Documentary linguists and risk communication: views from the virALLanguages project experience

Abstract: Linguists are seldom, if ever, engaged in work aimed at communicating risk to the general public. The COVID-19 global pandemic and its associated infodemic may change this state of affairs, at least for documentary linguists. Documenting languages may bring researchers in direct contact with communities speaking minority or marginalized languages and gain key insights into their communicative ecologies. By being both immersed in local networks and more or less knowledgeable about the community’s communicative habits, documentary linguists appear to be placed in a unique position to contribute to communicating risk in ways that are better tailored to the community and, therefore, potentially quite effective locally. Furthermore, adding work in risk communication to their agenda may also stimulate documentary linguists to find new models for “giving back” to the communities they work with. In order to provide a concrete example of how all this may play out in concrete terms, we illustrate the virALLanguages project.

Keywords: COVID-19; crisis and risk communication; intercultural communication; language documentation

1 Introduction

You’re a human being, and your time as a human being should be socially useful. It doesn’t mean that your choices about helping other people have to be within the context of your professional training as a linguist. Maybe that training just doesn’t help you to be useful to other people. In fact, it doesn’t. (Noam Chomsky as reported in Olson et al. 1991: 30)

One of the earliest responses to the worldwide spread of COVID-19 has been at the level of crisis and emergency risk communication (CERC) through the multiplication of translations of preventive measures in the world’s languages – what McCulloch (2020) referred to as “history’s biggest translation challenge”. Risk communication – that is, “the flow of information and risk evaluations back and forth between academic experts, regulatory practitioners, interest groups, and the general public” (Leiss 1996: 86) – is a domain of language work that linguists do not normally frequent and whose theories and practices they have not contributed to shaping (see, e.g., Bourrier 2018: 3). On the surface, then, it would seem that linguists willing “to be useful to others” through work within risk communication targeting the COVID-19 pandemic would be doing so outside of “the context of their professional training as linguists”, thus exemplifying Chomsky’s argument above. By contrast, in this article we argue that Chomsky’s point is misplaced at least for a specific kind of linguist – those who, as part of their work, document (endangered) languages in their broader sociocultural contexts.
and embrace a view of language documentation that includes what Anthony Woodbury (2011) called the “documentation of contemporary communicative ecologies”.1

There are two main reasons that we think legitimize our claim. First, building mutually beneficial relationships with the target community through one’s own work has gradually become integral, rather than optional, to the profession of documentary linguists (see, e.g., Czaykowska-Higgins 2018; Dobrin and Berson 2011; Rice 2012). Second, an important goal of documentary linguists is to gain insights into their target community’s language ideologies and, more broadly, communicative habits. This kind of knowledge may allow them to approach risk communication from a perspective that, being sensitive to local realities, might favor the development of campaigns that are better tailored to and, therefore, have more chances of success in a particular community.

To put this in perhaps clearer terms: Because of its magnitude, the COVID-19 pandemic presents itself as a perfect case in which many people would like to be of help regardless of whether this is or is not part of their professional training. In this context, we claim that documentary linguists are uniquely positioned to “be useful to others” through work in risk communication in ways that would both fall within and benefit the exercise of their profession.

As a way to exemplify how this possibility may play out in concrete terms, we present here the virALLanguages project, an initiative that we ran as part of a wider international team of volunteers including academics, students, and members of communities speaking minority, marginalized, or endangered languages. VirALLanguages was launched on the LINGUIST List on 17 April 2020 – three weeks after it was first conceived – and none of the initial team members were experts in CERC. Many of the points we make in this article have emerged in hindsight, and only after we dealt more closely with CERC literature, as this was not possible while virALLanguages was still active. In the hope that our experience and acquired knowledge might foster discussion among (documentary) linguists about the role that they can play in the CERC domain, we decided to provide here a coherent narrative that includes discussions and intellectual connections we could not have foreseen in 2020, rather than a genuine but necessarily intellectually poorer report of what we did.2

After introducing the mission of virALLanguages (Section 2), we deal with its two main components: facilitating the production of public health messages in as many languages as possible (Section 3) and leveraging community resources to build trust in communicators (Section 4). In Section 5 we discuss concrete issues of project design.

2 The virALLanguages project mission

The mission of virALLanguages can be summarized in two goals: (1) to facilitate the translation of public health information useful to contain the spread of the SARS-CoV-2 virus into as many languages as possible, crucially including endangered ones so that ideally all the world’s languages are covered; and (2) to maximize the impact of messages by embedding them in community-specific ideologies, practices, and networks of trust—i.e. by making them as culturally-appropriate as possible. Before we illustrate how these goals were pursued in the project (Section 5), it is important to clarify their broader context and motivations.

3 Translating messages into as many languages as possible

By the end of 2020, COVID-19 preventive measures had been translated into about 700 languages, around 630 of which are under-resourced (see Figure 1). Many consider this to be an impressive figure and this is best

1 For the sake of convenience, it is to this specific subcategory of linguists that we will refer to in the article as “documentary linguists”, although we acknowledge that language documentation is practiced by researchers of quite diverse backgrounds.

2 Message production activities ended in December 2020, i.e., when the availability of vaccines in some countries required significant updates. Since then, virALLanguages was the main topic in a talk story session at ICLDC 7 and project activities were limited to preparing materials, getting feedback, and seeking funds for a new project phase (see Supplementary Material S.1).
appreciated if one compares it with the reductionist stance on linguistic diversity that is found in CERC scholarship and practice. To cite but one example, the US Centers for Disease Control and Prevention (CDC 2014: 62) estimate the number of languages that “are most commonly spoken in US households” at “about 30”, implicitly setting this figure as a goal for the number of languages into which translations should be made in order to get a satisfactory reach. However, no fewer than 350 languages are reported in the US Census (2015), a figure that many would deem to be a considerable underestimation (see, e.g., the Endangered Language Alliance’s report that there are over 650 languages spoken in New York City alone; ELA 2020).

To our knowledge, CERC scholars have never clearly articulated the scientific grounds for such a reductionist stance on linguistic diversity, so one is left with the impression that it is rooted exclusively in a logic of (economic) efficiency – like the “90–10 principle” which postulates that it would be inefficient to incur 90% of costs to address the last 10% of the risk (Lofstedt 2003). We would like to stress that producing content in the languages that are understood by most people in a given area is a strategy that implicitly rests on the assumption that people whose (minority or marginalized) L1 is not targeted will understand the message anyway thanks to their bi- or multilingual competences. This does not problematize possible reception differences between the use of one’s L1 or L2s.3

Available evidence indicates that messages received in one’s L1 are more likely to trigger a stronger emotional response (see, e.g., Citron et al. 2020; Rivière 2016), thus making the content both more memorable (Mohanty 2020) and more likely to result in actions and changes in behavior (Tyng et al. 2017) than would messages delivered in one’s L2. Seen from this angle, the figure of 700 languages becomes less impressive as it

---

3 We realize this distinction may not always be straightforward, especially in small-scale multilingual environments (see, e.g., Lüpke 2010, 2015). In any case, our use of L1 has no bearing on the nature of the language used beyond the fact that it is a particularly familiar way of speaking in the community. This is further stressed by our emphasis on the message's cultural appropriateness (see Section 5.2) which is, first and foremost, a matter of style of delivery and is in no way connected with considerations of language purism or standardization.
excludes about 90% of the nearly 7,000 languages currently spoken in the world (Eberhard et al. 2021): according to official figures, speakers of these languages amount to about 13% of the world’s population (Esene Agbara 2021), which is more than enough to cause large-scale public health emergencies during an ongoing pandemic. This makes the supposed “efficiency” of CERC’s traditionally reductionist stance on language diversity questionable in this specific case.

4 Infodemic, impact, and trust

An infodemic – that is, the uncontrolled proliferation of diverse and often unsubstantiated information relating to the pandemic – “makes it challenging for people to know what or whom to trust, especially when faced with conflicting claims or information” (Gruzd et al. 2021). Uncertainty, skepticism, and distrust of regulators and communicators make risk communication less effective, so that preventive measures like social distancing, wearing personal protective equipment, or restricting travel may not achieve their public health goal as the general population may decide to not adopt them (e.g., Okereke et al. 2021). This is particularly consequential for minorities and marginalized groups in postcolonial contexts.

First, a majority of such groups live in countries of the Global South, where public health infrastructure is generally limited and so the response to COVID-19 is realistically restricted to behavior change (e.g., Ataguba and Ataguba 2020) since available resources leave very little scope for medical treatment (through, e.g., ICUs or the use of experimental drugs).

Second, while trust in governments and state institutions is in decline globally (Acemoglu et al. 2020; Bertsou 2019), in postcolonial contexts one can often speak of deeply rooted distrust of state institutions (see, e.g., Cho and Kirwin 2007; IFRC 2020) especially among marginalized groups (see, e.g., Mizuno and Okazawa 2009 for Africa) and due to nation-building cultural assimilation policies that “simply ignored the diversity of large segments of the population, in particular racial and linguistic minorities” (Torbisco Casals, 2015: 469). This is why, in these contexts, minority groups’ unwillingness to abide by state-promoted health policies might be fundamentally due to “perceived intentions to harm, rooted in [a] history of domination and exclusion” (Hermesh et al. 2020: 16; see also Bastide 2018).

Third, let us consider the different channels through which one is exposed to an infodemic. Put roughly: there is a big difference between getting a piece of information from a public Facebook or Twitter account versus a private WhatsApp group chat as the latter usually comprises (1) a relatively limited number of participants and (2) people one knows directly and respects. Due to these features, a WhatsApp chat may reasonably replicate collective contexts such as meetings and, depending on how the group is organized, these may connect with preexisting (i.e., real-world) patterns of trust. A telling example is provided by Naima Abubakar, who studies women’s social media use in Nigeria’s Kano State. Referring to his research participants’ attitude to never question the truthfulness of the messages they receive through their church-based WhatsApp group chat, Abubakar said “unless someone else that they trust more comes with a different truth, they won’t stop believing the original information” (Hitchen et al. 2019; cf. also NatCen and Wellcome Trust 2020). Furthermore, WhatsApp has a number of technical advantages that make it the primary communication tool for a majority of the people living in the Global South, especially in Africa (Boyd 2019; Vogt 2020), thus ensuring that it is the channel through which most people are exposed to all types of information.

Based on these points, one might wonder if McCulloch’s (2020) claim that the COVID-19 pandemic is “history’s biggest translation challenge” should not instead be rephrased as “history’s biggest trust-building challenge”. VirALLanguages was designed starting from the centrality of the problem of trust in

---

4 Here we consider a treatment medical only if it matches the standards of Western medicine based on the scientific method. This is not to be taken as an implicit dismissal of traditional medicines as invariably ineffective against COVID-19.

5 Clearly, the scope of this specific claim is restricted to areas where WhatsApp is usable, leaving out the whole of China, for instance.
Table 1 below summarizes the points that have emerged so far, the guiding principles they engendered, and how they materialized in terms of project design choices. Each point (except point 0, which is our baseline) will be discussed in the remainder of this section.

5.1 Hyper-collaboration

One essential goal of virALLanguages is to function as a bridge between two poles: at one end, researchers and professionals in the domain of global health – that is, “collaborative trans-national research and action for promoting health for all” (Beaglehole and Bonita 2010) – and, at the opposite end, communities speaking marginalized languages, especially but not only from the Global South. To refer to this kind of networking connecting hitherto unconnected parties we use the term “hyper-collaboration” here, taken from the world of corporate innovation where it is used to emphasize the power of putting together “less obvious partners” (Kolk et al. 2018: 23).

VirALLanguages was built on existing collaborative networks: at one end, the Community for Global Health Equity (CGHE) of the University at Buffalo and, at the other end, communities from the Cameroonian...
Grassfields. CGHE Co-Director Katarzyna Kordas and staff Lisa Vahapoglu and Jessica Scates contributed contents for the virALLanguages reference text (see Section 5.2.2), provided detailed feedback on multiple versions of the project protocol (see Section 5.2.3), and, later on, collaborated at the level of dissemination (e.g., in a talk story session at the seventh International Conference on Language Documentation and Conservation) and contributed at various levels to designing what a virALLanguages 2.0 should look like (see Supplementary Material S.1) including through funding a temporary research assistant position for a PhD student in epidemiology and environmental health in 2021 who prepared up-to-date contents about vaccines for a possible second phase. As for partnerships with communities, starting with Cameroon was relatively easy thanks to the extensive and longtime collaboration networks established in the country by project co-founders Pierpaolo Di Carlo and Jeff Good (http://kpaam-cam.org). During a brief three-week initial testing period, we received very helpful feedback allowing us to refine certain aspects of the project protocol (https://virallanguages.org/protocol/) and technical guidelines (https://virallanguages.org/help-sheets/). Indonesia and northern Pakistan are the other two main areas where previous work experience of project team members (Bradley McDonnell and Pierpaolo Di Carlo, respectively) facilitated finding proactive collaborators (see Section 6).

5.2 Messages should be accurate and memorable

5.2.1 Verbatim translations and free messages

CERC guidelines stress that the closer a message is to the audience’s communication habits and styles, the more likely it is accepted (see, e.g., Hyer and Covello 2005: Annex 1). We have already reviewed (Section 3) the motivations for producing content in the audience’s primary language regardless of the size of the speaker community. Here we deal with the fact that if such content comes in the form of verbatim translations from a major language—thus potentially reflecting an overall different communicative tradition—it may fail the requirement to be also stylistically and, more in general, culturally appropriate for the intended audience (see Section 2).

In this regard, virALLanguages adopted a flexible strategy: while the project protocol was aimed at providing information and assistance for the production of culturally appropriate (i.e., formally free) messages, the documents made available to collaborators also included a set of five scripts (https://virallanguages.org/video-scripts/), each covering one of the key topics in COVID-19 prevention and sufficient to produce a relatively short (1–3 min) video or audio message. We made scripts available thinking of teams that might find it hard to produce original messages on a novel topic such as COVID-19 in a short span of time.

Messages produced by translating scripts are not ideal from the perspective of the message’s cultural appropriateness but, like verbatim translations, do ensure that the information conveyed would mirror, at least to a large extent, information whose accuracy has been established. It must be kept in mind that stimulating nonspecialists to produce culturally appropriate messages brings about one major risk: the information conveyed might not reflect the accurate information that was provided. This risk remains essentially undressed until one has a way to double-check messages, which requires competence in the language. Since this could be done for only a handful of the target languages, in virALLanguages this key problem was addressed at two levels: (1) composition and training of language teams; and (2) adoption of a systematic workflow.

5.2.2 Autonomous teams, with checks and balances

Those directly involved in message creation in the virALLanguages project should form a language team: as a rule, language teams comprise no less than two people, one of whom acts as spokesperson and the other as judge or reviewer. This has two main goals: on the one hand, the presence of two people increases the likelihood that what the speaker actually says complies with the information provided by the project (see below); on the other hand, it also increases the likelihood that at least two components (e.g., age groups, families, etc.) of the community collaborate in creating the message.
Each team is given a reference text (https://virallanguages.org/reference-text/), written in a major language they understand well (currently available in English, French, and Indonesian), which contains a condensed overview of the most vital health information and practices for preventing infection and the spread of the coronavirus. Crucially, the reference text also includes more general information aimed at improving the team members’ overall health literacy: this was considered key to providing message creators with enough background to allow them to produce culturally salient, original, and accurate messages. For instance, such background information might allow one to find similarities between the current emergency and past events or existing proverbs, thus making it possible to mention them in the message. After reading the reference text, language teams were given a short quiz (10–12 questions, in the language of the reference text) to test their overall comprehension of the information. It was only at this point that they reached the stage of creating messages.

5.2.3 Project workflow

The initial consideration was that collaborative work needs to be regulated by a protocol that is systematic enough to produce reliable results (i.e., the ultimate delivery of accurate information) no matter the content/language, but not so rigorous or inflexible that it becomes too burdensome for any ordinary person to take on. This resulted, ultimately, in a multistep process which was repeated, refined, and adapted through the quick implementation of consistent collaborator feedback throughout the life of the project. In each case, it started with the creation of a local language team which was then assigned to a coordinator (i.e., one of the linguists or activists volunteering in the project) who was able to walk the team through each step of the workflow (Figure 2; http://virallanguages.org/protocol).

5.3 Trusted communicators

As we saw in Section 4, there are major reasons to consider trust in communicators to be central in the contexts that the project was targeting. Therefore, recommendations were given that the speaker be a well-known and

---

*Participation in the project was voluntary, as our very limited funds (donations by project directors) allowed only for the reimbursement of internet or phone charges that the team would incur for sending and receiving large files (see Section 5.4).*
respected community member, and this materialized in teams that included spokespersons such as traditional leaders, local artists, academics, students, or medical practitioners renowned in the community, as well as community members from the diaspora (for concrete examples, see Supplementary Material S.3).

5.4 Diversified content distribution

Based on information gathered from language teams about locale-specific uses of communication media, we created compressed versions of all the finished products (i.e., versions with small file sizes) and sent those back to the language teams directly via email or WhatsApp. We then published the full-sized files on YouTube and Facebook and made all the files (i.e., both the full-sized and the compressed versions of each message) accessible to download from the Internet Archive (https://virallanguages.org/output/). This way we were able to support person-to-person sharing, social media sharing, and freely accessible downloads for those looking for another way to access and share the content.7

---

7 Following recommendations from our language teams, messages were also broadcast by local radio stations (e.g., messages in Mafa [ISO 639-3 code: maf] on Radio Écho des Montagnes in northern Cameroon), but this happened less often than language teams would have hoped because of the very limited funds available.
6 Conclusions

In less than eight months, the virALLanguages initiative produced more than 90 resources in 47 marginalized languages of Cameroon, Indonesia, and Pakistan. Interested readers can find web traffic data in the supplementary materials. However, it should be noted that local forms of multimedia content sharing (i.e., via downloaded files rather than links) are not captured by web traffic data, precluding definitive data on the accessing of the videos. Moreover, mass testing campaigns have not been carried out in the communities targeted by virALLanguages. This is to say that we are unable to provide any reliable figures on the impact of virALLanguages. It must be said, however, that the reports we received were positive and, in some cases, explicitly connected community interest in the messages to the fact that they were presented in the local language by recognizable individuals that they personally trusted.

As Figure 3 shows, many of the languages targeted by virALLanguages are spoken in compact areas. In regions such as the Cameroonian Grassfields or the northern areas of Pakistan, there were so many languages targeted that it would not be too much of a leap to imagine that, for some time at least, the circulation of a number of videos in local languages may have caused a sort of “echo effect” (further amplified in environments where high rates of individual multilingualism in local languages are widespread across the population), thus increasing their reach through word of mouth.

This allows us to make another key point: What virALLanguages did would not have been possible without the help of key collaborators who acted as “regional coordinators” – intermediaries between the communities, on the one hand, and the (geographically dispersed) project team, on the other. These are linguists (like Margaret Neh Chenemo, Ndokobai Dadak, Yanti, and Ika Nurfayani), language activists (like Zubair Torwali), and locally well-known personalities (Achuo Christopher Ikom) who have worked in the domain of language documentation in the past in one way or another. It is highly likely that most documentary linguists who are foreign to the region where they do fieldwork rely on collaborators who know a certain area well, have multiple connections to the local populations, and are locally respected. In projects like virALLanguages, the role of such collaborators has been simply irreplaceable.

Our initial point was that work in CERC might legitimately (and usefully) be added to the inventory of the possible ways documentary linguists can “return linguistic favors” (Wolfram 1993: 227) to the communities they work with, and we hope that what we said so far has corroborated that point. We would now like to stress that the opposite, too, is true: documentary linguists’ professional training can benefit CERC. CERC guidelines are rooted in a number of assumptions that seem to stem from (economic) efficiency-based logics only, and we are not aware of any real debate among scholars (let alone with communities) aiming to clarify what these assumptions may determine in actual terms. At the very least, we believe that if such a debate can be started at all, documentary linguists (and the communities they work with) might have something to say.

We need to acknowledge that the COVID-19 pandemic is a particularly complex crisis to deal with at the level of public health communications. For one thing, the novelty of the virus caused experts’ advice to change rapidly and, at times, in conflicting ways. This further stresses the importance of (1) creating networks of hyper-collaboration (see Section 5.1) and (2) integrating their outputs with governmental directives, rather than relying on the latter only. As an example, the first version of the virALLanguages reference text (dated 15 April 2020) dealt with the different kinds of face masks even though their use was not (yet) recommended by the WHO. We now know how fundamental this measure is for containing the spread of the SARS-CoV-2 virus: had we relied on the WHO guidelines only, we would have contributed to the diffusion of misleading information for some time. For a campaign that can aspire to produce only one or two messages in total, due to the difficulties in reaching local volunteers, this is a threat to the “do no harm” basic mandate.

COVID-19 is just one of the possible crises in which risk communication campaigns might benefit from the collaboration of documentary linguists. It is not hard to imagine other concerns that may impact marginalized communities, be they related to public health or environmental risk. We would therefore recommend asking community members about possible needs in this regard as a way to start exploring the possibility of returning
“linguistic favors” through the circulation of accurate information in communities speaking marginalized languages.

References


Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/lingvan-2021-0021).