



CEMPASUCHIL

Corpus sociolingüístico de español mensajado
para el análisis del habla universitaria

Lic. Miguel Alejandro Dorantes Cruz

Tlahulia Yamín Donohue Pérez

Ana Laura Del Prado Mota

Mayra Paulina Díaz Rojas



Justificación

En la actualidad no existe ningún corpus que permita estudiar la comunicación mediada por dispositivos móviles que se centre en jóvenes universitarios mexicanos.



Objetivo general

Crear un recurso lingüístico que permita el estudio y caracterización del habla de Ciudad Universitaria en mensajería instantánea.



Objetivos específicos


- Conformar una base de datos de historiales de conversaciones de WA en formato txt, así como de la información de los hablantes de acuerdo a las variables sociales del corpus.
- Ofrecer una muestra representativa del uso de lengua tecleada entre estudiantes de todas las facultades que se encuentran en Ciudad Universitaria.

Estado del arte

Existe el precedente de dos proyectos similares, realizados en Europa.



SoNaR: STEVIN Nederlandstalig Referentiecorpus



Holgado Lage y Recio Diego (2013) :
(Skype, Facebook y WhatsApp)

10 chats

Alcántara Plá (2014):

106 informantes, 32 chats, 176 000 palabras

García Arriola (2014):

16 chats

Vázquez Cano (2015):

101 401 palabras

Martín Gascueña (2016):

90 informantes, 100 chats

Sampietro (2016):

44 000 palabras, 3 151 mensajes

Gómez del Castillo (2017):

38 informantes, 529 chats, 3 872 mensajes, 20 404
palabras



Estudios con corpus de mensajería instantánea

- Las unidades discursivas (turnos, secuencias, actos de habla, interacciones, etc.)
- Variación lingüística desde un punto de vista diafásico, diastrático o diatópico
- Comunicación multimodal (verbal, icónica o híbrida)
- Uso de elementos ortotipográficos
- Intervención de emojis en la comunicación
- Uso didáctico de este tipo de comunicación para la competencia digital y la competencia lingüística



Metodología

Constó de tres fases:

1. Codificación de informantes : Por tratarse de un corpus sociolingüístico se definieron variables sociodemográficas bajo las cuales fueron agrupados los informantes; éstas se dividieron en dos tipos:

- Ejes de balance: Sexo y Facultad
- Criterios de post-estratificación: Edad, Orientación sexual, Lugar de nacimiento, Código Postal, Dominio de otras lenguas, Nivel educativo, Ocupación y Profesión u oficio.

2. Recolección de datos:

- “Amigo de un amigo”
- Recolección dirigida

3. Desarrollo informático

Facultad	Población total		Población meta	
	H	M	H	M
Arquitectura	3595	3501	36	35
Ciencias	3863	3745	38	37
Ciencias Políticas y Sociales	2904	4208	29	31
Contaduría y Administración	5631	5094	56	50
Derecho	2938	4180	29	42
Economía	2768	1346	28	13
Filosofía y Letras	3256	4523	32	45
Escuela Nacional de Trabajo Social	487	1678	5	17

Facultad	Población total		Población meta	
	H	M	H	M
Ingeniería	10076	2957	100	30
Medicina	2440	4325	24	43
Medicina Veterinaria y Zootecnia	1079	2240	10	22
Odontología	803	2059	8	20
Psicología	617	2035	6	20
Química	3377	3759	34	37



Desarrollo informático

Etapas

Base de datos

Formulario (xlsx) > JSON > Base de datos

Conversaciones (txt) > programa > user 1 /
user 2 > vincula la información de cada
usuario a la conversación que le corresponde
> Base de datos

```
{
  "_id" : ObjectId(""),
  "ID" : String,
  "tags" : [
    String,
    String,
  ],
  "user1" : {
    "age" : String,
    "gender" : String,
    "osexual" : String,
    "birthPlace" : String,
    "livePlace" : String,
    "languages" : String,
    "education" : String,
    "facult" : String,
    "major" : String,
    "ocupation" : String,
    "prof_of" : String,
    "relation" : String,
    "contact" : String,
    "email" : String,
    "lines" : String
  },
  "user2" : {
    "age" : String,
    "gender" : String,
    "osexual" : String,
    "birthPlace" : String,
    "livePlace" : String,
    "languages" : String,
    "education" : String,
    "facult" : String,
    "major" : String,
    "ocupation" : String,
    "prof_of" : String,
    "relation" : String,
    "lines" : String
  },
  "conversación" : String
}
```



Desarrollo informático

Etapa 2

App para consultar la base de datos

- se elaboró una interfaz de usuario para consultas en lenguaje python
- manual de instalación
- manual de usuario para aplicación de consultas

[GIL](#)[Acerca del corpus](#)[Integrantes](#)[Ligas de interés](#)[Otros](#)

CEMPASÚCHIL

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam maximus arcu sodales tincidunt iaculis. Sed faucibus eget eros non hendrerit. Fusce diam nibh, vehicula sed tellus a, dictum luctus augue. Quisque iaculis vitae dui eu tincidunt. Integer ultrices ipsum velit, et finibus neque vehicula non. Mauris fringilla augue at velit vulputate lobortis. Curabitur eleifend pretium condimentum. Etiam ex diam, tempor sit amet hendrerit a, auctor in mi. Vestibulum quis nisi et eros convallis fermentum. Donec vehicula metus libero, quis porta ex commodo eu. Cras hendrerit ante id lobortis lacinia.

Ut ultrices tellus dolor, eget sodales magna commodo nec. Nunc et consequat nibh. Integer laoreet posuere interdum. Duis laoreet urna nec nulla convallis dapibus. Pellentesque a dui suscipit, auctor nisi et, dictum mi. Ut at lectus ac metus luctus egestas. Sed ultricies dictum arcu, a congue diam interdum commodo. Nullam elementum mauris nec lorem molestie pellentesque. Vestibulum in sagittis lorem. In risus augue, sagittis eget tincidunt at, facilisis eget eros. Suspendisse id fringilla tortor, sit amet facilisis orci. Aliquam tempus ornare odio eget imperdiet. Phasellus sit amet dolor a erat ullamcorper cursus. Morbi dignissim mattis purus vitae maximus. Integer at rutrum arcu, consequat faucibus odio. Mauris vitae tincidunt orci. Aliquam erat eros, congue in orci ac, imperdiet pulvinar ex. Nunc ut venenatis nunc, id fringilla erat. Sed sed ipsum sed sapien congue interdum nec sed purus. Nunc diam quam, ullamcorper sit amet metus nec, cursus pharetra massa. Donec auctor lacus sit amet sem blandit, quis varius elit venenatis. Sed viverra scelerisque interdum. Donec et vehicula lorem. Etiam lacinia nibh sem, sit amet ullamcorper massa vulputate volutpat.

Nuestro corpus cuenta con:

Mujeres 127 – Hombres 169

[Manual para usar la página de consulta](#)

Elige los parámetros de la búsqueda

[Limpiar](#)

Edad Género O. Sexual L. Nacimiento Nivel educativo

Ocupación Otras lenguas Relación

Carrera Facultad

L. Residencia

Enviar



Conclusiones

Se lograron obtener 835 conversaciones con 1237 informantes de los cuáles 1121 son estudiantes de licenciatura de C.U.

De los informantes válidos 574 son mujeres y 547 son hombres.

- 
- Edad: 14 - 60 años
 - Orientación sexual: Heterosexual, Homosexual, Bisexual
 - Lugar de nacimiento:

Aguascalientes, Campeche, Chiapas, Chihuahua,, Ciudad de México, Coahuila, Durango, Estado de México, Guanajuato, Guerrero, Hidalgo, Jalisco, Michoacán, Morelos, Nuevo León, Oaxaca, Puebla, Querétaro, Quintana Roo, Sinaloa, Tabasco, Veracruz, Yucatán
Chile, Colombia, Costa Rica, Estados Unidos, Italia, Líbano.

- Otras lenguas:

Alemán, Búlgaro, Chino, Coreano, Francés, Griego moderno, Inglés, Italiano, Japonés, Náhuatl, Portugués, Ruso, Sueco, Árabe.

- Facultades

Facultad	Hombres	Mujeres	Facultad	Hombres	Mujeres
Arquitectura	32	41	Ingeniería	144	33
Ciencias	40	60	Medicina	27	50
CP y S	37	56	MVZ	16	31
Contaduría y Administración	71	52	Odontología	14	25
Derecho	25	50	Psicología	10	41
Economía	39	22	Química	52	39
Filosofía y Letras	33	54	Trabajo Social	6	20



Equipo

Coordinación general :

Lic. Miguel Alejandro Dorantes Cruz

Tlahuilia Yamín Donohue Pérez

Asesor:

Dr. Gerardo Sierra Martínez

Equipo de trabajo:

Ana Laura Del Prado Mota

Mayra Paulina Díaz Rojas

Mónica Gabriela Jasso Rosales

Eduardo Daniel Rivera Peñaloza

Desarrollo informático:

Paola Sánchez González



Agradecimientos

Agradecemos el apoyo para realizar este proyecto al **Grupo de Ingeniería Lingüística** de la UNAM y a la **Red Temática en Tecnologías del Lenguaje**.

Ling^{enierí}_{üístic}a

 RedTTL



Referencias

- Alcántara Plá, Manuel. "Las unidades discursivas en los mensajes instantáneos de wasap". *Estudios de Lingüística del Español* 35, 2014. 214–233.
- García Arriola, Manuel. "Análisis de un corpus de conversaciones en Whatsapp. Aplicación del sistema de unidades conversacionales propuesto por el grupo VALES.CO." Tesis. Universidade da Coruña. Facultade de Filoloxía, 2014.
- Holgado Lage, Anais, and Álvaro Recio Diego. "La oralización de textos digitales: usos no normativos en conversaciones instantáneas por escrito". *Caracteres. Estudios culturales y críticos de la esfera digital* 2.2, 2013. 92–108.
- Martín Gascueña, Rosa. "La conversación guasap". *Sociocultural Pragmatics* 4.1. Ed. Diana Bravo, 2014. Berlin: De Gruyter. 108–134
- Sampietro, Agnese. "Exploring the punctuating effect of emoji in Spanish whatsapp chats". *Lenguas Modernas* 47, 2016. 91–113.
- Vázquez-Cano, Esteban, Santiago Mengual-Andrés, and Rosabel Roig-Villa. "Análisis lexicométrico de la especificidad de la escritura digital del adolescente en WhatsApp". *RLA. Revista de Lingüística Teórica y Aplicada Concepción* 53.1, 2015. 83–105.