

Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis

Scott A. Crossley¹ · Kristopher Kyle¹ · Danielle S. McNamara²

© Psychonomic Society, Inc. 2016

Abstract This study introduces the Sentiment Analysis and Cognition Engine (SEANCE), a freely available text analysis tool that is easy to use, works on most operating systems (Windows, Mac, Linux), is housed on a user's hard drive (as compared to being accessed via an Internet interface), allows for batch processing of text files, includes negation and part-of-speech (POS) features, and reports on thousands of lexical categories and 20 component scores related to sentiment, social cognition, and social order. In the study, we validated SEANCE by investigating whether its indices and related component scores can be used to classify positive and negative reviews in two well-known sentiment analysis test corpora. We contrasted the results of SEANCE with those from Linguistic Inquiry and Word Count (LIWC), a similar tool that is popular in sentiment analysis, but is pay-to-use and does not include negation or POS features. The results demonstrated that both the SEANCE indices and component scores outperformed LIWC on the categorization tasks.

Keywords Sentiment analysis · Affect detection · Opinion mining · Natural language processing · Automatic tools · Corpus linguistics

The analysis of sentiment is an important component of a number of research disciplines, including psychology, education, sociology, business, political science, and economics.

✉ Scott A. Crossley
scrossley@gsu.edu

¹ Department of Applied Linguistics/ESL, Georgia State University, 25 Park Place, Suite 1500, Atlanta, GA 30303, USA

² Department of Psychology, Learning Sciences Institute, Arizona State University, Tempe, AZ, USA

Measuring sentiment features automatically in a text is thus of value, to better understand how emotions, feelings, affect, and opinions influence cognition, economic choices, learner engagement, and political affiliation. However, the freely available natural language processing (NLP) tools that measure linguistic features related to sentiment, cognition, and social order are limited. The best-known example of an available sentiment analysis tool is Linguistic Inquiry and Word Count (LIWC), which comprises a number of dictionaries that capture conscious and unconscious psychological phenomena related to cognition, affect, and personal concerns. LIWC has proven extremely useful in a number of different disciplines and has had a large impact on our understanding of how lexical elements related to cognition, affect, and personal concerns can be used to better understand human behavior. However, it has several shortcomings with regard to usability and to the facile and broad measurements of its dictionaries. First, LIWC is not freely available (it costs a modest fee). Second, the LIWC indices are based on simple word counts (some of which are populated by fewer than eight words), and the program does not take into consideration issues of valence such as negations, nor part-of-speech (POS) tags, both of which can have important impacts on sentiment analysis. In addition, the indices reported by LIWC are standalone and do not report on larger constructs related to sentiment.

This article introduces a new sentiment analysis tool called the Sentiment Analysis and Cognition Engine (SEANCE). SEANCE is a freely available text analysis tool that incorporates a number of freely available sentiment dictionaries. The tool is easy to use, works on most operating systems (Windows, Mac, and Linux), is housed on a user's hard drive, allows for batch processing of text files, includes text negation indices and a POS tagger, and reports on a number of component scores specifically developed to make text interpretation easier. In total, the tool reports on over 3,000 classic and

recently developed micro-indices and 20 macro-indices related to sentiment, cognition, and social-order analysis.

In this study, we demonstrate the utility of the sentiment, cognition, and social-order indices provided by SEANCE, with a focus on the domain of positive and negative reviews in two corpora across five domains. We examine the degree to which the features reported by SEANCE are able to predict whether a review is positive or negative, and compare this with the predictive ability of LIWC indices. The reviews used in this study include the 2,000 positive and negative movie reviews collected by Pang and Lee (2004) and the Multi-Domain Sentiment Dataset, which comprises 8,000 Amazon product reviews across four domains: books, DVDs, electronics, and kitchen appliances (Blitzer, Dredze, & Pereira, 2007). These reviews have served as a gold standard for many sentiment analysis investigations. The analyses conducted in this study allow us not only to introduce SEANCE and validate the tool (i.e., by testing its predictive validity in assessing positive and negative writing samples), but to also compare the tool to the current state of the art (LIWC) as well as to examine how lexical features in text are related to the affective state of that text.

Sentiment analysis

The automatic extraction of semantic information related to human feelings and opinions and the subsequent analysis of texts based on this information is categorized under a number of umbrella terms, including *subjectivity* (Langacker, 1985; Lyons, 1981), *opinion mining* (Pang & Lee, 2008), emotion (Collins, Ortony, & Clore, 1988; Ketai, 1975), *affect* (Batson, Shaw, & Oleson, 1992), and *sentiment analysis* (Pang & Lee, 2008). Sentiment is widely associated with feelings, emotions, and opinion, and the term *sentiment analysis* is commonly used as a general term related to extracting subjective information related to human feelings and opinions from natural language texts (Hutto & Gilbert, 2014; Liu, 2012; Pang & Lee, 2008). Sentiment analysis is a useful approach to a number of different problems posed across a number of different disciplines, such as psychology, education, sociology, business, political science, and economics (Hutto & Gilbert, 2014), as well as research fields such as NLP, data mining, and information retrieval (Zhang, Gan, & Jiang, 2014).

The foundations for sentiment analysis can be found in NLP techniques (Hutto & Gilbert, 2014), which can be used to determine the polarity of text segments (sentences, phrases, or whole texts) on the basis of a binary classification of positive or negative affect. Thus, what is being discussed is not the focus of sentiment analysis, but rather the sentiment toward the topics of discussion (Hogenboom, Boon, & Frasincar, 2012).

There are numerous applications of sentiment analysis. For instance, in question-answering systems, knowing the opinions of different sources can provide better answers to users

(Stoyanov, Cardie, Litman, & Wiebe, 2006; H. Yu & Hatzivassiloglou, 2003). In text summarization, sentiment analysis can be used to label and summarize reviews, articles, and blogs (Pang, Lee, & Vaithyanathan, 2002). Sentiment analysis is also helpful in automating decision making by helping organizations better understand the effects of specific issues on people's perceptions and responding to these effects appropriately through marketing and communication (Sauter, 2011). Sentiment analysis is also important to understanding financial markets (Schumaker, Zhang, Huang, & Chen, 2012; Yu, Duan, & Cao, 2013), corporate sales (Ghose & Ipeirotis, 2011; Yu, Liu, Huang, & An, 2012), economic systems (Ludvigson, 2004), medical discourse (De Choudhury, Gamon, Counts, & Horvitz, 2013), politics (Baron, 2005; Tumasjan, Sprenger, Sandner, & Welpe, 2010), and educational discourse (D'Mello & Graesser, 2012).

Approaches to sentiment analysis

Generally speaking, sentiment analysis uses bag-of-words vector representations to denote unordered collections of words and phrases that occur in a text of interest. These vector representations are used in machine-learning algorithms that find patterns of sentiment used to classify texts on the basis of polarity (generally positive or negative texts). Additionally, the vectors can contain information related to semantic valence (e.g., negation and intensification; Polanyi & Zaenen, 2006) and POS tags (Hogenboom et al., 2012). There are two basic approaches to developing these vectors. The first is domain-dependent (also referred to as a *text classification approach*), wherein the vectors are developed and tested within a specific corpus drawn from a specific domain (i.e., a movie review corpus). The second is domain-independent (also referred to as a *lexical-based approach*), in which vectors are developed on the basis of general lists of sentiment words and phrases that can be applied to a number of different domains (Hogenboom et al., 2012).

Domain-dependent approaches involve the development of supervised text classification algorithms from labeled instances of texts (Pang et al., 2002). The approach usually follows a three-step pattern. First, texts are queried for words and phrases (i.e., *n*-grams) that express sentiment. This is sometimes done on the basis of POS tags, but not always. The most successful features in such an approach tend to be basic unigrams (Pang et al., 2002; Salvetti, Reichenbach, & Lewis, 2006). Next, the semantic orientations of the words and phrases are estimated by calculating the pointwise mutual information (i.e., co-occurrence patterns) of the words within the corpus in order to classify the words on the basis of polarity (i.e., positive or negative). The occurrences of these words and phrases are then computed for each text in the corpus and

used as predictors in a machine-learning algorithm to classify the texts as either positive or negative (Turney, 2002).

Classifiers built using supervised methods are generally quite accurate in classifying texts on the basis of polarity within the domain for which they were developed (Bartlett & Albright, 2008; Boiy et al., 2007; Chaovalit & Zhou, 2005; Kennedy & Inkpen, 2006). The problem with such classifiers is that although they perform strongly for the domain in which they were trained, their performance strongly drops (almost to chance) when they are used in different domains (Aue & Gamon, 2005), topics, and even time periods (Read, 2005). For instance, Brooke (2009) extracted the 100 most positive and negative unigrams from the Polarity Dataset of 2,000 movie reviews. Although many of the unigrams were related to positive and negative terms, many were not. For instance, if the plot, director, or writer was mentioned, the review was more often negative. In contrast, unigrams related to the movie's ending or its flaws were predictive of positive movie reviews. Names were also predictive of negative reviews (as was also reported by Finn & Kushmerick, 2003, and Kennedy & Inkpen, 2006), as were words such as *video*, *TV*, and *series*. These terms, when used to examine polarity in different datasets, are less meaningful.

Knowing that domain-dependent methods do not perform well in other domains (Aue & Gamon, 2005), a number of methods have been proposed to create sentiment analysis approaches that offer greater portability. The most common approach is to leverage general word and phrase vectors that are categorized on the basis of associated sentiment and obtained from domain-independent sources, such as corpora, dictionaries, or the Internet (Andreevskaia & Bergler, 2008; Hogenboom, Hogenboom, Kaymak, Wouters, & De Jong, 2010). Like the domain-dependent methods, this approach uses lexicon-based vectors to calculate the orientation of documents on the basis of the aggregation of the individual word scores (Turney, 2002). Such approaches have gained attention in more recent research because their performance is robust across texts and domains (Heerschop, van Iterson, Hogenboom, Frasinicar, & Kaymak, 2011; Hogenboom et al., 2012; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011), and they can be easily enhanced with the inclusion of multiple dictionaries (Taboada, Brooke, & Stede, 2009).

A number of domain-independent sentiment dictionaries are publicly available for use, such as General Inquirer (Stone, Dunphy, Smith, Ogilvie, et al., 1966), SenticNet (Cambria, Havasi, & Hussain, 2012; Cambria, Speer, Havasi, & Hussain, 2010), SO-CAL (Taboada et al., 2009), and EmoLex (Mohammad & Turney, 2010, 2013). These dictionaries usually consist of word vectors that are manually annotated for corresponding polarities, semantic categories, social positioning, or cognitive perspective. Although these dictionaries perform worse than domain-specific models trained on sufficiently large corpora (Pang et al., 2002), they

outperform domain-specific classifiers in out-of-domain training sets or when the training sets are small (Andreevskaia & Bergler, 2008). They also perform well on a number of different domains and texts types. For instance, SO-CAL has proven robust for identifying sentiment in video game reviews (Brooke & Hurst, 2009) and blog postings (Murray and Carenini, 2009).

LIWC

Overview

LIWC is a sentiment analysis tool that is easy to use, transparent, fast, and accurate. As a result, it has been widely used by sociologists, psychologists, computer scientists, and linguists in a number of education and social media domains (Hutto & Gilbert, 2014; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007; Pennebaker, Francis, & Booth, 2001). LIWC is available for a small fee (\$90, at the time of writing) and, once purchased, is housed on the user's hard drive, allowing for secure data processing in the absence of an Internet connection. LIWC was designed to capture conscious and unconscious psychological phenomena related to cognition, affect, and personal concerns. The LIWC dictionary is proprietary and contains about 4,500 words. The word lists that comprise the LIWC dictionary include words that had been compiled from previous lists, thesauri, and dictionaries by researchers and confirmed as construct-relevant by three to six independent judges. The initial word lists were refined through corpus analysis, and new lists were added (Pennebaker et al., 2007; Tausczik & Pennebaker, 2010). LIWC has been used in hundreds of studies to investigate constructs such as social status (Sexton & Helmreich, 2000), deception (Newman, Pennebaker, Berry, & Richards, 2003), and individual differences (Mehl, Gosling, & Pennebaker, 2006). LIWC is not sensitive to POS tags and does not make use of valence markers such as negations. The LIWC software provides information on the percentage of words per text that are covered by its internal dictionary and the percentage of words per text in each of the 80 categories on which it reports. For a complete listing of the categories, see Pennebaker et al. (2007). A brief overview of the sentiment, cognition, and social-order categories reported by LIWC is included below.

Psychological processes

Psychological-processes categories form the heart of LIWC and comprise 32 word categories. These indices can provide information about the psychological states of writers. The psychological-processes category is subdivided into social, affective, cognitive, perceptual, and biological processes, as well as relativity (motion, space, and time) subcategories.

Each subcategory reports a number of variables, all based on word lists.

Personal concerns

LIWC reports on seven lexical categories related to personal concerns. These categories include lists of words related to one's personal life. The categories include work, achievement, leisure, home, money, religion, and death.

SEANCE

Overview

SEANCE is a sentiment analysis tool that relies on a number of preexisting sentiment, social-positioning, and cognition dictionaries. Like LIWC, SEANCE is housed on the user's hard drive, allowing users to work independently of outside servers, which allows for secure processing of sensitive data. Unlike LIWC, SEANCE is freely available and includes negation rules, POS tagging, and broad component scores. SEANCE is written in Python but is implemented in a way that requires little to no knowledge of programming, and it can be started by simply double-clicking the SEANCE icon. The SEANCE interface is an easy-to-use and intuitive graphical user interface (GUI) that requires the user to select an input folder containing the files of interest (in .txt format). The user then selects an output folder for the output file and enters a name for a .csv file into which SEANCE will write the results for each text (the default name is results.csv). The user then selects to process the texts, and a program status box informs the user of how many texts have been processed (see Fig. 1 for the SEANCE GUI). Instructions and explanations for using SEANCE, a user help file, and the program itself are available at www.kristopherkyle.com/seance.html.

SEANCE contains a number of predeveloped word vectors developed to measure sentiment, cognition, and social order. These vectors are taken from freely available source databases, including SenticNet (Cambria et al., 2012; Cambria et al., 2010) and EmoLex (Mohammad & Turney, 2010, 2013). In some cases, the vectors are populated by a small number of words and should be used only on larger texts that provide greater linguistic coverage, to avoid nonnormal distributions of data (e.g., the Lasswell dictionary lists [Lasswell & Namenwirth, 1969] and the Geneva Affect Label Coder [GALC; Scherer, 2005] lists). For many of these vectors, SEANCE also provides a negation feature (i.e., a contextual valence shifter; Polanyi & Zaenen, 2006) that ignores positive terms that are negated. The negation feature, which is based on Hutto and Gilbert

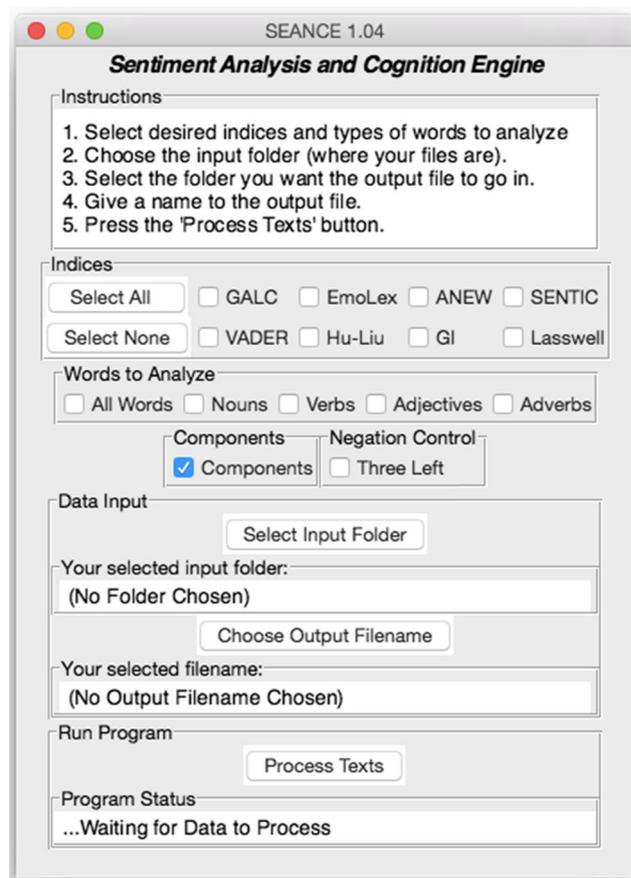


Fig. 1 Graphical user interface of SEANCE

(2014), checks for negation words in the three words preceding a target word. In SEANCE, any target word that is negated is ignored within the category of interest. For example, if SEANCE processes the sentence *He is not happy*, the lexical item *happy* will not be counted as a positive emotion word. This method has been shown to identify approximately 90 % of negated words (Hutto & Gilbert, 2014). SEANCE also includes the Stanford POS tagger (Toutanova, Klein, Manning, & Singer, 2003) as implemented in Stanford CoreNLP (Manning et al., 2014). The POS tagger allows for POS-tagged specific indices for nouns, verbs, and adjectives. POS tagging is an important component of sentiment analysis, because unique aspects of sentiment may be conveyed more strongly by adjectives (Hatzivassiloglou & McKeown, 1997; Hu & Liu, 2004; Taboada, Anthony, & Voll, 2006) or verbs and adverbs (Benamara, Cesarano, Picariello, Reforgiato, & Subrahmanian, 2007; Sokolova & Lapalme, 2009; Subrahmanian & Reforgiato, 2008). SEANCE reports on both POS and non-POS variables. Many of the vectors in SEANCE, for example, are neutral with regard to POS. This allows for SEANCE to accurately process poorly formatted texts that cannot be accurately analyzed by a POS tagger. We briefly discuss below the source databases used

in SEANCE. Table 1 provides an overview of the categories reported in SEANCE and the source databases that report on each category.

Source databases

General inquirer SEANCE includes the Harvard IV-4 dictionary lists used by the General Inquirer (GI; Stone et al., 1966). The GI lists are the oldest manually constructed lists still in widespread use and include 119 word lists organized into 17 semantic categories, containing over 11,000 words. These categories include semantic dimensions, pleasure, overstatements, institutions, roles, social categories, references to places, references to objects, communication, motivation, cognition, pronouns, assent and negation, and verb and adjective types. The lists were developed for content analysis by social, political, and psychological scientists. Greater detail on the categories and available word lists is available at www.wjh.harvard.edu/~inquirer/homecat.htm.

Lasswell SEANCE also includes the Lasswell dictionary lists (Lasswell & Namenwirth, 1969; Namenwirth & Weber, 1987), which are also included in the GI. Included are 63 word lists organized into nine semantic categories. These categories include power, rectitude, respect, affection, wealth, well-

being, enlightenment, and skill. Additional information on these categories and their supporting word lists is available at www.wjh.harvard.edu/~inquirer/homecat.htm.

Geneva affect label coder The GALC is a database composed of lists of words pertaining to 36 specific emotions and two general emotional states (positive and negative; Scherer, 2005). The specific emotion lists include anger, guilt, hatred, hope, joy, and humility.

Affective norms for english words The Affective Norms for English Words (ANEW) database (Bradley & Lang, 1999) includes affective norms for valence, pleasure, arousal, and dominance (Osgood, Suci, & Tannenbaum, 1957). Unlike the LIWC and GI word lists, ANEW word lists have associated sentiment scores that are positive if the score is above 5 and negative if it is below 5 (and neutral if the score is around 5). Bradley and Lang collected these norms by using the Self-Assessment Manikin system (Lang, 1980) to collect norms for 1,033 English words.

EmoLex EmoLex (Mohammad & Turney, 2010, 2013) consists of lists of words and bigrams that evoke particular emotions (e.g., anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). Additionally, EmoLex include lists of

Table 1 An overview of the text categories in SEANCE and their source databases

Category	GALC	EmoLex	ANEW	SenticNet	VADER	Hu-Liu	Harvard IV-4	Lasswell	Total
Action							2		2
Arousal	2		2				1		5
Arts and academics							2	1	3
Cognition				2			9	7	18
Communication							4	1	5
Dominance, respect, money, and power			2				4	24	30
Economics, politics, and religion							9	10	19
Effort							10	2	12
Evaluation							9		9
Feeling/emotion				1			2		3
Negative emotion words	18	5			1	2	7	6	39
Other affect					2			4	6
Physical							13	1	14
Positive emotion words	13	2		1	1	2	4	3	26
Quality and quantity							9		9
Reference							4		4
Social Relations	4	1					17	3	25
Surprise	1	2							3
Time and space							13	1	14
Valence/polarity			2	1		1			4
Total	38	10	6	5	4	5	119	63	250

words and bigrams that generally evoke negative and positive emotions. Word and bigram lists were compiled from entries in the *Macquarie Thesaurus* (Bernard, 1986) that were also frequent in the Google *N*-Gram Corpus (Brants & Franz, 2006), the WordNet Affect Lexicon (Strapparava & Valitutti, 2004), and the GI (Stone et al., 1966). Mohammad and Turney then used Amazon Mechanical Turk to determine which emotions (if any) were evoked by each word or bigram. The ten lists each include between 534 (for surprise) and 3,324 (for negative emotions) entries. EmoLex has been used to examine emotions in mail and e-mail (Mohammad & Yang, 2011) and to investigate emotion in fiction writing (Mohammad, 2012).

SenticNet SenticNet (Cambria et al., 2012; Cambria et al., 2010) is a database extension of WordNet (Fellbaum, 1998) consisting of norms for around 13,000 words with regard to four emotional dimensions (sensitivity, aptitude, attention, and pleasantness), based on work by Plutchik (2001) and norms for polarity. Unlike LIWC, GI, or ANEW, the SenticNet scores were calculated using semisupervised algorithms, and the scores are thus not a gold-standard resource. SenticNet was designed to build and improve upon SentiWordNet (Esuli & Sebastiani, 2006) using a number of data-refining techniques.

Valence aware dictionary for sentiment reasoning The Valence Aware Dictionary for Sentiment Reasoning (VADER) is a rule-based sentiment analysis system (Hutto & Gilbert, 2014) developed specifically for shorter texts found in social media contexts (e.g., Twitter or Facebook). VADER uses a large list of words and emoticons that include crowd-sourced valence ratings. Additionally, the VADER system includes a number of rules that account for changes in valence strength due to punctuation (i.e., exclamation points), capitalization, degree modifiers (e.g., intensifiers), contrastive conjunctions (i.e., *but*), and negation words that occur within three words before a target word. VADER has been used to accurately classify valence in social media text, movie reviews, product reviews, and newspaper articles (Hutto & Gilbert, 2014).

Hu–Liu polarity SEANCE includes two large polarity lists compiled by Hu and Liu (2004) for the purposes of sentiment analysis. The Hu–Liu word lists were developed specifically for product reviews and social texts. The positive word list includes 2,006 entries, whereas the negative word list includes 4,783 entries. Both lists were constructed through bootstrapping processes in WordNet. The Hu–Liu lists have been used to successfully predict whether product reviews were positive or negative (Hu & Liu, 2004; Liu, Hu, & Cheng, 2005).

SEANCE component scores

One potential pitfall with the SEANCE tool is the sheer number of indices that it reports. With the potential for each index to report results for all words, nouns, verbs, and adjectives, in addition to each of these having the potential to be negated, the SEANCE tool can report on almost 3,000 indices. For the uninitiated such a large number of indices can be unwieldy, and they are also potentially unnecessary because of the overlap between databases. Thus, we developed component scores derived from the SEANCE indices to provide users with more manageable options, if desired, and to investigate the potential of combining similar indices into larger, macrofeatures.

To compute the component scores, we adopted an approach similar to those of Graesser, McNamara, and Kulikowich (2011) and Crossley, Kyle, and McNamara (2015). We conducted a principal component analysis (PCA) to reduce the number of indices selected from SEANCE to a smaller set of components, each of which was composed of a set of related features. The PCA, based on the Movie Review Corpus, clustered the indices into groups that co-occurred frequently, allowing for a large number of variables to be reduced into a smaller set of derived variables (i.e., the components). This gave us two approaches to assessing sentiment: a microfeature approach (i.e., the indices individually) and a macrofeature approach (i.e., the indices aggregated into components).

We set a conservative cutoff for the eigenvalues to be included in a component (i.e., .40), to ensure that only strongly related indices would be included in the analysis. For inclusion in the analysis, all variables needed to be normally distributed. We then controlled for multicollinearity between variables (defined as $r \geq .90$), so that the selected variables would not measure the same construct. After conducting the factor analysis, we set a cutoff of 1 % for the variance that needed to be explained by each component included in SEANCE. Components that explained less than 1 % of the variance were removed. For the included component scores, we used the eigenvalues for each included index to create weighted component scores. In total, we developed 20 component scores, which explained 56 % of the variance in the Movie Review Corpus. The 20 components are summarized in Table 2.

Method

To validate the SEANCE indices and component scores, we examined the relations between the indices calculated by SEANCE for two corpora of reviews: the Movie Review Corpus and the Multi-Domain Sentiment Dataset. In both analyses, we use LIWC as a baseline to compare the reliability of the SEANCE results. The Multi-Domain Sentiment Dataset analysis afforded a chance to examine the generalization of the

Table 2 Description of component scores

Component	Label	Number of Indices	Key Indices
1	Negative adjectives	18	EmoLex negative adjectives, EmoLex disgust adjectives, EmoLex anger adjectives, GI negative adjectives, Hu–Liu negative adjectives
2	Social order	11	RC ethics verbs, GI need verbs, RC rectitude words
3	Action	9	GI ought verbs, GI try verbs, GI travel verbs, GI descriptive action verbs
4	Positive adjectives	9	Hu–Liu positive adjectives, Vader positive, GI positive adjectives, Lasswell positive affect adjectives
5	Joy	8	EmoLex joy adjectives, EmoLex anticipation adjectives, EmoLex surprise adjectives
6	Affect for friends and family	9	Lasswell affect nouns, Lasswell participant affect, GI kin noun, GI affiliation nouns
7	Fear and disgust	8	EmoLex disgust nouns, EmoLex negative nouns, NRC fear, NRC anger
8	Politeness	7	GI politeness, GI politeness nouns, Lasswell power
9	Polarity nouns	7	Polarity nouns, Pleasantness nouns, Aptitude nouns
10	Polarity verbs	4	Polarity verbs, Aptitude verbs, Pleasantness verbs
11	Virtue adverbs	5	Lasswell rectitude gain adverbs, GI concerns for hostility adverbs, Lasswell sureness adverbs
12	Positive nouns	4	Hu–Liu nouns
13	Respect	4	Lasswell respect nouns
14	Trust verbs	5	EmoLex trust verbs, EmoLex joy verbs, EmoLex positive verbs
15	Failure	5	Lasswell power loss verbs, GI failure verbs
16	Well being	4	Lasswell well-being physical nouns, Lasswell well-being total
17	Economy	4	GI names adjectives, GI economy adjectives, GI economy all
18	Certainty	6	GI quantity, GI overstatement, Lasswell if, Lasswell sureness nouns
19	Positive verbs	3	Hu–Liu positive verbs
20	Objects	4	GI objects, GALC being touched

SEANCE component scores (derived from the Movie Review Corpus) to a broad set of reviews across a variety of domains (book, DVD, electronics, and kitchen appliance reviews).

Corpora

Movie review corpus This corpus comprises 1,000 positive and 1,000 negative movie reviews collected by Pang and Lee (2004) from the Internet Movie Database (IMDB). The review polarity in this corpus is based on the numerical score given by the reviewer (e.g., a score of 3.5 or higher on a five-point scale is considered positive). This corpus (or portions of it) has been used for a number of sentiment analysis studies that have attempted to classify the reviews as either positive or negative (e.g., Hutto & Gilbert, 2014; Kennedy & Inkpen, 2006). The recent classification accuracies for these studies have ranged from .803 to .872 (Andreevskaia & Bergler, 2008; Kennedy & Inkpen, 2006; Pang & Lee, 2004) for domain-dependent classifiers, and from .581 to .764 for domain-independent classifiers (Kennedy & Inkpen, 2006; Taboada et al., 2006).

Multi-domain sentiment dataset This corpus (Blitzer, Dredze, et al., 2007) comprises 2,000 Amazon product reviews across four domains: books, DVDs, electronics, and kitchen appliances (for a total of 8,000 reviews). Each domain includes equal numbers of positive (earning more than three out of five stars) and negative (earning fewer than three stars) product reviews. The dataset has been used in a number of studies (Blitzer, Dredze, et al., 2007; Blitzer, Crammer, Kulesza, Pereira, & Wortman, 2007; Dredze, Crammer, & Pereira, 2008) to investigate domain-adaptive automatic polarity detection. Blitzer, Dredze, et al. (2007) achieved domain-specific prediction accuracies ranging from 80.4 % (books) to 87.7 % (kitchen appliances) by using unigram and bigram predictors. Cross-domain model prediction accuracies have ranged from 68.6 % accuracy (for the kitchen appliances model adapted to books) to 86.8 % accuracy (for the kitchen appliances model adapted to electronics). To our knowledge, no studies have been conducted with the Multi-Domain Sentiment Dataset using domain-independent predictors (such as those found in LIWC and SEANCE).

Statistical analysis

Our goal was to examine the differences between positive and negative reviews and to use these differences to create a model that would classify each review as either positive or negative. To accomplish this, we first conducted a multivariate analysis of variance (MANOVA), followed by a stepwise discriminant function analysis (DFA). We did this for each corpus and for the LIWC indices, the SEANCE indices, and the SEANCE component scores. Indices reported by these tools that lacked normal distributions were removed. We used the MANOVA to examine which indices reported differences between the positive and negative reviews (i.e., the LIWC and SEANCE indices were the predictor variables, and the positive and negative classifications were the criterion variables). To control for the Type I errors that result from multiple comparisons, we applied Bonferroni corrections to all of the MANOVA analyses. The MANOVA was followed by a stepwise DFA (Field, 2013; Jarvis, 2011) using the selected indices from each tool that had demonstrated significant differences between the negative and positive reviews after Bonferroni corrections and that did not exhibit multicollinearity ($r \geq .90$) with other indices in the set. In the case of multicollinearity, the index demonstrating the largest effect size was retained in the analysis. The DFA provided an algorithm to predict group memberships (i.e., whether the review was positive or negative) through a discriminant function coefficient. A DFA model was first developed for the entire corpus, and the indices from the model were then used to predict group memberships of the reviews, using leave-one-out cross-validation (LOOCV) to ensure that the model was stable across the datasets. To compare prediction accuracies across the three tools, we assigned each review that the DFA predicted correctly a score of 1, and those that were predicted incorrectly a score of 0, across the three feature sets (i.e., the LIWC indices, the SEANCE indices, and the SEANCE component scores). We then conducted a one-way analysis of variance (ANOVA) among the three conditions to examine differences in the prediction accuracies,

with the classification scores as the dependent variable and tools as the independent variable. For both the Movie Review Corpus and the Multi-Domain Sentiment Dataset, we conducted ANOVAs for the entire set. In addition, for the Multi-Domain Sentiment Dataset we conducted individual ANOVAs for the DVD reviews, the book reviews, the electronics reviews, and the kitchen appliance reviews.

Results

Movie review corpus

LIWC indices

MANOVA A MANOVA was conducted using the LIWC indices related to psychological processes and personal concerns as the dependent variables and the positive and negative movie reviews in the Movie Review Corpus as the independent variables. The indices were first checked for normal distributions and then assessed using Pearson correlations for multicollinearity. After these assumptions were checked, 27 indices remained that were used in the MANOVA. The MANOVA indicated that 14 of the variables demonstrated significant differences between the positive and negative movie reviews and reported a p value below .001.

DFA We used the 14 significant variables from the MANOVA analysis as our predictor variables in the DFA. For this analysis, the significance level for a variable to be entered or removed from the model was set at $p \leq .05$. The stepwise DFA retained nine variables (see Table 3 for the descriptive and MANOVA statistics for these variables), and we removed the remaining variables as nonsignificant predictors.

The results demonstrated that the DFA using these nine indices correctly allocated 1,379 of the 2,000 texts in the total set, $\chi^2(df=1, n=2,000) = 287.331, p < .001$, for an accuracy of 69 % (the chance level for this and all other analyses was

Table 3 Descriptive and MANOVA statistics for variables included in movie review DFA: LIWC indices

Index	Positive Review: M (SD)	Negative Review: M (SD)	F	p	η_p^2
Negative emotion	0.019 (0.008)	0.025 (0.009)	164.816	<.001	.076
Negations	0.011 (0.005)	0.013 (0.006)	115.467	<.001	.055
Positive emotion	0.040 (0.013)	0.036 (0.011)	81.124	<.001	.039
Discrepancies	0.009 (0.005)	0.011 (0.005)	78.190	<.001	.038
Inclusive	0.042 (0.010)	0.039 (0.009)	60.817	<.001	.030
Swear words	0.001 (0.002)	0.001 (0.002)	29.087	<.001	.014
Anger	0.007 (0.005)	0.009 (0.006)	29.158	<.001	.014
Assent	0.001 (0.001)	0.001 (0.001)	15.080	<.001	.007
Perceptual processes	0.018 (0.008)	0.017 (0.007)	11.593	<.001	.006

Table 4 Confusion matrix for DFA classifying movie reviews: LIWC

		Positive Review	Negative Review	Total
Whole set	Positive review	696	304	1,000
	Negative review	317	683	1,000
Cross-validated	Positive review	695	305	1,000
	Negative review	319	681	1,000

50 %). For the LOOCV, the discriminant analysis correctly allocated 1,376 of the 2,000 texts, for an accuracy of 68.8 % (see the confusion matrix reported in Table 4 for the results). The measure of agreement between the actual text types and those assigned by the model produced a weighted Cohen's Kappa of .379, demonstrating moderate agreement.

SEANCE indices

MANOVA A MANOVA was conducted using the SEANCE indices as the dependent variables and the positive and negative reviews in the Movie Review Corpus as the independent variables. After checking normal distributions and multicollinearity, 295 indices remained. These indices were used in the MANOVA. The MANOVA indicated that 290 of these variables demonstrated significant differences between the positive and negative movie reviews. Among these, 206 of the variables had a p value that was below .001.

DFA We used the 206 significant variables from the MANOVA analysis as our predictor variables in the DFA. The stepwise DFA retained 46 variables (see Table 5 for the descriptive and MANOVA statistics for these variables), and we removed the remaining variables as nonsignificant predictors.

The results demonstrated that the DFA using these 46 indices correctly allocated 1,799 of the 2,000 texts in the total set, $\chi^2(df=1, n=2,000) = 980.004, p < .001$, for an accuracy of 85.0 %. For the LOOCV, the discriminant analysis correctly allocated 1,683 of the 2,000 texts, for an accuracy of 84.2 % (see the confusion matrix reported in Table 6 for the results). The measure of agreement between the actual text types and those assigned by the model produced a weighted Cohen's Kappa of .700, demonstrating substantial agreement.

SEANCE component scores

MANOVA A MANOVA was conducted using the SEANCE indices as the dependent variables and the positive and negative movie reviews in the Movie Review Corpus as the independent variables. All indices were normally distributed, and no multicollinearity ($r < .899$) was reported between any of the variables. Thus, the 20 component scores were all used in the MANOVA. This

analysis indicated that all 20 of the variables demonstrated significant differences ($p < .001$) between the positive and negative movie reviews.

DFA We used the 20 significant variables from the MANOVA as our predictor variables in the DFA. The stepwise DFA retained ten variables (see Table 7 for the descriptive and MANOVA statistics for these variables), and we removed the remaining variables as nonsignificant predictors.

The results demonstrated that the DFA using these ten component scores correctly allocated 1,495 of the 2,000 texts in the total set, $\chi^2(df=1, n=2,000) = 492.394, p < .001$, for an accuracy of 74.8 %. For the LOOCV, the discriminant analysis correctly allocated 1,488 of the 2,000 texts, for an accuracy of 74.4 % (see the confusion matrix reported in Table 8 for the results). The measure of agreement between the actual text types and those assigned by the model produced a weighted Cohen's Kappa of .495, demonstrating moderate agreement.

ANOVA comparison between models for the movie review corpus

We conducted a one-way ANOVA between the models, using the accuracy scores as the dependent variable and the models as the independent variable. The ANOVA reported an overall significant effect, $F(2, 5997) = 74.690, p < .001, \eta^2 = .025$. Pairwise comparisons showed significant differences between all models ($p < .001$), indicating that the SEANCE indices model was significantly better at classifying the movie reviews ($M = .850, SD = .357$) than either the SEANCE component score model ($M = .748, SD = .435$) or the LIWC indices model ($M = .690, SD = .463$). In addition, the pairwise comparisons indicated that the SEANCE component score model was significantly better at predicting the movie review classifications than was the LIWC indices model.

Multi-domain sentiment dataset

LIWC indices

MANOVA A MANOVA was conducted using the LIWC indices related to psychological processes and personal concerns as the dependent variables, and the positive and negative

Table 5 Descriptive and MANOVA statistics for positive and negative movie reviews: SEANCE indices

Index	Positive Review: <i>M</i> (<i>SD</i>)	Negative Review: <i>M</i> (<i>SD</i>)	<i>F</i>	<i>p</i>	η_p^2
Hu–Liu positive (percentage, adjectives, negation controlled)	0.611 (0.150)	0.443 (0.154)	616.063	<.001	.236
Lasswell negative affect (adjectives, negation controlled)	0.024 (0.025)	0.055 (0.048)	328.01	<.001	.141
Hu–Liu negative	0.032 (0.011)	0.041 (0.012)	316.935	<.001	.137
Hu–Liu negative (adjectives, negation controlled)	0.013 (0.006)	0.018 (0.008)	266.978	<.001	.118
GI vice (negation controlled)	0.019 (0.008)	0.025 (0.009)	259.105	<.001	.115
VADER positive	0.148 (0.042)	0.122 (0.034)	238.378	<.001	.107
GI weakness	0.037 (0.010)	0.043 (0.010)	176.384	<.001	.081
NRC sadness (adjectives negation controlled)	0.060 (0.042)	0.087 (0.052)	157.155	<.001	.073
NRC joy (negation controlled)	0.031 (0.011)	0.026 (0.010)	129.216	<.001	.061
Hu–Liu positive (percentage, adverbs)	0.704 (0.317)	0.556 (0.337)	102.419	<.001	.049
GI doctrine	0.006 (0.005)	0.004 (0.003)	99.279	<.001	.047
GI solve	0.038 (0.011)	0.042 (0.011)	58.686	<.001	.029
NRC fear (adjectives, negation controlled)	0.058 (0.040)	0.074 (0.052)	58.193	<.001	.028
GI understated	0.041 (0.012)	0.045 (0.012)	57.347	<.001	.028
SenticNet polarity (verbs)	0.082 (0.057)	0.062 (0.062)	56.244	<.001	.027
GI polite (adjectives, negation controlled)	0.065 (0.053)	0.050 (0.045)	45.660	<.001	.022
Hu–Liu positive (average, nouns)	0.008 (0.005)	0.007 (0.004)	45.099	<.001	.022
GI emotions (nouns)	0.020 (0.016)	0.016 (0.013)	42.818	<.001	.021
ANEW valence (adjectives)	5.885 (1.070)	5.574 (1.100)	41.155	<.001	.020
SenticNet valence (negation controlled)	6.112 (0.475)	5.976 (0.487)	39.803	<.001	.020
GI virtue (verbs)	0.047 (0.024)	0.040 (0.024)	38.277	<.001	.019
Lasswell rectitude (adverbs, negation controlled)	0.072 (0.057)	0.089 (0.073)	33.823	<.001	.017
GI tool	0.029 (0.019)	0.034 (0.020)	32.074	<.001	.016
GI male	0.023 (0.015)	0.019 (0.012)	31.724	<.001	.016
GI yes (adverbs, negation controlled)	0.041 (0.051)	0.030 (0.041)	31.578	<.001	.016
GI passive (adjectives, negation controlled)	0.061 (0.039)	0.071 (0.044)	31.508	<.001	.016
Lasswell well-being	0.026 (0.009)	0.023 (0.009)	28.986	<.001	.014
Lasswell well being physical (negation controlled)	0.007 (0.005)	0.006 (0.005)	24.921	<.001	.012
GI quality (adjectives)	0.066 (0.041)	0.058 (0.038)	22.909	<.001	.011
GI space (adverbs)	0.069 (0.051)	0.080 (0.054)	22.550	<.001	.011
SenticNet pleasantness (nouns, negation controlled)	0.074 (0.073)	0.059 (0.062)	22.159	<.001	.011
GI emotion (verbs)	0.020 (0.016)	0.017 (0.015)	21.681	<.001	.011
SenticNet attention (nouns, negation controlled)	0.133 (0.058)	0.121 (0.054)	21.090	<.001	.010
Lasswell power gain	0.005 (0.003)	0.004 (0.003)	20.013	<.001	.010
SenticNet polarity (nouns)	0.082 (0.061)	0.071 (0.051)	19.036	<.001	.009
GI natural processes	0.028 (0.019)	0.024 (0.017)	18.900	<.001	.009
Lasswell respect (nouns)	0.021 (0.016)	0.018 (0.014)	18.653	<.001	.009
SenticNet aptitude (nouns)	0.104 (0.067)	0.093 (0.059)	17.260	<.001	.009
Lasswell power conflict (nouns)	0.016 (0.016)	0.014 (0.014)	14.885	<.001	.007
Lasswell enlightenment (verbs)	0.138 (0.043)	0.131 (0.044)	14.330	<.001	.007
GI understated (adjectives)	0.070 (0.039)	0.077 (0.043)	13.937	<.001	.007
Lasswell respect	0.004 (0.004)	0.004 (0.003)	13.520	<.001	.007
NRC trust (verbs)	0.060 (0.029)	0.055 (0.027)	12.861	<.001	.006
GI overstated (verbs, negation controlled)	0.067 (0.031)	0.072 (0.033)	12.474	<.001	.006
SenticNet valence (adjectives, negation controlled)	0.780 (0.404)	0.722 (0.346)	12.027	<.001	.006
GI humans	0.045 (0.014)	0.043 (0.013)	11.631	<.001	.006

Amazon reviews as the independent variables. After checking for normal distributions and multicollinearity, 15 of the

indices remained. These indices were used in the MANOVA, which indicated that 12 of the variables

Table 6 Confusion matrix for DFA classifying movie reviews: SEANCE indices

		Positive Review	Negative Review	Total
Whole set	Positive review	849	151	1,000
	Negative review	149	851	1,000
Cross-validated	Positive review	843	157	1,000
	Negative review	160	840	1,000

demonstrated significant differences between the positive and negative Amazon reviews and reported p values below .001.

DFA We used the 12 significant variables from the MANOVA as our predictor variables in the DFA. The stepwise DFA retained nine of the variables (see Table 9 for the descriptive and MANOVA statistics for these variables), and we removed the remaining variables as nonsignificant predictors.

The results demonstrated that the DFA using these nine indices correctly allocated 5,748 of the 8,000 texts in the total set, $\chi^2(df = 1, n = 8,000) = 1,530.852, p < .001$, for an accuracy of 71.9 %. For the LOOCV, the discriminant analysis correctly allocated 5,742 of the 8,000 texts, for an accuracy of 71.8 % (see the confusion matrix reported in Table 10 for the results). The measure of agreement between the actual text type and that assigned by the model produced a weighted Cohen's Kappa of .437, demonstrating moderate agreement.

For the individual domains in the dataset, the DFA based on the LIWC features correctly allocated 1,424 of the 2,000 book reviews, $\chi^2(df = 1, n = 2,000) = 359.696, p < .001$, Kappa = .424, for an accuracy of 71.2 %; 1,473 of the 2,000 DVD reviews, $\chi^2(df = 1, n = 2,000) = 447.480, p < .001$, Kappa = .473, for an accuracy of 73.7 %; 1,395 of the 2,000 electronics reviews, $\chi^2(df = 1, n = 2,000) = 332.450, p < .001$, Kappa = .395, for an accuracy of 69.8 %; and 1,456 of the 2,000 kitchen appliance reviews, $\chi^2(df = 1, n = 2,000) =$

416.412, $p < .001$, Kappa = .456, for an accuracy of 72.8 %. See Table 11 for an overview.

SEANCE indices

MANOVA A MANOVA was conducted using the SEANCE indices as the dependent variables and the positive and negative Amazon reviews as the independent variables. After checking for normal distributions and multicollinearity, 109 indices remained. These indices were used in the MANOVA, which indicated that 80 of the variables demonstrated significant differences between the positive and negative Amazon reviews and had p values below .001.

DFA We used the 80 significant variables from the MANOVA as our predictor variables in the DFA. The stepwise DFA retained 37 of the variables (see Table 12 for the descriptive and MANOVA statistics for these variables), and we removed the remaining variables as nonsignificant predictors.

The results demonstrated that the DFA using these 37 indices correctly allocated 6,219 of the 8,000 texts in the total set, $\chi^2(df = 1, n = 8,000) = 2,468.464, p < .001$, for an accuracy of 77.7 %. For the LOOCV, the discriminant analysis correctly allocated 6,196 of the 8,000 texts for an accuracy of 77.5 % (see the confusion matrix reported in Table 13 for the results). The measure of agreement between the actual text

Table 7 Descriptive and MANOVA statistics for positive and negative movie reviews: SEANCE component scores

Index	Positive Review: M (SD)	Negative Review: M (SD)	F	p	η_p^2
Negative adjectives	-0.133 (0.619)	0.615 (0.750)	592.426	<.001	.229
Positive adjectives	0.118 (0.150)	0.027 (0.134)	206.543	<.001	.094
Polarity verbs	0.342 (0.187)	0.272 (0.203)	64.960	<.001	.031
Social order	0.367 (0.097)	0.400 (0.103)	52.740	<.001	.026
Economy	0.168 (0.111)	0.135 (0.097)	49.756	<.001	.024
Fear and disgust	0.194 (0.073)	0.214 (0.070)	38.743	<.001	.019
Politeness	0.115 (0.066)	0.100 (0.047)	35.840	<.001	.018
Action	0.553 (0.112)	0.579 (0.113)	26.883	<.001	.013
Certainty	0.158 (0.034)	0.166 (0.033)	26.383	<.001	.013
Well being	0.095 (0.046)	0.087 (0.044)	15.675	<.001	.008

Table 8 Confusion matrix for DFA classifying movie reviews: SEANCE component scores

		Positive Review	Negative Review	Total
Whole set	Positive review	782	218	1,000
	Negative review	287	713	1,000
Cross-validated	Positive review	777	223	1,000
	Negative review	289	711	1,000

types and those assigned by the model produced a weighted Cohen's Kappa of .555, demonstrating moderate agreement.

For the individual domains in the dataset, the DFA based on the SEANCE indices correctly allocated 1,500 of the 2,000 book reviews, $\chi^2(df = 1, n = 2,000) = 501.929, p < .001$, Kappa = .500, for an accuracy of 75 %; 1,547 of the 2,000 DVD reviews, $\chi^2(df = 1, n = 2,000) = 598.634, p < .001$, Kappa = .547, for an accuracy of 77.4 %; 1,583 of the 2,000 electronics reviews, $\chi^2(df = 1, n = 2,000) = 679.860, p < .001$, Kappa = .583, for an accuracy of 79.2 %; and 1,605 of the 2,000 kitchen appliance reviews, $\chi^2(df = 1, n = 2,000) = 737.128, p < .001$, Kappa = .605, for an accuracy of 80.3 %. See Table 11 for an overview.

SEANCE component scores

MANOVA A MANOVA was conducted using the SEANCE component scores as the dependent variables and the positive and negative Amazon reviews as the independent variables. All indices were normally distributed, and no multicollinearity was reported between any of the variables. Thus, all 20 component scores were used in the MANOVA, which indicated that 16 of the variables demonstrated significant differences between the positive and negative Amazon reviews and reported p values below .001.

DFA We used the 16 significant variables from the MANOVA as our predictor variables in the DFA. The stepwise DFA retained 11 of the variables (see Table 14

for descriptive and MANOVA statistics for these variables), and we removed the remaining variables as non-significant predictors.

The results demonstrated that the DFA using these 11 component scores correctly allocated 5,963 of the 8,000 texts in the total set, $\chi^2(df = 1, n = 8,000) = 1,936.240, p < .001$, for an accuracy of 74.5 %. For the LOOCV, the discriminant analysis correctly allocated 5,960 of the 8,000 texts, for an accuracy of 74.5 % (see the confusion matrix reported in Table 15 for the results). The measure of agreement between the actual text types and those assigned by the model produced a weighted Cohen's Kappa of .491, demonstrating moderate agreement.

For the individual domains in the dataset, the DFA based on the SEANCE component scores correctly allocated 1,449 of the 2,000 book reviews, $\chi^2(df = 1, n = 2,000) = 412.867, p < .001$, Kappa = .449, for an accuracy of 72.5 %; 1,519 of the 2,000 DVD reviews, $\chi^2(df = 1, n = 2,000) = 538.735, p < .001$, Kappa = .519, for an accuracy of 76 %; 1,477 of the 2,000 electronics reviews, $\chi^2(df = 1, n = 2,000) = 455.751, p < .001$, Kappa = .477, for an accuracy of 73.9 %; and 1,518 of the 2,000 kitchen appliance reviews, $\chi^2(df = 1, n = 2,000) = 541.432, p < .001$, Kappa = .518, for an accuracy of 75.9 %. See Table 11 for an overview.

ANOVA comparison between models for the Amazon review corpus

We conducted one-way ANOVAs between the classification accuracies reported for the entire model and by the

Table 9 Descriptive and MANOVA statistics for positive and negative Amazon reviews: LIWC indices

Index	Positive Review: M (SD)	Negative Review: M (SD)	F	p	η_p^2
Positive emotion	0.060 (0.041)	0.036 (0.027)	991.205	<.001	.110
Negations	0.015 (0.017)	0.026 (0.022)	627.403	<.001	.073
Negative emotion	0.012 (0.016)	0.023 (0.024)	608.401	<.001	.071
Exclusion	0.024 (0.023)	0.034 (0.025)	302.267	<.001	.036
Affect	0.072 (0.041)	0.059 (0.035)	240.435	<.001	.029
Inclusive	0.051 (0.030)	0.044 (0.027)	134.379	<.001	.017
Tentativeness	0.021 (0.021)	0.025 (0.022)	65.392	<.001	.008
Social	0.062 (0.047)	0.056 (0.042)	29.464	<.001	.004
Certainty	0.016 (0.018)	0.014 (0.016)	22.578	<.001	.003

Table 10 Confusion matrix for DFA classifying Amazon reviews: LIWC

		Positive Review	Negative Review	Total
Whole set	Positive review	2,964	1,036	4,000
	Negative review	1,216	2,784	4,000
Cross-validated	Positive review	2,960	1,040	4,000
	Negative review	1,218	2,782	4,000

Table 11 Domain accuracy by feature set

Domain	LIWC	SEANCE Indices	Component Scores
Book reviews	.712	.750	.725
DVD reviews	.737	.774	.760
Electronic reviews	.698	.792	.739
Kitchen appliance reviews	.728	.803	.759

Table 12 Descriptive and MANOVA statistics for Amazon reviews: SEANCE indices

Index	Positive Review: <i>M</i> (<i>SD</i>)	Negative Review: <i>M</i> (<i>SD</i>)	<i>F</i>	<i>p</i>	η_p^2
Hu-Liu positive (percentage, negation controlled)	0.754 (0.245)	0.45 (0.284)	2,626.78	<.001	.247
VADER Positive	0.184 (0.101)	0.099 (0.066)	1,944.881	<.001	.196
Hu-Liu positive (average, negation controlled)	0.065 (0.044)	0.032 (0.025)	1,755.93	<.001	.180
VADER compound adjectives	0.594 (0.405)	0.145 (0.558)	1,692.333	<.001	.175
Hu-Liu positive (percentage, adjectives only)	0.761 (0.315)	0.460 (0.358)	1,591.082	<.001	.166
VADER positive adjectives	0.529 (0.279)	0.311 (0.259)	1,305.756	<.001	.140
SenticNet pleasantness (negation controlled)	0.15 (0.146)	0.046 (0.123)	1,212.68	<.001	.132
SenticNet pleasantness (adjectives, negation controlled)	0.256 (0.314)	0.063 (0.297)	801.264	<.001	.091
Hu-Liu proportion positive/negative (negation controlled)	2.145 (2.186)	1.075 (1.207)	733.563	<.001	.084
SenticNet polarity (adjectives)	0.26 (0.263)	0.111 (0.257)	655.855	<.001	.076
EmoLex Positive (negation controlled)	0.073 (0.043)	0.052 (0.033)	591.971	<.001	.069
SenticNet pleasantness (verbs negation controlled)	0.106 (0.228)	0.008 (0.175)	466.286	<.001	.055
SenticNet aptitude (verbs, negation controlled)	0.233 (0.207)	0.149 (0.182)	375.302	<.001	.045
SenticNet polarity (verbs)	0.146 (0.177)	0.083 (0.143)	307.745	<.001	.037
VADER neutral	0.774 (0.098)	0.807 (0.086)	264.036	<.001	.032
SenticNet attention (negation controlled)	0.202 (0.116)	0.162 (0.100)	261.415	<.001	.032
VADER neutral (adverbs)	0.772 (0.303)	0.858 (0.227)	209.216	<.001	.025
GI weak (negation controlled)	0.040 (0.030)	0.049 (0.034)	158.12	<.001	.019
GI action verbs	0.288 (0.158)	0.331 (0.159)	145.778	<.001	.018
GI strength	0.112 (0.046)	0.102 (0.040)	115.587	<.001	.014
GI hostile (negation controlled)	0.033 (0.027)	0.039 (0.028)	86.544	<.001	.011
SenticNet attention (verbs)	0.186 (0.173)	0.159 (0.135)	63.376	<.001	.008
GI understatement	0.040 (0.029)	0.045 (0.028)	63.097	<.001	.008
GI quantity	0.085 (0.042)	0.079 (0.037)	52.326	<.001	.006
GI overstatement (adjectives)	0.294 (0.227)	0.258 (0.212)	52.293	<.001	.006
GI passive (negation controlled)	0.069 (0.038)	0.075 (0.038)	44.781	<.001	.006
Lasswell rectitude (negation controlled)	0.027 (0.024)	0.030 (0.024)	38.741	<.001	.005
GI economy	0.031 (0.030)	0.035 (0.032)	35.842	<.001	.004
GI overstatement (negation controlled)	0.084 (0.043)	0.078 (0.040)	35.206	<.001	.004
GI time (negation controlled)	0.050 (0.035)	0.055 (0.038)	28.056	<.001	.003
GI social relations	0.049 (0.030)	0.052 (0.030)	23.277	<.001	.003
GI affiliation (negation controlled)	0.041 (0.031)	0.038 (0.029)	22.626	<.001	.003
GI communication (nouns)	0.138 (0.133)	0.150 (0.129)	16.683	<.001	.002
GI interpretation of action (nouns, negation controlled)	0.175 (0.129)	0.186 (0.131)	15.102	<.001	.002
GI work (verbs)	0.160 (0.136)	0.172 (0.134)	14.900	<.001	.002
GI power (verbs)	0.178 (0.134)	0.168 (0.113)	12.559	<.001	.002
SenticNet sensitivity (negation controlled)	-0.057 (0.049)	-0.061 (0.079)	5.129	<.001	.001

Table 13 Confusion matrix for DFA classifying Amazon reviews: SEANCE

		Positive Review	Negative Review	Total
Whole set	Positive review	3,007	993	4,000
	Negative review	788	3,212	4,000
Cross-validated	Positive review	2,995	1,005	4,000
	Negative review	799	3,201	4,000

model for each domain (books, DVDs, electronics, and kitchen appliances), using the accuracy scores as the dependent variable and the models as the independent variable.

The ANOVA reported an overall significant effect for the feature set used, $F(2, 23997) = 35.598, p < .001, \eta^2 = .003$. Pairwise comparisons showed significant differences between all models ($p < .001$), indicating that the SEANCE indices model ($M = .779, SD = .415$) was significantly better at classifying the Amazon reviews than either the SEANCE component score model ($M = .745, SD = .436$) or the LIWC indices model ($M = .719, SD = .449$). In addition, the pairwise comparisons indicated that the SEANCE component score model was significantly better at predicting the Amazon review classifications than was the LIWC indices model.

The ANOVAs for the model's domain-specific predictability reported significant effects for the domains of books, $F(2, 5997) = 3.799, p < .050, \eta^2 = .001$; DVDs, $F(2, 5997) = 3.792, p < .050, \eta^2 = .001$; electronics, $F(2, 5997) = 23.403, p < .001, \eta^2 = .008$; and kitchen appliances, $F(2, 5997) = 15.571, p < .001, \eta^2 = .005$. Pairwise comparisons showed a number of significant differences between the models ($p < .001$; see Table 16), indicating that the SEANCE indices model was significantly better at classifying the Amazon reviews than were the SEANCE component score model (except with DVD reviews) and the LIWC indices model. In addition, the pairwise comparisons indicated that the SEANCE component score model was significantly better at predicting the Amazon review classifications than was the LIWC indices model, with the exception of book and DVD reviews.

Table 14 Descriptive and MANOVA statistics for Amazon reviews: SEANCE component scores

Index	Positive Review: $M (SD)$	Negative Review: $M (SD)$	F	p	η_p^2
Negative adjectives	-0.709 (0.812)	0.534 (1.305)	2,615.76	<.001	.246
Positive adjectives	0.355 (0.474)	0.047 (0.357)	1,081.875	<.001	.119
Polarity verbs	0.243 (0.569)	-0.050 (0.596)	504.229	<.001	.059
Trust	0.228 (0.214)	0.163 (0.153)	239.083	<.001	.029
Positive nouns	-0.018 (0.565)	-0.207 (0.554)	226.572	<.001	.028
Polarity noun	0.502 (0.442)	0.371 (0.414)	186.300	<.001	.023
Affect: Friends and family	0.259 (0.182)	0.216 (0.134)	146.409	<.001	.018
Action	0.567 (0.281)	0.617 (0.279)	63.558	<.001	.008
Failure	0.046 (0.086)	0.060 (0.093)	50.759	<.001	.006
Well being	0.086 (0.117)	0.071 (0.098)	39.850	<.001	.005
Politeness	0.095 (0.103)	0.087 (0.100)	12.500	<.001	.002

Table 15 Confusion matrix for DFA classifying Amazon reviews: SEANCE component scores

		Positive Review	Negative Review	Total
Whole set	Positive review	2,841	1,159	4,000
	Negative review	878	3,122	4,000
Cross-validated	Positive review	2,839	1,161	4,000
	Negative review	879	3,121	4,000

Table 16 Pairwise comparison results for domain classifications

Domain	Differences
Book reviews	SEANCE indices > SEANCE component scores, LIWC
DVD reviews	SEANCE indices > LIWC; SEANCE component scores = LIWC, SEANCE indices
Electronics reviews	SEANCE indices > SEANCE component scores > LIWC
Kitchen appliances reviews	SEANCE indices > SEANCE component scores > LIWC

Discussion

This article introduces a new tool, SEANCE, which automatically analyzes text features related to sentiment, cognition, and social order, among numerous other features. The tool is domain-general, and the output allows users to develop theoretical inferences from their datasets. The findings from this study have helped provide predictive validity for SEANCE, by demonstrating the potential for the SEANCE indices to predict positive and negative reviews with a number of well-known and widely used sentiment analysis test corpora. In addition, the study provides evidence for lexical differences between positive and negative texts, providing insight into the linguistic underpinnings that are predictive of writers' emotional states. Our hope is that this freely available, user-friendly tool will provide wider access to a greater depth and breadth of lexical-based indices related to sentiment, cognition, and social order for researchers into discourse processing, language assessment, education, and cognitive science. The indices reported by SEANCE could be used to study the sentiment in a number of different discourse domains beyond those tested here (e.g., Web-based media, educational discourse, language assessment, and product reviews). In essence, researchers in any number of fields with an interest in examining sentiment in discourse structures could use SEANCE as a research tool.

We tested SEANCE against the most common tool used in sentiment analysis for behavioral studies (LIWC) and found that both the individual indices (i.e., the microfeatures) and the component scores (i.e., the macrofeatures) statistically outperformed LIWC in classic sentiment analysis tasks. For the Movie Review Corpus, SEANCE's microfeatures performed on a par with domain-dependent tools (cf. .847 for SEANCE vs. .803–.872 for previous classification accuracies; Taboada et al., 2009), and better than previous domain-independent classification accuracies, which had reported accuracies that ranged from .581 to .764 for the same corpus. The component scores proved less accurate than the microfeatures in the Movie Review Corpus, but still reported accuracies of around 75 %, putting the component scores on a par with the top end of previous classification accuracies for domain-independent tools. For the Multi-Domain Sentiment Dataset (i.e., Amazon reviews), the SEANCE indices performed slightly below previously reported domain-specific

algorithms based on *n*-grams (cf. 80 % from *n*-grams vs. 75 % from SEANCE for books, 82 % vs. 77 % for DVDs, 84 % vs. 79 % for electronics, and 88 % vs. 80 % for kitchen appliances) and on a par with or slightly better than (cf. 70.7 %–72.8 % vs. 75 % for books, 70.6 %–77.2 % vs. 77 % for DVDs, 70.8 %–82.7 % vs. 79 % for electronics, and 74.5 %–84 % vs. 80 % for kitchen appliances) previously reported domain adaption algorithms (i.e., algorithms that adapted predictive *n*-grams on the basis of frequency and mutual information from similar domains, such as the book and DVD review domains). The SEANCE component scores reported lower classification accuracies, but they were still on a par with previous domain adaption algorithms.

Overall, the findings of this study support the notion that negative and positive reviews in a variety of domains can be classified on the basis of a number of lexical features related to sentiment, cognition, and social order. Thus, the findings support the notion that emotional texts can be classified on the basis of the types of words selected by their authors. The LIWC analysis indicated that negative movie reviews contained more negative emotion words, negations, discrepancy terms, anger words, and exclusion terms, whereas positive movie reviews had more positive emotion words, inclusion terms, and terms related to perception processes. The LIWC findings from the Amazon.com reviews were similar, in that positive reviews also contained more positive emotion words along with fewer negations, negative emotion words, and exclusion terms. The DFA also indicated that positive Amazon reviews contained more affective, social, and certainty words, along with fewer exclusion terms.

The SEANCE microfeature analysis of the movie reviews supported the LIWC findings, but indicated that terms used as adjectives were the most predictive features of positive and negative texts and that reversing polarity based on negation was also an important component of predicting positive and negative opinions and emotions. This was especially true for the Hu–Lui positive and negative emotional categories, which were the strongest predictors in both the DFA models. Furthermore, the results indicated that other POS tags, such as adverbs, verbs, and nouns, are also important discriminators for sentiment analysis tasks. In addition, the SEANCE microfeature analysis reported that a number of features indirectly related to sentiment were important predictors of positive and negative movie reviews. These features indicated that

positive movie reviews contained more words related to organized systems (doctrines), more dominance words, more polite terms, more male and human terms, more words related to power gain and solving, more terms related to natural processes, more enlightenment words, more words related to well-being, and more respect terms. Negative movie reviews, on the other hand, contained terms related to understatement, more tool terms, more spatial terms, and more terms of understanding. These findings were generally upheld in the analysis of the Amazon reviews, in that negation and POS tags were important components of predicting positive and negative reviews, and that a number of nonsentiment features related to strength of propositions, economy, social relations, time, communication, quantity, overstatements, communication, power, and action were predictive of both positive and negative Amazon reviews.

Our macrofeature analysis demonstrated the power of combining like indices into specific components. Of the 20 components we developed, each demonstrated significant differences between the positive and negative movie reviews. In addition, ten of the 20 component scores were significant predictors of positive movie reviews in a DFA analysis. Like the microfeature analysis, these components indicated that adjectives were the strongest predictors of sentiment, and that components directly related to positive and negative sentiment were the strongest predictors of movie review type. In addition, components that were indirectly related to sentiment were strong predictors. For instance, negative movie reviews reported higher social-order, action, and certainty component scores than did positive movie reviews. Conversely, positive movie reviews reported higher economy, politeness, and well-being component scores than did negative movie reviews. Similar findings were reported for the component scores in the Amazon review analysis. This analysis indicated that 16 of the component scores showed significant differences between positive and negative Amazon reviews, and 11 of these component scores were predictors in the DFA. These component scores indicated that positive reviews contained fewer negative adjectives, more positive adjectives and verbs, more well-being terms, and were more polite. In addition, positive reviewers had fewer action terms. Three components that differed from the movie review analysis were also included for the Amazon reviews, indicating that positive reviews contained more trust terms, more affective terms related to friends and family, and fewer failure terms.

Combined, these findings provide support for the notion that word vectors related to positive and negative emotions are the strongest predictors of review types in both the Movie Review Corpus and the Amazon corpus, indicating that writers rely on emotion words to convey affect. In addition, the findings indicate that valence features such as negation (Hutto & Gilbert, 2014; Polanyi & Zaenen, 2006), along with POS tags (Hogenboom et al., 2012), are important

components of sentiment analysis and should be included in sentiment tasks. The findings indicate that writers may localize emotions more often in adjectives followed by verbs and adverbs. Although writers do use nouns to convey emotional content, nouns that contain emotions are less predictive than adjectives, verbs, and adverbs. Interestingly, a number of lexical features indirectly related to sentiment analysis were shown to be significant predictors of positive and negative reviews in both the test corpora, indicating that writers may not rely solely on emotional terms. These indirect assessments of sentiment were weaker predictors than traditional sentiment indices related to positive and negative terms, but their inclusion indicates the possibility to better understand sentiment features. A few of the features reported in this study are likely domain-specific. For instance, political movies appear more likely to be reviewed positively than movies centering on the legal system (as can be seen in the word vectors related to political and legal terms). In a similar fashion, reviews that discuss defined roles and men are more likely to be positive, and Amazon reviews that contain words related to the economy (most likely those related to price) are more likely to be negative. However, a number of word vectors in SEANCE that indirectly assess sentiment and are domain-independent were strong predictors of review types in this analysis. For instance, vectors related to dominance, respect, and power (e.g., power conflict, power gain, strength, and weakness), evaluation (e.g., overstatement, understatement, and virtue), quality and quantity (e.g., increase and quality), action, and temporality and spatiality (e.g., space and time) all reported significant differences between positive and negative reviews in both the movie review and Amazon corpora. Thus, a number of domain-general lexical features that are not specifically emotional are used by writers when producing positive and negative texts. Overall, the findings appear to allow for theoretical inferences about how language features related to sentiment, cognition, and social order are predictive of affect, and thus provide a better understanding of how writers use specific types of words to convey affect to readers.

Conclusion

This study introduces and demonstrates the use of a new sentiment analysis tool, SEANCE, which is freely available to researchers and provides an automated approach to the examination of discourse in terms of sentiment, cognition, and social order. The two evaluations presented in this study afford strong evidence for the utility of SEANCE. Nonetheless, we plan to extend this foundational research by conducting additional validation studies using other metrics of sentiment beyond movie reviews (i.e., product reviews, blogs, tweets, and human discourse). These studies will provide evidence for the effectiveness of the SEANCE indices and component scores

in other domains, to ensure domain independence. In addition, future studies will focus on developing new indices of sentiment for inclusion in SEANCE and additional valence features. Such indices will be based on sentiment dictionaries that are currently available, those that become available over time (e.g., the Warriner norms; Warriner, Kuperman, & Brysbaert, 2013; Westbury, Keith, Briesemeister, Hofmann, & Jacobs, 2015), or previous dictionaries that are updated. We also plan to add valence features, as necessary, to examine discourse features such as intensification. In the future, we will use these new indices and features to examine and test SEANCE on domains beyond movie reviews.

We presume that SEANCE will facilitate research on sentiment analysis in discourse studies, language assessment, business management, education, and cognitive science (among other disciplines). We foresee SEANCE being used to examine the effects of negative and positive text on readers; to investigate affective educational states such as engagement, motivation, and arousal; to assess the effects of emotions in language assessment; and to control stimuli in behavioral studies. Outside of academic research, SEANCE could also aid businesses and industry in assessing product responses, advertising, and consumer sentiment. The study also provides evidence supporting the notion that valence and POS tags are important elements of sentiment analysis, and that word vectors indirectly associated with sentiment can provide valuable information about positive and negative language.

Author note This research was supported in part by the Institute for Education Sciences (IES) and the National Science Foundation (NSF) (Grant Nos. IES R305A080589, IES R305G20018-02, and DRL-1418378). The ideas expressed in this material are those of the authors and do not necessarily reflect the views of the IES or NSF.

References

- Andreevskaia, A., & Bergler, S. (2008). When specialists and generalists work together: Overcoming domain dependence in sentiment tagging. In *Proceedings of 46th Annual Meeting of the Association for Computational Linguistics: Human language technologies* (pp. 290–298). Stroudsburg, PA: Association for Computational Linguistics.
- Aue, A., & Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. In *Proceedings of Recent Advances in Natural Language Processing*. Retrieved from http://research.microsoft.com/pubs/65430/new_domain_sentiment.pdf
- Baron, D. P. (2005). Competing for the public through the news media. *Journal of Economics and Management Strategy*, 14, 339–376.
- Bartlett, J., & Albright, R. (2008). *Coming to a theater near you! Sentiment classification techniques using SAS Text Miner*. Paper presented at the SAS Global Forum 2008, San Antonio, TX.
- Batson, C. D., Shaw, L. L., & Oleson, K. C. (1992). *Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions*. Thousand Oaks: Sage.
- Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., & Subrahmanian, V. (2007). Sentiment analysis: Adjectives and adverbs are better than the adjectives alone. In *Proceedings of International Conference on Weblogs and Social Media (ICWSM'2007)*.
- Bernard, J. (Ed.). (1986). *The Macquarie thesaurus*. Sydney: Macquarie Library.
- Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Wortman, J. (2007). Learning bounds for domain adaptation. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), *Advances in neural information processing systems 20* (pp. 1–8). Cambridge: MIT Press.
- Blitzer, J., Dredze, M., & Pereira, F. (2007). *Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification*. Paper presented at the conference of the Association of Computational Linguistics, Prague, Czech Republic.
- Boiy, E., Hens, P., Deschacht, K., & Moens, M. F. (2007). Automatic sentiment analysis in on-line text. In D. Chan & B. Martens (Eds.), *Proceedings of the 11th International Conference on Electronic Publishing* (pp. 349–360). Göttingen: ELPUB.
- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings (Technical Report No. C-1)*. Gainesville: University of Florida, NIMH Center for Research in Psychophysiology.
- Brants, T., & Franz, A. (2006). *Web It 5-gram version 1*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania. Retrieved from <https://catalog.ldc.upenn.edu/LDC2006T13>
- Brooke, J. (2009). *A semantic approach to automated text sentiment analysis (Doctoral dissertation)*. Burnaby: Simon Fraser University.
- Brooke, J., & Hurst, M. (2009). Patterns in the stream: Exploring the interaction of polarity, topic, and discourse in a large opinion corpus. In *Proceedings of the ACM conference on information and knowledge management, 1st international workshop on topic-sentiment analysis for mass opinion measurement*
- Cambria, E., Havasi, C., & Hussain, A. (2012). SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis. In G. M. Youngblood & P. M. McCarthy (Eds.), *Proceedings of the 25th Florida artificial intelligence research society conference* (pp. 202–207). Palo Alto: AAAI Press.
- Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). SenticNet: A publicly available semantic resource for opinion mining. In C. Havasi, D. Lenat, & B. Van Durme (Eds.), *Commonsense knowledge: Papers from the AAAI fall symposium* (pp. 14–18). Palo Alto: AAAI Press.
- Chaovalit, P., & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. In J. F. Nunamaker Jr. & R. O. Briggs (Eds.), *Proceedings of the 38th annual Hawaii international conference on system sciences* (p. 112c). Piscataway: IEEE Press.
- Collins, A., Ortony, A., & Clore, G. L. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). To aggregate or not? Linguistic features in automatic essay scoring and feedback systems. *Journal of Writing Assessment*, 8, 80. Retrieved from www.journalofwritingassessment.org/article.php?article=80
- D'Mello, S. K., & Graesser, A. (2012). Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, 5, 304–317.
- De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the Seventh International Conference on Weblogs and Social Media* (pp. 128–137) Palo Alto, CA: AAAI Press.
- Dredze, M., Crammer, K., & Pereira, F. (2008). Confidence-weighted linear classification. Article presented at the International Conference on Machine Learning, Helsinki, Finland.
- Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation* (pp. 417–422). Paris, France: European Language Resources Association.

- Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. Cambridge: MIT Press.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th ed.). London: Sage.
- Finn, A., & Kushmerick, N. (2003). Active learning selection strategies for information extraction. In F. Ciravenga & N. Kushmerick (Eds.), *Proceedings of the International Workshop on Adaptive Text Extraction and Mining* (pp. 18–25). Retrieved from staff www.dcs.shef.ac.uk/people/F.Ciravegna/ATEM03/accepted.html
- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23, 1498–1512.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40, 223–234.
- Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics* (pp. 174–181). Stroudsburg, PA: Association for Computational Linguistics.
- Heerschoop, B., van Iterson, P., Hogenboom, A., Frasincaer, F., & Kaymak, U. (2011). Analyzing sentiment in a large set of web data whereas accounting for negation. In *Advances in intelligent web mastering-3* (pp. 195–205). Berlin: Springer.
- Hogenboom, A., Boon, F., & Frasincaer, F. (2012). A statistical approach to star rating classification of sentiment. In J. Casillas, F. J. Martínez-López, & J. M. Corchado (Eds.), *Management intelligent systems: First international symposium* (pp. 251–260). Berlin: Springer.
- Hogenboom, A., Hogenboom, F., Kaymak, U., Wouters, P., & De Jong, F. (2010). Mining economic sentiment using argumentation structures. In *Advances in conceptual modeling—Applications and challenges* (pp. 200–209). Berlin, Germany: Springer.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In W. Kim & R. Kohavi (Eds.), *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177). Washington, DC: ACM Press.
- Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In E. Adar & P. Resnick (Eds.), *Proceedings of the eighth international AAAI conference on weblogs and social media* (pp. 216–225). Palo Alto: AAAI Press.
- Jarvis, S. (2011). Data mining with learner corpora: Choosing classifiers for L1 detection. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 127–154). Amsterdam: John Benjamins.
- Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22, 110–125.
- Ketai, R. (1975). Affect, mood, emotion, and feeling: Semantic considerations. *American Journal of Psychiatry*, 132, 1215–1217.
- Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Norwood: Ablex.
- Langacker, R. W. (1985). Observations and speculations on subjectivity. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 109–150). Amsterdam: Benjamins.
- Lasswell, H. D., & Namenwirth, J. Z. (1969). *The Lasswell value dictionary*. New Haven: Yale University Press.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5, 1–167.
- Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In A. Ellis & T. Hagino (Eds.), *Proceedings of the 14th international conference on World Wide Web* (pp. 342–351). Stroudsburg: Association for Computational Linguistics.
- Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic Perspectives*, 18, 29–50.
- Lyons, J. (1981). *Language and linguistics*. Cambridge: Cambridge University Press.
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System demonstrations* (pp. 55–60). Stroudsburg, PA: Association for Computational Linguistics.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90, 862–877.
- Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53, 730–741.
- Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases: Using mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26–34). Stroudsburg: Association for Computational Linguistics.
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29, 436–465.
- Mohammad, S. M., & Yang, T. W. (2011). Tracking sentiment in mail: How genders differ on emotional axes. In A. Balahur, E. Boldrini, A. Montoyo, & P. Martínez (Eds.), *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis* (pp. 70–79). Stroudsburg: Association for Computational Linguistics.
- Murray, G., & Carenini, G. (2009). Predicting Subjectivity in Multimodal Conversations. *Proceedings of the 2009 conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1348–1357).
- Namenwirth, J., & Weber, R. (1987). *Dynamics of culture*. Boston: Allen & Unwin.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29, 665–675.
- Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana: University of Illinois.
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In D. Scott (Ed.), *Proceedings of the 42nd annual meeting on association for computational linguistics*. Stroudsburg: Association for Computational Linguistics. doi:10.3115/1218955.1218989
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2, 1–135.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing* (pp. 79–86). Stroudsburg: Association for Computational Linguistics.
- Pennebaker, J. W., Francis, M., & Booth, R. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahwah, NJ: Erlbaum.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC 2007. Retrieved from www.liwc.net/LIWC2007LanguageManual.pdf
- Plutchik, R. (2001). The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89, 344–350.

- Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In J. G. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing attitude and affect in text: Theory and applications* (pp. 1–10). Dordrecht: Springer.
- Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop* (pp. 43–48). Stroudsburg.
- Salveti, F., Reichenbach, C., & Lewis, S. (2006). Opinion polarity identification of movie reviews. In J. G. Shanahan, Y. Qu, & J. Wiebe (Eds.), *Computing attitude and affect in text: Theory and applications* (pp. 303–316). Dordrecht: Springer.
- Sauter, V. (2011). *Decision support systems for business intelligence*. Hoboken: Wiley.
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44, 695–729.
- Schumaker, R. P., Zhang, Y., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53, 458–464.
- Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: The links between language, performance, and workload. *Human Performance in Extreme Environments*, 5, 63–68.
- Sokolova, M., & Lapalme, G. (2009). Classification of opinions with non-affective adverbs and adjectives. In *Proceedings of the 7th international conference on recent advances in natural language processing* (pp. 416–420).
- Stone, P., Dunphy, D. C., Smith, M. S., Ogilvie, D. M., & Associates. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge: MIT Press.
- Stoyanov, V., Cardie, C., Litman, D., & Wiebe, J. (2006). Evaluating an opinion annotation scheme using a new multi-perspective question and answer corpus. In *Computing attitude and affect in text: Theory and applications* (pp. 77–91). Dordrecht, The Netherlands: Springer.
- Strapparava, C., & Valitutti, A. (2004). WordNet affect: An affective extension of WordNet. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the fourth international conference on language resources and evaluation* (pp. 1083–1086). Stroudsburg: Association for Computational Linguistics.
- Subrahmanian, V. S., & Reforgiato, D. (2008). AVA: Adjective-verb-adverb combinations for sentiment analysis. *Intelligent Systems*, 23(4), 43–50.
- Taboada, M., Anthony, C., & Voll, K. (2006). Methods for creating semantic orientation dictionaries. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (pp. 427–432). Stroudsburg, PA: Association for Computational Linguistics.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37, 267–307.
- Taboada, M., Brooke, J., & Stede, M. (2009). Genre-based paragraph classification for sentiment analysis. In P. Healey, R. Pieraccini, D. Byron, S. Young, & M. Purver (Eds.), *Proceedings of the SIGDIAL 2009 conference* (pp. 62–70). Stroudsburg: Association for Computational Linguistics.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24–54. doi:10.1177/0261927X09351676
- Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 173–180). Stroudsburg, PA: Association for Computational Linguistics.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM)* (vol. 10, pp. 178–185).
- Turney, P. D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424). Stroudsburg, PA: Association for Computational Linguistics.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191–1207. doi:10.3758/s13428-012-0314-x
- Westbury, C., Keith, J., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2015). Avoid violence, rioting and outrage; Approach celebration, delight, and strength: Using large text corpora to compute valence, arousal, and the basic emotions. *Quarterly Journal of Experimental Psychology*, 68, 1599–1622.
- Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55, 919–926.
- Yu, H., & Hatzivassiloglou, V. (2003). Toward answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In M. Collins & M. Steedman (Eds.), *Proceedings of the 2003 conference on empirical methods in natural language processing* (pp. 129–136). Stroudsburg: Association for Computational Linguistics.
- Yu, X., Liu, Y., Huang, X., & An, A. (2012). Mining online reviews for predicting sales performance: A case study in the movie domain. *IEEE Transactions on Knowledge and Data Engineering*, 24, 720–734.
- Zhang, H., Gan, W., & Jiang, B. (2014). Machine learning and lexicon based methods for sentiment classification: A survey. In X. Yuan & X. Meng (Eds.), *Proceedings of the 11th Web information system and application conference* (pp. 262–265). Piscataway: IEEE Press.