PAKISTAN JOURNAL OF LANGUAGE AND LITERARY STUDIES
Volume 2, Issue 1 (Spring 2026)
Pages: 52-67

Evaluating Forced Alignment in an Under-Documented Shakargarh Punjabi Speech Corpus

Adnan Arif[1]

ABSTRACT

This study evaluated the applicability and usefulness of forced alignment for developing a searchable speech corpus of under-documented Shakargarh Punjabi, a local Punjabi variety spoken in Narowal District, Punjab, Pakistan. Positioned at the intersection of corpus linguistics, speech technology, and language documentation, the study addressed the problem that low-resource speech recordings often remain insufficiently transcribed, time-aligned, searchable, and reusable for linguistic analysis. The design adopted was a mixed method that consisted of quantitative descriptive evaluation and qualitative error analysis. Primary spoken data were generated from 24 recordings of 12 native speakers using semi-structured interviews, natural conversations and short personal accounts that produced 120 rated utterances (19.79 minutes, 1,565 words). The recordings were subsequently transcribed in Shahmukhi Punjabi; with a consistent Roman transliteration based on Montreal Forced Aligner, annotated in ELAN and manually verified in Praat. Successful utterances were those in which the ABDev remained under 50ms and the segment mismatch was not present. 74 (61.7%) utterances were successfully aligned initially and 96 utterances (80.0%) were successfully aligned after project specific pronunciation-lexicon refinement. The number of OOV cases dropped from 62 to 18 instances while the OOV rate dropped from 3.96 to 1.15 OOV/100 words. The remaining problems primarily were those items which were not present in the standard language, connected speech, and vowel reduction. The findings show that semi-automatic alignment, supported by local lexicon refinement and manual validation, can materially improve the searchability and documentary value of Shakargarh Punjabi speech data.

Keywords: *forced alignment; Shakargarh Punjabi; speech corpus; language documentation; low-resource languages*

[1] PhD Scholar, School of English Language, Technology and Literature, Faculty of Languages and Literature, Minhaj University Lahore, Pakistan

Corresponding author:
Mr. Adnan Arif, PhD Scholar English Linguistics, Minhaj University Lahore, Pakistan
Email: 2023F-mulphd-eng-007@mul.edu.pk

1. Introduction

The interface between language documentation and corpus linguistics has assumed a prominent position in recent scholarly literature. Today, however, the documentation of a modern language is not merely marking the moment of recording on a phonetic device and archiving it for posterity; what is increasingly emphasized is the transformation of language data into analyzable resources, ones that are easy to reuse and search. Beyond the large written corpora, too, corpus linguistics has grown to now embrace high-quality transcribed and sequentially annotated spoken corpora—copies of the spoken text correlated with the audio file. This association plays a much bigger role for under-documented and low-resource speech varieties, in which there are recordings, but their lack of systematic documentation and accessibility prevents them from being used to their full potential.

Thus, forced alignment is a crucial methodology in speech corpus construction. Forced alignment is the procedure of connecting an audio recording with its written transcriptions and for tagging phonetic or word segments with time boundaries. This process makes it possible to convert raw audio data to a searchable corpus, offering the basis for further phonetic investigation as well as investigation at the lexical and discourse levels. Technological interventions in the creation of resources, in annotation and mobilization of resources are noted in the literature to be important, especially for communities with fewer corpora or digital systems, such as speech communities (Berez-Kroeker et al., 2023). But also corpus-based projects show that documentation collection is of limited use and that a documentation collection that is retrieved and temporally aligned, can be used to systematically retrieve, compare and analyze linguistic data (Janssen & Seifart, 2025; Paschen et al., 2020).

Shakargarh is a tehsil of the district Narowal in Punjab province of Pakistan, where the present study was done. Narowal is a border district and Shakargarh is one of the tehsils in it. The geopolitically and historically important area of Shakargarh Tehsil borders Pakistan and India and was historically a part of Gurdaspur before being separated from it during the partition and handed over to the Pakistani territory. Official sources further mention that Kartarpur is also an important locality of Shakargarh which further emphasizes the cultural and linguistic significance of this border area with complex historical development (District Narowal, Government of the Punjab, n.d.-a, n.d.-b, n.d.-c).

In the present research, Shakargarh was taken as a specific speech-community area besides its geographical position. The Punjabi spoken locally remains poorly represented in searchable corpus form. This scarcity creates scope for community-based research grounded in actual speech data rather than generalized descriptions of Punjabi. Such work is relevant to applied linguistics and corpus linguistics because it enables investigation of spoken language as it is used within a particular community. At the same time, local varieties may not correspond closely to existing speech models, pronunciation lexicons, or alignment systems developed for better-resourced languages.

The study therefore examined forced alignment in an under-documented Shakargarh Punjabi speech corpus. It first established the research problem, gap, objectives, and questions, and

then evaluated a completed pilot corpus through a mixed quantitative and qualitative workflow. Particular attention was given to the extent to which local pronunciation, dialect-specific vocabulary, connected speech, and pronunciation-lexicon refinement affected alignment performance and the documentary usability of the resulting corpus.

2. Research Problem

Low-resource and under-documented speech varieties are often recorded, but most of the time these recordings do not develop into searchable and analyzable corpora. As a result, useful speech data still remain difficult to access, segment, and analyze in a systematic way. Forced alignment provides an effective tool for matching speech recordings to transcripts at the word or segment level, which facilitates corpus building, phonetic research, and long-term language documentation. However, in low-resource settings, it tends to be less effective because of small datasets, natural speech recording conditions, and mismatches between local speech varieties and the models on which alignment tools rely. The case of the Shakargarh Punjabi variety is especially important because this variety is not sufficiently documented, and it has not been developed into a searchable corpus for linguistic research.

3. Research Gap

Forced alignment has been demonstrated to be an important tool for increasing the resources for under-resourced languages, such as in corpus compilation and speech annotations (Leinonen et al., 2022; Tosolini & Bowern, 2025). However, much of current literature is focused on the general aspect of low-resource or cross-linguistic settings – and not specifically on Shakargarh Punjabi. The researcher is not aware of any published work which has attempted to quantify the forced-alignment results on any Punjabi speech corpus from Shakargarh, or consider manual evaluation or lexicon refinement as methods that may help record the local variety. There is thus a gap between existing low-resource forced-alignment work and its empirically assessed use with this poorly-documented Punjabi-speaking community.

4. Research Objectives

- To identify the challenges affecting alignment accuracy in this under-documented Shakargarh Punjabi variety.
- To evaluate the usefulness of forced alignment for corpus development and language documentation of this selected variety of Punjabi.
- To assess the performance of forced alignment on a Shakargarh Punjabi speech corpus.

5. Research Questions

- What are the main challenges affecting alignment accuracy in this under-documented Shakargarh Punjabi variety?
- How effectively does forced alignment perform on a Shakargarh Punjabi speech corpus?
- How useful is forced alignment for corpus development and language documentation of Shakargarh Punjabi?

6. Literature Review

In the current study into language documentation, documentation is no longer taken to mean simply the collection of linguistic materials, but as a more inclusive process of recording, annotating, preserving, making accessible and re-using language materials. Berez-Kroeker et al. (2023) believe that in the last few years the process of writing a document has been closely linked to technologies of creating and mobilizing resources. This view is applicable because documentation is considered a process going back to the recorded language events and reconstructing them to make them useful for documentation. It applies directly to the present study, since the Shakargarh Punjabi corpus was designed not merely as a collection of recordings but as a searchable and analyzable speech resource. Documentation becomes especially valuable when linguistic data are organized for long-term academic and community use.

A similar trend can be observed in time-aligned documentation corpora. In their analysis of the DoReCo project, Paschen et al. (2020) illustrate how the data presented in language documentation can be converted into a time-aligned cross-linguistic reference corpus. It is also worth mentioning that their contribution demonstrates that the value of documentation corpora can be increased considerably when audio, transcription, and annotation are organized in a manner that facilitates retrieval and comparison. Despite the fact that DoReCo is a huge collaborative enterprise and the present research is conducted on a considerably smaller scale, the principle is exactly the same: unprocessed recordings can be much more useful once they are aligned and systematized. This observation justifies the need to study forced alignment as part of corpus development in Shakargarh Punjabi.

The question of searchability has also been further developed in another work by Janssen and Seifart (2025), who show that documentary corpora can be made searchable with the help of TEITOK implementation through the integration of audio, transcription, translation, and interlinear information. They are of particular relevance because their study emphasizes the role of accessibility and search functions in language documentation. This is to say that documentation deals not only with the preservation of speech, but also with making it retrievable and processable. In the case of the current study, the literature helps explain the importance of alignment; forced alignment is valuable not only because it saves time but also because it can transform a local collection of recordings into a more usable corpus that can be used in future analysis and documentation.

Recent research on forced alignment as such provides further support for its relevance. Leinonen et al. (2022) examine the application of semiautomatic speech alignment in under-resourced languages and find that alignment can indeed be used beneficially even in low-data situations. Their work is especially useful since it does not present alignment as a flawless or entirely automated solution; instead, it demonstrates that cross-language forced alignment can be improved when researchers combine alignment tools with domain knowledge and validation. This is especially applicable to the situation of an under-documented form such as Shakargarh Punjabi, where performance can be affected by gaps in transcription, vocabulary

peculiar to the dialect, and local phonological features. Their study, therefore, justifies the careful and critical use of forced alignment rather than unreflective technological application.

This argument is further reinforced by Ahn et al. (2024), who conducted a study of phone alignment in the case of Panãra. They show that the results of the alignment process may differ considerably depending on whether researchers approach it through language-specific methods or cross-linguistic methods. According to their findings, the structure of the phone inventory and the modeling strategy are important. This is particularly relevant to the present study because it implies that the effectiveness of forced alignment on Shakargarh Punjabi cannot be assumed in advance. Although a tool may be available, its outcomes will depend on the suitability of the acoustic and phonological assumptions of the model to the target variety. That is, forced alignment is not only a technical process, but also a linguistic issue.

A closely related insight is offered by Tosolini and Bowern (2025), when they compare multilingual and cross-lingual approaches using Montreal Forced Aligner for different low-resource related languages. Their findings indicate that a larger baseline model can still be applied conveniently to previously unseen languages, though they assert that performance has to be evaluated carefully. This investigation is particularly relevant in that it offers a direct answer to the question of the conditions under which forced alignment can be applied in low-resource settings. Most importantly, it is relevant to the current study as it indicates that the usefulness of forced alignment is limited not only by the software itself but also by the way models are trained, adapted, and evaluated in relation to real speech data.

Further evidence comes from current research on minority and non-dominant speech communities. Passmore et al. (2025) discuss forced alignment in two English-based Pacific Creoles and demonstrate that even in minority-language conditions, alignment may be helpful when tested critically against the linguistic and acoustic peculiarities of the corpus. Their contribution leads the discussion outside a particular language family and indicates that the issue is broad: although forced alignment can help streamline the process of corpus generation and annotation, the output should be approached with care in relation to the reality of the speech community. This directly relates to the case of Shakargarh Punjabi, where local variation and under-documentation suggest that alignment cannot be taken at face value.

There is also literature that helps to indicate a wider applied-linguistic relevance to the current study. In language pedagogy, pronunciation study, discourse analysis, and sociolinguistic description, searchable speech corpora can also be useful. This could then be followed by a corpus of Shakargarh Punjabi speech that would enable comparative analysis of various varieties of Punjabi, language maintenance in the local area, and the creation of future educational or descriptive materials. In this sense, the present study is not purely technical; it stands at the intersection of corpus linguistics, speech technology, and community-based applied linguistics, with the primary issue being how to make under-documented types of speech more visible and analyzable with the help of tools.

Collectively, the literature points to four key points. To start with, searchability, accessibility, and reuse are all becoming important in language documentation. Second, time-aligned corpora are extremely useful in comparison with unstructured recordings. Third, forced alignment has obvious potential in low-resource contexts, although the key to its success is the condition of the data, linguistic adaptation, and evaluation procedures. Fourth, regardless of the increased research in the area of low-resource alignment, no apparent study has investigated forced alignment in a Shakargarh Punjabi speech corpus. These findings are relevant to the current research in that they demonstrate the relevance of the present study in the context of modern research on corpus development and documentation technologies.

The study was guided by a documentation-as-reusable-resource perspective. Himmelmann (2006) conceptualizes language documentation as the creation of lasting multipurpose records of linguistic practices, while more recent work emphasizes accessibility, mobilization, and technological reuse (Berez-Kroeker et al., 2023). In corpus terms, this orientation means that documentary recordings gain additional research value when transcription, annotation, and temporal alignment make them searchable and comparable (Paschen et al., 2020). Forced alignment was consequently not considered to be an ultimate goal, but as a corpus-enabling procedure, which to be of benefit, had to enhance access to, and analysis of, authentic community speech. This theoretical orientation led to the sampling, the documentation value annotation and the error analysis and interpretation in the successive sections.

7. Research Methodology

The sequential mixed-methods design was used in this study where quantitative evaluation of alignment performance is followed by qualitative analysis of the error sources in linguistic and technical aspects. The quantitative strand measured successful and problematic utterances, out-of-vocabulary (OOV) instances, boundary deviation, segment mismatch, and change after pronunciation-lexicon refinement. The qualitative strand interpreted recurrent alignment problems in relation to dialect-specific vocabulary, connected speech, vowel reduction, overlap, recording noise, and transcription inconsistency. This design matched the research questions because the study required both measurable evaluation of the aligner and linguistically informed explanation of why particular forms succeeded or failed (Leinonen et al., 2022; Tosolini & Bowern, 2025). Figure 1 summarizes the workflow from community-based participant selection and recording through ELAN transcription, MFA alignment, Praat verification, error analysis, and assessment of documentation value. The figure also highlights that manual checking was integrated into, rather than separated from, the computational workflow.

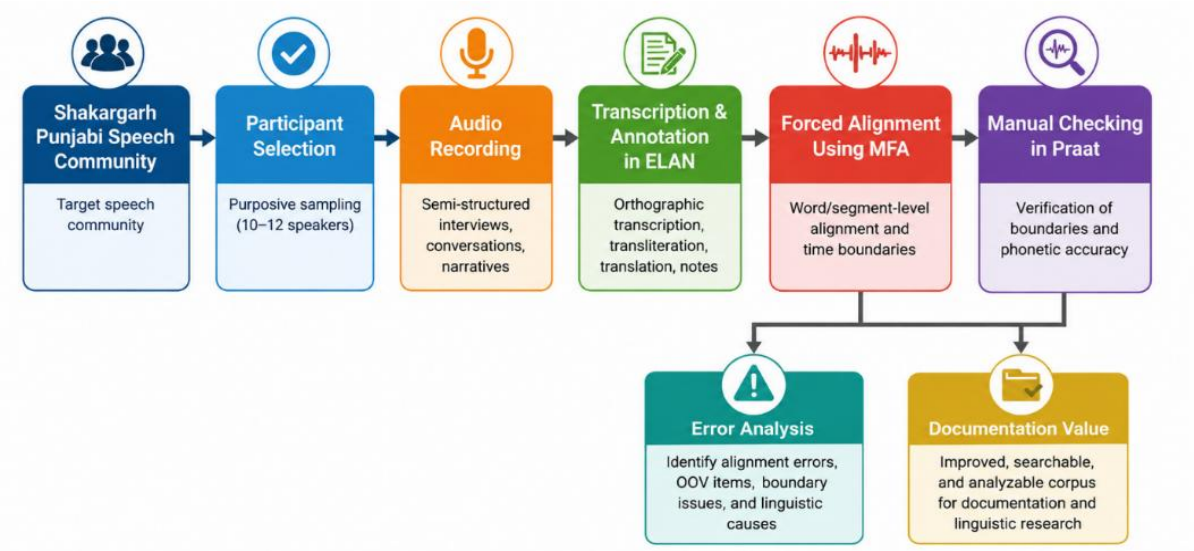


Figure 1

Research workflow for evaluating forced alignment in a Shakargarh Punjabi speech corpus

8. Sampling Method and Size

Non-probability purposive sampling was applied to draw the sample of native speakers of Adult Shakargarh Punjabi, who were long time residents of Shakargarh and areas close to it, in the Narowal District, and who regularly spoke the language. Six women and 6 men, 22-63 years old, and from both urban and rural areas were included in the sample. There were 24 recording files, each from two different speakers. The design was based on the central importance of detailed transcription, forced alignment, development of the pronunciation-lexicon and manual boundary verification, which led to the decision to keep the corpus at pilot scale. Signal wording was assessed on 120 sentences split evenly into semi-structured interviews, natural conversation and personal accounts (40 utterances per speech context). In documentation with limited resources, targeted sampling can be justified when the size of the sample depends either on the lack of a large, consistent corpus, or on the lack of consistent transcription and verification of a large corpus. (Berez-Kroeker et al., 2023; Leinonen et al., 2022).

9. Data Collection Tools

Primary spoken data was gathered using semi-structured interviews, short personal narratives and natural conversations. The use of this combination not only enabled the guided speech but also relatively spontaneous speech so that forced alignment was able to be evaluated under different speech condition. Speaker code, age group, gender, residence, speech context and recording conditions were documented on a participant information sheet and a consent form and on a metadata sheet. All recordings were transferred to WAV format to retain best quality of recording for transcription and alignment. Time-aligned multi-tier transcription/annotation of ELAN was conducted (Max Planck Institute for Psycholinguistics, 2026). Montreal Forced Aligner (MFA), a trainable alignment system based on Kaldi, was used to align audio with transcripts (McAuliffe et al., 2017). Manual inspection of word and

segment boundaries and verification of decisions regarding phonetic alignment were carried out using Praat (Boersma & Weenink, 2026). These tools facilitated both corpus compilation and alignment, verification and later, error analysis

10.Data Analysis Plan

The analysis combined quantitative descriptive statistics with qualitative error analysis. The 120 evaluated utterances represented 19.79 minutes of speech and 1,565 running words. Shahmukhi transcripts were paired with standardized Roman transliteration, and a project-specific pronunciation lexicon was developed for local forms that were absent or inadequately represented in existing resources. MFA output was manually inspected in Praat. An utterance was classified as successful when the absolute boundary deviation was 50 ms or less and no segment mismatch remained; an utterance was classified as problematic when the boundary deviation exceeded 50 ms and/or a segment mismatch was present. This criterion was applied consistently before and after lexicon refinement. The successful utterances, problematic utterances, frequency of OOVs, rate of OOVs, types of residual errors, and success levels of utterances per speech context were calculated through quantitative analysis. Finally, the qualitative stage translated the linguistic sources of the error marks left out secondary (dialect variation, under-completion, connected speech, overlap, background noise, transcript-audio mismatch) into English. This procedure ensured that performance was numerically captured, while maintaining the need to keep the linguistic interpretation for corpus-based tasks in low resource settings (Ahn et al., 2024; Leinonen et al., 2022; Tosolini & Bowern, 2025).

11.Ethical Considerations

Ethical guidelines for conducting community based linguistic research were followed. Before recording, the participants were told the reason of the research and informed consent was obtained both in written form and verbally. Participants chose to participate, and confidentiality was ensured without using names and instead using speaker codes. The audio files, transcripts, metadata and annotation files were maintained securely, and only utilized for suitable and legitimate academic purposes. Local forms of speech were expressed carefully and verified with speakers, if needed, in order to avoid arbitrating the interpretations made on the data gathered from the community into a standardized form. Both proper citation and original analysis and responsible archiving was observed throughout the study. Both the attention given to documenting with care and the location-by-sort system of controlled access and responsible reuse with others align with the current practice of language documentation (Berez-Kroeker et al., 2023).

12.Data Analysis

12.1. Overview of Data Analyzed

A total of 12 speakers took part in the completed pilot, with 24 recordings and 120 evaluated utterances of a total duration of 19.79 mins and 1,565 words. The source of the data was equally from semi structured interviews and for short personal narratives from natural conversation. Table of the material used was transcribed in Shahmukhi Punjabi and subsequently translated into Roman characters. Dialect-specific vocabulary, reduced/non-standard pronunciation, connected speech, overlap and other features typical of spontaneous speech were included in the dataset to assess the forced alignment in this low-resource environment. After developing a community-specific pronunciation-lexicon, the experts reviewed the model generating the local pronunciation-lexicon and made adjustments to its pronunciation rules. The aligned result of the refined pass after community-specific pronunciation-lexicon development was compared with the result of the initial pass, enabling the effectiveness of local community-specific pronunciation-lexicon development on alignment performance to be evaluated.

A sample excerpt from the corpus illustrates the linguistic nature of the collected data:

”اوہ پائی، کیہہ کرے کرنا ایں یار؟ ایتھے کون کسے دا ہُندا اے؟ کوئی کسے دا نئیں، پا جی۔“

Romanized form:

“Oh paai, keeh karay karna ain yaar? Aithay kaun kisay da hunda ae? Koi kisay da nai, paa ji.”

The excerpt represents naturally occurring Shakargarh Punjabi rather than standardized Punjabi and therefore provides a realistic test of alignment performance. The recurring manuscript forms were standardized as پائی / paai, ایں / ain, نئیں / nai, and نِکھاکھرا / nikhakhra so that the same lexical items could be traced consistently across the analysis, findings, discussion, and conclusion.

12.2. Method of Analysis

Analytical framework was a mix between the quantitative descriptive analysis and the qualitative error analysis. The quantitative measures consisted of: utterance-level alignment status, absolute boundary deviation, segment mismatch, OOV frequency, and performance before and after pronunciation-lexicon refinement. A qualitative analysis was used in order to determine and explain the linguistic and technical sources of the problematic alignments observed and the effect of local phonology, vocabulary, pace and talk-in on output. A numerical accuracy rating by itself would not capture the reason for failure or success of an aligner with an under-documented variety.

Past low-resource alignment research also indicates that the quality of the alignment could vary considerably with certain factors; for example, due to dialect variation, pronunciation differences, phone inventories, or spontaneous speech conditions (Ahn et al., 2024; Leinonen et al., 2022; Tosolini & Bowern, 2025). Descriptive statistics and linguistic interpretation were used, however, as complementary elements of the evidence in the present analysis.

12.3. Analytical Procedure

The analysis was done in three interrelated steps. Initially, recordings were sorted and made into WAV format. The transcription of each utterance was done manually in Shahmukhi Punjabi script and a standard Roman transliteration was prepared for pronunciation tracking and for constructing the lexicon. The Utterance boundaries, speaker turns were then set up for time-aligned annotation in ELAN.

Secondly, a project specific pronunciation lexicon was created and evolved from common locally recurrent forms. Special consideration was paid to words that were differently pronounced or differently spelt in relation to the standard, such as پائی / paai, ایں / ain, نئیں / nai, نِکھاکھرا / nikhakhra, ایتھے / aithay, and combinations like اَپنا تن ای / apna tan ee. The local lexical entries included in the working lexicon were verified with the participation of native speakers.

These kinds of forms directly tested for the effect of local lexical alignment. For instance, the item نِکھاکھرا / nikhakhra was sensitive to OOV, and the connected phrase اَپنا تن ای / apna tan ee was sensitive both in terms of the lexicon and reduction. Thus the lexicon was a controlled version of an adaptation layer on top of the community transcription than replacement for the original speech data onto the general purpose aligner.

Thirdly, the corpus was processed using MFA and the boundaries checked in Praat. Absolute boundary deviation and segment mismatch was noted for every single utterance. Critically one segment mismatch, OR a boundary deviation of more than 50 ms, was considered problematic for a successful alignment. This operational definition enabled the manual process of verification to be repeated and provided for comparison of initial and refined alignment passes directly.

The quantitative analysis tabulated successful and problematic utterances, OOV occurrences, rate of incorrect OOV per 100 words, and error categories that were not explained by the OOV categories. This was followed by a qualitative analysis of the language sources of residual problems. Of particular interest was the handling of dialect-specific/out-of-vocabulary items, boundary ambiguity of connected speech, vowel reduction/non-standard pronunciation, overlapping speech, background noise, and transcription inconsistency.

The corpus included successful and problematic micro-level examples. In UTT007, the sequence of connected words (apna tan ee kam aanda ae) moved from an absolute BD of 135 ms to a final BD of 14 ms after lexicon refinement without any remaining mismatch and hence was classified as successful. In contrast, UTT008, bearing the sequence paani neeviñ thaañ ee charrhda ae, had an issue that arose even after the process of refinement: the boundary deviation went up from 59 to 90ms while a segment mismatch remained. The corresponding Shahmukhi forms were then kept in the corpus annotation and verified against the original audio. The reasons why numerical thresholds were needed to determine alignment, along with linguistic interpretation of local connected speech, are revealed in these paired examples.

12.4. Reliability and Trustworthiness

Repeat listening, standardized Roman forms, the systematic use of the 50-ms boundary criterion and transcript checking were used for reliability. The output for each utterance was examined using Praat, and the diagnoses of errors made were analyzed with regard to the audio material, the transcripts and the local pronunciation evidence. The consistency of status assignment was done both prior to and following the lexicon refinement process, minimizing the likelihood that improvements were due to variable manual rulings.

The trustworthiness was still further supported by comparing the Shahmukhi with the Romanized forms, and by having the forms used by native speakers confirmed within the projects. The interpretive checking process was still needed because the material in the corpus was community speech that was not put out in a controlled laboratory setting, but for expressions of local meaning or discourse forms or reductions.

12.5. Ethical Handling During Analysis

Ethical procedures outlined in section 11 were adhered to during analysis and archiving, with codes for speakers retained in all of the analysis files while no personally identifiable information was included in the examples.

13.Findings

13.1. Overview of Findings

Forced alignment resulted in a measurable improvement following pronunciation-lexicon refinement. A total of 74 utterances of a possible 120 were successful (61.7%) and 46 problematical (38.3%) during the initial pass for MFA. Following the refinement process 96 utterances (80.0%) fulfilled the success criterion while 24 utterances (20.0%) represented problematic ones (18.3 percentage points more fulfilling the success criterion). The number of OOV went down from 62 to 18 and the rate of OOV went down from 3.96 to 1.15 number of OOV per 100 words (down by 71.0%). The results demonstrate that local lexicon refinement substantially enhanced corpus-level alignment, but some linguistic and acoustic issues still remain. Initial and refined performance measures are summarized in Table 1.

Table 1. Forced-alignment performance before and after pronunciation-lexicon refinement

Indicator	Initial	After refinement	Change
Successful utterances	74 (61.7%)	96 (80.0%)	+18.3 percentage points
Problematic utterances	46 (38.3%)	24 (20.0%)	−18.3 percentage points
OOV instances	62	18	−71.0%
OOV rate per 100 words	3.96	1.15	−2.81

13.2. Findings Related to Objective 1

The goal of the first objective was to determine the main issues that impact alignment accuracy. After the refinement process, 24 utterances remained problematic, of which the most frequent error types were dialect-specific/OOV items (29.2%) and connected-speech boundary ambiguity (29.2%). Five (20.8%) were listed as due to vowel reduction and/or non-standard pronunciation, three (12.5%) as due to overlapping speech, and one (4.2%) each due to either background noise or transcription inconsistency. This distribution illustrates that the other errors were not caused by a lack of one specific technical skill only, but by a

combination of local lexical structure, spontaneous speech and recording conditions. The distribution of the residual is given in Table 2.

Table 2. Residual primary misalignment errors after lexicon refinement

Primary error type	Frequency	% of problematic utterances
Dialect-specific/OOV item	7	29.2%
Connected speech/boundary ambiguity	7	29.2%
Vowel reduction/non-standard pronunciation	5	20.8%
Overlapping speech	3	12.5%
Background noise	1	4.2%
Transcription inconsistency	1	4.2%

These are exemplified in the transcribed examples. Alignments like پائی / paai, ایں / ain, نئیں / nai, and نِکھاکھرا / nikhakhra were mentioned several times in connection with OOV-sensitive and/or dialect-sensitive alignments. Similarly, connected sequences like اَپنا تن ای کم آندا اے / apna tan ee kam aanda ae also exhibited reduction and compression of boundaries; rapid delivery in conversations was also complicating boundaries, as was overlap. The patterns account for the failure of the pronunciation-lexicon refinement to account for many occurrences while still maintaining errors due to connected speech and acoustic competition.

13.3. Findings Related to Objective 2

Overall performance and variation by speech context were evaluated in the second objective. Semi-structured interview utterances had the highest success rate at 90.0% (36/40) followed by personal utterances 85.0% (34/40). Only 65.0% of natural-conversation utterances were successfully aligned (26/40), indicating that natural conversation was quite difficult to align. Qualitative observation testifies to the fact that spontaneous, connected speech with varying pacing and overlap etc., decreases the clarity of boundaries; guided or more carefully paced speech increases the clarity. The refined performance by speech context is detailed in table 3.

Table 3. Refined alignment performance by speech context

Speech context	Evaluated n	Successful n	Success rate
Semi-structured interview	40	36	90.0%
Natural conversation	40	26	65.0%
Personal narrative	40	34	85.0%

The impact of lexicon refinement could also be observed at utterance level. UTT007 initially had a problem with boundary deviation but after refinement it was reduced to 14ms from 135ms and was classified as successfully aligned. On the other hand, UTT008 was still a problem at 90 ms, and exhibited a segment mismatch, indicating that lexical adaptation can help resolve certain local-form issues, but not fully offset connected-speech and acoustic complexity. Manual checking in Praat was therefore necessary to identify true improvement from residual alignment errors.

13.4. Findings Related to Objective 3

The third objective explored usefulness for corpus development and for language documentation. The rise in percentage of successful utterances from 61.7% to 80.0% and the 71.0% decrease in the number of OOV instances was evidence that forced alignment materially facilitated the ability to organize speech into searchable time-aligned units. All the other 20.0% problematic utterances were also analytically productive since errors could be explicitly written down, rather than simply left as "alignment" output.

So it created more than automated segmentation workflow. It established a structured record of the audiotapes, the Shahmukhi transcription, the Roman transliteration of the text, entries of local pronunciation of the text, alignment status, and diagnosis of the error. This hierarchical structure of organization contributes to making the corpus useful in future phonetic, lexical, morphosyntactic and sociolinguistic studies, at the same time maintaining the relationship with the documentary record of authentic community discourse.

14.Discussion

The results illustrate that forced alignment can be useful to develop an under-documented Shakargarh Punjabi speech corpus, however, the performance of forced alignment is strongly dependent on linguistically and speech-contextually sensitive factors. The results of successful utterances indicate that the quality of alignment was not determined by the quality of the baseline model, but by the fact that lexicon was refined as well, the ratio of successful utterances has increased by 18.3 percentage points after that process. Local pronunciation and the use of local vocabulary explicitly mattered, and there appears to have been an additional difficulty with connected and spontaneous speech which is indicated by a lower success rate for natural conversation (65.0%) than for interviews (90.0%) or narratives (85.0%).

One of the very obvious reasons for difficulty was the use of dialect specific words. Even with general resources of pronunciation, forms like پائی / paai, ایں / ain, نئیں / nai, and نِکھاکھرا / nikhakhra were not consistently represented and dialect-specific/OOV items remained high (29.2%). Characteristic of adaptation and validation in under-resourced alignment (Leinonen et al., 2022), and in line with Tosolini and Bowern (2025), who demonstrate that multilingual models might still need to be language- or variety-adjusted. The reduction in the number of OOV instances (from 62 to 18) offers an immediate, quantifiable proof of that interpretation.

Connected speech was also crucial. Another 29.2% of the remaining problematic utterances were due to connected-speech boundary ambiguity, with vowel reduction/non-standard pronunciation making up 20.8%. This is consistent with the results of Ahn et al. (2024), who find that the phonological structure and the strategy of modeling influence the results of alignment. The present results in a context-specific context support this, as even native speech, which was processed with lexicon refinement, proved to be more difficult than guided speech to make an appropriate alignment of its boundaries, which suggests that beyond lexicon refinement there are still other aspects of boundary ambiguity that need to be resolved.

Forced alignment, despite these constraints, was undoubtedly documentary. Four of the five evaluated utterances fulfilled the operational success criterion, providing a considerably larger amount of material that also can be searched, displayed, and used in a time-aligned manner. This is part of a larger argument in support of greater long-term usability of documentary resources through aligned corpora (Berez-Kroeker et al., 2023; Paschen et al., 2020). Thus, we can see that forced alignment was not just a useful technical tool, but also a production process as part of the construction of a corpus that allowed for an increase in the amount of recorded community speech.

The results also make it clear the pronunciation lexicon's role and manual verification. Lexical refinement, specific to the project, lowered the OOV rate by 71.0%, but seven utterances (29.2%) remained problematic because of OOV items, which means that a fully automatic workflow would have swallowed important remaining errors. These, therefore, had to be combined for methodologically equivalent use, as was the combination of local lexicon development and Praat verification. By following a trainable approach for creating alignments for the popular MFA framework, this recent pilot shows why it is essential for users to adapt to their local domain and not rely on a given “baseline” dictionary.

The study also affirms the connection of the corpus linguistics and language documentation. Language documentation aims at long-term, multi-purpose documentation of practices while corpus linguistics brings systematic organization, retrieval and analysis (Himmelmann, 2006). These objectives were partially met by using forced alignment in order to convert the oral recordings of Shakargarh Punjabi into structured, time referenced corpus data, but retaining the connection to the original recordings (oral baseline). This is especially true of a local variant, which is largely spoken and which has hardly appeared in retrievable digital materials.

Overall, the three research objectives were met. The study established the main alignment error sources, provided quantitative results of the performance with and without the pronunciation-lexicon refinements, and presented the usefulness of the documents by considering a semi-automatic workflow. The evidence suggests forced alignment is effective, but it requires a combination of local linguistic knowledge, consistent transcription, explicit errors and manual confirmation to obtain the best results.

15.Conclusion

Based on this study the conclusion is drawn that with local knowledge of the Punjabi pronunciation as well as manual validation, forced alignment could play a valuable role in the development of a searchable Shakargarh Punjabi speech corpus. After pronunciation-lexicon refinement, the successful alignment rate jumped to 80.0% from 61.7%, and the number of OOVs was reduced by 71.0%. However, it is not a purely automatic process as evidenced by residual errors involving dialect-specific forms, connected speech, vowel reduction and overlap. The most effective method was thus to have semi-automatic procedures (Shahmukhi transcription and standardized Roman transliteration based on the MFA procedures and project-specific lexicon) followed by verification carried out by Praat software, with a 50-ms boundary criterion. This workflow resulted in a more systematic, searchable and reusable corpus resource from a pilot collection of community speech, and offers an empirical foundation for low-resource Punjabi documentation.

16.Study Limitations

The study had its limitations on the pilot level – 12 speakers, 24 recordings, 120 evaluated utterances, 19.79 minutes of speech. Though this data set was adequate to make controlled comparison between initial and refined alignment, the results should not be generalized for all

speech situations and speakers of Shakargarh Punjabi. The analysis also relied on manual transcription, Roman transliteration, as well as verification by researchers and online alignment tools were not available for Shakargarh Punjabi. Such restrictions are crucial since performance could differ in the case of larger corpora, another, more variable set of speakers, or different recording settings.

17.Recommendations and Future Directions

Especially future studies should include more speakers, age groups, villages and interactional situations, with the same explicit verification criteria. Project-specific pronunciation lexicon should be expanded by ongoing native-speaker validation along with systematic notation of pronunciation variations. Comparative testing with other acoustic or multilingual alignment model hypotheses would be useful to explore whether the achievements shown here are unique to the MFA pipeline discussed here or whether they are generalizable to other systems. An aligned corpus, with more data, can then inform phonetic analyzes, lexical profiling, morphosyntactic analyzes, sociolinguistic variation analyzes, and community-oriented digital documentation of Shakargarh Punjabi.

References

Ahn, E. P., Chodroff, E., Lapierre, M., & Levow, G.-A. (2024). The use of phone categories and cross-language modeling for phone alignment of Panãra. In Proceedings of Interspeech 2024 (pp. 1505–1509). ISCA.

Berez-Kroeker, A. L., Gabber, S., & Slayton, A. (2023). Recent advances in technologies for resource creation and mobilization in language documentation. Annual Review of Linguistics, 9, 195–214.

Boersma, P., & Weenink, D. (2026). Praat: Doing phonetics by computer [Computer software]. University of Amsterdam.

District Narowal, Government of the Punjab. (n.d.-a). District profile.

District Narowal, Government of the Punjab. (n.d.-b). Our history.

District Narowal, Government of the Punjab. (n.d.-c). Important places.

Himmelmann, N. P. (2006). Language documentation: What is it and what is it good for? In J. Gippert, N. P. Himmelmann, & U. Mosel (Eds.), Essentials of language documentation (pp. 1–30). Mouton de Gruyter.

Janssen, M., & Seifart, F. (2025). Searchable language documentation corpora: DoReCo meets TEITOK. In É. Le Ferrand, E. Klyachko, A. Postnikova, T. Shavrina, O. Serikov, E. Voloshina, & E. Vylomova (Eds.), Proceedings of the Fourth Workshop on NLP Applications to Field Linguistics (pp. 58–64). Association for Computational Linguistics.

Leinonen, J., Partanen, N., Virpioja, S., & Kurimo, M. (2022). Semiautomatic speech alignment for under-resourced languages. In A. K. Ojha, S. Ahmadi, C.-H. Liu, & J. P. McCrae (Eds.), Proceedings of the Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-Resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference (pp. 17–21). European Language Resources Association.

Max Planck Institute for Psycholinguistics. (2026). ELAN (Version 7.1) [Computer software]. Author.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. Proceedings of Interspeech 2017, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Paschen, L., Delafontaine, F., Draxler, C., Fuchs, S., Stave, M., & Seifart, F. (2020). Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo). In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 2657–2666). European Language Resources Association.

Passmore, S., Roque, L. S., Gillespie, K., Nath, S. K., Davey, K., Mullan, K., Cawley, T., Biggs, J., Billington, R., Evans, B., Thieberger, N., Evans, N., & Barth, D. (2025). English-based acoustic models perform well in the forced alignment of two English-based Pacific Creoles. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 31172–31183). Association for Computational Linguistics.

Tosolini, A., & Bowern, C. (2025). Multilingual MFA: Forced alignment on low-resource related languages. In J. Lachler, G. Agyapong, A. Arppe, S. Moeller, A. Chaudhary, S. Rijhwani, & D. Rosenblum (Eds.), Proceedings of the Eighth Workshop on the Use of Computational Methods in the Study of Endangered Languages (pp. 100–109). Association for Computational Linguistics.