

Analyse différentielle des discours de prévention du VIH : textes institutionnels et textes informels en français et en vietnamien

Océane Hô-Dinh¹, Mathieu Valette²

¹ INALCO – oceane.ho-dinh@inalco.fr

² INALCO – mvalette@inalco.fr

Abstract

Using the computer-based analysis of a French and a Vietnamese corpus, we offer an interpretative framework aiming to map out the linguistic differences of the way information is processed and knowledge organised or produced in relation to a similar topic-AIDS prevention. Our corpus is made up of official government documents and unofficial ones – e.g. forums – that share the same goal of prevention. We will examine where and how knowledge fails to be mentioned in official documents by analysing the relationships between stakeholders and concepts, stakeholders and time schemes.

Keywords: contrastive analysis, comparable corpora, corpus semantics, institutional discourse, social web

Résumé

À partir de l'analyse textométrique de deux corpus (l'un français, l'autre vietnamien), nous proposons une grille interprétative destinée à circonscrire d'un point de vue linguistique les différences de traitement de l'information et d'organisation ou de production des connaissances, pour une même thématique (la prévention médico-sanitaire du VIH), dans les textes institutionnels et dans les textes informels issus de forum de discussion aux objectifs similaires. Nous aborderons le repérage et la caractérisation de connaissances absentes des textes institutionnels, du point de vue des rapports entre acteurs et concepts et entre acteurs et temporalité.

Mots-clés : analyse contrastive, corpus comparables, sémantique de corpus, textométrie, discours institutionnel, web social

1. Introduction

Les connaissances sanctionnées qu'exposent les textes institutionnels sont aujourd'hui concurrencées et parfois disqualifiées par les textes informels produits sur le Web social. Certains forums de discussion, notamment, parce qu'ils sont régis par des procédures de modération exigeantes, ne sont pas les avatars modernes des conversations dites du Café du Commerce auxquels on voudrait parfois les réduire ; ils sont des lieux d'élaboration et de co-construction de savoirs experts fondés sur un partage d'expériences singulières qui pourrait relever d'une *sagesse des foules* (*Wisdom of Crowds*, cf. (James Surowiecki, 2008)). À titre d'exemple, les acteurs de l'informatique médicale admettent que les connaissances capitalisées par les populations concernées par une pathologie (patients, familles de patients), lesquelles se regroupent en communauté d'internautes, complètent l'expertise des médecins traitants (identification des maladies rares, effets secondaires, addictions, pratiques à risque, etc.). Dans cette perspective, notre projet est de questionner la construction, la sanction et l'élaboration des connaissances formelles et instituées au prisme des connaissances construites dans les textes informels.

A partir de l'analyse textométrique de deux corpus (l'un français, l'autre vietnamien), nous tâcherons de proposer une grille interprétative destinée à mieux circonscrire d'un point de vue linguistique les différences d'organisation ou de production des connaissances, pour une même thématique (la prévention médico-sanitaire du VIH), dans des textes institutionnels et dans des textes issus de forums de discussion aux objectifs similaires. Nous aborderons le repérage et la caractérisation de connaissances absentes des textes institutionnels, du point de vue des rapports entre acteurs et concepts d'une part, et entre acteurs et temporalité d'autre part.

1.1. Contexte applicatif de l'étude

Nous nous intéressons ici, dans une perspective contrastive, aux cas de la prévention à la contamination par le VIH en France et au Vietnam.

En France, les progrès et la qualité des soins permettent aux personnes porteuses du VIH de ne plus mourir du SIDA. L'effet pervers est un relâchement de la vigilance : l'épidémie continue de s'étendre au rythme de 6000 nouveaux cas détectés par an. La prévention demeure nécessaire notamment contre les pratiques les plus à risque, mais aussi en termes de détection de la séropositivité pour éviter les contaminations fortuites. Les acteurs de la santé produisent un discours de prévention et de dépistage continu.

Au Vietnam, le SIDA est considéré comme un fléau social. Le nombre de nouvelles infections par an est 50 fois plus élevé qu'en France pour une population 1,37 fois plus importante seulement. Très préoccupées, les autorités informent la population des dangers de l'épidémie, mais les tabous sont encore puissants. Le pays fait face à des tensions entre modernité et maintien des traditions. Héritée du confucianisme, la famille notamment occupe une place très importante et l'État maintient un contrôle de la société (censure des médias, verrouillage des discours, relayé par la famille et l'école). Internet et les technologies de l'information et de la communication se développent à grande vitesse, rapidement adoptées par la jeune génération qui y voit un moyen de libération de la parole, facilitée par un anonymat.

2. Hypothèse et méthode

Dans la lignée de (Slodzian et al., 2009), nous faisons l'hypothèse que les textes institutionnels médicaux sont stables quant à leurs prescriptions car fondés sur les pratiques normalisées et consensuelles d'une médecine scientifique internationale. Les textes de prévention sont, de fait, homogènes et peu productifs car ils sont adossés à des connaissances médicales dont les progrès sont lents. Cependant, la réappropriation des discours institutionnels sur les forums de discussion, implique hétérogénéité et productivité. Ainsi, les textes informels complètent les informations dispensées par les textes institutionnels en nous renseignant sur la variété des pratiques, peu couvertes par les textes institutionnels. Nous montrerons ainsi en quoi les textes informels apportent une vision du domaine différente et complémentaire des textes institutionnels, qui reflète plus précisément les préoccupations des internautes.

2.1. Corpus et prétraitement

Les deux sous-corpus institutionnels (français et vietnamien) sont constitués d'articles et de foires aux questions (FAQ) tirés de sites officiels d'information sur le VIH/SIDA et les deux sous-corpus informels (français et vietnamien) sont constitués de messages postés par des internautes sur des forums de discussion.

Le corpus institutionnel français est constitué de 830 textes (560 000 mots) extraits de Sida Info Services¹. Le corpus informel français est constitué de 2 222 messages (700 000 mots) postés sur différents forums (doctissimo.fr, aufeminin.com, ados.com, etc.) sélectionnés sur les thèmes du VIH, des maladies et infections sexuellement transmissibles (MST/IST), de la séropositivité, des rapports sexuels à risque, etc. Le corpus institutionnel vietnamien est constitué de 70 textes (52 000 mots), extraits du site HIV Online. Ce site issu d'une initiative gouvernementale donne accès à des informations sur la prévention du VIH. Il résulte d'un programme lancé en 2004, le programme de « Lutte en ligne contre le VIH/SIDA ». Fin 2004, le site a été étendu aux questions de la jeunesse et la sexualité. Le corpus informel vietnamien est constitué de 560 messages (56 000 mots) postés sur le forum du même site (forum.hiv.com.vn).

Le corpus vietnamien a été segmenté à l'aide de l'outil *vnTokenizer* (Lê Hồng, 2008) et le corpus français, a été traité au moyen de l'analyseur morphosyntaxique *TreeTagger* (Schmid, 1995) inclus dans TXM (Heiden et al., 2010), le logiciel de textométrie utilisé pour nos analyses. Le corpus français fait donc l'objet d'un prétraitement plus approfondi (segmentation, étiquetage morphosyntaxique) de sorte que nous disposons, pour ce corpus d'une annotation en lemmes et en parties du discours. L'opération de lemmatisation en vietnamien est sans objet puisqu'il n'y a pas de morphologie en vietnamien (les formes sont invariables). Comme pour toutes les langues d'Asie du Sud-Est, la caractérisation en parties du discours du vietnamien est une tâche complexe car très contrainte par le contexte d'apparition de la forme et aucun outil ne permet d'en obtenir une représentation satisfaisante. Nos analyses ne seront donc pas menées de manière strictement similaire sur nos sous-corpus français et vietnamien.

2.2. Méthodologie

Notre méthodologie repose sur l'interprétation aux moyens d'outils théoriques de résultats d'analyses statistiques différentielles. Pour celles-ci, nous recourons aux algorithmes de spécificités de (Lafon, 1980) et de cooccurrences (calcul de spécificités sur les contextes) implémentés dans le logiciel de textométrie TXM (Heiden et al., 2010). Notre démarche consiste à :

- (1) analyser les spécificités générales des sous-corpus afin de déterminer les grandes tendances en termes de vocabulaire,
- (2) identifier un certain nombre d'unités lexicales caractéristiques de la problématique générale (et non d'un sous-corpus donné), en étudier les cooccurrences statistiquement significatifs,
- (3) sélectionner les plus pertinents d'entre eux pour les qualifier en termes de corrélats, c'est-à-dire éléments d'un thème sémantique (Rastier, 2001 : 211-213).

L'enjeu sera de repérer et de caractériser les connaissances absentes des textes institutionnels, du point de vue des rapports entre acteurs et concepts et entre acteurs et temporalité.

¹ <http://www.sida-info-service.org>

3. Analyses des corpus

3.1. Distanciation et intimité : des univers de référence distincts

Les spécificités macroscopiques mettent en évidence une opposition majeure entre (i) ce qui relève du conceptuel, du monde des références, en un mot, de l'*ontologie* propre au discours médical, dans les corpus institutionnels, (ii) un univers de pratiques, d'expériences et de témoignages, qui relèveraient d'une *praxis* de la maladie ou du risque pathologique, dans les corpus informels (cf. figure 1).

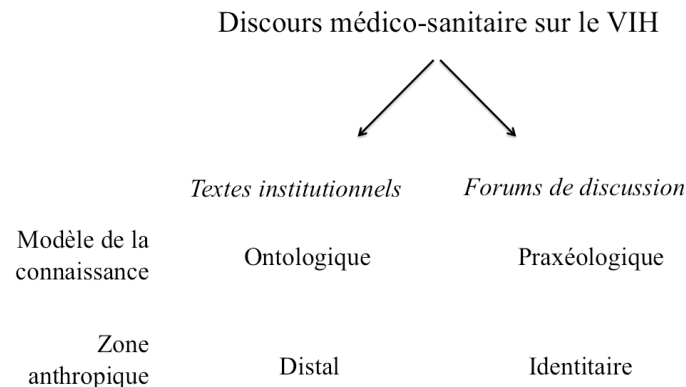


Figure 1. Articulations conceptuelle et anthropique des discours de prévention médico-sanitaire

Cette opposition fondamentale entre textes institutionnels et textes informels issus du web social a déjà été mise en évidence par (Slodzian et al., 2009) à propos du discours de prévention du tabagisme et semble ici se confirmer : le discours institutionnel est caractérisé par des opérations de catégorisation, d'abstraction et de généralisation, tandis que le discours informel se caractérise, à l'inverse, par la description de pratiques concrètes et variées sur un mode narratif et particularisant. Par exemple, dans le corpus vietnamien, les textes institutionnels désignent de grandes catégories de population, comme *người bệnh* (« les malades ») ou *bệnh nhân* (« les patients »), tandis que les textes informels privilégient l'expression de la première et de la deuxième personne : *em* (« petit·e frère/sœur », qui, dans l'usage vietnamien, signifie « moi » ou « je », avec une connotation déférente), *mình* (« moi », « je » dans le langage courant), *anh chị* (« grand·e·s frères/sœurs », i.e. « vous ») (cf. figure 2).

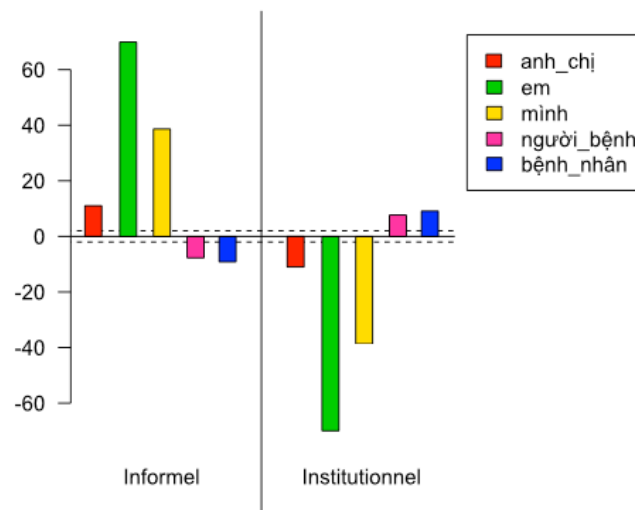


Figure 2. Les acteurs dans le corpus vietnamien

3.1.1. Zones anthropiques distales, proximales, identitaires

On observe par ailleurs que les acteurs des textes institutionnels évoluent dans une *zone anthropique distale*², c'est-à-dire qu'ils construisent une distance entre l'énonciateur et l'objet, il est question d'aide aux malades, de populations à risques, de prévention contre de nouvelles contaminations, mais ni de l'*ethos* du malade ni des conséquences psychologiques de l'exposition, de la peur de l'exposition, ou de la contamination. À l'inverse, les acteurs des textes informels évoluent dans une zone anthropique identitaire (*j'ai pris un risque, je flippe*) et proximale (*aidez-moi, vous me conseillez quoi ?*). L'analyse des variables morpho-syntaxiques, dans le corpus français, corrobore cette opposition anthropique : la deuxième personne du pluriel (vouvoiement) et la troisième personne du pluriel sont caractéristiques des textes institutionnels, tandis que les première et deuxième personnes du singulier sont propres aux textes informels (cf. figure 3).

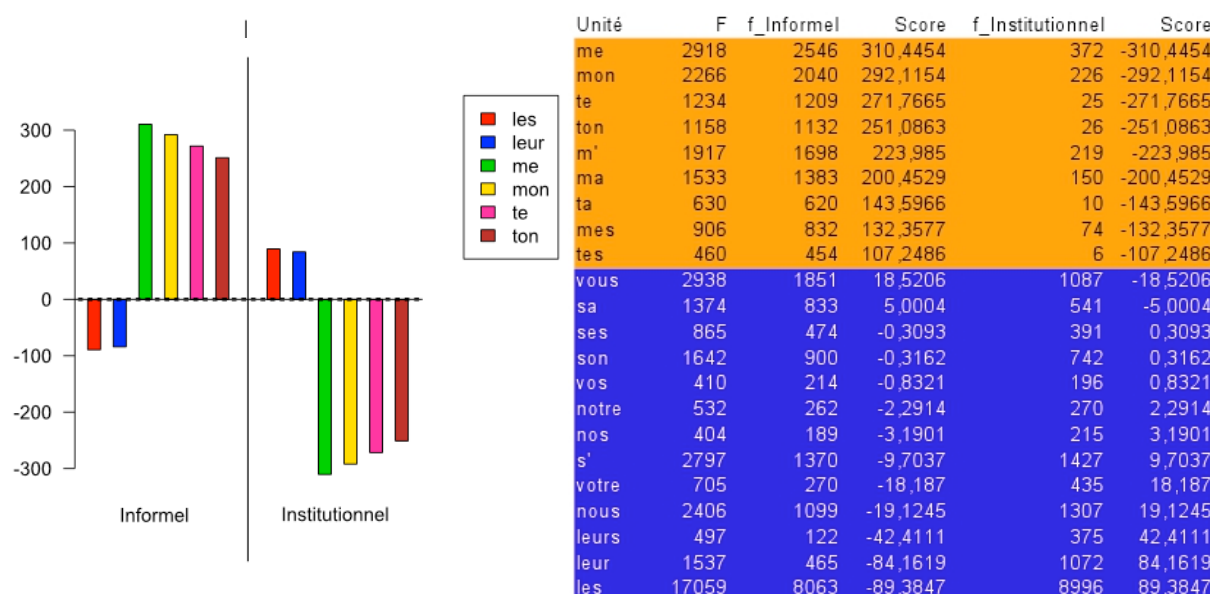


Figure 3. Les pronoms personnels et les adjectifs possessifs dans le corpus français

L'opposition marquée entre un univers distal, abstrait et désincarné et un univers intime, concret et lié à des pratiques, permet d'identifier des thèmes sémantiques contrastifs. On prendra ci-dessous l'exemple de deux d'entre eux : les effets secondaires et la sérologie.

Les cooccurents des « effets secondaires » *tác dụng phụ*, dans le corpus institutionnel où le concept est peu abordé sont *chóng mặt* « vertiges » ou *thuốc mới* « nouveaux médicaments ». Au contraire les effets secondaires sont très présents dans les textes informels et montrent l'importance des perceptions intimes (identitaires). Les cooccurents expriment là encore des procès : *mọc* « apparaître », *gây* « provoquer », *thấy* « ressentir », « se sentir » et des éléments à valeur thymique *ghê gớm* (« horrible »).

La sérologie, dans le corpus institutionnel, est exprimée de façon très distanciée, montrant là encore la fonction objectivante des genres privilégiés dans les discours institutionnels ; par exemple : *tải lượng vi rút* « charge virale », *lây* « contaminer », etc. Dans le corpus informel l'ancrage énonciatif est fondamentalement différent et indique que la sérologie est quelque chose de vécu, d'ancré dans la réalité : on a (i) des marqueurs de spatialité : *cơ sở* « centre », *bệnh viện* « hôpital », *Tràng An* (un grand hôpital privé au Vietnam), *đi* « aller », *tại, chỗ*

² Nous empruntons à (Rastier, 2001) le concept de zone anthropique.

(locatifs), etc. ; et (ii) des marqueurs de temporalité : *6 tuần* « 6 semaines », *3 tháng* « 3 mois », *sau* « après », *giai đoạn* « période », etc.

3.1.2. *Ontologie et praxéologie*

L'examen des spécificités macroscopiques du corpus français confirme les observations faites sur le corpus vietnamien, non seulement au niveau des formes graphiques, mais aussi au niveau des catégories syntaxiques. On constate ainsi une sous-représentation notable des verbes dans le corpus institutionnel (67 000 occurrences contre 113 000 dans le corpus informel). Cette observation corrobore vraisemblablement l'opposition entre l'*ontos*, dont la catégorie grammaticale privilégiée est le substantif, et la *praxis*, souvent caractérisés par les verbes exprimant un procès. Dans le corpus français informel, on relève de nombreux verbes d'action spécifiques au domaine (*passer, protéger, immuniser, choper*, etc.).

Dans le corpus vietnamien informel, on ne peut guère interpréter l'usage qui est fait des verbes en raison de la morphologie isolante du vietnamien. En revanche, les verbes d'action propres au domaine que nous relevons en abondance sont, eux, très spécifiques au corpus informel (*rác* « déchirer », *xét nghiệm* « faire un examen (sérologie) », *sản sinh* « produire » (des anticorps), etc.). Nous notons également une forte présence des verbes à valeur thymique : *yên tâm* « se tranquilliser », *lo, lo lắng* « s'inquiéter », *mong* « espérer », etc.

Dans les textes institutionnels français, non seulement les verbes sont moins représentés, mais ils ne font pas partie des mêmes classes sémantiques, expression d'un discours volontariste ou interventionniste : *permettre, proposer, faciliter*, etc. ; *réaliser, agir, travailler, constituer, créer*, etc. ; *engager, consacrer, mobiliser, financer*, etc. ; *lutter, intervenir*, etc. ; *améliorer, renforcer, évoluer, augmenter, assurer*, etc. Ces verbes sont propres, selon nous, aux genres textuels « interventionnistes ». En revanche les verbes spécifiques au domaine sont rares dans le corpus (*entraîner* (des conséquences, des effets secondaires), *traiter* (une maladie)). Tout se passe comme si nous nous situions au plus haut niveau d'abstraction possible. Les mêmes observations peuvent être réalisées à propos du corpus vietnamien institutionnel : *điều trị* « traiter », *chăm sóc* « surveiller », *tránh* « éviter », *hỗ trợ* « aider », *phòng* « lutter », *tiếp cận* « avoir accès », etc.

3.2. *Tabous, déviance et écart à la norme dans le corpus vietnamien*

L'analyse thématique du corpus vietnamien met en évidence un hiatus marqué au niveau des événements à risque abordés dans les forums de discussions et dans les textes institutionnels, alors que dans le corpus français, ces différences sont nettement moins marquées. Tout ce passe comme s'il existait, dans le contexte vietnamien, une relation de complétion entre les différents genres textuels : dans les textes institutionnels, les risques de transmission sont la sexualité en général (*tình dục*), le partage de seringues (*dùng chung bơm kim tiêm*), la transmission de la mère à l'enfant (*mẹ truyền qua con*), ou la transfusion sanguine (*bị truyền máu*), tandis que dans les forums de discussion, les risques de transmissions évoqués sont les rapports sexuels (*quan hệ, QHTD* — à opposer à la sexualité en générale, cf. *supra*), les prostituées (*cô, GDM, cô gái*), et les accidents de protection, principalement le préservatif (*bao cao su, BCS*) qui sera déchiré, périmé, absent, etc.

Par exemple, en contexte institutionnel, les cooccurrents de *bao cao su* « préservatif » le décrivent suivant la norme implicite relative à son utilisation : fonction « protection » (*an toàn*), accessibilité « distributeur automatique » (*máy bán bao cao su tự động*), usage (*sử dụng bao cao su đúng cách* « littéralement « utiliser correctement le préservatif »). En

contexte informel, sont exprimées des situations s'écartant de cette norme : surreprésentation des verbes signifiant *se déchirer, utiliser, retirer, être en contact, périmé*, etc.

L'écart à la norme est également révélé par l'utilisation fréquente d'euphémisations, dans les textes vietnamiens issus des forums de discussion, pour aborder des thématiques socialement taboues et donc souvent absentes des textes institutionnels. La lecture humaine du corpus permet, par exemple, d'identifier les euphémismes suivants :

- rapport sexuel

<i>đời sống tình *ục</i>	« vie ***uelle »
<i>ngủ</i>	« dormir »
<i>có lỗi "trót đại"</i>	« commettre la faute »
<i>nếm trái cấm</i>	« goûter le fruit défendu »
- prostitution

<i>đi "bóc bánh"</i>	« sortir du chemin »
<i>đi ăn bánh rán trả tiền</i>	« aller manger un beignet »
- homosexualité

<i>tgt3 (thế giới thứ ba)</i>	« troisième monde »
<i>cộng cao đẳng</i>	« communauté des marins »
- masturbation

<i>TD (thủ dâm)</i>	« masturbation »
---------------------	------------------

D'une manière générale, l'informel fait la part belle à la métaphore et aux expressions imagées comme l'attestent les exemples ci-dessous :

- préservatif

<i>bao cao su</i>	« sac en caoutchouc »
<i>bao</i>	« sac »
<i>dùng "áo mưa"</i>	« utiliser un vêtement de pluie »
- rapport sexuel

<i>khi đang hành</i>	« pendant qu'on voyage »
----------------------	--------------------------
- pénis

<i>cái của quý</i>	« la précieuse possession »
<i>"cậu nhỏ"</i>	« petit père »
<i>"sung"</i>	« arme à feu »
<i>chim</i>	« oiseau »
- vagin

<i>bướm</i>	« papillon »
<i>"cô bé"</i>	« petite dame »

Plus généralement dans la langue vietnamienne, le processus d'abréviation des expressions fréquemment utilisées dans un domaine est systématiquement adopté. On aura notamment recours à l'acronymie pour remplacer des expressions figées, comme par exemple :

- *BCS* pour *bao cao su* (« préservatif »)
- *NCH* pour *người có HIV* (« personne atteinte du VIH », « séropositif »)

Et même :

- *H* pour *HIV* (« VIH »).

Mais le recours plus singulier à des procédés langagiers imagés et détournés, très caractéristique du corpus informel vietnamien, vise, selon nous, à effectuer une mise à distance purement lexicale (métaphorisation, euphémisation), au sein même de la zone

anthropique proximale – distance qui permet de contourner le caractère tabou d'un sujet de discussion, alors que ce caractère relève précisément de sa dimension identitaire ou proximale. En d'autres termes, la distance produite par ces procédés permet finalement de s'affranchir des barrières séparant les sujets tabous de la verbalisation.

3.3. La perception de la temporalité

Les rapports entre les acteurs et la temporalité enrichissent notre grille d'analyse des discours institutionnels à la lumière des textes informels. D'un point de vue générique, la surreprésentation de l'imparfait dans le corpus informel français confirme que les genres narratifs y sont privilégiés : l'imparfait est le temps verbal du récit, du témoignage, du retour d'expérience, de la description des actions successives du locuteur, qui les raconte ensuite. « *J'étais un peu ivre* », « *toi ou ton mari était porteur avant d'être en couple* », « *La question me turlupinait* », « *On m'annonçait que j'étais séropositive* », etc. S'ajoutent à ces critères verbaux divers marqueurs substantivaux (*hier soir, ce matin*, etc., et dans le corpus informel vietnamien, *rồi* (marque du passé), *ngày sau* « le lendemain », etc.). À l'inverse, le participe présent est caractéristique du corpus institutionnel : « *limitant ainsi le risque de transmission* », « *la sexualité des femmes vivant avec le virus* », « *les hommes ayant des relations sexuelles avec des hommes* », etc.³ En bref, la temporalité ne semble pas exprimée dans les textes institutionnels.

A la vérité, dans les corpus français comme vietnamien, les textes institutionnels substituent à la temporalité deux états atemporels auxquels correspondent deux publics cibles distincts : les personnes non contaminées d'une part et les personnes séropositives d'autre part. Les deux états correspondent donc à la période avant contamination, sur laquelle porte le discours de prévention, et à la période après contamination, auquel correspond un discours médical, établissant une description de la pathologie. A titre d'exemple, en vietnamien,

- l'état préventif (discours de prévention) actualise des syntagmes tels que *Ai sẽ là người nhiễm HIV?* « qui peut attraper le VIH ? », *Con đường lây nhiễm* « les voies de contamination », *Sử dụng bao cao su đúng cách* « utilisation correcte du préservatif », *hành vi nguy cơ* « pratique à risque » ;
- l'état pathologique (discours médical relatif à la pathologie) actualise des syntagmes tels que type *Đã nhiễm HIV* « contaminé par le VIH, *ăn uống tốt* « manger sainement », *tập thể dục* « faire du sport », *điều trị ARV* « traitements antirétroviraux », *Người bệnh là nạn nhân* « les malades sont des victimes ».

Cette observation cruciale est rendue possible par l'examen différentiel du corpus informel : la temporalité n'y est pas perçue de la même manière. En effet, entre ces deux états, il existe une période charnière, *entre l'instant précis de la prise de risque, et celui aussi précis du résultat de la sérologie*. Cette période d'*entre-deux*, superbement ignorée des discours institutionnels, cristallise toutes les inquiétudes et interrogations des personnes concernées. C'est cette période que couvrent et investissent dans une large mesure les forums de discussion (cf. figure 4).

Lui correspond un foisonnement d'échanges saturés de marqueurs dysphoriques. Si on analyse plus en détail cet *entre-deux*, on y décèle un enchaînement d'intervalles temporels très précis. Par exemple, l'instant de la prise de risque met un terme à l'état préventif, mais

³ Il est à noter que le patron syntaxique [NOM + VER:pper] (substantif suivi d'un verbe au participe présent) collecte 936 occurrences dans le corpus institutionnel français, aucune dans le corpus informel.

n'entraîne pas pour autant un passage à l'état pathologique. Le temps se distend fortement à ce moment-là et les marqueurs de temporalité abondent. Ils ressortissent à deux temps :

- le temps réel (ou chronologique) : il se décompose objectivement en phases médicales (d'une part le traitement post-exposition à prendre dans les 48 heures après l'exposition ; puis les semaines d'incubation où la contamination n'est pas encore détectable, avant la possibilité de procéder à un examen sanguin qui indiquera l'état sérologique et qui déterminera ainsi le passage vers l'état pathologique ou le retour à l'état préventif.
- le temps vécu, où la mesure prend une grande importance. L'infection n'étant pas détectable pendant les premières semaines (six à douze en général), les sujets sont confrontés à une période d'attente incompressible mais d'une durée variable d'une situation à l'autre, les plongeant dans un état d'impuissance et d'incertitude. Cet état est marqué par les signes de la dysphorie d'une part, mais aussi une abondance de commentaires-exutoires. Celle-ci sera observable dans le corpus par la forte présence d'un vocabulaire de la mesure (*semaine, jours, après, đủ ba tháng* « trois mois entiers », *bao giờ* « à quel moment ? »), etc.), et des verbes porteurs de procès et d'itération (*refaire, commencer, retourner*, etc.).

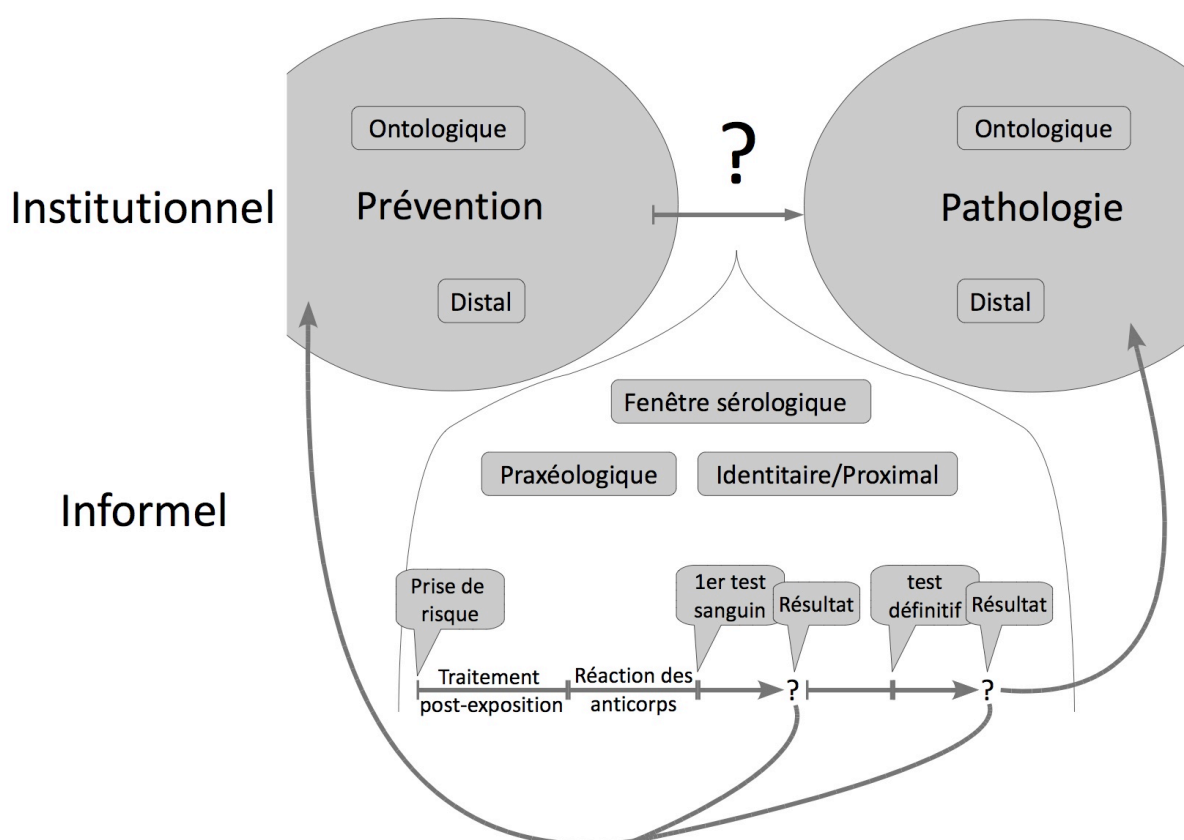


Figure 4. Les temporalités dans le discours sur le VIH

Les histogrammes des figures 5 et 6 montrent la distribution de quelques unités lexicales désignant la temporalité dans les corpus informels et institutionnels.

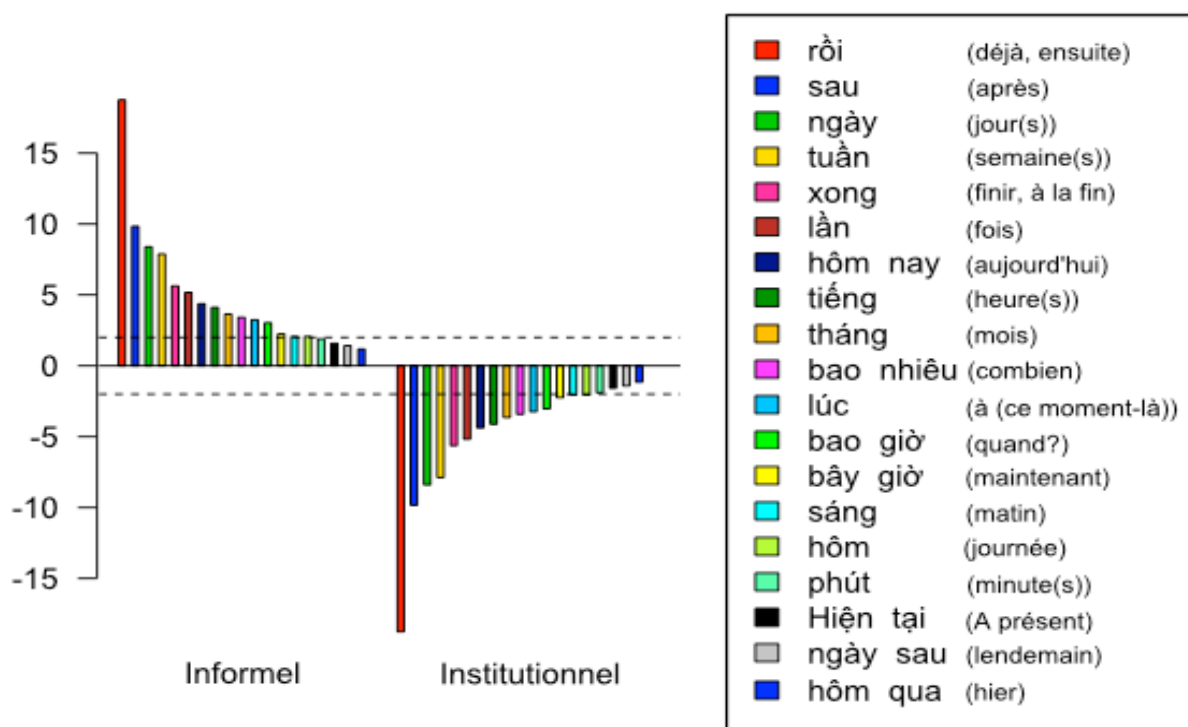


Figure 5. La temporalité dans le corpus vietnamien

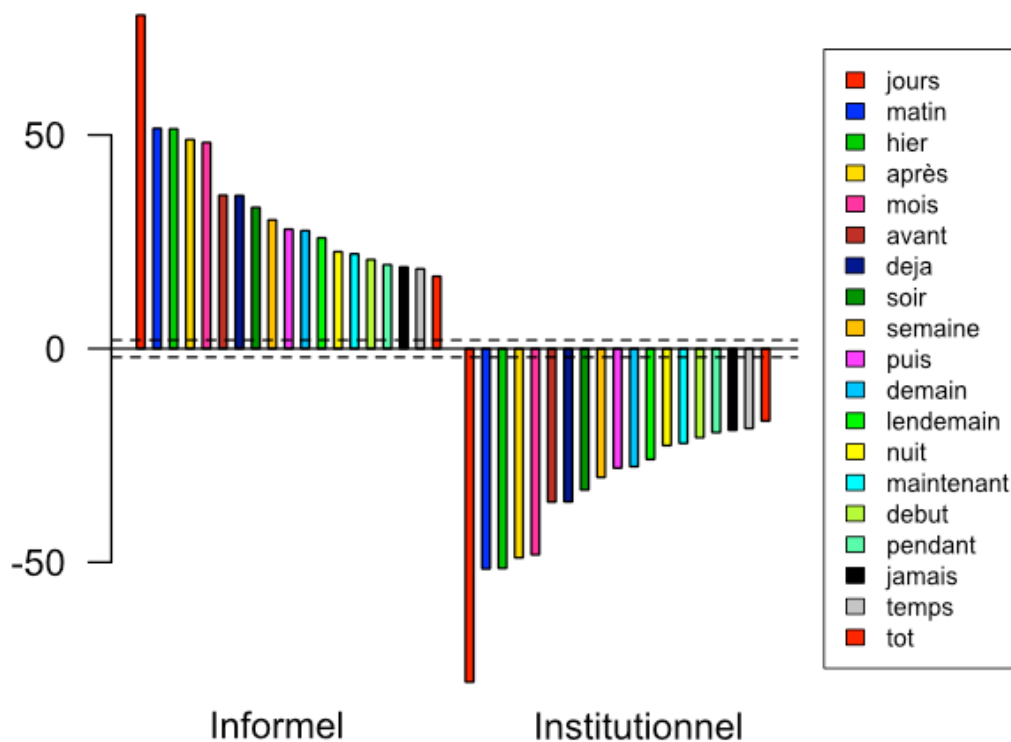


Figure 6. La temporalité dans le corpus français

4. Conclusion

Les discussions sur les forums occupent un espace informationnel laissé vacant par le discours institutionnel, qui, dans le domaine médical de la prévention, par exemple, ne traite pas du temps vécu, de l'attente et ne prend pas en considération les inquiétudes du public. En d'autres termes, le discours institutionnel apparaît lacunaire.

En analysant de manière contrastive les textes institutionnels et les textes informels, nous entendons discerner ce que ne couvre pas le discours institutionnel, ce qui lui échappe et en quoi il apparaît insuffisant et lacunaire pour répondre aux attentes du public et, à terme, étudier comment il est interprété, récupéré, réapproprié par le public auquel il s'adresse.

Dans les textes institutionnels, les connaissances existent en amont des productions textuelles, elles relèvent de savoirs scientifiques, médicaux, construits dans d'autres situations énonciatives, d'autres pratiques sociales (par exemple, dans des articles scientifiques). Les textes institutionnels exposent des univers conceptuels peu ou pas temporalisés, détachés des pratiques et des conditions d'énonciation initiales leur correspondant. A ces univers désincarnés, abstraits, les forums de discussion apportent des compléments aujourd'hui considérés avec attention, tant par les institutionnels que par le public : comportements, préoccupations. Il est probable que la mesure de la pertinence et de la confiance attribuable aux textes informels para-institutionnels deviendra rapidement un enjeu de science.

Références

- Heiden S., Magué J.-P. et Pincemin B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In I. C. Sergio Bolasco, editor, *JADT 2010*, vol.(2), pp. 1021-1032. Logiciel disponible sur : <http://textometrie.ens-lyon.fr/>.
- Lafon P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots* (n°1): 127-165.
- Lê Hồng P., Nguyễn T. M. H., Roussanaly A. et Hồ T. V. (2008). A hybrid approach to word segmentation of Vietnamese texts. *LATA 2008*.
- Rastier F. (2001). *Arts et sciences du texte*, Paris, PUF.
- Schmid, H., 1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proc. of the ACL SIGDAT-Workshop*. Dublin, Ireland.
- Slodzian M. et Valette M. (2009). Connaissances prescrites ou connaissances décrites ? L'apport de la sémantique des textes. In Khaldoun Zreik, dir., *Patrimoine 3.0 (Actes du 12e Colloque International sur le Document Électronique, Montréal (CIDE.12))*, Europia Productions, Paris, pp. 129-141.
- Surowiecki J. (2008). *La Sagesse des foules* (traduction de *The Wisdom of Crowds*, 2004), Éditions Jean-Claude Lattès, Paris.

