

Catalan stress is not quantity-sensitive: evidence from wordlikeness judgements

Scott James Perry, Anja Arnhold

University of Alberta

This paper evaluates two contrasting theoretical accounts of lexical stress (hereafter stress) in Catalan by presenting experimental data. The descriptive facts of Catalan stress are not in contention: stress occurs in a three-syllable window along the right edge of the word, and native words do not have antepenultimate stress when the penultimate syllable is heavy/bimoraic (Hualde, 2013). Theoretical accounts have dealt with a small set of loanwords that deviate from the native pattern – *Washington* ['wa.ʃɪŋ.tən], *Amsterdam* ['am.ster.ðəm], and *Manchester* ['man.tʃes.tər] – in different ways. For example, Torres-Tamarit and Pons-Moll (2019) account for these words with their default set of constraints, which can generate antepenultimate stress with a heavy penult. By contrast, Wheeler (2005) claims that these words are truly exceptional, and that the grammar does not generate this stress pattern. We address this debate empirically by investigating the following research question: Do the wordlikeness judgements of native speakers support either of the theoretical positions introduced above?

We designed a wordlikeness experiment that elicited judgements from native speakers of Catalan (N=60) using the online experimental platform *wordlikeness* (Chen & Myers, 2017). Participants listened to a set of pseudowords and decided for each pseudoword whether or not it sounded like a possible word in Catalan. Six four-syllable pseudowords were created in each of five syllable structure conditions: one with only light syllables (CV) and the remaining four with a heavy (CVC) first, second, third or fourth syllable. The pseudowords were recorded in a sound attenuated booth by a native speaker of Catalan who produced each pseudoword 4 times, stressing each syllable in turn to create four different stress patterns (initial, antepenultimate, penultimate, and final stress) for a total of 120 stimuli (5 syllable structures X 6 items X 4 stress positions). This design experimentally controlled for phonotactic probability, a factor that has been shown to affect wordlikeness judgements (e.g., Bailey et al., 2001). A second native speaker of Catalan independently confirmed the placement of the stressed syllable in each item. Additionally, we examined the duration, mean pitch, and mean intensity of all vowels acoustically in Praat (Boersma & Weenink, 2016). As expected, stressed syllables had longer durations, higher pitch, and higher intensity than their unstressed counterparts, differences which were statistically significant as evidenced by linear mixed-effects modelling.

Each participant judged 60 items balanced across conditions with the order fully randomized. To examine how wordlikeness judgements vary based on the different stress position and syllable structure conditions, we analyzed our data with Bayesian logistic mixed-effects regression in the *brms* package in R (Bürkner, 2017; R Core Team, 2013). We chose Bayesian modelling as it offers numerous advantages over frequentist modelling, such as being able to compute the probability that a hypothesis is true, and to differentiate between absence of evidence and evidence of absence (McElreath, 2018; Vasisht et al., 2018).

In our analysis, the binary wordlikeness judgement (yes/no) was the dependent variable. An initial model was fit where syllable structure and stress pattern were fixed effects with an interaction specified. The optimal model was found through backwards selection, with model comparisons based on the leave-one-out information criterion. Participant and item were random intercepts in all models. Six chains were run with 40,000 iterations on each. Due to the absence of empirical research on this topic, we implemented the weakly informative priors recommended by Gelman et al. (2008).

Posterior distributions of parameter estimates from the optimal model are presented in *Figure 1*. This model had stress pattern as the only fixed effect. Final and penultimate stress received the highest estimates in the model, meaning words with these stress patterns were most likely to be rated as sounding like possible Catalan words. Antepenultimate stress was rated worse than final and penultimate stress, but better than initial stress.

Both of the theoretical accounts of Catalan discussed above assumed that Catalan stress was at least partially quantity-sensitive. If this corresponds to psychological reality, we should expect there to be an interaction of stress position and syllable structure in our models predicting wordlikeness response. This was not found. The present study presents experimental data which supports the fact that the Catalan stress system is not quantity-sensitive in the minds of speakers, mirroring what has been found for Spanish (Alvord, 2003).

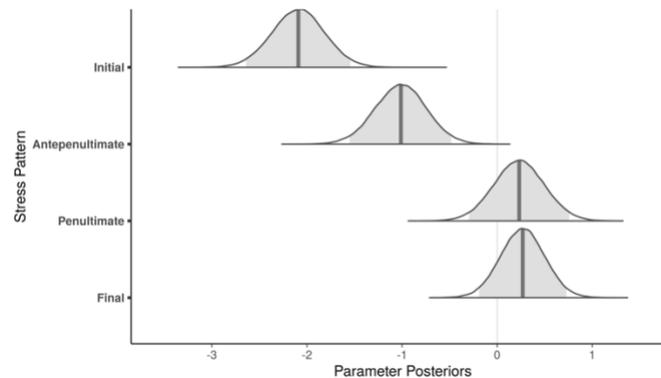


Figure 1. Posterior distributions for fixed-effect parameter estimates from optimal model. The mean parameter estimate is indicated by the dark-gray line at the center of each distribution. The light-gray shaded area of each posterior distribution corresponds to a 95% credible interval.

References

- Alvord, S. M. (2003). The psychological unreality of quantity sensitivity in Spanish: experimental evidence. *Southwest Journal of Linguistics*, 22(2), 1-13.
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods?. *Journal of Memory and Language*, 44(4), 568-591.
- Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer.[Computer program]. Version 6.0. 19. Online: <http://www.praat.org>.
- Bürkner, Paul-Christian. "brms: An R package for Bayesian multilevel models using Stan." *Journal of Statistical Software* 80.1 (2017): 1-28.
- Chen, T. Y., & Myers, J. (2017). Worldlikeness: a web-based tool for typological psycholinguistic research. *University of Pennsylvania Working Papers in Linguistics*, 23(1), 4.
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4), 1360-1383.
- Hualde, J. I. (2013). *Catalan*. Routledge.
- McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.
- Team, R. C. (2013). R: A language and environment for statistical computing.
- Torres-Tamarit, F., & Pons-Moll, C. (2019). Enclitic-induced stress shift in Catalan. *Journal of Linguistics*, 55(2), 407-444.
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71, 147-161.
- Wheeler, M. (2005). *The phonology of Catalan*. Oxford University Press on Demand.