Extending ASR Evaluation Resources for Modern Greek Dialects

Chara Tsoukala[*1], Stavros Bompolas[*2], Antigoni Margariti[3], Konstantina Panagiotou[4], Maria Elisavet Plaiti[4], Nefeli Tzanakaki[3], Petros Karatsareas[5], Angela Ralli[2,6], Antonios Anastasopoulos[2,7], Stella Markantonatou[1,2]

[1]ILSP/Athena R.C., Greece, [2]Archimedes/Athena R.C., Greece, [3]NKUA, Greece,
[4]University of Crete, Greece, [5]University of Westminster, UK,
[6]University of Patras, Greece, [7]George Mason University, USA

Correspondence: chara.tsoukala@athenarc.gr

Abstract

Recent progress in Automatic Speech Recognition (ASR) has primarily benefited high-resource standard languages, while dialectal speech remains challenging and underexplored. We present an expanded benchmark for low-resource Modern Greek dialects, covering Aperathiot, Cretan, Lesbian, and Cappadocian, spanning southern, northern, and contact-influenced varieties with varying degrees of divergence from Standard Modern Greek. The benchmark provides dialectal transcriptions in the Greek alphabet, following SMG-based orthographic conventions, while preserving dialectal lexical and morphophonological forms. Using this benchmark, we evaluate state-of-the-art multilingual ASR models in a zero-shot setting and by further fine-tuning per dialect. Zero-shot results reveal a clear performance gradient with dialectal distance from Standard Modern Greek, with best WERs ranging from about 60-70% for southern dialects to over 80% for Lesbian and nearly 97% for Cappadocian. Fine-tuning substantially reduces error rates (up to 47% relative WER improvement), with Cappadocian remaining the most challenging variety (best WER 68.17%). Overall, our results highlight persistent limitations of current pretrained ASR models under dialectal variation and the need for dedicated benchmarks and adaptation strategies.

1 Introduction

Recent advances in Automatic Speech Recognition (ASR), driven by end-to-end architectures and large-scale pretraining, have led to substantial improvements in performance for high-resource languages and standard language varieties (Babu et al., 2021; Radford et al., 2022; Omnilingual ASR Team et al., 2025). However, these improvements do not readily extend to dialectal (e.g., Torgbi et al., 2025) and other non-standard forms of speech (e.g., Koenecke et al., 2020), where recognition accuracy degrades significantly. Dialectal variation poses challenges at multiple linguistic levels, including phonology, lexicon, morphology, syntax, and orthography, while the scarcity of annotated resources further limits the effectiveness of contemporary ASR models (Blaschke et al., 2025). As a result, speakers of regional and underrepresented varieties are still granted limited access to reliable speech technologies.

Large multilingual and self-supervised ASR models such as Whisper (Radford et al., 2022), XLS-R (Babu et al., 2021), and Omnilingual ASR (Omnilingual ASR Team et al., 2025) have become standard baselines for low-resource and cross-dialectal scenarios. Nevertheless, recent benchmarking efforts consistently show that even state-of-the-art models underperform on dialectal and accented speech compared to standard varieties, and typically require adaptation to achieve acceptable accuracy (Shi et al., 2024; Chen et al., 2025). Systematic evaluation across dialects is therefore essential to understand model behavior and remaining limitations.

Greek presents a particularly challenging test case due to its rich dialectal diversity (Kontosopoulos, 2008). Standard Modern Greek (SMG) co-exists with numerous regional and contact-influenced varieties that differ substantially in their linguistic properties, many of which are low-resource or endangered. While recent work has introduced an initial benchmark for Greek dialectal ASR (Vakirtzian et al., 2024), coverage remains limited, especially for varieties shaped by extensive language contact.

Building on this line of research, we present an expanded benchmark for ASR on low-resource Modern Greek (MG) dialects, incorporating four distinct varieties: Aperathiot, Cappadocian, Cretan, and Lesbian. Using this benchmark, we evaluate state-of-the-art ASR models from the Whis-

[*]Equal contribution.

Proceedings of the Thirteenth Workshop on NLP for Similar Languages, Varieties and Dialects, pages 210–222
March 29, 2026.

per and XLS-R families under zero-shot and fine-tuned conditions, and additionally assess Omnilingual ASR in a zero-shot setting.

Our contributions are threefold: (i) an expanded ASR benchmark covering previously unevaluated Greek dialects from both northern and southern groups, including the critically endangered Cappadocian; (ii) a systematic evaluation of multilingual ASR models with and without dialect-specific adaptation; and (iii) an empirical analysis demonstrating a strong relationship between dialectal distance and ASR performance, with word error rates ranging from approximately 60% for southern varieties to over 95% for structurally divergent and contact-influenced dialects.

2 Related Work

ASR for dialectal and non-standard speech remains a persistent challenge across many languages. A broad body of work shows that models trained primarily on standard or high-resource varieties degrade substantially when applied to regional dialects, even when those dialects are closely related. This pattern has been documented across typologically and sociolinguistically diverse settings, including Catalan (Hopton and Chodroff, 2025), Arabic (Khalafallah et al., 2024; Nasr et al., 2023), Japanese (Imaizumi et al., 2020; Toyama et al., 2024; Takahashi et al., 2024), German and Swiss German (Blaschke et al., 2025; Sicard et al., 2023), Scottish English (Torgbi et al., 2025), Telugu and other Indian languages (Aditya Yadavalli and Ganesh Mirishkar and Anil Kumar Vuppala, 2022; Bhardwaj et al., 2021; Alumäe et al., 2023), Tibetan (Qin et al., 2022), Pomak (Tsoukala et al., 2023), and numerous African languages surveyed by Imam et al. (2025).

Recent benchmarking efforts further confirm that dialectal and accented speech remains difficult even for state-of-the-art multilingual models. The ML-SUPERB 2.0 benchmark and challenge (Chen et al., 2025; Shi et al., 2024) has evaluated ASR and language identification across more than 200 languages, accents, and dialects, revealing substantial performance disparities and consistent degradation with non-standard varieties. More recently, Omnilingual ASR (Omnilingual ASR Team et al., 2025) has expanded multilingual coverage to over 1,600 languages, highlighting the growing emphasis on scale, although its robustness to dialectal variation within individual languages remains largely unexplored.

A complementary line of work addresses these issues through dialect-specific data creation and modeling. New corpora and evaluation resources have been introduced for Arabic dialects (Khalafallah et al., 2024; Nasr et al., 2023), Punjabi and Telugu (Bhardwaj et al., 2021; Aditya Yadavalli and Ganesh Mirishkar and Anil Kumar Vuppala, 2022), German dialects (Blaschke et al., 2025), and low-resource varieties such as Tibetan and Pomak (Qin et al., 2022; Tsoukala et al., 2023), highlighting the central role of dedicated datasets in dialectal ASR research.

More recent work has focused on adapting end-to-end and self-supervised ASR models to dialectal speech. Increased dialectal diversity during fine-tuning has been shown to improve robustness for Catalan (Hopton and Chodroff, 2025), while dialect-aware modeling and adaptation of large pretrained models have returned consistent gains for Japanese (Imaizumi et al., 2020; Toyama et al., 2024; Takahashi et al., 2024), Swiss German (Sicard et al., 2023), Scottish accents (Torgbi et al., 2025), and dialect-rich Indian languages (Alumäe et al., 2023). Nonetheless, performance remains sensitive to the specific variety, data availability, and evaluation choices.

Despite this growing body of work across diverse language families, MG remains underrepresented. Existing ASR research has largely focused on SMG, with limited attention to dialectal variation. While domain-specific adaptation has been explored, for example in medical dictation (Georgilas and Stafylakis, 2025), systematic evaluation across MG dialects remains largely unexplored. To date, the benchmark introduced by Vakirtzian et al. (2024), covering Aivaliot, Eastern Cretan, Griko, and Messenian, constitutes the only dedicated effort in this direction.

We extend this line of work by introducing four additional varieties, including Cappadocian, a typological outlier shaped by extensive contact with Turkish, and providing systematic evaluation across a broader range of ASR models.

3 Dialectal Scope of the Study

This study examines four MG varieties: Aperathiot (spoken on the island of Naxos), Cappadocian, Cretan, and Lesbian. According to Trudgill (2003, 59–60), Aperathiot, as part of the Naxos varieties, belongs to the Northern

Figure 1: Geographic distribution of the Greek dialects examined in this study; squares indicate data collection sites.

Cyclades group, Cretan to the southern group, and Lesbian to the northern group of MG dialects. Cappadocian falls outside this classification due to its development in isolation and under strong Turkish influence (Karatsareas, 2011, 45). Together with SMG, these varieties form a heterogeneous dialectal landscape (Figure 1) well suited for evaluating ASR under conditions of dialectal variation.

Aperathiot (Glottocode: `cycl1238`) is spoken primarily in Apeiranthos, a mountainous village in central Naxos. Linguistically, it forms a distinct dialectal enclave, differing markedly from other varieties of Naxos across all linguistic levels (Oikonomidis, 1952, 216, 272–273). Its origins have been linked to historical Cretan settlement, based on linguistic and cultural parallels, though this remains debated (Xefteri, 2009). Increased mobility and tourism in recent decades have reduced intergenerational transmission, heightening the urgency of documentation and technological support.

Cappadocian (ISO 639-3: `cpg`; Glottocode: `capp1239`) refers to a group of MG dialects historically spoken in Cappadocia (present-day Turkey). Prolonged contact with Turkish, combined with long-term isolation from other Greek varieties, triggered extensive structural innovation, making Cappadocian a typological outlier within MG. Beyond pervasive lexical borrowing, Turkish influence is reported in morphosyntax and clause structure, including nominal and verbal morphology, argument marking, and constituent order. These innovations are widely discussed in the contact-linguistic literature (see Dawkins, 1916; Karatsareas, 2011; Thomason and Kaufman, 1988, 93–94, 215–222 for discussion and examples). Following the population exchange of the early 1920s, speakers were resettled in Greece and rapidly shifted toward SMG. Although long considered extinct, native speakers were identified in the mid-2000s through fieldwork by Janse and Papazachariou. Today, only one variety survives, Mišiotika, which is generally regarded as being heavily influenced by Turkish (Dawkins, 1916, 209; Bompolas, 2023, 165–170). Mišiotika also exhibits phonological features resembling *northern vocalism*, discussed below in relation to Lesbian. However, these features are assumed to have developed independently and do not place Cappadocian within the traditional northern-southern division of MG dialects (Dawkins, 1916, 192–193). Mišiotika is spoken by only a few hundred, mostly elderly speakers and is classified as critically endangered (UNESCO, 2010).

Cretan (Glottocode: `cret1244`) is spoken across Crete and in diaspora communities. Its development reflects long-term geographic isolation and successive periods of foreign rule, including Arab, Venetian, and Ottoman domination (Pangalos, 1955; Kontosopoulos, 2008, 28–41). Cretan is traditionally divided into Eastern and Western subvarieties, with a boundary that roughly coincides with the administrative division between the prefectures of Rethymno and Heraklion (Pangalos, 1955, 143–151). The Eastern variety is generally described as more homogeneous, whereas the Western variety exhibits greater internal diversity (Kontosopoulos, 2008, 36). Unlike most MG dialects, Cretan remains robust and widely used. Our benchmark includes data from both subvarieties, extending previous resources that focused exclusively on Eastern Cretan (Vakirtzian et al., 2024).

Lesbian is the only northern MG dialect examined in this study (Glottocode: `nort2600`). It is characterized by the so-called *northern vocalism*, including vowel raising and deletion: unstressed mid vowels /e/ and /o/ are raised to [i] and [u], respectively, while unstressed high vowels /i/ and /u/ are deleted. These features clearly distinguish Lesbian from southern dialects, including SMG (Chatzidakis, 1905). The dialect has been shaped by prolonged contact with Italo-Romance and Turkish, primarily affecting the lexicon and morphology (Ralli, 2015, 2019a,b; Alexelli, 2021). Population movements between Lesbos and nearby Asia Minor (e.g., Ayvalik and Moschonisia), followed by refugee resettlement

after 1922, have contributed to notable intra-dialectal variation. Unlike many MG dialects, Lesbian remains vital and continues to function as the primary means of everyday communication on the island.

4 Datasets

Our benchmark consists of naturalistic speech datasets (narratives, conversations, and everyday-life stories) recorded from speakers of each variety, with transcriptions manually verified by trained linguists. Corpus statistics are reported in Table 1.

Aperathiot The Aperathiot dataset was collected in 2025 in the village of Apeiranthos and consists of narratives, conversations, and everyday-life stories from four native speakers (1 male, 3 female), ranging from middle-aged to elderly. Initial transcriptions were generated using Whisper Large-v3 and subsequently manually corrected by a trained linguist who is also a native speaker of the dialect.

Cappadocian The Cappadocian dataset consists of conversations, narratives, and everyday life stories recorded during fieldwork conducted in 2011 by twelve native or heritage speakers (8 male, 4 female) of Mišiotika in a village in Northern Greece. Speakers' ages range from 19 to 94 years; the group includes both individuals born in Misti (Cappadocia) prior to the population exchange and later-generation speakers born and raised in Greece. The village where the recordings were collected is marked with a square on the map in Figure 1. The audio recordings were manually segmented into utterances and transcribed by two trained native speakers; in addition, parallel translations into SMG were provided as an auxiliary resource (and are not used as the reference transcription in our ASR evaluation).

Cretan The Cretan dataset combines previously published material with newly collected recordings. The previously published material consists of approximately two hours of already-processed transcribed radio broadcasts representing the Eastern Cretan variety, originally published by Vakirtzian et al. (2024)[1] and recorded between 1998 and 2001 by Radio Mires in the Messara region of Heraklion. To extend dialectal coverage beyond the Eastern variety, we collected additional recordings in 2025 from two elderly native speakers (1 male, 1 female) from Western Crete, consisting of natural conversations and narratives. The villages where the new Western Cretan data were recorded are marked with squares on the map in Figure 1. The transcriptions of the new data were generated using Whisper Large-v3 and manually corrected by trained native-speaker linguists. All transcriptions follow the orthography of SMG and adhere closely to the conventions used by Vakirtzian et al. (2024) in the Eastern Crete corpus to ensure comparability of subsets.

Corpus	Tokens	Utterances	Audio Duration Original	Audio Duration Processed
Aperathiot[2]	8,830	798	1h 24m 17s	1h 3m 27s
Cappadocian[3]	11,715	2,357	3h 29m 42s	1h 17m 5s
Cretan[4]	36,594	6,897	4h 59m 47s[5]	3h 56m 54s
Lesbian[6]	11,652	2,294	2h 29m 14s	1h 6m 12s
Total	**68,791**	**12,346**	**12h 23m 0s**	**7h 23m 38s**

Table 1: Summary Statistics of the Speech Corpora.

Lesbian The Lesbian dataset was collected through fieldwork conducted in 2023–2024 and consists of narratives, conversations, and everyday-life stories from eleven native speakers (5 male, 6 female) originating from eight villages in northern Lesbos. Initial audio-aligned transcriptions were generated using Whisper Large-v3 and subsequently manually corrected by a trained linguist who is also a native speaker of the dialect. Transcription follows the benchmark policy described below (Section 4.1): dialectal forms are rendered in the Greek alphabet using SMG-based grapheme–phoneme correspondences, with systematic adaptations to capture characteristic phonological patterns of northern vocalism. Specifically, the raising of /o/ (orthographic o, ω) is transcribed as ov, and the raising of /e/ (orthographic $\alpha\iota$, ϵ) as ι. The deletion of unstressed high vowels /u/ and /i/ (corresponding to the relevant Greek orthographic representations) is marked by an apostrophe only in word-final position; word-internal deletions are not represented, in order to preserve compatibility with standard orthographic conventions.

[1] huggingface.co/datasets/ilsp/cretan-speech-corpus

[2] huggingface.co/datasets/ilsp/aperathiot-speech-corpus

[3] huggingface.co/datasets/ilsp/cappadocian-speech-corpus

[4] huggingface.co/datasets/ilsp/cretan-extended-speech-corpus

[5] For Cretan, "Original" reflects the Eastern corpus as published plus unprocessed Western recordings.

[6] huggingface.co/datasets/ilsp/lesbian-speech-corpus

4.1 Transcription Policy and Benchmark Scope

The benchmark provides dialectal transcriptions (e.g., Kuparinen, 2025), not dialect-to-standard renderings (e.g., Blaschke et al., 2025; Ducceschi and Franzini, 2025; see also Dimakis et al., 2025 on dialect normalization for Greek in NLP). All varieties are transcribed using the 24-letter Greek alphabet and an orthography anchored in SMG grapheme-phoneme correspondences; i.e., we employ SMG-based spelling conventions as an orthographic substrate while preserving dialectal lexical and morphophonological forms. This choice is motivated by the fact that the dialects covered here do not have a widely adopted standardized orthography; therefore, we follow an ad-hoc pronunciation-oriented spelling practice that is common in modern dialectal texts for these varieties, aiming for consistency across corpora and readability for Greek-literate users. Importantly, we do not provide a parallel normalized to SMG reference transcription; accordingly, the benchmark is intended for dialectal ASR (recovering the dialectal wording in SMG-based Greek orthography), rather than dialect-to-standard ASR/speech-to-standard translation.

4.2 Data Anonymization

All speech data were anonymized prior to release, at both the audio and transcription levels. For Aperathiot and Cretan, identifying information was removed by muting audio segments and deleting aligned transcriptions, while for Cappadocian and Lesbian, sensitive segments were excluded during transcription. This process was performed manually by trained linguists to ensure speaker privacy while preserving linguistic integrity.

5 Experiments

5.1 Models

To assess the robustness of current ASR systems to dialectal variation in MG, we evaluate a range of widely used multilingual speech recognition models under both inference-only and fine-tuned settings. Our benchmark focuses on models from the Whisper (Radford et al., 2022) and XLS-R (Babu et al., 2021) families, as well as Omnilingual ASR (Omnilingual ASR Team et al., 2025), all of which have emerged as standard reference points for low-resource and cross-dialectal ASR.

Inference-only evaluation All dialects are evaluated under zero-shot (inference-only) conditions using pretrained models without dialect-specific adaptation. We report results for representative models from the Whisper and XLS-R families, as well as Omnilingual ASR, enabling comparison across architectures, model sizes, and levels of language-specific adaptation. Specifically, we evaluate Whisper Large-v3, Whisper Large-v2, Whisper-medium, XLS-R-53-greek, XLS-R-300-greek, and Omnilingual ASR.

XLS-R is a multilingual speech encoder trained on approximately 56k hours of audio from 53 languages. In our experiments, we use Greek-adapted XLS-R variants, which provide stronger baselines for MG and closely related varieties than the original multilingual checkpoints. Whisper, in contrast, is a large-scale multilingual sequence-to-sequence model trained on substantially larger and more heterogeneous data: Whisper Large-v2 was trained on roughly 680k hours of weakly supervised audio, while the most recent Large-v3 extends training to approximately 1 million hours of weakly labeled audio and an additional 4 million hours of pseudo-labeled data generated using Large-v2. Finally, Omnilingual ASR is a large-scale multilingual model supporting over 1,600 languages and is evaluated exclusively under inference-only conditions.

Fine-tuning In addition to inference-only evaluation, we examine the impact of dialect-specific adaptation by fine-tuning XLS-R-53-greek and Whisper-medium for 35 epochs. Fine-tuning is performed separately for each dialect included in the benchmark.

Larger Whisper variants (Large-v2 and Large-v3) were not fine-tuned due to their high computational cost and the increased risk of overfitting given the limited amount of available dialectal data. Omnilingual ASR was likewise evaluated only in a zero-shot setting, as its scale and training setup make dialect-specific fine-tuning impractical in low-resource scenarios. Similarly, XLS-R-300-greek was not fine-tuned, since preliminary experiments showed no clear gains over the smaller XLS-R-53-greek variant under low-resource conditions.

5.2 Preprocessing

All datasets were processed using a unified preprocessing pipeline to ensure consistency across dialects. All preprocessing, fine-tuning, and evalu-

ation scripts are available at https://github.com/athena-ilsp/greek-dialects-asr. Audio recordings were converted to mono WAV files at a sampling rate of 16 kHz. Text normalization consisted of lowercasing the text and removing punctuation.

5.3 Segmentation

ASR models require relatively short audio segments for both training and evaluation. All recordings were therefore segmented into utterances with a maximum duration of 30 seconds. For Cappadocian, where time-aligned transcriptions were already available, segmentation was performed automatically using the existing annotations.

For the remaining dialects, which include newly collected material, audio recordings were first transcribed using Whisper, and the resulting timestamp information was converted into Praat TextGrid files (Boersma and Weenink, 2001). Next, these files were manually reviewed and corrected by trained native-speaker linguists, who refined both segment boundaries and transcriptions. Non-speech material, such as long pauses, untranscribed portions, and, where applicable, the utterances of the fieldworker, was removed. As a result, the effective audio duration was reduced for some datasets, as reflected in Table 1.

5.4 Dataset Construction

Following segmentation, a separate dataset was constructed for each dialect using the resulting audio–text pairs. Utterances exceeding 30 seconds were excluded. The material was then split into training, development, and test sets using an 80/10/10 split to enable supervised adaptation and evaluation.

5.5 Fine-tuning

Fine-tuning experiments were carried out using XLS-R-53-greek and Whisper-medium with model selection based on validation WER (best checkpoint loaded at the end of training). XLS-R-53-greek was trained on an NVIDIA GeForce RTX 3090 GPU for up to 35 epochs with early stopping, using a learning rate of 3×10^{-4}, batch size of 8, and gradient accumulation of 2. Whisper-medium was trained on an NVIDIA A100 with a learning rate of 10^{-5} and the same batch configuration; training used step-based evaluation and checkpointing every 1000 steps up to a maximum of 10000 update steps. Parameter-efficient fine-tuning methods (e.g., LoRA) are left for future work and would enable extending fine-tuning to larger checkpoints under the same compute budget.

6 Results

This section reports the performance of the evaluated ASR models on Greek dialectal speech. We first present inference-only (zero-shot) results obtained with pretrained models, followed by results after dialect-specific fine-tuning. Performance is measured using Word Error Rate (WER) and Character Error Rate (CER), computed on normalized text without punctuation.

6.1 Inference-only Evaluation

ASR performance under inference-only conditions, measured by WER and CER, is reported in the upper parts of Tables 2-5. Across dialects, WER and CER exhibit distinct but complementary patterns that reflect both the degree of dialectal divergence from SMG and the ability of different model architectures to capture subword structure under mismatch.

For Aperathiot and Cretan, both southern dialects sharing core phonological and morphosyntactic properties with SMG, error rates are consistently lower than for the remaining varieties. Whisper Large-v3 achieves the best zero-shot performance in both cases (Aperathiot: WER 61.85%, CER 33.08%; Cretan: WER 70.24%, CER 41.74%), indicating that large multilingual sequence-to-sequence models are relatively robust when dialectal variation remains close to the standard language.

The Omnilingual ASR model performs competitively for Aperathiot, slightly outperforming Whisper Large-v3 at both the word and character level (WER 60.22%, CER 31.04%), but degrades more substantially on Cretan (WER 86.31%, CER 49.45%). In contrast, the Greek-adapted XLS-R models yield WERs exceeding 100% in both dialects, despite somewhat lower CERs, suggesting that while some subword patterns are captured, word-level reconstruction fails under zero-shot conditions.

For Lesbian, the only northern dialect in the benchmark, both WER and CER increase markedly across all models. Whisper Large-v3 again performs best (WER 80.87%, CER 57.70%), while Omnilingual ASR shows comparable character-level accuracy (CER 56.07%) but

Model	Checkpoint	WER (%)	CER (%)
Large-v3	pretrained	61.85%	33.08%
Large-v2	pretrained	69.68%	40.84%
Whisper-medium	pretrained	70.32%	39.73%
XLS-R-53-greek	pretrained	104.18%	91.89%
XLS-R-300-greek	pretrained	101.08%	90.48%
Omnilingual ASR	pretrained	60.22%	31.04%
XLS-R-53-greek[7]	epoch 34	45.64%	16.45%
Whisper-medium[8]	step 2000	**37.41**%	**15.54**%

Table 2: Aperathiot Model Performance Comparison.

Model	Checkpoint	WER (%)	CER (%)
Large-v3	pretrained	96.65%	59.72%
Large-v2	pretrained	105.94%	70.41%
Whisper-medium	pretrained	117.13%	91.25%
XLS-R-53-greek	pretrained	109.4%	97.48%
XLS-R-300-greek	pretrained	103.1%	94.03%
Omnilingual ASR	pretrained	101.24%	59.72%
XLS-R-53-greek[9]	epoch 16	**68.17**%	**30.66**%
Whisper-medium[10]	step 4000	77.79%	47.80%

Table 3: Cappadocian Model Performance Comparison.

Model	Checkpoint	WER (%)	CER (%)
Large-v3	pretrained	70.24%	41.74%
Large-v2	pretrained	78.90%	50.67%
Whisper-medium	pretrained	81.69%	55.92%
XLS-R-53-greek	pretrained	106.29%	94.81%
XLS-R-300-greek	pretrained	102.55%	93.04%
Omnilingual ASR	pretrained	86.31%	49.45%
XLS-R-53-greek[11]	epoch 6	55.04%	22.08%
Whisper-medium[12]	step 2000	**37.02**%	**17.02**%

Table 4: Cretan Model Performance Comparison.

Model	Checkpoint	WER (%)	CER (%)
Large-v3	pretrained	80.87%	57.70%
Large-v2	pretrained	92.02%	57.70%
Whisper-medium	pretrained	87.89%	60.86%
XLS-R-53-greek	pretrained	107.96%	97.02%
XLS-R-300-greek	pretrained	102.03%	93.6%
Omnilingual ASR	pretrained	95.05%	56.07%
XLS-R-53-greek[13]	epoch 22	71.79%	32.97%
Whisper-medium[14]	step 4000	**54.73**%	**26.70**%

Table 5: Lesbian Model Performance Comparison.

substantially higher WER (95.05%). The elevated CER values across models indicate systematic difficulties in modeling vowel reduction and deletion associated with northern vocalism, which affects not only word segmentation but also character-level alignment. The persistent gap between southern and northern dialects thus emerges clearly in both evaluation metrics.

Cappadocian represents the most challenging case under inference-only evaluation. All pretrained models exhibit extremely high WERs, reflecting severe word-level mismatch. Whisper Large-v3 attains the lowest WER (96.65%) and a comparatively lower CER (59.72%), closely matched at the character level by the Omnilingual ASR (CER 59.72%) but with higher WER (101.24%). Other Whisper variants and the Greek-adapted XLS-R models perform progressively worse, with CER values exceeding 70% and in some cases approaching 95%. The divergence between WER and CER for Cappadocian is particularly informative: although most models fail to recover correct word forms, lower CERs for Whisper Large-v3 and Omnilingual ASR suggest partial preservation of phonotactic or subword structure even under extreme mismatch. This outcome is expected given that Cappadocian is a typological outlier within the Greek dialectal continuum: beyond prolonged contact with Turkish, it has undergone extensive structural changes across all linguistic levels, resulting in substantial lexical, phonological, and morphosyntactic divergence from the data used to pretrain current ASR models.

Overall, inference-only results point to a strong interaction between dialectal distance and error type. Varieties closer to SMG yield lower WER and CER, whereas increased phonological divergence and contact-induced change lead to sharp degradation, most visibly at the word level. At the same time, systematic WER–CER gaps indicate that some architectures preserve limited subword structure even when word-level reconstruction fails. Interpretation of zero-shot scores must also account for the benchmark target: our references encode dialectal lexical and morphophonological forms rendered in SMG-based orthography (cf. Section 4.1), so performance reflects not only acoustic mismatch but also lexical and orthographic mismatch with decoders biased toward standard-language distributions. Models exposed predominantly to SMG text may therefore normalize toward SMG-like outputs or fail to reproduce dialect-specific spellings, inflating WER under inference-only settings.

Taken together, these observations suggest that

[7] huggingface.co/ilsp/xls-r-53-greek-aperathiot
[8] huggingface.co/ilsp/whisper-aperathiot-asr
[9] huggingface.co/ilsp/xls-r-53-greek-cappadocian
[10] huggingface.co/ilsp/whisper-cappadocian-asr
[11] huggingface.co/ilsp/xls-r-53-greek-cretan-extended
[12] huggingface.co/ilsp/whisper-cretan-extended-asr
[13] huggingface.co/ilsp/xls-r-53-greek-lesbian
[14] huggingface.co/ilsp/whisper-lesbian-asr

current pretrained ASR systems do not adequately handle Greek dialectal speech—especially structurally divergent varieties—without dialect-aware adaptation, motivating the fine-tuning experiments presented in the following section.

6.2 Fine-tuning

Dialect-specific fine-tuning leads to substantial and consistent improvements across all varieties (Tables 2-5, lower rows), confirming that even limited amounts of dialectal data are sufficient to markedly reduce both WER and CER.

For the southern dialects (i.e., Aperathiot and Cretan), fine-tuning Whisper-medium yields the strongest gains. In Aperathiot, WER is reduced from 60.22% under the best inference-only setting (Omnilingual ASR) to 37.41%, corresponding to a 39% relative improvement, while CER drops to 15.54%. Cretan shows the largest absolute improvement: Whisper-medium achieves 37.02% WER and 17.02% CER, representing a 47% relative WER reduction compared to the best zero-shot model (Whisper Large-v3 at 70.24%). These results indicate that dialects structurally close to SMG benefit strongly from lightweight adaptation, with fine-tuning substantially closing the gap between dialectal and standard speech recognition.

For Lesbian, fine-tuning also yields clear improvements, though performance remains lower than for southern dialects. Whisper-medium reduces WER from 80.87% to 54.73% (32% relative improvement) and CER from 57.70% to 26.70%. The remaining gap reflects persistent phonological divergence due to northern vocalism, which continues to challenge word-level recognition even after adaptation, despite substantial gains at the character level.

Cappadocian remains the most challenging variety after fine-tuning. The best-performing adapted model is XLS-R-53-greek, achieving 68.17% WER and 30.66% CER, corresponding to a 29% relative WER reduction compared to the best zero-shot result (Whisper Large-v3 at 96.65%). Notably, for Cappadocian, XLS-R-53-greek outperforms Whisper-medium after fine-tuning, reversing the pattern observed in the other dialects. This suggests that the wav2vec-style architecture may be better suited to extreme out-of-distribution settings when sufficient dialect-specific acoustic evidence is provided, particularly at the subword level, as reflected in the larger CER reductions.

Across all dialects, fine-tuning consistently narrows the gap between WER and CER, indicating improved alignment between acoustic modeling and lexical prediction. However, the magnitude of improvement varies systematically with dialectal distance: while adaptation largely mitigates mismatch for southern dialects, varieties involving deep structural divergence and extensive contact-induced change remain substantially more difficult.

Overall, these results demonstrate that fine-tuning is essential for robust ASR on Greek dialects, but also that model architecture and dialect typology jointly shape the limits of adaptation under low-resource conditions.

7 Discussion and Outlook

Across both inference-only and fine-tuned settings, dialectal distance from SMG is a strong predictor of ASR difficulty, with systematic differences between southern, northern, and contact-influenced varieties. Southern dialects (Aperathiot, Cretan) are consistently easier, while the northern Lesbian and especially Cappadocian remain substantially harder, even after adaptation.

A key result is the growing divergence between WER and CER as dialectal distance increases. For the southern dialects, both metrics are comparatively lower, suggesting that models recover not only phonotactics but also a meaningful portion of the word inventory. For Lesbian and Cappadocian, WER rises sharply—often approaching or exceeding 100% in zero-shot evaluation—while CER remains noticeably lower. This is clearest in Cappadocian, where Whisper Large-v3 and Omnilingual ASR both reach 59.72% CER but still exhibit very high WER (96.65% and 101.24%), indicating that subword structure is partially captured even when word-level recognition breaks down under severe lexical and structural mismatch.

Model comparisons further point to architectural effects. Whisper Large-v3 is the strongest zero-shot Whisper model across dialects, while Omnilingual ASR generalizes unevenly: it slightly outperforms Large-v3 on Aperathiot, but degrades on Cretan, where Large-v3 is clearly better. In contrast, both Greek-adapted XLS-R models perform poorly in zero-shot conditions across dialects, and scaling from XLS-R-53-greek to XLS-R-300-greek yields no clear improvement, suggesting that model scaling alone does not resolve strong dialectal mismatch.

Fine-tuning improves all varieties but with uneven returns. Whisper-medium reaches ≈37% WER on both Aperathiot and Cretan (about 40–47% relative reduction), while Lesbian (54.73%) and Cappadocian (68.17%) remain substantially harder. Notably, Cappadocian is the only case where fine-tuned XLS-R-53-greek outperforms fine-tuned Whisper-medium, suggesting that CTC-style adaptation may be more robust than seq2seq decoding under extreme out-of-distribution conditions (for similar results and discussion in dialectal/low-resource settings, see Williams et al., 2023; Adnan and Hassani, 2025; see also Barcovschi et al., 2023; Vásquez-Correa and Álvarez Muniain, 2023). This suggests architectural differences in handling extreme out-of-distribution data, a question worth exploring in future work.

Our findings align with prior Greek dialect ASR benchmarking (Vakirtzian et al., 2024). Their zero-shot Whisper Large-v3 result on Eastern Cretan (58.42% WER; single speaker) is lower than ours on combined Eastern–Western Cretan (70.24%), plausibly reflecting greater internal variation in Western Cretan (Kontosopoulos, 2008) without changing the overall pattern. Similarly, our results for the dialect of Lesbos match the high-error profile reported for the closely related Aivaliot variety (zero-shot WER $>$ 100%), reinforcing that northern vocalism and related SMG-mismatch phenomena pose a systematic challenge for pretrained ASR. Beyond ASR, similar distance-sensitive transfer effects have been observed in morphosyntactic parsing, with stronger transfer from SMG to Cretan than to Lesbian (Bompolas et al., 2025; Vakirtzian et al., 2025). Contact-driven divergence remains the hardest case: Vakirtzian et al. (2024) report WER $>$ 100% for Griko (Italo-Romance contact plus Latin orthography), while Cappadocian (contact with Turkish plus extensive structural change) is the most challenging variety in our benchmark even after fine-tuning. Unlike Aivaliot, where Turkish influence is largely lexical (Vakirtzian et al., 2024), Cappadocian shows multi-level restructuring that amplifies lexical and subword mismatch.

Overall, performance ranges from WERs near 37% for fine-tuned southern dialects to nearly 70% for Cappadocian, highlighting the need for dialect-specific benchmarks and adaptation strategies. For the most divergent varieties, where orthographic and lexical gaps persist, approaches beyond straightforward fine-tuning (e.g., dialect-aware pretraining or modeling) may be necessary, particularly for endangered dialects where ASR can support documentation and revitalization.

Limitations

As with most dialectal ASR studies, our results are shaped by practical constraints related to data availability, transcription conventions, and computational resources. The limitations outlined below contextualize the reported WER and CER scores and highlight directions in which future work could improve reliability, comparability, and linguistic interpretability.

Orthography and transcription variability. None of the examined dialects has a fully standardized orthography, and transcriptions therefore necessarily reflect local conventions and annotator decisions. Although we aimed to keep transcriptions as close as possible to SMG orthography, including systematic choices for northern vocalism in Lesbian, residual inconsistencies remain and may inflate both WER and CER, particularly for varieties with larger lexical and phonological gaps from SMG. This issue is especially salient for Cappadocian, where extensive contact-induced and internal restructuring places additional pressure on spelling conventions, and where multiple transcribers may adopt slightly different practices. Moreover, inter-annotator agreement was not formally measured, which limits our ability to quantify transcription consistency and disentangle annotation-related variability from model-induced errors.

Heterogeneous transcription pipelines. In addition to orthographic variability across dialects, the benchmark currently includes subcollections produced via different annotation workflows (e.g., direct manual transcription vs. manual correction of ASR-bootstrapped drafts). Even when the same transcription policy is targeted, workflow differences may yield subtle systematic effects (e.g., segmentation preferences or persistence of model-like spellings). Users who wish to train on the benchmark as a single pooled dataset should therefore treat transcription workflow as a potential confound and consider (i) training and evaluating per dialect, (ii) adding workflow metadata as a control variable, and/or (iii) applying post-hoc normalization of segmentation and tokenization conventions. Extending the benchmark with parallel

re-annotation under a single unified pipeline is an important direction for future work.

Data size, speaker coverage, and domain differences. While the benchmark expands dialectal coverage, the amount of processed audio remains limited for some varieties (e.g., approximately one hour for Aperathiot and Lesbian, and about 1h17m for Cappadocian), and speaker diversity is uneven across datasets. For instance, the Western Cretan subset is based on recordings from a small number of speakers, whereas the Eastern Cretan data originate from radio broadcasts. These factors constrain the robustness of fine-tuning results and may limit generalization to broader speaker populations and communicative contexts within each dialect.

Limited error diagnosis beyond aggregate metrics. Evaluation is based on aggregate WER and CER computed over normalized text (lowercased, punctuation removed). While this enables consistent quantitative comparison, it does not reveal which specific dialectal phenomena, such as phonological alternations, morphological variation, lexical borrowing, or word boundary effects, contribute most to recognition errors. A more detailed qualitative error analysis could provide deeper insight into model behavior and better inform future adaptation strategies (see, for example, Parsons et al., 2023; Blaschke et al., 2025).

Speaker overlap across splits. The training, development, and test partitions are constructed at the utterance level, which implies that the same speakers appear in all splits. This split strategy is common in low-resource ASR benchmarks, but it can lead to optimistic scores because models may partially leverage speaker-specific characteristics observed during training. As a result, these results primarily reflect within-speaker generalization under dialectal mismatch; evaluating cross-speaker robustness would require speaker-disjoint splits, which we leave for future benchmark extensions and ablation experiments.

Computational constraints. Computational resources limited the range of adaptation strategies explored. In particular, large models such as Whisper Large-v2/Large-v3 and Omnilingual ASR were evaluated only in zero-shot settings, and fine-tuning was restricted to Whisper-medium and XLS-R-53-greek. As a result, we did not investigate more computationally demanding approaches such as continued pretraining, larger-scale hyperparameter optimization, or dialect-aware model variants, which may be necessary for highly divergent varieties such as Cappadocian.

Ethical Considerations

All speech data used in this study were collected with informed consent from participants and anonymized prior to inclusion in the benchmark. Personal names and identifiable information were removed at both the audio and transcription levels through manual review by trained linguists, and no sensitive personal data are released. In the case of the Cappadocian recordings, which were collected during earlier fieldwork, consent for research use and subsequent reuse of the data was obtained by the original fieldworker.

Several of the examined varieties are low-resource or endangered (notably Cappadocian), and the goal of this work is to support linguistic documentation and technological inclusion rather than deployment in real-world applications. The reported ASR models exhibit high error rates for several dialects, and their outputs should not be interpreted as suitable for practical use or decision-making involving speakers of these varieties.

Finally, dialectal ASR systems may reproduce or amplify existing linguistic biases if deployed without adequate adaptation and evaluation. We therefore frame this benchmark as a research resource aimed at understanding model limitations and guiding future work, rather than as an endorsement of current systems for dialectal speech processing.

Acknowledgments

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.

References

Aditya Yadavalli and Ganesh Mirishkar and Anil Kumar Vuppala. 2022. Multi-Task End-to-End Model for Telugu Dialect and Speech Recognition. In *Interspeech 2022*, pages 1387–1391.

Renas Adnan and Hossein Hassani. 2025. Which one Performs Better? Wav2Vec or Whisper? Applying both in Badini Kurdish Speech to Text (BKSTT). *arXiv:2508.09957*.

Vasileia Alexelli. 2021. *Chartografisi tis glossikis poikilias tis Lesvou [Mapping the linguistic variety of Lesbos]*. Ph.D. thesis, University of Patras, School of Humanities and Social Sciences, Department of Philology, Linguistics Section.

Tanel Alumäe, Jiaming Kong, and Daniil Robnikov. 2023. Dialect Adaptation and Data Augmentation for Low-Resource ASR: TalTech Systems for the MADASR 2023 Challenge. *Preprint*, arXiv:2310.17448.

Arun Babu, Changhan Wang, Andros Tjandra, Kushal Lakhotia, Qiantong Xu, Naman Goyal, Kritika Singh, Patrick von Platen, Yatharth Saraf, Juan Pino, Alexei Baevski, Alexis Conneau, and Michael Auli. 2021. XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale. *Preprint*, arXiv:2111.09296.

Andrei Barcovschi, Rishabh Jain, and Peter Corcoran. 2023. A comparative analysis between Conformer-Transducer, Whisper, and wav2vec2 for improving the child speech recognition. *Preprint*, arXiv:2311.04936.

Vivek Bhardwaj, Vinay Kukreja, Navjeet Kaur, and Nandini Modi. 2021. Building an ASR System for Indian (Punjabi) language and its evaluation for Malwa and Majha dialect: Preliminary Results. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–5.

Verena Blaschke, Miriam Winkler, Constantin Förster, Gabriele Wenger-Glemser, and Barbara Plank. 2025. A Multi-Dialectal Dataset for German Dialect ASR and Dialect-to-Standard Speech Translation. In *Interspeech 2025*, pages 913–917. ISCA.

Paul Boersma and David Weenink. 2001. Praat: Doing phonetics by computer. Computer program.

Stavros Bompolas. 2023. *Computational dialectology in the linguistic varieties of Cappadocian, Pharasiot, and Silliot*. Ph.D. thesis, University of Patras, School of Humanities and Social Sciences, Department of Philology, Linguistics Section.

Stavros Bompolas, Stella Markantonatou, Angela Ralli, and Antonios Anastasopoulos. 2025. Crossing Dialectal Boundaries: Building a Treebank for the Dialect of Lesbos through Knowledge Transfer from Standard Modern Greek. In *Proceedings of the Eighth Workshop on Universal Dependencies (UDW, SyntaxFest 2025)*, pages 39–51, Ljubljana, Slovenia. Association for Computational Linguistics.

Georgios Chatzidakis. 1905. *Mesaionika kai Nea Ellinika [Medieval and Modern Greek]*, volume 12. Sakellarios, Athens.

William Chen, Chutong Meng, Jiatong Shi, Martijn Bartelds, Shih-Heng Wang, Hsiu-Hsuan Wang, Rafael Mosquera, Sara Hincapie, Dan Jurafsky, Antonis Anastasopoulos, Hung yi Lee, Karen Livescu, and Shinji Watanabe. 2025. The ML-SUPERB 2.0 Challenge: Towards Inclusive ASR Benchmarking for All Language Varieties. In *Interspeech 2025*, pages 2093–2097.

Richard MacGillivray Dawkins. 1916. *Modern Greek in Asia Minor: a study of the dialects of Sílli, Cappadocia and Phárasa with grammar, texts, translations and glossary*. Cambridge University Press, Cambridge.

Antonios Dimakis, John Pavlopoulos, and Antonios Anastasopoulos. 2025. Dialect Normalization using Large Language Models and Morphological Rules. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 23696–23714, Vienna, Austria. Association for Computational Linguistics.

Luca Ducceschi and Greta H. Franzini. 2025. Speech transcription from South Tyrolean Dialect to Standard German with Whisper. In *Interspeech 2025*, pages 1–5. ISCA.

Vardis Georgilas and Themos Stafylakis. 2025. Automatic Speech Recognition for Greek medical dictation. *Preprint*, arXiv:2509.23550.

Zachary Hopton and Eleanor Chodroff. 2025. The Impact of Dialect Variation on Robust Automatic Speech Recognition for Catalan. In *Proceedings of the 22nd SIGMORPHON workshop on Computational Morphology, Phonology, and Phonetics*, pages 23–33, Albuquerque, New Mexico, USA. Association for Computational Linguistics.

Ryo Imaizumi, Ryo Masumura, Sayaka Shiota, and Hitoshi Kiya. 2020. Dialect-Aware Modeling for End-to-End Japanese Dialect Speech Recognition. In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 297–301.

Sukairaj Hafiz Imam, Tadesse Destaw Belay, Kedir Yassin Husse, Ibrahim Said Ahmad, Idris Abdulmumin, Hadiza Ali Umar, Muhammad Yahuza Bello, Joyce Nakatumba-Nabende, Seid Muhie Yimam, and Shamsuddeen Hassan Muhammad. 2025. Automatic Speech Recognition (ASR) for African Low-Resource Languages: A Systematic Literature Review. *Preprint*, arXiv:2510.01145.

Petros Karatsareas. 2011. A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective. Publisher: Apollo - University of Cambridge Repository.

Haneen Bahjat Khalafallah, Mohamed Abdel Fattah, and Ruqayya Abdulrahman. 2024. Speech corpus for Medina dialect. *Journal of King Saud University - Computer and Information Sciences*, 36(2):101864.

Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14):7684–7689.

Nikolaos G. Kontosopoulos. 2008. *Dialektoi kai idiomata tis Neas Ellinikis [Dialects and Idioms of Modern Greek]*. Ekdoseis Grigori, Athens.

Olli Kuparinen. 2025. Automatic Dialectal Transcription: An Evaluation on Finnish and Norwegian. In *Interspeech 2025*, pages 2390–2394. ISCA.

Seham Nasr, Rehab Duwairi, and Muhannad Quwaider. 2023. End-to-End Speech Recognition For Arabic Dialects. *Arabian Journal for Science and Engineering*, 48(8):10617–10633.

Dimitrios V. Oikonomidis. 1952. *Peri tou glossikou idiomatos Aperathou–Naksou [On the Linguistic Idiom of Aperathos, Naxos]*. Typografeion Myrtidou, Athens.

Omnilingual ASR Team, Gil Keren, Artyom Kozhevnikov, Yen Meng, Christophe Ropers, Matthew Setzler, Skyler Wang, Ife Adebara, Michael Auli, Can Balioglu, Kevin Chan, Chierh Cheng, Joe Chuang, Caley Droof, Mark Duppenthaler, Paul-Ambroise Duquenne, Alexander Erben, Cynthia Gao, Gabriel Mejia Gonzalez, and 14 others. 2025. Omnilingual ASR: Open-Source Multilingual Speech Recognition for 1600+ Languages. *Preprint*, arXiv:2511.09690.

Georgios Emmanouil Pangalos. 1955. *Peri tou glossikou idiomatos tis Kritis, itoi diagramma grammatikis kai glossarion tou simerinou glossikou idiomatos tis Kritis [On the Linguistic Idiom of Crete: A Grammatical Outline and Glossary of the Contemporary Cretan Dialect]*. n.p., Athens.

Phoebe Parsons, Knut Kvale, Torbjørn Svendsen, and Giampiero Salvi. 2023. A character-based analysis of impacts of dialects on end-to-end Norwegian ASR. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 467–476.

Siqing Qin, Longbiao Wang, Sheng Li, Jianwu Dang, and Lixin Pan. 2022. Improving low-resource Tibetan end-to-end ASR by multilingual and multi-level unit modeling. *EURASIP Journal on Audio, Speech, and Music Processing*, 2022(1):2.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust Speech Recognition via Large-Scale Weak Supervision. *Preprint*, arXiv:2212.04356.

Angela Ralli. 2015. Strategies and Patterns of Loan Verb Integration in Modern Greek Varieties. In Angela Ralli, editor, *Contact Morphology in Modern Greek Dialects*, pages 73–88. Cambridge Scholars Publishing.

Angela Ralli. 2019a. Affixoids and Verb Borrowing in Aivaliot Morphology. In Angela Ralli, editor, *The Morphology of Asia Minor Greek*, pages 221–254. BRILL.

Angela Ralli. 2019b. Greek in Contact with Romance. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

Jiatong Shi, Shih-Heng Wang, William Chen, Martijn Bartelds, Vanya Bannihatti Kumar, Jinchuan Tian, Xuankai Chang, Dan Jurafsky, Karen Livescu, Hung yi Lee, and Shinji Watanabe. 2024. ML-SUPERB 2.0: Benchmarking Multilingual Speech Models Across Modeling Constraints, Languages, and Datasets. *Preprint*, arXiv:2406.08641.

Clement Sicard, Kajetan Pyszkowski, and Victor Gillioz. 2023. Spaiche: Extending State-of-the-Art ASR Models to Swiss German Dialects. *Preprint*, arXiv:2304.11075.

Naoki Takahashi, Shogo Miwa, Yuta Kamiya, Takumi Toyama, Raufun Nahar, and Atsuhiko Kai. 2024. Comparison of Large Pre-trained Models and Adaptation Methods for Japanese Dialects ASR. In *2024 IEEE 13th Global Conference on Consumer Electronics (GCCE)*, pages 811–814.

Sarah Grey Thomason and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*, 1. paperback print edition. Univ. of California Press, Berkeley.

Melissa Torgbi, Andrew Clayman, Jordan J. Speight, and Harish Tayyar Madabushi. 2025. Adapting Whisper for Regional Dialects: Enhancing Public Services for Vulnerable Populations in the United Kingdom. *Preprint*, arXiv:2501.08502.

T. Toyama, A. Kai, Y. Kamiya, and N. Takahashi. 2024. Adapting Large-Scale Pre-trained Models for Unified Dialect Speech Recognition Model. *Acta Physica Polonica A*, 146(4):413–418.

Peter Trudgill. 2003. Modern Greek dialects: A preliminary classification. *Journal of Greek Linguistics*, 4(1):45 – 63. Number: 1 Place: Leiden, The Netherlands Publisher: Brill.

Chara Tsoukala, Kosmas Kritsis, Ioannis Douros, Athanasios Katsamanis, Nikolaos Kokkas, Vasileios Arampatzakis, Vasileios Sevetlidis, Stella Markantonatou, and George Pavlidis. 2023. ASR pipeline for low-resourced languages: A case study on pomak. In *Proceedings of the Second Workshop on NLP Applications to Field Linguistics*, pages 40–45, Dubrovnik, Croatia. Association for Computational Linguistics.

UNESCO, editor. 2010. *Atlas of the Worlds Languages in Danger*. United Nations Educational, Scientific and Cultural Organization, Paris.

Socrates Vakirtzian, Vivian Stamou, Yannis Kazos, and Stella Markantonatou. 2025. Dialectal treebanks and their relation with the standard variety: The case of East Cretan and Standard Modern Greek. In *Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th*

Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025), pages 776–784, Tallinn, Estonia. University of Tartu Library.

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 *Proc. Interspeech 2024*, pages 3974–3978.

Juan Camilo Vásquez-Correa and Aitor Álvarez Muniain. 2023. Novel Speech Recognition Systems Applied to Forensics within Child Exploitation: Wav2vec2.0 vs. Whisper. *Sensors*, 23(4).

Aiden Williams, Andrea Demarco, and Claudia Borg. 2023. The Applicability of Wav2Vec2 and Whisper for Low-Resource Maltese ASR. In *2nd Annual Meeting of the ELRA/ISCA SIG on Under-resourced Languages (SIGUL 2023)*, pages 39–43. ISCA.

Maria Xefteri. 2009. Koinonioglossiki proseggisi tou idiomatos t' Aperathou Naksou (opos diatireitai stin Athina) [(A Sociolinguistic Approach to the Aperathou Dialect of Naxos as Preserved in Athens)]. In *Proceedings of the 4th Graduate Student Meeting of the Department of Philology, University of Athens*, Athens, Greece. Department of Philology, National and Kapodistrian University of Athens. Available at Academia.edu.