

Recolección online de grabaciones para el estudio de las variantes argentinas del español

Fernando Bugni

Directores:
Agustín Gravano,
Miguel Martínez Soler

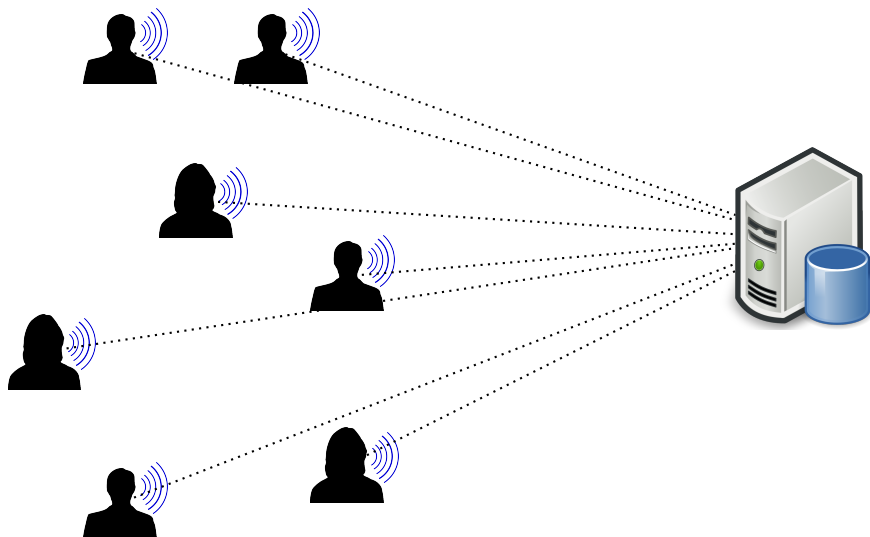
Departamento de Computación - Facultad de Ciencias Exactas -
Universidad de Buenos Aires

2014

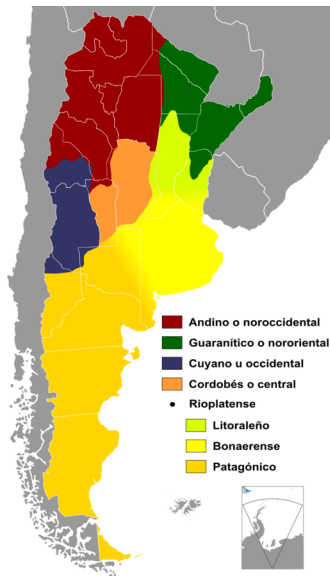
Recolectar grabaciones para estudios del habla



Recolectar grabaciones para estudios del habla



Variantes del español en Argentina



Caso de estudio: Buenos Aires y Córdoba



Diferencias entre Córdoba y Buenos Aires

- (1) Los hablantes de Córdoba estiran la sílaba anterior a la acentuada mientras los de Buenos Aires no lo hacen. Ejemplo: '*Espectacular*'
- (2) Los hablantes de Córdoba aspiran y elisionan la /s/ al finalizar una palabra. Esto no sucede en Buenos Aires. Ejemplo: '*Pájaros*'
- (3) Para hablantes de Córdoba, la /s/ antes de la /c/ o /t/ suenan más suaves que para hablantes de Buenos Aires. Ejemplo: '*Mosca*'
- (4) La 'c' antes de la 't' se pronuncia con menor frecuencia para hablantes de Córdoba que para hablantes de Buenos Aires. Ejemplo: '*Doctor*'
- (5) Para hablantes cordobeces la 'y' y 'll' se pasa a 'i'. No sucede esto para Buenos Aires. Ejemplo: '*lluvia*'
- (6) En hablantes cordobeces la /r/ no vibra mientras que en Buenos Aires pasa lo contrario. Ejemplo: '*Espárrago*'

Bibliografía:

- El español en la Argentina y sus variedades regionales - María Beatriz Fontanella de Weinberg
- Español en la Argentina - Elena Vidal de Battini

Diseño del experimento

- **Frases Comunes:** habla espontánea
- **Frases Amper:** reconocer palabra acentuada

Diseño del experimento

Frases Comunes

Pronunciar frases popularmente conocidas

- Objetivo: pronunciación espontánea
- Reglas a cubrir: 2 a 6

**‘En la pelea se conoce al soldado,
sólo en la **victoria** se conoce al **caballero**’**

- **‘victoria’** cubre la **regla 4** que nos propone medir la duración de la /c/ antes de la /t/.
- **‘caballero’** para la **regla 5** el fonema /ll/ se pasa a /i/

Diseño del experimento

Frase	Frase que cubre
'No hay dos sin tres'	Regla 2: 'dos', 'tres'
'Más difícil que encontrar una aguja en un pajar'	Regla 2: 'más'
'Más perdido que turco en la neblina'	Regla 2: 'más'
'No le busques la quinta pata al gato'	Regla 2: 'busques', Regla 3: 'busques'
'Se te escapó la tortuga'	Regla 3: 'escapó'
'Todos los caminos conducen a Roma'	Regla2: 'todos', 'los', 'caminos'
'Siempre que llovió, paró'	Regla 5: llovió
'La suegra y el doctor, cuanto más lejos, mejor'	Regla 2: más, lejos , Regla 4: doctor
'La belleza que atrae, rara vez coincide con la belleza que enamora'	Regla 5: belleza
'No esta mal ser bella, lo que está mal es la obligación de serlo'	Regla 5: bella
'Río revuelto, ganancia de pescadores'	Regla 3: pescadores, Regla 2: pescadores, Regla 6: río, revuelto
...	

Agrega 31 Frases populares para grabar

Diseño del experimento

Frases Amper

Pronunciar frases con una estructura fija variando acentuaciones

- Objetivo: cubrir acentuaciones
- Regla a cubrir: 1

Sujeto + “ salió ” + Adjetivo

- **Sujeto:** “El canapé”, “El repollo”, “El espárrago”.
- **Adjetivo:** “espectacular”, “delicioso”, “riquísimo”.

“**El canapé** salió **delicioso**”
 { palabra aguda palabra grave }

Agrega 9 frases Amper

Bibliografía: AMPER-ARGENTINA: VARIABILIDAD RÍTMICA EN DOS CORPUS - Jorge A. Gurlekian, Reina Yanagida,

Mónica Noemí Trípodí y Guillermo Toledo

Combinamos cada tipo de frase de f3rma aleatoria



Sistema de grabación online



¡Bienvenido/a!

Este proyecto consiste en **grabar una serie de frases a través de tu computadora**, para luego poder estudiar las características del habla de cada región (por ejemplo, la tonada o los sonidos empleados).

Requisitos para poder participar:

1. Tener una **buena conexión a Internet**; preferentemente, no wireless.
2. Tener un **buen micrófono**; preferentemente, no usar el micrófono incluido en una laptop.
3. Estar en un **ambiente silencioso**.

Si cumplís estos requisitos, por favor completá los siguientes datos para comenzar:

Sexo:

Lugar donde te criaste:

Lugar donde vivís actualmente:

Mes de nacimiento: 01-1990

¡Empezar!

Figura : Encuesta inicial del sistema

Sistema de grabación online

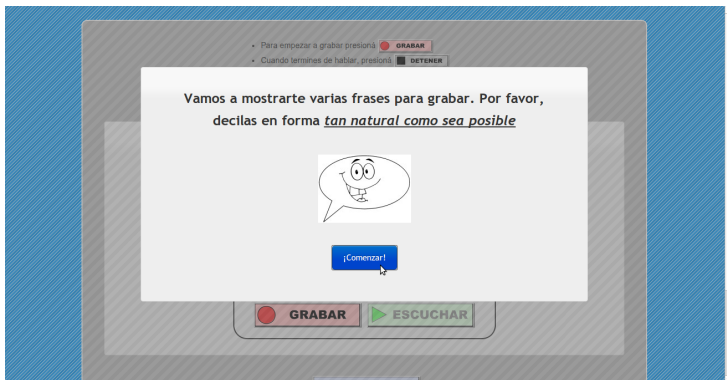


Figura : Instrucciones del experimento

Sistema de grabación online

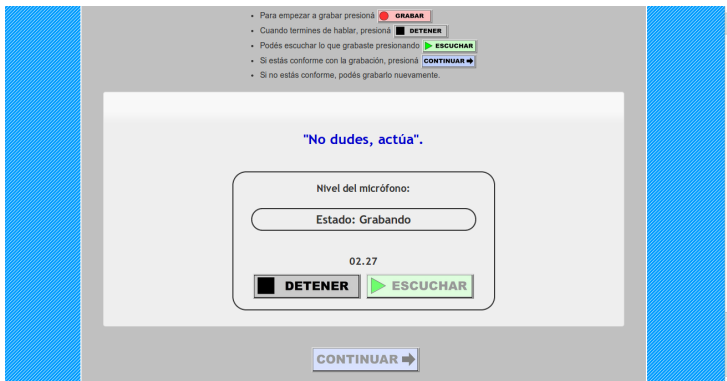


Figura : Grabando

Sistema de grabación online



Figura : Reproduciendo

Datos obtenidos

	Bs.As.	Cba.	Total
Conservar	220	90	310
Problemas en el habla¹	33	15	48
Mucho ruido de fondo	2	12	14
Sonido saturado	2	0	2

Tabla : Evaluación manual de las grabaciones

	Bs.As.	Cba.	Total
Todos los intentos	220	90	310
Último intento	181	79	260

Tabla : Cantidad de audios repetidos

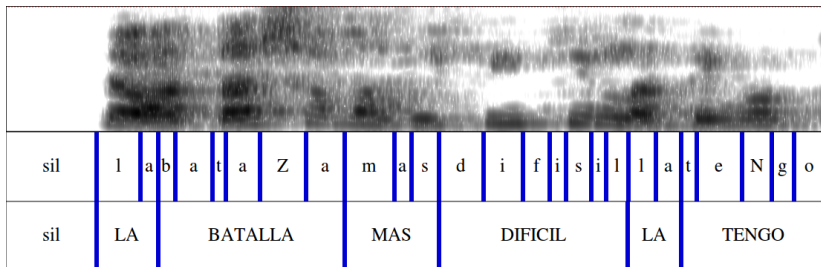
¹Problemas más comunes: entonación exagerada y error al pronunciar una frase

Extracción de información

¿Cómo extraer atributos (features) de un audio?

Etiquetamos el principio y el final de cada fonema.

ProsodyLab-Aligner: alineación forzada



Extracción de información

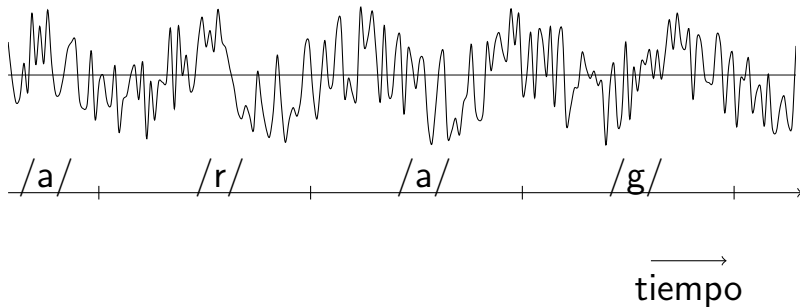
Definimos tipos de atributos:

- Atributos acústicos
- Atributos fonéticos
- Atributos silábicos

ProsodyLab, Python 2.7, Numpy, Pymatlab

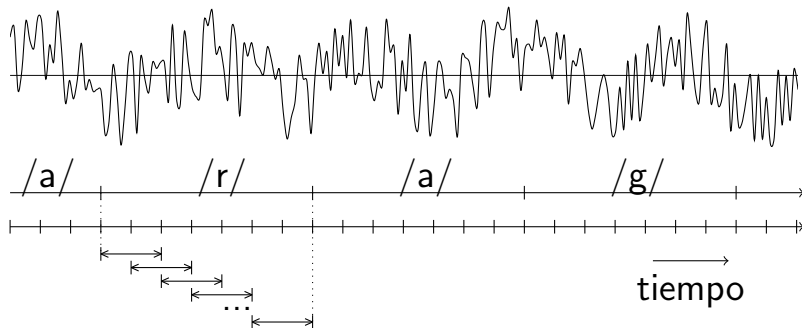
Extracción de información

Atributos acústicos: MFCC



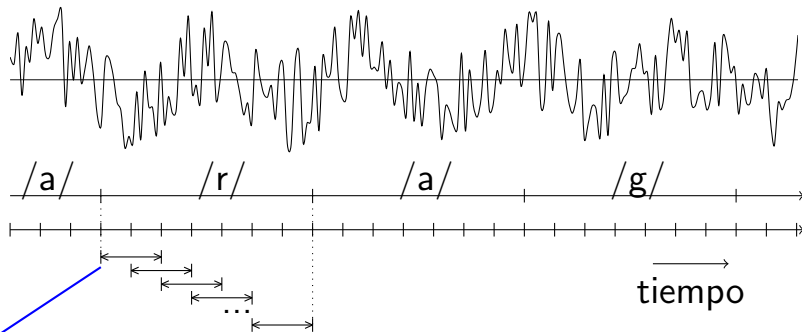
Extracción de información

Atributos acústicos: MFCC



Extracción de información

Atributos acústicos: MFCC

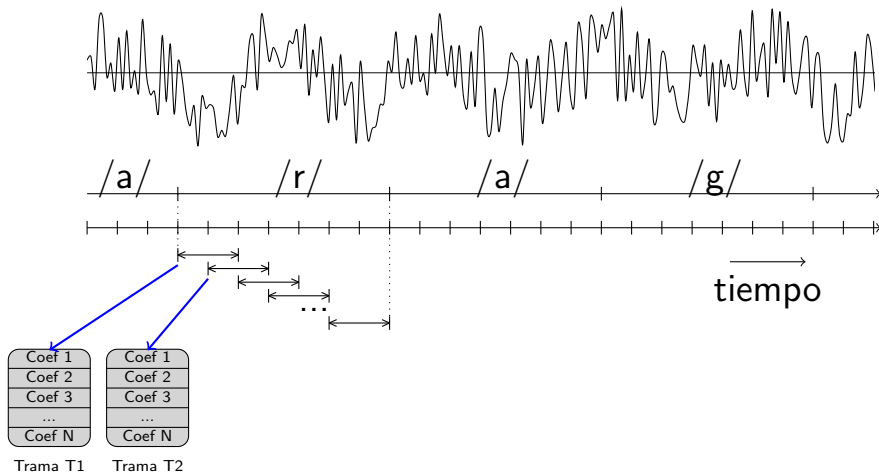


Coef 1
Coef 2
Coef 3
...
Coef N

Trama T1

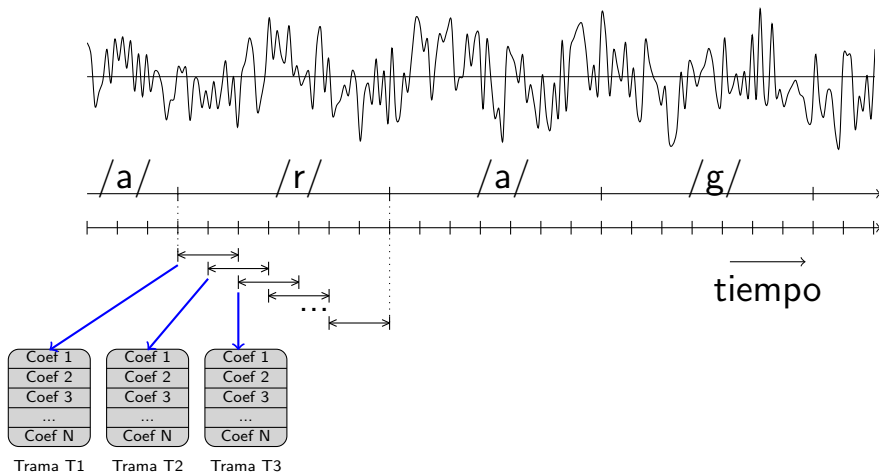
Extracción de información

Atributos acústicos: MFCC



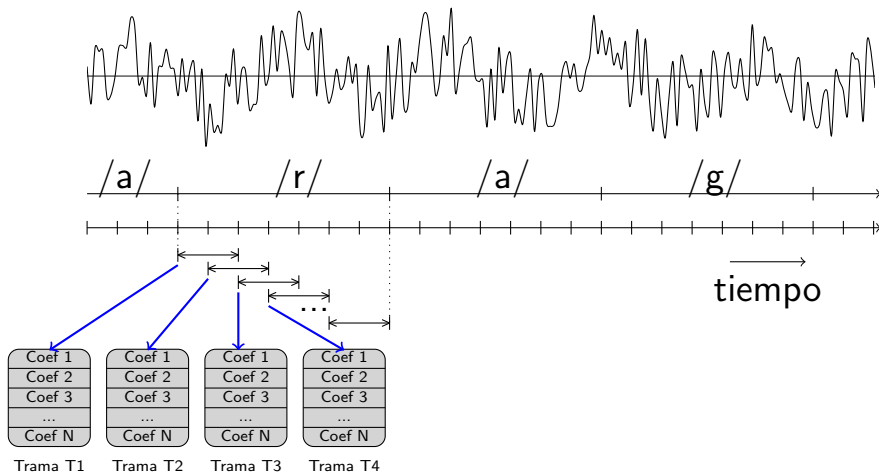
Extracción de información

Atributos acústicos: MFCC



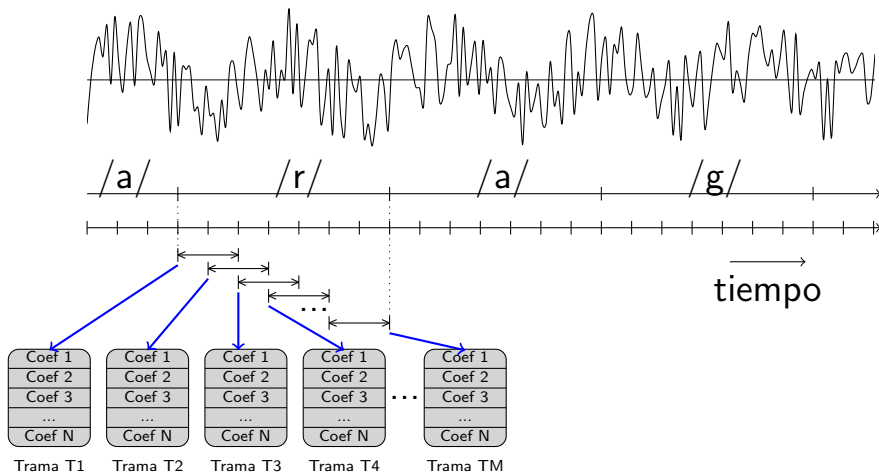
Extracción de información

Atributos acústicos: MFCC



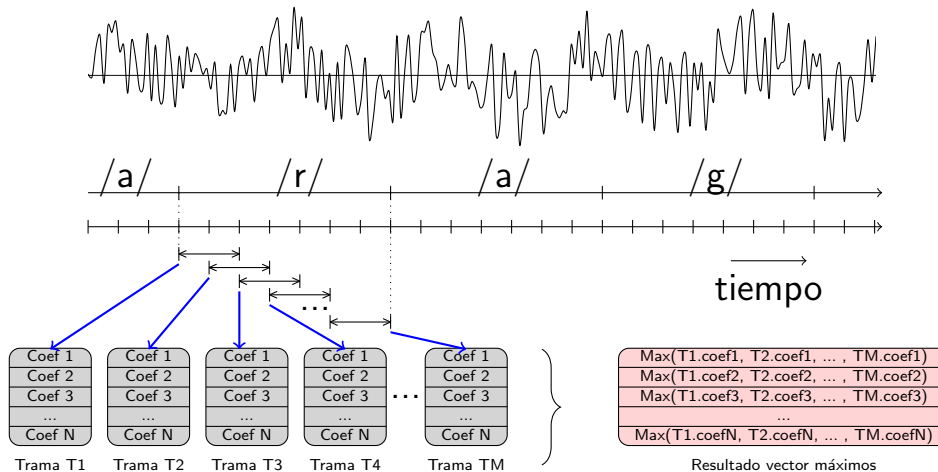
Extracción de información

Atributos acústicos: MFCC



Extracción de información

Atributos acústicos: MFCC



Extracción de información

Atributos fonéticos

- Duración de 'kt'
- Duración de 'sc'
- Duración de 'll'
- Duración de 'rr'
- Duración de 's' final
- Duración de cada fonema
- Duración de cada vocal
- Duración de cada consonante

Extracción de información

Atributos silábicos

- Duración de la sílaba acentuada
- Duración de la sílaba anterior a la acentuada

Análisis

¿Cómo podemos saber si teniendo en cuenta estos atributos podemos clasificar mejor a un cordobés?

Análisis

¿Cómo podemos saber si teniendo en cuenta estos atributos podemos clasificar mejor a un cordobés?

Dividimos nuestros datos en dos conjuntos: **Entrenamiento** y **Testeo**.

- Un *clasificador* se entrenará con los datos del conjunto de **Entrenamiento**
- Luego, intentará clasificar correctamente el conjunto de **Testeo**

Realizamos varias configuraciones similares a ésta, variando los elementos de Testeo (*folds*)

Ésta configuración se llama *Cross-validation*.

Análisis

Clasificadores

- Zero Rules
- RIPPER
- C4.5
- Support vector machine
- Naive Bayes

Cross-validations

- Clasificación por muestra
- Clasificación por hablante

Clasificación por muestra

Hablantes: 19 Buenos Aires, 8 Córdoba

 Hablante para train  Hablante para test









































	<i>Número de hablante</i>										
	1	2	3	4	5	6	7	...	25	26	27
Fold 1								...			
Fold 2								...			
Fold 3								...			
				...							
Fold 27								...			

Tabla : Esquema de validación cruzada

Clasificación por muestra

Veamos el clasificador Ripper para el fold 4. Este es el conjunto de reglas que generó:

- $(FON_rr_norm \leq -6,901) \text{ and } (ACU_AverageRR_7 \leq 11,23) \Rightarrow place = cba(18,0/3,0)$
- $(FON_ll_norm \leq -7,975) \text{ and } (ACU_AverageLL_6 \leq 4,308) \Rightarrow place = cba(15,0/0,0)$
- $else \Rightarrow place = bsas(222,0/49,0)$

	Zero Rule	Ripper	C4.5	SVM	NaiveBayes
Promedio	70	69	70	71	71

Tabla : Clasificación correcta en porcentaje

Clasificación por muestra

Detalles:

- C4.5 obtuvo la misma performance que Zero Rule y su árbol de decisión fue muy pobre
- Zero Rule obtuvo una buena performance y algunos clasificadores no pudieron obtener provecho de los atributos

Clasificación por muestra

Detalles:

- C4.5 obtuvo la misma performance que Zero Rule y su árbol de decisión fue muy pobre
- Zero Rule obtuvo una buena performance y algunos clasificadores no pudieron obtener provecho de los atributos

¿Qué sucedería si colapsáramos los atributos por hablante y equilibramos nuestros datos?

Clasificación por hablante

Hablantes equilibrados: 9 Buenos Aires, 8 Córdoba

 Hablante para train  Hablante para test









































	<i>Número de hablante</i>										
	1	2	3	4	5	6	7	...	15	16	17
Fold 1								...			
Fold 2								...			
Fold 3								...			
				...							
Fold 17								...			

Tabla : Esquema de cross-validation

Clasificación por hablante

Juntamos los atributos de cada hablante de la siguiente forma.

Atributos		A1	A2	A3	...	AN
Hablante 1	Audio1	1	?	2		2
	Audio2	?	?	1	...	?
	Audio3	2	?	3		?
Hablante 2	Audio1	1	?	?	...	?
	Audio2	1	2	?		?

Tabla : Datos original

esto pasaría a:

Atributos		A1	A2	A3	...	AN
Hablante 1	Audio1	1.5	?	2	...	2
Hablante 2	Audio1	1	2	?	...	?

Tabla : Datos modificados

Clasificación por hablante

	ZeroR	RIPPER	C4.5	SVM²	NaiveBayes
Promedio	53	53	76	94	76

Tabla : Clasificación correcta en porcentaje

²Corriendo T-test obtenemos $p < 0,05$

Clasificación por hablante

Detalles:

- Cada clasificación tiene 1 instancia para analizar
Matrices de confusión muy pobres.

Clasificadas como		Instancias de
Buenos Aires	Córdoba	
1	0	Buenos Aires
0	0	Córdoba

Selección de atributos de forma automática

¿Cuál es la importancia relativa de cada atributo?

Medimos cuanta información aporta cada atributo utilizando el algoritmo InfoGain.

Ganancia de Información	Atributo
0.07231	FON_consonant_norm
0.07217	FON_vowel_norm
0.03963	SIL_syllableAccent_normhd
0.03963	SIL_prevSyllableAccent_normhd
0.02332	FON_ll_norm
0.02285	FON_Sfinal_norm
0.02226	ACU_MinLL_1
0.02144	ACU_AverageLL_1

Tabla : Resultados de InfoGain

Conclusiones y trabajo futuro

- Armamos una plataforma para la recolección de grabaciones
- Caso de estudio: diferencia entre habla de Cba. y BsAs.
- Características del conjunto de datos y cómo repercute en sus resultados
- Grabaciones chequeadas entre los hablantes
- Desarrollo de varios filtros para evitar grabaciones con problemas
- Realizar clasificación en vivo a través de una página web
- Mejores modelos variando sus parámetros y nuevos atributos

¿Preguntas?

Más detalles...

Diferencias entre Córdoba y Buenos Aires

- **Regla 1: Los hablantes de Córdoba estiran la sílaba anterior a la acentuada mientras los de Buenos Aires no lo hacen**

*‘Especta**cu**lar’*

Sílaba acentuada en *‘-lar’*

La sílaba anterior *‘-cu-’* se alarga para hablantes de Córdoba

Diferencias entre Córdoba y Buenos Aires

- **Regla 2: Los hablantes de Córdoba aspiran y elisionan la /s/ al finalizar una palabra. Esto no sucede en Buenos Aires**

'Pájaros'

/s/ se acorta su duración en el hablante de Córdoba

Diferencias entre Córdoba y Buenos Aires

- **Regla 3: Para hablantes de Córdoba, la /s/ antes de la /c/ o /t/ suenan más suaves que para hablantes de Buenos Aires**

‘Mosca’

/s/ suena más suave para Córdoba que para Buenos Aires

Diferencias entre Córdoba y Buenos Aires

- **Regla 4: La 'c' antes de la 't' se pronuncia con menor frecuencia para hablantes de Córdoba que para hablantes de Buenos Aires**

'Doctor'

No debe sonar el fonema /c/

Diferencias entre Córdoba y Buenos Aires

- **Regla 5: Para hablantes cordobeces la 'y' y 'll' se pasa a 'i'. No sucede esto para Buenos Aires**

'lluvia'

Palabras con el fonema /y/ o /ll/ se pronuncian /j/

Diferencias entre Córdoba y Buenos Aires

- **Regla 6: En hablantes cordobeces la /r/ no vibra mientras que en Buenos Aires pasa lo contrario**

‘Espárrago’

Para Córdoba /r/ debe ser suave en comparación de Buenos Aires

Bibliografía:

- El español en la Argentina y sus variedades regionales - María Beatriz Fontanella de Weinberg
- Español en la Argentina - Elena Vidal de Battini

Diseño del experimento

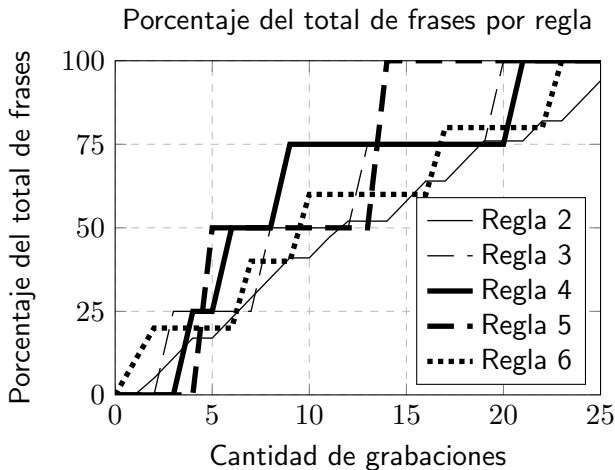


Figura : Porcentaje del total de frases grabadas por cada regla

Sistema de grabación online

Audio 6

Id: 6

Speaker: 2

Word: No está mal ser bella; lo que está mal es la obligación de serlo

Attempt: 1

Filename:



download: [bsas_u2_t32_a1](#)

Labels:

- ☐ Conservar
- ☒ Sonido saturado
- ☐ Mucho ruido de fondo
- ☐ Problema en el habla

Submit

Figura : Administrador

Extracción de información

Atributos acústicos: MFCC

Escala Mel: escala sobre la precepción auditiva humana

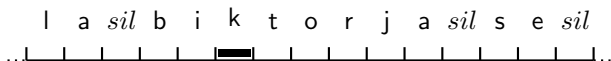
- (1) Frame the signal into short frames.
- (2) For each frame estimate the power spectrum (Fast Fourier Transform).
- (3) Apply the mel filterbank to the power spectra, sum the energy in each filter.
- (4) Take the logarithm of all filterbank energies.
- (5) Take the DCT of the log filterbank energies. (DCT=discrete cosine transform)
- (6) Keep DCT coefficients 2-13, discard the rest.

Script en Matlab llamado por Pymatlab

Extracción de información

Atributos fonéticos: cálculo duración de 'kt'

“en la pelea se konose al soldaDo solo en la biktorja se konose al kaBaZero”



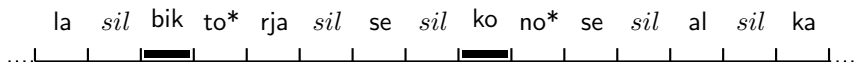
$$\frac{X - \mu}{\sigma}$$

- X es el valor a normalizar (por ej.: la duración de un fonema dado).
- μ es el promedio de duración de la unidad utilizada en la grabación.
- σ es el desvío estándar de la unidad utilizada en la grabación.

Extracción de información

Atributos silábicos: sílaba anterior a la acentuada

**“en la pelea se konose al soldaDo solo en la
biktorja se konose al kaBaZero”**



$$\frac{X - \mu}{\sigma}$$

- X es el valor a normalizar (por ej.: la duración de un fonema dado).
- μ es el promedio de duración de la unidad utilizada en la grabación.
- σ es el desvío estándar de la unidad utilizada en la grabación.

Selección de atributos de forma automática

Para cada atributo calcula la entropía de la clase y luego calcula la entropía³ de la misma sabiendo el valor de este atributo

$$InfoGain(Class, Attribute) = H(Class) - H(Class|Attribute)$$

- $H(Class)$ representa el valor de la entropía de la clase a predecir. Mide la incertidumbre asociada a la clase sin tener en cuenta el valor de ningún atributo en particular.
- $H(Class|Attribute)$ representa el valor de la entropía de la clase sabiendo el valor del atributo *Attribute*

³Cuán frecuente es una clase en una serie de muestras