

DOSSIER DE PRESSE

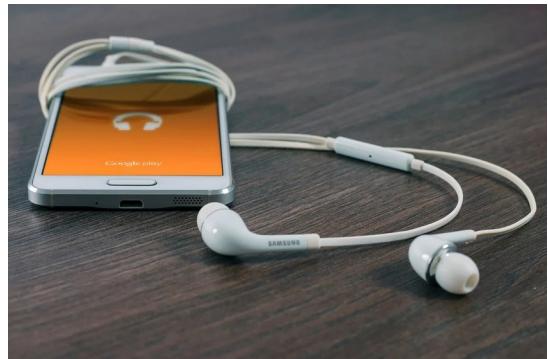


Votz

Synthèse vocale en occitan

Les ordinateurs aussi peuvent parler occitan ! La **synthèse vocale** est l'outil qui **met en voix** automatiquement un **texte écrit** et qui permet aux ordinateurs de donner des informations de façon orale.

Elle est utilisée pour des applications comme les **GPS**, les annonces dans les **transports en commun**, les logiciels de **lecture à l'écran** pour les **malvoyants** ou de **communication** pour les **personnes muettes**, les **services téléphoniques** automatiques (météo, heure)... Associée à la reconnaissance vocale, elle permet d'**échanger** oralement avec les **machines**, comme on le fait, par exemple, avec les **assistants personnels**.



Votz est la première **synthèse vocale en langue occitane**, pour l'occitan **gascon** et l'occitan **languedocien**. Elle a été réalisée dans le cadre de **Linguatec**, un **programme européen transfrontalier** qui a pour objectif le développement des technologies de la langue pour les langues aragonaise, basque et occitane.

- Pour essayer la synthèse vocale Votz : votz.eu

/ Pourquoi faire parler les machines en en occitan ?

Les technologies de la langue – reconnaissance vocale, synthèse vocale, traduction automatique ou encore analyse sémantique – sont un **enjeu vital pour les langues minorisées**. Pour se projeter vers une société de plus en plus numérisée, elles doivent disposer des ressources et des outils nécessaires pour que les locuteurs **échangent dans leur propre langue** à travers des interfaces. Plusieurs programmes ont été réalisés en ce sens pour la langue occitane : Linguatec (traduction automatique et synthèse vocale), BaTelOc (base textuelle occitane), ROLF (claviers prédictifs).

À l'heure actuelle, nous manquons d'**espaces où la langue occitane peut être entendue** et de locuteurs qui puissent la transmettre. Pourtant l'occitan en a besoin, autant pour sa **socialisation** que pour les **apprenants** de plus en plus nombreux qui ont besoin de l'avoir dans l'oreille. La synthèse vocale permettra d'amener la langue dans de nouveaux lieux, comme les transports en commun, et de transformer des situations de lecture en situations d'écoute (par exemple en synthétisant les actualités d'un site web).

/ Quelles applications utilisent la synthèse vocale ?

GPS et transports



Il serait très long d'enregistrer tous les noms de rue de toutes les villes d'un pays, ou toutes les stations de bus d'une agglomération. Ceux-ci sont lus automatiquement grâce à la synthèse vocale.

Aide pour les personnes handicapées



La synthèse vocale permet à des personnes malvoyantes de lire un écran d'ordinateur, ou à des personnes paralysées de communiquer oralement. Une des synthèses vocales les plus connues est celle du célèbre Stephen Hawking.

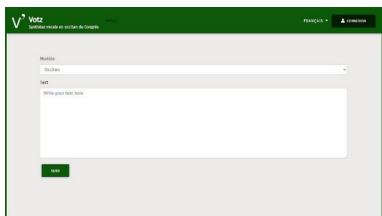
ChatBots vocaux ou assistants personnels



Plusieurs applications utilisent la reconnaissance vocale et la synthèse vocale pour communiquer à l'oral avec les utilisateurs : ChatBots téléphoniques, assistants personnels (comme Siri, Google Home ou Alexa).

/ Les applications Votz

La plateforme de synthèse



C'est l'interface publique pour utiliser Votz. Vous entrez un texte, vous choisissez la variété et vous cliquez sur un bouton pour obtenir un fichier audio avec votre contenu lu par notre synthèse vocale.

- Disponible sur votz.eu

L'API Votz



C'est le service de Votz pour les développeurs. Ils peuvent l'interroger automatiquement avec leurs propres programmes, et intégrer ainsi la synthèse vocale occitane dans leurs applications.

- À venir

Les applications



Certaines applications basées sur la synthèse vocale seront distribuées dès le mois de juin : plug-ins pour les navigateurs web, application mobile, modules pour l'intégration dans des sites internet.

- À venir

/ La synthèse vocale, comment ça marche ?

Votz fonctionne grâce à l'**intelligence artificielle**. Les réseaux neuronaux ont été entraînés avec **des heures de textes lus** par Domenja Lekuona (occitan gascon) et Clamenç Alet (occitan languedocien).

Pour plus de précision, le corpus d'entraînement et les phrases à synthétiser sont d'abord **phonétisés** grâce à l'API fonç du Congrès.

Pour être adaptée à la situation de diglossie de la langue occitane, *Votz* prend en compte la **prononciation française** d'un grand nombre de noms de personnes ou de lieux français.



I Linguatec

La synthèse vocale occitane *Votz* a été réalisée dans le cadre de **Linguatec**.

Le **projet européen EFA 227/16 LINGUATEC** « Développement de la coopération transfrontalière et du transfert de connaissances en technologies du langage » vise à développer, démontrer et diffuser de nouvelles ressources, des outils et des applications linguistiques innovantes qui améliorent le niveau de numérisation de l'**aragonais**, du **basque** et de l'**occitan**.



Il est porté par un consortium d'institutions et de centres de recherche **spécialisés dans le TAL** (Traitement Automatique de la Langue) pour les langues aragonaise, basque et occitane, qui pratiquent la **coopération transfrontalière** et le **transfert de compétences**.

Ce projet permettra la coopération technologique entre les langues pour développer de nouvelles ressources et des outils linguistiques, ainsi que de nouvelles applications en ligne pour faciliter l'utilisation des trois langues et un accès multilingue au contenu par les utilisateurs finaux.

II Les membres de Linguatec



I Les créateurs de *Votz*

II Le Congrès

Le Congrès était responsable de la réalisation de *Votz* et de la partie linguistique du travail. Il a constitué les données (corpus audio aligné) et a développé le phonétiseur occitan.



Le Congrès permanent de la langue occitane est l'organisme interrégional de régulation de l'occitan. Il œuvre dans les domaines de la linguistique et du TAL (traitement automatique de la langue).

Il produit des outils linguistiques numériques de référence (dictionnaires, conjugueurs, correcteurs orthographiques...), des applications pour le TAL (synthèse vocale, traduction automatique...) et des applications pour les mobiles (claviers prédictifs...).

Il a également des missions de régulation linguistique et de recherche scientifique appliquée.

Il est l'éditeur d'un multidictionnaire occitan (dicod'Òc) qui a chaque année plus d'un million de visites.

// Elhuyar

La fondation basque Elhuyar s'est occupée de la partie technique : entraînement des réseaux neuronaux, développement des interfaces et des applications.



L'objectif de la fondation Elhuyar Fundazioa est d'encourager, promouvoir et développer la science et la langue basques. Dans le domaine des technologies de la langue et de l'intelligence artificielle, les secteurs de la recherche sur lesquels travaille la Fondation sont, entre autres, la traduction automatique, la création automatique de ressources linguistiques, l'extraction de textes et les technologies de la parole.

Elhuyar dispose par ailleurs d'une solide expérience dans le domaine du leadership et de la participation à des projets européens de technologies de la langue. La Fondation est membre du Réseau basque de science et technologie.

// Intervenants

- Clamenç Alet : voix languedocienne
- Domenja Lekuona : voix gasconne
- Studio Arbus : enregistrement et traitement des sons
- Philippe Boula de Mareüil (Limsi) : expertise scientifique
- Nicolas Quint : expertise linguistique

Nous remercions les testeurs volontaires des deux synthèses vocales (universitaires, associatifs, conseillers pédagogiques, étudiants, institutionnels, journalistes, linguistes, artistes ou simples usagers du numérique occitan).

// Soutiens



/ Pour plus d'informations

Le Congrès permanent de la langue occitane
Château d'Este, Avenue de la Pléiade
64140 Billère

premsa.locongres.com/votz

info@locongres.org

05 59 13 06 40