

Lauren Hall-Lew\*, Claire Cowie, Catherine Lai, Nina Markl, Stephen Joseph McNulty, Shan-Jan Sarah Liu, Clare Llewellyn, Beatrice Alex, Zuzana Elliott and Anita Klingler

# The Lothian Diary Project: sociolinguistic methods during the COVID-19 lockdown

<https://doi.org/10.1515/lingvan-2021-0053>

Received March 30, 2021; accepted September 2, 2021; published online February 22, 2022

**Abstract:** The Lothian Diary Project is an interdisciplinary effort to collect self-recorded audio or video diaries of people's experiences of COVID-19 in and around Edinburgh, Scotland. In this paper we describe how the project emerged from a desire to support community members. The diaries have been disseminated through public events, a website, an oral history project, and engagement with policymakers. The data collection method encouraged the participation of people with disabilities, racialized individuals, immigrants, and low-proficiency English/Scots speakers, all of whom are more likely to be negatively affected by COVID-19. This is of interest to sociolinguists, given that these groups have been under-represented in previous studies of linguistic variation in Edinburgh. We detail our programme of partnering with local charities to help ensure that digitally disadvantaged groups and their caregivers are represented. Accompanying survey and demographic data means that this self-recorded speech can be used to complement existing Edinburgh speech corpora. Additional sociolinguistic goals include a narrative analysis and a stylistic analysis, to characterize how different people engage creatively with the act of creating a COVID-19 diary, especially as compared to vlogs and other video diaries.

**Keywords:** COVID-era sociolinguistics; data collection; diaries; ethics; sociolinguistics

## 1 Introduction

Sociolinguistic studies based on self-recorded data are growing. While much of the data collection in our field has been in-person, face-to-face, and driven by our own interview questions and tasks, the idea behind self-recordings is to obtain speech data less constrained in these ways. Self-recordings are logistically, analytically, and ethically challenging, but they offer a unique representation of speech production with innumerable benefits. Previous work argues that self-recordings can result in a wider range of styles than results from other speech elicitation methods (Boyd et al. 2015; Hall-Lew and Boyd 2017, 2020). They also happen to be an ideal means of data collection during a global pandemic. The otherwise difficult conditions of the COVID-19 lockdown allowed for a project specifically designed around the remote collection of speech. In this paper, we discuss the methodological and ethical aspects of the Lothian Diary Project, an interdisciplinary effort to collect self-recorded audio or video diaries of people's experiences of the COVID-19 lockdown and social distancing measures in and around Edinburgh, Scotland.

Self-recordings are sociolinguistically interesting because they are constrained by the researcher's questions and physical presence in ways that differ from a traditional interview. The Observer's Paradox is not

---

\*Corresponding author: Lauren Hall-Lew, Linguistics and English Language, University of Edinburgh, Edinburgh, UK,

E-mail: Lauren.Hall-Lew@ed.ac.uk

Claire Cowie, Catherine Lai, Stephen Joseph McNulty and Zuzana Elliott, Linguistics and English Language, University of Edinburgh, Edinburgh, UK

Nina Markl, Institute for Language, Cognition and Computation, University of Edinburgh, Edinburgh, UK

Shan-Jan Sarah Liu, Politics and International Relations, University of Edinburgh, Edinburgh, UK

Clare Llewellyn, Neuropolitics Research Labs, University of Edinburgh, Edinburgh, UK

Beatrice Alex, Edinburgh Futures Institute, University of Edinburgh, Edinburgh, UK

Anita Klingler, History, University of Edinburgh, Edinburgh, UK

done away with, but altered in interesting ways (Barnes and Hall-Lew to appear; Schøning and Møller 2009). Speakers have more freedom to select which topics to cover and how to present themselves on the recording, then on, say, a survey. For this project, self-recordings were more practical to collect than online interviews, for various reasons. We, the research team, and we, the residents of the Edinburgh community, were coming to grips with living in lockdown, and an interview schedule might have inhibited data collection in a rapidly evolving and uncertain situation. Also, only one of the eleven members of our research team is Scottish, and he (McNulty) is not from Edinburgh. Interviews introduce the complexity of linguistic convergence to the researcher's speech in a way that self-recordings do not, being both automatic and agentive, as opposed to the purely agentive convergence to an imagined audience. Lastly, the distorting effects of various conference call technologies on the acoustic signal (De Decker and Nycz 2011; Freeman and De Decker 2021) are avoided here by speakers recording directly to their own devices, which results in recordings sufficiently comparable to professional field-recording equipment (see Sneller et al. this issue).

## 2 The context: Edinburgh and the Lothians

The Lothian Diary Project is based in Linguistics and English Language at the University of Edinburgh with collaborators from a range of different disciplines across the university. The project is part of Edinburgh Speaks, a larger research project on linguistic variation and change in and around Edinburgh. Edinburgh Speaks began in 2012, with a primary focus on curating and digitally processing existing corpora and supporting student projects, which have largely examined social class and binary gender. Work on ethnic variation has focused on immigrants from Central and Eastern Europe, the largest resident group with non-Anglo or non-Celtic ancestry. Sociolinguistic studies in Edinburgh have understudied race, gender identity, religion, neighbourhood, political affiliation, and other factors that we know are important to local social life. The first COVID-19 lockdown raised the possibility of collecting a new and more diverse speaker sample for building a more holistic sociolinguistic model of the community.

Due to the project's base in Edinburgh Speaks, data collection was limited to the city of Edinburgh and its three surrounding counties, known as the Lothians (East Lothian, West Lothian, and Midlothian). Participation was open to any resident of Edinburgh and the Lothians between May 2020 and July 2021. In contrast to previous sociolinguistic studies in the area, participation was open to people of all backgrounds, including immigrants from any location, all language backgrounds, and all ages. Collectively, the Edinburgh metropolitan area (with over 900,000 inhabitants) includes urban, suburban, and rural spaces, but the population is heavily concentrated in the urban centre (2020 population of 524,930).<sup>1</sup> Based on records from 2018, the proportion of city residents between 25 and 44 years old is much higher than in the rest of Scotland (34 vs. 26%). The city also has a much higher proportion of residents over the age of 16 with "Census Level 4 or above qualifications, such as a university degree" – 41% as compared to a national average of 26%.<sup>2</sup> Edinburgh has a national reputation for being middle class or wealthy, and indeed the least deprived area in all of Scotland is the city neighbourhood of Stockbridge.<sup>3</sup> However, "[a]n estimated 77,600 people were in poverty in Edinburgh in the year prior to the coronavirus outbreak, including almost one in five of all children" and "housing costs in Edinburgh are a key driver of high poverty rates in the city".<sup>4</sup> The high housing costs are connected to Edinburgh being Scotland's capital city and home to five campus universities. These institutions also mean

<sup>1</sup> Data from April 2020, National Records Scotland, <https://www.nrscotland.gov.uk/files/statistics/council-area-data-sheets/city-of-edinburgh-council-profile.html> (accessed 12 March 2021).

<sup>2</sup> Data from 2011, Scotland's Census, <https://www.scotlandscensus.gov.uk/education-0> (accessed 12 March 2021).

<sup>3</sup> Data from January 2020, the Scottish Index of Multiple Deprivation, <https://www.gov.scot/news/scottish-index-of-multiple-deprivation-2020/> (accessed 12 March 2021).

<sup>4</sup> Data from September 2020, the Edinburgh Poverty Commission, "Poverty in Edinburgh – data and evidence" [https://edinburghpovertycommission.org.uk/wp-content/uploads/2020/09/20200930\\_Poverty\\_in\\_Edinburgh-Data\\_and\\_evidence.pdf](https://edinburghpovertycommission.org.uk/wp-content/uploads/2020/09/20200930_Poverty_in_Edinburgh-Data_and_evidence.pdf).

that the population is relatively diverse: approximately 19% of the city's population is non-British, and 23% were born overseas, compared to 8 and 10% nationally.<sup>5</sup> The number of Edinburgh city residents born overseas is larger than the population of either East Lothian or Midlothian counties (108,000, vs. 105,790 and 91,340, respectively), and 47% of this population are of European origin, the highest proportion in any UK city.<sup>6</sup> Approximately 8% of the city's population is non-white, the plurality being of Chinese and South Asian descent. We aimed to obtain a speech sample reflecting some of this demographical diversity.

The Lothian Diary Project seeks to understand how the wider Edinburgh speech community has been affected by COVID-19. Scotland went into lockdown for the first time on 23 March 2020, and it stayed in place until rules began easing from 10 May (when the restriction on being outdoors only once a day was lifted), through June and July. New restrictions were gradually introduced from July (e.g., masks became mandatory on public transport), and then began to vary across Scotland according to the severity of infection. Medical professionals identified variant B.1.1.9, the “UK variant”, in December 2020, and Scotland re-entered lockdown on 5 January 2021, as the Scottish Government announced the roll-out of vaccines. Lockdown rules began to ease again on 22 February (when young children returned to childcare and classrooms), and continued to ease up into the summer. The final data set has 195 contributions, roughly half ( $N = 91$ ) from May 2020 and July 2020, roughly another half ( $N = 88$ ) from May 2021 to July 2021, and a smattering ( $N = 16$ ) from the months in between. The data are a mix of then-current descriptions of lockdown experiences and retrospective descriptions of earlier lockdown experiences.

### 3 Community-centred project design

Conducting research during a time when social inequalities are amplified<sup>7</sup> motivated our team of privileged academics to design a community-focused methodology. Our attempt was to design a linguistic research project that could support community members both culturally and financially. The project paid £15 for each standard contribution and £20 for each contribution from someone unhoused or otherwise vulnerable. There were three options for payment, none of which involve payment apps e.g., (PayPal, which is not as commonly used in Scotland as in the US) nor pre-registration on a work site (Amazon Turk or Prolific Academic), nor the support of a major corporation (such as Amazon). The first option was if the participant has a bank account, in which case the payment could be made by direct debit; Betsy Sneller (pers. comm.) notes that this option is “bureaucratically unwieldy” and, indeed, this option was only possible with the support of our finance office. The second option was payment in the form of a gift card to a local business, which supported community businesses and let participants without bank accounts be easily paid virtually. A participant could list three local businesses of their choice and a member of the research team then emailed them a £15–20 gift card from one of the nominated businesses that offered e-vouchers. The third option was a donation to a local charity (community non-profit organization) of their choice.

From the first recruitment phase it was clear that we were lacking the voices of those who had been hardest hit by the pandemic. In August 2020 we launched a second phase of recruitment specifically aimed at partnering with local charities to recruit under-represented participants: people struggling with homelessness, people with disabilities, and caregivers of these and other groups. Accessing participants from more vulnerable or marginalized backgrounds using remote, self-recording methods posed a unique logistical and ethical challenge. Therefore, partnerships were formed with local charitable organizations who provided support or advocacy for these groups. Each partner charity was given a charity support code, and for each participant who

<sup>5</sup> Data from January 2020, National Records Scotland, “Population by Country of Birth and Nationality, Scotland, 2020”, <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/population-by-country-of-birth-and-nationality/jan-dec-20> (accessed 13 January 2022).

<sup>6</sup> Data from <https://worldpopulationreview.com/world-cities/edinburgh-population> (accessed 12 March 2021).

<sup>7</sup> Data from September 2020, “The Impacts of Covid-19 on Equality in Scotland”, <https://www.gov.scot/publications/the-impacts-of-covid-19-on-equality-in-scotland/> (accessed 13 January 2022).

used this code when participating in the Lothian Diary Project, £5 would be donated to the charity, in addition to the regular £15–20 payment to the participant.

Two charities proactively reached out to us prior to this effort. Forging partnerships with additional charities was achieved through professional networks and through an online interactive workshop held as part of the 2020 Festival of Social Science organized by the Economic and Social Research Council (ESRC) on 7–15 November. Approximately ten community organizations were represented at this event. A targeted social media campaign, making use of paid advertising, was also employed. As a result of these strategies, four additional charity partnerships were formed. Five additional charity partnerships were formed in the subsequent months through word-of-mouth. While such partnerships have provided valuable access to the target groups, they have, of course, been difficult to implement during the pandemic, when many charities are unable to gain easy access to their communities, and have an increased workload.

The project was designed to offer cultural contributions in addition to financial support. A subset of those diaries marked for public access were subtitled and made available for viewing on our website. We held public events and created an interactive map of places important to our contributors. We contributed a subset of these to be housed as an oral history archive (Becker 2010; Kendall and Wolfram 2016), with the Museums and Galleries division of Edinburgh City Council. In November 2021, we delivered a report and brief to Members of the Scottish Parliament.

## 4 Participant recruitment and data collection

Participant recruitment was achieved through local radio, local newspapers, press releases, leafleting campaigns, emails to schools, and through our social networks, as well as through charitable organisations. In November 2020 and May–June 2021, as part of the 2020 ESRC Festival of Social Sciences, an outdoor community space was rented to engage more directly with the public and allow more digitally excluded members to take part in the project on-site. Project team members handed out flyers and recorded four participants in an old police box near central Edinburgh that has been repurposed for community events. While this greatly aided the visibility of the project, the quality of the recordings taken outside unfortunately proved to be largely unsuitable for phonetic analysis (though transcripts should still be usable for other linguistic analysis). Understandably, many potential participants took a flyer but did not stay for the recording: most were too busy to stop, but some indicated that they felt too inhibited to record their lockdown experiences on-site. This reticence to speak in a public space shows one of the benefits of a self-recording made in a private space. This is interesting to consider in light of projects such as the QuakeBox corpus (Clark et al. 2016), which collected stories from residents of Christchurch, New Zealand, in the wake of the 2010–2011 earthquakes. The QuakeBox was a “a mobile recording studio built into a shipping container ... [and] ... positioned at various locations in and around the city of Christchurch” (15). Participants entered the QuakeBox and created a self-recording, with the researchers outside and out of earshot. It was a highly successful data collection effort, and raises the practical question of what kind of environment is needed for participants to feel comfortable sharing potentially painful experiences for research purposes.

Our aim was to collect as many unique contributions as possible. We also wanted to give each contributor the opportunity to frame their contribution in their own way, while at the same time obtaining enough information to make comparisons across the sample. Each participant submitted one recorded session that was paired with one set of responses to a survey. While other research on self-recordings has employed multiple recordings from each speaker (e.g., Sneller et al. this issue), each speaker in the Lothian Diary Project only participated once. In lieu of longitudinal comparisons, these snapshots in time are amenable to linguistic analyses of style and narrative. For example, on-going work is examining those styles and personae that characterize how different people engage with the act of creating a COVID-19 diary, especially as compared to vlogs and other video diaries (Frobenius 2014; Lee 2017; Starr et al. 2020; Voigt et al. 2016), before but also during the COVID-19 pandemic (Tateo and Dario 2020; Zhang and Zhao 2020).

The “participation” page of the Lothian Diary Project website<sup>8</sup> listed the five-step process to contributing. The first step was to create an audio or video diary. The website provided topic prompts (e.g., “How has your life changed during lockdown?”), and people varied in how closely they referred to this guidance. Some read the seven prompts aloud and then responded to them, like an interview, while others showed little indication of having seen the prompts at all. In cases where a child or vulnerable adult was the main speaker being interviewed by someone else, that interviewer typically asked those seven questions. A FAQ page listed, among other things, all information related to research ethics (e.g., how the data will be used). We also provided a page on “tips on making your recording” including hardware, software, file formats, and how to reduce ambient noise. The four remaining steps for contributing all took place on a single survey designed on Qualtrics. These were giving consent to use the data, uploading the diary file, completing the survey, and choosing the payment option.

The norms for research ethics are continually evolving and vary by country and region. Research in Scotland follows the General Data Protection Regulation 2016/679 for the European Union and European Economic Area, and the Lothian Diary Project took an oral history approach to this.<sup>9</sup> Even though most of the recordings are not interviews, the project website and Qualtrics form serve as a kind of proxy interviewer. Participants each gave consent as to how their contribution can be used, according to three levels: (1) use by other researchers, (2) use as part of an oral history archive, and (3) use for the public: radio, TV, and online. Participants were provided with the option to waive anonymity, which was provided in acknowledgement of their authorship in the content they produced and shared. Caregivers were discouraged from choosing non-anonymity for children and vulnerable participants, but the option was there and was taken. Every contribution from children and vulnerable adults was manually checked by 1–3 research team members for any sensitive information (see Supplementary Materials). If any contributions with sensitive information had been marked for de-anonymizing, the research team would have recontacted the caregiver to reaffirm consent. However, this was not the case in this data set.

After giving their consent, participants uploaded their pre-recorded video or audio diary to the survey, using a Box.com interface. Speed of upload depended entirely on their personal device and connection speed, and for some participants this was the most frustrating aspect of participation (see Sneller et al. this issue, who avoided the issue by hiring an app developer). The file was manually transferred by a member of the research team from Box.com to a protected server space for processing.<sup>10</sup> All diaries were backed up in original formats and as WAV files (mono, 16 kHz and mono/stereo 44 kHz). Transcripts of English language recordings were generated using custom automatic speech recognition models implemented in Kaldi (Povey et al. 2011), and these were all hand-corrected in ELAN (2020) by a member of the research team.<sup>11</sup> All non-English speech has been hand-transcribed and translated. All mono 16 kHz WAV files, MP4 files, and accompanying transcripts and surveys are held in the Lothian Diary Project repository on DataVault (Hall-Lew et al. 2021).<sup>12</sup> For those files marked as publicly shareable, video files and transcripts (but not survey data) are available on DataShare.<sup>13</sup>

After the upload, each participant completed a survey collecting demographic information and information about the COVID-19 lockdowns. The questions for the survey were adapted from questions created by the University of Edinburgh’s COVIDLife project.<sup>14</sup> The survey questions, many of which had already been piloted by the COVIDLife research team, went through two rounds of piloting with this project in 2020, including input from archivists at Museums and Galleries, Edinburgh, prior to public release (see Supplementary Materials). The questions were then piloted again in 2021 prior to the final stage of data collection.

<sup>8</sup> <https://lothianlockdown.org/participate/>.

<sup>9</sup> “Data protection for oral historians and organisations holding oral history interviews”, <https://www.ohs.org.uk/advice/data-protection/> (accessed 13 January 2022).

<sup>10</sup> <https://www.digitalresearchservices.ed.ac.uk/resources/datastore>.

<sup>11</sup> Thanks also to Ellie MacCallum, who assisted in 2020.

<sup>12</sup> Contact research team for access enquires. <https://doi.org/10.7488/7a22cc4b-87ec-4df3-a549-3b347fd4bca5>.

<sup>13</sup> <https://datashare.ed.ac.uk/handle/10283/3866>.

<sup>14</sup> <https://www.ed.ac.uk/generation-scotland/covidlife-volunteers/what-is-covidlife>.



Data collection was completed on 15 July 2021. The final phase recruited participants from under-represented demographic groups through new charity partnerships and new academic collaborations, particularly with those working with young people, the elderly, and their caregivers.

## 5 The diaries and the diary makers

The project received submissions from 195 participants, of varying lengths (range = 01:00–52:00 min, mean = 04:58, SD = 05:10). The sample covers a good representation of the different parts of the city of Edinburgh, but only a few of the participants were from the Lothian counties. Submissions are mainly in English or Scots, with some in Scottish Gaelic, Mandarin Chinese, Cantonese, and American Sign Language.

The demographics of the speaker sample are unbalanced, but there are many speakers who are not typically represented in sociolinguistic studies of Edinburgh. The proportion of Lothian Diary Project participants who have a disability (12%) is slightly less than the city population,<sup>15</sup> but higher than other sociolinguistic surveys where ability is not a reported factor at all. The sample has a greater-than-proportionate representation of people of colour (16 vs. 8% of the Edinburgh population). Immigrants are well represented (24 vs. 16% of the population), as are those who are SOGI<sup>16</sup>-diverse (19%, general statistics unknown). These characteristics create new opportunities to analyse and document the experiences of marginalized communities who are under-represented in other Edinburgh Speaks corpora. At the same time, the Lothian Diary Project speaker sample is also unbalanced due to over-representing those groups that already distinguish Edinburgh from the rest of Scotland. For example, there are more participants with post-secondary education (54%) than in the population as a whole (41%), and more speakers aged between 25 and 44 (42%) than in the city population (34%). With respect to COVID-19 experience, the current sample is also more privileged: most participants had stable wages and were neither care workers nor essential workers, and most had the same job and pay they had prior to COVID-19. With respect to gender and sex, we initially had far more contributions from women and girls than from men and boys (as expected; see, e.g., Gannon et al. 1971), but a targeted recruitment strategy in phase 3 resulted in a somewhat more proportionate balance (52% female, 40% male, 2% prefer not to say, 1% other).

The notable affluence of the Lothian Diary Project speaker sample results, in part, from social network differences between Edinburgh residents, and the socially stratified social practice of engaging in academic research. It also results, in part, from issues of digital access inherent to our methodology; submissions were made on a smartphone, tablet, or computer, and good internet access was required. It also presumably correlates with a better lockdown experience, so the sample currently has fewer recordings of the most negatively affected people. We had initially thought that video diaries would increase ease of access, for example for those with limited literacy; however, the inclusion of a written survey undermined such benefits.

Remote participant recruitment was also a problem. The research team was not on-site to answer questions, assist with technological problems, or prompt the speaker to continue talking. Potential participants may not have read the FAQs, and so any information included there that might have encouraged participation was missed. For example, although audio-only diaries were acceptable, this fact was on the FAQ page and not highlighted, and anecdotally we know that some people were put off by the video aspect. Lastly, COVID-19 itself was a problem: communities across the world have been coping with illness and death, and participation in non-medical research is rightly not people's top priority. Those suffering might have the stories that we most want to share with policymakers, but they do not have the time or energy to participate, and talking and thinking about their suffering in itself might be a painful experience.

On a more positive note, our methodology seems to have encouraged the participation of people with disabilities, racialized individuals, immigrants, and low-proficiency English/Scots speakers, all of whom have

<sup>15</sup> The percentage of “adults with long term limiting physical or mental health condition” in 2019 was 17.2% (<https://statistics.gov.scot/>, accessed 13 January 2022).

<sup>16</sup> Sexual orientation and gender identity.

been under-represented in previous studies of linguistic variation in Edinburgh, and are also more likely to be negatively affected by COVID-19. And thanks to our charity partners, we also do have contributions from the unemployed, unhoused, disabled, and caregivers, even if they remain a small part of the sample. Again, none of these groups have been represented in previous studies of linguistic variation in Edinburgh, despite their representing such a large portion of the population.

## 6 Lockdown diaries as linguistic corpora

The diary genre lends itself especially well to different kinds of deep qualitative work, in particular narrative analysis and style analysis. With respect to variationist sociolinguistic research, the data set's exceptional demographic and stylistic diversity present some challenges as well as new opportunities.

Most of the diary contributions contain “narratives of personal experience” (Labov and Waletzky 1967), many of them related to matters of life and death (Labov 2013). One approach which the team intends to take will be to focus on how COVID-19 affected subjective experience, drawing comparisons to parallel analyses in other regions (Jaspal 2021; Zhang et al. 2021). This will, for example, consider both the structures and topics that occur across COVID-19 narratives, examining differences between individuals and between social groups. Which aspects of narrative structure and content seem to be most variable or most consistent? How do these different linguistic practices construct speakers' identities, and do they suggest shifts in those identities due to COVID-19? This analysis will be informed by the wealth of work on diaries in a medical context (e.g., Ram-anathan 2009) as well as work on narrative medicine (Charon 2008, 2017). It will also draw on oral history projects from similar COVID-19 recordings which look to “the value of narrative as a mode for understanding this historic moment” (Kelly 2020: 249). Such projects aim to provide a fuller picture of the global pandemic, fleshing out insights from broader quantitative reports: “Although it can feel minute against the scale of COVID-19, capturing the oral history of the individual, family, or community experience of the coronavirus distills the pandemic down to something that has a surface and a quality. It does so in a manner that is profound and makes the intangible more tangible” (Sloan 2020: 196). A narrative analysis has both applied and theoretical possibilities, addressing questions of subjective experience at the same time as centring the value of subjectivity in sociolinguistic theory.

The Lothian Diary Project team also plan to analyse the linguistic and multimodal styles characterizing the video diaries. Different styles arise in this collection because there is no single shared style for recording oneself on video. The non-vulnerable adult participants had tremendous agency in terms of topic, framing, stance, etc.<sup>17</sup> It has been previously suggested that self-recordings might be sociolinguistically valuable because they are more likely to represent a more complete picture of a speaker's linguistic repertoire than other researcher-led elicitation techniques. However, rather than collecting a stylistic range, here each speaker is represented by a single style, and that “video diary” style is highly variable between individuals. We aim to describe and analyse the range of choices people make about how to frame their contribution and the resources they draw on to construct those styles, including linguistic but also gestural and postural variables. Impressionistic analysis suggests that many participants in the Lothian Diary Project orient to the task of diary making with an elevated sense of seriousness and formality. Perhaps this is brought on by the topic of a deadly pandemic, the framing of the diary as oral history, or the presumed audience being university researchers. These “serious” styles and stances contrast with some of the other styles found in the data set, including those that evoke the style of social media vloggers (e.g., Lee 2017) on, for example, YouTube or TikTok. One approach would be to contrast these video diary styles to those found in similar genres (e.g., Zhang and Zhao 2020). This would entail drawing on various theories of style (Coupland 2007; Eckert and Rickford 2001; Jaffe 2009), with particular attention to audience design (Clark and Murphy 1982), given the absence of a physical audience in most of the COVID-19 diaries.

<sup>17</sup> Some children and vulnerable adults did not record monologues but were instead interviewed by a caregiver or charity worker, resulting in more stylistically constrained speech.

The high level of demographic and stylistic variability, and a relatively small data set, does present a challenge for variationist analysis. Qualitative analysis of linguistic variation is more likely than quantitative. An analysis of place-indexed variation is an example, since the data revolve around a global phenomenon potentially redefining local social experiences, identities, and structures. We also anticipate seeing new indexes emerge that are specific to COVID-19. If so, these will be necessarily intertwined with the different stages of the pandemic. “Time” may be an important factor but not necessarily in the way we are used to seeing it in sociolinguistics. An example of quantitative work is a small-scale comparison to historical corpora of the Edinburgh speech community, focusing on a matched subsample of diarists. We are currently analysing a sample of diaries from white Scottish men, and comparing them to similar speakers Esling (1978) who have already been analysed for variation in voice quality, /t/-glottalling (Göbel 2020; Titheridge 2020), the *which~witch* merger (Kelly 2021), and (ing) variation and KIT vowel variation (Göbel 2021). Social descriptions of Edinburgh are rich in historical studies, which should make it possible to study the co-evolution of linguistic and social change. Studies of newer speaker groups (e.g., South Asians; Chinese) can similarly investigate the production of these and other variables from the previous literature, examining the extent to which a more diverse sample alters our representation of a speech community. The challenge to all of this is the variability between recordings that is not due to social differences, but differences such as the speaker’s location, distance from their microphone, amount of ambient noise, and length of the recording.

## 7 Conclusion

Sociolinguistic fieldwork was one of the countless things impacted by the COVID-19 pandemic, the constraints of lockdown and social distancing raising new challenges and opportunities for sociolinguistic methodology. With the Lothian Diary Project we attempted to address these challenges and explore these opportunities with a community-focused approach to the remote collection of self-recordings. The most significant challenge to data collection was reaching potential participants with limited digital access or digital literacy. Nonetheless, the project collected nearly 200 unique COVID-19 oral history narratives, many of which give voice to previously under-represented speakers from the wider Edinburgh community.

**Acknowledgment:** This work would not have been possible without the generous contributions from the diary contributors. Thanks also goes to the community partners who have helped to make this research possible: Carr Gomm, Dr Bell’s Family Centre, the Edinburgh Society of Model Engineers, LinkLiving, People First (Scotland), Redwoods Caring Foundation, Streetwork, Super Power Agency, the Thistle Foundation, West Lothian Financial Inclusion Network, and the Welcoming Association. Special thanks to Monty Roy of the Leith Walk Police Box pop-up space. The following were sources of financial support: the ESRC Festival of Social Sciences; the Being Human Festival; the University of Edinburgh’s College of Arts, Humanities, and Social Sciences Research Office; an ESRC Impact Acceleration Grant awarded to the University of Edinburgh; the School of Philosophy, Psychology, and Language Sciences’ Knowledge Exchange and Impact Office; and the Centre for Doctoral Training in Natural Language Processing, funded by the UKRI and the School of Informatics and School of Philosophy, Psychology & Language Sciences of the University of Edinburgh.

## References

- Barnes, Sonia & Lauren Hall-Lew. to appear. Situations of use: Self-recordings & oral histories. In Malcah Yaeger-Dror, Christopher Cieri & Katie Drager (eds.), *Dimensions of linguistic variation*. Oxford: Oxford University Press.
- Becker, Kara. 2010. *Regional dialect features on the lower east side of New York City: Sociophonetics, ethnicity, and identity*. New York: New York University Dissertation.



- Boyd, Zac, Zuzana Elliott, Josef Fruehwald, Lauren Hall-Lew & Daniel Lawrence. 2015. An evaluation of sociolinguistic elicitation methods. In *18th international conference on the phonetic sciences*. Glasgow: The University of Glasgow. Available at: [https://www.research.ed.ac.uk/portal/files/21806989/Boyd\\_etal\\_2015\\_ICPhS\\_final.pdf](https://www.research.ed.ac.uk/portal/files/21806989/Boyd_etal_2015_ICPhS_final.pdf).
- Charon, Rita. 2008. *Narrative medicine: Honoring the stories of illness*. Oxford: Oxford University Press.
- Charon, Rita. 2017. *The principles and practice of narrative medicine*. Oxford: Oxford University Press.
- Clark, Herbert H. & Gregory L. Murphy. 1982. Audience design in meaning and reference. *Advances in Psychology* 9. 287–299.
- Clark, Lynn, Helen MacGougan, Jennifer Hay & Liam Walsh. 2016. “Kia ora. This is my earthquake story”: Multiple applications of a sociolinguistic corpus. *Ampersand* 3. 13–20.
- Coupland, Nikolas. 2007. *Style: Language variation and identity*. Cambridge: Cambridge University Press.
- De Decker, Paul & Jennifer Nycz. 2011. For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics* 17(2). 51–59.
- Eckert, Penelope & John R. Rickford (eds.). 2001. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.
- ELAN. 2020. *ELAN (version 6.0) [Computer program]*. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Available at: <https://archive.mpi.nl/tla/elan>.
- Esling, John. 1978. The identification of features of voice quality in social groups. *Journal of the International Phonetic Association* 8(1). 18–23.
- Freeman, Valerie & Paul De Decker. 2021. Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *Journal of the Acoustical Society of America* 149(2). 1211–1223.
- Frobenius, Maximiliane. 2014. Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics* 72. 59–72.
- Gannon, Martin J., Joseph C. Nothorn & Stephen J. Carroll. 1971. Characteristics of nonrespondents among workers. *Journal of Applied Psychology* 55(6). 586–588.
- Göbel, Jessica. 2020. Don’t glottal stop me now. *Lifespans and Styles* 6(2). 32–43.
- Göbel, Jessica. 2021. *A sociophonetic study of (ing) variation and KIT vowel realization in Edinburgh English in the 1970s*. Edinburgh: The University of Edinburgh Honours Dissertation.
- Hall-Lew, Lauren, Claire Cowie, Stephen Joseph McNulty, Nina Markl, Shan-Jan Sarah Liu, Catherine Lai, Clare Llewellyn, Alex Beatrice, Nini Fang, Zuzana Elliott & Anita Klingler. 2021. The Lothian Diary Project: Investigating the impact of the COVID-19 pandemic on Edinburgh and Lothian residents. *Journal of Open Humanities Data* 7. 4.
- Hall-Lew, Lauren & Zac Boyd. 2017. Phonetic variation and self-recorded data. *University of Pennsylvania Working Papers in Linguistics* 23(2). 86–95.
- Hall-Lew, Lauren & Zac Boyd. 2020. Sociophonetic perspectives on stylistic diversity in speech research. *Linguistics Vanguard* 6. 20180063.
- Jaffe, Alexandra (ed.). 2009. *Stance: Sociolinguistic perspectives*. Oxford: Oxford University Press.
- Jaspal, Rusi. 2021. Identity threat and coping among British South Asian gay men during the COVID-19 lockdown. *Sexuality & Culture* 25(4). 1428–1446.
- Kelly, Fiona. 2021. *Variation and change in the which-witch merger in Edinburgh English*. Edinburgh: The University of Edinburgh Honours Dissertation.
- Kelly, Jason M. 2020. The COVID-19 oral history project: Some preliminary notes from the field. *The Oral History Review* 47(2). 240–252.
- Kendall, Tyler & Walt Wolfram. 2016. Engagement through data management and preservation: The North Carolina language and life project and the sociolinguistic archive and analysis project. In Karen P. Corrigan & Mearns Adam (eds.), *Creating and digitizing language corpora: Volume 3: Databases for public engagement*, 133–157. London: Palgrave Macmillan UK.
- Labov, William. 2013. *The language of life and death: The transformation of experience in oral narrative*. Cambridge, UK: Cambridge University Press.
- Labov, William & Joshua Waletzky. 1967. Narrative analysis. In June Helm (ed.), *Essays on the verbal and visual arts*, 12–44. Seattle: University of Washington Press.
- Lee, Sarah. 2017. Style-shifting in vlogging: An acoustic analysis of “YouTube voice”. *Lifespans and Styles* 3(1). 28–39.
- Povey, Daniel, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Silovsky Jan, Georg Stemmer & Karel Vesely. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*. Hong Kong, China: IEEE Signal Processing Society. Available at: <https://infoscience.epfl.ch/record/192584?ln=en>.
- Ramanathan, Vaidehi. 2009. Scripting selves, stalling last shadows: (Auto)biographical writing of Alzheimer patients and their caregivers. *Critical Inquiry in Language Studies* 6(4). 292–314.
- Schønning, Signe & Janus Spindler Møller. 2009. Self-recordings as a social activity. *Nordic Journal of Linguistics* 32(2). 245–269.
- Sloan, Stephen M. 2020. Behind the “curve”: COVID-19, infodemic, and oral history. *The Oral History Review* 47(2). 193–202.
- Starr, Rebecca Lurie, Tianxiao Wang & Christian Go. 2020. Sexuality versus sensuality: The multimodal construction of affective stance in Chinese ASMR performances. *Journal of Sociolinguistics* 24(4). 492–513.
- Tateo, Luca & Nadia Dario. 2020. Listen to us, grown-ups! A research on children’s diaries during the COVID-19 pandemic. *Formazione & Insegnamento* 18(3). 89–100.

- Titheridge, Zoé. 2020. Glottal replacement in retrospect. *Lifespans and Styles* 6(1). 28–37.
- Voigt, Rob, Penelope Eckert, Dan Jurafsky & Robert J. Podesva. 2016. Cans and cants: Computational potentials for multimodality with a case study in head position. *Journal of Sociolinguistics* 20(5). 677–711.
- Zhang, Carol X., Liang Wang & Jillian M. Rickly. 2021. Non-interaction and identity change in COVID-19 tourism. *Annals of Tourism Research* 89. 103211.
- Zhang, Leticia Tian & Sumin Zhao. 2020. Diaspora micro-influencers and COVID-19 communication on social media: The case of Chinese-speaking YouTube vloggers. *Multilingua* 39(5). 553–563.

---

**Supplementary Material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/lingvan-2021-0053>).