Factores influyentes en tipos de errores de concordancia para un corpus de cuatro aprendientes italianos de español L2 (material suplementario)

1. Selección de modelos

Se pueden usar las siguientes medidas para seleccionar modelos (Stroup, 2013; p. 193; Burnham & Anderson, 2010, caps. 2 y 4) [$\theta = \hat{\sigma}, \hat{\beta}$ es el vector de los coeficientes fijos y aleatorios estimados]:

(a) <u>Schwarz</u>: $BIC = -2L(\theta) + (p \times log(s))$ [s= número de grupos; y $p = p_{\sigma} + (p_{\beta} = rank[X])$; o sea el número de parámetros fijos más los aleatorios]. Menos es mejor.

(b) <u>Akaike</u>: $AIC = -2L(\theta) + 2p$. Menos es mejor.

(c) <u>Akaike corregido</u>: $AICc = -2L(\theta) + 2p(n^*/(n^* - p - 1))$ [$n^* = N$, tamaño muestral]; Corrige por muestra pequeña. Menos es mejor. Como heurística, se debería usar cuando[1]: $\frac{n}{p} < 40$.

(d) <u>Delta de Akaike</u>: $\Delta AIC = \Delta = AIC_i - AIC_{min}$. Indican la distancia del modelo al mejor de todos (el de menor AIC). $\Delta_i \leq 2$ indica evidencia substancial para el modelo i.

(e) <u>Pesos de Akaike</u> (ω_i): indican el peso de la evidencia en favor de que modelo sea el mejor de entre todos los modelos candidatos. Es decir, responde a la pregunta: ¿Cómo soportan los datos al modelo i con respecto al resto de los modelos? Se define como:

$$\omega_{modelo_i} = \frac{exp(-\Delta_i/2)}{\sum_{i=1}^{R} exp(-\Delta_i/2)}$$

donde $i = 1, \dots, R$ son los modelos considerados; y $\sum_{i=1}^{R} w_i = 1$.

(g) <u>Ratio de evidencia</u> ("Evidence Ratio", [ER]): Ratio entre el peso de Akaike del modelo i-ésimo y el peso de Akaike del j-*ésimo* modelo: $\frac{W(i)}{W(j)}$. Muchas veces resulta de interés establecer i como el índice del mejor modelo: $\frac{W(1)}{W(j)}$. Los ER son invariantes a los demás modelos, a parte de i y j. Responden a la pregunta: ¿Cuántas más veces apoyan los datos al (mejor) modelo i respecto del modelo j?

Una vez ordenados los modelos según alguno de los criterios, se puede reducir dicho conjunto por medio de un "conjunto de confianza" [*confidence set*] para el mejor modelo hallado. Burnham & Anderson (2010, p. 169) plantean tres alternativas: (i) sumar los pesos de Akaike

[1] Las medidas AIC y AIC_c convergen para n grande (manteniendo p constante). Es decir que cuando dicho ratio es suficientemente grande, tienden a seleccionar el mismo modelo. Entonces, en la práctica conviene usar siempre AIC_c.

de los modelos hasta alcanzar ≥ 0.95 (recuérdese que los pesos de Akaike suman *1*); (ii) tomar los modelos tal que $\Delta_i \leq 2$, ya que indican evidencia sustancial para el modelo i; (iii) establecer un corte usando ratios de evidencia (poniendo ahora el mejor modelo en el denominador), tal que[2]: $\frac{W(i)}{W(1)} > \frac{1}{8}$ $(\Delta_i = 2)$. Los autores prefieren el tercer criterio debido a su invariancia por adición o borrado de modelos del conjunto de confianza.

Resulta imperativo tener en cuenta la incerteza debida al proceso de selección de modelos. De R modelos considerados se selecciona el mejor modelo i. Sin embargo, ¿Si hubieran cambiado los datos, se elegiría igualmente el modelo i como el mejor o habría variabilidad de entre las muestras de datos en cuanto al modelo elegido?. Una forma de tener en cuenta dicha incerteza es estimar la probabilidad de que un determinado predictor x_j esté en el mejor modelo si se pudiera recoger una nueva muestra de datos. Se trata de una medida de importancia relativa de los predictores. Se lleva a cabo sumando los pesos de Akaike de los modelos en los cuales el predictor x_j está presente: $W_+ = w_i I_j(g_i)$; donde $I_j(g_i)$ es la función indicadora que es "1" si x_j está en el modelo g_i o cero, si no. Entonces, la importancia relativa es la proporción de modelos en los cuales la predictora está presente.

Si se diera el caso de que, por ejemplo, $w(i) > 0.9$, entonces el modelo i es un claro ganador. En dicho caso es válido hacer inferencia mediante la estimación de los coeficientes β_i y sus errores típicos serán condicionales al modelo seleccionado. Si embargo, muchas veces, especialmente si el conjunto de modelos a considerar es grande, los modelos con $\Delta_i \leq 2$ poseen pesos de Akaike similares o bien deltas de Akaike cercanos al cero. En este caso, β_i puede diferir en los modelos del conjunto considerado. Una solución es usar la información de todos los modelos involucrados mediante un promedio pesado de los coeficientes. En este caso, los errores típicos de los coeficientes estimados no son condicionales al modelo (ganador) en cuestión sino a todo el conjunto de modelos. Por lo tanto, dichos errores típicos "incondicionales" tienen en cuenta la varianza que proviene del proceso de selección de modelos. Para promediar los coeficientes se utilizó:

$$\bar{\beta}_j = \sum_{i=1}^{R} w_i \, I_j(g_i) \; \hat{\beta}_{j,i} = W_+ \, \hat{\beta}_{j,i}$$

donde:

$$I_j(g_i) = \begin{cases} 1 & x_j \in g_i \\ 0 & x_j \notin g_i \end{cases}$$

y la suma es sobre todos los modelos del conjunto: $i = 1, \dots, R$. En este estimador se usan todos los modelos ("full average"), y cuando la predictora x_j no estuviera presente en un determinado modelo entonces $\beta_j = 0$. Tiene la ventaja de "correr hacia cero" [*Shrinkage*] las estimaciones de parámetros presentes en "modelos malos". La varianza del estimador resulta:

---

[2] También podrían usarse: *0.135* $(\Delta_i = 4)$; *0.082* $(\Delta_i = 5)$; *0.05* $(\Delta_i = 6)$

$$\widehat{var}(\bar{\beta}_j) = \left[\sum_{i=1}^{R} w_i \sqrt{\widehat{var}(\bar{\beta}_j|g_i) + (\hat{\beta}_j - \bar{\beta}_j)^2}\right]^2$$

y su error típico: $\sqrt{\widehat{var}(\bar{\beta}_j)}$.

Se llevó a cabo selección de modelos (multinomiales generalizados) basado en medidas de información (Burnham & Anderson, 2010), optimizando la función de log-verosimilitud por medio de una red neuronal (utilizando el paquete *nnet* de *R* [Venables & Ripley (2002), p. 203]). Se decidió dividir el problema en dos grupos de variables, ya que utilizar todas las discretas implicaba una búsqueda exhaustiva aproximada de *4* millones de modelos. El primer grupo contenía en el modelo global a las variables predictoras: Fabs.SC.f, MORF.f, STEM.f, MOD, ES, ANIM, GRAMS, FAM.LEX.f, IMA.CONC.f, CUMRES.f, LDA. Mientras que el segundo contenía: POS, DIS, EST1, EST2, EST3, EST4, EST5, EST6, EST7, GRUPO6. Se usó AIC como medida de información para la selección. Se reporta la importancia relativa de las predictoras sobre *2048* modelos del grupo I (Cuadro *1*), y *1024* del grupo II (Cuadro *2*), sobre el total de modelos. Se observó que LDA, STEM.f, GRAMS, DIS, POS, EST4 poseían probabilidades debajo del *50* %. A continuación, se llevó a cabo el promedio de los coeficientes en el conjunto de "confianza" de los modelos (con la regla $\frac{W(i)}{W(1)} > \frac{1}{8}$). Las predictoras promediadas que resultaron significativas fueron: Fabs.SC.f, MOD, ANIM, FAM.LEX.f, ES, MORF.f, CUMRES.f, EST1, EST2, EST5, GRUPO6 (Cuadros *3* y *4*).

Tabla 1. Importancia relativa de las predictoras: grupo 1

	Names	x
1	Fabs.SC.f	1.00
2	ANIM	1.00
3	MORF.f	1.00
4	CUMRES.f	1.00
5	FAM.LEX.f	1.00
6	MOD	0.99
7	ES	0.90
8	IMA.CONC.f	0.72
9	LDA	0.44
10	STEM.f	0.22
11	GRAMS	0.11

Tabla 2. Importancia relativa de las predictoras: grupo 2

	Names	X
1	EST5	1.00
2	EST1	0.98
3	GRUPO6	0.96
4	EST2	0.93
5	EST6	0.82
6	EST7	0.75
7	EST3	0.62
8	EST4	0.48
9	DIS	0.27
10	POS	0.06

Tabla 3. Promedio de los coeficientes con FULL AVERAGE ($p < 0.05$): grupo 1

	Estimate	Std. Error	z value	Pr(> |z|)
1((Intercept))	-3.799	0.691	5.494	0.000
1(Fabs.SC.f1)	-1.171	0.320	3.661	0.000
1(MOD2)	0.646	0.309	2.089	0.037
2((Intercept))	-3.543	0.924	3.835	0.000
2(ANIM1)	1.833	0.431	4.248	0.000
2(FAM.LEX.f1)	-1.317	0.424	3.102	0.002
3((Intercept))	-0.724	0.294	2.461	0.014
3(CUMRES.f2)	0.470	0.173	2.711	0.007
3(ES2)	-18.804	8.832	2.129	0.033
3(FAM.LEX.f1)	-0.330	0.137	2.408	0.016
3(Fabs.SC.f1)	-0.402	0.168	2.395	0.017
3(MOD2)	0.353	0.175	2.018	0.044
3(MOD3)	0.482	0.199	2.425	0.015
3(MORF.f1)	-1.090	0.231	4.720	0.000
3(MORF.f2)	-0.691	0.304	2.272	0.023
4((Intercept))	-3.841	0.598	6.420	0.000
4(CUMRES.f1)	1.114	0.363	3.067	0.002

4(CUMRES.f2)	1.297	0.365	3.554	0.000
4(FAM.LEX.f1)	-0.464	0.230	2.019	0.043
4(MOD2)	1.074	0.293	3.659	0.000

Tabla 4. Promedio de los coeficientes con FULL AVERAGE ($p < 0.05$): grupo 1

	Estimate	Std. Error	z value	Pr(> |z|)
1((Intercept))	-2.904	0.324	8.949	0.000
1(EST11)	-1.693	0.534	3.169	0.002
2((Intercept))	-2.946	0.370	7.954	0.000
2(EST21)	-2.629	1.053	2.496	0.013
2(GRUPO66)	1.419	0.546	2.600	0.009
3((Intercept))	-1.845	0.176	10.506	0.000
3(GRUPO63)	0.544	0.167	3.256	0.001
3(GRUPO65)	0.974	0.300	3.245	0.001
4((Intercept))	-2.914	0.316	9.226	0.000
4(EST51)	-0.971	0.460	2.111	0.035
4(GRUPO62)	0.930	0.369	2.517	0.012

2. Modelo multinomial mixto bayesiano

Desde la perspectiva bayesiana los parámetros se consideran variables aleatorias y, como tales, llevan asociada una distribución, que a su vez, depende de un parámetro genérico ψ. Si se quiere hacer inferencia, la pregunta bayesiana sería: ¿cuáles son los valores del parámetro θ, que explican los datos Y (que están fijos)? Es decir que la respuesta es una *distribución* de dichos valores, que representa la incerteza en la estimación del parámetro θ. A esta distribución se la denomina "posterior" y se la denota: $p(\theta \mid y)$. Asigna diferente grado de probabilidad ("creencia") a los posibles valores del parámetro θ *luego* de ver la evidencia. El problema de inferencia se resuelve mediante el teorema de *Bayes*, que reza:

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta \mid \psi)}{p(y)}$$

En donde: (i) $p(y \mid \theta)$ es la verosimilitud de los datos y (o sea, la distribución de y variando los valores del parámetro y dejando fijos los datos), que depende de parámetro genérico θ; (ii) $p(\theta \mid \psi)$ es la distribución del parámetro θ, que depende del parámetro genérico ψ. $p(y)$ es la distribución marginal de los datos: $p(y) = \int_{\theta} p\,(y \mid \theta)p(\theta)d\theta$. Como ésta última no depende de θ (porque al integrar sobre θ se obtiene una constante), se la considera como constante de normalización. La idea es asignar, *antes* de ver los datos, una distribución "prior" al parámetro θ, que indica una creencia sobre los posibles valores de θ. Esta creencia es actualizada por la distribución de θ teniendo en cuenta la evidencia (datos). En cualquier caso, la distribución posterior siempre es un compromiso entre verosimilitud de los datos y el *prior* del parámetro. La distribución posterior puede resumirse según: (i) la moda; (ii) la media; (iii) la mediana; (iv) el intervalo de *credibilidad* del *95* %. Este último da los valores de θ que dejan a la derecha y a la izquierda una densidad de probabilidad del *0.025*. Es decir que indica la probabilidad de que el parámetro esté en un intervalo de credibilidad del 95 %.[3].

La distribución posterior para el modelo mixto es (Sorensen & Gianola, (2002), cap.14.2):

$$p(\beta, u, \sigma_u^2 \mid y) = p(y \mid \beta, u)p(\beta)p(u \mid \sigma_u^2)p(\sigma_u^2 \mid \psi)$$

Las variables aleatorias se distribuyen como sigue ($c = 1,2,3,4$): (1) Respuesta: $y_{ijc} \mid \beta, u \sim \mathcal{M}\big(N; (\pi_1, \pi_2, \pi_3, \pi_4)\big)$; (2) Coeficientes: $\beta_{kc} \mid \sigma_\beta^2 \sim \mathcal{N}\big(0, \sigma_\beta^2 = \sigma_e^2 + \pi^2/3\big)$; en notación vectorial: $\boldsymbol{\beta} \mid \sigma_\beta^2 \sim NMV\big(\mathbf{0}, \mathbf{B} \otimes \mathbf{V}_{\boldsymbol{\beta}}\big)$, **B**=I.; (3) Factores aleatorios: $u_j \mid \sigma_u^2 \sim \mathcal{N}(0, \sigma_u^2)$; donde: $\sigma_u^2 \sim \chi_u^2(1)$ [chi-cuadrada con 1 grado de libertad]; en notación vectorial: $\mathbf{u} \mid \sigma_u^2 \sim NMV(\mathbf{0}, \mathbf{A} \otimes \mathbf{V}_\mathbf{u})$; **A**=I; $V \sim W(\nu, S)$ [Wishart].

Las matrices de varianza-covarianza $\mathbf{V}_\mathbf{u}$ de los factores aleatorios son diagonales heterogéneas, con una varianza para cada efecto aleatorio. Los efectos aleatorios se consideran independientes. Por ejemplo, en el modelo más simple habrá cuatro efectos aleatorios de ordenada al origen, uno por cada categoría c, para *cada* grupo. Entonces la matriz tendrá en su diagonal las varianzas para cada uno de ellos. Es necesario estimar las distribuciones posteriores de los coeficientes (β), los factores aleatorios (u) y la varianza de los factores aleatorios (σ_u^2). La varianza de los errores no es un parámetro a estimar porque no es independiente de la media para los modelos (multi/bi)nomiales. Según recomendaciones de Hadfield (2010) se fijará la varianza de los errores en $V = \frac{1}{k}\big(I_{(k-1)} + J_{(k-1,k-1)}\big)$, donde k el el número de categorías de la respuesta, I es la matriz identidad y J, una matriz de *unos*. Siguiendo a Gelman (2006) se usó el siguiente prior para β: $\beta_k \sim N(0, \sigma_e^2 + \pi^2/3)$. Por otra parte, se tomará la sugerencia de Villemereuil et al. (2013) [ver también: Villemereuil (2012)] de utilizar *priors expandidos* para σ_u^2, y usando una distribución χ^2 (que pertenece a la familia *gama*) con un grado de libertad[4].

---

[3] En cambio, en estadística frecuentista el intervalo de confianza indica que, si se repitiera muchas veces el experimento con diferentes muestras, el 95 % de la veces el verdadero parámetro poblacional caerá en el intervalo. Como no es una variable aleatoria (y por lo tanto carece de distribución) esto no es lo mismo que decir que hay una probabilidad del 95 % de que el parámetro caiga dentro del intervalo.

[4] Se expande el efecto aleatorio u_i en dos componentes: $u_i = \alpha\eta_i$ con $\eta_i \sim \mathcal{N}\big(0; V_\eta\big)$; $\alpha \sim \mathcal{N}(0, V_\alpha)$; $V_\eta \sim W^{-1}(S, \nu)$ [Wishart inversa]. Implícitamente es lo mismo que definir $V_u = V_u/V_\alpha$, que se distribuye como una F de Fisher $\mathcal{F}(1, \nu)$. A medida que $\nu \to \infty$, dicha distribución se acerca a una $\chi_{(1)}^2$ (chi-cuadrado con un grado de libertad).

Generalmente estas varianzas se modelan como χ^2 inversas escaladas, pero la nueva distribución pone menos densidad de probabilidad cerca del cero, lo cual hace que la MCMC salga más fácilmente de la región del cero si se queda atascada.

La solución de la posterior $p(\beta, u, \sigma_u^2 \mid y)$ no es analítica explícita, sino que se hace por medios computacionales usando MCMC (Markov Chain Monte Carlo) y sus aplicaciones, como el muestreo de *Gibbs* (Blangiardo & Cameletti, 2015, cap. 4). El punto importante es que, satisfaciendo ciertas condiciones, la cadena llega ("converge") a una distribución π de estados que es invariante (no se modifica con el tiempo). En el contexto bayesiano, la distribución invariante es precisamente $p(\theta \mid y)$, la posterior. La idea es tomar una secuencia de valores de los parámetros de la posterior $(\theta^1, \theta^2, \ldots)$ hasta que la cadena alcance la distribución invariante; que será una aproximación a la posterior. La convergencia sucede luego de un (largo) número de interacciones, digamos $t > t_0$. Las interacciones hasta t_o se descartan, y este conjunto de interacciones descartadas se conoce como periodo de *burn-in*. Sucede que en dicho periodo los valores muestreados pueden estar más correlacionados entre sí. En cambio los valores muestreados de la posterior (invariante) deben ser independientes. Por ello, un diagnóstico frecuente de convergencia de la cadena consiste en avaluar la auto-correlación r a diferentes *lags* o intervalos entre valores y verificar que $r < 0.1$. Cuanto más parámetros tenga el modelo, más se tardará en lograr la convergencia.

Por último, resulta conveniente presentar una medida bayesiana de comparación de modelos: el criterio de información de devianza [DIC] (Gelman et al., cap. 6). La devianza se define como menos dos veces el logaritmo de la verosimilitud del modelo: $D(y, \hat{\theta}) = -2logp(y \mid \hat{\theta})$, donde $\hat{\theta}$ es una estimación puntual, por ejemplo la moda o la media de la posterior del (vector de) parámetro(s). La esperanza $D_{avg}(y) = E(D(y, \theta) \mid y)$ puede estimarse como $D_{avg}(y) = \frac{1}{\mathrm{L}} \sum_{k=1}^{L} D(y, \theta^{(l)})$, donde $\theta^{(l)}$ es la l-ésima muestra del parámetro, usando, por ejemplo, muestreo de Gibbs. Es decir que es un promedio de las devianzas calculadas sobre cada uno de los valores del parámetro obtenidos en la posterior. Usando estas dos fórmulas, se define el número efectivo de parámetros como: $p_D = \hat{D}_{avg}(y) - D(y, \hat{\theta})$. Entonces la medida DIC, que debe minimizarse para el mejor modelo, rezará:$DIC = 2\hat{D}_{avg}(y) - D(y, \hat{\theta}) = p_D + \hat{D}_{avg}(y) = D(y, \hat{\theta}) + 2p_D$.

3. Ajuste del modelo multinomial bayesiano

El modelo mixto multinomial bayesiano general para la concordancia i en el grupo j (el grupo está definido como *la sesión* k anidada en *el alumno* g) $[k = 1, \ldots, 12(14)$; $g = 1, \ldots, 4$; $j = 1, \ldots, 52$; $i = 1, \ldots, 1857]$, dado que la observación tiene la categoría $c = 1, \ldots, 4$ se describe como sigue:

(1) Distribuciones y *Priors*: $y_{ij} = c | \upsilon_{0ij}, \upsilon_{1ij} \sim multinomial\left(N, (\pi_{ij1}\pi_{ij2}\pi_{ij3}\pi_{ij4})\right), N = 1$; $\boldsymbol{\upsilon_0} \sim NMV(\mathbf{0}, \mathbf{G_0})$; $\boldsymbol{\upsilon_1} \sim NMV(\mathbf{0}, \mathbf{G_1})$; $\boldsymbol{\beta} \sim NMV(\mathbf{0}, \mathbf{V_\beta})$;

(2) Matrices de varianza: $G_{INTERCEPT} = V \otimes A$; $G_{FAM.LEX.f} = V \otimes A$; $A = I$; V=I (efectos aleatorios independientes).

$$V_{SESION:ID,INTERCEPT} = \begin{bmatrix} \sigma^2_{v_0 1} & & & \\ & \sigma_{v_0 2} & & \\ & & \sigma_{v_0 3} & \\ & & & \sigma_{v_0 4} \end{bmatrix} \quad ; \quad V_{SESION:ID;FAM:LEX} = \begin{bmatrix} \sigma^2_{v_1,FAM.LEX.f=0} & \\ & \sigma_{v_1,FAM.LEX.f=1} \end{bmatrix}$$

(3) Función de enlace: $logit = g\left(E\left[y_{ij} = c|v_{0ij}, v_{1ij}\right]\right) = log\left[\frac{\pi_{ijc}}{1-\pi_{ijc}}\right]$;

(4) Función inversa: $\pi_{ij} = g^{-1}(X\beta + Z_v v)$;

(5) Predictor lineal: $\eta_{ij} = \varphi_{ij} + v_{0ij} + v_{1ij}$. Donde: φ_{ij} representa los efectos fijos; v_{0ij} y v_{1ij} son los efectos aleatorios "entre".

(6) Tamaño muestral efectivo: $(Interactions - burn\,in)/thin = (4100000 - 100000)/2000 = 2000.$

(7) Modelo teórico: para $c = 1,2,3,4$, con $c = 0$ como categoría de referencia.

$$g\left(E\left[y_{ij} = c|v_{0ij}, v_{1ij}\right]\right) = log\left[\frac{\pi_{ijc}}{1-\pi_{ijc}}\right] = (\beta_{0c} + v_{0ic}) + (\beta_{1c} + v_{1ic})FAM.LEX.f$$
$$\beta_{2c}MOD_1 + \beta_{3c}MOD_2 + \beta_{4c}MOD_3 + \beta_{5c}CUMRES.f_1 + \beta_{6c}CUMRES.f_2$$
$$\beta_{7c}Fabs.SC.f + \beta_{8c}MORF.f_1 + \beta_{9c}MORF.f_2 + \beta_{10c}EST1 + \beta_{11c}EST5 + \beta_{12c}EST2$$
$$\beta_{13c}GRUPO6_2 + \beta_{14c}GRUPO6_3 + \beta_{15c}GRUPO6_4 + \beta_{16c}GRUPO6_5 + \beta_{17c}GRUPO6_6$$
$$\beta_{18c}ES_1 + \beta_{19c}ES_2 + \beta_{20c}ANIM + \beta_{21c}TIME$$

La Tabla 6 describe los parámetros del modelo a estimar, para *cada* modelo $c = 1,2,3,4$. Hubo *22* efectos fijos a estimar por modelo. Considérense en primer lugar los diagnósticos de convergencia. En lo que respecta a las varianzas de los factores aleatorios, las correlaciones en *lags > 2000* estaban debajo de *<0.1*[5]. Para el *lag 2000* (equivalente al *lag 1*) la autocorrelación fue mayor a 0.1 para $r_{intercept,c,lag=2000} = [1 = 0.33; 2 = 0.6; 3 = 0.21; 4 = 0.55]$; y <0.1 para FAM.LEX.f. En suma, la autocorrelación fue moderada (lag 1) para las varianzas en las categorías 2 y 4. En cuanto a los coeficientes β, la correlación como mucho rondó *0.3*. Por lo tanto, se considera que hubo una convergencia razonable de las

---

[5] Como se muestrea de la posterior cada 2000 valores, las equivalencias serían: lag 2000 = lag 1; lag 10000 = lag 5, lag 20000 = lag 10; lag 100000 = lag 50.

distribuciones posteriores[6]. Además, las posteriores de las varianzas resultaron muy asimétricas a derecha. Por lo tanto se usó la moda como medida para la estimación puntual.

Tabla 5. Parámetros del modelo logístico multinomial bayesiano

Parámetro	Descripción
$\boldsymbol{v_{0ic}}$	la desviación del grupo $\boldsymbol{i_c}$ de la ordenada al origen
$\boldsymbol{v_{1ic}}$	la desviación del grupo $\boldsymbol{i_c}$ de la media marginal de FAM.LEX.f.
$\boldsymbol{\beta_{0c}}$	la media basal marginal
$\boldsymbol{\beta_{1c}}$	el efecto de FAM.LEX.f, nivel 1 (referencia: FAM.LEX.f = 0)
$\boldsymbol{\beta_{2c}}$	el efecto de MOD, nivel 1 (referencia: MOD = 0)
$\boldsymbol{\beta_{3c}}$	el efecto de MOD, nivel 2 (referencia: MOD = 0)
$\boldsymbol{\beta_{4c}}$	el efecto de MOD, nivel 3 (referencia: MOD = 0)
$\boldsymbol{\beta_{5c}}$	el efecto de CUMRES.f, nivel 1 (referencia: CUMRES.f = 0)
$\boldsymbol{\beta_{6c}}$	el efecto de CUMRES, nivel 2 (referencia: CUMRES.f = 0)
$\boldsymbol{\beta_{7c}}$	el efecto de Fabs.SC.f, nivel 1 (referencia: Fabs.SC.f = 0)
$\boldsymbol{\beta_{8c}}$	el efecto de MORF.f, nivel 1 (referencia: MORF.f = 0)
$\boldsymbol{\beta_{9c}}$	el efecto de MORF.f nivel 2 (referencia: MOD = 0)
$\boldsymbol{\beta_{10c}}$	el efecto de EST1, nivel 1 (referencia: EST1 = 0)
$\boldsymbol{\beta_{11c}}$	el efecto de EST5, nivel 1 (referencia: EST1 = 0)
$\boldsymbol{\beta_{12c}}$	el efecto de EST2, nivel 1 (referencia: EST1 = 0)
$\boldsymbol{\beta_{13c}}$	el efecto de GRUPO6, nivel 2 (referencia: GRUPO6 = 1)
$\boldsymbol{\beta_{14c}}$	el efecto de GRUPO6, nivel 3 (referencia: GRUPO6 = 1)
$\boldsymbol{\beta_{15c}}$	el efecto de GRUPO6, nivel 4 (referencia: GRUPO6 = 1)
$\boldsymbol{\beta_{16c}}$	el efecto de GRUPO6, nivel 5 (referencia: GRUPO6 = 1)
$\boldsymbol{\beta_{17c}}$	el efecto de GRUPO6, nivel 6 (referencia: GRUPO6 = 1)
$\boldsymbol{\beta_{18c}}$	el efecto de ES, nivel 1 (referencia: ES = 0)
$\boldsymbol{\beta_{19c}}$	el efecto de ES, nivel 2 (referencia: ES = 0)
$\boldsymbol{\beta_{20c}}$	el efecto de ANIM, nivel 1 (referencia: ES = 0)
$\boldsymbol{\beta_{21c}}$	el efecto de TIME

Las varianzas de los factores aleatorios son los que se detallan en las matrices:

[6] El test de estacionariedad de Heidelberger y Welch (1981) propone como hipótesis nula que los valores muestreados de la posterior provienen de una distribución estacionaria. Ningún σ^2 de efectos aleatorios o β de efectos fijos rechazó dicha hipótesis nula.

$$V_{INTERCEPT} = \begin{bmatrix} 1.139 & & & \\ & 0.015 & & \\ & & 0.002 & \\ & & & 0.004 \end{bmatrix}; V_{FAM:LEX} = \begin{bmatrix} 0.001 & \\ & 0.562 \end{bmatrix};$$

Hubo $52 \times 4 = 208$ efectos aleatorios de ordenada al origen y $52 \times 2 = 104$ efectos aleatorios de FAM.LEX.f. Los desvíos típicos de los efectos aleatorios de ordenada al origen para cada categoría y para FAM.LEX.f resultaron:

$$\sigma_{0ic} = [c1 = 1.067; c2 = 0.125; c3 = 0.052; c4 = 0.068]$$
$$\sigma_{1ic} = [FAM.LEX.F(0) = 0.043; FAM.LEX.f(1) = 0.75]$$

Hubo más desvío en las categorías 1 y 2 respecto de las 3 y 4. El desvío de FAM.LEX.f(0) fue casi nulo. El coeficiente de ordenada al origen de *cada* categoría varió en términos del *logit* como $\beta_{0c} = [c1 = -2.28; c2 = -3.07; c3 = -0.72; c4 = -2.77]$ más / menos la desviación del efecto aleatorio "entre" de ordenada al origen de cada grupo. Análogamente, el efecto (coeficiente) de FAM.LEX.f varió en términos del logit en cada categoría como:

$\beta_{1c} = [c1 = -0.46; c2 = -1.36; c3 = -0.56; c4 = -0.89]$ más / menos la desviación del efecto "entre" de FAM.LEX.f de cada grupo.

En lo que atañe a los supuestos distribucionales de los factores aleatorios, se aplicó el test de *Shapiro-Wilks* para detectar desvíos groseros del supuesto de normalidad. Se tomó a la media como resumen de la posterior[7]. En cuanto a los efectos aleatorios de ordenada al origen, el supuesto no se cumplió para las categorías 1 y 2 ($p < 0.0001$) y se cumplió para las categorías 3 ($p = 0.77$) y 4 ($p = 0.45$). En los que atañe a los factores aleatorios de FAM.LEX.f se cumplió para ambos, aunque con menos evidencia para rechazar la hipótesis nula de normalidad para [FAM.LEX.f = 1] ($p = 0.054$), que para [FAM.LEX.f = 0] ($p = 0.39$). A continuación, el ajuste del modelo bayesiano total.

Tabla 6. Ajuste del modelo (todos los predictores)

	mean	l.CI	u.CI	Eff	pMCMC	OR	l.CI.OR	u.CI.OR
traitRES_CAT.1	-2.281	-3.443	-1.081	2000.000	0.001	0.102	0.032	0.339
traitRES_CAT.2	-3.071	-4.420	-1.890	1572.478	0.001	0.046	0.012	0.151
traitRES_CAT.3	-0.724	-1.432	-0.085	2000.000	0.037	0.485	0.239	0.919
traitRES_CAT.4	-2.772	-3.801	-1.786	1594.264	0.001	0.063	0.022	0.168
traitRES_CAT.1:TIME	-0.012	-0.044	0.019	1332.941	0.458	0.988	0.957	1.020
traitRES_CAT.2:TIME	-0.007	-0.043	0.029	1093.957	0.734	0.993	0.958	1.029
traitRES_CAT.3:TIME	0.009	-0.004	0.024	1870.334	0.184	1.009	0.996	1.024
traitRES_CAT.4:TIME	0.009	-0.012	0.029	1874.825	0.384	1.009	0.989	1.030

[7] La correlación de Pearson entre las estimaciones puntuales usando la media y la mediana estuvieron arriba de $r > 0.98$; con lo cual se consideró que las distribuciones eran simétricas. Por tanto, la media es un buen resumen de la distribución.

traitRES_CAT.1:MOD1	-0.169	-1.743	1.276	2000.000	0.847	0.845	0.175	3.582
traitRES_CAT.2:MOD1	-0.583	-2.443	1.412	1713.370	0.569	0.558	0.087	4.104
traitRES_CAT.3:MOD1	-0.489	-1.645	0.532	1687.632	0.382	0.613	0.193	1.702
traitRES_CAT.4:MOD1	-0.556	-1.936	1.012	1609.775	0.475	0.574	0.144	2.750
traitRES_CAT.1:MOD2	0.347	-0.287	0.935	1812.128	0.293	1.415	0.751	2.546
traitRES_CAT.2:MOD2	0.194	-0.755	1.124	1276.952	0.644	1.215	0.470	3.077
traitRES_CAT.3:MOD2	0.328	-0.032	0.706	2038.619	0.089	1.388	0.968	2.025
traitRES_CAT.4:MOD2	0.883	0.315	1.500	1596.517	0.001	2.419	1.370	4.480
traitRES_CAT.1:MOD3	-0.898	-1.814	-0.163	1173.143	0.032	0.407	0.163	0.850
traitRES_CAT.2:MOD3	-0.259	-1.443	0.758	1208.766	0.683	0.772	0.236	2.135
traitRES_CAT.3:MOD3	0.431	0.007	0.840	1738.849	0.045	1.538	1.007	2.316
traitRES_CAT.4:MOD3	0.472	-0.157	1.163	1655.423	0.149	1.603	0.855	3.201
traitRES_CAT.1:Fabs.SC.f1	-1.277	-2.036	-0.491	1596.650	0.001	0.279	0.130	0.612
traitRES_CAT.2:Fabs.SC.f1	0.071	-0.799	1.099	1278.310	0.900	1.073	0.450	3.002
traitRES_CAT.3:Fabs.SC.f1	-0.424	-0.799	-0.012	1851.367	0.035	0.654	0.450	0.988
traitRES_CAT.4:Fabs.SC.f1	-0.927	-1.485	-0.250	1530.948	0.003	0.396	0.226	0.778
traitRES_CAT.1:ANIM1	-0.121	-0.772	0.543	1447.283	0.738	0.886	0.462	1.721
traitRES_CAT.2:ANIM1	1.254	0.521	2.008	1190.987	0.003	3.503	1.684	7.445
traitRES_CAT.3:ANIM1	0.132	-0.192	0.467	1809.376	0.436	1.141	0.826	1.596
traitRES_CAT.4:ANIM1	0.519	-0.026	1.031	1673.390	0.063	1.680	0.974	2.803
traitRES_CAT.1:ES1	0.541	-0.647	1.677	1797.495	0.382	1.718	0.524	5.351
traitRES_CAT.2:ES1	0.066	-1.274	1.383	1680.962	0.933	1.068	0.280	3.989
traitRES_CAT.3:ES1	-0.308	-1.049	0.379	2000.000	0.426	0.735	0.350	1.461
traitRES_CAT.4:ES1	0.482	-0.519	1.454	1799.522	0.355	1.620	0.595	4.281
traitRES_CAT.1:ES2	-0.672	-2.717	1.237	1867.843	0.525	0.511	0.066	3.445
traitRES_CAT.2:ES2	0.111	-1.610	1.923	1505.873	0.872	1.117	0.200	6.842
traitRES_CAT.3:ES2	-2.007	-3.532	-0.624	1601.730	0.004	0.134	0.029	0.536
traitRES_CAT.4:ES2	0.046	-1.304	1.484	1548.543	0.952	1.047	0.272	4.409
traitRES_CAT.1:MORF.f1	-0.191	-1.150	0.779	2000.000	0.667	0.826	0.317	2.178
traitRES_CAT.2:MORF.f1	-0.962	-2.058	0.274	1537.687	0.100	0.382	0.128	1.315
traitRES_CAT.3:MORF.f1	-1.233	-1.714	-0.746	2000.000	0.001	0.291	0.180	0.474
traitRES_CAT.4:MORF.f1	-0.630	-1.342	0.097	1733.746	0.091	0.533	0.261	1.102
traitRES_CAT.1:MORF.f2	-0.342	-1.359	0.757	1867.880	0.540	0.710	0.257	2.132
traitRES_CAT.2:MORF.f2	-0.086	-1.293	1.039	1725.062	0.870	0.917	0.274	2.826
traitRES_CAT.3:MORF.f2	-0.949	-1.557	-0.287	1647.699	0.005	0.387	0.211	0.751
traitRES_CAT.4:MORF.f2	-0.497	-1.299	0.374	1724.647	0.264	0.608	0.273	1.454
traitRES_CAT.1:FAM.LEX.f1	-0.464	-1.105	0.173	1544.699	0.163	0.629	0.331	1.189
traitRES_CAT.2:FAM.LEX.f1	-1.362	-2.201	-0.475	1098.049	0.001	0.256	0.111	0.622
traitRES_CAT.3:FAM.LEX.f1	-0.563	-0.953	-0.141	2000.000	0.004	0.570	0.386	0.868
traitRES_CAT.4:FAM.LEX.f1	-0.897	-1.432	-0.360	1533.184	0.001	0.408	0.239	0.698
traitRES_CAT.1:CUMRES.f1	0.137	-0.573	0.862	1863.065	0.736	1.147	0.564	2.369
traitRES_CAT.2:CUMRES.f1	0.404	-0.443	1.269	1052.338	0.351	1.498	0.642	3.558
traitRES_CAT.3:CUMRES.f1	-0.038	-0.449	0.373	2000.000	0.855	0.962	0.638	1.452
traitRES_CAT.4:CUMRES.f1	0.704	0.025	1.322	1653.229	0.026	2.021	1.026	3.749
traitRES_CAT.1:CUMRES.f2	0.046	-1.058	1.130	2000.000	0.951	1.047	0.347	3.097
traitRES_CAT.2:CUMRES.f2	-0.510	-1.849	0.874	1403.991	0.468	0.601	0.157	2.396
traitRES_CAT.3:CUMRES.f2	-0.222	-0.834	0.433	2000.000	0.467	0.801	0.434	1.542
traitRES_CAT.4:CUMRES.f2	0.503	-0.379	1.401	1616.208	0.296	1.653	0.685	4.060
traitRES_CAT.1:EST11	-1.773	-2.720	-0.924	914.828	0.001	0.170	0.066	0.397
traitRES_CAT.2:EST11	-0.444	-1.288	0.497	1306.433	0.347	0.641	0.276	1.644

traitRES_CAT.3:EST11	-0.077	-0.476	0.314	2000.000	0.691	0.926	0.621	1.368
traitRES_CAT.4:EST11	-0.893	-1.539	-0.229	1130.814	0.008	0.410	0.215	0.796
traitRES_CAT.1:EST21	0.118	-0.641	0.818	1760.448	0.755	1.125	0.527	2.267
traitRES_CAT.2:EST21	-1.510	-2.773	-0.198	672.114	0.021	0.221	0.062	0.820
traitRES_CAT.3:EST21	0.272	-0.124	0.672	2000.000	0.205	1.313	0.884	1.959
traitRES_CAT.4:EST21	0.102	-0.481	0.722	1711.588	0.740	1.107	0.618	2.058
traitRES_CAT.1:EST51	-0.865	-2.059	0.175	1556.623	0.120	0.421	0.128	1.191
traitRES_CAT.2:EST51	-2.139	-3.540	-0.729	856.514	0.001	0.118	0.029	0.483
traitRES_CAT.3:EST51	-0.107	-0.715	0.427	2000.000	0.730	0.899	0.489	1.533
traitRES_CAT.4:EST51	-1.855	-2.756	-1.088	1372.963	0.001	0.156	0.064	0.337
traitRES_CAT.1:GRUPO62	-0.603	-1.757	0.401	1510.496	0.275	0.547	0.173	1.493
traitRES_CAT.2:GRUPO62	-0.520	-1.809	0.640	1549.728	0.396	0.594	0.164	1.897
traitRES_CAT.3:GRUPO62	0.170	-0.412	0.781	2000.000	0.565	1.186	0.663	2.183
traitRES_CAT.4:GRUPO62	0.664	-0.083	1.472	1576.769	0.100	1.943	0.920	4.357
traitRES_CAT.1:GRUPO63	-0.139	-0.987	0.661	1723.035	0.756	0.871	0.373	1.937
traitRES_CAT.2:GRUPO63	-0.504	-1.695	0.654	1029.938	0.405	0.604	0.184	1.923
traitRES_CAT.3:GRUPO63	0.074	-0.342	0.488	1647.641	0.747	1.077	0.711	1.629
traitRES_CAT.4:GRUPO63	-0.488	-1.150	0.204	1587.010	0.163	0.614	0.317	1.226
traitRES_CAT.1:GRUPO64	-0.218	-1.119	0.687	1710.223	0.661	0.804	0.327	1.987
traitRES_CAT.2:GRUPO64	-0.278	-1.364	0.945	1299.757	0.659	0.757	0.256	2.573
traitRES_CAT.3:GRUPO64	-0.008	-0.448	0.499	1813.817	0.970	0.992	0.639	1.647
traitRES_CAT.4:GRUPO64	-0.067	-0.854	0.638	1650.738	0.864	0.935	0.426	1.893
traitRES_CAT.1:GRUPO65	-0.586	-2.041	0.842	1737.677	0.435	0.557	0.130	2.321
traitRES_CAT.2:GRUPO65	-0.591	-2.236	0.893	1723.644	0.466	0.554	0.107	2.443
traitRES_CAT.3:GRUPO65	0.519	-0.169	1.146	2000.000	0.117	1.681	0.845	3.145
traitRES_CAT.4:GRUPO65	-1.207	-2.366	0.050	1367.219	0.036	0.299	0.094	1.051
traitRES_CAT.1:GRUPO66	-0.242	-1.545	1.177	1369.080	0.763	0.785	0.213	3.246
traitRES_CAT.2:GRUPO66	0.752	-0.351	2.051	1864.481	0.237	2.121	0.704	7.777
traitRES_CAT.3:GRUPO66	-0.251	-1.139	0.675	2000.000	0.605	0.778	0.320	1.964
traitRES_CAT.4:GRUPO66	0.138	-0.959	1.388	1836.897	0.797	1.148	0.383	4.005

mean = media de la posterior, l CI = Int. de credibilidad del 95 por ciento, extremo inferior, u CI = Int. de credibilidad del 95 por ciento, extremo superior; eff = tamaño muestral efectivo; pMCMC = p valor; OR = exp(mean), l CI OR = exp(l CI), u CI OR = exp(u CI).

4. Código de R

Con el objetivo de fomentar la replicación de los resultados de este trabajo, se brinda en adjunto el script de R usado para el análisis, junto a las bases de datos.

Referencias

Blangiardo, M. y Cameletti, M. (2016). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley and Sons.

Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515–533.

Gelman, A., Carlin, J., Stern, H. S. y Rubin, D. (2004). *Bayesian Data Analysis*. Chapman y Hall/CRC.

Hadfield, J. (2010). MCMCglmm: Markov chain monte carlo methods for generalised linear mixed models. Recuperado de *http://cran. r-project.org/web/packages/MCMCglmm/index.html*.

Heidelberger, P. y Welch, P. D. (1981). A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM*, 24(4), 233-245.

Sorensen, D. y Gianola, D. (2002). *Likelihood, bayesian and MCMC methods in quantitative genetics*. Springer.

Stroup, W. W. (2013). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press, Chapman Hall.

Villemereuil, P. (2012). Estimation of a biological trait heritability using the animal model: how to use the mcmcglmm r package. Recuperado de *http://devillemereuil.legtux.org/downloads/*

Villemereuil, P., Gimenez, O. y Doligez, B. (2013). Comparing parent-offspring regression with frequentist and bayesian animal models to estimate heritability in wild populations: a simulation study for gaussian and binary traits. *Methods in Ecology and Evolution*, 4, 260–275.