

Mining Twitter for New Words

Lance Hahn

Western Kentucky University

William Walters

Western Kentucky University

Taylor Blaetz

Western Kentucky University

Abstract: New lexical elements such as LOL are appearing in natural digital language at high frequencies. The usage of these elements suggests that they are being treated like real words. The first step in examining this type of element is to identify them. We gathered 2,798 messages within a 10-mile radius of a specific GPS location for a 10.5 hour period. The novel elements were identified by excluding words found within an English word list. The majority of the novel high-frequency elements were manually identified as abbreviations, slang, emoticons or shortened words. A larger follow-up data collection of over 580,000 messages within the same area, but over several months, produced correlated usage frequencies but, as expected, revealed more subtle frequency differences and challenges associated with the temporal aspect of the data. We conclude by describing how we are currently evaluating the mental processing of these novel elements through behavioral measures.