INTEROPERABLE DATA SHARING FRAMEWORK

Report of the project, Advancing SASL for 4IR Technological Development using Place Names

Funded by the Department of Sport, Arts and Culture (South Africa)

April 2022 – June 2025

HMVE COMBRINK | KIRSTEN DE VILLIERS | SARA SIYAVOSHI | PRISCILLA KECHE | NICK COETZEE
NKATEKO NKUNA | CHRISMI LOTH

sport, arts & culture
Department:
Sport, Arts and Culture
REPUBLIC OF SOUTH AFRICA

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS
INTERDISCIPLINARY CENTRE FOR DIGITAL FUTURES

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA
UFS
THE HUMANITIES
SOUTH AFRICAN SIGN LANGUAGE AND DEAF STUDIES

CONTENTS

FOREWORD

This research is the product of the project, Advancing SASL for 4IR Technological Development using Place Names, funded by the Department of Sport, Arts and Culture (South Africa), April 2022 – June 2025. It is a collaborative project between the Interdisciplinary Centre for Digital Futures and the Department of South African Sign Language and Deaf Studies, both at the University of the Free State (South Africa).

The following people made notable contributions to the sociolinguistic component of the project: Dr Chrismi Loth, Dr Patrick Sibanda, Dr Sara Siyavoshi, Ms Jani de Lange, Prof Annalene van Staden, Ms Emily Matabane, Ms Susan Lombaard, Prof Theodorus du Plessis, Mr Donovan Wright, Ms Lucia Mamotete Mapeshoane, Ms Kirsten de Villiers, Ms Gloria Motshoeneng, Mr Nhlanhla Simelane, Ms Anele Kotoyi and Ms Annemarie le Roux.

The following people made notable contributions to the computational component of the project: Dr Herkulaas Combrink, Mr Nkateko Nkuna, Ms Priscilla Keche, Mr Taylon Colbert, Mr Aviwe Matoti, Mr Jarryd Trip, Ms Valusha Oelofse and Mr Molemo Dibe.

The advancement of Human Language Technology (HLT) in underrepresented languages, particularly signed languages, remains a critical area of research and development. South African Sign Language (SASL), as the recognised Sign Language of the Deaf community in South Africa, presents unique challenges and opportunities in the realm of artificial intelligence, computational linguistics, and inclusive digital technologies. This report presents a comprehensive and interoperable data-sharing framework designed to address these challenges by open research, technological innovation, and community-driven development in SASL studies.

Over the past several years, our research team has actively explored multiple dimensions of SASL research and application, culminating in several impactful outcomes. Central to this endeavour has been our work with a Microsoft dataset, enabling us to analyse and expand the available linguistic resources for SASL. This research has played a pivotal role in furthering our understanding of SASL's computational representation, sign recognition, and translation methodologies. Moreover, our team has supported a critical master's study that explored key aspects of sign language processing, ensuring an academic foundation for future developments.

A significant milestone in this journey has been the development of the LebitsoApp, an initiative that demonstrates the power of sign language communication. The App serves as both a research tool and a practical solution for bridging research gaps between Deaf and hearing individuals, to empower more people to become excited about this research, and to enhance HLT for SASL. Beyond its immediate applications, the LebitsoApp also serves as a showcase of the underlying methodologies and technologies driving SASL research forward. Some of the deliverables of the project are showcased through the App.

Recognising the intersection between linguistic accessibility and economic empowerment, our team has also spearheaded an initiative around financial literacy which will be further developed for the Deaf community. The "Good Money Habits" project, initially designed as an entrepreneurial tool for broader communities, is the stepping stone for hearing and ultimately Deaf users, providing essential financial education in an accessible format so that basic money habits can be trained in a user-friendly manner. Future iterations of this initiative will see the development of a dedicated Deaf Good Money Habits App, ensuring that financial literacy tools are inclusive and impactful for all members of society.

To support the development of sign language processing models for HLT, we have successfully built a Generative Adversarial Network (GAN) engine capable of generating and labelling synthetic SASL data. This innovative approach addresses the scarcity of labelled SASL data, significantly enhancing the ability of machine learning models to recognise and translate SASL with greater accuracy. By incorporating synthetic data generation, we contribute to the ongoing refinement of AI models while maintaining ethical considerations regarding data privacy and inclusivity.

All of our work, including datasets, applications, and their corresponding code, is hosted on GitHub. This commitment to open-source principles ensures that researchers, developers, and the broader community can access, modify, and build upon our contributions. Extensive documentation accompanies each aspect of our research and development process, providing valuable insights and transparency to those interested in engaging with our work.

The overarching aim of this report is to detail the structure and components of an interoperable data-sharing framework for SASL research. By making our data and applications openly available, we seek to empower researchers, technologists, and the Deaf community to advance the field of Human Language Technology in the signed modality. This framework is designed to facilitate seamless collaboration, data integration, and innovation across multiple stakeholders, ensuring that future advancements in SASL processing are driven by inclusive and ethically sound principles.

In presenting this report, we hope to underscore the critical importance of accessible and inclusive technological advancements in SASL research. Through collaboration, innovation, and open data practices, we can collectively work towards a future where SASL, and sign languages globally, are fully integrated into the digital landscape, creating equal opportunities for communication, education, and economic participation for the Deaf community.

DR HMVE COMBRINK

THE UNIQUE CHALLENGES OF SASL RESEARCH

Sign languages, unlike spoken languages, exist in a visual-spatial modality. This fundamental difference presents significant challenges for computational processing, storage, and standardisation in the field. Unlike text-based datasets, which can be processed using well-established Natural Language Processing (NLP) methods, sign language data is typically captured through videos, requiring advanced computer vision techniques and machine learning models for analysis. Furthermore, SASL, does not have a universally agreed-upon written form for HLT and computer vision research, making data annotation and linguistic structuring more complex. However, in this area there is some consensus over different types of annotation methods for its linguistic properties.

A significant barrier to progress in SASL HTL research, however, is the lack of large-scale, high-quality datasets. Data scarcity results from various factors, including the limited number of fluent SASL signers available for HTL recording, the ethical and privacy concerns surrounding video data collection, and the absence of standardised HLT annotation frameworks. Unlike widely spoken languages, where vast amounts of text data exist across books, websites, and media, SASL HLT requires targeted and controlled data collection efforts.

Moreover, sign languages are not universally interchangeable. While there are similarities among different sign languages, each has its own grammar, lexicon, and cultural context. This linguistic diversity necessitates dataset-specific HLT models, further complicating the creation of generic sign language HLT systems. As a result, a collaborative and interoperable data-sharing framework is essential to accelerate research progress while ensuring inclusivity and representation.

The Need for Interoperability in SASL Data

Interoperability refers to the ability of different systems, datasets, and research outputs to work together cohesively. ***In the context of SASL research, interoperability is critical for several reasons:***

1 STANDARDISATION OF DATA FORMATS FOR HLT:
The variety of data formats used in SASL research, ranging from raw video recordings to annotated gloss-based representations, requires standardisation for HLT to ensure compatibility across different platforms and research initiatives. Without standardised formats of HLT data, researchers and developers face unnecessary obstacles in merging datasets and leveraging existing models.

2 CROSS-INSTITUTIONAL COLLABORATION:
Many institutions and research teams work in silos, limiting the potential for knowledge sharing and dataset expansion. By establishing an interoperable framework, we facilitate collaboration between universities, tech companies, and Deaf advocacy groups, ensuring that research efforts are not duplicated but rather build upon each other.

3 ETHICAL DATA SHARING:
Given the sensitive nature of video-based datasets involving signers, ethical considerations around privacy, consent, and data governance must be at the forefront. An interoperable framework provides guidelines and safeguards to ensure responsible data sharing without compromising individual privacy.

4 ENHANCING AI AND MACHINE LEARNING MODELS:
A significant challenge in sign language AI development is the need for large-scale, well-labelled datasets. By making datasets interoperable, we enable researchers to train more accurate machine learning models, leading to better real-world applications such as sign language recognition, translation, and avatar-based interpretation systems.

Existing Limitations in SASL Data Availability

Despite growing interest in sign language AI research, current datasets remain fragmented and limited in scope. ***The primary issues include:***

SMALL SAMPLE SIZES:	LACK OF DIVERSITY:	LIMITED ACCESS TO DATA:	INCONSISTENT ANNOTATION PRACTICES:	ETHICS AND CONFIDENTIALITY:
Many existing datasets only include a handful of signers, making them unrepresentative of the broader SASL-speaking community.	Variations in age, gender, and regional dialects within SASL are often underrepresented.	Many datasets are locked behind proprietary agreements or institutional restrictions, hindering open research and innovation.	There is no universal standard for labelling sign language data, making it difficult to integrate multiple datasets into a unified resource.	SASL data is the most identifiable of data, and both the collection and storage of SASL data poses identifiable and security challenges.

Addressing these challenges through an interoperable data-sharing framework will lay the foundation for more inclusive and impactful advancements in SASL research.

DR HMVE COMBRIN | P KECHE

SASL DATA SHARING FRAMEWORK

To ensure the responsible and ethical use of South African Sign Language (SASL) data, we propose an open yet structured data-sharing framework that aligns with the needs of SASL researchers, technology enthusiasts, and developers. This framework is designed to create innovation while maintaining transparency, fairness, and recognition for the original contributors of the dataset and associated applications.

Open Access with Clear Requirements

Anyone who wishes to use the datasets, applications, or models developed in this project must adhere to the following principles:

1 PUBLIC AVAILABILITY OF DERIVATIVE WORK:
Any work derived from this project—including models, applications, or modifications to the dataset—must be made publicly available. This includes the publication of trained AI models, research papers, and documentation.

2 RECOGNITION AND ATTRIBUTION:
Users must acknowledge and reference this project when using its datasets, applications, or AI models. Proper citation of the original work must be included in academic publications, software repositories, and commercial applications.

3 OPEN-SOURCE CODE REQUIREMENT:
Any software, model, or tool developed using this dataset must be made open-source. The full source code must be available under a compatible open-source license, ensuring transparency and fostering further improvements by the community.

4 COMMERCIALISATION CONTRIBUTIONS:
If any derivative application or service is commercialised, the entity must pay a fair-use licensing fee to the project entity. This ensures continued support for SASL research and development while maintaining fairness in commercial applications. Additionally, full transparency on recognition must be maintained, ensuring that users and stakeholders are aware of the origins of the dataset and tools..

Recommended Licensing Model: CC BY-NC-SA License

To legally enforce these principles while maintaining the openness of the project, we propose the adoption of the ***Creative Commons Attribution-Non-Commercial-Share Alike (CC BY-NC-SA) License.***

The CC BY-NC-SA License ensures the following:

ATTRIBUTION (BY):
Any person or entity using this dataset, models, or applications must provide proper credit to the original project. This means citing the original work in all relevant materials, publications, and applications.

NON-COMMERCIAL USE (NC):
The dataset and its derivatives can be freely used for research, education, and personal projects, but they cannot be used for commercial purposes without explicit permission and an approved commercial licensing agreement.

SHARE ALIKE (SA):
Any modifications, enhancements, or derivative works must be distributed under the same **CC BY-NC-SA** license. This ensures that all improvements remain open and accessible to the community, preventing proprietary restrictions on extended versions of the work.

Commercial Licensing Requirement

If an organisation or developer intends to commercialise an application or service that incorporates this dataset or its derivatives, they must:

OBTAIN A COMMERCIAL LICENSE:
A separate agreement must be made with the project entity to ensure a fair financial contribution toward ongoing research and development.

PROVIDE FULL RECOGNITION:
The commercialised product must visibly acknowledge the SASL Data Sharing Framework as the foundational source.

MAINTAIN TRANSPARENCY:
The commercialised application must openly disclose how the dataset and technology were integrated.

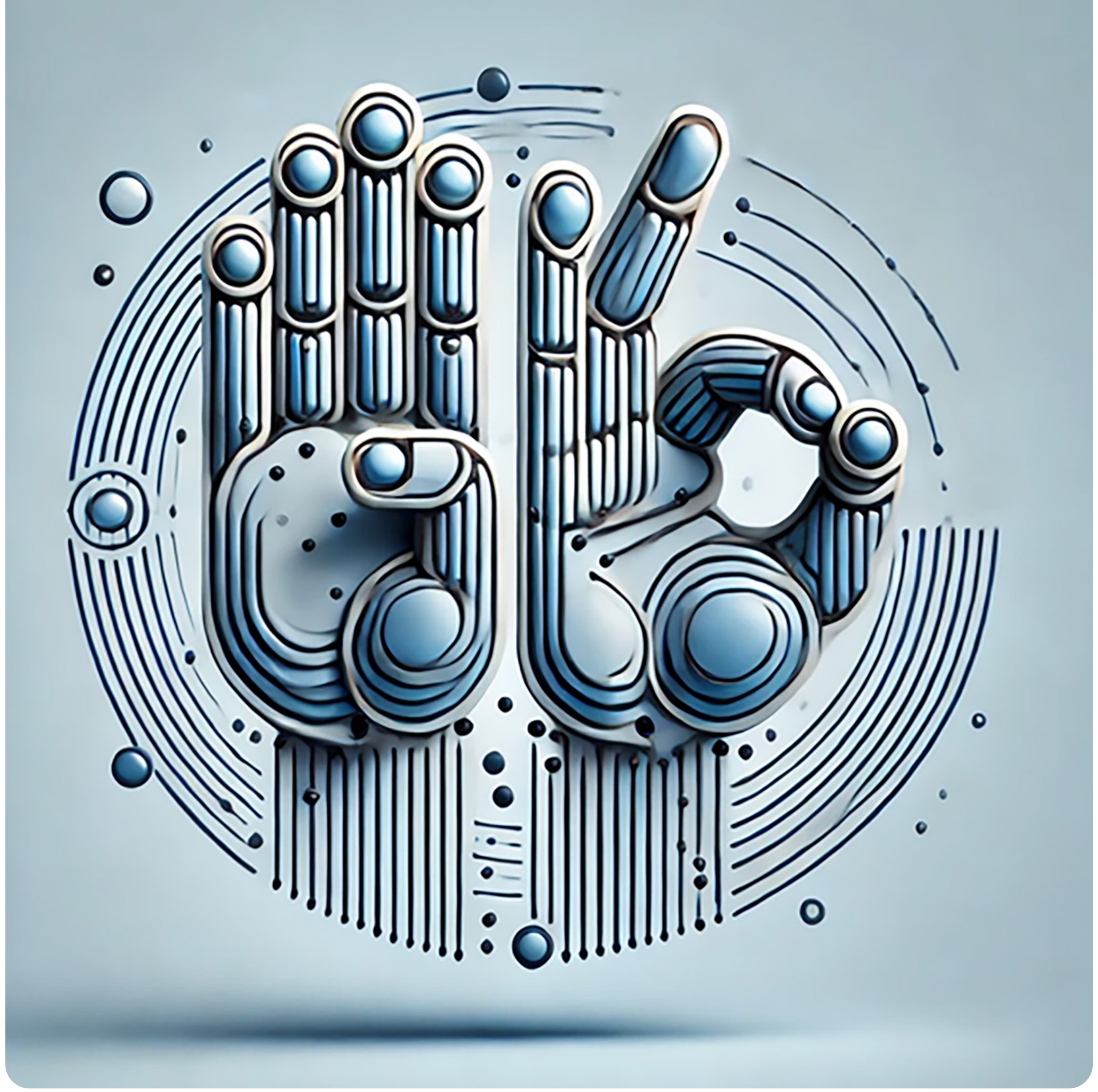
Implementation and Enforcement

To ensure compliance with these principles:

1. **ALL DATASETS, MODELS, AND APPLICATIONS** will include a licensing agreement explicitly outlining these terms.
2. **A MONITORING SYSTEM** will track usage and compliance, ensuring that all users properly attribute the project.
3. **ANY VIOLATIONS OF THE LICENSING TERMS** may result in restrictions on dataset access and potential legal action to protect the integrity of the project.

By using a ***CC BY-NC-SA License***, we ensure that research and development in SASL technology remain open and beneficial to all while protecting against commercial exploitation without fair compensation. This structured framework allows for ethical and responsible innovation, ensuring that contributions remain accessible, transparent, and supportive of the broader SASL community.

DR HMVE COMBRINK | P KECHE | DR C LOTH



FIRST SET OF INTEROPERABLE DATA: ASL-SASL PROJECT

The primary objective of this task was to analyse videos in American Sign Language (ASL) to assess whether the words, terms, or phrases presented in these videos are signed in a similar manner in South African Sign Language (SASL). This project used a dataset comprising 15,205 ASL videos, with Video links from Microsoft. From Microsoft, the agreement with the data was that this dataset functions under the license agreement of Computational Use of Data Agreement (C-UDA) v0.1

Computational Use of Data Agreement (C-UDA) v0.1

The **Computational Use of Data Agreement (C-UDA) v0.1** is a licensing framework designed to facilitate the computational use of data while maintaining proper attribution and ethical considerations. This agreement allows users to **use, modify, and distribute** data, provided they adhere to the specified terms.

Key Provisions

1. Provision of the Data

1.1. Users may utilise, modify, and distribute the provided data for **computational purposes,** as long as they comply with the agreement.

1.2. The **Data Provider** will not initiate legal action against users or downstream recipients, provided they follow the terms of the C-UDA.

2. Restrictions

2.1. The data may only be used for **computational analysis** and cannot be used for other unauthorised purposes.

3. Redistribution of Data

3.1. Users may redistribute the data if:

- **PROPER ATTRIBUTION** is maintained, ensuring that credit is given to the original data provider.
- **ALL DOWNSTREAM RECIPIENTS** are bound by the same agreement.

4. No Warranty and Limitation of Liability

4.1. The data provider **does not guarantee ownership rights** to the data.

4.2. The data is provided **"as is"**, with **no warranties or guarantees** regarding accuracy, completeness, or fitness for a particular use.

4.3. The provider holds **no liability** for any **damages, losses, or legal claims** arising from the use of the data.

5. Definitions

- **COMPUTATIONAL USE:** Using data for analysis, machine learning, or any other automated processing.
- **DATA PROVIDER:** The original source of the data.
- **DOWNSTREAM RECIPIENT:** Any person or entity receiving the redistributed data.
- **OUTPUT:** Results derived from data processing, excluding significant portions of the raw data itself.
- **UPSTREAM DATA PROVIDERS:** Original sources from which the Data Provider obtained the material.

Interpretation

The **C-UDA v0.1** is a permissive agreement that allows researchers and developers to use data freely for computational analysis while ensuring that **proper credit is given** and that modifications remain open. The agreement restricts commercial use without explicit permissions, safeguards against legal claims, and ensures transparency in data redistribution. However, it also places responsibility on users to acknowledge that the data is provided **without any warranty**, meaning they must validate and verify data reliability before application. This agreement aligns with the ethical use of **open-access datasets** in AI, machine learning, and computational research while maintaining **legal protection** for data providers.

Adaptation based on the Microsoft Agreement for the project by:

P KECHE, DR HMVE COMBRINK, N NKUNA, N COETZEE, A MATOTI

PHASE 1 METHOD/PROCESS: ANALYSIS OF ASL-SASL VIDEO DATASET

The purpose of this phase of the project was to analyse a large dataset of **15,205 American Sign Language (ASL) videos** to determine how closely their signs align with South African Sign Language (SASL) using phenology only. This process was fundamental in identifying similarities, differences, and areas where SASL lacks direct equivalents for ASL signs. The work was carried out by Deaf and SASL-signing Student Assistants, who played a critical role in evaluating linguistic and structural variations between the two sign languages. Each Student Assistant was assigned a list of videos, along with a **description of the sign** depicted in the video and a corresponding **Video link** to the ASL content. They were required to collect and document specific information regarding each sign, using a structured framework to ensure consistency across the dataset.

The following criteria were assessed:

1 DOWNLOADABLE/ACCESSIBLE STATUS:

The availability of each video was checked and categorised as either **"Yes"** (accessible) or **"No"** (not downloadable/inaccessible).

2 ASL-SASL SIMILARITY:

The Student Assistants determined whether the ASL sign had an equivalent meaning in SASL.

The classification was divided into three categories:

- **SAME:** The sign had the same meaning and form in both ASL and SASL.
- **SOMEWHAT THE SAME:** The sign shared similarities but had differences in execution or meaning.
- **NOT THE SAME:** The sign did not have a recognisable or equivalent form in SASL.

3 NUMBER OF HANDS USED:

Each sign was documented based on whether it involved **No hands, One hand, or Two hands.**

4 STUDENT ASSISTANT IDENTIFICATION:

Each assistant recorded their initials and surname to track their contributions and ensure accountability.

All the collected data was compiled into an online spreadsheet, where real-time updates and collaborative verification were performed.

All collected data was compiled into a shared sheet, as illustrated in Figure 1 below:

A	B	C	D	E	F
	YouTube link	Same in SASL as ASL?	Can the link be downloaded?	Number of Hands used	Name of person responsibl
match [light-a-MATCH]	www.youtube.com/watch?v=C3	SOMEWHAT THE SAME	YES	TWO HANDS	[illegible]
FAIL	www.youtube.com/watch?v=PIs	NOT THE SAME	YES	TWO HANDS	[illegible]
BOOK	www.youtube.com/watch?v=J7t	SAME	YES	TWO HANDS	[illegible]

Figure 1: Copy of the completed shared sheet

Following the initial analysis, the videos were categorised based on the results from the ASL-SASL similarity assessment into the three classifications: SAME, SOMEWHAT THE SAME, and NOT THE SAME.

Importance of the Analysis

This study was significant for several reasons:

1. **ENHANCING SASL RESEARCH:** The comparative analysis provided an extensive linguistic mapping of ASL and SASL, contributing to the broader study of **signed language linguistics.**
2. **DEVELOPING SASL LANGUAGE TECHNOLOGY:** Understanding ASL-SASL similarities helps in the development of **machine learning models** for sign recognition, automatic translation, and synthetic sign generation.
3. **ADDRESSING DATA SCARCITY:** Sign languages are underrepresented in computational linguistics, and **this project bridges the data gap** by offering structured insights into sign variations.
4. **SUPPORTING DEAF EDUCATION:** By identifying where SASL lacks equivalents, the study helps **curriculum developers** and **educators** enhance SASL resources.

Findings

The preliminary results indicated that:

- A **significant proportion of ASL signs did not match SASL** due to linguistic phonological differences.
- Some signs had **partial alignment** but required context-specific modifications.
- There were instances where **no direct equivalent existed in SASL**, highlighting areas where further development in SASL lexicon is needed.
- **Hand usage varied**, with some signs requiring adaptation to fit SASL's grammatical and spatial structures.

Justification for Importance

The **Microsoft dataset** provided a unique opportunity to conduct a large-scale comparative analysis between ASL and SASL, a **critical step in advancing sign language AI models**. Without structured data, most machine learning approaches for sign recognition rely on **unannotated or small-scale datasets**, limiting their ability to generalise across sign languages. By ensuring that **this dataset is meticulously analysed**, we create a **valuable linguistic resource** that serves as the foundation for **future computational sign language models.**

Furthermore, this study promotes **inclusive technological advancements**, ensuring that SASL is represented in the growing field of **HLT**. By **bridging linguistic gaps**, this research offers a more equitable approach to **AI-powered research** for the Deaf community. The ability to map **ASL to SASL systematically** also opens doors for **cross-linguistic sign language interpretation tools**, reducing communication barriers across **global Deaf communities**.

The structured approach used in this project ensures that the **ASL-SASL video dataset** is thoroughly analysed for computational and linguistic applications. Future work will involve the **expansion of the SASL dataset**, the **development of AI-driven sign language recognition**, and **continued advocacy for open-access sign language data to support innovation in human language technology**.

P KECHE | DR HMVE COMBRINK | DR S SIYAVOSHI

VALIDATION AND ACCURACY IN ASL-SASL ANALYSIS: SECOND TASK

The **Second Task** in Phase 1 of the project was essential for validating and refining the initial dataset analysis. After the **first task classified 15,205 ASL videos** into categories based on their similarity to South African Sign Language (SASL), the second phase was implemented to enhance **accuracy, consistency, and reliability**. By re-examining the data, this phase ensured that findings were not only correct but also **replicable**, strengthening the foundation for further research and computational applications.

Why the Second Task Was Necessary

Validation is crucial in any study, particularly when dealing with datasets intended for **machine learning models** and **linguistic mapping**. The **first task's initial classifications**, SAME, SOMEWHAT THE SAME, and NOT THE SAME, provided an essential first look at ASL-SASL similarities, but **human analysis is inherently prone to error.** Differences in interpretation, sign execution, and linguistic context necessitated **a secondary review** to confirm accuracy and correct inconsistencies.

This second task was particularly important for:

1. **CONFIRMING CONSISTENCY:** Ensuring that different Student Assistants provided the same classification for identical signs.
2. **ELIMINATING ERRORS:** Identifying and correcting **misclassified** or **inconclusive** results from the first task.
3. **IMPROVING DATA INTEGRITY:** Establishing a **high-confidence dataset** for future computational sign language research.
4. **REFINING CATEGORISATION:** Strengthening the dataset's precision by **reassessing ambiguous cases.**

Breakdown of the Second Task

The second task was structured into **three validation categories**, ***each corresponding to classifications from the first task:***

CATEGORY 1: SAME

- Videos initially marked as **SAME** were reassessed to verify that the ASL signs were indeed identical to their SASL counterparts.
- The **same evaluation criteria** from the first task were applied: downloadability, number of hands used, and ASL-SASL similarity.
- Any inconsistencies or misjudgements were flagged and corrected.

CATEGORY 2: SOMEWHAT THE SAME

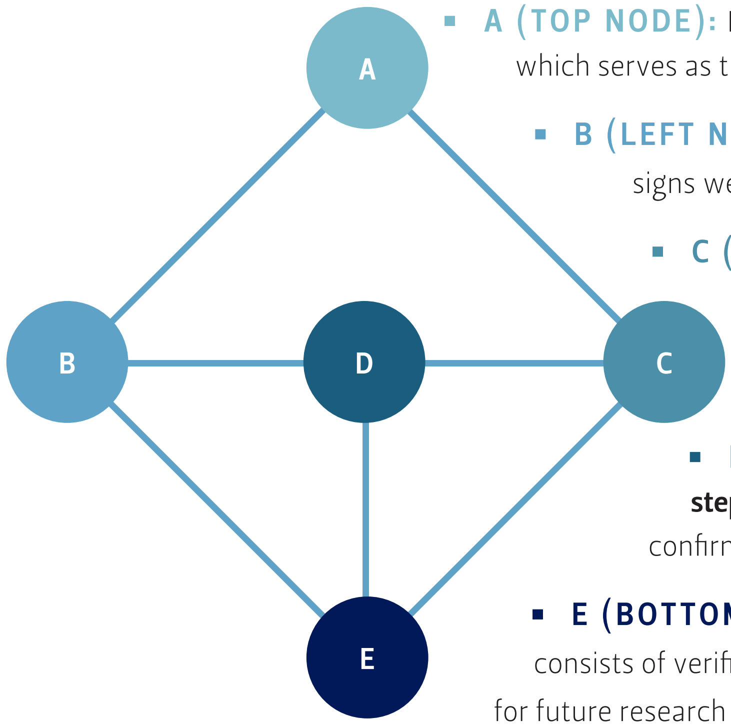
- Videos categorised as **SOMEWHAT THE SAME** were analysed again to determine if they should remain in this category or be moved to **SAME** or **NOT THE SAME**.
- This review helped establish whether perceived differences were **minor variations** or **significant enough** to warrant reclassification.
- This step was especially critical for identifying **regional or dialectal variations** in SASL that could influence classification.

CATEGORY 3: NOT THE SAME

- This validation ensured that signs classified as **NOT THE SAME** were indeed **linguistically and structurally distinct**.
- The process helped rule out any **overlooked similarities** that may have been missed in the first task.
- It also contributed to documenting cases where **SASL does not have a direct equivalent** for ASL signs, highlighting areas for future lexicon development.

PHASE PROCESS DETAILS

The second task in **Phase 1** was not merely a repetition of the first task, it was an essential **validation step** ensuring that findings were **accurate, consistent, and meaningful**. Without it, the dataset would be vulnerable to errors that could **hinder** technological advancements in **sign language AI, educational tools, and linguistic research**. By reinforcing the credibility of the dataset, this task **lays the foundation** for SASL's **digital and computational future**.



- **A (TOP NODE):** Represents the **initial dataset** containing ASL videos, which serves as the starting point for the classification process.
- **B (LEFT NODE):** Represents the **SAME** category, where ASL signs were found to have direct SASL equivalents.
- **C (RIGHT NODE):** Represents the **SOMEWHAT THE SAME** category, where ASL signs shared partial similarities with SASL signs but required further analysis.
- **D (CENTRAL NODE):** Represents the **validation step**, where classification results were reviewed to confirm accuracy and consistency across categories.
- **E (BOTTOM NODE):** Represents the **final refined dataset,** which consists of verified ASL-SASL classifications, ensuring high-quality data for future research and AI training.

Weekly data clean up and verification

Each week, a thorough data cleanup and verification process was conducted on the dataset to ensure that all accessed videos received comprehensive analysis feedback from the Student Assistants. This verification focused on all videos that had been previously marked as not downloadable or inaccessible by the assistants.

The data cleanup process involved reviewing the analysis feedback provided to ensure that the results met the established criteria for analysis. ***The key criteria included:***

- **ASL-SASL SIMILARITY:** Categorised as SAME, NOT THE SAME, or SOMEWHAT THE SAME.
- **NUMBER OF HANDS USED:** Classified as TWO HANDS, ONE HAND, or NO HANDS.
- **VIDEO LINK ACCESSIBILITY:** Indicated as YES (downloadable/accessible) or NO (not downloadable/inaccessible).
- **STUDENT ASSISTANT IDENTIFICATION:** Documented by the initials and surname of the assistant responsible for the analysis feedback.

Any inconsistencies identified in the analysis feedback were communicated to the respective Deaf and hearing Student Assistants to ensure corrections were made and that the feedback aligned with the acceptable criteria. Ongoing communication was maintained with the assistants to facilitate the smooth and timely completion of each task. Additionally, a backup of the dataset, maintained in the online spreadsheet, was downloaded weekly to ensure data security. This precaution was necessary due to the collaborative nature of the editing process, involving multiple users.

P KECHE

MONITORING AND EVALUATION: WEEKLY PROGRESS REPORT

This report summarised the weekly progress of the project and was submitted to the Project Lead. The report included detailed information on the analysis conducted on the videos, verification of links, Student Assistants' activities and data management activities as follows:

NUMBER OF HANDS USED	Video Analysis feedback ASL vs SASL			GRAND TOTAL
	NOT THE SAME	SAME	SOMEWHAT THE SAME	
NO HANDS		2		2
ONE HAND	2 966	1 859	312	5 137
TWO HANDS	3 697	2 929	776	7 402
Grand Total	**6 663**	**4 790**	**1 088**	**12 541**

Monitoring and evaluation items

- **NUMBER OF VIDEOS ANALYSED:** This section provides the total count of videos analysed during the week.
- **ANALYSIS RESULTS:** Detailed feedback from the student assistants based on their analysis.
- **DOWNLOADABLE/ACCESSIBLE LINKS:** Count of video links verified as not downloadable or inaccessible.
- **STUDENT ASSISTANT ACTIVITIES:** A breakdown of each student assistant's progress based on their analysis feedback.
- **DATA CLEAN-UP & VERIFICATION ISSUES:** This section highlights any data clean-up and verification issues identified during the week.
- **BACKUP FILE:** Confirmation of the backup file download.

PHASE 1 RESULTS

First Task

The analysis results from the first task = 15 205 videos. 2664 video links not downloadable/ inaccessible in the First Task.

NUMBER OF HANDS USED	Video Analysis feedback ASL vs SASL			GRAND TOTAL
	NOT THE SAME	SAME	SOMEWHAT THE SAME	
ONE HAND	199	1 549	71	1 819
TWO HANDS	303	2 495	155	2 953
Flagged***				4
Grand Total	**502**	**4 044**	**226**	**4 776**

Flagged*** 4 videos for further investigation as the provided term/phrase is not indicated/available in the video content. 14 video links not downloadable/inaccessible, however, these links were accessible during the First Task.

Second Task Category 1: SAME

4 790 videos which were analysed and results to be ASL-SASL SAME in the First Task were analysed again in the Second Task Category 1. The videos for Second Task Category 2 included the videos with the SOMEWHAT THE SAME results from the First Task = 1 088 and Second Task Category 1= 226 analysis.

Second Task Category 3: NOT THE SAME

Second Task C3 videos = 7 610, were made up of NOT THE SAME results from First Task = 6 663 + Second Task C1 = 502 + Second Task C2 = 445.

NUMBER OF HANDS USED	Video Analysis feedback ASL vs SASL			GRAND TOTAL
	NOT THE SAME	SAME	SOMEWHAT THE SAME	
ONE HAND	2 043	400	173	2 616
TWO HANDS	2 561	574	326	3 461
Flagged***				1
Grand Total	**4 604**	**974**	**499**	**6 078**

PHASE 2

Videos for Machine readable sign language

The ASL videos that were analysed and had the SAME results from Second Task C1&C2 were compiled for Phase SAME category = 4 630 videos (Second Task C1 SAME results 4 044 videos + 586 videos from Second Task C2 SAME results). Sign language students were assigned with re-recording the videos and 364 videos, from this combined Phase 2 SAME category were recreated. A unique code was developed for each of the videos in the Phase 2 SAME Category for reference. The recreated 364 videos were labelled accordingly using the unique code and description of ASL video.

Some of the issues identified with the dataset

DUPLICATES

- ASL video description provided = 12 527 duplicates were identified.
- Video links =11 853 duplicates were identified.

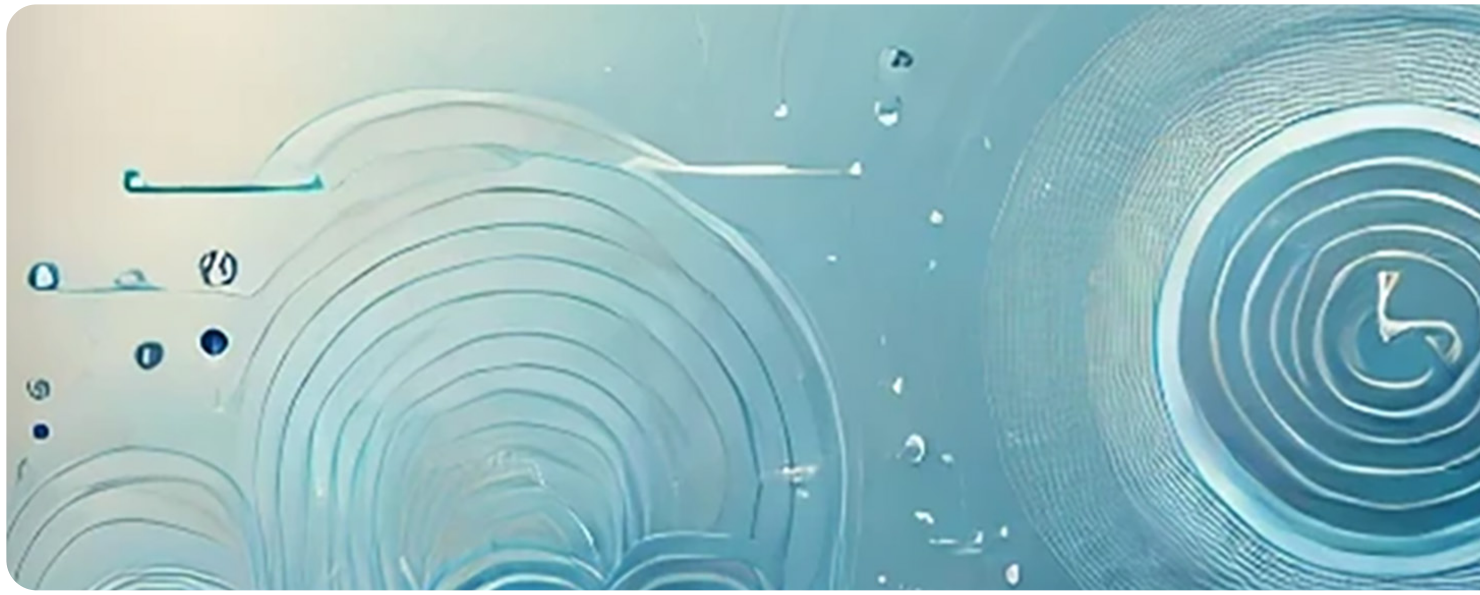
VIDEO DESCRIPTION

- 1 entry was blank, with no video description provided but video link was provided.

VIDEO LINKS PROVIDED

- 1 entry had a description provided but no video link was provided.
- 2 664 Video links were not downloadable/inaccessible in First Task.
- Some links were downloadable/accessible during First Task and became not downloadable/ inaccessible during Second Task.

P KECHE



INTERPRETATION OF ASL-SASL ANALYSIS RESULTS AND THEIR IMPORTANCE FOR SASL

Understanding the Results

The dataset analysis of 15 205 ASL videos provided key insights into the relationship between ASL and South African Sign Language (SASL). The classification of videos into **SAME, SOMEWHAT THE SAME, and NOT THE SAME** categories helped determine linguistic similarities and differences between the two sign languages. The results revealed that a **large portion of ASL signs do not have direct equivalents in SASL**, emphasising the importance of structured linguistic research in bridging these gaps. One of the main reason this dataset was used was due to data availability.

Key Findings:

- **66% of ASL signs were classified as NOT THE SAME in SASL.** This indicates that while ASL and SASL share a visual modality, they are linguistically distinct. Many signs do not translate directly, highlighting the need for **SASL-specific data collection** and linguistic development.
- **A total of 4 790 ASL signs were categorised as SAME.** These signs represent strong linguistic overlap and could be used to develop initial models for **SASL-ASL translation.**
- **Approximately 1 088 signs were classified as SOMEWHAT THE SAME.** These require further research to determine whether **regional variations or contextual changes** in SASL influence their similarity.
- **2 664 video links were inaccessible.** This posed a challenge to dataset completeness and highlighted the limitations of **external video repositories** for sign language research.

Importance of These Findings for SASL

The findings from this analysis are crucial for multiple reasons, particularly for advancing **SASL recognition, education, and accessibility.**

1. Strengthening SASL Linguistic Research	2. Enhancing AI-Based Sign Language Research	3. Addressing Educational Gaps for Deaf Learners	4. Supporting the Development of SASL-Based Assistive Technology
The lack of direct equivalency between ASL and SASL reinforces the need for **dedicated SASL linguistic documentation**. Unlike spoken languages, sign languages do not have written forms, making video-based datasets the primary resource for linguistic studies. The identification of gaps in SASL vocabulary allows **Deaf linguists and educators** to work toward standardisation and expansion of the language.	Developing **machine-readable SASL datasets** is key to advancing **AI-driven sign language research**. By confirming which ASL signs have direct SASL equivalents, and which do not, AI models can be **trained on verified data** rather than relying on assumptions. This improves accuracy in **sign language translation models, gesture recognition software, and digital communication tools** for the Deaf community.	For **Deaf education**, knowing which ASL signs lack direct SASL counterparts helps tailor **bilingual learning resources**. Many SASL users consume ASL-based educational content due to the scarcity of SASL materials. By identifying **mismatched signs**, educators can create **localised resources** that cater specifically to SASL learners rather than relying on adapted ASL content.	Assistive technologies such as **real-time sign language translation apps, SASL avatars, and interactive learning tools** require accurate datasets. The results of this study ensure that assistive technology developers focus on **verified SASL data** rather than assuming ASL models can be directly applied to SASL users.

DR HMVE COMBRINK

PHASE 2: TOWARDS A MACHINE-READABLE SASL DATASET

The findings from **Phase 1** informed the next phase of the project, which focused on developing a **machine-readable SASL dataset**. Videos that were classified as SAME in the second task (4 630 videos) were **recreated by sign language students** to form an **original SASL dataset**.

This approach was essential because:

- Many ASL videos were subject to copyright restrictions, preventing their direct use for **training SASL AI models.**
- Recreating videos in SASL ensures that the dataset is **authentic and tailored** to South African Deaf users.
- Assigning **unique codes** to each recreated SASL video enables structured data organisation and **future scalability**.

Challenges Identified and Their Impact on SASL Development

1. Duplicates in the Dataset

- **12 527 duplicate ASL video descriptions** were found.
- **11 853 duplicate video links** were detected. These duplicates needed to be removed to ensure the **integrity and efficiency** of the dataset. The presence of duplicates inflates dataset size without adding new information, making it inefficient for **AI training and linguistic analysis.**

2. Missing or Inaccessible Video Links

- **2 664 video links** were inaccessible in the first task.
- **Some previously accessible links became unavailable** in later stages. This underscores the importance of **data sovereignty**—relying on third-party platforms like Video makes sign language datasets vulnerable to content removal. Future SASL research should prioritise **independent data hosting**.

The results of this analysis provide **critical insights** into the linguistic structure of SASL and its relationship to ASL. The research highlights the importance of **building SASL-specific datasets, improving AI-driven recognition models, and expanding Deaf education resources**. The findings emphasize that **SASL cannot simply be derived from ASL**, and **dedicated efforts must be made to document, develop, and support SASL as a distinct linguistic system**. Future work should focus on expanding **the machine-readable SASL dataset, strengthening AI applications, and ensuring the sustainability of SASL research** to empower the **Deaf community**.

DR HMVE COMBRINK

THE LEBITSOAPP PROTOTYPE SHOWCASE

The **LebitsoApp** serves as a central hub for **SASL data collection, analysis, and dissemination**, encapsulating the fundamental outputs of the project. More than just an application, **Lebitso is link to different digital repositories of sign language data**, designed to inform the development of **SASL HLT** through structured research, AI-driven analytics, and user-driven engagement. The app plays a pivotal role in organising and sharing critical linguistic insights, **supporting both computational sign language advancements and linguistic standardisation efforts**.

Purpose and Motivation

The **LebitsoApp** was developed to **showcase the outputs of the project** while providing a **centralised platform** that integrates all components of **SASL research, technology, and linguistic development** into a single mobile application. The complexity of managing **code, raw data, analysed data, annotated data, and various types of structured information** necessitated a unified space where all elements of the project could be easily navigated and accessed.

Key Features of the LebitsoApp

1. SASL Data Repository and Research Tool

- Aggregates **SASL sign language data**, making it accessible for **researchers, linguists, and AI developers.**
- Functions as a **live dataset** that grows through user interaction, **feeding into AI models** for continuous improvement.

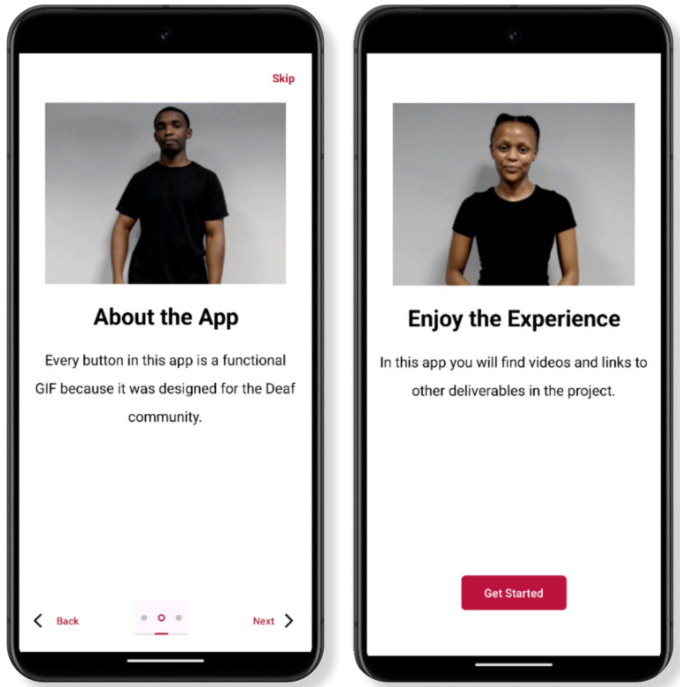
2. Sharing and Community Engagement

- Allows users to **contribute SASL data**, expanding the dataset through **verified community inputs**.
- Promotes knowledge exchange between **researchers, educators, and the Deaf community**, fostering **collaborative sign language development.**

3. Open-Source Development for Accessibility

- The **LebitsoApp**, along with all its **code, data, and development tools,** is **open-source.**
- This ensures that **anyone who wants to contribute to its improvement** can do so, making it a continuously evolving resource for **SASL technology**.

UX Design and Ongoing Research

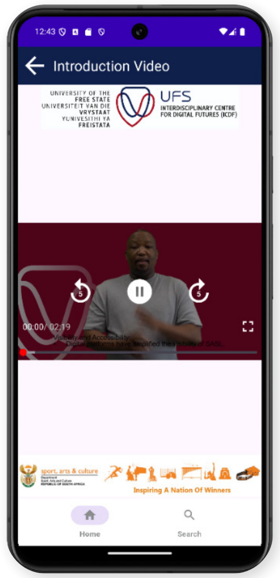


A significant aspect of the **LebitsoApp** is its **user experience (UX) design**, which has been carefully crafted to make the platform **Deaf-friendly**. However, **achieving a truly inclusive and effective design** requires ongoing refinement, as different perspectives and user needs **continuously evolve**.

The app's design was informed by consultations with **Deaf users and sign language researchers**, but true accessibility requires **iterative improvements** based on real-world user feedback. Because **sign language technology must be community-driven**, the open-source nature of Lebitso ensures that **UX enhancements, technological refinements, and new features** can be collaboratively developed. As stated, all of this is available on GitHub.

N COETZEE | T COLBERT | DR HMVE COMBRINK

The LebitsoApp is a fusion of key principles: community, research, inclusion, and innovation. These four pillars drive the development and purpose of the application. The primary goal is to incorporate these principles into showcasing research and innovation at the University of the Free State and potentially beyond the borders of the Free State. The app's user interface (UI) has been designed with the Deaf community in mind, ensuring that SASL is prominently represented. The videos are provided in SASL from a prominent figure not only in the Deaf community but in South Africa; and most interactive buttons feature SASL symbols created by communication science students one of which is a CODA.



Rapid Application Development (RAD) was used to produce the first iteration of the LebitsoApp. More time is needed to refine its elements to meet the standards and expectations of the Deaf community. Future iterations will prioritise user feedback to enhance the overall user experience (UX) and accessibility.

Development Environment and Technical Specifications

To ensure the LebitsoApp is both functional and easy to maintain, it is built using the following tools and frameworks:

Development Environment

- **IDE:** Android Studio (latest stable version)
- **Programming Language:** Java
- **Gradle Version:** Compatible with the Android Gradle Plugin
- **Build System:** Gradle

Android Specifications

- **Minimum SDK Version:** 27 (Compatible with most modern devices)
- **Target SDK Version:** 34 (Ensures compatibility with the latest Android features)
- **Compile SDK Version:** Latest stable version

Application Components

- Activities handle the UI and user interactions

User Interface (UI)

- Used to design and structure the app's screens using XML

Testing & Debugging

- JUnit and Mockito are used to test individual components of the app
- Espresso is used for automated testing of the user interface
- Logcat helps in debugging and identifying issues

Open-Source Approach

The LebitsoApp is an open-source project, with all source code available on the University of the Free State's GitHub page. The code is thoroughly documented, making it easy for developers and aspiring software engineers, especially within the Deaf community, to contribute. One of the primary reasons for making the app open source is to encourage individuals in the Deaf community to participate in software development and HLT development, a field in which they are currently underrepresented. By involving the Deaf community in the creation and improvement of the LebitsoApp, the project aims to inspire more Deaf individuals to pursue careers in Computer Science, Computer vision, data science, data analytics, and other STEM (science, technology, engineering, and mathematics) fields to innovate in this exciting domain. Ultimately, the goal is to create an application built both with and for the Deaf community.

How does the LebitsoApp do something different from that of other platforms?

LebitsoApp explores Machine Learning-driven text-to-avatar translation, data management, resources, and a community of practice that is inclusive, making sign language more accessible in new and innovative ways. What the project is trying to achieve is to make the LebitsoApp truly unique in its development approach, created *with* the Deaf community rather than just *for* them, ensuring it meets their needs. Unlike many closed systems, LebitsoApp is open source, allowing Deaf individuals and developers to contribute to its growth. It also prioritises accessibility by presenting research in SASL videos rather than relying solely on written explanations, making it more user-friendly for those with lower literacy levels. Additionally, while many platforms require constant internet access, LebitsoApp stores videos on the device, ensuring users can engage with content anytime, anywhere.

Bridging the Gap Between the Deaf Community, Research, and Human Language Technology

The LebitsoApp aims to serve as an accessible platform where members of the Deaf community can stay informed about ongoing and future research efforts in HLT. Research updates will be presented through SASL videos, making the information easily digestible.

Features of the LebitsoApp

Full-Screen Video Player

- Let's users watch videos in full screen for a better experience.
- Allows pausing, rewinding, fast-forwarding and scrubbing through the video.
- Includes subtitles or captions for translation of SASL.

Feedback Forms for Every Lesson

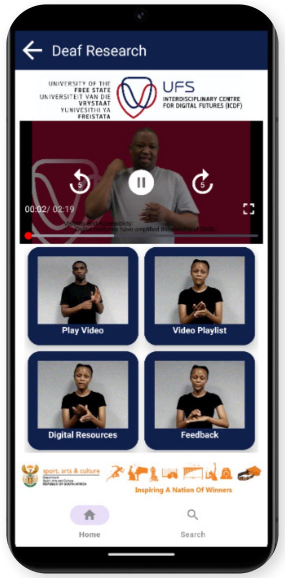
- Users can share their thoughts for each activity of the application to improve usability of the application.

Offline Access and Syncing

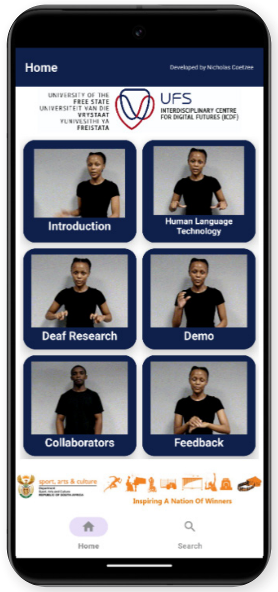
- Users can watch videos without internet access.
- Helps users keep interacting with the app even in an offline environment.

Easy-to-Use Interface

- Simple design for easy navigation.
- SASL GIFs for button translation.



Accessibility and User Experience (UX)



The LebitsoApp places a strong emphasis on accessibility, ensuring that users of all literacy levels can engage with its content effectively providing they can understand SASL. The app's user experience has been designed to accommodate Deaf users while also making it accessible to non-SASL users through subtitles and intuitive navigation. Ongoing research and user testing will help refine these features to ensure they provide maximum benefit to users.

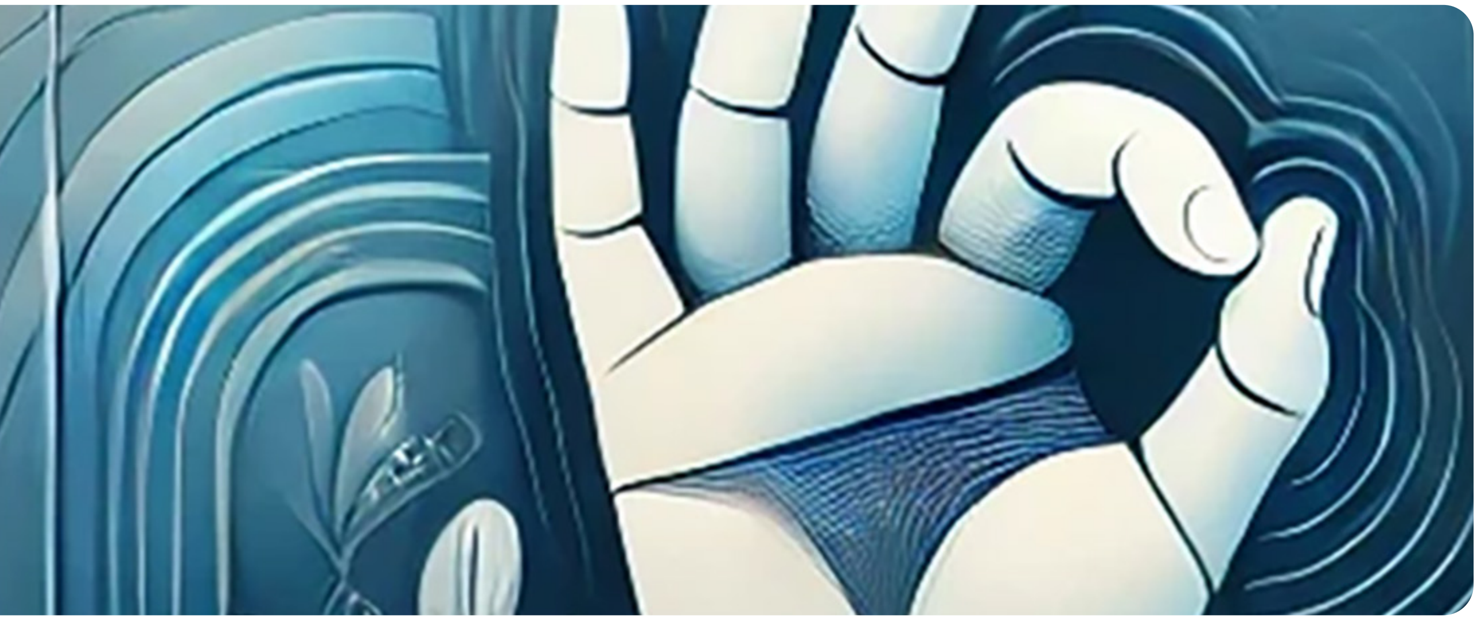
Key accessibility features include:

Important icons are accompanied by animated GIFs depicting the corresponding SASL signs, making navigation more intuitive for Deaf users. Each section of the app features explanatory videos in SASL, ensuring that information is conveyed in a manner that is both familiar and accessible to the Deaf community. The app is designed with a clean and uncluttered interface, presenting only the most essential information on each screen.

Videos are stored on the device to minimise the need for continuous data usage, allowing users to access content even without an internet connection. The LebitsoApp integrates ExoPlayer, which allows for customised video controls, including full-screen mode, forward and rewind functions, and improved usability tailored to Deaf users.

Usability Testing for LebitsoApp

For LebitsoApp, a usability test was conducted in a controlled environment, supported by user questionnaires. This approach allowed direct feedback to be collected from users and observe their interactions with the app. The goal was to evaluate if the LebitsoApp is user friendly and if the purpose of the app is successfully portrayed to the users. We used different methods to assess how effectively LebitsoApp meets users' needs. Our focus was on understanding how users interact with the app and whether they find the available materials engaging and useful.



Steps in the Usability Test

1. Pre-Test Questionnaire

Before testing, we asked participants to complete a short questionnaire. This helped us understand their background, experience with mobile apps, and initial expectations of LebitsoApp.

2. Usability Test Process

Participants were asked to complete specific tasks within the app while observers recorded important data such as:

- How easily users navigated the app.
- Users experience watching videos and reading content.
- Any difficulties or errors they encountered.

Who Participated?

We tested the app with 5–10 participants, which included staff from the University of the Free State. Due to time and resource limitations, we used convenience sampling. Participants were 18 years or older and had prior experience using mobile apps. Each participant provided informed consent before taking part.

Usability Test Protocol

The usability test followed these steps:

- **INTRODUCTION:** Explained the purpose of the test.
- **CONSENT FORMS:** Participants signed consent forms to confirm their voluntary participation.
- **PRE-TEST QUESTIONNAIRE:** A 5-minute survey collected background information.
- **EXPLORATORY TASK:** Each participant explored the application and went through the provided questionnaire.
- **DATA USE EXPLANATION:** We informed participants how their feedback would be used to improve the app.
- **CONCLUSION:** We wrapped up the session.

Post-Test Discussion

After completing the usability test, participants participated in an open forum where they could voice other concerns not mentioned on the questionnaire. The feedback collected helped us refine LebitsoApp to ensure a smoother and more engaging user experience.

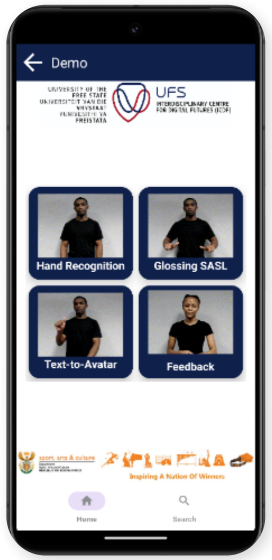
Further Testing and the Future

Further Testing will need to be done with users who are willing to take part in the applications development. Larger sample Sizes and participants specifically from the Deaf community will need to be incorporated to ensure that the app is meeting its purpose and proposed inclusiveness.

N COETZEE

Walkthrough of the LebitsoApp

Upon launching the LebitsoApp, users are greeted with an introductory sequence consisting of three informational slides that provide background on the application. Users can navigate these slides by swiping left or right or using the next and back buttons. If they are already familiar with the app, they can skip this section entirely.



The main menu serves as the navigation hub, featuring animated GIFs that translate the button text into SASL. The home screen layout is organised in a grid format for easy access to different sections. The Introduction activity presents an overview video explaining the app's purpose and functionality. Since all videos are stored on the device, users can access them without requiring an internet connection.

The Human Language Technology section provides an in-depth look at how technology can benefit the Deaf community. Users can explore online resources such as GitHub repositories and Media CMS, where research and literature on Machine Learning and sign language translation are available. A feedback button is provided for users to share suggestions or report issues. The Deaf Research section follows a similar format as the previous Human Language Technology activity but focuses specifically on Deaf-related research.

Finally, the Demo activity contains interactive web apps showcasing advancements in Human Language Technology. Users can experiment with tools (these are neatly packaged as web applications) such as text-to-avatar translation, which utilises Machine Learning algorithms to convert text input into a visual sign language representation. The focus is on making text-to-avatar translation a viable tool for Deaf communication.

Good Money Habits

Good Money is an Android app designed to educate entrepreneurs, particularly those in small businesses or the informal sector, on effective money management. The link to HLT is that whenever there is an innovation, entrepreneurial skills also include the financing and financial decision making around a small business. It helps users track and record their finances, make informed financial decisions, and ultimately grow their profits. By promoting better financial habits, the app empowers businesses to scale sustainably and achieve long-term success.

How the App Works

The Good Money Habits app is designed to help entrepreneurs develop strong financial management skills. It provides structured guidance on handling money effectively, ensuring better financial decision-making and business growth. The app begins with an introductory screen that explains its purpose, content structure, and how users should interact with the material for maximum benefit.

The core content is delivered through a series of short videos, each lasting 3 to 5 minutes, organised into four key parts:

1. **PART 1:** **Basics:** Why Good Money Habits (4 videos)
2. **PART 2:** Counting and recording money **INFLOWS** (Revenue) (4 videos)
3. **PART 3:** Counting and recording money **Outflows** (3 videos)
4. **PART 4:** Counting and recording **PROFIT** & The Risk of Customer Credit (4 videos)

For the best learning experience, users are strongly encouraged to watch the videos in order, as each build on key insights from the previous one. Additionally, spacing out the lessons across multiple sessions is recommended to enhance retention and practical application.

Basic Features

Full-Screen Video Player

1. Let's users watch videos in full screen for a better experience.
2. Allows pausing, rewinding, and fast-forwarding.
3. Includes subtitles or captions for better understanding.

Feedback Forms for Every Lesson

1. Users can share their thoughts after each lesson.
2. Helps improve the lessons based on feedback.
3. Users can rate lessons to show how helpful they are.

Offline Access and Syncing

1. Users can watch videos and lessons without the internet.
2. The app updates progress when the user is back online.
3. Helps users keep learning anytime, anywhere.

Easy-to-Use Interface

1. Simple design for easy navigation.
2. Lessons are well-organised for a smooth learning experience.

Additional Features

- **INTERACTIVE LEARNING & REWARDS**

1. Users can answer feedback questions for every lesson and apply knowledge.
2. Earn rewards for completing a series.
3. Motivational messages to keep users engaged.

- **GAMIFICATION & BADGES**

1. Users earn badges when they reach milestones.
2. Makes learning fun and rewarding.
3. Users can track their progress easily.

- **COMPLETION CERTIFICATE**

1. Users get a certificate after finishing all lessons and feedback questions.
2. Encourages users to complete the whole course.

Limitations

Hardware and Software Limitations

During the development of the Good Money Habits (GMH) app, it was ensured that it runs smoothly, performs well, and stays secure on different types of mobile devices.

Several key factors must be considered:

1. Mobile Operating System Compatibility

- The GMH app is designed for Android devices, as Java is a major programming language for Android development.
- Since Android comes in many versions, the app must be compatible with a wide range of devices, from older models to the latest smartphones.

2. Hardware Limitations

- The app should work well on both budget smartphones and high-end devices.
- It must be optimised for different hardware specifications, such as:
  - CPU (PROCESSING POWER) – Ensuring smooth video playback and fast performance.
  - MEMORY (RAM) – Making sure the app runs efficiently without slowing down the device.
  - STORAGE – Keeping the app lightweight so it doesn't take up too much space.
  - BATTERY USAGE – Reducing power consumption to avoid draining the battery quickly.

3. Network Connectivity

- The app will allow users to watch videos offline, so they can learn even without an internet connection.
- When online, it will sync progress and update content without using too much data.
- If there is no network connection, the app will show a message to inform users and suggest offline options.

Evaluation

Evaluating GMH is essential for determining how well the App meets its objectives. The purpose of an evaluation is to systematically collect and analyse data regarding an app's features, activities, and outcomes to identify areas of strength and those needing improvement. This process leads to better-informed decisions regarding further development and enhances the app's functionality and user experience.

Evaluation Techniques

Evaluating GMH involves three main approaches: evaluation in controlled settings with users, evaluation in natural settings involving users, and evaluation in settings that do not directly involve users.

- CONTROLLED SETTINGS DIRECTLY INVOLVING USERS: In this approach, users interact with GMH while being closely monitored. Evaluation techniques like user tests and experiments are used to control the participants' actions and environment, ensuring that external factors don't impact the results.
- NATURAL SETTINGS INVOLVING USERS: In this method, evaluations are conducted in the real-world context in which the app is used, such that business owners interacts with the App to identify if they find the content of the app useful and it helps them learn valuable information needed to grow their business.
- SETTINGS NOT INVOLVING USERS: This approach involves evaluating the app without direct user involvement. Methods like heuristic evaluation and walkthroughs allow evaluators to simulate user scenarios and predict how users would interact with GMH.

For GMH, a usability test conducted in a controlled setting was chosen, complemented by questionnaires. This technique was selected to gather direct feedback from users and observe their interactions with the app, helping assess how well GMH supports business owners improve to handle their money well. The evaluation strategy employed for GMH involves using several techniques to collect detailed information about how effectively the app meets the users' needs. This strategy focuses on evaluating how well GMH enables interact with the content and see if they users are interested in the lessons available on the App. Before conducting the usability tests, a pre-test questionnaire was distributed to gather demographic information and understand participants' familiarity with apps. The pre-test questionnaire also captured initial feedback and expectations. The usability test was conducted to observe participants using GMH to complete specific tasks, such as, navigating through the App, viewing the videos, reading the content. During the test, observers recorded both qualitative and quantitative data, such as task completion times, error rates, and user satisfaction levels.

- FORMATIVE USABILITY: This evaluation method was applied iteratively throughout the development process to identify usability issues early and implement improvements over time.
- SUMMATIVE USABILITY: A summative usability test was conducted after development to assess how well GMH met its predetermined goals, particularly the ease of navigating through the app.

The study involved a sample of nine participants, consisting of UFS staff members, and business owners from the Free State region. Convenience sampling was used due to time and resource constraints. Participants were aged 18 and above, and the selection criteria included prior experience with other apps used in their everyday life. Informed consent was obtained from all participants.

The usability test protocol followed these steps:

1. Introduction to the purpose of the evaluation.
2. Distribution of informed consent forms.
3. Completion of pre-test questionnaires (10 minutes).
4. Task assignments and guidance for thinking aloud during task execution (20 minutes per participant).
5. Completion of post-test questionnaires (15 minutes).
6. Discussion of how collected data would be used.
7. Conclusion and expressions of gratitude.

Post-Test Questionnaire

The post-test questionnaire aimed to gather participants' opinions and feedback on GMH's usability, including its Interaction and rewards, Gamification badges, and certification of completion features.

N NKUNA

BIBLIOGRAPHY

1. Marschark, M., Tang, G. and Knoors, H. eds., 2014. *Bilingualism and bilingual deaf education.* Oxford University Press.
2. Baker-Shenk, C.L. and Padden, C., 1979. *American Sign Language: A look at its history, structure, and community.* TJ Pub Incorporated.
3. Leigh, I., 2009. *A lens on deaf identities.* Oxford University Press.
4. Heinrich, S. and Wermter, S., 2018. Interactive natural language acquisition in a multi-modal recurrent neural architecture. *Connection Science, 30*(1), pp.99-133.
5. Camurri, A. and Volpe, G. eds., 2004. *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers* (Vol. 2915). Springer Science & Business Media.
6. Karam, M., 2006. *A framework for research and design of gesture-based human-computer interactions* (Doctoral dissertation, University of Southampton).
7. Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T. and Vogler, C., 2019, October. Sign language recognition, generation, and translation: An interdisciplinary perspective. In *Proceedings of the 21st international ACM SIGACCESS conference on computers and accessibility* (pp. 16-31).
8. Kumar Attar, R., Goyal, V. and Goyal, L., 2023. State of the art of automation in Sign Language: a systematic review. *ACM Transactions on Asian and Low-Resource Language Information Processing, 22*(4), pp.1-80.
9. Gillespie-Lynch, K., Greenfield, P.M., Feng, Y., Savage-Rumbaugh, S. and Lyn, H., 2013. A cross-species study of gesture and its role in symbolic development: implications for the gestural theory of language evolution. *Frontiers in psychology, 4,* p.160.
10. Crasborn, O., 2010. What Does" Informed Consent" Mean in the Internet Age?: Publishing Sign Language Corpora as Open Content. *Sign Language Studies, 10*(2), pp.276-290.
11. Napier, J., Leeson, L., Napier, J. and Leeson, L., 2016. *Sign language in action* (pp. 50-84). Palgrave Macmillan UK.
12. Hodge, G., 2014. *Patterns from a signed language corpus: Clause-like units in Auslan (Australian sign language)* (Doctoral dissertation, Macquarie University).
13. Ananthanarayana, T., Srivastava, P., Chintha, A., Santha, A., Landy, B., Panaro, J., Webster, A., Kotecha, N., Sah, S., Sarchet, T. and Ptucha, R., 2021. Deep learning methods for sign language translation. *ACM Transactions on Accessible Computing (TACCESS), 14*(4), pp.1-30.
14. Al-Qurishi, M., Khalid, T. and Souissi, R., 2021. Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. *IEEE Access, 9,* pp.126917-126951.

15. **Farooq, U., Rahim, M.S.M., Sabir, N., Hussain, A. and Abid, A.,** 2021. Advances in machine translation for sign language: approaches, limitations, and challenges. ***Neural Computing and Applications, 33***(21), pp.14357-14399.

16. **Natarajan, B., Rajalakshmi, E., Elakkiya, R., Kotecha, K., Abraham, A., Gabralla, L.A. and Subramaniyaswamy, V.,** 2022. Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation. ***IEEE Access, 10,*** pp.104358-104374.

17. **Amin, M., Hefny, H. and Ammar, M.,** 2021. Sign language gloss translation using deep learning models. ***International Journal of Advanced Computer Science and Applications, 12***(11).

18. **Schick, B., Marschark, M. and Spencer, P.E. eds.,** 2005. ***Advances in the sign language development of deaf children.*** Oxford University Press.

South African Sign Language (SASL) is undergoing a transformative shift through the integration of **technology, artificial intelligence, and open-source collaboration**. This report explores the development of a **comprehensive SASL data-sharing framework**, emphasising the role of the **LebitsoApp** as a central hub for **data collection, research, and technological innovation**. The project's vision is to foster inclusivity, preserve linguistic diversity, and propel SASL into the digital age.

A key component of this initiative is the **LebitsoApp**, a pioneering platform designed to consolidate **code, raw and annotated data, AI models, and linguistic resources** in a single accessible space. Through **open-source principles**, the app ensures that researchers, developers, and the Deaf community can collectively contribute to the refinement of SASL digital tools. User experience (UX) design is a crucial element, as the app continues to evolve based on community feedback to ensure **Deaf-friendly accessibility**.

The research also involves an extensive **comparative analysis of ASL-SASL**, using a **15,205-video dataset** to determine linguistic similarities and distinctions. Through **structured classification and validation**, the project has established a **high-confidence dataset** for **AI-powered SASL recognition, machine learning applications, and sign language standardisation**.

To support the **visual representation of SASL**, a series of **artistic and modern digital images** were generated. These images capture the essence of **SASL communication**, employing a **blue-themed aesthetic with abstract designs, thin lines, and 3D-rendered hands**. The visual elements not only enhance engagement but also symbolise **the intersection of language, culture, and technology.**

This report demonstrates how **data-driven research, AI integration, and artistic representation** can work together to advance **SASL accessibility, education, and recognition**. By ensuring **open collaboration, ethical data sharing, and continuous innovation**, this initiative paves the way for a more **inclusive and technologically advanced** future for SASL users and researchers alike.