A corpus of written Swedish texts and keystroke data from deaf and hard-of-hearing children and adolescents

Moa Gärdenfors
Department of Linguistics
Stockholm University
moa.gardenfors@ling.su.se

Robert Östling
Department of Linguistics
Stockholm University
robert@ling.su.se

Lisa Rudebeck
Department of Swedish Language and Multilingualism
Stockholm University
lisa.rudebeck@su.se

Mats Wirén
Department of Linguistics
Stockholm University
mats.wiren@ling.su.se

Abstract

Deaf and hard-of-hearing (DHH) children constitute a distinct subgroup of language learners, as they are acquiring a written language with reduced access to the corresponding spoken language. Also, many in the group have a sign language as their first language, thus making their visual-domain L1 another unique feature. We present an open access corpus of the written production of such learners, including 74 texts with associated keylogging data and correction annotation, making this the only such resource available for any language. Specifically, it provides a necessary component for CALL applications for DHH children.

1 Introduction

Deaf and hard-of-hearing (henceforth, DHH) children constitute a heterogeneous group with diverse linguistic backgrounds. Their language experiences vary depending on factors such as degree of hearing loss, age of identification, access to hearing technology, family language practices, educational setting, and much more (Gärdenfors, 2023a). DHH children grow up with diverse language experiences: some primarily use a spoken language, some acquire a sign language as their first language alongside a written language, while others are exposed to both signed and spoken languages. Consequently, DHH children differ considerably in their language development and literacy experiences. Despite this heterogeneity, research has consistently shown that, as a group, DHH children face greater challenges in writing than their hearing peers. At the same time, relatively little is known about their writing development, as research has traditionally focused more on reading than on writing (Mayer and Trezek, 2018).

The work reported here has its basis in the dissertation work of Moa Gärdenfors (2023b) on writing in DHH children, and provides the following novel contributions. First, an additional data collection has been made, and secondly, the entire dataset has been annotated, using tools and methodology from SweLL, a learner corpus for Swedish (Volodina et al., 2019, 2025). The resulting resource is called the Signwrite corpus (Signing Children and Adolescents' Writing Development; see Section 3). It is the first corpus for any language that includes both DHH children's writing and keylogging data and that is open access. As such, it is aimed at providing a necessary component for both CALL applications and research on second-language acquisition (SLA) for DHH children. We provide analyses of the data (Section 4) and conclude with ideas for future work (Section 5). The Signwrite corpus is publicly available under a CC-BY 4.0 license (Gärdenfors et al., 2026).

2 Background

2.1 Keylogging data and written text

Studies on writing development, whether in DHH subjects or in general, increasingly rely on data about the writing *process* in addition to the written *product* (text). The products are typically narratives elicited from subjects using a visual stimulus. For example, Wolbers et al. (2014) collected personal-narrative writing samples from DHH students in grade 6–8 who used American Sign Lan-

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/.

Proceedings of the 15th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2026)
pages 229–243, November 19–20, 2026, University of Tartu Library

guage (ASL) as their L1, and analysed features of this in the students' writing of English.

The writing process is analysed by means of keylogging (also known as keystroke logging), which captures the temporal behavior of the user by recording each keystroke. A frequently used keylogging system is ScriptLog (Johansson et al., 2018). This functions as a simple word processor without a spellchecker and records all keystrokes and mouse movements, timestamping each action to generate a temporal log of the writing process. This enables unobtrusive documentation of a subject's writing behavior while allowing researchers to replay the writing session. It is also possible to extract output for further analyses, including descriptive statistics on writing time, the number of inserted and deleted characters, pause duration and location and revisions. Furthermore, ScriptLog output can be exported to the keystroke-logging software Inputlog (Leijten and Van Waes, 2013) to enable more extensive pause and revision analyses.

Writing-process research began in the late 1970s and early 1980s, initially using methods such as think-aloud protocols and video recordings. With the increasing availability of personal computers, keystroke logging emerged in the 1990s as a more automated and fine-grained method for studying the temporal dynamics of writing, and several such systems were developed during this period (Wengelin and Johansson, 2023).

2.2 Keylogging data in CALL

A number of previous resources aimed at supporting CALL applications and containing keylogging data have been published. The Pro-TEXT corpus (Miletic et al., 2022) contains French texts produced by a variety of groups, ranging from young children to academics, annotated with syntactic information. The KUPA-KEYS corpus (Velentzas et al., 2024) consists of essays from English L1 and L2 writers, along with annotations of writer proficiency. The KLiCHe corpus (Tian et al., 2025) is an extensive resource containing nearly 5000 argumentative essays from adult writers in the United States, along with holistic scores. The GENIUS corpus (Schaller et al., 2026) is a similar resource containing short essays from German high school students, which have been annotated with a division into sections, as well as an analysis of the arguments and claims produced. Apart from analyzing the writing process, these resources have been used for CALL tasks such as prediction of essay scores (Tian et al., 2025), of students' overall grades (Schaller et al., 2026), and of general writing proficiency (Babalola et al., 2026).

We note that in none of the studies on CALL applications with keylogging data that we have been able to locate, the DHH group is represented to any significant degree or even mentioned. Among the 4992 participants from Tian et al. (2025), 29 different native languages are reported — none of them a signed language.

2.3 Writing in DHH children

As mentioned above, research on writing in DHH children is less extensive than research on reading, and studies that specifically address bilingualism, including sign language proficiency, are even more limited, see a systematic review by Zhang et al. (2024).

Schönström (2010, 2014) collected and analysed hand-written narratives produced by deaf students in grades 5 and 10, to investigate the development of Swedish as an L2. The results showed an implicational order in the students' development of Swedish that largely followed the grammatical learning order predicted by Processability Theory. The study also found a positive relationship between Swedish and Swedish sign language (STS) proficiency, suggesting that a well-developed first language may support the development of written Swedish.

Vizzi et al. (2023) investigated writing and spelling in 16 deaf young adults educated in bilingual settings. Compared with hearing controls, the deaf participants produced shorter and less informative texts, with fewer adjectives and subordinate clauses, and showed more semantic, morphological, and syntactic errors. In contrast, spelling errors were relatively infrequent, suggesting that their main difficulties concerned broader grammatical and lexical-semantic aspects of written language rather than spelling itself.

Wengelin (2002) investigated written-language production in nine congenitally deaf adults who regarded STS as their first language, alongside adults with reading and writing difficulties and hearing controls. The participants completed five different writing tasks, and their texts and

writing processes were analysed with regard to spelling, grammar, vocabulary, syntactic complexity, pauses, and editing. Wengelin (2002) represents an early study using keystroke logging to investigate written-language production in deaf adults. The results showed that the deaf participants had some spelling difficulties but, more prominently, grammatical difficulties, particularly with the syntactic rules of written Swedish. Their writing processes were also characterised by relatively linear text production, limited initial planning, and a low proportion of content-related editing.

Probably the first study of writing in DHH children using keylogging was conducted by Asker-Árnason et al. (2010). The results showed differences between children with cochlear implants (CI) and their hearing peers in both the writing process and the written product. Children with CI spent more time pausing, produced text more slowly, and showed lower narrative ability and fewer complex clauses. They also made more grammatical errors, although fewer spelling errors, and their texts had higher lexical density. The clearest group differences were found for pause time and lexical density (Asker-Árnason et al., 2012).

Gärdenfors (2023b) studied the writing of DHH children from both bilingual and cognitive writing perspectives, including analyses of the writing process using ScriptLog. The results showed that DHH children's written texts were largely comparable to those of hearing peers, especially at the age of 10–11, although they tended to produce shorter texts and showed differences in lexical density. Their writing processes revealed a slower, more revision-oriented writing style, likely reflecting increased cognitive demands during writing. Age, gender, hearing degree, and sign language proficiency were important predictors, while sign language skills did not hinder written language development and even seemed to support certain aspects of writing. Intragroup analyses further revealed that deaf children, with no functional hearing, and hard-of-hearing children, who still could make use of residual hearing, employed partly different writing strategies, reflecting their differing access to auditory and visual language input (Gärdenfors et al., 2019; Gärdenfors, 2021, 2023a,b; Gärdenfors and Johansson, 2023).

2.4 Corpora of writing in DHH children

In spite of the many studies of writing in DHH children, there are very few open access corpora in this domain. We are only aware of one other such case, namely, the Thai Deaf Corpus (TDC) by Traitruengsakul and Chuangsuwanich (2024), collected among 322 deaf students in grades 7–12. TDC is primarily aimed at CALL applications, and specifically grammatical error correction (GEC). However, it differs from Signwrite in two respects: First, the data collection for TDC involved each child writing isolated *sentences* (on average 85 per participant) rather than connected text. Secondly, TDC does not include keylogging data. It is thus not directly comparable to Signwrite.

2.5 The SweLL infrastructure

For annotation of Signwrite, we used the existing infrastructure of SweLL, a learner corpus of L2 Swedish (Volodina et al., 2019). The end product of the project, referred to as the SweLL-gold corpus, contains 502 essays collected from 321 (presumably hearing) adults (Volodina, Elena et al., 2025). The learners were recruited from schools and universities, largely through placement tests in SFI courses (Swedish for Immigrants) and university admission tests. They represented a variety of L1s, the most frequent being Arabic, Dari, Kurdish, Persian, Tigrinya, Turkish, English, Greek, Russian and Spanish.

A key aim of the SweLL project was to provide annotation of the data that facilitated both development of CALL applications and research in SLA. To this end, the project involved creating a parallel corpus consisting of the original texts along with *normalized* versions of these using a purpose-built tool, Svala (Wirén et al., 2019). A normalized (*target*) version is created by a human annotator by editing a copy of the original according to the norms of written standard Swedish (Bylin and Melander, 2023; Rudebeck and Sundberg, 2024, pp. 138–142) while retaining (the annotator's interpretation of) the meaning of the original. Also, the form of the original should not be unnecessarily altered. Normalizing a text thus involves balancing these conflicting goals (for guidelines, see Rudebeck et al. 2021). The original and target tokens are automatically aligned during editing, with Svala incrementally displaying a graph of this. Links may also be manually adjusted.

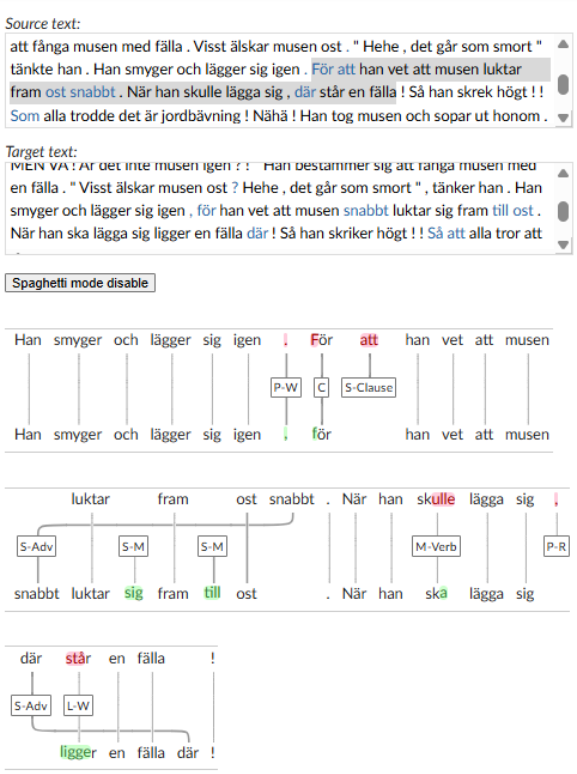


Figure 1: Svala in correction annotation mode (detail)

Once the text is normalized, a *correction annotation* is made, using a different mode in Svala (see Figure 1). (This is also known as error annotation; compare Lüdeling and Hirschmann 2015.) Here, the annotator selects relevant tokens in the source and target texts, or links that connect them, and assigns correction tags using a taxonomy developed in the SweLL project (for guidelines, see Rudebeck and Sundberg 2021). The taxonomy is grouped into orthographic, lexical, morphological, punctuation and syntactic corrections, each of which contains several specific correction tags (see Table 5). In the graph in Figure 1, correction tags are displayed on the links, given that corrections are relations between original and target tokens.[1]

3 The Signwrite corpus

3.1 Data collection

Participants in the Signwrite corpus were recruited from four groups of children and adolescents, defined as follows: deaf, hard-of-hearing, CODA, and hearing.

1. The deaf group refers to individuals who cannot effectively perceive or use spoken Swedish, do not use hearing technology, and primarily communicate through STS as their L1, with written Swedish as their L2.

2. The hard-of-hearing group refers to individuals who are deaf and use cochlear implants or have partial hearing loss and use hearing aids to enhance their auditory perception. This group communicates through spoken and written Swedish alongside STS (with exception for two individuals).

3. The CODA group refers to children of deaf adults, who are normal-hearing, but typically use STS at home with their deaf parents and are therefore often bilingual in a spoken and a signed language.

4. The hearing group refers to hearing children who do not have knowledge in STS.

Although CODA are not part of the DHH group, they increase the value of Signwrite by providing a highly relevant comparative baseline of children proficient in both STS and spoken Swedish. Likewise, hearing children constitute a very useful control group.

Signwrite is based on two data collections. The first was carried out as part of the dissertation work of Gärdenfors (2023b), involving 63 participants. Gärdenfors subsequently decided to make as much as possible of this data open access, and 44 of the participants consented to this. The second data collection was carried out by recruiting another four DHH children and adolescents, and by creating a longitudinal component. The latter was brought about by inviting the participants from the first data collection to take part in an additional session, taking place 5–7 years later. Altogether 26 DHH and CODA participants agreed to this.[2] Thus, the current Signwrite corpus comprises 74 texts and 74 keylogging files produced by 48 individuals.

3.2 Scenario

The two data collections described above were carried out using the exact same scenario. Each

[1] We refer broadly to normalization and correction annotation as "annotation". Normalization can be seen as a kind of annotation in the sense that it adds corrected tokens aligned with the original ones.

[2] The dissertation project was approved by Etikprövningsnämnden in Stockholm (2018/5:6), but for the prolongation project the approval needed to be expanded in order to make the data open access. For this purpose, a supplementary application was submitted to Etikprövningsnämnden, which was also approved (2024-02109-02).

Final text (Swedish original)	Linear input (Swedish original)
En snögig natt, där bor rosa pantern, han älskar sova. Det är bästa som han vet. Han sover jätte djupt. Men plötsligt hörde han någon knackar dörren. Han rusade ut och kikade om någon är där. Men han ser ingen.	<11.7>En<BACKSPACE3><2.1><BACKSPACE8>En snögig <BACKSPACE4>gig natten <2.9><BACKSPACE3>, där bor ett hus<BACKSPACE9>r rosa pantern <BACKSPACE3>n <BACKSPACE2>r n<BACKSPACE2>m <BACKSPACE2>n <BACKSPACE1>, han älskar sova <15.6><LEFT39><RIGHT1><LEFT1><2.2><RIGHT39><2.1>. Han sover djupt och djupt. <BACKSPACE1> Men<2.3>plötsligt hörde han någon knackar <BACKSPACE1>de dörren<15.7>. Han rusar <BACKSPACE2>de u<BACKSPACE1> ut och kikade om någon är där. Men han såg inte att <BACKSPACE11>er ingen.
Final text (English translation)	**Linear input (English translation)**
A snowgy night, there lives the pink panther, he loves sleep. It's best he knows. He sleeps very deeply. But suddenly he heard someone knocking the door. He rushed out and looked to see if someone is there. But he sees no one.	<11.7>A<BACKSPACE3><2.1><BACKSPACE8>A snowgy <BACKSPACE4>ogwy night <2.9> <BACKSPACE3>, there lives a house<BACKSPACE9> r pink panther <BACKSPACE3>n<BACKSPACE2>r n<BACKSPACE2>m <BACKSPACE2>n <BACKSPACE1>, he loves sleeping <15.6><LEFT39><RIGHT1><LEFT1><2.2> <RIGHT39><2.1>. He sleeps deeply and deeply. <BACKSPACE1>But<2.3>suddenly he heard someone knocki <BACKSPACE1>ing the door <15.7>. He rushed out <BACKSPACE2>de u<BACKSPACE1>out and looked to see if someone is there. But he did not see that<BACKSPACE11>ees no one.

Table 1: The upper left cell shows the beginning of a text by a 12-year-old DHH participant of the Signwrite data collection. The upper right cell shows the corresponding input to ScriptLog, that is, a step-by-step representation of all events during editing, including character input, cursor movements and pauses. The threshold for a pause was set at 2 seconds of inactivity, with durations shown inside angle brackets, for example, <11.7> (here rounded from ScriptLog's millisecond resolution). Angle brackets are also used for the backspace key and left and right arrow keys. Notation such as <BACKSPACE3> means that the backspace key was pressed 3 times in succession. The English translation attempts to mimic the participant's behavior in the Swedish original (including lexical, grammatical and spelling errors), and is here provided for illustrative purposes only.

		Gender		Swedish		STS		Word counts	
Hearing background	Age span	Male	Female	Written	Spoken	Yes	No	Total	Mean
Deaf	8–17	2	11	13	0	13	0	3404	261.8
Hard-of-hearing	8–21	7	25	32	32	30	2	11 021	344.4
CODA	10–18	6	14	20	20	20	0	8781	439.1
Hearing	9–11	6	3	9	9	0	9	2341	260.1
Total	8–21	21	53	74	61	63	11	25 547	345.2

Table 2: Participant metadata by hearing background, including age range, gender distribution, and knowledge of written Swedish, spoken Swedish, and STS. The total number of words (tokens excluding punctuation) is given, as well as the mean number of words per text.

session was conducted on a computer using ScriptLog, resulting in a keylogging file and a text file (see Table 1 for an example). The participants were shown a two-page Pink Panther cartoon and were asked to retell the story in written form. They were informed that they could interpret the cartoon freely, choose any name for the Pink Panther, and use their imagination when composing the story.[3] To minimize pressure and reduce the risk of unfinished texts, participants were given unlimited time to complete the task. Also, they were not informed in advance that their keystrokes would be recorded, but told about this only after completing the task. This decision was made to ensure that the writing process remained as natural as possible and to minimize any potential stress or self-consciousness associated with being monitored. Each child completed the writing task individually and without interruption.

3.3 Metadata

Table 2 summarizes personal metadata along with the number of words written in the original texts. Personal metadata includes information about hearing background, age, gender (self-reported at the first recording session) and language proficiency. As expected, the linguistic backgrounds differed across the groups. Nearly all deaf, hard-of-hearing, and CODA participants had knowledge of STS in addition to written and/or spoken Swedish. Everyone outside the deaf group had knowledge of spoken Swedish. In contrast, none of the hearing participants reported any knowledge of STS. As for the gender distribution, it is uneven with girls constituting 72% of the participants (53 girls and 21 boys). Table 3 shows the age distribution of the participants, which varied between 8 and 21 years. Table 4 shows the hearing background of the 27 participants in the longitudinal data collection.

3.4 Annotation

For the purpose of the normalization and correction annotation of Signwrite, we re-used the tools and guidelines from SweLL, as outlined in Section 2.5. This was facilitated by the fact that the annotator in Signwrite (the third author of this paper) also carried out the major part of these tasks in the construction of the SweLL corpus. Table 5 shows the outcome of the correction annotation of Signwrite in terms of tag frequencies.

[3] One advantage of this scenario is that it does not involve personal data in the sense of GDPR (Megyesi et al., 2018), thereby obviating the need for pseudonymization.

Age interval	Participants
8–9	9
10–11	25
12–13	10
14–15	4
16–17	10
18–21	16
Total	74

Table 3: Age distribution among participants in the corpus. Note that individuals in the longitudinal part are here counted twice.

Hearing background	Participants
Deaf	7
Hard-of-hearing	9
CODA	10
Total	26

Table 4: Overview of individuals who contributed longitudinal data. Each individual participated on two occasions with an intervening time of 5–7 years. The resulting 52 texts (twice the number of participants) are included in the 74 texts in Table 2.

We have not checked the inter-annotator agreement for Signwrite, but an analysis of the reliability of the SweLL correction annotation as measured on 50 essays (10% of the corpus) was made by Volodina et al. (2025, p. 272). Here, two annotators got an inter-annotator agreement of 0.88 according to Fleiss's kappa and 0.76 according to Krippendorff's alpha (Artstein and Poesio, 2008; Krippendorff, 2019).

4 Preliminary analyses

4.1 Comparison with SweLL

With Signwrite and SweLL, we are in a position to make comparisons of errors in written Swedish by subjects with different hearing status and language proficiency. Whereas the DHH and CODA subjects in Signwrite are proficient in STS and have Swedish as their strongest written language, the subjects in SweLL-gold are L2 learners of Swedish who most commonly identify another language as their strongest written language. (Hearing status was not documented in SweLL,

Orthography	O	Orthography (spelling)	717
1145	O-Cap	Upper/lower case	270
	O-Comp	Spaces and hyphens within compounds and between words	158
Lexicon	L-Der	Word formation (compounding or derivation)	64
551	L-FL	Non-Swedish word corrected to Swedish word	0
	L-Ref	Choice of anaphoric expression	277
	L-W	Word choice (other)	210
Morphology	M-Adj/adv	Adjective form corrected to adverb form	5
935	M-Case	Nominative vs genitive/accusative	33
	M-Def	Definiteness: articles; forms of nouns, adjectives	212
	M-F	Form corrected, same category (e.g. one plural suffix corrected to another plural suffix)	46
	M-Gend	Gender	50
	M-Num	Number	23
	M-Other	Other inflection-related correction	0
	M-Verb	Verb forms; auxiliaries	566
Syntax	S-Adv	Word order: adverbial placement	67
632	S-Clause	Basic clause structure	91
	S-Comp	Compound versus multi-word expression	3
	S-Ext	Extensive and complex syntactical correction	21
	S-FinV	Word order: placement of finite verb	33
	S-M	Word missing (added in target), other	179
	S-Msubj	Subject missing (added in target)	52
	S-Other	Other syntactical correction	0
	S-R	Word redundant (removed from target)	142
	S-Type	Change of phrase type/part of speech	29
	S-WO	Word order (other)	15
Punctuation	P-M	Punctuation missing (added in target)	988
1306	P-R	Punctuation redundant (removed from target)	56
	P-Sent	Sentence segmentation	31
	P-W	Wrong (i.e. changed) punctuation	231
Other	C	Consistency correction, necessitated by other correction	593
599	X	Unintelligible string	4
	Cit-FL	Non-Swedish word kept (i.e. no correction)	2
	Total		5168

Table 5: The correction taxonomy in Signwrite: main groups, specific correction tags, short descriptions of these, and their frequencies in Signwrite. Note that the taxonomy and annotation guidelines were adopted from SweLL. Further details about the taxonomy are provided by Rudebeck and Sundberg (2021).

but none of the participants reported knowledge of any sign language and they were presumably all hearing.)

For the comparison here, we have used a subset of Signwrite that contains texts written by DHH and CODA subjects who were 16 years or older (26 out of 74 texts), which is the same age cohort as the subjects of SweLL-gold. Furthermore, we have used a subset of SweLL-gold that includes texts solely with the labels ”narrative” and/or ”descriptive” (comprising 135 of 502 texts) to best fit the retellings of a cartoon in Signwrite.

The two groups of writers differ primarily in their language proficiency profiles. The Signwrite subset comprises deaf, hard-of-hearing, and CODA writers, nearly all of whom report proficiency in STS and have Swedish as their strongest written language. In contrast, the SweLL-gold subset comprises learners of Swedish as a second language, most of whom identify another language as their strongest written language. None of the SweLL-gold participants reports proficiency in a sign language. Hearing status is not documented for the SweLL-gold participants.

Table 6 juxtaposes correction annotation data from the SweLL-gold and Signwrite subsets. For the sake of comparison, the number of tag occurrences is related to the number of tokens in the original (not normalized) texts in each corpus. This relative number is presented as a percentage, referring to the number of tags in each category per 100 tokens in the corpus of original texts, in the respective corpus subsets.

The relative frequencies in Table 6 indicate that most correction tags occur more frequently in SweLL-gold than in Signwrite. The exceptions are found among corrections related to punctuation and sentence segmentation. Only three tags have a clearly higher frequency in the Signwrite subset than in the SweLL-gold subset, namely P-W, P-Sent, and C. While the C tag may be used for various corrections necessitated by another correction, their most frequent usage is to indicate that a lower case letter has been changed to upper case, as a consequence of the insertion of a sentence-segmenting punctuation mark in the normalized version of the text. Three tags have about the same frequency in both subsets, namely O-Cap, L-Ref, and P-M.

The higher frequency of corrections relating to punctuation and sentence segmentation may have to do with the material writing conditions. The Signwrite texts were written on computers, contrary to the texts in the SweLL-gold subset, which mainly includes texts which were handwritten and subsequently transcribed (but also some texts which were written on computers).

The correction category that stands out in this context is L-Ref (choice of anaphoric expression), which is slightly more frequent in the Signwrite subset than in the SweLL subset. Among the correction categories with higher or equal frequency in the Signwrite material, this is the only one which is not related to sentence segmentation and punctuation.

A typical L-Ref correction in the Signwrite material is the change of *han* ’he’ to a name or a noun. Such corrections are made when the local textual context supports an anaphoric interpretation of *han* ’he’, which, given the broader context and the cartoon itself, is unlikely to be the one intended by the writer. The higher frequency of corrections of this kind may have to do with the nature of the task. The cartoon may lead the writer to choose a pronoun that depends on a deictic association with an image, whereas the normalization is made without taking the cartoon into consideration. However, it is also possible that the higher frequency of corrections of anaphoric expressions may be attributed to factors related to the specific language experiences and proficiencies of DHH and CODA subjects. This indicates that anaphoric expressions may be an interesting aspect of the Signwrite data to further analyze.

4.2 Longitudinal data

The longitudinal data shows clear writing development across deaf, hard-of-hearing and CODA groups, although the developmental patterns differ. Deaf students show particularly strong improvements in spelling accuracy and syntactic complexity, alongside substantial increases in text length. Hard-of-hearing students display a more stable profile, with development mainly in lexical diversity and syntactic complexity. CODA students show strong growth in both lexical diversity and syntactic complexity, often reaching the highest levels at the second occasion. Overall, the findings suggest that the three groups develop along partly different trajectories, with deaf students showing more pronounced gains in formal accuracy, and the HoH and CODA groups show-

Category	SweLL-gold		Signwrite	
	Count	Percent	Count	Percent
O	1076	4.55	83	0.78
O-Cap	130	0.55	70	0.65
O-Comp	130	0.55	33	0.31
L-Der	88	0.37	25	0.23
L-FL	25	0.11	0	0.00
L-Ref	118	0.50	62	0.58
L-W	639	2.70	63	0.59
M-Adj/adv	20	0.08	1	0.01
M-Case	23	0.10	7	0.07
M-Def	564	2.38	67	0.63
M-F	98	0.41	7	0.07
M-Gend	206	0.87	17	0.16
M-Num	183	0.77	8	0.07
M-Other	25	0.11	0	0.00
M-Verb	550	2.32	158	1.48
S-Adv	114	0.48	21	0.20
S-Clause	203	0.86	21	0.20
S-Comp	17	0.07	2	0.02
S-Ext	67	0.28	0	0.00
S-FinV	158	0.67	5	0.05
S-M	405	1.71	45	0.42
S-Msubj	115	0.49	7	0.07
S-Other	14	0.06	0	0.00
S-R	264	1.12	38	0.35
S-Type	108	0.46	9	0.08
S-WO	38	0.16	1	0.01
P-M	371	1.57	153	1.43
P-R	89	0.38	15	0.14
P-Sent	6	0.03	6	0.06
P-W	70	0.30	110	1.03
C	219	0.93	129	1.20
X	59	0.25	0	0.00
Cit-FL	7	0.03	2	0.02
Total	6199	26.20	1165	10.88

Table 6: Comparison of corrections between comparable subsets (ages 16+, narrative/descriptive texts) of SweLL-gold and Signwrite. Percentages refer to the number of corrections relative to the total number of tokens. SweLL-gold tokens: 23657. Signwrite tokens: 10706.

ing greater refinement in lexical and syntactic aspects of writing.

These preliminary results are supported by correction annotation data. Table 7 presents the frequencies of the correction categories in the entire corpus with longitudinal data (including all students, i.e. deaf, hard-of-hearing, and CODA). The first two columns in Table 7 show the figures for each session. The last column shows the change in frequency between the two sessions, through a ratio calculated as the frequency for each tag in session 2, divided by the frequency for the same tag in session 1. A ratio below 1 indicates a decrease in the correction frequency between the two sessions, while a ratio above 1 indicates an increase. Generally, it may be assumed that the lower this ratio, the higher the change toward standard norms between the two sessions, that is, the greater decrease of "errors" of this particular kind.

In Table 7 orthographic and punctuation corrections have been merged into a common category. Within each general category, the category with the lowest ratio (session 2/session 1), indicating the greatest change toward norm-adherence, is presented first, followed by the second lowest, etc.

Almost all correction categories decrease in frequency between session 1 and session 2, indicating a general development toward adherence to standard norms. Among the general correction categories, the most significant development occurs within orthography and punctuation, followed by syntax. Within lexicon and morphology, the decrease in the correction frequency is somewhat lower. Interestingly, two correction categories are more frequent in session 2 than in session 1, namely M-Gend (gender) and P-W (changes of punctuation marks). (S-Comp and Cit-FL occur only in session 2, but only with 2 instances each.)

Table 8 compares the correction frequencies between the three sub-groups of students: deaf, hard-of-hearing, and CODA.[4] The comparison is made by dividing the frequency of the correction category (instances/100 tokens) in texts written by the specific group of students in one session with the frequency of the same correction category in all texts (by all students from all the three groups) in the same session. For example, the ratio 8.85 for S-FinV (placement of finite verb) among deaf students in session 1 indicates that this correction category is 8.85 times more frequent in texts written by deaf students than in all texts taken together in session 1. This ratio may be compared with 4.29, which is the ratio for the same correction category in texts by the same group (deaf students) in session 2. Although this type of correction (placement of finite verbs) remains much more frequent in texts by deaf students, the difference from texts by the other student groups has decreased between the two sessions, indicating that the deaf students have caught up some with the other students regarding correct placement of finite verbs in their texts.

In Table 8, black cells indicate a higher frequency for the correction category in texts by the specific student group than in the whole sample (ratio > 1.30), and white cells indicate a lower frequency in texts by the specific student group than in the whole sample (ratio < 0.70). Gray cells indicate approximately the same frequency in texts by the specific student group as in the whole sample.

Generally, the figures presented in Table 8 indicate that, with regard to syntax and lexicon, deaf students write texts with a lower degree of norm-adherence than the other student groups, but that this difference decreases between the two sessions. There are interesting differences between specific correction categories that call for a deeper analysis. For example, we find, again, that corrections related to the choice of anaphoric expressions (L-Ref) stand out. Contrary to the other L and S categories, the frequency of L-Ref does not decrease in texts by deaf students between session 1 and session 2.

Among corrections pertaining to morphology, orthography, and punctuation, the distribution patterns are more complex. Some of the M, P, and O categories, presented in Table 8, exhibit the same general pattern as that found within the S and L categories, that is, corrections are more frequent in texts written by deaf students than in texts written by other students, and the difference decreases between session 1 and session 2. This holds for M-Def (definiteness), M-Verb (verb inflection), and P-Sent (sentence segmentation). Two categories, P-M (punctuation added) and O-Cap (upper/lower case), are over-represented in texts by deaf students in session 1, but this over-representation is eliminated in session 2. In the

[4] Correction categories within the general category "other" are not included in Table 8, nor are categories with fewer than 15 instances in the two sessions taken together.

Correction category		session 1	session 2	session 2/session 1
Syntax	S-Ext	0.15	0.00	0.00
	S-FinV	0.24	0.04	0.18
	S-Msubj	0.30	0.09	0.29
	S-Clause	0.48	0.19	0.39
	S-WO	0.10	0.04	0.42
	S-Type	0.18	0.08	0.44
	S-M	0.91	0.48	0.53
	S-R	0.68	0.36	0.53
	S-Adv	0.30	0.22	0.75
	S-Comp	0.00	0.02	
	S total	3.34	1.53	0.46
Lexicon	L-Ref	1.41	0.60	0.43
	L-Der	0.34	0.24	0.71
	L-W	0.87	0.64	0.74
	L total	2.62	1.49	0.57
Morphology	M-Adj/adv	0.03	0.01	0.28
	M-Case	0.13	0.06	0.45
	M-Def	1.23	0.60	0.49
	M-Verb	2.91	1.58	0.54
	M-Num	0.11	0.09	0.76
	M-F	0.18	0.14	0.79
	M-Gend	0.20	0.24	1.24
	M total	4.78	2.72	0.57
Orthography and punctuation	O	3.37	0.79	0.23
	P-M	4.34	1.44	0.33
	O-Comp	0.74	0.27	0.36
	P-Sent	0.11	0.05	0.46
	O-Cap	1.21	0.60	0.50
	P-R	0.22	0.13	0.60
	P-W	0.74	0.94	1.27
	P and O total	10.73	4.22	0.39
Other	X	0.03	0.00	0.00
	C	2.45	1.13	0.46
	Cit-FL	0.00	0.02	
	Other, total	2.48	1.15	0.46
Total		23.95	11.11	0.46

Table 7: Correction frequencies (instances/100 tokens) in all texts in the longitudinal study

Category		Deaf		Hard of hearing		CODA	
		session 1	session 2	session 1	session 2	session 1	session 2
Syntax	S-FinV	8.85	4.29	0.00	0.00	0.09	0.00
	S-Ext	7.40		0.50		0.00	
	S-Type	6.53	1.91	0.44	1.27	0.24	0.27
	S-WO	6.47	4.29	0.00	0.00	0.61	0.00
	S-Msubj	5.10	2.14	0.43	0.57	0.56	0.72
	S-R	4.91	2.45	0.71	0.75	0.37	0.40
	S-Adv	4.78	2.31	0.51	0.55	0.56	0.65
	S-M	3.89	2.68	1.24	0.56	0.16	0.43
	S-Clause	3.35	1.56	0.63	0.65	0.78	0.98
Lexicon	L-Der	4.77	2.60	0.68	0.51	0.43	0.51
	L-W	3.08	2.14	0.92	0.77	0.61	0.55
	L-Ref	2.30	2.33	0.63	0.57	1.02	0.62
Morphology	M-Def	6.76	3.61	0.35	0.12	0.26	0.27
	M-Verb	4.53	2.39	0.63	0.98	0.52	0.24
	M-Gend	3.89	3.83	1.44	0.31	0.00	0.00
	M-Num	0.84	3.43	1.35	0.57	0.74	0.00
	M-Case	0.00	0.61	1.72	0.82	0.63	1.37
	M-F	0.00	2.68	1.02	0.18	1.20	0.75
Orthography and punctuation	P-Sent	8.41	2.86	0.23	0.95	0.00	0.00
	P-M	2.55	0.85	0.79	0.99	0.83	1.09
	O-Cap	2.13	0.55	0.68	1.96	1.01	0.45
	O	1.27	1.18	0.47	0.41	1.38	1.40
	O-Comp	0.77	0.69	0.62	0.55	1.36	1.55
	P-W	0.77	1.30	1.14	0.68	0.94	1.10
	P-R	0.44	0.29	1.89	2.10	0.39	0.48

Table 8: Correction category frequency (instances/100 tokens) within each student group and session, divided by the category frequency in the total sample (all three groups) in the same session. Black cells indicate a higher frequency for the correction category in the specific group than in the whole sample (ratio > 1.3). White cells indicate a lower frequency in the specific group than in the whole sample (ratio < 0.7). Gray cells indicate approximately the same frequency in the specific group as in the whole sample.

case of gender-related corrections (M-Gend), the over-representation in texts by deaf students remains stable across the two sessions. The rest of the M, O, and P correction categories exhibit patterns with even clearer deviations from the general pattern among the S and L corrections, as will be described in the following.

For two M categories, the over-representation in texts by deaf students *increases* between session 1 and session 2. This holds for M-Num (number-related morphology) and M-F, that is, instances when a form associated with a particular grammatical category (e.g. number) has been changed to another form (from another paradigm or declension) associated with the same grammatical category. In other words, most M-F corrections are corrections of errors usually analyzed as over-generalizations.

For a couple of correction categories, the frequency relation is reversed compared to the general pattern, that is, corrections are *less* frequent in texts written by deaf students than in the other texts. This holds particularly for M-Case (see below) and P-R (punctuation mark removed). Moreover, some correction categories are about as frequent in texts by deaf students as in all texts, namely O (spelling), O-comp (see below), and P-W (punctuation mark changed). O and O-comp are relatively more frequent in texts written by CODA students, while P-R is more frequent in texts written by students in the hard-of-hearing group.

Among these correction categories with an equal or lower frequency in texts by deaf students, M-Case and O-comp deserve special commenting. M-Case stands for case-related corrections. The most common M-Case correction is a change between the nominative form *de* 'they' and the accusative form *dem* 'them'. O-comp stands for corrections that have to do with spaces and hyphens in and between words. The most common O-Comp correction is the removal of a space between the two parts of a compound (concatenated in Swedish). Both of these correction types (changes between *dem* 'them' and *de* 'they', and removal of spaces in compounds) correspond to norms of written standard Swedish commonly known to be frequently broken, in many text types and by many types of writer. They are norms which are subject to debate, and which tend to be included in standard descriptions of correctness issues in written standard Swedish (e.g. Bylin and Melander (2023)). It is interesting that the current data seem to indicate that these types of norm deviation are relatively less common in texts written by deaf students than in texts written by other student groups.

5 Conclusions and future work

We have presented Signwrite, a corpus of written Swedish by DHH children and adolescents, complemented with material from CODA and hearing subjects. In addition to the texts, it includes keylogging data recording the writing process and corrected versions of the texts with error annotation. All of this is being made available under a CC-BY license, in formats suitable for use by researchers in writing development, CALL and the broader NLP field. It is the first corpus for any language that includes both DHH children's writing and keylogging data and that is open access. We have here also provided preliminary analyses of the data, suggesting the potential for more targeted contrastive analyses.

Given the relatively small size of the target population and the limited amount of available data, we aim to continue expanding the Signwrite corpus as new data become available. All future data collection includes informed consent for publication in the corpus, allowing for its continued growth. A larger dataset would enable more robust comparisons between learner groups and more detailed investigations of writing development over time.

In light of the significant impact the SweLL corpus has had on Swedish CALL applications, such as grammatical error detection, grammatical error correction, and proficiency level classification, our choice to make the Signwrite annotations follow the SweLL guidelines will enable future research on how to adapt these applications to the DHH group. The addition of keylogging data, which was not present in the SweLL corpus, will allow exploring methods for these CALL applications that depend on data about the writing process.

Acknowledgements

This work has been supported by Språkbanken, a Swedish national research infrastructure consortium jointly funded by its 10 partner institutions and the Swedish Research Council (current phase 2025–28, project id 2023-00161).

AI Disclosure

☐ Not applicable (no AI tools used)
☒ Assistive use
☐ Editing
☐ Generation of minor elements
☐ Translation
☐ Compression/expansion
☐ Generation from prompt

Details: Claude Code was used to modify the code generating Table 8 and Table 7 into a different format, and GPT-6 was used for converting bibliography items in other formats into BibTeX. The results have been manually verified by the authors.

References

Ron Artstein and Massimo Poesio. 2008. Survey article: Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Lena Asker-Árnason, Viktoria Åkerlund, Cecilia Skoglund, Ingela Ek-Lagergren, Åsa Wengelin, and Birgitta Sahlén. 2012. Spoken and written narratives in swedish children and adolescents with hearing impairment. *Communication Disorders Quarterly*, 33(3):131–145.

Lena Asker-Árnason, Tina Ibertsson, Malin Wass, Åsa Wengelin, and Birgitta Sahlén. 2010. Picture-elicited written narratives, process and product, in 18 children with cochlear implants. *Communication Disorders Quarterly*, 31(4):195–212.

Damilola Babalola, Naga Buddarapu, Caleb Scott, Piotr Mitros, Paul Deane, and Collin Lynch. 2026. Measuring writing proficiency through keystroke process data: A cross-school generalizability study. In *Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con): Works in Progress*, pages 358–368, Wyndham Grand Pittsburgh Downtown, Pittsburgh, Pennsylvania, United States. National Council on Measurement in Education (NCME).

Maria Bylin and Björn Melander, editors. 2023. *Språkrådet rekommenderar: Perspektiv, metoder och avvägningar i språkriktighetsfrågor*. Språkrådet, Institutet för språk och folkminnen.

Moa Gärdenfors. 2021. The writing process and the written product in bimodal bilingual deaf and hard of hearing children. *Languages*, 6(2).

Moa Gärdenfors. 2023a. Writing development in dhh students: A bimodal bilingual approach. *Journal of Deaf Studies and Deaf Education*, 28(2):211–225.

Moa Gärdenfors. 2023b. *Writing in deaf and hard-of-hearing children: A bimodal bilingual perspective on their written products and writing processes*. Ph.D. thesis, Stockholm University, Department of Linguistics.

Moa Gärdenfors and Victoria Johansson. 2023. Written products and writing processes in swedish deaf and hard of hearing children: an explorative study on the impact of linguistic background. *Frontiers in Psychology*, 14.

Moa Gärdenfors, Victoria Johansson, and Krister Schönström. 2019. Spelling in deaf, hard of hearing and hearing children with sign language knowledge. *Frontiers in Psychology*, 10.

Moa Gärdenfors, Robert Östling, Lisa Rudebeck, and Mats Wirén. 2026. Signwrite: A corpus of written Swedish texts and keystroke data from deaf and hard-of-hearing children and adolescents [Dataset]. DOI: 10.5281/zenodo.23803517.

Victoria Johansson, Johan Frid, and Åsa Wengelin. 2018. Scriptlog – an experimental keystroke logging tool. In *Proceedings of the 1st Literacy Summit*, page 51, Porto, Portugal. European Literacy Network (ELN).

Klaus Krippendorff. 2019. *Content Analysis: An Introduction to Its Methodology*, 4th edition. SAGE Publications, Los Angeles, CA.

Mariëlle Leijten and Luuk Van Waes. 2013. Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication*, 30.

Anke Lüdeling and Hagen Hirschmann. 2015. Error annotation systems. In Sylviane Granger, Gaëtanelle Gilquin, and Fanny Meunier, editors, *The Cambridge Handbook of Learner Corpus Research*, Cambridge Handbooks in Language and Linguistics, pages 135–158. Cambridge University Press.

Connie Mayer and Beverly J Trezek. 2018. Literacy outcomes in deaf students with cochlear implants: Current state of the knowledge. *The Journal of Deaf Studies and Deaf Education*, 23(1):1–16.

Beáta Megyesi, Lena Granstedt, Sofia Johansson, Julia Prentice, Dan Rosén, Carl-Johan Schenström, Gunlög Sundberg, Mats Wirén, and Elena Volodina. 2018. Learner corpus anonymization in the age of GDPR: Insights from the creation of a learner corpus of Swedish. In *Proceedings of the 7th workshop on NLP for Computer Assisted Language Learning*, pages 47–56, Stockholm, Sweden. LiU Electronic Press.

Aleksandra Miletic, Christophe Benzitoun, Georgeta Cislaru, and Santiago Herrera-Yanez. 2022. ProTEXT: an annotated corpus of keystroke logs. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1732–1739, Marseille, France. European Language Resources Association.

Lisa Rudebeck and Gunlög Sundberg. 2021. SweLL correction annotation guidelines. Technical report, Göteborgs universitet.

Lisa Rudebeck and Gunlög Sundberg. 2024. On the other side of the error tag: Predefined normalizations as a basis for correction annotation. In *Continuing Learner Corpus Research: Challenges and Opportunities*, pages 123–152. Ackerley, Catherine and Castello, Erik.

Lisa Rudebeck, Gunlög Sundberg, and Mats Wirén. 2021. SweLL normalization guidelines. Technical report, Göteborgs universitet.

Nils-Jonathan Schaller, Thorben Jansen, Lars Höft, Hannah Pünjer, and Andrea Horbach. 2026. GENIUS keylog corpus - a German high school student corpus with keystroke logging data. In *Proceedings of the Fifteenth Language Resources and Evaluation Conference*, pages 6919–6928, Palma de Mallorca, Spain. ELRA Language Resource Association.

Krister Schönström. 2010. *Tvåspråkighet hos döva skolelever: Processbarhet i svenska och narrativ struktur i svenska och svenskt teckenspråk*. Ph.D. thesis, Institutionen för lingvistik, Stockholms universitet.

Krister Schönström. 2014. Visual acquisition of swedish in deaf children: An l2 processability approach. *Linguistic Approaches to Bilingualism*, 4(1):61–88.

Yu Tian, Scott Crossley, and Luuk Van Waes. 2025. The klicke corpus: Keystroke logging in compositions for knowledge evaluation. *Journal of Writing Research*, 17(1):23–60.

Supachan Traitruengsakul and Ekapol Chuangsuwanich. 2024. Developing a Thai Grammatical Error Correction Tool for Deaf Students. *IEEE Access*, 12:153980–153999.

Georgios Velentzas, Andrew Caines, Rita Borgo, Erin Pacquetet, Clive Hamilton, Taylor Arnold, Diane Nicholls, Paula Buttery, Thomas Gaillat, Nicolas Ballier, and Helen Yannakoudakis. 2024. Logging keystrokes in writing by English learners. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 10725–10746, Torino, Italia. ELRA and ICCL.

Francesca Vizzi, Paola Angelelli, Marika Iaia, Anthony H. Risser, and Chiara Valeria Marinelli. 2023. Writing composition ability and spelling competence in deaf subjects: A psycholinguistic analysis of source of difficulties. *Reading and Writing*, 36:1201–1226.

Elena Volodina, Lena Granstedt, Arild Matsson, Beáta Megyesi, Ildikó Pilán, Julia Prentice, Dan Rosén, Lisa Rudebeck, Carl-Johan Schenström, Gunlög Sundberg, and Mats Wirén. 2019. The SweLL language learner corpus: From design to annotation. *Northern European Journal of Language Technology*, 6:67–104.

Elena Volodina, Arianna Masciolini, Béata Megyesi, Julia Prentice, Lisa Rudebeck, Gunlög Sundberg, and Mats Wirén. 2025. SweLL with pride: How to put a learner corpus to good use. In *Huminfra handbook: Empowering digital and experimental humanities / Gerlof Bouma, Dana Dannélls, Dimitrios Kokkinakis, Elena Volodina (eds.)*, pages 251–305. University of Tartu Library, Tartu.

Volodina, Elena, Granstedt, Lena, Megyesi, Beáta, Prentice, Julia, Rudebeck, Lisa, Sundberg, Gunlög, and Wirén, Mats. 2025. SweLL-gold [dataset]. DOI: 10.23695/2K47-Y432.

Åsa Wengelin. 2002. *Text Production in Adults with Reading and Writing Difficulties*. Ph.D. thesis, University of Gothenburg.

Åsa Wengelin and Victoria Johansson. 2023. Investigating writing processes with keystroke logging. In Otto Kruse, Christian Rapp, Chris M. Anson, Kalliopi Benetos, Elena Cotos, Ann Devitt, and Antonette Shibani, editors, *Digital Writing Technologies in Higher Education: Theory, Research, and Practice*, pages 405–420. Springer International Publishing, Cham.

Mats Wirén, Arild Matsson, Dan Rosén, and Elena Volodina. 2019. Svala: Annotation of second-language learner text based on mostly automatic alignment of parallel corpora. In *Selected papers from the CLARIN Annual Conference 2018, Pisa, 8-10 October 2018*, Linköpings universitet. Linköping University Electronic Press, Linköpings universitet.

Kimberly A. Wolbers, Lisa M. Bowers, Hannah M. Dostal, and Shannon C. Graham. 2014. Deaf writers' application of American Sign Language knowledge to English. *International Journal of Bilingual Education and Bilingualism*, 17(4):410–428.

Dongbo Zhang, Sihui Ke, Junhui Yang, and Hannah Anglin-Jaffe. 2024. Sign language in d/deaf students' spoken/written language development: A research synthesis and meta-analysis of cross-linguistic correlation coefficients. *Review of Education*, 12(3):e70016. E70016 RoE-2023-11-1047.R2.