Using virtual reality for linguistic fieldwork

Gonzalez Gonzalez, P.; Wal, G.J. van der; Berruti, C.; Doorn, V.T. van; Morozova, I.; Raesfeld Meyer, J.M.M. von; Vorisek, T.J.

Using Virtual Reality for Linguistic Fieldwork

Paz González
Leiden University

Jenneke van der Wal
Leiden University

Claudia Berruti
Leiden University

V.T. van Doorn
Leiden University

Irina Morozova
Leiden University

Jai von Raesfeld Meyer
Leiden University

Thomas Vorisek[1]
Leiden University

Abstract: This paper reports on the use of Virtual Reality (VR) technology for linguistic data collection. Traditional verbal and 2D visual stimuli can be quite restricted in the context they provide, but thanks to VR technology, we can now get much closer to a full and natural context if we present speakers with a 360° vivid environment; one in which the linguistic factors to be studied are carefully controlled. We piloted VR technology for fieldwork by creating 360° videos, and tested these to study past tense in Spanish, and the interaction of focus and evidentiality in Xitsonga/Xichangana. We provide a detailed report of this proof-of-concept project, documenting all steps in the process.

Keywords: Virtual Reality, linguistic fieldwork, 360° video stimuli

1 Introduction

In this proof-of-concept project, we employ Virtual Reality (VR) technology in linguistic fieldwork. In contrast to video, where there is a clear separation between the stimuli (video) and the subject, VR places the subject in the stimulus context (video) space itself. The potentially revolutionary impact of VR on linguistic research was described by Peeters (2019: 898) as "a game-changing method for the language sciences". However, its applications in linguistic fieldwork are so far unexplored. Part of traditional linguistic fieldwork consists of native speakers producing, translating and judging sentences. Crucially, these elicited data (i.e., the produced sentences, translations and judgements) require a context to place the utterances or to control the acceptability or unacceptability of an utterance. This context can be provided verbally by the researcher ('imagine you see someone fixing a car'), or by visual stimuli (as in Figure 1 and 2), which have gained interest over the past decade (see for example Burton & Matthewson 2015 on storyboards, but also our earlier work in Van der Wal 2016; González & Kleinherenbrink 2021 containing visual stimuli; Fuchs & González 2022 analysing translations; and the online BaSIS methodology and González' Spanish database).

Providing such a context is already a step in the direction of more natural and ecologically valid data when compared to out-of-context translations or (most) psycho- and neurolinguistic experiments. Nevertheless, the stimuli we use now only provide part of the context: we can never fully control what the speaker may be imagining in addition to the given verbal or 2D stimuli. The starting point of this project is that we can get much closer to a full and natural context if we present speakers with a 360° environment: one in which the linguistic factors to

[1] González and van der Wal are project leaders with an equal share on the research and paper writing and therefore alphabetically ordered as first authors; Berruti, Morozova and von Raesfeld Meyer were involved in the preparation and data collection and von Raesfeld Meyer in the transcription, van Doorn in recording and editing the videos and Vorisek in selecting the video equipment and further technical assistance.

be studied are carefully controlled by the researcher. A controlled vivid context for linguistic experiments is therefore a desirable tool in the elicitation of natural data. This is now possible thanks to VR (and augmented reality) technology, and in recent years has also become more affordable and accessible for researchers.

Thanks to a Small Research Grant awarded by the Leiden University Centre for Digital Humanities (LUCDH), we have been able to pilot this method. We have created 360° videos and tested these with participants wearing a VR headset (goggles and controllers). Concretely, we have created scenarios to study past tense in Spanish, referred to as the 'Past Tense project', and focus and evidentiality (applied to Xitsonga/Xichangana[2]), referred to as the 'Focus project'.

Figure 1: Visual stimulus from the BaSIS methodology (Van der Wal 2021) to elicit a range of possible foci, for example: 'What is the man doing? (VP focus), 'Is the man fixing a bicycle?' (corrective object focus), 'What is happening here?' (thetic), etc.

Figure 2: Visual stimulus for Spanish aspect data collection (González & Kleinherenbrink 2021)

[2] Xitsonga is spoken primarily in South Africa, and Xichangana is the mutually intelligible sister language spoken in Mozambique.

For the Past Tense project, we used VR scenarios to see which verb form speakers of different varieties of Spanish use, comparing for example Argentinian with Castilian Spanish. In which scenarios do speakers use *caminé* 'I walked', and in which *he caminado* 'I have walked'? What are the essential contextual and linguistic factors in the choice between these past verb forms? These research questions have already been investigated with more traditional methodologies (see González et al. 2019, among others).

For the Focus project, we used VR scenarios to test focus (what is the new or contrastive part of the message?) and evidentiality (what is the source of the information?). Controlling each of these factors, we investigated the core meaning of three different forms of the present tense in Xitsonga/Xichangana. For example, in answer to the question 'Where do you work?', the answer 'I work in town' must be *Ndz-i-tirha ádórópéni* and not *Ndz-a-tirha ádórópéni* (Sitoe 2001), which would be felicitous in another context. The verb forms and their semantic-pragmatic aspects form a unique combination, and one that has not been investigated previously.

In this paper, we will not report on the actual research findings for these two studies, but rather share our experiences in testing the method. VR has recently been employed in language documentation (Pentangelo 2020 reports on documenting Kanien'kéha with 360° videos) and in psycho- and neurolinguistics (we refer to Peeters 2019 for an overview of studies using VR in the language sciences), and we think it has great potential for the elicitation of semantic and morphosyntactic data too. Testing the VR approach as a proof-of-concept means that all linguists interested in linguistic data collection methodology can learn from the initial challenges we met in this project. In the rest of the paper we therefore document the process throughout the project, from the motivations (Section 2) and methodological decisions (Section 3) to the technical details of the scenario creation (Section 4) and video creation (Section 5), as well as the data collection and elicitation period (Sections 6 and 7). Section 8 presents preliminary findings and Section 9 briefly concludes.

2 Motivation for this method

The use of augmented and/or virtual reality in linguistic research is not common practice (see Peeters 2019), and we are convinced that it will make it easier than ever for consultants to put themselves in imaginary contexts (carefully controlled by the linguist), and produce and judge sentences accordingly. This means that the resulting data, elicited by immersion in a natural context, more closely approximate language as it is used in everyday life, instead of perhaps more formal or careful speech as sometimes can be the case when being interviewed by a researcher. By being fully immersed in the setting, a participant can feel more likely to engage with the situation as if they were there, and thus producing speech that is closer to what they would use in a natural context (not in an interview/elicitation session). At the same time, linguistic data collection in this way is more controlled than spontaneous speech, such as in a recording of a conversation. While spontaneous data are an essential part of data collection, especially for underdescribed languages and for aspects of linguistics that concern interaction and larger units of language (discourse analysis, information structure, etc.), it is equally necessary to obtain negative data (what is *im*possible in the grammar of language L?) as well as judgements in particular conditions to test hypotheses on language rules and use (what is grammatical or most appropriate in this context?) — see Matthewson (2004). Those elicited data need a specific context with fixed parameters, which can be created more easily in 3D. In Peeters' (2019: 895) words, VR "is argued to be capable of combining high ecological validity with high experimental control". Therefore, the collected data are expected to be more reliable (though this prediction needs systematic testing).

All concepts that require a presence of the speaker in the environment will potentially benefit most from the immersive VR stimuli. For 360° videos, we can think of any topic related to deixis, ranging from demonstratives to spatial positions (see Nölle et al. 2020, Nölle &

Spranger 2022), but also the category of evidentiality: in languages where direct vs indirect evidence is relevant, and especially those where visual evidence is contrasted with other evidence, the virtual environment can provide specifically the one or the other type of evidence. Future applications will reveal which other areas and topics lend themselves to the application of this methodology. For more extended VR stimuli, where participants can interact with the virtual environment (that is, beyond 360° videos), all aspects of interactive language use can be tested — Peeters (2019) refers to an unpublished thesis relevant to this topic by Tromp (2018).

We see possibilities for the following three types of data elicitation. First, we can ask the speaker to simply describe what they see and hear (spontaneous speech). Second, the speakers can answer questions about what they perceive during and/or after the video. Third, speakers can be asked to judge statements during or after the video. We piloted all three types in the Past Tense project and the Focus project.

3 Decisions in methodology

The decision between using interactional VR or 360° videos was made very quickly: the difference in cost of creating each was substantial. This was partly due to the difference in price between the type of camera needed for each sort of video, and partly due to the fact that hiring developers to create an actual environment in which participants can interact with the virtual space was very expensive. As such interaction did not seem necessary for our purposes (although this would of course make the experience even more lifelike), we decided on the 360° videos option. Whilst acquiring VR equipment (360° camera and VR headset) is generally expensive, a professional set-up for 2D videos would not be significantly cheaper if one were to buy the equipment for the project alone. The difference is of course that most institutes already have this equipment and will provide it, causing the costs to be relatively low. We hope that with the advancement of VR technology it will be adopted by more institutes and will be more readily available for researchers to use, thus bringing the costs down. In comparison to other traditional elicitation material, we suspect that creating photographic or artistic 2D material is generally less expensive in terms of financial and time investment, but this too depends on the desired professional quality.

Further methodological choices depended on the individual subprojects, which we describe here in some detail so that the considerations can be understood against that background. Again, the goal is to share the methodology and not the results of the subprojects in terms of content.

3.1 Past Tense project[3]

The Past Tense project was conceptualised by Paz González and concerns past tenses in Spanish, both European and American, as these contain a lot of variation (Fuchs and González 2022, among others). The three main past tense forms in Spanish are: Present Perfect, Preterit and Imperfect. For the purposes of this study, we will only focus on the Present Perfect and the Preterit. The Present Perfect in (1) is mostly used in hodiernal contexts, where it expresses anteriority with respect to the present, and focuses on the result of the event. It is more common in European than in American Spanish (see González & Verkuyl 2017, and González et al. 2019, for a description of different varieties). The Present Perfect is also used in particular contexts (relevance and resultativity, among others). A sentence such as in (1) would not be

[3] This section is an adaptation of similar sections in González & Verkuyl (2017), González & Quintana Hernández (2018), and González & Diaubalick (2019). We follow the same theoretical framework so some parts are identical.

easily found in Latin American Spanish and would therefore sound marked (dialectologically speaking).[4]

(1) Has com-ido una manzana. [Spanish]
have.2SG eat-PPTCP DET.INDEF apple
'You have eaten an apple.'

The Preterit in (2) presents an event as a discrete whole at some specific moment in the past (perfective aspect). In hodiernal contexts such as *Comí hoy* ('I ate today'), it is not regularly used in European Spanish but fully accepted in Latin American Spanish (Rojo & Veiga 1999).

(2) Com-iste una manzana. [Spanish]
eat-2SG.PRET DET.INDEF apple
'I ate an apple.'

There were several methodological choices to be made for this part. We needed our participants to describe scenarios in the past. However, while wearing the headset, the participants mostly spoke in present tense. This is why we decided to allow them to describe in the present tense while watching the video, and afterwards ask them what they saw, considering two different time frames: yesterday and this morning, as they seem to trigger different past tenses in different Spanish varieties. Moreover, different scenarios with different aspectual value of the events (inherent aspectual information) were created. With these methodological decisions, we were able to collect the data we were aiming for (see Section 8 for preliminary results).

3.2 Focus project

The Focus project was conceptualised by Jenneke van der Wal and concerns forms of the present tense in the Bantu languages Xitsonga/Xichangana and Cicopi, both spoken in the south of Mozambique. Both languages have three present tense verb forms, and their exact meaning and use remain to be completely understood. Aspect is one factor, with a difference between a present progressive and a general or habitual present; and information structure is another factor. Sitoe (2001) names the forms as 'present conjoint', 'present disjoint' and 'present exclusive'. The terms 'conjoint' and 'disjoint' refer to an alternation that has been linked to information structure across eastern Bantu languages (see Van der Wal 2017 for an overview). We refer to the forms here as the zero form (conjoint), a-form (disjoint), and o-form (exclusive), to remain neutral as to their meaning and function. For Xichangana, Sitoe (2001) provides the following illustration: the zero form *ndzi-tirha* (4a) places focus on the element following the verb, whereas the a-form *ndza-tirha* is used when that element is not in focus (4b).[5]

[4] Abbreviations and symbols used in the paper: ! = downstep, A = a-form of present tense ('disjoint'), AUX = auxiliary, CONN = connective, DEF = definite, DET = determiner, DIM = diminutive, DJ = disjoint, ESS = European Spanish Speakers, EXCL = exclusive, F = feminine, FV = final vowel, IMPF = imperfect, INDEF = indefinite, INF = infinitive, LASS = Latin-American Spanish Speakers, LOC = locative, M = masculine, NEG = negative, O = o-form of present tense ('exclusive'), OM = object marker, PFV = perfective, POSS = possessive, PRET = preterite, PRO = pronoun, PROG = progressive, PRS = present , PPTCP = past participle, PTCP = present participle, RED = reduplication, REL = relative, SG = singular, SM = subject marker, VR = vitual reality. Numbers refer to noun classes unless followed by SG or PL, in which case they refer to persons. High tones are indicated by an acute accent; low tones are generally left unmarked.
[5] Note, however, that the alternation in Xitsonga/Xichangana is primarily determined by constituency (see Zerbian 2007), and indirectly by information structure.

(4) a. Ndzi-tirh-a á-dórópé-ni, a-ndzí-tírh-í [Xichangana, zero]
1SG.SM-work-FV LOC?-town-LOC NEG-1SG.SM-work-NEG
káyá.
home.LOC
'I work *in town*, not at home.'

b. Ndz-a-tirh-a á-dórópé-ni, a-ndzí-tláng-í. [Xichangana, a-form]
1SG.SM-DJ-work-FV LOC?-town-LOC NEG-1SG.SM-play-NEG
'I *work* in town, I don't play.'

(Sitoe 2001: 6, glosses adapted)

Sitoe (2001: 230) describes the o-form as "focus on a single action to the exclusion of any other action, or on a particular entity expressed by the verbal complement to the exclusion of any other possible complement", with the examples in (5).

(5) a. Ndz-ó-tirh-a, a-ndzí-tláng-í. [Xichangana, o-form]
1SG.SM-EXCL-work-FV NEG-1SG.SM-play-NEG
'I am just working, I am not playing.'

b. Ndz-ó-svék-á nyáma, a-ndzí-svék-í [Xichangana, o-form]
1SG.SM-EXCL-cook-FV meat NEG-1SG.SM-cook-NEG
ntsúmbúlá.
cassava
'I am just cooking meat, I am not cooking cassava.'

(Sitoe 2001: 230, glosses adapted)

Apart from aspect and information structure, there might be a third factor. In the neighbouring language Cicopi, the same three present tense forms are found, and in the data for Cicopi, we found hints that evidentiality might be a factor too (Nhantumbo & Van der Wal to appear). Speakers explained the difference between the a-form and the o-form in terms of whether the speaker is an eye-witness, as illustrated in (6).

(6) a. K-á-phínd-a mŏ:vha. [Cicopi, a-form]
17SM-DJ-pass-FV 3.car
'A car is passing by.' (You see it.)

b. K-ó-phˈínd-a mŏ:vha. [Cicopi, o-form]
17SM-PROG-pass-FV 3.car
'A car is passing by.' (Someone else tells you.)

(Nhantumbo & Van der Wal database)[6]

For both of these languages, we want to know what determines which of the three present tense verb forms is used. For the pilot with VR methodology, we kept aspect constant by only presenting currently ongoing events, and varied the visibility of the action and the focus, as described in Section 4.

For this project too, there were a number of methodological choices to be made. First, we wanted participants to answer questions at particular moments in the video. We decided to present the questions auditorily rather than in written form, partly because we did not want to distract the participant from the visible events in the virtual space, but also because this makes

[6] The database will be archived through The Language Archive; until then, access to this online database can be given by the second author.

participation possible for people who are less literate. Second, we decided that the questions would be posed in the lingua franca, being English for Xitsonga and Portuguese for Xichangana. This would not prime the participants' responses for one or the other verb form, which would have been problematic if the question had been posed in Xitsonga/Xichangana.

We now turn to the detailed decisions in methodology regarding the creation of the scenarios (Section 4), the creation of the videos (Section 5) and the data elicitation (Section 6 and 7). A flowchart of our work process can be found in Appendix D.

4 Creating the scenarios

The first step in using VR for fieldwork is to devise the scenarios for the videos. Some decisions and considerations for the scenarios applied to both projects, and others only to one of the two. We first describe the general aspects, and then those relevant to the individual projects. It will be useful for the reader to have an impression of the scenarios we created to obtain the data for our research questions, which we therefore verbally describe and explain in this section. Stills from the videos are provided in Appendix A, and the videos themselves are freely available via https://video.leidenuniv.nl/playlist/dedicated/1_5d9xqz5l/.

The first consideration is that it was best to record the videos indoors. There were three main reasons for this choice. Firstly, outdoors the weather is unpredictable; secondly, it would be a big challenge to control the background and surroundings; and thirdly, there is less control over the light and sound. A further consideration was that random passers-by would possibly be captured. As a result, the events to be filmed had to also represent indoor scenarios.

A second point is that we tried to conceptualise the videos in such a way that they could potentially be used in different projects worldwide. Therefore, we aimed to create scenarios that would not be limited to one culture. Given many considerations involved (some more practical, others more conceptual, as this article discusses) we had to set priorities and make choices. Nevertheless, we restrained from using some obviously 'culture-specific' things, for example choosing rice with beans over sandwiches. Also, the activities that we have chosen to be represented in the videos are generic rather than culture specific (sawing, playing the guitar, washing dishes), hopefully recognisable and relatable for the majority of the potential audience. Furthermore, the actors in the videos were students coming from different ethnic backgrounds. Nevertheless, we suspect that using videos set in an environment that is as natural as possible for the speakers of each particular language will enhance the chances of success.

For all scenarios, we had to take into account the fact that the camera does not move (see further in Section 5), which meant that all scripted activities had to be visible from one viewpoint. This also had consequences for the location of filming, as all activities needed to be physically as well as logically/culturally possible. These imposed certain limits on the types of actions chosen. We opted to include multiple events within one video, rather than presenting participants with separate videos for separate actions, as this reduced the time needed for viewing and prevented a lot of hassle with editing, as well as stopping and starting the videos during the viewing.

For the Past Tense project, we needed eventualities that trigger a particular past inflected form. To elicit the present perfect, contexts of hodiernality/prehodiernality were created: first at a cafeteria for 'yesterday' and second in a classroom for 'today'. The same actors appeared in both contexts. In the cafeteria (first video) scene we see five activities (see also the video stills in Appendix A):

- Two boys discussing their holidays at a table with maps and pictures;
- A girl studying very hard at another table, with study books, taking notes very extensively;
- A girl walking in, wearing sports clothing, she fills her water and then runs off;
- A barista squeezing orange juice;

- A girl getting a cup of freshly made orange juice, spilling it on her shirt and trying to clean it with the help of the barista.

In the classroom (second video), the results of these actions are perceivable: the same people come into the classroom for an exam; a lecturer is also present in the room. We see the following:

- The two boys appear grumpy and unhappy since they did not study for the exam;
- The girl who was studying is happy and doing well;
- The sporty girl comes in on crutches, as she probably hurt herself the day before;
- The barista is simply writing the exam;
- The girl with a stained shirt notices the stain and then tries to hide it.

The idea was that the participants would see relevance relations between the two videos (yesterday running > today crutches; yesterday spilling of the juice > today a stain on t-shirt; yesterday studying hard > today happy with the test, and so forth). Moreover, some activities were durative (running, studying, squeezing juice) and others terminative (spilling of the juice), to control for different aspectual meanings. These videos did not contain sound.

The main challenge in this project was to make apparent to the participants that the activities took place in the past. A first solution was to show a rewind of the actions at the beginning, to visualise 'travelling to the past'. In a first run-through, in which we could see the whole video played backwards at a higher speed, this turned out to not be clear at all. A second solution was therefore adopted, which added a black screen with white words *ayer* 'yesterday' and *esta mañana* 'this morning' to inform the participants about the temporal situation of the events at the start of the fragments, and then showing the scenarios to the participant. While watching the fragments they described what they were seeing. After the viewing experience, the participant was asked to relate to us what happened in the two fragments, drawing attention to the 'yesterday' and 'this morning' timeframes. The instruction was given without using any past tense form, to avoid interference with their answers (see further Section 5.3 on editing the videos).

For the Focus project, we wanted to check visual vs non-visual evidence as a first factor, and focus on the verb or focus the object as a second factor. For the first factor, we needed activities that could be perceived visually as well as non-visually, and for the second factor, we needed transitive actions. In order to provide non-visual evidence of an event, each of the activities should have a recognisable sound to satisfy the non-visual evidence condition. With these requirements in mind, three activities were chosen: sawing a plank, playing the guitar, and washing dishes. We used a space which had a door to a side room, so that part of the activities would happen in the side room, invisible to the camera/participant. Obtaining the right focus on the verb or the object was achieved by asking a question targeting the verb ('what are they doing with the cup?') or the object ('what are they washing?'). As we were targeting the present tense in this project, we needed to ask the object/verb questions directly at the point when the action was happening.

Two scenarios were used. In scenario and video 1, we see person A eating rice and beans, and we hear the sound of sawing coming from the side room. The sawing stops, and out of the side room comes person B with a plank and saw. Person C crosses the room and goes into the side room. We hear the sound of a guitar (but cannot see it). Person A finishes eating and walks over to the sink to wash their plate.

In scenario and video 2, we are in the side room, and see person A eating and person B playing the guitar. Person A finishes eating, leaves the room and we hear the sound of washing up (but cannot see it). Person C comes into the room and we see them sawing a plank.

Each of the actions was thus presented once as direct eyewitness and once without direct visual evidence, and each of the actions was questioned to elicit focus either on the object or

on the verb. Table 1 presents an overview of the conditions for both videos. As mentioned, the aspect (progressive vs habitual) is kept constant.

Table 1: Overview of conditions and questions per video (Focus project)

	Question	Focus	Evidence
Video 1	What is s/he playing?	object	non-visual
	What is s/he doing with the plank?	predicate	non-visual
	Is s/he washing a cup?	object (contrast)	visual
Video 2	What is s/he doing with the guitar?	predicate	visual
	What is s/he sawing?	object	visual
	What is s/he washing?	object (new information)	non-visual

With these scenarios, we set out to do the recording and subsequent editing of the videos.

5 Creating the videos

We created three videos in total, one for the Past Tense project of nearly five minutes and two videos for the Focus project of just over two minutes each. We also recorded a try-out video at a random point in a university building, which later turned out to be useful in making the participants accustomed to the headset and distracting them between target videos. We highlight several points of attention with respect to the equipment, the actual recording, and the editing.

5.1 Selecting the equipment

In terms of hardware, we needed a 360° camera, SD cards, a monopod, and a VR headset (see overview in Table 2 at the end of the section).

For the camera, we took into consideration the following aspects, apart from the budget:

- The resolution should minimally be at 5.7K (5760*2880 pixels). Going over the 4k resolution might seem extreme but this is advisable since the image is stretched out in 360°.

- The light quality should be suitable for indoors recording as well as outdoors. There is a risk of underexposure (too little light) for a camera with a small sensor. The image should equally not be ‘noisy’ when filming indoors when all the lights are on.

- The camera should be user friendly, since it will be researchers or student assistants (and not professional camera people) operating the camera. It should be clear where the settings are and how to adjust them, as well as what is being filmed (a viewfinder). VR cameras usually do not have a viewfinder themselves but accompanying software could provide this (e.g., the viewfinder appears on a smartphone).

- The battery should be able to record 1.5–2hr minimum (although this is of course dependent on how much one wants to do in a day).

- The camera should be compatible with different software. If the accompanying software does not work properly, then it should be flexible and at least work with Adobe suite. This is to avoid frustration for the editor.

- The camera should be sturdy: it is a small camera so it can easily slip from the hand. In addition, the camera will have convex lenses, which will more easily scratch if handled in a less than careful way. Protector caps for the lenses are available. However, these might introduce blur and/or loss of light due to the extra filter.

- Audio is often an afterthought, so the audio being of at least reasonable quality is important. 'Prosumer' VR cameras usually have average (stereo) audio quality. 360° audio can further enhance the natural context of the subject and the stimulus. In our project, this unfortunately needed to be cut, both for budget reasons and since editing 360° audio is quite intensive work and not all VR headsets have support for it.

With these specifications, we opted for the Insta360 ONE X2. Note that this camera can record for a maximum of 30 minutes consecutively, then needs a break before continuing. The camera is quite small (it fits in one's hand) and therefore easy to travel with on fieldwork.

Figure 3: The camera set-up of the Insta360 on the monopod

For the SD cards, it is important that they are quick enough to capture video data without delay. For the Insta360, we used Class 10 | A2 | U3 | V30 | UHS-I, for which 256G gives 4.5hrs recording time with maximal resolution.

For the monopod, there is a trade-off between stability of the camera and the (in)visibility of the legs. The legs are not easy to edit out of the video, and the less visible they are, the more immersive the experience will be for the participant. However, smaller legs also means less stability. Note that there is a difference between a monopod and tripod: The monopod specifically only shows its 'feet' and not legs. Hence it is useful for 360° video recordings,

where the viewer in the end result only sees a black dot when looking straight down (see Figure 4).

Figure 4: The black dot from the monopod when looking down in the VR environment

For the headset, we first used the Oculus Rift S and later the Oculus Quest 2 (now rebranded as the Meta Quest 2). Oculus is currently the best supported headset, hence the choice for this brand. The Rift S is not wireless, which brought difficulties in the ease of moving and looking around (see further under Section 6); we therefore switched to the wireless Quest 2, which is the most recent model at the point of the study. Because of its portability and ease of use, this is very suitable to be taken to the field as well. The market for VR headsets is changing rapidly, because of this it is always advisory to look into which headsets are available and are still being supported.

In order for the researchers to see what the participant is seeing, a link is needed between a device (like a laptop) and the headset. Once linked, the Oculus computer app has a programme called mirror.exe, which shows everything the headset projects. The link can be established in two ways: either through a cable, which is not practical as it reduces mobility and is somewhat expensive for the required type of cable, or through the Oculus Air Link, via a Wi-Fi network that both devices need to be connected to. It turned out that where we were, the network of Leiden University is not strong enough to facilitate this link, but most private Wi-Fi networks should be fine (see Section 6 for a reflection on how this may work in the field). A solution for this problem is to cast from the Quest 2 to another device, this is a feature supported by all Oculus applications. This can be done to a mobile device which has the Oculus mobile app installed or to a computer with an internet browser.[7,8] The user now has an option to cast either to the Oculus app or through a web link with a code.[9] The mirroring/casting function may be helpful because it allows the researcher to follow where the participant is and thus to ask questions, nudge the participant to look somewhere else if needed, and give extra stimuli etc. at specific points in the video.

[7] When you have clicked on an application installed on the Oculus and it has opened, you can press the Oculus button on the right-hand controller and a pop-up screen will appear and display some options including a 'cast' option.

[8] Although an internet connection is also needed for casting, it does not need to be as strong as the Air Link connection requires, making it the preferred method when in an environment with a weaker, but still available, Wi-Fi network.

[9] For a more detailed explanation see the Meta site: https://www.meta.com/help/quest/articles/in-vr-experiences/oculus-features/cast-with-quest-2/.

Table 2: Overview of hardware and software used in our pilot

Hardware	Software
Insta360 one X2 camera	Insta360 mobile app
Sirui Monopod P-325FL	-
Sandisk Extreme miniSD V30 256GB	-
Meta Quest 2 VR headset	Meta client
Video editing PC	Adobe Creative Cloud (video editing) Insta360 Studio

5.2 Recording the videos

Having selected the right environment in preparing the scenarios (e.g., with a side room), it was time to set up the camera. Most modern consumer-grade 360° cameras are very easy to operate. A basic knowledge on how to operate a simple camera should be enough for anyone to use the camera. It is important for the camera person to easily manoeuvre in a digital environment, whether this be the camera, the VR headset, editing software, etc.

The battery of our camera was sufficient for two hours; nevertheless, it is advisable to recharge the camera after recording, if possible, especially if multiple recordings are planned on the same day.

As mentioned before, the camera best remains static to avoid confusion for the viewer, as it may destroy the immersion feeling if the camera suddenly moves. When placing the camera, care has to be taken that everything that needs to be visible is indeed visible (e.g., the actor eating with their right hand meant that the food was less visible), and what should not be visible is out of sight of the camera (e.g., letting the blinds down so that events outside are not distracting). It is important that the camera is at the right height: participants were seated in our set-up (see Section 6), whereas the camera was at about 1.5 metres during the recording. One participant therefore indicated having the feeling of hovering above the situation. If the video contains multiple scenes that have been filmed separately, you will want to have the camera at the same height to avoid disorientation. Ideally, each scene/scenario is filmed in one take, or otherwise it may again be disorienting for the viewer, having to readjust their place in virtual space.

Using the Insta360 X2, we recommend that the person doing the recording download the Insta360 app on a smartphone. The app can directly connect the phone to the camera using a Wi-Fi signal emitted by the camera. This allows the camera to be operated at a distance, and the camera person (who should not be present on the actual film set lest they also be caught on film) can see live what is being recorded. With the app, the recordings can also be reviewed on the mobile phone.

Directly after recording, it is recommended to have a very close look at the footage to see if the recordings fulfil all the requirements. It is better to have a couple of takes too many, rather than too few. In our case, we did not check immediately, which meant we had to ask the actors to come back another day for a retake of one video. Because the videos were very short (2–4 minutes), it only took us around 30–40 minutes to rehearse and do multiple takes of the scenes.

5.3 Editing the videos

The Past Tense videos were edited several times. In order to make clear that the events in the video happened in the past, as mentioned in Section 4, we first presented the events in rewind to create a sense of travelling to the past. However, this idea of 'rewinding' time was not sufficiently clear to the participants. The second adaptation was to add a black screen with white words ('yesterday', 'this morning') to inform the participants about the temporal situation of the events. The second part of the video was also shortened, as it was clear that there were not enough actions happening to justify the original length. A balance needs to be found between the extremes of overstimulation (too much happening in the video to keep track) and boredom of the participants – the right balance here will be dependent on what the intended participants are accustomed to, on average. In our experience, the participants indicated that they felt like they were in the virtual environment almost instantly, so longer videos would not be required or contribute significantly to the immersive experience. Apart from the earlier mentioned battery life, there is no technical restriction on the length of the video, but the videos being short made it not only easier to edit, but also significantly easier to record in a relatively short amount of time.

For the Focus project, we wanted participants to answer questions at particular moments in the video, since we wanted them to use the present tense. As mentioned, the questions were presented auditorily rather than in written form. We therefore recorded the questions, and then edited the wave-files into the video at the right point in time. This also allowed us to create an English and a Portuguese version of the video. We discovered that when embedding the questions into the video, care should be taken to provide a considerable amount of time between questions, allowing the participant to give an answer. This implies that if the video consists of several actions, these should be spread out over the video, as it is problematic to give a response while also trying to focus on the next event in the video. Sufficient time is therefore needed to avoid a dual-task effect. Furthermore, even though elicitation targets particular sentences and structures, responses are rarely given in just one sentence, thus requiring more time. While more time for the response can potentially be gained by pausing the video, this would require additional effort from the participant (for example, figuring out/recalling how to pause the video, even realising that the video should be paused) instead of focusing on the actual response.

One of the problematic issues with the raw recordings was the sound; especially when the sound came from another room (like the indirect evidence for the guitar playing or the sawing), it was not always picked up by the camera sufficiently clearly. We solved this by recording the sounds separately (holding a recorder next to the dish washing or sawing) and adding it to the video afterwards.

As for technical details, editing the videos is best done using the raw images from the SD card. This retains the highest possible image quality. Adobe Premiere Pro by itself was not enough to cut and edit, as the video file format (.insv) was not supported. The solution here is to use the Insta360 Studio plugin for Adobe Premiere Pro. This is an application functioning as a plugin for Adobe Premiere Pro and Final Cut Pro X. The Insta360 Studio can also be opened on its own to watch the recordings and make simple edits (e.g., to export documents as mp4 or trim the beginning and end of the video). However, it is still advisable to do the actual editing in Premiere Pro because the Insta360 Studio app has very limited options when it comes to editing. The editing process itself, once everything was set up, went very smoothly and is not unlike the editing of regular videos. This, not including revisions and render time, took around 30 minutes per video.

6 Preparing the data elicitation

There are a couple of practical things to consider before running the data elicitation, first for the physical environment, and second for the headset.

Placing participants in a virtual space means that they must also have the physical space, for example to move their arms and to turn around. We organised the viewing session in a (class)room, clearing a space of about 3x3 metres, with a rotating chair. The rotating chair provides the participant with a stable base, yet allowing them to look around in the virtual space. The room was quiet, so that there were no noises distracting the participant, and the recorder would only and clearly pick up the participant's speech.

With regard to the headset, we discovered that not all frames of eyeglasses will fit under the headset. Taking off the glasses is typically not an option for the participant, as they need to see everything clearly, so this factor should be taken into account when selecting the participants and the headset.

Using the Oculus Quest 2 headset, videos can be played with the standard app that comes with it, called Oculus TV. However, this turned out to have an autoplay function that cannot be switched off. This means that the next video will automatically start after the first has finished. Autoplay is very impractical, as it requires extra unnecessary effort from the side of the participant: since they see the video for the first time and do not know what to expect, moving on to the next target video without a pause is not only confusing but creates confounding factors. The solution is to use the app Skybox VR, in which autoplay can be switched off. Moreover, it allows the videos to pause and restart automatically once the headset is put on or taken off. The researcher can pick the right video themselves, start playing it and pass the headset to the participant. Once the headset is taken off, the video pauses, and as soon as the participant puts it on, the video continues. This, in turn, allows for a very smooth procedure. Moreover, the participant in this case does not need to hold the controllers or learn how to select and start the video. Note that this functionality also has implications for the recording and editing of the videos: the action in the video should not start straight away, since automatic pausing and restart risks taking a couple of seconds from the original video. This time interval should be taken into consideration.

Figure 5: Enough physical space and a rotating chair worked well

7 During the data elicitation

For both projects, after the participant was welcomed (including explanation of the procedure and signing the informed consent form),[10] they tried on the headset and with the help of the researcher adjusted it to fit comfortably. It is important to first have a little practice trial before starting the actual target video, because for many it might be their first experience with VR. Firstly, a practice trial ensures that they understand the concept (e.g., know that they can look around in all directions). Practising also allows the participant to get over the first emotions after experiencing VR (be it excitement, fear, etc.) and be able to perceive the target video as naturally as possible. Secondly, since the whole set up of the headset and the virtual space might be unsettling at the beginning, the trial allows the researcher to make sure that the participant is not confused during the actual task. Moreover, trying out the expected procedure ensures that the speaker is able to participate — therefore, the practice trial should be comparable to the target one, for example using sound and using questions if those are present in the target videos. The time needed for practice might be dependent on the age and prior experiences of the participant.

To capture the descriptions and answers that the participants produced, we first used the headset's built-in microphone. This did not result in sufficient quality, so we recommend using a separate recorder (in our case, the standard Zoom H5). This worked well, but had another confound: since the participant looks around, they will sometimes be facing away from the recorder resulting in poorer quality, or at least in variable clarity. Possible solutions would be taping the recorder to the chair or using a wireless clip-on microphone.

As already mentioned, a wireless headset is to be preferred over a 'wired' headset, because during the viewing, the participant is expected to look around, and if the headset is connected by wires, the movement will require the researcher to move around with the participant, holding the wires. We also noted that younger participants will spontaneously look around more than older participants: while the younger participants will find out on their own that they can see in 360°, the older participants seem to have a tendency of sitting still and watching in one direction, this way sometimes missing parts of the action in the videos. Possible solutions would be to remind participants during the practice trial to look in all directions, and to have the video start in the direction where the initial action happens (that is, if there is only one action, which is not the case in the recordings for our projects, where various actions happen simultaneously).

Taking off the headset between different parts of the data elicitation (e.g., different videos, different tasks) may have benefits and drawbacks. On the one hand it will make for a clear break between the tasks, allowing the participant to refocus afresh; and it allows the researcher to select and start the next video (but see the remarks on autoplay in Section 6). On the other hand, it will take the participant out of the virtual environment and requires readjusting the headset comfortably every time. The choice depends on the goals of the study but is certainly something that should be taken into consideration.

For the Past Tense project, the data elicitation consisted of four parts, with parts two to four being recorded.

1. The first part allowed the participants to get accustomed to the headset. They were asked to sit down and experience how it felt to have the headset on. The 360° fragment shows a university building which was well known to most of the participants. The participants were asked to state when they were ready for the target video.

[10] While in some way this type of data elicitation involves the same ethical aspects as traditional data elicitation (concerning well-being, data protection etc.), the use of VR may warrant a different ethical procedure, see for example Spiegel (2018) for potential risks in other applications of VR.

2. In part two, before putting the headset on, the participants were instructed to describe what they were going to see in the VR video. They were also presented with a document containing pictures of all the actors and their (fake) names so that they could refer to the actors by name. Once the participants put the headset on to watch the video, they started describing what they were seeing. Their descriptions were recorded. Once the video was finished, they could remove the headset, with the help of the researcher.

3. Part three started with a small instruction from the researcher. The participants were reminded that the first part of the video was yesterday and the second part was this morning. This was done to elicit past tense forms. The participants were then asked to renarrate what they had seen.

4. The fourth and last part consisted of a reflection and debriefing. The participants were asked to share their thoughts on the VR experience.

For the Focus project, there were five parts to the data elicitation, with parts two to five being recorded.

1. The first part was the trial, allowing participants to get used to the headset (see Section 6). We used the first part of the video for the Past Tense project for this part.

2. The second part was one of the target videos, for which participants were instructed that they would hear a question in English/Portuguese, and that they were to answer the question directly, in Xitsonga/Xichangana. After this task, they were asked some follow-up questions about the answers they gave, primarily for reasons of transcription, but also judgements about alternative ways of phrasing. For practical reasons (ease of removing headset for one but not the other participant), these questions were asked while still wearing the headset for one participant and without the headset for the other participant.

3. The third part was an intermezzo, consisting of a video of random activity in a university building, where participants were asked to describe what they saw. This functioned as a distractor between the two target videos, so that participants would not directly hear similar questions for similar situations.

4. The fourth part was the second of the target videos, with the same instructions. This had a more extensive follow-up about alternative ways of answering the questions and transcription (without the headset).

5. The fifth part consisted of a reflection on the experience and debriefing.

The Focus project was carried out with only two participants: one adult speaker of Xitsonga and one elderly speaker of Xichangana. We have tried to recruit speakers of Cicopi, as the evidence for a potential link with evidentiality came from this language, but the contact we had with one Cicopi speaker in the BeNeLux (through Facebook) unfortunately suddenly ended. For the Spanish study, the participants were 10 native speakers of various dialects of Spanish between the ages of 30 and 55 (Andean region, Chilean, European Spanish, Rio Plata region) and two L2 learners of Spanish with Dutch as L1, both in their twenties.

As the data were recorded in the same way as for 2D stimuli, for us there were no differences in the data management and processing for these projects. However, we can imagine that some

adjustments may be necessary if the video and the audio and the transcription data are combined, as for example in the data management software ELAN, considering that the videos do not only have a temporal dimension, but also a spatial dimension. We have not explored this further.

8 Preliminary results

A preliminary look at the Spanish data shows that VR technology is indeed successful in eliciting (variation in) past tenses (see an impression in Appendix B). The research question for this project was 'What are the essential contextual and linguistic factors in the choice between three Spanish past tense forms?'. We had to discard the data of one participant because she did not understand the instructions and instead of 'this morning' she constructed sentences with 'the next day', which influenced her choice of past tense forms. The other nine participants were three European Spanish speakers (ESS) and six Latin American Spanish speakers (LASS). The raw numbers indicate that the Perfect was used 29 times by ESS and five times by LASS. On the other hand, the Preterit was used 29 times by ESS and 105 times by LASS. Taking into account the difference in participant numbers per group, the Spaniards use the Perfect in 50% of their past tense representations, and the Latin Americans not even 5% — a convincing difference.

As for the Focus project, where we wanted to gauge the factors determining the choice between three present tense verb forms, a first look at the data suggests that the zero form can be used in practically each context (provided the verb is not clause-final), that is, with focus on the object or the predicate, and with direct or indirect visual evidence. In contrast, the a-form is restricted to predicate focus contexts. The use of the o-form seems to differ between Xitsonga and Xichangana, and visual evidence does not seem to be a factor for the use of the three verb forms. Because of the visual and non-visual evidence as presented in the videos, we did encounter alternative ways to indicate indirect evidence, for example a speaker indicating 'I don't see her, but I think she is washing a plate', or by means of a third person plural construction ('they are playing a song'), which is used crosslinguistically for a functional passive/impersonal construction. Needless to say, further investigation with more than two speakers is needed (into both variants Xitsonga and Xichangana), but given that differences in use of these present tense forms are intricate and hard to pinpoint, with the help of VR we see some first preliminary evidence and a possibility to disentangle this puzzle.

9 Evaluation

Having described the complete procedure, in addition to the experiences and advice already shared, in this section we close with some general reflections on our proof-of-concept project.

Considering that the recording, editing, and viewing of the videos in the VR headset requires electricity, this methodology is not suitable for (fieldwork) environments where there is no, weak, or unreliable electricity supply. We also mentioned using a Wi-Fi connection at various points in the process; while this is now possible through simcard hotspots/MiFi in an increasing number of places on earth, we want to stress that it is not required. During the recording, the connection with a smartphone can be made through a signal that is emitted by the camera, and the recording can alternatively be done by simply starting the video recorder, walking out of the space, and editing the first and last seconds out of the video. Loading the videos onto the headset can be done through a cable, and during the viewing, Wi-Fi can be used for casting what the participants see to a phone, tablet, or laptop, but this is also possible via a cable connection.

In general, participants feel at ease, and only one participant out of the 14 mentioned feeling uncomfortable with the headset. Most importantly, the participants indicated that the experience felt like they were present in the situation, so the main objective of using VR stimuli is obtained.

A quote from one of the Spanish speaking participants (own translation): 'I loved the experience; it is fascinating, the idea that you more or less enter their space'. This is a significant advantage that 360° videos have over regular fixed videos: it allows the participant to be present in an environment. This way the entire video will feel more like an actual experience to the participants rather than passive observation of a 2D fixed video. That participants felt part of the experience can also be seen in descriptions of the scenes as being relative to themselves, for example in Appendix C 'it looks like he is looking at me'. Being in the 3D environment also allows the participants to focus more on things they find interesting and that attract their attention, and this way they will feel like more active viewers or even participants in the events.

We found that the videos work naturally for descriptions of the present tense, although past tense can also be effectively elicited by retelling what was seen after the viewing. Furthermore, we discovered in the descriptions of the intermezzo/filler video that this was extremely useful in eliciting thetic sentences, i.e., presentational 'out of the blue' sentences such as 'there is a man walking towards the door' (see Appendix C).

Some participants are better able to perceive details (such as the stained shirt) whereas others mainly take in the big line of the story (such as the students taking an exam). Some of them talked the whole time, while others did not have much to say (and specifically said so). Not all scenarios were therefore read as we expected and intended. One learning point here is that, just like with theatre shows, the more obvious you make the action or expression, the bigger the chance it will get noticed. Especially facial expressions are easily missed, so slightly exaggerating them (but not beyond what is recognisable as 'natural') and taking the time to express them will increase the chances of recognition (for example, the students' reactions while taking the exam).

Within the limits of this pilot, we unfortunately cannot offer a comparison of VR with other (traditional) methodologies — the aim here was merely to test how VR technology can be implemented into linguistic data-collection and whether useful data can be collected by using VR stimuli. Looking directly at whether and how the proposed methodology outperforms (or not) other stimuli types, including 2D videos, was beyond the scope of our project (though this is one of the logical follow-up studies). Nevertheless, as the comments of our participants during the reflection sessions show, VR does seem to fulfil its task in making the experience realistic. It is also the case that the data collected as an outcome of using the technology gives a valuable insight into the context-dependent phenomena in question. Furthermore, the replicability of this methodology is high. Our impression, based on the amount of data collected, is that participants spoke more freely, and in the Spanish project, given the short time used for the collection, a larger amount of more targeted production data was elicited than with other types of elicitation tasks. These remain impressions, which need systematic testing in further research.

As mentioned, we expect the 360° immersive aspect to be a valuable tool in research into evidentiality as well as deixis and spatial aspects, where the position of the participant in the environment is an essential part of the meaning (see also the research by Nölle et al. 2020 and Nölle & Spranger 2022, using VR to investigate spatial reference strategies). A next step will be to use not just 360° videos, but actual VR environments, where participants can interact with the virtual environment. This will be a game-changer in the investigation of multimodal communication and interactional studies, where the role of speaker and hearer are crucial.

In conclusion, the innovative methodology of using 360° VR scenarios for fieldwork looks very promising. Participants enjoy the experience, and a first look at the collected data shows that we managed to elicit the type of data we were looking for.

Acknowledgements

We are grateful to the Leiden University Centre for Digital Humanities for awarding us the Small Grant that allowed us to carry out this pilot research. We also thank Maarten Mous for

planting the idea of VR for fieldwork, Sanne Arens for bringing us together, and Rob Goedemans of LUCDH for his advice in this project. We thank Ahmed Sosal, Daen van Zuijlen, Arthur Swarts, and Sophie Sierig for their participation as actors; as well as all the participants in the two projects. Finally, we thank two anonymous reviewers and Vera Hohaus as editor for their helpful comments on previous versions of this paper. All content and any errors are the responsibility of the authors alone.

References

Burton, Strang, and Lisa Matthewson. 2015. Targeting construction storyboards in semantic fieldwork. In M. R. Bochnak and L. Matthewson (eds.), *Methodologies in semantic fieldwork*, 135-156. Oxford: Oxford University Press.

Fuchs, Martín and Paz González. 2022. Perfect-Perfective variation across Spanish dialects: A parallel-corpus study. *Languages* 7(3). 166.

González, Paz and Henk J. Verkuyl. 2017. A binary approach to Spanish tense and aspect: On the tense battle about the past. *Borealis—An International Journal of Hispanic Linguistics* 6. 97–138.

González, Paz and Lucía Quintana Hernández. 2018. Inherent aspect and L1 transfer in the acquisition of Spanish as a Second language. *Modern Language Journal* 102(4). 611-625.

González, Paz and Tim Diaubalick. 2019. Task and L1 effects: Dutch students acquiring the Spanish past tenses. *Dutch Journal of Applied Linguistics* (DuJAL) 8(1). 24–40.

González, Paz, Margarita Jara Yupanqui, and Carmen Kleinherenbrink. 2019. The microvariation of the Spanish Perfect in three varieties. *Isogloss* 4. 115–33.

González, Paz, and Carmen Kleinherenbrink. 2021. Target variation as a contributing factor in TAML2 production. *Círculo de Lingüística Aplicada a la Comunicación* 87. 39–51.

Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4). 369–415.

Nhantumbo, Nelsa J. and Jenneke van der Wal. To appear. The expression of information structure in Cicopi. In J. van der Wal (ed.), The expression of information structure in Bantu. Berlin: Language Science Press.

Nölle, Jonas, Simon Kirby, Jennifer Culbertson and Kenny Smith. 2020. Does environment shape spatial language? A virtual reality approach. In A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd and T. Verhoef (eds.), *The evolution of language: Proceedings of the 13th international conference (Evolang XIII)*, 321–323. Nijmegen: Max Planck Institute for Psycholinguistics.

Nölle, Jonas and Michael Spranger. 2022. From the field into the lab: causal approaches to the evolution of spatial language. *Linguistics Vanguard* 8(1). 191-203. https://doi.org/10.1515/lingvan-2020-0007

Peeters, David. 2019. Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review* 26 (3). 894–900.

Pentangelo, Joseph. 2020. 360º video and language documentation: towards a corpus of Kanien'kéha (Mohawk). PhD dissertation, CUNY.

Rojo, Guillermo and Alexandre Veiga. 1999. El tiempo verbal. Los tiempos simples. In I. Bosque and V. Demonte (eds.), *Gramática descriptiva de la lengua española* 2. 2.867-2.935. Madrid: Espasa-Calpe.

Sitoe, Bento. 2001. *Verbs of motion in Changana*. Leiden: CNWS Publications.

Spiegel, James S. 2018. The ethics of virtual reality technology: Social hazards and public policy recommendations. *Science and Engineering Ethics* 24, 1537–1550. https://doi.org/10.1007/s11948-017-9979-y

Tromp, Johanne. 2018. Indirect speech comprehension in different contexts. Unpublished doctoral dissertation. Nijmegen: Max Planck Institute for Psycholinguistics.

Van der Wal, Jenneke. 2016. Diagnosing focus. *Studies in Language* 40(2). 259–301.

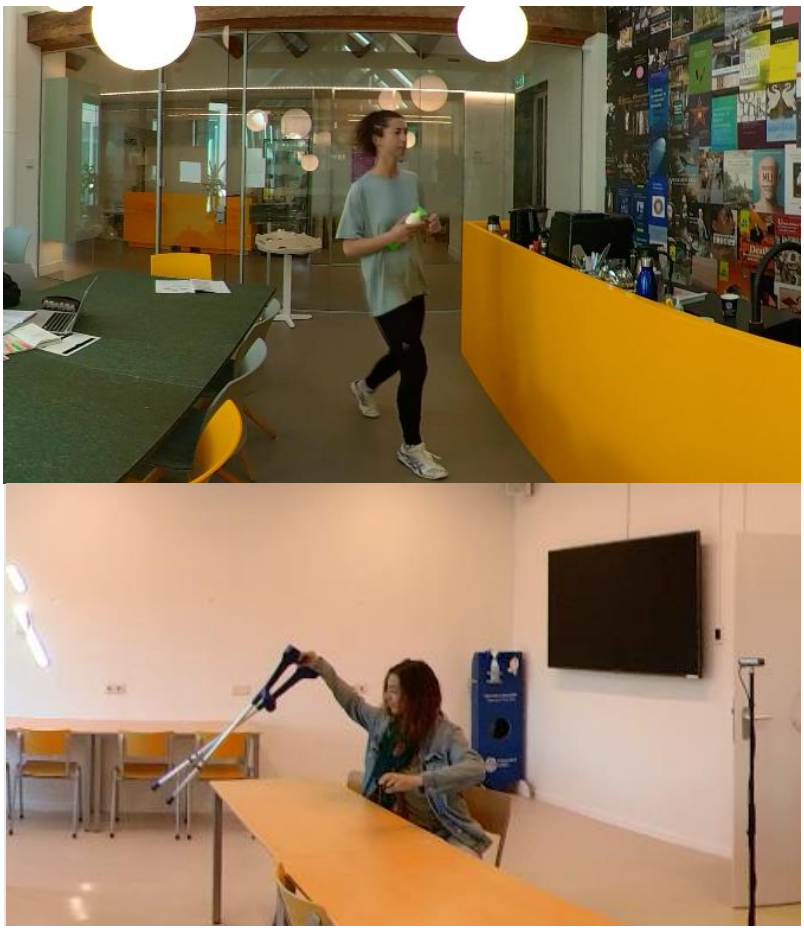
Van der Wal, Jenneke. 2017. What is the conjoint/disjoint alternation? In J. Van der Wal and L. M. Hyman (eds), *The conjoint/disjoint alternation in Bantu*, 14–60. Trends in Linguistics series. Berlin: Mouton de Gruyter.

Van der Wal, Jenneke. 2021. The BaSIS basics of information structure. Available via Leiden Repository: https://scholarlypublications.universiteitleiden.nl/handle/1887/3608096.

Zerbian, Sabine. 2007. A first approach to information-structuring in Xitsonga/Xichangana. *SOAS Working Papers in Linguistics* 15. 65–78.

Appendix A – Stills from the videos

Resultative: running > using crutches



Terminativity: squeezing (durative) vs spilling (terminative)

Evidentiality: indirect vs direct

Appendix B – Extract from the Spanish data

The video used for elicitation of these data can be found here:
https://video.leidenuniv.nl/playlist/dedicated/1_5d9xqz5l/

(1) En el primer vídeo he visto lo que
in DET.DEF.M first video have.1SG.PRS see.PPTCP PRO.M REL
pasó ayer en la cafetería de la facultad.
happen.3SG.PFV yesterday in DET.DEF.F cafeteria of DET.DEF.F faculty
'In the first video I have seen what happened yesterday at the Faculty cafeteria.'

(2) Había un par de chico-s sentados a la mesa
have.3SG.IMPF DET.INDEF.M pair of boy-PL sit.PPTCP-PL at DET.DEF.F table
leyendo un libro.
read.PTCP DET.INDEF.M book
'There were a couple of boys sitting at the table reading a book.'

(3) Otra chica que estaba estudiando con su libro, su
other.F girl REL be.3SG.IMPF study.PTCP with POSS.SG.M book POSS.SG.M
ordenador.
computer
'Another girl that was studying with her book, her computer.'

(4) Entró una chica rubia que creo que
enter.3SG.PFV DET.INDEF.F girl blond.F REL think.1SG.PRS COMP
era Laura.
be.3SG.IMPF Laura
'A blond girl entered, I think it was Laura.'

(5) Y luego también estaba Jai que fue a
and afterwards also be.3SG.IMPF Jai REL go.3SG.PFV to
por un café.
for DET.INDEF.M coffee
'And later Jai was there also, she went for a coffee.'

(6) Fue a pedir un café.
go.3SG.PFV to order.INF DET.INDEF.M coffee
'She went to order a coffee.

(7) Estaba también Juan a-l otro lado de la barra
be.3SG.IMPF also Juan at-DET.DEF.M other.M side of DET.DEF.F counter
'Juan was also at the other side of the counter.'

Appendix C – Extract from the Xitsonga data

Description of the intermezzo video (tones are not transcribed).

(1) Ni-vhon-a va-nhu… (ah) va-le ku-famb-eni.
1SG.SM-see-FV 2-person 2SM-be 15-walk-LOC
'I see people... eh, they are walking.'

(2) Ah vale receptio-ni.
eh 2SM-be reception-LOC
'Eh, they are at the reception.'

(3) Ahmm ku-na nuna loyi a-huma=ku hi nyango,
uhm 17SM-have 1.man 1.REL 1SM-exit?=REL 16 3.door
eya handle ngako w-a-ni-nangut-a.
1SM-go outside like 1SM-DJ-1SG.OM-look-FV
'Uhm, there is a man who is getting out the door, he is going outside, it looks like he is looking at me.'

(4) Ni-vhon-a va-nhu va-mbiri, un'wany-ana e-ya handle
1SG.SM-see-FV 2-person 2-two 1.other-DIM 1SM-go outside
e-nyangw-eni, un'wany-ana e-nghen-a hi nyango.
LOC-3.door-LOC 1.other-DIM 1SM-enter-FV 16 3.door
'I see two people, one is going outside the door, the other one is entering at the door.'

(5) Ni-tw-a bele... ngaku (ku-rila), ku-na (lesh)...
1SG.SM-hear-FV 9.bell like (15-cry) 17SM-have (whatsit)
bele leyi yi-ril-a=ku.
9.bell 9.REL 9SM-cry-FV=REL
'I hear a bell, it seems like it is ringing... there is a bell that is ringing.'

(6) Ah, kuna nuna, kambe e nsati... (a-ni-sh.. ah) nuna
eh 17SM-have 1.man but 1SM 1.woman NEG-1SG.SM-? eh 1.man
(or... ya a-ni-swi,) a-ni-swi-vhon-i kahle kore
or DM NEG-1SG.SM-AUX NEG-1SG.SM-AUX-see-NEG well COMP
e nuna kambe e nsati, mara w-a-dy-a
1SM 1.man or 1SM 1.woman but 1SM-DJ-eat-FV
or e-le ku-dy-eni, ethla a-va-na x-a ku-nwa…
or 1SM-be 15-eat-LOC also 1SM-be-with 7-CONN 15-drink
ah xitima tora e-kusuhi ka yena... ok e nuna
eh 7SM-put.out thirst LOC-near LOC 1.PRO ok 1SM 1.man
'There is a man... is it a woman? (I don't... eh) man... (or... I can't) I can't see if it is a man or a woman, but s/he is eating; s/he also has a drink... thirst quencher with him... ok it's a man.'

(7) enen ku-na va-nhu v-a-ku-tala la va-nga-tsham-a
and.then 17SM-have 2-person 2-CONN-15-be.many 2.REL 2SM-REL-sit-FV
hansi... (va-na)... ngaku va-le ku-vula-vul-eni, va-n'wany-ana
16-down 2SM-have like 2SM-be 15-talk-RED-LOC 2-other-DIM
va-le ku-tir-eni, va-n'wany-ana va-le ku-hlay-eni
2SM-be 15-work-LOC 2-other-DIM 2SM-be 15-read-LOC
'There are a lot of people who are sitting down... (they have)... It's like they're busy talking, some are working, some are reading.'

(8) Ok so, ku-na mi-nyango mi-mbiri...
ok so 17SM-have 4-door 4-two
'Ok so, there are two doors...'

Appendix D - Flowchart of our work process