First Steps in ASR for Cypriot Greek: Challenges and Insights

Vivian Stamou[1], Spyros Armostis[3], Antigoni Klimi[2]
Georgios Paraskevopoulos[2], Vassilis Katsouros[2] , Antonios Anastasopoulos[1,4]
[1]Archimedes/Athena R.C., Greece
[2]Institute for Language and Speech Processing/Athena R.C., Greece
[3]University of Cyprus, Cyprus
[4]George Mason University, USA
{vistamou, antigoni.klimi, g.paraskevopoulos, vsk}@athenarc.gr
armosti.spyros@ucy.ac.cy, antonis@gmu.edu

Abstract

This paper presents the first automatic speech recognition (ASR) system for Cypriot Greek, a non-standardized variety of Modern Greek with distinctive lexicogrammatical and graphematic characteristics. We adapt Whisper, a state-of-the-art multilingual ASR model, to Cypriot Greek through fine-tuning on the Mozilla Common Voice spontaneous speech dataset for Cypriot Greek. The phonological and lexical divergence between Cypriot Greek and Standard Modern Greek poses significant challenges for mainstream ASR, particularly under conditions of limited training data and dialectal variation. Results demonstrate that whisper-medium achieved a best word error rate (WER) of 37.85%, while whisper-large-v3 consistently outperformed it, reaching a minimum WER of 33.93%. In the light of these findings, increased model size, combined with targeted fine-tuning on normalized dialectical data, significantly improves recognition accuracy, indicating that careful handling of orthographic and dialectical variation provides an effective path for ASR adaptation to low-resource varieties.

Keywords: low resource ASR, Cypriot Greek, dialectal resources

1. Introduction

State-of-the-art language and speech technologies are strongly data-driven, and language varieties with larger quantities of available data typically receive priority (Chin et al., 2023; Hedderich et al., 2020; Joshi et al., 2020). As a result, the speakers of non-dominant dialects have access to less technological support than the speakers of the dominant variety (Blasi et al., 2022). Therefore, promoting technological development for low-resource languages is vital for supporting language inclusion and preservation while enabling communities of minority languages to participate fully in the digital world (Mohanty et al., 2024).

In this work, we present the first ASR model for Cypriot Greek, a Modern Greek dialect with high vitality. In Section 2 Cypriot Greek is introduced. The adopted methodology is presented in Section 3, while Sections 4 and 5 report the experimental setup and corresponding results, respectively. Section 6 provides a brief overview of related work, and Section 7 concludes the paper.

2. Cypriot Greek

Cypriot Greek (CG) dialect belongs to the South-Eastern Greek dialect group. There are more than 700,000 speakers of CG in Cyprus and Cypriot communities abroad (UK, USA, Australia, Canada and South Africa). CG is considered the non standard variety of Modern Greek with the highest vitality (Statistical Service of Cyprus, 2011). The long time of isolation from other Greek-speaking areas led to substantial differences between Cypriot and Standard Modern Greek (SMG), occasionally rendering the two mutually unintelligible due to a host of phonological, morphological, lexical and syntactic differences. Importantly, this divergence also extends to the phonetic realization of segments and their coarticulation patterns, as shown by (Themistocleous, 2019), who successfully distinguish CG from SMG based on acoustic features of sonorant consonants and coarticulatory cues from adjacent vowels, achieving up to 81% classification accuracy using deep neural networks.

The official languages of the Republic of Cyprus are Greek and Turkish (the latter introduced to the island with the Ottoman conquest in 1571). Cypriot Greek is used in oral communication, while SMG functions in formal contexts and is the instructional language in public primary and secondary education. So far, CG has no standardized writing system, while attempts to establish language policies have never been further formalized (Papapavlou, 2010). Crucially, the SMG alphabet does not represent the distinctive sounds of the dialect, prompting several proposals for an appropriate writing system (Armostis et al., 2014). Nonetheless, there exists a body of medieval literature (13th–16th centuries), such as ‘Leontios Machairas: Recital about the sweet land of Cyprus entitled “Chronicle” (Syme-

Proceedings of Workshop on Dialects in NLP — A Resource Perspective (DialRes) , pages 308–314
16 May 2026.

onidis, 2006) that represent regional linguistic features and sounds not present in SMG, such as postalveolar fricatives and affricates, and geminates. It is the case that CG writers often devise their own spelling conventions and preferences (e.g., different written renditions of [ˈʃ ɛ r i n] 'hand' as ⟨σχέριν⟩ or ⟨σ́έριν⟩; Themistocleous, 2010).

Before 1974, Cypriot Greek displayed considerable internal variation (18 regional varieties according to Newton 1972) along a continuum that ranged from rural local dialects (basilects) to the urban variety (acrolect) (Newton, 1972; Contossopoulos, 1969). This continuum underwent homogenization due to demographic shifts after the Turkish military occupation and greater exposure to SMG (Tsiplakou et al., 2006). Nowadays, CG is largely uniform, with rare regional variants rapidly disappearing at the phonological, morphosyntactic, and lexical level (Tsiplakou, 2014).

3. Methodology

In this section, we describe the methodology employed to adapt ASR models to Cypriot Greek. We first present the data preparation pipeline, which involved manual transcription and input normalization. We then outline the fine-tuning configurations used to adapt Whisper and XLSR to the target variety.

3.1. Data collection

For ASR training, we employ version 23.0 of the Mozilla Common Voice Spontaneous Speech dataset for Cypriot Greek ('el-CY'). This resource comprises 1,284 audio clips, corresponding to approximately 11 hours of recorded speech produced by 10 speakers, transcribed and validated. The dataset is publicly accessible via the Mozilla Data Collective[1], with transcription statistics provided in Table 1.

Metric	Value
Prompts	146
Duration	39,313,800 [ms]
Avg. Transcription Length	350
Avg. Duration	30.62 [s]
Valid Duration	38,861.24 [s]
Total Hours	10.92 [h]
Valid Hours	10.79 [h]

Table 1: Transcription statistics of the el-CY dataset.

Specifically, Cypriot Greek audio recordings were transcribed by eight native speakers of the dialect. As it is a non-standardized variety, it has no codified orthography (Armostis et al., 2014). As discussed in Section 2, however, there exists a long-standing tradition in dialectal lexicography and edited literary works of representing Cypriot Greek in writing using the Greek script, supplemented with diacritics on top of certain letters to denote consonants that do not occur in SMG, as illustrated in Table 2[2]. The transcriptions were graphematically homogenized by two linguists to reduce spelling variation in the corpus; this was based both on existing dictionaries (Yiangoullis, 2014) and on the spelling conventions outlined in Armostis (2022).

Furthermore, given the absence of a standardized orthography for Cypriot Greek and the substantial variability observed in spontaneous speech transcripts, we applied a deterministic, rule-based text normalization pipeline prior to model training. The objective of this normalization was to reduce transcriptional variation that is orthographically irrelevant for ASR performance, while explicitly preserving dialectal identity and linguistically meaningful contrasts.

The normalization pipeline operates entirely at the Unicode text level and follows a fixed sequence of transformations to ensure reproducibility across training and evaluation splits.

1. **Unicode normalization and casing:** All transcripts are lowercased and normalized to Unicode NFC (Normalization Form C) form. Unicode NFD (Normalization Form D) decomposition is used internally to enable precise handling of combining diacritics, ensuring a single canonical representation for visually identical characters.

2. **Punctuation and whitespace normalization:** Quotation marks and apostrophes are unified into canonical forms, while all dash-like symbols are removed. Excess whitespace is collapsed into single spaces and trimmed (e.g., *"εν-τάξει"* → *"εντάξει"*).

3. **Removal of non-linguistic artefacts:** Non-linguistic elements such as noise markers, parenthetical comments, URLs, email addresses, phone numbers, and emojis are removed. Paralinguistic expressions are mapped to explicit placeholders (e.g., *"χαχαχα"* → *"(γέλιο)"*) to ensure consistent acoustic alignment.

4. **Normalization of elongations and repetitions:** Orthographic elongations are collapsed

[1] https://datacollective.mozillafoundation.org/datasets/cmflnuzz3xq2ok28b4o1ojacv

[2] While the transcription follows the conventions shown in Table 2, a different normalization strategy was used for the ASR experiments; further investigation into the impact of these distinct orthographic representations is reserved for future work.

Grapheme(s) & Case			Phonetic	Example	Comment
Lower	*Proper*	*Upper*	Value		
⟨σ̌⟩	⟨Σ̌⟩	—	[ʃ]	«σ̌έριν», «ίσ̌ια», «σ̌ασ̌άρω»	
⟨ς̌⟩	—	⟨Σ̌⟩	[ʃ]	«ντούς̌», «γιαβάς̌»	word-finally
⟨σ̌σ̌⟩	⟨Σ̌σ̌⟩	⟨Σ̌Σ̌⟩	[ʃː]	«σ̌σ̌ίζω», «σ̌σ̌επάζω», «μελισ̌σ̌ιά», «μάσ̌σ̌αλλα»	
⟨ζ̌⟩	⟨Ζ̌⟩	—	[ʒː]	«πουζ̌ιάζω», «αμπαζ̌ούρ», «κκεράζ̌ια»	
⟨ξ̌⟩	⟨Ξ̌⟩	—	[kʃ]	«ξ̌ιουρίζουμαι», «ταξ̌ιά», «μοναξ̌ιά», «γιναξ̌ής»	
⟨ψ̌⟩	⟨Ψ̌⟩	—	[pʃ]	«ανιψ̌ιός», «κλεψ̌ιά», «καλουψ̌ής»	
⟨τζ̌⟩	⟨Τζ̌⟩	⟨ΤΖ̌⟩	[t͡ʃ]	«τζ̌ερίν», «χότζ̌ας», «ευτζ̌ή», «αυλατζ̌ιά»	
⟨ντζ̌⟩	⟨Ντζ̌⟩	⟨ΝΤΖ̌⟩	[nd͡ʒ]	«ντζ̌ίζω», «φεντζ̌άνιν», «καντζ̌έλλιν», «πουντζ̌ιάζω»	
⟨τσ̌⟩	⟨Τσ̌⟩	⟨ΤΣ̌⟩	[tʃ^h:]	«τσ̌έκκιν», «Λατσ̌ιά», «κουτσ̌ιά»	
⟨τς̌⟩	—	⟨ΤΣ̌⟩	[tʃ]	«σάντουιτς̌», «ματς̌», «κλατς̌»	word-finally
⟨ὀ⟩	—	⟨Ὀ⟩	[o]	«πὀν», «μὄδωκεν», «πὀννά»	from /u/ + /e/ coalescence

Geminates, i.e., long consonants, are indicated in spelling with a double letter: e.g., [m] is spelled as ⟨μμ⟩, as in ⟨σήμμερα⟩ [ˈsimːeɾa] 'today'.

Table 2: Graphemes not present in the alphabet of Standard Modern Greek.

while preserving meaningful doubles (e.g., *"εεεεν"* → *"εν"*, *"ααα"* → *"αα"*). Word-initial duplicates caused by emphasis are simplified (e.g., *"εεεεσύ"* → *"εσύ"*).

5. **Polytonic-to-monotonic Greek conversion:** Legacy polytonic spellings are normalized by removing obsolete diacritics while preserving tonos and diaeresis (e.g., *"μητσ̌ιοί"* → *"μητσοί")*. Ill-formed spacing accent characters are either reattached to the appropriate vowel or discarded.

6. **Mixed Greek–Latin character correction:** Mixed Greek-Latin spellings commonly observed in Cypriot transcripts are corrected via targeted mappings from Latin to Greek characters, reducing orthographic fragmentation without altering pronunciation (e.g., *"τζ̌ια" or* "τzια"→ "τζία").

7. **Dialect-preserving lexical canonicalization:** Cypriot-specific variants are mapped to canonical Cypriot forms rather than Standard Modern Greek (e.g., *"τζαι"*, *"τζε"* → *"τζ˘αι"*). Optional normalization of legal geminates is applied depending on the experimental configuration.

To quantify the impact of normalization, we compute WER and CER between original and normalized transcripts, obtaining 15.8% and 3.8%, respectively. This indicates minimal, primarily orthographic modifications.

3.2. Data segmentation

To prepare the Cypriot Greek speech corpus for ASR fine-tuning, we segmented the audio into sentence-level units aligned with the corresponding transcriptions. Forced alignment was performed using the open-source *ctc-forced-aligner* toolkit[3] with the multilingual *mms-300m-1130* alignment model[4]. For each recording, the audio waveform was divided into overlapping windows (30 s with 2 s context) to generate CTC (Connectionist Temporal Classification) emissions, which were then aligned to the normalized transcript without requiring a pronunciation lexicon. The aligner produced token-level spans with start and end timestamps, from which sentence boundaries were derived.

The script outputs a Kaldi-style segmented dataset consisting of `wav.scp`, `text`, `utt2spk`, and `segments` files. In addition, a quality metric (`utt2score`) was computed for each utterance, combining alignment coverage, average log-probability, and duration heuristics. Segments with zero-length spans were automatically discarded, while utterances longer than 30 s were removed to match Whisper's input constraints. This procedure resulted in a clean set of sentence-level utterances with accurate time alignment, substantially improv-

[3] https://github.com/MahmoudAshraf97/ctc-forced-aligner

[4] https://huggingface.co/MahmoudAshraf/mms-300m-1130-forced-aligner

ing dataset quality and making it suitable for robust fine-tuning.

3.3. Model training

To assess the performance of state-of-the-art ASR models on Cypriot Greek, we selected two prominent ASR systems: XLSR-53 (Babu et al., 2021) and Whisper (Radford et al., 2023). XLSR-53, one of the first large multilingual wav2vec 2.0 models, was pretrained on 56,000 hours of speech across 53 languages. Whisper was trained on a much larger corpus, with the `small` model trained on 680,000 hours of weakly supervised multilingual data. For XLSR-53, we used a version further fine-tuned on Greek to improve performance on the target variety. Despite these pretraining efforts, both models exhibit reduced performance on dialectal varieties, particularly Cypriot Greek, whose phonological and lexical differences from SMG poses challenges for mainstream ASR.

To address this gap, we fine-tuned four Whisper variants on the Cypriot Greek dataset: `whisper-small`[5], `whisper-medium`[6], `whisper-large-v3`[7], and `whisper-large-v3-turbo`[8], as well as the `XLS-R-greek` model[9].

4. Experimental Setup

All experiments were conducted using the Hugging Face Transformers library[10]. Fine-tuning was performed on both the original and normalized Cypriot Greek corpus, which was split into training and evaluation subsets using an 80/20 ratio to ensure a representative validation set. We note that the train–test split is not speaker-independent.

Training proceeded for 2,048 update steps with a learning rate of 5×10^{-6}, a maximum input duration of 30,s, and an output length capped at 225 tokens. An effective batch size of 64 was used; the encoder was kept frozen while the decoder was fully trainable. To optimize GPU memory usage, gradient checkpointing and mixed-precision training (fp16) were enabled.

5. Results

In this section, we present the performance of the fine-tuned Whisper models on the Cypriot Greek test set, using Word Error Rate (WER) and Character Error Rate (CER) as evaluation metrics. Tables 3 and 4 summarize the WER and CER before and after fine-tuning for the *whisper-small*, *whisper-medium*, and *whisper-large-v3* variants, as well as the fine-tuned XLS-R-greek model, evaluated on both normalized and original transcripts.

Model	WER (%)		CER (%)	
	Before	After	Before	After
Whisper-small	81.27	**52.38**	49.21	**32.33**
Whisper-medium	69.03	**37.85**	39.69	**26.71**
Whisper-large-v3	58.33	**33.93**	31.42	**21.15**
Whisper-large-v3-turbo	61.93	**38.12**	35.06	**22.24**

Table 3: ASR performance on the Cypriot Greek test set before and after fine-tuning with normalized transcripts.

Model	WER (%)		CER (%)	
	Before	After	Before	After
Whisper-small	81.28	**51.39**	49.15	**29.85**
Whisper-medium	69.17	**43.64**	38.95	**28.22**
Whisper-large-v3	58.61	**35.70**	31.29	**22.30**
Whisper-large-v3-turbo	62.16	**36.73**	34.95	**21.88**

Table 4: ASR performance on the Cypriot Greek test set before and after fine-tuning using the original transcripts. Word Error Rate (WER) and Character Error Rate (CER) are reported.

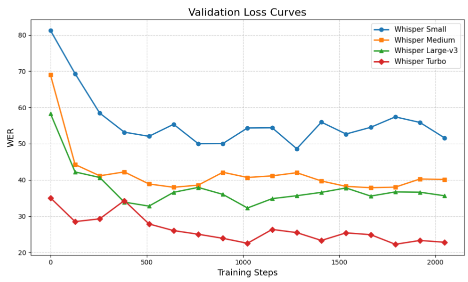

Figure 1: Validation WER curves across training steps for Whisper-small, Whisper-medium, Whisper-large-v3 and Whisper-large-v3-turbo.

As shown in Table 3 and Table 4, Whisper without fine-tuning achieves high WER scores for Cypriot Greek. We verify that this is due to the characteristics of the dialect as Whisper before fine-tuning achieves similar performance both on the original and the normalized transcripts, indicating that this is not due to transcription artifacts or rarely occurring characters and accents. Comparing Tables 3 and 4, normalization leads to consistent improvements for medium and large models (e.g., 5.8 absolute WER for Whisper-medium), while smaller models show less stable behavior, suggesting that normalization benefits models with sufficient capacity to exploit reduced orthographic variability.

[5] openai/whisper-small

[6] openai/whisper-medium

[7] openai/whisper-large-v3

[8] openai/whisper-large-v3-turbo

[9] lighteternal/wav2vec2-large-xlsr-53-greek

[10] https://github.com/huggingface/transformers/blob/main/examples/pytorch/speech-recognition/run_speech_recognition_seq2seq.py4

After fine-tuning with less than 20 hours of speech, all Whisper variants achieved significantly lower WER in both settings, with a $\sim 38\%$ and $\sim 40\%$ average relative WER reduction for the original and normalized transcripts, respectively. Among the variants, Whisper large v3 emerged as the superior model, achieving a WER of 33.93% and a CER of 21.15% on normalized transcripts. Performance consistently improves with model scale. However, although Whisper-large-v3-turbo is optimized for inference efficiency, it does not match the full large-v3 model in accuracy.

Figure 1 shows the validation WER across training steps. Larger models (*medium* and *large-v3*) not only achieve lower error rates but also converge more smoothly compared to *whisper-small*. While *whisper-small* exhibits notable fluctuations, peaking at 57% even late in training, *whisper-large-v3* stabilizes early, reaching its performance floor around step 1,000. This suggests that the higher parameter count in the large variant provides a more robust framework for dialectal transfer learning.

	Before		After	
Dataset	**WER**	**CER**	**WER**	**CER**
CY-original	84.07	48.75	**57.21**	**25.14**
CY-normalized	83.14	43.04	**46.96**	**17.88**

Table 5: ASR performance of wav2vec 2.0 XLS-R-greek on the Cypriot Greek test set before and after fine-tuning with original and normalized transcripts.

Table 5 reports the XLS-R-greek results. In the zero-shot scenario, WER exceeds 83% and CER is similarly high. Fine-tuning substantially reduces errors, with normalized transcripts leading to the best performance: 46.96% WER and 17.88% CER. Notably, XLS-R achieves WER reductions comparable to *whisper-medium* and competitive CER scores, but its WER remains higher than *whisper-large-v3*, suggesting that sequence-to-sequence modeling with an autoregressive decoder better captures word-level dependencies in dialectal speech.

Overall, the results highlight the importance of both model size and transcript normalization in ASR adaptation to low-resource dialects. Larger models capture dialectal variation more effectively, while normalization mitigates orthographically irrelevant variability. Fine-tuning on dialect-specific data is thus essential for achieving robust recognition of Cypriot Greek across different ASR architectures.

6. Previous studies

ASR models have achieved significant achievements in high-resource languages, English being the most represented instance of well-resourced language (Eisenstein et al., 2023; Pratap et al., 2024). Typical methods to deal with data scarcity issues involve transfer learning from high-resource languages, unsupervised pre-training with unpaired data, and synthetic data augmentation to improve model robustness and reduce overfitting (Conneau et al., 2020).

The most recent work on Greek dialects establishes a benchmark for six low-resource Greek varieties (Aivaliot, Aperathiot, Cappadocian, Cretan, Griko, and Messenian) and discusses the salient problems of limited data, orthographic differences, and language contact (Vakirtzian et al., 2024; Tsoukala et al., 2026). Initial experiments showed that state-of-the-art models and cross-lingual transfer methods struggle to adapt to these varieties. For instance, while ASR models handle SMG well (WER 11–14%), their performance drops drastically on Greek dialects, with WERs ranging from 28% (Cretan) to 100% (Griko). These difficulties arise from multiple factors, including significant phonological differences between SMG and the dialects, alongside variations in discourse register; dialectal speech is often recorded as oral speech rather than reading texts (Vakirtzian et al., 2024; Tsoukala et al., 2026). Consequently, varieties that exhibit higher phonological proximity to the high-resource SMG, such as the Cretan dialect, achieved the strongest performance.

7. Conclusions

This work presents the first ASR model for Cypriot Greek, addressing the challenge of data scarcity in ASR for low-resource dialects. Initial experiments indicate that Whisper struggles on non-standardized dialects (WER 60%) due to the variety-specific characteristics. This further highlights the challenge of recognizing non-standard dialects with out-of-the-box ASR models. Fine-tuning on less than 20 hours of dialect-specific speech substantially improves performance, reducing WER by 38% on original transcripts and 40% on normalized transcripts. The improvement demonstrates that even a relatively small amount of dialect-specific data can effectively adapt the model, with Whisper-large-v3 achieving the best results due to its greater capacity to capture complex speech patterns, a result that also benefits from Cypriot Greek's closer proximity to SMG as a southern Greek dialect. Moreover, these results highlight the importance of orthographic normalization when adapting ASR systems to dialectal varieties with inconsistent spelling conventions.

8. Future Work

In future iterations of this work, we plan to conduct a comprehensive comparative analysis between Cypriot Greek and Standard Modern Greek in order to quantify the "dialctal gap". Recent work on SMG (Paraskevopoulos et al., 2023) evaluated the performance of XLS-R on Standard Modern Greek (SMG) using the Mozilla Common Voice (CV) dataset. This dataset represents a "controlled" environment of read speech and consists of approximately 16.2 hours of total audio, with 12.27 hours specifically used for training. In this controlled read-speech environment, the model achieved a 29.33% WER.

In contrast, our fine-tuning results for Cypriot Greek reveal a notable performance disparity, which we hypothesize is due to the distinct phonological and morphological characteristics of the dialect, such as gemination and palatalization. Even after fine-tuning with dialect-specific data, our best-performing Whisper-large-v3 model achieved a WER of 33.93%, while our XLS-R-greek model reached 46.96% on normalized transcripts. By integrating these SMG benchmarks, we could: (i) identify the specific linguistic features of CG that contribute most to the increase in WER, (ii) evaluate whether specialized fine-tuning strategies, such as Mixed Multi-Domain Self-Supervision (M2DS2), can bridge this gap by anchoring target dialect representations to the source language, (iii) assess the cross-dialectal transferability of Whisper's zero-shot capabilities in comparison to fine-tuned self-supervised models like XLS-R.

9. Acknowledgements

This work has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.

AWS resources were provided by the National Infrastructures for Research and Technology GRNET and funded by the EU Recovery and Resiliency Facility.

10. Bibliographical References

Spyros Armostis. 2022. Η κυπριακή ελληνική τζ̆αι η γραφή της [=cypriot greek and its writing]. In Alexia Achilleos, Spyros Armostis, and Eleftheria Socratous, editors, ΑΠΟαποικιοΠΟΙΗΣΗ: Γλωσσοπλάσματα που μηχανές τζ̆αι πλάσματα *[=Decolonisation: Linguistic creations by machines and humans]*, pages 13–33. Ypogeia Skini, Limassol.

Spyros Armostis, Kyriaki Christodoulou, Marianna Katsoyannou, and Charalambos Themistocleous. 2014. Addressing writing system issues in dialectal lexicography: The case of cypriot greek. In Carrie Dyck, Tania Granadillo, Keren Rice, and Jorge Emilio Rosés Labrada, editors, *Dialogue on Dialect Standardization*, pages 23–38. Cambridge Scholars Publishing, Cambridge.

Thirunavukkarasu Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, and Michael Auli. 2021. Xls-r: Self-supervised cross-lingual speech representation learning at scale. *arXiv preprint arXiv:2111.09296*.

Damian Blasi, Antonios Anastasopoulos, and Graham Neubig. 2022. Systematic inequalities in language technology performance across the world's languages. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5486–5505, Dublin, Ireland. Association for Computational Linguistics.

Jessica Chin, Elena Talevska, and Mark Antoniou. 2023. Speech-in-speech recognition is modulated by familiarity to dialect. pages 3113–3116.

Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. 2020. Unsupervised cross-lingual representation learning for speech recognition. In *Proceedings of Interspeech*.

Nikolaos Contossopoulos. 1969. Συμβολή εις την μελέτην της Κυπριακής διαλέκτου [a contribution to the study of the cypriot dialect]. *Epetiris tu Kentru Epistimonikon Erevnon*, 3:87–109.

Jacob Eisenstein, Vinodkumar Prabhakaran, Christopher Rivera, Dorottya Demszky, and Divyansh Sharma. 2023. Md3: The multi-dialect dataset of dialogues. In *Proceedings of INTERSPEECH*, pages 4059–4063.

Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strotgen, and Dietrich Klakow. 2020. A survey on recent approaches for natural language processing in low-resource scenarios. In *North American Chapter of the Association for Computational Linguistics*.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Sushree Sangita Mohanty, Satya Ranjan Dash, and Shantipriya Parida, editors. 2024. *Applying AI-Based Tools and Technologies Towards Revitalization of Indigenous and Endangered Languages*. Studies in Computational Intelligence. Springer Singapore.

Brian Newton. 1972. *Cypriot Greek: Its Phonology and Inflections*. Mouton, The Hague.

Andreas Papapavlou. 2010. Language planning in action: searching for a viable bidialectal program. *Language Problems and Language Planning*, 34(2):120–140.

Georgios Paraskevopoulos, Theodoros Kouzelis, Georgios Rouvalis, Athanasios Katsamanis, Vassilis Katsouros, and Alexandros Potamianos. 2023. Sample-efficient unsupervised domain adaptation of speech recognition systems: A case study for modern greek. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 32:286–299.

Vineel Pratap, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Elkahky, Zhaoheng Ni, Apoorv Vyas, Maryam Fazel-Zarandi, Alexei Baevski, Yossi Adi, Xiaohui Zhang, Wei-Ning Hsu, Alexis Conneau, and Michael Auli. 2024. Scaling speech technology to 1,000+ languages. *Journal of Machine Learning Research*, 25(1):1–52.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine Mcleavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 28492–28518. PMLR.

Statistical Service of Cyprus. 2011. *Population Census 2011: Population by Language*. Statistical Service of Cyprus, Nicosia.

Christos Symeonidis. 2006. Ιστορία της Κυπριακής διαλέκτου. *Istoria tis Kypriakis dialektou. [History of the Cypriot Dialect]*. Kentro Meleton Ieras Monis Kykkou, Nicosia.

Charalambos Themistocleous. 2010. Writing in a non-standard greek variety: Romanized cypriot greek in online chat. *Writing Systems Research*, 2(2):155–168.

Charalambos Themistocleous. 2019. Dialect classification from a single sonorant sound using deep neural networks. *Frontiers in Communication*, 4.

Stavroula Tsiplakou. 2014. How mixed is a 'mixed' system?: The case of the cypriot greek koiné. *Linguistic Variation*, 14(1):161–178.

Stavroula Tsiplakou, Andreas Papapavlou, Pavlos Pavlou, and Marianna Katsoyannou. 2006. Levelling, koineization and their implications for bidialectism. In Frans Hinskens, editor, *Language Variation - European Perspectives III: Selected Papers from the 3rd International Conference on Language Variation in Europe (Amsterdam, 23-25 June 2005)*, pages 265–276. John Benjamins, Amsterdam.

Chara Tsoukala, Stavros Bompolas, Antigoni Margariti, Konstantina Panagiotou, Maria Elisavet Plaiti, Nefeli Tzanakaki, Petros Karatsareas, Angela Ralli, Antonios Anastasopoulos, and Stella Markantonatou. 2026. Extending ASR evaluation resources for Modern Greek dialects. In *Proceedings of the 13th Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 210–222, Rabat, Morocco. Association for Computational Linguistics.

Socrates Vakirtzian, Chara Tsoukala, Stavros Bompolas, Katerina Mouzou, Vivian Stamou, Georgios Paraskevopoulos, Antonios Dimakis, Stella Markantonatou, Angela Ralli, and Antonios Anastasopoulos. 2024. Speech recognition for greek dialects: A challenging benchmark. In *Proceedings of Interspeech 2024*, pages 3974–3978.

Constantinos Yiangoullis. 2014. Θησαυρός της μεσαιωνικής και νεότερης κυπριακής διαλέκτου: Ερμηνευτικός, Ετυμολογικός, Φρασεολογικός και Ονοματολογικός *[Thesaurus of Medieval and Contemporary Cypriot Greek dialect: Interpretative, Etymological, Phraseological, and Onomastic]*. (n.p.), Λευκωσία.