

My project proposes to test the null hypothesis; A word or concept to which it represents measured affect and arousal do not contribute to any significant change in the variance and frequency of its colexification. I believe this hypothesis will be rejected by the data when analysed as there is evidence of affective meaning being used to train predictive algorithms in the paper 'Colexification Networks Encode Affective Meaning' (Di Natale et al., 2021). This study shows evidence that they were able to use affective ratings of words using the lexicons from KKB and NRC VAD, and data from CLICS, OmegaWiki and FreeDict to train a neural network with the expected valances. The results showed that this ML trained algorithm was able to strongly predict affective words (Di Natale et al., 2021).

This finding of embedded affect compliments the research we have been presented in class showing the influence of cognitive efficiency in shaping colexification as word meanings evolve and perhaps could potentially be considered in future as a moderator for effects in the efficiency (Xu et al., 2020). Additionally, this analysis is important as research into affect and arousal in social cognition utilize the information construct for their adaptive contribution to decision making(Isbell et al., 2013). This could offer future avenues of study to test the idea that perhaps affect information, much like semantic meaning, might have a goldilocks zone where its emotional relatedness is just right for colexification but the arousal levels might be where the zone drops of as dangerous or very intense situations would require a distinctness similar to 'right' or 'left' which deters colexification (Brochhagen & Boleda, 2022).

The methods I plan to employ to test my hypothesis are extracting the top quartile of the most colexified words, and the bottom quartile of least colexified words and use the sets for comparison. I will then employ the arousal and affect lexicons to rate the affect and arousal separately for each group of data and perform a t-test comparing the bottom and top quartiles for affect and arousal separately to determine if there is significance.

To illustrate my data, I will do scatter plots of both affect and arousal measures for the quartiles against colexification values to visually assess any relationships. I will use a frequency distribution of valence of affect arousal measures to consider the shape of distributions and if that shape might represent data issues.

I will be doing all the work as I do not have a partner. I will do my best to start early and immediately seek assistance along the way as well as push regularly to my GitHub repository to chronicle my progress. My goal is to have all the methods for extracting information complete in 1-1 ½ weeks, and have the functions designed for testing. The following week my goal is to start applying the functions to the data and do my first visualizations of the relationships. In the third and fourth week I will formally test my hypothesis using t-scores against permutations reflecting significant p-values. I will then present the data in a formal write up and give appropriate visualizations to convey my results clearly.

