International Journal of Language, Literature and Culture (IJLLC)
ISSN: 2582-9823
Vol-3, Issue-5, Sep-Oct 2023
Journal DOI: [10.22161/ijllc](https://dx.doi.org/10.22161/ijllc)
Article CrossRef DOI: https://dx.doi.org/10.22161/ijllc.3.5.3
Peer-Reviewed Journal

Natural Language Understanding of Low-Resource Languages in Voice Assistants: Advancements, Challenges and Mitigation Strategies

Ashlesha V Kadam

Amazon, Seattle WA, USA.
Email: ashleshakadam1@gmail.com

Article Info

Received: 18 Aug 2023,
Received in revised form: 21 Sep 2023,
Accepted: 01 Oct 2023,
Available online: 08 Oct 2023

Keywords— Low resource languages, NLU, voice assistant, voice technology

Abstract

This paper presents an exploration of low resource languages and the specific challenges that arise in natural language understanding of these by a voice assistant. While voice assistants have made significant strides when it comes to their understanding of mainstream languages, this paper focuses on extending this understanding to low resource languages in order to maintain diversity of linguistics and also delight the customer. In this paper, the specific nuances of natural language understanding when it comes to these low resource languages has been discussed. The paper also proposes techniques to overcome some of the challenges in voice assistants understanding low resource language models. The proposed methods and future direction presented in this doc are poised to drive advancements in voice technology and promote inclusivity by ensuring that voice assistants are accessible to speakers of underrepresented languages.

I. INTRODUCTION

Voice Assistants (VAs) like Alexa, Siri and Google Assistant have become ubiquitous and well recognized household names[1]. Voice Assistants (VAs) are an application of Artificial Intelligence (AI) and Natural Language Processing (NLP) that recognize and understand human speech and respond to it in a way that is understandable by humans[2]. Voice lends a natural and intuitive interface for humans to interact with technology. This coupled with the ease with which voice assistants are accessible on mobile phones (e.g. Siri on iPhone) and smart speakers (e.g. Alexa on Echo devices) has helped with the adoption of VAs[3].

However, despite the advancements in the natural language understanding (NLU) of VAs, a significant gap persists in the support for low resource languages (LRLs). In the field of NLP, languages are classified as either being high-resource or low resource. LRLs are languages that have relatively less data available for training ML models[4]. There are limited data sets of these languages and their grammar and rules are under-described, making it challenging to train language models for high accuracy of interpretation. High resource languages (HRLs), on the other hand, are languages that have adequate data sources and are well described, making it easier to train models for interpreting these languages. Examples of LRLs are Belarusian, Pashto, Bengali and Kinyarwanda[5,6,7]. Examples of HRLs are English, Spanish, French and German[5]. This paper aims to bridge the gap between the performance of models with HRLs and with LRLs by addressing key challenges in the NLU of LRLs and proposing mitigation strategies for these challenges.

II. BACKGROUND

2.1 Voice Assistants

It is important to understand key components and steps in the end-to-end working of VAs before diving into the

LRL-specific challenges and solutions[8]. VAs can be activated using their specific wake words. For example, users of Amazon's VA say "Alexa" while those of Apple's VA say "Hey Siri" to invoke the respective VA. Post this invocation, the next task is translating the speech of the user to text tokens, i.e. the Automatic Speech Recognition (ASR)[9] stage that does speech-to-text (STT). The next stage is taking the output of the ASR stage, i.e. a string of tokens, and parsing them in order to understand the syntactic and semantic interpretation of this text string. This stage is called Natural Language Understanding (NLU), and the outcome is an understanding of the intention of the user[8]. Once the VA understands what the user is looking for, it can use multiple systems in the back-end to retrieve the right response to this query, including but not limited to the internet, a cloud platform connecting to specific servers in the back-end, an application, and more. When the response is retrieved, the VA again converts the response back into a format that is understandable to the user, like text to speech (TTS) of a response, text displayed in case of a multi-modal interface, or simply the desired action being executed (e.g. switching off the lights). See Fig. 1 for an overview of how a hypothetical voice assistant, Nova's, end to end functionality could be.

2.2 Low Resource Languages

It is also critical to understand more about LRLs before understanding how to improve their NLU performance on VAs. LRLs, as mentioned in the Introduction section, are languages that have less data sets available for training ML models. LRLs have several characteristics that make them particularly challenging to NLU practitioners. These challenges are described in the following section.

III. CHALLENGES IN NLU FOR LOW RESOURCE LANGUAGES

This section examines in detail the challenges when it comes to interpreting input in an LRL on a voice assistant.

3.1 Scarcity of Data

There is lack of adequate and robust data sets to train NLP models so they have high accuracy[10]. These languages are under-described, i.e., there are limited scientific papers that describe the grammar and rules of these languages[11]. These languages usually also have limited reliable sources like dictionaries, corpora or even books, thus making it challenging to train NLP models to be robust[12]. An extension of this is the lack of pre-trained models for LRLs.

3.2 Cross Linguistic Variability

Specific challenges arise due to diversity of linguistics and cross-linguistic variations of LRLs. LRLs might have multiple variations in the way they are written, spoken and understood. For example, different dialects, phrases, accents, colloquial adaptions, and more[10]. Cross-linguistic variations in syntax can pose a challenge for NLP researchers and practitioners, as different languages may have different word orders and grammatical structures[13].

3.3 Code Switching and Multilingualism

Code switching refers to the user alternating between two or more languages while conversing. Code switching can result in syntactic variations of a sentence since different languages might have different grammatical structures[14]. There could also be lexical variations because different languages might have different words for describing the same concept[14]. There might also be differences in interpretations due to cultural differences[15].

3.4 Commercial Viability

Zooming in specifically on the case of voice assistants, LRLs are often not commercially viable to invest in deeply as the market for these languages is usually small[16].

IV. TECHNIQUES TO OVERCOME NLU CHALLENGES WITH LOW RESOURCE LANGUAGES

This section describes the techniques that can be implemented to overcome the challenges presented in the previous section.

4.1 Leveraging Transfer Learning

Transfer learning involves re-using a pre-trained model as the starting point of a model for a new task instead of having to train a model from scratch. Transfer learning methods are popular when it comes to training a model for LRLs. Adapting pre-trained models to use for LRLs involves taking a model that is trained on an HRL and then fine-tuned to the LRL. A recent study[17] demonstrated that using a multilingual model to transfer knowledge from HRLs to LRLs by modeling the shared structure across languages can be effective in terms of improving the performance of NLP models for LRLs. Another popular technique is zero-shot learning, where a model is trained on an HRL, referred to as the source language, and is made to perform tasks in the target language[18], i.e. LRL.

4.2 Performing Data Augmentation

Data augmentation strategies can help overcome the challenges of developing NLU models for LRLs. Synthetic data generation is a popular technique that involves generating new data based on existing data, even if the latter is limited. This technique has been shown to be effective in terms of NLU model performance for LRLs[19]. Semi-supervised learning is another technique that uses a small amount of labeled data and relatively larger amounts of unlabeled data to train a model. This technique has been proved to be effective, as demonstrated by various studies[20]. Crowd-sourcing and crowd-annotation are another way to source data using humans as annotators for labeling[21]. Finally, bootstrapping with related languages is another way improve the NLU for LRLs[22].

4.3 Using Multilingual and Code-switching Models

Since multilingual models are trained on data from multiple languages, these models perform better than monolingual ones when it comes to understanding the nuances of each individual language, including when it comes to NLU for LRLs[23]. Code-switching models understand the phenomenon of code switching and have been effective in terms of improving NLU for LRLs[24].

4.4 Deploying Rule-based and Linguistic Approaches

These approaches leverage syntactic structure and linguistic rules to extract information from text. In the context of VAs, this can be extended to extracting meaning from the user utterance. Some methods are rule-based extraction where a domain-specific seed dictionary is used to identify key phrases, or syntactic-based extraction where the structure of the sentence is leveraged to derive understanding[25].

V. CASE STUDIES

This section covers case studies demonstrating the effectiveness of the NLU techniques described above in improving VA capabilities for specific low-resource languages.

A case study explored the different data-related challenges to improve language understanding in VAs for LRLs. This study proposed multiple data-augmentation strategies, including synthetic data generation to improve NLU[26]. Another paper focused on deep learning techniques for speech processing including data augmentation strategies via means like human re-phrasing, back-translation and dropout[27]. Another study explores the effectiveness of semi-supervised consistency training for NLU in LRLs and also uses similar data augmentation.

VI. FUTURE DIRECTION

Development of NLU for LRLs is an ongoing area of research. This section covers some avenues for future research in this area.

The most obvious avenue is investing in creation of linguistic resources for LRLs, such as annotated corpora, dictionaries and ontologies[28]. Collaboration with communities that use LRLs can provide better understanding along with the cultural context, history and other contextual attributes[29]. Additionally, NLU models should be developed for understudied dialects and indigenous languages to preserve linguistic diversity and richness.

VII. FIGURES AND TABLES

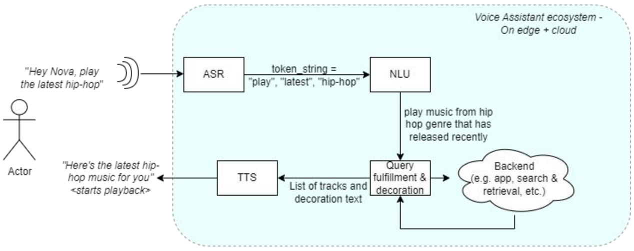

Fig. 1: End to end overview of how a hypothetical voice assistant Nova might look like[8]

VIII. CONCLUSION

In summary, this paper identified the top challenges for voice assistants to perform well when it comes to NLU of low resource languages. It also recommends strategies to overcome these challenges and insights to drive future research and development efforts to make voice assistants more accessible and inclusive.

REFERENCES

[1] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. ACM Trans. Comput.-Hum. Interact. 26, 3, Article 17 (June 2019), 28 pages. https://doi.org/10.1145/3311956

[2] Ashlesha Vishnu Kadam, Designing Thoughtful Experiences for Kids on Voice Assistants, International Journal of Artificial Intelligence & Machine Learning (IJAIML), 2(1), 2023, pp. 75-81. https://doi.org/10.17605/OSF.IO/HKTS8

[3] Hoy MB. Alexa, Siri, Cortana, and more: an introduction to voice assistants. Med Ref Serv Quart. 2018;37(1):81–8

[4] Alexandre Magueresse, Vincent Carles, Evan Heetderks, "Low-resource Languages: A Review of Past Work and Future Challenges", arXiv:2006.07264

[5] https://keyreply.com/blog/conversational-ai-low-resource-languages/

[6] https://vtiya.medium.com/is-your-language-a-low-resouce-language-c17da390909f

[7] https://blog.neurotech.africa/low-resource-languages-vs-conversational-artificial-intelligence/

[8] Ashlesha Vishnu Kadam, “Designing Thoughtful Experiences for Kids on Voice Assistants”, International Journal of Artificial Intelligence & Machine Learning (IJAIML) Volume 2, Issue 01, Jan-Dec 2023, pp. 75-81. DOI: https://doi.org/10.17605/OSF.IO/HKTS8

[9] https://www.miquido.com/blog/what-are-voice-assistants/

[10] Brown, Cozby, et al., “Research Methods in Human Development”, Page 37, Psychological Abstracts, Vol. 79, #19842, p. 2397

[11] https://homepage.cs.uri.edu/faculty/wolfe/book/Readings/Reading13.htm

[12] Duong, L., Kanayama, H., Ma, T., & Bird, S. (2018). A comparative study of language representation methods for low-resource languages. arXiv preprint arXiv:1808.08437

[13] Futrell, R., Gibson, E., & Fedorenko, E. (2022). Crosslinguistic word order variation reflects evolutionary pressures of dependency and information locality. Proceedings of the National Academy of Sciences, 119(23), e2122604119. doi: 10.1073/pnas.2122604119.

[14] Svetko, Y., & Callison-Burch, C. (2017). Opportunities and challenges in working with low-resource languages. In Proceedings of the 2017 Joint Summer Workshop on Machine Learning and Human Language Technologies (pp. 1-9). Carnegie Mellon University.

[15] Gardner-Chloros, P. (2009). Code-switching. Cambridge University Press

[16] Bender, E. M., & Koller, A. (2020). Low-Resource Natural Language Processing. arXiv preprint arXiv:2002.11794

[17] Haifeng Wang, Jiwei Li, Hua Wu, Eduard Hovy, Yu Sun, Pre-Trained Language Models and Their Applications,
Engineering, 2022

[18] https://www.v7labs.com/blog/transfer-learning-guide

[19] Thai, Jimerson, et al., “Synthetic Data Augmentation for Improving Low Resource ASR”, Rochester Institute of Technology, https://par.nsf.gov/servlets/purl/10161886

[20] Ragni, A. & Knill, Kate & Rath, Shakti & Gales, M.J.F.. (2014). Data augmentation for low resource languages. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. 810-814.

[21] Bohan Li, Yutai Hou, Wanxiang Che, “Data augmentation approaches in natural language processing: A survey”,
AI Open, Volume 3, 2022, Pages 71-90, ISSN 2666-6510, https://doi.org/10.1016/j.aiopen.2022.03.001.

[22] Catherine Gitau, VUkosi Marivate, “Textual Augmentation Techniques Applied to Low Resource Machine Translation: Case of Swahili”, https://arxiv.org/abs/2306.07414

[23] Xinyi Wang, “Data Efficient Multilingual Natural Language Processing”, Carnegie Mellon University, https://www.lti.cs.cmu.edu/sites/default/files/wang%2C%20cindy%20-%20Thesis.pdf

[24] Diwan, Vaideeswaran, et al., “Multilingual and code-switching ASR challenges for low resource Indian languages”, https://arxiv.org/abs/2104.00235

[25] Yang, L., Wang, Q., Yu, Z., Kulkarni, A., Sanghai, S., Shu, B., Elsas, J., & Kanagal, B. (2021). MAVE: A Product Dataset for Multi-source Attribute Value Extraction. arXiv preprint arXiv:2112.08663.

[26] Subendhu Rongali, “Low Resource Language Understanding in Voice Assistants”, https://scholarworks.umass.edu/dissertations_2/2717/

[27] Mehrish, A., Majumder, N., Bhardwaj, R., Mihalcea, R., & Poria, S. (2023). A Review of Deep Learning Techniques for Speech Processing. Information Fusion, 101869. doi: 10.1016/j.inffus.2023.101869.

[28] Gabriel Nicholas, Aliya Bhatia, “Lost in Translation – Large Language Models in non-English Content Analysis”, May 2023

[29] Dwivedi, Kshetri, et al., Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy, International Journal of Information Management, Volume 71, 2023, 102642, ISSN 0268-4012, https://doi.org/10.1016/j.ijinfomgt.2023.102642