

Adrian Leemann\*, Péter Jeszenszky, Carina Steiner, Melanie Studerus and Jan Messerli

# Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing

<https://doi.org/10.1515/lingvan-2020-0061>

**Abstract:** Linguistic data collection typically involves conducting interviews with participants in close proximity. The safety precautions related to the COVID-19 pandemic brought such data collection to an abrupt halt: Social distancing forced linguistic fieldwork into involuntary hibernation in many parts of the world. Such hardship, however, can inspire innovation. In this contribution, we present an approach that – we believe – enables a reliable switch from in-person, face-to-face interviews to virtual, online data collection. In this approach, participants remain at home and navigate a smartphone application, enabling high-quality audio recordings and multisensory presentation of linguistic material, while they are being supervised via videoconferencing (Zoom 2020 <https://zoom.us/> (accessed 11 August 2020)). The smartphone app and the infrastructure presented are open source, accessible, and adaptable to researchers' specific needs. To explore whether participants' experiences of in-person data collection are different from participation in a virtual setting, we conducted a study with 36 participants. Overall, findings revealed a substantial degree of overlap in interview experience, setting a methodological baseline for future work. We end this contribution by discussing the benefits and pitfalls of this new approach.

**Keywords:** linguistic fieldwork, pandemic, COVID-19, smartphone app, Swiss German

## 1 Introduction

In traditional dialectology, linguistic fieldwork often involves linguists going to remote areas to interview carefully selected speakers. Fieldwork interviews are typically interactive and done in close physical proximity (cf. Sakel and Everett 2012). Particularly in the era when many multi-volume linguistic atlases were created (e.g. Survey of English Dialects [SED], Linguistic Atlas of Swiss German [SDS]), it was common for researchers to stay for several days at a time in various locations in order to collect thousands of tokens of regional variants from NORMs (**n**on-mobile **o**lder **r**ural **m**ales) or NORFs (**n**on-mobile **o**lder **r**ural **f**emales). The SDATS project (Swiss German Dialects Across Time and Space,<sup>1</sup> Leemann et al. 2020a) was initially designed in this tradition of data collection in physical proximity. SDATS investigates variation and change in German-speaking Switzerland using data from 1000 speakers across 125 survey sites – with the data being collected by five

---

1 [www.sdats.ch](http://www.sdats.ch)

---

**\*Corresponding author: Adrian Leemann**, Center for the Study of Language and Society, University of Bern, Muesmattstrasse 45, 3012 Bern, Switzerland, E-mail: [adrian.leemann@csls.unibe.ch](mailto:adrian.leemann@csls.unibe.ch)

**Péter Jeszenszky, Carina Steiner, Melanie Studerus and Jan Messerli**, Center for the Study of Language and Society, University of Bern, Muesmattstrasse 45, 3012 Bern, Switzerland, E-mail: [peter.jeszenszky@csls.unibe.ch](mailto:peter.jeszenszky@csls.unibe.ch) (P. Jeszenszky), [carina.steiner@csls.unibe.ch](mailto:carina.steiner@csls.unibe.ch) (C. Steiner), [melanie.studerus@csls.unibe.ch](mailto:melanie.studerus@csls.unibe.ch) (M. Studerus), [jan.messerli@csls.unibe.ch](mailto:jan.messerli@csls.unibe.ch) (J. Messerli)

linguists. *SDATS*, as well as probably the vast majority of linguistics projects involving fieldwork, came to an abrupt halt with the outbreak of COVID-19 in early 2020.

The first case of COVID-19 in Switzerland was reported on 25 February 2020. The government responded with the following key measures: On 28 February, events with more than 1000 participants were banned. On 1 March, case-based self-isolation was mandated; on 16 March, social distancing was encouraged, and lockdown was ordered (Flaxman et al. 2020). The University of Bern, where *SDATS* is hosted, took the following measures: On 27 March, it switched to ‘emergency operations’, which meant that research operations were completely shut down with approved exceptions only. *SDATS* had its last (52nd) in-person interview on 13 March. In mid-April 2020, measures were eased by the University and *SDATS* was granted permission to resume data collection in early May. However, places to record participants were prohibitively difficult to find, as schools, universities, and other public buildings denied access; in addition, fieldworkers were not able to travel long distances by train. In the worst-case scenario, if one of the group members had been infected with COVID-19, they could potentially pass the virus on to participants across the country. Adding to the grim outlook, the World Health Organization predicted potential further peaks of infections (Reuters 2020).

After a pause in data collection, we – and very likely many other linguists involved in fieldwork – started thinking of solutions that could enable a continuation of linguistic fieldwork even if another wave or peak of COVID-19 were to emerge. We were looking for a solution to decrease physical proximity to a minimum while attempting to maintain the same interview experience and recording quality. We decided to switch from in-person to online interviews. In this contribution, we present a novel approach to collecting linguistic data through a smartphone application specifically developed for the project’s purposes (*SDATS*, Leemann et al. 2020b) while participants are simultaneously supervised via video calls (Zoom 2020) (henceforth referred to as ‘online interviews’). Our approach, we believe, should hold up against a wide range of predicted and unanticipated developments over the course of the current pandemic as well as in future emergency situations. The app is open source (GNU GPL-3.0) and can be adapted by other linguists for their specific purposes. One of the core benefits of today’s smartphones is their multimodality, featuring screens, speakers, high quality microphones, cameras, and a variety of sensors – all of which can be used for data collection in linguistic research. Most importantly, they are ubiquitous, especially among younger people; more than two thirds of Swiss people, for instance, have smartphones (71.7%, cf. Newzoo 2018).

This paper contains five sections. In Section 2, we present the process behind developing the app. In Section 3, we present an experiment checking the validity of this quasi-personal, online approach: we compare the experience of participants who took part in the different elicitation modes (in-person and online). Finally, we discuss the findings and address benefits and pitfalls of this new paradigm as well as ethical considerations.

## 2 From in-person to online interviews

The aim of developing the smartphone app was to replicate the in-person data collection experience as closely as possible, but with physical proximity and travel on public transport out of the equation. The tasks that comprise the *SDATS* study and which needed to be replicated in the app are shown in Table 1.

**Table 1:** *SDATS* tasks.

Task	Description
Phonetic, lexical, and morpho-syntactic items	Elicitation of 315 phonetic, morphosyntactic, and lexical items via picture and text prompts
Read speech (Standard)	Elicitation of read speech in standard German (56-word text)
Read speech (dialect)	Elicitation of read speech in dialect (266-word text)
Spontaneous speech	Elicitation of spontaneous speech via semi-structured sociolinguistic interviews (10 min)
Imitation task	Testing of speech imitation capacity (cf. e.g. Christiner and Reiterer 2015). Participants are asked to repeat phrases in a language unknown to them; stimuli presented via speakers
Speech perception task	Perception of Yanny/Laurel (cf. Pressnitzer et al. 2018); stimuli presented via speakers
Dialect mapping task	Participants are asked to draw dialect regions (cf. Preston 2010)

Before COVID-19, tasks were presented and recorded on the SpeechRecorder software (cf. Draxler and Jänsch 2004) using a Behringer U-Phoria audio interface, a Beyerdynamic TGL58TG lavalier condenser microphone, and a Beyerdynamic MPA-PVA phantom power adapter. Recording quality was set to 44.1 kHz with 16-bit quantization rate. Figure 1 shows the typical setup *SDATS* used before the pandemic. Each recording session lasted around 2 h.

The *SDATS* study presented via SpeechRecorder was coded as a string of .HTML files connected by .XML files. This became the basis for the straightforward development of the open source (GNU GPL-3.0), cross-platform *SDATS* smartphone app on the Xamarin (2020) platform, which utilizes .NET and C# and runs on iOS and Android devices (Leemann et al. 2020b). Figure 2 shows two sample prompts as displayed in SpeechRecorder during in-person data collection (left), alongside the same prompts as rendered in the smartphone app (right); the top panel shows a picture prompt eliciting a lexical variable, while the bottom prompt elicits a verb paradigm.

In both the computer-based and smartphone approaches, participants are assigned a user ID (four-character string) that is entered before starting the survey. Each audio file, recorded with a sampling rate of 44.1 kHz as an uncompressed .WAV, is encrypted and transferred to our server at the University of Bern. This enables real-time monitoring of proper capture and transmission of recordings, allowing us to check whether recordings are clipped in amplitude by inspecting the waveform displayed on top of a web-interface (Figure 3).

The reason why the audio files are transmitted one by one is because uncompressed .WAV files are large in size (one-word recordings, for example, are typically between 200 and 800 KB; while the size of the 10-min spontaneous interview typically ranges between 50 and 100 MB). This allows us to avoid smartphone buffering problems. Collecting linguistic data with smartphone applications is not new (cf. forthcoming *Linguistics Vanguard* special collection “Using smartphones to collect linguistic data” edited by Adrian Leemann and Nanna Hilton). What is new in the paradigm proposed, however, is the supervision of each participant for the entirety of the recording session via Zoom videolink (Zoom 2020). In this approach, the fieldworker is able to follow along with the participant’s progress and intervene if necessary. A PowerPoint file is open on the researcher’s own computer, showing one survey item per slide (Figure 4, left); they can simultaneously see the actions of the interviewee via Zoom (Figure 4, top right) and monitor the quality of the incoming audio files (Figure 4, bottom right).

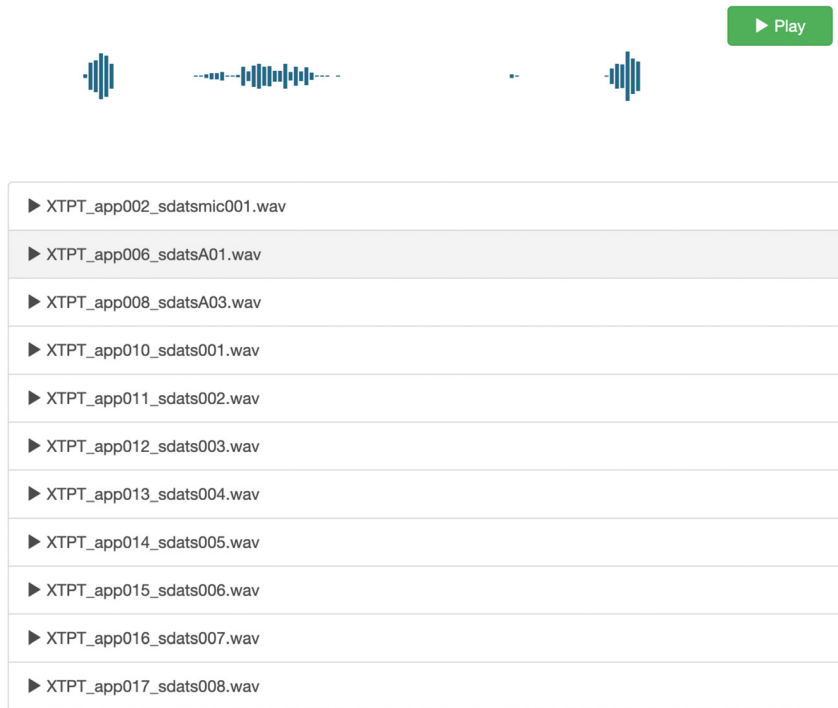
The dialect mapping task (cf. Table 1), where participants are asked to draw dialect regions of German-speaking Switzerland, was carried out manually on a piece of paper before the outbreak of COVID-19 (Figure 5, top panel). We



**Figure 1:** Typical recording setup before the outbreak of COVID-19; study participant on the left, researcher monitoring the recording session on the right.



Figure 2: SpeechRecorder screenshots (left) and smartphone screenshots of the same prompts (right).



**Figure 3:** Web-interface for monitoring the capture of audio files (bottom) and for inspecting and listening to waveforms (top).

are now using an online interface to carry out the mapping task in a web-browser (Figure 5, bottom panel), which allows participants to draw dialect regions with a computer mouse. The digital background map is the same as the one used in the in-person interviews. The regions drawn are automatically georeferenced, enabling us to collect latitude and longitude data in electronic format.

### 3 Comparing in-person versus online interviews

To compare participants' experience of in-person and online interviews, we conducted an online survey using a between-subjects and within-subjects design.

#### 3.1 Participants

Three cohorts ( $n_{\text{total}} = 36$ ) participated in the study (see Table 2). In the between-subjects comparison, 16 people who participated in the in-person interviews completed our survey, of which five were age 65+ and 11 were 20–35 (Cohort I); another 16 people who participated in the online interview took part, with the same age distribution (Cohort II). For the within-subjects comparison, four people – two older and two younger – participated in both in-person and online interviews and took the survey (Cohort III).

At the time of submission, we had 52 responses in the study from participants who could have been considered in Cohort I. To keep the results of the in-person and online modes comparable, we randomly sampled these respondents so as to equal the number in Cohort II. The number of older and younger participants was kept equal across both modes.

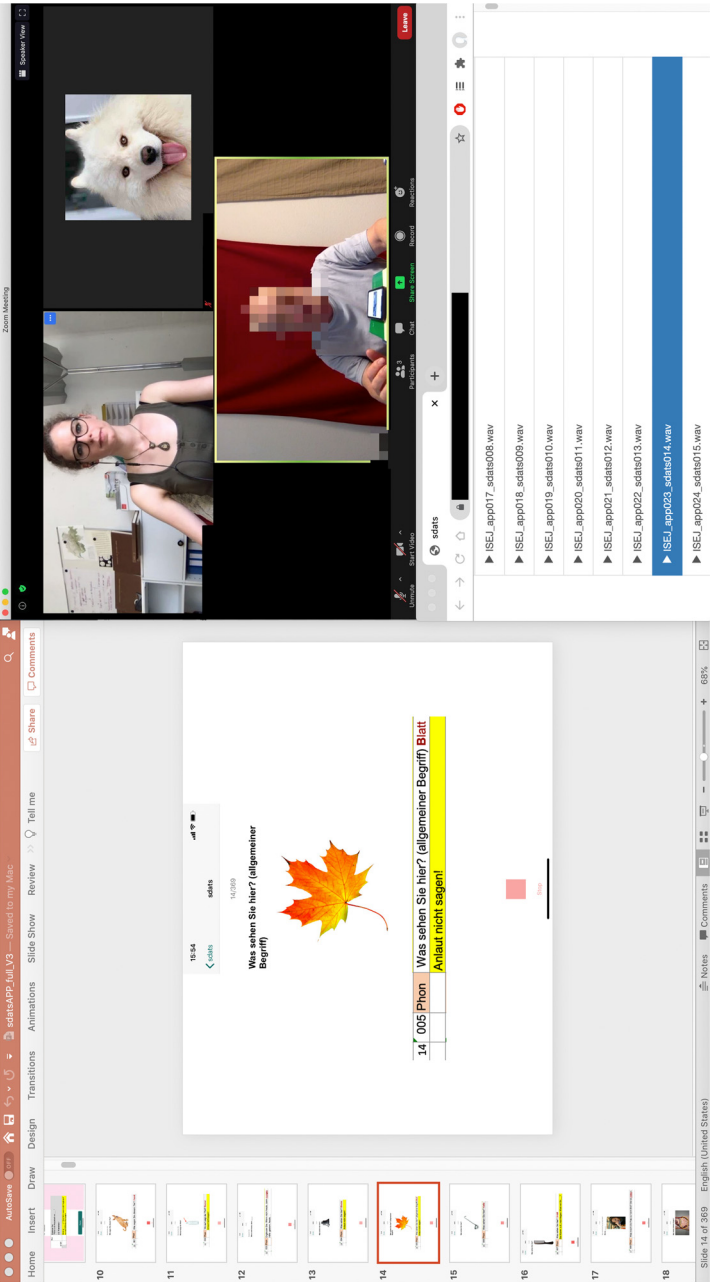
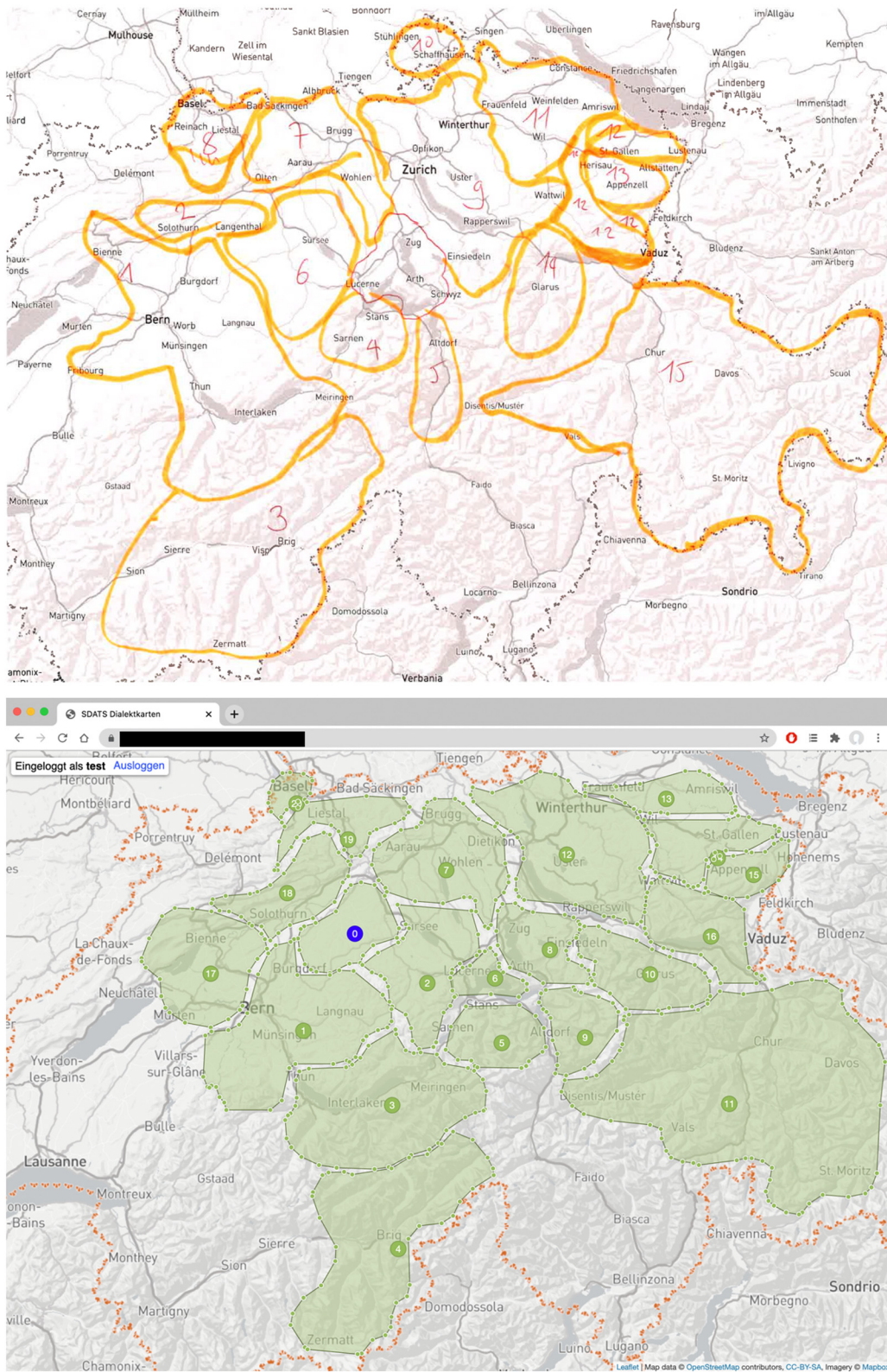


Figure 4: Fieldworker’s screen: PowerPoint file with prompts (left), Zoom session with participant (top right), interface for audio files (bottom right).

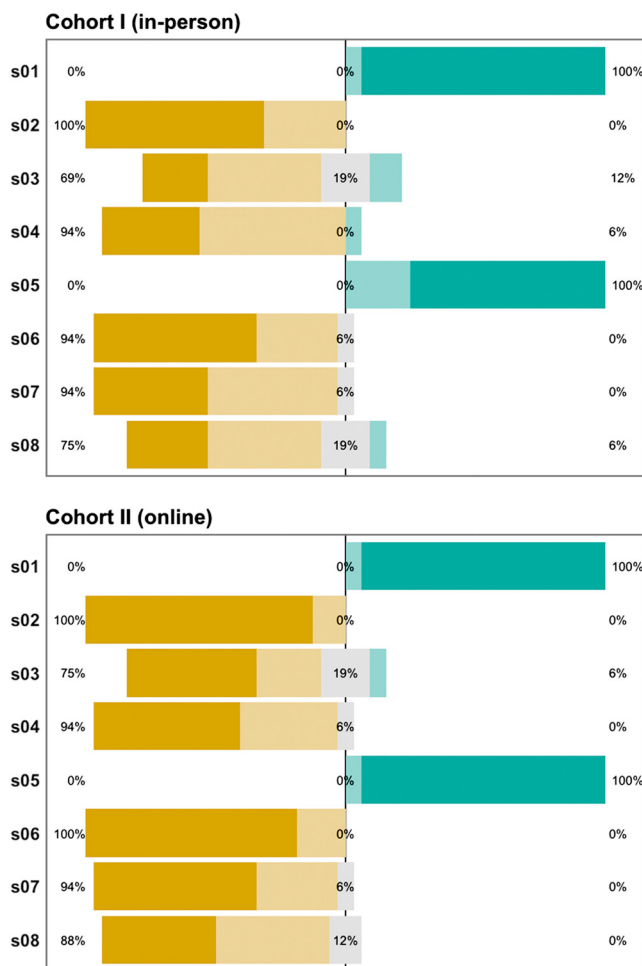




**Figure 5:** Dialect mapping task – done by hand before the outbreak of COVID-19 (top panel) and now done on a web-interface (bottom panel).

**Table 2:** Participants in the study.

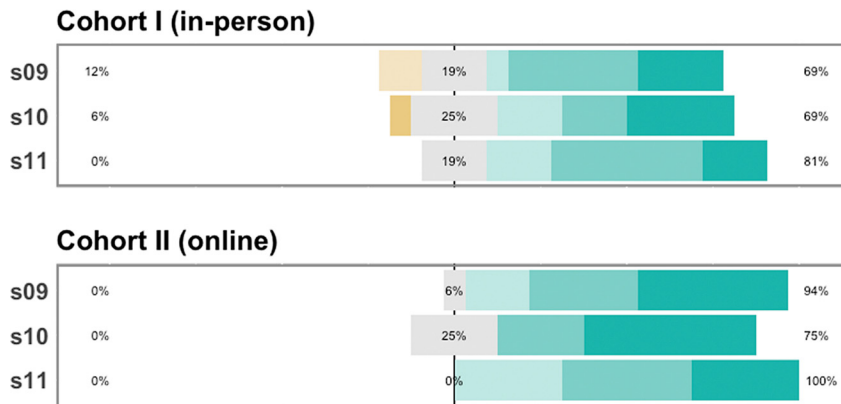
Design	Age	In-person	Online
Between-subjects	Age 65+	Cohort 1 (Feb/Mar 2020) n=5 (2f, 3m)	Cohort II (June 2020) n=5 (1f, 4m)
	Age 20–35	n=11 (8f, 3m) Cohort III (t1=Feb/Mar, t2=June 2020)	n=11 (6f, 5m)
Within-subjects	Age 65+	1f, 1m	
	Age 20–35	1f, 1m	

**Figure 6:** Stacked bar chart of responses to statements s01–s08: Cohort I (in-person, top) versus Cohort II (online, bottom). See Appendix for statements.

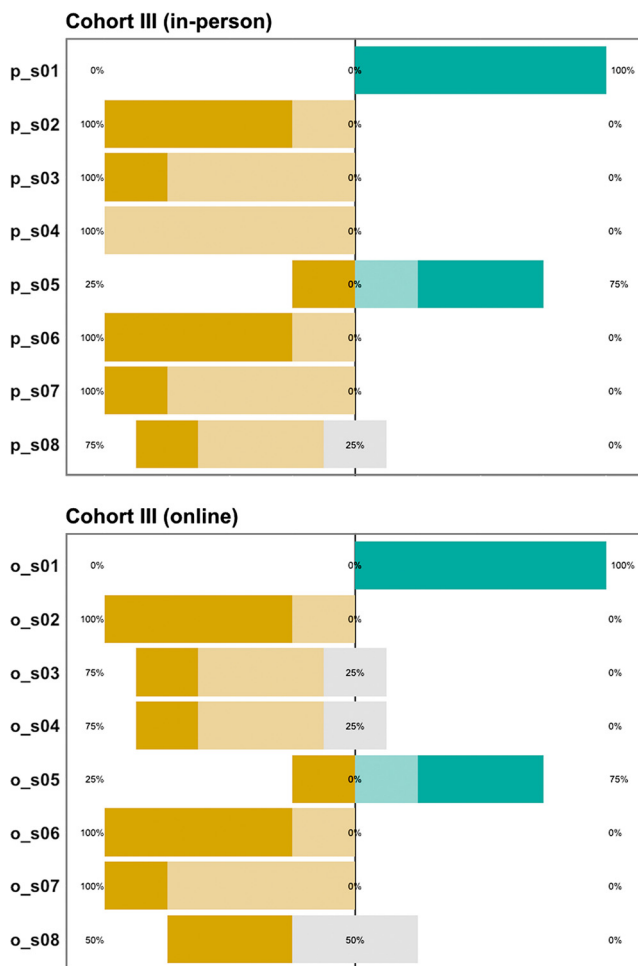
### 3.2 Material and procedures

The online survey was created using Questback (2020) and contained a total of 11 statements and questions (see Appendix). The statements to which participants were asked to indicate agreement or disagreement were geared towards eliciting responses on clarity of instructions, anxiety levels, difficulty, degree of formality, and perceived support and supervision by the interviewer. A link to the survey was distributed via email and interviewees participated on a voluntary basis. Open questions on general impressions, potential problems, and suggestions for improvement were analyzed qualitatively. Responses to statements were collected using



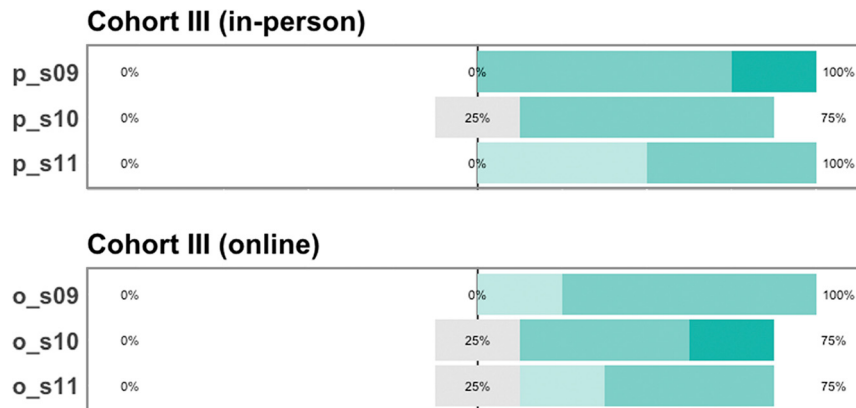


**Figure 7:** Stacked bar chart of responses to semantic differentials s09–s11: Cohort I (in-person, top) versus Cohort II (online, bottom). See Appendix for statements.



**Figure 8:** Stacked bar chart of responses to statements s01–s08: Cohort III (in-person, top) versus Cohort III (online, bottom). See Appendix for statements.

Likert scales and were initially analyzed visually and descriptively. To test for group differences between Cohorts I and II, Mann–Whitney U tests were calculated. All analyses were run in R (R Core Team 2020; Likert package: Bryer and Speerschnieder 2016; cf. technical report, data, and scripts at <https://osf.io/s9z4q/>).



**Figure 9:** Stacked bar chart of responses to statements s09–s11: Cohort III (in-person, top) versus Cohort III (online, bottom). See Appendix for statements.

### 3.3 Results

Figure 6 shows the degree to which participants from Cohorts I and II ( $n = 16$  for each cohort, between-subjects design) agreed or disagreed with each statement on a five-level polarized Likert scale (s01–s08; see Appendix for full statements). Cohort I responses are displayed in the top panel, with Cohort II responses in the bottom panel. Neutral responses are marked in grey, disagreement in brown, and agreement in turquoise, with darker shades indicating stronger (dis-)agreement. Percentages indicate the proportion of participants who disagreed (left), gave neutral responses (middle), or agreed with each statement (right). For instance, in Cohort I, 69% disagreed with statement s03 ('I felt observed during the task'), 19% opted for a neutral answer, and 12% agreed.

Figure 7 indicates how participants from Cohorts I and II rated the interview atmosphere on a seven-level Likert response to semantic differentials (s09–s11; see Appendix). Percentages on the left side indicate the proportion of participants who preferred the adjective with a negative connotation, percentages of neutral responses are shown in the middle, and percentages of positive responses are shown on the right. For example, for s09 (artificial vs. natural) the interview situation was evaluated as rather artificial by 12% of the participants in Cohort I, while 19% opted for a neutral position and 69% evaluated the situation as rather natural (the darker the turquoise, the more positive the perception of the interview situation).

Overall, the results in Figures 6 and 7 suggest a similar perception of the interview situation for both cohorts; responses differ only marginally in that participants in Cohort II (online) reported stronger opinions than those in Cohort I (in-person). Cohort II (online) participants disagreed more strongly with negative statements (02, 03, 04, 06, 07, 08) and, in general, reported higher agreement with positive statements (05, 09, 10, 11). Participants in this cohort appear to have felt less stressed and observed, perceived the tasks as less difficult, and indicated that they accommodated less to the dialect of the interviewer. They further indicated feeling more strongly supported by the interviewer and evaluated the atmosphere as more natural, relaxed, and easy-going. Mann–Whitney U tests were run for each item to test for potential group differences. None of the differences, however, reached statistical significance (cf. <https://osf.io/s9z4q/>).

Responses from Cohort III ( $n = 4$ ), who participated in both modes in a within-subject design, are shown in Figures 8 and 9.

This small cohort evaluated the in-person interview as slightly better than the online interview. Their responses suggest that they felt slightly more nervous and observed and found the tasks more difficult in the online mode (statements 03, 04, 08). They further perceived the online mode as less easy-going than the in-person atmosphere. Note, however, that no statistical tests were run as this cohort only consisted of four participants.

## 4 Discussion

In this section, we discuss three key issues: (1) How we interpret the patterns found in Section 3, (2) benefits and pitfalls of the method proposed, and (3) ethical considerations.

### 4.1 Key findings of the study

The results of our survey reveal two interesting trends: the between-subjects study indicates substantial overlap in terms of how in-person and online interviews were perceived. These results reflect the findings of Archibald et al. (2019), who report equally high scores in the usefulness of online versus in-person interviews for qualitative research (see Section 4.2.1). Surprisingly, though, the online-only cohort trended toward perceiving the interview as less stressful, less difficult, and more relaxed than the in-person interview cohort did (Figures 6 and 7). There are several possible explanations for these trends. First, this could be due to the more frequent use of videoconferencing in times of COVID-19, for both older and younger people. This increased frequency of use may in turn boost the degree to which people feel comfortable in this virtual setting. Second, participants could have felt safer being physically distanced from the interviewer. This may have tipped them toward expressing an even more positive evaluation of the online interview setting than they would express under normal circumstances regarding an in-person interview; that is, the feeling of doing the ‘right’ thing by participating online during a global emergency may have increased positive evaluations. Another unanticipated finding is that the small cohort of within-subjects participants, i.e. those who took part in both in-person and online interviews, appear to slightly favor in-person interviews. In considering why this might have been, we note that the two older participants were told that they were the first ones taking part in the in-person interviews and seemed to be particularly ecstatic about this. The young woman was interviewed by her own sister, which is also likely to have created a particularly comfortable situation during the in-person interview.

### 4.2 Benefits and challenges of online data collection

The following section summarizes the main benefits of the type of data collection proposed, but more importantly, reflects on potential challenges. The latter include rapport building, Internet connectivity, computer literacy, recording quality, and room acoustics.

#### 4.2.1 Benefits

There are a number of benefits that a combination of smartphone app and supervision via Zoom offers. For one, it is cost-effective, especially if research is being conducted with participants from remote regions. The protocol proposed is thus particularly useful for under-resourced researchers conducting interviews from afar. Secondly, it is logistically efficient. For in-person interviews, the researcher has to travel to selected locations, where rooms with proper sound conditions have to be arranged. Moreover, non-overlapping appointments with at least two or three participants in the region are typically arranged for a day trip to be efficient. The smartphone app plus Zoom combination no longer requires travel. This lack of travel was one of the main reasons why some participants in Cohort III preferred the online experience to the in-person interview, as attested in the qualitative feedback. These benefits have also been noted by Archibald et al. (2019) in a study on the usefulness of Zoom for qualitative research. The majority of their participants preferred Zoom-based interviews to in-person interviews due to ease of access, time-effectiveness, not having to be mobile (which expands the accessibility of this paradigm, as people who are handicapped/disabled are easier to include in ways they might have been excluded from traditional dialectological research), and fitting in better with busy work schedules. Thirdly, the digital paradigm saves time spent on cleaning the data, as app recordings contain

only the item in question, whereas the pre-COVID-19 setup recorded the entire interview, sliced into different larger files. Besides, having participants draw perceived dialect areas on a web-interface saves the time of having to manually digitize regions of hand-drawn maps post-elicitation.

Finally, and probably most importantly, the proposed paradigm allows us to continue to do fieldwork during the COVID-19 pandemic. At the same time, the paradigm is applicable to in-person settings as well: assuming the pandemic is ‘resolved’, the same app can still be used in the field, with the Zoom session replaced by the interviewer in a face-to-face setting. This allows for keeping recording setups maximally similar.

#### 4.2.2 Challenge 1: Rapport building

There are several challenges involved in online linguistic data collection. One potential challenge involves building rapport with interviewees over the course of the interview. Rapport, or empathy, can be crucial to a successful linguistic interview, as researchers want participants to trust them and feel comfortable enough to open up to them (cf. King and Horrocks 2010). Studies are inconclusive as to whether videoconferencing enables the same degree of rapport as interviewing in person (cf. Cater 2011). While participants can see and hear the interviewer, it is impossible for them to make eye contact (although Apple did at one point try out ‘FaceTime Attention Correction’ in iOS 13/Beta 3, which makes it seem as though people are looking directly into the camera during FaceTime video calls).

#### 4.2.3 Challenge 2: Internet connectivity

By far the largest drawback of the online method is that the paradigm is predicated on stable Internet connections. This may be prohibitively difficult for some research in language documentation. For example, in Papua New Guinea, where more than 800 languages are spoken (Eberhard et al. 2020), Internet penetration stands at about 12% (Datareportal 2020), meaning that online data collection is of very limited use. In addition to stable Internet, devices also need to be available and powered (for an analysis of power consumption in smartphones, see Carroll and Heiser 2010). Archibald et al. (2019) and Deakin and Wakefield (2014) further point out that dropped calls or issues in video connectivity are to be avoided, as both the interviewee and interviewer run the risk of not being able to interpret non-verbal cues.

#### 4.2.4 Challenge 3: Computer (i)literacy

Archibald et al. (2019) report that the vast majority of participants in their study on the use of Zoom videoconferencing for qualitative data collection had issues joining a Zoom session (88%, i.e. 14 of 16 practice nurses in South Australia). This brings up the question of computer literacy skills and how these vary within and between communities. Tsai et al. (2017) report that children and middle-aged adults outperform the elderly in gestural operation of smartphones. It remains to be seen how the younger and older cohorts differ in terms of usability of the *SDATS* app. Half of the participants planned for inclusion in *SDATS* ( $n = 500$ ) are 65+. So far, when looking at the open questions from our study, some of the older participants who took part in the online mode reported having substantial technical difficulty with the virtual setup and would have preferred an in-person interview for this reason. In addition, a substantial number of potential participants declined our request for an online interview due to their lack of technical equipment or skills.

#### 4.2.5 Challenge 4: Recording quality

Aside from connectivity-related problems, the main technical issue in online data collection is variation in the recording quality of smartphones. This is a major issue for *SDATS*, as the project is heavily focused on phonetic variation and change (for projects focused on lexical, morphosyntactic, or other non-phonetic variation, Sections 4.2.5 and 4.2.6 are not particularly pertinent). There have been a number of studies on the effects of smartphones on acoustic features. De Decker and Nycz (2011) examined vowel measurements based on data

retrieved from Ediol, iPhone, MacBook, and Mino devices. They found that recordings on a first-generation iPhone were suitable for formant analysis, as the overall shape of the vowel space does not seem to be affected substantially. More recently, Rathcke et al. (2017) studied the effects of different equipment on F1 and F2 measurements in real-time sociolinguistic analyses, reporting that formant measurements using LPC tracing are sensitive to the technical specifications of the microphones used at different points in time. Their key recommendation is that linguists should keep the recording equipment controlled in order to maximize comparability between recording sessions. In terms of voice quality measures, Grillo et al. (2016) tested the effect of different devices (head-mounted condenser microphone, iPhones 5 and 6s, and Samsung Galaxy S5) on measures of voice quality (e.g. F0, jitter, shimmer, HNR). Results yielded no significant differences across devices. Jannetts et al. (2019) performed similar analyses to Grillo et al. (2016), but report somewhat more sobering findings, namely that acoustic parameters can be measured based on smartphone recordings, but reliability is heavily dependent on the parameter investigated. Measurements of mean F0 and SCPP (smoothed cepstral peak prominence) proved relatively robust across devices, but jitter and shimmer did not. The findings of Jannetts et al. (2019) and Rathcke et al. (2017) indicate that ideally, device specifications should be the same for all participants. This poses a problem for projects such as *SDATS*, as there are currently more than 20 major vendors within the smartphone industry (Metodieiev 2019). To account for this variation in devices, *SDATS* collects device and operating system specifications before the participant begins their online interview. This information can subsequently be factored into statistical modelling.

To avoid smartphone recordings altogether, one could opt for cloud recording on Zoom itself. We believe this is unfavorable for three reasons. First, diversity in a participant's laptop/computer/webcam/and microphone would exponentially increase the amount of between-participant variation in recording quality. Secondly, Zoom applies audio compression algorithms that lead to inferior audio quality. Bulgin et al. (2010), for instance, studied the effect of Skype's audio compression (similar to Zoom's) on vowel formant measurements, reporting that Skype distorts the vowel space in non-linear ways; they advise against its use for the collection of audio data. Thirdly, this would require a thorough review of Zoom's terms and conditions with participants in order to obtain properly informed consent. It may have been more difficult to gain participants' consent to record via Zoom, as the company has received bad press over its handling of security issues in the first months of 2020 (cf. The Guardian 2020).

#### 4.2.6 Challenge 5: Room acoustics

Finally, room acoustics – which are bound to vary between online participants – are another issue which can greatly impinge on acoustic analyses. De Decker (2016), for instance, found that white noise is most harmful for LPC-based formant measurements. However, formant measurements based on enveloped speech (e.g. a 60 Hz hum, roughly the sound of a fan), were reported to accurately reflect measurements without noise present. Most recently, Hansen and Pharaon (in prep.) studied the effect of different microphones and room acoustics on vowel measurements, reporting that differences between simultaneous recordings of the same vowel were smaller in recordings made “in the wild” than in studio and anechoic chamber. They further found an effect of microphone and acoustic setup on vowel quality, with the measurement of high back vowels being particularly affected. All of these provisions need to be taken into consideration when instructing participants in how to set up for the online interviews. We try to mitigate these adverse effects by telling participants to drape their rooms with fabrics, if possible (cf. Figure 4, top right panel) and through frequent checks of the audio recording quality during the interview.

### 4.3 Ethical considerations

In the paradigm proposed, participants need to be able to view and consent to the terms and conditions of the software used for recording, given that these are legally binding. In the case of *SDATS*, where the audio track is recorded independently of the Zoom video feed, no further action needs to be taken aside from the privacy and consent forms the participants already sign before the interview session (GDPR-compliant, vetted by legal



services at the University of Bern). The audio recordings in *SDATS* are encrypted and transferred item-by-item to servers that are physically located at the University of Bern.

## 5 Conclusions

The main goal of the current contribution is to present a case for switching from in-person to online linguistic data collection during ongoing emergencies like the COVID-19 pandemic, which has been particularly disruptive for early career researchers who are affected by the impossibility of in-person fieldwork. We have presented an open source app which enables remote linguistic data collection with physical proximity out of the equation. The approach proposed is not meant to replace in-person interviews altogether, but rather to complement existing data collection paradigms. Our study, which compares participant experiences of in-person versus online data collection, indicates a high degree of overlap in experiences. We address numerous limitations of this new paradigm, with core weaknesses revolving around technical aspects such as variation in recording quality across smartphones and participants' varying computer literacy skills. Notwithstanding these limitations, we believe this approach to be sustainable and useful in the face of pandemics such as COVID-19 – as long as participants have stable Internet connectivity, a smartphone, and a laptop or desktop computer with a webcam. This research has thrown up a number of questions in need of further investigation. For one, it will be worthwhile pursuing whether there are measurable differences in linguistic behavior that might stem from a difference in elicitation mode (in-person vs. online). Cohort II (online), for example, reported they had accommodated less to the interviewers. How this can be corrected in subsequent analyses must be addressed in future studies. Considerably more future work is further needed to determine the degree to which varying smartphone specifications affect different acoustic measurements, as it is likely that many linguists will have to switch (or already have switched) to smartphones as an alternative mode of data collection. This contribution makes an effort to expand the accessibility and adaptability of a smartphone app to the scientific, and in particular linguistic, research community. Researchers interested in customizing the app and infrastructure to their purposes are encouraged to get in touch with the first author.

**Acknowledgements:** We thank Daniel Wanitsch at [www.ibros.ch](http://www.ibros.ch) for the development of the *SDATS* app and Thomas Kettig (University of Hawai'i at Mānoa), two anonymous reviewers, and the associate editor for useful comments and edits on earlier versions of this manuscript.

## Appendix

The survey covered the following topics and sets of questions:

- How well do you remember the interview? [Polarized 7-level Likert scale]
- Agreeing with statements [Polarized 5-level Likert scale]:
  - s01 The interviewer's explanations were clear
  - s02 I felt stressed because of the interviewer's presence
  - s03 I felt observed doing the task
  - s04 I thought the tasks were difficult
  - s05 The interviewer was able to help me when I was stuck
  - s06 I found the structure of the survey complicated
  - s07 I think I may have accommodated to the dialect of the interviewer
  - s08 I was nervous during the interview
- General impressions of the interview

- s09 Artificial versus natural [Polarized 7-level Likert scale]
- s10 Serious versus relaxed [Polarized 7-level Likert scale]
- s11 Stressed versus easy-going [Polarized 7-level Likert scale]
- Open questions:
  - Open question about general impressions of the interview [Text box]
  - Open question about whether there were technical problems during the interview [Text box]
  - Open question about which aspects could be improved in the survey [Text box]
- Indication of age and gender

Additional questions were tailored for Cohorts II & III:

Cohort II (online interview):

- How often do you participate in videoconferencing calls (Skype, Zoom, or similar)?
- How user-friendly was the app? [Text box]
- What was your general impression of conducting the interview with a smartphone app while being monitored on Zoom? [Text box]

Cohort III (in-person & online interview):

- How often do you participate in videoconferencing calls (Skype, Zoom, or similar)?
- How user-friendly was the app? [Text box]
- What was your general impression of conducting the interview with a smartphone app while being monitored on Zoom? [Text box]
- Which formant, in-person or online, did you prefer? Motivate your choice [Text box]

Additional information on Figures 8 and 9: As Cohort III answered the same questions for both modes, these are labelled with an initial ‘p\_’ for statements referring to the in-person interview (e.g. ‘p\_s01’), and with an initial ‘o\_’ for the online interview, respectively (e.g. ‘o\_s01’).

## References

- Archibald, Mandy M., Rachel C. Ambagtsheer, Mavourneen G. Casey & Michael Lawless. 2019. Using Zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods* 18. 1–8.
- Bryer, Jason & Kimberly Speerschneider. 2016. likert: Analysis and visualization Likert items. R package version 1.3.5. <https://CRAN.R-project.org/package=likert> (accessed 11 August 2020).
- Bulgin, James, Paul De Decker & Jennifer Nycz. 2010. Reliability of formant measurements from lossy compressed audio. Poster presented at *The British Association of Academic Phoneticians Colloquium*, University of Westminster, 29–31 March.
- Carroll, Aaron & Gernot Heiser. 2010. An analysis of power consumption in a smartphone. *USENIX Annual Technical Conference* 14. 21–21.
- Cater, Janet K. 2011. SKYPE – A cost-effective method for qualitative research. *Rehabilitation Counselors & Educators Journal* 4. 10–17.
- Christiner, Markus & Susanne Maria Reiterer. 2015. A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience* 9(482). (accessed 11 August 2020). <https://doi.org/10.3389/fnhum.2015.00482>.
- Datareportal. 2020. <https://datareportal.com/reports/digital-2020-papua-new-guinea> (accessed 11 August 2020).
- De Decker, Paul. 2016. An evaluation of noise on LPC-based vowel formant estimates: Implications for sociolinguistic data collection. *Linguistics Vanguard* 2(1). <https://doi.org/10.1515/lingvan-2015-0010> (accessed 11 August 2020).
- 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.
- Deakin, Hannah & Kelly Wakefield. 2014. SKYPE interviewing: Reflections of two PhD researchers. *Qualitative Research* (14). 1–14.
- Draxler, Christoph & Klaus Jänsch. 2004. SpeechRecorder – A universal platform independent multi-channel audio recording software. *Proceedings of the Fourth Conference on Language Resources and Evaluation* 559–562.

- Eberhard, David M., Gary F. Simons, & Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the world*, 23 ed. Dallas, TX: SIL International. Online version: <http://www.ethnologue.com> (accessed 19 August 2020).
- Flaxman, Seth, Swapnil Mishra, Axel Gandy & Samir Bhatt. 2020. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Imperial College London. <https://doi.org/10.25561/77731> (accessed 11 August 2020).
- Grillo, Elizabeth U., Jenna N. Brosious, Staci L. Sorrell & Supraia Anand. 2016. Influence of smartphones and software on acoustic voice measures. *International Journal of Telerehabilitation* 8(2). <https://doi.org/10.5195/ijt.2016.6202> (accessed 11 August 2020).
- Hansen, Gert Foget & Nicolai Pharo. in prep. *Microphones and formant estimates*.
- Jannetts, Stephen, Felix Schaeffler, Janet Beck & Steve Cowen. 2019. Assessing voice health using smartphones: Bias and random error of acoustic voice parameters captured by different smartphone types. *International Journal of Language & Communication Disorders* 54(2). 292–305.
- King, Nigel & Christine Horrocks. 2010. *Interviews in qualitative research*. Los Angeles: Sage Publications Ltd.
- Leemann, Adrian & Nanna Hilton (to appear). Using smartphones to collect linguistic data. *Special Collection in Linguistics Vanguard*.
- Leemann, Adrian, Péter Jeszenszky, Carina Steiner, Melanie Studerus & Jan Messerli. 2020a. SDATS Corpus – Swiss German dialects across time and space. Retrieved from [osf.io/s9z4q](https://osf.io/s9z4q) (accessed 19 August 2020).
- Leemann, Adrian, Péter Jeszenszky, Carina Steiner, Melanie Studerus & Jan Messerli. 2020b. <https://play.google.com/store/apps/details?id=ch.sdats> and <https://apps.apple.com/ch/app/sdats/id1516597765?ign-mpt=uo%3D4>.
- Metodiev, Boris. 2019. <https://www.strategyanalytics.com/access-services/devices/mobile-phones/smartphone-country-share/market-data/report-detail/global-smartphone-vendor-asp-and-revenue-share-by-region-q2-2019?Related> (accessed 11 August 2020).
- Newzoo. 2018. <https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users/> (accessed 11 August 2020).
- Pressnitzer, Daniel, Jackson E. Graves, Claire Chambers, Vincent De Gardelle & Paul Egré. 2018. Auditory perception: Laurel and yanny together at last. *Current Biology* 28(13). R739–R741.
- Preston, Dennis R. 2010. Perceptual dialectology in the 21st Century. In Christina Ada Anders, Markus Hundt & Alexander Lasch (eds.), *‘Perceptual Dialectology’*. *Neue Wege der Dialektologie*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110227529.1> (accessed 11 August 2020).
- Questback GmbH. 2020. *EFS survey*. Version EFS 2020 Cologne: Questback GmbH.
- R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/> (accessed 11 August 2020).
- Rathcke, Tamara, Jane Stuart-Smith, Bernard Torsney & Jonathan Harrington. 2017. The beauty in a beast: Minimising the effects of diverse recording quality on vowel formant measurements in sociophonetic real-time studies. *Speech Communication* (86). 24–41. <https://doi.org/10.1016/j.specom.2016.11.001>.
- Reuters. 2020. <https://www.reuters.com/article/us-health-coronavirus-who-peak/who-warns-of-second-peak-in-areas-where-covid-19-declining-idUSKBN2311V> (accessed 11 August 2020).
- Sakel, Jeanette & Daniel L. Everett. 2012. *Linguistic fieldwork: A student guide*. Cambridge: Cambridge University Press.
- SDS = Sprachatlas der deutschen Schweiz 1962–2003. Bern (I–VI)/Basel (VII–VIII): Francke.
- SED = Survey of English Dialects 1962. Harold Orton & Eugen Dieth. Leeds: E. J. Arnold & Son.
- The Guardian. 2020. <https://www.theguardian.com/technology/2020/apr/02/zoom-technology-security-coronavirus-video-conferencing> (accessed 11 August 2020).
- Tsai, Tsai-Hsuan, Kevin C. Tseng & Yung-Sheng Chang. 2017. Testing the usability of smartphone surface gestures on different sizes of smartphones by different age groups of users. *Computers in Human Behavior* 75. 103–116.
- Xamarin. 2020. <https://dotnet.microsoft.com/apps/xamarin> (accessed 11 August 2020).
- Zoom. 2020. <https://zoom.us/> (accessed 11 August 2020).