**Appendix S1.**

# V about n

(VP !<# VBN) <2 ((PP <1 (IN < about)) !<< (NP << CD))

# V across n

(VP !<# VBN) <2 (PP <1 (IN < across))

# V after n

(VP !<# VBN) <2 (PP <1 (IN < after))

# V against n

(VP !<# VBN) <2 (PP <1 (IN < against))

# V among n

(VP !<# VBN) <2 (PP <1 (IN < among))

# V around n

(VP !<# VBN) <2 ((PP <1 (IN < around|round)) !<< (NP << CD))

# V as n

(VP !<# VBN) <2 (PP <1 (IN < as) !<< (NP << (JJ|RB < much|little|few|many)) &!< (NP << (NN < result)))

# V between n

(VP !<# VBN) <2 (PP <1 (IN < between))

# V for n

(VP !<# VBN) <2 (PP <1 (IN < for) !<< (ADVP < (RB < long)))

# V into n

(VP !<# VBN) <2 (PP <1 (IN < into))

# V like n

(VP !<# VBN &!<# (VBP < dont)) <2 (PP <1 (IN < like))

# V of n

VP !<# VBN) <2 (PP <1 (IN < of))

# V off n

(VP !<# VBN) <2 (PP <1 (IN < off))

(VP < (PRT < (RP < off)))

# V over n

(VP !<# VBN) <2 ((PP <1 (IN < over)) !<< (NP << CD))

# V through n

(VP !<# VBN) <2 (PP <1 (IN < through))

# V toward n

(VP !<# VBN) <2 (PP <1 (IN < towards|toward))

# V after n

(VP !<# VBN) <2 (PP <1 (IN < under) < NP)

# V with n

(VP !<# VBN) <2 (PP <1 (IN < with))

# V in n

VP < ((PP <# (IN < in)) < (NP <# (/^NN/ !< (order|case|charge|front|favour|favor|addition

|general|fact|following|touch))))

**Appendix S2.**

Table S1. Results of the ZINB model for verb-VAC combinations in L1 German learners, with raw counts and log10(corpus\_size) as the offset variable

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Predictors | Estimates | Incidence Rate Ratios (IRR) | 95% CI for IRR | *z* | *p* |
| **Count Model** |  |  |  |  |  |
| (Intercept) | -6.124 | 0.002 | 0.001–0.004 | -19.013 | <0.001\* |
| BNC freq | 0.001 | 1.001 | 1.001–1.001 | 19.723 | <0.001\* |
| Group [GA2] | -0.02 | 0.976 | 0.460–2.072 | -0.063 | 0.949 |
| Group [GB1] | 0.776 | 2.173 | 1.063–4.443 | 2.128 | 0.033\* |
| Group [GB2] | 1.019 | 2.771 | 1.374–5.589 | 2.848 | 0.004\* |
| Group [GC1] | 0.534 | 1.705 | 0.735–3.954 | 1.244 | 0.214 |
| **Zero-inflated model** |
| (Intercept) | -3.394 | 0.034 | 0.015–0.075 | -8.293 | <0.001\* |
| BNC freq | -0.018 | 0.982 | 0.975–0.990 | -4.455 | <0.001\* |
| Group [GA2] | -1.580 | 0.206 | 0.079–0.539 | -3.220 | 0.001\* |
| Group [GB1] | -1.749 | 0.174 | 0.072–0.421 | -3.874 | <0.001\* |
| Group [GB2] | -1.929 | 0.145 | 0.060–0.350 | -4.307 | <0.001\* |
| Group [GC1] | 0.232 | 1.261 | 0.440–3.612 | 0.432 | 0.666 |
| Theta  | 0.348 |  |  |  |  |
| AIC | 2383.78 |  |  |  |  |

\* *p* < 0.05

Table S2. Results of the ZINB model for verb-VAC combinations in L1 Spanish learners, with raw counts and log10(corpus\_size) as the offset variable

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Predictors | Estimates | Incidence Rate Ratios (IRR) | 95% CI for IRR | *z* | *p* |
| **Count Model** |  |  |  |  |  |
| (Intercept) | -7.303 | 0.001 | 0.000–0.001 | -24.217 | <0.001\* |
| BNC freq | 0.002 | 1.002 | 1.002–1.002 | 12.218 | <0.001\* |
| Group [SA2] | 0.828 | 2.289 | 1.191–4.401 | 2.484 | 0.013\* |
| Group [SB1] | 1.401 | 4.058 | 2.177–7.563 | 4.410 | <0.001\* |
| Group [SB2] | 1.903 | 6.704 | 3.421–13.138 | 5.544 | <0.001\* |
| Group [SC1] | 1.317 | 3.732 | 1.618–8.608 | 3.088 | 0.002\* |
| **Zero-inflated model** |
| (Intercept) | -4.988 | 0.007 | 0.003–0015 | -11.926 | <0.001\* |
| BNC freq | -0.020 | 0.980 | 0.971–0.990 | -3.953 | <0.001\* |
| Group [SA2] | -0.927 | 0.396 | 0.149–1.053 | -1.856 | 0.063 |
| Group [SB1] | -1.284 | 0.277 | 0.105–0.728 | -2.602 | 0.009\* |
| Group [SB2] | 0.340 | 1.405 | 0.564–3.505 | 0.730 | 0.465 |
| Group [SC1] | 2.159 | 8.663 | 2.651–28.308 | 3.574 | <0.001\* |
| Theta  | 0.359 |  |  |  |  |
| AIC | 2521.06 |  |  |  |  |

\* *p* < 0.05