

Relations entre types de violation des libertés syndicales garanties par les conventions de l'OIT : Une analyse statistique implicative des résultats d'une fouille de texte

Matthias Studer*, Gilbert Ritschard*
Lucio Baccaro****, Irini Georgiou***, Nicolas S. Müller*, Djamel A. Zighed****

*Département d'économétrie, Université de Genève, Suisse
{matthias.studer, gilbert.ritschard, nicolas.muller}@metri.unige.ch
<http://www.unige.ch/ses/metri/>
** MIT, Cambridge MA

***International Institute of Labour Studies, IILS (OIT), Genève
****Laboratoire ERIC, Université de Lyon 2, France
abdelkader.zighed@univ-lyon2.fr

Résumé. Au travers d'une analyse de violations observées de conventions de l'OIT, le but de ce papier est de montrer comment l'analyse implicative complémente avantageusement des analyses exploratoires plus classiques. Plus précisément, nous nous intéressons aux types de violations relevées par les experts chargés d'observer le respect des Conventions n° 87 et n° 98 de l'OIT sur les droits syndicaux. Les données sont des prédictions obtenues à l'aide d'un apprentissage fondé sur la fouille de texte. Nous comparons essentiellement trois méthodes soit l'analyse statistique implicative, l'analyse factorielle des correspondances et la classification automatique des individus. Nous discutons les apports de chacune de ces méthodes.

1 Introduction

Dans cet article, nous montrons l'apport de la statistique implicative dans l'étude des systèmes de dialogue social. Par système de dialogue social, nous entendons les systèmes sociopolitiques par lesquels les travailleurs ont la liberté de créer les organisations de leur choix, de négocier collectivement les conditions de travail et de participer au travers de leurs associations à l'élaboration et à l'application des politiques qui les concernent (Baccaro, 2003)¹. Nous nous centrons ici sur l'analyse du respect des conventions n° 87 et n° 98 de l'Organisation Internationale du Travail (OIT). Ces conventions fixent un ensemble de conditions requises à l'émergence d'un système qui permet la négociation libre entre organisations de travailleurs et d'employeurs. À ce titre, elles sont considérées comme fixant un ensemble de droits fondamentaux du travail². La convention n° 87 traite de l'indépendance des syndicats et de la protection des syndicats vis-à-vis de l'État. Elle garantit également les droits d'action syndicale. La convention n° 98 traite des relations entre syndicats et employeurs (notamment ingérences et discrimination antisyndicale) ainsi que des conditions-cadres nécessaires à l'émergence des négociations collectives. L'étude du respect de ces conventions nous permettra donc d'aborder la constitution d'un terreau légal propice à l'émergence d'un système de dialogue social.

L'article est structuré de la manière suivante. Dans un premier temps, nous présentons les données à notre disposition sur le respect des conventions n° 87 et 98. Nous cherchons à comprendre, dans une optique descriptive, les relations entre les types de violations des conventions précitées. Pour ce faire, nous cherchons tout d'abord à identifier des types de systèmes légaux par le biais d'une classification automatique. Nous complétons l'étude par une analyse factorielle des correspondances multiples et enfin une analyse par les graphes d'implication. Nous exploitons également différentes méthodes de regroupement hiérarchique des variables afin d'affiner notre perception des relations de dépendance entre violations. Finalement, nous concluons en faisant le point sur les divers enseignements de l'analyse réalisée, tant du point de vue de notre problématique du dialogue social que sur le plan méthodologique.

2 Présentation des données

Chaque année, le CEACR (« Committe of Experts on the Application of Conventions and Recommendations ») de l'OIT publie un volume de près de 800 pages rassemblant les rapports sur le respect des

¹ Étude réalisée avec le soutien financier de la Fondation RUIG (www.ruig-gian.org).

² Plus précisément, depuis l'adoption en 1998 de « la déclaration de l'OIT relative aux principes et droits fondamentaux au travail »

conventions de l'OIT par les pays signataires de celles-ci. Les données que nous utilisons dans cet article concernent exclusivement les conventions no 87 et 98. Elles résultent d'une procédure d'étiquetage automatique de ces rapports à l'aide de techniques fondées sur la fouille de texte. Cette procédure d'analyse textuelle automatique permet de prédire la probabilité de la présence dans chaque rapport de chacun de six types de violation de la convention 87 et cinq de la convention 98. Le tableau ci-dessous décrit les différents types de violations considérés, seuls ceux effectivement observés dans les rapports ayant été retenus.

Convention 87

- v3 *Droit d'établir et de rejoindre les organisations de travailleur.* Aucune catégorie de travailleurs ne peut être exclue du droit d'établir et/ou de rejoindre un syndicat, excepté l'armée et les forces de police.
- v4 *Pluralisme syndical.* La loi doit mettre en place un système permettant le pluralisme syndical, c'est-à-dire la présence de plusieurs syndicats.
- v6 *Élection des représentants/ Critère d'éligibilité.* Les représentants syndicaux doivent pouvoir être élus en toute liberté sans restriction de métiers ou de nationalité, par exemple
- v7 *Organisation des activités syndicales.* Les syndicats doivent être libres d'organiser leurs activités et de formuler leurs programmes sans interférence de l'État. Les syndicats doivent bénéficier d'une protection adéquate de leur droit de propriété et de leur indépendance administrative.
- v8 *Enregistrement des organisations de travailleurs.* L'établissement et l'enregistrement des syndicats ne doivent pas être soumis à des formalités ou à des contrôles excessifs. Les syndicats doivent avoir le droit de décider de leur règlement interne.
- v9 *Restrictions des droits à l'action syndicale.* Le droit à l'action syndicale ne peut être interdit ou restreint excepté en cas de crise nationale grave, pour les travailleurs des services essentiels dans le sens strict du terme et les fonctionnaires exerçant une autorité au nom de l'État.

Convention 98

- w1 *Discrimination antisyndicale.* La législation doit mettre en place des sanctions suffisamment dissuasives contre la discrimination syndicale.
- w2 *Actes d'ingérences.* La législation doit mettre en place des sanctions suffisamment dissuasives contre l'ingérence des employeurs dans les organisations syndicales.
- w5 *Exclusion du droit à la négociation collective.* Tous les travailleurs sans distinctions doivent pouvoir bénéficier du droit à la négociation collective.
- w6 *Désignation de l'agent de négociation/Critères de représentativité des syndicats.* Les droits à la négociation collective des syndicats ne doivent pas être sujets à des conditions excessives. La désignation du syndicat le plus représentatif doit être établie selon des critères objectifs.
- w8 *Restriction sur les domaines négociables/approbation et enregistrement des conventions collectives.* Les parties doivent avoir le droit de déterminer les domaines négociables. L'approbation et l'enregistrement des conventions collectives ne doivent pas pouvoir être refusés, exceptés pour des problèmes de forme ou si leurs termes sont en conflit avec les minimums standard fixés par le droit du travail.

TAB. 1 – Description des types de violation des conventions 87 et 98

La procédure de fouille de texte suivie avait pour but de dériver des règles de prédictions qui puissent être utilisées par des experts du domaine non spécialistes de fouille de texte. Le choix a ainsi porté sur une approche où les règles peuvent être mises en œuvre sur des textes bruts non prétraités (pas de marquage grammatical, pas de lemmatisation, ...). Le marquage grammatical a cependant été exploité dans la phase d'apprentissage pour identifier des concepts descripteurs des textes. Ceux-ci, définis chacun sous forme d'une liste de termes représentatifs, ont en effet été établis en regroupant, en étroite collaboration avec les experts du domaine, les expressions caractéristiques des textes. Les descripteurs ont été quantifiés avec leur *tf \times idf*, et les règles de prédiction ont été obtenues par le biais d'arbres d'induction. Pour plus de détails, voir Ritschard et al. (2007).

Les prédictions obtenues avec les règles ainsi dérivées ne sont évidemment pas sans erreurs. Le tableau 2 présente à titre d'exemple les taux d'erreur obtenus pour chaque type de violation de la convention n° 87.

Violation	Taux d'erreur en apprentissage	Taux d'erreur en validation croisée	Erreur standard de la validation croisée	Nombre d'erreurs sur l'échantillon test (taille 21)
v3	14.10%	n.d.*	n.d.*	3
v4	5.13%	5.13%	2.50%	0
v6	12.82%	14.1%	3.94%	4
v7	15.38%	n.d.*	n.d.*	7
v8	7.69%	7.69%	3.01%	4
v9	2.56%	2.56%	1.79%	2

* non disponible puisque le premier nœud a été choisi théoriquement

TAB. 2 – Taux d'erreur de classification des rapports du CEACR sur la convention n° 87.

Les taux d'erreurs restent, en particulier pour les violations v3, v6 et v7, relativement conséquents, et il en va de même pour ceux des violations de la convention n° 98. Cela est évidemment gênant du point de vue de la fiabilité de prédictions individuelles pour un pays donné. Toutefois, nous utilisons ici les prédictions issues de la fouille de texte pour faire une analyse synthétique de l'ensemble des pays, dans le but de dégager en particulier des « régimes » de dialogues sociaux. Dans cette perspective agrégée, on peut raisonnablement penser que les erreurs, en se compensant les unes les autres, n'affectent que de façon limitée les tendances que nous cherchons à mettre en évidence. Notons, par ailleurs, que ces données constituent l'une des premières sources d'information aussi détaillée sur le sujet.

Les mentions de violation de l'application de la convention n° 87 montrent une certaine inertie. Lorsqu'un problème est identifié par le CEACR, il peut s'écouler quelque temps avant que celui-ci soit reporté dans les rapports. Ainsi, lorsqu'un problème est énoncé, ceci ne signifie pas que le problème est apparu à ce moment, mais plutôt que le CEACR a décidé pour diverses raisons de le mentionner. De plus, le CEACR ne fait pas nécessairement un rapport par année et par pays, mais plutôt, dans la situation la plus courante, un rapport tous les deux ans. L'absence de rapport sur une longue période signifie généralement l'absence de problème pendant celle-ci. Ainsi, il n'est pas possible de se baser uniquement sur les rapports publiés une année donnée pour avoir un compte rendu de la situation. Nous avons donc décidé de faire porter notre analyse sur l'ensemble des rapports publiés pendant la période 1997-2002. Nous avons assigné à chaque pays et pour chaque type de violation, la probabilité maximale (prédite) de la violation pendant la période 1997-2002. Pour limiter le nombre de faux positifs, nous avons choisi de prédire la présence de la violation j dans un pays donné lorsque cette probabilité maximale dépassait les 80 %. Le tableau 3 montre les pourcentages de pays pour lesquels nous avons prédit chacune des violations. Les pourcentages sont calculés sur la base du nombre de pays concernés, c'est-à-dire ceux qui ont signé la convention.

Violation	Pourcentage des pays
v3 Droit d'établir et de rejoindre les organisations de travailleur	36.9%
v4 Pluralisme syndical	22.7%
v6 Élection des représentants/ Critère d'éligibilité	31.9%
v7 Organisation des activités syndicales	23.4%
v8 Enregistrement des organisations de travailleurs	28.4%
v9 Restrictions des droits à l'action syndicale	61.0%

TAB. 3 – Pourcentages des pays concernés par chaque type de violation pour la convention 87.

La situation est la même pour les rapports de la convention n° 98, nous avons donc suivi une procédure identique. Nous distinguons dans cette convention cinq types de violation différents. D'une manière générale, il semble y avoir moins de violations de cette convention que de la précédente. En effet, les pourcentages de pays concernés par chaque type de violation sont beaucoup plus bas. De même, le pourcentage de pays ne subissant aucune des violations est plus élevé dans le cas de la convention 98 (60 %) que pour la 87 (34 %) alors même que plus de pays ont signé la convention 98.

Violation	Pourcentage des pays
w1 Discrimination antisyndicale	25.6%
w2 Actes d'ingérences	15.4%
w5 Exclusion du droit à la négociation collective	18.6%
w6 Désignation de l'agent de négociation/Critères de représentativité des syndicats	10.3%
w8 Restriction sur les domaines négociables/approbation et enregistrement des conventions collectives	6.4%

TAB. 4 – Pourcentages des pays concernés par chaque type de violation pour la convention 98.

3 Analyses

Nous nous proposons de comparer les résultats de plusieurs méthodes de statistique exploratoire, dont l'ASI, afin d'en souligner la complémentarité. Ainsi, nous procédons premièrement à une analyse par classification automatique des pays dans le but d'identifier des ensembles logiques ou systèmes de violations caractéristiques des groupes de pays ainsi mis en évidence. Nous présentons ensuite une analyse factorielle des correspondances multiples pour essayer de comprendre les dimensions qui fondent cette logique. Nous discutons également une analyse sous forme de graphe implicatif qui fournit un éclairage complémentaire permettant de mieux structurer les relations entre types de violations. Enfin nous examinons ces liens à la lumière de regroupements hiérarchiques orientés et non orientés.

Nous focalisons notre présentation sur les résultats de l'analyse des violations de la convention no 87 qui donne les résultats les plus intéressants, et commentons beaucoup plus brièvement ceux obtenus pour la convention no 98.

3.1 Typologie des pays

Afin de mieux cerner les systèmes légaux existants, nous avons choisi d'effectuer une analyse de classification automatique des individus, les pays dans notre cas, avec l'intention de constituer une typologie. Cette analyse a pour but de regrouper les pays pour lesquels la fouille de texte indique des situations semblables en termes de violations de la convention. Ainsi, nous cherchons à identifier des systèmes. Nous avons effectué les analyses avec SPSS séparément pour les conventions n° 87 et 98, puisque les pays n'ont pas nécessairement signé les deux conventions. Nous présentons ici l'analyse pour la convention n° 87. Nous avons effectué cette analyse en deux étapes. Premièrement, nous avons extrait les facteurs principaux à l'aide d'une analyse factorielle des correspondances multiples (Greenacre, 1993, Lebart et al., 2000). Deuxièmement, nous avons identifié les groupes en utilisant une méthode d'agglomération hiérarchique avec le critère de Ward (Anderberg, 1973; Lebart et al. 2000) sur la base des six dimensions factorielles. L'étude du schéma d'agglomération nous montre que la part la plus importante de réduction de l'inertie intra-classe est obtenue avec une partition en quatre groupes. Pour celle-ci, l'inertie intra-classe est réduite de plus de 60 %.

Le tableau 5 donne les pourcentages des membres de chaque groupe qui, toujours au vu de la fouille de texte, sont concernés par les violations. Remarquons que les quatre groupes semblent relativement équilibrés puisque le plus important regroupe 36 % des cas contre 13 % pour celui qui contient le moins de pays.

Violation	Groupe				Total	ν de Cramer
	Aucune violation	Restriction sur le pluralisme syndical	Restrictions sur l'action	Restrictions organisationnelles et sur l'action		
v3 Droit d'établir et de rejoindre les organisations de travailleur	0%	38.9%	51.1%	81.5%	36.9%	0.64
v4 Pluralisme syndical	0%	100.0%	0.0%	51.9%	22.7%	0.85
v6 Élection des représentants/ Critère d'éligibilité	3.9%	33.3%	22.2%	100.0%	31.9%	0.74
v7 Indépendance administrative et organisation des activités syndicales	2.0%	0.0%	11.1%	100.0%	23.4%	0.89
v8 Enregistrement des organisations de travailleurs	0%	44.4%	26.7%	74.1%	28.4%	0.60
v9 Restrictions des droits à l'action syndicale	0%	88.9%	95.6%	100.0%	61.0%	0.94
Pourcentage de pays dans le groupe	36.2%	12.8%	31.9%	19.1%	100%	
Nombre moyen de violations	0.06	3.06	2.07	5.07	2.04	

TAB. 5 – Résultats de la classification automatique.

Les valeurs du ν de Cramer nous indiquent que la décomposition en quatre groupes est fortement liée à chaque type de violation. En fait, on voit bien que les pourcentages de pays concernés par chaque violation varient tous fortement d'un groupe à l'autre. Les violations v9, v7, v6 et v4 montrent la plus forte association avec la partition retenue.

L'interprétation du premier groupe est assez directe. En effet, il regroupe les pays qui ne subissent aucune violation. Les groupes qui suivent nous permettent d'identifier trois types de systèmes légaux qui violent les droits fondamentaux définis dans la convention n° 87.

Le second groupe rassemble les pays connaissant des violations du pluralisme syndical sans toutefois subir de contraintes excessives sur l'organisation de leurs activités. Ces pays semblent également connaître légèrement plus fréquemment des conditions excessives dans les procédures d'enregistrement de nouveaux syndicats. Ceci semble relativement logique puisqu'il s'agit d'une des méthodes pour mettre en place un monopole syndical. Il existe ainsi une forme de contrôle des associations de travailleurs.

Le troisième groupe regroupe les pays posant des restrictions des droits à l'action syndicale (violation v9) ainsi que, dans une moindre mesure, sur le droit d'établir et rejoindre des syndicats. D'autres types de violations peuvent coexister, mais les pourcentages sont généralement inférieurs à la situation générale. Il existe ainsi une forme de contrôle sur les possibilités d'action des syndicats.

Finalement, dans le dernier groupe, nous observons des pourcentages élevés pour chaque type de violation. Le nombre moyen de violations se situe à cinq sur un maximum de six. Si l'on compare ce groupe à la situation générale, les violations v6 (élection des représentants) et v7 (indépendance administrative et organisation des activités syndicales) semblent les plus présentes. Ainsi, dans ce groupe de pays, les gouvernements semblent privilégier une forme de contrôle de l'administration et des activités internes aux syndicats. Cette forme de

contrôle semble généralement aller de pair avec les autres types de violations qui, si l'on excepte la violation v4 du pluralisme syndical, affichent tous les pourcentages les plus élevés dans ce groupe.

Ces groupes nous permettent donc de distinguer trois types de systèmes légaux qui violent la convention n° 87 : la mise en place d'un monopole syndical, les restrictions sur l'action syndicale et une forme de contrôle des activités internes aux syndicats.

3.2 Analyse factorielle des correspondances multiples

Nous avons construit les groupes sur la base des facteurs d'une analyse des correspondances multiples. Les figures 1 et 2 présentent leurs positions par rapport aux types de violation dans l'espace des trois premières dimensions. Les cercles pleins indiquent la position des violations et les vides celle de leur absence. Les losanges permettent de localiser les groupes (clusters). Finalement, nous avons également ajouté les positions moyennes de chaque région des pays signataires de la convention n° 87. Le graphique de la figure 1 permet de considérer la première dimension par rapport à la deuxième, celui de la figure 2 affiche la troisième dimension sur l'ordonnée.

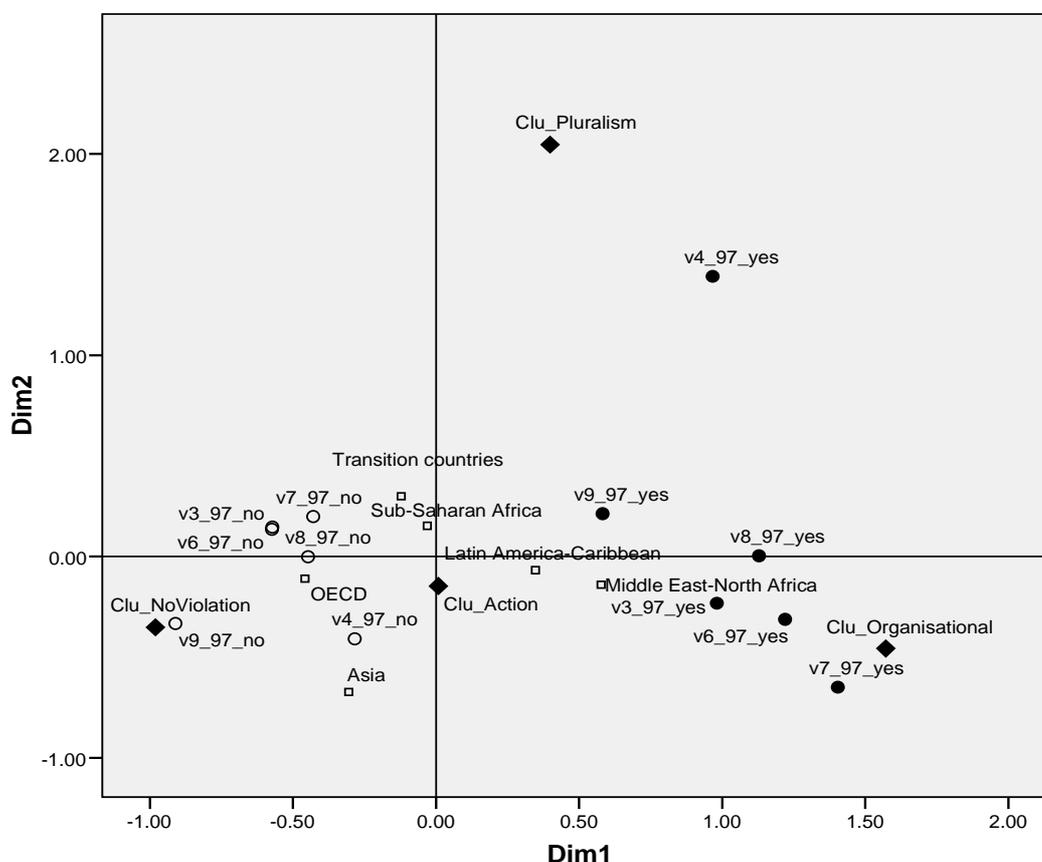


FIG. 1 – Représentations des violations dans l'espace des deux premières dimensions.

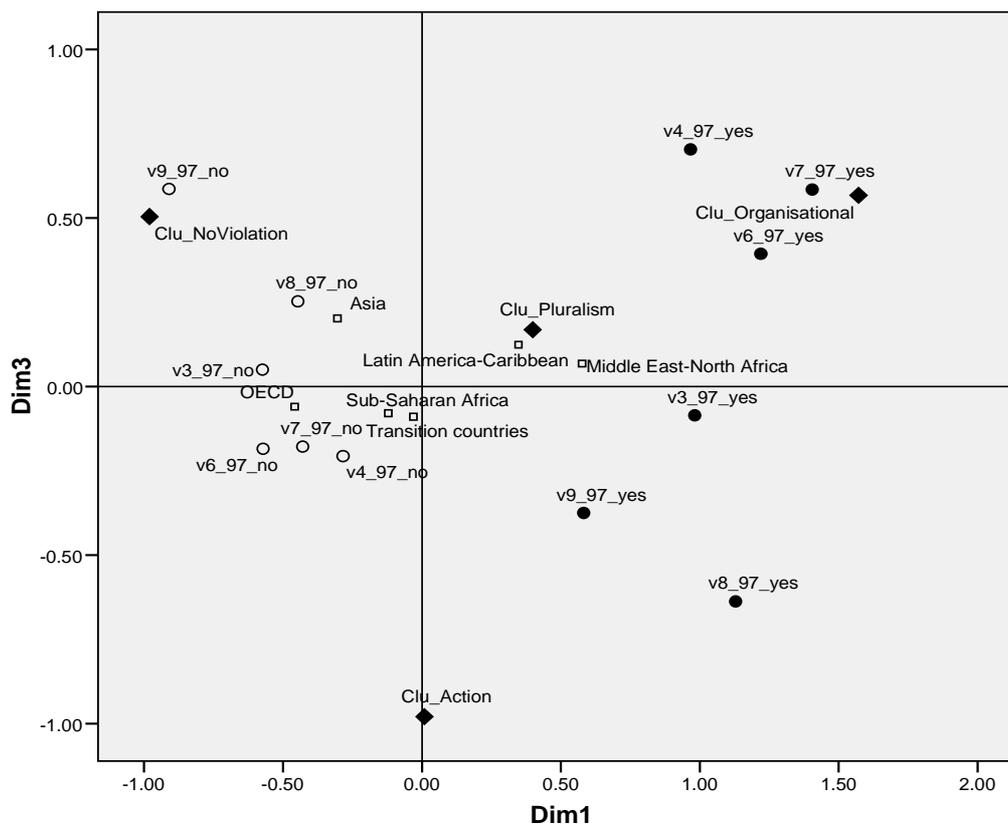


FIG. 2 – Représentations des violations dans l'espace de la première et la troisième dimension.

On remarque ainsi que la première dimension oppose la présence (à droite) à l'absence de violation. Ainsi, le cluster sans violation est le point le plus à gauche et s'oppose au cluster avec le plus de violations (restriction organisationnelle). Cette dimension est donc très fortement corrélée au nombre total de violations différentes subies par un pays ($\rho=0.998$; $p<0.0001$). Elle pose ainsi un « ordre » d'apparition des violations. La violation v9, la plus courante, se trouve le plus à gauche alors que la plus rare (violation v7) se trouve le plus à droite. Cette dimension reproduit la plus grande partie de la variation observée (52.9 %). La seconde dimension en explique une part nettement moindre (14.1 %). Elle semble mettre en évidence une opposition entre les violations v3, v6 et v7 et la violation v4. Ainsi, le cluster « pluralisme » est opposé aux autres selon cet axe. Dans la figure 2, on remarque que la troisième dimension (11.8 % de la variation) met en évidence le cluster « restriction de l'action syndicale » vis-à-vis des autres.

Il est intéressant d'observer les positions moyennes des régions. Deux raisons peuvent nous amener à penser qu'il existe un lien entre les violations et les régions. Premièrement, il peut exister des formes de cultures politiques selon les régions qui devraient se répercuter sur les types de systèmes légaux. Deuxièmement, il est reconnu que les syndicats de certaines régions, notamment de l'Amérique latine, font plus facilement recours à l'OIT dans le cadre de leurs actions politiques (Kucera, 2004). Ainsi, on peut s'attendre à plus de mentions de violations dans les rapports. C'est ce que l'on observe, puisque ces pays (ici regroupés avec les pays caribéens) ont une position moyenne positive sur la première dimension qui est, comme nous l'avons vu, fortement corrélée au nombre de violations. On observe une situation similaire, quoique plus marquée, pour les pays du Moyen-Orient et de l'Afrique du Nord qui semblent ainsi subir, en moyenne, plus de violations de la convention n° 87 que les autres groupes régionaux. Ces régions semblent plus proches des violations « organisationnelles » que les autres groupes. Il est également intéressant de noter la position des pays de l'OCDE qui se placent à gauche (proche du cluster sans violation). Ceci ne constitue pas une réelle surprise et était attendu. On peut relever également que les pays asiatiques semblent subir moins fréquemment des violations de la convention 87. Cela ne

concerne évidemment que les pays qui ont signé la convention et qui sont relativement peu nombreux dans cette zone asiatique.

3.3 Graphe implicatif

Pour mieux comprendre les relations entre les types de violations qui ressortent des analyses précédentes, il est intéressant d'examiner les graphes d'implication de la figure 3 obtenus avec le logiciel CHIC (Couturier et Gras, 2005, Couturier et al., 2006). Ces graphes ont été construits avec la mesure d'implication « entropique » qui est plus discriminante. Dans le graphe de gauche, on peut observer les relations entre les violations. Nous retrouvons ici des implications entre violations qui correspondent aux liens relevés dans le cadre de l'analyse en cluster. Par exemple, la violation n° 7 implique les violations 6 et 9, ce qu'on peut mettre en relation avec le fait que ces trois violations formaient la base du cluster « restrictions organisationnelles et sur l'action ». Toutefois, le sens des relations est ici clairement affiché ce qui les rend plus directement interprétables. Ainsi, les restrictions de l'action syndicale constituent les violations les plus courantes et, dans bien des cas, la présence d'un autre type de violation semble impliquer la violation v9. Ce résultat apparaissait de manière bien moins claire dans l'analyse factorielle des correspondances multiples, puisque cette violation affichait une valeur relativement basse sur le premier axe, fortement corrélé avec le nombre moyen de violations. On remarque également immédiatement qu'il ne semble pas exister de systèmes à proprement parler entre les autres types de violations. C'est ce que l'on trouvait dans le cluster « restriction de l'action syndicale ». Notons toutefois que l'on peine à identifier des groupes qui permettraient de dégager des types, ou la présence de méta-règles.

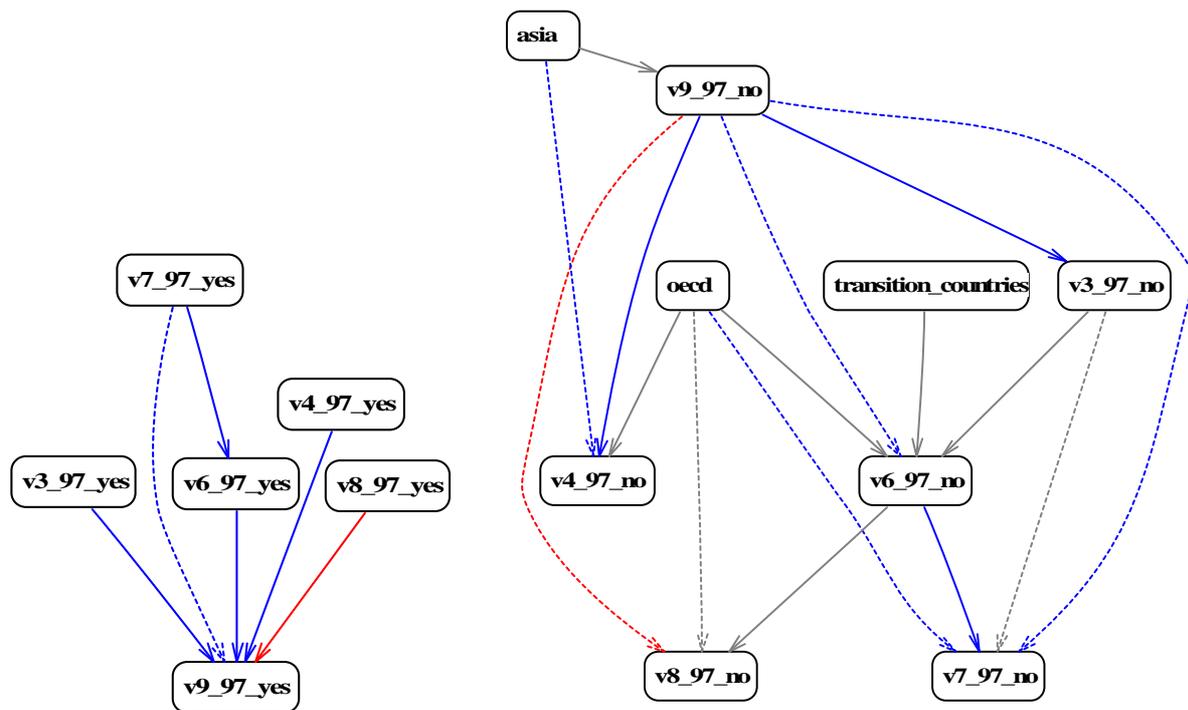


FIG. 3 – Graphe implicatif des violations de la convention n° 87. Seuils retenus 99% (rouge), 95% (bleu). Dans le graphe de droite 75% (gris).

Aucune relation des présences de violations avec les régions n'est significative aux différents seuils retenus. Nous avons donc créé un nouveau graphe sur la base des absences de violation (graphe de droite). Dans celui-ci, nous avons retenu des seuils plus bas en ajoutant en gris les relations au seuil de 75%. Si l'on en reste à la présence-absence de violations, les deux graphes sont très similaires, celui de droite offrant une version « inversée » de celui de gauche. Ainsi, si la présence de la violation v7 (organisation des activités) implique la violation v9 (droit sur l'action syndicale), l'absence de la violation v9 implique l'absence de la violation v7 dans le graphe de droite. Ce résultat est trivial puisqu'il est construit sur la base de données exactement inverses.

Cependant, outre les relations entre absences de violation, le graphe de droite permet de mettre en lumière des liens intéressants avec les régions. Ainsi, faire partie des pays membres de l'OCDE semble impliquer qu'on ne connaisse pas de restrictions excessives sur l'organisation et l'administration interne des syndicats (violation v7). On observe également, mais au seuil de 75 % seulement, une relation similaire avec l'absence de monopoles syndicaux (violation v4), les procédures d'enregistrement (violation v8) ou les procédures de désignation des représentants syndicaux (violation v6). Ainsi, on retrouve les résultats précédents, mais de manières beaucoup plus précises, puisque rattachés à des violations spécifiques. Les pays en transitions (ex-union soviétique) ne semblent pas mettre en place des restrictions sur les critères d'élection des représentants syndicaux. Notons finalement que les pays signataires asiatiques (relativement peu nombreux) ne semblent pas subir de violations du droit à l'action syndicale ou de monopoles syndicaux. Les relations entre les régions et l'absence de violations viennent conforter ce que nous avons observé dans le cadre de l'analyse factorielle des correspondances multiples. De fait, ce deuxième graphique nous précise le sens des relations que nous observions pour les valeurs négatives de la première dimension de l'analyse factorielle des correspondances.

Ce deuxième graphe nous montre l'importance de prendre en considération l'« absence » dans les graphes implicatifs. En effet, dans le cas qui nous intéresse, nous pouvons donner un sens à « l'absence de violation ». Dès lors, les relations qui sont liées à cette absence nous intéressent également puisqu'elles dénotent une forme d'implication « négative » (dans le sens de la corrélation négative).

Chemin	Pays des Caraïbes et Latino-Américains	Moyen-Orient et Afrique du Nord	Asie	OCDE
v7_97_yes → v6_97_yes → v9_97_yes	*	**		
v7_97_yes → v6_97_yes		**		
v7_97_yes → v9_97_yes		**		
v6_97_yes → v9_97_yes	**			
v4_97_yes → v9_97_yes				
v3_97_yes → v9_97_yes		*		
v8_97_yes → v9_97_yes		*		
<i>Absence de violation</i>				
v9_97_no → v4_97_no			**	*
v9_97_no → v6_97_no → v7_97_no			**	*
v9_97_no → v6_97_no			**	*
v6_97_no → v7_97_no				**
v9_97_no → v8_97_no			**	**
v9_97_no → v3_97_no			**	

Une étoile indique une typicalité avec un risque à 10 %, deux étoiles à 5 % et trois étoiles à 1 %.

TAB. 6 – *Typicalité des variables supplémentaires par chemin.*

Notons qu'on ne retrouve pas les régions dans le graphe implicatif des présences de violation. Il aurait fallu introduire pour cela des variables du type « non OCDE » qui ne font pas tellement sens. Toutefois, si l'on considère les régions comme variables supplémentaires, on constate que deux régions sont typiques des implications entre violations et plus généralement de « chemins de violations ». Autrement dit, certaines implications entre violations apparaissent plus fréquemment dans certaines régions que dans d'autres. Les pays du Moyen Orient et de l'Afrique du Nord de même que les pays des Caraïbes et Latino-américains semblent relativement typiques du chemin « v7 → v6 → v9 ». Ceci aussi correspond aux observations faites dans l'analyse factorielle des correspondances multiples sur le lien entre ces trois variables et ces régions. On voit cependant ici que ce sont plus les chemins d'implication entre ces violations que les violations elles-mêmes qui sont concernés.

Similairement, si l'on observe les relations entre absences de violations, on remarque que les pays signataires d'Asie ainsi que ceux de l'OCDE semblent être moins sujets à certaines implications entre violations. Ainsi, les pays de l'OCDE sont typiques de la situation où l'absence de la violation v6 va de pair avec l'absence de la v7, il

en va de même entre les absences des violations v9 (action syndicale) et v8 (procédures d'enregistrement des syndicats). L'interprétation de la typicalité des pays d'Asie est plus problématique puisque comparativement peu de pays ont signé la convention. L'explication est peut-être que les pays asiatiques tendent à signer la convention qu'à partir du moment où il n'existe plus de problèmes légaux avec celle-ci.

En regardant les typicalités des chemins, nous avons de nouveau mis en évidence l'importance de prendre en considération les négations, les « absences » dans notre cas, à partir du moment où celles-ci font sens. En effet, les variables supplémentaires peuvent être typiques de chemins entre négations. Les relations que nous avons mises en évidence ne sont pas les mêmes que dans l'analyse entre « présences ».

3.4 Regroupement hiérarchique des violations

Nous avons également procédé à plusieurs types de regroupement hiérarchique des violations. La figure 4 ci-dessous montre l'arbre cohésitif pour des violations de la convention n° 87. On retrouve ici aux deux premiers niveaux les deux principales implications du graphe implicatif. Ainsi, la présence de la violation v7 (organisation des activités) semble impliquer celle de la 6ème (élection). Il en va de même entre les violations v8 (procédure d'enregistrement des syndicats) et v9 (droit à l'action syndicale). L'arbre nous apporte cependant une information supplémentaire puisqu'il met en évidence une méta-règle, la relation entre l'implication de la violation v7 vers la v6 qui implique la présence de la troisième violation. Ce résultat concorde avec l'analyse en cluster des pays (tableau 5), où l'on avait observé que tous les membres du cluster « restrictions organisationnelles et sur l'action syndicale » subissent simultanément les violations v6 et v7. Les pays membres de ce cluster subissent également la violation v3 (rejoindre et établir des syndicats) le plus souvent. Compte tenu de la relative proximité de v3 et v9, que l'on observe dans le graphe d'implication, mais aussi dans les figures 1 et 2, on aurait pu s'attendre éventuellement à voir v3 former une meta-règle avec v8→v9. Compte tenu de la structure hiérarchique, ceci n'est évidemment plus possible après que v3 ait été regroupé avec v7→v6. Ces contraintes de la présentation hiérarchique peuvent parfois masquer des enseignements intéressants. A titre d'exemple, mentionnons que si l'on enlève la variable v3 dans l'arbre de la figure 4, on voit apparaître à la place de la méta-règle (v7→v6)→v3, la méta-règle significative (v7→v6)→(v8→v9). Ce type d'effet masqué n'est évidemment pas propre à l'arbre cohésitif, mais concerne toutes les présentations arborescentes.

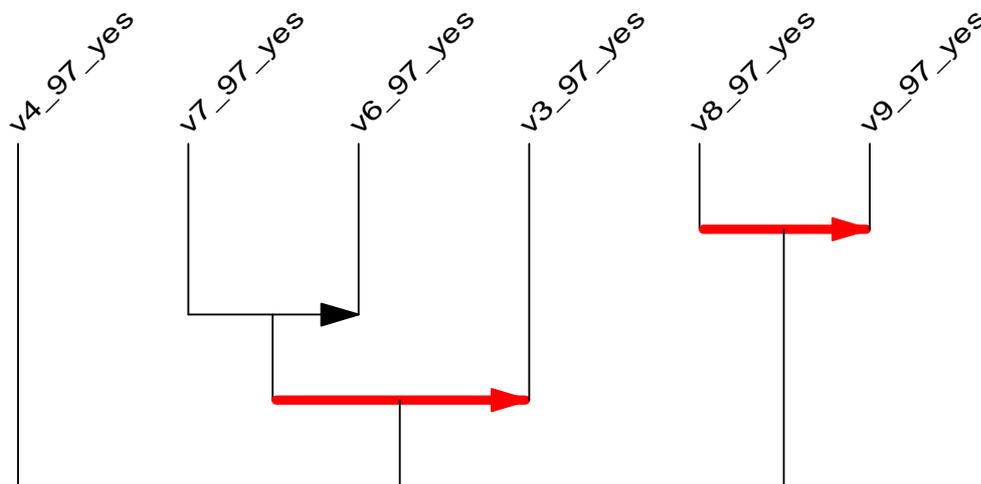


FIG. 4 – Arbre cohésitif des violations de la convention n° 87.

La figure 5 présente l'arbre des similarités des violations de la convention 87. Ici encore, les relations deux à deux sont conformes à ce qui a été mis en évidence auparavant. On note que la violation v3 s'associe ici à la conjonction des violations v8 et v9, et non pas comme dans l'arbre cohésitif avec v6 et v7.

À titre de comparaison, nous avons également réalisé avec SPSS une classification automatique des variables (c'est-à-dire des violations). Nous avons utilisé une procédure d'agglomération hiérarchique selon le critère de la distance moyenne, les proximités entre variables étant évaluées par la corrélation de Pearson. Le résultat est donné à la figure 6. Ici encore, les regroupements deux à deux sont les mêmes, la violation v3 s'associant ensuite avec v6 et v7 comme dans l'arbre cohésitif. Par rapport aux arbres hiérarchiques produits par CHIC, notons

qu'on peut ici évaluer la qualité des regroupements effectués selon la longueur des branches. Toutefois, il est plus difficile d'interpréter les associations entre groupes étant donné que ceux-ci sont des sortes d'agrégation de distances deux à deux entre variables.

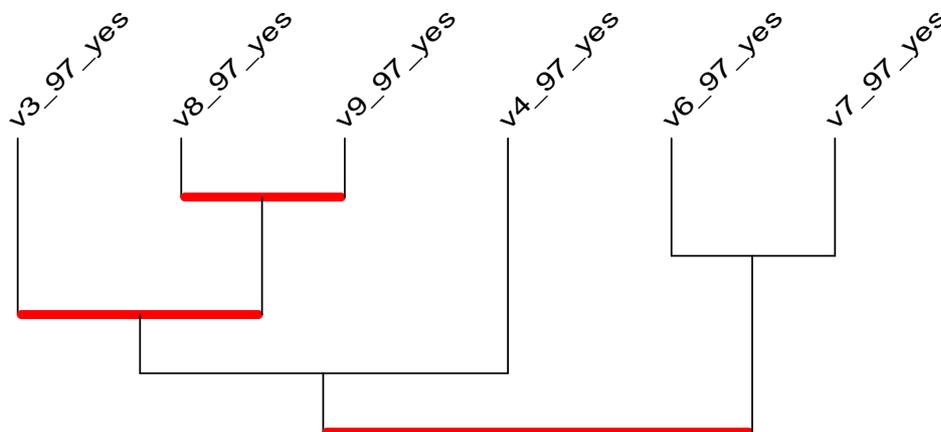


FIG. 5 – Arbre des similarités des violations de la convention n° 87.

Étonnement, le résultat obtenu à la figure 6 ressemble plus à l'arbre cohésitif (figure 4) construit sur la base d'associations asymétriques, qu'à l'arbre des similarités (figure 5). On peut ainsi se demander dans quelle mesure les regroupements obtenus résisteraient à de petites perturbations des données. Plus généralement, cela soulève la question de la stabilité et de la robustesse des solutions obtenues.

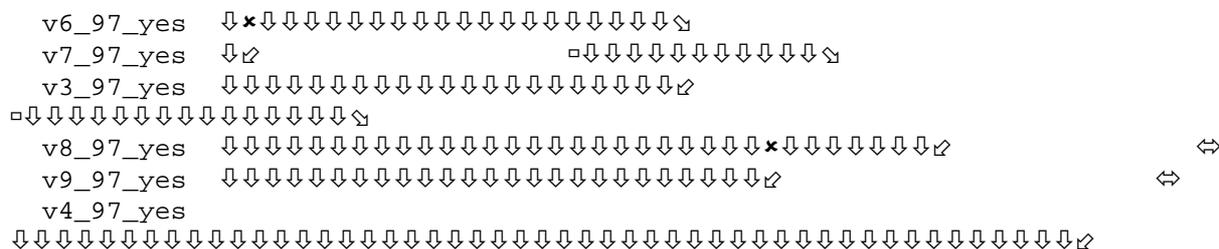


FIG. 6 – Dendrogramme de la classification automatique des violations de la convention n° 87 (agglomération hiérarchique selon le critère de la distance moyenne, distance fondée sur la corrélation de Pearson).

3.5 Violations de la convention n° 98

Nous avons effectué les mêmes analyses pour la convention n° 98 qui traite des relations entre syndicats et employeurs ainsi que des conditions-cadres nécessaires à la négociation collective. Toutefois, nous n'avons pu identifier aucune organisation claire des violations entre elles. Ainsi, les groupes issus de la classification automatique semblent principalement s'articuler autour de la présence d'un seul type de violation par groupe. De même dans l'analyse factorielle des correspondances multiples, on a à peu près une dimension par type de violation. Dans ce cadre, le graphe implicatif à l'avantage de montrer très clairement cette absence de relation entre les violations puisque aucune implication n'est affichée hormis à des seuils très bas.

4 Conclusion

Nous avons dans cet article tenté de dégager des logiques d'articulation entre types de violations du droit syndical garanti par les conventions no 87 et 98 de l'OIT. Nous nous sommes basés pour cela sur une source de données unique qui est le résultat d'une fouille automatique des rapports de la Commission des experts chargée d'observer l'application des conventions (CEACR).

Nos résultats montrent comment l'analyse de statistique implicative vient compléter avantageusement les analyses de données classiques, et en particulier la classification automatique des cas et l'analyse factorielle des correspondances multiples. En particulier, l'analyse implicative nous apporte une information précieuse sur la direction des liens. C'est grâce à elle que nous avons vu émerger la logique systémique que nous cherchions pour décrire d'articulation de nos violations. Ainsi, nous avons observé que les violations qui touchent l'organisation interne des syndicats semblent former un système. Nous avons également pu observer que les différents types de violations semblent impliquer une forme de restriction du droit à l'action syndicale.

Notre analyse nous a aussi amenés à relever deux aspects importants sur le plan méthodologique. La première est l'importance à accorder aux négatifs des variables considérées. Dans notre cas, les variables étant la présence des diverses violations, leurs négatifs sont les absences. Si a priori on peut s'attendre à observer simplement des implications inverses entre les négatifs, il apparaît cependant que leur prise en compte fait apparaître des liens avec d'autres variables, les régions dans notre exercice, non liées aux valeurs positives. Le second point concerne la stabilité des représentations obtenues. En effet, notre discussion, notamment des regroupements de violations, nous laisse penser que certains résultats pourraient être sensibles à de faibles perturbations des données. Il nous semblerait utile alors de pouvoir disposer d'un indicateur de la robustesse de la structure obtenue.

Références

- Anderberg, M. (1973). *Cluster Analysis for Application*. New York: Academic Press, 1973.
- Baccaro, L., J.-M. Bonvin, J.-P. Laviec, P. O'Donovan, G. Ritschard, and D. A. Zighed (2003). Social dialogue regimes: An investigation in the structural determinants and socioeconomic outcomes of negotiated regulation. Research proposal supported financially by the Geneva International Academic Network (GIAN).
- Couturier, R., A. Bodin, et R. Gras (2006). CHIC v3.7 Classification Hiérarchique Implicative et Cohésive. Guide d'utilisation, Ecole Polytechnique, Université, Nantes.
- Couturier, R. et R. Gras (2005). CHIC : traitement de données avec l'analyse implicative. In S. Pinson and N. Vincent (Eds.), *Extraction et Gestion des Connaissances (EGC 2005)*, Volume E-3 of *Revue des nouvelles technologies de l'information RNTI*, pp. 679–684. Cépaduès.
- Gras, R. (2005). Panorama du développement de l'ASI à partir de situations fondatrices. In R. Gras, F. Spagnolo, and J. David (Eds.), *Actes des Troisièmes Rencontres Internationale ASI Analyse Statistique Implicative*, Volume Secondo supplemento al N.15 of *Quaderni di Ricerca in Didattica*, pp. 9–33. Palermo : Università degli Studi di Palermo.
- Gras, R., S. Ag Almouloud, M. Bailleul, A. Laher, M. Polo, H. Ratsimba-Rajohn, et A. Totahasina (1996). *L'implication statistique : Nouvelle méthode exploratoire de données*. Recherches en didactique des mathématiques. Grenoble : La pensée sauvage.
- Greenacre, M. (1993). *Correspondence analysis in practice*. London: Academic Press, 1993.
- Kucera, D. (2004), Measuring trade union rights: a country-level indicator constructed from coding violations recorded in textual sources, Policy Integration Department, Statistical Development and Analysis Unit, International Labour Office, Working Paper n°50.
- Lebart, L., A. Morineau, et M. Piron (2000). *Statistique exploratoire multivariée* (Troisième ed.). Paris: Dunod, 2000.
- Ritschard, G., D. A. Zighed, L. Baccaro, I. Georgiou, V. Pisetta, and M. Studer (2007). Mining expert comments on the application of ILO conventions on freedom of association and collective bargaining. Internal report, University of Geneva.

Summary

The aim of this paper is to illustrate through an analysis of observed violations of ILO's Conventions, how implicative statistics advantageously complements more classical exploratory data analyses results. More specifically, we are interested in the study of the kind of violations reported by the Committee of experts on the application of ILO Conventions on freedom of association and collective bargaining. Data considered are predictions obtained using text mining based learning. We compare statistical implicative analysis with multiple correspondence analysis and case clustering.