

CESA and CoBiVa Training Handbook

Developed by:

**Ryan M. Bessett
University of Texas Rio Grande Valley**

**Ana M. Carvalho
University of Arizona**

**Katherine Christoffersen
University of Texas Rio Grande Valley**

About the handbook and the corpora

El corpus del Español en el Sur de Arizona (CESA):

This corpus aims at documenting and disseminating Spanish varieties spoken in Arizona. Graduate and undergraduate students, under PI's supervision, collect, transcribe, and analyze interviews with local Spanish speakers. This digital oral corpus provides material for multiple linguistic analyses of local bilingual Spanish and subsequent comparisons with other varieties. In order to be able to document and study Arizona Spanish as it is actually spoken, sociolinguistic interviews are conducted with native speakers of this variety. Volunteers from the Spanish-speaking communities are asked to sit for an interview for approximately 60 minutes to discuss themselves and their community. The interviews are conducted wherever the participant prefers, either in his/her home or in a public space. Interviews are conducted in Spanish, although participants switch to English whenever they want to. Participants are interviewed and digitally recorded in individual sessions that aim at eliciting spontaneous speech. The speech data are transcribed, anonymized, and stored in a password protected online file. The project is IRB-protected and all measures are taken to assure the participants' anonymity.

El corpus bilingüe del valle (CoBiVa):

This project is the first comprehensive digital documentation, description, and analysis of bilingualism in the Rio Grande Valley. This digital archive of sociolinguistic interviews provides a collection of oral narratives documenting local histories, cultural events, and language attitudes. This corpus is meant to inform scholars on bilingualism, sociolinguistics and border languages and educate the community on issues of dialect awareness, thereby fostering an appreciation of local language varieties in the Rio Grande Valley.

Through the development of a corpus of naturally occurring bilingual speech in the Rio Grande Valley, our goal is to accomplish the following objectives:

- Foster respect for linguistic diversity and admiration for bilinguals of the border region of the Rio Grande Valley
- Legitimize the local variety by incorporating its usage in the university classroom
- Promote dialect awareness within the university and broader community
- Provide data for numerous and diverse analyses of local language varieties, fostering a scientific approach for language analysis that is detached from prejudice and hegemonic notions of language
- Train students in sociolinguistic research methodology including data collection, transcription, and analysis
- Establish and enhance connections between the university and the local community
- Facilitate student research, presentations, and publications based on this IRB-approved local data

How to cite the handbook and the corpora

To cite the CESA and CoBiVa Training Handbook:

Bessett, Ryan M., Carvalho, Ana M., and Christoffersen, Katherine. (2022). *The CESA and CoBiVa Training Handbook*. https://bit.ly/CESA_CoBiVa_Handbook

To cite the CESA corpus:

Carvalho, Ana M. (2012-). *Corpus del Español en el Sur de Arizona (CESA)*. University of Arizona. cesa.arizona.edu.

To cite the CoBiVa corpus:

Christoffersen, Katherine and Bessett, Ryan. M. (2019). *Corpus Bilingüe del Valle (CoBiVa)*. University of Texas Rio Grande Valley. <https://utrgv.edu/cobiva>

Table of Contents

Part 1: Interview Protocol

Interview Protocol (Spanish)	p. 6-9
Interview Protocol (English)	p. 10-13
Interview Procedures (Condensed)	p. 14-15
ACTFL Proficiency Guidelines (2012)	p. 16-20
Sociolinguistic Interview Plan.....	p. 21-22

Part 2: Forms to complete

Bilingual Language Profile (Spanish)	p. 24-25
Bilingual Language Profile (English)	p. 26-27
Demographic information about the Interviewee.....	p. 28
Demographic information about the interviewer.....	p. 29
Field notes.....	p. 30
Consent form in English (CoBiVa).....	p. 31
Consent form in Spanish (CoBiVa).....	p. 32
Consent form in English (CESA).....	p. 33
Consent form in Spanish (CESA).....	p. 34-35

Part 3: Transcription Protocol

Organizing the files.....	p. 37
Transcription software.....	p. 38
Transcription instructions.....	p. 39-41
How to transcribe using stream.....	p. 42-43
Revising transcriptions from stream with R.....	p. 44-52

Part 1:

Interview Protocol

Procedimientos para las entrevistas

Antes de reunirte con el/la participante:

1. Piensa en temas de conversación para sacar durante la entrevista.
2. Crear un plan
 - Vas a tener que ser flexible y adaptarte a tu participante, pero hay que tener un plan general
 - No queremos preguntas relacionadas al lenguaje hasta el final de la entrevista
 - Si el participante piensa que el enfoque es el lenguaje, influirá su manera de hablar
3. Hay que conseguir una grabadora (una aplicación del smart phone puede servir)
 - Trata de usarla un par de veces
 - Asegúrate que la sabes usar y que va a funcionar bien
 - Escucha la grabación
 - Hay que entender claramente lo que fue grabado

Al llegar a la entrevista:

1. Antes de empezar a grabar, platica un poco con él/ella, haz que se sienta cómodo(a)
 - Especialmente importante si no conoces muy bien a tu participante
2. Formulario de consentimiento
 - Dáselo a tu participante con cuidado
 - Explícale que es apenas una formalidad
 - Que no haría falta si no fuera por los tramites de la universidad, etc etc
 - Es importante no asustarle a tu participante o hacerle sentir en una situación formal

Empezar la entrevista:

1. ¿Cómo debe ser la conversación?
 - El participante habla mucho y el entrevistador habla muy poco
 - La conversación sigue de una manera más o menos natural
 - Ryan:** ¿Y de niña qué juegos jugabas?
 - Hillo003:** ¿Te los vas a saber?
 - Ryan:** A lo mejor no, pero luego me puedes explicar...

Hillo003: Ah ok, pues era uno que se llamaba la bebe leche..

Ryan: Ah esto es lo que contó tu mamá también.

Hillo003:

-Hay que "engage" el participante.

-Mirarle a los ojos

-Estar metido en la conversación

-Tomar la posición de que quieres aprender de la persona, de quién es, de qué sabe

-Esta persona es la persona más interesante del mundo

-(Pero: NO EXAGERES. Se nota cuando parece falso y eso no ayuda)

-Problema: La paradoja del observador

-"To obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed" (Labov 1972: 113).

-Es nuestra tarea de disminuir los efectos de la paradoja

-Trata de hacer la entrevista lo más natural posible

-Trata de mantener la entrevista lo menos formal y más cómoda que sea posible

-Siempre seguir los consejos bajo de la sección de "Hay que "engage" el participante"

-Según Labov, durante los primeros 20 minutos (más o menos) el participante es más consiente de el entrevistador como un observador, pero después, empieza una conversación más natural. Ahí es donde podemos ver menos efectos de la paradoja.

-Por eso es fundamental mantener un ambiente casual y siempre buscar temas de interés personal del participante

-También por eso hacemos las preguntas demográficas al inicio de la entrevista

-Ya que la sección es más formal, podemos preguntar cosas más formales y sacar ideas para después

2. La entrevista debe durar una hora

3. Durante la entrevista puedes tomar unos apuntes de cosas que salen/ocurren que parecen de algún interés particular

-Es mejor notar esas cosas mentalmente y apuntarlas justo después de la entrevista

-Si pasas toda la entrevista apuntando cosas, no vas a lograr que el participante hable de una manera informal

4. Recuerda que tu plan de temas/preguntas debe ser flexible

- Hay que seguir los temas que le interesen al participante
- Hay que crear preguntas para seguir otras (muchas veces necesario para sacarle la plática a tu participante)

Ryan: ¿Y has cruzado la frontera?

Hillo003: Sí

Ryan: ¿Y cuantas veces?

Hillo003: Pues más de veinte yo creo

Ryan: ¿Y que te gusta hacer al otro lado?

Hillo003: De compras jaja.

Ryan: ¿Como qué?

Hillo003: Pues ropa. Voy y compro ropa. Me ha gustado ir hacer otras cosas pero igual no tengo tiempo porque siempre voy como dos días o tres días y es que como he querido ir como al zoológico o ir a otras partes así a fuera. He ido al Gran Canyon también y así, y es muy bonito pero, es cuando tienes que tener mas días, pero por lo regular cuando voy es con lo de compras.

5. ¿Qué haces si el participante no sabe contestar?

Ryan: Ok. Y ¿puedes recordar de un evento muy importante de tu vida?

Hillo 020: ¿Un evento muy importante? ¿Evento como qué?

Ryan: Como algo que te pasó muy bueno..quizás malo...

Hillo 020: Un evento...pues no es tan un evento... Yo estuve en un... concurso académico de la historia de México y soy una de las que si me das un libro para aprendérmelo de memoria con puntos y comas..yo me lo aprendo y así soy... Y estuve en la historia de México... Estuve en la historia de México cursando al nivel estatal y gané. Y pues pero pues, la historia de México. Imagínate. La cultura hasta Fox hasta ahí...muy... Tengo mm... Tengo la memoria..Tengo muy buena xxx de memoria. Y este fue un evento que porque a mí.. Yo estaba en una escuela y en ese escuela donde estaba los maestros apostaban... eh....si ganas te voy a dar esto, si ganas te voy a dar el otro...Entonces fueron muchas de que me tenían muy presionada y cuando logré no pensé que lo iba a lograr pero...era de los cinco mejores del estado en historia..

6. Diferentes tipos de participantes

- Algunos participantes por naturaleza hablan mucho y no paran de hablar (te facilita la tarea)
- Otros, no quieren contestar mucho y tienes que tratar de encontrar un tema que les anime
- Tenía un participante que contestaba "sí" o "no" o en un par de palabras para TODO. Se quedó viendo un calendario en el

cuarto que tenía imágenes de carros. Le pregunté ¿Cuál modelo de carro es tu preferido? y habló por unos cinco minutos sin parar.

Después de la entrevista:

1. Agradecerle a tu participante por el tiempo en que pasó en participar
2. Pide que tu participante rellene el documento "Bilingual Language Profile table (for speaker)"
3. Pide que tu participante rellene el documento "Demographic information, participant (DI)"

En algún momento:

Tienes que rellenar el documento "Demographic information for the interviewer"

-Hay que referirse a los criterios de ACTFL para documentar tu nivel de competencia en el español

[-http://actflproficiencyguidelines2012.org/speaking](http://actflproficiencyguidelines2012.org/speaking)

Lista de los formularios que hay que entregar:

1. Bilingual Language Profile (BLP)
2. Demographic information, participant (DI)
3. Demographic information, interviewer (II)
4. Fieldnotes (FN)
5. Consent form (in language the participant prefers)

Procedures for interviewing

Before meeting with the participant:

1. Think about the possible conversation topics for the interview
2. Create a plan
 - You will need to be flexible and adapt to your participant, but you do need a plan
 - We want to AVOID LANGUAGE related questions until the END of the interview
 - If the participant suspects the focus of the interview is about language, it will influence the way he/she speaks
3. You will need a recording device (an application on your smart phone will work)
 - Test the recorder/phone app a few times before the interview
 - Ensure that you know how to use it and that it will record
 - Listen to the recording
 - Make sure you can clearly make out what is recorded
 - Transfer the file from your device to a computer

When you arrive for the interview:

1. Before recording, talk for a little while with your participant, make him/her feel comfortable
 - This is especially important if you do not know the participant
2. Consent form
 - Give it to your participant CAREFULLY
 - Explain to him/her that it is just a formality
 - That it wouldn't be necessary except for the policies of the university
 - It is very important to make sure your participant doesn't see this as a formal situation

Starting the interview:

1. What should the conversation look like?
 - The participant speaks a lot and the interviewer speaks little
 - But the interview should follow a normal conversation flow
 - Ryan:** ¿Y de niña qué juegos jugabas?
 - Hillo003:** ¿Te los vas a saber?
 - Ryan:** A lo mejor no, pero luego me puedes explicar...

Hillo003: Ah ok, pues era uno que se llamaba la bebe leche..

Ryan: Ah esto es lo que contó tu mamá también.

Hillo003:

- You need to engage the participant.
 - Make eye contact
 - Be active in the conversation
 - You want to learn from the participant, their life, their experiences
 - This is the most interesting person in the world!!
 - (But: DON'T EXAGGERATE. Its easy to tell if you are being fake and this wont help)

- Problem: The observers paradox
 - "To obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed" (Labov 1972: 113).
 - Our job is to lessen the effects of the observers paradox
 - Try to make the interview as natural as possible
 - Try to make the interview as informal and as comfortable as possible
 - Always follow the instructions under the section "you need to engage the participant"
 - Según Labov, during the first 20 minutes (more or less) the participant is the most aware of the interviewer as an observer, but later, the conversation becomes more natural. This is when we can see the least effects of the paradox
 - For this reason, it is important to keep the environment a casual one and to find topics that are interesting to the participant
 - This is also why we do the demographic questions at the start of the interview, since this is the most formal part, we can ask the more formal questions and get ideas of what to ask later

2. The interview should last about an hour

3. During the interview you want to take notes about what the participant says or how the participant says them

- It is best to take these notes mental and write them down right after the interview
 - If you are taking notes during the interview, the participant will not speak in an informal way
- You can include these in the "Field Notes (FN)" document

4. At the end of the interview (the last 10 minutes) you can ask questions

about language.

5. Remember to be flexible with your plan

- You have to find themes that are of interest to the participant
- You need to ask follow up questions (to get the participant to talk)

Ryan: ¿Y has cruzado la frontera?

Hillo003: Sí

Ryan: ¿Y cuantas veces?

Hillo003: Pues más de veinte yo creo

Ryan: ¿Y que te gusta hacer al otro lado?

Hillo003: De compras jaja.

Ryan: ¿Como qué?

Hillo003: Pues ropa. Voy y compro ropa. Me ha gustado ir hacer otras cosas pero igual no tengo tiempo porque siempre voy como dos días o tres días y es que como he querido ir como al zoológico o ir a otras partes así a fuera. He ido al Gran Canyon también y así, y es muy bonito pero, es cuando tienes que tener mas días, pero por lo regular cuando voy es con lo de compras.

6. What do you do if the participant doesn't give a good response?

Ryan: Ok. Y ¿puedes recordar de un evento muy importante de tu vida?

Hillo 020: ¿Un evento muy importante? ¿Evento como qué?

Ryan: Como algo que te pasó muy bueno..quizás malo...

Hillo 020: Un evento...pues no es tan un evento... Yo estuve en un...concurso académico de la historia de México y soy una de las que si me das un libro para aprendérmelo de memoria con puntos y comas..yo me lo aprendo y así soy... Y estuve en la historia de México... Estuve en la historia de México cursando al nivel estatal y gané. Y pues pero pues, la historia de México. Imagínate. La cultura hasta Fox hasta ahí...muy... Tengo mm... Tengo la memoria..Tengo muy buena xxx de memoria. Y este fue un evento que porque a mí.. Yo estaba en una escuela y en ese escuela donde estaba los maestros apostaban... eh....si ganas te voy a dar esto, si ganas te voy a dar el otro...Entonces fueron muchas de que me tenían muy presionada y cuando logré no pensé que lo iba a lograr pero...era de los cinco mejores del estado en historia..

7. Different types of participants

- Some participants will talk a lot and make your job really easy
 - Husband after I interviewed his wife
- Other participants will not answer in many words and you have to find a theme that interests them
 - I had a participant that answered “yes” or “no” or in just a couple of words to every question (even open-ended

questions). He kept looking at this calendar on the wall that was filled with images of cars. So I asked “what was the make and model of your first car?” and he talked non-stop for the next five minutes.

After the interview:

1. Thank your participant for the time they spent talking to you
2. Ask your participant to fill out the "Bilingual Language Profile table (for speaker)"
3. Ask your participant to fill out the "Demographic information, participant (DI)"

At some point:

You need to fill out the "Demographic information for the interviewer"

-For non-native speakers (you learned Spanish in school) you will need to use the ACTFL guidelines to self-report your Spanish competence

-<http://actflproficiencyguidelines2012.org/speaking>

List of forms you need to turn in:

1. Bilingual Language Profile (BLP)
2. Demographic information, participant (DI)
3. Demographic information, interviewer (II)
4. Fieldnotes (FN)
5. Consent form (in language the participant prefers)

Sociolinguistic Interview Procedures (Condensed)

Before meeting with the participant:

1. Complete **Sociolinguistic Interview Plan**.
 - AVOID LANGUAGE related questions until the END of the interview.
 - Participant should agree to spend 1-1.5 hours talking with you.
2. Complete the **Demographic Information of the Interviewer (DI)** document.
3. Complete **Practice Recording**.
 - You will need a recording device (an application on your smart phone will work).

When you arrive for the interview:

1. Before recording, talk for a little while with your participant, make him/her feel comfortable
2. **Consent form**
 - Explain to your participant that it is just a formality
 - It is very important to make sure your participant doesn't see this as a formal situation

During the interview:

1. How long should the interview be?
The interview should be at least 1 hour long.
2. What should the interview be like?
 - The interview should follow a normal conversation flow.
 - Ask follow up questions. You are having a conversation, but trying to learn a little more about the participant.
 - Be engaging. (Make eye contact. You want to learn from the participant, their life, their experiences. This is the most interesting person in the world!! But... DON'T EXAGGERATE. It's easy to tell if you are being fake, and this won't help.
 - Note: Some participants will talk a lot. And some will talk a little. With some, you may need to work harder to find a topic that's interesting to them.
3. Should I take notes or read questions during the interview?
 - No. This is why your preparation is so important. You do not want to make this seem like a formal interview, so plan on topics/questions beforehand.
 - Take mental notes about what the participant says or how the participant says them for the "Field Notes (FN)" document after the interview. But do not take actual notes during the interview.
4. When can I ask the questions about language?
 - Wait to ask questions about language until the last 10 minutes.

After the interview:

1. Thank your participant for the time they spent talking to you
2. Ask your participant to fill out the **Bilingual Language Profile (English or Spanish)**.
3. Ask your participant to fill out the **Demographic information for the Participant (DP)**.

4. Complete the **Field Notes (FN)**.
5. Submit the BLP, DP, FN, signed Consent and audio file.

ACTFL PROFICIENCY GUIDELINES (2012) – SPEAKING¹

Superior:

Speakers at the Superior level are able to communicate with accuracy and fluency in order to participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives. They discuss their interests and special fields of competence, explain complex matters in detail, and provide lengthy and coherent narrations, all with ease, fluency, and accuracy. They present their opinions on a number of issues of interest to them, such as social and political issues, and provide structured arguments to support these opinions. They are able to construct and develop hypotheses to explore alternative possibilities.

When appropriate, these speakers use extended discourse without unnaturally lengthy hesitation to make their point, even when engaged in abstract elaborations. Such discourse, while coherent, may still be influenced by language patterns other than those of the target language. Superior-level speakers employ a variety of interactive and discourse strategies, such as turn-taking and separating main ideas from supporting information through the use of syntactic, lexical, and phonetic devices.

Speakers at the Superior level demonstrate no pattern of error in the use of basic structures, although they may make sporadic errors, particularly in low-frequency structures and in complex high-frequency structures. Such errors, if they do occur, do not distract the native interlocutor or interfere with communication.

Advanced High:

Speakers at the Advanced High sublevel perform all Advanced-level tasks with linguistic ease, confidence, and competence. They are consistently able to explain in detail and narrate fully and accurately in all time frames. In addition, Advanced High speakers handle the tasks pertaining to the Superior level but cannot sustain performance at that level across a variety of topics. They may provide a structured argument to support their opinions, and they may construct hypotheses, but patterns of error appear. They can discuss some topics abstractly, especially those relating to their particular interests and special fields of expertise, but in general, they are more comfortable discussing a variety of topics concretely.

Advanced High speakers may demonstrate a well-developed ability to compensate for an imperfect grasp of some forms or for limitations in vocabulary by the confident use of communicative strategies, such as paraphrasing, circumlocution, and illustration. They use precise vocabulary and intonation to express meaning and often show great fluency and ease of speech. However, when called on to perform the complex tasks associated with the Superior level over a variety of topics, their language will at times break down or prove inadequate, or they may avoid the task altogether, for example, by resorting to simplification through the use of description or narration in place of argument or hypothesis.

¹ American Council on the Teaching of Foreign Languages. (2012). *ACTFL Proficiency Guidelines 2012*. Alexandria, VA: ACTFL, INC. <https://www.actfl.org/resources/actfl-proficiency-guidelines-2012>.

Advanced Mid:

Speakers at the Advanced Mid sublevel are able to handle with ease and condense a large number of communicative tasks. They participate actively in most informal and some formal exchanges on a variety of concrete topics relating to work, school, home, and leisure activities, as well as topics relating to events of current, public, and personal interest or individual relevance.

Advanced Mid speakers demonstrate the ability to narrate and describe in the major time frames of past, present, and future by providing a full account, with good control of aspect. Narration and description tend to be combined and interwoven to relate relevant and supporting facts in connected, paragraph-length discourse.

Advanced Mid speakers can handle successfully and with relative ease the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine situation or communicative task with which they are otherwise familiar. Communicative strategies such as circumlocution or rephrasing are often employed for this purpose. The speech of Advanced Mid speakers performing Advanced-level tasks is marked by substantial flow. Their vocabulary is fairly extensive although primarily generic in nature, except in the case of a particular area of specialization or interest. Their discourse may still reflect the oral paragraph structure of their own language rather than that of the target language.

Advanced Mid speakers contribute to conversations on a variety of familiar topics, dealt with concretely, with much accuracy, clarity and precision, and they convey their intended message without misrepresentation or confusion. They are readily understood by native speakers unaccustomed to dealing with non-natives. When called on to perform functions or handle topics associated with the Superior level, the quality and/or quantity of their speech will generally decline.

Advanced Low:

Speakers at the Advanced Low sublevel are able to handle a variety of communicative tasks. They are able to participate in most informal and some formal conversations on topics related to school, home, and leisure activities. They can also speak about some topics related to employment, current events, and matters of public and community interest.

Advanced Low speakers demonstrate the ability to narrate and describe in the major time frames of past, present, and future in paragraph-length discourse with some control of aspect. In these narrations and descriptions, Advanced Low speakers combine and link sentences into connected discourse of paragraph length, although these narrations and descriptions tend to be handled separately rather than interwoven. They can handle appropriately the essential linguistic challenges presented by a complication or an unexpected turn of events.

Responses produced by Advanced Low speakers are typically not longer than a single paragraph. The speaker's dominant language may be evident in the use of false cognates, literal translations,

or the oral paragraph structure of that language. At times their discourse may be minimal for the level, marked by an irregular flow, and containing noticeable self-correction. More generally, the performance of Advanced Low speakers tends to be uneven.

Advanced Low speech is typically marked by a certain grammatical roughness (e.g., inconsistent control of verb endings), but the overall performance of the Advanced-level tasks is sustained, albeit minimally. The vocabulary of Advanced Low speakers often lacks specificity. Nevertheless, Advanced Low speakers are able to use communicative strategies such as rephrasing and circumlocution.

Advanced Low speakers contribute to the conversation with sufficient accuracy, clarity, and precision to convey their intended message without misrepresentation or confusion. Their speech can be understood by native speakers unaccustomed to dealing with non-natives, even though this may require some repetition or restatement. When attempting to perform functions or handle topics associated with the Superior level, the linguistic quality and quantity of their speech will deteriorate significantly.

Intermediate High:

Intermediate High speakers are able to converse with ease and confidence when dealing with the routine tasks and social situations of the Intermediate level. They are able to handle successfully uncomplicated tasks and social situations requiring an exchange of basic information related to their work, school, recreation, particular interests, and areas of competence.

Intermediate High speakers can handle a substantial number of tasks associated with the Advanced level, but they are unable to sustain performance of all of these tasks all of the time. Intermediate High speakers can narrate and describe in all major time frames using connected discourse of paragraph length, but not all the time. Typically, when Intermediate High speakers attempt to perform Advanced-level tasks, their speech exhibits one or more features of breakdown, such as the failure to carry out fully the narration or description in the appropriate major time frame, an inability to maintain paragraph-length discourse, or a reduction in breadth and appropriateness of vocabulary.

Intermediate High speakers can generally be understood by native speakers unaccustomed to dealing with non-natives, although interference from another language may be evident (e.g., use of code-switching, false cognates, literal translations), and a pattern of gaps in communication may occur.

Intermediate Mid:

Speakers at the Intermediate Mid sublevel are able to handle successfully a variety of uncomplicated communicative tasks in straightforward social situations. Conversation is generally limited to those predictable and concrete exchanges necessary for survival in the target culture. These include personal information related to self, family, home, daily activities, interests and personal preferences, as well as physical and social needs, such as food, shopping, travel, and lodging.

Intermediate Mid speakers tend to function reactively, for example, by responding to direct questions or requests for information. However, they are capable of asking a variety of questions when necessary to obtain simple information to satisfy basic needs, such as directions, prices, and services. When called on to perform functions or handle topics at the Advanced level, they provide some information but have difficulty linking ideas, manipulating time and aspect, and using communicative strategies, such as circumlocution.

Intermediate Mid speakers are able to express personal meaning by creating with the language, in part by combining and recombining known elements and conversational input to produce responses typically consisting of sentences and strings of sentences. Their speech may contain pauses, reformulations, and self-corrections as they search for adequate vocabulary and appropriate language forms to express themselves. In spite of the limitations in their vocabulary and/or pronunciation and/or grammar and/or syntax, Intermediate Mid speakers are generally understood by sympathetic interlocutors accustomed to dealing with non-natives.

Overall, Intermediate Mid speakers are at ease when performing Intermediate-level tasks and do so with significant quantity and quality of Intermediate-level language.

Intermediate Low:

Speakers at the Intermediate Low sublevel are able to handle successfully a limited number of uncomplicated communicative tasks by creating with the language in straightforward social situations. Conversation is restricted to some of the concrete exchanges and predictable topics necessary for survival in the target-language culture. These topics relate to basic personal information; for example, self and family, some daily activities and personal preferences, and some immediate needs, such as ordering food and making simple purchases. At the Intermediate Low sublevel, speakers are primarily reactive and struggle to answer direct questions or requests for information. They are also able to ask a few appropriate questions. Intermediate Low speakers manage to sustain the functions of the Intermediate level, although just barely.

Intermediate Low speakers express personal meaning by combining and recombining what they know and what they hear from their interlocutors into short statements and discrete sentences. Their responses are often filled with hesitancy and inaccuracies as they search for appropriate linguistic forms and vocabulary while attempting to give form to the message. Their speech is characterized by frequent pauses, ineffective reformulations and self-corrections. Their pronunciation, vocabulary, and syntax are strongly influenced by their first language. In spite of frequent misunderstandings that may require repetition or rephrasing, Intermediate Low speakers can generally be understood by sympathetic interlocutors, particularly by those accustomed to dealing with non-natives.

Novice High:

Speakers at the Novice High sublevel are able to handle a variety of tasks pertaining to the Intermediate level, but are unable to sustain performance at that level. They are able to manage successfully a number of uncomplicated communicative tasks in straightforward social situations. Conversation is restricted to a few of the predictable topics necessary for survival in

the target language culture, such as basic personal information, basic objects, and a limited number of activities, preferences, and immediate needs. Novice High speakers respond to simple, direct questions or requests for information. They are also able to ask a few formulaic questions.

Novice High speakers are able to express personal meaning by relying heavily on learned phrases or recombinations of these and what they hear from their interlocutor. Their language consists primarily of short and sometimes incomplete sentences in the present, and may be hesitant or inaccurate. On the other hand, since their language often consists of expansions of learned material and stock phrases, they may sometimes sound surprisingly fluent and accurate. Pronunciation, vocabulary, and syntax may be strongly influenced by the first language. Frequent misunderstandings may arise but, with repetition or rephrasing, Novice High speakers can generally be understood by sympathetic interlocutors used to non-natives. When called on to handle a variety of topics and perform functions pertaining to the Intermediate level, a Novice High speaker can sometimes respond in intelligible sentences, but will not be able to sustain sentence-level discourse.

Novice Mid:

Speakers at the Novice Mid sublevel communicate minimally by using a number of isolated words and memorized phrases limited by the particular context in which the language has been learned. When responding to direct questions, they may say only two or three words at a time or give an occasional stock answer. They pause frequently as they search for simple vocabulary or attempt to recycle their own and their interlocutor's words. Novice Mid speakers may be understood with difficulty even by sympathetic interlocutors accustomed to dealing with non-natives. When called on to handle topics and perform functions associated with the Intermediate level, they frequently resort to repetition, words from their native language, or silence.

Novice Low:

Speakers at the Novice Low sublevel have no real functional ability and, because of their pronunciation, may be unintelligible. Given adequate time and familiar cues, they may be able to exchange greetings, give their identity, and name a number of familiar objects from their immediate environment. They are unable to perform functions or handle topics pertaining to the Intermediate level, and cannot therefore participate in a true conversational exchange.

Sociolinguistic Interview Plan

For your sociolinguistic interview plan, you will need to plan the arrangements for the interview and the questions/topics you will discuss. Complete Part 1 by answering the questions and Part 2 by listing the questions/topics you will discuss.

Part 1: Answer the following questions for your sociolinguistic interview.

1. Who will you interview?
Remember, this should be someone who has lived most of their life here in the RGV, or at least the past 10 years. They should be a part of the RGV community.
2. What language will you conduct the interview in?
You are welcome to code-switch during the interview, but it'll probably be helpful to think of a primary language to use, the language you would normally use with this participant. We'd like to encourage the use of Spanish.
3. When might you conduct the interview?
You have been approved and may start your interview. Review the Tagliamonte (2003) reading and the training materials beforehand. Practice and review topics/questions beforehand.
4. Where might you conduct the interview?
Conduct the interview off campus in a relatively quiet location (ie, coffee shop, home, etc). NOT at the university
5. How will you record the interview?
Your phone, laptop or an audio recorder will work. Do a short test recording first. Make it as unobtrusive as possible. Listen to the audio recording to ensure good audio quality. Practice transferring the file to your computer and sending via email/uploading to Blackboard.

Part 2: Choose at least 10 general questions/topics to discuss during the interview and at least the first 8 language-related questions (for the last 10 minutes only). *You may choose more. Topics may be from the "Potential Interview Questions" list or not.*

General Questions/Topics (Feel free to follow the conversation.)

- 1.
- 2.
- 3.

- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.

Language (only last 10 minutes- Please use and copy starred questions here.)

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10.

Part 2: Forms to complete

BILINGUAL LANGUAGE PROFILE: SPANISH-ENGLISH²

SPEAKER'S ID: _____

I. Historial lingüístico	español	inglés
1. ¿A qué edad empezó a aprender las siguientes lenguas?		
2. ¿A qué edad empezó a sentirse cómodo usando las siguientes lenguas?		
3. ¿Cuántos años de clases (gramática, historia, matemáticas, etc.) ha tenido en las siguientes lenguas (desde la escuela primaria a la universidad)?		
4. ¿Cuántos años ha pasado en un país/región donde se hablan las siguientes lenguas?		
5. ¿Cuántos años ha pasado en familia hablando las siguientes lenguas?		
6. ¿Cuántos años ha pasado en un ambiente de trabajo donde se hablan las siguientes lenguas?		

II. Uso de lenguas	español	inglés
7. En una semana normal, ¿qué porcentaje del tiempo usa las siguientes lenguas con sus amigos?		
8. En una semana normal, ¿qué porcentaje del tiempo usa las siguientes lenguas con su familia?		
9. En una semana normal, ¿qué porcentaje del tiempo usa las siguientes lenguas en la escuela/el trabajo?		
10. Cuando se habla a usted mismo, ¿con qué frecuencia se habla a sí mismo en las siguientes lenguas?		

² Adapted from Birdsong, D., Gertken, L.M., & Amengual, M. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. <<https://sites.la.utexas.edu/bilingual/>>.

11. Cuando hace cálculos contando, ¿con qué frecuencia cuenta en las siguientes lenguas?		
--	--	--

III. Competencia	<i>(0=no muy bien 6=muy bien)</i>
12. a. ¿Cómo habla en español?	
b. ¿Cómo habla en inglés?	
13. a. ¿Cómo entiende en español?	
b. ¿Cómo entiende en inglés?	
14. a. ¿Cómo lee en español?	
b. ¿Cómo lee en inglés?	
15. a. ¿Cómo escribe en español?	
b. ¿Cómo escribe en inglés?	

IV. Actitudes	<i>0=no estoy de acuerdo 6=estoy de acuerdo</i>
16. a. Me siento “yo mismo” cuando hablo en Español.	
b. Me siento “yo mismo” cuando hablo en Inglés.	
17. a. Me identifico con una cultura Hispanohablante.	
b. Me identifico con una cultura Anglohablante.	
18. a. Es importante para mi usar (o llegar a usar) español como un hablante nativo.	
b. Es importante para mi usar (o llegar a usar) inglés como un hablante nativo.	
19. a. Quiero que los demás piensen que soy un hablante nativo de español.	
b. Quiero que los demás piensen que soy un hablante nativo de inglés.	

BILINGUAL LANGUAGE PROFILE: SPANISH-ENGLISH³

SPEAKER'S ID: _____

I. Linguistic History	Spanish	English
1. At what age did you start learning the following languages?		
2. At what age did you start feeling comfortable speaking the following languages?		
3. How many classes (grammar, history, math, etc.) have you taken in the following languages (thinking from elementary school through college)?		
4. How many years have you lived in a country/region where the following languages are spoken?		
5. How many years have you spoken the following languages with your family?		
6. How many years have you worked at a job where the following languages are spoken?		

II. Language Use	Spanish	English
7. In a normal week, what percent of the time do you use the following languages with your friends?		
8. In a normal week, what percent of the time do you use the following languages with your family?		
9. In a normal week, what percent of the time do you use the following languages at school/work?		
10. When you are speaking to yourself, with what frequency do you speak (think in) the following languages?		

³ Adapted from Birdsong, D., Gertken, L.M., & Amengual, M. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. <<https://sites.la.utexas.edu/bilingual/>>.

11. When you do math in your head, with what frequency do you speak (think in) the following languages?		
---	--	--

III. Competency	<i>(0=not very well 6= very well)</i>
12. a. How well do you speak Spanish?	
b. How well do you speak English?	
13. a. How well do you understand Spanish?	
b. How well do you understand English?	
14. a. How well do you read in Spanish?	
b. How well do you read in English?	
15. a. How well do you write in Spanish?	
b. How well do you read in English?	

IV. Attitudes	<i>0=disagree 6=agree</i>
16. a. I feel myself when I speak in Spanish.	
b. I feel myself when I speak in English.	
17. a. I identify with a Spanish speaking culture.	
b. I identify with an English speaking culture.	
18. a. It is important for me to use (or start to use) Spanish like a native speaker.	
b. It is important for me to use (or start to use) English like a native speaker.	
19. a. I want people to think of me as a native speaker of Spanish.	
b. I want people to think of me as a native speaker of English.	

Demographic Information of the Participant	
Speaker's ID	
Place of birth	
Place where s/he was raised	
Year of birth	
Last grade level completed	
Where education took place	
Years in the US	
Years in RGV	
Current neighborhood	
Previous neighborhood(s)	
Current profession	
Desired profession (students)	
Profession in birth country	
Household yearly income	
Number of people in the household	
Parents' languages	
Mother's profession	
Father's profession	
Mother's educational level	
Father's educational level	
Mother's birth place	
Father's birth place	
Languages other than English and Spanish	
Sex	
Ethnicity (self)	
Phenotype	
Religion	
Political affiliation	
Close family in Mexico (or anywhere else)	
Frequency of visits to Mexico (often, occasional, rare, none)	
Visitors from Mexico (often, occasional, rare, none)	
Married? (L used with spouse)	
Spouse's ethnicity	
Spouse's L1	
Children?	
L used with children	

Demographic Information of the Interviewer	
Interviewer's ID	
Interviewer's name	
Interviewer's sex	
Interviewer's age	
Interviewer's ethnicity and phenotype	
Interviewer's L1	
Interviewer's proficiency in Spanish	
Interviewer's dialect(s) of Spanish	
Interviewer's relationship with the interlocutor	
Interviewer's educational level	

Field notes	
Speaker's ID	
Day and time when the interview was conducted	
Place where interview was conducted	
Attitude of the interviewee towards the interview	
Attitude of the interviewee towards the interviewer	
Formality of the interview	
Other people present and their relation to the participant	
Silence level	
Comfort level	
Other relevant information about the interview	
Notes about language used in the interview:	

The University of Texas Rio Grande Valley Informed Consent Form

[The Corpus of Spanish and English in the Rio Grande Valley]

Investigators: *Professor Ryan M. Bessett, PhD. and Katherine O'Donnell Christoffersen, Ph.D..*

Background: We are conducting a research study on the language and culture in the Rio Grande Valley.

Procedure: You will be asked about yourself and your community for approximately 60 minutes, and to answer questions about your use of Spanish and English. The conversation will be recorded and archived, in addition to a transcription of the conversation, non-identifiable demographic information about yourself, and your responses to your use of Spanish and English. The archives will be made available for researchers interested in analyzing Rio Grande Valley Spanish and English.

Risks or Possible Discomforts Associated with the Study: There are no anticipated risks associated with your participation in this study

Benefits of Participation: There is no direct benefit to your participation in this study, but the knowledge gained may be of benefit to others, including those interested in how Spanish is spoken in the United States today.

Voluntary Participation: Your participation in this study is voluntary; you may discontinue your participation at any time without penalty. If for any reason you decide that you would like to discontinue your participation, simply tell the researcher that you wish to stop.

Anonymity and/or Confidentiality: All identifiable information about yourself will be removed from the audio recording and marked as "XY" in the transcription.

Who to Contact for Research Related Questions: For questions about the research itself, or to report any adverse effects during or following participation, contact the researcher, [Ryan Bessett] at [ryan.bessett@utrgv.edu].

Who to Contact Regarding Your Rights as a Participant: This research has been reviewed and approved by the Institutional Review Board for Human Subjects Protection (IRB). If you have any questions about your rights as a participant, or if you feel that your rights as a participant were not adequately met by the researcher, please contact the IRB at (956) 665-2889 or irb@utrgv.edu.

Signatures: By signing below, you indicate that you are voluntarily agreeing to participate in this study and that the procedures involved have been described to your satisfaction. The researcher will provide you with a copy of this form for your own reference. In order to participate, you must be at least 18 years of age. If you are under 18, please inform the researcher.

Participant's Signature

____/____/____
Date

The University of Texas Rio Grande Valley Formulario de consentimiento

[El Corpus del español y el inglés en el valle del Rio Grande]

Investigadores: *Profesor Ryan M. Bessett, PhD. y Katherine O'Donnell Christoffersen, Ph.D.*

Material de base: Estamos realizando un estudio sobre la lengua y cultura en el valle del Rio Grande.

Procedimiento: Le preguntarán sobre usted y su comunidad durante unos 60 minutos y usted contestará preguntas acerca de su uso del español y el inglés. Se grabará la conversación. En base de datos en internet protegido por una contraseña, se guardará la conversación junto con una transcripción de la conversación, información inidentificable sobre usted y sus respuestas a su uso del español y el inglés. Los archivos estarán disponibles a investigadores interesados en analizar el español y el inglés del valle del Rio Grande. Para acceder a los archivos, se requerirá una solicitud que incluyen las credenciales y el CV del solicitante con una descripción del motivo del interés del usuario. Los investigadores concederán acceso a los que quieren utilizar el base de datos para estudios lingüísticos y para la enseñanza.

Riesgos o incomodidades asociados con el estudio: No se anticipa ningún riesgo debido a su participación en este estudio.

Beneficios al participar: No se asocia ningún beneficio directo a su participación al estudio, sin embargo el conocimiento adquirido puede beneficiar a otros, incluyendo a la gente con interés en el español hablado en los Estados Unidos.

Participación voluntaria: Su participación en el estudio es voluntario; se puede suspender su participación en cualquier momento sin ninguna sanción. Por cualquier motivo, si usted desea suspender su participación, sólo hay que informarle al investigador que quiere parar.

Anonimidad y/o confidencialidad: Se eliminará toda la información identificable sobre usted del audio de la grabación y se la marcará como "XY" en la transcripción.

Contacto para preguntas relacionadas al estudio: Para cualquier duda sobre la investigación o para declarar cualquier tipo de efecto adverso durante o después de participar, comuníquese con el investigador [Ryan Bessett] a [ryan.bessett@utrgv.edu].

Contacto sobre sus derechos como participante: Esta investigación ha sido revisada y aprobada por el Consejo de Revisión Institucional para la Protección de los Participantes Humanos (IRB). Para cualquier duda sobre sus derechos como participante o si sus derechos como participante no fueron respetados por parte del investigador, por favor de comuníquese con el IRB a (956) 665-2889 o irb@utrgv.edu.

Firma: Al firmar abajo, usted indica que está participante en este estudio de su propia voluntad y que se le ha advertido de los procedimientos involucrados. El investigador le proporcionará una copia de este formulario para su referencia. Para participar, tiene que tener al menos 18 años. Si tiene menos de 18, por favor de informarle al investigador.

Firma de participante

_____/_____/_____
Fecha

Consent form: Variation in Arizona Spanish

I, _____, have been asked to participate in this research study, which has been explained to me by _____.

This study is being conducted by Professor Ana M. Carvalho and her students at the University of Arizona as part of the project “Spanish in Arizona”. I am being asked to participate in this study because I am a speaker of Arizona Spanish.

I understand that the purpose of this study is to document and analyze the Spanish varieties spoken in Arizona.

I understand that I will be asked about myself and my community for approximately 60 minutes, and to answer questions about my use of Spanish and English. I also understand that the conversation I provide will be recorded and archived, in addition to a transcription of the conversation, non-identifiable demographic information about myself, and my responses about my use of Spanish and English. The archives will be made available for researchers who are interested in analyzing Arizona Spanish.

My participation is voluntary. I may decline to answer any question, I may choose to stop the interview at any time, and I may demand that the whole or parts of the interview be deleted on spot and/or not added to the corpus. I have been allowed to ask questions regarding this study. I will be given a signed copy of this form for my records.

I understand that this study is not direct benefit to me but the knowledge gained may be of benefit to others, including those interested in how Spanish is spoken in the United States today.

If I have any questions regarding this study I may contact Professor Ana M. Carvalho by phone at 520-751-1224 or 520-621-3639, by mail to Department of Spanish and Portuguese, Modern Languages 545, University of Arizona, Tucson, AZ 85721-0067 or by e-mail at anac@email.arizona.edu. If I have any questions or concerns about my treatment as a participant in this study, I may contact the Human Subjects Protection Program, Office for the Responsible Conduct of Research, by calling (520) 626-6721 at the University of Arizona.

I understand that in any research that results from this interview, neither my name nor any information from which I might be identified will be published or shared in any form. Recordings of my interview will be kept permanently archived in a secure location. I understand that no personal damaging information will be requested or included, and that any information I give in the interview will be kept as anonymous as possible. I understand that portions of the interviews or transcriptions may be used for presentation at professional conferences or classrooms for teaching purposes and on the internet.

I have read the above information and have sufficient information to decide to participate in this study.

_____ I volunteer to participate in this study

_____ I give permission for the audio-recordings, demographic and linguistic information about me, and interview transcripts to be archived permanently for future research and teaching purposes.

Name of the participant: _____ Signature: _____

Formulario de consentimiento: Variación en el español de Arizona

A mí, _____, se me ha pedido participar en un estudio de investigación; el estudio se me ha explicado por _____.

El estudio se lleva a cabo por la profesora Ana M. Carvalho y sus estudiantes de la Universidad de Arizona como parte del proyecto “Español en Arizona”. Se me ha pedido participar porque soy hablante de español de Arizona.

El estudio examina donde se habla español en Arizona y la variación que se da en la lengua. Se me ha explicado que me pedirá que yo hable sobre mi mismo/misma y mi comunidad durante aproximadamente 60 minutos, además de contestar preguntas sobre mi uso de español y inglés. Se grabarán en audio y se archivarán mis relatos. Esos archivos, además de la transcripción de la entrevista, mis características demográficas no-identificables, y mis respuestas en cuanto a mi uso del español y inglés, estarán disponibles para investigadores que estén interesados en analizar el español que se habla en Arizona.

Mi participación es voluntaria. Puedo negarme a contestar cualquier pregunta y puedo terminar la entrevista cuando quiera. Además, puedo exigir que se borre parte de la grabación o toda la entrevista sin tardanza, y/o que no se la incluya en el corpus. Se me ha permitido hacer preguntas aclaratorias con respeto a esa investigación, y se me dará una copia firmada de este documento. Entiendo que no hay beneficio para mí directamente pero que el conocimiento logrado con la investigación puede ser de beneficio para otros, incluso aquellos con interés en el español contemporáneo hablado en los Estados Unidos.

Si tengo preguntas sobre este estudio puedo contactar a la Profesora Ana M. Carvalho por teléfono (520-751-1224 o 520 621-3639), por correo en el Departamento de Español y Portugués, Modern Languages 545, University of Arizona, Tucson, AZ 85721-0067, o por e-mail en anac@email.arizona.edu. Si tengo preguntas o preocupaciones acerca del tratamiento hacia mi persona como participante en este estudio, puedo llamar la Oficina de Protección de Participantes de Investigaciones en la Universidad de Arizona en el número (520) 626-6721.

Entiendo que en cualquier investigación que resulte de esa entrevista, no se publicará ninguna información que pueda identificar a mi persona. Las grabaciones se mantendrán archivadas indefinitivamente en un lugar seguro. Entiendo que no se obtendrá información general que me pueda perjudicar y la información que doy en la entrevista se mantendrá lo más anónima posible. Entiendo que partes de la entrevista o transcripción se podrán usar para presentaciones en congresos profesionales y como demostraciones en cursos de español o en el internet.

He leído la información anterior y es suficiente para tomar la decisión de participar en este estudio. _____ Quiero participar en este estudio.

_____ Doy mi permiso para que se archiven indefinidamente las grabaciones, transcripciones, información demográfica y lingüística sobre mí, para futuras investigaciones y fines educativos.

Nombre del/ de la participante: _____ Firma:

Nombre del/de la investigador/a: _____ Firma:

Fecha: _____

Part 3:

Transcription Protocol

TRANSCRIPTION PROTOCOL

Organizing the files:

1. You will receive a folder with the following files:
 - a. Sound File
 - b. Demographic information
 - c. Bilingual Language Profile
 - d. Interviewer Information
 - e. Field Notes

2. You will also receive a number for the interviewer (i.e. CESA050, CoBiVa050)

3. Each file should be labeled using the CESA/CoBiVa number, round number, and your last name:
 - a. For CESA
 - i. Sound File: CESA#_SF_R1_LASTNAME

 - ii. Demographic information: CESA#_DI_R1_LASTNAME

 - iii. Bilingual Language Profile: CESA#_BLP_R1_LASTNAME

 - iv. Interviewer Information: CESA#_II_R1_LASTNAME

 - v. Field Notes: CESA#_FN_R1_LASTNAME

 - vi. Transcription: CESA#_TR_R1_LASTNAME

 - b. For CoBiVa
 - i. Sound File: CoBiVa#_Audio_R1_LASTNAME

 - ii. Demographic information: CoBiVa#_DIP_R1_LASTNAME

 - iii. Bilingual Language Profile: CoBiVa#_BLP_R1_LASTNAME

 - iv. Interviewer Information: CoBiVa#_DIP_R1_LASTNAME

 - v. Field Notes: CoBiVa#_FN_R1_LASTNAME

 - vi. Transcription: CoBiVa#_Transcript_R1_LASTNAME

Transcription Software:

How to Transcribe with ExpressScribe
(Insert Time Stamps and/or Transcribe)

Caution: Save Frequently!

1. Go to <https://www.nch.com.au/scribe/index.html>
2. Click on the 'Get free version of Express Scribe here.'
3. Set up your 'hot keys' for the keys that you prefer. (Click on Options or Preferences to get to 'hot key' set up.)
4. Add a key for 'copy time.'
5. Use the hot keys to slow, rewind, fast forward, pause, and play the audio.
6. Type into a text editor.
7. Click the hot key for 'copy time' and then ctrl+v to paste into your text editor.
8. Reformat according to WEBVTT format and submit.

- f. Mark Other Language using **(OL)** before the word(s) that are switched. Include borrowings (such as biles, parqueadero) and code-switches.
- g. Use standard spelling in the transcription. This is for the purpose of consistency, searchability, readability, and in order to avoid ‘eye dialect.’ (For example, write pues instead of pos.)
- h. For morphological variation, you may write that as pronounced. (For example, fuiste, fuistes). **Ask if you have a question.**
- i. Do not make notes about phonological, lexical or morphosyntactic features within the transcription. You may note these in the ‘Fieldnotes.’ Note interesting and/or frequent phonological, lexical or morphosyntactic features in the ‘Fieldnotes.’ (For example, no pronunciation of the ‘s’, code-switching for ‘so’, etc) This is **not** required. The fieldnotes are also for making notes about the context.
- j. Represent short interjections (2-3 words) without starting a new turn. See example below for /yeah/

00:00:00.00 --> 00:00.05.00

<v PAR> So I didn’t know what we were going to do like (OL) no sé /yeah/ But there we were.

In this example, ‘yeah’ is a remark from the interviewer. Since it is a conversation between two individuals, it is not necessary to switch to a separate line and speaker code each time.

- k. Mark pauses in the following manner:
 - , Short breath (May also be used for clarity)
 - . Slightly longer breath (May also be used for clarity)
 - ... Slightly longer pause (Under 1 minute)
 - [2.0] Pause over 1 second. This example is for 2 seconds.

Do **not** spend too much time on this. Just do the best you can.

- l. Represent interjections consistently. (For example, hm, mhm, uh, ehm, aha, um.) Again, do not worry about this much. Just put what it sounds like to you.

5. WEBVTT Format:

- a. WEBVTT Format allows the transcripts to be time-aligned and clickable on the website. See examples below.
- b. Before each line of typed speech, include the time stamp (start and stop) in the format as seen above- 00:00:00.000 --> 00:00:18.000 (You may use ExpressScribe to get the timestamps).
- c. For each line of speech, you must attribute the speech to the correct speaker as seen above.

- i. For the Participant, use <v PAR>
 - ii. For the Investigator, use <v INV>
- d. See more details in Transcription Instructions on the following pages and Transcription Training materials/videos.

How to Transcribe Using Stream⁴:



NATIONAL ENDOWMENT FOR THE HUMANITIES

1. Convert your audio into a video using VLC.
 - a. Check if your device has VLC Media Player. If not, download here: <https://www.videolan.org/vlc/index.html> (You don't need to donate. It's optional.)
 - b. Click on the Media tab.
 - c. Then, click Convert/Save from the drop-down menu.
 - d. On the Open Media window, click 'Add.'
 - e. Browse to select the correct audio file.
 - f. At the bottom right side of the Window next to Cancel, there is a dropdown menu with 'Convert/Save,' click 'Convert.'
 - g. In the 'Convert' window, next to Profile, use the dropdown window to select 'Video- H.265 + MP3 (MP4).'
 - h. In the same file, next to 'Destination File' click 'Browse' to decide where you want to save the file.
 - i. Rename the file as COBIVA001_Video.
 - j. Click 'Save.'
 - k. Then, click 'Start' to start the file conversion.
 - l. See short video tutorial.
2. Go to Stream by accessing Microsoft Office 365 through your university's apps.
3. Click on the Stream App. It is a pink arrow. Note that you may have to click 'Other Apps.'
4. Click +Create.
5. Click Upload the Video file that you just created.
6. Drag the file to the screen or browse to upload the file.
7. Let the processing bar finish, and then click 'Publish.'
8. Go to 'My Content' and then 'Videos.'

⁴ This section of work was funded by the National Endowment of the Humanities:

Christoffersen, Katherine; Bessett, Ryan; & Carvalho, Ana. 2021. Bilingual Voices in the U.S./Mexico Borderlands. National Endowment of the Humanities Award. University of Texas Rio Grande Valley and University of Arizona. <https://www.neh.gov>



NATIONAL ENDOWMENT FOR THE HUMANITIES

9. Then, click on the pencil next to the video to ‘Update Video Details.’
10. Under ‘Details,’ make sure that under video language, it has the correct language selected Spanish OR English. (You have to choose one.)
11. Also, check that under ‘Options,’ ‘Autogenerate a Caption File’ has a check mark, and the box is pink to show that it is selected.
12. Under ‘Permissions,’ de-select ‘Allow everyone in your company to view this video.’
13. The video will take some time to process, but please check to try to estimate as closely as we can how long the captioning takes. Sometimes the caption file has been created before it is available to the right of the video. If there is a CC (closed captions), you will already have access to the caption file under ‘Update Video Details.’
14. When the video has processed, click on ‘My Content’ and then ‘Videos.’ Then select the video to listen to and edit the transcript that shows up in the right side ‘transcript’ box. Edit the transcript and add the participant codes. (You can copy and paste them if it is faster.) Note: Alternatively, you can try editing and adding participant codes after you download the file.
15. When the video has processed, you can click the pencil next to the video for ‘Update Video Details’ and click ‘Download file’ next to Captions under ‘Options.’ If you haven’t yet, add the participant codes and edit the transcript.
16. Once you have the transcript, follow the steps in the next section to clean up the transcription file.

Revising transcriptions from Stream with R⁵:

Step 1 & 2 (March 2021)

R is a free software environment and programming language for statistical computing and data analysis. It is also great for data manipulation, including regular expressions for identifying, extracting, and replacing patterns in texts – a convenient tool for revising transcripts!

Setup

To begin, you will need to download two things: **R** and **RStudio**.

1. [Download R](#) for your system; open the file and install

- **Mac:** click the .pkg file link under “Latest release”

R for Mac OS X

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) [here](#). Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old](#) directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

Package binaries for R versions older than 3.2.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting (<https://cran-archival.r-project.org>) accordingly.

R 4.0.3 "Bunny-Munnies Freak Out" released on 2020/10/10

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type `openssl sha1 R-4.0.3.pkg` in the *Terminal* application to print the SHA1 checksum for the R-4.0.3.pkg image. On Mac OS X 10.7 and later you can also validate the signature using `pkgutil --check-signature R-4.0.3.pkg`

Latest release:

R-4.0.3.pkg (notarized and signed)
SHA1 hash: 840209a1f1812a6c36c7b38a0e7118b1bc (ca. 85MB)

R 4.0.3 binary for macOS 10.13 (High Sierra) and higher, signed and notarized package. Contains R 4.0.3 framework, R_app GUI 1.73 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 6.7. The latter two components are optional and can be omitted when choosing "custom install". they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

Note: the use of X11 (including `tcltk`) requires [XQuartz](#) to be installed since it is no longer part of OS X. Always re-install XQuartz when upgrading your macOS to a new major version.

Important: this release uses Xcode 10.1 and GNU Fortran 8.2. If you wish to compile R packages from sources, you will need to download and GNU Fortran 8.2 - see the [tools](#) directory.

- **Windows:** click the “base” link then click the download link on the subsequent page

R for Windows

Subdirectories:

base
[contrib](#)
[old contrib](#)
[Rtools](#)

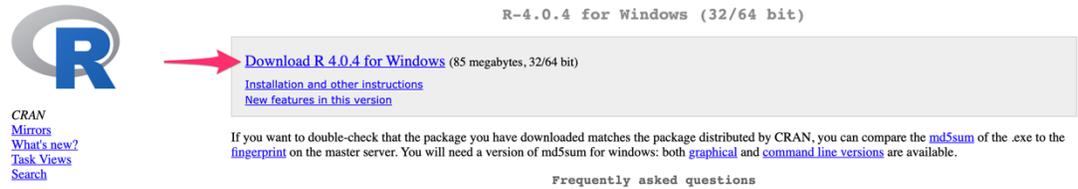
Binaries for base distribution. This is what you want to **install R for the first time**.
Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.
Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

⁵ This section of work was funded by the National Endowment of the Humanities and completed by Jessica Draper:

Draper, Jessica; Christoffersen, Katherine; Bessett, Ryan; & Carvalho, Ana. 2021. Bilingual Voices in the U.S./Mexico Borderlands. National Endowment of the Humanities Award. University of Texas Rio Grande Valley and University of Arizona. <https://www.neh.gov>



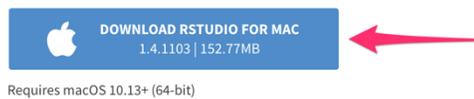
NATIONAL ENDOWMENT FOR THE HUMANITIES



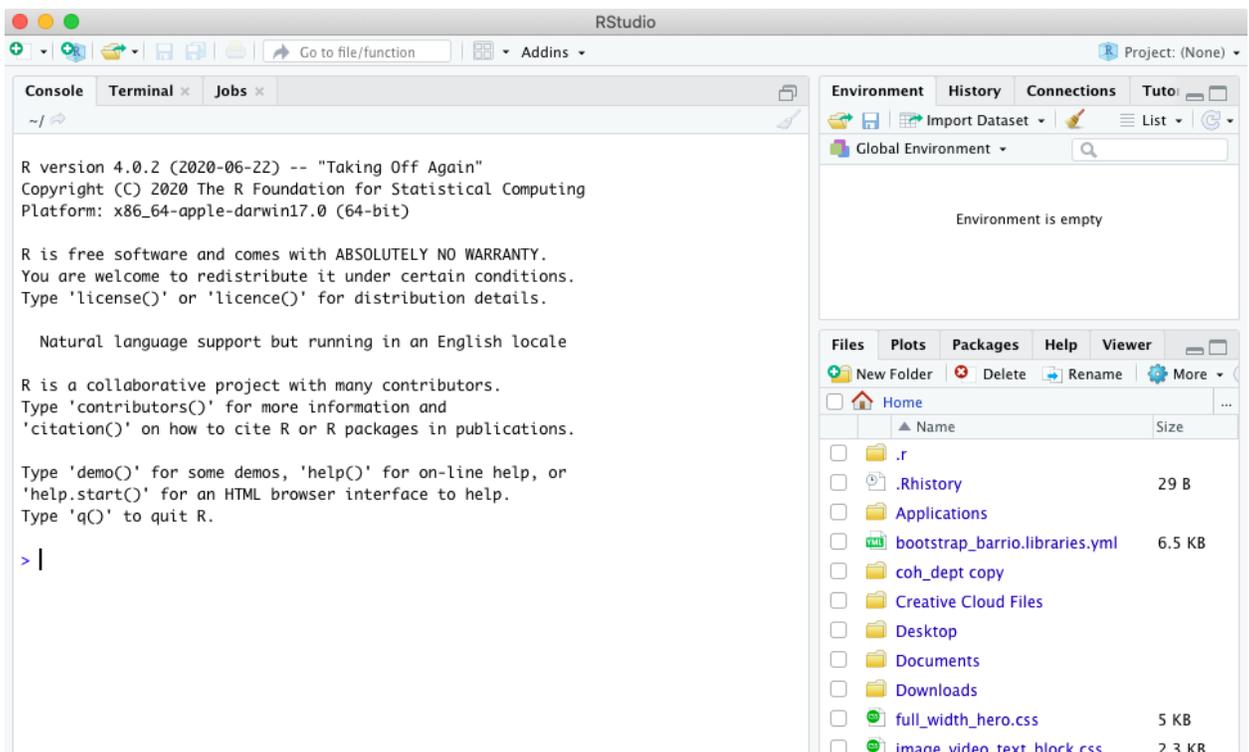
2. Once R is installed, [download RStudio](#) (Mac or Windows); open the file and install

RStudio Desktop 1.4.1103 - [Release Notes](#)

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



3. Once **RStudio** is installed, try to open the application. It should look something like this:



4. Before running any scripts, organize your transcript and .R files in your file directory in the following way:

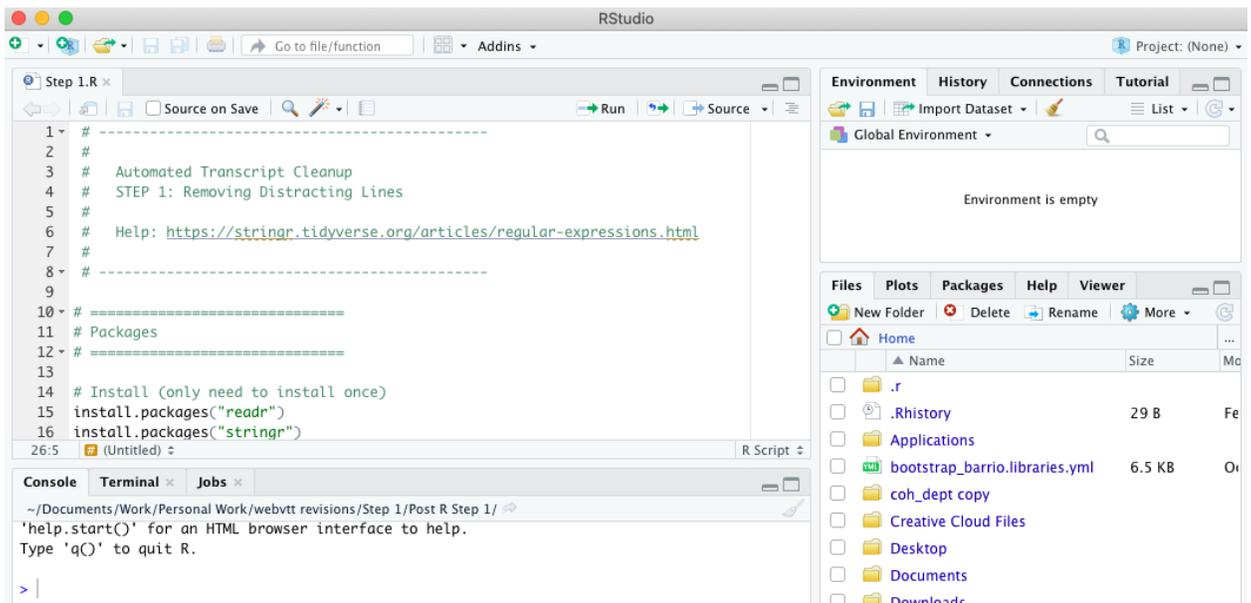
- .R and .txt files live in the same parent folder (e.g. “Documents/webvtt revisions” or along these lines)
- The .txt files live in their own “Step 1” folder
- The “Step 1” folder has an empty subfolder (e.g. “Post R Step 1” or “After Step 1”) for saving post-R revised files
- This structure is used for Step 2 as well

In the end, the folder structure should look something like:



Running Step 1.R

1. Once the .txt files have been added to the Step 1 folder, open **Step 1.R** in RStudio. It should look something like:



2. What you see in the Step 1.R file is the R programming language. Green text indicates comments, not programming code. Comments are skipped over when the code is run. They are created using the preceding “#”

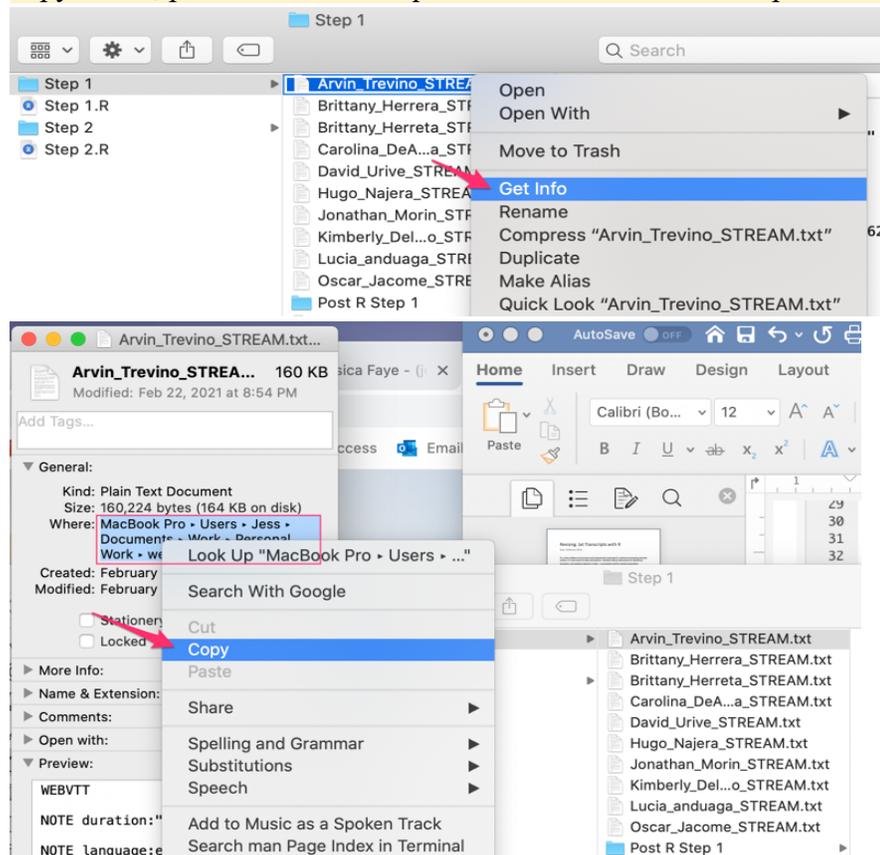
3. Before running this code:

- a. Update the “working directory” path on line 29 to reflect your local file location where your Step 1 .txt files are stored (*path should end with “Step 1”*)

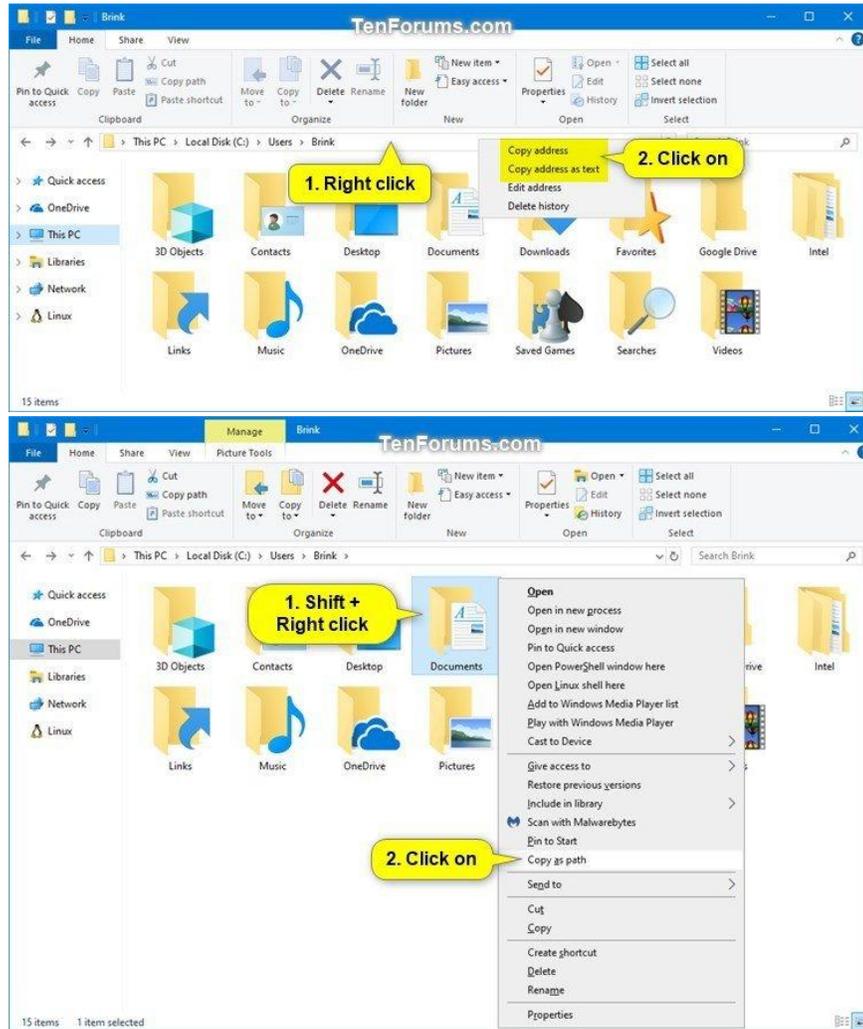
```

23 # =====
24 # Begin Editing
25 # (highlight from HERE to end)
26 # =====
27
28 # (!) YOUR working directory (wd):
29 your_wd <- "/Users/Jess/Documents/Work/Personal Work/webvtt revisions/Step 1"
30
31 setwd(your_wd)
32 getwd() # Confirm correct wd
33
    
```

- i. To find this location on a **Mac**, right-click on a .txt file within the Step 1 folder and select “Get Info.” Then highlight the “Where” text and click copy. Then, paste between the quotation marks in the R script on line 29.



- ii. To find this location in **Windows**, open up the Step 1 folder in File Explorer. In the address bar above, right click and select “Copy address” **OR** press shift and then right click on a .txt file and select “Copy as path.” Paste between the quotation marks in the R script on line 29. Then, **change all backslashes (\) to forward slashes (/)**.



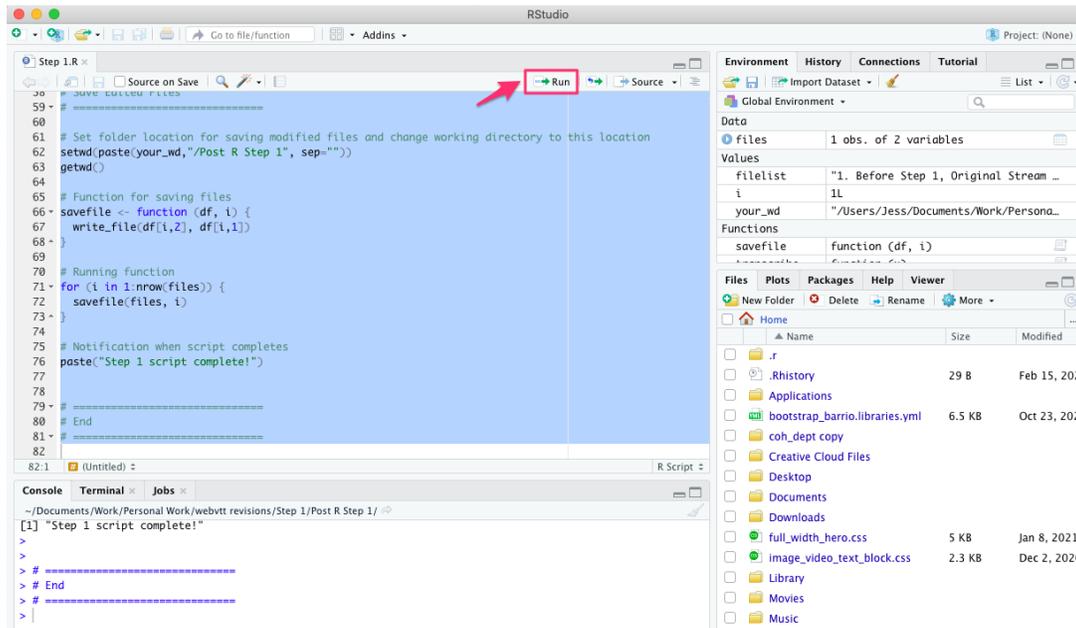
- b. On line 62, the code uses your working directory location and adds a subfolder to indicate where the new revised transcripts should be saved. Make sure this subfolder name matches the one you created earlier (e.g. “Post R Step 1”)

```

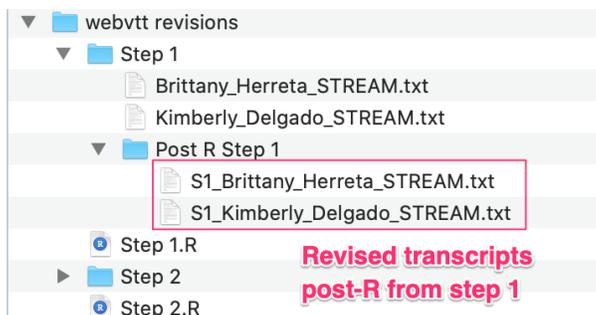
56
57 # =====
58 # Save Edited Files
59 # =====
60
61 # Set folder location for saving modified files and change working directory to this lo
62 setwd(paste(your_wd, "/Post R Step 1", sep=""))
63 getwd()
64

```

- Now you are ready to run the R script! Highlight the entire code and click “Run” in the top right. You’ll know it’s finished when you see “End” in the bottom console area.



- Now open the “Post R Step 1” folder from your file directory. It should now include all .txt files from its parent folder but renamed with the prefix “S1_”. When you open these files, the “NOTE Confidence” lines and random letters/numbers should be removed.



- Save the .R file with your updated working directory information (File > Save). You won’t have to update the file directories again for this step.
 - Optionally, before saving, you can comment out lines 15-16 by adding a “#” before the line, making it green. This means the code will skip over the install commands next time the code is run, shortening the time it takes to run the script. Once these packages have been installed the first time, you typically don’t need to

install them again (unless you get an error telling you otherwise). Going forward, you only need to highlight from line 23 down before running the code.

<pre> 10 # ===== 11 # Packages 12 # ===== 13 14 # Install (only need to install once) 15 install.packages("readr") 16 install.packages("stringr") 17 </pre>	<pre> 10 # ===== 11 # Packages 12 # ===== 13 14 # Install (only need to install once) 15 #install.packages("readr") 16 #install.packages("stringr") 17 Commented out, no longer part of code </pre>
---	--

7. Repeat this process as necessary!

Running Step 2.R

1. Once the speaker tags have been added to the .txt files, add them to the Step 2 folder.
 - a. Be sure to review these text files to make sure there are no obvious errors or issues; the code should catch most formatting issues, but there may be some it could miss if it's not setup in this general format:

```
00:00:10.050 --> 00:00:10.920
<v INV> Example text of a single line
```

```
00:00:12.980 --> 00:00:17.930
<v PAR> Example text of a line
with a break. This is okay, the code
will fix this to remove line breaks.
```

```
00:00:17.930 --> 00:00:22.880
continuing example text from the line above; the code
will collapse timestamps within a speaker's dialogue
```

2. Open **Step 1.R** in RStudio

3. Before running this code:

- a. Just like in Step 1, update the “working directory” path on line 31 to reflect your local file location where your Step 2 .txt files are stored (*path should end with “Step 2”*). See Step 1 again for more detail on how to find the path.
- b. On line 162, the code uses your working directory location and adds a subfolder to indicate where the new revised **Step 2** transcripts should be saved. Make sure this subfolder name matches the one you created earlier (e.g. “Post R **Step 2**”)

4. Now you are ready to run the R script! Highlight the entire code and click “Run” in the top right. You’ll know it’s finished when you see “End” in the bottom console area.
5. Now open the “Post R Step 2” folder from your file directory. It should now include all .txt files from its parent folder but renamed with the prefix “S2_”.
6. **When you open these S2_.txt files, speaker dialogues should be collapsed within a single timestamp, alternating between speakers and the text should wrap; no “00:00:000” timestamps should exist.**

If you see a problem, review the original .txt files to make sure there is nothing in there that might be breaking the code from running smoothly.

For example, if you do see “00:00:000” timestamps throughout the text in the resulting S2_.txt files, this means there is a formatting issue in at least one of the original .txt files that the code ran through in the Step 2 folder. If even just one file has a formatting issue, it will break the timestamp function code for all files in the Step 2 folder.

To find the file with the formatting issue, after running the code in R, scroll up in the bottom console area until you see **red error text**. It will not tell you specifically which file the error is in, but it should include a partial text sample that you can use as a clue to find the .txt file that is causing trouble.

```

178
179 # =====
180 # End
181 # =====
182
182:1 # (Untitled)
R Script

Console Terminal Jobs
~/Documents/Work/Personal Work/webvtt revisions/Step 2/Post R Step 2/
>
> # Running function over text column of "files" data frame
> files$Text <- mapply(assign_timestamps,files[,2],files[,3])

Error in `<-data.frame`(`*tmp*`, "text", value = c("\r\n<v INT> Ok awesome I think we're recording. Um, Ok I just want to say thank you for doing this because it's quite an interesting time and its really hard to get people to do this type of stuff. Thank you, uh ok so the first question gonna start with something simple, general, and hopefully, you know, see where that goes. /OK/ OK so, where were you born?\r\n\r\n", : replacement has 24 rows, data has 26
Show Traceback
Rerun with Debug

>
> # =====
> # Save Edited Files

```

Once you’ve identified the correct .txt file, review the entire file for any formatting issues. It will likely be related to how the timestamps are formatted (e.g. an extra “-->” or odd spacing or characters that deviate from the desired pattern). When found, fix the issue(s) manually, save the file, and run the R script again. Check the console for any further errors and, if present, repeat this process again.

7. When finished running the code, save the .R file with your updated working directory information (File > Save). You won't have to update the file directories again for this step.
 - a. Optionally, just like in Step 1, before saving, you can comment out lines 15-17 by adding a “#” before the line, making it green. This means the code will skip over the install commands next time the code is run, shortening the time it takes to run the script.
8. Repeat this process as necessary!