Part 2 Social media research

Kyungjin Hwang

Korea University, Department of Spanish Language and Literature

Index

- What is Social Media?
- Language of Social Media
- Social Media Research Methodology
- Social Media Research with Phonology

Chapter 1

What is social media?

What is social media?

What is social media?

Definition

- Social media is a collective term for websites and applications that focus on communication, community-based input, interaction, content-sharing and collaboration.

- The social web are popularized terms used to signal a shift toward the internet as an interpersonal resource rather than solely an informational network.

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

What is social media?

Types of social media

- Its centre is ‘user-generated content‘.

- Types of social media

    a. Self-publication by users of multimedia content such as blogs (websites displaying entries in reverse chronological order),
    b. Vlogs (video blogs such as those posted regularly by millions of users on YouTube)
    c. Microblogs (streams of small character-constrained posts).

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

What is social media?

Microblogging

- Users can post their opinion with limited characters and attach various media files, hashtags, photos, and links.

- For example, on Twitter, the timeline of users allows them to see the tweets published by the accounts they are following and, through the search of some words or hashtag, they see the tweets in real time published by people from all over the world.

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

What is social media?

Characteristics of different social media platforms

	Twitter	Facebook	Instagram	Youtube
Main Contents	Mainly words	Mainly words	Fotos	Videos
Direction	Unidirectional	Unidirectional	Unidirectional	Unidirectional
Communication	Possible (retweets, likes, mentions)	Possible (share, comments)	Possible (likes, comments)	Possible (comments)
Private Use	Possible but you can see other contents by searching	Possible	Possible	Almost impossible
Characteristics	real-time	group	hashtags	income
Number of users (million)[1]	397	2,853	1,386	2,291

1. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

What is social media?

What we can investigate through social media as linguists

- Social Linguistics
  - code-switching
  - identity with language
    - ex) Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach
- Syntax
  - What are the syntactic characteristics of language used on social media?
- Second Language Acquisition
  - How can social media be applied to foreign language learning?
    - The effect of using WhatsApp messenger in learning English language among university students
- Phonology
  - What are the phonological features that appear on social media?

Chapter 2

Language of social media

Language of social media

Characteristics

- Because of the character limitations imposed on microposts, many people consider how to effectively communicate their opinions or thoughts in constrained environments.
- Unlike everyday language, people can emphasize or express their opinion more efficiently through photos, quotes, likes, and hashtags.
- As opinions are expressed quickly and instantly through cell phones, many non-grammatical expressions, abbreviations, memes and typos appear.

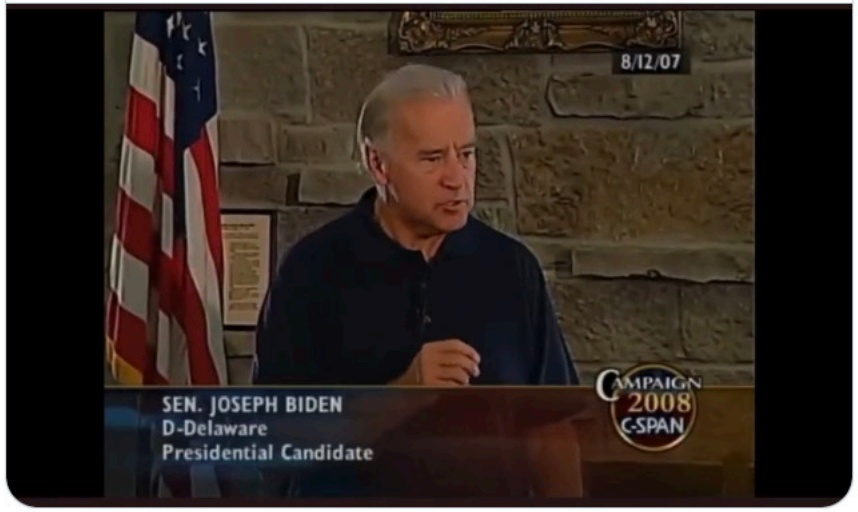
Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

Language of social media

Ted Cruz @tedcruz · 13h

.@JoeBiden is right.

That's exactly what he did.

> **Senator Melissa Melendez** @sen... · 2d
>
> Well isn't this interesting.
>
> 
>

melanycicenia
Disneyland California

Liked by **natalycicenia** and **59 others**

melanycicenia Disney, where dreams become reality💫🤍

#disneyland #california #la #happyday #blessed #travelphotography #picoftheday #tuesday #tuesdayvibes #anaheim #picoftheday

Language of social media

Specific use of language in social media

1) Express personal thoughts and feelings with emotional language
   - Expression of opinion through expression of consent
   - Emphasize emotions through emoticons, hashtags, etc.
   - Tends to use more emotional words than everyday conversations
2) Complain about their everyday existence
   - More freely than in everyday language
   - Using negative language directly to the target of criticism
3) Contribute to a micro-meme

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

Language of social media

Specific use of language in social media

4) Engage in humour

- There is a trend of humor used exclusively on social media.
- It is mainly used for linguistic play.

5) Express political opinion

- More freely than in everyday language
- A lot of political debate, criticism, sarcasm, meme…

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

Language of social media

How we can decide the topics that we will investigate

- Cueva, D. S. (2014). El Code Switching en las redes sociales: La expansión de lengua, cultura e identidad (Doctoral dissertation, State University of New York at Stony Brook).
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one , 8 (9), e73791.
- Lee, C., & Chau, D. (2018). Language as pride, love, and hate: Archiving emotions through multilingual Instagram hashtags. Discourse, Context & Media, 22, 21-29.
- Ta'amneh, M. A. A. A. (2017). The effect of using WhatsApp messenger in learning English language among university students. International Research in Education , 5 (1), 143-151.

Chapter 3

Social media research methodology

Social media research methodology

1. Using researcher’s account

- This means using your own accounts in research
- Advantages: When doing sociolinguistic research, you can easily study interactions online. It is easy to understand the sociolinguistic background of the research subjects.
- Disadvantages: The subjectivity of the researcher can be heavily involved, and the researcher's background or linguistic habits can greatly affect the research results.

Social media research methodology

1. Using researcher's account

- [Undergraduate report assignment]

  Analysis of the use of emoticons by Spanish people

- Methodology

  Using WhatsApp, the pattern of emoticon usage was analyzed through conversations with Spanish-speaking friends. In addition, using Facebook, I analyzed the pattern of emoticon usage by watching the writing and commenting of Spanish-speaking friends that I knew.

Social media research methodology

2. Using corpus

- Studying with corpus that people have already created.
- ex) HERMES corpus for Twitter, Tweets 2011

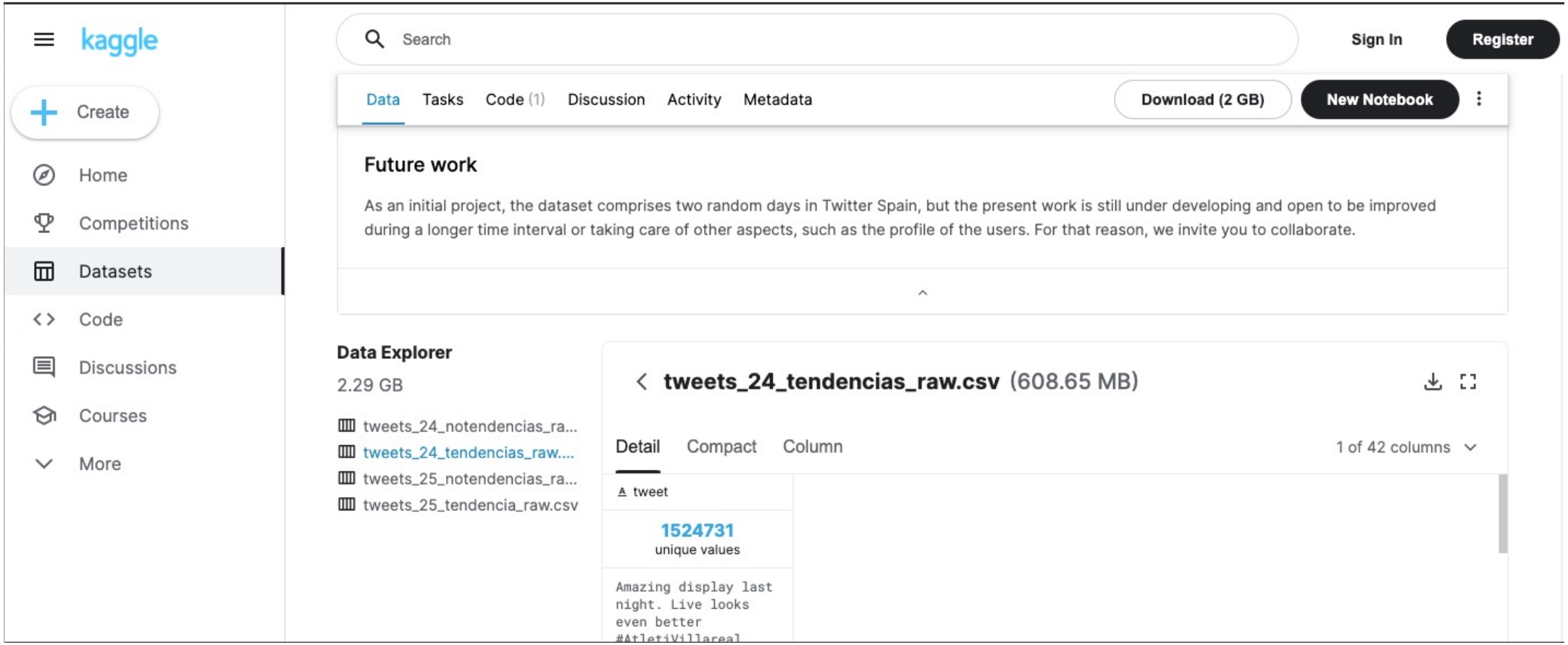
- Advantage: You can easily get a large amount of data.

- Disadvantages: There are limitations when you want to study online language use for a specific topic, or when a researcher wants to establish a specific time period, specific scope, specific region, etc.

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* (Vol. 6). A&C Black.

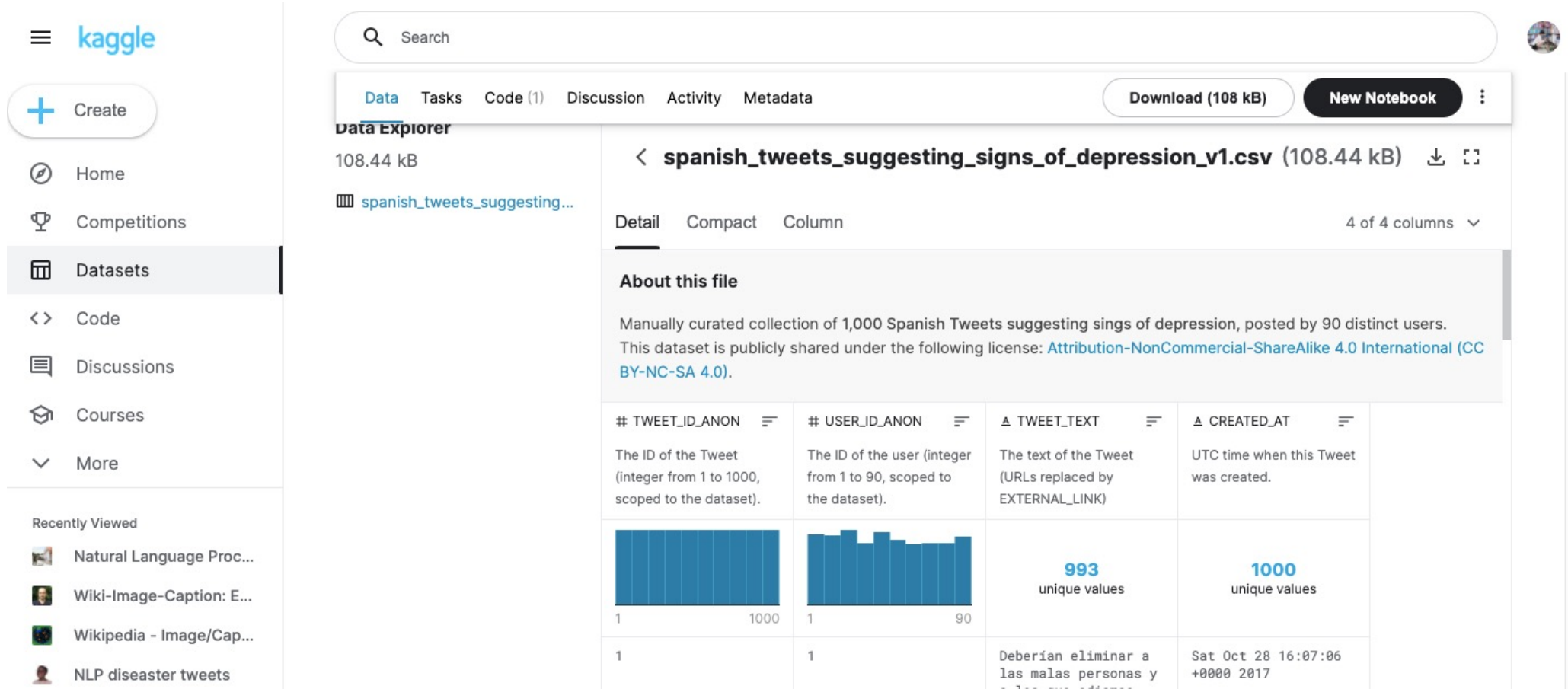
Social media research methodology

2. Using corpus https://www.kaggle.com/pilarlc/tweets-trends-and-no-trends?select=tweets_24_tendencias_raw.csv



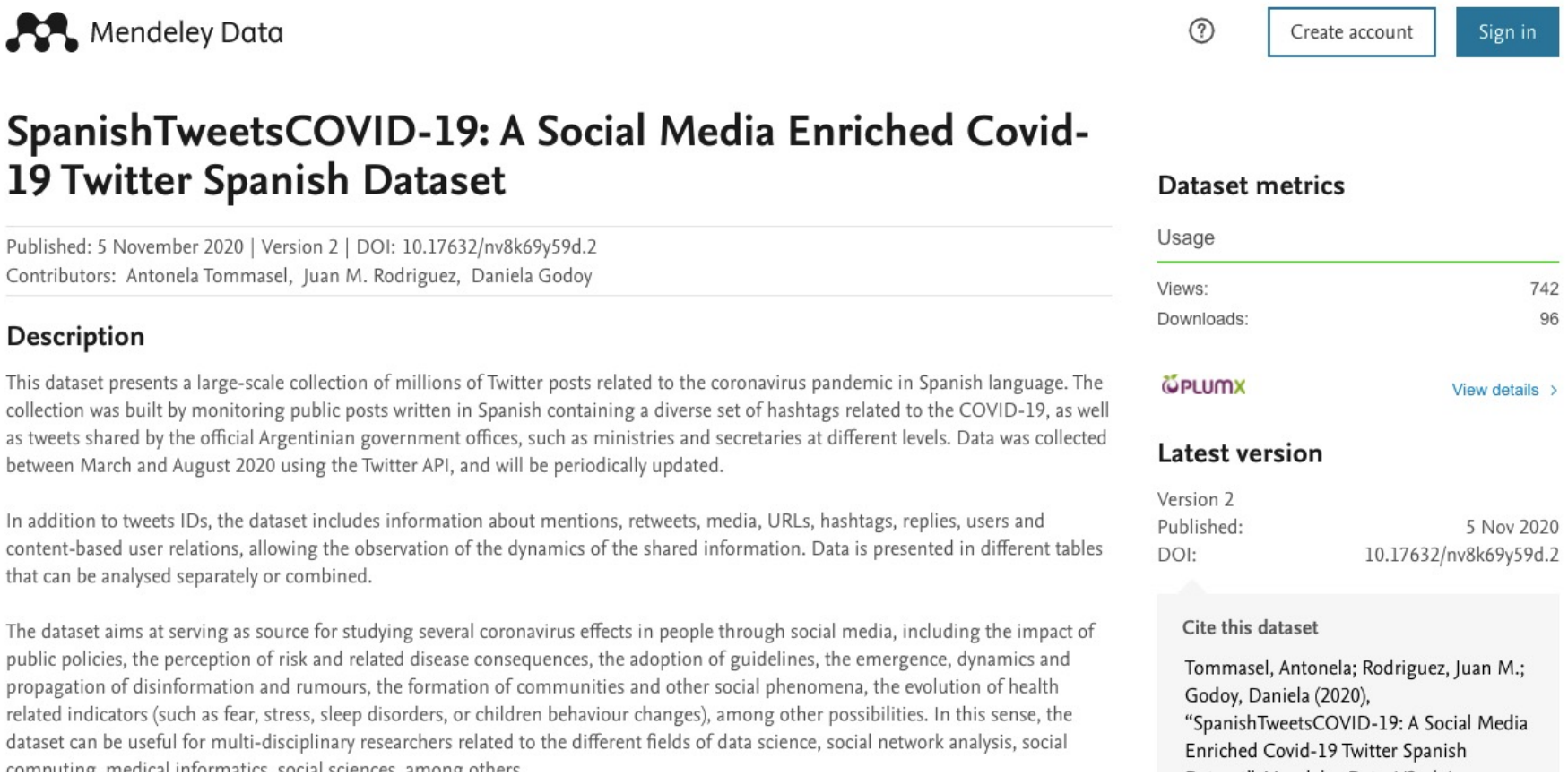
Social media research methodology

2. Using corpus https://www.kaggle.com/francescoronzano/spanish-tweets-suggesting-depression/version/1



Social media research methodology

2. Using corpus https://data.mendeley.com/datasets/nv8k69y59d/2



Mendeley Data

Create account Sign in

SpanishTweetsCOVID-19: A Social Media Enriched Covid-19 Twitter Spanish Dataset

Published: 5 November 2020 | Version 2 | DOI: 10.17632/nv8k69y59d.2
Contributors: Antonela Tommasel, Juan M. Rodriguez, Daniela Godoy

Description

This dataset presents a large-scale collection of millions of Twitter posts related to the coronavirus pandemic in Spanish language. The collection was built by monitoring public posts written in Spanish containing a diverse set of hashtags related to the COVID-19, as well as tweets shared by the official Argentinian government offices, such as ministries and secretaries at different levels. Data was collected between March and August 2020 using the Twitter API, and will be periodically updated.

In addition to tweets IDs, the dataset includes information about mentions, retweets, media, URLs, hashtags, replies, users and content-based user relations, allowing the observation of the dynamics of the shared information. Data is presented in different tables that can be analysed separately or combined.

The dataset aims at serving as source for studying several coronavirus effects in people through social media, including the impact of public policies, the perception of risk and related disease consequences, the adoption of guidelines, the emergence, dynamics and propagation of disinformation and rumours, the formation of communities and other social phenomena, the evolution of health related indicators (such as fear, stress, sleep disorders, or children behaviour changes), among other possibilities. In this sense, the dataset can be useful for multi-disciplinary researchers related to the different fields of data science, social network analysis, social computing, medical informatics, social sciences, among others

Dataset metrics

Usage

Views:	742
Downloads:	96

PLUMX View details >

Latest version

Version 2

Published:	5 Nov 2020
DOI:	10.17632/nv8k69y59d.2

Cite this dataset

Tommasel, Antonela; Rodriguez, Juan M.; Godoy, Daniela (2020), "SpanishTweetsCOVID-19: A Social Media Enriched Covid-19 Twitter Spanish

Social media research methodology

2. Using corpus

Tweets2011

https://trec.nist.gov/data/tweets/

As part of the TREC 2011 microblog track, Twitter provided identifiers for approximately 16 million tweets sampled between January 23rd and February 8th, 2011. The corpus is designed to be a reusable, representative sample of the twittersphere – i.e. both important and spam tweets are included.

The Tweets2011 corpus is unusual in that what you get is a list of tweet identifiers, and the actual tweets are downloaded directly from Twitter, using the open-source twitter-tools. However, to obtain the lists of tweets to be downloaded (i.e. the "tweet lists"), a data usage agreement must be signed. Once signed, the agreement must be emailed back to NIST, who will provide you with a username/password to download the tweet lists (in the form of a .tar.gz file).

Obtaining the collection

Download and sign the TREC 2011 Microblog Dataset Usage Agreement. Please note that this agreement requires you to also act within the terms of the Twitter terms of service, and in particular you agree not to redistribute the data and to delete tweets that are marked deleted in the future. The twitter-tools provides support for removing deleted tweets from your copy of the corpus.

Email the signed agreement, as a PDF file, to Angela Ellis <angela.ellis@nist.gov>. In the body of your email,

1. Be clear that you are requesting the Tweets2011 dataset
2. Include your name,
3. your email address, and
4. the name of your organization.

We will respond to your request with a URL, a username, and a password with which you can download the tweet lists. **Please allow seven business days for a response**.

Once you have downloaded and decompressed the tweet lists from NIST, you should obtain and run the corpus downloader. For further instructions on downloading and using the twitter-tools corpus downloader, see twitter-tools.

You **MUST NOT** re-distribute the tweet lists or the corpus obtained by using the tweet lists, as this breaks both the Tweets2011 corpus license agreement and the Twitter Terms of Use. Note that it can take several days to download your copy of the Tweets2011.

Social media research methodology

3. Creating new account for research

- A researcher creates an account for research and uses that account for research.
- Advantages: Researchers can choose their own research subjects and can easily collect linguistic data on specific topics.
- Disadvantages: It can take a lot of time for researchers to create accounts, find research subjects, and collect research subject language data.

Cueva, Daniel Stephan(2014), “El Code Switching en las redes sociales: La expansión de lengua, cultura e identidad”, Ph. D. Dissertation. State University of New York.

Social media research methodology

3. Creating new account for research

“

Los tipos de datos que fueron recolectados se clasificaron de diversas maneras.

Primero, fue buscar participantes en mis cuentas de Facebook, Twitter, e Instagram donde se viera movimientos de intercambio entre lenguas con otras personas y donde existiera una comunicación entre los participantes internos y externos.

Participantes internos son aquellos que forman parte de mis “amistades” en mis redes sociales. Por otro lado, participantes externos son aquellos que son “amistades” de mis amistades pero no forman parte de mi grupo interno.

”

Cueva, Daniel Stephan(2014), “El Code Switching en las redes sociales: La expansión de lengua, cultura e identidad”, Ph. D. Dissertation. State University of New York.

Social media research methodology

3. Creating new account for research

“

Segundo, fue buscar variedad en nacionalidad, es decir, participantes que sean de diferente países.

Tercero, decidí tener a participantes que no sobre salgan del rango de edad, es decir, quería que estén entre los 20 hasta los 26 años de edad. La razón que decidí mantener este rango de edades es porque presiento que el centro donde se genera el CS proviene en los jóvenes ya que son ellos los que más frecuencia están en las redes sociales publicando y compartiendo sus intereses.

Cuarto, decidí buscar participantes externos entre mis participantes internos donde el apellido muestre distinción de cultura, es decir, que el apellido muestre si son anglohablantes e hispanos.

”

Cueva, Daniel Stephan(2014), “El Code Switching en las redes sociales: La expansión de lengua, cultura e identidad”, Ph. D. Dissertation. State University of New York.

Social media research methodology

4. Crawling with python

- It is to use a crawling method to scrape data using a programming tool such as Python.
- Advantages: You can easily set a specific topic, a specific region to collect a lot of data in a short time.
- Disadvantages: It is difficult to know the background of the study subjects.

Kyungjin Hwang. (2021). El cambio de código entre español e inglés observado en Twitter. Asociación Coreana de Hispanistas, 98, 95-118.
2 Kyungjin Hwang. (2021). Expresiones emociales en español en línea, a través de análisis de sentimiento con Machine Learning, XCongreso Asiático de Hispanistas

Social media research methodology

4. Crawling with python

Practice!

Kyungjin Hwang. (2021). El cambio de código entre español e inglés observado en Twitter. Asociación Coreana de Hispanistas, 98, 95-118.
2 Kyungjin Hwang. (2021). Expresiones emociales en español en línea, a través de análisis de sentimiento con Machine Learning, XCongreso Asiático de Hispanistas

Social media research methodology

4. Crawling with python

Step 1. Register as developer in Twitter and Youtube

[Twitter]

Guide: https://dev.to/sumedhpatkar/beginners-guide-how-to-apply-for-a-twitter-developer-account-1kh7

Website: https://developer.twitter.com/en/apply-for-access

[Youtube]

Guide: https://github.com/Kyungjin-Hwang/211117UCLALecture

Kyungjin Hwang. (2021). El cambio de código entre español e inglés observado en Twitter. Asociación Coreana de Hispanistas, 98, 95-118.
2 Kyungjin Hwang. (2021). Expresiones emociales en español en línea, a través de análisis de sentimiento con Machine Learning, XCongreso Asiático de Hispanistas

Social media research methodology

4. Crawling with python

Step 2. Writing Code

https://colab.research.google.com/

Step 3. Crawling and sava data

Step 4. Analyze

Kyungjin Hwang. (2021). El cambio de código entre español e inglés observado en Twitter. Asociación Coreana de Hispanistas, 98, 95-118.
2 Kyungjin Hwang. (2021). Expresiones emociales en español en línea, a través de análisis de sentimiento con Machine Learning, XCongreso Asiático de Hispanistas

Social media research methodology

How we investigate social media language

- First, set the subject, scope, and topic you want to study.
  Case 1)
  Topic: Expression of sentiments by Spanish speakers in Spain
  Subjects: Youtube comments written by Spanish speakers.

  Case 2)
  Topic: Express identity by bilingual speakers of Spanish and English
  Subjects: People who speak mainly Spanish but also use English

- Second, choose a research method that suits the above.
  Case 1)
  Python Crawling

  Case 2)
  Create account for research

Chapter 4

Social media research methodology with phonology

Social media research with phonology

What we can analyze

- Phonemic confusion

  seseo (s/z/c+e,i: e.g., empesar instead of empezar), yeísmo (ll/y: llendo instead of yendo), <b> vs. <v> (e.g., vendito instead of bendito, confusion between haber vs. a ver), <g>+e,i vs. <j> (e.g., jelatina instead of gelatina), <l> vs. <r> in Puerto Rican Spanish (e.g., veldá instead of verdad).

- Simplification

  exclusion of silent sound <u> in qu+e,i by converting <q> to <k> (e.g., ke pasa, kiero)

- Phonetic strengthening

  <we> or <güe> instead of <hue> (e.g., webo instead of huevo, güérfano instead of huérfano)

- Weakening

  <hue> or <we> instead of <bue> (e.g., hierbahuena instead of hierbabuena)

- Influence of English

Social media research with phonology (in-progress)

Topic

A comparative study of Spanish vowel and syllable reduction on social media in Spain and the United States

Research Background and Purpose

Spanish is widely spoken not only in Spain but also in the United States. However, unlike Spain, it is thought that English has a lot of influence on Spanish in the United States. Therefore, this study intends to study the vowel and syllable reduction phenomenon that occurs in Spanish online in Spain and the United States, and to analyze the causes.

Social media research with phonology (in-progress)

Topic

A comparative study of Spanish vowel and syllable reduction on social media in Spain and the United States

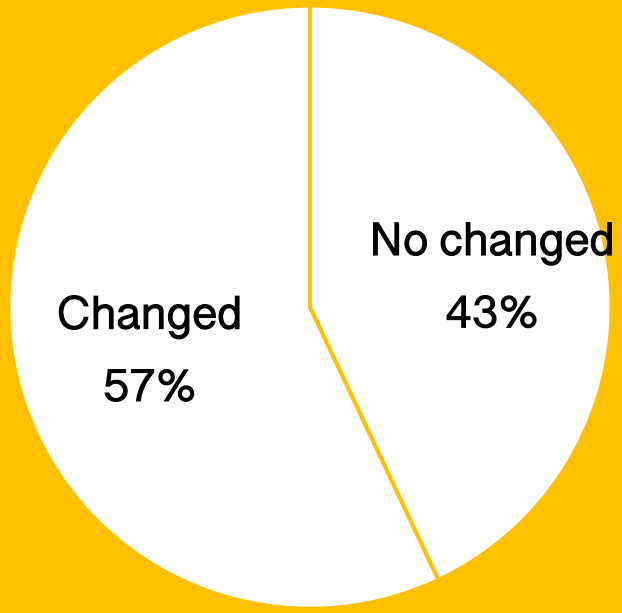
Methodology

This study aims to collect Spanish tweets written in Spain and the United States using the Twitter crawling technique. Afterwards, I want to select non-standard Spanish words used in each tweet through data analysis techniques, and then select and analyze words that are used in vowel or syllable abbreviations among them.

Social media research with phonology (in-progress)

IN SPAIN

(TWEETS WITH CHANGED VOWEL AND NO CHANGED VOWEL PER WHOLE TWEETS COLLECTED)

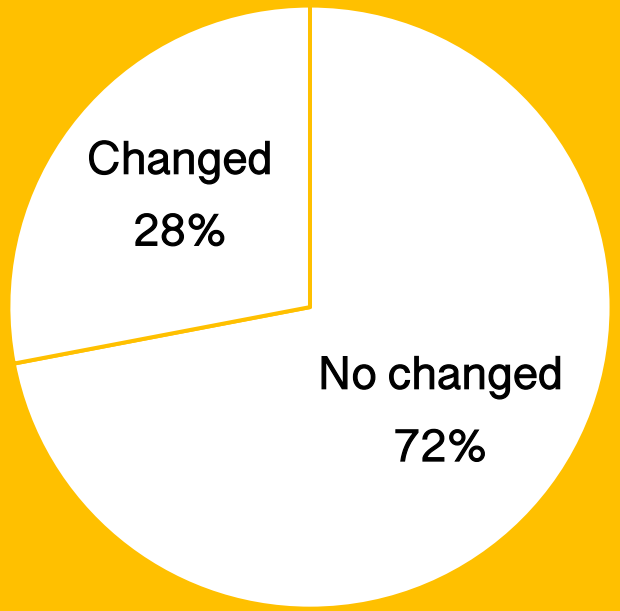


Number of tweets collected: 8074

Social media research with phonology (in-progress)

IN USA

(TWEETS WITH CHANGED VOWEL AND NO CHANGED VOWEL PER WHOLE TWEETS COLLECTED)



Number of tweets collected: 1231

Social media research with phonology (in-progress)

In Spain

Type	Original Word	Word in Twitter	Type	Original Word	Word in Twitter
Vowel simplification	que	ke	Vowel delete	quien	kn
	aquí	aki		no te preocupes	ntp
Vowel exaggeration	claro	claaaaaaro		me da igual	mdi
	bueno	bueeeeeno		¿Qué haces?	K acs
Syllable reduction	más o menos	maso		fin de semana	fds
	vacaciones	vacas		porque	xq

Vowel delete	Vowel simplification	Vowel exaggeration	Syllable reduction
52%	14%	26%	8%

Social media research with phonology (in-progress)

In USA

Type	Original Word	Word in Twitter	Type	Original Word	Word in Twitter
Vowel simplification	que	ke	Vowel delete	de	d
	quien	kien		porque	pq
Vowel exaggeration	gracias	graaaaacias		te	t
	bueno	bueeeeeno		quién	kn
Syllable reduction	chicos y chicas	chic@s		también	tb
				te quiero	tq

Vowel delete	Vowel simplification	Vowel exaggeration	Syllable reduction
33%	22%	41%	4%

Social media research with phonology (in-progress)

Analyze (in-progress)

1. Is there any influence of English in United States?
   - It seems that there is not much influence of English in United States
2. What is main difference?
   - The vowel changes and syllable reductions observed in Spain are more complex than in the United States.

Social media research with phonology (in-progress)

Analyze (in-progress)

3. What did it cause?

In the case of Spain, in a predominantly Spanish-speaking society, people speak Spanish perfectly and learned it as mother tongue. However, in the case of the United States, there are a lot of the second or third generation of Hispanic immigrants. Although their mother tongue could be Spanish, they live in an English-dominated environment, they will be not fluent Spanish speakers rather than the Spanish people.

Q & A