INTEROPERABLE DATA SHARING:

South African Sign Language Computer Readable Annotation for Artificial Intelligence

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 | PATRICK SIBANDA | NHLANHLA SIMELANE

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

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 Importance of a Computer-Readable Annotation System for South African Sign Language (SASL)

This guide outlines the justification for our investigation of a novel annotation system for Human Language Technology, specific to South African Sign Language (SASL). We needed these systems to assist us in processing the data and also to store the data with context. This allowed us to build much more context to all the models we develop. Machine learning and artificial intelligence alone is not enough without human input, human context, and human conceptualisation. All of our annotation-based research directed us to capture, manage and govern our data in a much more coherent manner for future research. The value of the human language technology project, alongside the Department of Sport, Arts, and Culture, has been a fundamental contribution in this domain, and field, for both the development and visibility of SASL.

DR HMVE COMBRINK

CONTENTS

Introduction to the text

Language is the foundation of human communication, enabling the exchange of ideas, emotions, and knowledge. While spoken languages have been extensively documented and analysed using computer-readable annotations, signed languages, which serve as the primary means of communication for Deaf communities worldwide, have not been as thoroughly represented in computational models. The growing field of computational linguistics has enabled vast progress in machine-readable annotations for various spoken languages, but signed languages require a fundamentally different approach due to their unique multimodal nature, which includes hand configurations, facial expressions, and movement sequences.

The development of a mathematical notation system for South African Sign Language (SASL) represents a significant leap forward in making signed languages machine-readable and systematically structured for digital processing. Although various computational annotation frameworks have been created for other sign languages, SASL remains largely underrepresented in this field. This gap highlights the urgent need for a specialised notation system that can accurately capture the nuances of SASL in a structured and interpretable format.

In this introduction, we explore the significance of developing a computer-readable annotation system for signed languages, focusing on the advancements made in computational sign language annotation and the limitations of current approaches. We also discuss why it is crucial to develop a specialized annotation framework for SASL and examine some of the challenges associated with implementing such a system.

Why Computer-Readable Annotation Systems Matter for Signed Languages

The documentation of signed languages is inherently more complex than that of spoken languages due to the reliance on visual and spatial elements rather than phonetic sequences. Unlike spoken languages, which can be transcribed using phonetic alphabets such as the International Phonetic Alphabet (IPA), signed languages require a visual-spatial representation to capture these dynamic features. ***Computer-readable annotation systems enable sign language to be structured in a way that can be used for:***

1. **MACHINE LEARNING APPLICATIONS:** Training artificial intelligence (AI) models for sign language recognition, translation, and synthesis.
2. **LINGUISTIC RESEARCH:** Studying sign language syntax, morphology, and semantics systematically.
3. **EDUCATION AND ACCESSIBILITY:** Developing digital tools to aid in teaching sign language and improving accessibility for the Deaf community.
4. **DATA STORAGE AND RETRIEVAL:** Creating databases of signed language corpora that can be used for research, learning, and preservation.

Without a standardised computer readable annotation system, computational models struggle to process signed languages effectively. Developing a mathematical framework that structures SASL in a machine-readable manner ensures that its features are preserved, analysed, and used in AI-driven applications.

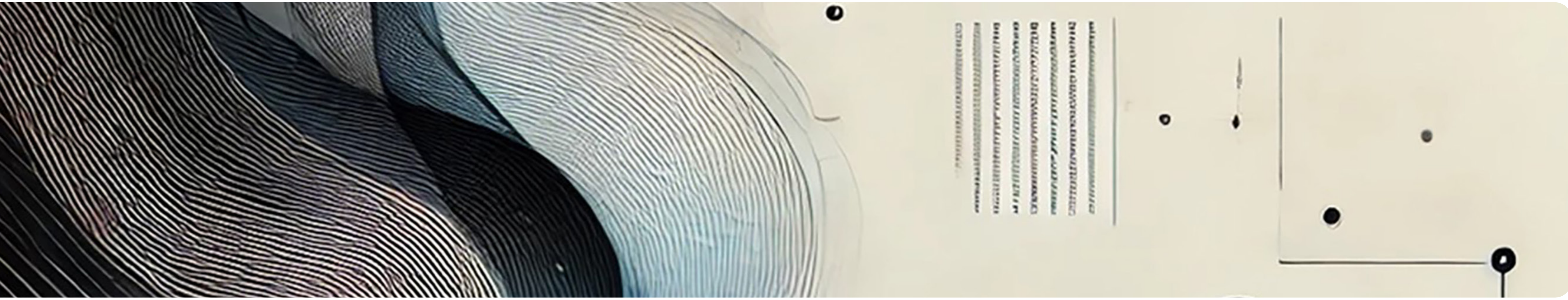
Existing Approaches to Computer-Readable Annotation for Signed Languages

Several sign language annotation systems have been developed over the years to facilitate linguistic analysis and machine processing of sign languages. ***Some of the most notable include:***

1. **STOKOE NOTATION:** One of the earliest sign language transcription systems, developed by William Stokoe in the 1960s. It represents handshape, location, and movement but lacks the ability to encode complex multimodal aspects like facial expressions and hand interactions.
2. **HAMBURG NOTATION SYSTEM (HAMNOSYS):** A phonetic transcription system developed at the University of Hamburg, which provides a detailed representation of sign language phonology, including movement, orientation, and location.
3. **SIGNWRITING:** A visual writing system that captures the movements and spatial aspects of signed languages in a way that resembles written scripts. While useful for literacy and teaching, it lacks the structured mathematical form needed for computational applications.
4. **ELAN (EUDICO LINGUISTIC ANNOTATOR):** A digital annotation tool widely used in linguistic research for annotating video recordings of sign languages. While useful for research, ELAN does not provide a compact, mathematical, or AI-ready notation.

Although these systems have significantly contributed to sign language research, they are often not designed with computational scalability in mind. Many existing sign language annotation frameworks focus on linguistic documentation rather than AI integration, making it difficult to train machine learning models directly from annotated data.

DR HMVE COMBRINK



The Need for a Computational Notation System for SASL

SASL has historically been underrepresented in sign language computational research. Existing databases and AI-driven sign language translation models primarily focus on American Sign Language (ASL) and British Sign Language (BSL), leaving SASL with limited resources for digital processing and machine learning integration.

Developing a structured and computationally viable notation for SASL is essential for several reasons:

1. **PRESERVATION AND STANDARDIZATION:**
A formalized mathematical notation system ensures that SASL is documented consistently and can be preserved for future research.

2. **AI AND MACHINE LEARNING INTEGRATION:**
A structured annotation format enables training AI models for SASL recognition and synthesis, improving accessibility and communication tools.

3. **EDUCATION AND ACCESSIBILITY:**
Standardised digital representation of SASL can support educational tools, helping both Deaf and hearing individuals learn the language more effectively.

4. **LINGUISTIC ANALYSIS:**
Providing a computationally structured dataset facilitates linguistic studies of SASL, enabling comparisons with other signed languages.

Pitfalls and Challenges of This Approach

While this approach offers a promising path forward, several challenges must be considered:

1. **COMPLEXITY OF SIGN LANGUAGE REPRESENTATION:**
Signed languages are inherently multimodal, involving facial expressions, eye gaze, and spatial positioning, which are difficult to capture in a purely mathematical notation.

2. **VARIABILITY IN SIGNING:**
Signers may have different styles, speeds, and variations in executing the same sign, making standardisation challenging.

3. **COMPUTATIONAL BURDEN:**
A detailed annotation system may require large datasets and computational resources for AI training, making scalability an issue.

4. **COMMUNITY ADOPTION:**
Any computational notation system must be accepted and validated by the Deaf community and linguistic researchers to ensure accuracy and usability.

5. **LACK OF LARGE ANNOTATED DATASETS:**
SASL currently lacks extensive corpora for training AI models, which means that an annotation system alone is not enough, data collection must also be prioritised.

Developing a computational annotation system for SASL is an essential step toward making SASL machine-readable and accessible for research, AI applications, and education. While existing sign language annotation systems have laid a strong foundation, they have yet to be fully adapted for SASL in a way that prioritises computational scalability while maintaining linguistic integrity. However, addressing the challenges of multimodal representation, signer variability, and dataset availability will be crucial to ensuring the system's success. By investing in computational annotation frameworks for SASL, we pave the way for greater accessibility, research opportunities, and technological advancements that benefit the Deaf community and beyond.

DR HMVE COMBRINK

Machine Learning Applications: Training Artificial Intelligence (AI) Models for Sign Language Recognition, Translation, and Synthesis

Machine learning (ML) has revolutionized numerous fields by enabling computers to learn from data and make predictions without explicit programming. In the context of signed languages, ML plays a crucial role in the development of automated sign language recognition, translation, and synthesis. However, the unique multimodal nature of sign languages presents significant challenges for ML-based systems. Unlike spoken languages, which primarily rely on auditory signals, sign languages involve a combination of hand movements, facial expressions, and spatial positioning. This complexity requires a well-structured approach to machine learning that can effectively process and interpret signed language data.

This section explores the critical components necessary for training AI models for sign language applications. We begin by discussing the types of ML approaches that are most relevant for sign language processing, followed by the importance of large-scale training data. Finally, we emphasize the necessity of a structured annotation system to ensure accurate data labelling before any model training can take place.

Some of the types of Machine Learning models for Signed Languages

To develop AI models capable of recognising and translating signed languages, various machine learning techniques must be used. Developing AI-driven systems for sign language recognition, translation, and synthesis requires specialized ML techniques that can interpret these complex patterns effectively. Three of these approaches in this regard are Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTMs), and Transformer-based models using different types of attention mechanisms.

Convolutional Neural Networks (CNNs) for Sign Language Processing

Convolutional Neural Networks (CNNs) have been widely used in the field of computer vision, particularly in image and video recognition tasks. Their application in sign language recognition is crucial as they enable AI models to detect and interpret essential visual elements of signed communication. CNNs function by processing images through multiple layers, extracting features such as edges, shapes, textures, and patterns. When applied to sign language, CNNs are particularly effective in detecting hand shapes, facial expressions, and movement trajectories from video frames.

A typical CNN model for sign language recognition begins by analysing a series of video frames that capture a signer's gestures. The first convolutional layers detect basic image features such as contours and edges of the hands. As data passes through deeper layers, the model starts recognizing complex structures like individual handshapes, the orientation of the palm, and the relative position of the hands to the body. Pooling layers help reduce computational complexity by summarizing key features, ensuring the model remains efficient while maintaining accuracy.

One of the key advantages of CNNs is their ability to generalize patterns across different signers, making them highly adaptable to variations in signing style, hand size, and lighting conditions. This robustness is particularly important in real-world applications where sign language recognition systems must function under diverse conditions. Additionally, CNNs can be combined with pose estimation techniques, which track key points of the hands and face, allowing for an even more refined understanding of sign language expressions. However, despite their effectiveness, CNNs alone are insufficient for full sign language interpretation, as they struggle with temporal dependencies, i.e., how a sign evolves over time. This is where Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks become essential.

Recurrent Neural Networks (RNNs) and LSTMs for Sequential Sign Language Processing

Sign language communication is inherently sequential, meaning that a series of movements, rather than isolated static gestures, convey meaning. While CNNs excel at recognising individual frames, they do not account for how one frame transitions into the next over time. Recurrent Neural Networks (RNNs) and their improved variant, Long Short-Term Memory (LSTM) networks, are designed specifically to process sequential data, making them essential for sign language translation and synthesis.

RNNs work by maintaining a memory of previous inputs, which allows them to analyse how gestures unfold over time. This capability makes them highly effective in processing sign sequences, such as fingerspelling, where each letter of a word is signed in succession, or complex gestures that change meaning depending on the context of preceding signs. However, standard RNNs suffer from the vanishing gradient problem, which limits their ability to retain long-term dependencies. This limitation is particularly problematic in sign language, where meaning often depends on gestures made several seconds apart.

LSTMs address this issue by introducing gated mechanisms that control how much information is retained or forgotten at each time step. This allows the model to maintain context across longer sequences, improving its ability to recognise and interpret full sentences in sign language. For instance, an LSTM model trained on sign language data can distinguish between a sequence of signs that form a single phrase versus isolated, unrelated signs. By integrating LSTMs with CNNs, AI systems can process both the spatial features of a sign (handshape, orientation, facial expressions) and its temporal evolution, creating a more holistic understanding of sign language. However, a lot of research is needed in this domain to make the technology functional, also, there is a need to contextually train models to make this possible.

Another advantage of LSTMs is their ability to handle variations in signing speed. Unlike spoken language, where words are typically produced at a steady pace, sign language can be signed at different speeds depending on the signer and context. LSTMs allow AI models to adapt to these variations, making them more effective in real-world applications, such as live sign language translation. As promising as this sounds, another layer of complexity is needed to contextually understand how signed languages can benefit from AI, and research in this domain, and that is with transformers and other attention mechanisms.

Transformers and Attention Mechanisms for Sign Language Translation

While CNNs and LSTMs have significantly improved sign language recognition, the most cutting-edge advancements in the field have come from Transformer-based models. Originally developed for natural language processing (NLP), Transformer models such as BERT and GPT have expanded a lot of research in AI-driven language understanding by introducing self-attention mechanisms that allow models to focus on the most relevant parts of an input sequence.

Transformers are particularly beneficial for sign language translation because they enable AI models to consider the entire context of a sentence rather than processing signs in isolation. The self-attention mechanism allows the model to determine which gestures and facial expressions are most relevant at any given moment, improving the accuracy of translations. This is especially important for sign languages, where meaning is often conveyed through non-manual signals such as eyebrow raises, head tilts, and mouth movements. The field is not yet at the point to perform these tasks fully effectively yet, but it is at the stage where meaningful research to expand the domain can be used to test ideas effectively.

How it works is a Transformer model trained on sign language data operates by encoding video input into numerical representations that capture both spatial and temporal features. Instead of analysing each sign in a linear sequence like LSTMs, Transformers can process an entire sentence simultaneously, identifying dependencies between signs regardless of their order. This parallel processing capability makes Transformers highly efficient and accurate in sign language translation tasks. However, to do this requires a lor of processing power, and a lot of ML training.

Another significant advantage of Transformers is their ability to generate high-quality sign language synthesis. In sign language avatars, AI-generated digital representations of signers, Transformer models can be used to predict natural, human-like signing gestures based on text input. This has the potential to create accessibility tools by enabling automatic real-time sign language interpretation for television broadcasts, online content, and digital communication platforms. Again, the field is not yet at the point where this knowledge can be seamlessly applied to all signed languages.

Despite their powerful capabilities, Transformer models require large amounts of high-quality training data to perform effectively, also, a lot of computational power to train models. This presents a major challenge for sign language AI applications, as annotated sign language datasets are still relatively scarce compared to text-based datasets. Additionally, the complexity of sign language means that models must be trained to recognise subtle variations in gestures, requiring robust annotation systems to ensure accurate labelling. For this, a lot of sociolinguistic research is needed – to the extent that more resources than spoken and written languages need.

DR HMVE COMBRINK, P KECHE

The Need for Large-Scale Training Data

One of the biggest challenges in developing AI models for sign language recognition and translation is the requirement for vast amounts of training data. Unlike spoken languages, where millions of text-based examples are readily available, signed language datasets are limited, making it difficult to train robust models.

The effectiveness of ML models depends on diverse, high-quality training data that accurately represents the variations, complexity, and multimodal nature of sign language. However, several obstacles hinder the development of such datasets. This section explores three primary challenges: the scarcity of publicly available sign language datasets, the variability in signing styles across different signers and regions, and the inherent multimodal complexity of sign languages that require detailed annotations.

DR HMVE COMBRINK

Limited Public Datasets

One of the most pressing challenges in sign language machine learning is the limited availability of large, annotated datasets. Unlike spoken languages, which benefit from extensive text and speech corpora, sign language lacks a comparable wealth of data. The reasons for this scarcity are multifaceted and deeply rooted in historical, technological, and societal factors.

Lack of Digitized Sign Language Resources

Historically, sign languages have not been systematically documented in digital formats at the same scale as spoken languages. Written text has been the dominant mode of communication for linguistic data storage, but sign languages, being visual-gestural in nature, do not have a widely adopted written form. As a result, early documentation efforts focused on written descriptions and photographs rather than video-based datasets that could be used for ML training.

High Costs and Resource Intensiveness of Data Collection

Creating a high-quality sign language dataset requires substantial resources, including video recording equipment, professional sign language interpreters, and linguistic experts to annotate the data accurately. Unlike speech recognition, which can leverage vast amounts of freely available text and audio from sources such as books, podcasts, and news articles, sign language datasets must be carefully curated through costly and time-consuming video recordings. These recordings then require labour-intensive manual annotation, further increasing the cost and limiting the availability of large-scale datasets.

Privacy and Ethical Concerns

Another barrier to large-scale sign language data collection is the issue of privacy. Unlike written text, which can be anonymized easily, sign language data consists of video recordings of real people, raising concerns about consent, privacy, and ethical data usage. Many members of the Deaf community may be hesitant to have their images stored and analysed by AI systems, especially when the purpose and security of the data are not clearly defined. Establishing strong ethical guidelines for data collection and storage is critical for building trust and encouraging broader participation in dataset creation efforts.

Limited Publicly Available Datasets

While some academic and commercial projects have created sign language datasets, many remain restricted due to licensing and proprietary constraints. Publicly available datasets for sign language research are often small, covering only basic vocabulary and lacking the depth needed for robust ML model training. Additionally, different projects use inconsistent annotation formats, making it difficult to integrate multiple datasets into a unified model.

Variability in Signing Styles

Sign language is not a monolithic language; it varies significantly across different regions, communities, and individual signers. This variability introduces several challenges when collecting data for ML applications.

Regional and National Variations

Sign languages are not universal—each country or region has its own distinct sign language. Even within a single sign language, dialectal differences exist based on geography, culture, and history. For example, American Sign Language (ASL) and British Sign Language (BSL) are completely different languages with unique vocabularies and grammar rules. South African Sign Language (SASL) also exhibits significant regional variations. Training a single ML model that can accommodate these differences requires data that represents a broad range of signers and dialects, which is difficult to compile.

Individual Signing Styles

Even within a specific sign language, individual signers exhibit variations in signing speed, expressiveness, and articulation. Some signers may use highly fluid, expressive signing, incorporating non-manual markers (e.g., facial expressions and head movements), while others may use more compact, minimal gestures. ML models need to be robust enough to handle this variability without misinterpreting signs. This requires training on large datasets that include diverse signing styles, which is currently a major challenge due to the limited availability of such datasets.

Different Signing Contexts

Sign language use varies depending on the context. Formal presentations, casual conversations, and educational settings all influence signing style. For example, signing in a legal or medical setting may be more precise and formal, while signing in a social setting may be more relaxed and involve slang or regional gestures. ML models need to be trained on diverse contexts to ensure accurate recognition and translation across different real-world applications.

Impact of Signer Identity on Model Accuracy

Research has shown that AI models trained on a limited subset of signers may struggle when exposed to new individuals. Factors such as hand size, body proportions, and signing speed can all impact recognition accuracy. To build inclusive and effective ML models, training datasets must represent signers of different ages, genders, ethnic backgrounds, and signing abilities. However, collecting such diverse data is a major logistical and ethical challenge.

DR HMVE COMBRINK

Multimodal Complexity of Sign Language Data

Unlike spoken languages, which primarily rely on phonetic sounds, sign languages convey meaning through a combination of multiple channels, including hand movements, facial expressions, body posture, and spatial positioning. Capturing and processing this complexity presents significant technical and computational challenges.

1. The Role of Facial Expressions and Non-Manual Markers

Facial expressions play a crucial role in sign language communication. They are not simply supplementary but integral to meaning. For example, in ASL and SASL, eyebrow movements can indicate questions or negation, while mouth shapes can modify the meaning of a sign. ML models must be trained to recognize and interpret these subtle cues, which adds an additional layer of complexity to the data annotation and processing pipeline.

2. Challenges in Capturing Three-Dimensional Movement

Unlike spoken words, which are linear and sequential, sign language involves three-dimensional spatial movement. Hand gestures move in various directions, interact with different parts of the body, and change in orientation. Capturing these nuances requires high-quality video recordings and advanced computer vision techniques, such as 3D pose estimation and depth analysis. However, many current datasets rely on standard 2D video recordings, limiting the accuracy of ML models in capturing full sign articulation.

3. Synchronization of Multiple Sign Language Components

A single sign often involves coordinated movements of both hands, along with facial expressions and body posture. For ML models to accurately interpret a sign, they must analyze all these elements in synchrony. This requires precise temporal alignment of video frames, pose data, and facial tracking, which significantly increases the computational complexity of sign language recognition models. Ensuring accurate synchronization across different modalities remains a key challenge in dataset development.

4. Annotation Complexity and Standardization Issues

To train ML models effectively, sign language datasets must be annotated with detailed information about hand shapes, movement trajectories, and facial expressions. However, there is no universally accepted annotation standard for sign language data, leading to inconsistencies between datasets. Some annotation systems focus only on hand configurations, while others include full-body tracking. Developing a standardized annotation framework that captures all relevant aspects of sign language while remaining computationally efficient is essential for advancing AI applications in this field.

The collection of high-qual ty sign language data for ML applications is a challenging but essential task. Limited public datasets, variability in signing styles, and the multimodal complexity of sign language all contribute to the difficulty of building robust AI models for sign language recognition, translation, and synthesis. Addressing these challenges requires significant investment in data collection infrastructure, ethical considerations for participant privacy, and advancements in annotation methodologies. As researchers work toward overcoming these obstacles, the potential benefits of AI-driven sign language technology—enhancing accessibility, bridging communication gaps, and preserving linguistic diversity—remain immense. Continued collaboration between the Deaf community, linguists, and AI researchers is crucial to ensuring that future ML models are inclusive, accurate, and capable of effectively interpreting the richness of sign language communication.

DR HMVE COMBRINK

Strategies for Data Collection

Researchers have explored alternative data collection strategies to improve dataset quality and quantity. Three promising approaches that can significantly enhance sign language datasets are crowdsourcing and community contributions, synthetic data generation using Generative Adversarial Networks (GANs) and probabilistic modelling, and data augmentation techniques. These methods address the shortage of annotated sign language data while improving the robustness and generalisability of ML models.

Crowdsourcing and Community Contributions

Crowdsourcing has become a powerful tool for generating large datasets in numerous AI applications, and it holds great potential for sign language data collection. Engaging the Deaf community in data collection efforts not only improves dataset diversity but also ensures cultural and linguistic accuracy.

Involvement of the Deaf Community in Data Collection

Sign languages are living languages shaped by their users. Direct involvement of the Deaf community in data collection ensures that the dataset captures authentic signing variations, including regional dialects, informal signing styles, and real-world conversational structures. Deaf contributors can provide a range of signs that reflect actual language use, which is critical for improving model accuracy.

Community-led initiatives can take various forms, such as:

Video contributions from Deaf individuals:	Collaborations with Deaf organizations and institutions:	Incentivized participation:
Encouraging Deaf signers to submit video recordings of common phrases, narratives, and interactive dialogues.	Partnering with schools, advocacy groups, and linguistic research centres to build ethically sourced and culturally inclusive datasets.	Offering financial or non-financial incentives (such as community recognition or digital rewards) to encourage contributions from a broad range of signers.

Crowdsourced Annotation Platforms

Collecting raw sign language videos is just the first step; annotation is necessary to make the data useful for ML models. Crowdsourced platforms, where volunteers and experts collaboratively label sign language data, can speed up the annotation process. ***Some possible approaches include:***

Gamification of annotation tasks:	AI-assisted annotation:	Sign language dictionaries and captioning projects:
Turning data labelling into an interactive task where users validate, correct, or categorize signs.	Using preliminary ML models to generate sign labels, which human annotators can then refine.	Encouraging the Deaf community to contribute sign meanings, alternative expressions, and contextual variations.

Crowdsourcing expands dataset accessibility while ensuring that AI models reflect the diversity and authenticity of sign language as used by native signers. For this to be effective, communication and community buy-in is needed.

DR HMVE COMBRINK

Synthetic Data Generation Using Generative Adversarial Networks (GANs)

Since real-world sign language datasets remain scarce, synthetic data generation presents an innovative solution to bridge the data gap. Generative Adversarial Networks (GANs) have emerged as a powerful AI technique for producing high-quality, realistic synthetic data, including sign language gestures.

How GANs Work for Sign Language Data

GANs consist of two neural networks, the generator and the discriminator, that work against each other to create realistic data. The generator creates synthetic sign language video sequences or static image, while the discriminator evaluates whether the generated samples are realistic compared to real-world sign language data. Over time, the generator improves, producing synthetic sign language data that closely resembles human signing. This type of approach can be used for static and dynamic signs with a temporal element to it.

Applications of GANs in Sign Language Recognition

GANs have multiple applications in sign language research:

Augmenting datasets with synthetic signers:	Filling gaps in underrepresented signs:	Enhancing data privacy:
AI-generated avatars can create new signing samples that expand dataset diversity by modelling different signing styles, speeds, and handshapes.	If certain signs are missing from real-world datasets, GANs can synthesize plausible versions of those signs.	Synthetic data generation eliminates concerns about personally identifiable video recordings while still providing meaningful training data.

Challenges and Considerations

Despite their potential, GANs face some challenges when applied to sign language generation:

Ensuring accurate motion dynamics:	Capturing multimodal cues:	Reducing model bias:
Unlike static images, sign language consists of dynamic movements that must appear natural. GANs must accurately model hand transitions, fluidity, and coarticulation (how signs change depending on surrounding signs).	Generating hand gestures alone is insufficient—GANs must also incorporate facial expressions, body positioning, and signing context.	If GANs are trained on a limited dataset, they may generate synthetic data that does not fully represent the diversity of real-world signing.

Despite these challenges, GANs provide a scalable and promising solution for enriching sign language datasets when real-world data is limited.

DR HMVE COMBRINK

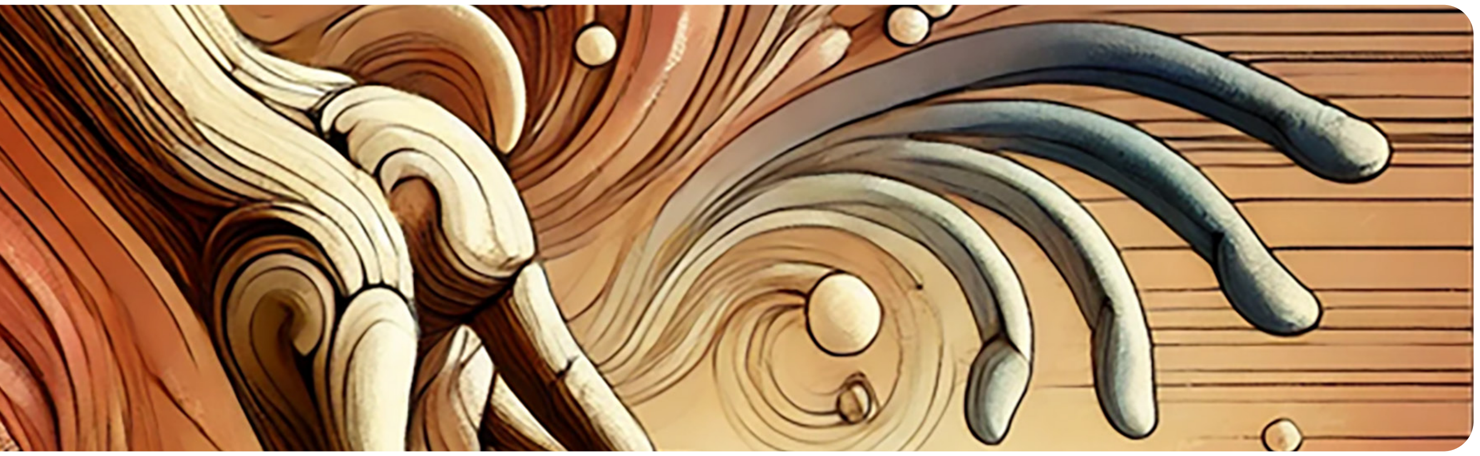
Data Augmentation to Enhance Dataset Diversity

Data augmentation is a well-established technique in ML that improves dataset robustness by applying transformations to existing data. Since collecting new sign language data is resource-intensive, data augmentation provides a cost-effective way to expand datasets and improve model generalization.

Common Data Augmentation Techniques for Sign Language

Several augmentation methods can be applied to sign language videos without altering the semantic meaning of signs. ***These include:***

Rotation:	Slightly rotating video frames to simulate natural variations in hand positioning.
Scaling:	Adjusting the size of the signer's hands to mimic variations in hand size across individuals.
Mirroring (Flipping):	Horizontally flipping video sequences to teach models to recognize signs regardless of left- or right-handed dominance.
Time Warping:	Slightly modifying the speed of sign execution to help ML models learn to interpret both fast and slow signing.
Background Alteration:	Changing backgrounds in videos to improve robustness against different environmental settings.
Adding Noise:	Introducing slight visual distortions to make models more resilient to video quality variations.



Benefits of Data Augmentation

Data augmentation offers multiple advantages in sign language ML applications:

Increases dataset size without additional data collection:	By generating multiple variations of existing data, researchers can effectively expand datasets.
Improves model generalization:	ML models trained on augmented data perform better on unseen signers and real-world scenarios.
Reduces overfitting:	Horizontally flipping video sequences to teach models to recognize signs regardless of left- or right-handed dominance.
Time Warping:	Prevents models from memorizing specific training samples by exposing them to diverse variations.

Considerations and Limitations

While data augmentation is highly beneficial, it must be used carefully to avoid distorting sign meaning. Excessive transformations (such as extreme rotations or distortions) could misrepresent signs and lead to incorrect model learning. Additionally, augmentation techniques must be adapted to the specific characteristics of sign language rather than blindly applying methods designed for other vision-based tasks.

Expanding sign language datasets is essential for developing accurate and inclusive ML models, and three key strategies—crowdsourcing, synthetic data generation, and data augmentation—offer viable solutions to this challenge. Crowdsourcing engages the Deaf community to contribute authentic signing data while fostering community involvement and cultural representation. Generative Adversarial Networks (GANs) provide an innovative approach to generating synthetic sign language data, helping to fill dataset gaps and improve privacy protections. Data augmentation techniques enhance dataset diversity by introducing variations that improve model robustness and generalization.

Each of these approaches comes with its own challenges, from ethical considerations in crowdsourcing to ensuring the realism of GAN-generated signs and maintaining the integrity of augmented data. However, when combined, these strategies offer a comprehensive solution to the longstanding issue of limited sign language datasets. By leveraging these methods, researchers and developers can create more effective sign language recognition, translation, and synthesis models, ultimately improving accessibility and communication for Deaf communities worldwide.

As technology advances, the integration of these data collection and enhancement techniques will play a crucial role in the future of AI-driven sign language applications. Collaboration between the Deaf community, AI researchers, and linguists will be key to ensuring that datasets are representative, ethical, and capable of supporting the next generation of sign language technologies.

DR HMVE COMBRINK

The Need for an Annotation System for Data Labelling

Before training any ML model, it is crucial to have a structured annotation system that accurately labels sign language data. Without proper annotations, models cannot learn meaningful representations, leading to poor accuracy and misinterpretation of signs. To create a structured and interpretable mathematical representation of South African Sign Language (SASL) shorthand annotations, we define the format as follows:

Understanding the Notation System

Sign language is a visual-gestural language that conveys meaning through hand movements, facial expressions, and body positioning. For computational applications, it is essential to represent these features in a structured way so that they can be processed by AI models. The mathematical notation for SASL provides a way to systematically encode these elements, allowing for efficient data processing, storage, and retrieval.

The fundamental structure of this notation is expressed in the following equation (equation 1):

$$S = G \times \left[\frac{D^2}{D_1} \mid \frac{L^2}{L_1} \Big| \frac{S^2}{S_1} \Big| \frac{P^2}{P_1} \Big| \frac{M^2}{M_1} \Big| \frac{R^2}{R_1}\right] \qquad (1)$$

where (equation 2):

$$G = (H \times C \times N) \qquad (2)$$

In this notation, the variable G represents the global features of the sign, consisting of three elements: H, which indicates whether the sign is one-handed (1) or two-handed (2); C, which denotes whether there is contact between the hands (0 for no contact, 1 for contact); and N, which signifies the presence of mouthing, where 1 represents a word and 2 represents a letter.

SYMBOL	MEANING
H	Hand usage: 1 = One-handed, 2 = Two-handed
C	Contact between hands: 0 = No contact, 1 = Contact
N	Mouthing: 1 = Word, 2 = Letter

Each fraction in the equation describes how different hand features interact. The denominator represents the dominant hand (D_1), while the numerator represents the non-dominant hand (D^2). If the numerator is zero, the sign is strictly one-handed, meaning the non-dominant hand does not participate.

Each fraction within the bracket illustrates how the non-dominant hand modifies or interacts with the dominant hand. The vertical bars (|) serve as dividers, distinguishing each category of features. It is important to note that the number for dominant is at the bottom of the letter, and non-dominant expressed as an exponent.

If we breakdown each section of the equation, we therefore have the defining characteristics of each component as:

SYMBOL	MEANING
D	Hand presence: 1 = Dominant hand, 2 = Non-dominant hand
L	Location of the sign: 0 = No specific location, 1 = Neutral position, 2-10 = Specific body locations (forehead, shoulder, chest, etc.)
S	Handshape: Sign alphabet letters (A-Z) and predefined hand configurations
P	Palm orientation: 1 = Out, 2 = In, 3 = Side, 4 = Down, 5 = Up
M	Movement variation: Numeric values representing different movements (e.g., side, up, down, circular motion, flicks, etc.)
R	Repetition: 0 = No repetition, 1 = Repeated movement

Making the Notation More Accessible

Suppose we have a shorthand notation for a sign.

This can be broken down into its sequential mathematical expression:

$$H, C, N, D_1, L_1, S_1, P_1, M_1, R_1, D^2, L^2, S^2, P^2, M^2, R^2$$

Let's say for our example, we have the sequence:

(1,0,1), (1,1,B,1,4,1), (2,0,0,0,0,0)

When converted into the formal notation, we obtain (equation 3):

$$S = (1,0,1) \times \left[\frac{2^2}{1_1} \Big| \frac{0^2}{1_1} \Big| \frac{0^2}{B_1} \Big| \frac{0^2}{1_1} \Big| \frac{0^2}{4_1} \Big| \frac{0^2}{1_1}\right] \qquad (3)$$

This structured representation provides clarity in sign processing, making it easier for ML models to interpret and analyse different sign language expressions accurately. If we capture data in this way, it is an added layer of context we can add to training a model to provide it with visual data that can improve the quality of the sign.

Advantages of This Approach

The SASL notation system offers several key benefits:

Logical Organisation:	By grouping all sign components within a structured bracket system, it becomes easier to analyse and compare different signs computationally.
Explicit Hand Interactions:	The fraction notation clearly indicates how the non-dominant hand modifies or interacts with the dominant hand, providing additional clarity in interpretation.
Scalability:	The system is designed to be extensible, meaning new features—such as signing speed or facial expressions—can be integrated without disrupting the framework.
AI-Readiness:	Since the format is structured mathematically, it can be directly converted into machine-readable formats such as JSON or XML for database storage and ML applications.
Efficient Parsing:	The notation is compact yet descriptive, allowing for seamless integration into AI-driven sign language recognition models.

Developing a structured, computationally viable notation system for SASL is a critical step toward making sign language more accessible to AI-driven applications. By clearly defining each component of a sign, hand usage, location, handshape, palm orientation, movement, and repetition, this system provides a framework for interpreting sign language in a machine-readable format. As AI technologies continue to advance, having a well-defined notation system ensures that sign language recognition and synthesis models can operate with greater accuracy and efficiency. Moreover, by making this system accessible to the broader public, we bridge the gap between linguistic researchers, AI developers, and the Deaf community, fostering a collaborative approach toward improving sign language accessibility worldwide.

DR HMVE COMBRINK, K DE VILLIERS, DR SARA SIYAVOSHI, DR CHRISMI LOTH

BIBLIOGRAPHY

1. Camgoz, N.C., Koller, O., Hadfield, S. and Bowden, R. 2020. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10023-10033).
2. Camgoz, N.C., Koller, O., Hadfield, S. and Bowden, R. 2020. Multi-channel transformers for multi-articulatory sign language translation. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16* (pp. 301-319). Springer International Publishing.
3. Cheng, K.L., Yang, Z., Chen, Q. and Tai, Y.W. 2020. Fully convolutional networks for continuous sign language recognition. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16* (pp. 697-714). Springer International Publishing.
4. Huang, J., Zhou, W., Zhang, Q., Li, H. and Li, W., 2018, April. Video-based sign language recognition without temporal segmentation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).
5. Ji, S., Xu, W., Yang, M. and Yu, K. 2012. 3D convolutional neural networks for human action recognition. I*EEE transactions on pattern analysis and machine intelligence*, 35(1), pp.221-231.
6. Pigou, L., Dieleman, S., Kindermans, P.J. and Schrauwen, B. 2015. Sign language recognition using convolutional neural networks. In *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13* (pp. 572-578). Springer International Publishing.
7. Parton, B.S. 2006. Sign language recognition and translation: A multidisciplined approach from the field of artificial intelligence. *Journal of deaf studies and deaf education,* 11(1), pp.94-101.
8. Simonyan, K. and Zisserman, A. 2014. Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems,* 27.
9. Taylor, G.W., Fergus, R., LeCun, Y. and Bregler, C. 2010. Convolutional learning of spatio-temporal features. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI 11* (pp. 140-153). Springer Berlin Heidelberg.
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems,* 30.
11. Wang, L., Duan, X., Zhang, Q., Niu, Z., Hua, G. and Zheng, N. 2018. Segment-tube: Spatio-temporal action localization in untrimmed videos with per-frame segmentation. *Sensors,* 18(5), p.1657.
12. Weissmann, J. and Salomon, R. 1999, July. Gesture recognition for virtual reality applications using data gloves and neural networks. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)* (Vol. 3, pp. 2043-2046). IEEE.
13. Pigou, L., Dieleman, S., Kindermans, P.J. and Schrauwen, B. 2015. Sign language recognition using convolutional neural networks. In *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13* (pp. 572-578). Springer International Publishing.
14. Bowden, R., Zisserman, A., Kadir, T. and Brady, M. 2003. Vision based interpretation of natural sign languages. In *Proceedings of the 3rd International Conference on Computer Vision Systems – ICVS 2003 (Graz, Austria).* Springer-Verlag Berlin.
15. Bird, J.J., Ekárt, A. and Faria, D.R. 2020. British sign language recognition via late fusion of computer vision and leap motion with transfer learning to American sign language. *Sensors,* 20(18), p.5151.

During the course of our research, we realised the need for a novel annotation system for Human Language Technology, specific to South African Sign Language (SASL). The purpose of this system is to aid the processing of data as well as its storage with context. This allowed us to build much more context to all the models we develop. Machine learning and artificial intelligence alone is not enough without human input, human context, and human conceptualisation. All of our annotation-based research directed us to capture, manage and govern our data in a much more coherent manner for future research.

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