

Quantité vs qualité dans le *Wiktionnaire* : de la diversité... à la rigueur

Franck Sajous
CLLE - CNRS
Université de Toulouse 2

Lors de la seconde édition du colloque *Dictionnaires et culture numérique dans l'espace francophone*, Sajous et al. (2018) s'intéressaient au principe de neutralité de point de vue dans le *Wiktionnaire* et mettaient en évidence les écarts du dictionnaire à ce principe. Ces écarts se manifestaient notamment par : 1) l'utilisation d'axiologiques subjectifs dans les définitions de vins et de cépages et 2) l'alternance entre marques de distance et de prise en charge dans le domaine de l'occultisme. Le jour même du colloque, un wiktionnariste présent dans l'assistance supprimait les évaluatifs des trois définitions de vins citées. Un an plus tard, lors de la publication de l'article faisant suite au colloque, ce même wiktionnariste modifiait, seul, les définitions constituant le jeu de données fourni par Sajous et al. (2019) : suppression d'évaluatifs dans le domaine du vin et introduction systématique de marques de distance dans celui de l'occultisme. Pour les deux domaines, il reste dans le dictionnaire des entrées, non mentionnées par Sajous et al. (2019), dans lesquelles on trouve encore axiologiques subjectifs et marques de prise en charge. Dans le domaine de l'occultisme, le contributeur à l'origine des modifications a visiblement pris, seul, l'initiative d'ajouter les marques de distance : je n'ai pas trouvé trace d'une discussion concernant cette démarche dans la communauté des contributeurs (Wikidémie, journal des contributeurs, etc.). Pourtant, une volonté d'homogénéisation aurait aussi bien pu conduire à la décision inverse (suppression systématique des marques de distance). Dans le domaine du vin, le wiktionnariste aurait pu tenter de mobiliser « les foules » pour, collaborativement, passer en revue les entrées correspondantes et réviser les définitions qui le méritaient, selon des critères discutés conjointement. Cela n'a pas été le cas.

Faisant écho à des observations récurrentes, ces constats me paraissent symptomatiques du fonctionnement du *Wiktionnaire* (à l'opposé de certains poncifs qui décrivent trop hâtivement la « sagesse des foules » et une « démocratie de la connaissance ») : absence de révision systématique des entrées, décisions concernant la ligne éditoriale prises sans concertation (quand d'autres génèrent des discussions sans fin impliquant de nombreux wiktionnaristes) et collaboration relative. Depuis 2018, des faits politiques et sociétaux (mouvement des gilets jaunes, réforme des retraites, violences policières, débats sur différents types de discriminations, sans parler de la pandémie de COVID-19) ont généré l'inclusion d'entrées dont la définition pourrait difficilement être consensuelle. Mais, justement, faute d'être discutées et révisées, elles ne font souvent pas plus l'objet de

consensus que de polémiques entre contributeurs. Dans le *Wiktionnaire*, c'est souvent, pendant de longues périodes et pour de nombreuses définitions, le premier qui a parlé qui a raison. Une fois une entrée définie, les contributeurs suivants complètent l'article en ajoutant des traductions, images, etc., plutôt qu'ils ne révisent la définition. Or celle-ci a pu être formulée par un contributeur aussi bien que copiée d'un journal officiel, ou reprise sans discernement d'une « petite phrase » d'un élu ou commentateur politique. Selon les cas, une étude pointilliste du *Wiktionnaire* pourrait faire passer ce dernier tour à tour pour un dictionnaire « réac », un relai de la propagande officiel ou, à l'inverse, pour un dictionnaire militant.

Cette communication propose d'interroger le sens de l'adjectif « collaboratif » lorsqu'il qualifie un dictionnaire comme le *Wiktionnaire*. Nous illustrerons qualitativement les conséquences de révisions trop peu nombreuses, ou inadéquates, puis examinerons de manière quantitative le taux de révision des articles du *Wiktionnaire* et la diversité de ses contributeurs. Nous montrerons que les métriques de « rigueur » et de « diversité » telles que définies par Lih (2004) à propos de la Wikipédia et utilisées par Wolfer et Müller-Spitzer (2016) pour décrire les éditions anglaise et allemande de *Wiktionary* sont à réinterpréter. Nous montrerons notamment que le nombre de révisions d'une page n'est pas nécessairement un indicateur de « rigueur ». Qualifier la nature des révisions et déterminer quels sont les éléments macro- et microstructurels édités est également nécessaire à l'étude du processus d'édition collaborative.

Références

- Lih, A. (2004). Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. In *Proceedings of the 5th International Symposium on Online Journalism*, Austin, Texas.
- Sajous, F., Hathout, N. et Josselin-Leray, A. (2018). Wiktionnaire, un dictionnaire crowdsourcé trop neutre ? *Colloque international Dictionnaires et culture numérique dans l'espace francophone*. 16-17 octobre 2018, Milan, Italie.
- Sajous, F., Hathout, N. et Josselin-Leray, A. (2019). Du vin et devin dans le Wiktionnaire : neutralité de point de vue ou neutralité et point de vue ? *Études de linguistique appliquée*, 194(2), pp. 147-164
- Wolfer, S. and Müller-Spitzer, C. (2016). How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions. *Lexicos*, 26.