

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

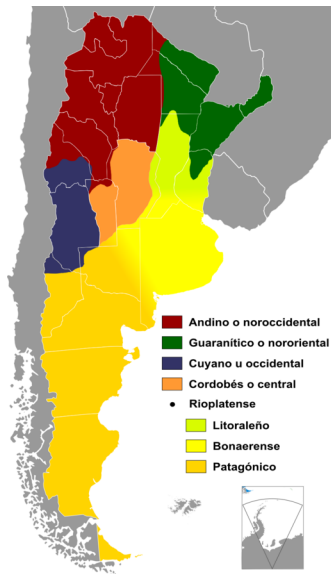
Fernando Bugni

Directores:  
Agustín Gravano,  
Miguel Martínez Soler

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

2014

# Variantes del español en Argentina



# 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

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

# Diseño del experimento

## Frases Comúnes

Pronunciar frases popularmente conocidas

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

# Diseño del experimento

**‘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 paro'	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 esta 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

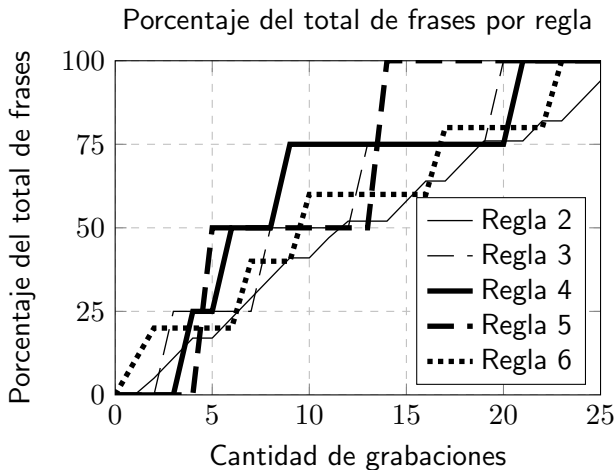


Figura : Porcentaje del total de frases grabadas por cada regla

# 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".

Bibliografía: AMPER-ARGENTINA: VARIABILIDAD RÍTMICA EN DOS CORPUS - Jorge A. Gurlekian, Reina Yanagida, Mónica Noemí Trípodi y Guillermo Toledo

# Diseño del experimento

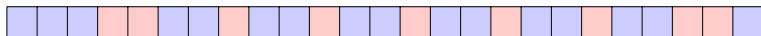
**“El **canapé** salió **delicioso**”**

- Canapé: palabra aguda
- Delicioso: palabra grave

**Agrega 9 frases amper**

# Diseño del experimento

Trazas: combinación de frases



■ Frases comunes

■ Frases Amper

Intercalado: 1 ó 3 Frases comunes cada una Amper



# 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 : Explicación del experimento

# Sistema de grabación online



Figura : Grabando

# 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

# Datos obtenidos

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

Tabla : Evaluación manual de las grabaciones

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

Tabla : Cantidad de audios repetidos

# Extracción de información

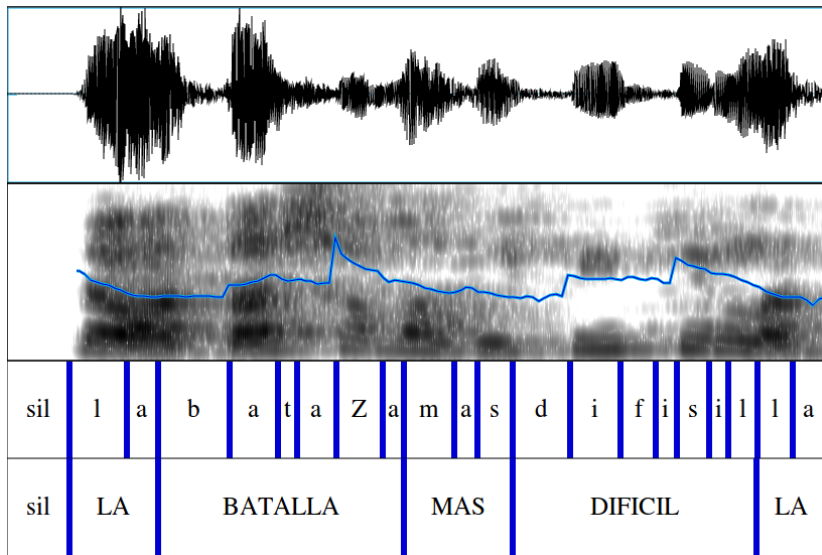
¿Cómo extraer atributos (features) de un audio?

Etiquetamos en qué momento se escuchó cada fonema.

## ProsodyLab-Aligner

Utilizando **Hidden markov models** alinea cada audio

# Extracción de información

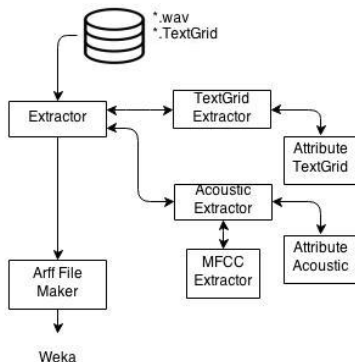


# Extracción de información

Definimos tipos de atributos:

- Atributo acústico
- Atributo fonético
- Atributo silábico

ProsodyLab, Python 2.7, Numpy, Pymatlab





# Extracción de información

## Atributos acústicos

**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

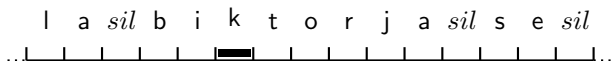
## Atributos fonéticos

- 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 fonético: 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).
- $\mu$  es el promedio de duración de la unidad utilizada en la grabación.
- $\sigma$  es el desvío estándar de la unidad utilizada en la grabación.

# Extracción de información

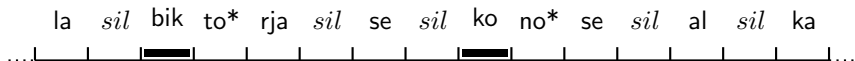
## Atributos silábicos

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

# Extracción de información

Atributos silábico: 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).
- $\mu$  es el promedio de duración de la unidad utilizada en la grabación.
- $\sigma$  es el desvío estándar de la unidad utilizada en la grabación.

## Clasificadores

- Zero rules
- RIPPER
- C4.5
- Support vectors machines
- Naive Bayes

## Cross-validations

- Grupos de hablantes
- Dejando un hablante fuera promediando los atributos
- Dejando un hablante fuera promediando los atributos desconocidos

# Grupo de hablantes

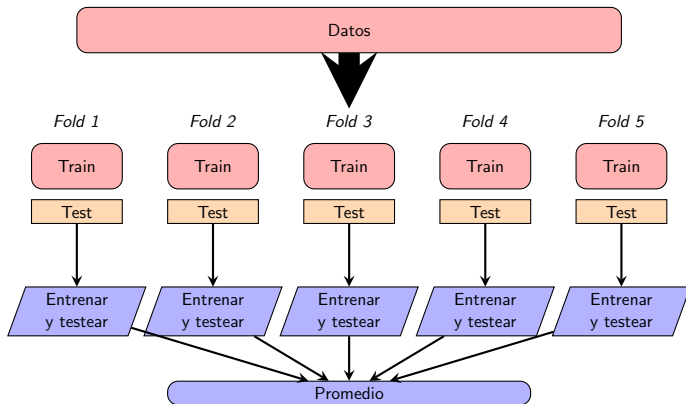


Figura : Esquema de test 5-folds



# Grupo de hablantes

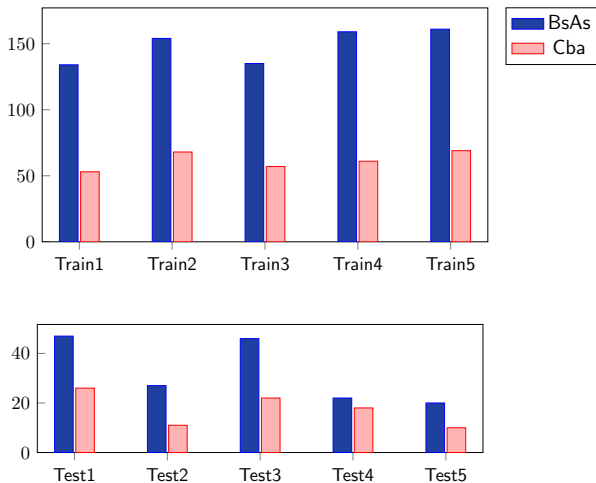


Figura : Cantidad de audios de Buenos Aires y Córdoba según cada grupo de Train y Tests

# Grupo de hablantes

	<b>Zero Rules</b>	<b>RIPPER</b>	<b>C4.5</b>	<b>SVM</b>	<b>NaiveBayes</b>
<b>Fold 1</b>	64	61	64	73	63
<b>Fold 2</b>	71	68	71	76	71
<b>Fold 3</b>	67	54	45	75	67
<b>Fold 4</b>	55	52	55	67	80
<b>Fold 5</b>	66	70	66	70	70
<b>Promedio</b>	64	61	60	72	70

**Tabla :** Clasificación correcta en porcentaje

	<b>Student Test</b>	<b>Wilcoxon Test</b>
<b>ZeroR y Ripper</b>	0.8438	0.87
<b>ZeroR y C4.5</b>	0.9772	0.813
<b>ZeroR y NaiveBayes</b>	0.2113	0.1692
<b>ZeroR y SVM</b>	0.03125	0.004545

**Tabla :** Resultados de cada test representado en p-valor

# Grupo de hablantes

## Detalles:

- Tenemos en cada fold más hablantes de Buenos Aires que de Córdoba. Esto se ve reflejado en el clasificador Zero Rules con un porcentaje de aciertos mayor al 50 %  
Necesitamos **equilibrar los hablantes** para analizar si se esta sacando provecho a los clasificadores.

# Grupo de hablantes

## Detalles:

- Tenemos en cada fold más hablantes de Buenos Aires que de Córdoba. Esto se ve reflejado en el clasificador Zero Rules con un porcentaje de aciertos mayor al 50 %  
Necesitamos **equilibrar los hablantes** para analizar si se esta sacando provecho a los clasificadores.
- Analizando la salida de cada clasificador, observamos que C4.5 nos da un árbol sólo de 1 hoja.  
Inclusive vemos que tiene igual porcentaje que Zero Rules.  
Analizando C4.5 vemos que su algoritmo puede generar árboles de pocos atributos si en los datos hay muchos **missing values**.

# Dejando un hablante fuera promediando los atributos

Hablantes equilibrados: 8 Buenos Aires, 8 Córdoba

 Hablante para train  Hablante para test









































	<i>Número de hablante</i>										
	1	2	3	4	5	6	7	...	14	15	16
Fold 1								...			
Fold 2								...			
Fold 3								...			
				...							
Fold 16								...			

Tabla : Esquema de cross-validation

# Dejando un hablante fuera promediando los atributos

Juntamos los atributos de cada hablante de la siguiente forma.

Atributos		A1	A2	A3	...	AN
<b>Hablante 1</b>	<b>Audio1</b>	1	?	2		2
	<b>Audio2</b>	?	?	1	...	?
	<b>Audio3</b>	2	?	3		?
<b>Hablante 2</b>	<b>Audio1</b>	1	?	?	...	?
	<b>Audio2</b>	1	2	?		?

Tabla : Original

esto pasaría a:

Atributos		A1	A2	A3	...	AN
<b>Hablante 1</b>	<b>Audio1</b>	<b>1.5</b>	<b>?</b>	<b>1.667</b>	...	<b>2</b>
<b>Hablante 2</b>	<b>Audio1</b>	<b>1</b>	<b>2</b>	<b>?</b>	...	<b>?</b>

Tabla : Modificado

# Dejando un hablante fuera promediando los atributos

	<b>ZeroR</b>	<b>RIPPER</b>	<b>C4.5</b>	<b>SVM</b>	<b>NaiveBayes</b>
<b>Promedio</b>	53.33	60	60	93.33	80

Tabla : Clasificación correcta en porcentaje

	<b>Student Test</b>	<b>Wilcoxon Test</b>
<b>ZeroR y Ripper</b>	0.3351	0.3828
<b>ZeroR y C4.5</b>	0.2908	0.3864
<b>ZeroR y NaiveBayes</b>	0.05191	0.06472
<b>ZeroR y SVM</b>	0.004282	0.009828

Tabla : Resultados de cada test representado en p-valor

# Dejando un hablante fuera promediando los atributos

Detalles:

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

Buenos Aires	Córdoba	
1	0	Buenos Aires
0	0	Córdoba



# Dejando un hablante fuera promediando los atributos desconocidos

Atributos		A1	A2	A3	...	AN
<b>Hablante 1</b>	<b>Audio1</b>	1	?	2		2
	<b>Audio2</b>	?	?	1	...	?
	<b>Audio3</b>	2	?	3		?
<b>Hablante 2</b>	<b>Audio1</b>	1	?	?	...	?
	<b>Audio2</b>	1	2	?		?

Tabla : Original

Se cambia a ...

Atributos		A1	A2	A3	...	AN
<b>Hablante 1</b>	<b>Audio1</b>	1	?	2		2
	<b>Audio2</b>	<b>1.5</b>	?	1	...	<b>2</b>
	<b>Audio3</b>	2	?	3		<b>2</b>
<b>Hablante 2</b>	<b>Audio1</b>	1	<b>2</b>	?	...	?
	<b>Audio2</b>	1	2	?		?

# Dejando un hablante fuera promediando los atributos desconocidos

	<b>ZeroR</b>	<b>RIPPER</b>	<b>C4.5</b>	<b>SVM</b>	<b>NaiveBayes</b>
<b>Promedio</b>	50	72.44	73.48	77.19	74.62

Tabla : Clasificación correcta en porcentaje

	<b>Student Test</b>	<b>Wilcoxon Test</b>
<b>ZeroR y Ripper</b>	0.06537	0.1284
<b>ZeroR y C4.5</b>	0.06156	0.1111
<b>ZeroR y NaiveBayes</b>	0.03916	0.06111
<b>ZeroR y SVM</b>	0.02936	0.03522

Tabla : Resultados de cada test representado en p-valor

# Dejando un hablante fuera promediando los atributos desconocidos

Matríz de confusión mejores

Buenos Aires	Córdoba	
33	1	Buenos Aires
0	0	Córdoba

# Selección de atributos de forma automática

Para cada atributo calcula la entropía de la clase y luego calcula la entropía<sup>1</sup> 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*

---

<sup>1</sup>Cuán frecuente es una clase en una serie de muestras

# Selección de atributos de forma automática

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

Tabla : Resultados de InfoGain

¿Preguntas?