



# Combien y a-t-il de mots dans une langue ?

Georgette DAL

Professeure de Linguistique – Université SHS-Lille 3

**Victor Dame** : Georgette Dal, vous êtes professeure en linguistique, selon vous, combien y a-t-il de mots dans une langue ?

**Georgette Dal** : Pour répondre à cette question, il faut d'abord savoir ce qu'on compte et où on le compte. Et, ensuite on pourra peut-être répondre à la question « Combien y a-t-il de mots dans une langue ? ».

Donc tout d'abord, qu'est-ce qu'un mot ? On peut en effet considérer qu'il y a deux sortes de mots dans les langues en général. Il y a des mots pleins – techniquement, on va parler de *lexèmes* – : il va s'agir de noms, d'adjectifs et de verbes. Et des mots outils – on parlera techniquement de *grammèmes* –, que sont toutes les autres catégories de mots. Je simplifie un peu ici, mais c'est l'idée.

Ensuite, pour se limiter simplement aux noms, aux verbes et aux adjectifs – aux *lexèmes* donc –, il faut savoir si l'on va considérer comme un mot à part entière chaque variation que ce nom, cet adjectif ou ce verbe va connaître dans un contexte phrastique, ou si toutes ces variations sont à ranger sous un même mot, qui sera le nom, l'adjectif ou le verbe considéré. Un exemple : sachant qu'en français, le paradigme – la conjugaison, si vous préférez – des verbes comporte 48 cases, pour ce qui est des temps simples, il va falloir décider si je considère qu'on a, par exemple, un seul verbe *chanter* en français et que les formes *chante*, *chanterait*, *chantassions*, etc., sont des variantes de cet unique mot *chanter*, ou alors si chacune de ces 48 variations compte pour un mot à part entière.

Même en réduisant aux *lexèmes*, d'autres décisions sont encore à prendre. Que va-t-on ranger sous cette étiquette de *lexème* ? Est-ce que je vais prendre en compte les sigles ? Par exemple, en français, considérera-t-on que ONG, que VTT sont des *lexèmes*, sont des mots à part entière ? Il faut également décider de ce que l'on va faire des emprunts. Par exemple, est-ce qu'on va considérer que *geek*, emprunté à l'anglais, que *tilapia*, emprunté au tswana, sont des mots à part entière ? Il faut aussi prendre des décisions relativement à ce qu'on appelle les homonymes ou polysèmes. Par exemple, est-ce que je vais considérer qu'en français, il y a deux verbes *fuir* différents selon que je dis *Le robinet fuit* ou *fuir la foule*, ou vais-je considérer que c'est un seul mot *fuir*, susceptible d'emplois différents ?

**VD** : Si je comprends bien, la notion de mot est moins simple qu'elle n'y paraît de premier abord, et, avant de se poser la question du nombre de mots que compte une langue, il faut être au clair sur la notion de mot, tout simplement.

**GD** : Voilà, exactement, c'est tout à fait ça. Et c'est précisément une des tâches que s'assignent une partie des linguistes, en particulier les sémanticiens et les morphologues.

Deuxième temps dans ma réponse maintenant, qui est sur le segment « dans une langue ». D'abord, « dans **une** langue ». Telle que la question est formulée, elle peut donner à croire faussement que toutes les langues ont le même nombre de mots. Or ce n'est pas le cas, on le sait bien. Une langue, **les** langues devrais-je dire, sont des produits sociaux, sont des produits culturels, sont des produits historiques. Si bien que, s'il y a du sens à se poser la question, au mieux on peut se poser la question relativement à **une** langue en particulier : Combien de mots en hindi ? Combien de mots en tagalog ? Combien de mots en espagnol ?, par exemple.

Ensuite, « dans une **langue** ». Maintenant, il faut prendre comme décision ce qu'on va intégrer dans la notion de langue. Est-ce que je vais intégrer dans la langue l'ensemble de ses dialectes, l'ensemble de ses patois, l'ensemble de ses argots, l'ensemble de ses jargons, de ses langages cryptés ? Par exemple, pour le français, est-ce que je vais considérer que *zarbi* est un mot à part entière, différent du mot *bizarre*, ou vais-je le traiter comme une variante de *bizarre* ? Pour la langue, il faut également décider de ce que l'on va retenir comme registre de langue. Va-t-on se limiter au langage dit « soutenu » ? Ou y intégrera-t-on les registres dits « populaire » ou « familier » ? Enfin, dernier point sur lequel il faut prendre une décision : prendra-t-on en compte l'ensemble des mots, toutes époques confondues ? Ou se limitera-t-on à l'époque contemporaine, lors du comptage des mots ?

**VD** : Tout cela fait beaucoup de décisions à prendre, mais, malgré tout ce qui vient d'être dit, est-ce qu'on peut répondre à la question ? Si non de manière générale, est-ce qu'on a une idée du nombre de mots qu'il y a en français ? Est-ce qu'il ne suffit pas de se référer au dictionnaire ?

**GD** : Ma réponse sera à la fois « Oui » et « Non ».

Oui, ça peut être utile, et les dictionnaires, en particulier les dictionnaires commerciaux, peuvent être de bonnes ressources pour déterminer le nombre de mots d'une langue. À condition que l'on consulte le dictionnaire le plus exhaustif possible, sachant que même le dictionnaire le plus complet n'intégrera pas l'ensemble des termes spécialisés qu'on peut trouver dans une langue. Pour ce qui est du français, on va pouvoir, par exemple, prendre comme dictionnaire de référence le *Trésor de la Langue Française*, qui est un dictionnaire qui a d'abord été édité sous une forme papier, ensuite sous forme CD-ROM et maintenant qui est consultable en libre accès en ligne. Et le *TLF* comporte à peu près 100 000 mots, où la notion de mot est ici utilisée de façon un petit peu élastique, puisqu'on va considérer comme mots aussi les préfixes, les suffixes, ce type de choses. On peut également prendre comme objet de référence, pour déterminer le nombre de mots d'une langue, des projets de recherche, sur lesquels travaillent en général des linguistes. Cela fait partie de leurs tâches, en tout cas d'un certain nombre d'entre eux. Par exemple, on va pouvoir se fonder sur un

projet comme le *Trésor Général des Langues et des Parlers Français*, qui dénombre à ce jour à peu près 1 200 000 mots. On peut également se fonder sur des dictionnaires collaboratifs comme le *Wiktionnaire*, dont l'objet est précisément de dénombrer le nombre de mots de toutes les langues du monde. Fin décembre 2015, le *Wiktionnaire* enregistrait pour le français à peu près 1 500 000 mots. Voilà. Donc ma réponse était « oui ».

Mais ma réponse va être « non », les dictionnaires ne sont pas de bonnes ressources. Tout simplement parce que toutes les langues du monde n'ont pas donné lieu à la rédaction de dictionnaires. Or, ces langues qui ne sont pas décrites dans les dictionnaires ont quand même des mots. Ensuite, une langue, c'est un objet vivant – je mets de côté ici les langues mortes, précisément. C'est un objet vivant qui, en tant que tel, peut sans arrêt s'enrichir, et chaque parleur d'une langue, du français par exemple, est le créateur de sa propre langue, sans arrêt : dans ses phrases, bien évidemment, puisqu'on passe son temps à créer de nouvelles phrases. On peut supposer qu'un certain nombre des phrases qui ont été prononcées là, en quelques minutes, sont de nouvelles phrases. Mais aussi dans la production de mots, si bien que c'est en recourant non seulement aux emprunts, qui est une source possible, mais également aux règles morphologiques du système linguistique qu'a intériorisées chaque parleur du français que, sans cesse, on crée de nouveaux mots. Et ce, y compris pour des mots qui ont l'air d'être attestés dans des dictionnaires ou bien installés dans le lexique, tout simplement parce que les paroleurs ne connaissent pas forcément ce qui est enregistré ou non dans les dictionnaires.

Je vais prendre un exemple plutôt chez les enfants, mais on pourrait avoir des exemples similaires chez les adultes : si un enfant de quatre ans dit « J'aime pas qu'on encage les animaux », il est assez peu probable qu'il ait consulté le dictionnaire préalablement, se soit dit « Ah, *encager* est dans le dictionnaire, je vais l'utiliser ». Non, il a simplement utilisé ses ressources linguistiques personnelles pour créer le mot *encager*, indépendamment ou non de son attestation.

Donc, si finalement il faut vraiment répondre à la question – et je pense qu'il faut vraiment répondre à la question –, ma réponse sera simplement que chaque langue vivante comporte un nombre infini de mots.