] Dans une vidéo expliquant la procédure pénale (`https://www.youtube.com/watch?v=MoyngoudBhU`), Shauket Rafique Bajwa, avocat à Lahore, reproche ainsi aux officiers de police de transformer les déclarations des déposants en un jargon incompréhensible, exemples à l'appui (à partir de 13min30).

que l'énoncé a été retravaillé. Dans les exemples ci-dessous, que l'on trouve parfois dans les FIR enregistrées dans des commissariats ruraux, le suffixe *-am* marque un possesseur à la 1re personne du singulier en persan. En ourdou, comme en hindi, les possessifs correspondants sont des unités autonomes qui s'accordent au déterminé et se placent devant eux. Le mot *barādar* est un emprunt au persan qui n'appartient pas à la langue commune.

(46) *sāsur-am, barādar-am,*
beau-père.1s, frère.1s
« mon beau père, mon frère »

Synthèse pour le hindi et l'ourdou

Les textes de la procédure hindi (en Inde) et ourdou (au Pakistan) possèdent des points communs quant à l'organisation générale de leur structure, ce qui témoigne de leur passé commun, et d'une absence d'évolution des documents depuis l'Indépendance. Ils partagent l'usage abondant des tournures passives, artificielles et considérées comme administratives, que l'on retrouve aussi en anglais. Les termes qu'ils utilisent appartiennent à des registres linguistiques perçus comme élevés et puisent dans les langues classiques de référence, sanskrit ou arabe et persan, et peu compréhensibles du grand public. Le problème est rendu d'autant plus opaque que, comme nous l'avons vu au chapitre 1, l'ourdou est la langue maternelle d'une très petite minorité de Pakistanais, et que la très grande majorité restante le maîtrise à un niveau inégal.

D'autre part, si l'on constate une influence de la terminologie pré-indépendance d'origine arabo-persane dans les textes indiens en hindi, les textes en ourdou, eux, ne contiennent pas d'éléments sanskrits dans leur terminologie de spécialité. Les termes arabo-persans sont en effet présents depuis des siècles dans la culture juridique et administrative commune au deux pays, tandis que la terminologie sanskrite est de diffusion beaucoup plus récente. Enfin, l'ourdou judiciaire semble plus perméable au lexique anglais que le hindi. Il n'hésite pas à intégrer des désignations et fonctions directement translittérées, que le grand public a plus facilement l'occasion d'entendre.

2.3.4 *Conclusion*

La traduction judiciaire s'est constituée récemment en champ d'étude, au croisement de la traductologie et du droit, afin de répondre à des problématiques spécifiques comme

la fonction de la traduction dans la mission demandée par la justice ou la légitimité de l'interprétation du droit par des non-juristes, en l'occurrence les traducteurs-interprètes. Ces questions théoriques sont soulevées par la pratique de terrain des traducteurs-interprètes qui doivent compter sur leur connaissance fine de la langue, sur une connaissance globale et comparative des procédures pénales dans les trois pays, et sur leur capacité à adapter le message au destinataire de la traduction. Les procédures pénales font l'objet d'un genre textuel divisé en sous-genres (documents ayant un effet juridique et procès-verbaux) dont l'étude permet de caractériser les tendances lexicale, phrastique et énonciative. Ces caractéristiques guideront le traducteur-interprète dans l'étape de reformulation vers le destinataire, qui pourra proposer un message plus idiomatique. L'aspect plus naturel de l'expression juridique lui conférera en outre un pouvoir solennel plus important. Le traducteur-interprète doit toutefois conserver à l'esprit l'obstacle de l'opacité terminologique, afin de rendre son message compréhensible et d'en conserver les effets du genre.

Poursuivant sur la lignée des caractéristiques du genre judiciaire, le chapitre suivant s'attachera à poser les bases de la constitution de ressources lexicales et phraséologiques pour la traduction. À travers des analyses et traitements informatisés, nous ferons ressortir de l'ensemble du corpus collecté les termes associés à la procédure pénale, ainsi que leurs cooccurrents et phraséologies. Il sera ainsi possible à la fois de créer des lexiques et des concordanciers multilingues dans lesquels on trouvera des alignements de termes, de concepts ou de contextes. Ces différentes ressources permettront de rechercher un terme ou le contexte pour valider des équivalents de traduction.

Chapitre 3

Constitution et traitement informatique du corpus multilingue

Dans le chapitre 2, nous nous sommes attachée à décrire les situations d'intervention et les rôles dévolus au traducteur-interprète, amené à traduire des textes de procédure ou à interpréter dans les différentes phases de celle-ci. Nous y avons vu que le traducteur et l'interprète, afin de mener à bien le processus traduisant qui consiste à comprendre le message, le déverbaliser puis le reformuler pour le restituer à un destinataire précis, se devaient –sans devenir des spécialistes chevronnés du droit– de connaître dans les grandes lignes le processus judiciaire et ses protagonistes. Il leur est ainsi possible de comprendre l'intention du producteur du message, son but, et d'en restituer les effets en adaptant la reformulation pour le destinataire. Les effets d'un texte, comme évoqué dans le précédent chapitre, sont liés à son genre textuel, caractérisé aux niveaux du lexique, de la phraséologie et de l'énonciation. Nous avons pu comparer ces traits dans les trois langues, français, hindi et ourdou, pour les textes juridictionnels.

Dans ce troisième chapitre, nous partirons de la collecte du corpus en trois langues pour exposer les étapes de la chaîne de traitement qui ont permis son analyse lexicale et phraséologiques selon la langue, et le sous-genre du judiciaire. La partie (3.1) présentera tout d'abord la notion de corpus retenue dans cette étude et ses objectifs pour notre recherche. La partie (3.2) détaillera la composition du corpus et ses genres : textes normatifs (qui disent la loi) hautement techniques, textes juridictionnels mêlant des passages très techniques et de la langue non spécialisée, textes explicatifs et textes déterminologisés de la presse judiciaire utilisant des tournures et du lexique de spécialité ou des expressions en langue commune relevant du champ sémantique judiciaire.

Dans la partie (3.3), nous évoquerons tout d'abord l'état de l'art de la recherche sur le traitement des corpus en hindi et en ourdou (3.3.2), afin de passer en revue les outils de collecte de corpus et d'analyse textuelle disponibles en ces langues. Nous aborderons ensuite les différentes étapes du traitement informatique de notre corpus qui

ont été nécessaires à l'alignement et à l'extraction des termes et unités phraséologiques. Le résultat de ces analyses fait l'objet du chapitre suivant.

3.1 Définition et objectifs de la constitution du corpus

Deux mouvements animent l'exploitation des corpus en linguistique de corpus. Le premier, basé sur les corpus (« corpus-based »), consiste à vérifier des modèles et des hypothèses dans un corpus. Le second, guidé par le corpus (« corpus-driven » (Bonelli, 2002), « laisse parler » les données pour concevoir ensuite des modèles et des théories. La présente étude aura recours à ces deux approches. D'une part, l'approche basée sur le corpus servira à vérifier la réalisation sous forme d'occurrences terminologiques de concepts extraits d'un repérage thématique. En complément, l'approche guidée par le corpus fournira les cooccurrents et unités phraséologiques trouvés grâce à l'identification de ces termes ou mots-clés. Quelle que soit l'approche considérée, le corpus constitue bien un point de départ pour la recherche d'équivalents traductionnels.

Dans cette section, nous présenterons tout d'abord la définition retenue du « corpus », car il en existe de nombreuses (3.1.1), puis nous expliquerons dans quel(s) but(s) le corpus a été constitué (3.1.2).

3.1.1 *Qu'est-ce qu'un corpus ?*

Le terme « corpus » a longtemps appartenu au champ religieux, puis au champ juridique. Au XIX^{e} siècle, « corpus » ou « corpus juris » désigne encore la collection du droit romain[1]. Dans la langue courante, un corpus est défini au sens large comme un « recueil réunissant ou se proposant de réunir, en vue de leur étude scientifique, la totalité des documents disponibles d'un genre donné, par exemple épigraphiques, littéraires, etc. »[2].

En sciences du langage, la dimension de « collection » n'implique pas nécessairement l'exhaustivité. Mellet propose cette définition : « recueil, formé d'un ensemble de données sélectionnées et rassemblées pour intéresser une même discipline » (Mellet, 2002). Le corpus prend la forme d'un « ensemble d'éléments sur lequel se fonde l'étude d'un phénomène linguistique » (Dalbera, 2002). Sinclair donne la définition suivante : « Un

[1] Dictionnaire Littré, 1863.

[2] TLFI (dictionnaire informatisé de la langue française), `http://stella.atilf.fr/Dendien/scripts/tlfiv5/visusel.exe?11;s=2757470295;r=1;nat=;sol=0;` site consulté le 28/03/2019.

corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage. » (Sinclair, 1996, p.4)[3] Les critères de sélection et de représentativité distinguent ainsi un corpus d'une simple collection (Habert *et al.*, 1997). On visera à obtenir un sous-ensemble de la réalité sur lequel pratiquer des analyses dont les conclusions pourront s'appliquer à un ensemble plus vaste. Dans cette étude, le corpus est un ensemble de textes construit à des fins d'analyse particulières. Il n'est délibérément pas représentatif de la langue en général.

3.1.1.1 Exhaustivité et représentativité

La représentativité d'un corpus est une question centrale lors de son élaboration. Dans le cas d'une recherche de phénomènes s'appliquant à l'ensemble de la langue (usages grammaticaux, lexicaux, habitudes langagières), on aura recours à un corpus de référence qui fournit des informations exhaustives sur la description de la langue (Habert *et al.*, 1997; Sinclair, 1996) et devra représenter par échantillonnage la variété des genres et du lexique.

Notre étude n'a pas de visée comparative sur l'ensemble des variétés des trois langues. Le corpus est un corpus de spécialité, au sens où le définit Sinclair (1996): il se limite à un domaine, celui de la procédure pénale. Il ne regroupe cependant pas l'intégralité des textes judiciaires, ni même ne constitue un échantillonnage en proportions représentatives de la totalité des textes ayant trait à la procédure pénale. Il a été construit à des fins d'analyse de phénomènes particuliers, on y a nécessairement introduit des présupposés, sa constitution n'est pas objective (Mellet, 2002; Dalbera, 2002).

Ainsi, dans notre recherche, on souhaite repérer d'une langue à l'autre des termes susceptibles de fournir des équivalents de traduction dans divers genres textuels judiciaires. On cherche aussi à enrichir une base lexicale dans laquelle le traducteur pourra puiser en fonction des contextes ou vérifier ses choix. On orientera donc la collecte des textes afin de couvrir les problèmes de traduction du judiciaire liés aux textes juridictionnels. La sélection des éléments du corpus portera ainsi pour chacune des langues sur des textes normatifs afin d'obtenir des définitions terminologiques attestées, des textes de la procédure judiciaire pour couvrir le plus d'usages de terrain possible, et des textes de vulgarisation qui permettront d'effectuer le passage de la connaissance théorique des concepts judiciaires vers une connaissance à la portée de destinataires non spécialistes. Notre corpus de spécialité est par conséquent subdivisé en trois sous-corpus, un pour

[3]Traduit de l'anglais dans (Habert *et al.*, 1997).

chaque langue, dans lesquels se retrouvent les différents genres textuels.

Pour une recherche donnée, on s'efforce de constituer un corpus représentatif échantillonné à l'aide de faits pertinents : il s'agit d'un ensemble construit (Dalbera, 2002) et non d'une sous-représentation du réel. Notre corpus étant limité au domaine judiciaire, c'est donc bien un corpus spécialisé, bien que tous les sous-corpus qui le composent n'appartiennent pas à la langue de spécialité.

3.1.1.2 Disponibilité des données

La contrainte de la disponibilité des données vient encore infirmer l'idéal d'un corpus parfaitement représentatif et équilibré. En effet, la linguistique de corpus s'appuie sur des données numériques. La collecte du corpus passe généralement par une phase de recension des données disponibles sur internet, ou par une numérisation des documents qui sont en version papier. L'inégalité de la disponibilité de documents satisfaisant à ces critères entraîne des difficultés supplémentaires pour constituer un corpus représentatif.

Le problème se pose pour la comparaison des trois sous-corpus considérés dans cette étude. Bien que le hindi et le français comptent plusieurs centaines de millions de locuteurs (442 millions pour le hindi et 285 pour le français, langues première et seconde confondues)[4] et peuvent de ce seul fait sembler très bien représentés sur internet, il n'en est rien. Le nombre de locuteurs n'est pas corrélé au développement informatique d'une langue. Les raisons en sont évidentes : si, en France, le taux d'équipement informatique ou d'accès à internet est très élevé, il est beaucoup plus faible en Inde ou au Pakistan, du seul fait des disparités de revenus entre les populations[5]. En Inde et au Pakistan, de très nombreuses institutions ne sont pas encore totalement informatisées et l'archivage papier y reste la norme. Comme nous le verrons dans l'état de l'art de la linguistique informatisée (3.3.2), la généralisation des standards Unicode pour le codage des caractères (en UTF-8) et le développement de ressources et de logiciels de traitement des corpus sont beaucoup plus récents dans les langues d'Asie du Sud qu'en français. Pour rendre compte de ces disparités, (Tròn *et al.*, 2005) introduit la notion de « densité » des données informatisées comme étant la « disponibilité de matériel archivé numériquement ». Le niveau de densité d'une langue répond à deux critères : d'une part le nombre de locuteurs, et d'autre part sa représentativité sur internet. La quantité et la variété des outils et ressources numériques

[4]Source : Chandramouli (2011) et la page hindi du site `http://www.ethnologue.com`.

[5]71% en France contre 8% en Inde, source : baromètre Calvet, `http://wikilf.culture.fr`.

doivent être proportionnelles à cette représentativité[6]. Cette classification ne prend pas en compte l'outillage informatique dans les langues, paramètre que propose (Berment, 2004, p.4) dans une classification en langues peu dotées, moyennement dotées ou très bien dotées, en fonction d'une note qui prend en compte la disponibilité des techniques informatiques, des ressources et le taux de satisfaction des utilisateurs.

Le français souscrit aux critères d'une langue très bien dotée, avec une forte présence sur internet tant quantitative que qualitative et des outils numériques développés. La fonctionnalité statistique de Wikipédia, qui possède des statistiques de l'activité de consultation et de rédaction des pages internet[7], offre un indice intéressant sur la dotation numérique des langues. On y apprend que plus de deux millions d'articles sont écrits en français, et que le taux d'éditeurs francophones est de 17 pour un million de locuteurs. En comparaison, le hindi, avec sa communauté de locuteurs parmi les cinq plus importantes du monde, totalise 131 000 articles et seulement 0,4 éditeur pour un million de locuteurs. Elle ne souscrit donc pas aux critères de la première catégorie, en tout cas pour ce qui est des données textuelles. L'ourdou, quant à lui, appartient nettement à la catégorie des langues moyennement dotées : avec 66 millions de locuteurs natifs, il totalise 144 000 articles (dont 60% générés par des robots) et un éditeur pour un million de locuteurs[8]. De plus, les articles sont souvent moins longs et moins complets dans ces deux langues qu'en français. Ces disparités de disponibilité des ressources se retrouvent dans le nombre de tokens[9] de chacun des sous-corpus que nous avons collectés (cf tableau 3.1).

langue	**nombre de tokens (sans corpus de référence)**
français	1 508 180
hindi	699 405
ourdou	698 016

Tab. 3.1: Nombre de tokens des corpus par langue.

Dans la typologie des corpus de (Sinclair, 1996), le corpus possède des caracté-

[6]Les seuils sont fixés comme suit : **langues à haute densité** : plus de 100 millions de locuteurs ; **langues à densité moyenne** : entre 500 000 et 100 millions de locuteurs ;**langues à faible densité** : mois de 500 000 locuteurs.

[7]L'indice est pertinent car Wikipédia n'est censuré dans aucun de ces trois pays et qu'il y demeure l'encyclopédie la plus populaire.

[8]source : `https://stats.wikimedia.org/FR/Sitemap.htm`.

[9]Un token est la plus petite unité signifiante dans la segmentation du corpus.

ristiques par défaut, qui sont susceptibles de varier. Ainsi, un corpus doit être de taille importante et regrouper des productions langagières authentiques. Notre corpus ne satisfait pas au premier critère, puisque ses sous-corpus totalisent moins de 3 millions de tokens. Il répond en revanche au second, les textes collectés étant des productions tirées de sites non traduits, ou des documents sources destinés à être traduits.

(Auroux, 1998[10]) décrit le corpus comme un « observatoire » qui permet de lier les données empiriques aux données théoriques. En somme, le corpus est avant-tout représentatif de lui-même et les conclusions que l'on en tire doivent lui être limitées.

3.1.1.3 Corpus parallèle, corpus comparable

Un corpus parallèle est un ensemble constitué de textes originaux et de leurs traductions dans une autre langue, ou uniquement de versions traduites d'un texte original non représenté. Il existe entre les textes sources et les textes cibles une relation d'équivalence traductionnelle au niveau des chapitres, des paragraphes, voire des phrases (Kraif, 2014). L'inconvénient d'un corpus parallèle est, comme dans les travaux de traduction en général, la tendance à l'appauvrissement des tournures idiomatiques, du lexique, ainsi qu'à l'artificialité dans les versions cibles.

Le corpus comparable, quant à lui, désigne un ensemble de textes dont les deux sous-ensembles ne sont pas des traductions l'un de l'autre. Ils sont en revanche apparentés au niveau sémantique (ils parlent de thèmes identiques). L'avantage du corpus comparable est que ses sous-ensembles sont écrits chacun dans un idiome sans rechercher d'équivalent de traduction. L'appauvrissement lexical ou grammatical constaté lors d'une traduction a par conséquent moins de risques de se produire. En revanche, l'alignement des textes ne peut être complet puisqu'il ne s'agit pas d'équivalents traductionnels des textes, ni même des paragraphes ou des phrases. Ce type de corpus permet de repérer des tournures idiomatiques.

Transversalement aux sous-corpus de langue, notre corpus se subdivise en sous-corpus parallèles ou comparables, chacun se caractérisant par un genre textuel (normatif, juridictionnel ou non spécialiste du judiciaire). Les textes des sous-corpus parallèles constituent des bi- ou tritextes[11] qui sont des traductions d'un même original. C'est notamment le cas du code pénal indien (IPC) avec ses versions en anglais, langue pivot, en hindi et en ourdou, du code de procédure pénale (CrPC) ou de la Loi sur les preuves

[10]Sylvain Auroux, La raison, le langage et les normes, Paris : PUF, 1998 (pp.166-169 notamment).

[11]Textes alignés bilingues ou trilingues.

(Evidence Act, 1872).

Nos sous-corpus comparables sont constitués des codes pénaux et de procédure pénale de l'Inde et de la France, de décisions juridictionnelles, d'articles de presse judiciaire et de vulgarisation juridique.

C'est dans les limites de ce cadre méthodologique que, prenant comme point de départ la langue observée en situation réelle d'énonciation, nous chercherons à constituer des ressources linguistiques dans le domaine de spécialité du judiciaire. D'une part, on extraira des candidats-termess à partir de l'alignement thématique de textes formant des sous-corpus comparables, selon la méthodologie exposée en 4.4.1.1. Ces termes formeront la base d'un lexique de spécialité dans chacune des langues étudiées. À partir de ces candidats-termes, des analyses textométriques permettront d'identifier des cooccurrents et des , grâce à des calculs de fréquence et de distance d'association. En observant comment ces termes, leurs cooccurrents et les unités phraséologiques articulent les concepts dans les différents systèmes judiciaires, on fera émerger des équivalents de traduction entre les langues et les systèmes, formels ou fonctionnels.

3.1.2 *Objectifs de l'analyse de corpus*

La constitution du corpus de cette étude sert trois objectifs : la validation et la formulation d'hypothèses concernant les emplois terminologiques et phraséologiques (3.1.2.1), la production de ressources lexicales et l'alignement de textes multilingues (3.1.2.2), l'observation des usages théoriques confrontés aux usages réels (3.1.2.3. La finalité de l'ensemble est la production de ressources pour aider aux choix de traduction.

3.1.2.1 Valider et formuler des hypothèses à partir du corpus

De l'époque de sa naissance, chez les taxinomistes aux XVII^{e} et XVIII^{e} siècle, jusqu'aux années 1990, la terminologie est normalisatrice : on interroge des experts qui dressent des listes des concepts et des termes du domaine de spécialité. Les désignations sont imposées verticalement aux domaines techniques.

Le début des années 1990 est marqué par le recours au traitement informatique des données textuelles et les informaticiens commencent à travailler en collaboration avec les linguistes. On définit des réseaux conceptuels auxquels sont associés des textes et des termes justifiant l'organisation du réseau (Anderson *et al.*, 1992), c'est la naissance des *bases de connaissance terminologiques* (BCT) qui rapproche la linguistique de la terminologie.

Dans les années 2000, une approche inverse de la démarche normalisatrice voit le

jour : on décide de partir du corpus pour remonter vers la terminologie, afin de faire ressortir les points de vue et les usages réels, tels qu'ils sont observés en contexte dans le domaine de spécialité (démarche guidée par les corpus ou *data driven*). On cherche désormais à construire une représentation à partir de textes plutôt que d'imposer des normes théoriques. A cette fin, on doit dégager les éléments textuels pertinents et les organiser dans un réseau de relations (Pédauque, 2006).

Les deux approches ne sont cependant pas exclusives. Dans notre corpus, le petit nombre de tokens, résultant des difficultés de la collecte, ne nous permet pas de nous reposer uniquement sur l'une ou l'autre. Nous aurons tantôt besoin de valider des hypothèses terminologiques par l'observation du corpus, tantôt de tirer des règles de formation des collocations ou des équivalents de traduction à partir de celui-ci.

3.1.2.2 Extraction lexicale et alignement multilingue des corpus

Comme expliqué en 3.1.1.2, deux typologies transversales s'appliquent à notre corpus. Une première est basée sur la langue, avec des sous-corpus pour le français, le hindi et l'ourdou, tandis que la seconde regroupe des genres textuels, indépendamment des langues. La première typologie fournit, dans ce que nous appellerons le corpus « langues », les caractéristiques pour chaque langue des différents sous-genres du judiciaire. La seconde typologie, fondée sur le découpage en genres, nous donne un corpus de « genres », qui permet la comparaison entre les langues des traits lexicaux, phraséologiques et énonciatifs des sous-genres. Ce corpus est constitué de textes alignés parallèles et d'autres comparables.

L'ensemble des informations obtenues par l'analyse de ces deux typologies fournit d'une part des termes et collocations qui alimenteront des lexiques multilingues avec des contextes, et d'autre part favorise la recherche d'équivalents de traduction par l'alignement des multitextes.

Nous verrons dans la suite de ce chapitre comment le lexique a été extrait, et comment la taille du corpus a limité la qualité de l'extraction automatique.

Pour compléter l'extraction des termes, nous avons également eu recours à l'identification des thèmes de la procédure pénale décrite au chapitre 2 (2.2), de ses réseaux notionnels, et des termes qui en illustrent les notions. Nous avons ensuite recherché les occurrences de ces termes et les réseaux sémantiques correspondants aux notions dans les sous-corpus de langue. Le retour à l'approche guidée par le corpus s'est fait dans un deuxième mouvement, avec l'observation des cooccurrents et des phraséologies autour de ces termes, pour affiner l'alignement des multitextes. Lors de cette étape, les terminologies normatives ont servi d'ancrages lexicaux : les correspondances lexicales extraites servent

de points d'ancrage pour l'alignement des phrases (Kraif, 2006)[12].

3.1.2.3 Observation des usages théoriques et des usages réels

Par ailleurs, nous avons vu au chapitre 1 (1.2), dans la section consacrée aux politiques linguistiques, que l'Inde et le Pakistan avaient élaboré, à l'aube de leur indépendance nouvelle, des terminologies pour de nombreux domaines dont la justice en recourant à des fonds lexicaux dans des langues classiques hors de la compréhension de la population générale. En étudiant les usages dans les corpus de chacune des langues, on pourra tirer des observations sur la pertinence de l'utilisation des termes de la nomenclature judiciaire officielle.

Cette prise en compte du texte dans l'extraction terminologique est au cœur de la terminologie textuelle (Bourigault et Slodzian, 1999). Selon cette théorie, développées par le groupe TIA (*Terminologie et Intelligence Artificielle*), le point de départ de l'étude terminologique est l'analyse des corpus textuels.

> Il n'y a pas UNE terminologie, qui représenterait le savoir sur le domaine, mais autant de terminologies que d'applications dans lesquelles ces terminologies ont été utilisées. Ces terminologies diffèrent quant aux unités retenues et à leur description selon l'application visée. (Bourigault et Slodzian, 1999, p.30)

Nous retiendrons cette approche dans l'extraction des termes car l'hétérogénéité de notre corpus vérifie la théorie : en fonction du sous-genre observé, la terminologie judiciaire varie. Elle s'adapte également à la fonction du texte. S'il n'est pas ici question de développer des applications comme évoqué dans l'article, le besoin d'adaptation au destinataire selon la situation, déjà préconisée dans la démarche de traduction, n'en reste pas moins vrai.

Pour que notre analyse terminologique soit la plus complète possible, il est en effet nécessaire de prendre en compte les différents sous-genres de textes : l'exposé formel de la norme (textes normatifs), dans laquelle on retrouvera le plus de termes issus des terminologies officielles ; l'application de la norme (textes juridictionnels) et enfin l'explication déterminologisée du droit (billets sur les forums, textes de vulgarisation en langue simple, presse judiciaire). Dans ces deux dernières catégories textuelles, l'effort est porté vers la compréhension du message par le plus grand nombre. C'est aussi là une des contraintes de la traduction judiciaire : pour être fidèle en sens, elle doit être compréhensible par le destinataire du texte.

Le corpus, de taille modeste, ne se prête pas à tous les types de traitement

[12] Cf. ci-après 4.2

automatique, en particulier ceux fondés sur des approches statistiques. À partir de son exploitation, il est possible d'aligner des textes multilingues, de constituer des lexiques et des concordanciers qui seront interrogeables par l'utilisateur via une application en ligne.

3.2 Présentation du corpus

Les sources de collecte des textes sont diverses : les deux principales sont internet et les documents à traduire réunis tout au long de notre pratique professionnelle au service de la justice. Les critères de sélection des textes correspondent aux objectifs de traduction demandés au professionnel par la justice, ainsi qu'aux besoins d'acquisition de connaissance du domaine judiciaire par les traducteurs-interprètes. L'objectif du corpus étant de construire des outils d'aide à l'accès au droit et à la traduction, les textes juridictionnels, qui intéressent tant les magistrats que les justiciables, en forment un sous-genre dont nous avons étudié les caractéristiques en 2.3. Comme il est nécessaire au traducteur comme aux autres destinataires d'utiliser un même référent juridique et conceptuel, un autre sous-genre (normatif) est composée de textes de loi. Enfin, le traducteur-interprète devant mettre l'expression du message à la portée d'un destinataire susceptible d'être non initié au droit, il est également utile d'observer les usages d'un sous-genre qui n'appartient pas à la langue de spécialité mais en emprunte les codes : il concerne les textes de la presse judiciaire et de vulgarisation (textes parajuridiques). L'ensemble de ces sous-genres se décline dans les trois langues d'étude (français, hindi et ourdou). Pour l'alignement des textes parallèles ou comparables, nous avons eu recours également à l'anglais comme langue pivot.

3.2.1 *Textes parajuridiques : un sous-genre entre pédagogie et vulgarisation*

Une des dimensions de notre recherche est de comparer les usages normatifs des terminologies officielles avec ceux auxquels les locuteurs initiés ou non initiés ont recours en situation réelle. Le genre juridique n'est pas uniquement formé de langue de spécialité, il comprend également une part de langue commune, non spécialisée, qui parle du judiciaire. Ces productions langagières, à défaut d'être proprement juridiques, renseignent sur l'état des connaissances des non initiés en matière de justice. Elles sont intéressantes à étudier pour pouvoir adapter la traduction au bagage linguistique du destinataire non spécialiste.

Harvey (2013)[13] distingue les textes *juridiques*, qui doivent créer des effets de droit, des textes *parajuridiques* qui expliquent le droit à des non-juristes et relèvent du discours de l'initié vers des non initiés. Selon lui, cette dimension de vulgarisation permet de documenter un point de droit tel que son application ou l'exposé de sa norme. C'est également dans cette perspective informative à destination des non initiés qu'une partie de ce corpus est constituée de textes de « vulgarisation » : extraits de forums, de sites informatifs, et articles de presse judiciaire. Ces textes, moins techniques, contiennent un lexique du genre judiciaire et des tournures propres (« l'assassin présumé »). Ces moyens de vulgarisation, en particulier la presse, influencent la conception que le non initié a de la justice et son discours sur celle-ci. Ils constituent pour lui des moyens d'accès au domaine de spécialité judiciaire.

L'ensemble parajuridique du corpus regroupe des textes de presse judiciaire extraits automatiquement à partir de sites internet ainsi que des billets postés sur des forums juridiques. Les textes de presse judiciaire on été extraits entre octobre 2017 et mai 2018[14]

- pour le français, des rubriques « justice » ou « carnet de justice » des sites du Monde (`www.lemonde.fr`) et du Point (`www.lepoint.fr`),
- pour le hindi, de la rubrique « crime » du quotidien indien Amar Ujala (`http://www.amarujala.com/crime`),
- pour l'ourdou, de la rubrique « crime court » du quotidien pakistanais Nawai Waqt (`http://www.nawaiwaqt.com.pk/crime-court`).

Le volume des textes de presse ourdou est le plus grand (cf. tableau 3.2), avec une longueur moyenne par texte de quelques phrases. Il en va de même pour la presse indienne, au contraire de la presse française où les textes sont beaucoup moins nombreux mais beaucoup plus longs, raison pour laquelle deux sources françaises ont été choisies au lieu d'une seule.

Les autres sous-genres du parajuridique sont représentés de manière inégale par langue. Aucun forum n'a été retenu pour l'ourdou en raison de la complexité d'aspiration des données, et le volume du forum français est près de cinq fois supérieur au volume en hindi. De plus, nous n'avons pas retenu de textes de vulgarisation en dehors de la presse et des forums en français, tandis que cette catégorie est très représentée en hindi.

[13] Article paru dans Meunier *et al.* (2013), pp.31-42.

[14] L'adoption de la loi RGPD (loi sur la protection des données personnelles) en juin 2018 a obligé les sites hors Union européenne à fermer temporairement leur accès pour les utilisateurs situés dans l'Union, mettant fin à l'extraction pour les journaux indiens et pakistanais.

En effet, les réponses des forums français étaient très proches dans leur style de textes de vulgarisation et renvoyaient souvent vers des sites d'explication du droit aux non initiés. Par rapport aux textes de vulgarisation, les forums français avaient l'avantage de présenter la formulation d'un discours sur le judiciaire par des non initiés.

1.3.8

catégorie de textes	**français**	**hindi**	**ourdou**	**total**
parajuridique	308 352	454 239	556 093	1 318 684
presse	241 933	371 313	553 810	1 167 056
forums	66 419	12 024	0	78 443
vulgarisation autre	0	70 902	2 283	73 185
textes juridictionnels	145 168	3 835	8 293	157 296
textes normatifs	746 308	223 292	88 565	1 058 165
total	1 508 180	681 366	655 234	3 398 590
corpus de référence	14 millions	365 millions	+ 5 millions (phrases)	

TAB. 3.2: Composition du corpus total (nombre de tokens). Les entrées par ligne donnent la composition du corpus par « genres », les entrées en colonnes par « langues ».

Dans le cas du champ judiciaire, une dimension supplémentaire à retenir est le caractère hétéroclite des documents à traduire (Monjean-Decaudin, 2012) : certains documents sont de nature juridique ou judiciaire tandis que d'autres intègrent ces deux aspects.

3.2.2 *Texte juridique ou judiciaire ?*

Comme nous l'avons précisé au chapitre 2 (2.3.1), tout type de document et de genre peut participer de la procédure judiciaire, et tous les documents de la procédure n'entraînent pas d'effet juridique. Pour être qualifié de juridique, le document doit comporter des termes de la langue de spécialité du droit.

En excluant les textes parajuridiques déjà décrits, qui ne sont pas performatifs et n'entrent pas dans la procédure judiciaire, ce corpus réunit deux autres sous-genres de textes. D'une part le sous-genre des textes normatifs, textes juridiques prescrivant la norme ou textes de loi[15], et d'autre part le sous-genre juridictionnel, qui comprend des documents de nature juridico-judiciaire appartenant à la procédure judiciaire tout en

[15]Pour les juristes, le terme « texte » ne désigne d'ailleurs que les textes de cette catégorie.

ayant des effets de droit (citations, ordonnances, jugements, textes du droit pénal…) et des textes judiciaires n'entraînant pas directement d'effet de droit (procès verbaux d'audition ou d'interrogatoire) :

- textes **normatifs** : textes de loi ;
- textes **juridictionnels** : textes de la procédure pénale (cf. chapitre 2, 2.3.1).

Les textes normatifs, ou textes de loi, ont été extraits de sites internet. `www.legifrance.fr` et `www.justice.gouv.fr` constituent les principales sources pour le sous-corpus français (code pénal et code de procédure pénale) ainsi que pour la version en anglais du code pénal français. Les textes juridiques en hindi proviennent du site gouvernemental du département législatif `http://legislative.gov.in` et de liens accessibles depuis celui-ci, tout comme leurs version en anglais. Le CrPC en version numérique provient du site de l'OCDE (`http://www.oecd.org/`). Les textes de loi indiens en ourdou proviennent également du site gouvernemental indien déjà cité. Le code pénal pakistanais en ourdou est extrait du site `http://www.pakistani.org`.

Pour les forums juridiques (sous-genre parajuridique), les sources en français sont les sites `https://forum-juridique.net-iris.fr` et `https://www.legavox.fr`. Net-iris propose des informations juridiques sur différents thèmes et y associe des forums. Legavox présente des thèmes plus affinés. Les membres qui répondent aux billets sont généralement des juristes ou des avocats. Les forums visités pour le hindi sont `https://teesarakhamba.com/` ou `https://indorepolice.blogspot.com/`. Teesara khambha (le « troisième pilier ») est le plus grand site d'information juridique et judiciaire en hindi. Il publie des articles et les internautes viennent y chercher des conseils de juristes sur le forum. Le blog de la police d'Indore publie les rapports au jour le jour des affaires criminelles résolues. Le style très factuel et administratif des billets se situe entre le rapport de police et l'article de presse judiciaire.

Les textes de vulgarisation en hindi proviennent de `http://adhikarexpress.com` et de `https://urdu.arynews.tv/`, `https://www.urduweb.org` pour l'ourdou. Adhikar Express est un autre site d'information juridique en hindi qui publie de nombreux articles de vulgarisation sur les droits des citoyens. Urduweb est un site généraliste qui traite des questions de société et possède une page juridique.

Les textes juridictionnels proviennent pratiquement tous de notre pratique professionnelle. Ils nous ont été confiés par des juridictions françaises ou des particuliers souhaitant faire traduire en français des actes de procédure initiée en Inde ou au Pakistan. Certains des actes en ourdou proviennent d'un juriste pakistanais qui nous a servi

d'informateur, tandis que d'autres ont été téléchargés du site de la Cour suprême d'Islamabad (http://www.supremecourt.gov.pk). Ces textes étaient disponibles pour certains en version numérique, pour d'autres en version papier et ont nécessité des prétraitements informatiques afin d'être exploitables[16].

3.2.3 *Corpus de référence*

Le corpus de référence du français est tiré d'articles divers du journal l'Est républicain de 2003, téléchargeable depuis le site du CNRTL https://www.cnrtl.fr/corpus/estrepublicain/. La totalité du corpus étant trop importante pour être traitée par le logiciel de textométrie, seuls 14 millions de tokens ont été retenus.

Le corpus de référence monolingue en ourdou, provient du projet LINDAT/CLARIN, site http://hdl.handle.net/11858/00-097C-0000-0023-65A9-5maintenu par l'European Language Resources Association de Reykjavík et date de 2014[17]. Il compte 5 464 575 phrases et provient de données aspirées sur internet en 2012.

Le corpus qui a servi de référence en hindi, est également hébergé sur le site de LINDAT/CLARIN, à l'adresse http://hdl.handle.net/11858/00-097C-0000-0023-6260-A. Il contient essentiellement des articles de presse publiés sur internet entre 2010 et 2013[18]. Dans les deux cas, les corpus ont dû être scindés pour être exploité dans les analyses textométriques. Près d'un million de tokens ont été retenus dans chaque cas.

Le corpus réunit donc des textes purement juridiques et normatifs, des textes juridico-judiciaires (textes juridictionnels), des textes judiciaires et des textes parajuridiques. Chaque langue possède de plus un corpus de référence qui servira à contraster les sous-corpus spécialisés. Un autre critère de typologie de ce corpus repose sur le degré de spécialité des textes.

3.2.4 *Langue de spécialité et langue commune*

Le corpus présenté est un corpus de spécialité (cf. 3.1.1.1) puisque, loin de reprendre les caractéristiques d'ensemble de la langue, il aborde un domaine et un genre spécifique, celui du judiciaire. Cela ne signifie cependant pas qu'il ne contient que des textes en

[16]Le récapitulatif des textes et de leurs sources est disponible en Annexe B, en fin de manuscrit.

[17]Cf. bibliographie, Jawaid *et al.* (2014).

[18]Cf. bibliographie, (Bojar *et al.*, 2014)

langue de spécialité. Nous l'avons vu au chapitre 2 (2.3), un énoncé judiciaire peut comporter des passages plus ou moins techniques. Certains, qui reprennent des énoncés de non initiés, contiennent d'ailleurs très peu de termes. De plus, des lexèmes de la langue commune[19] peuvent tout-à-fait entrer dans la composition d'unités polylexicales renvoyant exclusivement au domaine judiciaire. C'est le cas de « demande de libération », « mise en examen » ou « classement sans suite ».

Par ailleurs, il est possible de dresser une typologie des textes de spécialité en fonction de l'appartenance du lexique et de la phraséologie à la langue commune, à la langue administrative (emploi de tournures passives impersonnelles, « fait à », « aux fins de ») ou exclusivement à la langue de spécialité (« depuis temps non prescrit », « attendu que », « par ces motifs ») (Pecman, 2005) qui se retrouvent dans les énoncés du genre judiciaire. L'étude des caractéristiques d'un genre textuel et, dans le cas présent, des sous-genres qui le composent, ne peut se limiter à l'étude de la terminologie de spécialité. En prenant en compte les variations des sous-genres, on doit accepter l'idée d'une variété de terminologies adaptées en fonction des besoins. La terminologie n'est ainsi plus normative et fixe comme dans la tradition wüsterienne (Bourigault et Slodzian, 1999, p.31), et s'ouvre aux autres catégories grammaticales que les substantifs : elles intègrent les adjectifs, verbes, conjonctions qui appartiennent à la phraséologie et ne sont pas déconnectées de la situation énonciative.

3.3 Traitement informatique d'un corpus multilingue

Les outils informatiques permettent de constituer le corpus puis de l'exploiter. Si la constitution et le traitement outillé du corpus français sont des questions largement traitées (Habert *et al.*, 1997; Née, 2017; Poudat et Landragin, 2017), cette problématique reste en revanche bien moins connue pour le hindi et l'ourdou.

Notre corpus comporte en tout trois langues, dont deux à la structure grammaticale très proche, trois graphies et deux directions de lecture. Cette diversité impose de concevoir

[19]Nous avons bien conscience des débats autour des dénominations *langue générale*, *langue commune* ou *langue courante*. Il ne s'agit pas, dans cette étude, de porter un regard normalisateur sur les pratiques langagières de l'ensemble des locuteurs de l'une des trois langues et il ne faut pas entendre ces expressions comme des références à une langue standard, d'autant que nous avons précisé au chapitre 1 à quel point le hindi-ourdou formait un continuum linguistique variable. Nous utilisons ces expressions au sens de Pecman (2004, 2012) et de Fontenelle (1994) par opposition à langue spécialisée et normalisée incarnée, dans le cas qui nous occupe, par la langue utilisée dans l'administration.

des opérations de traitement du corpus adaptées à ces contraintes.

Après une brève introduction à la chaîne de traitement, nous nous appuierons sur l'état de l'art de la linguistique de corpus en hindi et ourdou afin de prendre connaissance des outils à disposition pour aider à la constitution des corpus et à leur analyse. Nous expliquerons ensuite comment nous avons mis en place une chaîne de traitement en utilisant les ressources disponibles et, le cas échéant, comment nous avons surmonté les difficultés de collecte et d'exploitation des données.

3.3.1 *Le projet de la chaîne de traitement*

En linguistique de corpus, une chaîne de traitement permet d'exploiter les données du corpus. Ces données se présentent dans le cas de cette étude sous différents formats : des données « papier » manuscrites ou typographiées, des documents images numérisés, des pages web, ou encore des documents tapés à l'ordinateur puis transformés en PDF. Tous ces documents sont, par ailleurs, déclinés en trois langues et trois graphies différentes.

La première étape de la chaîne de traitement consiste à rassembler et normaliser toutes ces données dans un format exploitable par l'informatique. Une série d'opérations vise ensuite à nettoyer le corpus, puis le scinder en unités signifiantes (segments puis tokens) que l'on pourra annoter (phase d'étiquetage) et analyser statistiquement (phase d'analyse textométrique). On pourra également aligner différents sous-corpus pour en tirer des correspondances traductionnelles. Ces différents résultats font l'objet d'applications en ligne.

La chaîne de traitement, avec ses opérations traitées indépendamment les unes des autres, est schématisée à la figure 3.1 et décrite ci-après. Les langues y sont traitées séparément.

Dans un souci d'économie de moyens, nous avons recherché les outils qui offraient le maximum de compatibilité entre les langues, quitte à délaisser des outils légèrement plus performants en terme de qualité.

Le corpus est d'une part d'un très petit volume en comparaison de ce que peut traiter une chaîne de traitement automatique du langage, d'autre part la chaîne des opérations n'est pas entièrement automatique. Des programmes informatiques indépendants prétraitent et analysent le corpus. L'aspiration des données internet, leur nettoyage et leur annotation ont nécessité différents outils qu'il a fallu adapter aux langues étudiées. Des outils de textométrie pour la recherche statistique des termes et cooccurrents ont ensuite constitué la base de l'analyse des corpus.

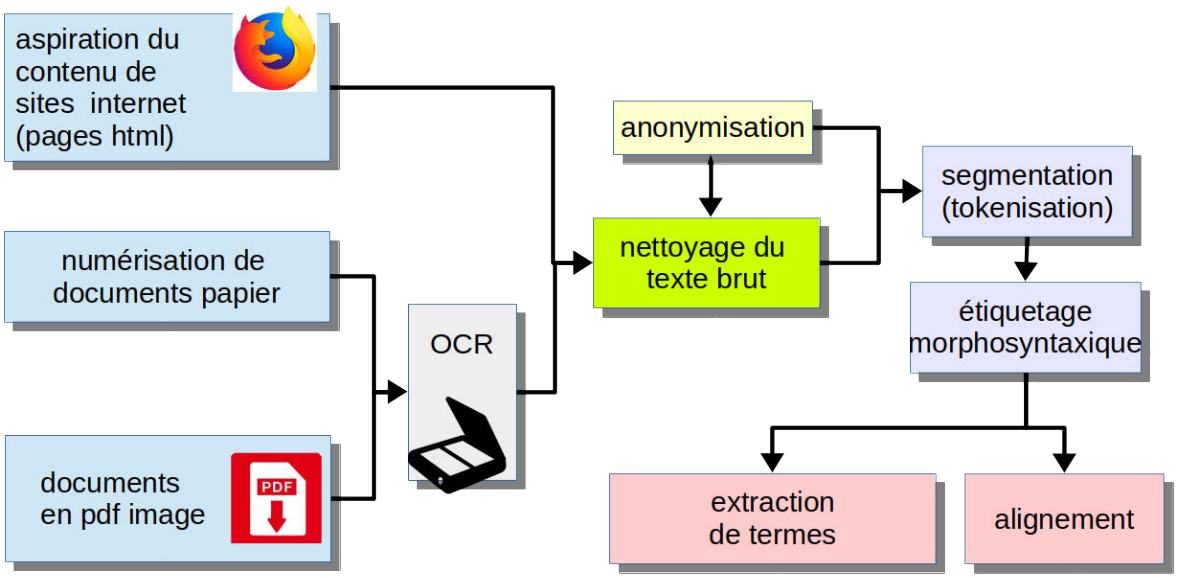

FIG. 3.1: Schéma de la chaîne de traitement du corpus

3.3.2 *Ressources et outils disponibles en hindi et ourdou pour le traitement des corpus*

Dans cette sous-section, nous décrivons les opérations pratiquées sur le corpus, en gardant en référence à chaque étape les ressources et outils disponibles en hindi et en ourdou. La description de la recherche de corpus déjà constitués en hindi et en ourdou, ainsi que d'outils pour les analyser, permet de comprendre par ailleurs les difficultés qui ont entouré la constitution et l'exploitation des sous-corpus en ces langues. Nous n'évoquerons pas ici l'état de l'art en France (pour le français) car la question a été traitée dans de nombreux ouvrages[20]. En revanche, la linguistique de corpus en hindi et en ourdou possède une histoire beaucoup plus récente et moins décrite.

3.3.2.1 Les acteurs de la recherche en linguistique de corpus hindi-ourdou

Les travaux sur le hindi et l'ourdou dans le domaine du traitement informatique des corpus ont été menés à la fois dans les pays d'origine de ces langues, et dans d'autres centres de recherche situés à travers le monde. L'Inde et le Pakistan ont travaillé sur l'ourdou tantôt séparément, tantôt conjointement. L'Inde a mené ses recherches sur le traitement de l'ourdou dans une perspective d'interéchangeabilité avec les autres langues

[20]Voir Née (2017) et Poudat et Landragin (2017), cf. références dans la bibliographie.

du pays (standardisation des outils, des étiquettes et des systèmes de translittération). De son côté, le Pakistan, qui n'est pas confronté au problème des graphies multiples de son voisin indien, a développé des outils à partir de l'ourdou pour les diffuser vers les autres langues.

La linguistique de corpus et les activités de traitement automatique des langues débutent en Inde dans les années 1980[21]. En 1990-91 naît le projet TDIL (Technology Development for Indian Languages) par le Department of Information Technology (DIT) du gouvernement indien. Le Central Institute of Indian Languages (CIIL) (Central Institute of Indian Languages), qui œuvre depuis des décennies à la promotion de l'apprentissage des langues vernaculaires en éditant des matériels pédagogiques non seulement de l'anglais vers les autres langues mais aussi d'une langue vernaculaire vers une autre, collabore sur des projets avec des universités internationales. Dans un effort de rendre accessible au plus grand nombre les savoirs des principaux domaines de connaissance, un autre organe public indien, la National Translation Mission (NTM) est chargée de traduire dans les 22 langues répertoriées à l'annexe 8 de la Constitution (cf. 1.2.2.1.3) les textes du savoir. Elle travaille en collaboration avec les C-DAC Pune, C-DAC Mumbai, IIIT Hyderabad et IIT Kanpur, qui ont développé des outils pour la traduction automatique.

Au Pakistan, l'ingénierie linguistique pour l'ourdou a été développée par le Centre for Research in Urdu Language Processing (CRULP), aujourd'hui hébergé au Centre for Language Engineering (CLE) à Lahore. L'institution, pionnière dans le développement de logiciels et d'outils informatiques en ourdou, est très active en recherche TAL. Son site[22] indique que des workshops sont régulièrement organisés depuis 2001 sous l'égide de l'AFNLP (Asian Federation of Natural Language Processing)[23]. Le Pakistan organise d'ailleurs depuis 2007 une « Conference on Language and Technology » (CLT) tous les deux ans, à laquelle participent des universités du pays et le CLE.

Des projets internationaux de développement et de coopération entre les pays d'Asie du Sud se mettent en place dans les années 2000 (Hock et Bashir (2016)). Le projet (PAN), qui concerne la localisation dans les pays asiatiques, a montré des signes d'activité jusqu'en 2012[24]. Il a permis de développer l'utilisation des langues régionales à la place de l'anglais pour développer la navigation sur internet dans les zones rurales (projet Dareecha).

[21] L'ouvrage de Hock et Bashir (2016) a guidé l'organisation de cette description.

[22] `cle.org.pk/alr10/`.

[23] Cf. leur site : `wwwafnlp.org`.

[24] Aucun billet posté sur le site `www.panl.10n.net` n'est ultérieur à 2012.

3.3.2.2 Les corpus disponibles

Les premiers corpus informatiques en hindi et en ourdou ont été constitués à partir des années 1990, tant en Inde et au Pakistan qu'en dehors de ces pays.

Hors Asie du Sud, Becker et Riaz (2002) constituent un corpus extrait du site de la BBC en texte brut, à étiqueter manuellement. Les universités de Lancaster et Sheffield, en collaboration avec le CIIL (Central Institute of Indian Languages)de Mysore en Inde, mettent au point le corpus 2000[25]. Il s'agit d'un corpus monolingue pour 14 langues d'Asie du Sud, et d'un corpus parallèle de l'anglais vers cinq autres des ces langues[26] Des dépôts de corpus électroniques en langues indiennes sont créées dans des universités européennes comme à l'Institut de Göttingen[27].

En Inde, le projet Technology Development for Indian Languages (TDIL) (Technology Development for Indian Languages), déjà évoqué précédemment, se place sous l'égide du Department of Information Technology (DIT) du gouvernement indien afin de développer des corpus numériques dans toutes les langues majeures de l'Inde. Il entend viser tous types de public. Le premier corpus collecté à cette époque est en anglais indien (Kolhapur Corpus of Indian English (KCIE, 1991). Il s'agit alors de comparer cette variante de l'anglais avec l'anglais britannique et américain. Le corpus, encodé dans la norme ISCII (un encodage développé pour les langues indiennes en référence à l'ASCII américain, cf. 3.3.3.1.1), est peu utilisé car la conversion d'ISCII vers le standard maintenant largement répandu Unicode n'est pas aisée (cf. 3.3.3.1.2).

Les travaux indiens s'accélèrent en 2012. Un nouveau projet du gouvernement, baptisé Indian Languages Corpora Initiative (ICLI), entend développer un corpus parallèle en 12 langues indiennes dans les domaines du tourisme et de la santé[28]

Au Pakistan, un premier corpus numérique ourdou en Unicode (cf. infra, 3.3.3.1.2) est constitué par Becker et Riaz en 2002[29]. Formé de 7000 articles de presse, il est formaté

---

[25]Site internet du projet : `http://www.emille.lancs.ac.uk/about.php`.

[26]Lien vers le corpus EMILLE :`http://www.lancs.ac.uk/fass/projects/corpus/emille/`.

[27]Voir le site`http://gretil.sub.uni-goettingen.de/`. On y trouve de très nombreux textes en sanskrit et en tamoul, moins en hindi.

[28]Le corpus est disponible sur : `http://www.tdil.mit.gov.in/Default.aspx` (autorisation à demander au TDIL`http://tdil-dc.in/index.php?option=com_vertical&parentid=58&lang=en`).

[29]Becker et Riaz (2002), *op. cit.*

avec des balises XML[30], et enrichi d'étiquettes et de métadonnées en anglais. Dans le corpus EMILLE évoqué plus haut, Baker *et al.* (2003) ont rassemblé 1 640 000 tokens en ourdou écrit. Cette partie du corpus a été étiquetée, annotée et alignée avec un corpus parallèle en hindi.

En 2008, un partenariat entre le CRULP (Pakistan) et IIIT Hyderabad (Inde) rassemble un corpus de 40 000 tokens, annoté avec 12 classes d'entités nommées[31].

En 2009, Ali et Ijaz (*op. cit.*) collectent un corpus comportant 19,3 millions de tokens, en extraient les mots-stop (mots grammaticaux, mots vides, formes verbales très courantes jugées non signifiantes pour l'analyse) (Ali *et al.*, 2017) et le lemmatisent. Le CLE rassemble entre 2011 et 2013 un corpus étiqueté sous licence.

Ces corpus sont constitués généralement d'articles de presse tirés de pages web. Aucun de ces corpus ne concerne le domaine juridique, encore moins le judiciaire.

3.3.3 *Les outils de traitement à disposition et leur insertion dans la chaîne de traitement*

En tout premier lieu, la difficulté a été de normaliser toutes les données numérisées, dans toutes les langues, dans un format appelé Unicode, qui applique l'encodage standard UTF-8. Cette normalisation permet par la suite une manipulation universelle des données.

3.3.3.1 D'un encodage hétérogène des documents à leur normalisation en Unicode

Certains textes numériques disponibles dans un format PDF, notamment pour les sources les plus anciennes en ourdou et en hindi (textes de loi, jugements), se sont avérés inexploitables directement. En effet, ces textes avaient été saisis au moyen de traitements non standard spécifiques à ces langues (par exemple le logiciel InPage[32]), utilisent une représentation interne des caractères et de leurs innombrables ligatures non-standard et opaque, c'est-à-dire une représentation dans un format propriétaire (Baker *et al.*, 2003, 2004).

N'ayant aucun moyen de connaître le codage des ligatures, il nous a été impossible

[30] Langage de mise en forme de contenus, cf. infra 3.3.3.2.

[31] Ce corpus est disponible à l'adresse suivante : `http://ltrc.iiit.ac.in/ner-ssea-08/index.cgi?topic=5`.

[32] InPage (https ://filehippodownload.net/download-inpage-2019.html) est un logiciel de PAO spécifique à l'ourdou et à sa variante d'écriture nastaliq. Toujours utilisé dans le monde de l'édition en ourdou, son intérêt est l'esthétisme de son rendu des caractères nastaliq. Il impose cependant des formats non standardisés et non éditables, l'exportation ne pouvant alors se faire qu'en format image.

malgré nos efforts de casser ce code et réaliser une application ad-hoc de conversion de format PDF en texte[33]. Bien sûr, avec de tels documents, le copier-coller est inopérant.

L'exploitation de ces documents pour notre corpus a nécessité le recours à l'OCR (cf. 3.3.3.3). Cette opération s'est avérée extrêmement fastidieuse et coûteuse en temps :

– conversion des documents PDF en image JPEG,
– reconnaissance optique des caractères de l'image,
– correction des innombrables fautes de reconnaissance (jusqu'à 10%)

3.3.3.1.1 Le problème de la représentation numérique des caractères

Les textes sont constitués de caractères, qu'ils soient syllabiques, alphabétiques ou idéographiques. Certains sont sont visibles ou imprimables, d'autres non (par exemple espace, saut de ligne). Ceux-ci sont représentés en machine par des valeurs numériques parmi celles définies dans la page de code, le jeu des caractères des écriture(s) disponibles.

L'affichage ou l'impression se fait au travers d'une police de caractères qui répertorie et décrit les glyphes, c'est-à-dire les diverses formes graphiques que peuvent prendre chaque caractère (normal, gras, italique) et les trace. Si la police ne contient pas le glyphe d'un caractère, elle affichera un caractère de substitution (en général un carré ou un point d'interrogation).

Lors du développement de l'informatique, la problématique de l'écriture numérique a longtemps été délaissée. En effet les fabricants d'ordinateurs ont d'abord adopté des codages de caractères qui leur étaient propres. Ils s'en est suivi une multitude de codages généralement incompatibles, rendant l'échange de documents difficiles. Un premier consensus est venu dans les années 1960 avec un jeu de caractères de base, l'ASCII (American Standard Coded Information Interchange) définissant 128 caractères : les 26 lettres latines majuscules et minuscules, les 10 chiffres et diverses ponctuations notamment. Ni les accents, ni les diacritiques n'étaient pris en compte, sans parler des autres écritures.

La fin des années 1980 voit les premières normalisations régionales avec l'apparition de jeux de caractères mono, bi voire multilingues[34]. Le codage des caractères devient un enjeu commercial et politique. Sont ainsi apparus progressivement :

– des jeux de caractères « bi-écritures », augmentant l'ASCII de 128 caractères :

[33]Il existe bien des applications de conversion de PDF en texte (PDF2txt) mais celles-ci ne marchent pas avec des codages « propriétaires » tels celui d'InPage.

[34]La Chine, Taïwan, le Japon, la Corée, du fait de la complexité de leurs écritures ont très tôt adopté des solutions plus multilingues.

- la série MS Windows 1250 à 1258 pour les variantes régionales de l'écriture latine, le grec, le cyrillique, l'arabe, l'hébreu, le thaï ;
- la série ISO-8859-1 à 16 à quelques détails près compatible avec la série MS Windows.

– des jeux « multi-écritures »
  - SHIFT-JS au Japon, BIG 5 à Taïwan, GB en Chine continentale ;
  - ISCII[35] pour les langues et écritures de la péninsule indienne.

Avec le développement croissant d'internet et la globalisation, cette profusion de codes nationaux et régionaux s'est révélé un frein aux échanges. Un nouveau standard a donc émergé : Unicode.

3.3.3.1.2 Qu'est-ce qu'Unicode ?

Unicode est un standard universel de représentation des caractères des écritures du monde qui s'est progressivement étoffé et implanté dans le milieu informatique depuis le début des années 1990[36]. Il est le fruit d'une réflexion collective des fabricants et organismes de normalisation désireux, à l'aube de la globalisation, de s'affranchir de la multitude de jeux de caractères tant nationaux que propriétaires, pour aboutir à une norme unique permettant la manipulation et l'échange de données et documents réellement multilingues.

Nous nous en tiendrons aux principales caractéristiques d'Unicode.

– C'est un répertoire des caractères du monde (il y en a aujourd'hui 137 929) regroupés par systèmes d'écritures.
– Il répertorie des caractères « abstraits » et non des glyphes, c'est-à-dire leur forme simple ou combinée en ligatures (c'est le rôle de la police de caractères d'afficher le bon glyphe).
– Il impose un codage biunivoque, non ambigu : à chaque caractère, lui est affecté un nombre unique (le point de code) et vice-versa.
– Il définit des principes de composition des caractères accentués ou dotés de diacritiques.
– Il impose que l'ordre d'énonciation, et non l'ordre graphique, soit celui de l'enregis-

[35] En 1991, le Bureau of Indian Standards a tenté d'imposer la norme ISCII, sur le modèle d'ASCII. Très peu utilisé par les développeurs indiens, cette norme est peu présente sur internet.

[36] La version actuelle est la 12.1 : `https://www.unicode.org/charts/` permet de visualiser la richesse des écritures et caractères supportés.

trement des caractères dans un document[37].

- Les caractères sont catégorisés en : lettres, nombres, idéogrammes, « diacritiques », ponctuations, espace, symboles, …).
- Il propose une description formelle des caractères (nom officiel, catégorie, type et sens d'écriture, équivalences, exemple de glyphe …).
- Il définit divers modes possibles de représentation interne (physique) des points de code en machine ou au sein des fichiers : c'est l'encodage (encoding).

Dans les codages non Unicode, une même valeur de code pouvait désigner divers caractères. Par exemple, la position 233 représente « é » en ISO-8859-1 (latin) et « Щ » en ISO-8859-5 (caractères cyrilliques). En Unicode, ces deux caractères auraient des points de code différents, évitant ainsi toute confusion :

- é LATIN SMALL LETTER E WITH ACUTE U+00E91 (=233)
- Щ CYRILLIC CAPITAL LETTER SHCHA U+0429 (=1065)

3.3.3.1.3 Et UTF-8 ?

La norme Unicode spécifie que tel caractère correspond à tel point de code, une valeur comprise entre 0 et 1 114 112. Il ne présage en rien de sa représentation physique sous forme binaire dans les fichiers ou en mémoire. Cependant, il préconise divers modes d'encodage binaire de ses caractères, plus ou moins gourmands en octets[38] et plus ou moins pratique ou transparents :

- Un encodage de longueur fixe sur 2 ou 4 octets, ayant l'avantage d'être 'lisible', mais consommant plus d'octets (UTF-16 ou UTF-32).
- Un encodage de longueur variable UTF-8, sur 1 à 4 octets. Son avantage est d'être naturellement insensible à l'architecture des machines et d'être souvent plus compact. Il est par contre opaque, au sens que le point de code qu'il représente n'est pas aisément lisible. Ce format d'encodage se généralise notamment sur internet. Si en 2010, UTF-8 encodait déjà plus de 50% des pages web, certaines études estiment que c'est aujourd'hui plus de 90%[39]. C'est dans ce format UTF-8 que nous avons donc normalisé tous les textes de notre corpus.

---

[37]Certaines voyelles, dans les langues indiennes, s'écrivent avant la consonne qui les soutient mais se prononcent après. Il en va de même de certaines consonnes dans des ligatures.

[38]L'octet, constitué de 8 bits, est l'unité fonctionnelle d'organisation de la mémoire des ordinateurs. Un octet peut encoder 256 valeurs différentes, deux octets 65 536 et quatre octets 4 294 967 296.

[39]Cf. `https://fr.wikipedia.org/wiki/UTF-8`.

Le hindi, et plus encore l'ourdou, sont passés tardivement à l'UTF-8 et aux normes Unicode. Dans les années 2000, presque chaque site internet imposait de télécharger une ou plusieurs polices propriétaires pour être lu.

Plusieurs raisons expliquent la multiplicité des systèmes d'encodage. Aux débuts de la linguistique de corpus, peu de traitements de texte acceptaient la norme Unicode. Même lorsque les langues indiennes étaient encodées sous cette norme, peu de systèmes d'exploitation et de navigateurs internet étaient alors capables de les prendre en charge.

Malgré l'existence aujourd'hui de logiciels ou outils en ligne de conversion de documents vers UTF-8, certains documents, en hindi notamment, utilisant des polices propriétaires. Ces polices ont nécessité après leur conversion de grands efforts (rédaction de scripts ad-hoc et corrections manuelles), tant pour le transcodage des caractères que pour la prise en compte des ligatures (combinaisons de consonnes dont la forme peut changer) et de la voyelle « *i* », placée avant le groupe consonantique qu'elle suit dans l'ordre d'énonciation (ex : *ki* s'écrit i+k : इ + क = कि)[40], contrevenant à l'ordre d'énonciation requis par Unicode.

3.3.3.2 Conversion des fichiers en formats interopérables

Pour faciliter l'exploitation, la maniabilité voire l'échange des textes, nous avons choisi d'encoder leurs caractères en UTF-8. Pour de semblables raisons, nous privilégierons comme formats du corpus les formats qui précisent la mise en page et la mise en forme du contenu. En traitement de corpus, les formats privilégiés sont le texte brut et XML (Extensible Markup Language, « langage de balisage extensible »). Nous réduirons donc tous nos textes collectés dans le corpus à l'un ou l'autre de ces formats.

Ces deux formats se suffisent à eux-mêmes, il ne contiennent aucune indication de formatage (police, mise en forme…). Pour pouvoir les lire et les manipuler, il suffit d'en connaître l'encodage (UTF-8). Ces format ont l'avantage d'être manipulables par tous types de systèmes d'exploitation (interéchangeabilité) et de logiciels (interopérabilité[41]).

On dit d'un texte qu'il est brut s'il ne contient que les caractères d'un texte, sans aucune marque explicitement structurante.

Le second format, XML, permet de structurer sémantiquement le texte selon un formalisme « à chevrons » encapsulant des fragments de textes entre des balises ouvrantes

[40]Cette particularité pose toujours problème dans l'importation d'un texte extrait d'un pdf en texte brut.

[41]Source : Wikipédia.

et fermantes. XML est un métalangage informatique de balisage générique qui est un sous-ensemble du Standard Generalized Markup Language (SGML). Sa syntaxe est dite « extensible » car elle permet de définir différents langages avec chacun leur vocabulaire et leur grammaire, comme XHTML, XSLT, RSS, SVG, TEI. L'objectif initial de XML est de faciliter l'échange automatisé de contenus complexes (arbres, texte enrichi, etc.). Il est interopérable et autodocumenté.

Un utilisateur pouvant définir aisément au gré de ses besoins, ses propres jeux de balises et leur grammaire, le format XML permet d'annoter un texte avec tout un ensemble de métadonnées. On peut ainsi baliser un texte en fonction de sa structure logico-sémantique (titre, contenu, notes, etc.), de la catégorie grammaticale de ses mots, ou encore identifier des noms de personne ou de lieux (entités nommées) dans un fichier à anonymiser.

Un dernier avantage est sa facilité à être transformé à la volée en un document HTML pour une meilleure lisibilité humaine de son contenu, au moyen de feuilles de style XSLT.

3.3.3.3 Reconnaissance optique des caractères

De nombreuses administrations en Inde et au Pakistan ne sont pas numérisées. De fait, de nombreux documents de procédure en hindi et ourdou sont encore manuscrits[42].

Une partie des textes de ce corpus provient de documents papiers scannés. Il s'agit de textes de la procédure française issus notre pratique professionnelle, et de très nombreux documents numérisés en hindi et en ourdou mis en ligne. La numérisation des documents se fait dans un format image (pdf, jpg ou autres) non éditable. Ces images doivent d'abord être transformées en texte pour une exploitation informatique.

La reconnaissance optique des caractères (*Optical Character Recognition* ou Reader) est le processus ou l'outil qui permet le passage d'un format image à un format texte numérique. L'OCR tente de reconnaître des caractères présents dans une image numérisée. Il convertir ainsi un contenu graphique fixe donné en entrée en contenu éditable à la sortie.

Le programme de reconnaissance optique segmente une partie de l'image et isole des ensembles de pixels qu'il associe à un caractère en fonction de ses caractéristiques typographiques, ainsi que des règles dictionnairiques d'une langue donnée. Pour cette raison,

[42] En ourdou notamment, un agent de police est spécialement chargé de recueillir les dépositions et de les transcrire.

il n'existe pas de programme OCR possédant des règles de reconnaissance universelles. De même, un programme d'OCR reconnaissant une langue dans un type d'alphabet donnera des résultats très décevants pour une autre langue utilisant le même alphabet. Il est donc primordial de trouver un programme OCR spécifique à la langue étudiée afin de transformer les fichiers images en données textuelles exploitables.

3.3.3.3.1 OCR pour l'ourdou

L'ourdou est la langue qui a posé le plus de problèmes lors de ce processus. D'une part en raison de la rareté des programmes OCR disponibles pour cette langue[43], et d'autre part à cause de la complexité de la graphie de l'ourdou qui rend difficile le processus de reconnaissance.

Un OCR adapté pour des langues comme l'arabe ou le persan donnera de très mauvais résultats en ourdou en raison de la différence des règles lexicales de ces langues et des diacritiques propres à l'ourdou. D'autre part, comme nous l'avons évoqué au chapitre 1 (cf. 1.2.2.3), l'ourdou demeure l'une des rares langues à utiliser la variante nasatliq de l'alphabet arabe, dont la particularité est que les glyphes qui composent les mots s'écrivent liés les uns aux autres sur une ligne diagonale descendante[44]. En fonction des glyphes environnants qui composent la ligature, un glyphe donné possède jusqu'à une quinzaine de formes possibles (Akram et Hussain, 2019) (cf. figure 3.2). Il se retrouve parfois superposé à un autre glyphe au sein d'une même ligature (fig. 3.3, a) ou appartenant à une autre ligature (b). De plus, les diacritiques du glyphe peuvent figurer au même niveau que le glyphe suivant dans une tranche analysée par l'outil d'OCR, provoquant des erreurs d'interprétation et de sortie du texte. Une autre source de difficulté est l'absence d'espace séparateur entre les mots. Cette absence ne pose pas de problème dans un contexte de lecture « manuelle » car dans l'alphabet ourdou les lettres finales des mots possèdent une forme qui les identifie comme telles. L'œil humain parvient sans problème à discerner les limites des mots, ce qui n'est pas toujours le cas de la machine.

[43]La limitation du libre accès aux outils en ourdou est un problème récurrent : l'OCR utilisé a nécessité l'achat d'une licence, et la diffusions des corpus annotés par le principal centre de recherche pakistanais, le CLE, est payante, raison pour laquelle il nous a fallu trouvé des méthodes alternatives pour constituer notre corpus de référence et entraîner des étiqueteurs morphosyntaxique, tâches chronophages.

[44]La réalité est un peu plus complexe puisque certaines lettres se trouveront toujours écrites sur une ligne de base tandis que celles qui l'entourent seront réparties sur cette ligne diagonale descendante

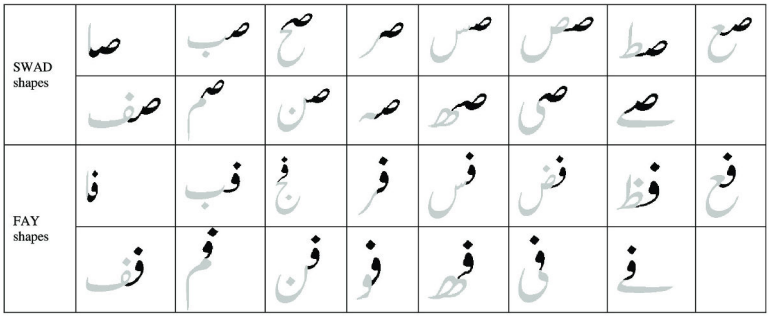

FIG. 3.2: Variantes pour les glyphes ص et ف, repris de (Sarmad 2019, p.3)

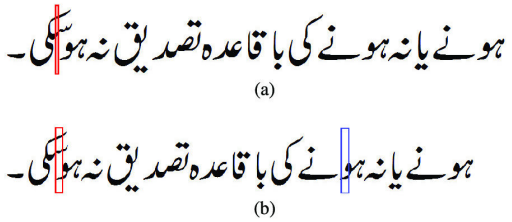

FIG. 3.3: Superposition de glyphes en ourdou, repris de (Sarmad 2019, p.4)

L'outil OCR développé par le Center for Research in Urdu Language Processing (CRULP) à Lahore, est limité à la reconnaissance des textes en nastaliq, typographiés en caractères de taille 14 à 44. Les textes en naskh ne sont pas pris en compte, or il arrive souvent qu'un texte judiciaire comporte les deux variantes pour des effets de mise en relief (pour des citations, cf. fig.(3.4)). Il se fonde sur des méthodes de reconnaissance des segments et des ligatures (Javed et Hussain, 2013).

Le site indien du projet TDIL, qui répertorie les outils de TAL pour les langues indiennes[45], propose aussi un outil OCR pour l'ourdou dont les résultats sont comparables. La sortie texte comporte dans les deux cas de nombreuses erreurs qu'il faut corriger manuellement (environ 10%, taux d'erreur très élevé).

Pour les textes manuscrits (des pièces de procédure, en particulier des rapports de police), l'étape de la saisie informatique est inévitable. Le lecteur se fera une idée de la difficulté de l'interprétation d'un texte manuscrit en se reportant à la figure 3.5.

[45]Cf. sous-partie 3.3.2 sur l'état de l'art.

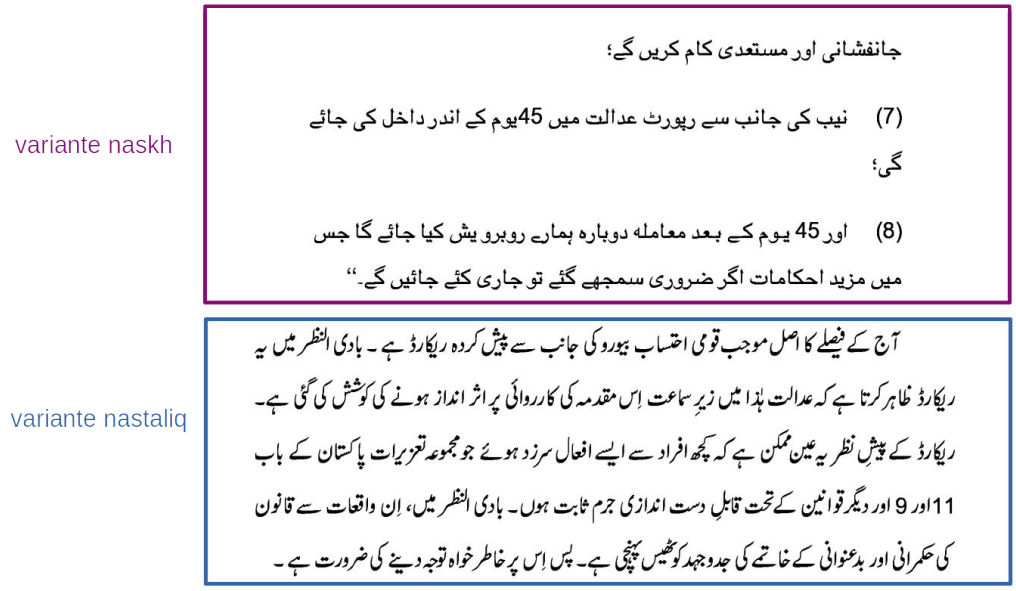

Fig. 3.4: Texte mélangeant les graphies naskh et nastaliq. Les passages en naskh sont des citations de précédentes procédures (extrait d'une décision de la Cour suprême pakistanaise).

3.3.3.3.2 OCR en hindi

La reconnaissance optique du hindi a fait l'objet d'initiatives privées ou publiques. Côté public, le TDIL propose e-Aksharayan, un outil OCR en ligne décliné en 22 langues indiennes. Il comportait cependant des bugs et nous n'avons pas pu l'utiliser.

Un autre OCR, payant, développé à Mumbai en 2011 par Karishma Enterprises (Hindi OCR, Professional Edition), fonctionne plutôt bien[46]. Il est possible de l'entraîner à reconnaître les caractères dans différentes polices : à partir d'un corpus d'entraînement, une fonction permet de corriger manuellement les caractères reconnus parmi un choix de possibilités proposées par le logiciel, ou d'assigner au caractère reconnu la valeur souhaitée du caractère hindi. Il est possible de traiter de très nombreuses pages en une seule fois. Enfin, le site I2OCR propose un outil en ligne[47] fonctionnant pour tout une variété de langues dont le hindi. Les résultats en sont très satisfaisants, même pour des textes multicolonnes.

[46]Le taux de reconnaissance est de 99% pour des documents imprimés de bonne qualité. Il reconnaît aussi des documents manuscrits après apprentissage.

[47]Cf. `http://www.i2ocr.com/free-online-hindi-ocr`.

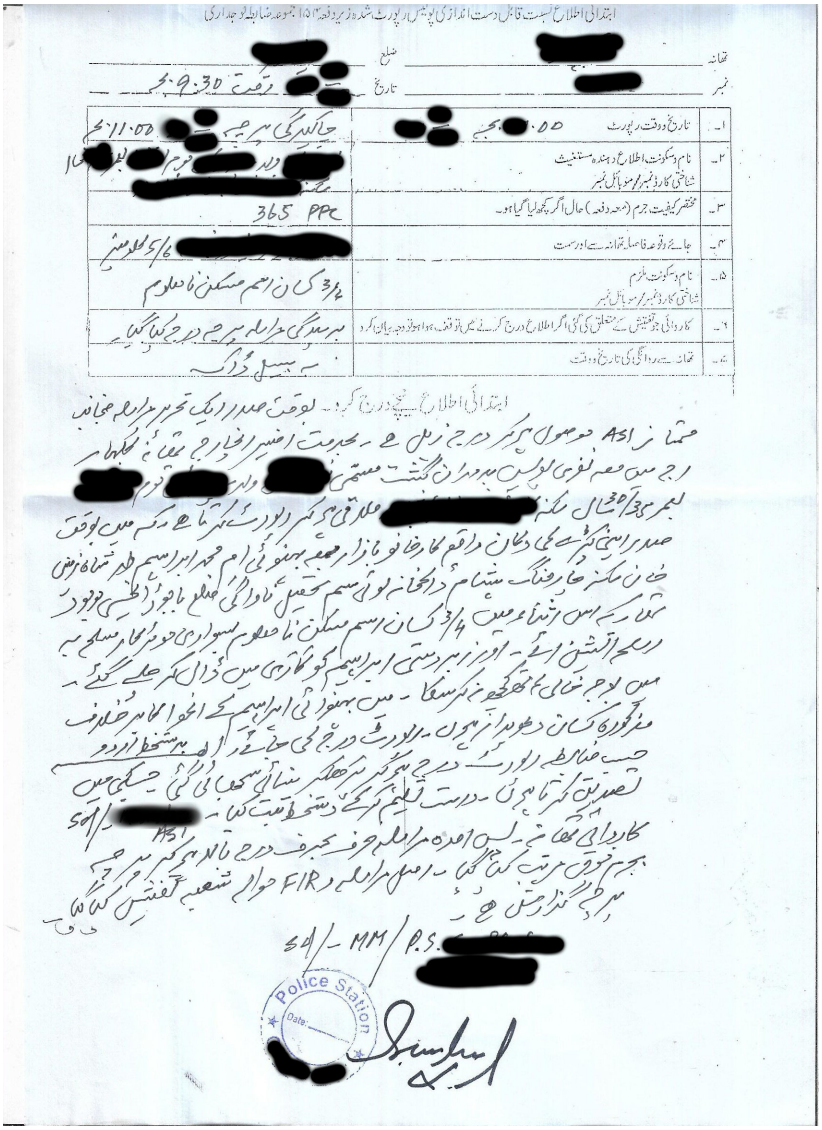

FIG. 3.5: Exemple de FIR manuscrit en ourdou (Pakistan).

3.3.3.4 Aspiration de pages internet

Le web offre une source quasi inépuisable pour la constitution de corpus. On y trouve en effet, et souvent en accès libre, des documents officiels, des forums ou des blogs, des articles de presse, des ressources lexicales en ligne, etc., dans un format standard non propriétaire et pérenne : HTML[48] Toutes ces pages en ligne peuvent aisément être aspirées automatiquement, c'est à dire rapatriées sur une machine, sous la forme de fichier HTML puis traitées par des scripts spécialisés afin d'en extraire les contenus utiles. Ces

[48]HTML (*HyperText Markup Language*) est un langage de balisage conçu pour décrire des contenus (textuels, graphiques, etc.) de pages web ainsi que leur mise en page. Sa dernière version est HTML5. Il permet de structurer logiquement voire sémantiquement le contenu des pages.

contenus seront, selon les cas, récupérés soit comme du texte brut, soit comme du texte structuré en XML.

La limite de l'aspiration reste juridique : tout ce qui est rendu public sur internet n'est pas à libre disposition, on doit s'assurer que les données sont libres de droit ou demander au propriétaire l'autorisation de les exploiter.

Pour notre corpus, un script (voir annexe B.4) a permis de relever quotidiennement le contenu des pages de la presse judiciaire et de le récupérer en texte brut ou en XML. D'autres pages ont été aspirées sans automatisation (forums juridiques et textes de vulgarisation). L'ensemble du contenu de ces fichiers, aspiré avec ou sans script, a été concaténé afin de former des fichiers globaux par sous-genre (presse, forums, vulgarisation).

3.3.3.5 Nettoyage du contenu numérisé

Arrivé à cette étape, le corpus est constitué de textes numérisés issus de documents et de pages internet. Avant tout traitement ultérieur (segmentation, annotation, etc.), le nettoyage est une étape cruciale qui permet d'obtenir un corpus propre, le plus exploitable possible. Il importe de s'assurer de la qualité des textes, c'est-à-dire de veiller à ce qu'il sont bien formés et ne comportent par d'éléments parasites (bruit) – souvent causés par le recours à l'OCR - susceptibles de nuire à son exploitation et de fausser les résultats. Une phase de nettoyage semi-automatique est ainsi nécessaire pour :

- éliminer des éléments parasites (en-têtes ou pieds de pages, liens, menus, légendes, etc.),
- reconstituer la continuité des phrases et paragraphes en supprimant des sauts de ligne,
- normaliser les caractères (par exemple les espaces, les apostrophes, les guillemets) ou convertir des entités HTML[49] en caractères.

Il est à noter que s'il est relativement simple avec des scripts ad-hoc de filtrer automatiquement dans les pages web aspirées les éléments indésirables (menus, liens, etc.), ce filtrage est d'autant plus simple et fiable que ces pages sont récentes et suivent les standards actuels de la programmation HTML (HTML5 et feuilles de styles CSS).

Cette nécessité de nettoyage est aussi valable lors de traitement ultérieur : pour corriger manuellement des annotations erronées attribuées automatiquement, des erreurs de segmentation, ou éliminer des résultats non pertinents.

[49]Une entité HTML est un mode « verbeux » de représentation de caractère, par exemple *é* sera littéralement représenté par *é*, ou numériquement par *&#00E9;*

3.3.3.6 Segmentation, tokenisation et étiquetage

Segmentation et tokenisation découpent le texte en unités qu'il est ensuite possible d'étiqueter (ou d'annoter).

La *segmentation* consiste à délimiter un texte en différentes unités exploitables. On peut segmenter un texte en parties, sous-parties, paragraphes ou phrases, selon le type d'analyse que l'on souhaite pratiquer.

Lorsqu'elle se situe à un niveau inférieur, la segmentation prend le nom de *tokenisation*, c'est-à-dire « segmentation en tokens ». Brique de base de l'exploitation de corpus, le token est une unité signifiante. Pour un corpus écrit, un token représentera une unité de sens (morphème) ou lexicale (lexème ou lexie).

La reconnaissance des fins de phrase et de la ponctuation, à l'aide de règles écrites dans un programme, aide à repérer les limites des tokens ou segments. Cette étape devient complexe dans le cas de la recherche de la limite de mots lorsque le système d'écriture ne prévoit pas d'espace entre deux unités de sens (en japonais, en thaï), ou au contraire lorsque plusieurs morphèmes séparés par des espaces constituent une unité lexicale. On retrouve une difficulté similaire en français avec les mots composés ou des unités polylexicales comme « état civil », « c'est-à-dire » ou « s'il vous plaît ».

La tokenisation en hindi est relativement simple au niveau lexical, les unités étant séparées par des espaces. Les postpositions, affixées ou non, peuvent être inclues dans le token ou former un token indépendant, selon l'analyse envisagée.

(47) मुझसे
mujhse
moi.INST
« à moi, par moi » (1 token)

(48) मुझ से
mujh se
moi INST
« à moi, par moi » (2 tokens, ou bigramme)

En ourdou, la tokenisation est plus difficile en raison de l'absence occasionnelle d'espace entre les mots, de la représentation graphique d'un mot en plusieurs morphèmes séparés par des espaces, et de la variabilité des représentations graphiques d'un même mot. Par exemple, pour le nom de ville suivant :

(49) *Islam Abad* اسلام آباد

« Islamabad » (l'unité lexicale est un bigramme.)

La locution postpositive *ke lie* (« pour ») peut être représentée en deux éléments ou un seul :

(50) *ke lie* کے لئے (bigramme.)

kelie کیلئے (monogramme.)

De même, les morphèmes suffixes de dérivation peuvent être physiquement détachés de l'adjectif de base mais font partie de la même unité lexicale. Ces problèmes peuvent être réglés par un programme de segmentation qui prend ces règles en compte, notamment en répertoriant les suffixes potentiels. Par exemple :

(51) *z̲immah dār* ذمہ دار

responsabilité SUF.ADJ/nom (N)

« responsable »

Au niveau des formes verbales, la différence entre les modèles hindi et ourdou de segmentation est de taille. Prenons par exemple les formes du futur. Pour le hindi comme pour l'ourdou, on distingue le radical du verbe, suivi d'un morphème de personne, suivi d'un morphème de genre et de nombre.

Exemple :

(52) جائے گا

jā-e gā
aller.2/3-S MS
« (tu, il) iras, ira » (3 morphèmes, 2 tokens)

La séquence se décompose en *jā* جا(radical « aller »), *e* ے (infixe de personne), *gā* گا (suffixe de nombre et de genre pour le futur), tandis qu'en hindi, les trois morphèmes sont réunis en un seul token :

(53) जाएगा
jā-e-gā
aller.2/3-S-M

« (tu, il) iras, ira » (3 morphèmes, 1 token)

Cette différence graphique a des conséquences sur l'étiquetage des tokens et empêche d'utiliser un système uniforme pour les deux langues.

3.3.3.7 Étiquetage morphosyntaxique en hindi et ourdou

L'étiquetage morphosyntaxique, ou POS tagging en anglais (de POS, part-of-speech, « partie du discours » ou catégorie morphosyntaxique), est une forme d'annotation qui associe à chaque token une étiquette morphosyntaxique : catégorie grammaticale, genre, nombre… Le nombre d'informations annotées peut varier. Les programmes d'étiquetage, appelés étiqueteurs, peuvent également préciser le *lemme* (forme canonique du token).

Les applications de l'étiquetage morphosyntaxique sont multiples : préparation à l'analyse syntaxique (pour lever des ambiguïtés), lemmatisation ou indexation, repérage des patrons morphosyntaxiques à partir desquels on pourra dégager la terminologie d'un corpus de spécialité, synthèse vocale à partir d'un texte (Yvon, 2007), repérage des entités nommées, etc.

Lors de l'étiquetage peuvent survenir des ambiguïtés, car une forme graphique peut appartenir à plusieurs catégories morphosyntaxiques. Certains étiqueteurs procèdent par des algorithmes statistiques ou par association de règles afin de les lever. Par exemple, dans le célèbre exemple suivant en anglais : « fruit flies like a banana »

fruit	NOM	ADJ
flies	VERB :3s	NOM :pl
like	PREP	VERB
a	DET	
banana	NOM	

Les trois premiers tokens possèdent chacun deux étiquettes. « fruit » peut être interprété comme un nom ou un qualificatif de « flies », « flies » comme un nom ou un verbe à la 3^{e} personne du singulier, « like » comme une préposition ou un verbe. La phrase peut prendre par conséquent deux sens possibles :

- Le fruit vole comme une banane.
- Les mouches drosophiles aiment une banane.

La plupart des outils d'analyse de corpus pour le français nécessitent un étiquetage morphosyntaxique et de nombreux étiqueteurs sont disponibles en cette langue. Certains

logiciels de traitement des corpus en intègrent un ou plusieurs. Il n'en existe cependant pas pour toutes les langues. Dans ce cas, il est possible d'entraîner un étiqueteur.

De par sa position préliminaire dans la chaîne de traitement, l'étiquetage morpho-syntaxique a fait l'objet de nombreuses recherches dans les principales langues écrites. Pour certaines langues, la question apparaît comme « relativement » résolue avec une recherche arrivée à maturité qui aboutit à des performances d'au moins 97,5% de segments correctement étiquetés[50].

Les travaux sur l'étiquetage des langues du sous-continent ont démarré assez tardivement en Asie du Sud et des standards spécialisés ont commencé à émerger au milieu des années 2000.

Pour mener à bien sa tâche, l'étiqueteur doit posséder un jeu d'étiquettes des parties du discours adapté à la langue étudiée. Malgré la grammaire commune au hindi et à l'ourdou, les usages graphiques ainsi que le développement d'une recherche séparés en Inde et au Pakistan n'ont pas permis de trouver un consensus sur l'étiquetage. Il existe différents jeux d'étiquettes des catégories morphosyntaxiques (tagsets) disponibles (ceux développés par Bharati et al. (2006) pour le hindi, ou Muaz *et al.* (2009) pour l'ourdou, au Pakistan.).

a. Étiquetage du hindi

Il existe en hindi quatre programmes d'analyse morphosyntaxique (ou *POS tagger*) qui utilisent tous la nomenclature fixée par (Bharati et al., 2006), elle-même dérivant de la liste du Penn tagset, conçue par l'université de Pennsylvanie en 1991 (Santorini, 1991). À partir de cette liste, de nouvelles catégories ont été créées afin de refléter au mieux la morphosyntaxe des langues de l'aire sud-asiatique. Par exemple, le Penn tagset propose différentes étiquettes pour les formes verbales (participes présents, passés, formes du présent, autres). Or dans les langues de l'aire sud-asiatique, les catégories de temps ou de formes participiales sont moins pertinentes que la décomposition des syntagmes verbaux. Ces derniers sont souvent composés d'un « verbe principal » (VM) et de « verbes auxiliaires » (VAUX). Après de nombreux débats, il a été convenu que même les verbes

[50]Il n'est pas possible de généraliser ce constat car de nombreuses langues posent des problèmes de segmentation. D'autres encore manquent de ressources qui permettraient d'enrichir les calculs statistiques. Enfin se pose aussi le problème de la spécificité du corpus : un étiqueteur n'obtient généralement pas les mêmes performances selon les types de corpus, pour une langue donnée.

non finis (les participes accomplis / inaccomplis, les coverbes, les radicaux servant de base aux explicateurs verbaux) seraient classés parmi les VM et que seul l'élément final, qui porte l'accord et la finitude, serait considéré comme l'auxiliaire et étiqueté VAUX (Bharati et al., 2006, 10).

Une autre catégorie introduite est celle des « noms spatiaux-temporels » (*Noun Spatial Temporal*, ou NST) qui permet de répondre au problème de la nature des locutions postpositives[51] du type *ke* + nom / adverbe de localisation :

– *ke andar* « à l'intérieur » (*andar* = « intérieur »),
– *ke sāth* « avec », (*sāth* = « compagnie »)

Les seconds éléments de ces locutions particulières ont d'autre part une existence indépendante où ils ne sont pas précédés d'une postposition au cas oblique (cf : *apnā sāth denā*, « apporter sont soutien », *kapṛe kā andar*, « l'intérieur du vêtement »), le plus souvent en tant qu'adverbes ou substantifs.

b. Étiquetage de l'ourdou

Le développement des programmes d'étiquetage en ourdou a suivi au Pakistan une voie parallèle, s'appuyant aussi sur les travaux américains ainsi que sur les recherches indiennes pour donner naissance à un étiquetage en 2009 (Muaz *et al.*, 2009).

Comme nous l'avons vu au paragraphe sur la tokenisation, l'ourdou et le hindi sont segmentés différemment. Bien que ces langues aient la même morphologie (usage des postpositions, flexions de cas, de genre et de nombre), les usages graphiques imposent des différences de taille, avec une tendance au synthétisme pour le hindi (les suffixes sont intégrés à la base lexicale, ainsi que certaines postpositions), et un mouvement contraire pour l'ourdou (séparation à l'écrit des morphèmes de suffixes et des postpositions). Pour un étiquetage morphosyntaxique au niveau des unités lexicales et non des morphèmes, l'ourdou pose davantage de difficulté en raison de l'éclatement graphique de l'unité lexicale en plusieurs morphèmes.

Si l'on suit la règle d'étiquetage des VAUX (dernier élément de la séquence verbale), l'étiquetage de l'exemple (52) en ourdou (« tu / il ira ») serait :

ja-e VM ga AUX

[51] Une postposition est un relateur grammatical qui se place après le mot. Presque tous les relateurs en hindi-ourdou sont de ce type.

On obtient deux étiquettes pour les trois morphèmes d'un seul verbe.

Ijaz et Hussain proposent de modifier le segmenteur en introduisant une liste de suffixes qui doivent être compris comme formant une unité avec le segment précédent (Ijaz et Hussain, 2007). Ce qui vaut par exemple pour des unités lexicales comme *ẓimmah-dārī* ذمہداری, écrit en deux segments *ẓimmah* ذمہ + *dārī* داری, ou pour les verbes au futur qui viennent d'être évoqués.

Concernant la catégorie des noms spatiaux temporels (NST, voir plus haut), elle est traitée en ourdou par l'étiquette NN, désignant les noms communs, ou ADV (adverbe), selon que le segment qui précède la postposition est un nom ou un adverbe. Les éléments de cette catégorie se comportent comme n'importe quel nom commun lorsqu'ils sont autonomes.

La catégorie des pronoms interrogatifs a été supprimée et ces derniers sont classés dans la catégorie des pronoms personnels (ex : *kaun*) ou des adverbes (ex : *kitnā*). L'étiquetage verbal distingue les infinitifs des autres formes verbales ainsi que quatre catégories de verbes auxiliaires. Une étiquette pour les *coverbes* (radical du verbe suivi du morphème *kar*) est introduite, SCK, plaçant ainsi les coverbes non plus dans la catégorie verbale mais dans celle des conjonctions de subordination.

Une autre nomenclature d'étiquettes, disponible sur le site de l'université de Lancaster[52], a été élaborée dans le cadre du projet EMILLE. Cette nomenclature est assez différente de celle de (Bharati et al., 2006) reprise pour de nombreux corpus et compte de très nombreuses étiquettes (plusieurs centaines) qui ne sont pas pertinentes pour la présente étude.

c. Personnalisation du jeu d'étiquettes

Afin de pouvoir comparer plus facilement les patrons morphosyntaxiques trouvés en hindi et en ourdou, certaines étiquettes ont été abandonnées et d'autre créées pour assurer une homogénéité. Il n'a cependant pas toujours été possible conserver partout cette uniformisation, en particulier pour tenir compte de la segmentation des formes verbales en ourdou ou des postpositions communément agrégées aux pronoms en hindi, tels des clitiques (cf. tableau 3.3).

[52]Voir le site : `http://www.lancaster.ac.uk/fass/projects/corpus/emille/U1tagset.htm`.

étiquette morpho-syntaxique	catégorie grammaticale
ADJ	adjectif
ADP	préposition
ADV	adverbe
AUX	auxiliaire (verbe fini en position finale d'une suite de formes verbales)
CONJ	conjonction de coordination
DET	déterminant (adjectifs démonstratifs, possessifs, indéfinis)
NOUN	nom (susbtantif)
NUM	nombre (en chiffre, en lettre, mais aussi numération en lettre de paragraphes, cardinal)
PRON	pronom
PROPN	nom propre
PSP	postposition
PUNCT	ponctuation
QO	ordinal
SCONJ	conjonction de subordination
SENT	ponctuation marquant une fin de phrase
SUF	suffixe (pour l'ourdou)
VALA	*vālā*, *vālī*, *vāle*
VERB	verbe fini ou non fini

Tab. 3.3: Jeu personalisé d'étiquettes morphosyntaxiques communes au hindi et à l'ourdou

En hindi, les pronoms ou démonstratifs suivis d'une postposition portent l'étiquette PRON (pronom) ou DET (déterminant) s'il s'agit d'un adjectif possessif. Dans les syntagmes verbaux, seul le dernier est marqué AUX lorsqu'il est fléchi. Un seul verbe fini est marqué comme VERB.

En ourdou, les suffixes entrant dans la composition des verbes, des substantifs ou des adjectifs sont marqués SUF et doivent être considérés avec le segment précédent. L'étiquette VALA désigne le suffixe adjectival ou verbal déclinable en genre et en nombre qui forme des composés avec des noms ou des verbes. Ce suffixe crée des relations diverses avec les noms ou verbes qui le précèdent, ce qui justifie la création d'une étiquette distincte.

Enfin, dans les deux langues, l'étiquette NST (noms spatio-temporels) a été abandonnée pour les raison évoquées plus haut d'indépendance du nom hors locution postpositionnelle. Les postpositions simples, en nombre très limité, sont marquées par PSP.

3.3.3.7.1 Phase d'apprentissage

Avant de s'appliquer à un corpus, un étiqueteur doit passer par une phase d'apprentissage qui lui permettra d'assigner une étiquette correct au segment. On compte trois approches d'apprentissage possibles : structurée par des règles (*rule-based*), stochastique (probabiliste) ou transformationnelle (hybride) [53].

La première approche consiste à définir manuellement un certain nombre de règles, notamment morphologiques, pour identifier les segments dans un syntagme. Elle est extrêmement coûteuse en temps et en main d'œuvre.

L'*approche probabiliste* choisit l'étiquette jugée la plus pertinente d'après le contexte du token, en fonction d'un corpus de référence et utilise les modèle de Markov cachés (Hidden Markov Models ou HMM). L'*approche transformationnelle*, quant à elle, combine les deux précédentes : une étiquette est choisie sur une base probabiliste. On vérifie si ce choix répond à des règles préétablies. Si ce n'est pas le cas, l'étiquette est modifiée. Les règles nouvellement apprises serviront à affiner l'étiquetage des segments suivants.

Plusieurs étiqueteurs morphosyntaxiques sont disponibles pour les langues indiennes comme Hybrid Postagger ou RDRPOSTagger. En ourdou, seul RDRPOSTagger est disponible. On peut aussi recourir à des étiqueteurs à entraîner à partir d'un corpus étiqueté manuellement (ou par tout autre moyen) et d'un lexique. Il faut toutefois prendre garde de choisir un étiqueteur qui soit portable et intégrable dans une chaîne plus vaste de traitement pour l'extraction de terminologie ou l'analyse des concordances en fonction des partie du discours.

Nous comparerons dans la séquence suivante différents outils d'étiquetage testés sur des corpus hindi et ourdou, dans la perspective de trouver l'outil le plus pratique pour traiter avec un minimum d'adaptations les trois langues de travail hindi, ourdou et français.

3.3.3.7.2 Evaluation des POS tagger

Pour évaluer les étiqueteurs, nous avons utilisé dans chacun des cas un processus d'apprentissage itératif. Un petit corpus d'essai (200 tokens) est étiqueté manuellement pour servir de modèle à l'étiqueteur. Ce dernier traite ensuite un corpus plus grand. On vérifie ensuite les étiquettes des 200 tokens suivants et on les corrige manuellement si

[53] Pour un état de l'art de la genèse des étiqueteurs, qui remonte aux années 1980, voir Màrquez et Rodrìguez (1998).

nécessaire. Le corpus modèle est ainsi augmenté de 200 tokens étiquetés. Le processus, appelé itération, est répété plusieurs fois pour permettre à l'étiqueteur de recalculer les probabilités et les règles d'étiquetage sur le corpus modèle sans cesse augmenté. Les résultats se présentent sous la forme de taux de précision (nombre de résultats corrects rapporté au nombre total de résultats).

a. Hybrid POS Tagger

L'étiqueteur HybridPOS tagger (Yoonus et Sinha, 2011) fonctionne pour le hindi et propose un étiquetage très détaillé : en plus des parties du discours, il indique le lemme, le suffixe, le genre, le nombre et le temps. Il est disponible pour d'autres langues indiennes mais pas pour l'ourdou, or l'objectif de l'étude est de trouver des outils utilisables dans les deux langues pour avoir la même qualité d'annotation dans les analyses ultérieures.

Le choix d'un étiqueteur à entraîner semble plus pertinent. Il permettra, en outre, de choisir un jeu d'étiquettes plus homogène, reflétant l'unicité de la grammaire de ces deux langues à l'origine commune.

b. RDRPOS Tagger

Cet étiqueteur développé par Dat Quoc Nguyen, Dai Quoc Nguyen, Dang Duc Pham et Son Bao Pham[54], est simple à installer et n'a besoin que d'un corpus modèle déjà étiqueté, préalablement segmenté. Avec ce petit nombre de paramètre, l'étiqueteur peut, après entraînement, être utilisé sur n'importe quelle langue. L'outil possède des modules de règles de dépendances universelles pour une quarantaine de langues, dont le hindi et tout récemment l'ourdou, et a déjà été entraîné pour une quinzaine de langues, dont le hindi (Nguyen et al, 2005). Dans la lignée de l'approche de Brill (1995)[55], il se fonde sur l'approche appelée « ripple down rules » (RDR) qui consiste à restructurer à chaque itération les interactions entre les règles déjà données en fonction des erreurs sous la forme d'une classification unique (Single Classification Ripple Down Rules, ou SCRDR). Le programme d'entraînement utilise une fenêtre de contexte de cinq tokens pour générer ses décisions. Bien que cet étiqueteur ait déjà été entraîné, il est nécessaire de l'entraîner

[54]RDRPOSTagger : A Ripple Down Rules-based Part-Of-Speech Tagger](http://www.aclweb.org/anthology/E14-2005). In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, EACL 2014, pp.17-20, 2014.

[55]Eric Brill. 1995. Transformation-based error-driven learning and natural language processing : a case study in part-of-speech tagging. Comput. Linguist., 21(4) :543–565.

à nouveau pour l'adapter au corpus spécialisé.

b.1. Test sur le corpus hindi

Après un bref entraînement de quatre itérations, le taux de précision est de 89,8 % (22 erreurs d'étiquetage sur 216 segments) sur un sous-corpus de textes normatifs (texte de loi). L'expérience est réitérée sur un corpus de presse généraliste. 69 étiquettes sont erronées sur 211 segments, en particulier dans les séquences verbales où seul le dernier segment verbal doit être étiqueté auxiliaire. Le taux de précision descend à 67,2 %, puis remonte à à 86,2 % (28 erreurs sur 204 segments), résultat proche de celui obtenu sur le corpus spécialisé.

L'essai, réitéré sur un corpus littéraire tiré de Frantext, donne un taux de précision légèrement plus bas (81,8 %, avec 41 erreurs sur 225 segments).

On constate que même après la modification des règles pour l'étiquetage des corpus non spécialisés, la réitération de l'étiquetage sur le corpus spécialisé initial donne de bons résultats : le taux de précision monte à 94,3 % (12 erreurs d'étiquetage seulement).

RDRPostagger génère son lexique à partir du corpus d'apprentissage et intègre les modifications apportées après chaque itération. La crainte que les règles modifiées d'une itération à l'autre ne fassent retomber le taux de précision en revenant à un corpus spécialisé ne semble pas fondée pour le hindi.

b.2. Test sur le corpus ourdou

Les premiers essais ont commencé à partir d'un corpus étiqueté en ourdou d'environ 10 000 tokens trouvé sur le site `http://www.cle.org.pk/software/ling_resources/UrduNepaliEnglishParallelCorpus.htm`[56]. Certaines étiquettes en ourdou ont été abandonnées ou fusionnées avec d'autres afin d'obtenir une nomenclature d'étiquetage normalisée.

Avant de commencer à entraîner l'étiqueteur, le corpus modèle est passé en revue pour en corriger les erreurs. Une première itération permet d'obtenir un résultat de 186 étiquettes correctes sur 226 ce qui représente un taux de précision de 79,5%. Les itérations suivantes donnent des taux de précision entre 82% et 89%.

Le résultat de l'étiquetage avec RDRPostagger est plutôt satisfaisant compte tenu du peu de ressources qu'il demande et de la rapidité d'apprentissage, même lorsqu'il

[56]Traductions en ourdou de mots et phrases en anglais.

existe des erreurs de segmentation. Cet étiqueteur n'est cependant pas intégré aux outils d'analyse de corpus dont nous souhaitons nous servir par la suite. Les fichiers produits lors de l'apprentissage (dictionnaire, fichier modèle standard), spécialisés pour les corpus juridiques, seront réutilisables pour d'autres travaux sur ces langues.

c. TreeTagger

Présentation Bien qu'il existe différents étiqueteurs morphosyntaxiques pour le hindi (RDR POStagger, hindi POS tagger, etc), l'avantage de TreeTagger est qu'il est intégré dans une grande variété de logiciels et d'applications d'extraction terminologique. Il est donc fort utile de pouvoir s'en servir pour des langues telles que le hindi et l'ourdou. Développé par Schmid (1994) dans les années 1990, il est disponible en libre accès[57]. La difficulté qui se présente avec le hindi et l'ourdou était qu'il n'existait pas en l'état de fichiers de paramètres disponibles et prêts à l'emploi pour ces langues. Le cas ayant été prévu par l'inventeur, il est possible d'entraîner TreeTagger à l'aide d'un corpus préétiqueté qui servira de modèle, et d'un lexique (lemmatisé ou non).

Fonctionnement Entraîner TreeTagger nécessite

- un programme *train-tree-tagger* ;
- un fichier de lexique, qui liste tous les lexèmes avec leur POS dans un fichier selon le format [lexème][tabulation][POS]
- un fichier open-class-tags qui permet à Tree Tagger d'étiqueter des tokens qu'il n'a jamais vus. Au contraire, des classes comme la ponctuation, les pronoms ou les conjonctions seront des classes limitées à un certain nombre.
- un fichier d'entrée, qui est un corpus déjà étiqueté à la main. Sous forme de colonnes, il comprend un token par ligne, suivi d'un POS et éventuellement d'un lemme, ce dernier n'étant pas obligatoire. Dans le cas d'homographes, tous les POS possibles doivent être notés sur la même ligne, suivis pour chacun de leur lemme (ou d'un « faux lemme » comme un tiret si l'on en n'a pas besoin).
- éventuellement un fichier de sortie.

La méthode utilisée par Tree Tagger est probabiliste mais ne s'appuie pas sur les modèles de Markov. Les transitions sont estimées à l'aide d'arbres de décision.

[57] Sur le site `http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/`.

Méthode d'apprentissage Pour l'apprentissage, nous avons procédé comme expliqué en 3.3.3.7.2. Nous avons pris au départ un corpus étiqueté à la main et de dimension modeste (200 mots). Après avoir généré un premier fichier de paramètres, nous avons sélectionné un corpus deux fois plus gros auquel nous l'avons appliqué. Nous avons ensuite corrigé les 200 dernières lignes (soit 200 mots) manuellement, et ainsi de suite pour affiner la précision des résultats. Un sous-programme perl a permis de générer au fur et à mesure un lexique lorsqu'il n'y en avait pas de disponible.

Dans le cas de nos langues, il n'existait pas de dictionnaire associant lexique et lemmes prêt à l'emploi. Nous avons recherché des listes de mots par catégorie grammaticale sur Wikipédia. La morphologie des mots du hindi-ourdou est assez simple (deux cas, deux nombres et deux genres, un système verbal analytique avec des formes peu nombreuses et peu d'irrégularités), ce qui a facilité la génération de toutes les formes possibles à partir d'un lemme pour enrichir le fichier lexique.

c.1. Test sur le corpus hindi

Le corpus d'entraînement est un fichier texte relatant un dépôt de plainte. Ce premier entraînement est axé sur les textes juridiques et judiciaires. Lors de la première itération, la précision est de 73 %. À la 2e itération, le taux monte à 81,5 %.

Une fois le fichier de paramètres généré, on exécute le programme en lui fournissant en argument le fichier d'entrée et éventuellement un fichier de sortie :

```
cmd/tree-tagger-hindi [chemin-corpus-à-étiquetter] > [chemin-fichier-de-sortie]
```

L'essai est poursuivi sur un corpus non étiqueté de presse, afin de tester un autre domaine. Le taux de précision plonge à 65,5 %.

À chaque itération, les 200 lignes corrigées sont intégrées au corpus modèle qui sera repris dans les entraînements suivants.

L'expérience d'entraînement est réitérée avec cette fois-ci un fichier de lexique lemmatisé. Le test est mené à nouveau sur le corpus de presse. Après plusieurs itérations, le taux de précision est de 88,5 %

Le test se poursuit sur un corpus du domaine judiciaire avec un taux de précision de 89,5 %.

Enfin, pour terminer ce test avec le hindi, le programme est appliqué à un éditorial de presse (600 segments). Le taux de précision redescend à 87 %.

c.2. Test sur le corpus ourdou

Un échantillon de 200 lignes d'un corpus en ourdou déjà étiqueté à l'aide d'un autre étiqueteur et corrigé à la main sert de point de départ à cette nouvelle phase de l'expérience. Le lexique qui sert à entraîner l'étiqueteur ne contient pas de lemmes, accélérant les calculs. Un premier entraînement est suivi d'un étiquetage sur un corpus de 1661 lignes, donnant un taux de précision à peine supérieur à 50 %, bien loin des 73 % de la version initiale en hindi.

Les entraînements ultérieurs montrent une amélioration substantielle de ce score. À l'itération suivante sur le même corpus, le taux de précision monte à 85,7 %. Après avoir entraîné une nouvelle fois l'étiqueteur, celui-ci est appliqué à un corpus d'origine différente. L'étiqueteur est à nouveau entraîné et le vocabulaire nouveau ajouté à la liste déjà constituée.

Après un travail de recherche ou de génération de lemmes visant à enrichir le lexique d'entraînement, l'étiqueteur est à nouveau entraîné sur un corpus modèle (974 segments) plus grand puisqu'il a bénéficié des phases de correction des essais antérieurs. L'étiqueteur est ensuite testé sur un corpus de 7000 lignes (ou segments). Le taux de précision monte à la fin de l'apprentissage à 83,6 %.

Corpus hindi												
	RDRPostagger					**TreeTagger**						
nb tokens	216	211	204	225	211	300	239	200	201	200	200	203
nb étiquettes correctes	184	142	176	184	199	219	195	131	153	177	179	176
nb étiquettes incorrectes	22	69	28	41	12	81	44	69	48	23	21	27
précision (%)	89,8 (1)	67,2 (2)	86,2 (3)	81,8 (4)	94,3 (5)	73 (1)	81,5 (1)	65,5 (2)	76 (2)(6)	88,5 (3)(6)	89,5 (1)(6)	87 (3)(6)

TAB. 3.4: Résultat de l'entraînement des étiqueteurs morphosyntaxiques. (1) test sur corpus juridique, (2) test sur corpus de presse, (3) test sur corpus de presse (réentraîné), (4) test sur corpus littéraire, (5) test sur corpus juridique (réentraîné), (6) avec lexique comprenant des lemmes.

c.3. Interprétation des résultats

c.3.1. Conclusion pour le hindi

Au bout d'un certain nombre d'itérations, l'étiqueteur RDRPostagger s'avère plus

fiable. Il est plus simple, plus rapide et demande moins de ressources, en particulier lexicales, ce qui convient à des langues pour lesquelles il est difficiles de trouver des ressources libres de droit ou fiables. Il obtient de très bons résultats sur des corpus test spécialisés ou non.

Corpus ourdou												
RDRPostagger						**TreeTagger**						
nb tokens	226	209	228	202	235	200	211	205	199	226	210	200
nb étiquettes correctes	186	187	196	170	193	108	181	140	125	156	159	166
nb étiquettes incorrectes	40	22	32	32	42	92	30	65	74	70	51	34
précision (%)	79,5 (1)(3)	89,5 (1)(3)	86,5 (2)	84 (2)(3)	82 (1)	51 (1)(3)	85,7 (2)	69 (1)(4)	62,8 (1)(3)(4)	69 (1)(3)(4)	75 (1)(3)(4)	83

TAB. 3.5: Résultat de l'entraînement des étiqueteurs morphosyntaxiques. (1) test sur corpus juridique, (2) test sur sous-corpus normatif, (3) après réentraînement, (4) avec lexique comprenant des lemmes.

c.3.2. Conclusion pour l'ourdou

Les résultats, tout d'abord meilleurs avec RDRPostagger au bout de quelques essais, stagnent autour de 82-83 % avec les deux méthodes, loin des taux de précision du hindi qui sont près de 10 points supérieurs. Plusieurs facteurs peuvent expliquer cet écart. Tout d'abord, le hindi est morphologiquement plus simple à étiqueter. D'une part, à la différence de l'ourdou, il n'y a pas de segmentation graphique en morphèmes dans un mot en hindi et d'autre part, toutes les voyelles (sauf le *-a* inhérent) sont indiquées, ce qui conduit à un nombre d'ambiguïtés bien moindre qu'en ourdou. De plus, en ce qui concerne TreeTagger, nous disposions de davantage de ressources lexicales pour le hindi (en particulier des listes de lemmes plus fournies trouvées sur Wikipédia). Enfin, la différence s'explique aussi par un meilleur préétiquetage pour le hindi dans le corpus d'entraînement. Le taux de précision pour l'ourdou pourrait par conséquent être amélioré par un étiquetage plus fin (par exemple en signalant les suffixes verbaux, nominaux ou adjectivaux qui devraient être rattachés au segment précédent) et par un enrichissement du lexique. Les taux de précision restent toutefois moindre dans tous les cas pour ces langues que pour celles dont les fichiers de paramètre sont déjà disponibles. Un entraînement plus long, sur une plus grande diversité de corpus, pourrait améliorer nos résultats.

3.3.3.8 Ressources lexicales et ontologies WordNet

Les ressources lexicales (ontologies, lexiques bilingues, dictionnaires, listes de lemmes, listes d'entités nommées) offrent des aides précieuses tout au long de la chaîne de traitement. Elles permettent d'affiner l'étiquetage, surtout si elles contiennent des annotations grammaticales, ou de faciliter l'anonymisation de corpus, dans le cas des listes d'entités nommées (noms propres, toponymes, etc.).

Les réseaux ontologiques WordNet, qui ont vu le jour à partir de l'anglais, tentent d'organiser ces ressources. Il en existe dans de nombreuses langues, avec des variations par rapport au système initial.

Un Indo WordNet a été construit pour 18 langues indiennes (assamais, bengali, bodo, gujarati, hindi, kannada, cachemiri, konkani, malayalm, manipuri, marathi, népali, oriya, pendjabi, sanskrit, tamoul, télougou et ourdou). Le WordNet hindi a été mis en ligne pour la recherche en 2006.

Le principe des réseaux WordNet est d'identifier les relations entre les concepts qui forment des réseaux sémantiques. Dans le cas où plusieurs mots dénotent un même concept, l'un d'eux est identifié comme l'« unité primaire » (*primary unit*) et porte un code d'identification. Tous les autres sont considérés comme étant ses synonymes (*synset*). L'unité primaire est le mot le plus générique et le plus représenté dans le corpus pour un concept donné. Les réseaux WordNet sont construits selon deux approches : l'approche par fusion (*merge approach*, « ME ») et l'approche par extension (*extension approach*, « EA »).

L'approche ME convient pour une langue donnée, sans tenir compte des langues voisines. On enregistre tous les sens des mots utilisés dans la langue. Lexicographes et lexicologues construisent des synsets pour chaque sens donné trouvé selon les principes de minimalisme, « portée » (*coverage*), « interchangeabilité » (*replaceability*).

L'approche EA, quant à elle, considère un groupe de langues voisines. Une langue est choisie comme langue source, on génère une liste de synsets complète d'après l'étude des sens des mots qui existent dans la langue. Ensuite des unités lexicales conceptuellement équivalentes sont extraites des lexiques des langues cibles pour être étiquetées avec les synsets de la langue source.

Certains principes président à l'alimentation de ces réseaux WordNets[58]

[58] D'après la présentation sur `http://www.ircc.iitb.ac.in/IRCC-Webpage/rnd/PDF/GlimpseIITBResearch/Indo_Word_Net.pdf`.

– Les synsets des langues cibles ne doivent pas être de simples translittérations des mots de la langue source, qui est ici le hindi ;
– Les concepts exprimés équivalents sont des unités mono ou polylexicales ;
– Pour fournir des équivalents conceptuels, quatre méthodes alternatives peuvent être utilisées, par ordre de préférence : mots reconnus par le dictionnaire, translittération, courtes périphrase ou mots composés, néologismes.
– Les entrées du dictionnaire sont intégrées en fonction de la fréquence d'occurrence dans la langue ;
– La translittération n'est utilisée que pour les termes scientifiques, techniques, médicaux qui n'ont pas d'équivalent dans les langues autochtones, ou lorsque la traduction risque d'être non compréhensible pour un usager moyen (Hock et Bashir, 2016, 758) ;
– On utilise des mots composés ou de courtes périphrases lorsque le concept existe dans la langue cible mais qu'il ne peut s'exprimer en un seul mot ;
– Les néologismes sont utilisés seulement si le concept dans la langue source est unique et n'existe pas dans la langue cible. Cela soulève la question de la standardisation et n'est pas encouragé ;
– Des listes distinctes sont créées pour chaque lot de synsets propres à la langue car des concepts existant dans la LC pourraient ne pas exister dans la langue source ;
– Les mêmes identifiants de synsets doivent être utilisés pour toutes les langues de Indo WordNet.

Au Pakistan, les résultats des recherches du CLE (Center for Language Engineering) ne sont pas encore disponibles en libre accès pour la recherche. Le centre a réalisé en 2007 un premier lexique numérique en ourdou.

Nous n'avons pas eu la possibilité d'utiliser des outils permettant d'accéder au réseau Indo WordNet et avons créé nous-même des listes lexicales. Nous avons récupéré les données disponibles dans des listes de lemmes et de mots classés par catégorie grammaticale sur Wikipédia et les Wiktionnaires. Toutes ces listes ont nécessité des corrections et des compléments. Nous avons par exemple recherché les noms et prénoms sur les pages Wikipédia qui répertorient les noms de castes et de sous-castes. Nous avons tenté de mettre en place des automates sous Unitex[59] et avons développé des modules pour l'utilisation du hindi et de l'ourdou sous Unitex susceptibles de servir à des travaux

[59]Logiciel d'analyse de corpus, cf. Paumier (2011).

ultérieurs[60].

3.3.3.9 Anonymisation des sous-corpus juridictionnels

Notre corpus comprend, nous l'avons évoqué en (3.2), un sous-corpus constitué de textes judiciaires émanant de véritables procédures : jugements, ordonnances, citations… Pour des raisons de confidentialité des données, il a fallu procéder à leur anonymisation, afin que les différents protagonistes ne soient pas reconnaissables. Ce processus implique le repérage des *entités nommées*. « Une entité nommée est une expression linguistique référentielle, souvent associée aux noms propres et aux descriptions définies.[61] » Ce peut être le nom d'une personne, d'un personnage réel ou fictif, d'un lieu, ou encore tout type d'expression permettant de repérer un contexte : un numéro de dossier, de téléphone, une adresse, un âge, un horaire, etc. Ces entités nommées, repérées généralement à partir de bases de données (listes de noms propres de personnes et de lieux[62]) et d'expressions régulières[63], sont identifiées par des balises qui indiquent la catégorie d'entité dont elles font partie (balises <LIEU>, <NOM>, <MEC(mis en cause)> …).

L'anonymisation des textes hindi et ourdou a été réalisée manuellement, en raison du petit nombre de tokens et des ambiguïtés trop grandes entre les nom propres et les noms communs. Il n'existe pas dans les écritures de ces langues de marque pour différencier des noms propres (comme les majuscules en écriture latine) et de nombreux noms communs peuvent servir de noms propres (noms de qualité, de fleurs, etc.).

3.3.4 *Conclusion*

Le genre judiciaire est très hétérogène. Il est constitué d'une multitude de sous-corpus qui forment des sous-genres plus ou moins spécialisés, répondant à des fonctions (information, performativité, pédagogie, création de la norme) et à des publics divers (spécialistes du droit ou non initiés). Nous nous sommes efforcée de rassembler des données pour alimenter des sous-corpus de chacun de ces sous-genre pour les trois

[60]Nous remercions Éric Laporte pour ses conseils pour la création de modules linguistiques.

[61]Source : WIkipédia `https://fr.wikipedia.org/wiki/Entit%C3%A9_nomm%C3%A9e`.

[62]Les listes utilisées proviennent de l'INSEE et Wikipédia. Nous avons également tenté de repérer les entités nommées à l'aide d'automates sur Unitex, mais les ambiguïtés entre certains noms étrangers et des mots de français courant étaient trop importantes et la tentative n'a pas fonctionné.

[63]Une expression régulière est une chaîne de caractères qui décrit un ensemble de chaînes de caractères possibles.

langues étudiées. L'informatique a contribué à cette collecte, ainsi qu'au traitement des divers corpus à des fins d'alignement et d'extraction de terminologies.

Le détail de la chaîne d'opérations qui a permis de traiter le corpus a montré, dans le domaine de la linguistique de corpus, le manque d'outils en libre accès pour traiter des langues proches génétiquement comme le hindi et l'ourdou, au fonctionnement très semblable mais à la forme dissemblable. L'une s'écrit à l'aide d'une graphie proche de la phonétique, de gauche à droite, tandis que dans l'autre chaque glyphe n'est pas associé à un son unique. Leur segmentation même est différentes. Ces différences nécessitent des adaptations dans les traitements, que nous nous sommes efforcée de rendre les plus cohérentes possibles (jeu d'étiquettes morphosyntaxiques harmonisé, entraînement de modules du même étiqueteur, développement de ressources lexicales parallèles), afin d'analyser le corpus avec les mêmes outils pour les trois langues.

L'enchaînement de ces opérations rend possible l'exploitation textométrique du corpus en vue de l'analyse du lexique et des phraséologies des différents sous-genres. Cette analyse permettra d'identifier les correspondances traductionnelles entre les langues.

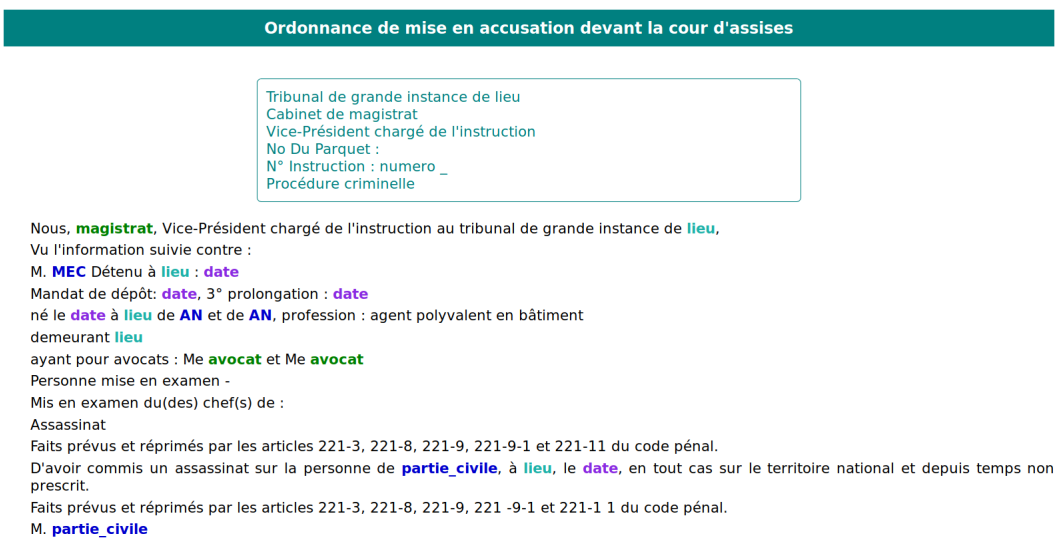

FIG. 3.6: Exemple d'anonymisation d'un texte. Extrait d'une ordonnance de mise en accusation, parties en-tête et exposé des faits.

Chapitre 4

Résultats des analyses textométriques et du phraséolexique judiciaire

Les chapitres 1 et 2 ont posé le contexte judiciaire et linguistique du corpus. Ils en ont expliqué l'histoire, le bagage culturel et l'hétérogénéité de la langue.

Le chapitre 3 a présenté le détail des genres textuels pour chaque langue, rendant compte d'une multiplicité de sous-genres à l'intérieur du genre judiciaire. Nous y avons expliqué comment tous les documents qui constituent le corpus ont été traités par l'informatique, en insistant sur la prise en charge des langues indiennes (hindi et ourdou).

Ce dernier chapitre présente le fruit des analyses que l'ensemble des traitements et réflexions théoriques précédents ont rendues possibles. Dans une première partie, nous analysons le corpus à l'aide d'une approche statistique, la textométrie, afin d'identifier les spécificités des sous-corpus. Cette étape est préalable à l'extraction des terminologies en contexte.

Dans la partie suivante, nous proposons d'aligner des corpus parallèles ou comparables, à l'aide des candidats-termes extraits qui servent de points d'ancrage, afin de relever les correspondances traductionnelles entre les langues.

La dernière partie présente une analyse plus poussée du lexique de genre, des sous-genres et des phraséologies qui les caractérisent, toujours en vue de faciliter la traduction. Les applications en ligne qui mettent à disposition ces résultats y sont également décrites.

4.1 Analyse textométrique

Une fois les textes des corpus nettoyés, enregistrés dans des formats interopérables, anonymisés et étiquetés, la phase d'analyse proprement dite peut commencer. Nous avons d'abord tenté différentes approches pour extraire les termes de spécialité judiciaire, puis recherché dans des concordanciers les cooccurrents qui serviront de base au recensement des unités phraséologiques.

4.1.1 *La textométrie, une approche à travers le texte et le contexte*

Peu connue dans la sphère anglo-saxone[1], l'approche textométrique naît dans les années 1970 et repose sur la statistique lexicale née dans les années 1950, la statistique textuelle et la lexicométrie. Benzécri et al. (1973) met au point des analyses factorielles et des classifications appliquées aux données linguistiques. Les nouveaux modèles de la discipline rendent compte de la significativité des données en contexte[2].

Dans le cadre de la présente recherche, l'analyse lexicométrique permet de faire ressortir par contraste avec un corpus standard le lexique spécifique d'un corpus spécialisé (calcul des spécificités) et sa phraséologie par la recherche des segments répétés ou des cooccurrents.

4.1.2 *Extraction terminologique*

L'approche textométrique a été adoptée pour procéder à l'extraction terminologique des corpus.

L'extraction terminologique consiste à rechercher des candidats-termes qui constituent potentiellement la terminologie du corpus. Le contraste de ce dernier avec un corpus de référence (cf. B), représentatif des usages de la langue en général, met en valeur les éléments surreprésentés ou sous-représentés dans chaque partie. Les éléments les plus représentés dans le corpus spécialisé par rapport au corpus de référence ont de fortes chances d'être des termes (avant validation, ce sont des candidats-termes).

4.1.2.1 La recherche des spécificités

La probabilité d'apparition d'un évènement *A* (mot ou expression) se calcule en comptant le nombre d'occurrences de *A* (*f* = fréquence de A) dans chaque corpus (partie *p*). On divise ensuite la fréquence *f* obtenue par le nombre total d'occurrences *t* (nombre de mots) par partie. La fréquence de *A* est ainsi pondérée, elle n'est plus dépendante de la taille des parties, c'est une fréquence relative. Or il a été démontré que cette fréquence ne suffit pas à rendre compte du mode de distribution du mot que l'on recherche (Lafon, 1984).

[1] Et par conséquent peu connue en Inde ou au Pakistan. Le monde de la recherche anglo-saxon a développé à partir de la statistique lexicale une approche basée sur l'indexation et le requêtage (text mining, recherche d'information, analyse sémantique latente, etc…).

[2] Informations tirées du Textométrie (2007), de l'ENS Lyon.

Le calcul de l'indice de spécificité repose sur la loi de probabilité hypergéométrique. L'indice de spécificité représente « la probabilité qu'une forme A apparaisse f fois dans une partie p de longueur t, la forme apparaissant F fois en tout dans l'ensemble du corpus dont la longueur totale est de T occurrences »[3].

Pierre Lafon (Lafon, 1980) a modélisé le calcul de l'indice de spécificité par l'équation suivante :

$$Prob_{specif}(card\{A \in V | A \in p\} \geqslant f_{obs}) = \sum_{f=f_{obs}}^{card\{A \in V | A \in p\}} Prob_{specif}(card\{A \in V | A \in p\} = f)$$[4]

Dans le cadre de cette recherche, les calculs d'indices de spécificité ont été réalisés avec le logiciel TXM[5]. Le corpus de la recherche a été divisé en langues, chaque corpus monolingue étant contrasté avec un corpus de référence[6]. Les unités retenues comme candidats-termes devront avoir un indice égal ou supérieur à 2. En dessous de ce seuil, l'unité n'est pas considérée comme spécifique au corpus. Elle se situe dans la zone de banalité (entre -2 et +2).

4.1.2.2 Limites de la recherche des spécificités

Une des particularités de la langue judiciaire est d'utiliser de très nombreuses unités lexicales de la langue commune en les resémantisant ou en les utilisant dans des collocations en composition avec d'autres mots tout aussi « communs », ce qui risque d'affecter les scores de spécificité. Par exemple, le nom « mise », qui fait partie de la langue courante, intervient dans la composition de nombreuses collocations dont il constitue la base (« mise en accusation », « mise en examen », « mise en cause », « mise à l'épreuve », « mise à l'isolement », « mise à exécution », « mise en liberté »…). Il peut être aussi étiqueté comme un participe passé et ne pas être reconnu lors de l'extraction de patrons à base nominale.

Les premiers résultats en français donnent, sur 129 617 entrées :

- 467 unités avec un score supérieur à 100,
- 2059 unités avec un score supérieur à 10,
- 6287 unités ayant un score supérieur à 2.

Ces résultats contiennent beaucoup de bruit, provenant de la ponctuation et des entités nommées des corpus non anonymisés (presse, forums, textes de loi). On y trouve aussi

[3]Projet Textométrie, ENS Lyon, 2007, cf. `http://textometrie.ens-lyon.fr/`.

[4]D'après Textométrie (2007), *op. cit.*

[5]Cf. Textométrie (2007), *op. cit.*

[6]Pour les sources des corpus de référence, voir l'annexe des ressources B.

les références aux articles de loi (nombres suivis ou précédés de lettres) et de nombreux mots vides[7]. Un nettoyage à l'aide de listes et d'expressions régulières, puis une relecture manuelle de la sélection, ramènent le nombre de candidats-termes à 2090. Les résultats contiennent aussi du silence, puisque des unités importantes comme « plaider », « empreintes » ou « médico-judiciaire » n'atteignent pas le seuil de 2. De plus, dans l'outil utilisé, le calcul des spécificités ne s'applique qu'à des monogrammes (séquences graphiques délimitées par des espaces ou des symboles de ponctuation).

4.1.3 *Recherche des segments répétés*

Afin de compléter la liste des candidats-termes monogrammes par des collocations, il est nécessaires de rechercher les segments répétés. Un segment répété est une séquence de mots non séparés par des symboles de ponctuations qui apparaît au moins deux fois dans le corpus (Lafon et Salem, 1983). Cette répétition porte à croire que le segment est significatif dans le texte, et en fait un candidat potentiel pour caractériser sa phraséologie. On calcule les segments répétés en fonction de leur fréquence par rapport à la fréquence de chaque segment du corpus. Leur longueur est variable (de 2 à n-grammes). Les segments trop longs contiennent souvent des répétitions internes, tandis que les bigrammes contiennent de très nombreuses séquences non pertinentes, comme des déterminants suivis d'un nom, ou un auxiliaire suivi d'un participe.

Pour rechercher les segments répétés, nous avons utilisé le logiciel en ligne iTrameur[8], capable de prendre en charge des langues aux scripts aussi variés que celles de cette étude. Il offre également l'avantage de rechercher les segments répétés en fonction des catégories morphosyntaxiques lorsque le corpus fourni est étiqueté.

Nous avons fixé un seuil minimal de 5 occurrences par segment et une longueur maximale de 10. Pour l'ourdou, les résultats montrent que la recherche de tokens trop longs s'avère inutile. Sur 562 segments retenus après nettoyage (élimination des séquences grammaticales incomplètes, des motifs morphosyntaxiques aberrants, des mots-vides), 513 sont des tri- ou tétragrammes. Aucun segment ne dépasse une longueur de 9. La démarche est la même pour l'analyse des segments répétés en hindi. Pour le français, nous avons

[7] Articles, verbes courants ou mots grammaticaux, qu'il ne faut pas rejeter en bloc car certains comme la conjonction « que », par exemple, entrent dans des collocations typiques du langage juridique.

[8] Développé par Serge Fleury, cette version en ligne reprend les principales fonctionnalités du Trameur, cf. `http://www.tal.univ-paris3.fr/trameur/iTrameur/` pour la plateforme et la documentation.

utilisé les fonctionnalités du logiciel TermSuite, qui avait également permis l'extraction des termes[9].

Après une révision manuelle, les segments répétés pertinents sont ajoutés à la liste des candidats-termes. Le projet initial prévoyait de les présenter ensuite à des personnes ressources (juristes) chargées de les évaluer. Des outils en ligne avaient été mis en place afin de pouvoir présenter ces candidats-termes avec leurs contextes d'occurrence dans le corpus[10]. Un trop petit nombre de ces candidats-termes ayant pu être évalué, nous avons décidé de conserver ces données pour corroborer les recherches des cooccurrents et des unités phraséologiques après alignement des corpus.

4.1.4 *Contraste entre genres textuels*

Le contraste entre les différentes parties du corpus est une application concrète de la recherche des spécificités d'un corpus et des segments répétés. Il permet d'identifier les unités lexicales et phraséologiques qui caractérisent un sous-genre par rapport à un autre. Nous avons déjà évoqué le caractère hétéroclite de notre corpus pour une langue donnée. Dans la partie sur les genres textuels liés à la procédure, nous avons observé que les textes juridictionnels possédaient à la fois des traits marqués de la langue administrative, avec des expressions figées et du lexique partagés par les autres branches administratives, et une partie plus spécialisée de ce phraséolexique typique du judiciaire, notamment avec des étymons d'origine arabe ou persane, hérités des périodes coloniale et moghole. Ces genres textuels, qui vont du parajuridique au genre normatif, présentent des proportions variables de lexique de langue générale et de langue très spécialisée du droit. Ces mesures de contraste permettent de mesurer la difficulté des textes judiciaires pour un non initié. L'étude qui suit a été réalisé en hindi, qui possédait un corpus avec davantage de formes de sous-genre que l'ourdou. Nous avons étudié le contraste entre la langue des forums et la langue des textes normatifs, puis entre la langue de la presse et ces mêmes textes normatifs.

---

[9] Rocheteau et Daille (2011).

[10] L'outil est en ligne sur le serveur de l'équipe de recherche à l'adresse suivante : `http://www.er-tim.fr/~stuck/BenedicteV2/html/choixLotDeTermes.html?idUtilisateur=v05`. La plateforme, en français et en anglais, a été conçue pour évaluer des termes en français, ourdou et hindi, mais peut supporter n'importe quelle langue. Faute de personne-ressource pour l'ourdou, les termes et leurs contextes n'ont pas été mis en ligne pour cette langue mais restent disponibles sur demande.

4.1.4.1 Contraste entre les sous-genres parajuridique et normatif en hindi

Cette série de contrastes oppose des textes parajuridiques aux textes normatifs (les lois). Afin de mesurer l'écart de lexique et de phraséologie entre les textes normatifs et les textes sur la justice en langue non spécialisée, nous avons procédé à plusieurs comparaisons entre sous-genres, en partant des sous-genres les plus éloignés en terme de spécialisation.

La première comparaison est un contraste entre le corpus de forums en hindi et celui des textes de lois. La langue des forums est une langue de non spécialistes qui posent des questions et souhaitent obtenir des conseils de la part de spécialistes du droit. C'est la langue du locuteur qualifié de moyen, bien qu'il s'exprime quand même par voie électronique, donc probablement urbain et assez éduqué.

Le second contraste est opéré entre la presse et les textes de loi. La presse, nous l'avons évoqué plus tôt, s'exprime dans une langue peu spécialisée en général, mais elle reprend une partie des codes de la langue normée, partagés également par la langue administrative (utilisation du passif, incorporation de néologismes opaques d'origine sanskrite), et intègre une partie du lexique spécialisé pour donner une « saveur » judiciaire (au moyen de persianismes, arabismes, et la reprise de termes judiciaires).

4.1.4.2 Contraste entre les forums et les textes de loi en hindi

Contraste entre les forums et les textes de loi en hindi

Lors de ce contraste, nous avons analysé les unités lexicales et phraséologiques les plus fréquentes et les plus spécialisées autour de thèmes similaires. Ces deux corpus sont de taille très inégale (plus de 225 000 tokens pour la partition « *Lois* », moins de 11 000 pour les « *Forums* »). Compter les fréquences absolues des unités n'est dans le cas présent pas pertinent. Il est en revanche intéressant de s'appuyer sur les spécificités et sur l'absence totale de lexies dans l'une ou l'autre partition.

Les thèmes abordés diffèrent entre ces deux partitions : les *Forums* abordent des problèmes pénaux comme civils, en particulier le mariage et le divorce, tandis que les *Lois* se réfèrent à une palette plus vaste d'infractions, mentionnent des articles, s'intéressent aux peines et à leur mise en œuvre. Le genre même du forum implique que l'on pose des questions, qu'on y cherche des réponses, et que le cadre énonciatif se déroule entre deux participants. Les 1re et 2e personnes y sont surreprésentées, tout comme les pronoms et adjectifs interrogatifs. Nous retirons donc de notre comparaison tout une liste d'unité lexicales trop spécifiques des deux genres (organisation des textes de loi, mots structurant

l'énonciation des problèmes et des réponses, thèmes du droit civil et déictiques de 1re et 2e personne pour les forums) pour étudier les unités lexicales et tournures qui se réfèrent à des situations et des thèmes similaires, voire des synonymes les unes des autres.

4.1.4.2.1 Distribution exclusive des lexèmes

Une série de lexèmes sont exclusivement utilisés dans la partition *Forums*. Ils possèdent des synonymes dans la partition législative qui sont, de leur côté, réservés à cette partition. On constate une réelle dichotomie entre ces sous-genres sur ces points, avec une utilisation dans les *Forums* de termes d'origine persane, arabe ou anglaise qui ont intégré la langue courante, contre des termes très sanskritisés réservés au législatif. Les unités lexicales suivantes sont exclusives à la partition *Forums*. Les scores de spécificité sont mentionnés entre parenthèses lorsque le nombre d'occurrences dans l'autre partition n'est pas nul.

terme	Forums	Lois	traduction
mukadmā	9	0	affaire, procès
kes	41	1	« case » affaire, procès
adālat	20	0	tribunal
kānūn	16	0	loi
gavāh	6	0	témoin
vakīl	12	0	avocat
gavāhī	10	0	témoignage
sazā	10	0	peine
sāl	16	0	année
vidhi	0	254	loi
daṇḍit	1	446 (4,6)	puni
prayojan	0	181	utilisation
anveṣaṇ	0	155	enquête
sākṣī	0	112	témoin
sākṣya	2	283 (3,68)	preuve, témoignage
adhīn	0	1213	dépendant de
vicāraṇ	0	387	procès, jugement

Tab. 4.1: Termes exclusifs à la partition *Forums*. Fréquence et score de spécificité.

D'autres lexies apparaissent dans les deux partitions, pour lesquelles il existe des synonymes exclusivement dans partition non spécialisée. Par exemple :

lexème	lois / forums	spécificité (lois)	traduction
varṣa	107 / 13	1,0	année
nyāyālaya	1574 / 23	5,7	tribunal

Ces deux termes formés de composants sanskrits apparaissent dans les deux partitions, tandis que leurs synonymes *sāl* « année » et *adālat* « tribunal », aux étymons persan et arabe, n'apparaissent que dans la partition déspécialisée. Cette dichotomie se poursuit jusqu'aux mots grammaticaux ou aux locutions les plus courantes. Voici ceux typiques de la partition *Lois* :

unité	*Forums* (fréquence)	*Lois* (fréquence)	traduction
parantu	0	107	mais
kintu	1	59	mais
lekin	31	0	mais
tathāpi	0	8	(corrélatif) bien que
paścāt	0	228	après
bād	13	20	après
*yathā.**	0	19	selon (morphème de formation adverbiale)
ke anusaraṇ	0	34	suite à
prakār	4	255	manière
tarah	33	0	manière

4.1.4.2.2 Recherche des segments répétés dans les *Forums*

On tente, à travers la recherche des segments répétés, d'identifier des collocations signifiantes pour caractériser le contraste entre les deux sous-genres. Le seuil minimal de répétition est fixé à 5, tandis que la longueur de la séquence (en tokens) est limitée à 8, pour les raisons exposées plus haut dans ce chapitre (cf. 4.1.3).

La recherche des segments répétés n'apporte pas d'information intéressante quant à la phraséologie des textes des *Forums*. On y trouve une quinzaine de locutions, pour la plupart généralistes : *is kāraṇ* « pour cette raison », *ke ādhār par* « sur la base de, basé sur », lexicalement non significatives, à l'exception de *darj karvānā* « faire enregistrer ». De très nombreux autres segments sont des syntagmes verbaux courants.

4.1.4.2.3 Recherche des segments répétés dans les *Lois*

La recherche des segments répétés dans les textes de loi révèle une très forte fréquence de la locution postpositionnelle *ke adhīn*, littéralement « de-soumis », « en vertu de », qui apparaît 2 261 fois, et caractérise fortement ce genre : *dhārā ke adhīn* « en vertu de l'article », *saṁhitā ke adhīn* « en vertu du code », et l'unité plus longue

ke adhīn rahte hue « soumis à ». Ces unités phraséologiques ont pour correspondants sémantiques, dans les textes normatifs français, des séquences comme « en vertu de l'article », ou « en vertu des dispositions ».

Le passif est surreprésenté, avec des occurrences très nombreuses de participes accomplis suivis du verbe *jānā* conjugué, expression de la forme verbale passive : *kiyā jāegā* « sera fait » (587 occurrences), *kiyā gayā* « fut fait » (515 occurrences), *kiyā jānā* « être fait » (160), *samajhā jāegā* « sera considéré ». Ce trait est commun à toute la langue administrative.

L'individu sujet d'application de la loi est désigné par la collocation *jo koī* (pronom relatif + indéfini) « quiconque » (417 occurrences), ou le substantif tatsam (emprunté au sanskrit) *vyakti* « individu », spécifique du genre. On n'utilise pas ce dernier terme à l'oral, il n'appartient qu'au genre administratif et a été assigné sémantiquement à calquer le terme anglais « person ». Le pronom-adjectif indéfini *koi / kisi* (à l'oblique) « quelque, quelconque » entre dans la composition de très nombreuses collocations, indiquant la portée généralisante de l'énoncé (*jo koī*« quiconque », *kisī vyakti ko* « à toute personne », *kisī bhā̃ti kā* « de toute sorte »), caractéristique de la prescription normative.

On note enfin des collocations adverbiales concernant les circonstances d'infraction qui sont formées d'un substantif suivi de la postposition *se* qui indique l'instrumental : *ke āśay se* « avec l'intention de » (124 occurrences), *is āśay se* (même sens, 112 occurrences), *beīmānī se* « malhonnêtement, frauduleusement » (94 occurrences).

La différenciation des sous-genres forums et texte de lois s'opère par des choix lexicaux contraints et réservés, jusque dans le choix des mots grammaticaux. On observe que la partition déterminologisée est celle qui contient le plus d'occurrences de mots d'origine non sanskrite.

4.1.4.3 Contraste entre les articles de presse et les textes de loi en hindi

Les articles de presse reprennent des codes de la langue de spécialité judiciaire tout en conservant un fonds général non spécialisé. La partition *Presse* compte 363 397 tokens contre 223 404 pour les textes de loi. Le corpus de presse contient de nombreuses lexies désignant des catégories de personnes (femme, homme, jeune, étudiant, enfant, résident…) ou des lieux comme autant de protagonistes et de cadres spatiaux pour les faits divers décrits. Les articles comportent également des lexies se rapportant à l'incident, aux faits eux-mêmes, etc. Les analyses précédentes sont répétées sur ce corpus.

4.1.4.3.1 Recherche des termes exclusifs à une partition

Le tableau 4.2 présente les termes ayant une fréquence élevée qui n'apparaissent que dans l'une ou l'autre partition. Les chiffres indiquent la fréquence d'apparition.

terme	Presse	Lois	traduction
āropī	1452	0	accusé
sāl	635	0	année
lekin	503	0	mais
maut	432	0	mort
barāmad	318	0	retrouvé
duṣkarma	264	0	mauvaise action
kotwālī	248	0	quartier général de la police
talāś	248	0	recherche
korṭ	223	0	tribunal
urf	207	0	alias
khulāsā	191	0	révélation
mukadmā	188	0	affaire, procès
tahrīr	187	0	déposition écrite
sazā	176	0	peine
adālat	173	0	tribunal
rep	172	0	viol
kośiś	169	0	tentative
hirāsat	148	0	détention
inspekṭar	129	0	inspecteur
gaingrep	112	0	viol en réunion
kārit	0	416	fait
vidhi	0	409	loi
vicāraṇ	0	383	procès, jugement
vidhijña	0	377	juriste
daṇḍādeś	0	230	condamnation
paścāt	0	226	après

TAB. 4.2: Répartition des termes exclusifs aux partitions *Presse* ou *Lois*.

On retrouve dans la partition *Lois* la plupart des lexies déjà exclusives à cette partition en contraste avec les *Forums*. Cela signifie que la langue des textes normatifs se distingue à la fois de la langue de presse et de celle des forums. Ici encore, les lexies d'origine non sanskrite sont davantage représentées dans le corpus non spécialisé.

4.1.4.3.2 Recherche des termes spécifiques à une partition

Le tableau 4.3 indique que la langue de la presse comporte davantage de lexèmes d'origine non sanskrite parmi les lexèmes signifiants. Elle utilise également des termes indiquant un parti pris (« accusé », là où les lois parlent de *mis en cause*, « mauvaise action »).

terme	**Presse**	spécif.	**Lois**	spécif.	**traduction**
bād	1846	1000	20	-1000	après
pulis	4899	1000	255	-1000	police
thānā	734	148	2	-148	commissariat
māmlā	928	106	80	-106	affaire
sāmne	503	90	9	-90	devant
samakṣ	16	-93	275	93	devant
nyāyālaya	29	-1000	1574	1000	tribunal
adhīn	1	-1000	1378	1000	lié à
aparādh	60	-1000	1174	1000	infraction
majiṣṭreṭ	13	-1000	1014	1000	magistrat
kārāvās	26	-294	804	294	prison
avadhi	3	-272	666	272	durée
abhiyukta	13	-172	464	172	mis en cause

TAB. 4.3: Répartition par nombre d'occurrences des termes spécifiques à chaque partition (*Presse* ou *Lois*)

4.1.4.3.3 Recherche des segments répétés dans la langue de presse

7721 segments répétés d'une fréquence supérieure à 5 et d'une longueur comprise entre 2 et 8 tokens sont identifiés. On y retrouve une forte fréquence des locutions postpositionnelles liées au genre journalistique, dans lequel les citations prennent une place importante : *ke anusār*, *ke mutābik*, *ke ādhār par* « selon, d'après », « sur la base de ».

D'un point de vue lexical, le sous-genre de la presse se caractérise par des unités phraséologiques qui abordent trois champs sémantiques autour du fait-divers : la désignation des infractions, les actions entreprises par la police, et le lexique de la presse à sensation. Voici les unités phraséologiques les plus fréquentes :

Unités phraséologiques désignant des infractions

mārne kī dhamkī	(litt. « menace de tuer ») menace de mort
X kā ārop lagānā	accuser de X
golī mārkar hatyā	(litt.« meurtre par balle ») meurtre par arme à feu
farār ho jānā	s'enfuir

Unités phraséologiques relevant de la presse à sensation

māmlā sāmne āyā	l'affaire (de …) a été révélée
dil dahlā dene vālā / darindagī kā māmlā	affaire à fendre le cœur / terrible
haṛkãp mac gayā	se mettre à trembler de peur

Unités phraséologiques exprimant les actions de la police

jãc śurū karnā	commencer une enquête
giraftār karnā	arrêter
vārdāt ko anjām denā	(litt. « donner une conclusion à l'évènement ») mettre fin à l'évènement [avec deux lexèmes persans : *vārdāt* « évènement » et *anjām* « conclusion »]
mauqe par pahũcnā	(litt. « arriver sur l'occasion ») arriver sur les lieux
A ke khilāf māmlā / mukadmā darj karnā	enregistrer une affaire contre A
kī sūcnā milnā	être averti de
mRt ghoṣit karnā	déclarer mort
posṭmārṭam ke lie bhejnā	envoyer à l'autopsie
hirāsat mẽ lenā	interpeller, emmener en détention
kabze mẽ lenā	saisir
riporṭ darj karnā	rédiger un rapport / procès-verbal

Ces unités reprennent en partie le lexique des rapports de police (*riporṭ darj karnā* « rédiger un rapport », *kī sūcnā milnā* « être averti de », *mauqe par pahũcnā* « arriver sur les lieux »), avec une forte proportion d'unités lexicales et de locutions postpositionnelles d'origine arabe ou persane. On y trouve des synonymes de syntagmes plus sanskritisés présents dans le genre normatif (*abhirakṣā mẽ lenā* (norm.) / *hirāsat mẽ lenā* (presse) « arrêter, emmener en détention »). Le sous-genre de presse se situe à une place intermédiaire entre la langue des forums, où les participants écrivent selon leur compétence générale de la langue, et la presse, qui reprend des codes du rapport de police en y ajoutant des tournures qui en appellent à l'émotion du lecteur et manifestent un parti pris.

4.1.4.3.4 Conclusion

Les textes normatifs du hindi contiennent un important lexique d'origine sanskrite. Nous en avons expliqué les raisons au chapitre 1. Ce lexique est composé de termes calqués sur l'anglais et qui ne s'emploient pas naturellement à l'oral. À l'inverse, les unités polylexicales qui reflètent les usages déterminologisés, non normatifs, contiennent davantage de mots non sanskrits. L'usage de ces termes, forgés dans une langue de culture réinterprétée pour y tailler des calques de l'anglais, n'est pas à la portée de la compétence linguistique du non initié hindiphone. Il faudra envisager d'adapter la traduction avec des termes plus transparents, probablement plus ourdouisés, pour permettre aux non spécialistes d'accéder au savoir de ce domaine.

4.1.5 *Recherche des cooccurrents*

La recherche des collocations, et plus généralement des éléments de la phraséologie, passe par l'étude des cooccurrents. Les cooccurrents sont des unités textuelles qui se trouvent dans le voisinage d'une unité donnée à une fréquence supérieure à la fréquence attendue (Habert *et al.*, 1997). Pour calculer les cooccurrents ainsi que leur association au pôle (l'unité de référence, ou pivot), on peut prendre comme unité de contexte le paragraphe, la phrase ou définir une fenêtre de longueur fixe de k occurrences avant et k occurrences après le pivot[11]. On recherche la fréquence totale F du cooccurrent dans le corpus, sa fréquence locale f dans la fenêtre de contexte ainsi que son indice de spécificité[12] (Martinez, 2000).

La recherche des cooccurrents sera réalisée après l'alignement des corpus et l'extraction des étiquettes notionnelles. Elle permettra d'identifier les collocations et unités phraséologiques du genre judiciaire et, grâce à l'alignement multilingue des corpus, de trouver leurs équivalences ou correspondances traductionnelles.

[11] On évitera de parler de contextes gauche et droite, repères non pertinents dans le cas de l'ourdou. Le logiciel TXM, avec lequel les cooccurrents ont été recherchés, n'est d'ailleurs pas conçu pour inverser l'ordre d'affichage des contextes latéraux.

[12] Il existe d'autres mesures possibles, comme l'information mutuelle (IM), qui compare le nombre de cooccurrences observées au nombre de cooccurrences attendues sous l'hypothèse nulle (distribution aléatoire de mots dans le corpus), ou le taux de vraisemblance (log likelihood), qui évalue dans quelle mesure il est vraisemblable d'obtenir une observation par rapport à l'hypothèse nulle (Prochasson, 2009).

4.2 Alignement de corpus parallèles multilingues et équivalences traductionnelles

Au fil de la collecte des textes du corpus, nous avons constitué des sous-corpus de textes parallèles. Nous souhaitons les aligner pour obtenir des corpus parallèles multilingues (Kraif, 2006) ou multitextes (Harris, 1988) et repérer les traductions de termes et d'unités phraséologiques entre les langues, segments en relation d'équivalence traductionnelle[13]. Cette étape sera partiellement informatisée.

4.2.1 *Indices de surface et cognats*

Olivier Kraif définit l'alignement comme « [l]'opération consistant à mettre en correspondance les segments équivalents » (Kraif, 2006). Cette opération s'appuie sur des indices de surface comme la ponctuation, les nombres, la longueur des phrases ou la recherche de cognats (mots apparentés) pour établir des correspondances segment par segment. Dans notre cas, les textes parallèles multilingues sont des versions de textes de loi indiens en anglais, hindi et ourdou, la version anglaise étant l'original (cf. chapitre 1 1.1.1.2.1). Les cognats (mots apparentés) ne seront pas repérables entre les langues par les formes graphiques mais par le sens. Ce seront par exemple des noms de lieux, connus dans les trois langues (Inde, noms d'États ou de territoires) ou de fonction, empruntés à l'anglais et translittérés dans chaque écriture (« judge », « magistrate », « jail »).

4.2.2 *Exemple de recherche d'équivalence traductionnelle*

Le sens d'une phrase dépendant d'un contexte textuel, l'alignement des phrases ne produit pas des équivalences traductionnelles universelles. L'exemple qui suit présente une phrase extraite du code pénal indien (IPC) en anglais, hindi puis ourdou. Les relations d'équivalence traductionnelle sont signalées par les couleurs.

(54) …to use the lawful power of such public servant to the injury or **annoyance** of any person, […]
…aise lok sevak kī vidhipūrṇa śakti kā upayog kare jis upayog se kisī vyakti ko kṣati yā ***kṣobh*** *ho,* (…) [HI.]
…aise sarkārī mulāzīm ke jā'yaz ikhtiyār kā istaʿmāl kisī śakhṣ ko nuqṣān yā ***dukh*** *pahũcāne ke lie kare (…)* *[OU.] (art. 182-b)*

Le sens du verbe anglais *to use* est repris en hindi par deux occurrences du nom *upayog* « usage », dont une au sein d'un prédicat complexe avec le verbe *kare* « faire ».

[13] *Op. cit.*

Le but, *to the injury* en anglais, est traduit par deux structures syntaxiques différentes dans les autres langues : une relative indiquant la cause (*jis+se*), puis l'effet *kṣobh ho* (« nuisance soit ») en hindi, une locution postpositionnelle indiquant le but (*ke lie*) en ourdou précédée de la collocation *dukh pahūcānā* (« faire parvenir de la tristesse »). Ce seul exemple montre qu'il n'est pas possible d'aligner les phrases uniquement d'après leur structure syntaxique ou la fréquence du lexique.

En comparant cet exemple avec le suivant,

(55) …but knows it to be likely that in consequence of this information the police will make enquiries and institute searches in the village to the **annoyance** of the villagers or some of them.
…kintu vah yah jāntā hai ki yah saṁbhavya hai ki is ittilā ke pariṇāmsvarūp pulis us grām mẽ jā̃c karegī aur talāśiyā̃ legī, jisse grāmvāsiyõ yā unmẽ se kuch ko ***kṣobh*** *hogā.* [HI.]
…lekin jāntā hai ki us amar kā imkān hai ki is iṭṭila' ke natīje mẽ polīs us gāõ mẽ taḥqīqāt kare-gī aur talāśiyā̃ legī, jisse gāõ vālõ yā un mẽ se cand ko ***taqlīf*** *hogī.* [OU.] (art. 182-c)

on observe qu'en anglais et en hindi les mêmes lexèmes ont été utilisés pour désigner la nuisance (*annoyance*, *kṣobh*) tandis qu'en ourdou *taqlīf* (la difficulté, le dérangement) remplace *dukh* (la peine, la tristesse).

De nombreux logiciels d'alignement utilisent différentes techniques (indices de surface, recherche des cognats, affinement de la zone de recherche par des ancrages lexicaux ou méthode hybride) pour aligner automatiquement des langues sans qu'il soit nécessaire de les connaître. Les algorithmes sollicités nécessitent des corpus de grande taille (au moins un million de tokens dans chaque langue, (Kraif, 2006). Si nous ne disposons pas de multitextes aussi volumineux (les nôtres comptent tout au plus 100 000 tokens), la maîtrise des langues permet, après un alignement par paragraphe[14] puis par phrase, de repérer des points d'ancrage et de vérifier si l'alignement est correct.

4.3 Alignement de corpus comparables et recherche des équivalences traductionnelles

Lorsqu'il n'existe pas de corpus parallèle multilingue dans les combinaisons linguistiques souhaitées, l'on doit se tourner vers la recherche de ressources dans les corpus comparables. Comme vu dans la section 3.1.1.3, les textes qui les composent ne sont pas des traductions les uns des autres mais des textes similaires. Les critères de cette similarité sont variables et doivent être définis par le chercheur. Dans le cas de cette étude,

[14]Les paragraphes des textes de loi sont relativement nombreux et courts, ce qui offre une granularité assez fine.

les sous-corpus sont clairement établis par sous-genre du judiciaire : textes normatifs, juridictionnels ou parajuridiques. Pour débuter l'alignement, nous avons choisi des textes normatifs car, comme nous l'avons expliqué précédemment, ils contiennent des termes officiels qui serviront d'étiquettes notionnelles[15].

4.3.1 *Méthodes d'alignement de corpus comparables*

L'alignement de corpus comparables existe depuis les années 1990, notamment avec Rapp (1995) et la comparaison des motifs de cooccurrences : si dans un texte de langue l1 deux mots A et B cooccurrent plus souvent que par hasard alors dans un texte de langue l2 leurs traductions doivent également cooccurrer plus souvent. Cette méthode nécessite de gros corpus et peut être améliorée grâce à des *amorces linguistiques*, listes de mots déjà alignés qui vont contraindre les réorganisations possibles de chaque matrice en bloquant certaines combinaisons. Fung (1995) propose une méthode de mesure de l'hétérogénéité du contexte entre un mot et sa traduction dans un corpus non parallèle. Cette méthode permet la constitution de lexiques bilingues.

Enfin, la méthode de structuration des données en vecteurs de contexte (Fung, 1998) permet de comparer les contextes des termes. Chaque vecteur possède une tête, le terme, et des éléments cooccurrents lui sont associés dans une fenêtre de contexte. La mesure d'association entre ces éléments et le centre permet de détecter les collocations. Les vecteurs sont ensuite comparés deux à deux en utilisant des mesures de similarité[16]. Les vecteurs les plus proches sont candidats à la traduction.

Ces approches, qu'il est possible de combiner pour de meilleurs résultats, fonctionnent sur des corpus très volumineux car elles nécessitent un très grand nombre de calculs comparatifs. La méthode que nous avons utilisée repose sur une approche sémantique associée aux principes d'alignement déjà décrits pour les corpus parallèles multilingues. Après identification des thèmes dans des textes comparables, des points d'ancrage lexical permettent de réduire la fenêtre de recherche de l'équivalent de traduction. La recherche des cooccurrents, décrite ci-dessus, permet de détecter les collocations utiles au traducteur.

[15]En cela, la démarche adoptée se démarque de la théorie de Bourigault et Slodzian (1999), selon laquelle on doit abandonner toute la tradition de la terminologie qui s'appuie sur le classement des concepts. Nous pensons que ces étiquettes notionnelles, présentes dans le corpus, constituent des points d'ancrage utiles.

[16]Mesure du cosinus, de Jaccard ou d'Euclide.

4.3.2 *Alignement multilingue des codes pénaux indien et français*

Pour illustrer ce type d'alignement, nous avons tenté d'aligner les versions du code pénal français et du code pénal indien. Nous disposions pour cela de la version anglaise du code pénal français, ainsi que des versions anglaises, hindi et ourdou de l'Indian Penal Code (IPC). Quels que soient les pays, on suppose intuitivement que les thèmes abordés dans le domaine de la justice pénale seront similaires. Ces thèmes constitueront le point de départ de la recherche d'équivalents de traduction.

4.3.2.1 Alignement intermédiaire avec l'anglais pour langue pivot

Pour l'IPC, la version anglaise choisie est disponible sur le site officiel de l'OCDE[17] et la version hindi, qui est une traduction de l'original en anglais, provient du Département législatif (Legislative Department) du gouvernement central, chargé de traduire et de diffuser les textes de loi de l'anglais vers le hindi puis les autres langues de statut officiel. En ce qui concerne le Code pénal français (CP), l'original en français ainsi que sa traduction anglaise proviennent du site Légifrance. Dans la suite de cet exposé, le « Code pénal français » doit être compris comme le texte de loi en vigueur en France et non comme le texte en langue française.

Dans chaque version, les termes les plus fréquents et les plus spécifiques sont recherchés, à l'aide des méthodes exposées en 4.1.2 (extraction, calcul de l'indice de spécificité). Ils sont ensuite regroupés par thèmes, ou réseaux notionnels, et représentés par des étiquettes notionnelles qui indiquent les notions communes dans les deux textes. Les références des articles où apparaissent ces termes et notions sont notés. Ces articles ou passages d'articles sont ensuite alignés. Une fois ces éléments repérés, le texte anglais est remplacé par les versions française et hindi. Dans une phase ultérieure, les notions exclusives à l'une ou l'autre langue seront identifiées.

4.3.2.1.1 Extraction des unités lexicales les plus fréquentes

L'extraction se fait donc à partir de versions en anglais. Pour le CP, sur une liste de 5209 lexies dont on retire les chiffres, ponctuations, symboles, déterminants et autres mots outils, on obtient une liste de 806 lexies qui ont une fréquence de 10 occurrences ou plus. Il reste 680 formes pertinentes pour le champ pénal dont certaines sont les formes d'un lemme commun.

[17] Cf. `http://www.oecd.org/site/adboecdanti-corruptioninitiative/46814358.pdf`.

Pour la version anglaise du CPI, on obtient 4101 lexies, que l'on nettoie en fixant cette fois le seuil de fréquence minimale à 7 car le corpus indien compte 68 000 segments contre 101 000 pour la traduction anglaise du code pénal français. On obtient une liste de 634 formes uniques non lemmatisées.

Après comparaison des deux fichiers en anglais, on trouve 211 formes communes. On constate qu'il y a assez peu de vocabulaire commun pour des textes, certes issus de systèmes juridiques différents, mais traitant de mêmes thématiques. Cela peut s'expliquer par les variantes de l'anglais utilisé dans les deux versions : l'une est une variante pratiquée dans un contexte européen, tandis que l'autre est héritée d'un état de langue qui n'a pas changé depuis l'époque coloniale britannique.

4.3.2.1.2 Organisation de la structure textuelle

L'organisation du texte dans ces deux versions affecte la manière d'effectuer l'alignement ainsi que sa granularité. Ici, cette organisation est assez différente. On ne pourra pas aligner les textes à l'échelle du chapitre ou du paragraphe.

a. Le code pénal français (CP)

Entré en vigueur en 1994, le Code pénal français est régulièrement mis à jour et la version utilisée ici date du 12 avril 2019. Il se compose d'une partie législative, qui expose les infractions pénales et leurs peines, et d'une partie réglementaire concernant les décrets pris par le Conseil d'État. Chaque partie est subdivisée en sept Livres classant les infractions par domaine (les personnes, les biens, les infractions politiques, la guerre, autres domaines divers et dispositions pour l'Outre-Mer), à l'exception du premier Livre qui définit les principes d'application de la loi (limites de temps, d'espace et de responsabilité, nature des peines). C'est par exemple dans cette section que sont définies des notions comme la circonstance aggravante ou le sursis. Chaque livre est lui-même divisé en une hiérarchie de sous-parties jusqu'à l'élément le plus petit représenté par l'article.

b. L'Indian Penal Code (IPC)

La première version du Code pénal indien est entrée en vigueur en 1860. Elle a été amendée et révisée régulièrement, sa version actuelle datant de 2013[18]. L'IPC est divisé en 23 chapitres subdivisés en « sections » correspondant à des articles. Plusieurs

[18] Amendé par la Criminal Law (Amendment) Act, 2013.

articles peuvent être regroupés dans une même catégorie nommée qui ne portera pas de numéro. Par exemple, la catégorie « Du vol » regroupe les articles 378 à 382 du chapitre XVII. Contrairement au CP, l'IPC ne réserve pas une partie du texte pour définir les principes légaux. Il définit certains termes (chapitre II), les peines et exceptions (chapitre III et IV), mais l'incitation et la tentative se trouvent respectivement dans les chapitres V et XXIII. Le découpage des domaines d'infraction est également différent. On trouve des regroupements thématiques pour les atteintes au corps humain, à la propriété, des chapitres sur la contrefaçon, les infractions liées au mariage, à la contrefaçon, à la religion, à l'autorité publique, etc.

4.3.2.1.3 Répartition en réseaux notionnels

Le tableau suivant (4.4) présente un extrait de la comparaison de cette répartition thématique, avec quelques étiquettes notionnelles par thème. La première colonne indique les termes ou étiquettes notionnelles, la plupart tirés des lexies les plus fréquentes déjà listées. La deuxième colonne indique la référence à l'article dans le CP et la troisième la référence dans l'IPC.

a. Représentation des notions par des unités lexicales ou des réseaux de sémies

Lorsque les occurrences étaient trop fréquentes, les références n'ont pas toutes été citées car l'objectif était de repérer concrètement un ou plusieurs passages représentant textuellement une notion. Certaines notions dans une langue ne sont pas actualisées dans le texte similaire par des termes équivalents mais par un réseau de sémies (unités porteuses de sèmes apparentés à cette notion, de catégorie grammaticale différente, pouvant être associées à des tournures syntaxiques). En français, par exemple, les notions de « récidive » ou de « circonstance aggravante » sont représentées par un réseau de sémies. Autre exemple, la notion de détention, actualisée par les lexies « emprisonnement », « incarcération », « détention », ou en anglais des dérivés du lemme « imprison », est représenté au moins 657 fois dans le texte français et 575 fois dans le texte indien par des lexies du type « imprison.* »[19]. La notion de victime est représentée 94 fois par la lexie « victim » dans le Code pénal français et 16 fois dans l'IPC.

[19]Les caractères .* indiquent que l'on recherche toutes les formes commençant par *imprison*.

b. Appartenance à plusieurs réseaux notionnels

Pour certaines étiquettes notionnelles se pose le problème du classement dans un ou plusieurs réseaux. Par exemple la « minorité » est à la fois une circonstance aggravante en France lorsqu'elle concerne une victime et fait l'objet de dispositions particulières (concernant l'abandon ou la prostitution par exemple). C'est en revanche une circonstance atténuante pour la thématique de la sanction.

Parmi les concepts communs, il n'existe jamais une correspondance parfaite car, comme vu précédemment, le découpage du champ notionnel est intimement identifié à chaque culture, et deux cultures, fussent-elles proches, ont rarement une façon exactement similaire d'appréhender le réel.

Comme signalé plus haut, l'anglais utilisé en Inde n'est pas le même que celui de la traduction proposée sur Légifrance du Code pénal français. Il en a été tenu compte dans la recherche des notions communes. Par exemple, « abetment » était surreprésenté dans la traduction indienne, et « incitement » dans la française. Or les deux lexies renvoient à la notion d' « incitation » en français. Idem pour « wrongful detention », « confinement » (séquestration) ou « abduction » et « kidnapping » (enlèvement). La lexie « child » a été rattachée à « minor » comme désignation du concept de minorité.

c. Identification des unités lexicales et des thèmes

Au total, 108 unités lexicales dans les deux documents renvoyant à des notions comme les types d'infraction (vol, homicide, fraude, contrefaçon) ou les publics concernés (fonctionnaire, mineur, adulte) ont été identifiées.

On a recherché dans un premier temps la similitude dans la désignation des actes. Par exemple, si « robbery » (vol avec violence ou extorsion) n'a pas d'équivalent dans le CP français, la lecture des articles définissant ce terme nous apprend qu'il se trouve en partie pris en charge sémantiquement dans les circonstances aggravantes du vol (commis avec violence, menace, avec extorsion). Les thématiques les plus larges pourront ensuite être affinées en identifiant davantage d'unités lexicales ou notions grâce aux calculs des cooccurrences.

Les grands thèmes identifiés dans les codes pénaux sont les suivants :

1. **Types des peines**
   - règlement financier (paiement)
   - détention.
2. **Domaines d'infraction**
   - atteintes volontaires à la vie (types d'homicides volontaires)
   - atteintes involontaires à la vie (types d'homicides involontaires)

- atteintes à l'intégrité physique ou psychique de la personne (ce qui inclut le champ de la santé, les blessures volontaires ou involontaires, les agressions, la discrimination, la menace, la traite, l'enlèvement et la séquestration, le proxénétisme, l' « obscénité », la diffamation)
- crimes et délits contre les biens (vol, extorsion, corruption, fraude, blanchiment, destruction)
- mise en danger d'autrui (santé publique, négligence et délits routiers, actes de malveillance)
- cruauté envers les animaux.

3. **Circonstances de l'infraction**
   - incitation
   - tentative
   - intention
   - complicité.
4. **Protagonistes du système judiciaire**
   - types de victimes
   - mis en cause
   - jurés
   - magistrat
   - avocat.

Notions et étiquettes notionnelles	Code pénal français (CP, n° art.)	Code pénal indien (IPC, n° sect.)
	Type de peine	
sanction	Livre I, titre III, Ch 2 "Du régime des peines ».	Chapter III, sections 53 à 75
détention	657 occurrences (occ.) du champ sémantique	575 occ. du champ sémantique
détention à vie	131-1 (« types de peine ») (34 occ.)	53. "Punishments » (97 occ.)
probation	131-36-1 à 131-36-8, 221-9-1 (52 occ.)	–
	Domaine d'infraction	
atteinte volontaire	Livre II, titre II, Ch 1, section 1	sections 299 à 311
homicide	434-7	299
homicide volontaire	–	299
meurtre	221-1 (définition), 221-4 (meurtre, peine)	300 (définition), 302 (peine)
assassinat	221-3 (définition), 221-4 et 221-1-5 (incitation)	
atteinte involontaire	Livre II, titre II, Ch 1, section 2: atteintes involontaires à la vie	
	221-6 : homicide involontaire par maladresse / négligence	304A. Causing death by negligence
enlèvement	224-1 à 224-9	359 à 361. kidnapping 362 à 369. abduction 366A (procuration of minor girl)
viol	222-23 à 222-26	375, 376
- en réunion	pas d'occurrence, mais circonstance dans 222-24-6°	376D
aggression sexuelle	132-16-1, 132-45, 222-22, 222-27 à 222-32	–
harcèlement sexuel	222-33 à 222-33-1	354A
	Infractions contre les biens	
« robbery »	absent, mais occurrences de « vol aggravé"/ « extortion »	390. Robbery (theft / extortion) => variante du vol
« dacoity »	absent, mais occurrences de « vol aggravé"/ « extortion »	391. Dacoity (« robbery » par au – 5 personnes)
	Protagonistes du système judiciaire	
	victimes	
mineur	(110 occ.)	
	131-36-1 à 131-36-8 (suivi socio-judiciaire)	–
	222-14 (violence sur mineur)	–
	223-15 (abus de faiblesse)	–
	224-5 (enlèvement / séquestration)	361. Kidnapping from lawful guardianship
	223-13 (incitation au suicide d'un -)	305. Abetment of suicide of child or insane person
	225-7, 225-12 (prostitution de mineur)	366A, 372, 373

Tab. 4.4: Comparaison thématique des codes français et indien. Les mentions entre parenthèses donnent des informations contextuelles.

Les résultats montrent une inégale répartition des thèmes et des notions entre les

deux versions. Trois catégories de concepts se dégagent :

- les concepts communs aux deux codes,
- les concepts faisant l'objet d'articles distincts dans le code français mais pas dans l'IPC,
- inversement, les concepts faisant l'objet d'articles distincts dans l'IPC mais pas dans le code français.

Ces deux dernières catégories regroupent à la fois des concepts qui soit font l'objet d'une formalisation dans un ou plusieurs articles dans un seul texte et sont totalement absents du second (par exemple « peine de mort », « diffamation », dans l'IPC, « mise à l'épreuve », « suivi socio-judiciaire » dans le PC), soit sont formalisés dans un ou plusieurs articles dans un code et sont disséminés sous forme de réseau sémique à travers l'autre code (question de l'*incitation* dans les articles du CPI, circonscrite dans le Livre I du CP par exemple).

4.3.2.1.4 Alignement des versions française et hindi

Nous pouvons à présent remplacer la version pivot anglaise par la version version française du CP français et la version en hindi de l'IPC. Le tableau suivant (4.5) présente une sélection d'entrées pour lesquelles il existe une unicité ou du moins une grande proximité conceptuelle. Comme dans le tableau précédent, la première colonne rappelle la notion représentée par une étiquette notionnelle, la deuxième un contexte dans lequel cette notion a été repérée en français et la troisième un contexte pour cette même notion en hindi.

notion - étiquette notionnelle	**terme et contexte en français**	**terme et contexte en hindi**
sanction	Du régime des *peines*	*aparādhī is saṁhitā kē upbandhõ adhīn jin* daṇḍõ *sē daṇḍanīya hai vē yē haĩ [...]* « Les *peines* auxquelles s'expose les auteurs en vertu des dispositions du présent Code sont... »
détention à vie	La *réclusion criminelle* ou la *détention criminelle à perpétuité.*	*ājīvan kārāvās* « réclusion à perpétuité »
atteinte volontaire	Des *atteintes volontaires* à la vie	*jīvan kē liē* saṅkaṭakārī aparādhõ *kē viṣay mẽ* « Des infractions portant *atteinte à la vie* »
homicide	Le fait de receler ou de cacher le cadavre d'une personne victime d'un *homicide* ou décédée des suites de violences [...]	*aparādhik mānav vadh* « homicide volontaire »

notion - étiquette notionnelle	**terme et contexte en français**	**terme et contexte en hindi**
meurtre	Le fait de donner volontairement la mort à autrui constitue un *meurtre*. Il est puni de trente ans de réclusion criminelle.(def)	*ētasmin paścāt apavādit daśāō kō chōṛkar āparādhik mānav vadh* hatyā *hai yadi vo kārya, jiskē dvārā mr̥tyu kārit kī gaī hō, mr̥tyu kārit karanē kē āśay sē kiyā gayā hō, athvā […]* « Excepté dans les cas ci-après, l'homicide volontaire est un *meurtre* si l'action ayant causé la mort a été réalisée avec l'intention de cause la mort, ou […] »
	Le *meurtre* est puni de la réclusion criminelle à perpétuité lorsqu'il est commis : […] (sanction)	*jō kōī* hatyā *karēgā, vo mr̥tyu yā ājīvan kārāvās sē daṇḍit kiyā jāēgā aur jurmānē sē bhī daṇḍanīya hōgā* . « L'auteur d'un meurtre sera puni de mort ou de réclusion à perpétuité et également redevable d'une amende. » (sanction)
atteinte involontaire	Le fait de causer, dans les conditions et selon les distinctions prévues à l'article 121-3, par maladresse, imprudence, inattention, négligence ou manquement à une obligation de prudence ou de sécurité imposée par la loi ou le règlement, la mort d'autrui constitue un *homicide involontaire* puni de trois ans d'emprisonnement et de 45 000 euros d'amende.	upēkṣā dvārā mr̥tyu kārit karnā- jō kōī utāvalēpan sē yā upēkṣāpūrṇa kisī aisē kārya sē kisī vyakti kī mr̥tyu kārit karēgā, jō āparādhik mānav vadh kē kōṭi mẽ nahĩ ātā, *vo dōnō mẽ sē kisī bhā̃ti kē kārāvās sē, jiskī avadhi dō varṣ tak kī hō sakēgī, yā jurmānē sē, yā dōnō sē, daṇḍit kiyā jāēgā* . « *Homicide par négligence - Quiconque cause la mort d'une personne par un acte imprudent ou négligent ne relevant pas de l'homicide volontaire* sera puni d'une peine de détention allant jusqu'à 2 ans, ou d'une amende, ou des deux. »
enlèvement	Le fait, sans ordre des autorités constituées et hors les cas prévus par la loi, d'arrêter, d'*enlever*, de détenir ou de *séquestrer* une personne, est puni de vingt ans de réclusion criminelle.	vyapaharaṇ – vyapaharaṇ *dō kism kā hōtā hai ; bhārat mẽ sē* vyapaharaṇ *aur vidhipūrṇ saṁrakṣaktā mẽ sē* vyapaharaṇ. « *Enlèvement* - Il existe deux types d'*enlèvement*: à destination hors de l'Inde, ou pour soustraire à la garde d'un gardien légal. »
harcèlement sexuel	Le *harcèlement sexuel* est le fait d'imposer à une personne, de façon répétée, des propos ou comportements à connotation sexuelle ou sexiste qui soit portent atteinte à sa dignité en raison de leur caractère dégradant ou humiliant, soit créent à son encontre une situation intimidante, hostile ou offensante.	laiṅgik utpīṛan *aur* laiṅgik utpīṛan *kē liē daṇḍ (…) śārīrik saṁsparś aur agrakriyāẽ karnē , jinmẽ avāñchanīya aur laiṅgik sambandh banānē sambandhī spaṣṭ prastāv antarvalit hō ; yā laiṅgik svīkr̥ti kē liē kōī mā̃g yā anurōdh karnē ; yā kisī strī kī icchā kē viruddh aślīl sāhitya dikhānē ; yā laiṅgik ābhāsī ṭippaṇiyā̃ karnē vālā puruṣ laiṅgik utpīṛan kē aparādh kā dōṣī hōgā.* « *Harcèlement sexuel* sanction pour *harcèlement sexuel* (…) contact physique et avances impliquant des allusions sexuelles déplacées et explicites ; ou une demande ou une requête pour des faveurs sexuelles ; ou le fait de montrer des images pornographiques à une femme contre sa volonté ; ou le fait de faire des commentaires connotés, constitue un délit de *harcèlement sexuel.* »

Tab. 4.5: Alignement des version françaises et hindi des codes pénaux

On observe d'après ce tableau des variations dans la relation d'équivalence de traduction entre le français et le hindi. Dans certains cas, le parallélisme est total : les segments sont de même longueur dans les deux cas et l'on parle de *compositionnalité* (Kraif, 2006), comme pour « peine » / *daṇḍ*. Des variations peuvent cependant intervenir comme pour « réclusion criminelle à perpétuité » /*ājīvan kārāvās* (4 tokens contre 2, motifs : NN ADJ PREP NN contre ADJ NN). Ces variations peuvent prendre une très grande amplitude comme pour la notion d'atteinte involontaire à la vie, exprimée en français par la collocation « homicide involontaire », et en hindi par une phrase de définition.

Ces premiers résultats nous ont permis de constituer un lexique bilingue. La version ourdou du code étant un texte comparable déjà aligné avec les versions hindi et anglaise, il est possible d'obtenir des séquences contextuelles quadrilingues (français, hindi, ourdou et anglais) pour des entrées lexicales avec des motifs simples ou polylexicaux. La figure ci-après (4.1) présente la recherche d'un motif dans la version bitexte français-hindi sur la plateforme en ligne de l'équipe de recherche ER-TIM. Le lexique, organisé en entrées multilingues avec des définitions, permet d'indiquer à l'utilisateur les équivalents dans les autres langues, ainsi que leur définition.

FIG. 4.1: Recherche d'un motif dans un corpus comparable aligné

4.3.2.1.5 Extraction de définitions grâce à l'alignement multilingue

Autre développement de l'alignement multilingue, celui de l'extraction de définitions semi-automatisée à travers l'observation des segments répétés.

L'alignement précédent met en évidence les équivalents notionnels, les éléments de sens qui deviendront les équivalents de traduction fonctionnels. Avec le relevé des segments répétés, il met aussi à jour les différences stylistiques et la phraséologie propres aux deux langues. En voici deux exemples. Dans le premier, la recherche d'un motif sémantique commun permet de révéler la phraséologie :

(56) *aisē ādatan aparādhī jō aparādh kē bal par hī apnī ājivikā calātē (...)*
tels criminels par habitude REL à la force du crime EMPH son gagne-pain mènent

La notion de « criminel par habitude, ou criminel habituel » sonne étrangement à l'oreille du traducteur français, d'ailleurs une telle expression est inconnue du texte de loi français. En cherchant des motifs à partir du lemme *habituel* ou *habitude*, on trouve cependant trois tournures sémantiquement proches : « infraction commise de manière habituelle » / « se livrant habituellement à ... » / « personne commettant habituellement ». S'inspirant de cette phraséologie, un équivalent de traduction en français de la phrase hindi pourrait être :

> ' Personne dont les infractions commises *de manière habituelle* constituent l'unique moyen de subsistance.'

Le second exemple montre comment l'alignement permet d'extraire des définitions et d'enrichir un lexique. Dans la phase de recherche des segments répétés, le trigramme suivant a été repéré :

(57) का द्योतक है
kā dyotak hai
de désignateur est
« désigne »

Il apparaît 15 fois dans le corpus. Dans le concordancier, qui permet d'observer dans un contexte resserré les occurrences autour d'une forme ou d'une séquence donnée, la forme शब्द *śabda* en est un cooccurrent à faible distance, avec une cofréquence de 12 selon la séquence :

(58) *[terme]* शब्द *[définition]* का द्योतक है ।
[terme] śabda [définition] kā dyotak hai
[terme] mot [définition] de-désignateur-est

« Le mot [terme] désigne [définition]. »

Un court script permet ainsi de récupérer de manière semi-automatique un terme et une définition …

[terme]= [définition]

… pour enrichir un lexique monolingue qui pourra être mis en relation avec le lexique de la seconde langue de travail. Le script peut être amélioré pour repérer les définitions qui contiennent des relatives et n'en extraire que le terme et sa définition. Nous avons pu récupérer de cette manière quelques dizaines de termes et de définitions en hindi, ourdou et anglais afin de les comparer à des définitions similaire ou proches en français. Ces lexiques et définitions permettent de localiser les notions et d'affiner l'alignement de corpus comparables. Ces opérations d'alignement des textes et d'extraction du lexique amènent à effectuer un va-et-vient constant entre le texte, le contexte et le mot.

Dans la suite de notre exposé, nous présenterons les résultats de ces allers-retours constants mettant en valeur la phraséologie et le lexique du genre judiciaire. Auparavant, il est utile de rappeler l'état de la recherche et des outils disponibles pour le traitement des langues hindi et ourdou.

L'approche analytique que nous avons adoptée repose à la fois sur le corpus et est guidée par lui. En effet, nous repartirons des thèmes et concepts développés dans les procédures pénales et les textes normatifs pour dégager un réseau terminologique représenté par des candidats-termes, puis des collocations et des phraséologies, qui constituent le cœur et la difficulté de l'idiome de la langue judiciaire. L'approche déductive et manuelle est complétée par l'alignement des sous-corpus multilingues parallèles et comparables de manière thématique. Au sein de ces sous-corpus, les candidats-termes et leurs équivalents dans les différentes langues sont repérés, puis leur cooccurrents et les unités phraséologiques qui leur sont liées. Grâce à cette méthodologie, nous avons mis en place les bases du développement d'outils pratiques pour le traducteur-interprète, le juriste, le justiciable ou le travailleur associatif (4.1) : une application numérique qui permet de rechercher des termes dans des textes alignés multilingues et d'en afficher les équivalents dans les autre langues. Elle permet aussi au traducteur de vérifier dans un concordancier si le terme qu'il souhaite employer l'est à bon escient dans le contexte donné.

4.4 Étude du lexique et de la phraséologie

Dans les sections précédentes de ce chapitre, nous avons abordé la constitution du corpus (3.2) et ses objectifs (3.1.2) dans l'aide à la traduction. Nous avons présenté à travers la chaîne de traitement les différentes opérations, plus ou moins longues, nécessaires à la préparation des données pour leur analyse (3.1). Nous avons privilégié l'approche textométrique (4.1.2), qui permet des allers-retours entre le texte et le motif lexical, dans nos démarches d'alignement multilingue de corpus parallèles ou comparables (3.1.1.3). Nous avons extrait des listes de candidats-termes et de segments répétés. Enfin, nous avons dégagé les principaux thèmes des codes pénaux. Dans la sous-partie suivante, nous allons regrouper ces thèmes et ceux de la procédure judiciaire, puis leur associer des étiquettes notionnelles dans chaque langue. Autour de ces étiquettes s'agrégeront des candidats-termes, concrets, tirés des corpus. Nous étudierons pour un certain nombre d'entre eux leurs cooccurrents pour en tirer des unités phraséologiques (collocations, expressions figées, formules rhétoriques) du domaine judiciaire, que nous pourrons comparer thématiquement dans les trois langues de travail. Cette dernière section présente les résultats à l'issue de toutes ces étapes, avec l'analyse du lexique et de la phraséologie des sous-genres du judiciaire.

Une langue n'est pas une somme d'étiquettes qui identifient des concepts. Elle est une matière vivante, à travers des textes. Tout un ensemble de traits lexicaux présents dans ces textes caractérisent le genre textuel en dehors du champ terminologique : le lexique non terminologique qui appartient toutefois au lexique du genre, les marques d'énonciation, les marqueurs, qui structurent le discours ainsi que la phraséologie (Tutin, 2007). Sans ces éléments textuels, les termes ne pourraient prendre pleinement leur sens. Dans le corpus de cette étude, on distingue ainsi des éléments du lexique et de la phraséologie qui appartiennent à la langue générale, d'autres communs avec la langue administrative, et d'autres encore spécifiquement judiciaires. C'est toutefois l'association de ces trois niveaux qui caractérise le genre textuel judiciaire.

La phraséologie est un aspect à la fois incontournable et épineux de la traduction. Elle rassemble des séquences polylexicales déterminées (collocations, expressions figées, formules routinières,« tours de phrase »). À la différence d'une séquence de mots prise au hasard, les composants qui forment l'unité phraséologique apparaissent davantage que la moyenne en cooccurrence. Nous employons ici le terme d'unité phraséologique tel

qu'il a été défini par Zinglé (2003)[20] : « Par unité phraséologique nous entendons tout groupement de mots récurrent (expression, locution ou construction au sens large) auquel les locuteurs […] associent un sens précis, qui n'est pas nécessairement lié au sens de ses composantes prises individuellement. »

Le caractère protéiforme de la phraséologie la rend difficile à détecter par les traitements informatiques, d'autant qu'elle varie selon les domaines. En droit, par exemple, la conjonction « que » ou le pronom enclitique « en » sont très présents dans les formations phraséologiques (Khalifa, 2015), alors qu'ils seraient considérés comme de simples mots grammaticaux hors du champ de la langue de spécialité dans d'autres domaines (ex : « attendu que » suivi de la conjonction « que » en anaphore, « en tout cas depuis temps non prescrit », « en l'espèce », « en vigueur », « en vertu de ».

Après la phase d'extraction, il est difficile de dresser automatiquement une typologie des unités phraséologiques (Pecman, 2004). Leur composition morphosyntaxique et leur longueur varient, elles peuvent exercer des fonctions rhétoriques propres au genre textuel, marquer l'articulation du discours, être formées totalement ou partiellement de termes, constituées d'éléments contigus ou distants. En traduction, le statut d'unité phraséologique se constate aussi par l'impossibilité de traduire mot à mot ses éléments de la langue source vers la langue cible (Pecman, 2004).

À l'intérieur de la phraséologie, les collocations constituent des associations récurrentes et typiques de mots (Seretan *et al.*, 2003). Une collocation n'a pas de sens figuré, contrairement à une expression idiomatique (« tirer à boulets rouges », « filer un mauvais coton »). Ses composants ont chacun une signification hors de la composition et ne sont pas nécessairement contigus. L'un d'entre eux est la tête, aussi appelé base ou noyau, de la composition. Une collocation peut subir différentes transformations selon son degré de figement.

Malgré ces critères, Wehrli *et al.* (2009) rappellent que la limite entre une collocation et un autre type d'association aléatoire de mots reste floue et qu'aucune théorie linguistique n'est encore parvenue à théoriser ce type d'association de manière satisfaisante :

> « Les collocations se révèlent, par des observations de nature empirique dérivées grâce à l'analyse des corpus de texte, en tant qu'associations conventionnelles de mots établies par l'usage dans la langue. »

En nous appuyant sur ce constat, nous avons relevé des collocations propres au domaine judiciaire et d'autres types d'unités phraséologiques en observant les cooccurrents

[20]Définition reprise par Pecman (2004).

des candidats-termes. Cette observation a eu lieu d'une part au fil de la compréhension de la procédure judiciaire (termes désignant des lieux, des personnes, des types de documents ou des actions), et d'autre part lors de l'exploitation des thèmes alignés entre des textes comparables de même genre (exemple des codes pénaux français et indien).

Enfin, les éléments d'énonciation caractérisent également le genre : ils situent le cadre spatio-temporel, le statut de l'énonciateur, du co-énonciateur et l'objet d'énonciation du texte.

4.4.1 *Classement notionnel des lexèmes et unités phraséologiques (phraséolexique)*

Comme nous l'avons vu précédemment, l'analyse textométrique offre la possibilité d'extraire, au moyen de méthodes statistiques, les mots ou les segments les plus récurrents (ceux dont la fréquence rapportée au nombre total de mots ou de segments est la plus élevée) ou les plus spécifiques (ceux dont la fréquence d'occurrence est plus élevée dans un corpus donné par rapport à un corpus de référence). Ils permettent de dresser la terminologie d'un corpus.

Dans le cas de l'étude phraséologique, ces traitements présentent des points négatifs. Ils excluent des unités phraséologiques de faible fréquence, pourtant significatives du genre (Pecman, 2004). Ils ne permettent pas non plus de classer automatiquement les unités phraséologiques repérées de manière satisfaisante. En effet, les critères pour dresser une typologie de ces unités sont nombreux :

- la fréquence,
- la composition (Cornu, 1990): longueur de l'unité polylexicale, catégorie des éléments morphosyntaxiques –ou famille morphologique (Tutin, 2007)–, discontinuité,
- la fonction : locative, déterminative, rhétorique…,
- la structure syntaxique,
- le degré de figement,
- la transparence (Pecman, 2004) base et extension de l'unité phraséologique ont des sens différents hors de leur association (opacité : « mise en état de la procédure »), leur sens en cooccurrence est interprétable (transparence : « renvoi d'une affaire »),
- le contenu sémantique ou classement notionnel (Pecman 2007) : les unités phraséologiques sont classées selon leur lien à des étiquettes notionnelles qui renvoient à des sphères notionnelles.

Cette dernière catégorisation, sémantique, est très intéressante pour un rédacteur ou pour un traducteur-interprète qui approchera l'unité phraséologique en contexte par son

sens. Nous l'avons retenue, en y apportant quelques modifications, dans la méthodologie décrite ci-après.

4.4.1.1 Méthodologie

La méthodologie choisie pour classer l'ensemble du lexique et de la phraséologie (le phraséolexique (Pecman, 2004)) et les rendre accessibles au traducteur-interprète part du lexique. C'est le point de départ d'une approche notionnelle, champ de recherche ouvert par Van Campenhoudt (1996) avec le projet d'établissement d'un réseau notionnel interlinguistique (R.N.I.) qui tient compte de la diversité du découpage notionnel des langues pour proposer des équivalences notionnelles.

Précédemment, nous avons extrait les lexèmes les plus fréquents du corpus par langue, puis filtré en fonction de leur pertinence (phase d'extraction terminologique, cf. 4.1.2). Les unités lexicales décrivant la procédure pénale, que le traducteur-interprète doit connaître pour comprendre la discipline de spécialité, sont ajoutés à la liste, permettant ainsi de récupérer des unités phraséologiques de faible fréquence. Le corpus n'étant pas exhaustif, des termes peu fréquents peuvent en effet présenter un intérêt pour la traduction. D'autre part, un terme peu fréquent dans une langue peut correspondre à une unité lexicale plus fréquente dans une autre en raison des différences de représentation de la procédure judiciaire. Dans d'autres cas, on s'apercevra que les différentes sémies d'un mot correspondent à différentes sphères thématiques (par exemple « arrêt » renvoie à un type de décision dans « arrêt de la cour d'appel », ou aux mesures et sanctions dans « mandat d'arrêt »).

Les termes-candidats ainsi sélectionnés désignent des notions regroupées en réseaux notionnels. Nous avons identifié 6 de ces réseaux rattachés à la procédure pénale : « infractions », « lieux de la justice », « instruction », « procès », « mesures et sanctions », « protagonistes ». Chaque réseau notionnel est constitué de notions représentées par des étiquettes notionnelles (termes mots-clés) qui prennent dans le texte la forme d'unités lexicales ou phraséologiques (éléments du phraséolexique), cf. figure 4.2. Les réseaux étant interconnectables, certains termes, en fonction des sèmes mis en avant dans le contexte, peuvent entrer dans plusieurs d'entre eux et nous avons dû faire des choix (cf. 4.2. L'indexation des termes et des notions permet un renvoi vers ces réseaux, leur liste n'est pas exhaustive.

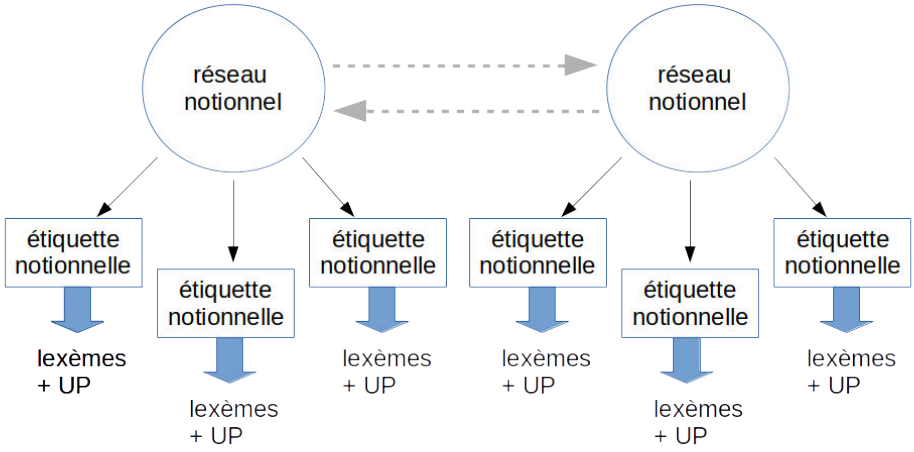

FIG. 4.2: Classement du lexique et des unités phraséologiques à partir des réseaux notionnels

4.4.1.1.1 Les réseaux notionnels

Par définition, la branche du droit pénal définit quelles sont les infractions à la loi et leurs peines. Deux réseaux notionnels y seront dévolus.

- les **infractions**: ce réseau reprend les notions liées aux infractions et à leurs circonstances (intention, complicité, armes).
- les **mesures et sanctions**: ce réseau groupe les mesures de prévention, de sûreté (détention, contrôle judiciaire), les condamnations et l'absence de sanction (relaxe, acquittement, etc.).
- l'**instruction**: ce réseau regroupe les notions liées à la procédure allant du signalement d'une infraction jusqu'aux portes du procès au tribunal.
- la **mécanique du procès**: ce réseau regroupe les notions liées à la procédure judiciaire dans la salle d'audience du procès, jusqu'au prononcé de la décision.
- les **lieux de la justice**: ce réseau répertorie les lieux de la justice en rapport avec la procédure judiciaire, du signalement de l'infraction jusqu'à la condamnation. Dans le cas des tribunaux, la désignation des lieux diffère selon le degré du système.
- les **protagonistes**: ce réseau concerne les notions renvoyant aux justiciables comme aux professionnels ou auxiliaires de justice. Alors que les parties (défense et accusation) conservent des fonctions pérennes, leurs désignations varient au cours des étapes de la procédure (plaignant, partie-civile, victime / prévenu, mis en cause, mis en examen, accusé, détenu…).

français	hindi	ourdou	anglais
nuisance, gêne	क्षोभ *kṣobh*	ایذا، تقلیف، نقصان، دکھ *dukh, nuqṣān, taqlīf, aiz̲ā*	annoyance
blessure	क्षति *kṣati*	ضرر، خضرت *kẖiżrat, żarar*	injury
perte	हानि *hāni*	نقصان *nuqṣān*	loss
dommage	अपहानि *apahāni*	نقصان *nuqṣān*	harm

Tab. 4.6: Lexèmes actualisant la notion « préjudice »

Ce travail de découpage notionnel est effectué dans les trois langues de travail.

4.4.1.1.2 Étiquettes notionnelles et unités lexicales

Chaque réseau notionnel regroupe des notions incarnées dans les textes par des termes et des unités phraséologiques. Les notions devant être verbalisées pour que le classement des termes soit compréhensible, nous reprenons la convention graphique consistant à les faire apparaître entre deux barres (ex : |atteinte à la vie| désigne l'étiquette notionnelle (Pecman, 2004). *atteinte à la vie* désigne le terme qui l'incarne). Les notions sont formulées dans les trois langues de travail pour permettre une recherche à partir de n'importe quelle des trois langues et ne pas imposer un découpage notionnel aux autres.

En effet, à une notion correspondent différents lexèmes en fonction des langues et des contextes. Pour reprendre l'exemple vu en 4.2 avec l'alignement de phrases où apparaissait *annoyance* en anglais et ses traductions en hindi et en ourdou, voici un tableau (4.6) qui récapitule les différents termes des codes pénaux indien et français construisant un réseau sémique autour de la notion de |préjudice|.

Il est donc nécessaire, pour tenir compte de ces découpages notionnels, de réserver une phase au travail lexicologique monolingue avant de comparer les résultats de ces travaux monolingues entre les langues.

Dans les différents réseaux, les termes ont été regroupés selon les affinités sémantiques en sous-domaines. Ainsi, dans les |infractions|, on trouvera les qualifications, les circonstances, les armes, etc.

Tous les termes sont indexés par langue avec un renvoi vers le réseau notionnel dans lequel ils apparaissent (cf. index de la terminologie en annexe C). Un tableau synoptique tente de rapprocher les notions et les termes entre les langues. Lorsqu'une notion n'existe pas dans une langue, elle est signalée entre parenthèses.

4.4.1.1.3 Collocations et unités phraséologiques

Dans un troisième temps, les cooccurrents des termes indexés sont recherchés afin d'identifier les unités phraséologiques dans lesquelles ils apparaissent. Ici, le terme constitue la forme-pôle à laquelle se rattachent des cooccurrents. Les cooccurrents retenus, calculés dans une fenêtre de contexte de 12 tokens, sont sélectionnés sur deux critères : leur indice de spécificité (lié à leur cofréquence d'apparition avec le terme principal, il signale que le terme apparaît spécifiquement avec le terme donné) et la distance qui sépare le cooccurrent de sa forme-pôle (la base, ou tête, de la collocation). Plus le cooccurrent est proche de celle-ci, avec un indice de spécificité élevé, plus l'unité phraséologique est susceptible de constituer une collocation restreinte. Si la distance augmente, la séquence a davantage de chances de constituer une tournure de phrase ou de marquer l'association thématique des deux éléments, sans communauté de syntagme[21]. Seuls quelques cooccurrents et unités phraséologiques ont été retenus pour chaque terme, pour des raisons matérielles.

Dans le cas du hindi et de l'ourdou, langues dans lesquelles les prédicats complexes, sous forme de locutions verbo-nominales (nom + verbe) ou verbo-adjectivales (adjectif + verbe), abondent, nous n'avons que rarement retenu les séquences les plus prévisibles, celles constituées par les verbes supports « faire » et « être ». Nous avons signalé les séquences dont les verbes supports sont plus rares et peuvent poser problème au traducteur-interprète.

La correspondance multilingue s'arrête au niveau des notions et termes, charge à l'utilisateur des tables d'aller, grâce aux renvois, d'une langue à l'autre pour trouver les équivalences d'unités phraséologiques.

4.4.1.1.4 Présentation des résultats

Les résultats se présentent sous la forme de tableaux.

– Dans le premier (4.4.1.1.4), les candidats-termes sont classés suivant les six grands réseaux : infractions, lieux de justice, instruction, mécanique du procès, mesures et sanctions, protagonistes. Ces candidats-termes se déclinent en trois langues[22]. Ce

[21]Par exemple dans le corpus français, l'unité « délit » a pour cooccurrent le plus plus spécifique « crime » (indice 1000), et une faible distance les lie (2,5). Cela s'explique par le fait que les deux unités sont des hétéronymes d'« infraction », et qu'elles sont distinguées dans la plupart des cas où elles cooccurrent : la séquence « crime ou délit », enrichie d'autres éléments, ne forme pas une unité phraséologique intéressante pour la traduction.

[22]Nous avons parfois eu recours à l'anglais pour accéder à des notions du système indo-pakistanais, toutefois

résultat reprend l'indexation des termes.

– La seconde série de tableaux (Annexe, section D) comporte pour chaque forme-pôle ses cooccurrents les plus fréquents et les plus liés. Nous retenons pour ces cooccurrents fortement liés les concordances les plus fréquentes, mettant ainsi en lumière les expressions et unités phraséologiques dans lesquelles le traducteur est le plus à même de trouver ces termes.

Dans le tableau synoptique (4.4.1.1.4), les termes entre parenthèses sont des traductions de termes d'autres langues qui ne relèvent pas de la terminologie de la langue cible. Par exemple, « désagrément » n'a pas d'occurrence dans les passages en langue spécialisée des textes français du corpus.

Les termes en hindi ou en ourdou signalés par un astérisque relèvent de la langue spécialisée normée et demandent une adaptation pour un public plus large car leur utilisation est limitée à la rédaction des textes normatifs ou juridictionnels.

cette langue reste un pivot et nous ne lui avons pas dressé de tableau.

français	hindi	ourdou
	Infractions	
acte	कार्य	فعل
fait	तथ्य	امر واقعہ
fait	तथ्य	واقعات؛واردات
crime	अपराध	جرم
délit	अपराध	جرم
contravention	अपराध	جرم
	Préjudice	
nuisance, (gène, désagrément)	क्षोभ*	،ایذا،تقلیف،نقصان،دکھ
perte	हानि	نقصان
blessure	क्षति*	ضرر،خضرت
dommage	अपहानि*	نقصان
	Domaine d'infraction	
homicide	मानव वध	قتل انسان
meurtre	हत्या ; कत्ल|कतल	قتل عمد
assassinat	(पूर्वचिन्तन* के साथ हत्या)	(پیش بین کے ساتھ قتل عمد)
extorsion	लूट ; उद्दापन*	سرقہ بالجبر
violence	हिंसा	تشدد
agression	आक्रमण* ; हमला	حملہ
harcèlement	तंग करना ; उत्पीड़न	اذیت
viol	बलात्संग * ; बलात्कार , रेप	زناءبالجبر؛زیادتی؛ریپ
(acte de malveillance)	रिष्टि*	نقصان رسانی
(assassinat lié à la dot)	दहेज मृत्यु	جہیز موت
(maltraitance liée à la dot)	दहेज प्रताड़ना	(جہیز سے متعلقہ تشدد)
menace	धमकी	دھمکی
enlèvement	व्यपहरण* ; अपहरण	بھگالے جانا؛اغوا
séquestration	परिरोध* ; उठाकर कैदी रखना	جس بیجا
proxénétisme	उपापन* ; दिलाई करना	دلائی
diffamation	मानहानि*	ہتک عزت کی
(obscénité)	अश्लील* (adj.)	فحش(adj.)
falsification	मिथ्याकरण*	جعل نویسی
contrefaçon	कूटरचना* ; कूटकरण*	جعل سازی؛جعلی بنانا
terrorisme	आतंकवाद	دہشت گردی
complot ; entente	षड्यंत्र*	سازش
	Circonstances	
incitation	दुष्प्रेरण*	اعانت(nf)
tentative	प्रयत्न	اقدام ;کوشش
vol	चोरी	سرقہ ;چوری
volontaire	साशय *	باالارادہ
préméditation	पूर्वचिन्तन*	پیش بین
intentionnellement	इस आशय से	ارادتاً
intention	आशय	نیت
consentement	सम्मति	رضامندی
flagrant délit	रंगे हाथ	رنگے ہاتھ
armé	—	مسلح
	Prévention	

français	hindi	ourdou
commissariat	थाना	تھانہ
poste de police	थाना	چوکی؛تھانہ
tribunal	न्यायालय	عدالت
cour d'appel	(अपीली न्यायालय)	(اپیلی عدالت)
maison d'arrêt	कारावास	جیل؛قید
prison	कारावास	جیل خانہ؛قید
juridiction	अधिकारिता	اختیارات
(haute cour)	उच्च न्यायालय	عالیہ
(cour suprême)	सर्वोच्च न्यायालय / उच्चतम न्यायालय	عظمہ
cour de cassation		

Tab. 4.8: Réseau notionnel *lieux de justice*

français	hindi	ourdou
interdiction	प्रतिषेध*	ممانعت؛امتناع
prévention	निवारण*	روک تھام
Armes		
arme	हथियार	ہتھیار؛اسلح
couteau	चाकू	چاقو
(grand couteau)	छुरी	چھری
pistolet	पिस्तौल	پستول/پسٹل
dague	किरपाण	کرپان
fusil	बन्दूक	بندوق
voleur	डाकू	ڈاکئ
trafiquant	तस्कर	تسکر

Tab. 4.7: Réseau notionnel *infractions*

français	hindi	ourdou
interpeller	गिरफ़तार करना	گرفتار کرنا
enquêter	अन्वेषण* करना	تحقیقات*
enquête	अन्वेषण* ; जाँच	تحقیقات*
élément	पता	معلومات
instruire	अन्वेषण* करना	تحقیقات*
instruction	अन्वेषण*	تفتیش
affaire	मामला ; मुकदमा	معاملہ؛مقدمہ
témoignage	साक्ष्य*	گواہی؛شہادت
déposition	अभिसाक्ष्य* ; तहरीर	تحریر*
indice	पता	سراغ

français	hindi	ourdou
preuve	साक्ष्य	ثبوت؛شہادت
mis en cause	अपराधी	ملزم
prévenu	अपराधी	ملزم
CR	अनुरोध–पत्र*	*مراسلہ استدعا
CRI	(अंतर्राष्ट्रीय अनुरोध–पत्र*)	*مراسلہ استدعا
mandat	वारण्ट	وارنٹ
plaignant	परिवादी* / परिवादनी*	مدعی؛سائل
témoin	साक्षी	گواہ
témoin assisté	(जज के सामने वकील के साथ प्रस–तुत साक्षी)	(ایسا گواہ جو جج کے سامنے اپنے وکیل کے ہمراہ پیش ہو)
mis en examen	अपराधी	ملزم
qualification	(अपराध का दंड संहिता में पाया जा–ना)	(مجموعہ تعزیرات میں واقعات کا جرم پایا جانا)
aide juridique	कानूनी / विधिक सहायता	قانونی امداد
aide juridictionnelle	(न्यायालय की ओर से कानूनी सहा–यता)	(عدالت کی جانب قانونی امداد)
déposant	अभिसाक्षी*	بیان دینے والا
réquisition	(पुलिस का निवेदन)	(پولیس کی طلب)
enquête préliminaire	प्रारंभिक जाँच	ابتدائی جانچ
plainte	परिवाद*	*استغاثہ
(acte d'accusation , réquisitoire)	आरोप पत्र	فردِ جرم
(traitable sans mandat du tribu-nal)	संज्ञेय*	قابلِ دست اندازی*
(non traitable sans mandat)	असंज्ञेय*	ناقابلِ دست اندازی*
(rapport d'enquête préliminaire)	चालान	چالان

Tab. 4.9: Réseau notionnel *instruction*

français	hindi	ourdou
détention provisoire	—	(عارضی حراست)
garde à vue	परिरोध* ; हवालात	پولیس کی حوالات
contrôle judiciaire	—	(عدالت کی دی ہوئی پابندیاں)
caution	ज़मानत ; जमानत	ضمانت
demande de libération	(छुड़वाने का अनुरोध)	رہاء کی درخواست
détention	कारावास* ; परिरोध* ; निरोध*	قید
(réclusion à perpétuité)	आजीवन कारावास*	عمر قید
réclusion	कारावास*	قید

français	**hindi**	**ourdou**
peine de mort	मृत्युदंड*	سزاِموت
verdict	निर्णय	فیصلہ
jugement	निर्णय	فیصلہ
décision	निर्णय	فیصلہ
arrêt	निर्णय	فیصلہ
relaxe	दोषमुक्ति* ; बरियत	بریت
acquittement	दोषमुक्ति* ; बरियत	بریت
relaxé	बरी	بری
acquitté	बरी	بری
amende	जुर्माना	جرمانہ
sursis	—	–
TIG	—	–
Non lieu	—	–
peine	दंड ; सज़ा	سزا
libération	(छुड़वाना) ; रिहाई	رہائی
(détention avec travaux forcés)	कठिन कारावास*	قیدبامشقت*
condamnation	दोषसिद्धि*	جرم کی سزا
condamnation (décision)	दंडादेश*	–
(détention à l'isolement)	एकांत परिरोध*	قیدتنہائی*
arrestation	गिरफ्तारी / गिरफ़्तारी	گرفتاری
interpellation	गिरफ्तारी / गिरफ़्तारी	گرفتاری
prison	जेल	جیل

TAB. 4.11: Réseau notionnel *mesures et sanctions*

français	**hindi**	**ourdou**
procès	(वाद का विचारण*)	مقدمہ؛ سماعت
affaire	वाद	مقدمہ؛ معاملہ
audience	सुनवाई	سماعت ؛ سنوائی کی تاریخ
déposition	अभिकथन*	بیان
accusé	अभियुक्त*	ملزم
défense	प्रतिवाद*	ملزم
la défense	प्रतिवादी*	ملزم
accusation	अभियोजन* [पक्ष]	مستغیث فریق
débat	बहस	بحث
plaidoirie	मौखिक बहस	غذرداری* ؛(فریق کی جانب وکیل کا بیان)
(requête en justice)	अर्ज़ी	*عرضی؛ درخواست؛ استدعا
Partie-civile		(مدعی)
décision	निर्णय	فیصلہ
poursuite	(विधि कार्यवाही)	قانونی کارروائی
(suivi d'une affaire)	पैरवी	پیروی
(juger une affaire)	विचारण* करना	(مقدمہ کا فیصلہ کرنا)
procédure	कार्यवाही	کارروائی
réquisition	(अभियोजक* के मौखिक तर्क)	(سرکاری وکیل کی قائل کرنے والی تحریر)
partie	पक्ष	فیریق
huis clos	बन्द कमरे में सुनवाई	بند کمرے میں سماعت
ministère public	लोक अभियोजक*	سرکاری وکیل؛(مقادمات میں ریاست کا نمائندہ)

TAB. 4.10: Réseau notionnel *mécanique du procès*

français	**hindi**	**ourdou**
juge	न्यायाधीश*	جج
juge d'instruction	(अन्वेषण न्यायाधीश* ; जाँच करनेवाला जज)	(تفتیش کنندہ جج / جانچ چلانے والا جج)
juge des libertés et de la détention	(छोड़ने या निरुद्ध* करनेवाला मजिस्ट्रेट ; छोड़ने या हिरासत में रखनेवाला मजिस्ट्रेट)	(رہائی اور حراست کا جج)
magistrat	मजिस्ट्रेट	مجسٹریٹ
policier	पुलिस / पुलिसवाला	پولیس / پولیس والا
(chef de commissariat)	थाना बरसाधक	تھانہ انچارج
(magistrat chargé des pouvoirs conférés par l'art. 30 du CP)	धारा 30 मजिस्ट्रेट	دفعہ 30 نجسٹریٹ
procureur de la République	लोक अभियोजक*	سرکاری وکیل ;(مقادمات میں ریاست کا نمائندہ)
procureur	अभियोजक*	سرکاری وکیل ;(مقادمات میں ریاست کا نمائندہ)
avocat	अधिवक्ता*, वकील	وکیل
avocat général	लोक अभियोजक	سرکاری وکیل ;(مقادمات میں ریاست کا نمائندہ)
expert	विशेषज्ञ*	ماہر
greffier	(मजस्ट्रेट का सहायक)	(عدالتی رجسٹرار)
huissier	(सम्मन पहुँचानेवाला अधिकारी ; सुनवाई पर न्यायालय की सहायता करनेवाला कर्मचारी)	(سمن پہنچانے والا ; سماعت پ پر عدالت کی مدد کرنے والا کارکن)
auxiliaire de justice	(न्याय से संबंधित कार्य करनेवाला सहायक)	(انصاف سے متعلقہ کام کرنے والا)
officier de police	पुलिस अधिकारी* / पुलिस अफ्सर	پولیس افسر

Tab. 4.12: Réseau notionnel *protagonistes*

4.4.1.1.5 Mise à disposition des résultats en ligne

Les applications en ligne évoquées en 3.1.2.3 et développée par François Stuck, ingénieur de recherche de l'équipe ERTIM (Équipe de Recherche Textes, Informatique, Multilinguisme), reprennent ces résultats pour les appliquer à des textes[23].

D'un côté, une application de « maintenance » permet de déposer les documents à fouiller et les lexiques.

Un répertoire *Gestion de document de recherche* conserve une collection de textes alignés ou monolingues (cf. 4.3). De ces alignements ont été extraits des termes et correspondances lexicales qui ont été enregistrés dans des lexiques quadrilingues (français, hindi, ourdou, anglais) au format xml.

Ces lexiques sont enregistrés dans un second répertoire, *Lexiques multilingues*. Les entrées peuvent être complétées par des données sur la catégorie morphosyntaxique, utiles pour désambiguïser un terme, le genre, la forme pluriel et une définition. Cette définition permet, dans le cas où le terme n'a pas de traduction satisfaisante dans une autre langue, de comparer les définitions de termes proches pour extraire une équivalence en contexte.

Dans une seconde application en ligne, baptisée *Occurrencier* (cf. 4.4), l'utilisateur entre sa requête (les expressions régulières sont permises) et la langue de recherche. Il choisit le texte à fouiller ainsi que le lexique qu'il souhaite utiliser (il est ainsi possible d'appliquer des lexiques plus ou moins spécialisés).

L'application recherche le motif dans le texte parallèle. Ces occurrences du motif sont comparées à des lexiques bilingues. Par exemple, le motif हत्या *hatyā* « meurtre » est recherché dans le code pénal indien trilingue hindi-anglais-ourdou. Ce motif est parallèlement recherché dans des lexiques bilingues hindi-anglais et hindi-ourdou. Les équivalents de *hatyā*, s'ils sont trouvés, sont mis en couleur dans le texte parallèle. Ici, *qatl* en ourdou et *murder* en anglais sont repérés. Un survol avec le curseur permet d'accéder à des informations supplémentaires comme la catégorie morphosyntaxique et la définition.

Dans le cas d'un texte multilingue, le résultat de la requête est présenté avec son contexte textuel. Lorsque le résultat d'une requête n'existe pas dans les lexique cas, l'utilisateur dispose au moins d'occurrences en contexte multilingue, ce qui lui permet de déduire des correspondances traductionnelles.

---

[23] Lien vers l'application *Occurrencier multilingue*, encore en cours de développement, mais fonctionnelle toutefois : `http://er-tim.fr/~stuck/BPA/occurencesMultilingues/ajaxJQ/index.php`.

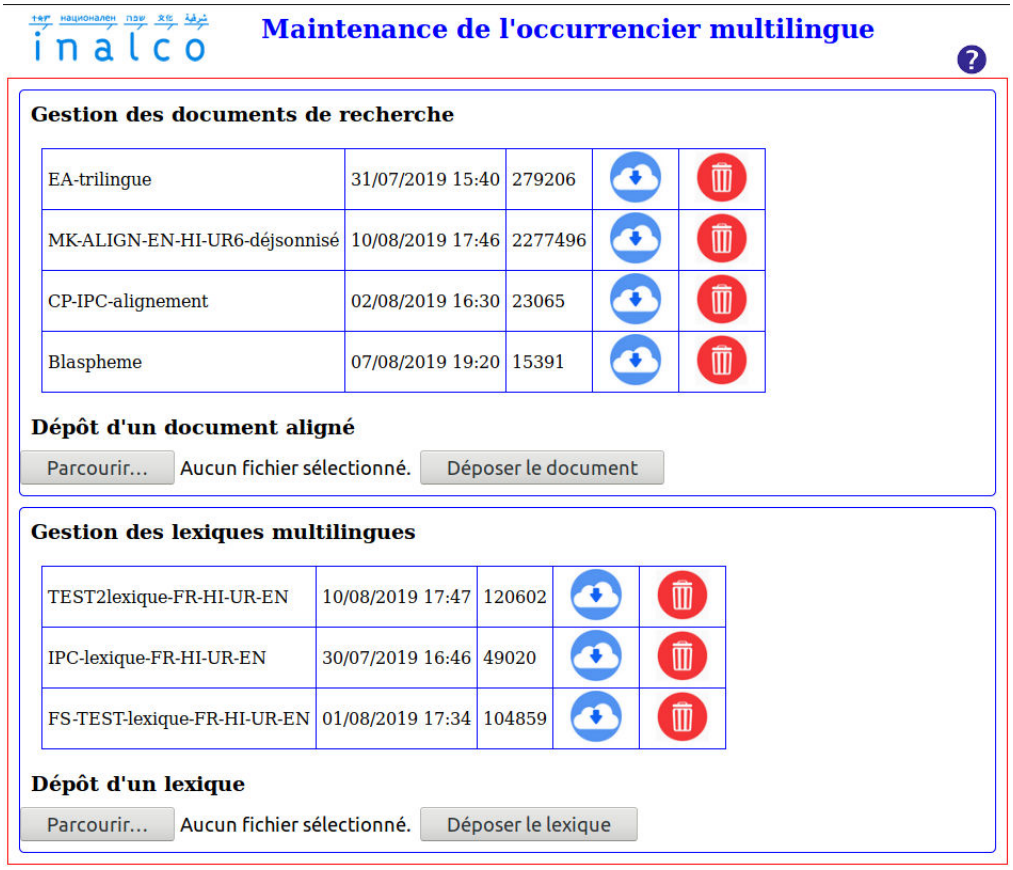

FIG. 4.3: Application *maintenance* pour le dépôt les documents à fouiller et les lexiques

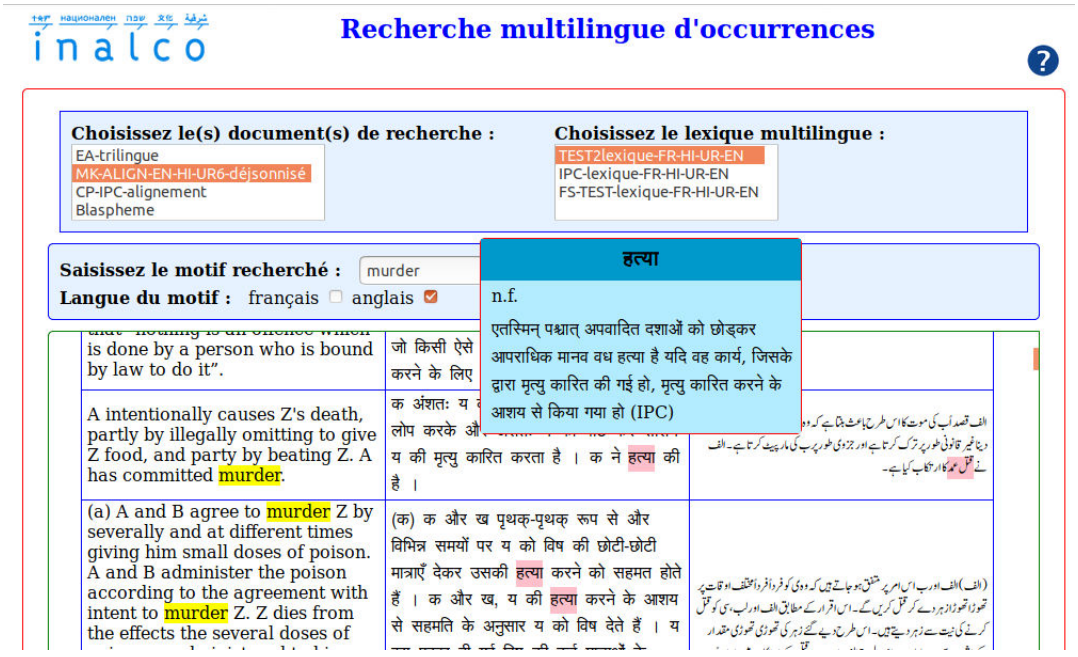

FIG. 4.4: Fonctionnement de l'*Occurrencier* pour la recherche de la traduction du motif *murder*.

L'indexation des termes identifiés au paragraphe précédent permet de compléter ces lexiques, de multiplier les ancrages lexicaux multilingues et d'ainsi faciliter la mise en relation des équivalences multilingues.

Un développement possible de cette application consiste à interroger un texte monolingue et obtenir des résultats pour les équivalences dans les autres langues.

Les occurrences des éléments du phraséolexique sont observables dans leur contexte d'énonciation grâce à un concordancier. Le traducteur-interprète peut aussi interroger une base de données contenant des textes alignés bilingues ou multilingues avec ces éléments afin de faire ressortir leurs équivalents dans les autres langues. Si ceux-ci n'ont pas été identifiés dans la base de données, le traducteur-interprète pourra toutefois observer le contexte d'apparition de l'élément recherché dans des contextes similaires dans les autres langues.

L'analyse du phraséolexique ne peut reposer entièrement sur la textométrie. Une lecture manuelle est nécessaire pour classer les lexèmes candidats en notions, puis en réseaux notionnels. Dans un second temps, les analyses textométriques permettent de repérer les cooccurrents, ces mots qui apparaissent de manière déterminée en lien avec un lexème candidat.

4.4.2 *Analyse des résultats et équivalences traductionnelles*

Les résultats présentés, constitués de ressources lexicales et d'alignement de textes, forment la base matérielle pour la recherche d'équivalences traductionnelles. On y observe plusieurs tendances.

Tout d'abord, comme dans tout travail de traduction, des variations du découpage notionnel apparaissent clairement d'une langue à l'autre, même entre les variantes hindi et ourdou, issues d'une même origine. Ces différences conceptuelles obligent le traducteur à adapter sa stratégie de traduction en fonction du contexte, afin de proposer une équivalence notionnelle, fonctionnelle, ou à défaut un emprunt à la langue cible suivi d'une périphrase explicative. Nous illustrons ce fait à travers quelques exemples (4.4.2.1).

D'autre part, ces résultats font ressortir les variations lexicales et phraséologiques entre les sous-genres du judiciaire et le genre parajuridique, qui parle de la justice dans une langue accessible aux non initiés. Si ces variations existent dans les trois langues, elles imposent pour de nombreuses unités lexicales ou phraséologiques en hindi et ourdou une dichotomie entre le genre de spécialité, associé à une langue standardisée et artificialisée

en partie commune au genre plus vaste de l'administratif, et les genres déterminologisés. En d'autres termes, on n'emploie pas les mêmes mots pour dire la même chose selon que l'on se place dans un genre judiciaire ou parajuridique, sans que cela influe sur le registre, qui n'est pas nécessairement plus familier dans le genre parajuridique (4.4.2.2).

Ces constatations sont illustrées par quelques exemples ci-après, à travers lesquels on commentera le choix du traducteur par rapport au genre du texte, à la fonction de celui-ci et au destinataire de la traduction.

4.4.2.1 Recherche d'équivalences traductionnelles

Dans le transfert de sens entre deux langues, trois cas se présentent au traducteur :

- le parallélisme des notions est total dans les langues source et cible, les catégories grammaticales correspondent : le transfert est simple et direct entre les langues ;
- les champs notionnels ne se recouvrent pas totalement, cas fréquent : on doit alors envisager, en fonction du contexte, si un équivalent fonctionnel existe dans la langue cible (une correspondance lexicale de sens proche ou contenant des sèmes présents dans l'unité à traduire). Nous avons vu au chapitre 2 qu'il était courant en jurilinguistique, comme en terminologie en général, de redéfinir les limites des concepts pour leur assigner des étiquettes stables dans des situations précises. Cette pratique est utilisée notamment dans les pays où la législation est multiple (comme au Canada) ou dans le contexte européen, avec l'apparition d'une langue juridique contrôlée où les termes représentent des concepts dépourvus des spécificités linguistiques et culturelles des pays membres. Cette démarche permet de trouver un consensus dans les désignations. Se retrouvent aussi dans cette catégorie les notions qui s'expriment par des structures syntaxiques ou des catégories grammaticales différentes (par exemple un substantif d'un côté pour une locution verbale de l'autre) d'une langue à l'autre.
- la notion n'existe pas dans l'une des langues. En ce cas, le traducteur dispose de la création néologique, déconseillée dans le cadre juridique si elle n'est pas validée par une norme, du calque ou de l'emprunt de la langue source, appuyé, au moins à la première occurrence, d'une périphrase définitoire.

4.4.2.1.1 Correspondances parallèles

Les correspondances parallèles sont celles pour lesquelles les unités à traduire correspondent tant par le sens que par la forme (longueur, catégorie morphosyntaxique). Il en ressort une série, d'après les résultats du tableau de synthèse (4.4.1.1.4), parmi

lesquelles :

- dans les infractions : le terme générique infraction, l'homicide, la violence, l'agression, la menace, la contrefaçon, le terrorisme ;
- les circonstances : intentionnalité, types d'outils, tentative ;
- pour l'instruction : la détention, le terme générique pour la décision, l'amende, la peine, l'arrestation ;
- dans les lieux de la justice : le tribunal, le terme générique désignant la prison ;
- dans le déroulement du procès : les parties, la procédure ;
- concernant les protagonistes, on retrouve les principales catégories : magistrat, policier, avocat, mis en cause, tant que l'on n'entre pas dans les subtilités des différentes procédures.

Ces correspondances parallèles représentent un socle important de termes relativement génériques, décontextualisés. Pour peu que l'on doive cibler un contexte précis, une procédure ou un degré de juridiction, ces désignation ne suffisent plus en raison d'une différence de découpage notionnel de ces champs, à l'intérieur desquels les réseaux de notions se restructurent autour du socle culturel du judiciaire. Il faut alors trouver, selon le contexte, des équivalents notionnels ou fonctionnels.

4.4.2.1.2 Correspondances partielles

Les exemples ci-après illustrent les différences de découpage notionnel entre les trois langues. Le premier aborde la question du contenu sémique différent dans les correspondances multilingues. Un exemple avait déjà été donné en 4.2 avec « préjudice », montrant la difficulté d'aligner des termes au niveau lexical sans contexte. Le premier des exemples ci-après montre les variations de sèmes contenus dans les termes *procès*, *samāʿat*, *mukadmā / muqadmah* et *vicāraṇ*, qui renvoient tous au |procès|. Le second exemple aborde le cas de la gestion l'hétéronymie entre les langues.

a. La notion de procès

La notion de procès illustre les différences de découpage notionnel entre le hindi et l'ourdou, bien que ces langues aient une origine commune.

En français, la notion de |procès| couvre la procédure qui se tient devant un tribunal en vue d'une décision. Elle est définie par la « difficulté de fait ou de droit

soumise à l'examen d'un juge ou d'un arbitre »[24]. Le lexème *procès* est, dans le corpus francophone, plus représenté dans le sous-corpus parajuridique, et particulièrement dans la presse (466 occurrences contre 39 dans les textes de loi) qui lui donne, en plus du sens évoqué ci-dessus, le sens d'« audience » (ex : « Ce nouveau procès, qui doit se tenir devant la cour d'appel (…) ». En français, le lexème intègre les sèmes de « procédure » et d'« évènement », actualisés en fonction des contextes.

En ourdou, la notion est représentée par les unités lexicales *samāʿat*, relativement spécialisée, et *muqadmah*, connue du grand public. Les deux lexies désignent l'audience au tribunal, mais seul *muqadmah* possède le sème de « procédure » dans sa totalité. Ce terme apparaît d'ailleurs dès le début de la procédure, à l'enregistrement de l'affaire par la police, dans le rapport mentionné dans le FIR. Ce n'est pas le cas de *samāʿat*, qui n'est présent que dans la partie qui se déroule au tribunal, et plus précisément en audience devant un juge.

En hindi, en revanche, c'est le processus de jugement d'une affaire devant le tribunal, et non l'évènement « audience », qui porte un nom : *vicāraṇ*, équivalent de l'anglais « trial », à partir duquel se construit le prédicat complexe *vicāraṇ karnā*, équivalent de l'anglais « to try », juger une affaire au tribunal, et utilisé selon le même schéma d'arguments : « to try a suit », *vād kā vicāraṇ karnā* (« faire le jugement d'une affaire »). Il n'y a pas de désignation en hindi recouvrant la notion française ou ourdou, et le processus qui se déroule à l'audience implique en hindi le sème « jugement », absent dans les autres langues.

Le lexème *vicāraṇ* appartient à la langue de spécialité tandis que *muqadmā*, qui désigne à la fois l'affaire et le procès, existe aussi en hindi et sera compréhensible du plus grand nombre. De plus, *muqadmā* conserve en hindi les mêmes sèmes qu'en ourdou (procédure et évènement). L'utilisation de ce dernier lexème sera plus aisée en raison de sa catégorie grammaticale similaire à celle de son équivalent français. Si l'on souhaite rester dans le genre spécialisé en hindi, il faudra adapter la syntaxe de la phrase pour y introduire le prédicat complexe ou le substantif et son complément. On notera aussi que |*muqadmā*| couvre une notion plus étendue que |procès| en français, puisqu'il part de l'ouverture de la procédure jusqu'à la décision du juge. On pourra, hors du cadre du tribunal, le traduire en français par « affaire, information, dossier ».

Réciproquement, le fait de juger une affaire au tribunal sera simplement traduit

[24] *Lexique des termes juridiques*, Dalloz, 2015.

en hindi comme en anglais par un substantif, alors qu'il faudra recourir à une périphrase en français et en ourdou.

b. L'hétéronymie

Nous avons vu au chapitre 2, lors de la description de la procédure pénale (2.2), que le français avait recours à des hétéronymes pour désigner aussi bien des fonctions que des actes selon le type de procédure (criminelle ou correctionnelle), le degré (première instance ou appel), ou les individus concernés (un témoin est *auditionné*, un prévenu *interrogé*, un *avocat général* et un *procureur de la République* remplissent la même fonction de représentation du ministère public dans des tribunaux différents). On utilisera en hindi-ourdou des termes génériques (cf. tableaux en annexe, en particulier D.6.).

français	hindi	ourdou
prévenu, mis en cause, accusé	*aparādhī*	*mulzim*
procureur de la République, avocat général, procureur général	*lok abhiyojan*	*sarkārī vakīl*
réclusion, détention	*kārāvās*	*qaid*
décision, arrêt, jugement, verdict	*nirṇaya*	*faiṣlah*

L'hétéronymie se trouve renversée en ce qui concerne les différents types de requêtes en hindi et en ourdou, ainsi que les différentes formes de présentation de la partie lésée, selon qu'elle signale son litige à la police ou au tribunal, conformément au système anglais (*plaintiff* / *petitioner*). Dans tous ces cas, le français utilisera le terme *plaignant* à l'ouverture de l'instruction, puis *accusation* / *partie civile* lors de l'audience au tribunal. *L*e champ de la |détention avant jugement| se caractérise par une hétéronymie en Inde et au Pakistan, entre *rīmāṇḍ*, *havālāt* (HI.)/ *ḥavālāt* (OU.), *nirodh* (HI.)/*ḥirāṣat* (OU.). *rīmāṇḍ*, de l'anglais « remand », désigne la détention décidée par le magistrat dans les 24 heures suivant l'arrestation d'un prévenu. Il n'y a pas de distinction claire entre les autres termes, qui peuvent renvoyer à la *garde à vue* ou à la *détention provisoire* du système français.

c. Le procureur de la République

En ourdou, le terme équivalent à « procureur de la République » pose difficulté dans la langue de spécialité. Le terme emprunté à l'anglais *prāsekyūṭar* (« prosecutor ») est attesté dans le corpus parajuridique et dans le juridictionnel, mais pas dans le

normatif. Une recherche dans un dictionnaire[25] nous mène à *sarkārī vakīl*, littéralement « gouvernemental avocat », terme attesté dans le corpus non spécialisé. Or ce terme est souvent utilisé en France par les interprètes pour désigner l'avocat commis d'office. En hindi, *sarkārī vakīl* désigne un représentant des intérêts d'un service public, pas forcément de la société dans son ensemble.

Une courte périphrase, reprise du dictionnaire, permet de lever l'éventuelle ambiguïté : *muqādmāt mẽ riyāsat kā numāindah* « représentant de l'État dans les affaires » apparaît alors plus approprié.

D'autres exemples de correspondance notionnelle partielle sont présentées dans les tableaux en annexe ainsi que dans le tableau synoptique des termes.

4.4.2.1.3 Absence de correspondance

a. Les catégories de magistrats

Les systèmes indiens et pakistanais ont recours à divers types de magistrats inconnus en France (de première, seconde ou troisième classe, exerçant des compétences spéciales au titre de l'article 30 du code pénal, etc.). Le traducteur-interprète doit envisager sa stratégie de traduction en fonction du contexte. Un calque peut suffire pour les magistrats de classe, une périphrase devrait pouvoir qualifier la fonction du « section 30 magistrate ».

Pour le greffier et l'huissier, nous avons également choisi de recourir à des périphrases expliquant leur rôle. Lorsque ces désignations apparaissent près de leur signature, nos recourons à un calque ou un équivalent fonctionnel plus court qu'une définition (« officier délivrant la citation pour l'huissier », « assistant du magistrat pour le greffier »).

Lors de la recherche d'un équivalent de traduction du *juge des libertés et de la détention* en hindi s'est posé le problème du choix des lexèmes car cette fonction du magistrat n'existe pas dans les systèmes indiens et pakistanais. *Liberté* se dit *āzādī* dans un registre ourdouisé en hindi, *svatantratā* dans un registre sanskritisé. Mais ces lexèmes ne conviennent pas dans le contexte judiciaire où l'on traite de détention ou de libération, tandis qu'*āzādī* et *svatantratā* renvoient à l'indépendance, à l'autonomie. Dans le corpus, les mots renvoyant à la sortie de détention sont les verbes *choṛnā* et *churvānā* (« laisser » et « faire abandonner », mais aussi « libérer » et « faire libérer »). Aucun substantif

[25]Dictionnaire anglais-anglais et ourdou (Basit, 2015).

n'est utilisé dans la langue administrative pour *libération*, et le terme qui exprime la détention, *nirodh*, est équivoque pour le grand public[26]. On trouve également à registre égal le syntagme verbal *niruddha karnā* « détenir, maintenir en détention ». Pour ce niveau de langue spécialisé, on obtient ainsi le terme polylexical *choṛne yā niruddha karne vālā majisṭreṭ*. Dans une variété de langue moins spécialisée, il est possible de puiser dans le fonds arabo-persan, comme le fait la presse judiciaire, en utilisant les substantifs *rihāī* et *hirāsat*, termes que nous retiendrons également en ourdou, pour les deux variétés de langue : *rihāī aur hirāsat/ḥirāsat karne vālā majisṭreṭ*.

b. La notion d'|assassinat|

Le hindi et l'ourdou ne possèdent pas de terme pour la notion française que recouvre *assassinat*. En français, on distingue par ce terme le *meurtre* (homicide volontaire) du *meurtre avec préméditation* (l'action a été pensée à l'avance, la volonté n'est pas née au moment du meurtre). Une telle notion n'est pas désignée dans le code pénal indien, ni en hindi, ni en ourdou, qui utilisent respectivement les termes *hatyā* et *qatl*, correspondant à *meurtre* en français. La circonstance de préméditation existe cependant dans ces deux langues (*pūrvacintan* « pré-réflexion » en hindi et *peś bīn* « pré-vision » en ourdou), permettant de forger des équivalents de traduction pour la notion d'*assassinat* par formation de collocations. Si ces extensions collocationnelles paraissent trop opaques pour un destinataire moyen, on leur substituera un coverbe qui précise l'antériorité d'un fait par rapport à un autre, et par conséquent la circonstance (*sockar* ou sa forme oralisée *socke*, « ayant réfléchi »).

c. Conclusion sur les équivalences notionnelles

Cette classification des types de correspondances est poreuse, il existe tout un continuum entre les degrés de correspondance, tout comme il en existe dans les variations du découpage notionnel. Ces correspondances sont d'ailleurs directement liées au problème des différences de genre textuel, comme l'a montré l'exemple de traduction du procureur de la République, ainsi que les exemples de la section suivante.

4.4.2.2 Prise en compte du destinataire, langue de spécialité ou parajuridique

On l'a vu précédemment, le genre judiciaire se caractérise par une langue de spécialité, mais aussi par un lexique du genre véhiculé dans les textes parajuridiques. Ce

[26]Dans la langue populaire, ce mot désigne un préservatif.

dernier constitue le fonds de connaissance en matière juridique et judiciaire du destinataire non initié.

Les résultats de l'analyse du corpus montrent que pour désigner une même notion, certains lexèmes sont utilisés dans le parajuridique (presse, forums, vulgarisation)[27] et d'autres dans les textes normatifs ou juridictionnels, comme dans l'exemple de |procès|.

Ce constat est d'autant plus manifeste dans les textes en hindi et en ourdou, où l'artificialisation de la langue administrative et de spécialité judiciaire opère une distinction très nette dans l'utilisation des lexèmes renvoyant à une même notion selon le sous-genre. Alors qu'elle tolère dans la plupart des cas un degré de perméabilité, cette distinction peut se révéler, comme en témoigne l'un des exemples ci-après. Ce phénomène renvoie directement à la sensation d'opacité générée par le choix du lexique dans le genre textuel.

4.4.2.2.1 Opacité dans les textes normatifs

On entend par transparence et opacité le fait que le lien qui unit un terme et son concept soient plus ou moins facilement décodables par le destinataire du message. Le lecteur ou le locuteur, qui ne perçoit pas de lien entre la motivation et la forme d'un terme, ne peut accéder à son sens. Dans le cas de la langue de spécialité en hindi, et plus généralement de la langue administrative, de nombreux termes, on l'a vu au chapitre 1 (1.3.1), ont été forgés par calque de traduction à partir de termes anglais. Les formants des langues classiques, en particulier du sanskrit, étaient censés éclairer le sens des néologismes. Or ces formants n'ont plus de signification claire pour la plupart des locuteurs de hindi moderne, et les traits notionnels qu'ils dénotent sont devenus indéchiffrables, d'autant qu'ils s'appuient sur un découpage notionnel anglais (dans le cas de ces calques). De fait, les textes administratifs, et par conséquent ceux de la justice normative ou juridictionnelle, comportent de telles occurrences incompréhensibles pour le lecteur / locuteur moyen de hindi.

Un phénomène similaire opère en ourdou, avec l'arabe et le persan comme langues de référence de formation néologique, bien que l'ourdou tende à intégrer davantage d'occurrences anglaises dans sa langue administrative et judiciaire officielle (cf. chapitre 2).

Le phénomène d'opacité est, semble-t-il, déjà pris en compte dans les textes normatifs, où l'on constate dans les définitions l'utilisation de certains lexèmes opaques,

[27] On exclura les usages erronés, comme de parler d'inculpation ou de détention préventive en France, ou de tueur récidiviste lorsque le fameux tueur n'a jamais été arrêté, encore moins condamné.

et dans les exemples de ces mêmes définitions des équivalents lexicaux plus accessibles.

C'est le cas en ourdou avec deux unités lexicales pour désigner le vol, *sarqah* et *corī*. *sarqah*, d'origine arabe et opaque pour le grand public, est utilisé dans les définitions du code pénal indien. On en dénombre 106 occurrences. Son synonyme d'origine indo-aryenne *corī*, dérivé du verbe *curānā* « voler » et du nom d'agent *cor* « voleur », possède en revanche une signification transparente. Il est présent dans les exemples (5 occurrences), peu dans les définitions (1 occurrence dans un titre d'article)[28]. *sarqah* est utilisé indifféremment dans les définitions et les exemples du code pénal.

4.4.2.2.2 Spécialisation lexicale judiciaire dans le parajuridique

Si des termes très spécialisés sont réservés aux textes normatifs ou juridictionnels, d'autres sont circonscrits en hindi-ourdou aux textes parajuridiques. Dans la presse judiciaire hindi, l'usage de termes très arabo-persianisés témoigne du lien d'association toujours fortement présent entre le persan et la langue de la justice. Outre que les termes en hindi ont souvent des doublets d'usage plus courant d'origine arabo-persane, la langue de la presse judiciaire fait du zèle en utilisant par exemple des adverbes au mode de formation arabe, procédé inconnu dans la grammaire habituelle du hindi comme de l'ourdou. Sur le nom *irādah* « volonté, intention » se forme ainsi par ajout d'une préposition persane *ġair* « sans » et d'une marque adverbiale arabe *an* l'adverbe *ġair irādatan*, « involontairement ». Sur le nom persan *ranjiś* « vexation », on forme de la même façon *ranjiśan*, au sens peu accessible au grand public. Ces éléments du lexique arabo-persan sont mélangés dans des collocations au résultat parfois surprenant :

(59) *ġair irādatan* *hatyā*
involontairement AR.-PS.) meurtre SKT.
« meurtre involontaire »

On le voit, la préférence pour des termes différents de ceux utilisés dans les textes normatifs ne reflète pas un registre de langue moins soutenu. Ces termes cooccurrent d'ailleurs souvent dans un même texte avec d'autres unités phraséologiques exprimant la même notion. Le cas suivant en est un exemple intéressant.

Dans le corpus hindi, le |viol| est en effet désigné par des unités lexicales différentes selon le genre. Tandis que le corpus de langue de spécialité utilise le terme *balātsaṅg*, formé de deux composants d'origine sanskrite *balāt*, « avec force », et *saṅg* « union », le

[28]La version pakistanaise du code pénal utilise davantage *sarqah*.

corpus parajuridique (y compris celui qui explique le droit en dehors des textes normatifs) utilise *balātkār*, formé également de composants sanskrits, « avec force », et *kār* « action ». Ce lexème n'appartient ni au genre normatif du judiciaire, ni au registre populaire. Dans la langue commune (non spécialisée), le terme *balātsaṅg* n'est jamais utilisé. *balātkār*, lui, est employé dans la langue de la presse en concurrence avec *rep*, de l'anglais « rape », beaucoup plus usité (172 occurrences de *rep* contre 77 pour *balātkār*). Ce dernier terme n'appartient toutefois pas à la langue parlée, dans laquelle on lui substitue des expressions euphémisantes (autour de la perte de l'honneur, de la transgression).

En ourdou, on observe le même phénomène. La collocation *zinā bāljabar* (littéralement « relation hors mariage avec force ») reste confinée à la langue de spécialité, tandis que la langue de la presse utilise un terme beaucoup plus vague, *zyādatī* (« excès », toujours péjoratif, ou transgression, de l'adverbe *zyādah*, plus), qui ne se comprend qu'en contexte. Par exemple, dans de très nombreux FIR, il est d'usage de terminer la déposition du plaignant avec une formule du type :

(60) *gustākẖānah kalmāt isteʿmāl karke usne sakẖt* ***zyādatī*** *kī,*
insultes utilisation faire.COV il.ERG sévère transgression a fait.PASSÉ
us ke kẖilāf qānūnī kārravāī kī jāe
lui-de contre légale procédure faire.SUBJ.PASS
« En m'insultant, il a commis une grave **transgression**, qu'il soit poursuivi par la justice »

En ce cas, nul lien avec le crime sexuel. En revanche, dans la presse judiciaire, ou dans les déclarations orales des plaignants, on trouvera le terme *zyādatī* employé ainsi :

(61) *« child court » ne zyādatī kes mẽ mulavaṣ 4 mulzimõ ko*
tribunal pour enfants ERG viol affaire dans impliqués 4 accusés OBL
barī kar dī
acquitté faire.PASSÉ
« Le tribunal pour enfants a acquitté les 4 accusés impliqués dans l'affaire de viol. »

Comme dans le corpus hindi, on trouve également des occurrences de *rep* dans les textes non spécialistes du droit.

En Asie du Sud, les agressions sexuelles, bien que largement relayées les pages des faits divers, restent un sujet socialement tabou. Dans la presse ou dans la langue quotidienne, on utilisera des stratégies de mise à distance, comme l'usage d'une langue étrangère (l'anglais), ou l'euphémisation (usage d'une expression moins forte ou moins

précise). Ce sont les mots que les justiciables utiliseront en justice. Traduire et interpréter pour ces justiciables, victimes comme auteurs ou présumés tels, implique de comprendre ces nuances et de les reconnaître dans le contexte d'énonciation.

4.4.2.2.3 Conclusion pour l'analyse des résultats et équivalences traductionnelles

À travers les tableaux qui synthétisent les réseaux notionnels, les termes et les cooccurrents qui y sont associés, l'analyse du lexique et des unités phraséologiques dessine les contours des découpages notionnels du domaine judiciaire spécifiques au français, au hindi et à l'ourdou. Elle montre que même dans le cas où les langues sont proches, des notions sont couvertes par des termes et des expressions qui ne décrivent pas la même réalité (cas de *procès* en hindi et ourdou).

Le travail de recherche d'équivalents traductionnels s'organise en fonction des degrés d'adéquation aux variations de ce découpage. Plus celui-ci se superpose d'une langue à l'autre, moins les adaptations sont nécessaires. Dans le cas de coïncidences partielles (extension sémantiques plus ou moins grande d'un terme, activations de sèmes différents) ou d'absence de coïncidence entre ces découpages, et par conséquent d'absence de superposition entre les termes, le traducteur-interprète développe des stratégies de traduction que l'on retrouve dans toutes les combinaisons linguistiques. Il cherche à transmettre le sens en s'interrogeant sur les sèmes portés par les mots, et sur ceux effectivement activés en contexte, sur leurs connotations. Au besoin, il utilise un terme fonctionnel. Il s'efforce de rendre son équivalent de traduction pragmatique, adapté à la finalité de la traduction et à la situation d'énonciation. Cet équivalent se doit donc d'être identifiable par le destinataire qui en décode le sens.

La recherche d'équivalence de traduction est étroitement liée à la compréhension du destinataire. Cette compréhension dépend de ses compétences linguistiques (maîtrise des caractéristiques du genre, de son lexique) et vaut pour l'initié au droit comme pour le non initié. Tout comme le justiciable ordinaire possède peu de connaissances en droit, les magistrats français sont peu familiers avec des notions culturelles présentes en Inde et au Pakistan comme les sujets tabous ou la large place faite au consensus dans la prise de décision, et à leurs modes d'expression. Négliger ce paramètre revient à risquer de formuler une traduction opaque, incompréhensible de part et d'autre, voire à passer à côté du sens du message.

La prise en compte de cette dimension socioculturelle dans la traduction va à l'encontre de la mission de littéralité imposée par la justice français au traducteur-interprète

mais s'attache à transmettre au mieux le sens à travers les langues et les cultures, finalité du processus traduisant.

4.4.3 *Conclusion*

Faisant suite aux aspects de contextualisation des premier et deuxième chapitres, et aux aspects techniques présentés au chapitre 3, ce quatrième et dernier chapitre présente l'analyse de corpus qui fournit une base matérielle attestée pour la recherche d'équivalents de traduction entre le hindi, l'ourdou et le français.

Nous avons expliqué l'utilité de collecter des documents variés pour constituer ce corpus. En effet, si les textes qui disent le droit posent les bases conceptuelles du domaine judiciaire, on s'aperçoit que les textes qui appliquent le droit sont constitués de multiples sous-genres et séquences qui comportent des termes en dehors de la langue de spécialité. D'autre part, la langue non spécialisée qui parle du judiciaire, qui possède son lexique, ses unités phraséologiques et ses connotations, offre un moyen d'accès pour le non spécialiste à ce domaine opacifié par les termes de spécialité. Or la justice, contrairement à des domaines techniques pointus, concerne l'ensemble de la communauté. Sa langue devrait donc être accessible à tous.

Cette recherche se veut une base pour des développements ultérieurs. Les résultats exposés en fin de chapitre et dans les annexes fournissent des ressources qui restent à développer pour la recherche d'équivalences de traduction. Une partie des documents n'a pu être exploitée faute de temps. Son analyse permettrait d'affiner davantage les caractéristiques des genres et d'enrichir les lexiques qui servent à l'alignement des textes. Les ressources présentées aideront le professionnel à choisir ses stratégies de traduction. Ce processus est une quête d'équilibre entre la transmission du sens et l'impression du genre. Ce travail d'équilibriste, nous espérons l'avoir démontré, est bien loin d'un simple transcodage et ses dimensions sont méconnues de la justice.

Conclusion

À travers cette thèse, nous avons démontré qu'il était possible de créer des ressources terminologiques et phraséologiques pour le genre juridique à partir d'un corpus de petite taille, dans des combinaisons lniguistiques disposant de peu de documentation.

Dans la première partie, nous avons posé le contexte culturel juridique et linguistique de trois pays. Cette description a mis en relief les différences de structures du judiciaire qui entraînent un découpage notionnel différent de ce champ. L'histoire juridique des trois pays a permis de comprendre comment s'étaient développés les concepts et termes de la justice au sein de leurs cultures respectives, et comment les termes se sont chargés de sèmes et de connotations au fil du temps.

La langue étant le moyen par lequel la justice se crée et s'exerce, la question du statut des langues se situe au cœur de notre problématique de traduction et de compréhension du système judiciaire. Nous avons vu que les politiques linguistiques déterminent le pouvoir accordé à une langue en privilégiant ou limitant son usage. Ainsi, dans des pays en réalité peu anglophones comme l'Inde et le Pakistan, où la maîtrise de l'anglais reste réservée à une élite et possède un capital symbolique fort, on constate que l'effort placé sur les traductions vers les autres langues prend toujours son origine dans l'anglais, qui finit par influencer le hindi standard, normé pour l'éducation, les médias et l'administration (calques, emprunts lexicaux et syntaxiques, notamment usage généralisé du passif).

La politique linguistique fixe non seulement la hiérarchie des langues dans les interactions, mais aussi la norme de ces langues. Inversement, elle décide des variétés jugées incorrectes de la langue. Elle contribue en cela à diriger le développement de la langue. Ce fut le cas du hindi, avec l'adoption d'une variété sanskritisée comme standard, artificielle car éloignée des usages populaires, surtout ceux parlés. Ce standard idéalisé coupe tout une partie de la population, peu éduquée, qui en maîtrise mal les codes, et dont les pratiques langagières sont déjà dévalorisées en public. Le même schéma se retrouve au Pakistan, avec la non reconnaissance officielle de langues parlées par la majorité de la population. Il s'est également appliqué à la France avec l'éradication des parlers autres

que le français, jugés inférieurs.

La politique linguistique d'un pays multilingue comme l'Inde est d'ailleurs contradictoire dans les faits, avec d'un côté une reconnaissance du plurilinguisme et des mesures de protection des langues minoritaires, et de l'autre un encouragement voire une pression sociale à utiliser des modes d'expression standardisés, artificiels, et l'anglais, langue d'élite.

Ces politiques ont par conséquent un impact direct sur le moyen pour les citoyens d'accéder au droit. En Inde, des initiatives gouvernementales ou associatives de vulgarisation du droit existent, avec notamment la création de matériel didactique dans un hindi plus simple. Lorsqu'on observe ces documents, on remarque que les mises en situation sont effectivement imagées et dans une langue simple. Toutefois, les explications techniques demeurent formulées dans une langue artificielle très proche de celle des textes normatifs. Nous avons expliqué en quoi cette situation d'inadéquation entre langue officielle et langue réellement pratiquée n'était pas nouvelle en Inde et au Pakistan, ce qui pose hors de ces pays un paradoxe intéressant sur l'étendue de l'assistance linguistique. En effet, un justiciable pendjabi pakistanais n'aura pas la possibilité d'être entendu ni lu dans son pays dans sa langue première, puisque celle-ci n'y est pas reconnue dans les actes officiels. En revanche, il aura le droit de s'exprimer dans sa langue dans une juridiction française, où son droit à être entendu dans une langue qu'il comprend s'applique.

Dans la deuxième partie, nous sommes entrée au cœur de la procédure judiciaire pour en cerner les acteurs, les phases, et comprendre les rôles remplis par les traduteurs-interprètes, experts ou non. Nous avons constaté le décalage entre la mission imposée par la justice et le procédé réel impliqué dans la traduction et l'interprétation qui nécessite de privilégier la transmission du sens sur la littéralité.

Après cette introduction aux notions judiciaires, nous avons associé des genres textuels à la procédure, afin d'identifier les caractérisitiques des différents sous-genres textuels que le traducteur ou l'interprète est amené à traduire au cours de sa pratique. Cette analyse a permis de constater les différences de genre entre les langues, préparant le traducteur-interprète à adapter le contenu des textes en conséquence.

Ici aussi se pose au praticien de la langue la question de l'imitation des traits du genre. En conservant ces caractéristiques, il risque de brouiller le message. C'est le cas dans les textes juridictionnels avec les expression figées et archaïques, ou les nombreuses unités phraséologiques dans les textes en hindi et en ourdou qui sont hautement persianisées, arabisées ou sanskritisées. Ces unités rendent le sens du texte opaque mais l'identifient dans son genre. La preuve en est que ces codes sont repris jusque dans la presse judiciaire

qui, en hindi, n'hésite pas à reprendre les traits typiques d'une langue très influencée par l'arabe et le persan. Le traducteur-interprète, qui possède désormais des informations sur la procédure pénale ainsi que sur les usages par genre, est armé pour faire ses choix sur des bases étayées.

Dans la troisième partie, nous avons détaillé l'ensemble des aspects techniques de cette recherche, depuis la constitution des corpus attestant de genres textuels variés jusqu'à la réflexion sur le processus de recherche d'équivalents traductionnels.

L'analyse de notre corpus excluait le recours à des traitements bien rôdés sur de gros corpus (comme le calcul des vecteurs d'association lors de l'alignement de corpus comparables) ou l'utilisation d'outils performants sur d'autres langues (notamment les outils d'extraction terminologiques conçus pour des langues en script latin, en cyrilliques ou en caractères idéographiques).

Le traitement du corpus a nécessité la levée de nombreux obstacles qui témoignent du chemin encore à parcourir pour doter le hindi, et plus encore l'ourdou, d'un véritable accès démocratisé à des ressources variées et informatives en ligne. Les difficultés rencontrées dans la collecte des documents et leur exploitation ont nécessité la création de ressources et d'outils libres de droit pour traiter au mieux les données. Nous avons ainsi constitué des listes de lemmes et de dictionnaires en hindi et en ourdou afin de pouvoir utiliser des logiciels de traitement des langues naturelles comme Unitex[29], pour entraîner des étiqueteurs morphosyntaxiques ou anonymiser des textes.

La linguistique de corpus est un domaine récent en hindi et en ourdou, en pleine évolution. Elle va de paire avec la numérisation croissante en Inde et au Pakistan, qui demeure toutefois limitée à un public urbain et anglophone. Il est ainsi possible d'effectuer de plus en plus de démarches via internet (demande de carte d'identité au Pakistan, dépôt de plainte en Inde), mais la totalité des fonctionnalités des interfaces utilisateurs n'est en général disponible qu'en anglais. Nous avons vu au cours de la collecte de corpus que les pages disponibles dans les autres langues contiennent moins de données et renvoient à moins de documents. Lorsqu'elles le font, ces documents comportent souvent des problèmes d'interopérabilité (format rendant le contenu non éditable) qui n'existent pas avec les versions anglaises, facilement manipulables. Le hindi et l'ourdou sont pourtant des

[29]Le logiciel Unitex permet, à l'aide de dictionnaires et de grammaires, de représenter des phénomènes liguistiques à l'aide de graphes et d'automates. Le logiciel est disponible sur `https://unitexgramlab.org/fr`. Cf. la documentation en bibliographie (Paumier, 2011).

langues pratiquées par des centaines de millions de locuteurs. Elles ont intérêt à travailler sur des moyens d'intégrer des outils d'analyse déjà existants plutôt qu'à en développer des spécifiques, ce qui leur permettrait une meilleure visibilité dans la discipline.

À travers notre chaîne de traitement, nous avons expérimenté les possibilités offertes d'une part par l'extraction terminologique, d'autre part par l'alignement de corpus parallèles ou comparables dans la recherche de correspondances multilingues. Ces expériences ont permis de développer des concordanciers prenant en charge les trois langues, les trois systèmes qui leur sont liés et deux directions de lecture. Cette partie de développement informatique, réalisée en collaboration avec l'équipe de recherche, a abouti également à l'élaboration d'applications en ligne permettant de retrouver un motif dans son contexte, ses équivalents dans les autres langues ainsi que des informations lexicogrpahiques d'après des documents du corpus.

Autre volet consécutif à la constitution des corpus, leur analyse lexicale et phraséologique. Les réseaux notionnels ainsi dégagés et leur actualisation dans le lexique et la phraséologie de textes classés par sous-genres démontrent encore une fois les variations du découpage notionnel d'une langue à l'autre, fussent-elles proches. Ces variations excluent une traduction littérale, « fidèle » au sens de la justice française, et obligent le traducteur-interprète à une quête constante d'équilibre entre sens du message, transmission de la fonction à travers les caractéristiques du genre, et prise en compte du public visé.

Outre la démonstration faite qu'une langue ne peut être réduite à un simple code, nous espérons avoir fait prendre conscience de l'importance du contexte dans la compréhension du message, à travers ce qui est écrit et ce qui ne l'est pas, et dans sa restitution. Tout comme le traducteur-interprète doit se former à la procédure judiciaire pour bien la traduire, il est primordial que les professionnels ou auxiliaires de justice (magistrats, avocats, experts, policiers…) soient sensibilisés au travail du traducteur-interprète, et en particulier au processus traduisant.

Cette recherche n'est qu'un commencement. Elle pose la base de développement dans les ressources de linguistique de corpus en hindi et en ourdou. Il est possible d'améliorer la plateforme de ressources multilingues en enrichissant le lexique. À partir de celui-ci on créera davantage d'amorces lexicales pour extraire des unités phraséologiques, avec supervision humaine. Ces nouvelles entrées lexicales doivent être complétées par des définitions multilingues afin de renforcer la possibilité de consulter les définitions de termes qui semblent ne pas avoir d'équivalent, mais que les traits sémiques pourraient, dans des contextes déterminés, transformer en candidats intéressants pour une équivalence

fonctionnelle.

Par ailleurs, cette recherche ouvre également des perspectives pour la formation à la traduction. Les étudiants peuvent travailler sur des versions alignées de textes multilingues afin d'y trouver des équivalents de traduction idiomatiques (Kübler, 2014) ou de comprendre les différentes stratégies qui mènent aux correspondances identifiées.

Enfin, ce travail ouvre sur des questions sociolinguistiques quant à l'accès au droit à travers la langue. En choisissant de rendre transparent sa traduction pour le destinataire non francophone, le traducteur place celui-ci dans une situation plus favorable que celle des francophones non initiés au droit, qui comprendraient moins bien les textes dans une variété de langue trop spécialisée. Cela pose la question des efforts de simplification de la langue administrative qui se développent partout en Europe et dans le monde (le « plain language », dans certains États des États-Unis, en Suisse, etc.), jusque dans les juridictions françaises où les magistrats ont pour consigne de simplifier la rédaction de leurs documents (abandon de la formule « attendu que », tournures à la première personne du pluriel remplacées par la troisième personne).

Par ailleurs, une extension de ce travail peut être réalisée sur un corpus oral afin d'étudier les interactions entre justiciables non initiés au droit et professionnels de la justice. Une telle étude permettrait de documenter le décalage des usages langagiers entre les deux catégories et ses conséquences. En effet, si le discours écrit de la justice parvient à se simplifier, il est nécessaire que le discours oral judiciaire prenne conscience des compétences linguistiques des justiciables et se mette à leur portée, sans se vider de son contenu juridique, afin de garantir une garantie universelle à la compréhension du droit.

Annexe A

Abréviations et glossaire

bâtonnier Chef et représentant du conseil de l'Ordre chargé d'administrer chaque barreau.

BCT Base de connaissance terminologique.

Bṛhas. Bṛhaspati.

candidats-termes Mots susceptibles d'être retenus comme termes par un analyste et permettant de fournir des étiquettes de concepts.

CESDH Convention européenne de sauvegarde des droits de l'Homme.

chalan Rapport de fin d'enquête de police dans la procédure en Inde et au Pakistan, avant avis du magistrat.

chef de prévention / d'accusation Motif pour lequel une personne est mise en cause / accusée.

cibliste En traductologie, démarche qui privilégie l'adaptation au contexte d'arrivée (public, culture).

CIIL Central Institute of Indian Languages.

CLE Center for Language Engineering.

collocation Type de séquence polylexicale déterminée.

commission rogatoire (CR) Document par lequel un juge d'instruction mandate des services judiciaires et leur délègue des pouvoirs pour les besoins de l'enquête.

common law Unification du droit coutumier de Grande-Bretagne, La common law a ultérieurement été complétée par l'equity qui en tempérait les jugements. Les deux fonctionnent aujourd'hui à l'unisson, sans avoir fusionné.

conseil Autre désignation de l'avocat.

contravention Degré le plus bas dans l'échelle des infractions.

corpus Ensemble de textes construit à des fins d'analyse de phénomènes particuliers. Pour une recherche donnée, on s'efforce de constituer un corpus représentatif échantillonné à l'aide de faits pertinents.

CRI Commission rogatoire internationale.

crime Infractions les plus graves, faisant l'objet d'une procédure criminelle.

CrPC Criminal Procedure Code.

CRPC Comparution sur reconnaissance préalable de culpabilité.

CRULP Center for Research in Urdu Language Processing.

CSTT Committee for Scientific and Technical Terminology ou textitvaigyānik evam taknīkī āyog.

data driven corpus (corpus guidé par les données) Ce sont les données qui permettent la formulation d'analyses sur le corpus.

délit Infractions faisant l'objet d'une procédure correctionnelle.

densité Pour une langue, disponibilité du matériel archivé numériquement (Varga et al., 2005).

DGLFLF Délégation générale à la langue française et aux langues de France.

discours Acte de langue produit dans le cadre d'une situation d'interaction sociale. Il s'articule sous forme de texte dans une "sphère d'usage" caractérisée par un domaine thématique.

DIT Department of Information Technology .

diyat En droit islamique, compensation ou « prix du sang ».

encodage Représentation d'une information par une suite binaire. L'encodage assigne à chaque caractère dans un répertoire (ou registre de caractères) un point de code (ou position) représenté par un nombre. Il lie ainsi un nombre dans l'ordinateur et un caractère.

ETI Expert traducteur-interprète.

étiquetage morphosyntaxique Identification d'un token en fonction de ses propriétés morphosyntaxique (en général de sa catégorie grammaticale).

faujdar (faujdār) Dans l'administration moghole, fonctionnaire en charge du maintien de l'ordre au niveau du district et responsable de l'ensemble des thanas.

FIJAIS Fichier judiciaire national automatisé des auteurs d'infractions sexuelles ou violentes.

fiqh (fiqḥ) Droit islamique.

FIR First information report.

formation de jugement Membres qui composent le tribunal amené à juger une affaire.

genre Interaction sociale qui définit le cadre d'énonciation du texte.

glyphe Forme graphique définie pour afficher un caractère.

hadd (*ḥadd*) En droit islamique, infraction relative à une prescription religieuse.

html (hypertext markup language) Format de mis en forme et de mise en page des documents pour le web.

huissier audiencier Auxiliaire de justice présent à l'audience, qui tient la liste des affaires (rôle).

huissier de justice Auxiliaire de justice chargé de délivrer ou notifier des actes judiciaires.

ICLI Indian Languages Corpora Initiative.

indice de spécificité Probabilité qu'une forme A apparaisse textitf fois dans une partie textitp de longueur textitt la forme apparaissant textitF fois en tout dans l'ensemble du corpus dont la longueur totale est de textitT occurrences » citepTextometrie2007.

interjeter Lorsqu'un recours est formé devant la cour d'appel, on emploie l'expression « interjeter appel ».

ISCII Format développé pour les langues indiennes sur le modèle de l'ASCII américain.

JAP Juge d'application des peines.

JEX Juge de l'exécution.

jirga (*jirgah*) (Pakistan) Assemblée non élue d'aînés et de notables qui se réunit pour rendre la justice dans la société pashtoune (province du KPK, FATA, Balouchistan).

JLD Juge des libertés et de la détention.

judiciaire Ce qui se rapporte à la Justice, à l'administration de la Justice.

jurisprudence Ensemble des décisions rendues antérieurement.

justiciable Personne ayant affaire à la justice pour demander réparation d'un tort ou pour répondre d'une mis en cause.

Kātyā. Kātyāyaṇa.

KCIE Kolhapur Corpus of Indian English.

kotwal (*kotvāl*) « policier », désigne aujourd'hui un officier de police.

lemme Forme canonique d'un mot.

Lois de Huddood (Pakistan) 5 lois concernant les catégories d'infraction pour lesquelles une une peine est prévue par le Coran ou la Sunna comme un droit de dieu.

madrasa École d'études religieuses islamiques pouvant s'ouvrir à d'autres sujets que l'exégèse des textes religieux.

magistrat du parquet Magistrat représentant les intérêts de la société (ex. procureur de la République).

magistrat du siège Magistrat qui rend une décision (ex : juge d'instruction, JI, JAP, JLD…).

maulvi Lettré religieux musulman.

MIA Middle Indo-Aryan, moyen indo-aryen.

Ministère public Autre désignation du Parquet. Il représente les intérêts de la société.

mise en accusation Décision prise, par le juge d'instruction ou par la chambre de l'instruction, à l'encontre d'une personne mise en examen pour crime, de la renvoyer devant la cour d'assises pour y être jugée.

Mitā. Mitākṣarā.

NIA New Indo-Aryan, langues actuelles de la famille indo-aryenne.

Nyaya Panchayats (*nyāya pañcāyat*) Système contemporain de justice alternatif au niveau du groupement de village, issu de l'institution traditionnelle du panchayat. Le consensus y est le mode privilégié de résolution des litiges.

OCR (optical character recognition / reader) Reconnaissance / lecture optique des caractères. Permet de lire le contenu d'une image et de le transformer en texte.

OIA Old Indo-Aryan, indo-aryen ancien.

OMA Ordonnance de mise en accusation.

OMD Ordonnance de maintien en détention.

ONL Ordonnance de non-lieu.

OPJ Officier de police judiciaire.

ordonnance Décision rendue par un juge unique.

ORTC Ordonnance de renvoi devant le tribunal correctionnel.

PAN Pan Asia Network.

pandit (*paṇḍit*) Lettré brahmane.

partie civile Individu ayant personnellement souffert d'une infraction et exerçant une action contre l'auteur.

phraséolexique Ensemble du lexique et de la phraséologie d'un genre.

plaidoirie Prise de parole d'un avocat visant à exposer l'argumentaire à l'appui de sa demande.

police (informatique) Répertoire de glyphes.

PPC Pakistan Penal Code.

prévenu Personne mise en cause dans une affaire correctionnelle.

procureur de la Republic Magistrat représentant les intérêts de la société à l'instruction ou en audience correctionnelle.

qazi (*qāzī*) Magistrat, aujourd'hui personne versée dans le droit civil musulman, chargé notamment de célébrer les mariages.

qisas (*qiṣāṣ*) En droit islamique, infraction relative à un homicide.

qualification Identification d'un fait dans un texte légal.

réclusion Peine de prison supérieure à 10 ans.

relaxe Reconnaissance de la non culpabilité d'un prévenu (procédure correctionnelle).

réquisitoire 1) Décision du procureur de confier l'affaire à un juge d'instruction ou d'abandonner l'information. 2) Lors de l'audience, prise de parole du procureur (ou équivalent selon les cours) après la clôture des débats pour justifier sa demande.

Rigveda (*ṛgveda*) Premier des quatre Védas.

rôle Liste des affaires prévues à l'audience.

saisine Prise en charge d'une plainte par le magistrat.

segmentation Délimitation d'un texte en différentes unités exploitables. On peut segmenter un texte en parties, sous-parties, paragraphes ou phrases.

segments répété Séquence de mots non séparés par des symboles de ponctuations qui apparaît au moins deux fois dans le corpus (Lafon et Salem, 1983).

smriti (*smṛti*) « Mémorisation » ensemble de textes transmis d'une génération de brahmanes à l'autre.

sourciste En traductologie, démarche qui privilégie le texte d'origine.

SPIP Service pénitentiaire d'insertion et de probation.

ṣulah Conciliation.

TAL Traitement automatique des langues.

tazir (*taʿzīr*) En droit islamique, infraction qui ne relève ni de la catégorie ḥadd, ni du qiṣāṣ.

TDIL Technology Development for Indian Languages.

texte Actualisation du discours dans un cadre énonciatif donné (genre). Il est réglé par divers plans d'organisation : sémantique, énonciative et illocutoire.

TGI Tribunal de grande instance.

thana (*thānā / thānah*) Poste de police.

token Unité signifiante d'un corpus.

tokenisation Segmentation au niveau inférieur, en unités signifiante (généralement unité de sens ou lexicale).

Tribunal Fédéral de la Charia (Federal Shariat Court) (Pakistan) Tribunal religieux fondé sur le droit islamique instauré afin de vérifier si les lois en vigueur sont en accord avec les principes de l'islam.

Unicode Standard pour l'encodage en UTF-8 qui permet l'échange de textes dans toutes les langues possibles, et d'intégrer dans un même texte différentes langues. Le standard établit un répertoire universel des caractères. Il définit également une documentation sur les usages des caractères en fonction de leur contexte d'occurrence.

unité phraséologique Tout groupement de mots récurrent (expression, locution ou construction au sens large) auquel les locuteurs [...] associent un sens précis, qui n'est pas nécessairement lié au sens de ses composantes prises individuellement (Zinglé 2003).

UTF-8 (Universal Character Set Transformation Format - 8 bits) Système inventé en 1992 par Kenneth Thompson dans lequel chaque caractère se voit attribuer un point de code unique, auquel correspond une séquence de 1 à 4 octets.

vakil (*vakīl*) Avocat.

Védas Ensemble de 4 textes sacrés révélés, premier monument littéraire de l'hindouisme.

verdict Décision d'un jury d'assises sur la culpabilité ou l'innocence d'un accusé.

xml (extent markup langage) Format de mise en forme des documents informatiques à l'aide de balises.

Yājñ. Yājñvalkya.

zamanat (*żamānat*) Libération sous caution.

Annexe B

Ressources utilisées pour la constitution des corpus

Les tableaux suivants récapitulent les noms des textes formant les corpus, leur lien url le cas échéant, ainsi que le nombre de tokens qui les composent.

B.1 Ressources en hindi

Les documents à caractère juridictionnel anonymisés (1 jugement et FIR) ne sont pas diffusables.

n°	**titre du document**	**nombre de tokens**
01	Code pénal indien (IPC) [mise à jour de 2013] भारतीय दण्ड संहिता, 1860 http://legislative.gov.in/important-legislation	82 652
02	Code de procédure pénale (CrPC) भारतीय प्रक्रिया संहिता, 1973 https://drive.google.com/file/d/0B468FAd4231SbVNHcW9hTkFYNFU/view	107 967
03	Loi sur les preuves (Evidence Act, 1872) भारतीय साक्ष्य अधिनियम, 1872 http://legislative.gov.in/sites/default/files/H187201_1.pdf	6 997
04	Advocate Act, 1961 (loi sur les conseillers juridiques) अधिवक्ता अधिनियम, 1961 http://legislative.gov.in/centralact/1961	17 002
05	Loi sur les inculpés अपराधी परिवीक्षा अधिनियमम, 1958 http://legislative.gov.in/centralact/1958-0	3 051
06	Loi sur l'identification des prisonniers बन्दी शनाख़्त अधिनियम, 1920 http://legislative.gov.in/centralact/1920	1 003
07	Loi sur la prévention de la dot दहेज प्रतिषेद्ध अधिनियम, 1961 http://legislative.gov.in/centralact/1961	2 507
08	Loi sur les prisonniers बन्दी (न्यायालय में उपस्थिति) अधिनियम, 1955 http://legislative.gov.in/centralact/1955-0	1 767
09	Special criminal court Act, 1950 विशेष दण्ड न्यायालय (अधिकारिता) अधिनियम, 1950 http://legislative.gov.in/centralact/1950	346
10	Jugement 1	1 267
11	Jugement 2	858
12	Dépôt de plainte	1 710

n°	titre du document	nombre de tokens
13	Blog de la police d'Indore (1) et (2) https://indorepolice.blogspot.com/	61 203
14	Charte des prisons (Uttarakhand) http://slsa.uk.gov.in/contents/view/4/20/5-informative-literature	2 784
15	Aide juridique (Hong-Kong) http://adhikarexpress.com/rights_page.php?catid=83&pfld=pages_content	234
16	Aide juridique (Inde) क़ानूनी उपाय http://adhikarexpress.com/rights_page.php?catid=83&pfld=pages_content	3 220
17	Histoire du droit (Inde) विधिक इतिहास https://teesarakhamba.com/	3 461
18	Forum de questions juridiques तीसरा खम्भा www.tisrakhambha.com	12 024
19	Articles de presse judiciaire (Amar Ujala - अमर उजाला) http://www.amarujala.com/crime	371 313
20	Corpus monolingue de référence, projet LINDAT-CLARIN http://hdl.handle.net/11858/00-097C-0000-0023-6260-A	365 millions

B.2 Ressources en ourdou

Les documents à caractère juridictionnel anonymisés (FIR, témoignage, mandats d'arrêt) ne sont pas diffusables.

n°	titre du document	nombre de tokens
01	Indian Penal Code مجموعہ تعزیرات بھارت http://legislative.gov.in/important-legislation	78 312
02	Evidence Act 1872 (loi sur les preuves) بھارتی شہادت ایکٹ،1872 http://legislative.gov.in/urdu	6 741
03	توہین رسالت اور اسلام (article sur le blasphème) https://www.urduweb.org	1 709
04	تعزیراتِ پاکستان کے تحت مختلف جرائم کی سزائیں (article sur les peines criminelles au Pakistan) https://urdu.arynews.tv	666
05	تحفظ حقوق نسواں بل (loi de protection des droits des femmes) https ://www.dunyapakistan.com/65247/	4621
06	Jugement1 (décision de la Cour suprême pakistanaise) http://supremecourt.gov.pk/web/user_files/File/C.M.A.3325_of_2012_dt_31-1-2013_ENG_URDU.pdf.pdf	2 892
07	JugementMQM (décision de la Cour suprême pakistanaise Crl.O.P. 96/2012) http://supremecourt.gov.pk/	1 918
08	FIR1 (First information repport)	551
09	FIR2 (First information repport)	539
10	FIR3 (First information repport)	299
11	FIR4 (First information repport)	440

n°	titre du document	nombre de tokens
12	FIR5 (First information repport)	716
13	Témoignage (déclaration sous serment)	445
14	Mandat 1 (mandat d'arrêt)	190
15	Mandat 2 (mandat d'arrêt)	146
16	Power of Attorney (formulaire)	124
17	Nawai Waqt (articles de faits-divers) http://www.nawaiwaqt.com.pk/crime-court	548 095
18	Corpus étiqueté monolingue de référence, projet LINDAT-CLARIN http://hdl.handle.net/11858/00-097C-0000-0023-65A9-5	5 464 575 phrases

B.3 Ressources en français

Les documents à caractère juridictionnel anonymisés (OML, OMA, ORTC, OMD, ORDML, CRI, COPJ, avis FIJAIS) ne sont pas diffusables.

n°	titre du document	nombre de tokens
01	Code pénal www.legifrance.fr	129 448
02	code de procédure pénale www.legifrance.fr	604 222
03	OMA1 Ordonnance de mise en accusation	13 188
04	OMA2 Ordonnance de mise en accusation	3 186
05	OMA3 Ordonnance de mise en accusation	5 470
06	OMD1 Ordonnance de maintien en détention	400
07	OMD2 Ordonnance de maintien en détention	364
08	OML1 Ordonnance de mise en liberté	980
09	ONL Ordonnance de non-lieu	2363
10	ORDML Ordonnance de rejet de demande de mise en liberté	874
11	ORTC2 Ordonnance de renvoi devant le tribunal correctionnel	9 344
12	ORTC3 Ordonnance de renvoi devant le tribunal correctionnel	5 826
13	ORTC4 Ordonnance de renvoi devant le tribunal correctionnel	40 514
14	ORTC5 Ordonnance de renvoi devant le tribunal correctionnel	2 651
15	ORTC6 Ordonnance de renvoi devant le tribunal correctionnel	11 347
16	ORTC7 Ordonnance de renvoi devant le tribunal correctionnel	34 862
17	ORTC8 Ordonnance de renvoi devant le tribunal correctionnel	2 409
18	ORTC9 Ordonnance de renvoi devant le tribunal correctionnel	7 001
19	CRI Commission rogatoire internationale	5 837
20	COPJ Convocation par officier de police judiciaire	487
21	FIJAIS Avis du fichier FIJAIS	1 192
22	Forum juridique de Net Iris https://forum-juridique.net-iris.fr	66 419

n°	**titre du document**	**nombre de tokens**
23	Forum juridique de Legavox `https://www.legavox.fr`	
24	Vocabulaire juridique `www.justice.gouv.fr`	3 313
25	Corpus d'articles de l'Est Républicain (2003) `https://www.cnrtl.fr/corpus/estrepublicain/`	14 millions

B.4 Programmes d'extraction des articles de presse judiciaire

Le script ci-après est l'un des scripts en bash qui a servi à l'aspiration des articles de presse judiciaire du quotidien ourdou Nawai Vaqt. Les scripts pour les autres langues ou journaux reprennent la même trame, avec des modifications pour l'extraction du corps de l'article en fonction de l'organisation de la page HTML.

Le programme, installé sur un serveur distant, aspirait quotidiennement et à heure fixe le contenu des pages pour les sauvegarder dans des répertoires horodatés, aux formats texte, xml et html.

```
#==================================
#!/bin/bash

# Programme d'extraction quotidienne des pages html de "Nawai Vaqt" (ourdou) et
# de leur contenu textuel de la rubrique "faits divers" de différents journaux.
#==================================
# Le répertoire de stockage des résultats se trouve dans "crime-court".

# Création de noms des répertoires horodatés :

repertoire=$(date|cut  -d" " -f2-4|sed -r 's/(.+) (.+) (.+)/\1_\2_\3/'|tr -d ',')

# Création des répertoires par type de fichier (xml, html, txt) :

mkdir -p $HOME/crime-court/$repertoire/xml
mkdir $HOME/crime-court/$repertoire/html
mkdir $HOME/crime-court/$repertoire/txt

#Téléchargement des pages html et stockage dans les répertoires horodatés :

for url in $(curl "http ://www.nawaiwaqt.com.pk/crime-court" |
grep -oE "http ://www\.nawaiwaqt.com.pk/crime-court[^\"]+" |uniq) ;
do
    wget -O $HOME/crime-court/$repertoire/html/$(basename $url) $url ;
#Enregistrement de la liste des urls dans un fichier hebdomadaire :
    echo $url >> $HOME/crime-court/url-hebdo ;
done

#Isolement du corps des articles des pages html :
for fic in $(ls crime-court/$repertoire/html) ;
do
echo $fic ;
grep -E -A 2 "<\!-- Details page top news content margin for empty image -->"
 crime-court/$repertoire/html/$fic |
 sed -r 's/<\/p>|<p>|<p class="details_empty_image_margin">|<\!.+>//g' ;
done
#Création d'un fichier xml à partir des articles nettoyés et déclaration
# des entités posant problème à l'affichage (virgule inversée, hamza
# sur le vao, guillemets et quotes) :

echo "<?xml version=\"1.0\" encoding=\"UTF-8\"?>
\n<!DOCTYPE articleourdou [
\n<!ELEMENT article (#PCDATA)>
\n<!ENTITY lsquo \"&#x060C;\">
\n<!ENTITY rsquo \"&#x8217;\">
\n<!ENTITY iquest \"&#x0624;\">
\n<!ENTITY nbsp \" \">
```

```
\n<!ENTITY rdquo \"&#x8221;\">
\n<!ENTITY ldquo \"&#x8220;\">
\n<!ENTITY para \"&#x0624;\">
\n<!ENTITY hellip \"&#x133;\">
]>\n<articles>\n" > crime-court/$repertoire/xml/ARTICLES-$repertoire.xml ;
for fic in $(ls crime-court/$repertoire/html) ;
do
    echo $fic ; grep -E -A 2 "<\!-- Details page top news content margin for
    empty image -->" crime-court/$repertoire/html/$fic |
    echo "<article num=\"$fic\">
    $(sed -r 's/<\/p>|<p>|<p class="details_empty_image_margin">|<\!.+>|<br \/>//g')</article>"
     >> crime-court/$repertoire/xml/ARTICLES-$repertoire.xml ;
done ;
echo "</articles>" >> crime-court/$repertoire/xml/ARTICLES-$repertoire.xml ;

#Stockage du contenu textuel des urls dans des répertoires horodatés :

for url in $(curl "http ://www.nawaiwaqt.com.pk/crime-court" |
grep -oE "http ://www\.nawaiwaqt.com.pk/crime-court[^\"]+" |uniq) ;
do lynx -dump -nolist > $HOME/crime-court/$repertoire/txt/$(basename $url) $url ;
done ;

#Recherche et enregistrement des titres des articles :

for titre in $(ls $HOME/crime-court/$repertoire/html) ;
do
grep -A 3 "<div class=\"title\">"  $HOME/crime-court/$repertoire/html/$titre |
egrep -o "<span>(.+)</span>"|sed -r 's/<span>(.+)<\/span>/\1/' >> TITRE ;
done ;
```

Annexe C

Index thématique des termes judiciaires

Index par langue. Les termes sont rangés par ordre alphabétique et renvoient à une sphère thématique, éventuellement un sous-domaine. Les unités entre barres verticales (| |) sont des variantes orthographiques. Les unités séparées par une oblique sont des variantes sémantiques. Cet index existe aussi en version xml.

C.1 Index des termes en français

accusation procès
accusé procès
acquittement sanction ; absence de peine
acte de malveillance infractions ; biens
agression infractions ; personne
aide juridique procès
amende sanction ; peine
arme infractions ; circonstances
arrestation instruction ; action
arrêt sanction ; décision
assassinat infractions ; personne
audience procès
auteur protagoniste ; partie
aux fins de instruction ; idiome
auxiliaire de justice protagoniste
avocat protagoniste ; aux
avocat général protagoniste ; magistrat
bâtonnier protagoniste ; aux
blessure infractions ; préjudice
caution sanction ; restriction
centre pénitentiaire lieu
commissariat lieu
commission rogatoire instruction ; document
comparution instruction ; action
complice protagoniste ; partie
condamnation sanction ; peine
consentement infraction ; circonstances
contradictoire instruction
contravention infractions ; hétéronyme
contrôle judiciaire sanction ; restriction
cour lieu
couteau infractions ; circonstances
crime infractions ; hétéronyme
dague infraction ; circonstances
décision sanction ; décision
défense procès
déférer instruction ; action
délit infractions ; hétéronyme
déposition instruction ; document
détention sanction ; peine
détention provisoire sanction ; restriction
diffamation infractions ; personne
dommage infractions ; préjudice
enlèvement infractions ; personne
enquête instruction ; action
expert protagoniste ; aux
extorsion infractions ; personne
fait infractions ; hétéronyme
fouille instruction ; action
fuite instruction ; action
fusil infraction ; circonstances
garde à vue sanction ; restriction
greffier protagoniste ; aux
harcèlement infractions ; personne
homicide infractions ; personne
Huis-clos procès
incitation infractions ; circonstances
inculpation instruction ; action
inculper (lemme) instruction ; action
instruction instruction ; action
intention infractions ; circonstances
interception instruction ; action
interpellation instruction ; action
juge protagoniste ; magistrat
jugement sanction ; décision
juré protagoniste ; partie
juridiction lieu
jury procès
libération sanction ; absence de peine
magistrat protagoniste ; magistrat
maison centrale lieu
maison d'arrêt lieu
menace infractions ; personne
meurtre infractions ; personne
ministère public procès
mis en examen protagoniste ; partie
mise en examen instruction ; action
non-lieu sanction ; absence de peine

nuisance infractions ; préjudice
OMA instruction ; document
ORTC instruction ; document
partie protagoniste ; partie
partie civile procès
peine sanction ; peine
perquisition instruction ; action
perte infractions ; préjudice
pièce instruction ; document
pistolet infractions ; circonstances
plaidoirie procès
plaignant protagoniste ; partie
plainte instruction ; document
poursuite procès
prévenu protagoniste ; partie
prison lieu
procès procès
procès procès
procureur de la République protagoniste ; magistrat
procureur général protagoniste ; magistrat
proxénétisme infractions ; personne
recherche instruction ; action
recherché instruction ; action
récidive infractions ; circonstances
réclusion sanction ; peine
relaxe sanction ; absence de peine
réquisition procès
saisine instruction ; action
scellé instruction ; document
séquestration infractions ; personne
statue instruction ; action
sursis sanction ; peine
surveillance instruction ; action
témoin protagoniste ; partie
tentative infractions ; circonstances
TIG sanction ; peine
trafiquant infractions ; auteur
tribunal lieu
urgences médico-judiciaires lieu
verdict sanction ; décision
verser au dossier instruction ; action
victime procès
viol infractions ; personne
violence infractions ; personne
vol infractions ; biens
voleur infractions ; auteur
volontaire infractions ; circonstances

C.2 Index des termes en hindi

परिरोध sanction ; peine
अधिकारी protagoniste ; police
अधिवक्ता protagoniste ; aux
अपराध infractions ; personne
अपराधी protagoniste ; partie
अपहानि **apahāni** infractions ; préjudice
अफसर protagoniste ; police
अभियुक्त procès
अभियोजन पक्ष procès
अश्लील infractions ; personne
असंज्ञेय मामला instruction ; procédure
आक्रमण infractions
आतंक infractions ; personne
आरोपी protagoniste ; partie
आशय infractions ; circonstances
उत्पीड़न infractions ; personne
एकांत परिरोध sanction ; peine
कचहरी lieu
कठिन कारावास sanction ; peine
कत्ल|कतल infractions ; personne
कानूनी सहायता / विधिक सहायता procès
कारावास sanction ; peine
कार्य infractions ; hétéronymes
कूटरचना ; कूटकरण infraction ; biens
कोतवाल protagoniste ; police
कोर्ट lieu
क्षति **kṣati** infractions ; préjudice
क्षोभ **kṣobh** infractions ; préjudice
गवाह protagoniste ; partie
गिरफ्तार instruction ; action
गिरफ्तारी|गिरफ़्तारी sanction ; restriction
चाकू infractions ; circonstances
चालान instruction ; document
चोरी infractions ; biens
छापा instruction ; action
छुरी infractions ; circonstances
ज़मानत|जमानत sanction ; restriction
ज़मानतीय|जमानतीय|जमानती instruction ; procédure
जाँच instruction ; action
जूरी procès
जूरी–सदस्य protagoniste ; aux
जेल lieu
जेलर protagoniste ; prison
डकैती infractions ; personne
डाकू infractions ; circonstances
तलाश instruction ; action
तलाशी instruction ; action
तस्कर infractions ; personne
तहरीर instruction ; document
तामील instruction ; action
तंग infractions ; personne
थानेदार protagoniste ; police
दहेज infractions ; personne
दुष्प्रेरण infractions ; circonstances
दोषमुक्ति sanction ; absence de peine
दोषसिद्ध **doṣasiddha** sanction ; peine

दोषसिद्धि **doṣasiddhi** sanction ; peine
दोषी protagoniste ; partie
दंड sanction ; peine
दंडादेश sanction ; peine
पूर्वचिन्तन infraction ; circonstances
धमकी infractions ; personne
निर्णय / फ़ैसला procès
नुकसान infractions ; préjudice
न्यायाधीश protagoniste ; magistrat
न्यायालय lieu
पक्ष protagoniste ; partie
परिवाद instruction ; document
परिवादी / परिवादनी protagoniste ; partie
पिस्तौल infractions ; circonstances
पीड़ित procès
पैरवी procès
पंचायत lieu
प्रतिवाद procès
प्रतिवादी, प्रत्यर्थी procès
प्रतिषेध infractions
निवारण infractions
प्रतीपवाद procès
प्रयत्न infractions ; circonstances
प्रारंभिक जाँच instruction ; action
बन्दूक infractions ; circonstances
बरी sanction ; absence de peine
बलात्कार infractions ; personne
बलात्संग infractions ; personne
मकदमा instruction ; procédure
मजिस्ट्रेट protagoniste ; magistrat
मानव वध infractions ; personne
मानहानि infractions ; personne
मामला instruction ; procédure
मिथ्याकरण infraction ; biens
याचिका procès
यौन अपराध infractions ; personne
रपट instruction ; document
रिपोर्ट instruction ; document
रिष्टि infractions ; biens
रिहाई sanction ; absence de peine
रंगे हाथ infractions ; circonstances
लूट; अद्यापन infractions ; biens
लोक अभियोजक protagoniste ; magistrat
वकील protagoniste ; aux
वाद procès
विचारण procès
व्यपहरण / अपहरण infractions ; personne
षड्यंत्र infractions ; biens
सज़ा|सजा **sazā** sanction ; peine
सह–अपराधी protagoniste ; partie
साक्षी protagoniste ; partie
साथी protagoniste ; partie
साशय infractions ; circonstances
सुनवाई procès
संज्ञेय मामला instruction ; procédure
हत्या infractions ; personne
हथियार infractions ; circonstances
हमला infractions ; personne
हवालदार protagoniste ; police
हवालात sanction ; restriction
हानि **hāni** infractions ; préjudice
हिरासत sanction ; peine
हिंसा infractions ; personne

C.3 Index des termes en ourdou

Attention : pour l'ordre alphabétique, la première colonne de l'index est celle de gauche.

instruction ; document ابتدائی طلاع/FIR
infractions ; personne اذیت;تنگ
infractions ; circonstances ارادتاً
instruction ; action استدعا
instruction ; action استغاثہ
infractions ; circonstances اسلحٰ
instruction ; action اشتہاری
infractions ; circonstances اعانت
infractions ; personne اغوا
protagoniste ; police افسر
infractions ; circonstances اقدام
infractions ; hétéronyme امر واقعات
infractions ; préjudice ایذا
infractions ; circonstances بالارادہ
procès بحث
instruction ; action برآمد
sanction ; absence de peine **barī** بری
sanction ; absence de peine بریت
infractions ; circonstances بندوق
infractions ; personne بھاگ لے جانا;لے بھاگنا
instruction ; document بیان
infractions ; circonstances پستول
protagoniste ; police پولیس
infractions ; personne پیش بین
instruction ; action تحقیقات
infractions ; personne تشدد
instruction ; action تعمیل
instruction ; action تفتیش
infractions ; personne تقلیف
instruction ; action تلاش
instruction ; action تلاشی
lieu تھانہ
protagoniste ; police **thānedār** تھانیدار
protagoniste ; magistrat **jaj** جج
infractions ; personne جس بیجا/جس بے جا
infractions ; biens جعل نویس

فردِجرم instruction ; document

فریق،فریقین protagoniste ; partie

فعل۔افعال infractions ; hétéronyme

فیصلہ procès

قانونی مدد / قانونی امداد procès

قتل infractions ; personne

قتل انسان infractions ; personne

قتل عمد infractions ; personne

قید lieu

قید sanction ; peine

قید بامشقت sanction ; peine

قید تنہائی sanction ; peine

کچہری lieu

کورٹ lieu

کوشش infractions ; circonstances

گرفتار instruction ; action

گرفتاری instruction ; action

گرفتاری instruction ; action

گشت instruction ; action

گواہ protagoniste ; partie

گواہی instruction ; action

مجرم protagoniste ; partie

مجسٹریٹ instruction ; protagoniste

مجسٹریٹ protagoniste ; magistrat

مدعاعلیہ protagoniste ; partie

مدعی protagoniste ; partie

مرتکب protagoniste ; partie

مستغیث protagoniste ; partie

معائنہ instruction ; action

معاملہ instruction ; procédure

مقدمہ instruction ; procédure

ملزم procès

ملزم protagoniste ; partie

نالش protagoniste ; magistrat

نقصان infractions ; préjudice

نقصان رسانی infractions ; biens

ہتھیار infractions ; circonstances

واردات infractions ; hétéronyme

واقعات infractions ; hétéronyme

وکیل/وکلا protagoniste ; aux

جہیز infractions ; personne

جیل lieu

جیلر **jelar** ; prison ; protagoniste

جیوری protagoniste ; aux

چاقو infractions ; circonstances

چالان instruction ; document

چھاپہ instruction ; action

چھری infractions ; circonstances

چوری infractions ; biens

چوکی lieu

حراست sanction ; peine

حملہ infractions ; personne

حوالات instruction ; action

حوالات sanction ; restriction

حوالدار protagoniste ; police

خصرت infractions ; personne

دریافت instruction ; action

دکھ infractions ; personne

دھمکی infractions ; personne

رپٹ instruction ; document

رپورٹ instruction ; document

رضامندی infractions ; circonstances

رہائی sanction ; absence de peine

ریمانڈ instruction ; action

زنابالجبر:زناءبالجبر infractions ; personne

زیادتی infractions ; personne

سائل protagoniste ; partie

سازش infractions ; biens

سرقہ infractions ; biens

سزا sanction ; peine

سزائے موت sanction ; peine

سماعت procès

شکایت instruction ; action

ضرر infractions ; préjudice

ضمانت sanction ; restriction

ضمانتی instruction ; procédure

عدالت lieu

عمرقید sanction ; peine

فحش infractions ; biens

فرار instruction ; action

Annexe D

Tableaux monolingues des cooccurrences et unités phraséologiques par réseau notionnel

Les tableaux ci-après reprennent langue par langue les termes, cooccurrents et unités phraséologiques notables par réseau notionnel. Le corpus d'occurrence du terme ainsi que celui d'occurrence de l'unité phraséologique sont indiqués dans les deuxième et quatrième colonnes. Lorsqu'une unité est utilisée exclusivement dans certains corpus, il en est fait mention entre parenthèses à côté. Le signe + à côté d'une unité phraséologique ou d'un terme précise que ces derniers sont surreprésentés dans le corpus mentionné. Des remarques précisent si l'utilisation d'un terme est limitée à un contexte précis. Dans les tableaux en hindi et en ourdou, la dernière colonne présente des équivalences de traduction en français.

Enfin, signalons que pour chaque langue, le classement est alphabétique, à partir de la colonne des termes.

D.1 Infractions

Les tableaux ci-après reprennent les termes, cooccurrents et unités phraséologiques en français, hindi et ourdou pour le réseau notionnel des infractions.

français

Terme	Remarque, notamment corpus d'occurrence	Cooccurrent	Unité phraséologique (corpus spécifique)
acte de malveillance	presse		
agression	CP, CPP	sexuelle	agression sexuelle
arme	tous textes	menace usage autorisation	usage ou menace d'une arme (CP) arme soumise à autorisation
armé	tous textes		
assassinat	CP, CRI	complicité	complicité d'assassinat
blessure	CP	mort entraîné	la mort ou la blessure de (CP) avoir entraîné une blessure (CP)
contravention	tous textes	prévue réprimée classe	contravention de la 5e classe contravention prévue à l'article contravention réprimée conformément à l'article
couteau	tous textes		
crime	tous textes	prévu puni indirect	produit direct ou indirect d'un crime
dague	tous textes		
délit	tous textes	commis prévu flagrant	délit commis délit prévu délit flagrant
diffamation	tous textes		diffamation envers diffamation non publique
dommage	tous textes	léger corporel causer occasionner réparation résulter	dommage léger dommage corporel dommages (matériels) causés occasionner le dommage réparation du dommage dommage résultant, il en est résulté un dommage
enlèvement		séquestration	enlèvement et séquestration / ou séquestration
extorsion	CP, CPP, presse	fonds en bande organisée	extorsion de fonds extorsion en bande organisée

français

fait	tous textes	commis	les faits (ont été) commis
		prévu	faits prévus et réprimés par les articles
		réprimé (169/5)	faits (qui lui sont) reprochés
		reproché	l'auteur des faits
		auteur	faits de l'espèce (CPP) / en l'espèce les faits
		espèce	(juridictionnel)
harcèlement	tous textes	moral	harcèlement moral
		sexuel	harcèlement sexuel
homicide	tous textes	volontaire	homicide volontaire
		involontaire	homicide involontaire
incitation	CP, CPP, presse	constituer	constitue une incitation
		commettre	incitation à commettre une infraction
		consommation	incitation à la consommation de (presse)
intention	juridictionnel		sans intention de
			manifester son intention
intentionnel	juridictionnel	élément	l'élément intentionnel
menace		terroriste	menace terroriste
		mort	menace de mort
meurtre	CP, CRI	préméditation	meurtre commis avec préméditation
nuisance	CPP	déroulement	nuire au bon déroulement (CPP)
		nature	de nature à nuire à (CPP)
		dignité	nuire à la dignité (CPP)
		intérêts	nuire aux intérêts (CPP)
perte	CPP	connaissance	perte de connaissance
		valeur	perte de la valeur (d'un bien)(CPP)
pistolet	tous textes		
proxénétisme	tous textes	aggravé	proxénétisme aggravé
récidive	tous textes	légale	récidive légale
		état	état de récidive
séquestration	tous textes	enlèvement	enlèvement et séquestration / ou séquestration
tentative	tous textes	délit	tentative du délit / tentative de l'infraction
		punie	tentative (…) est punie des mêmes peines
		constituée	tentative est constituée
		manifestée	tentative manifestée par un commencement d'exécution
		poursuivi	poursuivi pour tentative de
		complicité	complicité de tentative de
trafiquant	tous textes		

français

viol	presse	présumé	viol présumé (presse)
violence	CPP, juridictionnel	physique avec volontaire	violence physique violence sexuelle avec violence violence volontaire
vol	CP, CPP, juridictionnel	aggravé violence recel organisé	vol aggravé vol avec violence recel de vol vol organisé
volontaire	CP, CPP	homicide comparution atteinte violence	homicide volontaire comparution volontaire atteinte volontaire (à la vie) violence volontaire

hindi - infractions

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
अपराध aparādh	Tous textes	दंडनीय संज्ञान अज़मानतीय यौन	दंडनीय अपराध अपराध का संज्ञान अज़मानतीय अपराध यौन अपराध (IPC, presse)	infraction répréhensible qualification d'une infraction infraction sans possibilité de libération délit / crime sexuel
अपहानि apahāni	Lois	उत्पन्न स्वेच्छया	अपहानि उत्पन्न करना / होना अपहानि स्वेच्छया कारित करना	occasionner / subir un dommage occasionner volontairement un dommage
अश्लील aślīl	Tous textes	हरकतें	से अश्लील हरकतें	actes obscènes
आक्रमण ākramaṇ	Tous textes	पर	पर आक्रमण करना	agresser
आतंक ātaṅk		xxx	xxx	xxx
आशय āśay	Tous textes, une seule occurrence dans la presse	कपटपूर्ण सामान्य विमर्शित	के आशय से कपटपूर्ण आशय से सामान्य आशय विमर्शित आशय से	avec l'intention de avec une intention frauduleuse avec une intention commune avec l'intention délibérée
उत्पीड़न utpīṛan	Tous textes, seulement en exemple dans IPC	लैंगिक यौन दहेज	लैंगिक उत्पीड़न यौन उत्पीड़न दहेज उत्पीड़न	harcèlement sexuel harcèlement sexuel harcèlement pour la dot
उपापन ; दिलाई upāpan ; dilāī	Lois	लड़की	लड़की का उपापन	proxénétisme
कत्ल/कतल katal/katl	presse			meurtre
किरपाण kirpāṇ	xxx	से काटना		poignarder à l'aide d'une dague
कूटरचना ; कूटकरण kūṭaracanā ; kūṭakaraṇ	IPC, CrPC	छल	छल के प्रयोजन से कूटरचना	falsification, contrefaçon par la ruse

hindi - infractions

क्षति kṣati	Lois	पहुँचाना कारित शारीरिक	क्षति पहुँचाने से क्षति कारित होना शारीरिक क्षति	occasionnant une blessure occasionner une blessure blessure corporelle
क्षोभ kṣobh	Lois	उत्पन्न कारित	क्षोभ उत्पन्न करना (IPC) क्षोभ कारित करना (IPC) क्षोभ कारित होना	occasionner une nuisance occasionner une nuisance nuisance être occasionnée
चाकू cākū	presse / blog	हमला गोदना	चाकू से हमला चाकू से गोदकर हत्या करना	attaquer avec un couteau meurtre à coups de couteau
चोरी corī	Tous textes	करना	सामान चोरी करना	voler un bien
छुरी churī	IPC	वध	छुरी से वध करना	meurtre à coups de couteau
डकैती ḍakaitī	Tous textes			vol avec violence et en réunion (au moins 5 pers.)
डाकू ḍākū	IPC			voleur / brigand (IPC)
तंग taṅg	Tous textes, seulement en exemple dans IPC	करना	तंग करना	harceler
तस्कर taskar	presse	मानव	मानव तस्कर	trafiquant d'êtres humains
दहेज dahej	FIR, IPC, presse	मांग, मांगना प्रतिषेध प्रताड़ित	दहेज की मांग, दहेज मांगना दहेज प्रतिषेध अधिनियम दहेज के लिए प्रताड़ित	demander une dot loi d'interdiction de la dot maltraitée pour la dot
दुष्प्रेरण duṣpreraṇ		पूरवक toutes infractions	दुष्प्रेरण पूर्वक आपराधिक षडयंत्र (Blog)	entente criminelle en vue d'inciter à
धमकी dhamkī	tous textes	मारना	जान से मारने कि धमकी	menace de mort
निवारण nivāraṇ	Tous textes	अपराध	अपराधों का निवारण	prévention
नुकसान nuksān	Tous textes	नुकसानग्रस्त आंशीक स्थायी	वस्तु को नुकसानग्रस्त करना आंशिक नुकसान स्थायी नुकसान	endommager un bien dommage partiel dommage permanent
परिरोध parirodh	Lois	सदोष	सदिष परिरोध	détention illégale

hindi - infractions

पिस्तौल pistaul	Tous textes	तानना चलाना	पर पिस्तौल तान देना (blog) पिस्तौल चलाना (IPC)	braquer une arme sur tirer au pistolet
पूर्वचिन्तन pūrvacintan	IPC, CrPC	बिना	पूर्वचिन्तन बिना	sans préméditation
प्रतिषेध pratiṣedh	Lois	दहेज आदेश	दहेज प्रतिषेध प्रतिषेध करनेवाला आदेश	interdiction la dot ordonnance interdisant …
प्रयत्न prayatna	Lois	toutes infractions		tentative
बन्दूक bandūk	Tous textes	चलाना	बन्दूक चलाना	tirer au fusil
बलात्कार balātkār	parajuridique	सामूहिक	सामूहिक बलात्कार	viol en réunion
बलात्संग balātsaṅg	CrPC, IPC	सामूहिक	सामूहिक बलात्संग	viol en réunion
मानव वध mānav vadh	IPC	आपराधिक	आपराधिक मानव वध	meurtre
मानहानि mānhāni	IPC, CrPC ; presse			diffamation
मिथ्याकरण mithyākaraṇ	IPC, CrPC	लेखा	लेखाओं का मिथ्याकरण	falsification des écritures comptables
रंगे हाथ raṅge hāth	presse	पकड़ना गिरफ्तार करना	किसी को रंगे हाथ पकड़ना / गिरफ्तार करना	arrêter en flagrant délit
रिष्टि riṣṭi		द्वारा	करने द्वारा रिष्टि	mischief
लूट lūṭ	Tous textes	रिष्टि आबरू	लूट रिष्टि (IPC) की आबरू लूट लेना (presse)	extorsion viol
व्यपहरण ; अपहरण vypaharaṇ ; apaharaṇ	Tous textes pour अपहरण, lois pour व्यपहरण.			enlèvement
षड्यंत्र ṣaḍyantra	IPC, CrPC	आपराधिक	आपराधिक षड्यंत्र	association de malfaiteurs
सम्मति sammati	IPC, CrPC	देना	सम्मति देना	donner son consentement
साशय sāśay	CrPC, IPC	लोप प्रवंचना + autres infractions	साशय लोप साशय प्रवंचना	omission intentionnelle tromper volontairement
हत्या hatyā	Tous textes	इरादतन रंजिशन हत्या	गैर इरादतन हत्या (presse) रंजिशन हत्या (presse)	meurtre involontaire (incohérent)

hindi - infractions

हथियार hathiyār	Tous textes	अवैध धारदार	अवैध हथियार (tout) धारदार हथियार (blog, presse)	arme illégale arme aiguisée
हमला hamlā	Tous textes, seulement en exemple dans IPC	पर	पर हमला करना	agresser
हानि hāni	Lois	पहुँचाना कारित सदोष	हानि पहुँचाना हानि कारित होना (IPC, CrPC) सदोष हानि पहुँचाना / सदोष हानि उठाना	occasionner une perte perte être occasionnée occasionner / subir une perte intentionnelle
हिंसा hiṁsā	Tous textes	घरेलू प्रयोग शिकार यौन पंथीय, संप्रादयिक	घरेलू हिंसा हिंसा का प्रयोग (forum) हिंसा की शिकार (presse) यौन हिंसा (presse) पंथीय हिंसा (IPC)	violence domestique usage de la violence victime de violence violence sexuelle violence sectaire / communautaire

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
اذیت; تنگ aẕiat ; tang	Lois + (aẕiat) presse + (tang)	ذہنی پہنچانا	ذہنی اذیت (presse) اذیت پہنچانا	harcèlement moral harceler
ارادتاً irādatan	IPC	تعون مدد ارتکاب	ارادتاً مدد ارتکاب میں ارادتاً تعون\مدد	aide intentionnelle (IPC) aide / soutien intentionnelle dans la commission [d'un délit / crime]
اسلح aslaḥ	presse			arme
اعانت aʿānat	IPC	حملہ بغاوت قصوروار	عانت کرنا عانت کا قصوروار بغاوت میں عانت، بغاوت کے ارتکاب میں عانت	inciter instigateur inciter / incitation à la révolte
اغوا aġvā	tous textes	کرنا	اغوا کرنا	enlever
اقدام aqdām	Lois	قتل	قتل کے اقدام پر اقدام قتل	lors d'une tentative d'homicide tentative d'homicide
اقدام aqdām	IPC	tous types d'infraction		tentative de + infraction
امتناع imtanāʿ	IPC	حکم	حکم امتناعی	injonction
بالارادہ bālirādah	IPC	ضرر پہنچانا + types d'infractions	بالارادہ ضرر پہنچانا	occasionner (intentionnellement) une blessure
بندوق bandūq	tous textes	چلانا	بندوق چلانا	fusil
بھاگ لے جانا; لے بھاگنا bhāg le jānā ; le bhāgnā	IPC			enlever

پستول pistaul	tous textes	فائر کرنا بٹ مسلح	پستول فائر کرنا (tous genres) پستول کا بٹ مسلح پستول	tirer au pistolet crosse du pistolet armé d’un pistolet
پیش بین peś bīn	IPC	بلا	بلا پیش بین	sans préméditation
تشدد taśadud	IPC, presse	نشانات بہیمانہ جنسی	تشدد کے نشانات (presse) بہیمانا تشدد (presse) جنسی تشدد (presse, parajuridique)	parajuridique
تقلیف taqlīf	Tous textes	امر باعث	امر باعث تقلیف	nuisance (IPC)
جس بیجا jas bejā	IPC	رکھن	جس بیجا میں رکھنا	séquestrer
جعل نویسی jaʿl navīsī	IPC			falsification
جہیز jahez	IPC	موت	جہیز موت	assassinat pour la dot
چاقو cāqū	IPC	ہلاک	چاقو سے ہلاک کرنا (IPC)	tuer au couteau
چھری churī	presse	مارنا	چھری مارنا (presse)	donner un coup de couteau
چوری corī	presse	شدہ بجلی	چوری شدہ سامان (presse) بجلی چوری کرنا	bien volé vol d’électricité
حملہ ḥamlah	Tous textes	پر	پر حملہ کرنا	agresser (qqn)
دکھ dukh	IPC, presse			gêne (IPC)
دلائی dilāī	IPC	لڑکی	لڑکی کی دلائی	proxénétisme
دھمکی dhamkī	tous textes	جان	جان سے مارنے کی دھمکی مار دینے کی دھمکی	menace de mort (presse)

ourdou - infractions

رضامندی rażāmandī	tous textes	بغیر	رضامندی کے بغیر	sans consentement
زنا باجبر; زناء باجبر zinā bāljabar /zinā' bāljabar	Lois			viol
زیادتی zyādatī	presse	اجتماعی	زیادتی اجتماعی	viol en réunion (presse)
سازش sāziś	tous textes	مجرمانہ	سازش مجرمانہ	association de malfaiteurs
سرقہ sarqah	IPC	بالجبر مشترکہ	سرقہ بالجبر مشترکہ طور پر سرقہ	extorsion / vol avec violence vol en réunion [kośiś plutôt utilisé ds les exemples, mais pas exclusivement.]
ضرر żarar	Lois	بالارادہ شدید پہنچانا ہلاکت	بالارادہ ضرر (شدید) پہنچانا ہلاکت یا ضرر (شدید) پہنچانا	occasionner (intentionnellement) une blessure grave occasionner la mort ou une blessure (grave)
فحش faḥaś	IPC, presse	حرکات	فحش حرکات	actes obscènes
فعل / افعال fa'l / af'āl	Evidence Act, IPC	اعنات	فعل / افعال میں اعنات	initier un fait
قتل انسان qatl-e-insān	Tous textes	مستلزم سزا	قتل انسان مستلزم سزا	homicide volontaire
قتل عمد qatl-e-'amd	Evidence Act, IPC			meurtre
کوشش kośiś	presse, CPI	tous types d'infraction		tentative
مسلح musallaḥ	tous textes	ہتھیار	ہتھیار سے مسلح ہونا (IPC)	être armé
ممانعت mumāna't	IPC			interdiction

ourdou - infractions

		رسانی	نقصان رسانی	malveillance
نقصان	Lois	پہنچنا	نقصان رسانی کا ارتکاب	acte de malveillance
nuqṣān		پہنچانا	نقصان پہنچانے کی نیت	intention d’occasionner une nuisance
		ارتکاب	نقصان پہنچانے کا ارادہ	intention d’occasionner une nuisance
نقصان رسانی	IPC	کر کے	نقصان رسانی کر کے	par acte de malveillance
nuqṣān rasānī				
ہتھیار	IPC	مسلح	ہتھیار سے مسلح ہونا (IPC)	être armé
hathiyār				
واردات	presse (faits)	ملوث	واردات میں ملوث	impliqué dans les faits
vārdāt				
واقعات	tous textes	تنقیحی متعلقہ	تنقیحی واقعات	faits dont il est question
vāqiʿāt			متعلقہ واقعات	faits relatifs à

D.2 Mesures et sanctions

Les tableaux ci-après reprennent les termes, cooccurrents et unités phraséologiques en français, hindi et ourdou pour le réseau notionnel des mesures et sanctions.

Terme	Remarque, notamment corpus d'occurrence	Cooccurrent	Unité phraséologique (corpus spécifique)
acquittement	CP, CPP, presse, forums, FIJAIS	décision plaider	décision / arrêt d'acquittement plaider l'acquittement (presse)
amende	CPP	forfaitaire encourent	amende forfaitaire encourir une amende de
arrêt	CP, CPP CP, CPP, presse ambiguïté avec mandat d'arrêt	mandat rendu	mandat d'arrêt arrêt rendu
caution	CPP, presse, ORTC, forums	admise	caution admise par le comptable
condamnation	CP, CPP	décision entraîné	décision de condamnation ayant entraîné sa condamnation
contrôle judiciaire	tous textes	sous placement	placement sous contrôle judiciaire
décision	CP, CPP	définitive Motivée	décision définitive Décision motivée
détention	CP, CPP	criminelle libertés centre	détention criminelle juge des libertés et de la détention centre de détention
détention provisoire		placement / placer	placement / placé en détention provisoire
garde à vue	tous textes	prolongation placement / placé	prolongation de la garde à vue placement en garde à vue
jugement	CP, CPP CP, CPP CPP, ORTC	juridiction rendre définitif	juridiction de jugement rendre un jugement / jugement rendu jusqu'au jugement définitif
libération		conditionnelle	demande de libération conditionnelle
non-lieu	CPP CPP, forums forums, presse	partiel ordonnance bénéficier	non-lieu partiel ordonnance de non-lieu bénéficier d'un non-lieu
peine	CPP CPP CP	privative réduction prononcée encourue	peine privative de liberté réduction de peine peine prononcée / prononcé à titre de peine peine encourue / encouru à titre de peine

réclusion	CPP presse CP	criminelle encourt perpétuité	réclusion criminelle encourt la réclusion réclusion criminelle à perpétuité
relaxe	presse	générale	relaxe générale
sursis	CP, CPP	épreuve assortie révocation	sursis avec mise à l'épreuve peine assortie du sursis révocation du sursis
TIG	CP, CPP, forums, presse	effectuer	effectuer un travail d'intérêt général
verdict	Presse, forums	annonce rendre	annonce du verdict rendre son verdict / verdict rendu

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
दोषमुक्ति doṣamukti	IPC	निर्णय	दोषमुक्ति का निर्णय सुनाना	prononcer un acquittement / une relaxe
कारावास kārāvās	Lois	अवधि आजीवन	कारावास की अवधि आजीवन कारावास कारावास का दंडादेश कठिन कारावास (IPC)	durée d'emprisonnement prison à perpétuité condamnation à un emprisonnement détention avec travaux forcés
रिहाई rihāī	charte, blog, presse	जेल	जेल रिहाई	libération / sortie de prison
हवालात havālāt	uniquement presse	में डालना में रोकना	हवालात में डालना में रोकना	placer en détention
हिरासत hirāsat	presse et blog	में लेना न्यायिक पुलिस	हिरासत में लेना न्यायिक हिरासत पुलिस हिरासत	placer en détention détention judiciaire garde à vue
अभिरक्षा abhirakṣā	CrPC, IPC	में लेना न्यायिक पुलिस	अभिरक्षा में लेना न्यायिक अभिरक्षा पुलिस अभिरक्षा	placer en détention détention judiciaire garde à vue
गिरफ्तारी / गिरफ़्तारी giraftārī	tous textes	वारण्ट	गिरफ्तारी का वारण्ट	mandat d'arrêt
दंड daṇḍ	tous textes	मात्रा दंडनीय	दंड की मात्रा दंड से दंडनीय	quantum de la peine passible d'une peine
ज़मानत / जमानत zamānat	tous textes	छोड़ रिहा जमानत तुड़वाना	ज़मात पर छोड़ना ज़मात पर रिहा करना जमानत तुड़वाकर जेल जाना	libérer sous caution libérer sous caution annuler la libération conditionnelle
दोषसिद्ध doṣasiddha	Loi sur l'identification des prisonniers, CrPC	अपराध	अपराध के लिए दोषसिद्ध करना	condamner pour une infraction
दोषसिद्धि doṣasiddhi	CrPC, IPC, Loi sur l'identification des prisonniers	अपील पर	अपील पर दोषसिद्धि	condamnation en appel
दंडादेश daṇḍādeś	CrPC, IPC	कारावास पारित करना	कारावास का दंडादेश दंडादेश पारित करना	condamnation à la détention / réclusion condamner
सज़ा / सजा sazā	uniquement dans le parajuridique			peine
बरी barī	presse, forum	बरी होना / करना		relaxé / acquitté, relaxer / acquitter

D.3 Instruction

Les tableaux ci-après reprennent les termes, cooccurrents et unités phraséologiques en français, hindi et ourdou pour le réseau notionnel de l'instruction.

Français

Terme	Remarque, notamment corpus d'occurrence	Cooccurrent	Unité phraséologique (corpus spécifique)
aide juridique	CPP + forums	bénéficier	bénéficier de l'aide juridique
arrestation	tous textes	lieu	lieu d'arrestation
aux fins de	tous textes		
commission rogatoire		internationale agissant sur exécution	commission rogatoire internationale agissant sur commission rogatoire (entendu) sur CR exécution d'une CR
comparution	CPP, forums	devant immédiate première préalable	comparution devant le juge / la juridiction comparution imédiate (interrogatoire de) première comparution comparution sur reconnaissance préalable de culpabilité
contradictoire		débat signifier	débat contradictoire jugement contradictoire à signifier
déférer	tous textes	devant	être déféré devant un magistrat / une juridiction
déposition	CPP, forums, presse	témoin relire	déposition d'un témoin (invité à) relire sa déposition
élément		probant matériel	
enquête	843 occ, tous textes	préliminaire de flagrance menée	enquête préliminaire enquête de flagrance enquête menée
fouille	tous textes	intégrale	fouille intégrale
fuite	tous textes	prendre favoriser	prendre la fuite favoriser la fuite
inculpation	Presse (1 occ)	chef	chef d'inculpation
inculper	Presse (1 occ)		
instruction		juge chambre	juge d'instruction chambre de l'instruction
interception	CPP		
interpellation	tous textes	lors	lors de son interpellation

Français

mise en examen	tous textes textes juridictionnels	personne chefs	personne mise en examen mis en examen des chefs de
OMA	CPP, OMA		
ORTC	ORTC, CPP, presse		
perquisition	tous textes	effectuée découverte	perquisition effectuée à la perquisition amenait la découverte
pièce	CP, CPP	produite	pièce produite par / devant
plainte	tous textes	porter déposer contre	porter plainte déposer plainte plainte contre
recherche	tous textes	mandat vaines	mandat de recherche procès-verbal en vaines recherches
recherché	tous textes	fichier	fichier des personnes recherchées
saisine	tous textes	chambre juge	saisine de la chambre saisine du juge
scellé	tous textes	sous ouverture	placé sous scellé ouverture des scellés
statue	tous textes	sur contradictoirement juge d'instruction / cour	statuer sur statuer contradictoirement le juge d'instruction / la cour statue
surveillance	tous textes	électronique placement	surveillance électronique placement sous surveillance électronique
verser au dossier	tous textes	procédure copie	versée au dossier de la procédure copie est versée au dossier

Hindi

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
कानूनी सहायता / विधिक सहायता kānūnī sahāytā / vidhik sahāytā	*vidhik :* textes de loi (Avocats), *kanuni* : parajuridique, 1 occ. de chaque dans la Charte.	कानूनी विधिक		aide juridique / aide juridictionnelle
गिरफ्तार giraftār	tous textes			arrêté, interpeller
चालान cālān	presse	काटना करना	चालान काटना चालान करना	rédiger un rapport d'enquête préliminaire
छापा chāpā	presse	मारना	छापा मारना	perquisitionner
ज़मानतीय / जमानतीय / जमानती zamānatīya / jamānatīya / jamānatī	*forme zamānatī* parajuridique, *zamānatīya* lois	गैर वारण्ट	गैर जमानती (presse, blog) ज़मानतीय / जमानती वारण्ट तामील करना	sans possibilité de libération sous caution délivrer un mandat de libération sous caution
जाँच jāc	Loi	प्रारंभिक जाँच		enquête préliminaire
तलाश talāś	blog, presse			recherche
तलाशी talāśī	tous textes	अभियान लेना	तलाशी अभियान (presse) तलाशी लेना	fouille
तहरीर tahrīr	presse (1 occ.), FIR	देना मुकद्मा	घटना की तहरीर देना तहरीर पर मुकदमा दर्ज करना	faire une déposition enregistrer une affaire suite à une déposition
तामील tāmīl	tous textes			mis en œuvre
परिवाद parivād	CrPC, forum, loi contre la dot	लिखित	लिखित परिवाद	plainte écrite
मामला māmlā	tous textes	दर्ज करना (presse blog) संज्ञेय CrPC (2 occ.) असंज्ञेय CrPC (1 occ.)	मामला दर्ज करना संज्ञेय मामला असंज्ञेय मामला	ouvrir une information affaire de la compétence de la police affaire de la compétence du tribunal

Hindi

मुकदमा mukadmā	Loi Aparadh, presse	दर्ज करना कराना	मुकदमा दर्ज करना	ouvrir une information
रपट rapaṭ	presse	दर्ज करना		enregistrer un rapport / un PV
रिपोर्ट riporṭ	tous textes	रिपोर्ट दर्ज करना		enregistrer un rapport / un PV

Ourdou

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
فردِ جرم fard-e-jurm	IPC, CrPC	عائد کرنا عائد کرنا	فردِ جرم	qualifier des faits / mettre en examen
FIR / ابتدائی اطلاع رپورٹ FIR / ibtadāī riporṭ	FIR			rapport d'information rpéliminaire
استدعا isted'ā	tous textes	کرنا درخواست مراسلہ	استدعا کرنا درخواست استدعا مراسلہ استدعا (seulement pour l'étranger)	faire une requête requête commission rogatoire internationale
استغاثہ isteġāṯah	tous textes	کرنا	استغاثہ کرنا	porter plainte
اشتہاری iśtihārī	tous textes			recherché
برآمد barāmad	IPC, presse	ہونا	برآمد ہونا	a été trouvé
بیان bayān	tous textes	جھوٹا بیان	جھوٹا بیان حلفاً بیان	fausse déposition / déclaration déclaration sur l'honneur
تحقیقات taḥqīqāt	tous textes	مکمل	تحقیقات مکمل ہونے پر	à la clôture de l'instruction
تعمیل ta'mīl	tous textes	کرنا	تعمیل کرنار	mettre en œuvre
تفتیش taftīś	tous textes			instruction
تلاش talāś	EA, jugement			recherche
تلاشی talāśī	IPC, presse			fouille

چالان cālān	presse	کرنا	چالان کرنا	rédiger un rapport préliminaire / enregistrer une affaire
چھاپہ chāpā	presse	مارنا	چھاپہ مارنا	perquisitionner
حوالات ḥavālāt	presse			détention / garde à vue
دریافت daryāft	presse, 1 occ ds IPC			enquête
دست اندازی dast andāzī	IPC, CrPC		قابلِ دست اندازی ناقابلِ دست اندازی	relevant de la compétence de la police relevant de la compétence du tribunal
رپٹ rapaṭ	presse			rapport / procès-verbal
رپورٹ riporṭ	tous textes			rapport / PV
ریمانڈ rīmānḍ	remand	جسمانی	جسمانی ریمانڈ	détention / garde à vue
سراغ surāġ	presse			indice
شکایت śikāyat	tous textes	کنندہ	شکایت کنندہ (presse, vulg)	plaignant
شہادت śahādat	tous textes			témoignage, preuve
ضمانتی żamānatī	presse			relatif à la caution
فرار farār	tous textes			en fuite
قانونی مدد / قانونی امداد qānūnī madad / imdād	3 occ. dans la presse, 2 *imdad*, 1 *madad*.			aide juridique / aide juridictionnelle
گرفتار giraftār	IPC			arrêté, interpellé

Ourdou

گرفتاری giraftārī	tous textes	وارنٹ	وارنٹ گرفتاری	mandat d’arrêt
گشت gaśt	FIR, presse	دوران	دوران گشت	pendant une patrouille
گواہی gavāhī	IPC	جھوٹی	جھوٹی گواہی	faux témoignage
معائنہ muʿāinah	jugement, presse, général. Peu spécialisé			examen (enquête)
معاملہ mʿāmlah	tous textes, peu spécialisé			affaire
معلومات maʿlūmāt	tous textes, peu spécialisé			élément, information
مقدمہ muqadmah	essentiellement dans la presse.	درج خلاف بر آمد	کے خلاف مقدمہ درج کرنا	ouvrir une affaire / une information contre …

D.4 Mécanique du procès

Les tableaux ci-après reprennent les termes, cooccurrents et unités phraséologiques en français, hindi et ourdou pour le réseau notionnel de la mécanique du procès.

Français - procès

Terme	Remarque, notamment corpus d'occurrence	Cooccurrent	Unité phraséologique (corpus spécifique)
accusation	tous textes	mise objet	mise en accusation devant la juridiction objet de l'accusation
accusé	tous textes	reconnu	accusé reconnu coupable / non coupable
audience	tous textes	se présenter huis clos publique renvoi fond	se présenter à l'audience comparaître à l'audience audience à huis clos audience de renvoi / renvoi d'audience audience sur le fond
défense	tous textes	exercice droits moyens	droits de la défense exercice des droits de la défense moyens de sa défense
huis-clos	CPP	audience	audience à huis clos procès à huis clos
jury	CPP, presse	populaire assises président	jury populaire jury d'assises président du jury
ministère public	tous textes (lois +)	réquisition requête	les réquisitions du ministère public à la requête du ministère public
partie civile	tous textes	constitution	constitution de partie civile
plaidoirie	CPP, ORTC, presse, forums	entendu plaidoirie	me a été entendu en sa plaidoirie plaidoirie de l'avocat
poursuite	tous textes	aux fins de / des fins de acte de poursuite des infractions	aux fins de poursuite à des fins de poursuite acte de poursuite poursuite des infractions de
procès	presse	renvoyer ajourné	renvoyer un procès ajourner un procès (dans la langue de spécialité, on dira que l'audience / l'affaire a été ajournée / renvoyée)
procès	tous textes (parajuridique +)	renvoyé	procès renvoyé

Français - procès

réquisition	tous textes	prendre ministère public	prendre ses réquisitions
victime	tous textes	infraction plainte	victime de l’infraction plainte de la victime

Hindi - procès

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
निर्णय / फ़ैसला nirṇay / faislā	loi / tous textes	न्यायालय सुनाना	न्यायालय का निर्णय फ़ैसला सुनाना	nirnay partout, faisla uniquement ds la presse
पैरवी pairvī	FIR, presse	केस / मामला	केस / मामले की पैरवी करना	
वाद vād	loi	लाना संस्थित	किसी पर वाद लाना वाद संस्थित करना	maintenir des poursuites au nom de Initier / lancer des poursuites
अभियुक्त abhiyukta	loi			défense
अभियोजन abhiyojan	presse	पैरवी	अभियोजन पक्ष की ओर से पैरवी	poursuite par l'accusation
पक्ष pakṣa	tous textes	अभियोजन	अभियोजन पक्ष की ओर से पैरवी	poursuite par l'accusation
पीड़ित pīṛit	CrPC Presse	व्यक्ति की शिकायत	पीड़ित व्यक्ति पीड़ित की शीकायत	victime plainte de la victime
प्रतिवाद prativād	loi, juridictionnel	करना	प्रतिवाद करना	se défendre
सुनवाई sunvāī	Loi sur les avocats, CrPC, presse	मामला	मामले की सुनवाई	audience du procès
जूरी	IPC	सदस्य	जूरी-सदस्य	juré / membre du jury
विचारण vicāraṇ	IPC, CrC (CrPC)	विषय आगे चलाने	विचारण का विषय विचारण आगे चलाना	affaire à juger poursuivre un procès
याचिका yācikā	tous textes	दया इजाज़त दायर करना को ठुकरा देना	दया याचिका इजाज़त याचिका याचिका दायर करना याचिका को ठुकरा देना	requête requête écrite enregistrer une requête rejeter une requête
बहस bahas	CrPC	मौखिक	मौखिक बहस (CrPC)	plaidoirie

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
ملزم mulzim	tous textes	مبینہ، مذکورہ کے خلاف	مبینہ ملزم مذکورہ ملزم ملزم کے خلاف (سخت) کارروائی	le mis en cause susmentionné le mis en cause susmentionné poursuites (sévères) à l'encontre du mis en cause
سماعت samaʿt	tous textes	مقدمہ زیرِ ناقابل قابل کے روسے قانون	سماعت مقدمہ (presse) زیرِ سماعت ناقابل سماعت قابل سماعت قانون کی روسے سماعت	audience du procès en attente d'audience / du délibéré non audiençable audiençable poursuite dans le cadre de la loi
بحث baḥaṯ	loi	حتمی ابتدائی	حتمی بحث ابتدائی بحث	plaidoirie finale plaidoirie introductive
فیصلہ faiṣlah	presse	محفوظ کرنا	فیصلہ محفوظ کرنا	confirmer une décision

D.5 Lieux de justice

Les tableaux ci-après reprennent les termes, cooccurrents et unités phraséologiques en français, hindi et ourdou pour le réseau notionnel des lieux de la justice, du dépôt de plainte jusqu'à la peine de prison.

Français - lieux

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)
centre pénitentiaire	CPP, presse		
commissariat	tous textes	de police	commissariat de police
cour	CP	d'appel d'assises ressort statuer	cour d'appel cour d'assises le ressort de la cour d'appel la cour statue sur
juridiction	CP, CPP	saisie jugement devant	juridiction saisie juridiction de jugement devant une juridiction
maison centrale	forums, CPP, presse	incarcéré	incarcéré dans une maison centrale
maison d'arrêt	presse, CPP, juridictionnel	quartier des mineurs centre	quartier des mineurs de la maison d'arrêt maison d'arrêt du centre pénitentiaire
prison	CPP, forum	ferme sursis surveillants	prison ferme prison avec sursis (forums, presse) surveillants de prison (presse)
tribunal	CP	correctionnel de grande instance devant	tribunal correctionnel tribunal de grande instance devant le tribunal
urgences médico-judiciaires	ONL		

Hindi -lieux

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
कचहरी kacahrī	12 occurrences, presse			tribunal (bâtiment abritant le tribunal)
कोर्ट korṭ	vulgarisation	सुप्रीम	सुप्रीम कोर्ट	Cour suprême
जेल jel	tous textes	बन्द भेजना सेन्ट्रल	जेल में बन्द जेल भेज देना सेन्ट्रल जेल	emprisonné / détenu envoyer en prison maison / prison centrale
न्यायालय nyāyālaya	tous textes	उच्च सर्वोच्च उच्चतम	उच्च न्यायालय (Loi sur les avocats / CrPC) सर्वोच्च न्यायालय (CrPC, histoire du droit) उच्चतम न्यायालय (adhivakta / CrPC / presse)	Haute Cour Cour suprême Cour suprême
पंचायत pañcāyat	presse, sans précision (Nyaya Panchayat ou autre)			Panchayat (judiciaire)

Ourdou - lieux

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
عدالت ʿadālat	tous textes	عظمیٰ عالیہ	عدالتِ عظمیٰ عدالت عالیہ	Cour suprême Haute Cour
تھانہ thānah	FIR, presse	SHO	تھانہ SHO	officier en charge du commissariat
جیل jel	dans l'IPC, transcription de l'anglais. Sinon, *qaid*			prison
چوکی caukī	presse	انچارج	چوکی انچارج	responsable du poste de police
قید qaid	IPC	عمر محض	عمر قید قید محض	réclusion à perpétuité simple détention (sans travaux forcés)
کچہری kacahrī	presse, synonyme de *ʿadālat*	ضلع جوڈیشل مجسٹریٹ	ضلع کچہری کچہری میں جوڈیشل مجسٹریٹ کی عدالت کچہری کی عدالتیں	tribunal de district tribunal du magistrat judiciaire, dans l'enceinte du complexe judiciaire tribunaux dans le complexe judiciaire
کورٹ korṭ	tous textes	ڈسٹرکٹ سپریم (tous textes) سیشن (presse)	ڈسٹرکٹ کورٹ سپری کورٹ سیش کورٹ	tribunal de district Cour suprême cour d'assises

D.6 Protagonistes

Les tableaux ci-après reprennent les termes, cooccurrents et unités phraséologiques en français, hindi et ourdou pour le réseau notionnel des protagonistes de la justice (justiciables, professionnels du droit et auxiliaires de justice).

Français - protagonistes

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)
auteur	tous textes	infraction faits	auteur de l'infraction / des faits
avocat	tous textes	assister office barreau représenter	se faire assister par un avocat (COPJ, CPP) se faire représenter par un avocat (CPP) avocat désigné d'office (CPP)
avocat général	presse, CPP	réquisitoire	
bâtonnier	tous textes	avocats office ordre	bâtonnier de l'ordre des avocats (CPP) désigné d'office par le bâtonnier (CPP)
complice	tous textes	crime	complice d'un crime
expert	tous textes		expert près la cour d'appel
greffier	CPP		
juge	tous textes	instruction libertés application	juge d'instruction juge des libertés et de la détention juge d'application des peines
juré	CPP		
magistrat	CPP	saisi chargé instructeur	magistrat saisi magistrat chargé magistrat instructeur
mis en examen	tous textes, presse +		
partie	CPP	civiles	parties civiles
plaignant	CPP	requête	requête des parties
prévenu	tous textes, CPP+	comparution	comparution du prévenu
procureur de la République	tous textes	réquisition	réquisitions prises par le PR (CPP)
procureur général	CPP		procureur général près la cour d'appel
témoin	CPP	assisté	témoin assisté

Hindi - protagonistes

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
अधिकारी adhikārī	jugement, FIR, loi	पुलिस भारसाधक अन्वेषण पीठासीन	पुलिस अधिकारी भारसाधक अधीकारी अन्वेषण अधिकारी पीठासीन अधिकारी	officier de police officier en charge officier enquêteur officier responsable
अधिवक्ता adhivaktā	loi sur les avocats+, presse+, jugement	शासकीय	शासकीय अधिवक्ता	procureur
अपराधी aparādhī	blog police	गिरफ़्त	अपराधी पुलिस की गिरफ़्त में होना	le mis en cause a été appréhendé par la police
अफसर afsar	presse, vulgarisation	पुलिस	पुलिस अफसर	officier de police
आरोपी āropī	CrPC	दोषसिद्ध	अपराधी को दोषसिद्ध करना	condamner un prévenu / accusé
कोतवाल kotvāl	presse			officier de police
गवाह gavāh	Forum, presse	पेश करना बनना	गवाह पेश करना गवाह बनना	présenter un témoin devenir témoin
जूरी jūrī	IPC	सदस्य	जूरी-सदस्य	juré / membre du jury
जेलर jelar	IPC, CRPC, presse			gardien de prison
थानेदार thānedār	presse (2 occ.)	पूर्व	पूर्व थानेदार	ancien responsable du commissariat
दोषी doṣī	Loi, presse	अपराध पाना	अपराध करने का दोषी उस व्यक्ति को (अपराध का) दोषी पाया है	auteur, coupable
न्यायाधीश nyāyādhīś	tous textes	ज़िला (अपर) सेशन पीठासीन	ज़िला न्यायाधीश सेशन न्यायाधीश पीठासीन न्यायाधीश / न्यायाधीश का पीठासीन	juge de district juge des assises président de la formation de jugement

Hindi - protagonistes

पक्ष pakṣa	presse	पीड़ित अभियोजन	पीड़ित पक्ष अभियोजन पक्ष	partie civile accusation
परिवादी / परिवादनी parivādī / parivādanī	CrPC, FIR, presse	पैरवी	परिवादनी का पैरवी करना	assurer la défense
मजिस्ट्रेट majisṭreṭ	IPC / CrPC	न्यायिक	मजिस्ट्रेट के समक्ष न्यायिक मजिस्ट्रेट	devant le magistrat magistrat judiciaire
लोक अभियोजक lok abhiyojak	tous textes	नियुक्त	लोक अभियोजन नियुक्त करना	désigner un procureur
वकील vakīl	presse, loi sur les avocats, forums	सरकारी	सरकारी वकील	avocat représentant un organisme public
सह-अपराधी sah-aparādhī	Loi, 3 occ en tout			complice, co auteur
साक्षी sākṣī	CrPC, Loi avocats	परीक्षा बनना हाज़िर होना	साक्षी का परीक्षा	interrogatoire du témoin
साथी sāthī	presse, blog			complice
हवालदार havāldār	presse			sergent

Ourdou - protagonistes

Terme	Corpus d'occurrence du terme	Cooccurrent	Unité phraséologique (corpus spécifique)	Équivalent traductionnel de l'U.P.
وکیل / وکلا vakīl / vaklā	tous textes	درخواست گزار موقِف اختیار کرنا استدعا کرنا	درخواستگزار (کا) وکیل وکیل کا موقِف اختیار کرنا وکیل کا استدعا کرنا	l'avocat du requérant Le point de vue choisi par l'avocat La requête de l'avocat / l'avocat plaide que
افسر afsar	tous textes	تفتیشی	تفتیشی افسر	officier enquêteur
پولیس polīs	tous textes	افسر	پولیس افسر	officier de police
تھانیدار thānedār	presse			chef de commissariat
جج jaj	tous textes	سیشن ایڈیشنل	سیشن جج ایڈیشنل جج	juge des assises juge adjoint / assesseur
جیلر jelar	IPC			gardien de prison
جیوری	IPC	اہل	اہل جیوری	membre du jury
حوالدار ḥavāldār	presse	ڈیوٹی	ڈیوٹی حوالدار	sergent de permanence
سائل sāil	FIR, presse, témoignage			déposant, plaignant
فریق، فریقین fareq, fareqīn (une seule occ de pl. rég. *fareqõ*)	tous textes	دلائل	فریقین کے دلائل (presse) حتمی دلائل	arguments des parties argument final
گواہ gavāh	tous textes			témoin
مجرم mujrim	tous textes	اشتہاری کرنا	مجرم اشتہاری کرنا	émettre un avis de recherche contre

Ourdou - protagonistes

مجسٹریٹ	presse	جوڈیشل کچہری	جوڈیشل مجسٹریٹ مجسٹریٹ کی کچہری	magistrat judiciaire tribunal du magistrat
مدعاعلیہ muda‘ā ‘alaiyah	3 occ. dans la presse			mis en cause
مدعی muda‘ī	presse	مقدمہ کا مدعی پیروی		requérant
مرتکب murtakib	tous textes		کا مرتکب + infraction	auteur de + infraction
مستغیث mustaġīs	FIR, droits des femmes	تحریری	تحریری درخواست مستغیث	plaignant
ملزم mulzim	tous textes	مسمی زیر حراست حراستی (presse)	ملزم مسمی زیر حراست ملزم حراستی ملزم	mis en cause dénommé mis en cause détenu mis en cause détenu
نالش nāliś	Lois, juridictionnel			juge

Index

densité, 160

puga, 15

abhiyojan, 114
acquittement, 108
adalat khana, 18
adālat murafiah, 18
AFNOR, 81
ahadī, 18
Ain-i-Akbari, 20
amin, 17
amorces linguistiques, 220
Anglo-Hindu law, 20
Anglo-Muhammadan law, 20
apabhraṃśa, 49
approche probabiliste, 194
approche transformationnelle, 194
ardhatatsam, 76
arrêt, 108
arthavivāda, 14
audition de première comparution, 103
avis à partie, 104
avocat général, 108

banyan, 19
barqandazi, 18
barreau, 111
bitexte, 2
brahmane, 11
brahmanes, 11
brāhmī, 46
BSTT, 57
bâtonnier, 111

chalan, 113
chaudhry, 17
chaukidar, 18
chefs de prévention, 103
Chief Justice, 28
code de procédure pénale, 9
code pénal
  Indian Penal Code (IPC), 9
  indien, 9
collocation, 231
commission d'expertise, 112
conseil, 111
contradictoires, 103
contravention, 101
contrôle judiciaire, 103
cour d'assise, 25
crime, 101
criminal law, 41
Criminal Procedure Code, 25
CrPC, voir Criminal Procedure Code

daftarī, 17
Dareecha, 174
data driven, 164
daṇḍ saṁhitā, 25
deśī, 50
devanagari, 21
dharma, 10
Dharmashastra, 20
dharmasūtra, 11
diyat, 32
droit islamique, 16
  fiqh, 16
dubash, 19
délit, 101
détention provisoire, 103
díkē, 38

EMILLE, 175
enquête, 100

fard-e-jurm, 113
fas, 38
FATA, 36
faujdārī, 18
FIJAIS, 129
formation de jugement, 101, 104
Fort William College, 78

gaṇa, 15
garantie de représentation, 103
gavāh, 113
Government of India Act, 31

hadd, 17
Hidden Markov Models, 194
hindavī, 49
hinduī, 49
hindvī, 49
HTML, 185
huissiers audienciers, 111
huissiers de justice, 111
ḥirāsat, 18

indice de spécificité, 207
information judiciaire, 101
interjeter, 27
interjeter appel, 107
isolement, 108
isteġāṣah, 114
istiġasah, 18

jirgah, 36
juge, 37
juge d'instruction, 101
jurisprudence, 37
juriste, 37
jus, 38
justiciable, 109

Kashi Nagari Pracharani Sabha, 72

katib, 17
khari boli, 78
khaṛī bolī, 49
kotwal, 18
kṣatriya, 11
kula, 15

lemme, 189
lois de Huddood, 34
Lois de Hudood, 31
lok sabhā, 25

madrasa, 20
magistrat
magistrate, 25
magistrate court, 25
mahakma-e-adalat, 17
mahakma-e-shariat, 17
maison d'arrêt, 103
mandat de dépôt, 103
manu saṁhita, 12
Ministère public, 110
mirdaha, 17
mis en examen, 102
mise en accusation, 107
Mohajir, 66
mucalkā navīs, 17
muddaʿī, 113
munshi, 22
muqaddam, 17

Nyaya Panchayat Bill, 28
nyāya, 15
nyāya pañcāyat, 25, 27

objets complexes„ 119
Ombudsman, 28
ordonnances, 121

Pan Asian Networking, 174
panchayat, 13, 27
Panchayat Appelate Authority, 28
parivādī, 113
parloirs, 108
parquet, 109
partie civile, 107
paṭel, 17
pañcamaṇḍalī, 15
peśkar, 17
phraséolexique, 234
phraséologie, 231
plaidoirie, 106
plainte, 100
procureur de la République, 110
procédure correctionnelle, 42, 101
procédure criminelle, 101
puga, 15

qaza, 17
Qazi-ul Quzat, 17
qisas, 17, 32
qāzī-e-sarkār, 17
qāzī-e-ṣubah, 17

Raj, 8
relaxe, 106
rule-based, 194
réclusion, 108
réquisition, 112
réquisitoire, 101
réseaux notionnels, 234
rôle, 111
rājya sabhā, 25
rājā, 13
rājādharma, 13
ṛgveda, 11

sabhā, 14
saisine, 101
sarkār, 17
sarwar, 18
segmentation, 187
segments répétés, 208
Sessions Court, 25
sipāhî, 18
siège, 109
smṛti, 12
Suddar Diwani Adawlat , 19
Sudder Nizamat Adawlat, 19
sulah, 32
sāhasa, 14
sākṣī, 113
sūtra, 12

tadbhava, 76
tajwīz s̱ānī, 18
tatsam, 76
taʿzīr, 17
Three Language Formula, 57
thémis, 38
thānā, 18
tokenisation, 187

unité phraséologique, 231
unités phraséologiques, 163
urfī, 18

vaiśya, 11
vakālat nāmah, 18
vakīl-e-sharaʿī, 18
verdict, 108
vidhān sabhā, 25
vyavahāra, 13
véda, 11

Waz-i-Istelaahaat, 79

zamānat, 18, 114
zinā, 35
zones tribales, 36

étiquetage, 187
étiquetage morphosyntaxique, 189

ādeś, 133
ādivāsī, 45
ārop patr, 113

śreni, 15
śruti, 11
śudra, 11

ʿadil, 18
ʿadālat-ul aliah, 18

Table des figures

1 Translittération de la devanagari xi

1.1 Répartition des systèmes juridiques dans le monde 8
1.2 Carte de l'Asie du Sud 9
1.3 L'Asie du Sud avant 1947 23
1.4 Carte des États et territoires indiens 24
1.5 Organisation de la justice pénale en Inde 26
1.6 Carte politique du Pakistan 30
1.7 La justice pénale pakistanaise 33
1.8 Les juridictions pénales en France 42
1.9 Tableau comparatif des juridictions pénales 44
1.10 Répartition des langues en Asie du Sud 50
1.11 Caractère क combiné aux voyelles 52
1.12 Variantes nastaliq et naskh en ourdou 60
1.13 Carte linguistique du Pakistan 64

2.1 Tableau synoptique des grandes lignes de la procédure pénale 118
2.2 Pluriels communs ou soutenus en ourdou 145
2.3 Exemple de FIR 150

3.1 Schéma de la chaîne de traitement du corpus 173
3.2 Variantes pour les glyphes ص et ف 183
3.3 Superposition de glyphes en ourdou 183
3.4 Texte mélangeant les graphies naskh et nastaliq 184
3.5 Exemple de FIR manuscrit en ourdou (Pakistan). 185
3.6 Exemple d'anonymisation d'un texte 204

4.1 Recherche d'un motif dans un corpus comparable aligné 228
4.2 Classement du lexique et des unités phraséologiques à partir des réseaux notionnels 235
4.3 Application *maintenance* pour le dépôt les documents à fouiller et les lexiques 246
4.4 Fonctionnement de l'*Occurrencier* pour la recherche de la traduction du motif *murder*. 246

Liste des tableaux

1 Translittération des caractères arabo-persans x

1.1 Termes juridiques utilisés en sanskrit entre la période antique et le XII[e] s. 14
1.2 Ligatures en caractères devanagari 52
1.3 Équivalents des préfixes en sanskrit et en grec / latin 73
1.4 Dérivation à partir de *vidhi / law* 73

2.1 Typologie des types de textes juridiques 120
2.2 Variations entre les usages et la terminologie officielle dans les textes juridictionnels en hindi 142
2.3 Emprunts anglais et terminologie ourdou (Pakistan) 146
2.4 Équivalences lexicales langue spécialisée / langue commune 151

3.1 Nombre de tokens des corpus par langue 161
3.2 Composition du corpus total 168
3.3 Jeu personalisé d'étiquettes morphosyntaxiques communes au hindi et à l'ourdou 193
3.4 Résultat de l'entraînement des étiqueteurs morphosyntaxiques (hindi) 199
3.5 Résultat de l'entraînement des étiqueteurs morphosyntaxiques (ourdou) . . 200

4.1 Termes exclusifs à la partition *Forums* 211
4.2 Répartition des termes exclusifs aux partitions *Presse* ou *Lois* 214
4.3 Répartition des termes spécifiques à chaque partition (*Presse* ou *Lois*) . . . 215
4.4 Comparaison thématique des codes français et indien 225
4.5 Alignement des version françaises et hindi des codes pénaux 227
4.6 Lexèmes actualisant la notion « préjudice » 236
4.8 Réseau notionnel *lieux de justice* 240
4.7 Réseau notionnel *infractions* 240
4.9 Réseau notionnel *instruction* 241
4.11 Réseau notionnel *mesures et sanctions* 242
4.10 Réseau notionnel *mécanique du procès* 243
4.12 Réseau notionnel *protagonistes* 244

Bibliographie

ISO 704 : *Principes et méthodes de la terminologie*. Organisation internationale de normalisation (ISO/TC 37), Genève, 1987.

Jean-Michel Adam : Genres, textes, discours : pour une reconception linguistique du concept de genre. *Revue belge de Philologie et d'Histoire*, 75(3) :665–681, 1997.

Ludwig W. Adamec : *Historical dictionary of Islam*. Scarecrow Press, Lanham, 2009.

Muhammad Basheer Ahmad : *The administration of justice in medieval India : a study in outline of the judicial system under the sultans and the badshahs of Delhi based mainly upon cases decided by medieval courts in India between 1206-1750 A.D.* The Aligarh Historical Research Institute, 1941. URL `https://archive.org/details/administrationof029109mbp`.

T. Ahmad, S. Urooj, S. Hussain et A. & al. Mustafa : The CLE Urdu POS Tagset, 2013.

Q. U. A. Akram et S. Hussain : Improving Urdu recognition using character-based artistic features of nastalique calligraphy. *IEEE Access*, 7:8495–8507, 2019. ISSN 2169-3536.

Daud Ali, Khan Wahab et Che Dunren : Urdu language processing, ulp. *Artificial Intelligence Review*, 47(279), 2017. ISSN 0269-2821.

A.S. Altekar : *State and Government in Ancient India*. Motilal Banarsidass, 2009.

T Anderson, A Avizienis, W Carter, A Costes, F Christian, Y Koga, H Kopetz, J Lala, J Laprie, J Meyer *et al.* : Dependability : Basic concepts and terminology. *Series : Dependable Computing and Fault-Tolerant Systems*, 5, 1992.

David Annoussamy : *Le droit indien en marche*. Société de législation comparée, Paris, 2001.

Anonyme : The Constitution of the Islamic Republic of Pakistan, 1973. URL `http://www.na.gov.pk/uploads/documents/1333523681_951.pdf`. téléchargé sur le site de l'Assemblée nationale pakistanaise, consulté le 26-06-2019.

Anonyme : *hindī sahāyikā*. Bharatiya Heavy Electronics Limited, 2012. URL `https://web.bhelhyd.co.in/files/misc/hindi.pdf`.

Kunchukuttan Anoop et Damani Om P. : A System for Compound Noun Multiword Expression Extraction for Hindi. Mumbai, India.

P.J. Antony et K.P. Soman : Part of Speech Tagging for Indian Languages : A Litterature Survey. *Internation Journal of Computer Applications*, 34(8), novembre 2011.

Anukriti.net : Anukriti.net - Post Graduate Diploma in Translation Studies, 2016. URL `http://www.anukriti.net/pgdts/course413/index.html`. Consulté le 30 novembre 2016.

Alyssa Ayres : *Speaking like a state : language and nationalism in Pakistan*. Cambridge University Press, Cambridge New York Melbourne, 2009. ISBN 978-0-521-51931-1.

Paul Baker, Andrew Hardie, Tony McEnery et BD Jayaram : Corpus data for South Asian language processing. *In Proceedings of the 10th Annual Workshop for South Asian Language Processing, EACL*, 2003.

Paul Baker, Andrew Hardie, Tony McEnery, Richard Z. Xiao, Kalina Bontcheva, Hamish Cunningham, Robert Gaizauskas, Oana Hamza, Diana Maynard, Valentin Tablan, Cristian Ursu, B. D. Jayaram et Mark Leisher : Corpus linguistics and South Asian languages : corpus creation and tool development. *Literary and Linguistic Computing*, 19(4) :509–524, 11 2004. ISSN 0268-1145.

Mikhail Bakhtine : *Esthétique de la création verbale*. Gallimard, Paris, 1984.

Christian Balliu : Cognition et déverbalisation. *Meta*, 52(1) :3–12, 2007.

Muhammad Abdul Basit : *Rafiq's Law Dictionary with pronunciation*. Federal Law House, Lahore - Rawalpindi, 2015.

Dara Becker et Kashif Riaz : A study in Urdu corpus construction. *In Proceedings of the 3rd workshop on Asian language resources and international standardization-Volume 12*, pages 1–5. Association for Computational Linguistics, 2002.

Normand Belair : Législation bilingue — Production de la seconde version. *L'Actualité terminologique / Terminology Update*, 20(6) :6–8, 1987.

Emile Benveniste : *Vocabulaire des institutions indo-européennes*, volume 2, « Pouvoir, droit, religion ». les Editions de Minuit, collection Sens Commun, Paris, 1969.

Jean-Paul Benzécri et al. : *L'analyse des correspondances*, volume II. Dunod, Paris, 1973.

Gilles Bergeron : Traduction et terminologie juridique : l'interprétation en milieu judiciaire. *Meta*, 47(2), juin 2002.

Vincent Berment : *Méthodes pour informatiser des langues et des groupes de langues "peu dotées"*. Thèse de doctorat, 2004. URL `http://www.theses.fr/2004GRE10046`. Thèse de doctorat dirigée par Boitet, Christian Informatique Grenoble 1 2004.

Daniela Berti : Pouvoirs locaux et contraintes judiciaires dans une affaire de viol en Inde. *Diogène*, (3) :139–165, 2012. URL `http://www.cairn.info/revue-diogene-2012-3-page-139.htm`.

Daniela Berti : *Regimes of Legality. Ethnography of Criminal Cases in South Asia*. OUP, 2015.

Daniela Berti et Gilles Tarabout : Religion et environnement dans les procédures judiciaires en Inde. *Les Cahiers de la justice*, pages 409–420, 2015.

Jean-Marie Bertrand : À propos de la Rhétorique d'Aristote (i, 1373b 1-1374b 23), analyse du processus judiciaire. iv. Épieikeia. *Cahiers du Centre Gustave Glotz*, (20) :7–27, 2009.

Ilhem Bezzaoucha et Khadidja Merackchi : Corpus-Based Approaches to Specialized translation : a new framework ? *Cahiers de Traduction*, 4(8) :15–23, 2017. ISSN 1111-4606. URL `https://www.asjp.cerist.dz/en/article/27906`.

Akshar Bharati et al. : Anncorra : Annotating corpora guidelines for POS and chunk annotation for Indian languages. dec 2006.

Esther Blais : Le phraséologisme. une hypothèse de travail. *Terminologies nouvelles*, (10), décembre 1993.

Ondřej Bojar, Vojtěch Diatka, Pavel Rychlý, Pavel Straňák, Vít Suchomel, Aleš Tamchyna et Daniel Zeman : HindMonoCorp 0.5, 2014. URL http://hdl.handle.net/11858/00-097C-0000-0023-6260-A. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

E Tognini Bonelli : Functionally complete units of meaning across English and Italian. *Lexis in Contrast*, pages 73–95, 2002.

Pierre Bourdieu : L'économie des échanges linguistiques. 1977. ISSN 0023-8368. URL https://www.persee.fr/doc/lfr_0023-8368_1977_num_34_1_4815.

Didier Bourigault et Monique Slodzian : Pour une terminologie textuelle. *Terminologies nouvelles*, 19(1999) :29–32, 1999.

Thorsten Brants : TnT: A Statistical Part-of-speech Tagger. *In Proceedings of the Sixth Conference on Applied Natural Language Processing*, ANLC '00, pages 224–231, Stroudsburg, PA, USA, 2000. Association for Computational Linguistics. URL https://doi.org/10.3115/974147.974178.

Jean-Paul Bronckart : Les genres de textes et leur contribution au développement psychologique. *Langages*, (153) :98–108, 2004.

Jawaid Bushra, Kamran Amir et Bojar Ondrej : A Tagged Corpus and a Tagger for Urdu. *In* Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk et Stelios Piperidis, éditeurs : *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 26–31, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.

Miriam Butt et Tracy Holloway King : Urdu and the Parallel Grammar project. *Association for Computational Linguistics*, pages 1–3, Aug 2002.

Georg Bühler : *On the origin of the Indian Brāhma alphabet*. Trübner, Karl J., 1898.

Statistics Canada : La communauté sud-asiatique au Canada. URL http://www.statcan.gc.ca/pub/89-621-x/89-621-x2007006-fra.htm. Consulté le 01 décembre 2016.

C. Chakravarti : The constitution and language politics of India. URL http://trove.nla.gov.au/version/34728402. Consulté le 01 décembre 2016.

C Chandramouli : Census of India 2011. *Provisional Population Totals. New Delhi : Government of India*, 2011.

Ross Charnock : Les langues de spécialité et le langage technique : considérations didactiques. *revue Asp*, (23-26) :281–304, 1999.

Matthieu Chemin : The impact of the judiciary on entrepreneurship : Evaluation of pakistan's « Access to Justice Programme ». *Journal of Public Economics*, (93) :114–125, 2009.

Dansingh Suganchand Choudhary et Praveen Kumar Dansingh Choudhary : *Hindi English Legal Dictionary - vidhi śabdāvalī*. Universal Law Publishing, New Delhi, 2010.

Anne Condamines et Aurélie Picton : Étude du fonctionnement des nominalisations déverbales dans un contexte de déspécialisation. *In 4eCongrès mondial de Linguistique Française*. Congrès mondial de Linguistique Française, SHS Web of Conference, EDP Sciences, 2014.

Gérard Cornu : *Linguistique juridique*. Montchrestien, Paris, 1990.

Gérard Cornu : *Vocabulaire juridique*. PUF, Paris, 2010.

Denis Creissels : *Unités et catégories grammaticales : réflexions sur les fondements d'une théorie générale des descriptions*. Publications de l'Université des langues et lettres, Grenoble, 1979.

Maria Da Graça Krieger : Terminographie juridique et spécificités textuelles. *Meta : journal des traducteurs*, 47(2) :233–243, 2002.

Jean Dabin : *Encyclopædia Universalis*, chapitre DROIT - Théorie et philosophie.

Jean-Philippe Dalbera : Le corpus entre données, analyse et théorie. *Corpus*, (1), Nov 2002. ISSN 1638-9808. URL https://journals.openedition.org/corpus/10.

Éliane Damette : Enseigner la traduction juridique : L'apport du français juridique, discipline passerelle entre droit, méthodologie juridique et linguistique. *In La traduction juridique : Points de vue didactiques et linguistiques*. CEL, CEL, 2013. URL https://hal-univ-lyon3.archives-ouvertes.fr/hal-00980068/document.

Sukla Das : *Crime and Punishment in Ancient India*. Abhinav Publications, 1990.

Commission de terminologie et de néologie en matière juridique : Rapport quadriennal 2003-2007, 2007. URL http://www.justice.gouv.fr/art_pix/1_1_rapp_terminologie.pdf.

Thierry Debard et Serge (dir.) Guinchard : *Lexique des termes juridiques 2014-2015*. Dalloz, Paris, 2014.

Emmanuel Didier : *Langues et langages du droit. Étude comparative des modes d'expression de la Common Law et du Droit Civil, en français et en anglais*. Wilson and Lafleur, Montréal, 1990.

John Duncan et Martin Derett : *Religion, Law and the State in India*. OUP, 1999.

Baudouin Dupret : *La charia aujourd'hui. Usages de la référence au droit islamique*. (dir.) La Découverte, Paris, 2012.

Muriel Fabre-Magnan : *Introduction au droit*. PUF, 2010.

H. Felber : *Manuel de terminologie*. UNESCO, Paris, 1987.

Pierre Fiala : Pour une approche discursive de la phraséologie. *Langage et société*, (42) :27–44, décembre 1987.

Michel Foucault : *L'archéologie du savoir*. Gallimard, 1969.

Béatrice Fracchiolla : Méthodologie pour l'analyse comparée avec Lexico3 et résultats : L'étude des pronoms chez les Verts français et italiens, 2006. URL https://www.academia.edu/340469/M%C3%A9thodologie_Pour_LAnalyse_Compar%C3%A9e_Avec_Lexico3_Et_R%C3%A9sultats_L%C3%A9tude_Des_Pronoms_Chez_Les_Verts_Fran%C3%A7ais_Et_Italiens. [Online ; accessed 29. Apr. 2019].

Marie-Anne Frison-Roche : La rhétorique juridique. *Hermès, La Revue*, 2(16) :73–83, 1995. URL https://www.cairn.info/revue-hermes-la-revue-1995-2-page-73.htmr.

Frost et Sullivan : *Local Language Information Technology Market in India*. sous l'égide du Ministry of Communications and Information Technology, Delhi, 2003.

Pascale Fung : Compiling bilingual lexicon entries from a non-parallel English-Chinese corpus. *In Third Workshop on Very Large Corpora*, 1995.

Pascale Fung : A statistical view on bilingual lexicon extraction : from parallel corpora to non-parallel corpora. *In Conference of the Association for Machine Translation in the Americas*, pages 1–17. Springer, 1998.

Félix Gaffiot : *Dictionnaire latin-français*. Hachette, Paris, 1934.

Shumit Ganguly : *the Crisis in Kashmir, Potents of War, Hopes of Peace*. Woodrow Wilson Center, Cambridge University Press, 1997.

Roger Garside : The robust tagging of unrestricted text : the BNC experience. 1996.

Partha Sarathy Ghosh : *The politics of personal law in South Asia*. Routeledge, New Delhi, 2007.

Daniel Gile : *Basic Concepts and Models for Interpreter and Translator Training*. John Benjamins, Amsterdam / Philadelphia, 1995.

M. Gopal et G.N. Jha : Tagging sanskrit corpus using bis pos tagset. *Information Systems for Indian Languages. ICISIL 2011. Communications in Computer and Information Science*, 139, 2011.

Ministry of Human Resource Development Government of India : Language Education. URL `http://mhrd.gov.in/language-education-9`. Consulté le 29 novembre 2016.

George Abraham Grierson : *Linguistic Survey of India*, volume VI et IX. Office of the superintendent of Government printing, Calcutta, 1904.

Mathieu Guidère : *Introduction à la traductologie*. De Boeck, collection Traducto, Bruxelles, 2008.

Avdhesh Mohan Gupta : *rābhaṣā sahāyikā*. Prabhat Prakashan, Delhi, 2008. Le chapitre 6 est consacré aux principes de création terminologique.

Kanika Gupta, Monojit Choudhury et Kalika Bali : *Mining Hindi-English Transliteration Pairs from Online Hindi Lyrics*. LREC Proceedings, Istanbul, 2012.

Vaishali Gupta, Nisheeth Joshi et Iti Mathur : Rule Based Stemmer in Urdu. *CoRR*, abs/1310.0581, 2013. URL `http://arxiv.org/abs/1310.0581`. Consulté le 07 Jun 2017.

Vishal Gupta : Hybrid Multilingual Key Terms Extraction System for Hindi and Punjabi Text. *Progress in Systems Engineering. Advances in Intelligent Systems and Computing*, 366, 2015.

Vishal Gupta et Gurpreet Singh Lehal : Automatic Keyword Extraction for Punjabi Language. *IJCSI International Journal of Computer Science Issues*, 8(3) :327–331, septembre 2011. ISSN (en ligne) : 1694-0814. URL `http://www.learnpunjabi.org/pdf/IJCSI-8-5-3-327-331%20(1).pdf`.

Jean-Claude Gémar : Les enjeux de la traduction juridique. Principes et nuances. 1998. URL `www.tradulex.com/Bern1998/Gemar.pdf`.

Jean-Claude Gémar : Le plus et le moins-disant culturel du texte juridique. langue, culture et équivalence. *Meta : journal des traducteurs*, 47(2) :163–176, 2002.

Jean-Claude Gémar : Aux sources de la « jurilinguistique » : texte juridique, langues et cultures. *Revue française de linguistique appliquée*, XVI(1) :9–16, 2011. URL https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2011-1-page-9.htm.

Secrétariat général du Gouvernement : *Guide de légistique, 3e édition*. la Documentation française, 2017. URL https://www.legifrance.gouv.fr/Droit-francais/Guide-de-legistique2.

Benoît Habert, Adeline Nazarenko et André Salem : *Les linguistiques de corpus*. Colin, 1997.

Malcolm Harvey : Traduire l'intraduisible. stratégies d'équivalence dans la traduction juridique. *Les Cahiers de L'ILCE*, 3:30–49, 2001.

Malcolm Harvey : What's so Special about Legal Translation ? *Meta : journal des traducteurs*, 47(2), 2002.

Fahim Muhammad Hasan, Naushad UzZaman et Mumit Khan : Comparison of different POS Tagging Techniques (n-gram, HMM and Brill's tagger) for Bangla. *SpringerLink*, pages 121–126, 2007.

Serge Heiden, Jean-Philippe Magué et Bénédicte Pincemin : TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement. *In* Sergio Bolasco, Isabella Chiari et Luca Giuliano, éditeurs : *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, volume 2, pages 1021–1032, Rome, Italy, juin 2010. Edizioni Universitarie di Lettere Economia Diritto. URL https://halshs.archives-ouvertes.fr/halshs-00549779.

Florence Herbulot : La Théorie interprétative ou Théorie du sens : point de vue d'une praticienne. *Meta*, 49(2) :307–315, 2004.

Lucas Hilgert, Lucelene Lopes, Artur Freitas, Renata Vieira, Denise Hogetop et Aline Vanin : Building Domain Specific Bilingual Dictionaries. *In* Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk et Stelios Piperidis, éditeurs : *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.

Hans Henrich Hock et Elena Bashir : *The Languages and Linguistics of South Asia : A Comprehensive Guide*. The world of linguistics 7. de Gruyter Mouton, 2016. ISBN 9783110423303,3110423308,9783110423389,3110423383,9783110427158,311042715X.

Faqir Hussain : The Judicial System of Pakistan. 2015. URL http://supremecourt.gov.pk/web/user_files/File/thejudicialsystemofPakistan.pdf. site consulté le 09-07-2019.

Fidelia Ibekwe-SanJuan : Extraction terminologique avec INTEX. 2001. URL http://www.univ-montp3.fr/~amela/TAL/Intex/Intex&Terminologie.pdf.

IIIT : IIIT Tagset guidelines.

Madiha Ijaz et Sarmad Hussain : Corpus based Urdu lexicon development. *In the Proceedings of Conference on Language Technology (CLT07), University of Peshawar, Pakistan*, volume 73, 2007.

Indian Scientific Translators Association : Journal of the Indian Scientific Translators Association : JISTA. *Journal of the Indian Scientific Translators Association : JISTA.*, 1972. ISSN 0253-8776. OCLC: 647964829.

InMyLanguage : InMyLanguage - اونٹاریو میں عدالتی نظام (court system) کیا ہے؟. URL http://english.inmylanguage.org/article.aspx?cat=ONCOURTS&docid=2272181&doclang=UR. Consulté le 01 décembre 2016.

A. Jabbar, S. Iqbal, M.U.G. Khan et al. : A Survey on Urdu and Urdu like langage stemmers and stemming techniques. *Artificial Intelligence Review*, 49(3) :339–373., mars 2018.

Christophe (dir.) Jaffrelot : *L'Inde contemporaine de 1950 à nos jours*. Fayard, Paris, 1997.

Mahabir Prashad Jain : *Outlines of Indian Legal History*. Tripathy, Mumbai, 1972.

Sobia Tariq Javed et Sarmad Hussain : Segmentation based Urdu nastalique ocr. *In Iberoamerican Congress on Pattern Recognition*, pages 41–49. Springer, 2013.

Bushra Jawaid, Amir Kamran et Ondřej Bojar : Urdu Monolingual Corpus, 2014. URL http://hdl.handle.net/11858/00-097C-0000-0023-65A9-5. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Julius Jolly : *Hindu Law and Customs [trad. par Ghosh Batakrishna et pub. 1975]*. Bharatiya Pub. House, Varanasi, 1927.

Nisheeth Joshi, Hemant Darbari et Ity Mathur : HMM Based POS Tagger for Hindi. *Computer Science and Information Technology*, pages 341–349, 2013.

Kalele et alii. : Enabling linguistic Idiosynchrasy in ANUVADAKSH. *Wishwabharata*, n° 41, juil-déc 2013, 2013.

Rohit Kansal, Vishal Goyal et Gurpreet Singh Lehal : Rule Based Urdu Stemmer. *In Proceedings of COLING 2012: Demonstration Papers*, pages 267–276, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL http://www.aclweb.org/anthology/C12-3034.

Raj Kapil : Régler les différends, gérer les différences : dynamiques urbaines et savantes à Calcutta au xviiie siècle. *Revue d'histoire moderne & contemporaine*, 55-2(2) :70–100, 2008. URL https://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2008-2-page-70.htm.

Romain Karpinski et Abdel Belaid : Rapport Evaluation des OCR, août 2016. URL https://hal.inria.fr/hal-01356824.

Stephen Kelly et David Johnston : *Betwixt and Between : Place and Cultural Translation*. Cambridge Scholars Publishing, 2009. ISBN 978-1-4438-1593-2.

Rania A. Khalifa : La phraséologie du discours juridique français. Étude de cas : les arrêts de la Cour de cassation. *Synergies Espagne*, (8) :49–64, 2015. URL https://gerflint.fr/Base/Espagne8/khalifa.pdf.

Lakshman M. Khubchandani : *Language, Culture and Nation Building : Challenges of Modernization*. Shimla Institute of Advanced Studies, Shimla, 1991.

Charles M. Kieffer : *L'avancée du persan vers l'est : le cas des isolats persanophones les plus orientaux en Afghanistan*.

Robert D. King : *Nehru and the language politics of India*. Oxford University Press, Delhi New York, 1997. ISBN 978-0-19-563989-6.

Olivier Kraif : Qu'attendre de l'alignement de corpus multilingues ? *Revue Traduire, 4e Journée de la traduction professionnelle*, (210) :17–37, 2006. URL https://hal.archives-ouvertes.fr/hal-01073711.

Olivier Kraif : *Corpus parallèles, corpus comparables : quels contrastes ?* Habilitation à diriger des recherches, Université de Poitiers, décembre 2014. URL https://hal.archives-ouvertes.fr/tel-01184585.

Ramesh Krishnamurthy : Corpus-driven Lexicography. *International Journal of Lexicography*, 21(3) :231–242, 2008. URL http://dx.doi.org/10.1093/ijl/ecn028.

Natalie Kübler : Mettre en œuvre la linguistique de corpus à l'université. *Recherches en Didactique des Langues et des Cultures*, 11(1) :1–35, 2014.

Pierre Lafon : Sur la variabilité de la fréquence des formes dans un corpus. 1980. ISSN 0243-6450. URL https://www.persee.fr/doc/mots_0243-6450_1980_num_1_1_1008.

Pierre Lafon et André Salem : L'inventaire des segments répétés d'un texte. 1983. ISSN 0243-6450. URL https://www.persee.fr/doc/mots_0243-6450_1983_num_6_1_1101.

Judith Lavoie : *Le discours sur la traduction juridique au Canada*, volume 47. les Presses de l'Université, Montréal, 2002.

Jacques Leclerc : L'aménagement linguistique dans le monde, 2019. URL http://www.axl.cefan.ulaval.ca/asie/pakistan-const.htm. Consulté le 26-06-2019.

Georges Legault : Fonctions et structure du langage juridique. *Meta*, 24(1) :18–25, 1979.

Pierre Lerat : Vocabulaire juridique et schémas d'arguments juridiques. *Meta : journal des traducteurs*, 47(2) :155–162, 2002.

Lionel A. Levert : La cohabitation du bilinguisme et du bijuridisme dans la législation fédérale canadienne : mythe ou réalité ? *Revue de la common law en français*, (n° 3) :127–136, 2000.

D. Lindemann, I. Manterola, R. Nazar et X. San Vicente, I.and Saralegi : Bilingual Dictionary Drafting : The Example of German-Basque, a Medium-Density Language Pair. *In Proceedings of the XVI EURALEX Conference*, pages 563–576, Bolzano, 2014. URL https://www.academia.edu/7657526/Bilingual_Dictionary_Drafting_The_example_of_German-Basque_a_Medium-Density_Language_Pair.

David Lindemann : Bilingual Lexicography and Corpus Methods. The Example of German-Basque as Language Pair. *In Proceedings of CILC 2013*, pages 249–257. Procedia – Social and Behavioral Sciences, Elsevier, octobre 2013. URL https://www.sciencedirect.com/science/article/pii/S1877042813041657.

Robert Lingat : Les sources du droit dans le système traditionnel de l'Inde. *Revue internationale de droit comparé*, 20(3) :590–592, 1968. ISSN 0035-3337. URL https://www.persee.fr/doc/ridc_0035-3337_1968_num_20_3_17211.

Maria Teresa Lino et Christina Deschamps : Langue juridique et créativité terminologique : une perspective français-portugais. 2016. URL http://hdl.handle.net/10362/34347. Consulté le 15 juillet 2019.

Rodrick A. MacDonald : Legal Bilingualism. *Revue de droit de McGill*, (42) :119–168, 1997.

A. M. Maciel : *Para o reconhecimento da especificidade do termo jurídico*. Thèse de doctorat, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001.

Dominique Maingueneau : *Discours littéraire. Paratopie et scène d'énonciation*, chapitre Typologie des genres de discours. Armand Colin, Paris, 2004. URL http://dominique.maingueneau.pagesperso-orange.fr/pdf/Typologie-des-genres-de-discours.pdf.

Charles Malamoud : *Cuire le monde : rite et pensée dans l'Inde ancienne*. La Découverte, 2016.

Arvind-pal Singh Mandair : *Religion and the specter of the West : Sikhism, India, postcoloniality, and the politics of translation*. Insurrections. Columbia University Press, New York, 2009. ISBN 978-0-231-14724-8.

Sabiha Mansoor, Shaheen Meraj et Aliya Tahir : *Language policy, planning, and practice : a South Asian perspective*. Aga Khan University Oxford University Press, Karachi, Oxford, 2004. ISBN 978-0-19-579965-1.

Sahiba Mansoor : *A Sociolinguistic Study*. Vanguard Books. URL http://www.vanguardbooks.com/browsetitle.php?isbn=9694022401&subject=Book%20Bargains. Consulté le 01 décembre 2016.

Manu : *the institutes of Manu : code of hindu laws*, volume 62, 63, chapitre Manav Dharma Shastra. Cosmo Publication, collection Rediscovering India, Indian Philosophy Library, New Delhi, 1990.

Claude Markovits : *Histoire de l'Inde moderne, 1480-1950 (dir.)*. Fayard, Paris, 1994.

William Martinez : Mise en évidence de rapports synonymiques par la méthode des cooccurrences. *Actes de JADT 2000 (5es Journées internationales d'Analyse statistique des Données Textuelles)*, pages 78–84, 2000.

Sylvie Mellet : Corpus et recherches linguistiques. introduction. *Corpus*, (1), Nov 2002. ISSN 1638-9808. URL https://journals.openedition.org/corpus/7.

Igor Mel'cuk, André Clas et Alain Polguère : *Introduction à la Lexicologie Explicative et Combinatoire, coll. Universités Francophones*. Duculot, Louvain-la-Neuve, 1995.

Mariette Meunier, Marion Charret-Del Bove et Eliane Damette : La traduction juridique : Points de vue didactiques et linguistiques. 2013. URL https://hal-univ-lyon3.archives-ouvertes.fr/hal-00980068/document.

Minsitry of Law, Justice and Company Affairs : *Legal Glossary / vidhi śabdāvalī*. Government of India (éd. et pub.), 2001.

Monier-Williams : *A Sanskrit-English Dictionary, etymologically and philologically arranged*. OUP (1e éd. 1899), Oxford, 1960.

Sylvie Monjean-Decaudin : Approche juridique de la traduction du droit, 2010.

Sylvie Monjean-Decaudin : *La traduction du droit dans la procédure judiciaire*. Dalloz, 2012. ISBN ISBN : 978-2-247-11911-0.

Annie Montaut : Le hindi en 1947 : la question de la langue nationale, ses origines et ses conséquences. *Cahiers du Sahib*, (5), 1997a.

Annie Montaut : Problématique d'ensemble : l'asie du sud multilingue. *Faits de langues*, (10), septembre 1997b. URL https://www.persee.fr/doc/flang_1244-5460_1997_num_5_10_1165.

Annie Montaut : *La langue libérée. Etudes de socio-lexicologie*, chapitre Comment une langue « libérée » et libératrice en vient à agir comme langue d'oppression : le cas du hindi, pages 137–150. Peter Lang, Berne, laurian, anne-marie édition, 2003.

Annie Montaut : *Le hindi*. Peeter, Louvain, 2012.

Annie Montaut : *Le Tour du monde des concepts*, chapitre le hindi, pages 229–286. Fayard, Paris, 2014.

Emmanuel Morin et Béatrice Daille : Comparabilité de corpus et fouille terminologique multilingue. *Traitement Automatique des Langues*, 47(1) :113–136, 2006. URL https://hal.archives-ouvertes.fr/hal-00474316.

Emmanuel Morin et Emmanuel Prochasson : Bilingual Lexicon Extraction from Comparable Corpora Enhanced with Parallel Corpora. *In* Association for Computational Linguistics, éditeur : *4th Workshop on Building and Using Comparable Corpora : Comparable Corpora and the Web*, pages 27–34, Portland, United States, juin 2011. URL https://hal.archives-ouvertes.fr/hal-00608475.

Marie-Françoise Mortureux : Les vocabulaires scientifiques et techniques. *Les enjeux des discours spécialisés, Carnets du Cediscor*, (3) :13–26, 1995.

Leena Moudgil : *Fair and speedy trial need of the time in criminal justice system*. Thèse de doctorat, Kurukshetra University, The address of the publisher, 2015. URL http://hdl.handle.net/10603/94159. An optional note.

Ahmed Muaz, Aasim Ali et Sarmad Hussain : Analysis and development of Urdu POS tagged corpus. *In Proceedings of the 7th workshop on Asian language resources*, pages 24–29. Association for Computational Linguistics, 2009.

Bharata Muni : *The Natyashastra*. Bibliotheca Indica. Asiatic Society of Bengal, 1951. ISBN 9783110423303,3110423308,9783110423389,3110423383,9783110427158,311042715X. [traduit en anglais par Manmohan Ghosh].

Lluìs Màrquez et Horacio Rodrìguez : Part-of-speech tagging using decision trees. *SpringerLink*, pages 25–36, Apr 1998.

A.D. Needham et Rajan R. Sunder : *the Crisis of Secularism in India*. Duke University Press, New Delhi, 2007.

Émilie Née : *Méthodes et outils informatiques pour l'analyse des discours*. Presses universitaires de Rennes, 2017.

Ministry of Panchayati Raj : The Nyaya Panchayats Bill, 2009. URL https://www.prsindia.org/uploads/media/draft/Nyaya%20Panchayats%20Bill,%202009.pdf.

Sébastien Paumier : Unitex-manuel d'utilisation, 2011.

Sylvia Pavel et Diane Nolet : *Handbook of terminology*. Canadian Government Publishing, Ottawa, 2001.

Mojca Pecman : L'enjeu de la classification en phraséologie. *Actes du congrès EUROPHRAS 2004*, pages 127–146, 2004.

Mojca Pecman : Les apports possibles de la phraséologie à la didactique des langues étrangères. *Alsic [En ligne]*, 8(2), 2005. URL http://journals.openedition.org/alsic/334.

J. Pienkos : La jurilinguistique et la traduction - Quelques réflexions d'un linguiste et d'un juriste. *In* Paul Nekeman, éditeur : *Translation, our future / La traduction, notre avenir, Proceedings of the XIth World Congress of FIT*,, Maastricht, 1990. Euroterm.

Céline Poudat et Frédéric Landragin : *Explorer un corpus textuel : Méthodes-pratiques-outils*. De Boeck Superieur, 2017.

Emmanuel Prochasson : *Multilingual alignment from specialised comparable corpora*. Theses, Université de Nantes, décembre 2009. URL https://tel.archives-ouvertes.fr/tel-00462248.

Roger T. Pédauque : *Le Document à la lumière du numérique : forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité*. C & F éditions, 2006.

Jean Quirion : La localisation, palimpseste de l'aménagement terminologique ? Stratégies d'implantation terminologique et marketing. *Meta : journal des traducteurs*, 51(4), 2006.

Tariq Rahman : *Language and politics in Pakistan*. Oxford University Press, Karachi/ New York, 1996. ISBN 978-0-19-577692-8.

Tariq Rahman : *Language, education, and culture*. Sustainable Development Policy Institute Oxford Uinversity Press, Islamabad Karachi, 1999. ISBN 978-0-19-579146-4.

Tariq Rahman : Language policy and localization in pakistan : Proposal for a paradigmatic shift. 2004. URL http://apnaorg.com/research-papers-pdf/rahman-1.pdf.

Tariq Rahman : Language, Religion and Politics. urdu in Pakistan and North India. (124) :93–112, 2008. ISSN 0997-1327. URL http://remmm.revues.org/6019.

Tariq Rahman : *Language policy, identity, and religion : aspects of the civilization of the Muslims of Pakistan and North India*. Chair on Quaid-i-Azam and Freedom Movement, National Institute of Pakistan Studies, Quaid-i-Azam University, Islamabad, 2010. ISBN 978-969-8329-13-6.

Vaidehi Ramanathan : *The English-vernacular divide : postcolonial language politics and practice*. Numéro 49 *in* Bilingual education and bilingualism. Multilingual Matters, Clevedon, Buffalo, 2005. ISBN 978-1-85359-770-1 978-1-85359-769-5.

Reinhard Rapp : Identifying word translations in non-parallel texts. *arXiv preprint cmp-lg/9505037*, 1995.

François Rastier : *Sens et textualité*. Hachette, Paris, 1989.

François Rastier : Linguistique interprétative et fondements sémiotiques de la traduction. XV(4 (2010)), 2010. URL http://www.revue-texto.net/index.php?id=2718. [Consulté le 4 avril 2019].

Ratanlal et Dhirajlal : *daṇḍ prakriyā saṁhitā, The Code of Criminal Procedure, 20ème édition.* Lexis Nexis Butterworths, Wadhwa Nagpur, 2011.

Jayanta Ray : A review of terminological work being done in Indian Languages. *In Term banks for tomorrow's world : Translating and the Computer 4.* Aslib, 1983. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.511.3534&rep=rep1&type=pdf. Consulté le 30 juin 2019.

Alain Rey : *La Terminologie, noms et notions.* Numéro 1780. PUF, Paris, 1992.

Ludo Rocher : Lawyers in classical hindu law. *Law and Society Review*, 3(2/3) :383–402, nov.1968-fév.1969.

Jérôme Rocheteau et Béatrice Daille : TTC TermSuite : A UIMA Application for Multilingual Terminology Extraction from comparable corpora. *5th International Joint Conference on Natural Language Processing (IJCNLP)*, pages 9–12, nov 2011. URL https://hal.archives-ouvertes.fr/hal-00819025/document.

Norbert Rouland : *Introduction historique au droit.* puf, 1988.

Muneer A. S., Hazaa, Nazlia, Omar, Fadl Mutaher, Ba-Alwi et Mohammed Albared : Automatic Extraction of Malay Compound Nouns Using a Hybrid of Statistical and Machine Learning Methods. *International Journal of Electrical and Computer Engineering (IJECE)*, 6(3) :925–935, juin 2016. ISSN 2088-8708.

Muhammad Saleem : Bilingual lexicography : Some issues with modern English Urdu lexicography – a user's perspective. *Linguistik Online*, 31(2), Sep. 2013. URL https://bop.unibe.ch/linguistik-online/article/view/545.

Beatrice Santorini : Part-of-Speech Tagging Guidelines for the Penn Treebank Project, 1991.

Asha Sarangi : Language and politics in India. 2009.

Susan Sarcevic : *New Approach to Legal Translation.* Kluwer Law International, La Haye /Londres /Boston, 1997.

Helmut Schmid : Probabilistic part-of-speech tagging using decision trees, 1994. URL https://www.semanticscholar.org/paper/Probabilistic-part-of-speech-tagging-using-decision-Schmid/bd0bab6fc8cd43c0ce170ad2f4cb34181b31277d. [Consulté le 29 avril 2019].

Violeta Seretan, Luka Nerima et Eric Wehrli : Extraction of multi-word collocations using syntactic bigram composition. 2003.

M. C. Setalvad : *The role of English law in India, Lionel Cohen Lectures, 12th series.* Hebrew University, Jérusalem, 1966.

Hibah Shabkhez : L'anglais, la décolonisation, et la question de l'identité linguistique pakistanaise. URL http://www.academia.edu/14196952/L_anglais_la_d%C3%A9colonisation_et_la_question_de_l_identit%C3%A9_linguistique_pakistanaise. Consulté le 01 décembre 2016.

Brajkishore Sharma : *vidhi kī śabdāvalī aur vidhi kā anuvād*, chapitre vidhi śabdāvalī kā nirmāṇ, pages 79–120. PHI Learning, New Delhi, 2009.

P. Gopal Sharma : *Hindi and the Composite Culture of India*, volume 11, Language and civilization change in South-Asia de *contribution to Asian Studies*. E. J. Brill, Leiden, clarence maloney édition, 1978. An optional note.

Shivaji Simha : *Evolution of the Smṛti law : a study in the factors leading to the origin and development of ancient Indian legal ideas*. Bhāratīya Vidyā Prakāśana, Varanasi, 1972.

John Sinclair : Preliminary recommendations on corpus typology. *EAGLES Document TCWG-CTYP/P*, 1996. URL http://www.ilc.pi.cnr.it/EAGLES/corpustyp/corpustyp.html).

Parman Singh : *Hindi Lexicography : Past, Present, and Future*. Lambert Academic Publishing, 2013.

Rajendra Singh et Ramakant Agihotri : *Hindi Morphology : a word-based description*. Motilal Banarsidass, Delhi, 1997.

M. Singh Gill, G. Singh Lehal et S. Sharma Joshi : Part of speech tagging for grammar checking of Punjabi. *The Linguistics Journal*, 4(1) :6–22, 2009.

Michel Sparer : Peut-on faire de la traduction juridique ? comment doit-on l'enseigner ? *Meta : Journal des traducteurs*, 47(2) :266–278, 2002.

S.N. Sridhar : Language variation, attitudes, and rivalry : The spread of Hindi in India. *In Language Spread and Language Policy : Issues, Implications, and Case Studies*. GURT 87', Georgetown University Press, 1987. URL https://repository.library.georgetown.edu/bitstream/handle/10822/555480/GURT_1987.pdf?sequence=1#page=318.

Nazia Suleman et Masroor Dr. Sibtain : Term formation : An historical perspective to coin new terms in the Urdu language. *Modern Journal of Language Teaching Methods*, 9(1), janvier 2016. Consulté le 16 juillet 2019.

Line Teillot et Pascale Ubransky : *Droit pénal général*. Bréal, Paris, 2008.

Projet Textométrie : Projet Textométrie, 2007. URL http://textometrie.ens-lyon.fr/.

Bernard Thiry : *Problèmes de jurilinguistique contrastive : les équivalences interlinguistiques en droit*. HEC Liège, Bruxelles, 2008.

Bholanath Tiwari : *Kośāvijñāna*. Shabdakara, Delhi, 1979.

U. N. Tiwari : *hindī bhāṣā kā udgam aur vikās*. Lokbharati Prakashan, Allahabad, 1998.

Viktor Tròn, Andràs Kornai, Gyòrgy Gyepesi, Làszlò Nèmeth, Pèter Halàcsy et Dàniel Varga : Hunmorph : Open Source Word Analysis. *In Proceedings of the Workshop on Software*, Software '05, pages 77–85, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics. URL http://dl.acm.org/citation.cfm?id=1626315.1626321.

Amy Bik May Tsui et James W. Tollefson : *Language policy, culture, and identity in Asian contexts*. Lawrence Erlbaum Associates, Mahwah, N.J., 2007. ISBN 978-0-8058-5693-4 978-0-8058-5694-1.

Agnès Tutin : Autour du lexique et de la phraséologie des écrits scientifiques. *Revue Française de Linguistique Appliquée*, pages 5–14, 2007. URL http://www.cairn.info/article.php?ID_ARTICLE=RFLA_122_0005.

Saba Urooj, Sarmad Hussain, Asad Mustafa, Rahila Parveen, Farah Adeeba, Tafseer Ahmed Khan, Miriam Butt et Annette Hautli : The CLE Urdu POS Tagset. *In* Calzolari (Conference Chair) Nicoletta, Choukri Khalid, Declerck Thierry, Loftsson Hrafn, Maegaard Bente, Mariani Joseph, Asuncion Moreno, Jan Odijk et Stelios Piperidis, éditeurs : *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 26–31, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.

Marc Van Campenhoudt : Recherche d'équivalences et structuration des réseaux notionnels : le cas des relations méronymiques. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(1) :53–83, 1996.

Prasad Vasudevanand : *ādhunik hindī vyākaraṇ aur racnā*. Bharati Bhavan, Patna, 1993.

Lawrence Venuti : *The Translation Studies Reader*. Routledge, 2000.

Hugues-Jehan Vibert, Jean-Pierre Jouvelot et Benoît Pin : Legivoc : connecter les droits dans un monde en mouvement. 1, n°1, 2013. URL `https://ojs.law.cornell.edu/index.php/joal/article/view/19/23`. Consulté le 15 juillet 2019.

Jean-Paul Vinay et Jean Darblenet : *Stylistique comparée du français et de l'anglais, 2^e éd.* Didier, Paris, 1977.

Raghu Vira : *A Comprehensive English-Hindi Dictionary of Governmental and Educational Words and Phrases*. International Academy of Indian Culture, 1962.

Anne Wagner et Jean-Claude Gémar : Les enjeux de la jurilinguistique et de la juritraductologie. *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, 28(1) :1–8, Mars 2015. ISSN 1572-8722. URL `https://doi.org/10.1007/s11196-015-9405-1`.

Tahir Wasti : *The Application of Islamic Criminal Law in Pakistan, Sharia in Practice*. Brill, 2009.

Eric Wehrli, Violeta Seretan, Luka Nerima et Lorenza Russo : Collocations in a rule-based MT system : A case study evaluation of their translation adequacy. 2009.

Benjamin Lee Whorf : *Linguistique et anthropologie*. Denoël, Paris, 1969.

Jerzy Wroblewski : Les langages juridiques : une typologie. 1988. ISSN 0769-3362. URL `https://www.persee.fr/doc/dreso_0769-3362_1988_num_8_1_983`.

Blanco Xavier : Les pragmatèmes : définition, typologie et traitement lexicographique. *Verbum*, (4) :17–25, 2013.

M Mohamed Yoonus et Samar Sinha : A hybrid POS Tagger for Indian Languages. *Language in India*, 11(9), 2011.

Almas Yousif et Khurshid Ahmad : Lolo : A System based on Terminology for Multilingual Extraction. *In* Association for Computational Linguistics, éditeur : *Proceedings of the Workshop on Information Extraction Beyond The Document*, page 56–65, Sydney, 2006.

Yvon : Etiquetage Morphosyntaxique - PDF. Rapport technique, École Nationale Supérieure des télécommunications, Département Informatique et Réseaux, Apr 2007. URL `https://docplayer.fr/14713909-Etiquetage-morphosyntaxique.html`. [Consulté le 2 avril 2019].

Jamshed ʿĀlam : *Daftarī zabān kā taʿāruf*. Muqtadirah-yi Qaumī Zabān, Islāmābād, 2002. ISBN 978-969-474-080-5.

Bénédicte Diot-Parvaz Ahmad

Production de ressources multilingues pour l'aide à la traduction du droit pénal en hindi, ourdou et français

Résumé

Comment concilier l'impératif d'assistance linguistique à toute personne ne parlant pas français et l'absence de ressources linguistiques standardisées pour traduire des combinaisons de langues génétiquement et culturellement distantes ? C'est le problème posé par la traduction du hindi et de l'ourdou en France dans le contexte judiciaire. Le hindi et l'ourdou, langues sœurs parlées en Inde et au Pakistan, ont des liens distants avec le français. Les systèmes judiciaires dont elles sont le moyen d'expression proviennent de l'héritage colonial britannique qui repose sur la common law. Ce travail propose, à travers l'analyse d'un corpus de documents variés, de créer des ressources terminologiques et phraséologiques afin d'aider le traducteur-interprète à trouver des équivalences de traduction multilingues. Dans un premier temps, nous abordons les différences entre les systèmes judiciaires et le statut des langues de travail dans les trois pays. Nous étudions ensuite leurs procédures judiciaires et observons comment elles s'inscrivent dans des genres définis par un lexique et une phraséologie plus ou moins accessibles aux non spécialistes. Enfin, nous proposons une méthode d'extraction des termes et d'alignement par sous-corpus afin de faire ressortir les équivalences terminologiques ou traductionnelles du genre judiciaire entre ces langues. Ce travail, qui met en lumière les relations entre le texte, le contexte et les mots, fournit aux professionnels de la traduction et de l'interprétation des ressources attestées, adaptées au domaine de spécialité et contextualisées.

Mots clés : traduction judiciaire, alignement de corpus, genres textuels, linguistique de corpus, lexicologie, phraséologie, Inde, Pakistan, hindi, ourdou

Abstract

Production of multilingual resources helping in the translation of the judicial process in Hindi, Urdu and French

Is it possible to reconcile the need for language assistance to all non-French speakers with the lack of standardized language resources for translating combinations of languages which are genetically and culturally remote? This is the issue raised by Hindi and Urdu translation in France, in the judicial context. Hindi and Urdu are sister languages spoken in India and Pakistan. They have remote links with French. The judicial systems in which they are used come from the British colonial heritage based upon common law. Through the analysis of a corpus of various documents, this work is aimed at producing terms and phraseological resources in order to assist the translator-interpreter in finding out translation equivalences between languages. First, we will explore the differences between the three countries' judiciaries, as well as the status of the Hindi, Urdu and French languages. Then, we will study the judicial proceedings in all three countries and examine how they are embedded into text genres. We will see to what extent their lexicons and phraseologies are adapted for non-specialists. Eventually, we will propose a method for term extraction and sub-corpus alignment in order to stress term or translation equivalences between these languages in the judicial genre. This work, which sheds light on the relations between text, words and context, provides actual field-specific resources for judicial translation and interpretation professionals.

Keywords : Judicial Translation, Corpora Alignment, Text Genres, Corpus Linguistics, Lexicology, Phraseology, India, Pakistan, Hindi, Urdu