Les Petites Affiches de Vaucluse depuis 1839

Ecrit par Laurent Garcia le 4 juillet 2024

L'Occitan fait son entrée dans l'outil de traduction de Google



Google vient d'intégrer 110 nouvelles langues à son outil de traduction Google Translate. Parmi elles, plusieurs langues régionales utilisées en France comme le Breton, le Corse et l'Occitan.

« L'intégration de l'Occitan, même si elle nous paraît tardive, témoigne de l'importance de son usage sur internet », se félicite la structure de défense de cette langue Assemblada Occitana qui rappelle que « grâce au travail de beaucoup de bénévoles, l'occitan est très présent sur internet, notamment dans Wikipédia et dans des applications diverses : WordPress, Windows… »
« Maintenant il faut améliorer la qualité de l'occitan traduit : il faut y garantir une grammaire authentique, la disponibilité des sept dialectes et la norme du Conseil de la langue occitane », insiste Assemblada Occitana.

S'il est difficile d'évaluer précisément le nombre de locuteurs, on estime tout de même qu'ils seraient près de 600 000 à parler couramment l'Occitan dans la partie Sud de la France. Le Provençal étant l'une des variantes encore utilisée par environ 200 000 personnes dans la région.

Les Petites Affiches de Vaucluse depuis 1839

Ecrit par Laurent Garcia le 4 juillet 2024

8% de la population mondiale concernée

Aujourd'hui, grâce à ces 110 nouvelles langues représentent plus de 614 millions de locuteurs (de quoi permettre à 8% de la population mondiale d'effectuer des traductions), Google Translate propose désormais 243 langues.

Afin de déterminer, l'intérêt d'une langue, Google évalue notamment le nombre de demandes reçues pour la prise en charge de la langue, le nombre de locuteurs de cette langue ainsi que la quantité de données disponibles pour former l'intelligence artificielle qui pourra faire ces traductions.

Objectif : 1 000 langues

« Il y a beaucoup de critères à prendre en compte lorsque l'on ajoute de nouvelles langues à Google Traduction, depuis les variétés de langues que nous proposons jusqu'aux orthographes spécifiques que nous utilisons, explique Constantin Foniadakis, porte-parole français de Google. Les langues intègrent énormément de variétés : variétés régionales, dialectes, normes orthographiques différentes. En fait, de nombreuses langues n'ont pas de forme standard et il est donc impossible de choisir la 'bonne' variété. Notre approche a consisté à donner la priorité aux variétés les plus couramment utilisées de chaque langue. Grâce à nos partenariats avec des linguistes spécialisés et des locuteurs natifs, nous continuons à faire de grands progrès. Et au fur et à mesure que la technologie progresse, nous allons continuer à prendre en charge encore plus de variétés de langues et de conventions orthographiques. »

A termes, Google a annoncé le lancement de la '1,000 Languages Initiative' (initiative 1 000 langues), un engagement à construire des modèles d'IA qui prendront en charge les 1 000 langues les plus parlées dans le monde, sur les 6 000 à 7 000 dialectes de la planète.