Multivariate count data analysis using Bayesian hierarchical multinomial-t compound regression: A demonstration With collocations

Bayesian Hierarchical Multinomial-t Compound

James Uanhoro [1] Silvia Aguinaga Echeverria [2]

[1]Ohio State University

[2]University of Navarra

uanhoro.1@osu.edu

April 13, 2021

Outline

Context for Presentation

Background Information

- Collocation is a system of words that tend to be found together, e.g. "make the bed", "do homework", ...
- Higher collocation use comes with greater language acquisition.

Background Information

- Collocation is a system of words that tend to be found together, e.g. "make the bed", "do homework", ...
- Higher collocation use comes with greater language acquisition.

Design

- Oral interviews with 20 intermediate level speakers (L2), 20 advanced level speakers (L2) and 20 native speakers of Spanish.
- Interview duration was consistent across speakers.
- Interview text was coded for seven types of collocations.

Background Information

- Collocation is a system of words that tend to be found together, e.g. "make the bed", "do homework", ...
- Higher collocation use comes with greater language acquisition.

Design

- Oral interviews with 20 intermediate level speakers (L2), 20 advanced level speakers (L2) and 20 native speakers of Spanish.
- Interview duration was consistent across speakers.
- Interview text was coded for seven types of collocations.

Statistical question

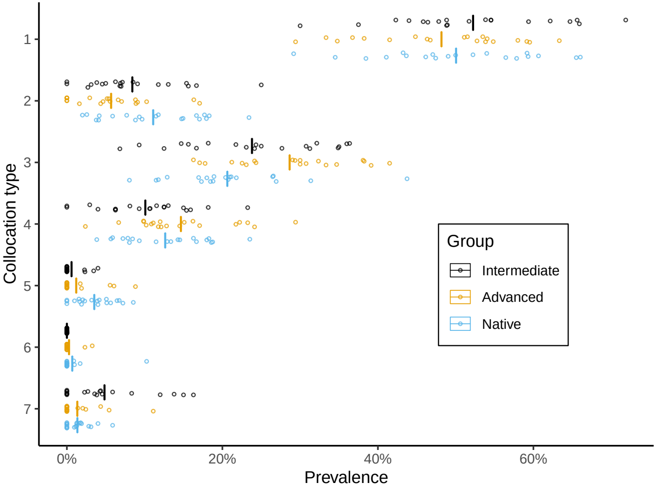
Did the three groups differ in their collocation use across the seven different collocations?

Table: Sample data

Person	Group	C1	C2	C3	C4	C5	C6	C7
i1	Intermediate	11	3	5	3	0	0	1
a1	Advanced	19	0	9	2	0	0	0
n1	Native	48	26	21	8	8	0	0

Number of times individual used collocation of a given type

Interest: Prevalence of different collocation types by group



Prevalence of collocation types (count of each collocation / total collocations) for each speaker.

Methods

Hierarchical multinomial

Rationale

- Multinomial has Poisson marginals (Townes, 2020)
- Hierarchical approach to regularize group coefficient estimation (Gelman et al., 2013)

Hierarchical multinomial

Rationale

- Multinomial has Poisson marginals (Townes, 2020)
- Hierarchical approach to regularize group coefficient estimation (Gelman et al., 2013)

Basic model:

$$lp_{gc} = \beta_c + \delta_{gc}, \quad p_{gc} = \frac{\exp\left(lp_{gc}\right)}{\sum_{c=1}^{7} \exp\left(lp_{gc}\right)}$$

$$use_i \sim \text{Multinomial}(p_{g1}, p_{g2}, \ldots, p_{g7})$$

use_i = count vector for individual i, β_c = collocation effect (7-levels), δ_{gc} = collocation By group interaction (21-levels)

Hierarchical multinomial

Rationale

- Multinomial has Poisson marginals (Townes, 2020)
- Hierarchical approach to regularize group coefficient estimation (Gelman et al., 2013)

Basic model:

$$lp_{gc} = \beta_c + \delta_{gc}, \quad p_{gc} = \frac{\exp\left(lp_{gc}\right)}{\sum_{c=1}^{7} \exp\left(lp_{gc}\right)}$$

$$use_i \sim \text{Multinomial}(p_{g1}, p_{g2}, \ldots, p_{g7})$$

use_i = count vector for individual i, β_c = collocation effect (7-levels), δ_{gc} = collocation By group interaction (21-levels)

Hierarchical priors:

$$\beta_c \sim \mathcal{N}(0, s_\beta), \quad s_\beta \sim t^+(3, 0, 1)$$

$$\delta_{gc} \sim \mathcal{N}(0, s_\delta), \quad s_\delta \sim t^+(3, 0, 1)$$

Hierarchical Dirichlet-multinomial

Rationale

- Multinomial fails to account for overdispersion
- Dirichlet-multinomial (negative binomial marginals, Townes, 2020) does

Hierarchical Dirichlet-multinomial

Rationale

- Multinomial fails to account for overdispersion
- Dirichlet-multinomial (negative binomial marginals, Townes, 2020) does

Model:

$$lp_{gc} = \beta_c + \delta_{gc}, \quad p_{gc} = \frac{\exp\left(lp_{gc}\right)}{\sum_{c=1}^{7} \exp\left(lp_{gc}\right)}$$

$$use_i \sim \text{DirichletMultinomial}([p_{g1}, p_{g2}, \ldots, p_{g7}] \times \kappa_g)$$

$$\kappa_g \sim \text{Gamma}(1, 0.1)$$

κ_g = overdispersion parameter permitted to vary by group

Retained same hierarchical priors from multinomial model.
Dirichlet-multinomial (marginal likelihood) coded in Stan.

Hierarchical multinomial-t compound

Rationale

- Poisson-normal compound to handle overdispersion (e.g. Hinde, 1982)
- Replace normal with t to handle outliers

Hierarchical multinomial-t compound

Rationale

- Poisson-normal compound to handle overdispersion (e.g. Hinde, 1982)
- Replace normal with t to handle outliers

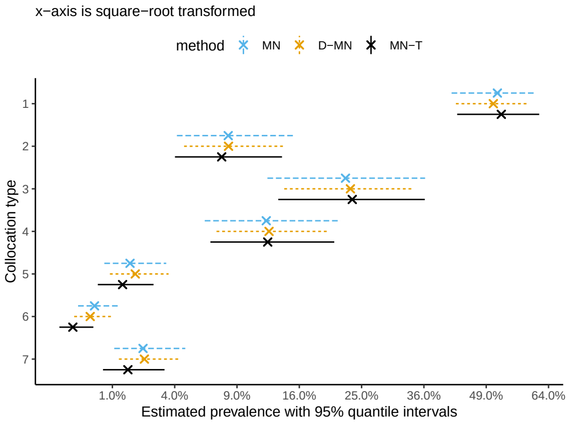
Model:

$$\begin{aligned} lp_{ic} &= \beta_c + \delta_{gc} + \gamma_{ic}, \quad p_{ic} = \frac{\exp\left(lp_{ic}\right)}{\sum_{c=1}^{7}\exp\left(lp_{ic}\right)} \\ use_i &\sim \text{Multinomial}(p_{i1}, p_{i2}, \ldots, p_{i7}) \\ \gamma_{ic} &\sim t(\nu, 0, s_c),\ \nu \sim \text{Gamma}(1, 0.1),\ s_c \sim t^{+}(3, 0, 1) \end{aligned} \tag{1}$$

Retained same hierarchical priors from multinomial model.

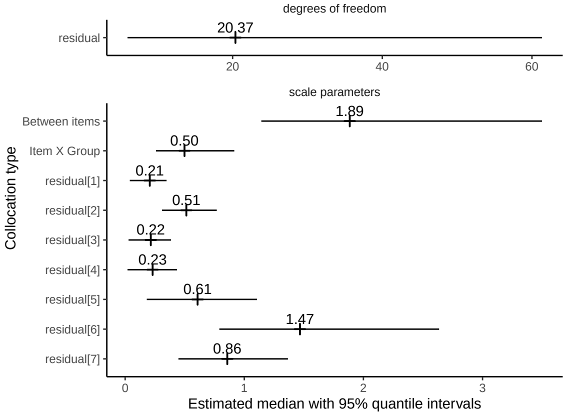
Results

Sampler: Stan (Carpenter et al., 2017), models passed both sampler-agnostic (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2020) and sampler-specific (Betancourt, 2018) diagnostics. 1,000 post-warmup iterations across 12 chains.

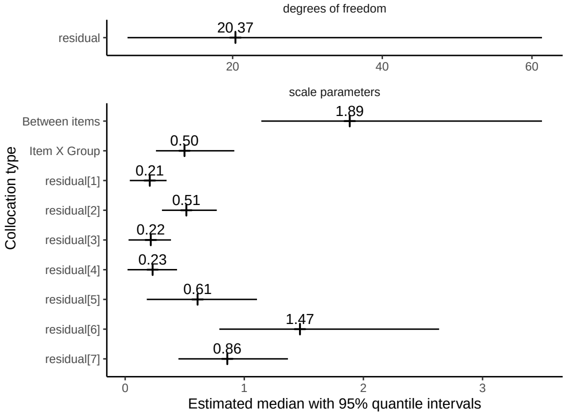


Estimated prevalence of collocation types by model. MN = multinomial, D-MN = Dirichlet-multinomial, MN-T = multinomial-t compound.

High-level model insights are about the same. Multinomial-t model has more parameters to learn from.



Degrees of freedom and scale parameters from multinomial-t model.

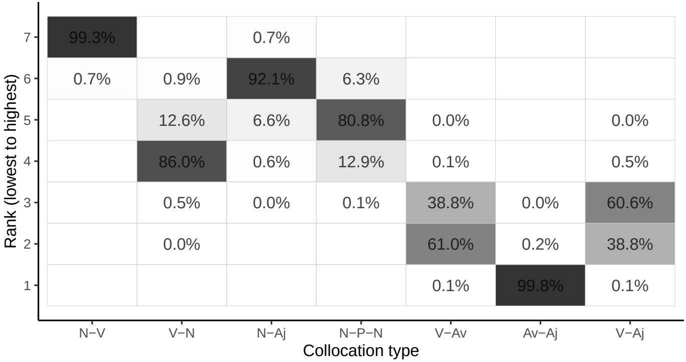


Degrees of freedom and scale parameters from multinomial-t model.

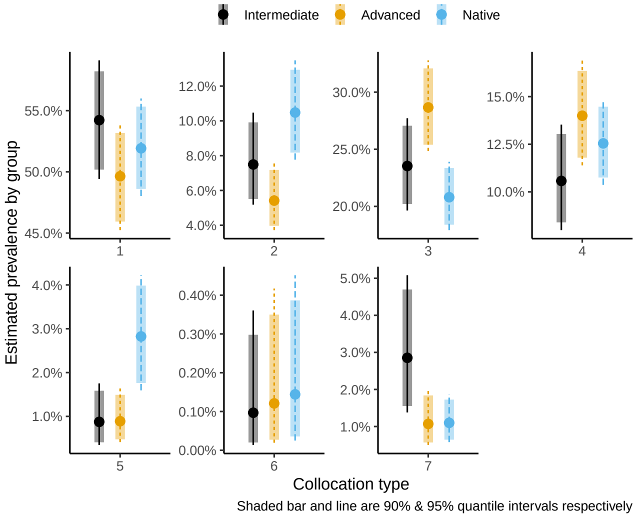
Notes

- Much of the variance is between items (collocation types), interaction accounts for less
- Residual variation differs markedly across collocation types
- Degrees of freedom is highly uncertain

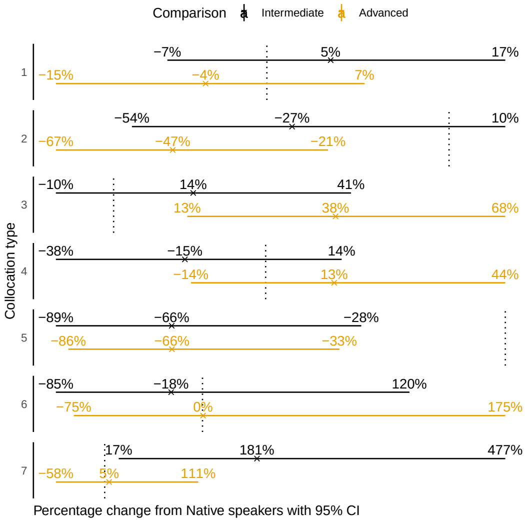
Ranking the average preference for collocations



Collocation use rate by group



Comparing L2 speakers to native speakers



In closing

In closing

Ongoing work with the generalized Dirichlet-multinomial

- Dirichlet-multinomial imposes restrictions on the correlation between the prevalences
- Generalized Dirichlet-multinomial eases these restrictions while doubling the number of parameters - "How would hierarchical estimation proceed?"

Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). doi: 10.18637/jss.v076.i01

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (No. 4). Chapman and Hall/CRC.

Hinde, J. (1982). Compound Poisson regression models. In R. Gilchrist (Ed.), *GLIM 82: Proceedings of the International Conference on Generalised Linear Models* (pp. 109–121). New York, NY: Springer New York.

Townes, F. W. (2020). *Review of probability distributions for modeling count data.* Retrieved from `https://arxiv.org/abs/2001.04343`

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2020, Jul). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*. Retrieved from http://dx.doi.org/10.1214/20-BA1221 doi: 10.1214/20-ba1221