



HAL
open science

Exploitation des similarités inter-dialectales pour la reconnaissance automatique de la parole en occitan

Malek Yaich, Vincent Colotte, Emmanuel Vincent, Sam Bigeard

► **To cite this version:**

Malek Yaich, Vincent Colotte, Emmanuel Vincent, Sam Bigeard. Exploitation des similarités inter-dialectales pour la reconnaissance automatique de la parole en occitan. 36èmes Journées d'Études sur la Parole, Jun 2026, Montpellier, France. ⟨hal-05604734⟩

HAL Id: hal-05604734

<https://hal.science/hal-05604734v1>

Submitted on 28 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Exploitation des similarités inter-dialectales pour la reconnaissance automatique de la parole en occitan

Malek Yaich Vincent Colotte Emmanuel Vincent Sam Bigeard

Université de Lorraine, CNRS, Inria, LORIA, France

malek.yaich@inria.fr, vincent.colotte@loria.fr,

emmanuel.vincent@inria.fr, sam.bigeard@inria.fr

RÉSUMÉ

Cet article présente une approche de fine-tuning pour la reconnaissance automatique de la parole (RAP), qui exploite les similarités linguistiques entre dialectes afin de faire face à la rareté et au déséquilibre des données dans les langues dialectalement riches et peu dotées. Cette approche est évaluée sur l'occitan, une langue composée principalement de six variantes aux ressources inégalement réparties. Pour les dialectes les mieux dotés, le languedocien et le gascon, un fine-tuning en deux étapes est appliqué afin d'améliorer les performances de reconnaissance en tirant parti des informations interdialectales. Pour les dialectes moins représentés, nous évaluons l'efficacité d'un transfert de connaissances depuis les variantes mieux dotées, en analysant la capacité du modèle à généraliser à des dialectes faiblement représentés. Les résultats montrent que l'exploitation de la proximité linguistique entre dialectes constitue une stratégie efficace pour améliorer les performances des systèmes RAP dans des contextes à faibles ressources et à forte variabilité dialectale.

ABSTRACT

Leveraging inter-dialectal similarities for automatic speech recognition in Occitan

This paper presents a fine-tuning approach for automatic speech recognition (ASR) that exploits linguistic similarities between dialects to address data scarcity and imbalance in dialectally rich, low-resource languages. The approach is evaluated on Occitan, a low-resource language composed of six main dialects with unevenly distributed resources. For the better-resourced dialects, Languedocien and Gascon, a two-step fine-tuning is applied to improve recognition performance by exploiting inter-dialectal information. For the less-represented dialects, we assess the effectiveness of knowledge transfer from the better-resourced variants, analyzing the model's ability to generalize to sparsely represented dialects. The results show that leveraging linguistic proximity between dialects is an effective strategy to enhance ASR performance in low-resource settings with high dialectal variability.

MOTS-CLÉS : Reconnaissance automatique de la parole, Langues peu dotées, Variabilité dialectale, Transfert de connaissances..

KEYWORDS: Automatic Speech recognition, Low-resource Languages, Dialectal Variation, Knowledge Transfer.

1 Introduction

Les systèmes de reconnaissance automatique de la parole (RAP) ont connu des avancées significatives grâce à l'émergence de modèles de grande taille pré-entraînés sur d'immenses volumes de données

multilingues (Baevski *et al.*, 2020; Keren *et al.*, 2025). Parmi ces modèles, Whisper (Radford *et al.*, 2023) se distingue par ses performances robustes sur un large éventail de langues, obtenues grâce à un apprentissage supervisé reposant sur plusieurs centaines de milliers d’heures d’audio diversifiés. Malgré cette couverture étendue, les performances restent toutefois très variables selon les langues et s’avèrent particulièrement limitées dans les contextes à faibles ressources.

Cette limitation est encore plus accentuée pour les langues présentant une forte variabilité dialectale. Les modèles RAP standards ne distinguent généralement pas les variantes dialectales, ce qui entraîne une dégradation notable des performances lorsqu’ils sont confrontés à des dialectes non représentés dans les données d’entraînement, comme c’est le cas pour l’arabe (Waheed *et al.*, 2023) et le farsi (Amoio *et al.*, 2025). Dans ces situations, le fine-tuning sur des dialectes spécifiques améliore la robustesse et la spécialisation des modèles.

Toutefois, cette stratégie repose sur la disponibilité de données transcrites de qualité pour chaque variante (Liang & Levow, 2025). Or, les ressources existantes sont souvent limitées et fortement déséquilibrées entre dialectes : les plateformes de collecte ouvertes, telles que Common Voice¹ (Ardila *et al.*, 2020) et ReVoc² (Lo Congrès, 2024), fournissent des données utiles, mais la distinction dialectale y est fréquemment absente ou inégale, ce qui limite l’apprentissage de représentations robustes pour l’ensemble des variantes.

Dans cette étude, nous nous intéressons à l’occitan, une langue romane parlée en France, en Espagne et en Italie, divisée en six dialectes principaux : auvergnat, languedocien, limousin, gascon, provençal et vivaro-alpin (Sibille, 2024; Bec, 1998; Quint, 2014). Bien que l’occitan soit inclus dans les données d’entraînement de modèles multilingues de grande taille tels que Whisper, les performances obtenues demeurent limitées. À titre indicatif, le modèle Whisper-large-v2 atteint un taux d’erreur sur les mots de 75,3 % lorsqu’il est évalué sur le corpus Fleurs (Conneau *et al.*, 2023), ce qui met en évidence les limites des modèles génériques lorsqu’ils sont appliqués à une langue dialectalement riche et peu dotée en ressources.

Plusieurs travaux ont étudié l’occitan et sa variation dialectale dans différentes tâches de traitement automatique de la langue. En traduction automatique, Nédey (2025) analyse le continuum dialectal occitan et souligne les limites actuelles dans la prise en compte de la diversité des variantes. Hopton & Aepli (2024) montrent que la proximité entre dialectes occitans permet d’apprendre des représentations robustes, même à partir d’une seule variante. Dans la RAP, les travaux se sont principalement concentrés sur le languedocien et le gascon (Morcillo *et al.*, 2024), laissant les autres dialectes largement sous-représentés et soulignant la nécessité de solutions adaptées aux langues minoritaires et dialectalement riches.

Dans ce contexte, nous nous proposons d’explorer comment pallier le manque de ressources pour les dialectes moins représentés de l’occitan en tirant parti des ressources disponibles. Nous cherchons notamment à exploiter les similarités linguistiques entre dialectes tout en tenant compte de la rareté et du déséquilibre des données, afin de mieux comprendre comment améliorer la robustesse et la généralisation des systèmes RAP dans un contexte de langue minoritaire et dialectalement riche.

La suite de l’article est organisée comme suit. La méthodologie proposée est détaillée en Section 2. Le protocole expérimental est présenté en Section 3, comprenant la description des corpus, du modèle utilisé et des configurations expérimentales. La Section 4 présente et discute les résultats. Enfin, la Section 5 conclut l’article et expose les perspectives de travaux futurs.

1. <https://commonvoice.mozilla.org/>

2. <https://contribuir.locongres.com/revoc/>

2 Méthodologie proposée

Certaines langues, telles que l’occitan, présentent une variabilité dialectale importante tout en conservant une forte proximité linguistique entre leurs variantes. Bien que ces variantes diffèrent sur les plans phonétique, morphologique et lexical, elles partagent une base linguistique commune.

Par exemple, selon [Hopton & Aepli \(2024\)](#), le mot *chambre* est généralement écrit *cambra* dans le dialecte provençal, et l’occlusive initiale /c/ se prononce comme un [k] vélaire. En revanche, dans les dialectes septentrionaux de l’occitan, tels que le limousin, une palatalisation du [k] initial est fréquente, se traduisant dans l’orthographe par *chambra*. Ces variations orthographiques et phonétiques illustrent que, malgré des différences de forme, les dialectes conservent des traits communs qui peuvent être exploités pour des approches d’adaptation linguistique.

Pour tirer parti de cette proximité inter-dialectale, nous adoptons différentes stratégies de fine-tuning adaptées à la distribution des ressources disponibles, comme illustré par la Figure 2.

Pour les variantes les mieux dotées en ressources, à savoir le languedocien et le gascon (voir 3.1), le fine-tuning est réalisé en deux étapes : dans un premier temps, le modèle Whisper est entraîné sur les données issues des autres variantes, ce qui permet d’exploiter les similarités entre dialectes et d’initier le modèle aux propriétés générales de la langue. Dans un second temps, le modèle est affiné sur les données correspondant à la variante cible, afin d’adapter les représentations apprises aux spécificités phonétiques et lexicales propres à cette variante.

Pour les autres dialectes moins dotés, nous évaluons l’efficacité d’un transfert de connaissances depuis les variantes mieux représentées. Concrètement, le modèle est entraîné uniquement sur les autres dialectes et testé sur la variante cible, permettant ainsi de mesurer la capacité du modèle à généraliser les connaissances acquises sur les dialectes principaux vers des dialectes faiblement représentés.

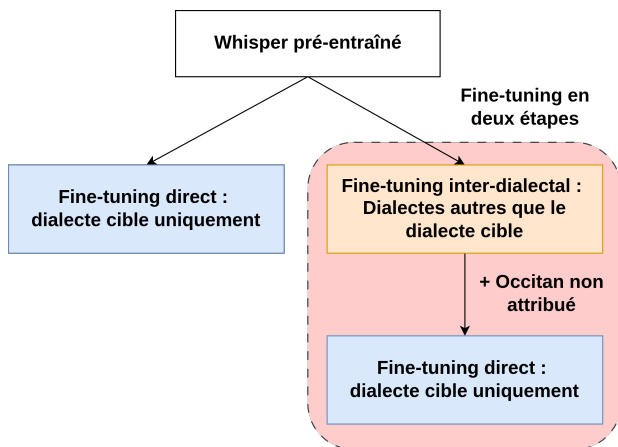


FIGURE 1 – Les différentes stratégies de fine-tuning testées : approche classique (à gauche) et approche proposée (à droite).

3 Protocole expérimental

Cette section décrit le cadre expérimental mis en place pour évaluer l’impact des différentes stratégies de fine-tuning sur la reconnaissance automatique de la parole pour les variantes dialectales de l’occitan.

3.1 Corpus

Les données utilisées proviennent de deux sources ReVOc et Common Voice :

ReVOc³ constitue une ressource spécialement conçue pour la reconnaissance automatique de la parole en occitan, développée et maintenue par Lo Congrès⁴. Il fournit des enregistrements audio alignés avec des transcriptions, explicitement associés à des variantes dialectales,

Common Voice⁵ (CV) est une collection ouverte de données vocales transcrites, couvrant un large éventail de langues. Nous avons utilisé la version 20.0, qui inclut des enregistrements en occitan, mais sans distinction explicite entre variantes dialectales. Malgré cette limitation, ces enregistrements constituent une source complémentaire utile pour enrichir la couverture linguistique et accroître la diversité acoustique des données disponibles.

Le Tableau 1 illustre la répartition des ressources disponibles par variante dialectale. Le gascon et le languedocien concentrent l’essentiel des données, tandis que les autres variantes disposent de volumes beaucoup plus limités.

Corpus	Variante dialectale	Durée
ReVOc	Gascon	51 h 43 min 58 s
	Languedocien	38 h 29 min 36 s
	Provençal	00 h 19 min 42 s
	Auvergnat	00 h 06 min 28 s
	Limousin	00 h 03 min 11 s
	Vivaro-alpin	00 h 00 min 36 s
CV	Occitan non attribué à une variante	01 h 09 min 01 s

TABLE 1 – Répartition des données audio occitanes selon les principales variantes dialectales.

3.2 Modèle de base

Les expériences sont réalisées à partir du modèle *Whisper-large-v3*⁶. L’entraînement est conduit avec un taux d’apprentissage de 1×10^{-5} et une taille de lot fixée à 8. Un arrêt anticipé (*early stopping*) basé sur la perte de validation est utilisé pour limiter le surapprentissage et sélectionner le modèle optimal. Lors de la phase d’inférence, un décodage par faisceaux (*beam search*) est appliqué avec une largeur de faisceau de 5. La langue cible est explicitement définie comme étant l’occitan, ce qui permet de contraindre le modèle à produire des transcriptions exclusivement dans cette langue.

3. <https://zenodo.org/records/11566430>

4. <https://locongres.org/fr/>

5. <https://datacollective.mozillafoundation.org/datasets?q=common+voice>

6. <https://huggingface.co/openai/whisper-large-v3>

3.3 Configurations expérimentales

Plusieurs stratégies de fine-tuning sont comparées afin d’analyser l’effet du transfert de connaissances entre dialectes sur les performances de la RAP :

- **Sans fine-tuning** : le modèle Whisper est évalué tel quel sur chaque variante dialectale, sans fine-tuning supplémentaire.
- **Fine-tuning direct** : le modèle est affiné uniquement sur les données de la variante cible.
- **Fine-tuning inter-dialectal** : le modèle est entraîné sur les données issues des autres variantes dialectales, à l’exclusion de la variante cible.
- **Fine-tuning en deux étapes** : le modèle est d’abord affiné sur les données des variantes autres que la variante cible, auxquelles sont ajoutées les données issues de CV lors de cette première phase, puis réentraîné sur les données de la variante d’intérêt.

Ces configurations permettent d’évaluer séparément l’impact de fine-tuning direct, du transfert inter-dialectal et de leur combinaison.

3.4 Données et découpage

Les expériences portent sur les six principales variantes de l’occitan : auvergnat, languedocien, limousin, gascon, provençal et vivaro-alpin (Quint, 2014). Pour chaque variante, les données disponibles sont exploitées selon les configurations décrites dans la Section 3.3. Les dialectes disposant de ressources suffisantes, en particulier le languedocien et le gascon, permettent une évaluation complète de l’ensemble des stratégies proposées. Dans ce cas, les données sont réparties en ensembles d’entraînement, de validation et de test, selon une proportion de 80 %, 10 % et 10 % respectivement.

En revanche, pour les variantes faiblement dotées, dont certaines disposent de moins de 20 minutes de données transcrites, la constitution d’un ensemble de test distinct n’est pas suffisante pour garantir une évaluation fiable des différentes stratégies. Par conséquent, seules les approches reposant sur le transfert inter-dialectal sont considérées pour ces variantes, l’ensemble des données disponibles étant utilisé à des fins d’évaluation.

4 Résultats et discussion

Le Tableau 2 présente les performances du modèle Whisper-large-v3 (1550 millions de paramètres) selon les différentes stratégies de fine-tuning évaluées, mesurées en taux d’erreur sur les mots (WER) pour les principales variantes dialectales de l’occitan.

La première partie du tableau correspond au fine-tuning du modèle Whisper-small (Morcillo *et al.*, 2024) (244 millions de paramètres) pour le gascon et le languedocien. Les deux études utilisent le même corpus ReVOC ; toutefois, les répartitions des données (apprentissage/validation/test) dans le présent travail diffèrent de celles de (Morcillo *et al.*, 2024). Cette référence permet d’évaluer l’impact de l’augmentation de la taille du modèle sur sa capacité à capturer les spécificités phonétiques et lexicales propres aux différentes variantes.

La significativité statistique des résultats a été évaluée à l’aide de la comparaison des intervalles de confiance à 95 % (Schenker & Gentleman, 2001), l’absence de chevauchement étant utilisée comme

une indication conservatrice de différence statistiquement significative.

Variante dialectale	Sans FT	FT direct	FT inter-dialectal	FT en deux étapes
Gascon (Morcillo <i>et al.</i> , 2024)	–	16.37	–	–
Languedocien (Morcillo <i>et al.</i> , 2024)	–	11.74	–	–
Gascon	81.11	11.04	43.77	10.83
Languedocien	73.34	08.76	26.33	08.40
Provençal	92.23	–	44.33	–
Auvergnat	94.72	–	72.61	–
Limousin	91.14	–	66.63	–
Vivaro-alpin	75.52	–	43.53	–

TABLE 2 – Taux d’erreur sur les mots (WER %) obtenu avec *Whisper-large-v3* selon différentes stratégies de fine-tuning pour les principales variantes occitanes. La partie supérieure du tableau rapporte les résultats publiés précédemment pour le gascon et le languedocien avec *Whisper-small* (Morcillo *et al.*, 2024). Les résultats mis en gras correspondent aux meilleures performances obtenues. Les valeurs en gras sur une même ligne ont la même significativité statistique (intervalles de confiance à 95%).

Variante	Type	Phrase
Gascon	Référence	entenèn pas mèi que las gravas de la pluja
	Prédiction	entenèm pas mei que las gravas de la ploia
Languedocien	Référence	e quand lo bèl temps es de la partida tan melhor
	Prédiction	e quan lo bel temps es de la partida tant melhor
Provençal	Référence	vieu ben que li siás ben
	Prédiction	vieu ben que l’i sesben
Auvergnat	Référence	sei ’nat vès clharmont par faire la comission
	Prédiction	sai un navèth charmon tà far la comission
Limousin	Référence	dins mon cas òm es jamai pro prudent
	Prédiction	dins mon cap un èis òmai pro prudent
Vivaro-alpin	Référence	la mando au diable
	Prédiction	la man deu diable

TABLE 3 – Exemples de transcriptions pour différentes variantes occitanes après fine-tuning inter-dialectal.

Sans fine-tuning, le modèle pré-entraîné affiche des performances limitées sur l’ensemble des variantes, avec des taux d’erreur élevés : 81,11 % de WER pour le gascon, 73,34% pour le languedocien et 75,52% pour le vivaro-alpin (l’interprétation de ce dernier reste limitée en raison d’un jeu de test très restreint de 36 secondes) et plus de 90% pour les trois autres dialectes. Ces résultats reflètent le manque de représentation de l’occitan dans les données d’entraînement initiales de Whisper.

Le fine-tuning direct sur les variantes dotées de ressources suffisantes (gascon et languedocien) permet une réduction significative du WER, confirmant l’efficacité de fine-tuning lorsque des données annotées sont disponibles. Les performances surpassent nettement celles de (Morcillo *et al.*, 2024), avec un gain relatif de 25.38 % pour le languedocien et de 32,56 % pour le gascon, soulignant

Variante	Type	Phrase
Gascon	Référence	entenèn pas mèi que las gravas de la pluja
	Prédiction	entenèn pas mei que las gravas de la pluja
Languedocien	Référence	e quand lo bèl temps es de la partida tan melhor
	Prédiction	e quand lo bèl temps es de la partida tant melhor

TABLE 4 – Exemples de transcriptions pour le gascon et le languedocien après fine-tuning en deux étapes.

également l’effet bénéfique de l’augmentation de la taille du modèle sur la capacité à capturer les spécificités phonétiques et lexicales de la langue.

Le fine-tuning inter-dialectal réduit le WER de manière significative sur toutes les variantes, avec un gain relatif compris entre 23.34% et 64.1%. Ces résultats confirment la proximité linguistique entre les dialectes et montrent que, même si le modèle n’est pas entraîné sur une variante spécifique, il peut exploiter les connaissances acquises sur les autres dialectes pour améliorer la reconnaissance. Le Tableau 3 fournit des exemples de transcriptions générées pour différentes variantes dialectales, illustrant la capacité du modèle à exploiter les similarités inter-dialectales.

La stratégie en deux étapes, combinant l’apprentissage inter-dialectal et le fine-tuning ciblé, offre les meilleures performances pour les dialectes les mieux dotés. Par rapport au fine-tuning direct, cette approche permet une réduction supplémentaire du WER de 1.9% pour le Gascon et de 4.11% pour le languedocien. Elle permet au modèle d’acquérir d’abord des représentations générales de l’occitan, en exploitant les similarités entre les dialectes, avant de se spécialiser sur une variante donnée. Des exemples de transcriptions produites par cette stratégie sont présentés dans le Tableau 4.

5 Conclusion

Dans cet article, nous avons présenté une approche exploitant les similarités entre variantes pour adapter les modèles aux dialectes et transférer les connaissances vers les dialectes moins représentés, dans le cadre de la reconnaissance automatique de la parole en occitan, langue dialectalement riche et peu dotée en ressources.

Les résultats mettent en évidence que l’exploitation des proximités linguistiques entre dialectes permet d’améliorer notablement les performances sur les variantes les mieux dotées tout en offrant une capacité de généralisation vers les dialectes faiblement représentés. Cette combinaison d’apprentissage global et d’adaptation locale s’avère ainsi une stratégie efficace pour les langues minoritaires et dialectalement riches.

Pour les travaux futurs, nous viserons une analyse fine des erreurs afin de mieux comprendre les difficultés spécifiques aux langues peu dotées. Nous explorerons également des mesures adaptées pour corriger automatiquement les transcriptions dans ces contextes à faibles ressources, afin de renforcer la robustesse et la généralisation des systèmes de RAP.

Remerciements

Ce travail s’inscrit dans le cadre du projet Défi Inria COLaF, financé par le Plan national de recherche en intelligence artificielle.

Les expériences présentées dans cet article ont été réalisées sur le banc d’essai Grid’5000, soutenu par un groupement d’intérêt scientifique hébergé par l’Inria et comprenant le CNRS, RENATER et plusieurs universités ainsi que d’autres organisations (voir <https://www.grid5000.fr>).

Références

- AMOOIE R., DE VRIES W., HAO Y., DIJKSTRA J., COLER M. & WIELING M. (2025). Enhancing standard and dialectal Frisian ASR : Multilingual fine-tuning and language identification for improved low-resource performance. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, p. 1–5.
- ARDILA R., BRANSON M., DAVIS K., KOHLER M., MEYER J., HENRETTY M., MORAIS R., SAUNDERS L., TYERS F. & WEBER G. (2020). Common voice : A massively-multilingual speech corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, p. 4218–4222, Marseille, France.
- BAEVSKI A., ZHOU Y., MOHAMED A. & AULI M. (2020). wav2vec 2.0 : A framework for self-supervised learning of speech representations. In *Advances in Neural Information Processing Systems*, volume 33, p. 12449–12460.
- BEC P. (1998). *La langue occitane*. Volume 1059 de Que sais-je ? Presses Universitaires de France, 6 édition.
- CONNEAU A., MA M., KHANUJA S., ZHANG Y., AXELROD V., DALMIA S., RIESA J., RIVERA C. & BAPNA A. (2023). FLEURS : Few-shot learning evaluation of universal representations of speech. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, p. 798–805. DOI : [10.1109/SLT54892.2023.10023141](https://doi.org/10.1109/SLT54892.2023.10023141).
- HOPTON Z. & AEPLI N. (2024). Modeling orthographic variation in Occitan’s dialects. In *Proceedings of the Eleventh Workshop on NLP for Similar Languages, Varieties, and Dialects (VarDial 2024)*, p. 78–88, Mexico City, Mexico. DOI : [10.18653/v1/2024.vardial-1.6](https://doi.org/10.18653/v1/2024.vardial-1.6).
- KEREN G., KOZHEVNIKOV A., MENG Y., ROPERS C., SETZLER M., WANG S., ADEBARA I., AULI M., BALIOGLU C., CHAN K., CHENG C., CHUANG J., DROOF C., DUPPENTHALER M., DUQUENNE P.-A., ERBEN A., GAO C., GONZALEZ G. M., LYU K., MIGLANI S., PRATAP V., SADAGOPAN K. R., SALEEM S., TURKATENKO A., VENTAYOL-BOADA A., YONG Z.-X., CHUNG Y.-A., MAILLARD J., MORITZ R., MOURACHKO A., WILLIAMSON M. & YATES S. (2025). Omnilingual ASR : Open-source multilingual speech recognition for 1600+ languages. arXiv preprint arXiv :2511.09690.
- LIANG S. & LEVOW G.-A. (2025). Breaking the transcription bottleneck : Fine-tuning ASR models for extremely low-resource fieldwork languages. In *Proceedings of the Fourth Workshop on NLP Applications to Field Linguistics*, p. 26–37, Vienna, Austria.
- LO CONGRÈS (2024). Revoc corpus. DOI : [10.5281/zenodo.11566430](https://doi.org/10.5281/zenodo.11566430).
- MORCILLO I., LETURIA I., CORRAL A., SARASOLA X., BARRET M., SÉGUIER A. & DAZÉAS B. (2024). Automatic speech recognition for Gascon and languedocian variants of Occitan. In

Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), p. 1969–1978, Torino, Italia.

NÉDEY O. (2025). La traduction automatique dialectale : état de l’art et étude préliminaire sur le continuum dialectal de l’occitan. In *Actes des 18e Rencontres Jeunes Chercheurs en RI (RJCRI) et 27ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL)*, p. 190–238, Marseille, France.

QUINT N. (2014). *L’occitan (superpack)*. Paris, France : ASSIMIL, 1^{re} édition.

RADFORD A., KIM J. W., XU T., BROCKMAN G., MCLEAVEY C. & SUTSKEVER I. (2023). Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*, p. 28492–28518. Place : Honolulu, Hawaii, USA.

SCHENKER N. & GENTLEMAN J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, **55**(3), 182–186.

SIBILLE J. (2024). Les dialectes occitans. In *Manuel de linguistique occitane*, p. 423–471. De Gruyter.

WAHEED A., TALAFHA B., SULLIVAN P., ELMADANY A. & ABDUL-MAGEED M. (2023). VoxArabica : A robust dialect-aware Arabic speech recognition system. In *Proceedings of ArabicNLP 2023*, p. 441–449, Singapore (Hybrid). DOI : [10.18653/v1/2023.arabnlp-1.38](https://doi.org/10.18653/v1/2023.arabnlp-1.38).