

## Lexicographie collaborative, variation et norme : le projet 10-nous

Marie Steffens  
Université de Liège / Universiteit Utrecht

Résumé : À côté des enquêtes sociolinguistiques classiques, la lexicographie collaborative, née avec le 21<sup>e</sup> siècle, envisage également la variation. Les mécanismes sur lesquels elle s'appuie pour la décrire et les résultats concrets qu'elle obtient à cet égard n'ont encore été que peu étudiés. Dans cette contribution, nous montrerons l'intérêt d'exploiter les connaissances et compétences des locuteurs pour la description scientifique de la variation lexicale. Nous proposerons un rapide état des lieux de la représentation de la variation dans les principales ressources collaboratives et nous présenterons *10-nous* (« Dis-nous »), projet de lexicographie collaborative accompagnée par une démarche scientifique destinée à garantir la précision et la pertinence des données recueillies auprès des locuteurs.

Abstract: In addition to the classical sociolinguistic inquiries, collaborative lexicography, born with the 21st century, describes variation. The mechanisms on which this description is based and the concrete results it achieves in this respect have not yet been fully studied. In this contribution, we will show the interest of exploiting the knowledge and skills of the speakers for the scientific description of the lexical variation. We will present a quick overview of the representation of the variation in the main collaborative resources and we will present *10-nous*, a collaborative lexicographic project accompanied by a scientific approach to guarantee the accuracy and the relevance of the data collected from the speakers.

Mots-clés : lexicographie collaborative, variation, enquête sociolinguistique

La pertinence de faire appel aux locuteurs pour décrire la variation des langues dans le temps, l'espace et la société est établie depuis le début du 20<sup>e</sup> siècle, avec la naissance puis le développement de la géographie linguistique (SWIGGERS 1999, LAUWERS *et al.* 2002).

Les enquêtes sociolinguistiques menées auprès des locuteurs ont notamment permis la production d'atlas linguistiques et de glossaires de qualité (ALW, DULONG et BERGERON 1980). Actuellement, les enquêtes en ligne facilitent le recueil rapide d'ensembles volumineux de données. Ces données, qui portent sur des lexèmes sélectionnés par les chercheurs, sont recueillies au moyen de questions (« Le midi, est-ce que vous *déjeunez* ou vous *dînez* ? », « Comment nommez-vous la viennoiserie contenant du chocolat : *pain au chocolat* ou *chocolatine* ? », etc.) ou au moyen d'images, avec généralement des choix de réponses (voir notamment le projet « Français de nos régions », AVANZI *et al.* 2016).

À côté de ces enquêtes, la lexicographie collaborative ou participative<sup>1</sup>, c'est-à-dire rédigée et validée par des locuteurs de profils divers (âge, sexe, profil socioéconomique, lieu d'appartenance, etc.), née avec le 21<sup>e</sup> siècle, envisage également la variation. Les mécanismes sur lesquels elle s'appuie pour la décrire et les résultats concrets qu'elle obtient à cet égard n'ont encore été que peu étudiés.

Dans les pages qui suivent, nous montrerons l'intérêt d'exploiter les connaissances et compétences des locuteurs pour la description scientifique de la variation lexicale. Nous proposerons un rapide état des lieux de la représentation de la variation dans les principales ressources collaboratives et nous présenterons *10-nous*, projet de lexicographie collaborative accompagnée par une démarche scientifique destinée à garantir la précision et la pertinence des données recueillies auprès des locuteurs.

---

<sup>1</sup> Pour une distinction entre lexicographie collaborative et lexicographie participative, voir Dolan dans ce volume.

## 1. Mettre le locuteur au centre de la description

Les ressources collaboratives accessibles en ligne, qui se sont multipliées ces quinze dernières années, permettent aux locuteurs d'être véritablement acteurs de la description de leur langue. Ces ressources numériques créées et alimentées par des locuteurs profanes sans formation en lexicographie sont fondées sur l'idée que tout locuteur d'une langue peut fournir des données utiles à la description du lexique de cette langue (essentiellement, définitions et exemples).

Ce faisant, les ressources collaboratives peuvent être exploitées dans l'enseignement des langues pour permettre aux élèves de développer leur conscience métalinguistique, de construire une réflexion critique autour de l'objet dictionnaire, d'appréhender les différentes réalités sociales liées à la langue et d'analyser les échecs de communication engendrés par la confrontation de différentes variétés (BEYER et al. à paraître). L'acquisition de ces compétences passe par la consultation des ressources lexicographiques, en format papier ou numérique, mais aussi par la contribution active à la description de la langue, que seules permettent les ressources collaboratives en ligne.

Donner l'opportunité aux locuteurs de devenir de réels descripteurs de leurs usages lexicaux, ce n'est pas seulement formateur pour eux, cela peut aussi être utile à la recherche scientifique, pour autant que les conditions nécessaires à un accompagnement efficace des locuteurs soient réunies de manière à les aider à convertir leurs représentations, leurs intuitions et leurs connaissances en données pertinentes, structurées et pleinement exploitables.

Les apports des locuteurs, ainsi accompagnés, à la lexicographie pourraient se manifester dans les cinq champs suivants : 1) la mise à jour constante des données, 2) le recueil de données sur des variantes peu décrites, 3) l'identification des normes locales, 4) la localisation des usages, 5) l'articulation de la description de la variation et de celle du lexique commun.

### 1.1. Données actualisées

L'apparition des outils numériques a non seulement favorisé l'accessibilité des ressources lexicographiques tant aux utilisateurs passifs qu'aux contributeurs actifs, mais elle est, de ce fait, également le moteur de bouleversements plus fondamentaux concernant leurs mécanismes éditoriaux et la temporalité dans laquelle ils s'inscrivent. Les contraintes de diffusion (éditions régulières voir annuelles) et d'espace (nombre de pages, taille et nombre des volumes), qui pèsent sur les mises à jour des dictionnaires papier ne s'appliquent plus aux ressources collaboratives en ligne ni aux ressources scientifiques qui s'adaptent presque en temps réel en fonction des remarques des utilisateurs (*Usito*). L'actualisation globale épisodique des dictionnaires papier cède ainsi la place à une actualisation partielle mais constante des ressources en ligne.

Cette actualisation constante rend possible dans les ressources collaboratives une veille néologique plus efficace que les dictionnaires classiques (Creese 2013). La lexicographie traditionnelle par souci de rigueur n'intègre les mots nouveaux qu'à partir du moment où leur usage est bien établi. Pour mieux comprendre les processus qui sous-tendent la naissance, la vie et la mort des lexèmes, il serait toutefois intéressant d'attester les mots éphémères, dont la période d'emploi peut n'être que très limitée. En encourageant les locuteurs à partager ces mots, à en documenter l'existence et à en signaler le déclin, la lexicographie collaborative pourrait garder la trace de la créativité des locuteurs.

## 1.2. Variantes sous-représentées

De nombreuses ressources collaboratives actuellement disponibles en ligne se concentrent sur le français populaire (*Blazz*, *Dico2rue*, *Le Dictionnaire de la zone*, *La Parlure* ou *Urbandico*, équivalent francophone de l'*Urban Dictionary*). Les lexèmes et les séquences figées qui y sont décrits sont en usage à l'oral ou dans des variétés familières non formelles du français écrit. Ils ne sont que peu ou pas envisagés dans les dictionnaires scientifiques et commerciaux.

Une des raisons de cette absence est la difficulté d'objectiver l'usage de ces lexèmes par des corpus oraux suffisamment étendus et représentatifs. La présence dans les ressources collaboratives d'extraits de chansons ou de films cités en guise d'exemples donne des indications, même sommaires, sur les conditions d'emploi de ces lexèmes et forme une base pour la constitution de tels corpus.

Une autre raison réside sans doute dans le caractère généralement normatif des dictionnaires professionnels. En tant qu'ouvrages de références, ils ne décrivent pas l'immense majorité des lexèmes argotiques, triviaux ou vulgaires. Ceux qui s'y trouvent sont clairement marqués comme tels et généralement anciens (voir, par exemple, *esgourde*, *pif*, *paluche*, *burne* et *panard* dans le *Grand Robert* en ligne).

Or il est essentiel de répertorier et de documenter les différentes variétés de la langue notamment pour en assurer un enseignement capable de préparer efficacement les apprenants à la réception et à la production d'énoncés dans de réels contextes de communication, nécessairement exposés à la variation.

L'objectif plutôt descriptif que normatif des ressources collaboratives étant tacitement partagé par leurs contributeurs, elles sont en mesure de prendre en charge la description des lexèmes absents des dictionnaires de référence (COTTER & DAMASO 2007, SMITH 2011).

## 1.3. Normes locales

La description de l'usage d'un lexème comprend son inscription dans un système de normes qu'il faut également caractériser : Ce lexème est-il d'usage courant ? Peut-il être utilisé dans un contexte formel ? Est-il critiqué ? Par qui ? Pourquoi ?

Les ressources collaboratives, par nature, ne sont pas des ouvrages de référence. Si elles ne sont pas conçues dans un but prescriptif et n'ont de facto pas suffisamment de légitimité pour atteindre un tel but, elles documentent toutefois la norme à différents niveaux.

Elles renseignent surtout sur la norme objective. La présence dans des ressources comme le *Wiktionnaire* de lexèmes qui ne sont pas (encore) recensés dans les dictionnaires classiques est un indice de leur emploi effectif par les locuteurs. Des mots farfelus sans usage réel pourraient toutefois, en théorie, être intégrés au *Wiktionnaire*. En pratique, des garde-fous efficaces sont mis en place pour éviter ce type de contributions. La communauté d'utilisateurs est, en effet, suffisamment active pour assurer un contrôle par les pairs des informations encodées et leur adéquation avec les critères d'acceptabilité d'une entrée, clairement énoncés sur le site ([https://fr.wiktionary.org/wiki/Wiktionnaire:Critères\\_d'acceptabilité\\_des\\_entrées](https://fr.wiktionary.org/wiki/Wiktionnaire:Critères_d'acceptabilité_des_entrées)). Par ailleurs, l'historique de contribution, librement accessible, indique le nombre de contributeurs impliqués dans la rédaction d'une page. Plus ce nombre est élevé, plus on peut être sûr que le lexème décrit est effectivement utilisé. En outre, les ressources collaboratives les plus construites, qui demandent aux utilisateurs d'évaluer les contributions (*j'aime/j'aime pas*) et qui, d'une manière ou d'une autre, tiennent compte de ces votes pour hiérarchiser les notices par popularité (ex. *Blazz*, *Dico2rue*, *Dico des mots*), donnent ainsi également des indications relatives, certes rudimentaires, sur la diffusion des lexèmes.

Les ressources collaboratives peuvent également renseigner sur les normes endogènes

locales qui ne sont pas nécessairement décrites dans des ouvrages de référence spécifiques (ex. *aboiteau* en français acadien, répertorié dans le *Wiktionnaire*, sans équivalent en français de référence au Québec ou dans une autre région francophone). La description de lexèmes standards locaux permet de cerner la hiérarchie prescriptive interne aux différentes variétés de français.

#### 1.4. Localisation des usages

Un autre atout des ressources collaboratives est la possibilité, jusqu'ici insuffisamment exploitée, d'utiliser le profil des contributeurs pour localiser précisément les usages qu'ils décrivent. À ce stade du développement de ces ressources, les contributeurs ont la possibilité de donner, dans la description des lexèmes qu'ils partagent, des indications permettant de situer l'emploi de ces lexèmes dans une zone géographique donnée. Le degré de précision de ces indications est laissé à l'appréciation des contributeurs et demeure souvent assez faible. Ainsi, les usages sont en général étiquetés *France, Belgique, Canada* (dans le meilleur des cas, *Québec*), alors que l'on sait bien que les réalités linguistiques dans les différentes provinces, régions et sous-régions de ces pays peuvent être très différentes.

Compte tenu de l'étendue et de la diversité de la communauté d'utilisateurs-contributeurs qu'une ressource comme le *Wiktionnaire* a pu réunir autour d'elle, les usages de ces différentes régions pourraient être localisés plus précisément en fonction des profils des contributeurs qui les partagent, pour autant que l'origine ou le lieu d'habitation soient des informations requises au moment de la création d'un compte de contribution.

#### 1.5. Variante et lexique commun

Les projets lexicographiques variationnels classiques sont le plus souvent différentiels et contrastifs (RÉZEAU 2001, MERCIER & VERREAULT 2002, THIBAUT 2004 et 2008, GLESSGEN & THIBAUT 2005, BERNET & RÉZEAU 2010). L'articulation de la description fine de la variation avec celle du lexique commun y est en général peu développée (POISSON 1999 et 2002, VIOLETTE 2006, GUÉRIN 2008). En amont de cette articulation, se posent les questions à l'objectivation du lexique commun : Quels sont les lexèmes réellement partagés, compris et utilisés dans le même sens par tous les francophones ? Y a-t-il un français international ? Et quelle est son étendue ? Plus largement, une approche intégrée du lexique, tenant compte de la combinatoire des lexèmes, peut permettre de vérifier l'hypothèse selon laquelle la grammaire reste relativement stable tandis que le lexique varie, de la nuancer et de documenter le français fondamental à l'échelle de la francophonie.

La prise en compte de l'ensemble du lexique, lexique commun et sociolectes, a fait ses preuves tant en linguistique historique et en philologie (CHAMBON 2006, GREUB 2002) qu'en linguistique synchronique (BARONIAN et MARTINEAU 2009). Une définition rigoureuse de cet ensemble est, en effet, nécessaire tant à la compréhension des mécanismes d'évolution de la langue qu'à la description de son fonctionnement actuel sur le plan phonétique, morphologique, syntaxique et lexical. Elle permet de représenter l'inscription des lexèmes dans des réseaux lexicaux mettant en évidence leur variation sémantique sur les différents axes, de quantifier précisément l'influence des autres langues dans ces réseaux et de mieux comprendre les mécanismes d'évolution sémantique.

Les ressources collaboratives actuellement disponibles peinent à concrétiser cette prise en charge globale. La seule exception notable est le *Wiktionnaire*. En visant la description de tout le lexique du français, en ce compris les mots anciens et/ou rares, cette ressource se positionne à la fois comme un équivalent, une synthèse et un complément des dictionnaires professionnels, dont elle épouse pour l'essentiel tant la forme que le fond et dans lesquels les

lexicographes amateurs empruntent souvent les matériaux utiles à leur contribution (FUERTES OLIVERA 2009, MEYER & GUREVYCH 2012).

Nous allons maintenant envisager plus avant la représentation de la variation dans les principales ressources en ligne, dont le *Wiktionnaire*.

## 2. Ressources collaboratives et variation

Les ressources collaboratives en ligne, par définition en constante mutation, sont très diverses (MURANO 2014, DOLAR 2017) : depuis la compilation de messages épars (blogues, forums tels que celui du Projet Babel), jusqu'aux ensembles plus structurés comme les nombreuses ressources centrées sur le français populaire (*Dico2rue*, etc.) et le *Wiktionnaire*, de loin la plus élaborée des ressources collaboratives en ligne.

L'échelle qualitative de la lexicographie numérique va des ressources qui laissent une totale liberté aux contributeurs de partager tout (et parfois n'importe quoi) sans révision *a posteriori* (*La Parlure*) aux dictionnaires professionnels qui consultent de plus en plus leurs utilisateurs sans toutefois leur donner le pouvoir de contribuer directement (*Usito*), en passant par les plateformes collaboratives qui proposent une validation des données par les pairs (rédaction collective des articles grâce au système wiki dans le *Wiktionnaire*, hiérarchisation des contributions sur la base du vote des autres utilisateurs dans *Blazz*).

Dans le cadre d'une première étude exploratoire, nous avons comparé le traitement de la variation dans les neuf principales ressources collaboratives monolingues francophones disponibles en ligne que sont *Blazz*, *Bob*, *Dico2rue*, *Dico des mots*, *Le Dictionnaire de la zone*, *La Parlure*, *Reverso*, *Urbandico*, *Wiktionnaire* (voir présentation et tableau récapitulatif en annexe). Ont donc été écartées de l'étude 1) les ressources qui ne sont pas collaboratives, même si elles sont plus ou moins à l'écoute de leurs utilisateurs (entre autres les ressources scientifiques comme la BDLP, *Base de données lexicographiques panfrancophone*, les dictionnaires scientifiques ou commerciaux comme *Usito* ou *Cordial*, etc.); 2) les dictionnaires traductifs éventuellement collaboratifs comme le dictionnaire du projet *Babel* ou le volet multilingue de *Reverso* ; 3) les ressources limitées de manière évidente à une seule région francophone comme le *Wikimanche* dont la partie linguistique se concentre sur le français parlé en Normandie ; 4) les forums traitant de questions linguistiques, comme le forum *Babel*, qui, en raison de leur structure, contiennent souvent des informations intéressantes et nombreuses mais éparses et donc difficiles à rassembler et à exploiter.

L'hétérogénéité sur le plan macrostructural (nombre et nature des unités traitées) des neuf ressources retenues rend difficile une comparaison fine. Il est toutefois possible de dégager de grandes tendances quant au traitement microstructural de la variation dans la lexicographie numérique collaborative en français.

Dans la plupart des ressources collaboratives, la microstructure est fixe et régie par les instructions de contribution (ABEL & MEYER 2013). Ces instructions sont cependant assez rudimentaires et ne prévoient en général l'encodage que d'une définition et d'un exemple. Les notices des neuf principales ressources collaboratives francophones, qu'elles traitent d'unités monolexicales ou de séquences figées, contiennent ainsi toutes au minimum une définition, généralement accompagnée d'un exemple (sauf *Reverso*). Dans la plupart de ces ressources, les exemples sont le plus souvent forgés (*Blazz*, *Dico2rue*, *Dico des mots*, *La parlure*, *Urbandico*). Il peut s'agir également de réelles attestations tirées de textes littéraires (*Bob*, *Wiktionnaire*), de chansons (*Dictionnaire de la Zone*, *Wiktionnaire*) ou de sites internet (*Bob*).

En ce qui concerne la prise en charge de la variation dans ces ressources, deux grands groupes de ressources peuvent être dégagés (voir tableau en annexe).

Les ressources du premier groupe proposent un réel marquage des unités décrites. Parmi ces ressources, on retrouve sans surprise le *Wiktionnaire* et *Reverso* qui sont les deux seules à

envisager le lexique de manière globale. On y retrouve aussi de manière un peu plus surprenante deux ressources centrées sur un usage particulier, *La Parlure* (français parlé) et *Bob* (argot). Il est important de noter que ce marquage n'apparaît vraiment dans *Reverso* et *Bob* qu'après révision. Par ailleurs, seuls le *Wiktionnaire* et *Reverso* proposent un marquage des variations diachronique, diatopique et diaphasique. Ce marquage ressemble beaucoup à celui que l'on trouve dans les dictionnaires classiques (*vielli, régional, familier, etc.*). *La Parlure*, seule ressource collaborative d'initiative québécoise, ne marque explicitement que la variation diatopique alors que *Bob*, limité à l'argot de France ne marque que les variations diachronique et diaphasique et ce d'une manière tout à fait originale. Après révision, les notices de *Bob* proposent en effet un code chiffré qui permet à la fois de situer les lexèmes sur une échelle d'argotité, du plus courant au plus argotique, et de mesurer l'évolution dans le temps de la position du lexème sur cette échelle. Toutes ces ressources documentent également, par le biais de remarques, les usages critiqués et les usages recommandés. Les marques présentes dans les ressources pleinement collaboratives (*Wiktionnaire* et *Parlure*), si elles sont souvent justes et éclairantes, ne sont pas nécessairement systématiques parce que laissées à l'initiative des contributeurs sans qu'un accompagnement particulier ne les incite à encoder spécifiquement ce type de données (ex. dans le *Wiktionnaire*, *sparadrap* dans le sens « tissu recouvert d'une sorte d'emplâtre agglutinatif » est marqué comme « désuet » mais pas comme régional, *ça lui fera les pieds* n'est pas marqué).

Les ressources du deuxième groupe ne proposent pas de marques mais laissent la liberté à leurs contributeurs de donner des indications, même sommaires, sur les conditions d'usage des lexèmes décrits. Ces indications prennent la forme de mentions indirectes, dans les définitions, autour des exemples ou dans les commentaires, sur l'endroit où un lexème est utilisé, sur le type de personnes qui l'emploient, sur le caractère ringard ou non de telle ou telle expression, etc. Ces textes libres contiennent parfois des jugements de valeur tranchés qui nous renseignent sur les normes locales en vigueur. Ainsi l'auteur de la définition la mieux classée pour *chocolatine* dans *Blazz* oppose-t-il vigoureusement les locuteurs du Sud de la France, clairement valorisés, aux locuteurs des autres régions. L'analyse de ces textes et de toutes les représentations épilinguistiques qu'ils véhiculent mériterait une recherche spécifique qui envisage des échantillons larges tirés d'une ou plusieurs ressource(s).

À ces deux groupes s'ajoute le *Dictionnaire de la Zone*, qui ne contient – pour autant que nous ayons pu en juger sans avoir consulté la totalité du contenu du dictionnaire – ni marques ni mentions indirectes, seulement des remarques de type encyclopédique (CELOTTI 2016). Le centrage sur un registre particulier, le français des banlieues françaises, explique sans doute cet état de fait, somme toute assez logique. La prise en compte de ce seul registre ne fait pas pour autant du *Dictionnaire de la Zone* un dictionnaire différentiel, les notices révisées faisant intervenir abondamment des données issues de dictionnaires scientifiques ou commerciaux notamment pour alimenter la rubrique étymologique. Dans la mesure où la révision tend également à rapprocher le contenu définitionnel des standards lexicographiques traditionnels, les éventuelles mentions indirectes proposées par les utilisateurs sont évacuées.

Ce rapide survol montre que les potentialités évoquées ci-dessus en termes de description fine des variantes, d'identification des normes locales et de localisation des usages sont exploitées de manière très inégale par les principales ressources collaboratives francophones. Très peu d'entre elles sont réellement ouvertes à la variation, spécialement diatopique, et proposent un marquage systématique efficace des variantes. Presque toutes sont, par contre, dotées de mécanismes d'évaluation de la qualité des données encodées par les utilisateurs, mais n'en tirent pas tout le parti nécessaire pour en assurer une exploitation maximale. Dans ce contexte, le *Wiktionnaire* apparaît comme le dictionnaire collaboratif le plus abouti, malgré ses imprécisions. Notre conviction est qu'il est possible de valoriser plus efficacement encore les compétences des locuteurs.

### 3. Projet *10-nous* : trois axes d'action

Le projet *10-nous*, lancé à l'Université de Liège en 2016, est un projet de lexicographie collaborative numérique accompagnée par une démarche scientifique rigoureuse destinée à créer les conditions nécessaires à la rationalisation de connaissances (méta-)linguistiques des locuteurs et à leur intégration dans une structure informatique. L'enjeu principal de la recherche est de définir les modalités permettant d'amener des contributeurs, sans formation lexicographique préalable, à alimenter une base en données linguistiques à la fois structurées, fiables, scientifiquement pertinentes et pleinement exploitables ultérieurement.

Pour que la ressource gratuite en ligne, la BOUF (Base OUverte Francophone), dont la création sous-tend le projet de recherche *10-nous*, puisse réaliser pleinement les potentialités de la lexicographie collaborative, trois grands axes de recherche sont poursuivis : la définition d'un protocole d'enquête, la validation des données encodées, la modélisation des connaissances.

#### 3.1. Protocole d'enquête

La clef de voûte du projet *10-nous* est le développement d'un protocole d'enquête permettant aux locuteurs francophones non linguistes d'encoder leurs représentations linguistiques en matière de lexique de façon suffisamment structurée pour permettre l'analyse automatisée des données. La démarche méthodologique adoptée vise dès lors à assurer l'équilibre entre l'accessibilité de la ressource à des contributeurs néophytes et la qualité scientifique des données encodées au travers de l'élaboration d'un questionnaire dynamique à la fois simple, structuré et souple, permettant à tous de proposer des témoignages variés et originaux mais cadrés quant à leur structure.

Ce questionnaire amènera le contributeur à se positionner comme le témoin d'un usage qu'il peut observer dans son milieu de vie en faisant appel à son expérience et à ses souvenirs (« c'est ma grand-mère qui dit *bayadère* », « je ne prononce pas le *t* dans *yogourt* », etc.). Sur la base de leurs pratiques linguistiques quotidiennes, les contributeurs seront amenés à fournir des éléments d'informations géographiques (« à Liège on dit *chiques* »), des informations formelles (genre, forme plurielle, prononciation), des informations grammaticales (« on dit *il s'attend à son départ* », « *s'attendre à* est un verbe »), des informations sémantico-référentielles (« les *bobettes* sont des sous-vêtements »), des informations sociolinguistiques (« *il n'y a pas personne* c'est familier au Québec », « à Liège, on dit *oufti* quand on est surpris »).

La principale innovation visée par ce questionnaire est son caractère dynamique et adaptatif. L'objectif est qu'il évolue continuellement au fil des interactions des utilisateurs avec l'interface en ligne en fonction de leurs réponses, des données déjà encodées dans la base mais aussi des données présentes sur le Web utilisé comme corpus (GATTO 2014). Il sera alors demandé aux contributeurs de commenter ou de valider des énoncés tirés du corpus : par exemple « à Québec, dit-on *la bus* ? », « Évaluez la validité des constructions suivantes : *il attend à son ami, il attend son ami, il attend pour son ami* » ou « Peut-on entendre cette phrase à Liège ? : *Tu cherches désespérément comment mettre de l'ambiance [dans ton party du temps des fêtes] ?* ».

La saisie de données en ligne permettra également, avec le consentement des contributeurs, de recueillir des métadonnées (date, localisation). Outre les données linguistiques, les contributeurs qui le souhaitent fourniront aussi leur profil sociolinguistique (sexe, âge, formation, profession, etc.). Une corrélation pourra alors être établie entre ce profil et la fiabilité des données fournies.

### 3.2. Validation et suivi

Pour évaluer la pertinence et de la qualité scientifique des données encodées dans la BOUF par les contributeurs non-linguistes, le projet *IO-nous* allie deux méthodes complémentaires.

La première est la validation des données par les utilisateurs eux-mêmes (validation par les pairs). Cette validation des données encodées par d'autres ou issues de corpus Web fait partie intégrante du processus de contribution de chaque utilisateur. Il sera ainsi demandé aux utilisateurs de se prononcer sur la qualité de définitions présentes dans la base, a fortiori quand plusieurs définitions sont proposées pour une même unité mono- ou polylexicale (*Blazz*). Un algorithme de calcul dynamique permettra de hiérarchiser les définitions, juxtaposées les unes aux autres, en fonction des évaluations. Par ailleurs, rien n'exclut que, parmi les pairs qui évaluent et valident les données, se trouvent des experts linguistes, capables le cas échéant de réorganiser des contributions (changer certaines informations de place, fusionner deux contributions, etc.) et de supprimer les contributions inadéquates.

La seconde méthode de validation des données passe par la mise en réseau de la BOUF avec d'autres ressources francophones, à caractère scientifique. Loin de se contenter d'emprunter du contenu aux dictionnaires professionnels, comme le font nombre de lexicographes amateurs dans diverses ressources collaboratives (*Bob* ou le *Wiktionnaire*, par exemple), la mise en réseau de ressources scientifiques avec une ressource collaborative, chacune d'entre elles gardant son intégrité, assurera l'existence des lexèmes répertoriés via d'autres sources et permettra de situer ces lexèmes dans un cadre plus large, historique et géographique, de façon fiable. Cette mise en réseau permettra en retour la mise à jour du contenu des ressources scientifiques.

### 3.3. Modélisation des connaissances

Pour permettre une exploitation maximale des données recueillies auprès des locuteurs, celles-ci seront extraites et visualisées de manière pleinement lisible par les utilisateurs en fonction de leur profil et besoins spécifiques. Le support numérique donne, en effet, la possibilité de générer des visualisations adaptées à des utilisations distinctes (MAZZIOTTA 2010). Cette adaptation aux utilisateurs issus des différentes communautés francophones est nécessaire pour éviter les écueils auxquels est confronté notamment le *Wiktionnaire*, dernier-né des dictionnaires panfrancophones (VINCENT 2011 et 2016). Tout l'enjeu de la création de tels dictionnaires est, en effet, de rendre compte de l'intégration des unités lexicales dans les réseaux qu'elles tissent avec d'autres dans une communauté linguistique donnée (ex. l'antonymie *fin/épais* en français européen vs. l'absence de relation antonymique entre *fin* « gentil » et *épais* « stupide » au Québec, l'usage de *écœurant* comme intensificateur ou comme antonyme de *bon*, etc.) tout en rapprochant ces réseaux pour en identifier les points communs et les différences. Or, dans le *Wiktionnaire*, l'écrasement des données inhérent au fonctionnement par wiki – dans lequel un utilisateur peut modifier une information encodée par un autre utilisateur et la remplacer par sa propre contribution – combiné à une interface centralisante pour toutes les données francophones, tendent à masquer cette double dimension. La principale difficulté à cet égard ne se pose pas au niveau des entrées, les variantes graphiques et phoniques ainsi que les différents usages régionaux étant localisés (plus ou moins) précisément, mais au niveau du contenu même des définitions. Ainsi, choisira-t-on *arachide* ou *cacahuète* pour définir *colle pistache*, qui désigne une friandise de la Réunion, ou *satay/saté*, condiment indonésien très populaire aux Pays-Bas ? Il arrive ainsi que des locuteurs issus de communautés différentes modifient à tour de rôle des éléments



d'une définition pour qu'elle corresponde mieux à leur propre usage, au détriment de celui des autres<sup>2</sup>.

Si l'on privilégie la juxtaposition des données à l'écrasement, en exploitant toutes les potentialités de l'hypertexte, il sera possible, par exemple, de relier *arachide*, *cacahuète* et *pistache* comme désignations culturellement marquées et localement situées d'une même réalité, sans rien changer à la définition de *colle pistache*, initialement encodée idéalement par un locuteur réunionnais selon son propre usage. Les descriptions qui donnent lieu à des conflits entre utilisateurs issus de régions différentes dans une ressource centralisée comme le *Wiktionnaire*, qui rassemble les différentes variantes dans une même interface, peuvent ainsi concourir à l'élaboration pixel par pixel d'une photographie riche et variée du français dans le monde dans une ressource qui compartimente mieux les données issues des différentes communautés linguistiques à la manière de la BDLP.

L'adaptation de la BOUF à ses différents utilisateurs quel que soit leur profil s'instancie donc dès le protocole d'enquête mais se manifeste aussi dans les outils d'exploitation de la base de données. L'interface de consultation doit s'adapter aux différents profils d'utilisateurs, experts ou non, québécois, belges, camerounais, suisses ou autre pour leur proposer des fonctionnalités répondant à leurs besoins (de la recherche simple à l'extraction de tout ou partie de la base de données) et pour leur donner accès à la description des usages de leurs communautés ou d'autres communautés au choix. Cette adaptation nécessite entre autres un moteur de recherche performant permettant d'accéder aux données au départ de n'importe quel point d'entrée : lexèmes décrits, éléments de définitions, exemples, variantes géolinguistiques et, plus largement, tout le système de marques proposé aux contributeurs pour décrire le registre de langue, la zone géographique et plus généralement les conditions d'usage des lexèmes.

L'autre innovation proposée par le projet *10-nous* est la visualisation sur carte de la répartition géographique des données. Cette fonctionnalité rendra possible une comparaison panfrancophone et donnera à voir les zones d'emploi des lexèmes. Par voie de conséquence, cette visualisation permettra ainsi de documenter le lexique commun mettant en lumière les lexèmes les mieux répartis à travers toute la francophonie. Cette objectivation sera d'autant plus intéressante qu'elle pourrait montrer que le lexique commun ne se superpose pas nécessairement au lexique français de référence. L'interface de cartographie permettra également l'intégration de données qualitatives obtenues à partir des votes des utilisateurs et destinés à hiérarchiser les données par pertinence. La visualisation sur carte de données linguistiques pose des questions méthodologiques cruciales quant aux moyens d'assurer une lisibilité maximale de la représentation de la variation (syntaxique, sémantique, pragmatique) de l'usage d'un ou plusieurs lexèmes. Pour répondre à ces questions, le projet *10-nous* pourra bénéficier de l'expertise développée à cet égard dans le cadre notamment de la rédaction de l'ALW et du projet « Français de nos régions ».

#### 4. Conclusion

Le projet *10-nous* est né de la constatation que l'immense potentiel descriptif de la lexicographie collaborative quant à la variation et à la norme n'est pas encore pleinement exploité faute d'un accompagnement adéquat des locuteurs nécessaire pour les amener à fournir des données ciblées, pertinentes et détaillées. Cet accompagnement prendra la forme d'un protocole d'enquête interactif et dynamique, accessible à des contributeurs sans formation particulière en linguistique et en informatique.

---

<sup>2</sup> Pour une réflexion plus approfondie sur cette question, voir Vincent, dans ce volume.

La définition d'un tel protocole pose des questions méthodologiques fondamentales auxquelles *10-nous* vise à répondre et ouvre la voie à des recherches innovantes dans les domaines de l'évaluation de la qualité des données issues de la production participative, de la modélisation des connaissances et de l'optimisation des interfaces humains-machines.

## 5. Références

BDLP : <http://www.bdlp.org/>  
Blazz : <http://www.blazz.fr/>  
Bob : <http://www.languefrancaise.net/Bob/Introduction>  
Cordial : <http://dictionnaire.cordial-enligne.fr/index.php?mot=>  
Dico des Mots : <http://dico-des-mots.com/>  
Dico2rue : <http://www.dico2rue.com/dictionnaire/>  
Grand Robert : <https://gr.bvdep.com/robert.asp>  
La Parlure : <http://www.laparlure.com/>  
Le Dictionnaire de la Zone : <http://www.dictionnairedelazone.fr/>  
Magène, Langue et chansons normandes : <http://magene.pagesperso-orange.fr/index.html>  
Projet Babel : <http://projetbabel.org/index.php>  
Projet PFC : <http://www.projet-pfc.net>  
Reverso : <http://dictionnaire.reverso.net/francais-definition/>  
TLFi : <http://atilf.atilf.fr/tlf.htm>  
Urban Dictionary : <https://www.urbandictionary.com/>  
Urbandico : <http://www.urbandico.com/>  
Usito : <https://www.usito.com/>  
WikiManche : <http://www.wikimanche.fr/Accueil>  
Wiktionnaire : [https://fr.wiktionary.org/wiki/Wiktionnaire:Page\\_d%E2%80%99accueil](https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil)

ABEL, Andrea & MEYER, Christian, « The dynamics outside the paper: user contributions to online dictionaries », in KOSEM I., KALLAS J., GANTAR P., KREK S., LANGEMETS M., TUULIK M. (eds), *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 2013, pp. 179-194.  
[http://eki.ee/elex2013/proceedings/eLex2013\\_13\\_Abel+Meyer.pdf](http://eki.ee/elex2013/proceedings/eLex2013_13_Abel+Meyer.pdf)

ALW : *Atlas linguistique de la Wallonie* (1953- ). Liège, Presses universitaires de Liège, 10 tomes parus.

AVANZI, Matthieu, BARBET, Cécile., GLIKMAN, Julie & PEUVERGNE, Julie, « Présentation d'une enquête pour l'étude des régionalismes du français », *Actes du Congrès Mondial de Linguistique Française, SHS Web of Conferences 27*, 03001, 2016.  
[https://www.shs-conferences.org/articles/shsconf/pdf/2016/05/shsconf\\_cmlf2016\\_03001.pdf](https://www.shs-conferences.org/articles/shsconf/pdf/2016/05/shsconf_cmlf2016_03001.pdf)

BARONIAN, Luc & MARTINEAU, France, *Le français d'un continent à l'autre. Mélanges offerts à Yves Charles MORIN*. Québec, Presses de l'Université de Laval, 2009.

BERNET, Charles. & RÉZEAU, Pierre, *Dictionnaire des expressions quotidiennes - On va le dire comme ça*. Paris, Le Livre de poche, 2<sup>e</sup> édition, 2010.

BEYER, Thomas, CORNÉLIS, Mélanie, RENDERS, Pascale et STEFFENS, Marie, « La plateforme 10<sup>Nous</sup> : une nouvelle ressource pour l'enseignement », *Actes du XIV<sup>e</sup> Congrès*

*Mondial de la Fédération Internationale des Professeurs de Français*, Liège (14-21 juillet 2016).

CELOTTI, Nadine, « Voies novatrices du *Dictionnaire de la Zone, tout l'argot des banlieues* en ligne : à l'écoute des voix sur la toile », in *Publifarum : Du labyrinthe à la toile / Dal labirinto alla rete*, 26, 2016.

[http://www.publifarum.farum.it/ezine\\_articles.php?publifarum=116d451d75f6d51bead3553b68a64b9d&art\\_id=375](http://www.publifarum.farum.it/ezine_articles.php?publifarum=116d451d75f6d51bead3553b68a64b9d&art_id=375)

CHAMBON, Jean-Pierre, « Lexicographie et philologie : réflexions sur les glossaires d'éditions de textes (français médiéval et préclassique, ancien occitan) », in *Revue de linguistique romane*, 70, 2006, pp. 123-141.

COTTER, Colleen. & DAMASO, John, « Online Dictionaries as Emergent Archives of Contemporary Usage and Collaborative Codification », *QMOPAL - Queen Mary's Occasional Papers Advancing Linguistics*, 2007.

<http://linguistics.sllf.qmul.ac.uk/linguistics/media/sllf-migration/departement-of-linguistics/09-QMOPAL-Cotter-Damaso.pdf>

CREESE, Sharon, « Exploring the Relationship between Language Change and Dictionary: Compilation in the Age of the Collaborative Dictionary » in I. KOSEM, J. KALLAS, P. GANTAR, S. KREK, M. LANGEMETS, M. TUULIK (2013) eds, *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 2013, pp. 392-406.

DOLAR, Kaja, *Les dictionnaires collaboratifs en tant qu'objets discursifs, linguistiques et sociaux*, thèse soutenue le 13 janvier 2017 à l'Université Paris Ouest Nanterre La Défense, 2017.

DULONG, Gaston & BERGERON, Gaston, *Parler populaires du Québec et de ses régions voisines : Atlas linguistique de l'Est du Canada*, OLF, 1980.

FUERTES-OLIVERA, Pedro, « The function theory of lexicography and electronic dictionaries: WIKTIONARY as a prototype of collective free multiple-language internet dictionary » in H. BERGENHOLTZ, S. NIELSEN & S. TARP (eds), *Lexicography at a crossroads: dictionaries and encyclopedias today, Lexicographical Tools Tomorrow*. Bern, Peter Lang, 2009, pp. 99-134.

GATTO, Maristella, *Web as corpus. Theory and practice*, New York, Bloomsbury, 2014.

GLESSGEN, Martin-Dietrich & THIBAUT, André (eds), *La lexicographie différentielle du français et le Dictionnaire des régionalismes de France. Actes du Colloque en l'honneur de Pierre RÉZEAU*, Strasbourg, Presses Universitaires de Strasbourg, 2005, pp. III-XVII.

GREUB, Yan, *Les régionalismes lexicaux du moyen français et la formation des français régionaux, d'après l'exemple d'un corpus de farces (1450-1550)*, sous la direction conjointe de CHAMBON J.-P. (Université Paul-Valéry [Montpellier-III], puis Université Paris-Sorbonne [Paris IV]) & ECKARD G. (Université de Neuchâtel) ; thèse soutenue le 21 juin 2002, à Neuchâtel.

GUÉRIN, Emmanuelle, « Le « français standard » : une variété située ? » in DURAND J., HABERT B. & LAKS B. (eds), *Actes du Congrès Mondial de Linguistique Française - CMLF'08*, Paris, Institut de Linguistique Française, 2008.

LAUWERS, Peter, SIMONI-AUREMBOU, Marie-Rose & SWIGGERS, Pierre, *Géographie linguistique et biologie du langage: autour de Jules Gilliéron*, Leuven, Peeters, 2002.

MAZZIOTTA, Nicolas, « Logiciel NotaBene pour l'annotation linguistique. Annotations et conceptualisations multiples », in *Recherches qualitatives. Hors-série « Les actes »* 9, 2010, pp. 83-94.

MERCIER Louis & VERREAULT Claude, « Opposer français “standard” et français québécois pour mieux se comprendre entre francophones ? Le cas du *Dictionnaire québécois français* », in *Le Français moderne*, 70/1, 2002, pp. 87-108.

MEYER Christian & GUREVYCH Iryna, « Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography » in GRANGER S. & PAQUOT M. (eds), *Electronic Lexicography*, Oxford: Oxford University Press, 2012, pp.259-291.

MURANO Michela, « La lexicographie 2.0 : nous sommes tous lexicographes ? », in *Cahiers de recherche de l'École doctorale en linguistique française*, 8, OpenstarTs, 2014, <http://hdl.handle.net/10077/10767>

POISSON Esther, « Compte rendu du *Dictionnaire universel francophone* », in *Dialangue*, 10, 1999, pp. 118-120.

POISSON Esther, « Français en usage au Québec et dictionnaires », in VERREAULT C., MERCIER L. et LAVOIE T. (eds), *Le français, une langue à apprivoiser : textes des conférences prononcées au Musée de la civilisation (Québec, 2000-2001) dans le cadre de l'exposition Une grande langue : le français dans tous ses états*, Québec, Presses de l'Université Laval, 2002, pp. 93-111.

RÉZEAU Pierre, *Dictionnaire des régionalismes de France*, Bruxelles, De Boeck/Duculot, 2001.

SMITH Rachel, « Urban dictionary: youth slang and the redefining of definition », *English Today*, 27(4), 2011, pp. 43-48.  
[https://de.ryerson.ca/DE\\_courses/uploadedFiles/6052\\_Arts/CSOC202/Modules/Module\\_11/urban%20dic.pdf](https://de.ryerson.ca/DE_courses/uploadedFiles/6052_Arts/CSOC202/Modules/Module_11/urban%20dic.pdf)

SWIGGERS Pierre, « La géographie linguistique de Jules Gilliéron : aux racines du changement linguistique », in *Cahiers Ferdinand de Saussure*, 51, 1999, pp. 113-132.

THIBAUT, André, *Dictionnaire suisse romand : Particularités lexicales du français contemporain*, Genève, Éditions ZOÉ (Nouvelle édition revue et augmentée préparée par Pierre KNECHT), 2004.

THIBAULT, André (eds), *Richesses du français et géographie linguistique: Recherches lexicographiques sur les variétés du français en France et hors de France* (t. 2), études rassemblées et supervisées par THIBAULT A., Bruxelles, Duculot, 2008.

VINCENT, Nadine, « Combien faut-il de dictionnaires pour décrire le français ? », in BERTRAND O. et SCHAFFNER I. (eds), in *Variétés, variations et formes du français*, Palaiseau, Éditions de l'École polytechnique, 2011, pp. 389-404.

VINCENT, Nadine, « La prise en compte de plusieurs variétés nationales dans un dictionnaire du français : exercice de lexicographie pratique », in MOLINARI C. et GAVINELLI D. (eds), *Espaces réels et imaginaires au Québec et en Acadie: enjeux culturels, linguistiques et géographiques*, Lingue Culture Mediazioni/Langages Cultures Médiation, Milan, Led, 2016.

VIOLETTE Isabelle, « Pour une problématique de la francophonie et de l'espace francophone : réflexions sur une réalité construite à travers ses contradictions », in *Francophonies d'Amérique*, 21, 2006, pp. 13-30.

<https://www.erudit.org/fr/revues/fa/2006-n21-fa1813513/1005362ar.pdf>

## Annexe 1

### 1) **Blazz**

- Lexique ciblé : français populaire actuel. Ouverture sur la francophonie (Québec et Congo).
- Macrostructure : regroupement des entrées ayant reçu le plus de votes des utilisateurs (*j'aime*), classement complémentaire par ordre alphabétique.
- Microstructure : définition et exemple (exemples forgés et attestations d'usage dans des chansons, des vidéos, etc.).
- Outils d'exploitation : recherche par mots-clés sur les entrées seulement.
- Évaluation de la qualité des contributions : hiérarchisation des définitions en fonction des votes des utilisateurs, rattachement à des communautés d'utilisateurs.
- Médias : liens vers des vidéos, possibilité de partage sur des réseaux sociaux.

### 2) **Bob**

- Lexique ciblé : argot, français populaire et familier.
- Macrostructure : affichage des entrées les plus récentes, possibilité d'afficher tous les lexèmes issus d'une même source ou les lexèmes rattachés à un même thème. Distinction entre notices validées, révisées, assumées par le responsable du site (« Gb ») et notices écartées. L'aspect collaboratif de la ressource est assuré par les « notices à Momo » (proposées par les utilisateurs) et par la possibilité d'enrichir les notices révisées.
- Microstructure : a minima, dans les notices à Momo, définition, souvent accompagnée d'un exemple le plus souvent forgé. Dans les notices révisées, attestations littéraires, renvois synonymiques, étymologies reprises de dictionnaires scientifiques et commerciaux (TLF, Robert, etc.), indication du nombre de source attestant un lexème (« fréquence »), valeur numérique plaçant le lexème sur une échelle d'argotité de 1 (« parfaitement académique ») à 10 (« très argotique »). Pour rendre compte de l'éventuelle évolution de cette valeur dans le temps, deux périodes sont distinguées sans que la découpe temporelle soit clairement définie (« registre ancien » et « registre moderne ») (<http://www.languefrancaise.net/Aide/Registre>).
- Outils d'exploitation : recherche par mots-clés sur les entrées et les définitions.
- Évaluation de la qualité des contributions : révision interne et possibilité pour les utilisateurs d'enrichir le contenu et de signaler des erreurs.
- Médias : quasiment inexistant hormis quelques images.

### 3) **Dico2rue**

- Lexique ciblé : français populaire actuel.
- Macrostructure : affichage des entrées sur la page d'accueil sans règle de classement apparente, affichage d'un mot au hasard, classement complémentaire par ordre alphabétique. Classement des dix entrées ayant reçu le plus de votes positifs.
- Microstructure : définition et exemple (exemples le plus souvent forgés).
- Outils d'exploitation : recherche par mots-clés sur les entrées seulement.
- Évaluation de la qualité des contributions : vote des utilisateurs (*yeah / bof*), sans que ces votes ne servent à la hiérarchisation des entrées, si ce n'est dans le classement des dix entrées ayant reçu le plus de votes positifs.
- Médias : liens vers des images ou des vidéos, vers les réseaux sociaux.

#### 4) ***Dico des mots***

- Lexique ciblé : français populaire actuel.
- Macrostructure : affichage des entrées par ordre chronologique de rédaction des plus récentes aux plus anciennes, classement complémentaire par ordre alphabétique. Classement des dix entrées ayant reçu le plus de votes positifs, des dix entrées ayant reçu le plus de votes négatifs.
- Microstructure : définition et, parfois, exemple forgé.
- Outils d'exploitation : pas de moteur de recherche.
- Évaluation de la qualité des contributions : vote des utilisateurs (*j'aime / j'aime pas*), sans que ces votes ne servent à la hiérarchisation des entrées, si ce n'est dans le classement des dix meilleures/pires entrées. Commentaires des utilisateurs.
- Médias : possibilité de partage sur les réseaux sociaux.

#### 5) ***Le Dictionnaire de la zone***

- Lexique ciblé : argot des banlieues françaises.
- Macrostructure : affichage du « mot du jour », affichage par ordre alphabétique dans le lexique (description des mots argotiques), dans le glossaire (équivalents argotiques de mots courants) et dans le répertoire d'expressions argotiques.
- Microstructure : définition, catégorie grammaticale, parfois prononciation en API et fichier audio à écouter, exemple forgé ou attesté par un extrait de roman ou de chanson, avec extrait à écouter, indications étymologiques.
- Outils d'exploitation : recherche par mots-clés sur les entrées et les définitions.
- Évaluation de la qualité des contributions : possibilité pour les utilisateurs de commenter les notices. Révision par le responsable du site, « Cobra le Cynique » (Abdelkarim Tengour). Publication de 2600 notices sous forme de livre papier en 2013.
- Médias : extraits audios, images, vidéos, possibilité de partage sur des réseaux sociaux.

#### 6) ***La Parlure***

- Lexique ciblé : français parlé, centré sur les usages québécois.
- Macrostructure : affichage des entrées par ordre chronologique de rédaction des plus récentes aux plus anciennes.
- Microstructure : définition et exemple forgé.
- Outils d'exploitation : recherche par mots-clés sur les entrées, les définitions et les exemples.
- Évaluation de la qualité des contributions : votes des utilisateurs (*pour / contre*), mais pas de hiérarchisation des entrées en fonction de ces votes et pas de révision.
- Médias : possibilité de partage sur des réseaux sociaux.

#### 7) ***Reverso***

- Lexique ciblé : tout le lexique, centration sur la France.
- Macrostructure : accès aux notices via le moteur de recherche uniquement. La partie collaborative du dictionnaire est affichée en dessous des notices du Cordial quand elles existent.
- Microstructure : définition, catégorie grammaticale, parfois prononciation au travers d'un fichier audio à écouter.
- Outils d'exploitation : recherche par mots-clés sur les entrées et les définitions.
- Évaluation de la qualité des contributions : votes des utilisateurs (*pour / contre*), mais pas de hiérarchisation des entrées en fonction de ces votes, validation par les responsables du site. Possibilité de commenter les notices.

- Médias : possibilité de partage sur des réseaux sociaux.

#### 8) *Urbandico*

- Lexique ciblé : français familier, populaire.
- Macrostructure : affichage des entrées par ordre chronologique de rédaction des plus récentes aux plus anciennes. Classement complémentaire par ordre alphabétique.
- Microstructure : définition et exemple (exemples le plus souvent forgés).
- Outils d'exploitation : recherche par mots-clés sur les entrées et les définitions.
- Évaluation de la qualité des contributions : votes des utilisateurs (*j'aime / j'aime pas*), mais pas de hiérarchisation des entrées en fonction de ces votes, si ce n'est dans le classement des cinq meilleures définitions.
- Médias : vidéos, images, possibilité de partage sur des réseaux sociaux.

#### 9) *Wiktionnaire*

- Lexique ciblé : tout le lexique, ouverture sur la francophonie.
- Macrostructure : « entrée en lumière » sur la page d'accueil, accès aux autres notices via le moteur de recherche, regroupement de lexèmes par champs lexicaux (listes alphabétiques qui renvoient aux notices).
- Microstructure : étymologie, catégorie grammaticale, prononciation en API avec parfois un fichier audio, définition(s) (le cas échéant, définitions ordonnées et caractérisées les unes par rapport aux autres), exemples le plus souvent attestés, dérivés, expressions, anagrammes, homophones, conjugaison des verbes, pluriel des noms.
- Outils d'exploitation : recherche par mots-clés sur les entrées.
- Évaluation de la qualité des contributions : évaluation par les pairs de type wiki (possibilité de modifier des informations encodées par d'autres), historique des modifications et possibilité de dialogue avec les différents contributeurs à une même notice.
- Médias : images, fichiers audios.



Annexe 2

	SF	Variation diachro	Étym	Variation diatop	Variation diastra	Norme	Évaluation qualité	Recherche mots-clés	Médias
<i>Blazz</i>	oui	B	B	B	B	B	votes + hiérarchie	entrée	vidéos + RS
<i>Bob/Momo</i>	oui	A/B	rubrique/B	C	A/B	A/B	révision	entrée + définition	rare images
<i>Dico2rue</i>	oui	B	B	B	B	B	votes	entrées	images, vidéos + RS
<i>Dico des mots</i>	oui	B	B	B	B	B	votes	/	RS
<i>Dictionnaire de la Zone</i>	oui	B	rubrique	C	C	C	révision	entrée + définition	images, vidéos, audios + RS
<i>La Parlure</i>	oui	B	B	A	B	B	Votes	entrée + définition + exemple	RS
<i>Reverso</i>	oui	A	rubrique	A	A	A	votes + révision	entrée + définition	RS
<i>Urbandico</i>	oui	B	B	B	B	B	votes	entrée + définition	images, vidéos + RS
<i>Wiktionnaire</i>	oui	A	rubrique	A	A	rubrique	wiki	entrée	images, audios

A = marquage clair

C = marquage ou mention explicites extrêmement réduits voire inexistantes

B = mention sporadique dans les définitions, dans les exemples, dans les commentaires