Börstell, C., Schembri, A., & Crasborn, O. (2024). Sign duration and signing rate in British Sign Language, Dutch Sign Language and Swedish Sign Language. *Glossa Psycholinguistics*, 3(1): X, pp. 1–15. DOI: https://doi.org/10.5070/G60111915

Sign duration and signing rate in British Sign Language, Dutch Sign Language and Swedish Sign Language

Carl Börstell, Department of Linguistic, Literary and Aesthetic Studies, University of Bergen, NO, carl.borstell@uib.no
Adam Schembri, Department of English Language and Linguistics, University of Birmingham, UK, a.schembri@bham.ac.uk
Onno Crasborn, Independent researcher, NL

In this article, we look at sign duration and signing rate in corpora of three sign languages – British Sign Language (BSL), Dutch Sign Language (NGT) and Swedish Sign Language (STS). We investigate whether token frequency and sociolinguistic variables (e.g., age, gender, region) influence the production rate of signing. Following Zipf's law of abbreviation, we see that a sign's duration is negatively correlated with its frequency. Both sign duration and signing rate are found to correlate with signer age, in that older signers have longer durations and lower rates than younger signers. Signers' gender, family (deaf or hearing) and age of exposure have no effect on duration or signing rate. For NGT and STS, there is no effect of region on either duration or rate. However, in the BSL data, duration and signing rate vary with region. The overall findings align with previous work on spoken languages, particularly that frequency and aging are correlated with word length and production rate, thus demonstrating such patterns across modalities of language.

1. Introduction

Regardless of language modality, production rate can vary: words can be reduced in duration/length and the number of linguistic units produced over a time span can be higher or lower. But what conditions this variability? In this article, we look at lexical frequency as well as sociolinguistic variables, such as age, gender, region and language background in relation to *sign duration*, defined as the length in time of individual signs produced in context (i.e., not isolated production) and *signing rate*, defined as the number of signs produced per time unit within an utterance. We use corpus data from three unrelated sign languages, British Sign Language (BSL), Dutch Sign Language (NGT; *Nederlandse Gebarentaal*) and Swedish Sign Language (STS; *svenskt teckenspråk*) to research these issues cross-linguistically in the signed modality.

Zipf (1949) posited the *law of abbreviation*, which states that a word's magnitude (i.e., its length) is negatively correlated with its token frequency: the more frequent a word is, the shorter it is expected to be. For example, frequent function words in English, such as *the* and *it*, are significantly shorter than low-frequency (e.g., domain-specific) words, such as *agglutinating* and *grammaticalization* (see, e.g., Bybee, 2007; Ernestus & Warner, 2011; Sigurd et al., 2004). This can be demonstrated with homophones, such as *time* and *thyme* (both pronounced /taɪm/), for which the more frequent *time* has significantly shorter duration than *thyme* when produced in speech (Gahl, 2008) – see also Wright (1979). The frequency–reduction correlation was initially attributed to the *principle of least effort*, such that language users will reduce the amount of articulatory effort to make communication more efficient (cf. Petrini et al., 2022). Work in usage-based linguistics attributes this type of reduction to routinization of production: phonetic strings that occur often are "practiced" more and thus quicker to produce (e.g., Bybee, 2007; Diessel, 2007; Lepic, 2019). Corpus-based cross-linguistic work on spoken languages has also attributed the frequency–reduction correlation to predictability and surprisal. The brain processes language based on contextual predictability, hence we are more likely to predict the next word of a sequence to be one that frequently follows the previous one(s), rather than one that rarely appears in that context (e.g., Gibson et al., 2019; Jurafsky et al., 2001; Mahowald et al., 2013; Piantadosi et al., 2011). In sign language communication, it has been shown that signs that are repeated in discourse are reduced in size and time when previously introduced, due to being established in the discourse and, thus, expected (Lepic, 2019; Stamp et al., 2024).

The length of a word is related to production rate, in that individual lexical items are produced faster or slower. The production rate (whether spoken or signed) may be influenced by many factors. Among spoken languages, there is plenty of research pointing to different sociolinguistic variables influencing production rate, e.g., age, gender and dialect (Byrd, 1994; Jacewicz et al., 2009; Keune et al., 2005). That speech rate is lowered as a function of age has been attributed to slower processing and control of the articulation apparatus (Horton et al., 2010; Jacewicz et al., 2009; Ramig, 1983). Speech rate is also part of the *perceived* age of a speaker, such that we identify lower rates as a property of aging (Skoog Waller et al., 2015). In American English, both gender and dialect have been found to correlate with speech rate, with men speaking at a higher rate than women, and certain regions having higher speech rates than others (Byrd, 1994).

To date, few studies have looked at production rate in sign languages. It has been argued that signing rate is lower than speech rate, measured as signs/words per minute, and that sign duration is reduced as signing rate increases (Grosjean, 1979; Wilbur, 2009). Higher signing rate can also lead to phonetic reduction, such as lowering the place of articulation of signs with higher locations in their citation form (Tyrone & Mauk, 2010). In a corpus study on STS, Börstell et al. (2016) found that sign duration decreases as a function of the sign's token frequency, conforming to the law of abbreviation. Additionally, Börstell et al. (2016) found an effect of age on sign durations, such that durations increase with signers' age, but found no effect of gender.

It is well known that there is substantial sociolinguistic variation in the lexicons of sign languages, conditioned by factors like age, gender and region (Bickford, 1991; Kimmelman et al., 2022; Lucas et al., 2009; Mudd et al., 2020; Safar, 2021), including the languages of this study (Börstell & Östling, 2016; Schembri et al., 2018; Stamp et al., 2014; Vermeerbergen et al., 2013). However, little is known about the potential effect of such sociolinguistic variables on duration and signing rate, other than what was found by Börstell et al. (2016) in terms of age (positive correlation) and gender (no effect). In many signing communities, there are perceptions about *"who signs in what way"* (e.g., Rowley & Cormier, 2023) – for example, some signers of BSL anecdotally report that Glasgow signers sign faster than others.

Here, we explore duration and signing rate on the basis of the following research questions:

1. Do token frequency and sociolinguistic factors (i.e., age, gender, age of exposure to sign language, deaf/hearing family, region) correlate with sign duration and signing rate in BSL, NGT and STS?
2. Do BSL signers from certain regions sign faster with regard to sign duration and signing rate compared to signers from other regions?

We expect token frequency to be negatively correlated with sign durations, based on the general law of abbreviation (Zipf, 1949) and previous work by Börstell et al. (2016) on STS. We expect age to be positively correlated with duration and negatively correlated with signing rate, on the assumption that older signers sign more slowly, following previous work by Börstell et al. (2016) on STS, as well as work on spoken languages (e.g., Horton et al., 2010). For the other variables, we do not have any predictions of possible effects nor their directionality, other than the possibility of Glasgow signers signing faster among BSL signers.

2. Methodology

For this study, we use corpus data from three sign languages: BSL (Schembri et al., 2017); NGT (Crasborn et al., 2015); and STS (Mesch et al., 2012, 2014; Öqvist et al., 2020). Sign language corpora are comparatively small, but the three corpora used here constitute some of the largest sign language corpora currently available (cf. Börstell, 2022b; Kopf et al., 2022). The data consists of ELAN annotation files (Crasborn & Sloetjes, 2008; Wittenburg et al., 2006) with segmentations for each sign produced in video-recorded sessions.

The data processing consisted of several steps. First, the ELAN (.eaf) annotation files and metadata files from the three sign language corpora were downloaded and read using R v4.4.1 (R Core Team, 2024) and the packages `signglossR` v2.2.6 (Börstell, 2022a) and `xml2` v1.3.6 (Wickham, Hester, & Ooms, 2023). Only narrative and conversational text types were included, in order to focus on naturalistic and cohesive signing, thus excluding interviews and lexical elicitation tasks. The data was further processed, analyzed and visualized using the packages `here` v1.0.1 (Müller, 2020), `tidyverse` v2.0.0 (Wickham et al., 2019), `afex` v1.3.1 (Singmann et al., 2024), `marginaleffects` v0.21.0 (Arel-Bundock et al., Forthcoming), `patchwork` v1.2.0 (Pedersen, 2024), `scales` v1.3.0 (Wickham, Pedersen, & Seidel, 2023) and `sf` v1.0.16 (Pebesma & Bivand, 2023). We use the gloss annotations from the data, which are the individually segmented and labeled signs (i.e., lexical items). For the BSL and NGT corpus data, double annotations for two-handed signs were removed, such that each sign is only represented once, whether one- or two-handed. Empty annotations without a gloss were excluded, and signs with a duration of 0 were adjusted to a minimum of 1 millisecond. The three corpora use different criteria for determining the start of signs: the BSL Corpus defines *start* as the beginning of a transitional movement away from the previous sign, whereas the NGT and STS corpora define it as the start of the articulation phase.[1] Thus, the durations and signing rates reported here can be compared *within* languages, but not necessarily *across* languages. After these steps, the resulting dataset contains approximately 50 hours of annotated data (BSL: 12.4 hours; NGT: 14.7 hours; STS: 23.2 hours), comprising 344,869 sign tokens (BSL: 59,925; NGT: 95,897; STS: 189,047) across 288 signers (BSL: 171; NGT: 75; STS: 42) representing different genders and ages (**Figures 1–2**) and geographic regions (**Figure 2**). The sign tokens were further grouped into *utterances*. Utterances were split at any pause between signs of $\geq$500 milliseconds, a threshold selected on the basis of the approximate range of pause durations reported for spoken languages (see Heldner & Edlund, 2010). Utterances were used to calculate *signing rate*, defined as the number of signs divided by the total duration of the utterance (scaled to signs per minute). Only utterances with a length of $\geq$2 signs were included, so as to differentiate the metric from individual sign durations. In total, 21,358 utterances were segmented and included (BSL: 3,047; NGT: 6,707; STS: 11,604).

[1] See annotation guidelines: BSL: https://bslcorpusproject.org/wp-content/uploads/BSLCorpus_AnnotationConventions_v3.0_-March2017.pdf; NGT: https://www.bslcorpusproject.org/wp-content/uploads/CorpusNGT_AnnotationConventions_v3_Feb2015.pdf; STS: https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-193356.

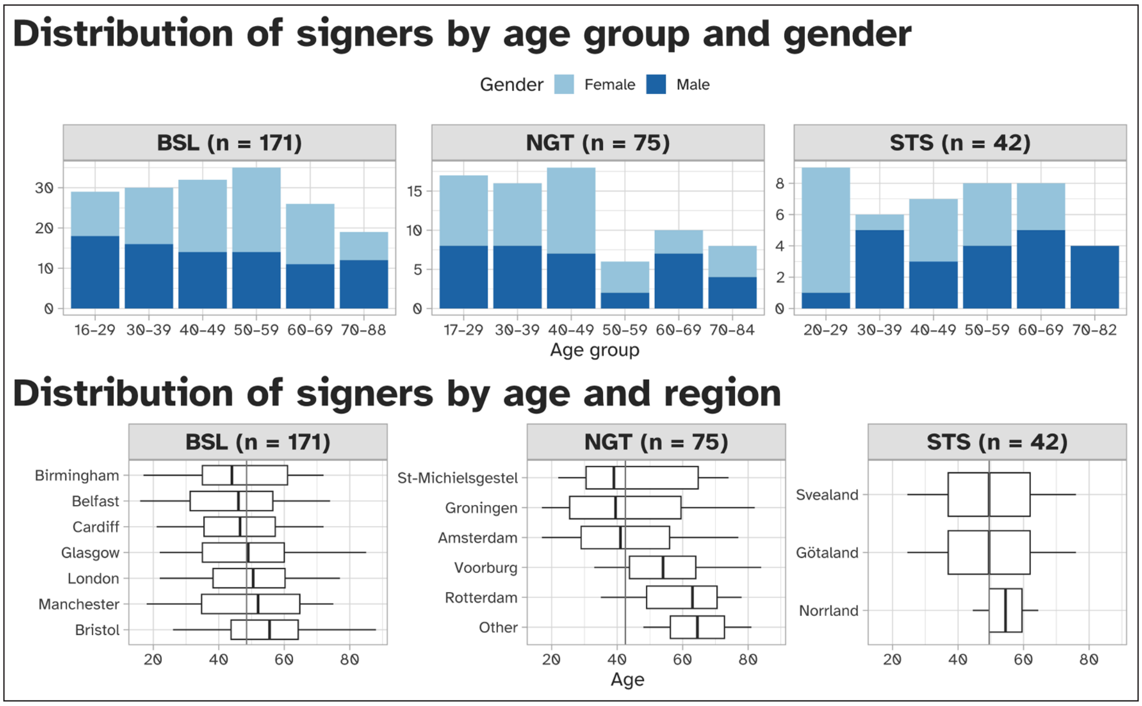

Figure 1: Distribution of signers by age, gender and region across corpora. Solid line shows corpus median.

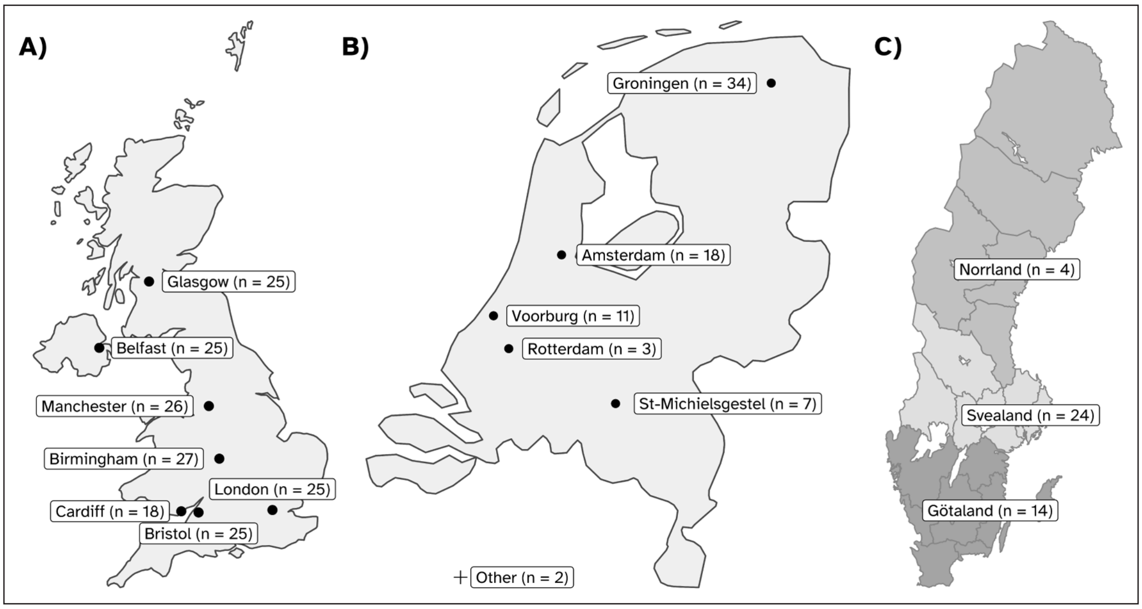

Figure 2: Geographic distribution of signers across the three corpora (number of signers per city/region in brackets): A) BSL in the United Kingdom; B) NGT in the Netherlands; C) STS in Sweden.

The data was combined with available metadata: for BSL, metadata about the signer and task is embedded in the file names; for NGT and STS, metadata is available in the repository of the respective corpus collection. Unlike BSL and NGT, signer age for STS is given as age group in approximately 10-year increments, and age for STS was thus estimated as the midpoint between the lower and upper bounds of each age group range. Age was treated as a continuous variable in the statistical modeling, but with less precision for STS. For BSL and STS, the metadata contains information about the family of the signer; NGT metadata contains information about the age of exposure to NGT. The variable *family* thus refers to whether or not the signer comes from a deaf or hearing family for BSL and STS (deaf family: BSL: n = 67 [39%]; STS: n = 15 [36%]). For NGT, *age of exposure* refers to the reported age of first exposure to sign language for each signer, defined as a continuous variable in years ($M = 2.79$; $SD = 2.83$).

3. Results

In order to evaluate the effects of individual variables, a mixed effects model was constructed for each language, with sign duration (log-scaled and z-scored) as the outcome and sign token frequency (log-scaled and z-scored), age (centered by language), gender, family[2] and region as fixed effects, with interaction between age and gender, and signer and sign as random effects. The fixed effects of these models were evaluated using likelihood ratio tests against a null model for each effect, showing that token frequency and age were the only significant effects across languages, and the only other variable that had a significant effect was region, for BSL only (**Table 1**). Based on these evaluations, a final model was constructed for all three languages together, with sign duration (log-scaled and z-scored) as the outcome and sign token frequency (log-scaled and z-scored) and age (centered by language) as fixed effects, with language, signer and sign as random effects. With this model, the effect of frequency was significant and negative ($\beta = -.295$; $t(13530) = -40.095$; $p <$

Table 1: Likelihood ratio tests of fixed effects for each language's sign duration model evaluation.

(a) BSL sign duration evaluation			**(b) NGT sign duration evaluation**			**(c) STS sign duration evaluation**		
Effect	χ^2	***p***	***Effect***	χ^2	***p***	***Effect***	χ^2	***p***
Frequency	214.97	***	Frequency	69.81	***	Frequency	1346.56	***
Age	14.43	***	Age	10.60	**	Age	11.46	***
Gender	.03	.865	Gender	.26	.611	Gender	.14	.712
Family	1.98	.160	Family	–	–	Family	3.63	.057
Region	15.09	*	Region	8.96	.110	Region	4.82	.090
Age × Gender	.18	.672	Age × Gender	.03	.857	Age × Gender	3.27	.071

[2] Family was not available for NGT. Age of exposure (mean-centered) was included in a separate model excluding gender in order to converge, but was non-significant compared to a null model (age of exposure: $\chi^2(1) = .56$; $p = .456$).

.001***) and age was significant and positive (β = .006; t(253.1) = 8.778; p < .001***) – that is, duration decreases with sign frequency, but increases with signer age.[3]

Turning to signing rate, a mixed effects model was constructed for each language, with signing rate (log-scaled and z-scored) as the outcome and utterance length (log-scaled), age (centered by language), gender, family (BSL and STS only), age of exposure (NGT only) and region as fixed effects, with interaction between age and gender, and signer as a random effect. Table 2 shows the statistics of the individual fixed effects as compared against a null model, using likelihood ratio tests for each effect. Age again shows a significant effect across languages, as does utterance length. As with the duration models, only the BSL model shows a significant effect of region. A final model was constructed for all three languages, with signing rate (log-scaled and z-scored) as the outcome and utterance length (log-scaled) and age (centered by language) as fixed effects, with language and signer as random effects. With this model, the effects of utterance length (β = –.06; t(21350) = –10.119; p <.001***) and age (β = –.01; t(227.1) = –7.028; p < .001***) were significant and negative – that is, signing rates decrease as utterance length and signer age increase.

Table 2: Likelihood ratio tests of fixed effects for each language's signing rate model evaluation.

(a) BSL sign duration evaluation			**(b) NGT sign duration evaluation**			**(c) STS sign duration evaluation**		
Effect	χ^2	***p***	***Effect***	χ^2	***p***	***Effect***	χ^2	***p***
Utt. length	32.51	***	Utt. length	27.25	***	Utt. length	177.27	***
Age	6.97	**	Age	7.76	**	Age	3.93	*
Gender	.09	.761	Gender	1.92	.166	Gender	.00	.967
Family	1.40	.237	AoE	2.23	.135	Family	2.67	.102
Region	26.91	**	Region	8.66	.124	Region	3.17	.205
Age × Gender	1.44	.231	Age × Gender	.37	.543	Age × Gender	.09	.762

Figure 3 shows the observed duration and signing rates across languages, by age group (for visualization purposes; models used continuous age).

Across the three languages, only BSL showed an effect of region on both sign duration and signing rate. **Figure 4** shows the predictions of duration and signing rate based on the individual BSL models, predicted for four ages from 20 to 80 years old. **Figure 4** shows that the Birmingham signers are among the top for both models, indicating shorter durations and higher signing rates than the others (i.e., faster signing). At the other end, Cardiff signers end up with the longest durations and second to lowest signing rates, despite Cardiff having younger than average signers in the data (cf. **Figure 1**).

[3] Comparing NGT and STS (with similar criteria for sign segmentation) in a model with language as a fixed effect, alongside frequency and age, shows no significant effect of language (β = – .054; t(106.7) = –1.308; p = .19).

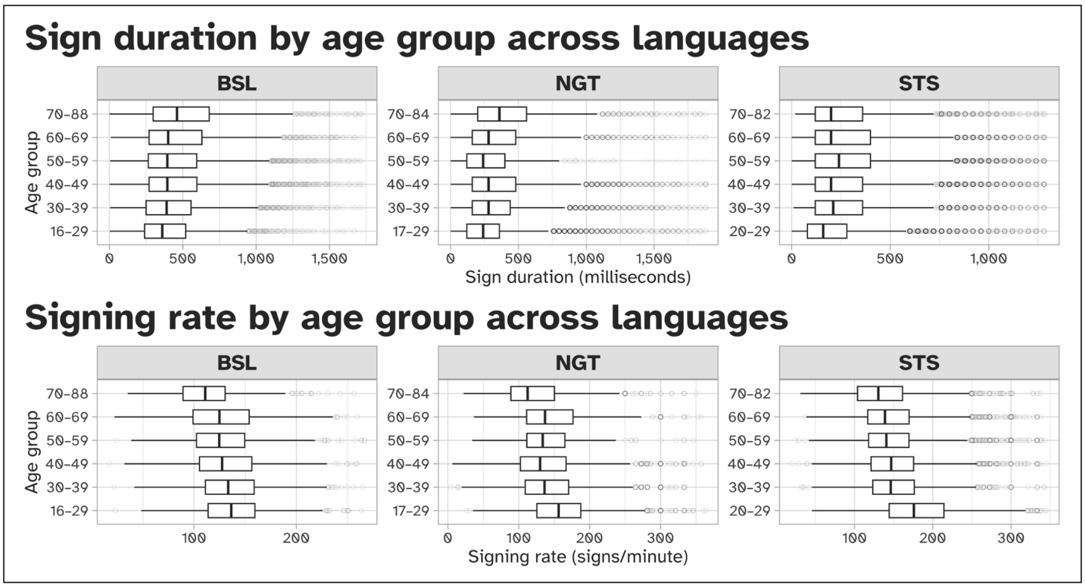

Figure 3: Sign duration and signing rates across language by age group. Durations and rates more than 3 *SDs* from the grand means by language are excluded.

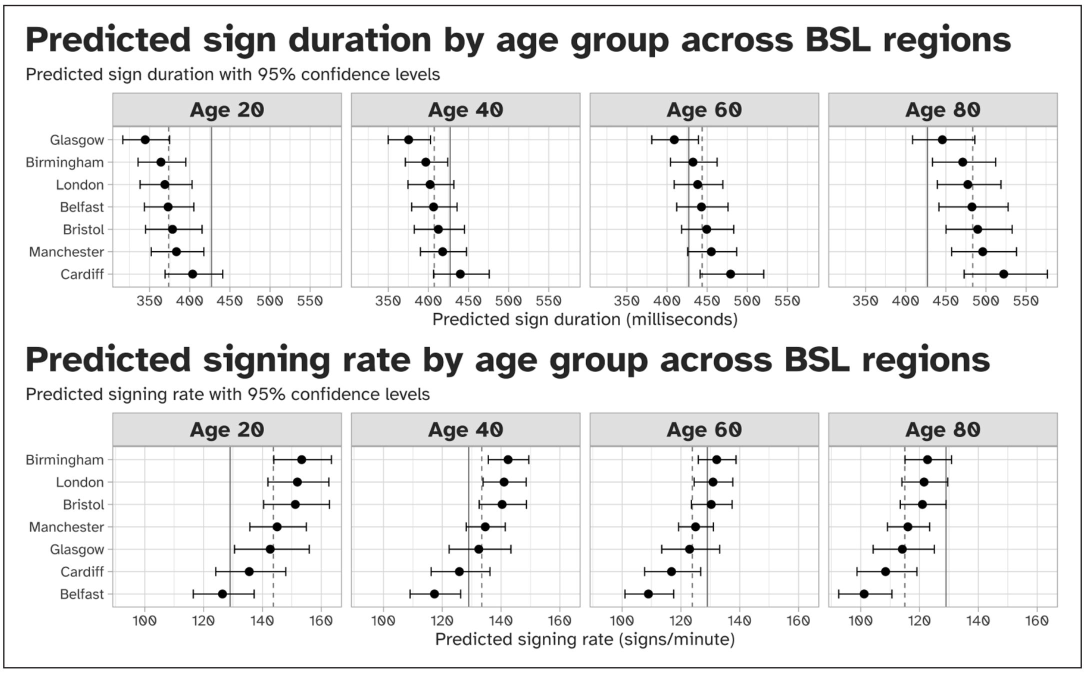

Figure 4: Predicted sign durations (top) and signing rates (bottom) in BSL by region and age. Whiskers show 95% confidence intervals. Solid lines show mean values across all ages and regions; dashed lines show the mean values across regions per age.

4. Discussion

In this study, we looked at aspects of sign duration and signing rate across three sign languages on the basis of corpus data.

First, we were able to corroborate previous work, in that a sign's token frequency is negatively correlated with its duration across all three languages. Although this finding is unsurprising, it shows how the law of abbreviation is found in the signed modality, as it has been found in spoken and written modalities.

Second, we showed that signer age correlates with sign duration (positively) and signing rate (negatively) across these languages, such that an increase in age results in longer sign duration and lower signing rate. This aligns with findings from spoken languages (e.g., Horton et al., 2010; Jacewicz et al., 2009; Ramig, 1983), and can be attributed to changes in both processing and physical production (i.e., articulation). An additional aspect to investigate in the future would be whether signing rate changes based on the age of the addressee and/or the difference in age between interlocutors in a dyad, as has been shown for spoken language (see Cohen Priva et al., 2017). That is, do younger signers accommodate their signing rate to older signers, or vice versa (cf. Stamp et al., 2016)? This question could not be addressed in this study, due to limitations in the size and stratification of the data, as it would require a more balanced sample of signers in different age group dyad combinations. In an ideal dataset for this research question, signers would be paired with younger, older and same age interlocutors in different conditions. This is not the case in our data, where most signers are paired with exactly one other signer. Furthermore, a more detailed investigation of pausing and the length and structure of sentences, utterances and turns may potentially show additional age-related differences across signers.

Third, whereas sociolinguistic variables, such as dialect and gender, have been found to influence speech rate in spoken languages (Jacewicz et al., 2009), and are known to affect lexical variation in sign languages (Bickford, 1991; Lucas et al., 2009; Schembri & Lucas, 2015), there was no significant effect of gender on duration or signing rate for any language in our study, which corroborates previous work by Börstell et al. (2016) on sign duration in STS. We additionally looked at signers' language background with regard to family and age of exposure to a sign language. While "native signer" status is sometimes claimed to influence sign language production/perception in linguistic studies, the issue of "nativeness" is particularly problematic for sign languages, where the majority of signers tend to come from hearing, non-signing families, and, thus, are not strictly "native signers" in the traditional definition (cf. Cheng et al., 2021; Quer & Steinbach, 2019). In this study, we can conclude that the datasets reflect the general trend in the corresponding deaf communities of having a majority of signers from hearing families, but that neither family nor age of exposure seems to influence sign duration or signing rate.

Lastly, region had an effect on sign duration and signing rate only for BSL. Here, it was found that Glasgow signers have shorter sign durations than average, as expected on the basis of anecdotal perceptions in the BSL community, but this prediction was not corroborated by the signing rate model. However, seeing as the signing rate dataset is much smaller than the sign duration dataset, there are potential issues with the stratification of data across all ages and regions. For example, the number of signs and utterances for each of the regions Belfast, Glasgow and Cardiff is an order of magnitude smaller than that of Birmingham, Bristol, London and Manchester. This means the data and model are less reliable for the regions with fewer data points, which is also reflected in the range of the confidence intervals in **Figure 4**. Thus, while region may be a significant factor for sign duration and signing rate in BSL, we are hesitant to make any definitive claims about individual regions based on these data and analyses alone. Additional data may facilitate a reanalysis of possible differences between duration and signing rates across regions. Another approach is to address *perceived* signing rate by experimentally varying duration of individual signs and overall signing rate in different combinations, and have participants view and provide ratings for different conditions. We leave these possibilities for future work.

In summary, apart from seeing frequency effects in sign durations, we also see an effect of aging on both sign duration and signing rate across three unrelated sign languages. Aging as an effect on production rate is well established in spoken language research (e.g., Horton et al., 2010; Jacewicz et al., 2009; Ramig, 1983), and can now be shown also across different sign languages. In the bigger picture, the results of this study shed new light on the question of production rate as a phenomenon across modalities. The spoken and signed modalities come with different affordances in terms of the transmission channel and the main articulators employed. This may, in turn, lead to differences, such as the possibility for more simultaneous expression and a higher degree of iconicity in the signed modality. In principle, this could – as one reviewer points out – lead to restrictions on the amount of reduction that is possible, but yet we find the same effects as for spoken languages, in that frequency leads to articulatory reduction (shorter duration). Thus, our findings point to important similarities in how language is shaped across modalities, regardless of the way in which it is expressed. This is seen in how words/signs of a language compare to each other with regard to frequency of use, with reduction as part of efficiency, expectedness and routinization in communication, as well as in how the physical aging of speakers/signers alters the way language is physically produced and processed.

Abbreviations

BSL = British Sign Language

NGT = *Nederlandse Gebarentaal* (Dutch Sign Language, a.k.a. Sign Language of the Netherlands)

STS = *Svenskt teckenspråk* (Swedish Sign Language)

Data accessibility statement

The original corpus data can be obtained through permission from the respective repositories in which they are stored: BSL: *BSL Corpus Project* (https://bslcorpusproject.org/cava/); NGT and STS: *The Language Archive* (https://archive.mpi.nl/tla). The geospatial data for the map of Sweden comes from Statistics Sweden (SCB). Scripts used for the data pre-processing, analyses and visualizations for this article can be found at: https://osf.io/m5tzk/.

Acknowledgements

We thank three anonymous reviewers for comments and suggestions that significantly improved this article.

This study was partly supported by a European Research Council Advanced Grant (ERC-ADG 885220 *SignMorph*) awarded to Adam Schembri.

Competing interests

The authors have no competing interests to declare.

Author contributions

The study was first devised by CB and AS and developed further by CB, AS and OC. The data was processed and analyzed by CB with support from AS and OC. Data visualizations were done by CB. The article was initially drafted by CB and revised in collaboration with AS and OC.

References

Arel-Bundock, V., Greifer, N., & Heiss, A. (Forthcoming). How to interpret statistical models using marginaleffects in R and Python. *Journal of Statistical Software*.

Bickford, J. A. (1991). Lexical variation in Mexican Sign Language. *Sign Language Studies*, *72*(Fall 1991), 241–276. DOI: https://doi.org/10.1353/sls.1991.0010

Börstell, C. (2022a). Introducing the signglossR package. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, J. Mesch & M. Schulder (Eds.), *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources* (pp. 16–23). European Language Resources Association (ELRA). https://www.sign-lang.uni-hamburg.de/lrec/pub/22006.pdf

Börstell, C. (2022b). Searching and utilizing corpora. In J. Fenlon & J. A. Hochgesang (Eds.), *Signed language corpora* (pp. 90–127). Gallaudet University Press. DOI: https://doi.org/10.2307/j.ctv2rcnfhc.9

Börstell, C., Hörberg, T., & Östling, R. (2016). Distribution and duration of signs and parts of speech in Swedish Sign Language. *Sign Language & Linguistics*, *19*(2), 143–196. DOI: https://doi.org/10.1075/sll.19.2.01bor

Börstell, C., & Östling, R. (2016). Visualizing lects in a sign language corpus: Mining lexical variation data in lects of Swedish Sign Language. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen & J. Mesch (Eds.), *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining* (pp. 13–18). European Language Resources Association (ELRA). https://www.sign-lang.uni-hamburg.de/lrec/pub/16004.pdf

Bybee, J. L. (2007). *Frequency of use and the organization of language*. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*(1–2), 39–54. DOI: https://doi.org/10.1016/0167-6393(94)90039-6

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *Frontiers in Psychology*, *12*, 715843. DOI: https://doi.org/10.3389/fpsyg.2021.715843

Cohen Priva, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *The Journal of the Acoustical Society of America*, *141*(5), 2989–2996. DOI: https://doi.org/10.1121/1.4982199

Crasborn, O., & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In O. Crasborn, E. Efthimiou, T. Hanke, E. D. Thoutenhoofd & I. Zwitserlood (Eds.), *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (pp. 39–43). European Language Resources Association (ELRA). https://www.sign-lang.uni-hamburg.de/lrec/pub/08022.pdf

Crasborn, O., Zwitserlood, I., van der Kooij, E., Sáfár, A., Ros, J., Nauta, E., van Zuilen, M., van Dijken, L., van Winsum, F., & de Meijer, A. (2015). *Corpus NGT gloss annotations (3rd release)*. Centre for Language Studies, Radboud Universiteit, Nijmegen. DOI: https://doi.org/10.13140/RG.2.1.2303.7525

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*(2), 104–123. DOI: https://doi.org/10.1016/j.newideapsych.2007.02.002

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*(3), 253–260. DOI: https://doi.org/10.1016/S0095-4470(11)00055-6

Gahl, S. (2008). "Time" and "Thyme" are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496. DOI: https://doi.org/10.1353/lan.0.0035

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. DOI: https://doi.org/10.1016/j.tics.2019.02.003

Grosjean, F. (1979). A study of timing in a manual and a spoken language: American Sign Language and English. *Journal of Psycholinguistic Research*, *8*(4), 379–405. DOI: https://doi.org/10.1007/BF01067141

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568. DOI: https://doi.org/10.1016/j.wocn.2010.08.002

Horton, W. S., Spieler, D. H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology & Aging*, *25*(3), 708–713. DOI: https://doi.org/10.1037/a0019424

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(2), 233–256. DOI: https://doi.org/10.1017/S0954394509990093

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Benjamins. DOI: https://doi.org/10.1075/tsl.45.13jur

Keune, K., Ernestus, M., Hout, R. V., & Baayen, R. H. (2005). Variation in Dutch: From written MOGELIJK to spoken MOK. *Corpus Linguistics and Linguistic Theory*, *1*(2). DOI: https://doi.org/10.1515/cllt.2005.1.2.183

Kimmelman, V., Komarova, A., Luchkova, L., Vinogradova, V., & Alekseeva, O. (2022). Exploring networks of lexical variation in Russian Sign Language. *Frontiers in Psychology*, *12*, 740734. DOI: https://doi.org/10.3389/fpsyg.2021.740734

Kopf, M., Schulder, M., & Hanke, T. (2022, June). The Sign Language Dataset Compendium: Creating an overview of digital linguistic resources. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, J. Mesch & M. Schulder (Eds.), *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources* (pp. 102–109). European Language Resources Association. https://aclanthology.org/2022.signlang-1.16

Lepic, R. (2019). A usage-based alternative to lexicalization in sign language linguistics. *Glossa: A Journal of General Linguistics*, *4*(1), 23. DOI: https://doi.org/10.5334/gjgl.840

Lucas, C., Bayley, R., & Valli, C. (2009). *Sociolinguistic variation in American Sign Language*. Gallaudet University Press. DOI: https://doi.org/10.2307/j.ctv2rh2959

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. DOI: https://doi.org/10.1016/j.cognition.2012.09.010

Mesch, J., Rohdell, M., & Wallin, L. (2014). *Annotated files for the Swedish Sign Language Corpus. Version 2.* Sign Language Section, Department of Linguistics, Stockholm University. http://www.ling.su.se/teckensprakskorpus

Mesch, J., Wallin, L., Nilsson, A.-L., & Bergman, B. (2012). *Dataset. Swedish Sign Language Corpus project 2009–2011 (version 1)*. Sign Language Section, Department of Linguistics, Stockholm University. https://teckenspraakskorpus.su.se

Mudd, K., Lutzenberger, H., De Vos, C., Fikkert, P., Crasborn, O., & De Boer, B. (2020). The effect of sociolinguistic factors on variation in the Kata Kolok lexicon. *Asia-Pacific Language Variation*, *6*(1), 53–88. DOI: https://doi.org/10.1075/aplv.19009.mud

Müller, K. (2020). *here: A simpler way to find your files*. https://CRAN.R-project.org/package=here R package version 1.0.1.

Öqvist, Z., Riemer Kankkonen, N., & Mesch, J. (2020). STS-korpus: A sign language web corpus tool for teaching and public use. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen & J. Mesch (Eds.), *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* (pp. 177–180). European Language Resources Association (ELRA). https://www.sign-lang.uni-hamburg.de/lrec/pub/20014.pdf

Pebesma, E., & Bivand, R. (2023). *Spatial data science: With applications in R* (1st ed.). Chapman; Hall/CRC. DOI: https://doi.org/10.1201/9780429459016

Pedersen, T. L. (2024). *patchwork: The composer of plots* [R package version 1.2.0]. https://CRAN.R-project.org/package=patchwork

Petrini, S., Casas-i-Muñoz, A., Cluet-i-Martinell, J., Wang, M., Bentz, C., & Ferrer-i-Cancho, R. (2022). The optimality of word lengths. Theoretical foundations and an empirical study [Version Number: 5]. DOI: https://doi.org/10.48550/ARXIV.2208.10384

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526–3529. DOI: https://doi.org/10.1073/pnas.1012551108

Quer, J., & Steinbach, M. (2019). Handling sign language data: The impact of modality. *Frontiers in Psychology*, *10*, 483. DOI: https://doi.org/10.3389/fpsyg.2019.00483

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders*, *16*(3), 217–226. DOI: https://doi.org/10.1016/0021-9924(83)90035-7

Rowley, K., & Cormier, K. (2023). Accent or not? Language attitudes towards regional variation in British Sign Language. *Applied Linguistics Review*, *14*(4), 919–943. DOI: https://doi.org/10.1515/applirev-2020-0144

Safar, J. (2021). Whats your sign for TORTILLA? Documenting lexical variation in Yucatec Maya Sign Languages. *Language Documentation & Conservation*, *15*, 30–74. http://hdl.handle.net/10125/24970

Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2017). *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008–2017 (Third edition)*. University College London. https://bslcorpusproject.org

Schembri, A., & Lucas, C. (Eds.). (2015). *Sociolinguistics and Deaf communities*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781107280298

Schembri, A., Stamp, R., Fenlon, J., & Cormier, K. (2018). Variation and change in varieties of British Sign Language in England. In N. Braber & S. Jansen (Eds.), *Sociolinguistics in England* (pp. 165–188). Palgrave Macmillan. DOI: https://doi.org/10.1057/978-1-137-56288-3_7

Sigurd, B., Eeg-Olofsson, M., & van de Weijer, J. (2004). Word length, sentence length and frequency – Zipf revisited. *Studia Linguistica*, *58*(1), 37–52. DOI: https://doi.org/10.1111/j.0039-3193.2004.00109.x

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2024). *afex: Analysis of factorial experiments* [R package version 1.3-1]. https://CRAN.R-project.org/package=afex

Skoog Waller, S., Eriksson, M., & Sörqvist, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology*, *6*. DOI: https://doi.org/10.3389/fpsyg.2015.00978

Stamp, R., Dachkovsky, S., Hel-Or, H., Cohn, D., & Sandler, W. (2024). A kinematic study of phonetic reduction in a young sign language. *Journal of Phonetics*, *104*, 101311. DOI: https://doi.org/10.1016/j.wocn.2024.101311

Stamp, R., Schembri, A., Evans, B. G., & Cormier, K. (2016). Regional sign language varieties in contact: Investigating patterns of accommodation. *Journal of Deaf Studies and Deaf Education*, *21*(1), 70–82. DOI: https://doi.org/10.1093/deafed/env043

Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., & Cormier, K. (2014). Lexical variation and change in British Sign Language. *PLoS ONE*, *9*(4). DOI: https://doi.org/10.1371/journal.pone.0094053

Tyrone, M. E., & Mauk, C. E. (2010). Sign lowering and phonetic reduction in American Sign Language. *Journal of Phonetics*, *38*(2), 317–328. DOI: https://doi.org/10.1016/j.wocn.2010.02.003

Vermeerbergen, M., Twilhaar, J. N., & Herreweghe, M. V. (2013). Variation between and within Sign Language of the Netherlands and Flemish Sign Language. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 680–699, Vol. 3). De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110261332.680

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., Fran, C., Cois, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. DOI: https://doi.org/10.21105/joss.01686

Wickham, H., Hester, J., & Ooms, J. (2023). *xml2: Parse XML* [R package version 1.3.6]. https://CRAN.R-project.org/package=xml2

Wickham, H., Pedersen, T. L., & Seidel, D. (2023). *scales: Scale functions for visualization* [R package version 1.3.0]. https://CRAN.R-project.org/package=scales

Wilbur, R. B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech*, *52*(2–3), 245–285. DOI: https://doi.org/10.1177/0023830909103174

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk & D. Tapias (Eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556–1559). https://aclanthology.org/L06-1082/

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition*, *7*(6), 411–419. DOI: https://doi.org/10.3758/BF03198257

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press.