

Informativité, neutralité et point de vue dans une offre dictionnairique hétérogène : vers une complémentarité ?

Franck Sajous, Nabil Hathout

▶ **To cite this version:**

Franck Sajous, Nabil Hathout. Informativité, neutralité et point de vue dans une offre dictionnairique hétérogène : vers une complémentarité ?. Revue Française de Linguistique Appliquée, 2017, XXII (1), pp.27-39. halshs-01533250v2

HAL Id: halshs-01533250
https://shs.hal.science/halshs-01533250v2

Submitted on 1 Oct 2018

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.

Informativité, neutralité et du point de vue dans une offre dictionnairique hétérogène : vers une complémentarité ?

Franck Sajous et Nabil Hathout, CLLE-ERSS, CNRS & Université de Toulouse 2

Résumé : *Dans cet article, nous nous intéressons aux dictionnaires comme outils de décodage du sens, du point de vue de l'utilisateur, et plus particulièrement à la relation entre neutralité et informativité des définitions. Les dictionnaires écrits par des lexicographes et ceux écrits « par les foules » sont examinés à travers le prisme de ces deux caractéristiques, pour être finalement envisagés comme des ressources complémentaires.*

Abstract : *This article considers how different dictionaries help decode meaning, based on the user perspective. Definitions taken from expert-written and crowdsourced dictionaries are compared, with a focus on their neutrality and their informativeness. Both of the two types of dictionaries are finally judged complementary resources for the user.*

Mots-clés : dictionnaires en ligne, neutralité, informativité, approvisionnement par les foules
Keywords : online dictionaries, neutrality, informativeness, crowdsourcing

1 Introduction : les dictionnaires en ligne

La numérisation des dictionnaires a débuté par la rétroconversion de dictionnaires papier vers des fichiers informatiques et des bases de données. Une fois indexées, les données dictionnairiques peuvent être éditées, enrichies et mises à jour, constituant un nouveau point de départ vers des dictionnaires sous différents formats : CD-ROM, sites web, applications mobiles ou de nouvelles éditions imprimées. La numérisation du Trésor de la Langue Française (TLF) a par exemple permis la vente de ce dictionnaire sous forme de CD-ROM, avant d'être mis à disposition du public dans une version en ligne. Nesi (1999) décrit les possibilités offertes par ces nouveaux supports, comme celles de copier-coller des articles, les imprimer ou les mettre à disposition de multiples utilisateurs, via des réseaux locaux, tout en ne possédant qu'une licence mono-utilisateur. Elle questionne la nature du meilleur compromis pour les éditeurs entre une totale protection de leurs produits et un accès intégral à leurs dictionnaires électroniques. Une décennie plus tard, tous les éditeurs, au moins pour la langue anglaise, mettent à disposition gratuite des internautes une version en ligne de leur dictionnaire. Cette évolution ne relève pas d'un choix volontariste de leur part, mais, comme l'explique Gao (2012), de la concurrence des « agrégateurs de dictionnaires » apparus au milieu des années 1990 (`Dictionary.com`, `OneLook.com` ou `TheFreeDictionary.com`). Ces « pots-pourris » de dictionnaires mélangent du contenu glané sur le web et des extraits d'anciennes éditions de dictionnaires tombés dans le domaine public ou de dictionnaires dont les licences d'utilisation ont été acquises. Avec à sa

tête des lexicographes professionnels et un chiffre d'affaire conséquent, *WordNik* est en passe de devenir *« un poids lourd de la lexicographie en ligne »* (Gao, *ibid.*). Cette mise en concurrence tiendrait aux attentes limitées des utilisateurs, ainsi qu'à l'association « web = gratuité » ancrée dans l'esprit des internautes : le public, prêt à acheter un dictionnaire papier, un CD-ROM ou une application mobile, n'est en revanche pas prêt à payer pour un site web [1]. La question de la qualité est secondaire : dans une vision à peine caricaturale, Google et les agrégateurs font bien l'affaire. Les autres concurrents, également gratuits, sont les dictionnaires « écrits par les foules » tels qu'Urban Dictionary et Wiktionary. Les agrégateurs et les dictionnaires dits collaboratifs ont bouleversé le modèle économique des maisons d'édition. Comme le formule Kilgarriff (2005), *« si les dictionnaires sont gratuits, qui va les acheter ? »* Pour les dictionnaires de français, l'offre « en ligne » est moindre : en 2008, Larousse publie un dictionnaire en accès gratuit alors que les éditions Le Robert font le choix d'un abonnement payant. Les articles du TLF, mis en ligne en 2002, ont été rédigés entre 1971 et 1994 et ne peuvent prétendre décrire le français contemporain. De la même manière, on peut reconnaître une grande qualité au Dictionnaire de l'Académie Française, mais sa 9^{e} édition, non encore achevée, est déjà obsolète. Il suffit de consulter l'entrée *mobile* pour s'en convaincre. En 2012, MacMillan Dictionary (MMD) annonce l'arrêt de la publication de son dictionnaire au format papier, au profit exclusif de sa version en ligne. L'annonce est alors commentée positivement par son éditeur en chef Michael Rundell, mais aujourd'hui, la seule rétribution des dictionnaires en ligne par la publicité semble inquiéter les lexicographes dont les « rêves » (De Schryver, 2003) ont laissé la place à de funestes prophéties.

Nous étudions dans la suite de l'article différents dictionnaires de langue générale accessibles en ligne comme outils de décodage du sens lexical. Après avoir introduit les systèmes de *crowdsourcing* utilisés en lexicographie, nous comparons les dictionnaires issus de ce mode de production avec ceux écrits par les lexicographes. Dans un précédent article, nous avons étudié la complémentarité des deux types de dictionnaires en termes de nomenclature (Sajous et al., 2014). Nous nous proposons d'examiner ici leurs caractéristiques en termes de neutralité et d'informativité en nous posant les questions suivantes : les dictionnaires écrits par les foules présentent-ils une neutralité de point de vue ? Qu'en est-il de ceux écrits par les lexicographes ? La neutralité est-elle toujours souhaitable, *i.e.* cette contrainte permet-elle toujours de renseigner efficacement le lecteur sur le sens des mots ?

2 Le crowdsourcing en lexicographie

Le terme générique *crowdsourcing* regroupe différents modes d'accomplissement de tâches ou de production de contenu par les foules. « Les foules » désigne généralement un nombre relativement important de participants sans compétence particulière relativement à la tâche assignée.

1. Voir à ce sujet le billet *Is the web killing the dictionary?* de M. B. Mechura :
`https://multikulti.wordpress.com/2013/09/28/is-the-web-killing-the-dictionary/`

En lexicographie, deux systèmes sont utilisés : les wikis, popularisés par l'encyclopédie Wikipédia, et *Mechanical Turk* (*MT*), moins connu du grand public, initialement mis au point par l'entreprise de commerce électronique Amazon. Ce système consiste à faire accomplir par une multitude de contributeurs (les *turkers*) des tâches simples et courtes (*micro-tâches*) contre une rétribution minimale. Ces deux systèmes s'opposent sur plusieurs points. Dans les wikis, les contributeurs éditent collaborativement un contenu dont ils ont une vision globale. Inversement, *MT* ne repose sur aucune collaboration : le résultat final, souvent inconnu des *turkers*, est obtenu par agrégation de l'ensemble des contributions. Les participants ne sont souvent pas conscients de la finalité des tâches qu'ils accomplissent. Nous décrivons dans un premier temps différentes mises en œuvre de *MT* en lexicographie, qui diffèrent par le degré d'implication des participants et par la complexité des tâches qui leur sont confiées. Nous décrivons ensuite quelques projets lexicographiques fondés, en totalité ou en partie, sur le *crowdsourcing*.

2.1 *Mechanical Turk* : des mises en œuvre diverses

Une première réponse des maisons d'édition à la chute des ventes de dictionnaires a été la réduction des coûts de conception. Dans ce contexte, l'accomplissement par les foules de tâches peu qualifiées est un moyen d'économiser le temps (coûteux) des lexicographes dont l'expérience est plus utile pour identifier les différents sens d'une entrée polysémique ou rédiger des définitions. Par exemple, Kosem et al. (2013) extraient de manière automatique des collocations destinées à un dictionnaire du slovène et utilisent *MT* pour la détection des faux-postifs parmi des collocations automatiquement extraites. La difficulté de cette tâche excède largement celle prévue par *MT*, mais, dans cette expérience, les *turkers* sont des étudiants formés en lexicographie. On peut dans ce cas se demander s'il s'agit encore de *crowdsourcing*, lorsque *crowd* (les « foules ») désigne un nombre relativement restreint de participants dont la qualification est tout sauf quelconque. La notion de *MT* existait en lexicographie bien avant qu'Amazon ne l'invente, et bien avant l'arrivée d'Internet. En effet, le programme de lecture mis en place par James Murray, éditeur de l'Oxford English Dictionary, est une mise en œuvre précoce de ce système. Son appel de 1879 demandait aux lecteurs bénévoles de relever dans des œuvres littéraires prédéfinies des attestations de mots sélectionnés et de les envoyer à l'éditeur par voie postale. En 1994, *« la semaine de la sortie de Netscape 1.0 »*, des locuteurs du Swahili ont été sollicités dans le cadre du projet Kamusi pour traduire dans leur langue des listes de mots anglais. Les phases suivantes étendant le projet à d'autres langues ont permis de produire un dictionnaire multilingue (Benjamin, 2015). Plus récemment, plusieurs maisons d'édition britanniques ont doté leur dictionnaire en ligne d'un « dictionnaire ouvert », dans lequel les internautes peuvent proposer de nouvelles entrées dont ils rédigent les définitions, ou des modifications de définitions existantes. MMD fait état de 4 000 nouvelles entrées ajoutées à son *Open Dictionary* [2] depuis 2009, dont la moitié ont été promues au rang d'entrées à part entière du dictionnaire « régu-

2. `http://www.macmillandictionary.com/open-dictionary/`

lier ». Cette proportion, importante compte tenu de la liberté laissée aux internautes, contraste avec ce que l'on peut trouver dans des dictionnaires tels qu'Urban Dictionary (cf. section 2.2.2). Il se pourrait d'ailleurs que ce soit précisément la connaissance des dérives de ce dictionnaire qui a inspiré aux éditeurs du MMD les consignes données aux internautes : *« ne soumettez pas svp de mots que vous ou vos amis avez inventé, d'obscénités, de noms de personnes que vous n'appréciez pas et ce que vous voudriez leur faire [. . .] Tout cela va directement à la poubelle (virtuelle) »*. On peut voir dans l'initiative des dictionnaires ouverts (mais contrôlés) la reconnaissance par les lexicographes d'une certaine compétence lexicale des foules, qui ne sont plus utilisées dans ce cas uniquement pour accomplir des basses besognes. On peut y voir également une volonté d'impliquer l'utilisateur dans le processus d'écriture de « son » dictionnaire, et de le fidéliser. La satisfaction intellectuelle éprouvée en voyant son entrée intégrer un « vrai dictionnaire » constitue la motivation du contributeur bénévole. Les entrées qu'il propose sont non-seulement celles dont il remarque l'absence, mais plus précisément celles qu'il voudrait voir figurer dans son dictionnaire. L'utilisateur-contributeur transforme ainsi le dictionnaire en un objet qui lui convient mieux, qui lui « ressemble ». De plus, si l'informatique facilite la détection de créations lexicales par les lexicographes, la néologie sémantique et les changements de connotation sont plus complexes à automatiser. L'aide des internautes peut dans ce cas se révéler utile en amont de l'analyse du lexicographe. Enfin la diversité des domaines plus ou moins spécialisés pénétrant la vie quotidienne, dont le vocabulaire a vocation à intégrer la nomenclature des dictionnaires, rend utile le recours à la pluralité des expertises des internautes, dont les connaissances complètent la compétence lexicographique des professionnels des dictionnaires.

2.2 Les dictionnaires écrits par les foules

2.2.1 Wiktionary – `https://www.wiktionary.org/`

Wiktionary [3] est un projet de dictionnaire collaboratif multilingue, fondé comme Wikipédia sur le principe du wiki : tout internaute peut éditer n'importe quel article et chaque modification est publiée immédiatement. La version anglaise Wiktionary, lancée en décembre 2002, compte aujourd'hui plus de cinq millions d'entrées. La version française Wiktionnaire, démarrée en mars 2004, en compte plus de trois millions. Une « version de langue » désigne la langue de rédaction des articles : le français est la langue de rédaction du Wiktionnaire, dont la nomenclature inclut essentiellement des entrées françaises, mais également des mots d'autres langues. Le dictionnaire se veut non-normatif et a vocation à accueillir aussi bien les mots désuets et rares que les néologismes, ceux de la langue générale que des domaines spécialisés, ainsi que des variations diatopiques (régionales et nationales). Le Wiktionnaire n'est pas un « pur produit » du *crowdsourcing* : 17% de ses entrées ont en effet été initialement importées du Littré ou de la 8^{e} édition du Dictionnaire de l'Académie Française. Comme pour la Wikipédia, la quête de la neutralité de point de vue est un principe essentiel des consignes éditoriales de ce dictionnaire.

3. Wiktionary désigne à la fois l'ensemble du projet (toutes les éditions de langue) et sa version anglaise. Voir (Sajous et al., 2011, 2014; Sajous & Hathout, 2015) pour des descriptions plus approfondies de ces dictionnaires.

2.2.2 Urban Dictionary – `http://www.urbandictionary.com/`

Initié en 2000 par Aaron Peckam, Urban Dictionary est un projet conçu initialement en langue anglaise comme un anti-dictionnaire parodiant l'agrégateur `dictionary.com`. Selon son fondateur, cité par Gao (*op. cit.*), *« ce n'est plus aujourd'hui une simple parodie. C'est un dictionnaire que les parents et les enseignants utilisent pour comprendre la génération suivante »*. Son système de contribution diffère de celui en vigueur dans Wiktionary. Il n'y a pas à proprement parler de collaboration dans Urban Dictionary : tout internaute peut ajouter une nouvelle entrée, ou une nouvelle définition à une entrée existante, cette nouvelle définition n'ayant pas nécessairement vocation à décrire un sens supplémentaire. Un internaute ne peut pas éditer la définition rédigée par un autre auteur, mais seulement voter pour ou contre celle-ci. Pour une entrée donnée, les différentes définitions sont affichées par score de pertinence décroissant (le nombre de votes favorables moins le nombre de votes défavorables). La figure 1 montre les définitions classées au 3e et au 6e rang pour l'entrée *dictionary*.

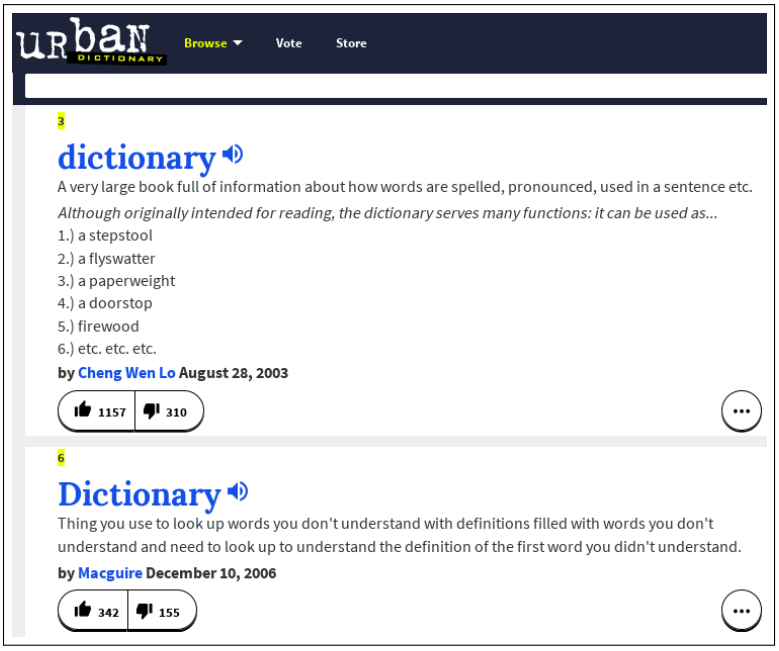


FIGURE 1: Définitions n°3 et n°6 de *dictionary* dans Urban Dictionary

Concernant la ligne éditoriale, Urban Dictionary se situe également à l'opposé de Wiktionary : le point de vue personnel y est encouragé, comme l'explique A. Peckam dans une interview accordée au New-York Times : *« Every single word on here is written by someone with a point of view, with a personal experience of the word in the entry* [4] *»*. Qualifié souvent de *slang dictionary*, Urban Dictionary contient des mots d'argot (pris dans un sens large), de *pop culture* et de

4. Street Smart: Urban Dictionary. New-York Times, 5 juillet 2009.

subcultures (dont les traductions « culture populaire » et « sous-cultures » ne seraient qu'imparfaites). Il contient également des termes de domaines spécialisés, comme *pseudo-code* en informatique et *hypernym* en linguistique, ou des mots du langage courant, comme *dictionary* (fig. 1) et *republican* (cf. section 3.1). Comme le déplorent certains de ses utilisateurs, Urban Dictionary est devenu un lieu de distraction (entrées spécialement fabriquées à but humoristique), de dénigrement de personnes ou de communautés (on parle de *namecalling* ou *nameflaming*) et d'auto-flatterie, par inclusion d'une entrée correspondant à son propre nom ou prénom et rédaction d'une définition avantageuse (Damaso & Cotter, 2007). La définition suivante montre la perception qu'ont les utilisateurs de l'évolution d'Urban Dictionary :

> **Urban Dictionary**, définition n° 1/910 (18/03/2005), votes : 48 128 pour, 6 520 contre
> A place formerly used to find out about slang, and now a place that teens with no life use as a burn book to whine about celebrities, their friends, etc., let out their sexual frustrations, show off their racist/sexist/homophobic/anti-(insert religion here) opinions, troll, and babble about things they know nothing about. (Urban Dictionary)

Nous illustrons en section 3 comment, malgré un pourcentage élevé de « déchets », Urban Dictionary peut se révéler utile en fournissant des informations absentes des autres dictionnaires.

2.2.3 La Parlure – `http://www.laparlure.com/`

Équivalent français d'Urban Dictionary lancé en 2009 et décrit dans (Dolar, 2014), La Parlure affiche une nomenclature beaucoup plus restreinte (4 453 entrées [5]). Les contributeurs peuvent, comme pour Urban Dictionary, créer une nouvelle entrée en précisant facultativement son origine géographique, ou ajouter une définition à une entrée existante. Bien que les internautes puissent voter pour ou contre une définition, les différentes contributions pour une même entrée suivent un ordre chronologique inverse et non un score de pertinence. Une nouvelle contribution n'exprime pas nécessairement un point de vue supplémentaire : elle peut apporter une précision ou répondre aux définitions précédentes, l'entrée prenant la forme d'un fil de discussion. Deux tiers des entrées sont polylexicales (locutions souvent enregistrées dans un but humoristique). La répartition de l'origine géographique des entrées est la suivante : 61 % pour le Canada, 14 % pour la France, 3 % d'autres origines diverses, 22 % des entrées n'ayant pas d'origine spécifiée. Une observation qualitative montre que La Parlure, qui aurait pu utilement recenser les mots de la « pop culture » francophone, avec une éventuelle visée diatopique, suit la même dérive qu'Urban Dictionary : prévalence de mots fabriqués, d'humour potache et d'expressions plus ou moins vulgaires, la plupart du temps à caractère sexuel.

2.2.4 Le Jargon Français – `http://jargonf.org/`

Le sous-titre « dictionnaire d'informatique francophone » qualifie mieux le site `JargonF.org` que son titre principal « le jargon français ». Compilé depuis 1994 par Roland Trique, informati-

5. Les chiffres mentionnés pour La Parlure et `JargonF.org` sont fondés sur une indexation des sites web correspondants réalisée par les auteurs de cet article en février 2017.

cien à l'Université de Rennes 2, ce dictionnaire compte plus de 15 000 entrées. L'infrastructure utilisée est un wiki placé sous l'autorité de l'administrateur du site : les ajouts et modifications suggérés par les internautes sont soumis à son contrôle éditorial. Bien qu'écartant les glossaires terminologiques de notre étude, nous avons choisi de mentionner ce projet, intéressant à plusieurs égards. D'une part, la nomenclature est considérable pour un dictionnaire écrit quasi-exclusivement par une seule personne. D'autre part, le fait que le wiki ne soit pas ouvert montre la nécessité perçue par l'administrateur de maintenir un contrôle éditorial. Enfin, l'informatique, pris au sens large (discipline scientifique, outils logiciels et matériels, internet, messagerie, domaines d'application) est un domaine qui a largement pénétré le quotidien. Si l'on trouve dans ce dictionnaire des termes techniques comme *rasterisation* (conversion d'une image *vectorielle* au format *bitmap*), dont certains sont présents à la fois en « version originale » et en français (*e.g. namespace* et *espace de nommage*), d'autres dépassent le pur champ de l'informatique et relèvent des nombreux domaines d'application de cette discipline : ainsi, *bitcoin, cryptomonnaie, prépaiement, micropaiement* relèvent du domaine bancaire, *propale* (proposition commerciale) relève de celui de la prestation de service, etc. Certaines entrées plus spécifiques à l'informatique, comme *compilateur* ou *compiler* figurent également dans les dictionnaires de langue générale. Il arrive dans ce cas que le site `JargonF.org` donne une définition plus précise (plus juste) des entrées :

> **compilateur** Programme de traduction en langage machine de programmes écrits dans un langage évolué tel que le fortran, le cobol, etc. (Larousse en ligne, 22/02/2017)

> **compiler** *verbe transitif.* [programmation]. Analyser le code source d'un programme afin d'en produire une autre forme, par exemple : binaire exécutable, pseudo-code destiné à une machine virtuelle, code source exprimé dans un autre langage. (`JargonF.org`, 22/02/2017)

Il faudrait certainement actualiser la liste des *« langages évolués »* du Larousse, mais l'inexactitude vient surtout de la traduction (uniquement) *« en langage machine »*, à laquelle en toute rigueur il convient d'ajouter, comme dans `JargonF.org`, *« en pseudo-code »* ou *« dans un autre code source »*. La valeur ajoutée du site spécialisé ne tient pas uniquement à son degré de technicité. Examinons par exemple la définition de *micropaiement* dans le Larousse en ligne :

> **micropaiement** Mode de paiement permettant d'acheter des contenus ou services de faible valeur unitaire en passant par un site Web ou par un service téléphonique (Larousse en ligne, 22/02/2017)

Un *micropaiement* peut effectivement être défini comme le paiement d'une faible somme (*e.g.* pour rétribuer les *micro-tâches* du système *MT*, cf. section 2.1). Or la raison de la création de ce terme (nul besoin n'a été ressenti, semble-t-il, de forger un néologisme pour « paiement de sommes astronomiques ») pourrait résider dans le problème que nous révèle `JargonF.org`. Le site n'apporte pas de précision technique (*e.g.* sur une éventuelle sécurisation du protocole réseau utilisé), mais formule une critique (informative) sur une pratique bancaire jugée abusive :

> **micropaiement** *nom masculin.* [banque]. Paiement d'un montant relativement faible, pas plus d'un euro en règle générale. Les systèmes sûrs les autorisant sont rares tellement les

> banques sont gourmandes : elles voudraient nous faire croire qu'une transaction leur coûte très très très cher... (`JargonF.org`, 22/02/2017)

3 Neutralité et informativité dans les dictionnaires

Atkins & Rundell (2008) écrivent que la neutralité du lexicographe est un mythe et illustrent leur propos en mentionnant l'exemple des définitions de *apratheid* dans deux millésimes du Longman Dictionary of Contemporary English (LDOCE) :

> **apartheid** the keeping of separate of races of different colours in one country, esp. of Europeans and non-Europeans in South Africa. (LDOCE, 1978)
>
> **apartheid** (in South Africa) the system established by government of keeping different races separate, esp. so as to give advantage to white people (LDOCE, 1987)

Dans la définition de 1987, les lexicographes sortent de leur neutralité en formulant un jugement communément admis dans la société occidentale de l'époque, selon lequel l'apartheid avantageait *« particulièrement »* les blancs. Maintenir la réserve de 1978 aurait d'une part pu être perçu comme un jugement de normalité, une approbation de ce système. D'autre part, cette réserve privait le lecteur d'une information essentielle. Le dictionnaire, selon Dubois & Dubois (1971), reflète la norme culturelle conforme à l'idéologie de la classe sociale dominante, à laquelle vont se référer les jugements de valeur du dictionnaire. Or les jugements de valeur sont partout dans le dictionnaire : même les marqueurs descriptifs deviennent normatifs et prennent valeur de sanction. La marque *populaire* est assimilée au 19^{e} siècle aux « basses classes ». La marque *dialectal* a une valeur sociale distinctive, du fait de la valorisation de la langue centrale. Au 20^{e} siècle, les styles *soutenu* et *littéraire* deviennent à leur tour marqués, et les métalexicographes remettent en cause les marques diastratiques telles que *populaire* (Baider et al., 2011) : les locuteurs de toutes classes sociales puisent dans le même stock lexical et choisissent d'utiliser un vocable donné en fonction des situations de communication (marques *formel, informel, familier, soutenu*…) ou les « attitudes » (marques *ironique, enfantin, péjoratif*…). Mais les jugements trouvent surtout leur place dans l'établissement de la nomenclature (le rejet étant la plus grave des sanctions culturelles) et la rédaction des définitions. Dubois & Dubois (*op. cit.*, pp. 102-103) décrivent l'étonnement des lecteurs étrangers devant certaines absences dans les dictionnaires français :

> Si les tabous sexuels sont les plus connus, ils ne sont pas les seuls : il existe des tabous religieux et politiques. Le dictionnaire a tendance à exclure les termes qui désignent une doctrine (ou les concepts fondamentaux d'une doctrine) qui remet en question l'idéologie dominante ; […] les termes *marxisme, masochisme*, les acceptions politiques de *luttes de classes* et *cellule*, etc. sont tardivement enregistrées dans les dictionnaires du XX^{e} siècle. Lorsque ces termes deviennent des entrées, le lexicographe, médiateur d'une communauté qui impose son idéologie, introduit des *jugements de rejet* : ainsi, dans les définitions, il substitue le verbe péjoratif « prétendre » à « affirmer, dire », afin de manifester qu'il ne prend pas à son compte ces affirmations *(doctrine qui prétend…)* ; ce même verbe *prétendre* n'apparaît pas dans les définitions des systèmes ou des doctrines qui sont acceptées comme des « faits » par les locuteurs actuels : *christianisme, athéisme, nationalisme*, etc.

Les auteurs mentionnent les religions et les courants de pensée. Est-il possible pour le lexicographe de rester neutre quand il s'agit de qualifier une *religion* ou une *secte* ? De quels critères

dispose-t-il ? Si le mot *secte* n'a pas initialement de connotation péjorative, on ne peut aujourd'hui l'employer sans lui associer le sens nécessairement négatif de *sectaire* et *dérive sectaire*. L'*anglicanisme* est la *religion officielle de l'Angleterre* ? Nul doute n'est permis : c'est une religion. Pour d'autres « *organisations* » ou « *mouvements* » religieux ou spirituels, la référence pourrait être le cadre juridique. Le TLF peut bien qualifier la scientologie de secte : ce n'est pas le lexicographe qui porte un jugement, mais le législateur qui établit une liste officielle des sectes, en vigueur à l'époque de la rédaction du dictionnaire. Il est intéressant de lire dans un rapport de la commission d'enquête sur les sectes, enregistré à l'Assemblée Nationale le 22/12/1995, les difficultés des parlementaires à définir cette notion, adoptant une *« approche étymologique »*, puisant dans le Littré et le Robert, et parlant de *« phénomène difficile à définir »*, d'*« impossible définition juridique »*. La conclusion du préambule de ce rapport est que la commission *« préfère retenir le sens commun que l'opinion publique attribue à la notion »*. Cette lecture est éclairante, mais n'est pas de nature à opérationnaliser le jugement du lexicographe, d'autant plus que la liste officielle, jugée obsolète, est abandonnée en 2005. Comme on peut s'y attendre, les pratiques sexuelles font également l'objet de jugements qui varient dans le temps… et l'espace. L'étude de Boisson (2000) sur la définition des pratiques « déviantes » montre effectivement une réprobation plus forte dans les dictionnaires espagnols où persistent encore des marqueurs lexicaux comme *« pêché contre nature »* dénotant un jugement fondé sur la religion. Dans une France laïque, si la morale a pendant une période remplacé la religion dans ce domaine, il serait difficilement acceptable de trouver dans un dictionnaire récent un qualifif autre que médical (*perversion* relevant de la psychiatrie) ou juridique (pratiques interdites par la loi). Béjoint (2015) confirme ce qu'écrivaient avant lui Dubois & Dubois : *"Definitions are ideologically as close to the dominant values of the society as possible, to accommodate the largest number of users. Modern DGUs [dictionaries for general users] want to be scientific, and they succeed in being impersonal; some would say that they are boring"*. Si les dictionnaires de langue générale gagnent en « scientificité », le problème n'est pas tant qu'ils deviennent « ennuyeux », mais on peut se demander s'il est toujours possible au lexicographe d'être informatif tout en adoptant une posture de neutralité qui ménage toutes les susceptibilités.

3.1 Noms de partis politiques : le cas de *Republican*

Lors de la conférence Elex 2015, le lexicographe Michael Rundell, président de session, annonçait avoir trouvé un moyen de dresser rapidement une liste exhaustive d'insultes en anglais : il s'agissait de chercher *Republican* dans Urban Dictionary. Cette anecdote, relatée également dans (Rundell, 2012), met en avant les dérives faisant de ce dictionnaire un défouloir. Or on peut légitimement se demander ce qui pousserait un internaute à chercher dans Urban Dictionary ce mot courant, décrit par les dictionnaires conventionnels tels que le LDOCE, comme le montre la définition suivante :

> **re.pu.bli.can**[1] /rɪˈpʌblɪkən/ *n* [C] **1** someone who believes in governement by elected representatives only, with no king or queen **2 Republican** a member or supporter of the Re-

> publican Party in the US → **Democrat 3 Republican** someone from Northern Ireland who believes that Northern Ireland should become part of the Republic of Ireland, not in the United Kingdom → **loyalist** (LDOCE, 2009)

La première acception désigne de manière générale les personnes opposées à la monarchie. La troisième fait référence aux républicains d'Irlande du Nord. La seconde acception, qui nous intéresse, définit un républicain comme *« un membre ou un partisan du Parti Républicain aux Etats-Unis »*, et renvoie à l'entrée *Democrat*. Comparons maintenant avec les premières définitions données dans Urban Dictionary :

> **Republican**, définition n° 1/330 (27 nov. 2002), votes : 28 105 pour, 17 848 contre
> An individual who believes that the white male Christian God should be the only object of worship on the planet, that power and wealth should remain in the hands of 1% of the world's population while the remaining 99% starve, that health care should be privatized so the poor can't afford basic medication [...] (Urban Dictionary)

> **Republican**, définition n° 2/330 (17 août 2004), votes : 12 597 pour, 7 302 contre
> Someone who supports the rights of the unborn, but won't fund stem cell research that could help the millions who are already here [...] (Urban Dictionary)

> **Republican**, définition n° 3/330 (28 nov. 2004), votes : 7 609 pour, 2 369 contre
> (Noun) Member of the United States' Republican Party, first emerging in the mid 1800s as an anti-slavery movement. Success of candidate Abraham Lincoln in the Presidential election of 1860 sparked the American Civil War. The party has since moved on to other issues. Republicans today are characterized by typically right-wing, conservative beliefs including but by no means limited to pro-business policies, lack of dependence on the government in daily life, religious convictions, and desire to outlaw what they deem socially undesirable actions.
> (Adjective) Typical or characteristic of a Republican or the Republican party.
> (Urban Dictionary)

> **Republican**, définition n° 4/330 (15 déc. 2002), votes : 6 408 pour, 3 889 contre
> One who causes almost %50 of this nations problems.
> (For the cause of the other %50 see: Democrat) (Urban Dictionary)

Comme on pouvait s'y attendre, les deux premières définitions sont orientées, sans être toutefois rédigées dans une langue ordurière. La quatrième définition a vocation humoristique. La troisième, qui distingue les deux parties du discours, définit simplement le nom *Republican* comme un membre du Parti Républicain américain, ce dernier faisant l'objet d'une définition encyclopédique : naissance du parti et description de ses valeurs, dans un style plus factuel que porteur de jugements : aile droite, conservateur, religieux, en faveur de l'entreprenariat, de la répression, etc. Admettons que cette dernière définition est plus informative que la seconde acception du LDOCE. Pour orientées qu'elles soient, les deux premières définitions ne sont pas inutiles : elles apportent des informations complémentaires concernant les positions du parti sur la répartition des richesses, la religion, le système de santé, le droit à l'avortement, etc. L'opacité pour les apprenants non « acculturés » des noms des partis ou courants politiques est emblématique de la

frontière entre la connaissance de la langue et celle de la culture : *Republican*/*Républicain* a-t-il le même sens aux Etats-Unis qu'en Espagne au long du 20^{e} siècle et particulièrement en 1936, ou en France à partir de 1792 et avant/après le « hold-up sémantique [6] » de l'UMP en 2015 ? Un républicain ne peut-il être démocrate, et vice-versa ? Les partis Républicain et Démocrate étant adversaires, les notions de république et de démocratie sont-elles antinomiques (bien qu'en général présentées comme compatibles par « l'école de la République ») ? Comment interpréter *travailleur, travailliste, populaire, populiste, social-démocrate* ? Qu'est-ce qu'une *démocratie autoritaire* ? On peut trouver réponse à ce type de questions, dans une certaine mesure, dans les « notes culturelles » du *Longman Dictionary of English Language and Culture (LDOELC)*. Ce dictionnaire contient, outre l'entrée *republican*, une entrée *Republican Party* :

> **Republican Party, the** one of the two main political parties in the US, which is generally regarded as more CONSERVATIVE than the Democratic Party. [. . .] The Republican Party tends to support the owners of businesses and industry, and to oppose high government spending on WELFARE.[. . .] It represents a wide range of opinions, from fairly conservative to very RIGHT WING. (LDOELC, 2002)

Ce type de définition encyclopédique est informatif, mais la question pour le lexicographe est de trouver le juste compromis entre être suffisamment informatif et ménager les différentes susceptibilité des lecteurs.

3.2 Contresens, marketing et « éléments de langage »

Allophone ou apprenant étranger, il arrive au locuteur attentif et curieux de noter des différences entre sa compréhension d'un vocable et l'usage qu'il en perçoit. S'il entend dans un débat l'usage du mot *sémantique* dans le (contre)sens *blabla* (*« tout ça c'est juste de la sémantique »*) décrit par Le Draoulec et al. (2014), le dictionnaire ne lui sera probablement d'aucun secours, sauf à chercher la traduction *semantics* dans Urban Dictionary (démarche tout sauf immédiate) :

> **semantics**, définition n° 4/4 (1 avr. 2009), votes : 143 pour, 143 contre
> The study of the meaning of words.
> Often misused when quibbling about something someone said. In that context, the statement "That's only semantics" would be more aptly phrased as "You're just 'splitting hairs on word meanings." [...] The very concept of semantics is frequently disparaged by wishy-washy passive-aggressives who refuse to be accountable for their careless use of language or their deplorable lack of education. (Urban Dictionary)

De la même manière, découvrant dans un dictionnaire la signification du sigle *SUV* (*sport-utility vehicle*) on pourra percevoir la juxtaposition de *sport* et *utility* comme un oxymore (voiture de sport *et* utilitaire ?) que l'on ne s'éxplique pas :

> **SUV** /ˌes juː ˈviː/ noun [countable] American English
> **(sport-utility vehicle)** a type of vehicle that is bigger than a car and is made for travelling over rough ground (LDOCE en ligne, 22/02/2017)

6. « Républicains » : non au hold-up sémantique de Nicolas Sarkozy, *Libération*, 5 mai 2015

> **SUV** (*plural* SUVs)
> *Initialism of* sport utility vehicle (Wiktionary – entrée : 6/12/2004, consult. 22/02/2017)
>
> **sport utility vehicle** (*plural* sport utility vehicles)
> A passenger vehicle which combines the towing capacity of a pickup truck with the passenger-carrying space of a minivan or station wagon together with on- or off-road ability.
> (Wiktionary – entrée : 27/09/2007, consult. 22/02/2017)

Les définitions trouvées dans ces « dictionnaires neutres » entretiennent la confusion : un SUV, *« conçu pour rouler sur terrain accidenté »*, serait donc un véhicule tout-terrain ? Dans ce cas, comment un « 4x4 » pourrait-il être qualifié de voiture de sport ou d'utilitaire ? Et quelle différence alors entre un *SUV* et un *4x4, four-by-four* ou *4WD* (four-wheel drive) ? Un élément de réponse peut être trouvé dans Urban Dictionary :

> **SUV**, définition n° 1 (30 avril 2003), votes : 1 718 pour, 276 contre
> Sport Utility Vehicle. Neither a sport vehicle nor a utility vehicle. A whack, fakeass (and successful) attempt by the motor vehicle industry to lure in overpaid middle class workaholic moms who think that they need a 3 ton vehicle to carry their stupid kids to soccer practice.
> (Urban Dictionary)

Cette définition nous permet finalement de clarifier ce concept dans notre « réalité culturelle » qui n'est pas si éloignée : un *SUV* est donc ce que certains appellent un *faux 4x4 de ville…* Mais une telle définition critiquant un « coup marketing » de l'industrie automobile pourrait-elle trouver sa place dans le dictionnaire d'une maison d'édition ? La politique empruntant volontiers les techniques de « communication » de la publicité, certains néologismes de ce domaine méritent également un décryptage. Examinons la définition de *TVA sociale* dans le Wiktionnaire :

> **TVA sociale** \te.ve.a.sɔ.sjal\ *féminin*
> *(France)* TVA utilisée pour financer la protection sociale.
> (Wiktionnaire – entrée : 22/06/2007, consult. 22/02/2017)

Rien de choquant dans cette taxe ainsi définie, si l'on considère favorablement la protection sociale. Mais comment comprendre qu'un universitaire fiscaliste rebaptise cette taxe *TVA antisociale* [7] ? Le Wiktionnaire, par sa neutralité, n'apporte aucune réponse. Il faut consulter un dictionnaire spécialisé [8] pour trouver les principaux arguments pour ou contre cette taxe, notamment le fait qu'elle *« viendrait renforcer la TVA qui est un impôt injuste car il est payé par tout le monde au même taux, sans tenir compte du niveau des revenus »*. La quête de neutralité n'est pas seulement difficile à satisfaire : il semble qu'elle soit contre-productive en termes d'informativité pour certaines entrées, et qu'une ligne éditoriale semblable à celle d'Urban Dictionary soit plus à même d'apporter une aide au décodage du sens et à la perception des connotations.

3.3 Wiktionnaire : neutralité et choix d'exemples citationnels

Une des consignes éditoriales du Wiktionnaire, comme mentionné plus haut, est la recherche de la neutralité de point de vue. Dans le projet encyclopédique Wikipédia, l'existence de polé-

7. TVA "sociale", antisociale ? Jacques Le Cacheux, *Le Monde*, 12/01/2012
8. *Toupictionnaire*, le dictionnaire politique : `http://www.toupie.org/Dictionnaire/Tva_sociale.htm`

mique est souvent indiquée par des formules du discours rapporté comme *« selon les partisans de X,… ses détracteurs estiment en revanche que… »*. L'article encyclopédique ne tranche pas mais permet au lecteur de prendre conscience qu'une notion est non consensuelle. Libre à ce lecteur de pousser plus loin son investigation. Or une définition lexicographique ne permet pas de tels développements et la brièveté de certaines gloses, comme celle de *TVA sociale*, rend ces dernières peu informatives. L'analyse des exemples qui suivent les gloses du Wiktionnaire montre que le choix des exemples citationnels n'est pas anodin. Ils permettent d'exprimer un jugement sans entamer la neutralité apparente du rédacteur de l'article, qui rédige la seule glose : la responsabilité du point de vue exprimé est celle de l'auteur de la citation. Reprenons l'exemple de *SUV*, dont le français emprunte le sigle, et examinons sa définition dans le Wiktionnaire :

> **SUV** \ɛs.y.ve\ *masculin*
> *(Automobile)* Voiture bicorps initialement tout-terrain souvent à roues motrices dont l'espace utile commun aux passagers et aux bagages est modulable
> • *Comme on doit regretter, à Boulogne-Billancourt, de ne pas avoir disposé plus tôt d'un* ***SUV*** *comme le Renault Captur dans la gamme, tant ce vrai faux 4x4 fait un carton! –* (Cristian DAVID, « Renault Captur : le SUV version sourire », *L'Express*, 26/02/2014)
> (Wiktionnaire – entrée : 9/08/2009, consult. 22/02/2017)

La glose peut laisser l'internaute à ses interrogations. La formule *« vrai faux 4x4 »* de l'exemple sera en revanche évocatrice pour certains lecteurs. Elle ne rompt pas avec le principe de neutralité du Wiktionnaire, puisque c'est un journaliste de l'Express qui l'écrit… Concernant la *scientologie*, c'est une *« doctrine »*, un *« mouvement scientifico-religieux »* :

> **scientologie** \sjɑ̃.to.lo.ʒi\ *féminin*
> Doctrine de la Scientologie, mouvement scientifico-religieux fondé aux États-Unis en 1954 par L. Ron. Hubbar qui promeut une méthode appelée « dianétique » […]
> • *La* ***scientologie*** *se caractérise par un fantasme de toute-puissance qui entretient une mystique de l'organisation propre à attirer des sujets fragilisés. –* (Encyclopædia Universalis) (Wiktionnaire – entrée : 7/10/2008, consult. 22/02/2017)

Certes, le composé *scientifico-religieux* peut être pris comme un doute émis sur le sérieux des fondements scientifiques de ce mouvement. Mais l'attaque vient de la citation tirée de l'encyclopédie Universalis, une « autorité », qui parle de *« fantasme de toute-puissance »* et du but d'*« attirer les sujets les plus fragiles »*. De la même manière, le Wiktionnaire ne définit pas le *créationnisme* par les dangers qu'il véhicule, mais l'attestation fournie, extrait du journal *Le Monde*, autre autorité, citant lui-même le Conseil de l'Europe, attire l'attention du lecteur sur *« les dangers du créationnisme dans l'éducation »* :

> **créationnisme** \kʁe.a.sjɔ.nism\ *masculin*
> Thèse selon laquelle la Terre, et par extension l'Univers, a été créée par un être supérieur, c'est-à-dire Dieu selon des modalités conformes à une lecture littérale de la Bible.
> • *Le Conseil de l'Europe souligne les dangers du* ***créationnisme*** *dans l'éducation. –* (*Le Monde*, 26 juin 2007) Wiktionnaire – entrée : 5/05/2008, consult. 22/02/2017

4 Conclusion

Dans son article « *I don't believe in word senses* », sur la tâche de désambiguïsation automatique en contexte, Kilgarriff (1997) interrogeait la pertinence d'un inventaire discret de sens et lui préférait une liste de concordances. Pour l'utilisateur de dictionnaires, le décodage du sens dénotatif et connotatif pourrait être vu comme l'établissement de la somme des définitions de différents dictionnaires. Dans cette perspective, les dictionnaires écrits par des lexicographes et ceux écrits par des amateurs, d'abord présentés comme concurrents potentiels (Zesch & Gurevych, 2010), puis complémentaires en termes de nomenclature (Sajous et al., 2014), pourraient le devenir en termes d'informativité. Wiktionary et, dans une certaine mesure, les *« open dictionaries »*, ont une propension à intégrer rapidement les néologismes, les régionalismes et les termes spécialisés de nombreux domaines. Urban Dictionary peut être utile, par la pluralité des opinions de ses contributeurs, à la compréhension de notions non consensuelles. La proportion de « déchets » dans sa nomenclature n'est pas nécessairement problématique : un dictionnaire en ligne n'est pas un manuel scolaire dont le vocabulaire, destiné à l'apprentissage, doit être rigoureusement sélectionné. On y trouve au contraire seulement les entrées que l'on recherche.

Farina (2016) évoque la difficulté du traitement lexicographique de changements lexicaux survenus en temps de bouleversements sociaux et de guerres. Elle donne notamment les exemples de la censure gouvernementale appliquée à un dictionnaire russe de l'ère pré-soviétique, ou plus récemment dans le contexte de l'annexion de la Crimée par la Russie, l'introduction d'étymologies erronées à dessein dans le Nouveau Dictionnaire Russe-Ukrainien. Elle mentionne enfin l'*« innovation linguistique descendante »*, digne de la littérature orwellienne, *i.e.* les efforts souvent fructueux pour imposer à la langue courante des changements planifiés en haut lieu via le relais de la presse officielle. Cette situation, certainement plus préoccupante que les usurpations lexicales du marketing, pourrait nous faire considérer la liberté laissée aux contributeurs de dictionnaires tels qu'Urban Dictionary et les possibilités de décryptage offertes à ses utilisateurs comme des aspects positifs d'une lexicographie… pas tout à fait « naïve ». Une perspective intéressante pourrait être la conception d'outils faisant émerger de ce type de ressource, selon des critères à définir, des définitions dont la dimension culturelle pourrait nourrir la réflexion des lexicographes professionnels et des enseignants de langue.

Références

Atkins, B. T. S. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Baider, F., Lamprou, E., & Monville-Burston, M. (2011). *La marque en lexicographie: états présents, voies d'avenir*. La lexicothèque. Lambert-Lucas.

Béjoint, H. (2015). Dictionaries for General Users: History and Development; Current Issues. In P. Durkin (Ed.), *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press.

Benjamin, M. (2015). Crowdsourcing microdata for cost-effective and reliable lexicography. In *Proceedings of the 9th International ASIALEX Conference.* Hong Kong.

Boisson, C. (2000). Définitions lexicographiques et pratiques sexuelles déviantes. In H. Béjoint & P. Thoiron (Eds.), *Le Sens en terminologie* (pp. 256–279). Presses Universitaires de Lyon.

Damaso, J. & Cotter, C. (2007). UrbanDictionary.com. Online dictionaries as emerging archives of contemporary usage and collaborative lexicography. *English Today*, 23(2), 19–26.

De Schryver, G. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2), 143–199.

Dolar, K. (2014). La place du nom propre dans les dictionnaires collaboratifs : le cas de La Parlure. In *Actes du 4e Congrès Mondial de Linguistique Française* (pp. 813–827). Berlin.

Dubois, J. & Dubois, C. (1971). *Introduction à la lexicographie : le dictionnaire*. Librairie Larousse, Paris.

Farina, D. M. T. C. (2016). Lexical change in times of upheaval and war - and the dictionary. In *Proceedings of the 17th EURALEX Congress* (pp. 767–776). Tbilisi.

Gao, Y. (2012). Online English Dictionaries: Friend or Foe. In *Proceedings of the 15th EURALEX International Congress* (pp. 422–433). Oslo.

Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, 31(2), 91–113.

Kilgarriff, A. (2005). If dictionaries are free, who will buy them? *Kernerman Dictionary News*, 13, 17–19.

Kosem, I., Gantar, P., & Krek, S. (2013). Automation of lexicographic work: an opportunity for both lexicographers and crowdsourcing. In *Proceedings of Elex 2013* (pp. 32–48). Tallin.

Le Draoulec, A., Péry-Woodley, M.-P., & Rebeyrolle, J. (2014). Glissements progressifs de "sémantique". *Le Discours et la Langue*, 1(6), 109–126.

Nesi, H. (1999). A user's guide to electronic dictionaries for language learners. *International Journal of Lexicography*, 12(1), 55–66.

Rundell, M. (2012). 'It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical. In *Proceedings of the 15th EURALEX International Congress* (pp. 47–92). Oslo.

Sajous, F. & Hathout, N. (2015). GLAWI, a free XML-encoded Machine-Readable Dictionary built from the French Wiktionary. In *Proceedings of eLex 2015* (pp. 405–426). Herstmonceux.

Sajous, F., Hathout, N., & Calderone, B. (2014). Ne jetons pas le Wiktionnaire avec l'oripeau du Web ! Études et réalisations fondées sur le dictionnaire collaboratif. In *Actes du 4e Congrès Mondial de Linguistique Française* (pp. 663–680). Berlin.

Sajous, F., Navarro, E., & Gaume, B. (2011). Enrichissement de lexiques sémantiques approvisionnés par les foules : le système WISIGOTH appliqué à Wiktionary. *TAL*, 52(1), 11–35.

Zesch, T. & Gurevych, I. (2010). Wisdom of Crowds versus Wisdom of Linguists - Measuring the Semantic Relatedness of Words. *Natural Language Engineering*, 16(01), 25–59.