Al-Bahir Journal for Engineering and Pure Sciences
BJEPS

Volume 5 | Issue 2 Article 7

Review of current Trends in Information Technology concerning Phonetic Similarity”

Zaid Rajih Mohammed
Faculty of Medical Sciences, Jabir ibn Hayyan University for Medical and Pharmaceutical Sciences, Najaf, Iraq

Ahmed H. Aliwy
Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

Review of current Trends in Information Technology concerning Phonetic Similarity”

Conflict of Interest

No conflict of interest to declare

Funding

No external funding was provided for this research

Author Contribution

Zaid Rajih Moahmmed: Read and collect all papers for review. Ahmed H. Aliwy: Conducted analysis and wrote the paper

Data Availability

Publicly available data

REVIEW

Review of Current Trends in Information Technology Concerning Phonetic Similarity

Zaid R. Mohammed [a,b,*], Ahmed H. Aliwy [b]

[a] Faculty of Medical Sciences, Jabir ibn Hayyan University for Medical and Pharmaceutical Sciences, Najaf, Iraq
[b] Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

Abstract

With the increasing availability of textual information in various languages via the Internet in homes and companies through Internet and intranet services, there is an urgent need for the technologies and tools necessary to process this information, phonetic representation, and voice interaction. For example voice to voice machine translation need to phonetic mapping and similarity among the languages especially for names and foreign words. This one example of the importance of phonetic mapping and similarity. This article aims to describe, in detail, the recent surge in interest and advancements in phonetic similarity (PS), phonetic representation, and phonetic mapping researches. PS and phonetic representation are a fundamental elements in information technology, supporting applications such as search engines, speech recognition and voice to voice MT systems. The importance of PS is demonstrated, the main characteristics of phonetic processing are highlighted, and the standardization aspects in converting text to phonetic representation are clarified. The current study presented a survey of previous studies for a period of time (2000/2024). By utilizing advanced AI algorithms and diverse linguistic datasets, new avenues for comprehending the dynamics and alterations of extinct languages over time, as well as, their pronunciation mechanisms, can be explored. This change presents a breakthrough in language learning and cross-cultural communication in addition to being a technical advancement. Additionally, various linguistic resources and approaches used in the PS field are explained. Also, the features of common tools are described, and standard evaluation metrics are illustrated. The article also reviews the current state of art for PS research and converting text to phonetic representation. Finally, we present our analysis and conclusions. Disseminating these findings is crucial for scholars focused on phonetic similarity and its related methodologies.

Keywords: Phonetic similarity, Phonetic representation, Text to speech, Artifitial intelligence

1. Introduction

The language, acting as the building block of human communication, takes on so many forms across the world. As starting point, the first used language was sign language and then it was developed to be spoken language and hence it was recorded as written symbols. Recently, all these forms are used but in vey different percentage. For example deaf persons use the sign language to communicate with another person [1]. Besides traditional ones, in the era of digital technologies, digital language emerged. Every language has its unique characteristics, power, and limitations to interpret the diversity and adaptability of human communication [2].

In spoken language, phonetics is a subfield of linguistics that is concerned with the physical nature of speech sounds. It takes into account the study of how sounds emanate from the human mouth through the air and how these are received by humans to make meaning of the auditory cues. Phonetic analysis is the process of analyzing spoken language and breaking it down into its constituent segments, called phonemes, and studying the articulatory processes that entail their production. Learning and understanding phonetic features

Received 18 July 2024; revised 14 September 2024; accepted 19 September 2024.
Available online 19 October 2024

* Corresponding author at: Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Iraq.
E-mail addresses: zaid.rajah@jmu.edu.iq (Z.R. Mohammed), ahmedh.almajidy@uokufa.edu.iq (A.H. Aliwy).

https://doi.org/10.55810/2313-0083.1077

allows a linguist to obtain knowledge in cross-linguistic comparability, which will enable them to discover patterns within sound variation and change. One of the most important components of this natural language processing is phonetic similarity, which has been causing more advancement in this field along with other recent advancements in varied programs of natural language processing, recognition of speech, and analysis of data [3]. This make extracting phonetic similarity among different languagesas an important task for understanding of linguistic elements to be used in many applications. Many researchers try to use many techniques with diffrerent test languages for doing this task. This study overviews how researchers used information technology while researching phonetic similarity, considering its applications, challenges, and prospects for the domain [4].

2. Phonetic similarity survey objective

A survey of phonetic similarity implies an analysis in which methods and techniques of measurement and analysis of the likeness between speech sounds or words in their pronunciation are considered in detail. In this work, a total of 108 different papers regarding phonetic similarity were read and analyzed. Additionally, it discusses the problems involved in phonetic similarity, the way linguistic variability in languages, dialects, or any other type of speech affects it, as well as the computational complexity of the analysis for large datasets. In short, a survey of phonetic similarity can turn out to be helpful in many ways, such as knowing about the state-of-the-art techniques or about future directions that would foster and bring speech processing and communication to the next level. Table 1 shows the distribution of papers according to the year of publication of the paper. Table 2 shows the distribution of papers according to the application of phonetic similarity mentioned in section 2. Table 3 shows the distribution of papers according to the number of natural languages adopted in the research, where Multi Languages refer to that paper using more than two languages, Bi Languages refer to that paper using two languages and others refer to languages appear at less once or twice such as (Italian, Thai, Indonesian, Kurdish, Turkish, and Urdu). The distribution of research articles that used different datasets varies due to the diversity of the data employed. Some researchers utilized specialized databases containing audio files and phonetic representations of letters, particularly in applications like Spelling Correction and similar programs. Conversely, another set of researchers depended on specialized dictionaries for phonetic transcription of words, utilizing them to train models that capable of pronouncing words across different languages as in G2P and Transliteration applications. Among the notable dictionaries is the CMU dictionary [5].

Table 1. Show the papers according to years.

No	Years	Papers
1	Before 2010	21
2	2010–2020	42
3	After 2020	45

Table 2. Show the papers according the phonetic similarity application.

No	Application	Papers
1	Cognate detection	30
2	G2P	22
3	ASR	22
4	NER	6
5	Spelling Correction	8
6	Transliteration	18
7	All Applications	2

Table 3. Show the papers according to the number of languages.

No	Language	Papers
1	Multi Languages	56
2	Bi Languages	14
3	Arabic	19
4	English	11
5	Persian	2
6	Others	6

3. Application and tasks of phonetic similarity

One area of natural language processing is phonetic similarity, which has multiple applications as it cuts across many scientific domains therefor it is used in wide area of applications across advances in artificial intelligence (AI), natural language processing (NLP), and modern speech identification and processing techniques. It also comes from the amazing developments progress that have been made in all areas of information technology, from algorithms to processing large amounts of data quickly and efficiently to using cloud computing and big data processing [6]. Because of this, processing and analyzing phonetic data has grown simpler, and features that may be applied to a variety of language applications have been extracted [7,8]. The following are some important applications and tasks of using phonetic similarity:

1- **Cognate detection**: is a key task in computational linguistics aimed for identifying words in different languages that share a common origin and meaning. Cognates are words that have evolved from a common ancestral word and retained similar phonetic and semantic properties across languages.

2- **Text-to-Speech (TTS)** is a disruptive technology wherein the written words are spoken out. In the recent past, it has developed at a great rate of improvement in terms of more intelligible speech, and the speech output is more human-like. The Grapheme to Phonem (G2P) is the core component for any TTS synthesis process as it produces more natural speech corresponding to the input of written text. This process plays an important part in converting the written text into its corresponding phonetic representation.

3- **Spelling correction** forms one of the most basic steps in natural language processing; it is the detection and automatic correction of all those words that are most likely to have been misspelled in written text. Mostly, the misspellings emanate from typographical errors, phonetic ambiguity, and even dialectal variations.

4- **Automatic speech recognition (ASR)** is one of the most revolutionary technologies that has been developed, allowing machines to convert spoken language into written text, thus allowing humans to interact with computers naturally.

5- **Multilingual named entity recognition (NER)** is a critical problem in natural language processing that identifies and classifies named entities, such as names of people, organizations, locations, dates, and so forth, in different languages. **This task can be done using phonetic similarity for the same name in different languages.**

6- **Transliteration** or **Language transliteration** is a process in which a written text in one writing system is converted into another, generally maintaining the pronunciation and other phonetic values of the original word langauge. It is found to be more useful in obtaining exact results in proper names, names, and loanwords for languages with different writing systems.

4. Phonetic transcription notations

Phonetic transcription, also called Phonetic notation, is a written symbols that represent speech sound. There are many phonetic transcriptions such as; International Phonetic Alphabet (IPA), SAMPA (Speech Assessment Methods Phonetic Alphabet), X-SAMPA (Extended SAMPA), ARPAbet, Hanyu Pinyin, Jyutping (Yale romanization of Cantonese), Hangul Phonetic Alphabet (South Korea and North Korea) and many others. The most used phonetic transcription is the International Phonetic Alphabet (IPA), Fig. 1 shows this phonetic transcription, as it ca n be seen it provides a unique symbol for each distinct sound (phoneme) in human languages.

5. Available software

There are many available programs, software and systems that are used in phonetics processing which

International Phonetic Alphabet (IPA) ˌɪntəˈnæʃn̩l fəˈnɛtɪk ˈælfəˌbɛt

Consonants (pulmonic)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Fig. 1. International Phonetic Alphabet (IPA) phonetic transcription.

are tailored to various linguistic and computational needs. These programs serve a vital role in fields such as linguistics, speech pathology, natural language processing, and computational linguistics. They are enabling the users to analyze, transcribe, and manipulate phonetic data with precision and efficiency. Table 4 shows some of these softwares with little descriptions. Beside these system there are many libraries the can be used for phonetic processing such as fuzzy, Phonetics and Epitran in Python, Natural and Fuzzy-Phonetics in JavaScript and many others.

The CJKI Arabic Romanization System, known as CARS, is a novel phonemic transcription method primarily designed to facilitate learning and assist linguists in examining the phonological aspects of Modern Standard Arabic (MSA). Its key attribute lies in its enhanced readability, making it particularly user-friendly for both learners and researchers alike [9].

Bakar et al. [10], showed an encoding within a Malay language corpus that written in Jawi script to ensure consistency and standardization of the corpora by employing the Buckwalter transliteration method to align similar characters.

Toma et al. [11], introduced a novel open access resource, the machine-readable phonetic dictionary for Romanian - MaRePhoR. It contains over 70,000 word entries, and their manually performed phonetic transcription.

Uroman is a tool for converting text in myriads of languages and scripts such as Chinese, Arabic and Cyrillic into a common Latin-script representation [12].

Epitran is a massively multilingual G2P system. To maximize its usefulness, it is written in Python and distributed as open source software, it supports 61 languages [13].

WikiPron, an open-source command-line tool designed for extracting pronunciation data from Wiktionary, a collaborative online dictionary available in multiple languages. Initially, it outlines the design and functionality of WikiPron before delving into the challenges encountered while scaling the tool to automatically generate a database comprising huge amount of pronunciations across 165 languages [14].

6. The available datasets

The construction of a dataset for phonetic similarity is a meticulous process that often blending expertise from linguistics, data science, and machine learning. Such a dataset typically can be used in different research field spanning varius linguistic domains, including phonetics, phonology, and speech processing. We list her some of the freely available dataset where Table 5 shows the summary.

- LEXiTRON-Pro [15] is a machine-readable pronunciation guide for Thai. It's crucial for Thai language processing, like speech recognition. Based on the International Phonetic Alphabet (IPA), it includes 21 single initial consonants, 17 cluster consonants, 24 vowels, and 5 tones as specified by NECTEC.
- PanPhon was developed by [16]. It has phonological features, which make using the extensive PanPhon database, produces superior outcomes compared to character-based models.
- CogNet was presented by [17], an extensive lexical database designed to offer cognates—words sharing common origins and meanings—across

Table 4. The available software in phonetic similarity.

No	Developer	software	Language	Application
1	[9]	CJKI	Arabic	Transliteration
2	[10]	Buckwalter transliterator	Malay	Transliteration
3	[11]	MaRePhoR	Romanian	Transliteration
4	[12]	Uroman	Multi Languages	Transliteration
5	[13]	Epitran	multi Languages	G2P
6	[14]	Wikipron	Multi Languages	Cognate

Table 5. Different dataset for multiple languages that are freely available.

No.	Dataset	Developer	Language	Application/Task
3	LEXiTRON-Pro	[15]	Thai	ASR
4	PanPhon	[16]	Multi Languages	Transliteration
1	CogNet	[17]	Multi Languages	Cognate
5	Openslr	[18]	Multi Languages	Cognate
2	CogNet	[19]	Multi Languages	Cognate
6	MohAli Dataset	[20]	Arabic-English	Transliteration

various languages. Currently, the database encompasses 3.1 million pairs of cognates spanning 338 languages and utilizing 35 distinct writing systems.

- The Multilingual LibriSpeech (MLS) dataset, presented by [18], offers a substantial resource for speech studies, drawn from audiobooks on LibriVox. It encompasses 8 languages, with approximately 44.5K hours in English and a combined total of around 6K hours across the other languages.
- CogNet (enhanced) is an enhanced version of CogNet by [19]. It is vast database of sense-tagged cognates—words sharing a common origin and meaning across different languages. CogNet is dynamic, with ongoing updates; it currently comprises more than 8 million cognate pairs across 338 languages and 35 writing systems, with future releases underway.
- MohAli dataset was developed by [20] in the form (English Word, English Phonetic, equivalent Arabic Word, and Arabic Phonetic). It was used in a semi-automated framework designed for generating a multilingual phonetic English-Arabic corpus, specifically tailored for application in multilingual phonetically and semantic similarity tasks.

7. Phonetic embedding

Phonetic embedding is a transformative concept in phonetic processing, revolutionizing the way we understand and analyze speech sounds. At its core, phonetic embedding involves converting phonetic representations, such as phonemes or acoustic features, into high-dimensional vectors. These vectors capture the nuanced relationships between speech sounds, effectively encoding their phonetic properties in a continuous space. Researchers leverage these embeddings to tackle various tasks, from speech recognition and synthesis to accent identification and language modeling and it captures the subtle phonetic characteristics and similarities between sounds. Some of the Phonetic embedding works will be showed in this section.

Artetxe et al. [21], demonstrated that each embedding model encompasses more information than initially evident. Through a linear transformation that modifies the similarity ranking of the model without relying on external resources, it can be customized to yield improved outcomes in various aspects. This offers a fresh insight into how embeddings encode diverse linguistic information. They used STS Benchmark dataset for Multi Languages and applied for NER task.

El-Geish [22] proposed the learning of encoders capable of mapping variable-length sequences, either acoustic or phonetic, representing words, into fixed-dimensional vectors within a common latent space. This enables the distance between two word vectors to reflect the similarity in their sounds. He used Librispeech dataset for Multi Languages and applied for ASR system.

Feng and Wang [23], explored the integration of a Word2vec model with an attention-based end-to-end speech recognition model. The study outlines the development of a phoneme recognition system utilizing the Listen, Attend, and Spell model. They used TIMIT dataset for Multi Languages and applied for ASR system.

Kanojia et al. [24], explored the identification of cognate word pairs across ten Indian languages alongside Hindi, employing deep learning techniques to ascertain their cognate status. The study investigates the utilization of Indo Wordnet as a promising tool for detecting such pairs through orthographic similarity-based approaches. They used Parallel Corpora based dataset for Multi Languages and applied for Cognate task.

Sharma et al. [25], introduced an innovative approach to computing phonetic similarity among words, inspired by human perception of sounds. This method is utilized to develop a continuous vector embedding space, which clusters words with similar sounds together. They used CMU dataset for Multi Languages and applied for Cognate task.

Zouhar et al. [7], introduced several innovative approaches that utilize articulatory characteristics to construct phonetically informed word embeddings. It also introduces a collection of phonetic word embeddings to promote their community development, assessment, and utilization. Despite the existence of various methods for acquiring phonetic word embeddings, there remains inconsistency in evaluating their efficacy. They used CMU dataset for Multi Languages and applied for All Tasks.

8. Techniques of producing phonetic representation

Phonetic representation technique is the technique that used to converte the written text (graphemes) or spoken sounds into phonetic transcripts (symbols) as a pronunciation of these words. There are various techniques used to produce phonetic representations, each with its own methods and applications. In this section, a list of works for producing phonetic transcripts will explore.

Altinok [26], described the architecture and implementation of a rule-based grapheme-to-phoneme converter for Turkish. They used TRmorph dataset for Turkish and applied for G2P task.

El-Imam [2], presented algorithms to implement the rules and the assessments of the output of converting Arabic text into sounds. They used Umm Alqura list dataset for Arabic and applied for G2P task.

Harrat et al. [27], presented approach to build a G2P converter for Algiers dialect. This approach is rule based; it gives perfect results for Arabic words and French words phonologically altered. They used Algiers dialect corpus for Arabic and applied for G2P Task.

Nehar [28],tried to suggest methods for pairing Arabic personal names. The initial method relies on aligning strings and phonetic transcription, employing tailored scoring functions to gauge similarity. The second method employs machine learning techniques to develop an appropriate model for this task. They used Author's names dataset for Arabic and applied for NER Task.

Bisani and Ney [29]presented a novel estimation algorithm known for its remarkable accuracy across various databases. It explores the impact of maximum approximation during training and transcription, the relationship between model size parameters, n-best list generation, confidence metrics, and phoneme-to-grapheme conversion. They used Special dataset for English and applied for G2P Task.

Sindran [30], in his thesis, He describes work carried out to develop an automatic phonetic transcription for Standard Arabic (SA) text using the rule and exception dictionary approach. The main contributions of this work are the development of a high-precision tool for the phonetic transcription software, a statistical analysis of SA linguistic units that will help in boosting the software performance, and developing a program to classify the meters of classic Arabic poems based on an accurate phonetic transcription result. The three applications developed as a result of developing such a tool for SA phonetic transcription are now described. They used Six corpora have been used in this work for Arabic and applied for ASR Task.

Ghio et al. [31],presented a measurement technique utilizing phonological transcriptions. An algorithm is employed to automatically and accurately calculate the distances between produced phonological forms and their expected counterparts, leveraging cost matrices derived from phoneme feature disparities. They used Speakers dataset for English and applied for Cognate Task.

Hixon et al. [32],investigated different metrics for evaluating the automatic pronunciation techniques of three grapheme-to-phoneme packages using a comprehensive dictionary. Introducing two novel metrics, it utilizes a newly devised weighted phonemic substitution matrix, derived from substitution frequencies in a trusted collection of alternative pronunciations. They used CMU dataset for English and applied for G2P Task.

Jucksriporn and Sornil [33],introduced a method for determining word distances based on phonetic similarities of homophones, rather than relying on spelling. The technique was tested using a database of names belonging to Thai individuals and locations. They used Private dataset for Thai and applied for ASR Task.

Masson M and Carson-Berndsen [34],investigated pronunciation differences among non-native speakers, utilizing both theoretical concepts and real-world data. It proposes two approaches to analyze these variations: one rooted in phonetic and phonological theories, and the other utilizing a text-to-speech system. They used TIMIT corpus for English and applied for ASR task.

Kantor and Hasegawa-Johnson [35], studied how to generate phonetic pronunciations from printed word forms, which is important for applications like text-to-speech systems. It describes a method that uses hidden Markov models (HMMs) to generate pronunciations for out-of-vocabulary words and word fragments from the Fisher speech corpus. They used CMU dataset for English and applied for G2P task.

Nahar et al. [36],used the Hidden Markov Toolkit (HTK) to develop a phoneme recognizer for identifying Arabic phonemes. It ran a series of experiments, varying factors such as the number of Hidden Markov simulate (HMM) states and Gaussian mixtures used to simulate Arabic phonemes to find the best configuration. They used KAPD dataset for Arabic and applied for G2P task.

Arab and Azimizadeh [37],introduced a Persian Letter-To-Sound conversion system using Classification and Regression Tree. Such a system is crucial for Text-To-Speech (TTS) technology as it's impractical to compile every word with its pronunciation in a lexicon. They used Persian Linguistic Database for Persian and applied for G2P task.

Khan [38],investigated various acoustic features such as pitch, energy, spectrum flux, zero-crossing, entropy, and MFCCs. It used Sequential Forward Selection to identify the most suitable features and employed the K-Nearest Neighbors classifier to detect mispronunciations in Arabic phonemes. They used private Dataset for Arabic and applied for Spelling Correction Task.

Maqsood et al. [39],described a CAPT method meant to help Pakistani people improve their pronunciation of difficult Arabic sounds. The system uses an articulatory phonetic framework with machine learning classifiers such as Naïve Bayes and K-NN to detect mispronunciations accurately. They used private Dataset for Arabic and applied for Spelling Correction task.

Nazir et al. [40],provided a novel approach to detecting mispronunciations that employs phone grouping and probabilistic error assessment. Unlike traditional approaches, it clusters phonemes based on mistake likelihood, lowering the number of classifiers necessary while conserving memory and time. The study assesses the Support Vector Machine (SVM) classifier's performance on a dataset of 28 Arabic phonemes. They used Speakers dataset for Arabic and applied for Spelling Correction task.

Johnson [41],introduced a matrix of phone-distance measures, valuable for managing extensive conversational speech databases. It outlines the process of creating this matrix from discrepancies in transcriber labeling and explains its application in aligning phonetic transcriptions of spoken discourse with their citation forms through a dynamic time warping (DTW) algorithm. They used Private dataset for English and applied for G2P task.

Ibrahim et al. [42],tried to identify key input features that enhance speech recognition. The proposed method employs a genetic algorithm for optimal feature selection. Initially, a baseline model utilizing feedforward neural networks is constructed. This model serves as a benchmark to compare the results of the proposed feature selection approach against a method that incorporates all elements of a feature vector. They used KAPD dataset for Arabic and applied for ASR task.

Galescu and Allen [43],introduced a statistical model corresponding author bidirectional transformation between spelling and pronunciation, regardless of language. It relied on joint graphene/phoneme units obtained from automatically aligned data. They used CMU dataset for English and applied for G2P task.

Suyanto et al. [44],introduced an Indonesian Grapheme-to-Phoneme (G2P) model called NGTSP, which combines n-gram methodology with a stemmer and phonotactic rules to address existing challenges. Through a 5-fold cross-validation analysis involving 50,000 Indonesian words, the NGTSP model demonstrates a significantly lower Phoneme Error Rate (PER). They used 50 k Indonesian words and applied for and applied for G2P task.

Halpern [45],discussed the orthographic variation of Arabic personal names, particularly the challenges arising from transcribing them into the Roman script. It outlines the methodology behind compiling the Database of Arabic Names (DAN), a comprehensive lexical resource containing millions of Arabic names and their variations in both romanized and fully vocalized Arabic. They used Private dataset for Arabic and applied for NER task.

Mahmudi and Veisi [46],introduced a G2P conversion approach grounded in Kurdish language phonological rules, departing from traditional reliance on pronunciation dictionaries and data-driven techniques. By precisely applying prioritized constraints, refine the process to exclude less desirable options, ensuring the selection of a single accurately formed pronunciation for each word. They used Private dataset for Kurdish and applied for G2P task.

Bhagat and Hovy [47],introduced two innovative phonetic models aimed at generating a variety of potential spellings for a given name. The methods showcased a threefold enhancement over the standard approach, producing four times as many quality name variations as a human could while still maintaining a commendable level of accuracy. They used CMU English and applied for NER task.

Algabri et al. [48], in their study, innovative deep learning methods were suggested for constructing a highly efficient and adaptable Computer-Assisted Pronunciation Training (CAPT) system aimed at detecting and diagnosing mispronunciations (MDD) and providing articulatory feedback for non-native Arabic speakers. They used Arabic speaker dataset for Arabic and applied for Spelling Correction task.

Asif et al. [49],created a framework employing Deep Neural Networks (DNN) to categorize Arabic short vowels. The model was built entirely from the ground up, involving: (i) the compilation of a fresh audio dataset, (ii) the design of a neural network structure, and (iii) the refinement and enhancement of the model through multiple iterations to attain superior accuracy in classification. They used Audio of 85 individual from 42 males, 43 females dataset for Arabic and applied for Spelling Correction task.

Nazir et al. [50],introduced various approaches for detecting mispronunciations, including a method utilizing Convolutional Neural Network features (CNN_Features) and another technique based on transfer learning. They used Private dataset for Arabic and applied for Spelling Correction task.

Ziafat et al. [51],divided the subject into two stages. Initially, it trained the model to identify individual letters, with a concentration on Arabic alphabet categorization. It then trained the machine to measure pronunciation accuracy, with a focus on Arabic alphabet pronunciation classification. They used audio samples of the Arabic alphabet and applied for and applied for Spelling Correction task.

Barman and Boruah [52]provided a modified Long Short Term Memory network (LSTM) model, a Recurrent Neural Networks (RNN) version explicitly designed for instant messaging applications. The primary goal is to predict the following word(s) based on the user's current words. They used Private dataset for Arabic and applied for G2P task.

Behbahani et al. [53],delineated the grapheme-to-phoneme conversion process as a sequential labeling task and employs modified Recurrent Neural Networks (RNNs) to construct an intelligent and cohesive model for this objective. They used FarsDat dataset for Persian and applied for G2P task.

Jakobi [54], in his thesis, he addresses the challenge of multilingual grapheme-to-phoneme (G2P) conversion. It involves becoming acquainted with the existing data landscape and integrating phonetic attributes into model inputs. This is achieved by extending the input phonemes with encoded phonetic features. He used CMU dataset for English and applied for G2P task.

Qamhan et al. [55],investigated the efficacy of employing spectrograms as the acoustic feature alongside DPFs constructed using two distinct deep learning methodologies: the deep belief network (DBN) and the convolutional recurrent neural network (CRNN). The research focuses on the application of this approach to Modern Standard Arabic (MSA). The proposed acoustic-to-phonetic converter incorporates multi-label modeling. They used KAPD dataset for Arabic and applied for ASR task.

Alashban and Alotaibi [56],examined, identified, and evaluated the pronunciation of specific terms in languages similar to Arabic. It introduces a deep learning framework, namely the Bidirectional Long Short Term Memory (BLSTM), for distinguishing and categorizing spoken Arabic and its similar languages sourced from the Mozilla speech corpus. They used Mozilla speech corpus for Arabic and applied for ASR task.

Yao and Zweig [57],investigated how Sequence-to-sequence translation techniques can be applied to the grapheme-to-phoneme task, which presents distinct challenges. In this scenario, both the input and output vocabularies are limited, and simple n-gram models perform effectively, with recognition given only for completely accurate outputs. They used CMU dataset for English and applied for G2P task.

Zia et al. [58],introduced a grapheme-to-phoneme conversion tool tailored for Urdu. It creates a pronunciation lexicon compatible with speech recognition systems by processing a list of Urdu words. Utilizing an LSTM-based model trained on a curated lexicon comprising a huge amount of words, the tool predicts word pronunciations. They used Lexicon dataset for Urdu and applied for ASR task.

Appendix A show the summary of all works that were explored in this section.

9. Techniques of phonetic mapping

Phonetic mapping, in our methodology, is a method, algorithm or technique that used to transform phonetic representation from one language to another. There are many techniques used for phonetic mapping under many categories. In this research, the techniques are classified into seven main categories and five sub-categories which are; (i) Rules-based Techniques, (ii) Alignment Techniques, (iii) Phonetic Distance Metric Techniques, (iv) Statistical and Probabilistic Techniques, (v) Classical Machine Learning Techniques, (vi) Neural Network and Deep Learning Techniques such as(ANN, CNN, RNN, LSTM, Cognate Transformer), and (vii) Hybrid Techniques. Appendix B show the summary of all works that will be explored in next sections.

9.1. Rules-based techniques

Rule-based approaches offer a structured framework for analyzing and describing these processes, often based on principles of phonology and phonetics. By identifying patterns and regularities in sound alterations, linguists and speech researchers can better understand the underlying mechanisms driving language production and perception.

(Alshuwaier & Areshey, 2011) proposed an English-Arabic transliteration model using phonetic rules and pronunciation. They used CMU dataset for Arabic language and applied for Transliteration task.

Meng et al. [59],presented a named entity transliteration technique for English-Chinese cross-lingual spoken document retrieval. They used TDT Collection dataset of English-Chinese languages.

Yousef [60],presented a framework for cross languages name mapping between English and Arabic that was proposed and implemented. They used Higher Education sector in Egypt dataset for Arabic-English languages and applied for NER task.

Rao [61],used rules-based techniques for Phonetic matching between Hindi and Marathi or in cross-language as part of information retrieval system. They used Private dataset dataset for Hindi and Marathi languages and applied for IR system.

9.2. Alignment techniques

Alignments in phonetic processes refer to the ways in which sounds, syllables, or words are positioned relative to each other during speech production. These alignments can have significant implications for phonological patterns and linguistic analysis. Understanding alignments is essential for uncovering the intricacies of speech production and phonological patterns across languages.

Blum and Lis [62],introduced a technique for automatically deducing correspondence patterns from phonetically aligned cognate sets. It outlined a method for refining phonetic alignments in comparative linguistics before deriving correspondence patterns. They used Lexibank collection 2022 dataset for Multi Languages and applied for Cognate task.

Karimi et al. [63],presented a novel algorithm designed for converting English text into Persian through transliteration. It employs a specialized alignment algorithm tailored specifically for transliteration purposes. They used corpora of word pairs in English and Persian dataset Persian-English and applied for Transliteration.

Kocharov [64],tried for enhancing the Levenshtein algorithm to improve the alignment of phoneme transcriptions from spoken utterances. It aims to find the best alignment between sequences of phonetic symbols. They used CORPRES dataset for Multi Languages and applied for Cognate task.

Kondrak [65],introduced an algorithm that merges various methods designed for comparing sequences with a scoring system for calculating phonetic likeness using multi-valued attributes. They used data set of 82 cognates for Multi Languages and applied for Cognate task.

List [66],used a concept where the core of his approach is to integrate various methods used in historical linguistics and evolutionary biology to create a novel framework that accurately reflects the key elements of the comparative method. He used linguistics dataset for Multi Languages and applied for Cognate task.

List et al. [67], introduced an innovative framework that merges cutting-edge methods for automated sequence comparison with original techniques for phonetic alignment analysis and the detection of sound correspondence patterns. This integration enables the supervised reconstruction of word forms in ancestral languages. They used CLDF dataset of Multi Languages and applied for Cognate task.

List et al. [68],outlined how cognate reflexes were predicted using multilingual wordlists, detailing both the methodology and findings. The training and unexpected data utilized standardized wordlists spanning various language families. They used SIGTYP-2022 dataset of Multi Languages and applied for Cognate task.

Sofroniev and Coltekin [69], investigated various data-derived vector depictions of IPA-encoded sound segments, aiming to align sound sequences effectively. It evaluates different representations by measuring their alignment accuracy, particularly within the domain of computational historical linguistics. They used BDPA dataset of Multi Languages and applied for G2P task.

9.3. Phonetic distance metric techniques

Phonetic distance is the distance between two phonetic representations, i. e how they are similar or different it can be Edit Distance (Levenshtein Distance), Dynamic Time Warping (DTW)or Weighted Edit Distance. In this section some of the works under this category will be presented.

Ahmed et al. [70], presented a Phonetic Edit Distance (PED), which calculates the replacement cost in the inner loop of edit distance computation by comparing the articulatory properties of letters from two distinct languages. It compares the letters in a sophisticated manner based on their articulatory qualities as stated in the International Phonetic Alphabet (IPA). They used UD corpora dataset for Multi Languages and applied for ASR system.

Al-Dhlan [71],explored the impact of word cognates systems. The model employed can automatically detect cognates in either Arabic or English and display them in a comprehensive list. It automatically identifies cognates in both languages using one of two functions: levenshtein( ) or similar_text( ). They used private dataset for Arabic-English languages and applied for Cognate task.

Dautriche et al. [72],developed a method to evaluate lexicons compared to phonotactically-controlled baselines, acting as benchmarks to gauge the anticipated arrangement of wordforms solely based on phonotactics. The results, supported by diverse metrics such as minimal pairs, average Levenshtein distance, and various network properties, provide evidence of the effectiveness of this approach. Dataset for CELEX pronunciations dataset for Multi Languages and applied for Cognate task.

Droppo and Acero [73],Proposed adding a third metric to evaluate the similarity between a word's pronunciation in its transcription and the output of a less constrained phonetic recognition system. Reveals the learnability of phonetic string edit distance using collected data and stresses the importance of incorporating context into the model for peak performance. They used private dataset for Multi Languages and applied for ASR system.

Eden [74], in his thesis, he explored three distinct methods for assessing cross-language phonological distance. Firstly, it examines the utilization of phonological typological parameters. Secondly, it evaluates the cross-entropy of phonologically transcribed texts. Lastly, it assesses the phonetic similarity of non-word pronunciations by speakers from diverse language backgrounds. He used Private dataset for Multi Languages and applied for G2P task.

9.4. Statistical and probabilistic techniques

The most well known statistical probabilistic models, that used in many NLP applications and tasks, are Hidden Markov Models (HMMs), Entropy, and Gaussian Mixture Models (GMMs). Hidden Markov Models (HMMs) have shown to be extremely useful tools for phonetic processing. In this concept, phonetic processing is defined as the analysis and recognition of speech sounds. HMMs are especially well-suited to this task because they can represent the dynamic aspect of speech production, where phonemes flow easily from one to the next. HMMs have been effectively used for a variety of phonetic applications, including voice recognition, speaker identification, and speech synthesis. Their capacity to simulate the probabilistic correlations between phonetic units makes them effective instruments for comprehending and analyzing speech signals.

Beinborn et al. [75],introduced a method using character-based machine translation to automatically generate cognates. It shows that the approach can identify production patterns from noisy data and is effective across different language pairs and alphabets, with successful outcomes observed in tests including English-Russian, English-Greek, and English−Farsi pairs. They used MT engine Moses dataset for Multi Languages and applied for Cognate task.

MacSween and Caines [76], presented an automated system for detecting cognates, treating it as an inference problem for a statistical model. The model incorporates both observed data (word pairs that may be related) and hidden variables representing the cognate status of these pairs, as well as global parameters defining sound correspondences across languages. They used LexStat dataset for Multi Languages and applied for Cognate task.

Rama and List [77],presented an automatic approach to tree inference from large-scale datasets. Two new methods for cognate detection and a fast way for Bayesian inference of phylogeny are introduced. Experiments show that these methods can analyze the time it takes to constitute language families in a matter of minutes, which is a big step from its previous reputation of time-consuming methods. They used list dataset for Multi Languages and applied for Cognate task.

Bhargava and Kondrak [78],recommended utilizing Hidden Markov Models (HMMs) for word-related tasks and assesses their efficacy in multiple cognate alignment and cognate set matching. It discovers that HMMs perform well in both tasks, outperforming average and minimum edit distance techniques in cognate set matching. They used private dataset for Multi Languages and applied for Cognate task.

Shao [79], in his thesis, he addressed the challenge of transliterating names from various source languages into Chinese by constructing a Hidden Markov Model (HMM)-based machine transliteration system. It explores four different approaches to enhance the baseline system and evaluates these modified systems across 14 language. He used NEWS 2018 dataset for Multi Languages and applied for Transliteration task.

9.5. Classical machine learning techniques

There are wide varieties for Machine learning algorithms, that can be used for many application and task including phonetic mapping, such as support vector machines, KNN, decision trees, and many others. The problem of phonetic mapping can be taken as classification problem, and in such case, they are often employed to classify phonetic units, such as phonemes from one language, into predefined categories in another language. These algorithms learn patterns from phonetic-labeled data and subsequently generalize their knowledge to do mapping accurately.

Lam et al. [80],introduced a fresh named entity matching model that incorporates semantic and phonetic indicators. The study explores three learning algorithms to derive similarity data from training examples for basic phoneme units. They used private dataset formulti Languages and applied for NER task.

Parrish [81],introduced an innovative technique for capturing sound similarities that are challenging to represent solely through orthographic or phonemic data. It demonstrates that similarity measurements within the resulting vector space effectively predict phonetic similarities, performing well on established tests. They used CMU dataset for Multi Languages and applied for Cognate task.

Jager [82],introduced a novel method employing support vector machines to integrate various cutting-edge techniques for phonetic alignment and cognate detection iCorresponding author one cohesive system. Through training and assessing this approach on a diverse range of gold-standard data, it demonstrates its superiority over current methodologies. They used ABVD dataset for Multi Languages and applied for Cognate task.

Mani et al. [83],investigated two trainable approaches centered on pronunciation analysis. The first method, cross-lingual, employs an automated name-matching system with criteria based on human-conducted phonological comparisons of two languages. The second method is monolingual, using automatic comparisons of phonological representations within each language pair. They used Special Dataset dataset for Multi Languages and applied for ASR system.

Freeman et al. [84],introduced a method to address the challenge of matching personal names written in English with their counterparts in Arabic script. Traditional string comparison methods struggle with this task due to inconsistent transliteration practices in both languages and the absence of short vowels in Arabic script. They used private dataset for Arabic-English languages and applied for Transliteration task.

Davis [85],introduced a system designed to transliterate text between two Persian dialects that employ incompatible writing systems. Additionally, the system functions as a means to enable the exchange of computational linguistic resources between the two languages. This is particularly important due to the unequal distribution of resources between Tajik and Farsi. They used private dataset for Multi Languages and applied for Transliteration task.

Loots and Niesler [86],analyzed pronunciations in American, British, and South African English dictionaries. It conducted three main analyses: evaluating the accuracy of grapheme-to-phoneme conversion across accents, comparing pronunciations between accents, and applying decision trees to convert pronunciations between accents. They used CMU dataset for Multi Languages and applied for G2P task.

Knight and Graehl [87],presented and assesses a technique for conducting reverse transliterations through automated means. The method employs a generative model, encompassing multiple stages within the transliteration procedure. They used CMU dataset for English-Japanese languages and applied for Transliteration task.

Stalls and Knight [88],tackled the inverse challenge: retrieving the original Roman script of a foreign name or loanword from Arabic text. Arabic poses unique hurdles due to its lack of written vowels and phonetic contextual influences. They used Private dataset for Arabic-English and applied for Transliteration.

9.6. Neural network and deep learning techniques

The modern approaches and application used neural networks and deep learning utilizing from the tremendous development in processing speed. This section provides a detail of the allocation of many papers based on ANN and deep learning techniques and their respective fields (such as ANN, CNN, RNN, LSTM, Phonetic Embedding, and Cognate Transformer). This section discusses the utilized model and presents the findings.

9.6.1. Artificial Neural Network

Artificial Neural Networks (ANNs) integrated with phonetic processes represent a fascinating fusion of machine learning and linguistics. This integration holds immense potential in diverse applications, from voice-controlled assistants to language translation systems, revolutionizing how humans interact with technology and enhancing accessibility for individuals with speech-related challenges.

Fourrier [89], in his dissertation, presented methodically study the applicability of machine translation inspired neural networks to historical word prediction, relying on the surface similarity of both tasks. They used PLexGen dataset for Multi Languages and applied for Cognate task.

Li and MacWhinney [90],introduced a novel phonological pattern generator named PatPho, enabling connectionist models to generate precise phonological representations of English vocabulary. They used CELEX database for Multi Languages and applied for ASR system.

Marjou [91],discussed Allosaurus, which investigates how phones and phonemes interact in multilingual audio modeling. It shows a significant 17% improvement in phone recognition accuracy for previously unknown languages. They used allosaurus dataset for Multi Languages and applied for ASR system.

Libovicky and Fraser [8],developed a trainable neural model for calculating edit distance between strings, which can be used for applications like string comparison and transformation. It was evaluated on a variety of tasks, including cognate identification, transliteration, and grapheme-phoneme conversion. They used CMU dataset for Multi Languages and applied for all the mentioned tasks.

Marjou [92],used an Artificial Neural Network (ANN) model called Orthographic Transparency

Estimation with an ANN (OTEANN) to determine the predictability of speech from printed words. He used Special Dataset dataset for Multi Languages and applied for G2P task.

9.6.2. CNN

Convolutional Neural Networks are a type of deep learning algorithm known for extracting hierarchical features from input data using convolutional layers. For the most part, CNNs allow state-of-the-art limits in phonetic processing and make them possible for an advance in the understanding and interaction of speech. CNNs generally play a very big role in phoneme recognition, speaker identification, and speech synthesis.

Goswami et al. [93],proposed a novel language-agnostic weakly supervised deep cognate detection framework for under-resourced languages using morphological knowledge from closely related languages. They used UniMorph datasets dataset for Multi Languages and applied for Cognate task.

Rama [94],presented phoneme level Siamese convolutional networks for the task of pair-wise cognate identification. It represent a word as a two-dimensional matrix and employ a siamese convolutional network for learning deep representations. He used Lexical database dataset for Multi Languages and applied for Cognate task.

Yolchuyeva [95], in his dissertation, introduced an innovative sequence-to-sequence (seq2seq) architecture centered on Convolutional Neural Networks (CNNs). It incorporates an end-to-end CNN for grapheme-to-phoneme (G2P) conversion, featuring residual connections. Furthermore, a model was created that used a convolutional neural network as the encoder, with and without residual connections, and a Bidirectional Long Short-Term Memory (Bi-LSTM) as the decoder. He used CMU dataset for Multi Languages and applied for G2P task.

9.6.3. RNN

Recurrent Neural Networks (RNNs) are at the core of phonetic processing, since they untangle delicate details in speech patterns that no other tool can unravel. The architecture design of RNNs, being implemented with feedback loops to allow the infusion of past information into the present, is very powerful in capturing the temporal dependencies in spoken language. RNNs find a place in phonetics, especially for speech recognition; they deal with dynamics of speech signals in a smooth way.

Cheng et al. [96],integrated phonetic data into neural networks through two methods: firstly, by generating additional data via forward and back-translation, utilizing a phonetic approach; secondly, by pre-training models on a phonetic task prior to transliteration learning. They used NEWS 2018 dataset for English-Chinese languages and applied for Transliteration task.

Hartmann [97],presented a computational approach using deep neural networks to forecast phonetic attributes of historical sounds when their precise qualities are uncertain. Additionally, it aims to assess the consistency of reconstructed historical phonetic features. They used Wiktionary dataset for Multi Languages and applied for Cognate task.

He et al. [98],examined the impact of integrating phonetic attributes into English-to-Chinese transliteration within the multi-task learning (MTL) framework. The study introduces a new dataset for English-to-Chinese transliteration and proposes a novel evaluation metric that accounts for various potential transliterations of a given source name. They used NEWS 2018 dataset for English-Chinese languages and applied for Transliteration task.

Huda et al. [99],introduced a cost-effective approach for extracting articulatory features (AFs). The method involves utilizing two multilayer neural networks (MLNs). The initial MLN, termed MLNLF-DPF, translates local features (LFs) from an input speech signal into discrete AFs. Subsequently, the second MLN, referred to as MLNDyn, regulates the dynamics of the AFs generated by MLNLF-DPF. They used the Acoustic Society of Japan (ASJ) Continuous Speech Database for Multi Languages and applied for ASR system.

Younes et al. [100],concentrated on transliterating the Tunisian dialect. It introduces a Sequence-to-Sequence approach rooted in deep learning to transliterate user-generated Tunisian dialect text on social media platforms. This method handles both Latin to Arabic and Arabic to Latin transliterations at the word level. They used Special Dataset for Multi Languages and applied for Transliteration task.

Yusuf et al. [101],presented an enhanced neural ASR-independent keyword search model that attains competitive performance while upholding an efficient and streamlined pipeline. This is achieved by multilingual pre-training and a thorough model examination. In this approach, both inquiries and documents are encoded using a pair of recurrent neural network encoders, which are then concatenated using a dot-product. They used IARPA Babel corpus for Multi Languages and applied for ASR system.

9.6.4. LSTM

Long Short-Term Memory (LSTM) networks provide a powerful tool for phonetic processing because they have an elaborate structure to model sequences with complicated dependencies through time. In the domain of speech recognition and phonetic analysis, LSTM networks hold their ground in the description of the subtle temporal patterns which are part of spoken language. Through their capacity to retain information and allow its selective updating over extended sequences, LSTM architectures facilitate the description of phonetic intonation, rhythm, and articulatory dynamics.

Datta et al. [102],introduced a training approach for a grapheme-based speech recognizer, which can be trained solely using data. Utilizing LSTM networks and cross-entropy loss during training, the graphene output acoustic models investigated here are highly applicable in real-world scenarios. They can be decoded using traditional ASR stack components like language models and FST decoders. They used IARPA BABEL speech corpus for Multi Languages and applied for ASR system.

Mu et al. [103],introduced a DDNN (double-layer deep neural network) framework comprising speech-text alignment and recognition modules. The first layer of the alignment module employs a novel Viterbi algorithm approach to optimize speech-text alignment. The second layer pioneers the utilization of deep learning networks in the encoding segment of Attention for speech evaluation and scoring. They used special corpus for Japanese and applied for ASR system.

Rosca and Breuel [104],illustrated that neural sequence-to-sequence models achieve cutting-edge or near-cutting-edge performance on established datasets. In a bid to enhance the accessibility of machine transliteration, the paper releases a new Arabic to English transliteration dataset and our trained models as open source. They used google-transliteration dataset for multi Languages and applied for Transliteration task.

Sokolov [105],presented a novel neural G2Pmodel trained comprehensively. It utilizes a combination of universal symbol sets inspired by Latin alphabets and shared feature representations across various languages. This model proves valuable in scenarios where languages have scarce resources or when dealing with code-switching and foreign terminology. They used CMU dataset for Multi Languages and applied for G2P task.

Tian et al. [106],suggested an improved neural transliteration model incorporating memory mechanisms to utilize phonemic details from both English and Arabic. More precisely, within the memory component, the phonemic representations linked with individual English letters are assessed and utilized selectively to direct the transliteration from English to Arabic. They used EANames corpus with English and Arabic names and applied for Transliteration task.

9.6.5. Cognate Transformer

The Cognate Transformer is the next revolutionary NLP technology, building further on the fusion of deep learning and principles of cognitive science. Unlike classical transformer structures, the Cognate Transformer relies on cognates—words that have similar meanings in different languages—to increase cross-lingual comprehension and correctness of translations. The Cognate Transformer is a new frontier in NLP promising a future where the whole world is connected through language, breaking down the way we interact across linguistic borders.

Akavarapu and Bhattacharya [107],introduceed a transformer-based paradigm influenced by computational biology for automating cognate discovery. It demonstrates that given adequate supervision, this method outperforms existing strategies and improves with further supervision, emphasizing the importance of using labeled data. They used private dataset for Multi Languages and applied for Cognate task.

Mahesh Akavarapu and Bhattacharya [108], described the adaption of the MSA Transformer, a protein language model, to the problem of automated phonological reconstruction. The MSA Transformer is trained on multiple sequence alignments, therefore it can handle aligned cognate words. This updated model is known as the Cognate Transformer. They used private dataset for Multi Languages and applied for Cognate task.

Dehak [109],explored the universality of certain representations and the enhancement of individual phonetic units within a multilingual context. To achieve this goal, a varied selection of phonetically diverse languages is chosen, and various experiments, including monolingual, multilingual, and cross-lingual (zero-shot) assessments, are conducted. He used IARPA BABEL speech corpus for Multi Languages and applied for ASR system.

9.7. Hybrid techniques

Hybrid techniques can be made by various method such as Combining Methods or Multimodal Techniques. In this section some of the hybrid models will be explored.

Kanojia [110], in his thesis, investigated two key challenges: recognizing cognates and identifying false friends. It focuses on distributional semantics and emphasizes the significance of shared vocabulary among closely related languages. Cognates refer to variations of the same lexical form across different languages. He used private dataset for Multi Languages and applied for Cognate task.

Kondrak [111],suggested techniques to identify and measure three key aspects of cognates: recurring sound patterns, phonetic resemblances, and semantic connections. The aim is to detect cognates and their similarities from word lists of language pairs known to be related. He used Algonquian of Multi Languages and applied for Cognate task.

Farooq and Hain [112], presented a novel method, Hybrid DNN-HMM, to investigate cross-lingual acoustic-phonetic similarities. It involves comparing posterior distributions from various monolingual acoustic models with a target speech signal, achieved through training deep neural networks as mapping networks to enable direct comparison of these distributions. They used Multilingual LibriSpeech dataset and applied for ASR system.

Farooq and Hain [113],introduced a new method for combining multilingual models. This method involves training a model to understand similarities between languages' acoustic-phonetic features. Traditionally, hybrid DNN-HMM ASR systems have relied on manually crafted lexicons, but they aim to eliminate this need by expanding the idea of trainable cross-lingual mappings for end-to-end speech recognition. They used IARPA BABEL speech corpus and applied for ASR system.

10. Evaluation metrics

There are many evaluation metrics can be used for the evaluation of Phonetic similarity technique such as Phonemic Error Rate (PER), Word Error Rate (WER), Phoneme Precision and Recall Mean Similarity Score for phonetic transcripts and many others. In this section two of these methods, Phonemic Error Rate (PER) and Phoneme Precision and Recall, will be explained.

10.1. Phonemic Error Rate (PER)

Phonemic Error Rate (PER) is a metric used to evaluate the accuracy of phonemic transcriptions and can be used for phonetic mapping in cross-language. It measures the difference between a reference phonemic transcription and the predicted phonemic transcription. PER is defined as the sum of the number of substitutions (S), insertions (I), and deletions (D) between the predicted and the referenced transcription divided by the total number of phonemes in the reference transcription (N). If we have a predicted phonetic transcripts (PPT) and referenced phonetic transcripts (RPT) then the PER will be as shown the following equation:

$$PER = \frac{S + D + I}{N}$$

Where:

S: Number of phonemes that substituted to the PPT to be the RPT.
D: Number of phonemes that deleted to the PPT to be the RPT.
I: Number of phonemes that inserted to the PPT to be the RPT.
N: Total number of phonemes in the RPT.

10.2. Phoneme Precision and recall

F-measure, Precision and Recallit is another famose evaluation metrics for many applications and task. Phoneme Precision and Recall are used to evaluate the performance of phonetic alignment and transcription systems. Precision and recall are particularly useful for understanding both the correctness and completeness of phonetic alignments.

Precision measures the proportion of correctly identified phonemes out of all phonemes that the system identified while Recall measures the proportion of correctly identified phonemes out of all phonemes in the reference transcription. The F1 score is the harmonic mean of precision and recall. If we deals with phonetic mapping as classification problem and we have a predicted phonetic transcripts (PPT) by our system to be compared to the referenced phonetic transcripts (RPT) then the we have three terms in this scenarioss; True Positives (TP), False Positives (FP), False Negatives (FN) and True Negatives (TN) where:

True Positives (TP): Phonemes correctly identified by the system that match the reference.

False Positives (FP): Phonemes identified by the system that do not match the reference (incorrect phonemes).

False Negatives (FN): Phonemes present in the reference transcription that the system failed to identify.

Recall, Precision and F1 score can be estimated as following:

$$P = \frac{Tp}{Tp + Fp}$$

$$R = \frac{Tp}{Tp + Fn}$$

$$F1\ measure = \frac{2\,PR}{P+R}$$

11. Discussion

There is certainly something very interesting in exploring phonetic congruence at the level of sounds themselves or at the level of integrated phonetic units among different languages. Also, discovering the graphene to phoneme and phonetic mapping techniques will help us to develop phonetic processing system with standardization among the world languages.

Based on a review of previous studies, it can be identified that phonetic similarity and related research areas cut across statistics and data science; therefore, these studies have adopted such methodologies and applied them to produce the best results. We used seven main categories under Rules-based Techniques, Alignment Techniques, Phonetic Distance Metric Techniques, Statistical and Probabilistic Techniques, Classical Machine Learning Techniques, Neural Network and Deep Learning Techniques, and Hybrid Techniques. Neural Network and Deep Learning Techniques has 5 sub-categories; ANN, CNN, RNN, LSTM, and Cognate Transformer. Also, it clear that the pre-2010 decade was almost entirely based on research activities focused on letter-to-G2P sound conversion, transliteration, and limited cognate investigation, driven by statistical methodologies and rule-based methodologies driven by experts. The decade of 2010–2020 saw a marked shift that was driven by technology and the rise of artificial intelligence. The earlier methodologies were revised with an infusion of AI, leading to breakthrough advances, especially in the field of G2P applications and new tools, such as multilingual voice search for IR and NER.

Following 2020, study focused on producing sounds using phonetic similarity between the sounds of extinct or ancient languages while utilizing deep learning techniques and modern transformers like Cognate Transformer. Phonetic similarity is becoming into a dynamic and scalable field. It is feasible to explore further into phonetic similarities between languages using AI algorithms and relevant linguistic information, providing insights into the dynamics and evolutions of languages. With AI poised to transform linguistics research and technology, there appear to be plethora of intriguing opportunities in the fields of intercultural communication, language acquisition and interdisciplinary collaboration.

12. Conclusion

The phonetic similarity is one of the basic elements in the field of information technology, on which different applications from search engines to speech recognition systems are designed. The collection and distribution of such findings will be highly important to scholars concerned with phonetic similarity and its related methodologies. As technology marches forward, the role of phonetic similarity in shaping advanced solutions of IT is going to have a major increase in the years to come, ushering in the era when linguistic subtleties will integrate seamlessly into digital transactions.

Disclosure of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix

Appendix A: list of Techniques of producing phonetic representation

No.	Papers	Dataset	language	Application
1	[26]	TRmorph	Turkish	G2P
2	[2]	Umm Alqura list	Arabic	G2P
3	[27]	Algiers dialect corpus	Arabic	G2P
4	[28]	Author's names	Arabic	NER
5	[29]	Special Dataset	English	G2P
6	[30]	Six corpora have been used in this work	Arabic	ASR
7	[31]	Speakers	English	Cognate
8	[32]	CMU	English	G2P
9	[33]	Private	Thai	ASR
10	[34]	TIMIT corpus	English	ASR
11	[35]	CMU	English	G2P
12	[36]	KAPD	Arabic	G2P
13	[37]	Persian Linguistic Database	Persian	G2P
14	[38]	Private Dataset	Arabic	Spelling Correction
15	[39]	Special Dataset	Arabic	Spelling Correction
16	[40]	Speakers	Arabic	Spelling Correction
17	[41]	Private Dataset	English	G2P
18	[42]	KAPD	Arabic	ASR
19	[43]	CMU	English	G2P
20	[44]	50 k Indonesian words	Indonesian	G2P
21	[45]	Private Dataset	Arabic	NER
22	[46]	Special Dataset	Kurdish	G2P
23	[47]	CMU	English	NER
24	[48]	database was developed for Arabic speaker	Arabic	Spelling Correction
25	[49]	Audio was 85 individual from 42 males, 43 females.	Arabic	Spelling Correction
26	[50]	Special Dataset	Arabic	Spelling Correction
27	[51]	audio samples of the Arabic alphabet	Arabic	Spelling Correction
28	[52]	Special Dataset	Arabic	G2P
29	[53]	FarsDat	Persian	G2P
30	[54]	CMU	English	G2P
31	[55]	KAPD	Arabic	ASR
32	[56]	Mozilla speech corpus	Arabic	ASR
33	[57]	CMU	English	G2P
34	[58]	Lexicon	Urdu	ASR

Appendix B: list of Techniques of Estimation Phonetic Similarity with summary

No.	Papers	dataset	language	Application/Task	Technique category
1	[114]	CMU	Arabic	Transliteration	Rules-based
2	[59]	TDT Collection	English-Chinese	Transliteration	Rules-based
3	[60]	Higher Education sector in Egypt	Arabic-English	NER	Rules-based
4	[100]	Private dataset	Hindi and Marathi	IR system	Rules-based
5	[62]	Lexibank collection 2022	Multi Languages	Cognate	Alignment
6	[63]	corpora of word pairs in English and Persian	Persian-English	Transliteration	Alignment
7	[64]	CORPRES	Multi Languages	Cognate	Alignment
8	[65]	data set of 82 cognates	Multi Languages	Cognate	Alignment
9	[66]	Linguistics	Multi Languages	Cognate	Alignment
10	[67]	CLDF	Multi Languages	Cognate	Alignment
11	[68]	SIGTYP 2022	Multi Languages	Cognate	Alignment
12	[69]	BDPA	Multi Languages	G2P	Alignment
13	[75]	MT engine Moses	Multi Languages	Cognate	Statistical
14	[109]	IARPA BABEL speech corpus	Multi Languages	ASR	Transformer
15	[110]	Private dataset	Multi Languages	Cognate	Hybrid
16	[111]	Algonquian	Multi Languages	Cognate	Hybrid
17	[76]	LexStat.	Multi Languages	Cognate	HMM
18	[77]	list	Multi Languages	Cognate	Bayesian inference
19	[70]	UD corpora	Multi Languages	ASR	Phonetic Edit Distance (PED
20	[71]	Special Dataset	Arabic-English	Cognate	Levenshtein or similar text.
21	[72]	CELEX pronunciations	Multi Languages	Cognate	Levenshtein distance.
22	[73]	Private dataset	Multi Languages	ASR	phonetic string edit distance
23	[74]	Private dataset	Multi Languages	G2P	cross-language phonological distance cross-entropy
24	[80]	Private dataset	Multi Languages	NER	ML
25	[81]	CMU	Multi Languages	Cognate	VSM + statistical
26	[78]	Private dataset	Multi Languages	Cognate	HMM
27	[79]	NEWS 2018	Multi Languages	Transliteration	Hidden Markov Model
28	[112]	Multilingual LibriSpeech	Multi Languages	ASR	Hybrid
29	[113]	IARPA BABEL speech corpus	Multi Languages	ASR	Hybrid
30	[82]	ABVD	Multi Languages	Cognate	ML – SVM
31	[83]	Special Dataset	Multi Languages	ASR	ML + Alignment
32	[84]	Private dataset	Arabic-English	Transliteration	ML
33	[85]	Private dataset	Multi Languages	Transliteration	ML
34	[86]	CMU	Multi Languages	G2P	ML
35	[87]	CMU	English-Japanese	Transliteration	ML (generative model)
36	[88]	Private dataset	Arabic-English	Transliteration	ML
37	[89]	PLexGen	Multi Languages	Cognate	ANN
38	[90]	CELEX database	Multi Languages	ASR	ANN
39	[91]	allosaurus	Multi Languages	ASR	ANN
40	[8]	CMU	Multi Languages	All Tasks	ANN
41	[92]	Special Dataset	Multi Languages	G2P	ANN
42	[93]	UniMorph datasets	Multi Languages	Cognate	CNN
43	[94]	Lexical database	Multi Languages	Cognate	CNN
44	[95]	CMU	Multi Languages	G2P	CNN
45	[96]	NEWS 2018	English-Chinese	Transliteration	RNN
46	[97]	Wiktionary	Multi Languages	Cognate	RNN
47	[98]	NEWS 2018	English-Chinese	Transliteration	RNN
48	[99]	the Acoustic Society of Japan (ASJ) Continuous Speech Database	Multi Languages	ASR	RNN
49	[100]	Special Dataset	Multi Languages	Transliteration	RNN
50	[101]	IARPA Babel corpus	Multi Languages	ASR	RNN
51	[102]	IARPA BABEL speech corpus	Multi Languages	ASR	LSTM
52	[103]	special corpus	Japanese	ASR	LSTM

(*continued on next page*)

(continued)

No.	Papers	dataset	language	Application/Task	Technique category
53	[104]	google-transliteration	multi Languages	Transliteration	LSTM
54	[105]	CMU	Multi Languages	G2P	LSTM
55	[106]	collect a corpus EANames with English and Arabic names	Arabic-English	Transliteration	LSTM
56	[107]	Private dataset	Multi Languages	Cognate	Transformer
57	[108]	Private dataset	Multi Languages	Cognate	Transformer

References

[1] Chomsky N. Language and mind. Cambridge University Press; 2006.

[2] El-Imam YA. Phonetization of Arabic: rules and algorithms. Comput Speech Lang 2004;18:339–73.

[3] AMaH Mahdi, Sabah Asaad, editors. The current state of linked data-based recommender systems. 2021 2nd Information Technology to Enhance e-learning and Other Application (IT-ELA). IEEE; 2021.

[4] Ladefoged P, editor. The measurement of phonetic similarity. International Conference on Computational Linguistics COLING 1969; 1969. Preprint No 57.

[5] Weide R. The CMU pronunciation dictionary, release 0.6. Pittsburgh, PA: Carnegie Mellon University; 1998.

[6] Gayar NaS El, Yee Ching. Computational linguistics, speech and image processing for Arabic language. World Scientific; 2018.

[7] Zouhar Ve, Chang K, Cui C, Carlson N, Robinson N, Sachan M, et al. Pwesuite: Phonetic word embeddings and tasks they facilitate. arXiv preprint arXiv:230402541. 2023.

[8] Libovicky J, Fraser A. Neural string edit distance. arXiv preprint arXiv:210408388 2022.

[9] Halpern J, editor. CJKI Arabic romanization system. Abu Dhabi: the international Symposium on Arabic transliteration standard: challenges and solutions; 2009.

[10] Bakar JA, Omar K, Nasrudin MF, Murah MZ, Ahmad CW, editors. Implementation of Buckwalter transliteration to Malay corpora. 2013 13th international conference on intellient systems design and applications; 2013.

[11] Toma S-A, Stan A, Pura M-L, B'arsan T, editors. MaRePhoR—an open access machine-readable phonetic dictionary for Romanian. 2017 international conference on speech technology and human-Computer Dialogue (SpeD); 2017.

[12] Out-of-the-box universal romanization tool uroman. In: Hermjakob U, May J, Knight K, editors. Proceedings of ACL 2018, system demonstrations; 2018.

[13] Epitran: Precision G2P for many languages. In: Mortensen DR, Dalmia S, Littell P, editors. Proceedings of the Eleventh international conference on language resources and evaluation (LREC 2018); 2018.

[14] JLLaLFEAaMEGaYL-SaSMaAWaADMaK Gorman, editor. Massively multilingual pronunciation modeling with WikiPron. International conference on language resources and evaluation; 2020.

[15] Chootrakool P, Wuttiwiwatchai C, Kosawat K. A large pronunciation dictionary for Thai speech processing. Proc of ASIALEX 2009.

[16] Panphon: A resource for mapping IPA segments to articulatory feature vectors. In: Mortensen DR, Littell P, Bharadwaj A, Goyal K, Dyer C, Levin L, editors. Proceedings of COLING 2016, the 26th international conference on computational linguistics. Technical Papers; 2016.

[17] KaB Batsuren. G'abor and Giunchiglia, Fausto and others. Cognet: a large-scale cognate database. ACL 2019 the 57th Annual Meeting of the Association for computational linguistics: Proceedings of the Conference. Association for Computational Linguistics; 2019.

[18] Pratap V, Xu Q, Sriram A, Synnaeve G, Collobert R. Mls: A large-scale multilingual dataset for speech research. arXiv preprint arXiv:201203411 2020.

[19] Batsuren K, Bella Ga, Giunchiglia F. A large and evolving cognate database. Comput Humanit 2022:1–25.

[20] Mohammed ZR, Aliwy AH, editors. English-Arabic phonetic dataset construction. BIO Web of Conferences; 2024.

[21] Artetxe M, Labaka G, Lopez-Gazpio I, Agirre E. Uncovering divergent linguistic information in word embeddings with lessons for intrinsic and extrinsic evaluation. arXiv preprint arXiv:180902094 2018.

[22] El-Geish M. Learning joint acoustic-phonetic word embeddings. arXiv preprint arXiv:190800493 2019.

[23] Feng X, Wang L, editors. Application of Word2vec in phoneme recognition. Proceedings of the 2020 12th international conference on machine learning and computing; 2020.

[24] Kanojia D, Patel K, Bhattacharyya P, Kulkarni M, Haffari G. Utilizing wordnets for cognate detection among indian languages. arXiv preprint arXiv:211215124 2021.

[25] Sharma R, Dhawan K, Pailla B. Phonetic word embeddings. arXiv preprint arXiv:210914796 2021.

[26] Altinok D. Towards Turkish ASR: Anatomy of a rule-based Turkish g2p. arXiv preprint arXiv:160103783 2016.

[27] Harrat S, Meftouh K, Abbas M, Sma"ili K. Grapheme to phoneme conversion-an Arabic dialect case. In: The 4th international workshop on spoken language technologies for under-resourced languages (SLTU'14); 2014: spoken Language technologies for under-resourced languages; 2014.

[28] Nehar A, Bellaouar S, Ziadi D, Omar KM. Arabic Personal Name Matching: Names Written using Latin Alphabet. J Comput Sci 2021;17:776–88.

[29] Bisani M, Ney H. Joint-sequence models for grapheme-to-phoneme conversion. Speech Commun 2008;50:434–51.

[30] Sindran F. Automatic phonetic transcription of standard Arabic with applications in the NLP domain. Ph.D. Dissertation]; 2021.

[31] How to compare automatically two phonological strings: Application to intelligibility measurement in the case of atypical speech. In: Ghio A, Lalain M, Giusti L, Fredouille C, Woisard V, editors. 12th conference on language resources and evaluation (LREC 2020); 2020.

[32] Hixon B, Schneider E, Epstein SL, editors. Phonemic similarity metrics to compare pronunciation methods. INTERSPEECH; 2011.

[33] Jucksriporn C, Sornil O. Thai Phonetic Distance Using Phonetic Transcription. Int J Inf Process Manag 2014;5:133.

[34] Investigating Phoneme Similarity with Artificially Accented Speech. In: Masson M, Carson-Berndsen J, editors. Proceedings of the 20th SIGMORPHON workshop on computational research in phonetics, phonology, and Morphology; 2023.

[35] Kantor A, Hasegawa-Johnson M. Hmm-based pronunciation dictionary generation. New tools and methods for very large scale phonetics research. University of Pennsylvania; 2011.

[36] Nahar K, Al-Muhtaseb H, Al-Khatib W, Elshafei M, Alghamdi M. Arabic Phonemes Transcription using Data Driven Approach. Int Arab J Inf Technol 2015;12.

[37] Construction of a Persian Letter-to-Sound Conversion System. In: Arab MM, Azimizadeh A, editors. Proceedings of the third workshop on computational approaches to Arabic-Script-based languages (CAASL3); 2009.
[38] Khan M. Selection of discriminative features for Arabic phoneme's mispronunciation detection. Pakistan J Sci 2015; 67.
[39] Maqsood M, Habib H, Anwar S, Ma Ghazanfar. A Comparative Study of Classifier Based Mispronunciation Detection System for Confusing. Nucleus 2017;2:114−20.
[40] Nazir F, Majeed MN, Ghazanfar MA, Maqsood M. An Arabic mispronunciation detection system based on the frequency of mistakes for Asian speakers. Mehran Univ Res J Eng Technol 2021;40:279−97.
[41] Johnson K. Aligning phonetic transcriptions with their citation forms. Acoust Res Lett Online 2004;5:19−24.
[42] Ibrahim AB, Seddiq YM, Meftah AH, Alghamdi M, Selouani S-A, Qamhan MA, et al. Optimizing arabic speech distinctive phonetic features and phoneme recognition using genetic algorithm. IEEE Access 2020;8:200395−411.
[43] Bi-directional conversion between graphemes and phonemes using a joint n-gram model. In: Galescu L, Allen JF, editors. 4th ISCA Tutorial and research workshop (ITRW) on speech synthesis; 2001.
[44] Suyanto S, Sunyoto A, Ismail RN, Rachmawati E, Maharani W. Stemmer and phonotactic rules to improve n-gram tagger-based indonesian phonemicization. J King Saud Univer -Computer Infor Sci 2022;34:3807−14.
[45] Lexicon-driven approach to the recognition of Arabic named entities. In: Halpern J, editor. Proceedings of the second international conference on Arabic language resources and tools; 2009.
[46] Mahmudi A, Veisi H. Automated grapheme-to-phoneme conversion for central kurdish based on optimality theory. Comput Speech Lang 2021;70:101222.
[47] Bhagat R, Hovy EH, editors. Phonetic models for generating spelling Variants. IJCAI; 2007.
[48] Algabri M, Mathkour H, Alsulaiman M, Bencherif MA. Mispronunciation detection and diagnosis with articulatory-level feedback generation for non-native arabic speech. Mathematics 2022;10:2727.
[49] Asif A, Mukhtar H, Alqadheeb F, Ahmad HF, Alhumam A. An approach for pronunciation classification of classical Arabic phonemes using deep learning. Appl Sci 2021;12:238.
[50] Nazir F, Majeed MN, Ghazanfar MA, Maqsood M. Mispronunciation detection using deep convolutional neural network features and transfer learning-based model for Arabic phonemes. IEEE Access 2019;7:52589−608.
[51] Ziafat N, Ahmad HF, Fatima I, Zia M, Alhumam A, Rajpoot K. Correct pronunciation detection of the arabic alphabet using deep learning. Appl Sci 2021;11:2508.
[52] Barman PP, Boruah A. A RNN based Approach for next word prediction in Assamese Phonetic Transcription. Procedia Comput Sci 2018;143:117−23.
[53] Behbahani YM, Babaali B, Turdalyuly M. Persian sentences to phoneme sequences conversion based on recurrent neural networks. Open Computer Science 2016;6: 219−25.
[54] Jakobi DN. Grapheme-to-Phoneme mapping in text [Master's thesis]. University of Zurich; 2022.
[55] Qamhan MA, Alotaibi YA, Seddiq YM, Meftah AH, Selouani SA. Sequence-to-sequence acoustic-to-phonetic conversion using spectrograms and deep learning. IEEE Access 2021;9:80209−20.
[56] Alashban AA, Alotaibi YA. A Deep Learning Approach for Identifying and Discriminating Spoken Arabic Among Other Languages. IEEE Access 2023;11:11613−28.
[57] Yao K, Zweig G. Sequence-to-sequence neural net models for grapheme-to-phoneme conversion. arXiv preprint arXiv:150600196 2015.
[58] Zia HB, Raza AA, Athar A. PronouncUR: An urdu pronunciation lexicon generator. arXiv preprint arXiv: 180100409 2018.
[59] Generating phonetic cognates to handle named entities in English-Chinese cross-language spoken document retrieval. In: Meng HM, Lo W-K, Chen B, Tang K, editors. IEEE workshop on automatic speech recognition and understanding, 2001 ASRU'01; 2001.
[60] Yousef AH. Cross-language personal name mapping. arXiv preprint arXiv:14056293 2014.
[61] Rao SMCaS. Rule-Based Phonetic Matching Approach for Hindi and Marathi. Int J Res Soc Sci 2011;1:26−41.
[62] Blum F, List J-M. Trimming phonetic alignments improves the inference of sound correspondence patterns from multilingual wordlists. arXiv preprint arXiv:230317932 2023.
[63] Collapsed consonant and vowel models: New approaches for English-Persian transliteration and back-transliteration. In: Karimi S, Scholer F, Turpin A, editors. Proceedings of the 45th annual Meeting of the Association of computational linguistics; 2007.
[64] Phoneme alignment using the information on phonological processes in continuous speech. In: Kocharov D, editor. Proceedings of the Tenth international conference on language resources and evaluation (LREC'16); 2016.
[65] Kondrak G, editor. A new algorithm for the alignment of phonetic sequences. 1st Meeting of the North American Chapter of the Association for Computational Linguistics; 2000.
[66] LexStat: Automatic detection of cognates in multilingual wordlists. In: List J-M, editor. Proceedings of the EACL 2012 joint workshop of LINGVIS & UNCLH; 2012.
[67] List J-M, Forkel R, Hill NW. A new framework for fast automated phonological reconstruction using trimmed alignments and sound correspondence patterns. arXiv preprint arXiv:220404619 2022.
[68] The SIGTYP 2022 shared task on the prediction of cognate reflexes. In: List J-M, Vylomova E, Forkel R, Hill N, Cotterell R, editors. Proceedings of the 4th workshop on research in computational linguistic Typology and multilingual NLP; 2022.
[69] Phonetic vector representations for sound sequence alignment. In: Sofroniev P, Coltekin C, editors. Proceedings of the Fifteenth workshop on computational research in phonetics, phonology, and Morphology; 2018.
[70] Ahmed T, Suffian M, Khan MY, Bogliolo A. Discovering lexical similarity using articulatory feature-based phonetic edit distance. IEEE Access 2021;10:1533−44.
[71] HaaKA Al-Dhlan. Auto-Extracting Method of Cognates Words in Arabic and English Languages. Int J Adv Stud Comput Sci Eng 2017:1−13.
[72] Dautriche I, Mahowald K, Gibson E, Christophe A, Piantadosi ST. Words cluster phonetically beyond phonotactic regularities. Cognition 2017;163:128−45.
[73] Droppo J, Acero A, editors. Context dependent phonetic string edit distance for automatic speech recognition. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing; 2010.
[74] Eden SE. Measuring phonological distance between languages. 2018.
[75] Cognate production using character-based machine translation. In: Beinborn L, Zesch T, Gurevych I, editors. Proceedings of the sixth international joint conference on natural language processing; 2013.
[76] An expectation maximisation algorithm for automated cognate detection. In: MacSween R, Caines A, editors. Proceedings of the 24th Conference on computational natural language learning; 2020.
[77] Rama T, List J-M, editors. An automated framework for fast cognate detection and Bayesian phylogenetic inference in computational historical linguistics; 2019.
[78] Multiple word alignment with profile hidden Markov models. In: Bhargava A, Kondrak G, editors. Proceedings of human language technologies: the 2009 annual conference of the North American chapter of the association for computational linguistics, companion volume: student research workshop and doctoral consortium; 2009.

[79] Shao Y. Machine Transliteration of Names from Different Language Origins into Chinese. Uppsala University [Master's Thesis]; 2014.
[80] Learning phonetic similarity for matching named entity translations and mining new translations. In: Lam W, Huang R, Cheung P-S, editors. Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval; 2004.
[81] Poetic sound similarity vectors using phonetic features. In: Parrish A, editor. Proceedings of the AAAI conference on artificial intelligence and interactive digital entertainment; 2017.
[82] Using support vector machines and state-of-the-art algorithms for phonetic alignment to identify cognates in multilingual wordlists. In: Jager G, List J-M, Sofroniev P, editors. Proceedings of the 15th conference of the {E}uropean Chapter of the Association for computational linguistics: Volume 1, Long papers. Valencia, Spain: Association for Computational Linguistics; 2017.
[83] Mani I, Yeh A, Condon S. Learning to match names across languages. Multi-source. Multilingual Information Extraction and Summarization. 2013. p. 53–71.
[84] Cross linguistic name matching in English and Arabic. In: Freeman A, Condon S, Ackerman C, editors. Proceedings of the human language technology conference of the NAACL, main conference; 2006.
[85] Davis CI, editor. Tajik-farsi Persian transliteration using statistical machine translation. LREC; 2012.
[86] Loots L, Niesler T, editors. Data-driven phonetic comparison and conversion between south african, british and american English pronunciations. INTERSPEECH; 2009.
[87] Knight K, Graehl J. Machine transliteration. Comput Ling 1998;24:599–612.
[88] Stalls BG, Knight K, editors. Translating names and technical terms in Arabic text. Computational Approaches to Semitic Languages; 1998.
[89] Fourrier Ce. Neural Approaches to Historical Word Reconstruction. 2022.
[90] Li P, MacWhinney B. PatPho: A phonological pattern generator for neural networks. Behav Res Methods Instrum Comput 2002;34:408–15.
[91] Li X, Dalmia S, Li J, Lee M, Littell P, Yao J, et al., editors. Universal phone recognition with a multilingual allophone system. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP); 2020.
[92] Marjou X. OTEANN: Estimating the transparency of orthographies with an artificial neural network. arXiv preprint arXiv:191213321 2019.
[93] Goswami K, Rani P, Fransen T, McCrae JP. Weakly-supervised Deep Cognate Detection Framework for Low-Resourced Languages Using Morphological Knowledge of Closely-Related Languages. arXiv preprint arXiv:231105155 2023.
[94] Siamese convolutional networks for cognate identification. In: Rama T, editor. Proceedings of COLING 2016, the 26th international conference on computational linguistics. Technical Papers; 2016.
[95] Yolchuyeva S. Novel NLP Methods for Improved Text-To-Speech Synthesis. Ph.D. Dissertation]; 2021.
[96] Cheng S, Ding Z, Yan S. English-to-chinese transliteration with phonetic back-transliteration. arXiv preprint arXiv: 211210321 2021.
[97] Predicting historical phonetic features using deep neural networks: A case study of the phonetic system of Proto-Indo-European. In: Hartmann F, editor. Proceedings of the 1st international workshop on computational approaches to historical language change; 2019.
[98] English-to-chinese transliteration with phonetic auxiliary task. In: He Y, Cohen SB, Linguist A, editors. Proceedings of the 1st conference of the Asia-Pacific Chapter of the Association for computational linguistics and the 10th international joint conference on natural language processing. Suzhou, China: Association for Computational Linguistics; 2020.
[99] Huda MN, Hasan MM, Hassan F, Kotwal MR, Muhammad G, Rahman CM. Articulatory feature extraction for speech recognition using neural network. Int Rev Comp Software 2011;6:25–31.
[100] Younes J, Souissi E, Achour H, Ferchichi A. A sequence-to-sequence based approach for the double transliteration of Tunisian dialect. Procedia Comput Sci 2018;142:238–45.
[101] Yusuf B, Černock`y J, Saraçlar M. End-to-End Open Vocabulary Keyword Search With Multilingual Neural Representations. In: IEEE/ACM Transactions on Audio, Speech, and Language Processing; 2023.
[102] Datta A, Zhao G, Ramabhadran B, Weinstein E. LSTM Acoustic Models Learn to Align and Pronounce with Graphemes. arXiv preprint arXiv:200806121 2020.
[103] Mu D, Sun W, Xu G, Li W. Japanese Pronunciation Evaluation Based on DDNN. IEEE Access 2020;8:218644–57.
[104] Rosca M, Breuel T. Sequence-to-sequence neural network models for transliteration. arXiv preprint arXiv:161009565 2016.
[105] Sokolov A, Rohlin T, Rastrow A. Neural machine translation for multilingual grapheme-to-phoneme conversion. arXiv preprint arXiv:200614194 2020.
[106] Tian Y, Lou R, Pang X, Wang L, Jiang S, Song Y, editors. Improving English-Arabic transliteration with phonemic memories. Findings of the Association for computational linguistics: EMNLP 2022; 2022.
[107] Akavarapu V, Bhattacharya A. Automated Cognate Detection as a Supervised Link Prediction Task with Cognate Transformer. arXiv preprint arXiv:240202926 2024.
[108] Mahesh Akavarapu V, Bhattacharya A. Cognate Transformer for Automated Phonological Reconstruction and Cognate Reflex Prediction. arXiv e-prints 2023. arXiv–2310.
[109] That Sounds Familiar: An Analysis of Phonetic Representations Transfer Across Languages. In: P Dehak, editor. 21st annual conference of the international speech communication Association. Shanghai: ISCA; 2020.
[110] Kanojia D. Investigations into distributional semantics for cognate detection and phylogenetics. 2021.
[111] Kondrak G, editor. Identification of cognates and recurrent sound correspondences in word lists. Traitement Automatique des Langues, vol. 50; 2009. Num{\'e}ro 2: Langues anciennes [Ancient Languages].
[112] Farooq MU, Hain T, editors. Investigating the impact of cross-lingual acoustic-phonetic similarities on multilingual speech recognition. Interspeech 2022-23rd Annual Conference of the International Speech Communication Association; 2022.
[113] Farooq MU, Hain T. Learning cross-lingual mappings for data augmentation to improve low-resource speech recognition. arXiv preprint arXiv:230608577. 2023.
[114] Translating English names to Arabic using phonotactic rules. In: Alshuwaier F, Areshey A, editors. Proceedings of the 25th Pacific Asia conference on language, information and computation; 2011.