The SMILE Continuous DSGS Corpus: A Resource for Longitudinal Exploration of Continuous Swiss German Sign Language

Alessia Battisti*, Katja Tissi†, Sandra Sidler-Miserez†, Sarah Ebling*
*University of Zurich
Andreasstrasse 15, 8050 Zurich
{battis, ebling}@cl.uzh.ch

†University of Teacher Education in Special Needs
Schaffhauserstrasse 239, 8050 Zurich
katja.tissi@hfh.ch | sandysidler@gmail.com

Abstract

This paper presents the *SMILE Continuous DSGS Corpus*, a longitudinal dataset that allows for investigating how hearing adults acquire Swiss German Sign Language as a second language. It includes recordings of sign language learners and native signer controls collected at four points over a period of 18 months and annotated for manual and non-manual components, errors, and sentence-level acceptability. The resource provides high-quality, synchronized video suitable for both linguistic and automatic sign language processing research, for example, supporting studies of interlanguage development and training of automatic sign language recognition models. We present here an exploratory analysis of the learner subcorpus using Bayesian mixed-effects modeling. The corpus and accompanying annotations are available for research purposes under a Creative Commons license (CC BY-NC-SA 4.0).

Keywords: Swiss German Sign Language dataset, automatic sign language assessment, sign language learner analysis, Bayesian mixed-effects modeling, multimodal annotation, sign language processing

1. Introduction

Sign languages are full-fledged natural languages with phonological, morphological, and syntactic structures expressed through manual and non-manual articulations. While research on sign language linguistics has advanced substantially over the past two decades, the development of standardized, annotated resources still lags behind that of spoken languages[1]. The emergence of large-scale corpora such as the British Sign Language (BSL) Corpus (Schembri, 2008), the German Sign Language (*Deutsche Gebärdensprache*, DGS) Corpus (Konrad et al., 2020), and the Corpus of the Sign language of the Netherlands (*Nederlandse Gebarentaal*, NGT) (Crasborn et al., 2006-2017) has changed linguistic and computational research by providing empirical foundations for studying grammatical variation and use within deaf[2] communities. These resources have also supported the creation of digital tools for lexicography as well as catalyzed progress in automatic sign language processing (SLP), fostering advances in recognition and generation.

Despite this progress, comparable resources remain scarce for sign languages such as Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*, DSGS), the first sign language of deaf and hard-of-hearing persons in German-speaking Switzerland. DSGS is used by approximately 7,000 native or early signers and about 13,000 hearing individuals who engage with the language for various purposes (Boyes Braem, 2024). DSGS has gained institutional visibility in education, media, and accessibility contexts, yet still lacks a comprehensive corpus suitable for both linguistic research and computational modeling.

To address this need, the SMILE Continuous DSGS Corpus was created within a four-year interdisciplinary project aimed at advancing sign language learning and assessment technologies. The resource combines DSGS productions of both hearing adult (L2) and native (L1) signers, recorded across multiple tasks and data collection phases. It provides continuous signing data accompanied by fine-grained manual and non-manual annotations, error labeling, and sentence-level acceptability judgments. This combination of linguistic richness and technical precision makes the corpus a dual-purpose resource: it enables in-depth analysis of sign language structure and learning while simultaneously serving as high-quality training/fine-tuning data for SLP tasks.

The recordings and annotations were designed to capture the variability of both L1 and L2 productions, offering realistic input for training models that can generalize across signer profiles, fluency levels, and communicative styles.

[1] In this paper, "spoken language" denotes any non-signed language, expressed either orally or in writing.

[2] Following more recent conventions (e.g., Kusters et al. (2017); Napier and Leeson (2016)), we use the term *deaf* inclusively to refer to all members of the sign language community, irrespective of audiological status.

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 17–30
16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

From a linguistic perspective, the resource supports the investigation of interlanguage (Selinker, 1972). The longitudinal design allows for tracing changes in articulatory precision, grammatical accuracy, and fluency over time, while parallel native data provide a benchmark for assessing progress and defining target competence.

This paper introduces the SMILE Continuous DSGS Corpus, which is made available for research purposes under a Creative Commons license[3]. It also presents a first exploratory analysis of the learner data, focusing on the way in which error patterns and fluency evolve over time.

The remainder of this paper is structured as follows: Section 2 reviews previous work on sign language corpora and learner-oriented resources. Section 3 describes the methodology and steps involved in creating the SMILE Continuous DSGS Corpus. Section 4 presents a first interlanguage study conducted on the collected data. Finally, Section 5 concludes with a discussion of limitations and directions for future work.

2. Related Work

Despite progress in the development of sign language corpora, sign languages remain severely under-resourced, particularly in the context of SLP (De Sisto et al., 2024; Holmes et al., 2023; Joshi et al., 2020; Sayers et al., 2021). Existing datasets have been created for a variety of purposes, including linguistic research and technological development, and therefore differ widely in their design, annotation granularity, and availability (Ebling et al., 2018). Compared to spoken language corpora, most sign language datasets are considerably smaller, often rely on manual transcription and annotation, and lack standardized practices, which hinders large-scale data sharing and machine learning applications (Tanzer and Zhang, 2024; Schulder et al., 2023).

Broadly speaking, sign language resources can be categorized into three types: (i) single-sign datasets primarily used for SLP, (ii) continuous signing datasets primarily used for SLP, and (iii) linguistic research corpora.

The first category, single-sign datasets primarily used for SLP, includes resources such as PopSign ASL (Starner et al., 2023), BOSPHORUS-Sign22k (Özdemir et al., 2020), and AUTSL (Sincan and Keles, 2020). These are essential for lexical modeling and recognition but focus on isolated signs, neglecting grammatical phenomena expressed across larger temporal units and often making use of non-manual articulators. The second category, continuous datasets mainly created for SLP, includes interpreted or translated content such as PHOENIX-14T (Camgöz et al., 2018) or BOBSL (Albanie et al., 2021). While they provide valuable parallel data, these datasets are often limited to narrow domains (e.g., weather forecasts) and reflect interpreter language rather than spontaneous signing. Moreover, many lack detailed linguistic annotation and metadata, limiting their usefulness for fine-grained linguistic analysis. From a technological perspective, advances in deep learning and computer vision have driven the creation of large-scale continuous SLP datasets such as MS-ASL (Joze and Koller, 2019), WLASL (Li et al., 2020), and YouTube-SL-25 (Tanzer and Zhang, 2024). These resources have enabled significant progress in automatic recognition and translation (Bragg et al., 2019), but they generally focus on lexical tasks, lack detailed linguistic annotation, and seldom include learner or non-native data.

In contrast, corpora developed primarily for linguistic research, such as the DGS Corpus (Konrad et al., 2020), Corpus NGT (Crasborn and Zwitserlood, 2008), and the BSL Corpus (Schembri et al., 2014), provide richly annotated, representative samples of native signers. These resources are invaluable for describing sign language structure and variation, yet they are often not designed for computational reuse. Licensing restrictions and format heterogeneity reduce their applicability for SLP research.

These combined issues have led to a divide between datasets that are optimized for large-scale use but linguistically shallow and those that are linguistically rich but not immediately usable for computational research.

Sign language learner data is even more scarce. Neither the comprehensive overviews by Kopf et al. (2021, 2022) nor existing learner corpus registries include any corpus devoted to sign language learners. Nonetheless, several studies have collected data for linguistic analyses (Chen Pichler, 2011; Ferrara and Nilsson, 2017; Gulamani et al., 2020; Schönström and Leeson, 2015; Kurz et al., 2023). Among these, the Swedish Sign Language L2 Corpus (STSC-L2) compiled by Mesch and Schönström (2018) stands out as the first longitudinal resource documenting adult learners of Swedish Sign Language (*Svenskt teckenspråk*, STS) over four phases of instruction. It has supported research on the development of mouthing and depicting signs (Mesch and Schönström, 2020; Schönström and Mesch, 2022). Similarly, the Corpus of German Sign Language as L2 (Oviedo et al., 2018) includes large-scale assessment data from adult learners, but the recordings are primarily exam-based and only partially longitudinal.

For DSGS, the dataset introduced by (Ebling et al., 2018) provides a foundational contribution

[3] https://www.swissubase.ch/en/catalogue/studies/21056

containing both L1 and L2 data. However, its recordings are limited to isolated signs rather than continuous sequences, and thus do not capture the interplay between manual and non-manual articulations or the temporal evolution of learner fluency.

3. Corpus Creation

The SMILE Continuous DSGS Corpus was developed within a four-year interdisciplinary project aimed at advancing research and technology in sign language learning and assessment. Its design was guided by two complementary goals: (1) to create a longitudinal, linguistically annotated resource of signing by both L2 and L1 signers of DSGS; (2) to provide a machine-readable benchmark suitable for training and evaluating SLP models.

It consists of productions in a semi-controlled setting over a period of 18 months, providing the first longitudinal record of DSGS learning. The corpus was collected between March 2022 and November 2023 at the University of Zurich.

3.1. Participants

The corpus includes productions from two groups of participants.

The L1 subcorpus contains recordings from ten adult native or early signers[4], all of whom use DSGS as their primary means of communication. This subset serves as a control subcorpus.

The L2 subcorpus includes hearing adults acquiring DSGS as a second language in a second modality (M2L2; cf. Chen Pichler, 2019). Participants ranged in age from 22 to 55 years and had all completed at least one formal DSGS course by the time of the first recording. Among them, 14 participants were enrolled in the DSGS interpreter training program at the University of Teacher Education in Special Needs (Interkantonale Hochschule für Heilpädagogik, HfH), while eleven learners were recruited from public DSGS courses.

All participants provided informed consent and completed an online metadata questionnaire available in both German and DSGS. Deaf participants received financial compensation, while learners either obtained ECTS credits within their training program, received monetary compensation, or obtained individualized performance feedback after each data collection phase.

3.2. Elicitation Tasks

To examine developmental trajectories, L2 participants were recorded across four data collection phases, spaced approximately six months apart, resulting in four comparable datasets (Phase 1-Phase 4). Each phase included the same set of core elicitation tasks, allowing for longitudinal comparisons across identical stimuli, while a small number of new items were added at each phase to gradually increase linguistic complexity.

At each data collection, seven tasks were administered. They were designed to elicit a wide range of linguistic features, specifically with respect to the use of non-manual components, within controlled and semi-spontaneous contexts.

The tasks were presented in DSGS along with written German instructions for the L2 group, using on-screen prompts.

Table 1 summarizes the seven tasks presented to the participants. Controlled items, such as from a *Sentence Repetition Task* (SRT) (Haug et al., 2020), featured identical items across phases, or a mix of repeated and new items, as in the *Sentence Transformation Task* (STT). Narrative tasks (e.g., video retelling) were alternated to maintain participant engagement and minimize memorization effects. In addition, L2 participants were invited to an interview (Haug et al., 2019), which was conducted individually with a deaf expert in DSGS teaching and research.

3.3. Data Collection Setup

All recordings were conducted in a controlled studio environment to ensure consistent visual conditions suitable for both linguistic analysis and computer vision processing.

Each recording session was filmed using five synchronized cameras of two types (cf. Figure 1). Three FLIR Blackfly S machine vision cameras (front, front-left, front-right) provided high-resolution frontal views of the signers' faces and upper bodies for detailed analysis of manual and non-manual components. These cameras recorded at 60 fps and with 1024p resolution. Two auxiliary webcams, positioned laterally and overhead, captured the depth and trajectory of hand and arm movements for complementary linguistic analysis. These recordings were captured at 30 fps.

Uniform controlled lighting was used to minimize shadow interference, particularly in the face and hand regions. A fixed recording distance and signer position were maintained across all sessions, marked on the floor and backdrop to guarantee consistent framing across participants and data collection phases.

All cameras were hardware-synchronized via a unified trigger cable to eliminate temporal lag and monitored through a single recording interface, an adaptation of the capture system used in (Ebling et al., 2018). Each session was logged using standardized metadata templates documenting partici-

[4] Early signers are individuals who acquired DSGS during early childhood and use it as their dominant language.

Task	Elicitation
Route description	Spatial descriptions, pointing signs, fingerspelling
Sentence Transformation Task	Grammatical structures, non-manual components
Picture retelling	Description and comparison
Clip retelling	Description, long sentences
Video retelling	Long sentences, non-manual components
Sentence Repetition Task	Language comprehension, processing, and production skills
Proficiency interview	Language comprehension, processing, and production skills

Table 1: Summary of the tasks performed in each recording session by the participants.

pant identifier, task type, and session date. Timestamped metadata were automatically generated to ensure reliable mapping between the raw video recordings and the corresponding task metadata files. Each recording session lasted approximately 45–60 minutes. To avoid potential order effects, the presentation order of individual stimuli within each task was randomized for every participant and session.

All recordings were saved in MP4 format (H.264 codec).

3.4. Annotation and Validation

A detailed description of the annotation conventions, data validation procedures, and reliability measures is available in Battisti et al. (2024). Building on this framework, two tasks produced by both L1 signers and L2 learners who participated in all four data collection phases were annotated accordingly. The annotations were carried out in iLex (Hanke and Storz, 2008) using a multilayered structure. Each file includes annotation tiers for:

- manual components: glosses as well as handshape, movement, location, orientation;
- non-manual component: facial expression, eyebrow position, head movement, eye gaze, mouthing;
- errors: deviations from canonical forms for manual and non-manual components;
- acceptability: judgments of comprehensibility and naturalness at the level of individual errors and entire sentences.

All annotations were produced by deaf expert annotators and project members with linguistic expertise. Annotation quality was ensured through double annotation and internal review cycles, followed by inter-annotator agreement (IAA) testing on a representative subset. IAA reached $\kappa > 0.7$ across the main tiers (manual and non-manual components, error types, and acceptability), indicating moderate to substantial reliability. All validated annotations are included in the released corpus, aligned with the corresponding metadata and video recordings.

3.5. Corpus Profile

Table 2 provides a quantitative overview of the SMILE Continuous DSGS Corpus, detailing the number of participants, sessions, hours, and annotated items across phases. The corpus comprises 153 video sessions collected from 35 participants, spanning over four data collection phases. Each phase includes all elicitation tasks, resulting in approximately 67 hours and 15 minutes of recorded material in total, of which 13 hours and 59 minutes correspond to effective signing time after removing pauses and preparation segments.

Within the L2 subset, the annotated portion currently contains 1,723 sentences, 11,883 gloss tokens, and 27,656 non-manual component instances. This yields an average of 6.9 glosses and 16 non-manual components per sentence. The overall annotation includes 4,074 error tokens, divided into 1,968 manual and 2,106 non-manual errors. Approximately two-thirds of the learner sentences (1,145 items) contain at least one annotated error.

The L1 control subset consists of around nine hours of recordings produced by ten L1 participants performing the same tasks as those assigned to the L2 group.

4. Interlanguage Study

This section presents a first empirical analysis of the L2 learner subset of the SMILE Continuous DSGS Corpus, focusing on error patterns, fluency development, and non-manual articulation across four data collection phases. The aim is to examine how adult hearing learners of DSGS develop competence over time and to assess how the corpus can support quantitative modeling of interlanguage.

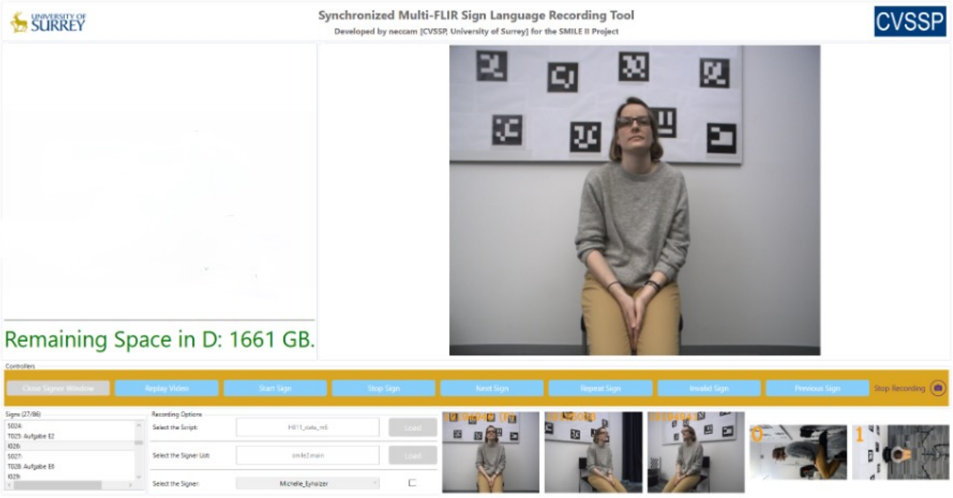


Figure 1: Sample of the recording interface with the input shown to the participant on the left and the multi-view perspectives at the bottom.

	(M1)L1	M2L2 (P1)	M2L2 (P2)	M2L2 (P3)	M2L2 (P4)
Signers	10	22	18	14	14
Items	1,089	757	720	602	631
Total hours	09:43:19	14:53:58	15:30:43	13:35:35	13:31:33
Eff. signing	02:56:46	03:06:18	02:53:33	02:22:57	02:40:15
Sentences	460	250	313	336	364
Tokens	3,191	1,414	1,820	2,377	3,081
Non-manuals	6,514	3,089	4,270	5,813	7,970
Errors (MC)	-	395	402	476	695
Errors (NMC)	-	351	465	575	715

Table 2: Corpus profile: participants, hours, items, and annotation coverage across phases P1|2|3|4.

More specifically, the goal is to model how accuracy and fluency in DSGS production change over time and to identify which articulatory components, manual or non-manual, show the most progress or persistence of errors. In other words, how do the frequency and distribution of errors in manual and non-manual components change over an 18-month learning period?

4.1. Data

As described in Section 3.4, two tasks were fully annotated for the present study: the STT and the SRT (see Section 3.2). The analysis presented in this section focuses on the L2 annotated subset. The subset comprises 1,263 learner sentences with token-aligned error tiers, yielding 4,074 error tokens in total: 1,968 affecting manual components and 2,106 affecting non-manual components (cf. Table 2 in Section 3.5).

For each learner and phase, the annotations record the number of glosses, non-manual components, and error instances per sentence. Since all learners completed the same tasks at each phase, the dataset has a hierarchical structure with repeated measures: multiple observations (sentences) are nested within learners, and each learner is observed across four phases.

4.2. Methods

Descriptive Analysis. We first examined the frequencies and distributions of (i) non-manual component occurrences, (ii) component-level error types and their co-occurrences across manual and non-manual domains, and (iii) sentence-level error categories. These distributions offer an overview of interlanguage profiles; however, as relative frequencies do not account for sentence length, learner-specific baselines, or task-related variation, they should be interpreted with caution (Gries and Wulff, 2021; Von Stutterheim et al., 2021).

Fluency Investigation. To approximate fluency, we computed signing speed as average time per gloss: sentence duration in seconds divided by gloss count (Kanto and Haapanen, 2019; Haug et al., 2024). We modeled speed differences across phases using linear mixed-effects models (`lmer`, Bates et al., 2015), followed by Tukey-adjusted pairwise contrasts with `emmeans` (Lenth, 2024).

Error-count Modeling. To examine phase effects on error production while accounting for repeated measures, we fitted Bayesian generalized mixed-effects models with `brms` (Bürkner, 2017). Manual-component and non-manual-component errors were modeled separately due to their weak correlation ($\rho = 0.24$) and to simplify the models for interpretation. Phase was entered as a fixed effect. Random intercepts were specified for learners and tasks, with by-learner random slopes for phase to capture individual learning trajectories (Baayen et al., 2008; Gries, 2021; Meteyard and Davies, 2020).

Distributional families were selected via leave-one-out cross-validation (Vehtari et al., 2017): a negative binomial for manual component errors to address overdispersion (McElreath, 2020; Winter and Bürkner, 2021) and a hurdle Poisson for non-manual-component errors to accommodate the 21% zero-inflation. Offsets standardized counts: $\log(\text{glosses})$ for the manual component model (*mc model*); $\log(\text{NMC-tokens})$ for the non-manual component model (*nmc model*). We used weakly informative priors encoding an expectation of declining error rates across phases while allowing substantial variability across learners and tasks; sensitivity checks indicated robust inferences.

4.3. Results: Descriptive Analysis

Use of Non-manual Components. As shown in Figure 2, the frequency and variety of non-manual components increased steadily over time. Later phases display higher median counts and reduced variability, indicating that learners gradually incorporated a broader range of articulators into their signing, particularly mouth gestures and eye gaze. In contrast, eyelid and nose movements remained infrequent across all phases.

Error Distributions. Visualizations of the error distributions at the manual, non-manual, and sentence levels are provided in Appendix A.

For manual components (Figure 5), the *movement* parameter consistently show the highest frequency (33.1%), followed by its combination with *location*. Overall, movement-related errors, alone or in combination, account for 72.1% of all manual-component errors. Manual error profiles remain stable across phases.

For non-manual components, errors involving the *eyebrows* are the most frequent (35.2% individually; 65.5% when co-occurring with other components), with *eyebrow* + *head* combinations forming a common pairing (16.1%) (Figure 6). Across phases, *mouthing* emerges as the second most frequent error category in Phase 3, overtaking head-movement errors.

At the sentence level (Figure 7), *fluency*-related issues are most prevalent (approximately 24%), followed by *facial expression* (12%) and *sentence construction* (10%). Fluency errors decreased across phases, from 43% in Phase 1 to 26% in Phase 4. In this context, *fluency errors* corresponded to annotator judgments indicating disruptions in the temporal progression of signing, such as prolonged pauses, visible hesitation, repetitions, or restarts, that hindered the smooth delivery or continuity of the utterance (Haug et al., 2024). Co-occurrence patterns between phonological and syntactic errors highlight the interaction of linguistic domains in the learners' developing interlanguage. For example, inaccurate handshape or movement realization (phonological error) often co-occurred with incorrect use of non-manual markers required for syntactic contrast.

4.4. Results: Modeling

Signing Speed. As shown in Figure 3, average signing speed, measured as time per gloss, improved by decreasing from Phase 1 (median $\approx$1.46 s) to Phase 4 (median $\approx$1.12 s), with significant contrasts for Phase 4-Phase 1 ($\beta = 0.15$, $p = 0.003$) and Phase 2-Phase 1 ($\beta = 0.04$, $p = 0.02$); Phase 3-Phase 2 and Phase 4-Phase 3 are not credibly different. This trajectory indicates that learners rapidly gained fluency during early acquisition stages before plateauing once a functional pace was reached.

Manual Component Errors. The *mc model* examined changes in manual component errors across phases while controlling for sentence length and random variation among learners and tasks. As shown in Figure 4a, manual errors decreased steadily from Phase 1 to Phase 3 ($\beta_{P3} = -0.33$, $\text{CrI}_{95\%}$=[-0.52,-0.15]), followed by a stabilization in Phase 4 ($\beta_{P4} = -0.24$, $\text{CrI}_{95\%}$=[-0.56,-0.06]). Credible intervals (CrI) are broad, reflecting strong individual and, to a lesser extent, task variability, but the posterior mass for Phase 4 is predominantly below zero (94%), indicating fewer errors than at baseline.

An investigation of the random effects revealed moderate inter-learner variability, particularly in the first and last phases, suggesting that some participants either progressed faster or maintained higher baseline accuracy. Task-related variability was comparatively small. Posterior parameter estimates are reported in Appendix B.

Non-manual Component Errors. The *nmc model* showed a similar non-linear trajectory but

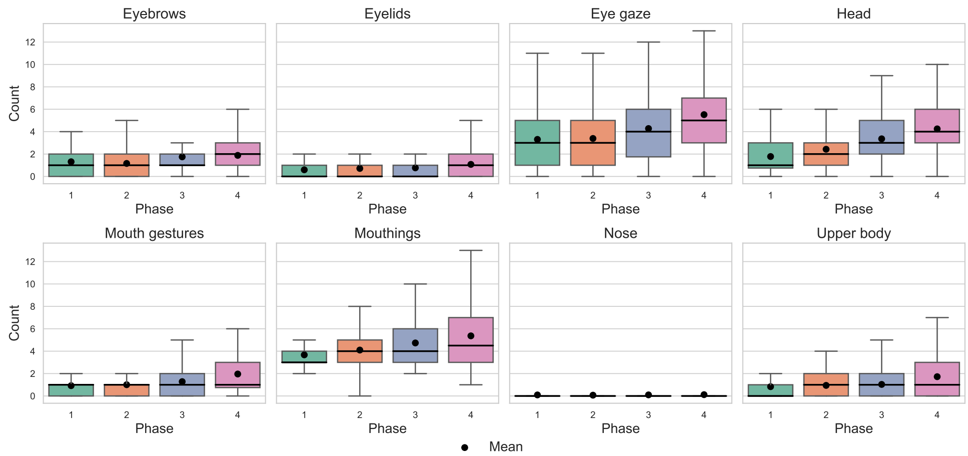


Figure 2: Distribution of individual non-manual components across data collection phases. Mouth gestures and eye gaze increase over time, while eyelid and nose movements remain rare.

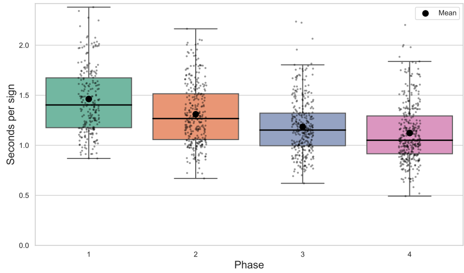
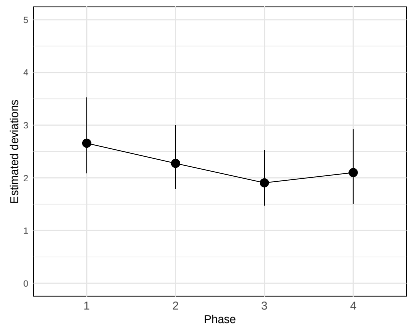


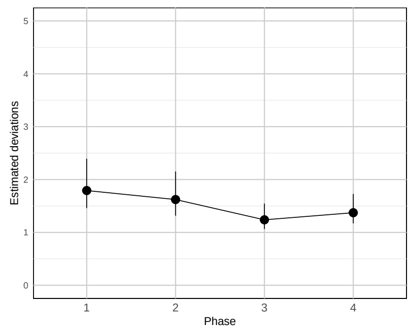
Figure 3: Average signing time per gloss across four phases. Median duration per gloss decreases over time, indicating increasing fluency.

with a steeper mid-phase decline (Figure 4b). Non-manual errors decreased sharply between Phase 1 and Phase 3 ($\beta_{P3} = -0.70$, CrI$_{95\%}$=[-1.03,-0.37]) and stabilized thereafter ($\beta_{P4} = -0.46$, CrI$_{95\%}$=[-0.79,-0.15]).

Unlike manual components, non-manual variability remained high across all phases, reflecting persistent differences in how learners used facial expressions and head movements to mark grammatical and discourse functions. This variability likely stems from individual differences in expressiveness, as well as from pedagogical factors; non-manual features often receive less systematic feedback and are introduced later or less explicitly in instruction compared to manual parameters. Learners accounted for most of the random effect variance, suggesting that non-manual coordination follows highly individualized developmental paths. Phase alone explained about 9% of the variance, rising to 15% when learner and task effects were included.



(a) Manual components



(b) Non-manual components

Figure 4: Conditional effects of phase on expected error counts for manual (a) and non-manual (b) components. Both show decreasing trends that stabilize in later phases.

4.5. Discussion

This study examined how L2 learners of DSGS developed manual and non-manual articulations over four data collection phases using a combination of descriptive and Bayesian mixed-effects analyses. The main goal was to identify whether error counts declined and whether fluency and non-manual component use improved over time.

The results confirm that interlanguage development in sign language is neither linear nor uniform. In early phases, learners relied on manual articulation, producing few non-manual components and many movement-related errors. This pattern reflects an initial focus on lexical retrieval and basic articulatory control rather than multi-channel integration. The predominance of movement and location errors is consistent with earlier studies of DSGS (Ebling et al., 2021) and aligns with cross-linguistic evidence from ASL and BSL learner corpora (Rosen, 2004; Ortega and Morgan, 2015). These errors can be interpreted as early-stage motor simplifications rather than conceptual misunderstandings (Marshall et al., 2021).

Between Phases 2 and 3 of our study, a marked restructuring occurred: manual accuracy improved, fluency errors declined, and the frequency of non-manual components increased. This phase may represent a consolidation of articulatory routines. The observed acceleration in signing speed supports this interpretation, showing that learners begin to produce sentences with greater temporal cohesion.

Non-manual components showed the steepest improvement but also the greatest variability. Eyebrow and head movement errors remained frequent across all phases, while mouthing errors peaked in Phase 3. This pattern suggests that learners temporarily rely on spoken-language strategies, such as overusing mouth gestures, before converging toward target-like grammatical facial expressions (Mesch and Schönström, 2020; Kimmelman et al., 2020). Persistent variability in non-manual components likely stemmed from perceptual constraints, i.e., signers could not directly monitor their own facial configurations (Marshall et al., 2021), and from pedagogical practices that prioritized manual precision in early training.

The models showed that the largest reduction in both manual and non-manual errors occurred between Phases 1 and 3, followed by a plateau. The wide credible intervals and learner-level variance highlighted heterogeneous trajectories rather than a uniform trend. This variability points to individual differences in motor learning, feedback, and exposure, aligning with usage-based theories of second-language acquisition (Chang and Zhang, 2021; Lambelet and Berthele, 2015).

Overall, the findings suggest that using and coordinating non-manual components is a late-acquired skill in the DSGS learning process. As learners gain control of articulatory timing, their interlanguage becomes more stable. These insights can thus support a pedagogical focus on integrating facial grammar and prosody earlier in instruction and motivate data-driven feedback tools.

5. Limitations and Future Work

Several limitations constrain the generalizability of this analysis. The annotated subset covers only two tasks and a limited sample of learners, which restricts statistical power and the range of linguistic phenomena examined. Manual annotation, particularly for non-manual components, can result to be inherently subjective due to overlapping articulations and fuzzy temporal boundaries. Although Bayesian modeling mitigates small-sample uncertainty, the wide credible intervals reflect substantial learner-level variance. Future analyses should incorporate additional predictors, such as learners' prior exposure, contact with deaf signers, and task complexity, to better explain this variability.

Future work will extend annotation to spontaneous and narrative tasks, such as video retelling, include automatic tracking of facial landmarks to complement manual labels, and refine the annotation scheme based on validation results.

6. Conclusion

This paper presented the *SMILE Continuous DSGS Corpus*, a longitudinal resource that includes continuous productions by adult learners of DSGS, as well as productions by native and early signers. Combining controlled recordings, detailed annotation, and rich metadata, the corpus provides a foundation for linguistic and computational research on sign language learning and processing. A first analysis of learner interlanguage using Bayesian mixed-effects modeling revealed a non-linear developmental pattern: a marked improvement in manual and non-manual accuracy between the early and mid-learning phases, followed by stabilization and strong inter-learner variability.

Methodologically, the study illustrates how probabilistic modeling can yield interpretable developmental insights even from a small-scale longitudinal dataset.

The SMILE Continuous DSGS Corpus thus offers both a benchmark for modeling sign language learning and a resource for developing data-driven feedback in sign language education.

7. Data and Code Availability

The *SMILE Continuous DSGS Corpus* is available for research purposes through SwissUBase[5] under the Creative Commons license CC BY-NC-SA 4.0: SMILE Continuous DSGS Corpus L1 https://doi.org/10.48656/ch11-5k92; SMILE Continuous DSGS Corpus L2 https://doi.org/

[5]https://www.swissubase.ch/en/catalogue/studies/21056/21315

10.48656/rk0q-3g09. All analysis scripts will be made available upon request to the corresponding author.

8. Acknowledgments

This work was funded through the Swiss National Science Foundation (SNSF) Sinergia project "Scalable Multimodal Sign Language Technology for Sign Language Learning and Assessment II" (SMILE II) (grant agreement no. CRSII5_193686).

9. Bibliographical References

Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. 2021. BBC-Oxford British Sign Language Dataset. ArXiv:2111.03635 version: 1.

Rolf Harald Baayen, Douglas J. Davidson, and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Alessia Battisti, Katja Tissi, Sandra Sidler-Miserez, and Sarah Ebling. 2024. Advancing annotation for continuous data in Swiss German sign language. In *Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, pages 1–12, Torino, Italia. ELRA and ICCL.

Adriane Boyd, Jirka Hana, Lionel Nicolas, Detmar Meurers, Katrin Wisniewski, Andrea Abel, Karin Schöne, Barbora Štindlová, and Chiara Vettori. 2014. The MERLIN corpus: Learner language and the CEFR. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1281–1288, Reykjavik, Iceland. European Language Resources Association (ELRA).

Penny Boyes Braem. 2024. Gebärdensprache. In Elvira Glaser, Johannes Kabatek, and Barbara Sonnenhauser, editors, *Sprachenräume der Schweiz*, volume 1 of *Sprachen*, pages 185–209. Narr Francke Attempto, Tübingen.

Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, pages 16–31, New York, NY, USA. Association for Computing Machinery.

Paul-Christian Bürkner. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28.

Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural Sign Language Translation. In *Proceedings of CVPR 2018*.

Necati Cihan Camgöz, Ahmet Alp Kındıroğlu, Serpil Karabüklü, Meltem Kelepir, Ayşe Sumru Özsoy, and Lale Akarun. 2016. BosphorusSign: A Turkish Sign Language Recognition Corpus in Health and Finance Domains. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1383–1388, Portorož, Slovenia. European Language Resources Association.

Pengyun Chang and Lawrence Jun Zhang. 2021. A CDST Perspective on Variability in Foreign Language Learners' Listening Development. *Frontiers in Psychology*, 12.

Deborah Chen Pichler. 2011. Sources of Handshape Error in First-Time Signers of ASL. In Gaurav Mathur and Donna Jo Napoli, editors, *Deaf around the world: the impact of language*. Oxford University Press, Oxford; New York.

Deborah Chen Pichler. 2019. Effect of minimal sign language instruction on hearing learners' constituent order. In *Presented at The 13th International Conference on Theoretical Issues in Sign Language Research (TISLR)*, University of Hamburg, Germany.

Onno Crasborn and Inge Zwitserlood. 2008. The Corpus NGT: an online corpus for professionals and laymen. In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 44–49, Marrakech, Morocco. European Language Resources Association (ELRA).

Nivja H. de Jong. 2018. Fluency in Second Language Testing: Insights From Different Disciplines. *Language Assessment Quarterly*, 15(3):237–254.

Nivja H. de Jong. 2023. Fluency in Speaking as a Dynamic Construct. *Language Teaching Research Quarterly*, 37:179–187.

Mirella De Sisto, Vincent Vandeghinste, Caro Brosens, Myriam Vermeerbergen, and Dimitar Shterionov. 2024. Xsl-horeco and gost-parc-sign: Two new signed language - written language parallel corpora. In Krister Lindén, Thalassia Kontino, and Jyrki Niemi, editors, *Selected papers from the CLARIN Annual Conference 2023*. Linköping Electronic Conference Proceedings, Leuven, Belgium.

Sarah Ebling, Necati Cihan Camgöz, Penny Boyes Braem, Katja Tissi, Sandra Sidler-Miserez, Stephanie Stoll, Simon Hadfield, Tobias Haug, Richard Bowden, Sandrine Tornay, Marzieh Razavi, and Mathew Magimai-Doss. 2018. SMILE Swiss German Sign Language Dataset. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association.

Sarah Ebling, Katja Tissi, Sandra Sidler-Miserez, Cheryl Schlumpf, and Penny Boyes Braem. 2021. Single-parameter and parameter combination errors in L2 productions of Swiss German Sign Language. *Sign Language & Linguistics*, 24(2):143–181.

Lindsay Ferrara and Anna-Lena Nilsson. 2017. Describing spatial layouts as an L2M2 signed language learner. *Sign Language & Linguistics*, 20(1):1–26.

Stefan Th. Gries and Stefanie Wulff. 2021. Examining individual variation in learner production data: A few programmatic pointers for corpus-based analyses using the example of adverbial clause ordering. *Applied Psycholinguistics*, 42(2):279–299.

Stefan Thomas Gries. 2021. *Statistics for linguistics with R: a practical introduction*, 3rd revised edition edition. Mouton Textbook. de Gruyter Mouton, Berlin.

Sannah Gulamani, Chloë Marshall, and Gary Morgan. 2020. The challenges of viewpoint-taking when learning a sign language: Data from the 'frog story' in British Sign Language. *Second Language Research*, 38(1):55–87.

Thomas Hanke and Jakob Storz. 2008. iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)*, pages 64–67, Marrakesh, Marocco. European Language Resources Association.

Tobias Haug, Aaron Olaf Batty, Martin Venetz, Christa Notter, Simone Girard-Groeber, Ute Knoch, and Mireille Audeoud. 2020. Validity evidence for a sentence repetition test of Swiss German Sign Language. *Language Testing*, 37(3):412–434.

Tobias Haug, Nivja H. de Jong, Franz Holzknecht, Katja Tissi, Sandra Sidler-Miserez, Alessia Battisti, Regula Perrollaz, Sarah Ebling, Sabine Reinhard, and Sarah Caminada. 2024. Development and validation of a fluency rating scale for Swiss German Sign Language. *Frontiers in Education*, 9.

Tobias Haug, Heidi Stocker, and Daniela Nussbaumer. 2019. Die Entwicklung von Instrumenten zur Überprüfung von kognitiven Fähigkeiten, gebärdensprachlicher Kompetenz und Dolmetschleistung von Gebärdensprachdolmetscherinnen. *Das Zeichen*, 111:130–143.

Ruth Holmes, Ellen Rushe, Mathieu De Coster, Maxim Bonnaerens, Shin'Ichi Satoh, Akihiro Sugimoto, and Anthony Ventresque. 2023. From Scarcity to Understanding: Transfer Learning for the Extremely Low Resource Irish Sign Language. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 2000–2009, Paris, France.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Hamid Reza Vaezi Joze and Oscar Koller. 2019. MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language. ArXiv:1812.01053.

Laura Kanto and Ulla-Maija Haapanen. 2019. Fluency in Sign Language. In *Fluency in L2 learning and use*, pages 96–110. Multilingual Matters.

Vadim Kimmelman, Alfarabi Imashev, Medet Mukushev, and Anara Sandygulova. 2020. Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *PLOS ONE*, 15(6):e0233731. Publisher: Public Library of Science.

Maria Kopf, Marc Schulder, and Thomas Hanke. 2021. Overview of Datasets for the Sign Languages of Europe | ZFDM Repository. Technical report, University of Hamburg, Germany.

Maria Kopf, Marc Schulder, and Thomas Hanke. 2022. The Sign Language Dataset Compendium:

Creating an overview of digital linguistic resources. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 102–109, Marseille, France. European Language Resources Association.

Kim B. Kurz, Geo Kartheiser, and Peter C. Hauser. 2023. Second language learning of depiction in a different modality: The case of sign language acquisition. *Frontiers in Communication*, 7.

Annelies Maria Jozef Kusters, Dai O'Brien, and De Meulder. 2017. Innovations in deaf studies: Critically mapping the field. In Annelies Kusters, Maartje De Meulder, and Dai O'Brien", editors, *Innovations in Deaf Studies*, pages 1–53. Oxford University Press, United Kingdom.

Amelia Lambelet and Raphael Berthele. 2015. *Age and Foreign Language Learning in School*. Palgrave Macmillan UK, London.

Russell V. Lenth. 2024. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.3.

Dongxu Li, Cristian Rodriguez Opazo, Xin Yu, and Hongdong Li. 2020. Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1448–1458.

Chloe Marshall, Aurora Bel, Sannah Gulamani, and Gary Morgan. 2021. How are signed languages learned as second languages? *Language and Linguistics Compass*, 15(1).

Richard McElreath. 2020. *Statistical rethinking: a Bayesian course with examples in R and Stan*, 2 edition. CRC texts in statistical science. Taylor and Francis, CRC Press, Boca Raton.

Johanna Mesch and Krister Schönström. 2018. From Design and Collection to Annotation of a Learner Corpus of Sign Language. In *Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, pages 121–126, Miyazaki, Japan. European Language Resources Association.

Johanna Mesch and Krister Schönström. 2020. Use and acquisition of mouth actions in L2 sign language learners: A corpus-based approach. *Sign Language & Linguistics*, 24(1):36–62.

Lotte Meteyard and Robert A. I. Davies. 2020. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112:104092.

Jemina Napier and Lorraine Leeson. 2016. *Sign Language in Action*, 1 edition. Number 2946-2258 in Research and Practice in Applied Linguistics (RPAL). Palgrave Macmillan London.

Gerardo Ortega and Gary Morgan. 2015. Phonological Development in Hearing Learners of a Sign Language: The Influence of Phonological Parameters, Sign Complexity, and Iconicity: Phonological Development in Sign L2 Learners. *Language Learning*, 65(3):660–688.

Alejandro Oviedo, Thomas Kaul, Leonid Klinner, and Reiner Griebel. 2018. The Cologne Corpus of German Sign Language as L2 (C/CSL2): Current development stand. In *Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, pages 163–166, Miyazaki, Japan. European Language Resources Association.

Russel S. Rosen. 2004. Beginning l2 production errors in asl lexical phonology. *Sign Language & Linguistics*, 7.

Dave Sayers, Rui Sousa-Silva, Sviatlana Höhn, and et al. 2021. The Dawn of the Human-Machine Era - A forecast of new and emerging language technologies. Report for EU COST Action CA19102 'Language In The Human-Machine Era'.

Adam Schembri. 2008. British Sign Language Corpus project: Open access archives and the observer's paradox. In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 165–169, Marrakech, Morocco. European Language Resources Association (ELRA).

Marc Schulder, Sam Bigeard, Thomas Hanke, and Maria Kopf. 2023. The Sign Language Interchange Format: Harmonising Sign Language Datasets For Computational Processing. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 1–5, Rhodes Island, Greece. IEEE.

Krister Schönström and Ingela Holmström. 2022. L2M1 and L2M2 Acquisition of Sign Lexicon: The Impact of Multimodality on the Sign Second Language Acquisition. *Frontiers in Psychology*, 13.

Krister Schönström and Lorraine Leeson. 2015. Signed language L2 learners' corpora and sign language assessment. Presentation at Interkantonale Hochschule für Heilpädagogik. Zurich.

Krister Schönström and Johanna Mesch. 2022. Second language acquisition of depicting signs: A corpus-based account. *Language, Interaction and Acquisition*, 13(2):199–230.

Larry Selinker. 1972. INTERLANGUAGE. *International Review of Applied Linguistics in Language Teaching*, 10(1-4):209–232.

Özge Mercanoğlu Sincan and Hacer Yalim Keles. 2020. AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods. *IEEE Access*, 8:181340–181355.

Thad Starner, Sean Forbes, Matthew So, David Martin, Rohit Sridhar, Gururaj Deshpande, Sam S. Sepah, Sahir Shahryar, Khushi Bhardwaj, Tyler Kwok, Daksh Sehgal, Saad Hassan, Bill Neubauer, Sofia Anandi Vempala, Alec Tan, Jocelyn Heath, Unnathi Kumar, Priyanka Mosur, Tavenner Hall, Rajandeep Singh, Christopher Cui, Glenn Cameron, Sohier Dane, and Garrett Tanzer. 2023. PopSign ASL v1.0: An Isolated American Sign Language Dataset Collected via Smartphones. In *Neural Information Processing Systems*.

Garrett Tanzer and Biao Zhang. 2024. YouTube-SL-25: A Large-Scale, Open-Domain Multilingual Sign Language Parallel Corpus. ArXiv:2407.11144 version: 1.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

Christiane Von Stutterheim, Monique Lambert, and Johannes Gerwien. 2021. Limitations on the role of frequency in L2 acquisition. *Language and Cognition*, 13(2):291–321.

Bodo Winter and Paul-Christian Bürkner. 2021. Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass*, 15(11):e12439.

Oğulcan Özdemir, Ahmet Alp Kındıroğlu, Necati Cihan Camgöz, and Lale Akarun. 2020. BosphorusSign22k Sign Language Recognition Dataset. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 181–188, Marseille, France. European Language Resources Association.

10. Language Resource References

Onno Crasborn, Inge Zwitserlood, Johan Ros, and Annemieke van Kampen. 2006-2017. *Corpus NGT*. The Language Archive.

Reiner Konrad, Thomas Hanke, Gabriele Langer, Dolly Blanck, Julian Bleicken, Ilona Hofmann, Olga Jeziorski, Lutz König, Susanne König, Rie Nishio, Anja Regen, Uta Salden, Sven Wagner, Satu Worseck, Oliver Böse, Elena Jahn, and Marc Schulder. 2020. *MEINE DGS – annotiert. Öffentliches Korpus der Deutschen Gebärdensprache, 3. Release / MY DGS – annotated. Public Corpus of German Sign Language, 3rd release*. Universität Hamburg.

Adam Schembri, Jordan Fenlon, Ramas Rentelis, and Kearsy Cormier. 2014. *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2014*. University College London.

A. Appendix A: Error Distributions

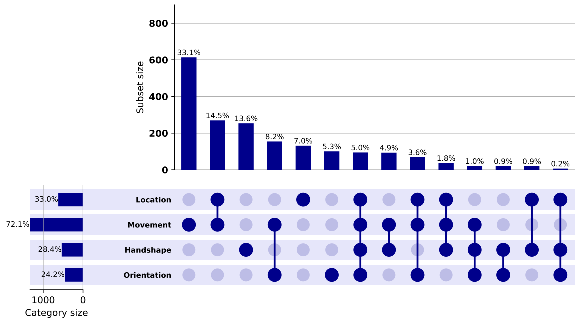


Figure 5: UpSet plots summarizing co-occurring error categories in manual components.

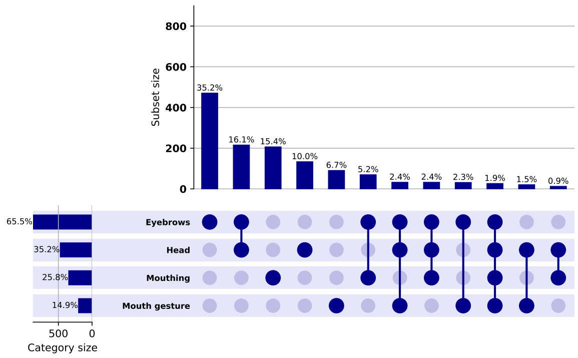


Figure 6: UpSet plots summarizing co-occurring error categories in non-manual components.

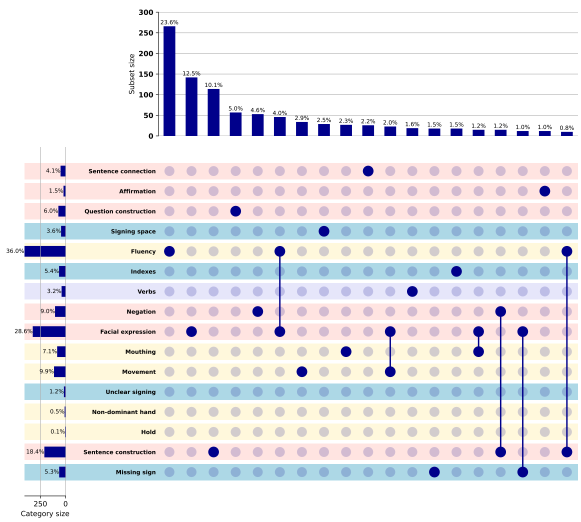


Figure 7: UpSet plots summarizing co-occurring error categories at sentence level.

B. Appendix B: Posterior Estimates

	mc model	
Parameter	*Mean± SE*	*95% CrI*
Fixed effects		
α_{P1}	-1.02 $\pm$ 0.13	[-1.27, -0.75]
β_{P2}	-0.16 $\pm$ 0.10	[-0.34, 0.03]
β_{P3}	-0.33 $\pm$ 0.09	[-0.52, -0.15]
β_{P4}	-0.24 $\pm$ 0.15	[-0.56, -0.06]
Random effects		
$\sigma_{id:P1}$	0.36 $\pm$ 0.09	[0.22, 0.57]
$\sigma_{id:P2}$	0.16 $\pm$ 0.10	[0.01, 0.38]
$\sigma_{id:P3}$	0.11 $\pm$ 0.08	[0.00, 0.31]
$\sigma_{id:P4}$	0.50 $\pm$ 0.13	[0.30, 0.80]
$\sigma_{ex:P1}$	0.36 $\pm$ 0.06	[0.26, 0.49]
$\sigma_{ex:P2}$	0.07 $\pm$ 0.06	[0.00, 0.22]
$\sigma_{ex:P3}$	0.12 $\pm$ 0.09	[0.01, 0.33]
$\sigma_{ex:P4}$	0.09 $\pm$ 0.07	[0.00, 0.24]
	nmc model	
Parameter	*Mean± SE*	*95% CrI*
Fixed effects		
α_{P1}	-2.01 $\pm$ 0.19	[-2.32, -1.59]
β_{P2}	-0.16 $\pm$ 0.18	[-0.53, 0.19]
β_{P3}	-0.70 $\pm$ 0.17	[-1.03, -0.37]
β_{P4}	-0.46 $\pm$ 0.16	[-0.79, -0.15]
Random effects		
$\sigma_{id:P1}$	0.50 $\pm$ 0.14	[0.29, 0.83]
$\sigma_{id:P2}$	0.57 $\pm$ 0.15	[0.32, 0.91]
$\sigma_{id:P3}$	0.38 $\pm$ 0.17	[0.05, 0.74]
$\sigma_{id:P4}$	0.45 $\pm$ 0.13	[0.23, 0.75]
$\sigma_{ex:P1}$	0.15 $\pm$ 0.07	[0.01, 0.29]
$\sigma_{ex:P2}$	0.12 $\pm$ 0.09	[0.00, 0.35]
$\sigma_{ex:P3}$	0.31 $\pm$ 0.14	[0.04, 0.60]
$\sigma_{ex:P4}$	0.14 $\pm$ 0.09	[0.01, 0.35]

Table 3: Posterior mean, standard error, and 95% credible interval (CrI) for the $phase$ predictor and the random effects in the manual (*mc*) and non-manual (*nmc*) error models.