

Qu'est-ce qu'un tag? Entre accès et libellés, l'esquisse d'une caractérisation.

Alexandre Monnin

► **To cite this version:**

Alexandre Monnin. Qu'est-ce qu'un tag? Entre accès et libellés, l'esquisse d'une caractérisation. : Entre accès et libellés, l'esquisse d'une caractérisation.. IC 2009 : 20es Journées Francophones d'Ingénierie des Connaissances, " Connaissance et communautés en ligne ", May 2009, Hamamet, Tunisie. pp.41, 2009. <hal-00377410>

HAL Id: hal-00377410

<https://hal-paris1.archives-ouvertes.fr/hal-00377410>

Submitted on 21 Apr 2009

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.

Qu'est-ce qu'un tag ? Entre accès et libellés, l'esquisse d'une caractérisation.

Alexandre Monnin¹

¹Laboratoire EXeCO, Université Paris I Panthéon - Sorbonne,
Alexandre.Monnin@malix.univ-paris1.fr

Résumé : Nous esquissons ici un essai de caractérisation des tags en les distinguant rigoureusement de toutes les expressions linguistiques dont ils sont potentiellement le support via l'analogie avec les étiquettes matérielles. Celle-ci nous permet de mettre en évidence deux types de relations : les relations d'accès, d'ordre technique, et de référence, langagières. Le recours à la sémantique formelle conduit à distinguer, de ce point de vue, mots, mots clef, descripteurs et vedettes matières. La conclusion, dans tous les cas, est l'absence de sémantique propre au tag. Enfin, dans la lignée de ce qui précède, nous soulignons l'indétermination du rapport à la ressource, tant du point de vue de l'accès que de la référence.

Mots-clés : Tag, Tagging, Folksonomies, Référence, Accès, Langage, Web Social.

1 Introduction.

Dans sa définition initiale des folksonomies, Thomas Vander Wal¹ soulignait leur complète dépendance vis-à-vis des tags, notant par-là même leur présence comme en échos lointain au fait qu'il fallut attendre prêt de deux ans après la création de Muxway, l'ancêtre de del.icio.us, pour qu'émergeât une appellation correspondant aux résultats socialement partagés de la pratique du tagging. Pourtant, en dépit de cette primauté unanimement admise, force est de constater la relative absence de toute caractérisation précise des tags, les chercheurs ayant souvent préféré étudier les enjeux afférant aux folksonomies. Sans doute faut-il lire dans ce diagnostic l'effet d'une saturation due au vocabulaire existant, le recours incessant à la notion pour le moins confuse et protéiforme de « mot clef » ayant achevé d'obscurcir les discussions autour du tagging en y associant pêle-mêle les balises <meta> des pages HTML, les requêtes en langage naturel² formulées par l'entremise des moteurs de recherche ou encore les langages documentaires et leurs multiples déclinaisons lexicales : mots

¹ Cf. Vander Wal, T. (2007), Folksonomy Coinage and Definition, *vanderwal.net* : « *Folksonomy is the result of personal free tagging of information and objects (...) for one's own retrieval. The tagging is done in a social environment (usually [Ndr : je souligne] shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information* ». On notera que l'aspect collaboratif n'y est pas explicitement posé comme une condition nécessaire à la constitution d'une folksonomie.

² Nous préférons parler de « langage naturel » plus que de langue dans la mesure où la notion de langue ressortit davantage à la linguistique qu'à la philosophie (et, *a fortiori*, qu'à la philosophie du langage).

clef, nous l'avons dit, mais aussi vedettes matière ou descripteurs. Esquisser une caractérisation des tags suppose de prendre en compte le contexte technique les ayant vu naître, avant tout lié au Web et à ses technologies en constante évolution, de même que les multiples déclinaisons auxquelles cette dynamique a donné naissance. C'est à la seule condition d'accorder suffisamment d'attention à ce milieu technique qu'il sera possible de dégager, par contraste, la part proprement « symbolique »³ des tags, pour enfin penser l'entrecroisement de ces deux dimensions.

2. Le tag : étiquette matérielle et libellé, entre accès et référence.

2.1. Les trois dimensions canoniques.

S'il ne faut pas chercher de définition canonique précise du tag, paradoxalement, du fait du rapprochement opéré de bonne heure⁴ entre folksonomies et ontologies⁵, longtemps perçues, de prime abord dans un rapport exclusif de pure et simple opposition, nous disposons d'ontologies susceptibles de nous éclairer quant à ses propriétés constitutives. A cet égard, un quasi consensus se dégage de la littérature dévolue à cette question. Une ontologie du tag comme celle de Richard Newman par exemple (c'est également vrai de la *TagOntology* de Thomas Gruber⁶) se propose avant tout de décrire un processus individuel de tagging en opérant une distinction entre :

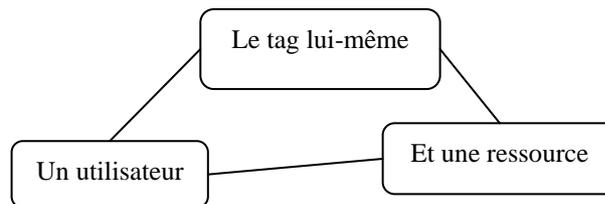


Fig. 1 – Les trois axes de la définition/réification courante du tag.

³ Nous entendons ici *symboliques* au sens large, reprenant à notre compte la problématique ouverte par le philosophe de la technique Gilbert Hottois qui oppose symbolicité et technique, sans toutefois nier l'importance de l'activité symbolique pour les technosciences. Celle-ci se situe cependant sur un autre plan selon Hottois qui précise, avec des échos annonciateur du Web Sémantique : « Le langage des technosciences se veut purement objectif. En termes logiques, ce devrait être un langage purement extensionnel ou référentiel, dépourvu de toute étendue de sens. Un langage qui étiquette le réel afin d'offrir des prises et de permettre l'organisation collective des relations opératoires [Ndr : je souligne] – techniques, mathématiques – au réel. Un tel étiquetage est conventionnel et instrumental : ses mots n'apportent rien – ni sens ni supplément de maîtrise – à la maîtrise technomathématique des objets, des opérations et des processus qu'ils désignent. » Cf. Hottois (1996), *Entre symboles et technosciences*, p. 91, Champ Vallon. Aussi, et bien que nous nous penchions davantage sur les premiers, la relation symbolique ici envisagée n'exclut-elle, de notre point de vue, ni les signifiants linguistiques, ni les représentations iconiques. Nous réservons pour de futurs travaux des développements plus précis sur ce point.

⁴ (NEWMAN 2005).

⁵ Nous ne posons pas de coupure radicale entre ontologies philosophiques et informatiques pour de multiples raisons. Il suffira de n'en retenir ici qu'une seule : si les ontologies informatiques prétendent modéliser, et non décrire de façon définitive, la réalité, nombreux sont les philosophes qui conçoivent l'ontologie sous pareil angle déflationniste (à ceci près que leurs modèles sont langagiers et non artefactuels, et leur méthodes d'ordre logique plus qu'épistémique). Cf. (MONNIN & FELIX 2009).

⁶ (GRUBER 2005).

Une telle tripartition, à laquelle on serait bien en peine de chercher des alternatives radicales dans la littérature dévolue à ces questions, pourrait cependant sembler limitée en ce qu'elle oblitère la nature duale du tag : à la fois instance *matérielle* (à l'instar de l'étiquette concrète à laquelle son nom est attaché⁷) mais également *symbolique*. Associer ces deux aspects c'est oublier que le lien symbolique usuel entre mots et choses ne nécessite aucunement de se voir implémenté d'une quelconque manière. Nul besoin d'avoir recours à des moyens d'ordre techniques pour qu'un mot atteigne son objet, aucun artefact n'y pourvoira ; autrement dit, la référence ressortit à la seule sémantique.

À l'inverse, chaque site qui emploie des folksonomies définit, selon ses besoins propres, les règles encadrant le tagging (qui a le droit de tagger ? quoi ?, comment ?, etc.), complétant, *de facto*, la relation de *référence* par une relation associant matériellement (la dépendance de cette relation vis-à-vis d'un réseau informationnel physique en atteste) le tag à une ressource, fondée sur la notion d'*accès*⁸. Ses tenants et aboutissants sont à chercher du côté du design informationnel des interfaces et de la réalité technique des réseaux, en particulier l'architecture du Web, et non plus simplement de l'analyse du langage.

Or, l'absence de la dichotomie référence/accès au cœur de l'ontologie de Newman, ne va pas sans entraîner de sérieuses conséquences. Au premier rang desquelles, celle-ci : une telle ontologie vaut autant par ce qu'elle précise et explicite (d'où son incorporation à d'autres ontologies plus vastes à l'instar de SIOC⁹) que pour ce qu'elle effusque et qu'il nous semble impératif de restituer.

Une vision du tag tournant invariablement autour de trois axes, sans que la relativité des interfaces ne modifie cette donnée essentielle, apparaîtra pour le moins discutable du point de vue de la *description*. En revanche, une fois implémentée, elle fournira un support adéquat pour réaliser (*prescrire*) une interopérabilité entre services utilisant le tagging, implémentant, par ce fait même, une définition unifiée du tag indépendamment de toute autre considération.

2.2. Du tag au machine-tag : entre accès et libellé.

La relation d'accès diffère de la relation de référence en ceci qu'elle est indissolublement d'ordre causal, et par conséquent matériel, et relie, par une relation d'ostension, l'utilisateur à la chose taggée. L'ostension, comme n'ont cessé de le

⁷ Un site pionnier du point de vue de l'utilisation des tags en France, Babelio.fr, traduit expressément « tag » par « étiquette ».

⁸ (HAYES 2006) souligne avec force la nécessité de bien dissocier chacune de ces deux dimensions dans sa discussion du statut des URIs. Il nous semble essentiel d'intégrer ces distinctions dans la perspective qui est la nôtre, à savoir proposer une caractérisation des tags susceptible de restituer ces deux aspects ; d'une part le mot ou la suite de caractères, éventuellement doté d'une signification ; d'autre part le tag, étiquette « matérielle » ménageant un accès à la ressource encadré par des limitations très précises, imputables au système informatique en place (qu'il s'agisse du Web ou d'un logiciel comme *Photo Gallery* pour l'OS Vista de Microsoft – cf. infra). La question du rapport entre URI et ressource ayant fait l'objet de nombreuses discussions, il apparaît utile de s'en inspirer pour penser les rapports – en particulier d'accès ou de référence – entre le tag, la ressource (l'objet au sens large) et la ressource informationnelle.

⁹ Semantically-Interlinked Online Communities, <http://sioc-project.org/>

montrer les philosophes, en particulier depuis Wittgenstein¹⁰, se caractérise par un rapport d'indétermination intrinsèque à l'objet, ou, pour préciser les choses d'une manière plus conforme aux usages et à la réalité technologique du Web, à la ressource ainsi désignée.

Le libellé n'est rien d'autre, quant à lui, que la suite de caractères inscrite à *même le tag*, lui-même conçu à la manière d'une étiquette. En ménageant un accès à la ressource (informationnelle ou non), cette étiquette permet à l'utilisateur d'associer à celle-ci le texte¹¹ qu'il désire. Il devient dès lors loisible d'indexer, d'évaluer, de partager ou encore de retrouver des objets qui échappaient jusqu'alors à ces possibilités d'annotations. Précisément ce que permettait depuis longtemps, dans l'univers analogique, le traditionnel post-it : produire une surface matérielle accueillant du texte là où celle-ci faisait défaut. Illustration de ce constat, l'application *Lignes de temps*, développée au sein de l'IRI¹², autorise un accès technique à des séquences filmiques qu'il devient possible d'annoter (à l'instar, sur un versant cette fois-ci collaboratif, de la toute récente *VideoTagGame* de Yahoo!¹³).

Ultime avatar de cette logique de mise à disposition de nouveaux supports *ad hoc* d'inscription des métadonnées, l'application *Gallery* de Windows Vista¹⁴, permettant de tagger ses photos localement, propose non plus d'accoler une étiquette à une ressource numérique mais de l'y injecter directement. A la différence du post-it, souvent appelé à jouer le rôle de pense-bête mais dont la perte menace d'entraîner avec elle celle de l'objet étiqueté, *Gallery*, en conservant la trace de tous les tags présents dans le système, garantit un accès pérenne à l'ensemble des ressources taggées. La logique n'est plus celle du raccourci, susceptible de pointer dans le vide pour peu que le chemin d'accès vers la ressource visée se soit modifié, mais de *l'incorporation* au système des images, dès lors accessibles au même titre que les tags eux-mêmes ; tags et ressources formant un nouvel ensemble disposant de ses caractéristiques propres, immunisé contre la perte¹⁵.

Rien, et l'analogie avec les étiquettes matérielles le confirme, ne contraint l'utilisateur à inscrire sur le tag une chaîne de caractère formant un mot – sans parler de sa forme lemmatisée, et ce, en dépit des contraintes syntaxiques minimales qu'imposent les différents systèmes existants¹⁶. Il appert de ce constat l'impossibilité d'assimiler les tags à quelque formes linguistiques précises que ce soit ; d'où une réalité incomparablement plus complexe que ne le laissait percevoir le précédent

¹⁰ Wittgenstein, L., (1953), *Recherches Philosophiques*, Gallimard.

¹¹ Nous entendons le mot « texte » dans une acception élargie incluant tous types de signes ou de codes imaginables.

¹² <http://www.iri.centrepompidou.fr/>

¹³ <http://sandbox.yahoo.com/VideoTagGame/>

¹⁴ Gene Smith présente le point de vue des concepteurs de cette application recueilli au cours d'entretiens, cf. Smith, G. (2008), *Tagging : People-Powered Metadata for the Social Web*, 1st ed., New Riders.

¹⁵ D'où les affinités que de telles application entretiennent vis-à-vis de la problématique dite du PIM (*Personal information management*).

¹⁶ DeLicio.us sépare les tags par des espaces ce qui oblige à former des phrases ou des expressions composées à l'aide, entre autres, de tirets ou de barres d'espacement, là où d'autre système séparent les tags par de simples virgules voire créent pour chaque tag un formulaire unique laissant toute latitude aux usagers dans le choix de leurs expressions (Zotero).

schéma à la structure ternaire où l'entité « tag », pour réifiée qu'elle fût en ses relations avec un *utilisateur* et une *ressource*, n'était paradoxalement pas interrogée, bien que sertie d'attributs destinés à la rendre manipulable¹⁷. Ce faisant, la figure unitaire du tag cède la place à une entité biface, à la fois réalité matérielle attachée à une ressource de par leur insertion à toutes deux dans un système technique donné, et relation sémiotique et langagières portée par des items de natures variables.

Avec, au surplus, et du fait de cette absence de contrainte qui constitue l'essence même du tagging, un mélange éventuel des deux : ce sont les fameux *triple tags* ou *machine-tags*, popularisés par Flickr¹⁸. Structuration légère des contenus produits par les utilisateurs, ils articulent trois dimensions : un espace de nom, un prédicat et une valeur associée.

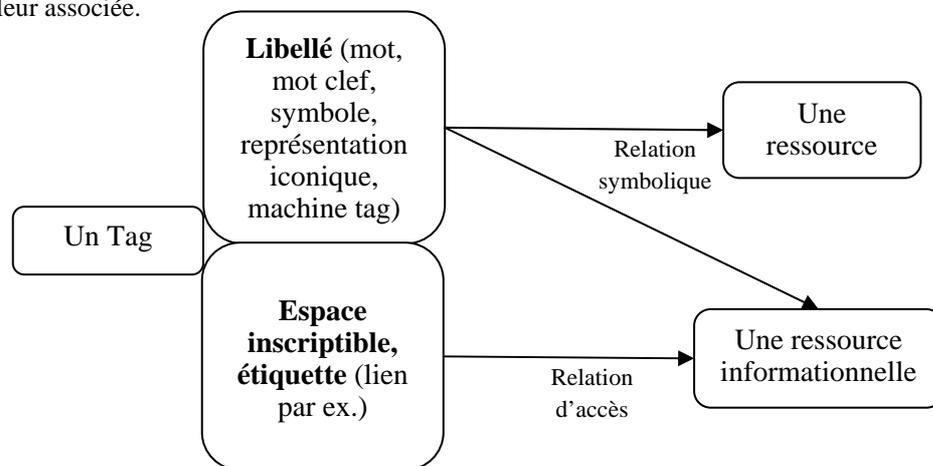


Fig. 2 – La bipartition du tag.

namespace:predicate=value¹⁹

¹⁷ Un relecteur anonyme de cet article écrit : « il s'agit d'une réification de l'acte de tagging qui devient référençable et donc annotable par des relations (ex : tagging#1 est intervenu avant tagging#2) et des attributs (ex : a pour ressource, a pour tag, a pour auteur, a pour date, a pour statut, a pour langue, a pour portée, etc.) ». Nous le suivons sur ce point tout en notant avec lui que le problème de la définition du tag reste entier. Ajoutons également qu'une ontologie du tag se doit sans doute de préciser ses liens avec une ontologie de la ressource dans la lignée des débats sur les URIs auquel il a déjà été fait allusion. Cf. en particulier le travail en cours de Valentina Presutti, Harry Halpin et Aldo Gangemi sur l'ontologie de la ressource (<http://ontologydesignpatterns.org/ont/web/irw.owl>).

¹⁸ On trouve aussi les « dctags », pour Dublin Core tags, cf. Johnston, P. (2006), eFoundations : dctagging, *eFoundations* (<http://efoundations.typepad.com/efoundations/2006/10/dctagged.html>),

Johnston, P. (2007a), dctagging revisited, *eFoundations* (<http://efoundations.typepad.com/efoundations/2007/09/dctagging-revis.html>),

Johnston, P. (2007b), Flickr Machine Tags and API changes, *eFoundations*. (http://efoundations.typepad.com/efoundations/2007/01/flickr_machine_.html).

¹⁹ Le choix de chaque facette est laissé à la discrétion des utilisateurs. Les machine-tags employés en liaison avec des services en ligne (le premier d'entre eux fut geo.licio.us), sont interprétés de manières spécifique par ces applications et les libellés choisis assimilable à du code informatique. Les *triple tags*, dont ils s'inspirent, matérialisent une idée lancée ultérieurement sous forme de boutade, celle d'employer des tags

En scindant l'information de la sorte, et en la répartissant conséquemment au sein d'une base de données spécifiquement dédiée à chacune de ces trois facettes, il devient possible de poser des requêtes sur un espace de nom donné, un prédicat, une valeur, ou l'une quelconque des combinaisons associant ces trois aspects.

Ces tags représentent un cas limite qui illustre néanmoins la dualité accès/référence inhérente au modèle proposée dans la figure n°2. Historiquement, deux cas de figure se présentent. Dans le premier, le machine-tag, baptisé *triple tag* à l'origine, est utilisé en guise de libellé : sa syntaxe, calquée sur les langages de type XML et influencée par le développement des micro-formats et de RDF, permet d'exprimer des relations complexes à l'aide d'un seul et unique tag. Celui-ci, une fois inséré dans une URL, donne alors accès à une ressource quelconque²⁰.

Dans le second cas de figure, le machine-tag n'est associé à aucune URL en ce qu'il se suffit à lui-même, en vertu de sa syntaxe interprétable par les machines, pour ménager un accès à la ressource numérique par l'intermédiaire des requêtes lancées via les API des sites où furent développées des applications susceptibles de traiter ce type d'informations (c'est la cas de Flickr qui, sous l'impulsion des utilisateurs qui usaient déjà des machine-tags en guise de simples libellés, ajouta à son API de nouvelles fonctionnalités permettant de les parser, *reconduisant ainsi du même coup la dichotomie accès/référence au niveau du libellé lui-même*)²¹.

Aucune règle expresse ne commandant l'inscription d'un mot ou d'un signe humainement compréhensible, rien n'interdit non plus le recours à du code informatique en guise de libellé. Mieux, une vague syntaxe que des utilisateurs reconstruisaient aisément sous forme d'énoncés du langage naturel donna naissance à des micro-formats à partir du moment où les machines furent reprogrammées pour

pour tagger d'autres tags, des « métatags » en somme, auxquels correspondent les facettes d'un triple tag. Cf. (MONNIN 2009) pour un rapide historique de la chose.

²⁰ A titre d'exemple, del.icio.us a employé cette convention pour signaler des liens pointant vers des ressources multimédia (images, vidéos), voire documentaires au sens plus traditionnel du terme (fichiers .doc ou PDF) – les valeurs des prédicats en question ne se limitaient cependant pas à de simples extensions de fichiers : les mashups réalisés à partir de mp3 étaient par exemple désignés par : « system:filetype:mp3+mashups ».

Cf. Schachter, J. (2005), casting the net wider, *delicious blog* (http://blog.delicious.com/blog/2005/06/casting_the_net.html).

²¹ Le fait que les utilisateurs de Flickr, précurseurs en la matière, qui comptaient parmi leurs tags des triple tags, durent les ajouter à nouveau au système pour que celui-ci fût en mesure de les interpréter comme des machines tags est significatif à cet égard, cf. Catt, D. (2007), Flickr Ramps up Triple Tag (Machine Tags) Support., *geobloggers.com* (<http://geobloggers.com/2007/01/24/offtopic-ish-flickr-ramps-up-triple-tag-support/>). Les machines-tags furent essentiellement conçus, dans un premier temps, pour lancer des recherches via une API. Comme l'écrivait Aaron Straup Cope en charge de cette question chez Flickr au moment de leur lancement, « For the moment, *machine tags* are principally an API "thing" ».

(<http://www.flickr.com/groups/api/discuss/72157594497877875/>). Cf. aussi sur le même sujet Keith, J. (2007), something:somethingelse=somethingspecific, *echoloquation* (<http://echoloquation.com/post/6939803/something-somethingelse-somethingspecific>).

On notera toutefois qu'ils sont toujours utilisés comme des triple tags insérés dans des URLs. Pour une chronologie inverse, « top down », voir les tags « for:username » permettant d'envoyer un lien à l'utilisateur de son choix, que del.icio.us mit en place en juillet 2005, cf. Schachter, J. (2005a), tags for two, *delicious blog*. (http://blog.delicious.com/blog/2005/07/tags_for_two.html) et la critique de Knight, J. (2005), del.icio.us : tags for two, *del.icio.us: tags for two* (<http://jk3.us/2005/07/09/delicious-tags-for-two/>).

l'« interpréter » de manière idoine. Bien qu'extrêmement sommaire, de pseudo-syntaxe calquée sur des langages informatiques existants, elle acquit *ipso facto* le statut parallèle de syntaxe informatique de plein droit. Sous cet angle, les machine-tags résultent bien d'un développement logique qui tire partie de la caractéristique définitionnelle fondamentale des tags.

3. Quelle sémantique pour les libellés ?

D'ordinaire, les tags sont assimilés à des mots clefs sans autre forme de procès. Or, il faut rappeler ici les conditions d'usage précises des notions, nombreuses, issues de la bibliothéconomie, auxquelles ils ont été comparés. Sans cela, le risque serait grand d'obscurcir tout discours visant à cerner avec précision la sémantique non des tags eux-mêmes mais des libellés²². La pratique documentaire usuelle présente des définitions très précises, examinons-en quelques-unes.

3.1. Les descripteurs.

La norme internationale ISO 2788, déjà citée par Manuel Zacklad²³, définit le thésaurus comme le « *vocabulaire d'un langage d'indexation contrôlé organisé formellement de façon à expliciter les relations a priori entre les notions (par exemple relation générique-spécifique)* ». Première conséquence immédiate de cette définition, si le but de l'indexation demeure de rendre compte du contenu d'un document, l'absence de relation référentielle confère à l'indexeur la possibilité de recourir à un nombre indéfini de descripteurs. Les descripteurs ne sauraient, dans ces conditions, et de manière univoque, correspondre à des sujets. Ceux-ci se laissent en revanche appréhender au moyen d'un faisceau de notions, toutes puisés dans le vocabulaire d'un thésaurus.

Celui-ci résulte de l'articulation d'un réseau de relations limitées entre les descripteurs qui le composent (relations de généralité et de spécificité ou d'association pour ce qui est des principales). Le choix de chacun d'entre eux présupposant que soient spécifiés les rapports qu'il entretient avec les autres, celui-ci s'effectue soigneusement en amont²⁴. Un descripteur n'existe par conséquent, qu'imbriqué dans un vocabulaire contrôlé définissant avec précision la place revenant à chaque terme. Sous cet angle, la signification des descripteurs se conçoit bien davantage sous un angle inférentiel, nourri par différentes relations de sens internes au lexique, telles l'hyponymie, l'hyperonymie ou la synonymie, que référentiel.

²² Sur la page de discussion consacrée à l'article « Tag (metadata) » de Wikipedia, Joshua Schacter, créateur de Muxway en 2002 et del.icio.us un an plus tard, précise les raisons qui l'ont poussé à distinguer les tags des mots clef et à forger, pour ce faire, un terme nouveau: « *While keywords are not new, I believe that tagging is a larger concept than just assigning keywords to things, however - I feel that it also includes the retrieval of the set of used terms/keywords/whatever upon view of the items. Additionally, I am reasonably sure that I named this.* »

http://en.wikipedia.org/w/index.php?title=Talk:Tag_%28metadata%29&action=edit§ion=14

²³ (ZACKLAD 2007).

²⁴ Quant aux termes non retenus, les « non-descripteurs », les requêtes les prenant pour point de départ sont basculées sur des termes du lexique sélectionnés pour jouer le rôle de descripteurs.

Certes, on objectera qu'à chaque terme est corrélié un concept qui en fournit la dénotation. Pourtant, comme le souligne Manuel Zacklad à nouveau, « les thésaurus s'appuient sur une caractérisation des concepts qui les font au moins pour partie dépendre des langues et des mises en discours ». Cette « dépendance du concept vis-à-vis du système de la langue et la variabilité intrinsèque induite par cette dépendance » conduisent à nuancer la portée référentielle concédée aux descripteurs pour la simple et bonne raison que les deux niveaux, conceptuels et langagiers, ne se conçoivent pas l'un sans l'autre, ou dans un rapport classique de représentation supposant l'autonomie revenant en propre à des domaines hétérogènes²⁵. Ils s'apparentent au contraire à l'unique face d'un anneau de Moebius, là où l'ontologie s'attache au contraire à garantir une caractérisation relativement indépendante du niveau conceptuel, outillant et opérationnalisant la rupture avec son ancrage originaire dans la langue vernaculaire.

3.2. Les vedettes matières.

Une manière de désigner une vedette-matière consiste à parler de sujets, traduction de l'anglais *subject heading*. Un sujet n'est pas un simple mot. Entité lexicale complexe extraite d'un langage documentaire artificiel, et non du langage naturel, il s'agit avant tout d'un syntagme résultant de la coordination de plusieurs descripteurs. L'écart entre langage naturel et langages artificiels de la bibliothéconomie se traduit, au plan sémantique et référentiel, par une hétérogénéité des modèles²⁶ servant d'appui à une appréhension de la signification des entités lexicales en jeu. Comme le note Helen Svevoni²⁷, traditionnellement, l'extension d'un mot fait référence à la classe des entités dénotée par ce mot ; tout l'enjeu d'une sémantique formelle, dans la lignée de ce qu'Emon Bach²⁸ a nommée « métaphysique du langage naturel », étant de spécifier le statut desdites entités. A l'inverse, la référence d'un syntagme tenant lieu de vedette matière ne s'entendra qu'en liaison avec un modèle divergent. Aussi un sujet, dans un contexte documentaire, fait-il expressément référence à une classe de documents portant sur un même contenu, à savoir le *thème* qui les unit²⁹. Au-delà de

²⁵ Une manière de nuancer cette affirmation serait de souligner la mise en concurrence des termes lorsqu'il s'agit d'en sélectionner un pour *représenter* un concept. Comme le souligne Zacklad toujours, « les concepteurs d'un thésaurus vont se fixer sur une expression linguistique, le descripteur, et le considérer « toute chose étant égales par ailleurs » comme le meilleur représentant du concept visé. » Toutefois, les équivalences ainsi dégagées, l'étant sur la base de relations de synonymie ou de traduction (dans une optique respectivement intra ou inter-linguistique), vis-à-vis desquelles la philosophie a formulé de nombreux avertissements, elles relèvent avant tout du *postulat de signification*. L'exemple fameux de Quine le montre, « célibataires » et « non-mariés », bien que couramment tenus pour synonymes, ne le sont pas en vérité, ce que seul un souci pour une caractérisation précise des extensions des prédicats « célibataires » et « non-mariés » met en relief, et que recouvre, à l'inverse, la postulation d'une équivalence allant de soi entre intensions.

²⁶ Un « modèle », au sens de la théorie des modèles (et de sa reprise par les linguistes dans le sillage des travaux pionniers de Richard Montague).

²⁷ (SVEVONIUS 2000), chapitre 8.

²⁸ (BACH 1989).

²⁹ La norme AFNOR NF Z 44-070 précise ceci : « Chaque vedette-matière correspond à un seul sujet, simple ou complexe. Un même document peut avoir plusieurs sujets donnant lieu à la rédaction de plusieurs vedettes-matières ».

la difficulté notoire grevant toute tentative visant à dégager des significations lexicales dans le cadre d'une sémantique formelle, toute appréhension d'un langage quelconque prétendant lui adjoindre une interprétation articulée sur un modèle unique éclaire davantage les conditions de production et les choix ayant présidé à l'élaboration de ce langage qu'elle ne tient lieu de vérité logique. Les tags, en revanche, s'abstraient aisément de cette contrainte. Il suffit de mentionner l'existence des « *to-do* » tags, indiquant des actions à accomplir en liaison avec une ressource sans qu'ils n'en reflètent aucunement le contenu, pour que s'évanouisse toute tentative de les assimiler purement et simplement à des vedettes matières.

3.3. Les mots clef.

Il est difficile, voire impossible, de s'accorder aujourd'hui sur une définition consensuelle du mot clef, celui-ci connaissant un extraordinaire succès depuis son emploi pour signifier des requête effectuées (en plein texte) via des moteurs de recherche, jusqu'à sa caractérisation par les normes documentaires. C'est à cet usage bien établi, et à lui seul, que nous nous référerons pour l'occasion. « Mot ou groupe de mots choisi soit dans le titre ou le texte d'un document, soit dans une demande de recherche documentaire, pour en caractériser le contenu », selon la définition de la norme AFNOR NF Z 47-102, le mot clef, bien qu'issu du langage naturel, se conçoit la plupart du temps comme extrait directement d'un document analysé. Or, dans le cas précis qui nous occupe, les types de documents soumis au tagging varient dans des proportions suffisamment importantes pour qu'il ne soit tout simplement pas possible de définir les libellés uniquement en ces termes : qu'ont de commun, en effet, une photo, un événement, un plan séquence ou un enregistrement audio ? S'agissant de ce qui retient ici notre attention, leur nature de document non-textuel n'offre guère de prise à l'extraction directe de mot clef. Inversement, la force du tagging, face à ce type d'objets, est précisément de nous permettre d'ajouter du texte (entendu au sens large, cf. *supra*). A une logique d'extraction à laquelle se prêtent tout particulièrement les documents de nature textuelle, succède une autre logique, *expressive* et non seulement descriptive, que sont incapables de capter les applications proposant des « tags » générés automatiquement. Tagger c'est aussi ajouter un contenu absent d'un document (ou d'une ressource) ; en d'autres termes, lui adjoindre un contenu *extrinsèque*.

3.4. Le libellés : un espace vide et non un mot clef, un descripteur ou une vedette-matière.

Le contraste le plus saisissant du point de vue d'un indexeur professionnel au vu des normes encadrant sa pratique, fut certainement le passage d'une indexation contrôlée, *a priori*, à des formes plus libres. En l'absence d'élaboration préalable de vocabulaires contrôlés et sans préjuger des tentatives en cours pour produire des ontologies à partir de l'effort collectif de constitution d'une folksonomie, il convient de tenir les libellés des tags non pour des termes, des sujets, voire de simples mots (clefs ou non) mais tout cela à la fois – parmi une kyrielle d'autres choses encore. Au

nombre desquelles on pourra mentionner, en guise de chaînons de caractère faisant office de libellés : les machines-tags, les URL, les émoticônes et autres représentations iconiques, les signes à la signification plus ou moins idiosyncrasique, etc. Rien n'interdit en effet de doter un simple signe tel @, par exemple, d'une signification idiosyncrasique (idiolecte) affirmée en dehors de tout système socialement constitué de la langue (sociolecte). Quant aux machine-tags, ils constituent un cas limite où du code informatique, venant nourrir des applications en ligne pour la constitution de nouveaux services, fournit une information compréhensible (au moins dans certains cas) par des êtres humains, du fait d'une syntaxe ressortissant aux langages artificiels tout en s'avérant pour partie copiée sur le langage naturel.

Si l'on s'accorde à souligner l'extrême liberté avec laquelle il devient possible de choisir ses libellés, liberté expliquant en grande partie le succès du tagging, reste alors à en tirer l'unique conclusion qui s'impose. Contrairement aux vedettes matières ou aux descripteurs dont la sémantique est attachée, d'une manière contrainte, soit à un modèle spécifique, soit à un lexique intégralement ordonné par des relations de sens en vue d'éliminer toute ambiguïté, les libellés inscrits sur les tags sont susceptibles d'accueillir des entités contrastées, linguistiques ou non, interdisant *de facto* une intelligence globale de la sémantique sous-jacente à leur usage. En d'autres termes, le *libellé* d'un tag est un *espace vide inscriptible* susceptible d'accueillir les entités symboliques voire informatiques de son choix³⁰. *En tant que tel, il est dès lors dénué de toute sémantique fixe* (une telle question ne se posant tout simplement pas au niveau du tag lui-même, compris sur son versant matérielle et technique).

4. Indétermination de la ressource, indétermination de l'objet.

Rashmi Sinha³¹, dans son analyse des fondements cognitifs du tagging, suggère d'y voir un assouplissement des procédures de catégorisation, fondées jusqu'alors sur l'utilisation des répertoires et jugées contraignante. Ces derniers exigeraient en effet un investissement cognitif très important pour déterminer où ranger une ressource. Obstacle qui disparaîtrait grâce au tagging, celui-ci éliminant cette barrière en vertu de sa supposée simplicité. Tout le monde n'en acquière pas pour autant les connaissances et le savoir-faire d'un indexeur professionnel. En revanche, l'indexation, par ce biais, est rendue plus accessible : elle autorise l'erreur et la redondance aussi bien que les hapax.

Une telle vision n'est pas sans exercer une certaine séduction qui tendrait à l'accréditer en retour. Seulement, en dépit de ses attraits, le tagging nous semble néanmoins comporter une difficulté d'un tout autre ordre. Nous l'avons dit, les tags instancient des relations d'accès et de référence. A des entités réelles ? A des ressources numériques ? Loin d'être évidente, la réponse renvoie en partie à la

³⁰ La limite dépend bien sûr des symboles, polices, etc. disponibles dans un système technique donné.

³¹ Rashmi Sinha est la fondatrice du site de partage de présentations *Slideshare* et l'auteur d'une analyse souvent citée dans la littérature consacrée aux tags sur les ressorts cognitifs du tagging. Cf. Sinha, R. (2005), A cognitive analysis of tagging, *Rashmi's blog*, <http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/>.

question du déréférencement des URIs. Les différents modes de visées des tags vers leurs objets suggèrent en effet, corrélativement, une multitude d'entités visées.

Les entités présentes sur le web quant à elles sont largement assimilables à ce que le philosophe Daniel Dennett nomme des « artefacts intentionnels »³². Le commerce que nous entretenons avec ces derniers est nécessairement affecté par une caractéristique qui leur échoit : celle d'être à *propos de*. Un texte peut être à propos de quelque chose de même qu'une image ou un enregistrement. Aussi, lorsqu'un utilisateur taggue une image, affirme-t-il quelque chose à propos d'une entité présente ou absente du Web (l'image vs ce qu'elle représente) et se donne-t-il la possibilité d'accéder à nouveau à une ressource numérique en ligne (la photo déposée sur le site à une adresse précise).

Cette relation d'accès elle-même est indéterminée³³ à l'instar de la description linguistique de l'objet, le passage d'une URI à une URL résultant d'une négociation entre agents. Une fois générée une page³⁴ (par exemple), reste à savoir ce que l'on décrit grâce aux libellés des tags : la page elle-même, notre rapport à la page, une partie de celle-ci ? Et laquelle ? Le billet d'un blog ou le commentaire particulièrement intéressant qui lui fait suite ? Le lien hypertexte lui-même ? Une action à accomplir liée à cette page ?, etc.

Le site *Upcoming.org* permet par exemple de créer une page dédiée à un événement et lui associe un identifiant unique. Or, alors qu'une requête HTTP ne livrera accès qu'à la page disponible sur le site et jamais à l'événement lui-même (qui, la plupart du temps, du fait de son caractère temporel passager, n'est pas encore advenu ou, au contraire, est déjà révolu), un tag, par l'intermédiaire de son libellé, en vertu de la relation symbolique qu'il établit, saura s'affranchir de ces limitations et référer à un événement passé, présent ou à venir. Mieux, seul l'utilisateur pourra établir avec certitude à quoi renvoient lesdits libellés de chacun de ses tags, un ensemble de tags pouvant servir à décrire des entités très différentes : une page, un lien, une partie d'un texte, une action, un jugement, une relation, une personne, un lieu, un événement, etc. Une telle indétermination cadre tout à fait avec l'absence de sémantique propre à la partie libellé des tags. Elle n'en exige pas moins un attachement scrupuleux au contexte, seul à même de rendre compte de l'utilisation singulière d'un tag car, si des relations sont *ipso facto* établies, elles n'en sont pas explicitées pour autant.

Conclusion.

Nous avons esquissé ici une caractérisation inédite des tags en les différenciant rigoureusement de toutes les chaînes de caractères dont ils sont potentiellement le support, et ce, quelque soit leur statut : mots, mots clef, descripteurs, vedettes matières, etc. En insistant sur la variabilité des inscriptions (libellés) qu'un tag

³² Dennett, Daniel C (1990), *The Interpretation of Texts, People and Other Artifacts*.

³³ On parle, s'agissant du phénomène de la *deixis*, de « *deferred ostension* », pour distinguer l'index du référent, ce qui est montré et ce à quoi il est fait référence. La monstration, de ce point de vue, se situe à mi-chemin entre l'accès technicisé et la référence linguistique : une zone où le langage s'incorpore les gestes.

³⁴ Qu'est-ce qu'une page au demeurant, la difficulté qu'il y a à s'accorder sur une réponse à cette question explique l'indétermination à laquelle nous faisons allusion au sujet de la relation d'accès elle-même.

(étiquette) est susceptible de recevoir, nous avons cherché à montrer que les premières n'étaient pas pourvues d'une sémantique fixe. Aussi n'est-il pas possible de partir du postulat qu'un tag réfère toujours à un concept en vertu de sa sémantique dénotationnelle. Un tel axiome sert évidemment le rapprochement entre ontologies et folksonomies en garantissant une commune appartenance à un même registre sémantique. On notera cependant qu'en laissant de côté nombre de libellés potentiels (des machines tags aux représentations iconiques), c'est une partie importante de l'intérêt des tags qui se trouve par-là même évacuée. Qu'il soit nécessaire d'opérationnaliser le rapprochement entre ontologies et folksonomie par de nécessaires concessions est une chose que l'on ne saurait contester³⁵. L'objection ne peut demeurer au plan théorique. Elle appelle par conséquent une réflexion sur l'apport des technologies du Web sémantique pour réintégrer tous les usages non-dénotationnels, expressifs, du langage. Mieux, outre l'approche guidée par un réalisme opérationnalisé dont les URIs sont le meilleur exemple, la réflexion menée autour des tags en appelle une autre, d'ordre sémiotique cette fois.

Références (sélection).

- BACH E. W. (1989), *Informal Lectures on Formal Semantics*, State University of New York Press.
- GRUBER, T. (2005), TagOntology - a way to agree on the semantics of tagging data, <http://tomgruber.org/>
- GRUBER, T. (2007), Ontology of Folksonomy: A Mash-up of Apples and Oranges, *Int'l Journal on Semantic Web & Information Systems*, 3(2).
- HAYES, P. (2006), In Defense of Ambiguity, Edinburgh, Scotland.
- LIMPENS F., GANDON F. BUFFA M. (2008), Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées : un état de l'art., *IC2008, 19èmes Journées Francophones d'Ingénierie des Connaissances*, Nancy, France.
- MONNIN A. (2009), From Game Neverending to Flickr. Tagging systems as ludic systems and their consequences. In: *Proceedings of the WebSci'09: Society On-Line*, 18-20 March 2009, Athens, Greece. (In Press)
- MONNIN A. & FELIX E. (2009), Essai de comparaison des ontologies informatiques et philosophiques : entre être et artefacts, *XVIèmes Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels de Rochebrune*, Megève, France.
- NEWMAN R. (2005), Tag Ontology, *holygoat*. <http://www.holygoat.co.uk/owl/redwood/0.1/tags/>
- NEWMAN R. (2005b), Tags, *holygoat.co.uk*. <http://www.holygoat.co.uk/blog/entry/2005-03-23-2>
- PASSANT A. (2007), Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs, *Proceedings of the First International Conference on Weblogs and Social Media (ICWSM)*, Boulder, Colorado, USA.
- PASSANT A. & LAUBLET P. (2008), Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data, *Proceedings of LDOW 2008*, Beijing.
- SVENONIUS E. (2000), *The Intellectual Foundation of Information Organization*, MIT Press (MA).
- ZACKLAD M. (2007), Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI), http://archivesic.ccsd.cnrs.fr/sic_00202440/en/

³⁵ (PASSANT & LAUBLET 2008) soulignent cet aspect en notant les limitations que leur application impose.