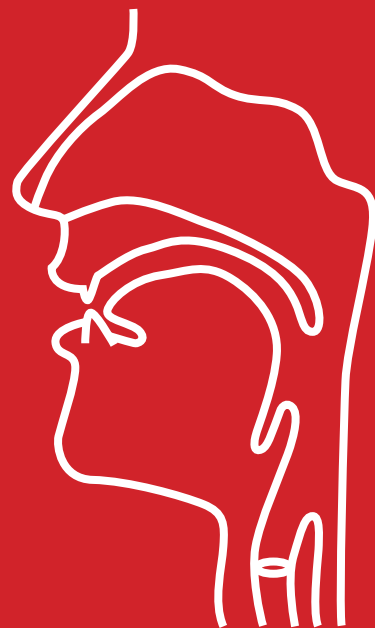
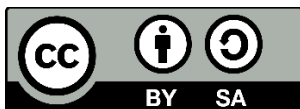


# Propuestas en fonética experimental: enfoques metodológicos y nuevas tecnologías

**Editores: Beatriz Blecua, Jordi Cicres, Marina Espejel  
y María J. Machuca.**





#### **Aviso legal**

Esta obra està sujeta a una licencia Reconocimiento 4.0 de Creative Commons. Está permitida la reproducción, distribución, comunicación pública y transformación para generar una obra derivada, sin restricción siempre que se cite al titular de los derechos (Universitat de Girona). La licencia completa se puede consultar en:  
<https://creativecommons.org/licenses/by-sa/4.0/deed.es>

© de los textos: los autores

Edita: Universitat de Girona-Servei de Publicacions

ISBN: 978 84 8458 590 9

Girona, Marzo de 2022

# Índice

PRESENTACIÓN .....	7
CATEGORIZACIÓN DE ERRORES PROSÓDICOS EN LA ENSEÑANZA DE LA PRONUNCIACIÓN L2 ASISTIDA POR ORDENADOR. Lourdes Aguilar, Eva Estebas-Vilaplana .....	11
LA IMPORTANCIA DE LA INTENSIDAD COMO PISTA ACÚSTICA EN LA IDENTIFICACIÓN DE ORACIONES INTERROGATIVAS. María Cristina Armero, Ignacio Moreno-Torres, Paolo Roseano. ....	16
INTRA- AND INTER-SPEAKER VARIATION OF MULTILINGUAL DATA: A FORENSIC PHONETICS APPROACH. José Vicente Benavent Cháfer .....	21
LA IMPORTANCIA DE LAS VOCALES EN LA ADQUISICIÓN Y LA ENSEÑANZA DEL ESPAÑOL HABLADO POR CHINOS. Yongfa Cao, Agnès Rius-Escudé .....	26
DOES METALINGUISTIC KNOWLEDGE ABOUT THE L2 ENHANCE THE EFFECTIVENESS OF L2 PERCEPTUAL TRAINING? Angélica Carlet, Juli Cebrian, Núria Gavaldà, Celia Gorba .....	31
LECTURA EN VEU ALTA, PAUSES SILENCIOSES I IOGA NIDRA. Josefina Carrera-Sabaté, Jesús Bach Marquès .....	36
LA ENTONACIÓN DEL ESPAÑOL DE LOS ÁNGELES. UNIDAD VS. DIVERSIDAD. Yolanda Congosto Martín .....	41
POSITIONALLY CONSTRAINED ACCENTEDNESS CUES. A PILOT STUDY ON BARI ITALIAN. Giovina Angela del Rosso .....	46
ALLOPHONES OF /tr/ AFTER CODA /s/ ASPIRATION IN WESTERN ANDALUSIAN SPANISH. María del Saz .....	51
SENTENCE ACCENT IN ENGLISH AS A FOREIGN LANGUAGE: EXCEPTIONS TO THE LAST LEXICAL ITEM RULE. María del Saz, Ximena Grau .....	56
ESTUDIO ACÚSTICO Y DIALECTOMÉTRICO DE LA DURACIÓN VOCÁLICA: RELACIÓN DE PROXIMIDAD DEL OCCIDENTE, CENTRO Y ORIENTE CUBANO. Chaxiraxi Díaz, Josefa Dorta, Carolina Jorge .....	61
LA PROSODIA DE ACTOS DEL HABLA EN EL ESPAÑOL DE ASUNCIÓN, BUENOS AIRES Y CIUDAD DE MÉXICO. Natalia dos Santos Figueiredo, Carolina Gomes da Silva .....	66
DETERMINING SPEAKERS' ORIGIN ON THE BASE OF THEIR INTONATION: A PRELIMINARY STUDY. Wendy Elvira-García, Paolo Roseano, Ana Ma. Fernández Planas .....	71
UMBRAL DE INTELIGIBILIDAD EN AUDIOMETRÍAS VERBALES: VALORES DE REFERENCIA EN NIÑOS NORMOYENTES DE 5, 6 Y 7 AÑOS. Wendy Elvira-García, Paolo Roseano, Ana Ma. Fernández Planas, Silvia Marro, Pilar Calvo, M. Antonia Claveria .....	76
APLICACIONES DEL ULTRASOUND SOUND IMAGE (UTI) A LOS ESTUDIOS DE FONÉTICA. Wendy Elvira-García, Paolo Roseano, Lourdes Romera Barrios, Victoria Marrero Aguiar .....	80

ChaLeT: DETERMINING THE CHANCE LEVEL IN PERCEPTION EXPERIMENTS BY MEANS OF BINOMIAL TESTS. Wendy Elvira-García, Paolo Roseano, Assumpció Rost Bagudanch .....	85
USO DE SISTEMAS DE CLASIFICACIÓN SUPERVISADA PARA LA EVALUACIÓN DE LA PROSODIA: APLICACIONES AL HABLA DE PERSONAS CON SÍNDROME DE DOWN. David Escudero, Mario Corrales-Astorgano, Yolanda Martín de San Pablo, Alfonso Rodríguez de Rojas, Valle Flores, César González-Ferreras, Valentín Cardeñoso, Lourdes Aguilar .....	90
UNA HERRAMIENTA DE FONÉTICA APLICADA: MATERIALES DE LOGOAUDIOMETRÍAS INFANTILES EN CATALÁN. Ana Ma. Fernández Planas, Paolo Roseano, Wendy Elvira-García, Silvia Marro, Pilar Calvo, M. Antonia Clavería .....	95
EVIDENCIA ACÚSTICA Y ESTADÍSTICA DE UNA CLASE NATURAL DE EYECTIVAS EN KAWÉSQAR. Mauricio A. Figueroa Candia, Daniela A. Mena Sanhueza .....	101
¿DÓNDE ESTÁ LA ENTONACIÓN? PROPUESTA DIDÁCTICA INTEGRADORA DE LA ENTONACIÓN EN EL NIVEL A1. Andrea García Coronado, Beatriz Bleuca .....	106
PROPUESTA METODOLÓGICA PARA LA EVALUACIÓN DE FACTORES PROSÓDICOS EN EXPOSICIONES ORALES DE ALUMNOS DE SECUNDARIA. Clara Luna García García de León, Juan María Garrido Almiñana .....	110
APROXIMACIÓN AL ANÁLISIS MELÓDICO DE LA ENTONACIÓN NEUTRA DEL ESPAÑOL HABLADO POR RUSOHABLANTES. Zoya Garmátina, Empar Devís .....	114
DISEÑO DE UNA BATERÍA DE PRUEBAS PARA LA EVALUACIÓN DE LA PROSODIA DEL ESPAÑOL EN EL MARCO DEL PROYECTO PROA. Juan María Garrido Almiñana, Eva Estebas-Vilaplana, María J. Machuca, Lorraine Baqué, Wendy Elvira-García, Victoria Escandell Vidal .....	118
INVESTIGATING THE ROLE OF COGNATE STATUS IN L2 PRODUCTION AND PERCEPTION. Celia Gorba, Susana Cortés, Juli Cebrian, Núria Gavaldà .....	122
PERCEPTION OF L1 SPEECH OF LATE CZECH-FRENCH BILINGUALS BY CZECH MONOLINGUALS. M. Hevrova, B. Köpke, T. Bořil .....	127
FRICATIVAS ALVEOLARES DEL EUSKERA: APROXIMACIÓN A SU CARACTERIZACIÓN ARTICULATORIA. Alexander Iribar, Rosa Miren Pagola, Itziar Túrrez .....	132
MODELOS MIXTOS Y CRITERIOS DE INFORMACIÓN EN FONÉTICA. Miguel Jiménez-Bravo .....	135
RELEVANCIA DE LA DURACIÓN EN EL ESPAÑOL DE CANARIAS: ETIQUETAJE Y DIALECTOMETRÍA A PARTIR DE DATOS ACÚSTICOS. Carolina Jorge Trujillo, Josefa Dorta Luis, Chaxiraxi Díaz Cabrera .....	140
DISTANCIA DEL COSENO ENTRE VOCALES CEPSTRALES COMO MÉTODO PARA CUANTIFICAR LA DIFERENCIA ACÚSTICA ENTRE MARCAS EN CONFLICTO. José María Lahoz-Bengoechea, Marianela Fernández Trinidad .....	145
SPEECH PRODUCTION IN NOISE BY CHINESE LEARNERS OF ENGLISH. Bin Li, Qianwen Guan, Zixia Fan, Mosi He .....	150
LA PRODUCCIÓN DE LAS OCLUSIVAS CHINAS POR HISPANOABLANTES. Zhao Liu, María J. Machuca .....	154
LA FUNCIÓN DE LAS HESITACIONES EN LA IDENTIFICACIÓN DEL HABLANTE. Joaquim Llísterri, María J. Machuca, Antonio Ríos .....	160

LAS TÉCNICAS DE NEUROIMAGEN Y LA FONÉTICA EXPERIMENTAL. Victoria Marrero Aguiar .....	165
DESEMPEÑO FONÉTICO-ACÚSTICO DE OCLUSIVAS ÁFONAS EN HABLANTES ADULTOS MAYORES DEL ESPAÑOL CHILENO, DE LA CIUDAD DE CHILLÁN, CHILE. Renato Martínez-Cifuentes, Jaime Soto-Barba, Mauricio Figueroa-Candia, Katia Sáez-Carrillo .....	170
LEXICAL EFFECTS IN THE USE OF WORD-LEVEL STRESS FOR WORD RECOGNITION BY ENGLISH AND KOREAN LEARNERS OF SPANISH. María Teresa Martínez-García, Dong-Jin Shin, Goun Lee .....	174
UN CAMBIO ENTONATIVO EN EL HABLA DE LOS JÓVENES DE MAÓ Y ALAIOR (MENORCA). Ignasi Mascaró, Paolo Roseano .....	178
LA ENTONACIÓN DE LAS PREGUNTAS ABSOLUTAS DEL ESPAÑOL DE SAN JUAN (ARGENTINA) EN EL HABLA ESPONTÁNEA. Miguel Mateo Ruiz .....	182
UNA APROXIMACIÓN AL ANÁLISIS PROSÓDICO DE LA CONVERSACIÓN COLOQUIAL. Miguel Mateo Ruiz, Francisco José Cantero Serena .....	187
P-CENTRES AND NUCLEAR VOWEL DURATION IN THE CONTEXT OF IRISH (GAELIC) INITIAL MUTATION. Connor McCabe .....	192
“JUST COMPARE THAT!” THE USE OF INSTANT AUDIO MESSAGES IN FORENSIC PHONETICS. Chiara Meluzzi, Sonia Cenceschi, Nicholas Nese .....	197
THE ROLE OF ATTENTION CONTROL AND STIMULI LEXICAL STATUS IN HIGH VARIABILITY PHONETIC TRAINING. Joan C. Mora, Mireia Ortega, Íngrid Mora-Plaza, Jonás Fouz-González .....	201
SECUENCIAS VOCÁLICAS EN FRONTERA DE PALABRAS EN DATOS DEL ESPAÑOL MEXICANO. João Paulo Moraes Lima dos Santos .....	206
PROSODIA DEL ASTURIANO CENTROMERIDIONAL: RASGOS CARACTERÍSTICOS Y ANOTACIÓN EN ASTUR_ToBI. Carmen Muñiz Cachón, Paolo Roseano .....	210
VARIACIÓN SOCIAL EN LA REALIZACIÓN DE LA FRICATIVA /s/ EN EL ESPAÑOL DE ANTIOQUIA, COLOMBIA. Diana Muñoz-Builes, Wendy Elvira-García .....	215
ESTUDIO PRELIMINAR DE LA VOCAL ROTIZADA EN EL HABLA ESPONTÁNEA DE ANDALUCÍA Y ASTURIAS. Núria Ortiz-de-Pinedo Sánchez .....	220
L2 PHONETIC ACCURACY DEVELOPMENT IN A FOREIGNLANGUAGE SETTING: A LONGITUDINAL STUDY. Alejandra Pesantez, Volker Dellwo .....	225
PERCEPTUAL ASSIMILATION OF TASHLHIYT CONSONANTS BY SPANISH-DOMINANT BILINGUALS. Lucrecia Rallo Fabra, Mustapha Achichaou, Michael D. Tyler .....	229
FUNDAMENTACIÓN ARTICULATORIA Y AERODINÁMICA DE LOS PROCESOS FONOLÓGICOS. LAS ASIMILACIONES. Daniel Recasens .....	234
LOS RASGOS PROSÓDICOS DE LOS NEOLECTORES: UN ESTUDIO MULTIDIMENSIONAL CON UN ENUNCIADO INTERROGATIVO. Sara Recio-Pineda, Alicia Sola, Francisco José Cantero Serena .....	240
EMERGÈNCIA FONÈTICA: CAL APRENDRE I ENSENYAR A PRONUNCIAR LA [ʎ]. Agnès Rius-Escudé, Francina Torras .....	244
RHYTHMIC TRANSFER IN GALICIAN LEARNERS OF ENGLISH. Rosalía Rodríguez Vázquez, Paolo Roseano .....	249

TraFo: AN ONLINE G2P TOOL FOR SPANISH. Francesco Rodriquez .....	253
EL ESTUDIO DE LOS VOCATIVOS DEL ESPAÑOL EN EL MARCO DEL PROYECTO AMPER-CAT. Lourdes Romera Barrios, Paolo Roseano, Ana Ma. Fernández Planas, Francesco Rodriquez, Diana Muñoz Builes, Carolina Fabra Navarro, Eugenio Martínez Celdrán .....	257
THE LINGUISTIC RHYTHM OF FRIULAN: FIRST DATA. Paolo Roseano, Wendy Elvira-García, Francesco Rodriquez .....	262
A PHONETICALLY BASED VALIDATION OF THE CLASSIFICATION OF FRIULIAN DIALECTS. Paolo Roseano, Franco Finco .....	266
IS CATALAN SYLLABLE-TIMED? AN ANSWER BASED ON CLUSTER ANALYSIS. Paolo Roseano, Patrícia Marsà Morales, Laura Alañá Vilas .....	270
ACTOS DE HABLA INDIRECTOS Y SU INTERPRETACIÓN EN EL MARCO PROSÓDICO Y GESTUAL (EN LA BASE DE INTERACCIONES EN ESPAÑOL DE APRENDICES RUSOS). Albina Sarymsakova .....	275
WORD-FINAL [Z] IN ENGLISH PLURALIA-TANTUM AND REGULAR-PLURAL NOUNS: A COMPARISON OF DURATION. Marcel Schlechtweg, Greville G. Corbett .....	279
MELODÍAS ENFÁTICAS EN PORTUGUÉS DE BRASIL. Raquel Sena Mendes, Dolors Font-Rotchés .....	283
UN ESTUDIO PILOTO DE EJERCICIOS DE LA PRONUNCIACIÓN DE LAS PALABRAS ÁTONAS Y TÓNICAS CON AYUDA DE MÚSICA. Hirotaoka Sensui, Takuya Kimura y Miyuki Takasawa .....	288
ANÁLISIS COMPARATIVO DEL F0 RANGO Y DE LA FLUIDEZ ORAL ENTRE LOS HABLANTES CHINOS Y NATIVOS DEL ESPAÑOL. Peizhu Shang .....	293
PROSODIC ORGANIZATION OF ENGLISH FOLK RIDDLES AND THEIR EMOTIONAL AND PRAGMATIC POTENTIALS. Larysa Taranenko .....	298
SENSIBILIDAD PROSÓDICA ACENTUAL DE NIÑOS CON TRASTORNO DEL DESARROLLO DEL LENGUAJE DE SEGUNDO GRADO DE PRIMARIA. Valeska Torres Bustos, Bernardo Riffo Ocares, Katia Sáez Carrillo .....	302
¿ES MI PREGUNTA UNA INTERROGATIVA EN ESPAÑOL? LA ENTONACIÓN DEL ESPAÑOL HABLADO POR POLACOS. Weronika Urbanik-Pęk .....	307
ADQUISICIÓN DE LA ENTONACIÓN DE ELE: LA PRODUCCIÓN DEL FOCO ESTRECHO POR LOS SINOHABLANTES. Weiqi Li .....	313
INTERROGATIVAS Y ENTONACIÓN EN EL ESPAÑOL HABLADO POR CHINOS. Tianshu Zhao, Dolors Font-Rotchés .....	318

# MODELOS MIXTOS Y CRITERIOS DE INFORMACIÓN EN FONÉTICA

Miguel Jiménez-Bravo

Universidad Complutense de Madrid  
miguel.jimenez.bravo@ucm.es

## ABSTRACT

This presentation, belonging to the monographic session *New methodologies for the study of phonetics*, deals with how the analysis of data in phonetics is moving towards the use of linear mixed models (LMM) in order to overcome some limitations associated to ANOVAs. This paradigm shift is allowing to tackle more ambitious research questions and to enrich the possibilities of analysis. Similarly, the use of information criteria have become common in hypothesis testing, and it has been suggested that together with the  $p$ -value, also other values such as the false positive rate should be reported.

La actual presentación, como parte de la sesión monográfica *Nuevas metodologías para el estudio fonético*, trata sobre cómo el análisis de los datos en fonética está cediendo al uso de modelos lineales mixtos (MLM) para solventar problemas asociados a ANOVA. Este cambio de paradigma está permitiendo abordar preguntas de investigación más ambiciosas y realizar análisis más complejos. Igualmente, el uso de criterios de información es cada vez más común en el proceso de validación de hipótesis y se ha propuesto que, junto al valor- $p$ , también se comuniquen otros, como la tasa de falsos positivos.

**Keywords:** statistics, phonetics, linear mixed models, information criteria,  $p$ -value

**Palabras clave:** estadística, fonética, modelos lineales mixtos, criterio de información, valor- $p$

## 1. INTRODUCCIÓN

El uso de la estadística inferencial para estudiar un determinado fenómeno –como complemento a la estadística descriptiva– permite generalizar conclusiones a partir de unos datos tomados de una muestra representativa. Para ello es posible aplicar diferentes tests estadísticos dependiendo de cuál sea el tipo de análisis requerido y de la naturaleza de los datos. Entre los análisis más frecuentes están los que permiten correlacionar diferentes grupos de datos (p.ej. coeficiente de correlación de Pearson) o comparar sus medias aritméticas para establecer si dos grupos de datos son diferentes (p.ej.  $t$ -Student). En la mayoría de los análisis resultan clave tanto las medidas de tendencia central, es decir, las que dan cuenta de cómo se distribuyen los datos en torno a un determinado valor, como las medidas de dispersión, que muestran cómo es la variabilidad de su distribución.

La idea básica es que cualquier análisis que permita generalizar conclusiones a partir de una muestra representativa siempre requiere la creación de un modelo, a una escala reducida, del fenómeno estudiado; es decir, un modelo estadístico, sea éste un modelo sencillo, como la media aritmética y su

desviación estándar, o más complejo, como el análisis de la varianza de un conjunto de datos (ANOVA).

A su vez, el uso de estos modelos estadísticos para realizar inferencias estadísticas es fundamental para confirmar o rechazar, con un cierto grado de certeza, cualquiera que sea la hipótesis inicial de la investigación, la *hipótesis nula*. Este grado de certeza, el valor- $p$ , equivale a la probabilidad de que un determinado resultado no sea fruto del azar y es un valor arbitrario y determinado de antemano.

Sin embargo, tanto el uso de ciertos modelos estadísticos para extrapolar los resultados obtenidos, como la búsqueda obstinada del valor- $p$  con el fin de confirmar una cierta hipótesis no siempre son el modo más acertado de aproximarse a los datos. En lo que sigue se abordarán las razones que han dado lugar a un cambio de paradigma en numerosas ciencias, entre ellas la lingüística.

## 2. LIMITACIONES DE ANOVA

En lingüística, diversos ámbitos –entre ellos la fonética experimental– han presenciado la llegada de un modo alternativo de analizar estadísticamente los datos (Quené y van den Bergh, 2008; Quené, 2013;

Torreira et al., 2014; Gahl, 2015; Adamou et al., 2017; Kirby y Sonderegger, 2018; Speelman et al., 2018). Los modelos estadísticos empleados tradicionalmente, entre ellos a menudo diferentes tipos de análisis de la varianza (ANOVA), han ido dando paso al uso de modelos lineales mixtos (MLM), especialmente en casos de medidas repetidas, datos de tipo categórico y diseños incompletos (Jaeger, 2008).

Uno de los problemas con los que suelen encontrarse los investigadores al recabar datos es la necesidad de registrar más de una respuesta por cada participante (o por palabra, si es esta la unidad de observación). Por ejemplo, en un hipotético estudio sobre la realización de la prominencia acústica en función de las diferencias que puede haber en el timbre vocálico, se registra la duración de la vocal tónica de un conjunto de pseudopalabras (Tabla 1).

**Tabla 1.** Tabla ficticia de datos para un experimento sobre la realización de la prominencia acústica con medidas repetidas.

Pseudopalabra	Vocal	Duración (seg)
1	[a]	0.045
1	[o]	0.034
1	[i]	0.029
2	[a]	0.035
2	[o]	0.031
2	[i]	0.028
3	[a]	0.037
3	[o]	0.041
3	[i]	0.034

Siguiendo la lógica del análisis de la varianza (ANOVA), que permite comparar las medias de más de dos grupos de datos, las observaciones repetidas registradas para cada pseudopalabra violan el principio de independencia de las observaciones – más exactamente, de los errores de las observaciones–, el cual constituye un pilar sobre el que se asientan los modelos de regresión lineal y ANOVA. Este principio puede, sin embargo, ser respetado si se emplea un análisis de la varianza para medidas repetidas (RM-ANOVA) (p.ej. Field et al., 2012, cap. 13).

No obstante, los modelos RM-ANOVA tienen serias limitaciones para analizar datos donde faltan observaciones (p.ej. Raaijmakers et al., 1999, Quené y van den Bergh, 2004). En estos casos, los datos

ausentes pueden reemplazarse por la media del participante, lo cual distorsiona las estimaciones; o pueden descartarse junto a todas las demás observaciones registradas para ese participante. De modo similar, se desaconseja utilizar RM-ANOVA en diseños experimentales incompletos, como por ejemplo los diseños entregrupos, donde los participantes no reciben todos los estímulos, o en casos donde de haya una aleatorización de estímulos mediante la técnica de cuadrado latino para contrarrestar efectos de aprendizaje.

Supongamos, sin embargo, que quisiera aplicarse el conjunto de estímulos del experimento anterior a un grupo de participantes en un estudio sobre la percepción de la prominencia en función del tipo de vocal y su duración. En este caso la variable dependiente sería categórica y no continua (Tabla 2), y como se ha señalado en varias ocasiones, la transformación arco-seno necesaria para analizar datos categóricos en ANOVA a menudo incurre en errores de estimación (p.ej. Jaeger, 2008; Warton y Hui, 2011).

**Tabla 2.** Tabla ficticia para un experimento sobre percepción de la prominencia acústica. La variable dependiente *prominencia* es categórica.

Partic.	Pseudopal.	Vocal	Dur.	Prom.
1	1	[a]	0.045	SI
1	1	[o]	0.034	NO
1	1	[i]	0.029	SI
1	2	[a]	0.035	NO
1	2	[o]	0.031	SI
1	2	[i]	0.028	SI

En este caso, además, el diseño multifactorial del experimento presenta otro inconveniente. Con los participantes como unidad de observación, los estímulos muestran claramente un jerarquía en la que cada vocal cuenta con tres niveles, uno para cada palabra. Es decir, suponiendo que hubiese 12 participantes a los que se les administrasen 3 pseudopalabras para cada vocal, se registrarían 9 respuestas por participante, dando un total de 108 observaciones (12x3x3). Tradicionalmente se ha buscado la forma de modelar la correlación que existe en las observaciones para los niveles de un mismo factor –puesto que estos tienden a obtener resultados más similares entre sí–, lo cual incrementa los errores de falso positivo, o errores de tipo I (Coleman, 1964; Clark, 1973). La solución más común a este problema ha sido analizar de forma separada la varianza que los estímulos aportan

al modelo para que ésta no afecte a las estimaciones; y lo mismo se ha hecho para la correlación en las observaciones registradas para los participantes. Por ello, tanto estímulos como participantes a menudo han sido tratados en el modelo estadístico como efectos aleatorios, y por tanto diferenciados de los efectos fijos, es decir, los efectos reales que se pretenden analizar, causados por la variable independiente.

### 3. MODELOS LINEALES MIXTOS (MLM) Y GENERALIZADOS (MLMG)

Frente a lo expuesto anteriormente, los modelos lineales mixtos (MLM) –en ocasiones también llamados modelos de regresión jerárquica o de regresión multinivel– son simplemente modelos de regresión lineal que no solo solventan de manera sencilla la mayoría de las limitaciones enumeradas, sino que además permiten declarar fácilmente los efectos aleatorios de estímulos y participantes.

Por ejemplo, en un estudio real, similar al ejemplo mencionado, se investigó la influencia de la gesticulación de los hablantes y el peso de los correlatos acústicos en la percepción de la prominencia (Jiménez-Bravo, 2020). Los participantes de un experimento tenían que determinar qué palabras de una serie de frases les resultaban prominentes, y el diseño multifactorial incompleto incluía, como variables independientes: dos modalidades perceptivas (auditiva y audiovisual); los valores acústicos de la vocal/sílaba tónica (frecuencia fundamental, intensidad y duración) homogeneizados en tres condiciones experimentales; y las diferentes fases en las que se dividían los gestos manuales realizados por el hablante al emitir la frase (Fig. 1).

Con este diseño experimental, habría sido imposible abordar el análisis estadístico de los datos mediante cualquiera de los análisis ANOVA. Por un lado, las respuestas repetidas para cada participante violan el principio de independencia de las observaciones e introducen una correlación para cada participante. Además, puesto que estos no podían juzgar la misma frase en ambas modalidades, los participantes no recibieron todos los estímulos, sino que un grupo recibió la mitad de los estímulos en la modalidad auditiva, mientras que el otro grupo los recibió en la modalidad audiovisual. Además, los estímulos presentan una estructura jerárquica de palabras anidadas dentro de frases, las cuales habían sido a su vez emitidas por diferentes hablantes.

Sí es posible, sin embargo, analizar los datos utilizando MLMs con el software estadístico R (2019), concretamente mediante el paquete *lme4*

(Bates et al., 2015), el cual incluye la función *glmer* para modelar respuestas categóricas de tipo binario como la del experimento mencionado. Esta función se basa en una variante de los MLMs llamada modelos lineales mixtos generalizados (MLMG), capaz de analizar distribuciones no normales utilizando un modelo de regresión logística. Por otro lado, y de acuerdo con la lógica de los MLMs, la formulación del modelo estadístico en R permite incluir la variabilidad que los estímulos y los participantes introducen en el modelo con el fin de que ésta no afecte a la estimación de los efectos fijos (Fig. 1).

```
prominencia ~ modalidad + condición + f0 + intensidad + duración + fase-gesto +
+ (1 | participante) + (1 | hablante/frase/palabra)
```

**Figura 1.** Ejemplo simplificado de la formulación del modelo estadístico en R. La notación (1 | ...) estima la variabilidad de participantes y estímulos como efectos aleatorios.

### 4. INFERENCIA: CRITERIO DE INFORMACIÓN Y VALOR-P

Como en otros modelos de regresión múltiple, es posible incluir más de una variable independiente en los MLMs y por ello es necesario averiguar cuáles de ellas tienen un efecto real en el fenómeno estudiado; es decir, ha de ajustarse el modelo para que éste describa la realidad de la mejor manera posible.

En el caso de los MLMs no existen fórmulas perfectas para realizar esta estimación, por lo que a menudo se realiza una comparación entre modelos basada en la ‘razón de verosimilitud’ (*likelihood ratio*). En esta comparación, el modelo difiere solamente en un único efecto –fijo o aleatorio– respecto al modelo con el que se compara –el modelo *nulo*– y el valor-*p* resultante de esta comparación determina la contribución que tiene el efecto en cuestión sobre la variable dependiente.

Sin embargo, varias críticas han señalado debilidades asociadas a este procedimiento (p.ej. Bolker et al., 2009) y han censurado la búsqueda obstinada del valor-*p*, lo cual se ha relacionado con la dificultad de reproducir resultados que otros investigadores previamente calificaron como ‘estadísticamente significativos’ (Cumming, 2014; Halsey et al., 2015; Colquhoun, 2017).

Como alternativa se ha propuesto el uso de procedimientos basados en criterios de información, los cuales no se limitan a comparaciones únicamente de dos modelos que difieren en un solo efecto, sino que permiten comparar los efectos de

varios de ellos (Burnham y Anderson, 2002). Por ejemplo, el criterio de información de Akaike (AIC) (1973), basado en el concepto de ‘distancia’ entre dos modelos (Kullback y Leibler, 1951), permite comparar varios modelos a partir del número de efectos incluidos en cada uno. El valor AIC, por tanto, es solamente informativo: clasifica los modelos para seleccionar el que mejor describe la realidad a partir de los datos.

Las recomendaciones en el uso de este procedimiento aconsejan incluir, partiendo de un modelo inicial, sólo los efectos que estén justificados por las preguntas de investigación, para ir desestimando progresivamente aquellos que no resulten significativos. También es importante no declarar un conjunto excesivo de efectos en los modelos construidos a partir de este modelo inicial y se recomienda, además, que aquellos modelos a menos de 2 puntos AIC del mejor modelo sean también incluidos en los resultados (Arnold, 2010; Grueber et al., 2011). El paquete *AICcmodavg* (Mazerolle, 2017) permite comparar con facilidad en R los modelos a partir de su valor AIC.

Finalmente, como se ha señalado, muchas veces resulta insuficiente dar cuenta únicamente del valor-*p* al comunicar los resultados. Dada la probabilidad de incurrir en un error de tipo I y considerar como significativo un efecto que en realidad no lo es, se ha propuesto, entre otras cosas, incluir también la tasa de falsos positivos, el tamaño del efecto y los intervalos de confianza (Colquhoun, 2017; Cumming, 2014). En este sentido, por ejemplo, la tasa de falsos positivos de un modelo puede calcularse fácilmente en R mediante el paquete *caret* (Kuhn, 2008), el cual permite crear una tabla de confusión que cruza las predicciones hechas por el modelo con los datos reales, para así valorar su precisión –o poder predictivo– (Kuhn y Johnson, 2013) (Tabla 3). En este sentido, como ironiza Andy Field (2012, p. 339): “Running a regression without checking how well the model fits the data is like buying a new pair of trousers without trying them on – they might look fine on the hanger but get them home and you find you’re Johnny-tight-pants.”

Por tanto, la tasa de falsos positivos, como uno de los elementos asociados a la validación del modelo estadístico se convierte en un valor que permite contextualizar el valor-*p* necesario en la inferencia. No obstante, este proceso de validación del modelo, en el que se cruzan las predicciones con los datos reales, puede realizarse bien utilizando el conjunto completo de datos bien dividiéndolos en dos subconjuntos –a menudo 70/30–, en lo que se conoce como validación cruzada (*cross-validation*).

Esta división de los datos permite ‘entrenar’ previamente el modelo con el conjunto menor de los datos y realizar las predicciones sobre un conjunto de datos diferente; en este caso, a falta de un conjunto de datos independiente a la investigación, la predicción se hace sobre la partición mayor, obtenida a partir de los datos originales.

**Tabla 3.** Matriz de confusión para un modelo sobre la percepción multimodal de la prominencia (Jiménez-Bravo, 2020). La precisión corresponde a la suma de las predicciones correctas dividida por el gran total. El valor 117 muestra los falsos positivos, equivalente a una tasa del 19%.

		Datos		
		No prom	Prom	Total
Predicción	No prom	500	250	750
	Prom	117	367	484
	Total	617	617	1234
Precisión		70%		

## 5. CONCLUSIONES

Este cambio de paradigma en el análisis estadístico surge de la limitación que los diferentes análisis de la varianza (ANOVA) presentan al analizar los datos. Ello hace posible abordar preguntas de investigación más ambiciosas y realizar análisis más complejos. En el ámbito de la lingüística, numerosos estudios fonéticos están haciendo uso de modelos lineales mixtos (MLM) o de su variante para distribuciones no normales, MLMs generalizados (MLMG). De forma similar, el criterio de información de Akaike (AIC) se hace enormemente útil al permitir comparar modelos con múltiples variables independientes con el fin de seleccionar aquel que mejor dé cuenta de los datos.

Por otro lado, el proceso de inferencia está estrechamente asociado al valor-*p*. No obstante, es necesario complementarlo con otros valores que ayuden a replicar los resultados. Entre ellos, destacan el tamaño del efecto y sus intervalos de confianza, así como la tasa de falsos positivos, que no es más que uno de los elementos propios de la validación del modelo estadístico. Este nuevo paradigma ha ido de la mano del desarrollo de R, cuya utilización está haciendo que la disciplina se enriquezca. Todo ello, sin embargo, plantea la necesidad de establecer criterios de buenas prácticas que permitan llevar a cabo tanto un correcto análisis de los datos como una adecuada comunicación de los resultados.

## 6. BIBLIOGRAFÍA

- Adamou, E., Gordon, M., Gries, S.T. (2018). Prosodic and morphological focus marking in Ixcatec (Otomanguan). In E. Adamou, K. Huade, & M. Vanhove (Eds.) *Information Structure in Lesser-described Languages*, (pp. 51–84). Amsterdam: John Benjamins.
- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.) *Proceedings of the 2nd International Symposium on Information Theory*, (pp. 267–281). Budapest: Akadémiai Kiadó.
- Arnold, T.W. (2010). Uninformative parameters and model selection using Akaike's Information Criterion. *Journal of Wildlife Management*, 74(6), 1175–1178.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1), 1–48.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3), 127–135.
- Burnham, K.P., & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: a Practice Information-Theoretic Approach*. Berlin/Heidelberg/New York: Springer-Verlag.
- Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359.
- Coleman, E.B. (1964). Generalizing to a language population. *Psychological Reports*, 14, 219–226.
- Colquhoun D. (2017). The reproducibility of research and the misinterpretation of *p*-values. *Royal Society Open Science*, 4.
- Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, 25(1), 7–29.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics using R*. London: SAGE.
- Gahl, S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics*, 49, 96–116.
- Grueber, C.E., Nakagawa, S., Laws, R.J., & Jamieson, I.G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology*, 24(4), 699–711.
- Halsey, L.G., Curran-Everett, D., Vowler S.L., & Drummond G.B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, 12(3), 179–85.
- Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Archives of General Psychiatry*, 59(4), 434–446.
- Jiménez-Bravo, M. (2019). *Multimodal perception of acoustic prominence in Spanish*. Tesis doctoral. UNED.
- Jiménez-Bravo, M. (2020). Multimodal Perception of Prominence in Spontaneous Speech: A Methodological Proposal using Mixed Models and AIC. *Speech Communication*, 124, 28–45.
- Kirby, J., & Sonderegger, M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics*, 70, 70–85.
- Kuhn, M., 2008. Building Predictive Models in R using the caret Package. *Journal of Statistical Software*, 28(5), 1–26.
- Kuhn M., & Johnson K. (2013). Measuring Performance in Classification Models. In: *Applied Predictive Modeling*, (pp. 247–273). New York: Springer.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1), 79–86.
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425.
- Quené, H. (2013). Longitudinal trends in speech tempo. *Journal of the Acoustical Society of America*, 133 (6), EL453–EL457.
- R Development Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Computer program. <https://www.R-project.org/>.
- Raaijmakers, J.G., Schrijnemakers, J.M., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416–426.
- Speelman, D., Heylen, K., & Geeraerts, D. (Eds.) (2018). *Mixed-Effects Regression Models in Linguistics*. Amsterdam: Springer.
- Torreira, F., Simonet, M., & Hualde, J.I. (2014). Quasi-neutralization of stress contrasts in Spanish. In *Proceedings of the 7<sup>th</sup> International Conference on Speech Prosody (SP2014)*, (pp. 197–201). Dublin, Ireland.
- Warton, D.I., & Hui, F.K.C. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, 92(1), 3–10.