



HAL
open science

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes industriels

Brigitte Juanals, Martin Lafréchoux, Jean-Luc Minel

► To cite this version:

Brigitte Juanals, Martin Lafréchoux, Jean-Luc Minel. KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes industriels. Institut Porphyre. TOTH 2012: Terminologie & Ontologie: Théories et applications, Jun 2012, Annecy, France. pp.1-20, 2012. halshs-00709019

HAL Id: halshs-00709019

<https://shs.hal.science/halshs-00709019v1>

Submitted on 16 Jun 2012

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.

Conférence Toth 2012

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes industriels

Brigitte Juanals*, Martin Lafréchoux*
Jean-Luc Minel*,

*MoDyCo, UMR 7114, Université Paris Ouest Nanterre La Défense – CNRS
92001 Nanterre
Brigitte.juanals@orange.fr
<http://www.modyco.fr>

Résumé. Cet article décrit la conception d'une méthodologie instrumentée, articulée en différentes étapes, qui vise à analyser les variations terminologiques dans des textes industriels. Nous cherchons notamment à identifier, dans les textes de normes, l'origine et les modes d'élaboration des concepts, ainsi que les variations diachroniques des termes qui les lexicalisent. La présentation du cadre théorique est suivie par la description de l'architecture de l'outil KONTRAST, et de la chaîne de traitements associée. En dernier lieu, des modes de représentation associés à des cas d'usage exposent les possibilités d'analyse linguistique et socio-pragmatique de cette méthodologie instrumentée. L'expérimentation est menée en utilisant un corpus issu du domaine de la continuité d'activité. La continuité d'activité désigne l'ensemble des procédures et mesures mises en place par une organisation (entreprise, institution, collectivité...) pour s'assurer que ses fonctions critiques seront préservées en cas de sinistre ou de désastre (catastrophe naturelle, attentat, risque sanitaire...). Ce travail s'inscrit dans une recherche plus large sur les textes et les acteurs de la normalisation industrielle internationale dans le domaine de la sécurité globale.

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

1. Introduction

Dans le champ de la normalisation internationale de la sécurité sociétale, des normes d'organisation – dites « normes de management » – de la sécurité sont édités (ou en cours d'édition) par des ONG habilitées à conduire des activités de normalisation. Ces textes industriels traitent de la sécurité et de la protection des personnes et des biens ; ils ont pour caractéristiques d'être internationaux, européens ou nationaux dans leur mode de production et de diffusion. Ils sont destinés à être mis en œuvre largement dans un espace désormais transnational et dans les organisations – publiques ou privées.

Les évolutions récentes des textes normatifs – au travers de normes dites de « management de la sécurité de l'information » – sont au cœur de notre analyse. A l'instar des normes techniques et des normes de qualité, ces normes contribuent, au travers de la certification et des bonnes pratiques, à structurer les pratiques, les modes d'organisation, d'évaluation de contrôle de l'information dans les organisations et à proposer des terminologies dans des langues de spécialité.

Notre objet d'étude – les normes de management de la sécurité de l'information dans le contexte international de la sécurité sociétale – nous a amenés à concevoir un cadre d'analyse spécifique dédié à l'analyse des textes normatifs. Notre objectif est de décrire et d'analyser ces textes afin d'appréhender les politiques communicationnelles et organisationnelles qu'ils préconisent au sein des organisations. Nous souhaitons également étudier le processus de construction terminologique qui est mis en œuvre par les comités techniques (TC) qui en ont la responsabilité.

Cette communication s'inscrit dans le cadre plus vaste du projet de recherche ANR « NOTSEG »¹. L'une des finalités est d'identifier, dans un secteur d'activité déterminé, les conditions et les mécanismes qui amènent l'introduction de nouvelles normes à participer à la cristallisation, à la communication et à la circulation de certaines préconisations de pratiques, de savoirs et de procédures, de modes d'organisation et d'agencements socio-techniques². Ce travail est structuré en cinq axes principaux, en vue de répondre aux questions suivantes : i) selon quels principes organisationnels et selon quelles modalités ces textes peuvent-ils contribuer à structurer les organisations et à encadrer les actions des individus qui y travaillent ou qui entretiennent des relations de travail avec ces dernières ? ii) Quelles sont les motivations et les stratégies de communication des organisations qui en sont les

¹ Normalisation et Sécurité Globale : la formulation du concept de sécurité globale dans la normalisation www.notseg.fr, (appel ANR-CSOG 2009).

² Pour compléter cette étude, il sera nécessaire d'analyser ultérieurement si ces nouvelles normes jouent effectivement un rôle dans des ré-agencements ou des évolutions des éléments cités ci-dessus, ce qui induirait une modification des connaissances et des pratiques professionnelles qu'ont les acteurs de leur domaine.

« auteurs-acteurs » ? iii) Quel est le degré de performativité de ces textes ? iv) Selon quels mécanismes les organisations sont-elles amenées à faire référence à ces normes et à les utiliser ? v) Quels sont, dans ces textes de normes, l'origine, les modes d'élaboration des concepts et les variations diachroniques des termes qui les lexicalisent ?

Cet article est focalisé sur le point v). L'avancement de nos recherches sur les autres aspects évoqués figure dans [Juanals et Minel, 2011 ; 2012], [Juanals, 2012], [Lafréchoux *et al.*, 2012a ; 2012b].

La section 2 décrit les travaux connexes à notre recherche. Dans la section 3, nous précisons le cadre théorique dans lequel s'inscrit notre réflexion. La section 4 décrit l'architecture de l'outil KONTRAST que nous avons conçu. La section 5 présente la chaîne de traitements mis en œuvre pour peupler la plate-forme, et la section 6 présente des modes de représentation associés à des cas d'usage. L'expérimentation est menée en utilisant un corpus de 18 normes issu du domaine de la continuité d'activité.

2. Aperçu de travaux de recherche connexes portant sur des terminologies et des ontologies métiers

En 2000, [Lelong B. et Mallard A., 2000] observaient déjà au sujet des normes techniques : « Que les normes techniques aient à voir avec le lien social et la communication, voilà qui peut surprendre » (p. 9). Les normes industrielles (normes techniques ou normes de management) constituent un objet d'étude relevant en général des sciences de gestion et des sciences de l'ingénieur. Il existe encore peu de travaux les concernant en Sciences de l'information et de la communication [Perriault J. et Vaguer C., 2010], [Juanals, 2010], [Juanals et al., 2011], mais elles peuvent être prises en compte dans des recherches plus larges (comme dans [Bazet I. et al., 2008]).

De nombreux travaux ont été menés dans le domaine de la terminologie et de l'ingénierie des connaissances. En tant que « pratique théorisée » [Roche, 2005], la terminologie a connu un développement rapide au cours des dernières décennies, notamment afin de répondre aux besoins des industries manufacturières [Condamines, 2005]. Cette forte opérationnalisation de la terminologie a favorisé l'émergence d'une relation « symbiotique » entre terminologie et ingénierie des connaissances [Condamines, 2005]. Les approches mêlant ontologie formelle et ressources terminologiques qui en résultent sont diverses ; elles dépendent fortement de leur contexte de réalisation.

Pour les industriels travaillant sur plusieurs sites ou avec de nombreux sous-traitants, dans les secteurs de l'automobile ou de l'aviation par exemple, l'emploi d'un vocabulaire de spécialité, partagé et contrôlé, s'impose. Récemment, [Omrane et al., 2011] ont présenté un prototype de système de gestion de règles métier

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

(Business Rules Management System, ou BRMS) fondé sur une ontologie OWL et sur des concepts exprimés selon le formalisme SKOS. Développé pour Audi dans le cadre du projet européen ONTORULE, le BRMS s'appuie sur une « ontologie lexicalisée » destinée à formaliser des règles métier écrites en langage naturel par des experts du métier. Ce système doit permettre d'accélérer le partage d'informations entre des départements qui utilisent parfois des termes différents pour désigner les mêmes procédures.

[Roche, 2007], [Damas et Tricot, 2010] ont proposé le modèle plus général de l'ontoterminologie, une « terminologie dont le système notionnel est une ontologie formelle ». Dans ce cadre, le terminologue doit travailler avec les utilisateurs et les experts du domaine pour établir une conceptualisation du domaine susceptible de faire consensus. Cette représentation est ensuite transposée sous forme d'ontologie, puis des termes de spécialité fournis ou validés par des experts sont associés à chaque concept. Dans le contexte dans lequel NOTSEG s'insère, le travail d'harmonisation terminologique fait partie du cœur de métier des instances de normalisation. C'est à elles qu'il appartient de désigner le mot 'juste' et de le définir, au cours d'un processus de rédaction qui fait intervenir différents experts du domaine et aspire à créer un consensus général. Notre travail se situe en aval de ces démarches terminologiques sur lesquelles nous n'avons pas prise.

A propos d'un sujet proche de celui de NOTSEG, [Gresser, 2010] a présenté une ontologie des risques financiers. Ce domaine recoupe en partie la continuité d'activité. Gresser fait le constat préalable de l'omniprésence des lexiques alphabétiques, et celui de leurs carences en termes d'utilité pratique, hormis pour un usage de référentiel. Son ontologie est basée sur une représentation conceptuelle du domaine qu'il a élaboré en se fondant sur ces référentiels. En tant qu'expert, Gresser a pu établir lui-même la conceptualisation du domaine utilisée comme base de son ontologie. Son travail d'unification a permis d'aboutir à des concepts univoques, et dont les relations ont pu être validées ou proposées une à une par des experts. La composante terminologique de son travail se présente sous la forme de deux « listes de vocabulaire » accompagnant l'ontologie, sous forme de glossaires alphabétiques. Ces différents choix ont permis à l'ontologie présentée d'être immédiatement opérationnelle. Notre approche se distingue de celle de Gresser sur plusieurs points. D'une part, KONTRAST intègre les données terminologiques dans l'ontologie. D'autre part, nous avons également travaillé à partir de termes de spécialité extraits de référentiels existants, mais sans chercher à les harmoniser. L'objectif n'était pas de produire une vision unifiée du domaine, mais de permettre de comparer les différentes visions existantes.

Dans le domaine médical, la multiplication des terminologies de spécialité est parfois problématique, notamment pour les tâches d'indexation. La solution la plus simple consiste à recourir au vocabulaire UMLS³. Cependant, selon [Roumier et al.,

³ Unified Medical Language System - <http://www.nlm.nih.gov/research/umls/>

2011], la volonté de recourir à un vocabulaire contrôlé unifié est « naïf », et pourrait avoir comme conséquence de « nuire à la santé des patients ». Ces auteurs proposent un système permettant de préserver les caractéristiques propres des différentes terminologies utilisées dans un système de recherche d'informations médicales. Leur système utilise une « terminologie d'interface » multilingue et interdisciplinaire, couplée à des terminologies spécialisées et généralement monolingues, destinées aux utilisateurs finaux. Dans une optique proche, [Grosjean *al.*, 2011] ont présenté EHTOP (European Health Terminology/Ontology Portal), un service offrant un accès unifié à trente-deux terminologies médicales, destinés à des utilisateurs humains ou à des agents logiciels. EHTOP intègre les différentes terminologies en les transposant dans un méta-modèle générique, puis en effectuant des « harmonisations sémantiques inter et intra-sémantiques ». Bien que les domaines et les objectifs poursuivis soient différents, la technique employée semble très proche de celle que nous avons adoptée.

[Charlet *et al.*, 2009] présentent leur approche de la création et de l'emploi de ressources onto-terminologiques en médecine au travers de trois applications concrètes. Les trois projets sont bâtis en suivant un ensemble de principes communs, résumés par la formule : le « primat de la terminologie sur le raisonnement ». Il s'agit de « prendre en compte le caractère majoritairement documentaire et linguistique de la médecine » en favorisant l'expression en langue naturelle des connaissances, quitte à réduire leur calculabilité et les possibilités d'automatisation offertes. Dans le premier projet, l'idée est de « proposer un système de représentation des connaissances avec lequel le pneumologue puisse interagir (...), en tenant compte de ses propres capacités de choix et d'interprétation mais en l'aidant dans sa tâche. » Le second projet présente une chaîne de traitements permettant d'extraire semi automatiquement des concepts à partir de formulaires d'observations et de guides de bonnes pratiques, grâce à l'utilisation d'outils s'appuyant sur des technologies issues du Traitement Automatique des Langues (TAL) comme Syntex et Upery. Dans sa dernière partie, l'article montre les difficultés à établir des correspondances entre les ressources terminologiques et les ontologies existantes : les différences de perspective et de granularité d'une discipline à une autre rendent très problématique l'alignement de concepts censément identiques. Les auteurs en concluent que les modélisations produites par chaque spécialité sont « irréductibles les unes aux autres et impossibles à combiner ». Nous partageons ces conclusions.

Dans les domaines où l'ingénierie des connaissances a connu le plus de succès, ces exemples illustrent des cas de conflits ou d'incohérences qui ont été résolus grâce à la présence d'une autorité centrale (la direction d'un groupe industriel) et d'une expertise reconnue (un expert mandaté par une organisation), ou grâce à un consensus réalisé autour d'un vocabulaire unifié préexistant (par exemple UMLS dans le domaine médical).

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

En revanche, dans le domaine de la normalisation industrielle internationale, les organes de normalisation n'ont pas vocation à imposer une définition, un terme ou un concept. Parmi les organismes de normalisation internationaux, l'ISO joue un rôle central et peut formuler des recommandations via des « guides » de structuration de normes ou d'emploi de termes spécifiques. Toutefois, en son sein, ce sont essentiellement les comités techniques et les groupes de travail qui sont chargés de l'écriture des normes internationales. L'écriture négociée de ces textes est en partie influencée par des actions émanant de groupes de pression.

Chaque organisme de normalisation reste libre de proposer son propre vocabulaire. L'originalité de notre travail a donc consisté à bâtir une ressource terminologique en nous basant sur plusieurs conceptualisations préexistantes, afin de pouvoir les représenter simultanément, et ce sans chercher à les synthétiser.

3. La construction d'un cadre d'analyse socio-pragmatique et linguistique

Notre travail a nécessité la construction d'un cadre théorique spécifique. D'une part, nous avons recours à une approche textuelle [Adam, 2011], discursive [Maingueneau, 1996] et pragmatique. D'autre part, nous prenons en compte les apports de la sociologie des sciences, qui est centrée sur les acteurs, leurs institutions et leurs politiques industrielles ou scientifiques [Latour, 2006]. La traçabilité informationnelle place les collectifs (en termes de culture, de pratiques professionnelles, d'intérêts industriels...) au cœur de la circulation des savoirs, ce qui souligne l'intérêt d'une approche socio-pragmatique et sociopolitique de ces problèmes. Dans cette direction, l'analyse des controverses scientifiques de [Callon, 1986] et [Latour, 2006] se focalise sur la posture d'affrontement des points de vue et des conflictualités qui sont situés au cœur des savoirs et des cultures. Dans des réseaux sociotechniques, les acteurs (individuels et collectifs, humains et non humains) négocient en permanence leurs intérêts, leurs problèmes, leurs identités, leurs langages, au sein de controverses. Les controverses impliquent une conception étendue des narrations qui suppose l'acceptabilité des divers types de discours énoncés à partir de points de vue différents. Chaque collectif défenseur de son système de discours – lui-même porteur d'enjeux politiques, idéologiques, scientifiques ou économiques – défend sa version d'un problème posé, ce qui peut générer des croisements, parfois des luttes, entre les différentes versions qui sont produites. Il n'est pas seulement question de subjectivité et de langage, mais tout autant de stratégies et d'intérêts, de représentations, de préjugés ou de sentiments. Dans le contexte de l'analyse de la circulation des savoirs [Juanals, 2008] et du suivi de controverses industrielles et scientifiques, la prise en compte de ces diversités et de ces différences implique l'élaboration d'une pensée de l'hétérogénéité.

La possibilité d'identifier les stratégies d'influence des organisations au moyen de l'analyse des textes normatifs est posée comme hypothèse de travail. Plus précisément, nous formulons l'hypothèse que l'analyse des traces linguistiques dans les normes peut être exploitée pour identifier des choix de modèles organisationnels et de contrôle structurants. L'analyse que nous proposons met en relation des textes normatifs avec les « auteurs-acteurs » qui participent à leur rédaction, et avec le contexte d'appartenance de ces acteurs. Ces choix de modèles organisationnels pourraient correspondre à des stratégies de communication dont les motivations restent à élucider. Dans cette perspective et en ce qui concerne notamment le processus d'édition des normes, nous remettons en question l'affichage institutionnel d'un fonctionnement dit de « consensus », bien que ce dernier soit présenté comme l'une des règles de fonctionnement des comités de normalisation pour l'élaboration collective des textes.

L'étude des textes normatifs et de leur contexte socio-organisationnel est réalisée concomitamment à des observations de terrain initiées en 2010. Elles sont menées selon des méthodes qualitatives, sous la forme d'observation participante dans des réunions de comités normalisation européenne et internationale (Comité européen de normalisation CEN/TC 391 Sécurité sociétale et du Citoyen, Comité international ISO/TC 223 Societal Security), et d'entretiens avec les experts qui y participent.

Les normes sont étudiées comme la cristallisation temporaire⁴ d'éléments qui relèvent de dimensions socioculturelle, juridique, industrielle et technique. L'activité d'énonciation est analysée à partir des traces linguistiques laissées dans les énoncés (les textes normatifs) comme autant d'indices de la subjectivité des scripteurs (les experts) ayant participé à leur élaboration. Il s'agit de saisir et de rendre compte de la diversité des positionnements et des points de vue nationaux des acteurs, qui s'expriment dans des régimes spatio-temporels et culturels hétérogènes.

4. Présentation de la méthodologie instrumentée dans la plate-forme KONTRAST

4.1 L'élaboration des ressources langagières du glossaire

En interaction avec le terrain, l'observation des comités de normalisation et les entretiens avec les experts ont progressivement fait émerger les normes et les acteurs situés au cœur des débats et des enjeux d'organisation et de marché. Selon une approche socio-pragmatique [Chateauraynaud, 2011] adaptée aux thèmes en construction et aux controverses dont nous traitons, les informations issues du terrain et des acteurs ont été privilégiées pour construire le corpus de textes.

⁴ Les normes étant révisées à intervalles réguliers, en moyenne au bout de deux ou trois ans.

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

Le corpus d'étude est composé actuellement d'un ensemble de dix-huit normes nationales et internationales, en anglais, dans le domaine de la continuité d'activité et de la gestion des risques et des crises. Il comporte des textes nationaux et internationaux en vue d'étudier la possibilité de confirmer ou d'infirmer, en procédant de manière comparative, l'hypothèse de jeux d'influence entre des textes normatifs. Plus précisément, nous cherchons à savoir si certains acteurs influents ont réussi à imposer des concepts, des termes, des définitions, des procédures, etc., lors de l'élaboration des normes internationales.

Dans l'analyse des textes des normes, nous avons centré l'attention sur les termes les définitions qui y sont associées ; ils sont principalement regroupés dans une section « Terms and Definitions » (T&D) que l'on retrouve dans toutes les normes. Chaque norme dispose ainsi de son propre vocabulaire, employé dans un objectif de « stipulation » [Depecker, 2005]. Le choix de ces termes et de la définition qui leur est donnée sont de première importance : d'une part, il est le reflet du système notionnel des experts et, d'autre part, conditionne la mise en œuvre concrète de la norme. Par ailleurs, on constate que la section « T&D » d'une norme cite ou reprend explicitement les définitions d'une norme préexistante. Ces réseaux de citations, qui ne sont pas toujours explicités, forment un système complexe d'emprunts et de références, tant au sein de chaque norme (réseau interne) qu'entre les normes du corpus (réseau externe) (cf figure 1).

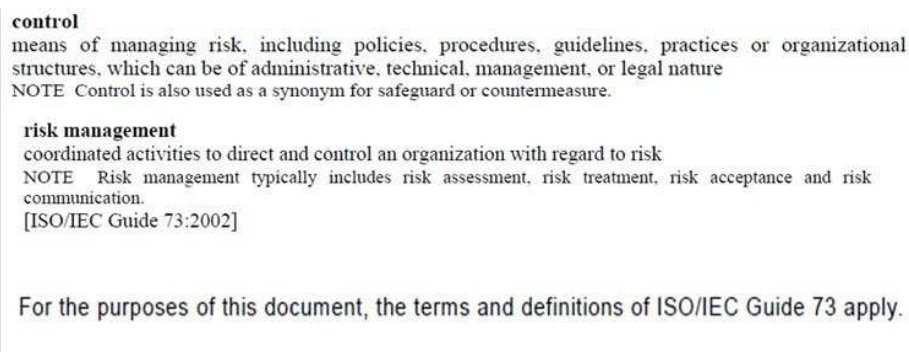


FIG. 1 – Exemples des différents systèmes utilisés pour indiquer des importations ou des compléments de définitions

De plus, la décision de centrer notre étude sur ces sections « T&D » a été motivée par l'observation des débats dans les réunions de normalisation. Nous avons pu assister à de longues discussions et à des négociations parfois véhémentes à leur sujet⁵. Tous les experts consultés se sont accordés à reconnaître l'importance

⁵ Les discussions et les textes intermédiaires au sein de ces comités de normalisation sont couvertes par des règles de confidentialité. Toutefois, notre travail d'observation participante dans ces comités est

stratégique des termes et des définitions associées qui sont retenus pour chaque norme. A l'issue du processus de rédaction, un ensemble de concepts, de définitions et de principes organisationnels sont ainsi réifiés.

Nous avons donc conçu notre outil de description et d'analyse de manière à ce qu'il puisse représenter simultanément plusieurs systèmes notionnels et terminologiques en parallèle. Notre approche s'appuie sur les propriétés opérationnelles des ontologies en RDF/OWL ; toutefois, elle n'a pas recours à la démarche terminologique qui a été décrite par [Roche, 2007]. C'est la raison pour laquelle nous ne reprenons pas le terme d'« ontoterminologie » de Roche, auquel nous préférons celui de glossaire contrastif ontologique.

L'un des livrables du projet NOTSEG est un glossaire structuré, synthétique et représenté en XML [Malik, 2011]. Elaboré manuellement à partir du corpus des 18 normes, il contient les différents glossaires et ressources terminologiques du corpus. KONTRAST a été construit à partir de ce glossaire. Le glossaire initial rassemble les termes et définitions des normes du corpus, le contenu des guides terminologique, et les relations entre les termes. Pour chaque terme, le glossaire indique les différentes définitions provenant de l'ensemble des normes du corpus, ainsi que les termes alternatifs éventuels. Les relations d'emprunts et d'héritage sont également indiquées lorsqu'elles sont explicites dans les textes et ont de ce fait pu être relevées lors de la constitution du glossaire. Toutefois, il n'existe pas de lien hypertexte entre ces termes et certaines relations d'emprunt ou de description mentionnées dans les textes des normes imprimées utilisant des sémiotiques différentes et implicites, n'ont pas été codées. Au total, ce glossaire contient 726 définitions de 291 termes issus des 18 normes du corpus. La DTD utilisée s'inspire de travaux sur les ressources terminologiques de spécialité en XML, notamment [Joseph, 2010].

4.2 Choix techniques et méthodologiques

Comme nous l'avons indiqué (cf. supra), nous formulons l'hypothèse que les sections « T&D » sont la manifestation du système notionnel utilisé par les experts qui les ont rédigées. La réalisation de KONTRAST vise à représenter le plus fidèlement possible les différentes conceptualisations exprimées par les définitions associées aux termes de spécialité. Nous n'avons donc pas cherché à choisir entre les différentes variations des définitions données à un même terme, mais plutôt à relier ces variations. Nous introduisons donc une représentation plus abstraite afin de pouvoir regrouper des notions et des termes, et de respecter les contraintes du langage formel utilisé (RDF) qui impliquent de désigner un individu⁶ par un identifiant unique référençable. Nous appuyant en partie sur OWL et SKOS pour construire KONTRAST, les deux structures abstraites choisies sont le

essentiel pour comprendre et acquérir des connaissances sur les sujets discutés. Ils constituent un socle d'informations permettant de mener des entretiens avec les experts.

⁶ Au sens où on l'entend en Ingénierie des connaissances.

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

« skos:Concept » et le « U-Concept » qui est une sous-classe de « skos:Collection ». L'emploi des guillemets pour la graphie « Concept » signale qu'il s'agit pour nous d'exploiter les fonctionnalités de SKOS, et non de prendre position dans le débat Concept *versus* Terme.

Les « skos:Concept » représentés dans KONTRAST sont donc issus des termes relevés dans les glossaires qui sont inclus dans les normes dans les sections « T&D » (cf. figure 2). Il s'agit de représenter une relation triadique entre une forme graphique, un contexte d'emploi et une définition. D'un point de vue formel, nous avons choisi de représenter les termes de spécialité, en utilisant leur graphie d'origine, comme des entités sans valeur sémantique ; ces termes sont appelés « U-Concept » dans KONTRAST. Enfin, le glossaire « T&D » de chaque texte de norme du corpus a été transposé dans KONTRAST sous la forme d'une instance de la classe « skos: ConceptScheme ». Concernant les identifiants des individus, c'est-à-dire les entrées du glossaire d'origine, nous avons choisi un principe de concaténation proposé par [Lafréchoux et al., 2012]. Nous avons préféré cette solution à l'alternative qui aurait consisté choisir des identifiants alphanumériques peu lisibles par un humain, comme dans [AGROVOC, 2011].

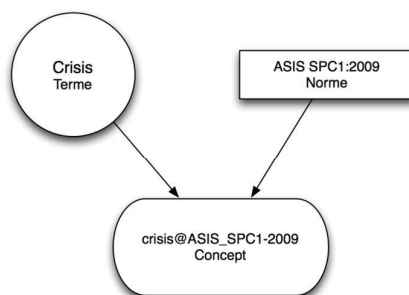


FIG. 2 – Dans KONTRAST, la classe « Concept » de SKOS est utilisée pour désigner l'emploi d'un terme dans le contexte d'une norme.

De manière complémentaire aux données linguistiques, KONTRAST contient des informations relatives aux instances de normalisation ayant rédigé ou publié les normes du corpus. La prise en compte de ces informations a pour objectif de permettre l'analyse des variations dans les définitions associées aux « skos:Concept » au moyen de leur mise en relation avec les instances de normalisation. Les définitions sont reliées à la partie organisationnelle par le biais des individus qui représentent les normes (cf. figure 3).

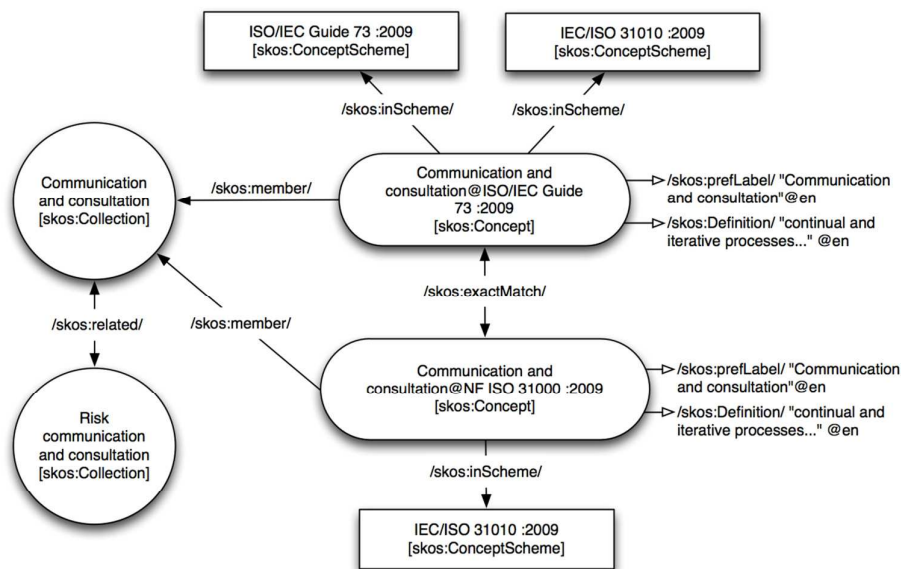


FIG. 3 – Exemple de triplets RDF dans KONTRAST.

5. Peuplement de la plateforme KONTRAST

La plateforme KONTRAST a été peuplée automatiquement à partir du contenu du glossaire structuré en XML (cf. supra). Les termes et les relations décrites dans le glossaire ont été transformés en utilisant des templates XSLT pour les transformer en RDF/OWL. Dans KONTRAST, chaque individu reçoit un identifiant unique, stable et référençable. Cet identifiant est calculé en appliquant une procédure unique de normalisation et de concaténation des informations du glossaire.

Les différentes références à la même entité au sein du glossaire (par ex. une norme) doivent donc conduire au calcul d'un même identifiant unique dans KONTRAST. Cependant, en dépit du soin apporté à la constitution du glossaire par le rédacteur humain, quelques variations ont échappé à la vigilance de celui-ci. Ces variantes ont été résolues en utilisant une fonction de normalisation en XSLT, qui permettait de régler les problèmes d'encodage, d'espaces multiples, et de caractères spéciaux ou accentués. Les incohérences restantes ont été corrigées manuellement après une recherche par expressions régulières.

Lors de la transposition du glossaire, des relations sont automatiquement extraites afin de lier les concepts ayant des définitions proches ou identiques, mais qui n'étaient pas explicitement référencées comme telles. Ces relations implicites n'ont pas été relevées manuellement lors de la constitution du glossaire. La

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

proximité de ces définitions est évaluée uniquement en fonction de leur texte, sans prise en compte de leur contenu sémantique.

Pour extraire les relations, nous effectuons une série de tests qui comparent les définitions correspondant aux mêmes termes, en allant du test le plus précis (identité parfaite - /skos:closeMatch/) au plus lâche (parenté - /skos:relatedMatch/). Lorsqu'un test réussit, la relation identifiée est transcrite sous forme de triplet RDF/OWL. Cette approche permet de maximiser la précision des résultats.

- /skos:exactMatch/

Une relation /skos:exactMatch/ correspond à une identité complète entre deux définitions. Pour extraire ces relations, nous avons calculé la distance de Levenshtein entre le texte de deux définitions. Dans 84 cas, les textes des définitions étaient strictement identiques (Levenshtein == 0).

- /skos:closeMatch/

Souvent, une définition est citée en totalité, mais avec de petites variations graphiques, les conventions typographiques différant d'un organisme de normalisation à l'autre. Ces relations de quasi-identité sont bien exprimées par /skos:closeMatch/.

Les variations formelles, par exemple le mode de numérotation des paragraphes, font augmenter considérablement la distance entre deux définitions par ailleurs identiques. Levenshtein ne constituait donc pas une approche satisfaisante. Pour le test 2, on normalise (*tokenization*, nettoyage de la ponctuation, passage en minuscule) le texte des définitions avant de les comparer.

- /skos:relatedMatch/

Deux définitions peuvent avoir une première partie identique, et n'être différenciées que par les notes qui suivent cette première partie. Dans KONTRAST, ces relations sont rendues par /skos:relatedMatch/. Le test 3 commence par analyser la structure de la définition pour en distinguer la partie principale et les notes, puis il applique le même traitement qu'au test précédent.

all-hazards	all-hazards-ISO_DIS_22300-2011	naturally occurring events, human induced events (both intentional and non-intentional) and technology caused events with potential impact on an organization, community or society and the environment on which it depends
all-hazards	all-hazards-NFPA_1600-2010	An approach for prevention, mitigation, preparedness, response, continuity, and recovery that addresses a full range of threats and hazards, including natural, human-caused, and technology-caused.
audit	audit-BCI_Good_Practice_Guidelines_2010	A systematic, independent and documented process for obtaining audit evidence and evaluating it objectively to determine the extent to which audit criteria are fulfilled. First-party audits are conducted by the organization itself for management review and other internal purposes, and may form the basis for an organization's declaration of conformity. Second-party audits are conducted by parties having an interest in the organization, such as customers, or by other persons on their behalf. Third-party audits are conducted by external, independent auditing organizations, such as those providing certification of conformity to a standard.
audit [BS EN ISO 9000:2005]"	audit-BS_25999-2-2007	systematic examination to determine whether activities and related results conform to planned arrangements and whether these arrangements are implemented effectively and are suitable for achieving the organization's policy and objectives.
audit	audit-ISO_DIS_22300-2011	systematic, independent and documented process for obtaining audit evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled

FIG.4 – Exemples de variations terminologiques

- En cours d'implémentation : renvoi et héritage

La structure des définitions offre d'autres régularités que nous souhaitons exploiter dans de futures versions de notre outil. D'une part, le texte d'une définition peut renvoyer explicitement à une autre définition de la même norme, avec des formes telles que « See also ». D'autre part, les définitions reprises explicitement comportent souvent une référence à la norme d'origine. Cette référence se fait généralement entre crochets droits, et pourrait être extraite automatiquement.

Toutes les données calculées dans les étapes précédentes sont représentées en RDF/OWL. Elles sont ensuite importées dans l'éditeur Protégé⁷. Cette importation permet de tester l'intégrité des données et de naviguer dans le « triplestore » en exprimant des requêtes dans le langage SPARQL.

⁷ (<http://protege.stanford.edu>)

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

6. Modes de représentation en fonction de deux catégories d’usagers

Le travail d’analyse des besoins et de caractérisation des profils d’acteurs, réalisé dans le projet ANR, nous a conduits à prendre en considération deux catégories d’usagers qui pourraient être intéressés par KONTRAST. En premier lieu, les usagers que nous nommons « Lecteurs » pourront naviguer dans le glossaire terminologique ontologique au moyen des fonctionnalités « d’utilisateur final » ; ils auront également à leur disposition plusieurs modalités de visualisations graphiques. En second lieu, les usagers que nous nommons « Explorateurs » utiliseront le langage SPARQL pour explorer les données, chercher des liens complexes afin de vérifier une hypothèse, ou évaluer et quantifier des faits terminologiques. Ils pourront également utiliser des raisonneurs (Fact++, Hermit, etc.) proposés dans Protégé ou d’autres plateformes pour construire une cohérence ou pour extraire une vérification en vue de favoriser la stabilisation terminologique d’un champ émergent.

6.1 Mode de représentation lié au cas d’usage par des « Lecteurs »

A l’heure actuelle, KONTRAST ne dispose pas de sa propre interface graphique. L’exploration visuelle repose sur des outils tiers capables de représenter visuellement un fichier OWL, par exemple RDF Gravity ou Ontograf⁸. Le fonctionnement de ces outils repose sur une recherche textuelle qui permet de filtrer les individus affichés jusqu’à ce que l’on obtienne les individus souhaités. L’utilisateur peut ensuite cliquer sur chaque nœud du graphe visualisé afin d’afficher ou de masquer ses relations avec d’autres individus. Les figures 5 et 6 illustrent ce type d’usage.

⁸ Les captures qui suivent ont été réalisées avec Ontograf dans l’éditeur Protégé 4.

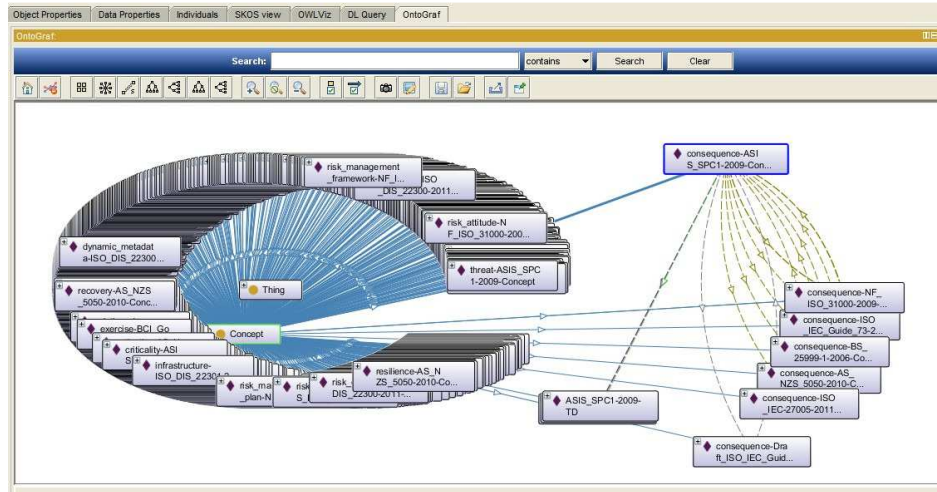


FIG. 5 – Visualisation d'un terme, de son contexte d'emploi et de ses relations avec des termes ayant la même graphie

Les classes, individus et relations sont colorés automatiquement afin de faciliter l'exploration. Un utilisateur pourra ainsi facilement comparer des individus ayant des propriétés communes, par exemple le même terme de spécialité ou le même contexte d'emploi (cf. figure 5).

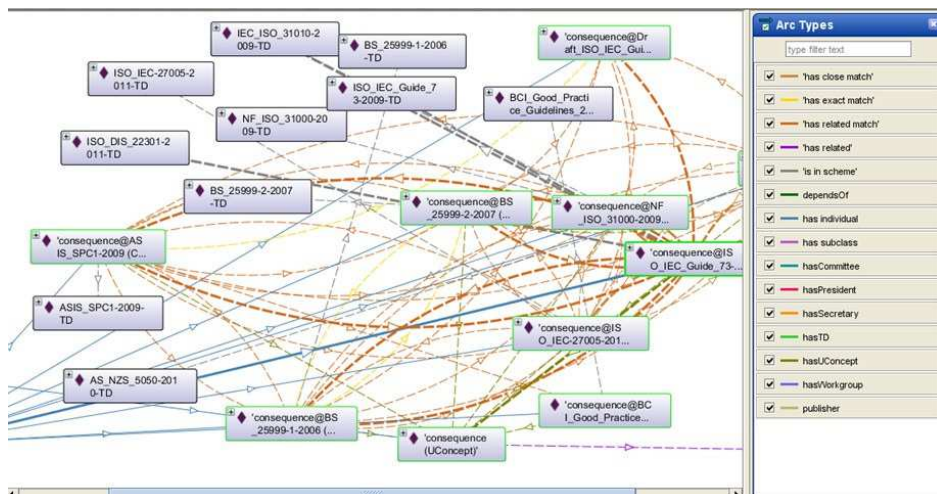


FIG. 6 – Visualisation de tous les termes ayant la graphie « consequence » et leurs relations.

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

6.2 Mode de représentation lié au cas d’usage par des « Explorateurs »

La plateforme KONTRAST modélise et représente, dans un même ensemble, deux types d’informations : les informations terminologiques relevées dans les normes du corpus et les informations organisationnelles concernant les instances de normalisation. Totalisant plus d’un millier d’individus (normes, concepts, institutions), cette plateforme constitue un outil puissant dédié à l’analyse de normes industrielles – dans le cas présent issues du domaine de la continuité d’activité.

Le mode de représentation (logiciel et graphique) lié au cas d’usage par un « explorateur » offre la possibilité de formuler des questions à partir d’hypothèses de regroupement, de recoupement ou de divergences entre des normes. Il repose sur l’hypothèse formulée dans le cadre du projet NOTSEG par [Juanals, 2011], selon laquelle il est possible de reconstituer les jeux d’influence entre les textes normatifs en comparant les textes des normes nationales avec le texte de la norme internationale éditée ou en cours d’édition. Dans une nouvelle norme, les réseaux de citations sont reconstitués par la détection des citations ou des reprises de définitions issues de normes ou de référentiels préexistants. Remonter le fil de ces héritages doit permettre d’identifier des familles de normes, et donc d’élucider les choix terminologiques ayant présidé à la rédaction des normes de notre corpus.

Nombre de termes de ce U_Concept	U_concept	
13	risk_assessment	risk_management-ASIS_SPC1-2009¶ Coordinated activities to direct and control an organization with regard to risk. [ISO IEC Guide 73:2002] - NOTE: Risk management generally includes risk assessment, risk treatment, risk acceptance, and risk communication. ¶
12	risk	
11	stakeholder	
10	risk_management	¶
9	risk_analysis	risk_management-AS_NZS_5050-2010¶
9	event	"Coordinated activities to direct and control an organization with regard to risk. [ISO Guide 73:2009, Risk Management—Vocabulary, definition 2.1.1]" ¶
9	business_impact_analysis	
9	consequence	¶
8	risk_treatment	risk_management-BS_25999-1-2006¶
8	resilience	structured development and application of management culture, policy, procedures and practices to the tasks of identifying, analyzing evaluating, and controlling responding to risk. ¶
8	incident	
7	residual_risk	
7	exercise	¶
7	top_management	risk_management-NF_ISO_31000-2009¶
7	risk_evaluation	coordinated activities to direct and control an organization with regard to risk (2.1). [ISO Guide 73:2009, definition 2.1] ¶
6	likelihood	

FIG. 7 – Exemple de résultat d’une requête SPARQL⁹

⁹ Le « mash-up » a été réalisé manuellement pour les besoins de l’article

Année	termes	definition
2005	resilience-ASIS_SPC1-2009	The adaptive capacity of an organization in a complex and changing environment.
2005	resilience-AS_NZS_5050-2010	Adaptive capacity of an organization in a complex and changing environment.
2005	resilience-BCI_Good_Practice_Guide	The ability of an organization to resist being affected by an incident.
2005	resilience-BS_25999-1-2006	ability of an organization to resist being affected by an incident
2005	resilience-BS_25999-2-2007	ability of an organization to resist being affected by an incident
2005	resilience-ISO_DIS_22300-2011	adaptive capacity of an organization in a complex and changing environment
2005	resilience-ISO_IEC_27031-2011	ability of an organization to resist being affected by disruptions
2005	resilience-ISO_IEC_Guide_73-2009	adaptive capacity of an organization in a complex and changing environment
2006	resilience-ASIS_SPC1-2009	The adaptive capacity of an organization in a complex and changing environment.
2006	resilience-BCI_Good_Practice_Guide	The ability of an organization to resist being affected by an incident.
2006	resilience-BS_25999-1-2006	ability of an organization to resist being affected by an incident
2006	resilience-BS_25999-2-2007	ability of an organization to resist being affected by an incident
2006	resilience-ISO_DIS_22300-2011	adaptive capacity of an organization in a complex and changing environment
2006	resilience-ISO_IEC_27031-2011	ability of an organization to resist being affected by disruptions
2006	resilience-ISO_IEC_Guide_73-2009	adaptive capacity of an organization in a complex and changing environment
2007	resilience-ASIS_SPC1-2009	The adaptive capacity of an organization in a complex and changing environment.
2007	resilience-BCI_Good_Practice_Guide	The ability of an organization to resist being affected by an incident.
2007	resilience-BS_25999-1-2006	ability of an organization to resist being affected by an incident
2007	resilience-BS_25999-2-2007	ability of an organization to resist being affected by an incident
2007	resilience-ISO_DIS_22300-2011	adaptive capacity of an organization in a complex and changing environment
2007	resilience-ISO_IEC_27031-2011	ability of an organization to resist being affected by disruptions
2007	resilience-ISO_IEC_Guide_73-2009	adaptive capacity of an organization in a complex and changing environment

FIG. 8 – Exemple de résultat d'une requête illustrant les variations pour le terme « Resilience »

L'exploration des relations représentées dans KONTRAST répond à ce besoin en permettant de mettre en évidence les chemins entre différents triplets du « triplestore ». Par exemple, prenons le cas d'un utilisateur qui cherche à tester l'hypothèse de l'influence d'une norme nationale sur une norme internationale. Il va donc rechercher des indices d'influence dans les textes. Dans cette perspective, il pourra rechercher et comparer des paires de concepts (l'un provenant d'une norme nationale et l'autre d'une norme internationale) ayant des définitions très similaires. Dans le cas évoqué, la dimension temporelle est un indice important puisque l'antériorité de la définition issue d'une norme nationale est indispensable pour pouvoir supposer une reprise possible dans la définition provenant d'une norme internationale. Cette question, formulée sous la forme d'une requête SPARQL, permet de retrouver des indices d'influence entre couples de normes [Lafréchoux, 2011].

7. Conclusion

Dans cet article, nous avons argumenté l'intérêt et décrit l'élaboration d'une méthodologie instrumentée qui exploite l'hétérogénéité des textes. Les outils de TAL et d'ingénierie des connaissances que nous avons décrits n'imposent pas de représentation unifiée des connaissances et des acteurs. Ils visent au contraire à conserver la diversité et la richesse des points de vue des textes tant nationaux

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

qu'internationaux¹⁰. Il s'agit d'en garder une représentation et d'en mesurer les évolutions dans la durée. Nous exploitons les possibilités offertes par les technologies numériques qui fournissent des outils de représentation iconiques et cartographiques et qui rendent possible l'appréhension et la traçabilité de cette diversité.

Les possibilités d'évolution de la plateforme KONTRAST sont nombreuses, notamment en termes d'automatisation de la chaîne de traitements. La relative rigidité de la langue des textes de normes permet d'imaginer la mise en place de différentes méthodes d'analyse au niveau syntaxique : extraction automatique de relations, de références, ou de thématiques, par exemple. D'autres caractéristiques para-textuelles peuvent aussi être exploitées automatiquement, par exemple la structure des normes ou les diagrammes qui les accompagnent. En termes d'usage, la création d'une interface utilisateur dédiée permettrait de rendre le système plus accessible aux utilisateurs de type « lecteur » ou « explorateur ». Nous souhaiterions notamment intégrer directement dans KONTRAST des liens avec les fichiers des textes de normes dans leur format d'origine (PDF), afin d'accompagner au mieux les investigations des lecteurs.

De manière plus large, la méthodologie et les outils que nous sommes en train de concevoir devront être utilisables dans d'autres domaines que la continuité d'activité, qui a constitué le corpus de cette étude.

Références

- Adam J.-M. [1999] (2011), *La linguistique textuelle: Introduction à l'analyse textuelle des discours*, Paris, Armand Colin, 3^e édition.
- Bazet I., Jolivet A., Mayère A., « Pour une approche communicationnelle du travail d'organisation: changement organisationnel et gestion des événements indésirables ». *Communication & Organisation*, n° 33.
- Callon M., (1986). «Éléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc», *L'Année sociologique*, n°36, pp.169-208
- Carroll, L., (1872). *Through the Looking-Glass*, Macmillan et Co, Londres, UK.
- Charlet, J., Baneyx, A., Steichen, O., Alecu, I., Daniel-Le Bozec, C., Bousquet, C., Jaulent, M.-C., « Utiliser et construire des ontologies en médecine, Le primat de la terminologie. », *Technique et science informatiques*, vol. 28, n°2/2009, pp. 145-171.

¹⁰ Ces derniers peuvent varier en fonction des comités techniques de normalisation.

- Chateauraynaud, F. (2011). *Argumenter dans un champ de forces. Essai de balistique sociologique*, Paris, Editions Pétra.
- Condamines, A. (ed). (2005), *Sémantique et corpus*, Paris, Hermès Science.
- Damas L., et Tricot C. (2010). L'ontoterminologie pour la recherche d'information sémantique, Actes de la conférence TOTh 2010, C. Roche éd., Annecy, Institut Porphyre, pp. 101-116.
- Depecker L. (2005). Contribution de la terminologie à la linguistique. *Langages*, 157(1), L. Depecker éd., pp. 6-13
- Gresser J.-Y. (2010). Ontologies des risques financiers — Continuité d'activité, gestion de crise, protection des infrastructures critiques financières, Actes de la conférence TOTh 2010, C. Roche éd., Annecy, Institut Porphyre, pp. 155-176.
- Juanals, B., Minel, J.L. (2011). Writing and Monitoring in International Standardization, Theoretical Choices and Methodological Tools, Proceedings of IMET2011, Florida, USA.
- Juanals, B., Minel, J.L. (2012). Monitoring the Standardization Writing Process, Theoretical Choices and Methodological Tools, (accepted), 12 p., *Journal of Systemics, Cybernetics and Informatics*, USA
- Juanals, B. (2012a). *Les acteurs et les textes de la normalisation internationale de la sécurité sociétale. Construction d'une approche communicationnelle*. Soumis à Sciences de la Société, Montréal.
- Juanals, B. (2012b) « Pour une approche communicationnelle des normes internationales de management de la sécurité sociétale », colloque international CMN2102, Communiquer dans un monde de normes, L'information et la communication dans les enjeux contemporains de la mondialisation, Roubaix (France).
- Juanals B., (2008), *La circulation médiatique des savoirs dans les sociétés contemporaines*, habilitation à diriger des recherches en Sciences de l'information et de la communication, Université Paris Diderot.
- Latour B. (2006), *Changer de société. Refaire de la sociologie* (Titre original : *Re-assembling The Social. An Introduction To Actor-Network Theory*, 2005), Paris, Ed. La Découverte.
- Lelong B. et Mallard A. (coord.) (2000), *La fabrication des normes*, revue Réseaux n° 102, France Télécom R&D/Hermès Sciences Publications, 2000.
- Lafréchoux, M., Juanals, B., Minel J.L. (2012a). *KONTRAST : création d'un glossaire contrastif à partir d'un corpus de normes internationales*. Soumis à JADT 2012, Liège Belgique.

KONTRAST, une méthodologie instrumentée pour analyser les variations terminologiques dans des textes normatifs

Lafréchoux, M., Juanals, B., Minel J.L. (2012b). *KONTRAST: a Terminological Ontological Glossary to Exploit Linguistic Variants*. Soumis à TKE 2012, Madrid, Espagne.

Maingueneau D. (1996), *Les termes clés de l'analyse du discours*, Paris, Seuil.

Malik, M., *Construction d'un glossaire contrastif*, Livrable ANR-NOTSEG (2011)

Omrane N., Nazarenko A. et Rosina P. (2011). *Lexicalized ontology for a business rules management platform: An automotive use case*, Proceedings of the 5th International Symposium on Rules (RuleML, Industry focused session), Fort Lauderdale, Florida

Roche C. (2007). *Le terme et le concept : fondements d'une ontoterminologie*, *Actes de la conférence TOTH 2007*, C. Roche éd., Annecy, Institut Porphyre, pp. 1-22

Summary

This paper describes the design of an instrumented method, linking up several steps, which aims to analyze terminological variants into technical texts. We try to identify, in texts of standards, the origin and the means of construction of concepts, as well as the diachronic variants of terms which lexicalize them. Theoretical framework presentation is followed by the architecture description of KONTRAST, and the associated workflow. Finally, several representations of using cases show the possibilities of linguistic and socio-pragmatic analysis of this method. An experimentation lead in the field of business continuity illustrates the whole argumentation.