*Readers: Social scientists in various fields. Most of them are faculty members of social sciences schools across the United States. They acquired their Ph.D. training in relevant fields. They have good critical thinking ability and are well-familiar with academic discussion and problem-solving processes. The readers may have no background in linguistics or computer sciences, but they are generally interested in the rising field of computational content analysis, which analyzing speech, text, images, and other human-generated contents to investigate the intrapersonal and interpersonal states. They are curious about the theoretical foundations and cost and benefit of the computational content analysis methods. They will use this information to evaluate if the new methods are proper to be employed in their own fields.*

Finding Experience in Text

To perceive the world, human develops languages (Gibbs, 2003). Languages are not only a product of speech but also an expression of human society’s cultural conventions (Saussure, 1959). Interpreted in the languages, these conventions interplay with human cognition, influence how people see themselves, other things, individuals, and the society (Carley, 1994; Levi-Strauss, 1976). Therefore, from the languages, human behaviors and the social facts can be observed (Saussure, 1959). This observation can also go beyond tangible realities. Through law, myth, and other linguistic products, attitudes of people, values of a society, and cultural structures behind all the tangibles and intangibles can be examined. Researchers analyzed linguistic patterns to inspect human conceptual thoughts (Lakoff & Johnson, 1980; Raymond W. Gibbs, 1994) and answer social questions (Evans & Aceves, 2016). At a macro level, scholars detect cultural, linguistics, and technology trends in a society (Michel, 2011); at a micro level, they identify individual opinions (Pang & Lee, 2008) perhaps from on-line posts for a stance in a debate (P Anand, M Walker, 2011) and from customer comments for features of a product (Duric & Song, 2012) and customer opinions corresponding to those features (Nan Tian, Yue Xu, Yuefeng Li, 2014; H. Zhang, Sekhari, Ouzrout, & Bouras, 2016).

To mine those customer comments, one common approach is sentiment analysis (M. Zhang & Ye, 2008). From product reviews or relevant texts, sentiment analysis extracts sentiment concepts toward a given topic, such as a writer’s positive or negative valance to a product (Agarwal & Bhattacharyya, 2005; Kim & Hovy, 2004; Yi, Nasukawa, Bunescu, & Niblack, 2003). From the meaning of each of single words and of higher level linguistic patterns shown in the texts, this method evaluates semantic intimacy of the sentiment concepts and the topic, for example, how much a writer like or dislike a product (Claveau & Monceaux-cachard, 2016). Advanced sentiment analyses review more complexed concepts, such as affects (Cambria, 2016) and product consumption experiences (Xiang, Schwartz, Gerdes, & Uysal, 2015). Human experiences are stored, exchanged, and construed with the use of languages (Halliday & Matthiessen, 1999). From languages, these experiences are ready to be discovered with a proper approach.

To label the types of experience and the specific types of experience conveyed in each video game, this research relies on the ideas of sentiment analyses and their linguistic foundations. To identify typical experiences, I curate an experience lexicon, which attempts to contain all potential words connected to human experiences. In studying various linguistic concepts, traditionally, researchers adopt similar approaches by manually curating concept specific lexicons, for example, a lexicon with vocabulary expressing positive sentiments and another with words representing negative ones (Das & Chen, 2007; Hurst & Nigam, 2004; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011; Yi et al., 2003). However, with more than two-hundred thousand English words extant (Oxford Dictionaries, 2017), the lexicon curation process is always extremely time-consuming. Manually curated lexicons also suffer from subjectivity problems because they depend largely on each scholar’s own judgment.

In this research, I build the experience lexicon through a modified approach, identifying the vocabulary of experiential concepts by the words’ definitions in accredited dictionaries. In a study of linguistic sexual equality, this approach was first adopted by Bolukbasi and his colleagues. They, for instance, searched female in all word definitions in a dictionary to locate those words contain the meaning of woman, such as mother, queen, or alumna (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016). In this research, I apply this inversed dictionary search in Webster’s Unabridged Dictionary to identify the experiential words—words related to an experience. Such as feel, emotion, and event, 23 key words are used in the search, resulting in two thousand experiential words constituting the lexicon. The experience lexicon is then expanded by WordNet, a respectful lexical database curating cognitive synonyms by scholars from Princeton University (Agarwal & Bhattacharyya, 2005; Ohana & Tierney, 2009). To ensure the curated lexicons complete, expansions by WordNet are common and proved useful (FL Cruz, JA Troyano, 2011; Verma & Pushpak Bhattacharyya, 2009) and often seen in sentiment analysis approaches (Andreevskaia & Bergler, 2006; Poria et al., 2012).

This research applies the experience lexicon in building a model for identifying the specific types of experience delivered in each video games. To make the result more interpretable, from the lexicon, the experiential words are clustered into experiential concepts within a Google News Word2Vec word embedding space. Word2Vec is a high-dimension dense matrix, which captures syntactic and semantic similarity of each word in a corpus (Jurafsky & Martin, 2016). In a Word2Vec space, each word has its own vector representing its location in the high dimensional space. The distance between word vectors expresses the syntactic and semantic similarities of the corresponding words (Mikolov, Chen, Corrado, & Dean, 2013). These similarities were applied to acquire semblance of deceases by decease names (Suárez-Paniagua, Segura-Bedmar, & Mart’inez, 2015), to identify tags for blog texts (Xue, Fu, & Shaobin, 2014), to cluster sentiments in movie reviews (Pouransari, 2014) and in customer comments of electronic commerce (D. Zhang, Xu, Su, & Xu, 2015). Applied in this study, the Google News Word2Vec is trained by Google in 2013, with a hundred billion words from the Google News service, representing the general usage and meaning of the modern English. This research attempts to cluster the experiential words in this Word2Vec space of general usage to identify general experiential concepts not restricted to the video gaming experiences.

To identify specific types of experience conferred in each video game, I apply expert reviews of the games, offered by video game information platforms with no direct business connection to the game owners. Product reviews provide product information based on product usage experience (Chen & Xie, 2008). They can be used to extract product features (Davril et al., 2013; Jiang & Pan, 2016; Samha, Li, & Zhang, 2014), forming product brand positioning (Kamakura & Moon, 2014). For customers, product review is a form of word-of-mouth (J. Q. Zhang, Craciun, & Shin, 2010), which influences readers’ consumption engagement (Karakaya & Barnes, 2010) and purchasing behaviors (Chevalier & Mayzlin, 2006; Jiménez & Mendoza, 2013). Sentiments occur during the product consumption can also be identified through product reviews (Dang, Zhang, & Chen, 2010; Meyliana & Budiardjo, 2014; Mukherjee & Bhattacharyya, 2012), which presents the consumption experiences of watching a movie (Verma & Pushpak Bhattacharyya, 2009; Vishwanathan, 2014), staying at a hotel (Xiang et al., 2015), visiting a doctor (Wiseman et al., 2015), or playing a video game (Zhu & Zhang, 2006). Instead of including all customer reviews, this study considers only the reviews written by video game experts, who provide reviews with better readability, more objective and deeper into the content of consumptions. The expert reviews are expected to capture, more effectively, experience nuances delivered in the games.

The expert video game reviews are then applied with the algorithm for a video game Word2Vec space. Within this Word2Vec space, I acquire the vector of each game by its corresponding review text (Le & Mikolov, 2014), and, for each game, acquire a score for each experiential concept, which characterizes the amount of that specific experience delivered in that video game.