Phonology **37** (2020). Supplementary materials

Incorporating tone in the modelling of wordlikeness judgements

Youngah Do
University of Hong Kong
Ryan Ka Yau Lai
University of California, Santa Barbara

Supplementary materials

Appendix A: List of stimuli

1	mɪk̚˧˥	21	wɛːt̚˩˧	41	kʷœːŋ˨
2	pʰyːu˧	22	pʰɵt̚˥	42	taː˨˩
3	kʷɛːk̚˧	23	hɐk̚˧	43	kɪk̚˧
4	juːy˧	24	pʰɔːi˨˩	44	lɵn˧
5	pɐi˨˩	25	kʷʰɵy˨˩	45	fyːp̚˧
6	kaːu˩˧	26	kʷiːy˧	46	kaːm˨
7	kʰyːi˥	27	kɔːt̚˧˥	47	piːp̚˩˧
8	lyːk̚˥	28	fɵt̚˥	48	tsʰiːp̚˨
9	kɵt̚˧˥	29	wyːŋ˩˧	49	myːŋ˨˩
10	kʷey˩˧	30	wɛːu˥	50	lɐn˩˧
11	tsʰɛːm˩˧	31	liːy˨	51	pʰɛː˨
12	hɛːm˥	32	hɐk̚˨˩	52	tɛːp̚˨˩
13	tsʰey˧˥	33	saːp̚˥	53	poy˩˧
14	kʰœːp̚˧˥	34	sɐi˨˩	54	wiː˩˧
15	tsɛːk̚˩˧	35	fɵt̚˧	55	taːy˧
16	jou˨˩	36	hɛːŋ˧	56	tsɔːt̚˥
17	jey˧	37	jaːm˥	57	puːm˨
18	tsʰei˨	38	pʰyːp̚˨	58	tʰʊk̚˧
19	fɛːn˥	39	tʰɔːt̚˩˧	59	tou˩˧
20	maːm˥	40	myːn˧˥	60	tʰɛːt̚˩˧

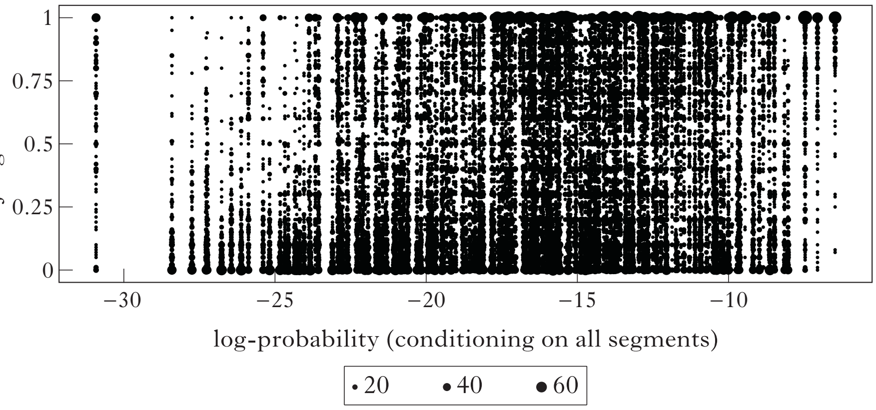
61	tɛːŋ˨˩	101	mɛːu˨˩	141	kʰiː˨˩
62	tʰuːy˨˩	102	lɪk̚˧	142	pʰɔːŋ˥
63	kɐŋ˨˩	103	tʰɐt̚˨	143	ʦʰuːt̚˥
64	puː˥	104	tœː˧	144	syːu˧
65	fɪk̚˧	105	pɵi˥	145	pɔːm˨˩
66	kɵt̚˧	106	mœːk̚˨˩	146	kʷœːu˩˧
67	kʷʰaːŋ˩˧	107	kuːn˨	147	hyːk̚˨
68	kʷyːp̚˧	108	pʰɛːk̚˩˧	148	ʦʰɵt̚˨˩
69	soy˧˥	109	kʷɔːn˥	149	ʦyːu˧˥
70	ʦʰɐp̚˩˧	110	kɪk̚˨˩	150	wɔːm˩˧
71	kʷoy˩˧	111	jɔːm˩˧	151	haːk̚˧˥
72	kʷʰɐi˧	112	pɔːt̚˧	152	fɐp̚˨˩
73	pœːu˥	113	paːp̚˧˥	153	fyːm˥
74	taːk̚˥	114	pʰɐk̚˧˥	154	ʦʰiːp̚˨˩
75	jɛːk̚˧˥	115	wɔːk̚˨˩	155	ʦʰœːp̚˨
76	mɐm˧˥	116	kʷʰyːi˧˥	156	fei˩˧
77	ʦʰuː˩˧	117	faːu˩˧	157	kʷɵi˨
78	syːu˩˧	118	kʷʰɵi˨˩	158	tʰaːk̚˨
79	kʷʰaːp̚˨˩	119	kʰɐ˥	159	tyːi˨
80	kʰɔːŋ˨	120	hɔːt̚˧˥	160	kʰaːp̚˨˩
81	kʰɔːm˧˥	121	kʰoy˧˥	161	mɵn˩˧
82	kɛːt̚˨	122	pʰuːp̚˨˩	162	kʷaːt̚˥
83	kʷʰiː˥	123	wiː˧	163	tɵn˨˩
84	sɛːn˧	124	waːu˩˧	164	miː˨
85	tʰɛːk̚˧˥	125	lɐy˨˩	165	fɔːi˨˩
86	fɐŋ˨	126	ʦɔːt̚˨˩	166	lyː˨
87	pʰɪk̚˨	127	joy˧˥	167	kʷʰaːu˨˩
88	pʰaːu˩˧	128	ʦuːn˥	168	pɐk̚˧
89	kʷɵi˥	129	wɵy˧˥	169	tʰɪk̚˩˧
90	hɛːŋ˧˥	130	faːm˨˩	170	kʰɛːk̚˨˩
91	fiːy˩˧	131	jɔːp̚˧	171	kʰoy˧
92	pɛːm˥	132	fyːu˩˧	172	pʰɔːn˧
93	ʦʰɛːu˩˧	133	kʷʰiːn˨˩	173	juːp̚˧
94	jœː˧	134	pyːu˥	174	ʦʰɛːn˥
95	fyːu˨˩	135	haːu˩˧	175	sey˧
96	lœːu˧	136	jɔːk̚˧	176	wyːn˧˥
97	pʰɛːk̚˨˩	137	kʰɐp̚˧	177	ʦyːt̚˨˩
98	pʰœː˨˩	138	kaːt̚˨˩	178	kʷʰɐŋ˨
99	fiːy˨	139	kʰɔːi˧˥	179	hyːp̚˩˧
100	tʰɔːt̚˧˥	140	tɐm˨˩	180	tuːm˩˧

181 pʰɛːt̚˨˩
182 jœː˧˥
183 ʦʰœː˨˩
184 pɵy˨˩
185 hʊk̚˨˩
186 fiːp̚˩˧
187 muːy˧
188 jɐk̚˩˧
189 waːt̚˨˩
190 sʊk̚˨˩
191 ʦɐ˧
192 kʷʰɵy˥
193 fɔːk̚˨˩
194 fɐt̚˩˧
195 kʰɛːŋ˥
196 tɐp̚˩˧
197 kʷyːŋ˥
198 mɔːi˧
199 pʰɵt̚˧˥
200 sɵp̚˥
201 hɵi˧˥
202 ʦʰuːm˨
203 kʷɛːk̚˨
204 wœː˨˩
205 mɛːn˧
206 tʰei˨
207 tʰɐ˥
208 tɐp̚˧
209 tʰɔːn˧
210 fɔːn˧
211 mʊŋ˧
212 ʦʰiːt̚˨
213 tɛːk̚˥
214 kɵn˥
215 jaːu˧˥
216 hyːi˨˩
217 tɵn˨˩
218 pœːp̚˨˩
219 ʦʰyː˥
220 ley˥
221 pɵn˧˥
222 kʷʰɐn˨˩
223 hyːk̚˧
224 jaːu˨˩
225 kʷʰœːp̚
226 kʰɐt̚˨˩
227 hyːp̚˧˥
228 pʰaːy˩˧
229 pyːu˧
230 kʷʰiːm˩˧
231 sɵp̚˨
232 sœː˧˥
233 kʷɛːn˨˩
234 waːu˧˥
235 wɵi˧
236 tɛːk̚˨˩
237 suːi˧
238 ʦɛːm˧
239 pyːt̚˧˥
240 miːp̚˨
241 kʷœːm˧˥
242 kɔːt̚˩˧
243 pɔːm˧˥
244 hyːm˥
245 kʷʰɔːk̚˩˧
246 kʷaːn˧˥
247 kʷiːm˧
248 tyːp̚˥
249 tei˩˧
250 wœː˧˥
251 ʦɐn˨˩
252 tʰɐp̚˩˧
253 saːu˨˩
254 kyːt̚˩˧
255 wʊk̚˩˧
256 ʦʰuːy˨
257 tʰuːy˥
258 pʰɵy˥
259 tʰɛːm˧˥
260 tʰɛːm˨˩
261 wiːy˨
262 jœːu˧˥
263 kʰœːm˧
264 pʰɛːn˨˩
265 tyːp̚˨
266 kʷɛːn˧
267 pʰœː˧˥
268 tʰɐn˨˩
269 kʰaːn˧
270 fey˩˧
271 kʷɐk̚˩˧
272 pɐŋ˧
273 pʰiːp̚˨
274 tʰɵn˨
275 pɐp̚˥
276 miːt̚˧˥
277 kʷɔːm˩˧
278 jɐk̚˥
279 kʰɐy˧
280 kʰɛːt̚˧
281 kʷʰaːt̚˧
282 sɛːt̚˧˥
283 mɐy˨˩
284 ʦyːk̚˧
285 tʰɐn˧˥
286 tʰɵi˧˥
287 tʰuːm˩˧
288 pʰɔːt̚˧

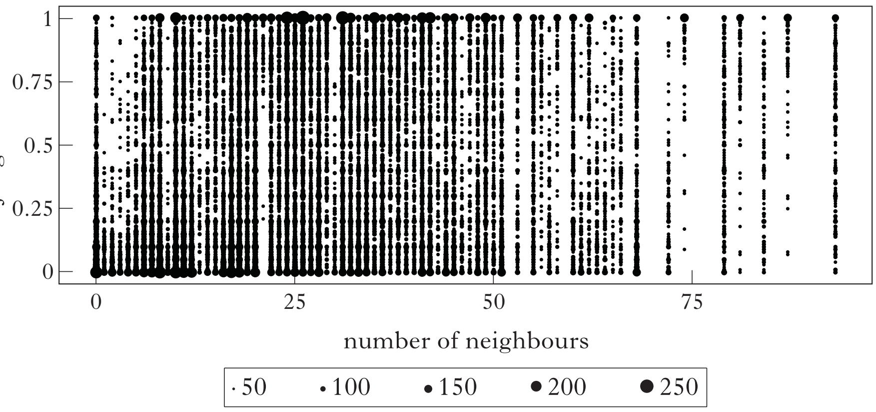
Appendix B: Scatterplot of predictors againts wordlikeness judgements

In the figures below, the size of the circles is proportional to the number of judgements.

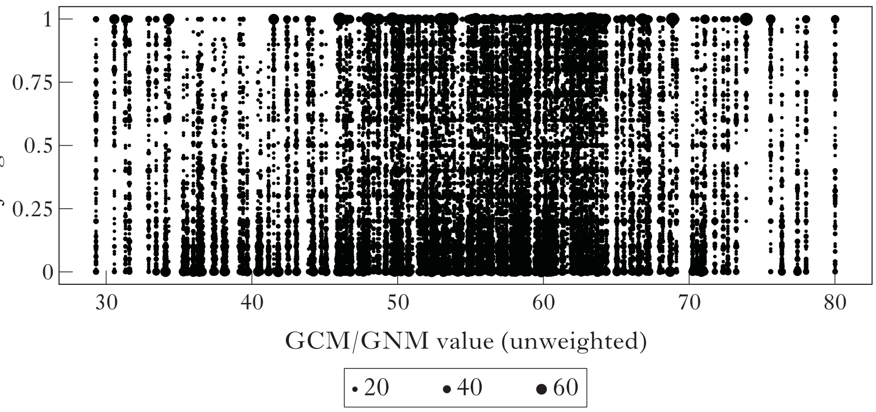
a.



b.



c.



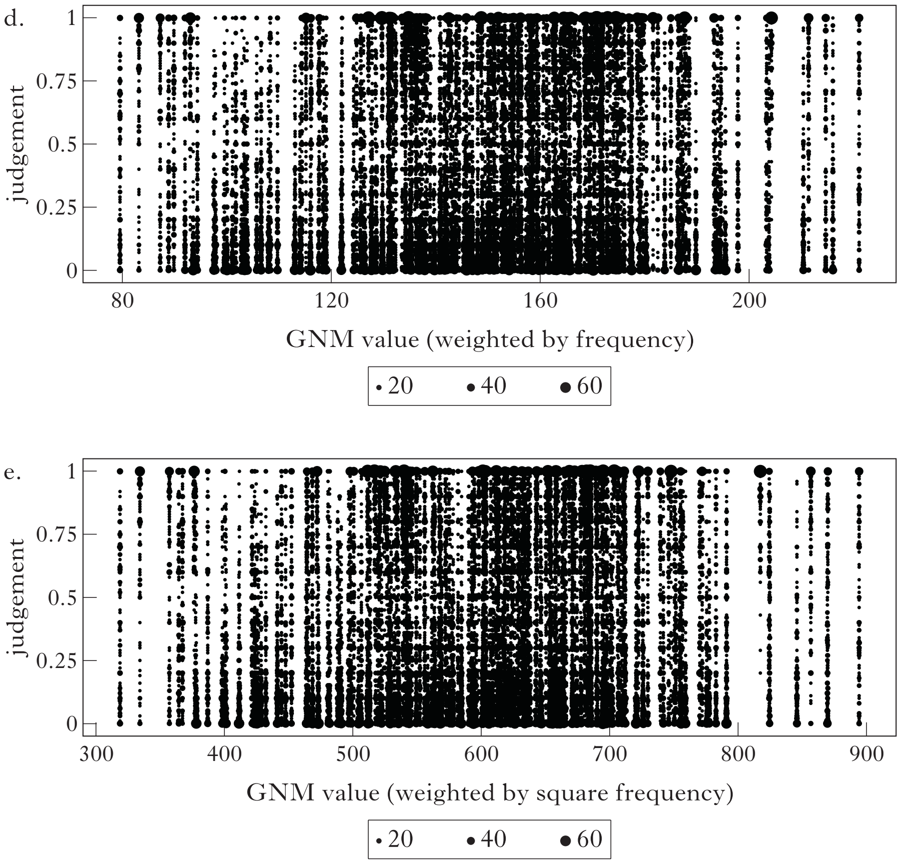


Figure 6

Scatterplot of judged wordlikeness against (a) log-probability; (b) the number of neighbours; (c) the third GCM value, i.e. the third GNM quality insensitive to frequency; (d) the second GNM quantity, i.e. GCM weighted by frequency (with B as a coefficient); (e) the first GNM quantity, i.e. GCM weighted by square frequency (with A as a coefficient).

Appendix C: Basics of the ZOIB model

The mixed-effect zero-one-inflated beta regression model (ZOIB; Ospina & Ferrari 2012) has three components: a Bernoulli-distributed (i.e. discrete probability distribution) component for predicting whether the judgement is zero (absolutely impossible), another Bernoulli-distributed component for predicting whether the judgement is one (absolutely possible) and a beta-distributed (i.e. continuous probability distribution) component for modelling the density of the gradient judgements (between 0 and 1). The three components' distributions are given in (7).

(7) $I(Y_{ij} = 0) \sim Bernoulli(logit^{-1}(\beta_{00} + (\beta_{01} + \alpha_{01i})x_{lp,j} + \alpha_{00i} + \gamma_{0j}))$
$I(Y_{ij} = 1) \sim Bernoulli(logit^{-1}(\beta_{10} + (\beta_{11} + \alpha_{11i})x_{lp,j} + \alpha_{10i} + \gamma_{1j}))$
$Y_{ij} \mid Y_{ij} \in 0, 1)) \sim Beta(\varphi\ logit^{-1}(\beta_{20} + (\beta_{21} + \alpha_{21i})x_{lp,j} + \alpha_{20i} + \gamma_{2j}),$
$\varphi\ (1 - \varphi\ logit^{-1}(\beta_{20} + (\beta_{21} + \alpha_{21i})x_{lp,j} + \alpha_{20i} + \gamma_{1j}))$

In the above formula, the means of the two Bernoulli distributions (0s and 1s) and the beta distribution (gradient judgements) depend on the same set of predictors, in this case the log-probability ($x_{lp,j}$). There are two population-level coefficients ('fixed effects' in frequentist terms) for each of the three parts of the model, namely the population-level intercept β_{00}, β_{10} and β_{20} and the population-level slopes β_{01}, β_{11} and β_{21}. There are also participant-level predictors ('random effects' in frequentist terms) that allow for variability across participants, including the three random intercepts α_{00i}, α_{10i} and α_{20i}, and the three random slopes α_{01i}, α_{11i} and α_{21i}. Finally, there is an item-level intercept.

The means of the Bernoulli distributions are related to the linear predictors through a logit link, as is the case for standard logistic regression. For the beta regression, the formula shown here is derived from a reparameterisation of the beta regression in terms of the mean and a precision parameter φ.

We will now look at the distributions of the model parameters in detail. Firstly, the group-level effects for each component come from bivariate normal distributions. The covariance matrix allows for correlations. There is a Lewandowski-Kurowicka-Joe (LKJ) prior with one degree of freedom (Lewandowski *et al.* 2009) on the lower Cholesky decomposition of the correlation matrix, and half-*t* priors (Gelman 2006) on the standard deviations, as in (8).

(8) $(\alpha_{c0i}, \alpha_{c1i}) \sim N(0, \sum_{\alpha c})\ for\ c \in \{0, 1, 2\}, i \in \{1, 2, \ldots I\}$
where $\sum_{\alpha c} = D_{ac}R_{ac}D_{ac}$, $R_{\alpha c} = L_{ac}L_{ac}^{T}$, $D_{ac\ c} = diag(\sigma_{\alpha c1}, \sigma_{\alpha c2})$,
$L_{ac} \sim LKJ(1)$, $\sigma_{\alpha c1}, \sigma_{\alpha c2} \sim half - t(3, 0, 2.5)$

The item-level intercept simply follows a univariate normal distribution, again with a half-t prior on its standard deviation, as in (9).

(9) $\gamma_{0j} \sim N(0, \sigma_{\Upsilon c})\, for\ c \in \{0, 1, 2\}, i \in \{1, 2, \ldots I\}, \sigma_{\Upsilon c} \sim half - t(3, 0, 2.5)$

There is a default standard normal prior on the 'fixed-effect' slopes, a t-distributed prior on the population-level intercept for the beta component, and a logistic-distributed prior on the population-level intercept for the logistic components, as in (10).

(10) $\beta_{c1} \sim N(0, 1)\, for\ c \in \{0, 1, 2\}$
$\beta_{01}, \beta_{02} \sim Logistic(0, 1)$
$\beta_{00} \sim t(3, 0, 2.5)$

Finally, there is a gamma prior on the precision parameter of the beta distribution, as in (11).

(11) $\varphi \sim \Gamma(0{,}01, 0.01)$

ADDITIONAL REFERENCES

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

Lewandowski, Daniel, Dorota Kurowicka & Harry Joe (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* **100**. 1989–2001.

Appendix D: Confidence intervals for the multiverse analysis

		T|O	T|N	T|C	T|S	T	NoT
a.	NN	(0.03, 0.08)	(0.05, 0.11)	(0.03, 0.08)	(0.02, 0.07)	(0.06, 0.12)	(0.03, 0.08)
	GCM	(0.05, 0.08)	(0.06, 0.09)	(0.04, 0.08)	(0.04, 0.07)	(0.06, 0.10)	(0.06, 0.09)
	none	(0.04, 0.08)	(0.05, 0.08)	(0.05, 0.08)	(0.06, 0.10)	(0.06, 0.10)	(0.06, 0.10)
b.	NN	(−0.03, 0.04)	(−0.03, 0.04)	(0.00, 0.07)	(−0.02, 0.04)	(−0.03, 0.05)	(0.00, 0.07)
	GCM	(−0.02, 0.03)	(−0.02, 0.03)	(−0.01, 0.04)	(−0.01, 0.03)	(−0.02, 0.03)	(−0.02, 0.03)
	none	(−0.01, 0.03)	(−0.02, 0.03)	(−0.00, 0.04)	(−0.02, 0.03)	(−0.02, 0.03)	(−0.02, 0.03)
c.	NN	(0.13, 0.44)	(0.30, 0.64)	(0.13, 0.45)	(0.06, 0.35)	(0.37, 0.73)	(0.13, 0.45)
	GCM	(0.27, 0.50)	(0.35, 0.58)	(0.25, 0.47)	(0.21, 0.43)	(0.37, 0.61)	(0.36, 0.59)
	none	(0.28, 0.51)	(0.28, 0.51)	(0.27, 0.48)	(0.37, 0.60)	(0.38, 0.62)	(0.38, 0.62)

Table V

Multiverse results for the 95% CI of log-probability effect on (a) gradient judgements, (b) 0 judgements, (c) 1 judgements.

<table>
<tr><th></th><th></th><th>T|O</th><th>T|N</th><th>T|C</th><th>T|S</th><th>T</th><th>NoT</th><th>none</th></tr>
<tr><td>a.</td><td>NN</td><td>(−0.00, 0.01)</td><td>(−0.01, 0.00)</td><td>(−0.00, 0.01)</td><td>(−0.00, 0.01)</td><td>(−0.01, 0.00)</td><td>(−0.00, 0.01)</td><td>(−0.01, 0.00)</td></tr>
<tr><td></td><td>GCM</td><td>(0.01, 0.10)</td><td>(0.01, 0.10)</td><td>(−0.00, 0.01)</td><td>(0.00, 0.02)</td><td>(0.00, 0.02)</td><td>(0.00, 0.02)</td><td>(0.00, 0.02)</td></tr>
<tr><td>b.</td><td>NN</td><td>(−0.01, 0.01)</td><td>(−0.01, 0.01)</td><td>(−0.01, 0.00)</td><td>(−0.01, 0.01)</td><td>(−0.01, 0.01)</td><td>(−0.01, 0.00)</td><td>(−0.01, 0.01)</td></tr>
<tr><td></td><td>GCM</td><td>(−0.01, 0.02)</td><td>(−0.01, 0.02)</td><td>(−0.01, 0.01)</td><td>(−0.01, 0.01)</td><td>(−0.01, 0.02)</td><td>(−0.01, 0.02)</td><td>(−0.01, 0.01)</td></tr>
<tr><td>c.</td><td>NN</td><td>(0.00, 0.07)</td><td>(−0.04, 0.04)</td><td>(−0.01, 0.07)</td><td>(0.01, 0.08)</td><td>(−0.05, 0.02)</td><td>(−0.01, 0.07)</td><td>(−0.05, 0.02)</td></tr>
<tr><td></td><td>GCM</td><td>(0.02, 0.11)</td><td>(0.02, 0.11)</td><td>(−0.01, 0.09)</td><td>(0.01, 0.11)</td><td>(0.01, 0.10)</td><td>(0.02, 0.11)</td><td>(0.02, 0.13)</td></tr>
</table>

Table VI

Multiverse results for the 95% CI of neighbourhood-density effect on (a) gradient judgements, (b) 0 judgements, (c) 1 judgements.