

INTRODUCTION

- Completed transcriptions and analyses of interview data from a study that I conducted in Seville, Spain in 2019.
- Chose to focus on these data because I had not yet touched them at the start of the semester.



ORIGINAL STUDY: 4 TASKS

Dialect mapping/language attitude evaluation

Semantic association task

False memory task

Brief sociolinguistic interview ** (data come from this)

VARIABLE OF STUDY: /S/ PRODUCTION IN WORD ONSET

	ALLOPHONE	PRESTIGE	FAMILIARITY
MADRID	[s]	National standard, overt prestige	Media/edu
SEVILLE CAPITAL	[s]	Regional standard, covert prestige	Local
SEVILLE OUTSKIRTS	$[\S^{\theta}]$	Stigmatized	Local

(Hernández-Campoy & Villena-Ponsoda, 2009; Regan, 2017; Penny, 2004)

RESEARCH QUESTIONS (TASK 4)

- 1. What are the general perspectives from Seville capital listeners about their own variety?
- 2. What kind of feedback do they receive from outside sources (other Spaniards, the media) about their variety?
- 3. Are the stereotypes about their variety changing or not?

Almost all born and raised in Seville capital

Age range: (coming soon ©)

Gender distribution: (coming soon ©)

5

METHODOLOGY: INTERVIEW QUESTIONS

<i>1</i> .	¿De dónde eres? ¿Has vivido siempre en España?	Where are you from? Have you always lived in Spain?
1.	¿Qué opinas de la variedad lingüística que existe dentro de Andalucía?	What is your opinion of the linguistic variety that exists within Andalusia?
1.	¿Has viajado fuera de Andalucía pero dentro de España? Cuando sales de la zona, ¿recibes algún comentario sobre tu acento?	Have you traveled outside of Andalusia, but within Spain? When you leave the region, do you receive any comments about your accent?
1.	¿Hay representación del acento andaluz en los medios de comunicación, las películas, etc?	Is there representation of the Andalusian accent in the media, movies, etc.?
1.	¿Crees que los estereotipos sobre los acentos están cambiando ahora, o que permanecen todavía?	Do you think that stereotypes about accents are changing now, or that they still remain?

DATA: TRANSCRIPTIONS

- Typed all transcriptions as .txt files while listening to the sound files
- All data are in Spanish
- Only included utterances produced by participants
- Files were saved as the participant code_Q1-5
- Broad transcriptions- did not include repetitions unless meaningful, restarts, fillers, etc.



865_Q3.txt

Dentro de España.

Dentro de España he estado en Madrid, he estado en Extremadura, he estado en Cataluña en Barcelona.

Y ya está, dentro de España no me he movido mucho, la verdad.

Pero me encantaría, porque tanto el euskera como el catalán me parecen dialectos preciosos y los acentos del centro de España son un poco para mí más chocante.

Pero me encanta la variedad lingüística que tenemos.

Para eso sí en casa en ningún sitio.

Sí, normalmente suele ser un acento que hace gracia, que parece cálido, siempre parece amable.

Y también depende de la zona donde te vayas y los perjuicios que encuentres pues te pueden calificar de lo que quieran calificarte, simplemente porque tienes un acento del sur más acentuado o menos.

Pero realmente se me reconoce por mi acento.

Lo que también se me confunde mucho son con acentos latinos.

Creo que el sur y las islas Canarias se acercan mucho más a lo latino.

A mí no sé por qué pero siempre me suelen asociar las islas Canarias.

Pero mi acento es muy sevillano.

Creo, vaya.

8

DATA: ORGANIZATION

- Created one large data frame with all transcriptions
- Made columns for 'word', 'line number', 'file name', 'participant', and 'question'
- Wanted to be able to analyze full data set, but also subset according to the question being answered
- Used unnest_tokens to put one word/row
- n=6,187 words
- head () results

	word <chr></chr>	line.number <int></int>	fname <chr></chr>	participant <chr></chr>	question <chr></chr>
1	de	1	./002_Q1.txt	002	Q1
2	sevilla	1	./002_Q1.txt	002	Q1
3	SÍ	2	./002_Q1.txt	002	Q1
4	opino	3	./002_Q2.txt	002	Q2
5	que	3	./002_Q2.txt	002	Q2
6	cuanto	3	./002_Q2.txt	002	Q2

6 rows

FREQUENCY RESULTS: ALL DATA

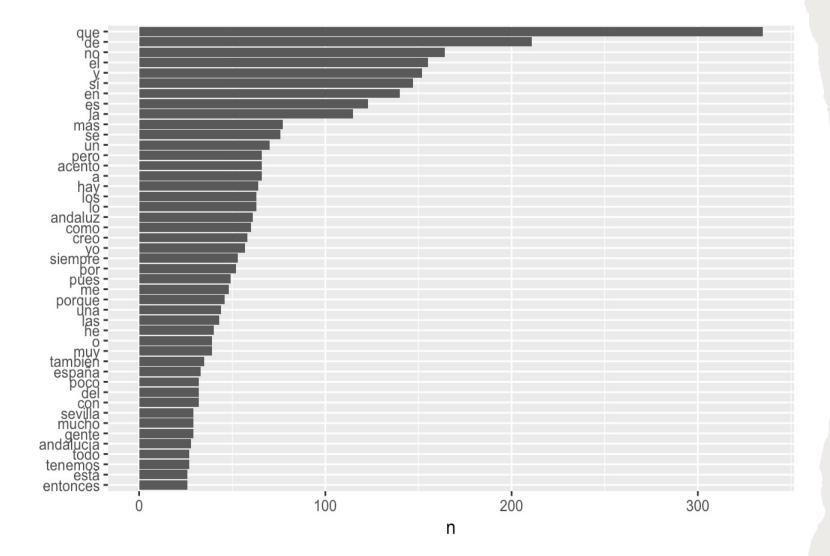
- Used count() to get frequency results
- Not very informativemost are function words!

Description: df [1,075 × 2]		
word <chr></chr>	n <int></int>	
que	335	
de	211	
no	164	
el	155	
у	152	
sí	147	
en	140	
es	123	
la	115	
más	77	

10

FREQUENCY RESULTS: ALL DATA

- More
 visualization here,
 similar results
- filter (n > 25)
- Need to remove "stop words" for more accurate analysis



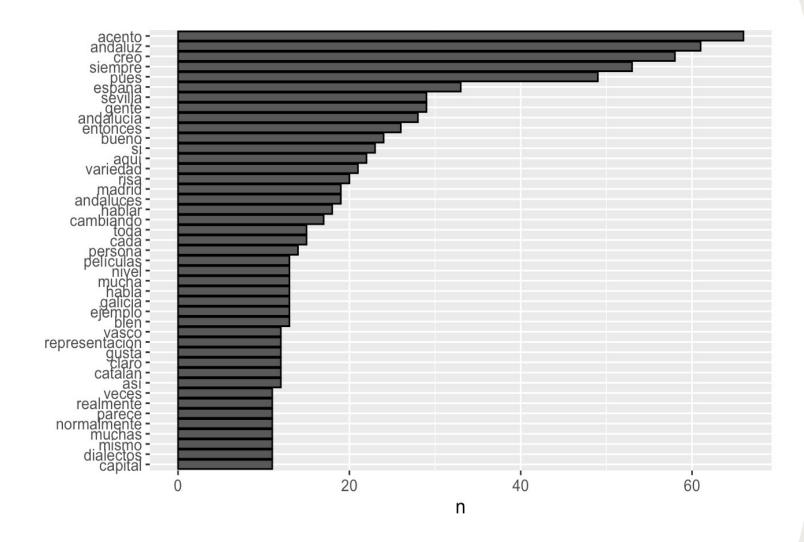
FREQUENCY RESULTS: STOP WORDS

- Used "tm" library for Spanish stop words
- 300 items
- Articles, prepositions, pronouns, a few highly frequent verbs
- Created a new data frame without these words

[[1]]				
[1]	"de"	"la"	"que"	
[6]	"y"	"a"	"los"	
[11]	"las"	"por"	"un"	
[16]	"no"	"una"	"su"	
[21]	"como"	"más"	"pero"	
[26]	"ya"	"o"	"este"	
[31]	"esta"	"entre"	"cuando"	
[36]	"sobre"	"también"	"me"	
[41]	"donde"	"quien"	"desde"	11
[46]	"durante"	"todos"	"uno"	
[51]	"contra"	"otros"	"ese"	
[56]	"ellos"	"e"	"esto"	
[61]	"algunos"	"qué"	"unos"	
[66]	"otras"	"otra"	"él"	
[71]	"estos"	"mucho"	"quienes"	
[76]	"cual"	"poco"	"ella"	
[81]	"algunas"	"algo"	"nosotros"	
[86]	"tú"	"te"	"ti"	
[91]	"ellas"	"nosotras"	"vosotros"	
[96]	"mío"	"mía"	"míos"	

OVERALL FREQUENCY RESULTS: NO STOP WORDS

- Better, more content words here
- Acento, or
 "accent" is the
 most frequent,
 followed by
 andaluz, referring
 to Andalusian
 variety
- Could have removed a few other words



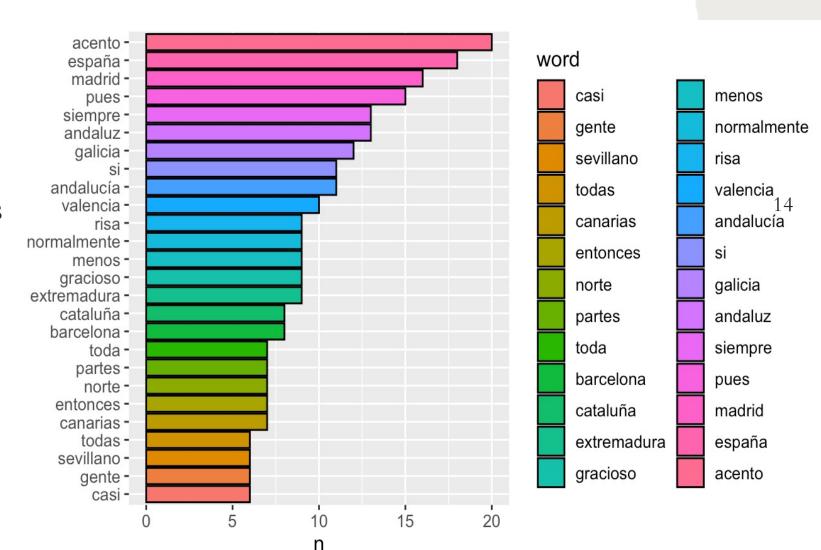
OVERALL FREQUENCY NO STOP WORDS

- Nice visualization options for frequency data
- library(wordcloud)
- Sentiment analysis
- Categorizes items as "positive" or "negative"
- https://www.jorgelopezperez.com/
 posts/analisis-de-texto-con-r-
 jugando-un-poco-con-tidytext/



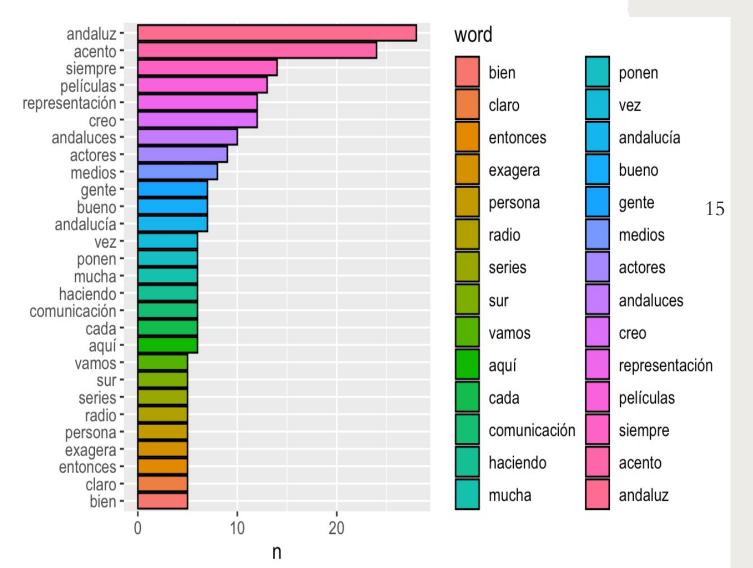
FREQUENCY BY QUESTION (Q3)

- Question about traveling within Spain/ receiving comments when leaving Andalusia
- Lots of region/city names
- Word gracioso,
 "fun/funny" was used frequently (n=9)
- Speakers felt that the tone is important with this use



FREQUENCY BY QUESTION (Q4)

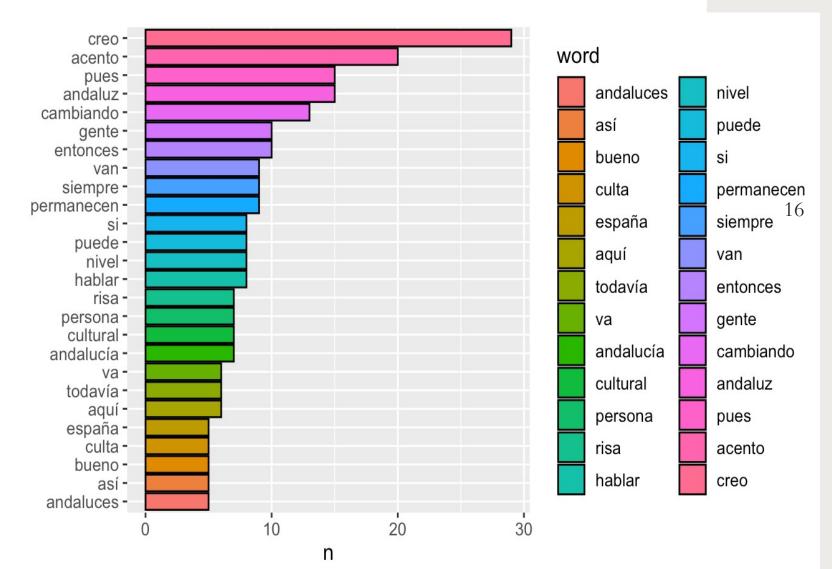
- Question about media representation
- Would need a full sentence to be able to code for the overall sentiment
- exagera, "exaggerate" was used 5 times
- *imita*, "imitate" was also used, though it is not represented here



FREQUENCY BY QUESTION (Q5)

- Question about whether the stereotypes about the accent are changing
- Cambiando,

 "changing" is more
 frequent than
 permanecen, "they
 remain"
- Unclear since no was removed as a stop word



LIMITATIONS

- "no" was included as a stop word, which would be important to have in a keyword analysis.
- Frequency analyses good for looking at overall trends, but it can miss less frequent, yet perhaps more salient, items
- *payaso*, "clown" lower frequency, but very impactful in context
- De Despeñaperros pa'rriba/hacia abajo: expression was used a few times, but never made higher frequency cutoffs

"Un andaluz sale de Andalucía y es un cateto, contamos chistes, y somos los payasos..."

"A person from Andalusia leaves the region and is a hick, we tell jokes and we're the clowns..."



