

Utilisation et évaluation du thésaurus belge 3BT adapté au Rwanda

Use and evaluation of the Belgian 3BT thesaurus adapted for Rwanda

M. De Jonghe¹, M. Jamouille² et D. Porignon³

¹D.M.G.-U.L.B., ²Médecine Générale, ³Cemubac, Ecole de Santé Publique U.L.B.

RESUME

Le Centre Hospitalier Universitaire de Kigali (CHUK) a besoin, depuis de nombreuses années, de donner une visibilité à son activité. La gestion de l'institution, la compréhension de la place de l'institution dans le système de soins de santé rwandais, le renforcement de la visibilité de l'institution sont autant d'objectifs de politique de santé. Cependant, aucun outil n'avait été mis en place à ce jour.

L'objectif est d'évaluer la faisabilité de l'utilisation du thésaurus 3BT comme instrument de collecte de données.

En août 2005, le thésaurus belge 3BT, bilingue, biclassifié (ICD-10/ICPC-2) adapté au CHUK a été mis en place. Critères de jugements primaires : qualité des encodages, opérationnalité du thésaurus.

Une analyse qualitative a été réalisée sur 899 encodages, dont 16 vides. Les occurrences faibles ($\leq 0,2\%$) de près de 25 % des encodages des diagnostics cliniques justifient l'utilisation et l'amélioration du thésaurus. Près de 45 % des encodages vides peuvent être complétés suite à la simple consultation des dossiers. Certains diagnostics sont manquants dans le thésaurus 3BT. 66 % de ces diagnostics ont des concepts similaires dans le thésaurus, mais non identifiables par un non-médecin. Enfin, un test d'encodage des données cliniques en double aveugle par un médecin habitué aux classifications et par un non-médecin habitué à l'encodage des diagnostics cliniques de dossiers de patients hospitalisés montre une parfaite similitude dans 70 % des encodages et une perte de précision dans près de 20 %. Aucune faute d'encodage n'a été relevée, même si elles ne peuvent être exclues.

En conclusion l'opérationnalité du thésaurus 3BT adapté au Rwanda est acceptable dans cette étude. Il facilite l'encodage normalisé de diagnostics cliniques, même par de non-médecins. La qualité des données enregistrées est suffisante pour pouvoir interpréter les résultats quantitatifs des encodages. Cette étude doit être répétée sur un plus grand nombre de données.

Rev Med Brux 2006 ; 27 : S 274-8

ABSTRACT

The Kigali university medical centre (CHUK) lacks visibility on its activities for many years. Institution management as well understanding the institution as the corner stone of the Rwandese health system and reinforcement of its visibility are all health policy major issues. Nevertheless, to this point no such tool has been developed nor implemented.

The objective is to assess the feasibility of the use of the thesaurus 3BT as data collection tool in a Rwandese health institution.

In august 2005, the thesaurus 3BT (Belgian, bilingual, biclassified ICD-10/ICPC-2) adapted to the CHUK has been implemented. Main issues : encoding quality and thesaurus operationality.

Qualitative analysis has been performed on 899 coding of which 16 empty. Low occurrences ($\leq 0.2\%$) of about 25 % of codings of clinical diagnosis show the need for using a thesaurus but also to upgrade it. Near 45 % of the empty codings could be fulfilled with a quick look to the original medical record. Some diagnoses are missing in the thesaurus 3BT. 66 % of which have similar concepts in the thesaurus although not identifiable by a lay person.

Finally, a clinical data blind coding test by a doctor used to classifications and by a lay person used to code clinical diagnosis from medical records of the hospitalised patients shows an exact similarity in 70 % of the coding and a loss of coding precision in 20 %. No coding error has been identified at this time.

In conclusion, operationality of the thesaurus is quite acceptable in this study. The thesaurus makes easy the coding of clinical diagnosis even by lay people. Quality of data is enough to be able to interpret the quantitative results of the coding process. This study has to be repeated on a wider basis.

Rev Med Brux 2006 ; 27 : S 274-8

Key words : thesaurus 3BT, ICPC-2, ICD-10, operationality of the thesaurus 3BT

ASPECTS METHODOLOGIQUES

Contexte

Au sein du Centre Hospitalier Universitaire de Kigali (CHUK), la problématique liée à la classification des pathologies a été identifiée depuis de nombreuses années. Elle n'a jamais fait l'objet d'une réelle démarche pour essayer de mettre en place un système répondant, dans la mesure du possible, à la fois aux exigences en matière de gestion et aux contraintes liées aux professionnels de santé.

Une expertise rwandaise de l'Ecole de Santé Publique de Butare a permis, sur base de la revue des dossiers et des informations disponibles au service des statistiques de l'hôpital, de préciser quels sont les diagnostics les plus fréquemment rencontrés au CHUK. Sur cette base, une adaptation de la Classification internationale des maladies (ICD-10) a été réalisée. L'élagage de la liste des pathologies a permis de réduire 12.000 entrées à quelque 3.500 entrées. Ce travail a été finalisé au début du mois de juillet 2005¹.

En août 2005, une mission d'observation a été effectuée dans le but de suivre la genèse de l'information médicale au sein de différents services du CHUK et le parcours de cette information jusqu'à l'établissement du rapport d'activité du CHUK. Un biais d'interprétation a été relevé puisque toute la richesse du vocabulaire professionnel était réduite dans les faits à 220 items reliés à un code propre au CHUK, mais sans relation aucune avec une classification de maladies quelle qu'elle soit.

Méthode

Dans ce contexte, il a été proposé d'introduire une terminologie normalisée étendue dénommée ici thésaurus.

Un thésaurus est un système terminologique incluant des synonymes. Le thésaurus belge 3BT est un réseau sémantique médical donnant accès à des libellés cliniques, diagnostics, problèmes de santé proches du vocabulaire utilisé quotidiennement par les médecins. La sélection d'un libellé induit l'encodage des codes ICPC et ICD ainsi que sa traduction dans d'autres langues.

L'intérêt d'un tel thésaurus est d'apporter un vocabulaire professionnel normalisé, contrôlé, le plus exhaustif possible, mais toujours dynamique. Le thésaurus belge, dénommé 3BT pour *Bilingual Biclassified Belgian Terminology* est particulier dans sa conception : il repose sur un identificateur conceptuel unique par ligne proposée dénommé IBUI*, ce qui permet de lier à chaque diagnostic un code ICPC-2, un code ICD-10 et ou d'assurer plusieurs traductions.

Le thésaurus doit permettre de traduire l'ensemble des diagnostics relatifs aux pathologies rencontrées telles qu'exprimées par les professionnels

de santé dans les dossiers " papier " des malades et de les mettre en relation avec les classifications CIM-10 (ICD-10)² et CISP-2 (ICPC-2)³. La version implémentée est ICPC-2-R⁴ qui correspond quasiment à la version française de la CISP-2. ICD et ICPC ont été reconnues par l'O.M.S. comme complémentaires et font partie toutes deux de la famille des classifications de l'O.M.S. (WHO-FIC). Le thésaurus 3BT a été développé et est mis à jour en Belgique par le Département de Médecine générale de l'U.L.B. et par le Département de Médecine Générale de l'Université de Gent sous la direction du Ministère Fédéral belge de la Santé Publique⁵. Il a été proposé de l'adapter au contexte rwandais afin de pouvoir l'utiliser au CHUK¹. Il est disponible pour les chercheurs sous forme d'un utilitaire développé par le CISP-Club" (Figure).

Interprétation des résultats

L'exploitation des encodages des diagnostics cliniques dépend :

- D'une bonne appréciation du problème de santé.
- D'une transcription correcte du ou des diagnostic(s) dans les dossiers.
- D'un encodage rigoureux, qui peut être évalué qualitativement.
- D'un encodage précis dans des champs bien définis : la qualité du logiciel informatique peut être évaluée.
- D'une analyse statistique rigoureuse et précise en fonction de questions clairement posées.
- D'une restitution à qui de droit en prenant le soin de montrer les limites des résultats.

Justification de l'utilisation de classifications internationales des pathologies au CHUK

L'analyse des données telles que présentées dans les rapports annuels du CHUK ne permet que très difficilement d'avoir une vue d'ensemble des activités de l'hôpital, si ce n'est à travers un périlleux exercice intellectuel de compilations des données partielles reprises dans chacun des chapitres de ce rapport et qui correspondent à la synthèse inégale des activités de chaque entité de l'hôpital".

L'ensemble de ces données, même si leur intérêt est évident pour certains responsables, n'en demeure pas moins très difficilement interprétable en termes de performances réelles de l'institution que ce soit d'un point de vue quantitatif épidémiologique et financier ou encore d'un point de vue qualitatif.

Les informations produites par le Service des Statistiques sont assez exhaustives quant à leur collecte systématique mais ne sont que partiellement

* IBUI : Identificateur Belge unique / Belgische unieke Identifier – Belgian Unique Identifier : concept de base avec identifiant numérique unique.

** www.cispclub.org.

*** 14 services administratifs, 10 services cliniques et 5 services techniques, soit un total de 29 entités.

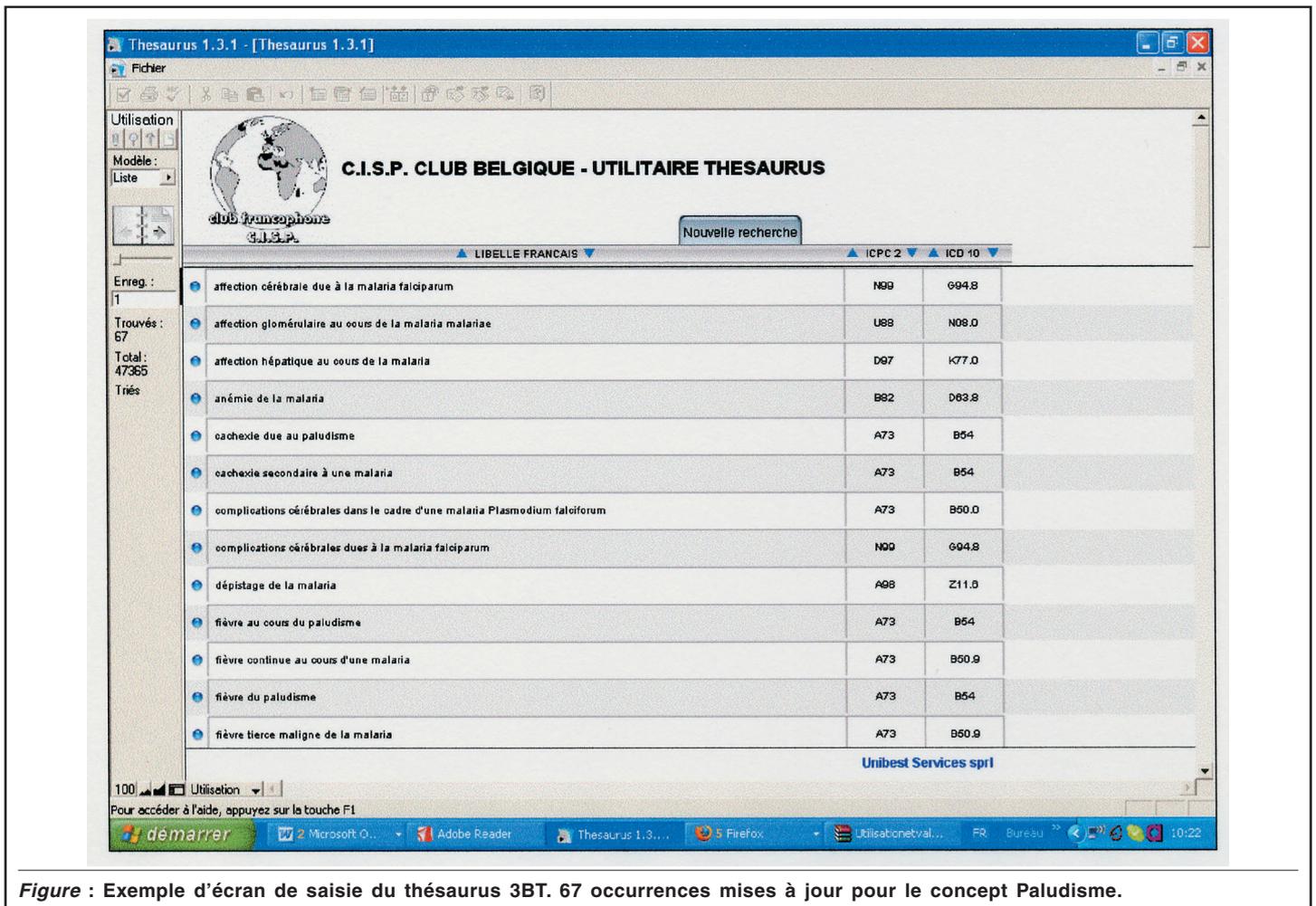


Figure : Exemple d'écran de saisie du thésaurus 3BT. 67 occurrences mises à jour pour le concept Paludisme.

reconnues par les responsables des services cliniques comme reflétant l'activité réelle des différents services.

Les données actuellement disponibles permettent de dégager des informations utiles sur le volume d'activités du CHUK, ainsi que sur la grande variabilité des performances entre les différents services de l'hôpital. Néanmoins, il apparaît également que ces mêmes informations ne permettent pas, à l'heure actuelle, d'avoir une idée précise et internationalement comparable sur les pathologies prises en charge et de faire des liens entre ces pathologies et la productivité réelle de l'hôpital en termes opérationnel (durée de séjour,...) et financier (coût par pathologie,...).

Cet écueil devrait être partiellement levé par la mise en place d'une politique de gestion systématique des informations de santé utilisant notamment les classifications internationales d'une ou plusieurs organisations internationales.

RESULTATS PRELIMINAIRES

Analyse du processus

Dans cette partie du travail, il s'agit d'évaluer l'apport du thésaurus belge 3BT pour l'encodage des diagnostics cliniques des patients hospitalisés. Il faut noter qu'il s'agit de la première mise à l'épreuve du thésaurus belge.

Impression subjective

Les deux encodeurs du service administratif du CHUK qui ne sont pas des médecins trouvent l'encodage des diagnostics cliniques facilité par l'utilisation du thésaurus 3BT. Ils se disent satisfaits, mais relèvent qu'il manque une série d'items parfois fort communs tels " cancer de la prostate ", " gastro-entérite aiguë ", etc. Ils nous remettent une liste comprenant 30 items manquants.

Analyse qualitative

Les premiers encodages de diagnostics cliniques à l'aide du thésaurus débutent début janvier 2006. Ces encodages sont obtenus de manière prospective à partir des diagnostics mentionnés sur les dossiers des patients hospitalisés dans les services de chirurgie, médecine interne et pédiatrie. Ils sont réalisés par une administrative du CHUK, habituée à la logique d'encodage d'informations cliniques depuis quelques années. Fin mars 2006, une première analyse des résultats est effectuée (Tableau 1). Le fichier comprend 899 données. 883 fiches sont remplies, 16 sont vides. Certains apparaissent une seule et unique fois (réurrence faible), d'autres apparaissent de nombreuses fois (réurrence forte).

Tableau 1 : Première analyse des résultats.

	N	N/données
Nombre de données	899	
Fiche remplie	883	98,22 %
Fiche vide	16	1,78 %
Réurrence la + forte d'un item	223	25,25 %
Réurrence la + faible d'un item	1	0,11 %
Nombre d'items apparaissant une seule fois	154	17,44 %
Nombre d'items apparaissant deux fois	30	6,80 %
Items à occurrence faible		24,24 %

Interprétation

Les récurrences très faibles de 24,4 % des encodages justifient l'utilisation mais aussi l'amélioration de la qualité du thésaurus 3BT.

Analyse des encodages vides (Tableau 2)

Les encodages vides sont des données manquantes au niveau des champs adéquats dans le logiciel. Il s'agit soit d'un oubli, soit de la volonté des administratifs de ne pas faire de faute lors de l'encodage des diagnostics cliniques à l'aide du thésaurus : ils n'ont donc pas trouvé dans le thésaurus un diagnostic normalisé équivalent au diagnostic transcrit sur le dossier du patient. Nous avons pu retrouver les dossiers des patients grâce au service

d'archivage et relever les diagnostics posés par les praticiens. Nous avons alors tenté de trouver un équivalent dans le thésaurus.

Interprétation

1,81 % d'appréciations cliniques non enregistrées ne modifieront pas fondamentalement les résultats des analyses micro-épidémiologiques qui seront réalisées au sein du CHUK. Il faut cependant remarquer que reprendre simplement les dossiers, relever les diagnostics portés et encoder les appréciations retrouvées permet de diminuer l'incertitude de près de 45 % en valeur relative et d'un peu moins de 1 % en valeur absolue. La contrepartie de ce gain de précision est une perte de temps non négligeable (dans ce cas précis : 3,5 heures pour 16 dossiers pour tout le processus de recueil de l'information médicale) !

Analyse des encodages manquants (Tableau 3)

Les encodages manquants sont des concepts qui n'ont pu être trouvés dans le thésaurus par les administratifs et pour lesquels ils ont relevé le diagnostic des praticiens dans une base séparée. L'analyse porte donc sur ce fichier séparé.

Interprétation

Le chiffre de 3,23 % d'intitulés cliniques non retrouvés dans le thésaurus pose la question de sa validité. Cependant, une première vérification permet

Tableau 2 : Analyse qualitative des encodages vides.

	N	N/vides	N/total encodages
Fiche vide/ patient	16		1,81 %
Dossier retrouvé	12	75,00 %	
Diagnostic mentionné	18		
Diagnostic non mentionné dans dossier	0	0,00 %	
Diagnostic mentionné et retrouvé dans thésaurus	8	44,44 %	0,89 %
Diagnostic mentionné non retrouvé mais concept apparenté existant	8	44,44 %	0,89 %
Diagnostic mentionné non retrouvé et concept inexistant dans thésaurus mais existant dans ICD	1	5,56 %	0,11 %
Diagnostic mentionné non retrouvé et concept inexistant dans le thésaurus et dans ICD	1	5,56 %	0,11 %

Tableau 3 : Analyse qualitative des encodages manquants.

	N	N/non trouvés	N/total encodages
Diagnostic non trouvé dans thésaurus par administratif	30		3,23 %
Diagnostic retrouvé dans thésaurus par médecin	4	13,33 %	0,43 %
Diagnostic non retrouvé dans thésaurus (= pas de correspondance complète)	26	86,67 %	2,80 %
Diagnostic non retrouvé mais concept apparenté présent	18	69,23 %	1,94 %
Diagnostic non retrouvé et concept inexistant dans thésaurus mais existant dans ICD-10	2	7,69 %	0,22 %
Diagnostic non retrouvé et concept inexistant dans thésaurus et inexistant dans ICD-10	2	7,69 %	0,22 %
Diagnostic non valide (procédure, anglais)	4	15,38 %	0,43 %
		100,00 %	

déjà de diminuer ce chiffre sous la barre des 3 % rien qu'en essayant à nouveau les recherches. Un médecin, habilité à rechercher et valider des concepts apparentés (néoplasie maligne = tumeur maligne = cancer) et à refuser l'intitulé d'une procédure en tant qu'appréciation diagnostique, permet de diminuer l'incertitude à moins de 0,5 % et même sous les 0,25 % s'il prend la peine de rechercher les concepts dans les tables de classifications elles-mêmes.

Analyse de terrain

Un médecin habitué aux classifications et une administrative du service d'encodage au sein du CHUK ont travaillé une demi-journée ensemble sur les mêmes dossiers. 32 diagnostics, bilan médical de l'hospitalisation, sont encodés, en double aveugle. Les résultats de ce travail sont présentés dans le Tableau 4.

Correspondance totale	69,70 %
Imprécision d'encodage	21 %
Faute d'encodage	0 %
Inconnu	9,09 %
	100,00 %

Interprétation

Près de 70 % de l'encodage est parfaitement similaire : le même ICD-10 est choisi par les deux encodeurs. Il y a près de 20 % d'imprécision d'encodage de l'administrative par rapport au médecin et 9 % de fiches laissées vides (= inconnus). La responsable administrative préfère diminuer la précision de l'information par rapport au diagnostic clinique écrit du praticien, plutôt que de commettre une faute. Les dossiers sont mis sur le côté et pourront être retrouvés pour pouvoir être retravaillés par après si nécessaire. Il faut souligner qu'aucune faute d'encodage n'a été relevée. Nous ne pouvons certainement pas conclure qu'aucune faute d'encodage n'est réalisée, mais on peut raisonnablement penser qu'elles seront limitées, et donc n'interféreront pas sur l'analyse générale globale de l'activité au sein du CHUK.

Conclusion du travail d'analyse qualitative de l'encodage des données au CHUK

- La qualité de l'encodage par les administratifs du CHUK est suffisante pour pouvoir interpréter les résultats.
- La relecture des inconnues par un médecin sensible à la logique classificatoire permet de diminuer les imprécisions à un taux inférieur à 0,25 % des encodages, ce qui est remarquable.
- Un meilleur système de recherche dans le thésaurus permettrait de gagner de la précision et du temps. Cela paraît important, même si l'enjeu n'est sans doute ni crucial, ni immédiat, pour la qualité des données cliniques enregistrées.

BIBLIOGRAPHIE

1. Annexe C, Rapport de la mission complémentaire relative à la classification des pathologies au niveau du CHUK, M. De Jonghe, Département de Médecine Générale, Faculté de Médecine, D. Porignon, Ecole de Santé Publique Université libre de Bruxelles, 13 au 20 août 2005
2. ICD-10. International Statistical Classification of Disease and Related Health Problems. Geneva (Switzerland), World Health Organisation, 1992
3. Jamouille M, Roland M, Humbert J, Brûlet JF, eds : Traitement de l'information médicale par la Classification Internationale des Soins Primaires, 2^{ème} version (CISP-2), assorti d'un glossaire de médecine générale, préparé par le Comité International de Classification de la WONCA. Buxelles, Care Edition, 2000
4. WONCA International Classification Committee. International Classification of Primary Care ICPC-2-R, revised second edition. Oxford University Press, 2005
5. Roland M, De Jonghe M, Kacenenelbogen N, Dendeau B : Convention " implémentation d'un thésaurus belge, codé ICPC-2 ", N° 605 QG / N04 - 365 – années 2004 – 2005 entre le Ministère de la Santé Publique et l'U.L.B.

Correspondance et tirés à part :

M. DE JONGHE
Rue du Viaduc 7
7500 Tournai

Travail reçu le 2 mai 2006 ; accepté dans sa version définitive le 1^{er} août 2006.