CHAPTER 21
Lexicons for Sentiment, Affect, and Connotation

“[W]e write, not with the fingers, but with the whole person. The nerve which controls the pen winds itself about every fibre of our being, threads the heart, pierces the liver.”

Virginia Woolf, Orlando

“She runs the gamut of emotions from A to B.”

Dorothy Parker, reviewing Hepburn’s performance in Little Women

“Festering’s always bad. There’s no good kind of festering.”

Adventure Time, Season 5

In this chapter we turn to tools for interpreting affective meaning, extending our study of sentiment analysis in Chapter 4. We use the word ‘affective’, following the tradition in affective computing (Picard, 1995) to mean emotion, sentiment, personality, mood, and attitudes. Affective meaning is closely related to subjectivity, the study of a speaker or writer’s evaluations, opinions, emotions, and speculations (Wiebe et al., 1999).

How should affective meaning be defined? One influential typology of affective states comes from Scherer (2000), who defines each class of affective states by factors like its cognitive realization and time course:

We can design extractors for each of these kinds of affective states. Chapter 4 already introduced sentiment analysis, the task of extracting the positive or negative orientation that a writer expresses in a text. This corresponds in Scheré’s typology to the extraction of attitudes: figuring out what people like or dislike, from affect-rich texts like consumer reviews of books or movies, newspaper editorials, or public sentiment in blogs or tweets.

Detecting emotion and moods is useful for detecting whether a student is confused, engaged, or certain when interacting with a tutorial system, whether a caller to a help line is frustrated, whether someone’s blog posts or tweets indicated depression. Detecting emotions like fear in novels, for example, could help us trace what groups or situations are feared and how that changes over time.

Detecting different interpersonal stances can be useful when extracting information from human-human conversations. The goal here is to detect stances like friendliness or awkwardness in interviews or friendly conversations, for example for summarizing meetings or finding parts of a conversation where people are especially excited or engaged, conversational hot spots that can help in meeting summarization. Detecting the personality of a user—such as whether the user is an extrovert or the extent to which they are open to experience—can help improve conversa-
CHAPTER 21 • LEXICONS FOR SENTIMENT, AFFECT, AND CONNOTATION

| Emotion: | Relatively brief episode of response to the evaluation of an external or internal event as being of major significance. (angry, sad, joyful, fearful, ashamed, proud, elated, desperate) |
| Mood: | Diffuse affect state, most pronounced as change in subjective feeling, of low intensity but relatively long duration, often without apparent cause. (cheerful, gloomy, irritable, listless, depressed, buoyant) |
| Interpersonal stance: | Affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange in that situation. (distant, cold, warm, supportive, contemptuous, friendly) |
| Attitude: | Relatively enduring, affectively colored beliefs, preferences, and predispositions towards objects or persons. (liking, loving, hating, valuing, desiring) |
| Personality traits: | Emotionally laden, stable personality dispositions and behavior tendencies, typical for a person. (nervous, anxious, reckless, morose, hostile, jealous) |

Figure 21.1 The Scherer typology of affective states (Scherer, 2000).

21.1 Defining Emotion

One of the most important affective classes is emotion, which Scherer (2000) defines as a “relatively brief episode of response to the evaluation of an external or internal event as being of major significance”.

Detecting emotion has the potential to improve a number of language processing...
tasks. Automatically detecting emotions in reviews or customer responses (anger, dissatisfaction, trust) could help businesses recognize specific problem areas or ones that are going well. Emotion recognition could help dialog systems like tutoring systems detect that a student was unhappy, bored, hesitant, confident, and so on. Emotion can play a role in medical informatics tasks like detecting depression or suicidal intent. Detecting emotions expressed toward characters in novels might play a role in understanding how different social groups were viewed by society at different times.

There are two widely-held families of theories of emotion. In one family, emotions are viewed as fixed atomic units, limited in number, and from which others are generated, often called basic emotions (Tomkins 1962, Plutchik 1962). Perhaps most well-known of this family of theories are the 6 emotions proposed by Ekman (see for example Ekman 1999) as a set of emotions that is likely to be universally present in all cultures: surprise, happiness, anger, fear, disgust, sadness. Another atomic theory is the Plutchik (1980) wheel of emotion, consisting of 8 basic emotions in four opposing pairs: joy–sadness, anger–fear, trust–disgust, and anticipation–surprise, together with the emotions derived from them, shown in Fig. 21.2.

![Plutchik wheel of emotion](image)

The second class of emotion theories views emotion as a space in 2 or 3 dimensions (Russell, 1980). Most models include the two dimensions valence and arousal, and many add a third, dominance. These can be defined as:

- **valence**: the pleasantness of the stimulus
- **arousal**: the intensity of emotion provoked by the stimulus
- **dominance**: the degree of control exerted by the stimulus

In the next sections we’ll see lexicons for both kinds of theories of emotion. Sentiment can be viewed as a special case of this second view of emotions as points in space. In particular, the **valence** dimension, measuring how pleasant or unpleasant a word is, is often used directly as a measure of sentiment.
A wide variety of affect lexicons have been created and released. The most basic lexicons label words along one dimension of semantic variability, generally called “sentiment” or “valence”.

In the simplest lexicons this dimension is represented in a binary fashion, with a wordlist for positive words and a wordlist for negative words. The oldest is the General Inquirer (Stone et al., 1966), which drew on content analysis and on early work in the cognitive psychology of word meaning (Osgood et al., 1957). The General Inquirer has a lexicon of 1915 positive words and a lexicon of 2291 negative words (as well as other lexicons discussed below). The MPQA Subjectivity lexicon (Wilson et al., 2005) has 2718 positive and 4912 negative words drawn from prior lexicons plus a bootstrapped list of subjective words and phrases (Riloff and Wiebe, 2003). Each entry in the lexicon is hand-labeled for sentiment and also labeled for reliability (strongly subjective or weakly subjective). The polarity lexicon of Hu and Liu (2004) gives 2006 positive and 4783 negative words, drawn from product reviews, labeled using a bootstrapping method from WordNet.

### Positive
- admire, amazing, assure, celebration, charm, eager, enthusiastic, excellent, fancy, fantastic, frolic, graceful, happy, joy, luck, majesty, mercy, nice, patience, perfect, proud, rejoice, relief, respect, satisfactorily, sensational, super, terrific, thank, vivid, wise, wonderful, zest

### Negative
- abominable, anger, anxious, bad, catastrophe, cheap, complaint, condescending, deceit, defective, disappointment, embarrass, fake, fear, filthy, fool, guilt, hate, idiot, inflict, lazy, miserable, mourn, nervous, objection, pest, plot, reject, scream, silly, terrible, unfriendly, vile, wicked

Slightly more general than these sentiment lexicons are lexicons that assign each word a value on all three affective dimensions. The NRC Valence, Arousal, and Dominance (VAD) lexicon (Mohammad, 2018a) assigns valence, arousal, and dominance scores to 20,000 words. Some examples are shown in Fig. 21.4.

<table>
<thead>
<tr>
<th>Valence</th>
<th>Arousal</th>
<th>Dominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>vacation</td>
<td>.840</td>
<td>.962</td>
</tr>
<tr>
<td>delightful</td>
<td>.918</td>
<td>.840</td>
</tr>
<tr>
<td>whistle</td>
<td>.653</td>
<td>.337</td>
</tr>
<tr>
<td>consolation</td>
<td>.408</td>
<td>.120</td>
</tr>
<tr>
<td>torture</td>
<td>.115</td>
<td>.046</td>
</tr>
</tbody>
</table>

The NRC Word-Emotion Association Lexicon, also called EmoLex (Mohammad and Turney, 2013), uses the Plutchik (1980) 8 basic emotions defined above. The lexicon includes around 14,000 words including words from prior lexicons as well as frequent nouns, verbs, adverbs and adjectives. Values from the lexicon for some sample words:
For a smaller set of 5,814 words, the NRC Emotion/Affect Intensity Lexicon (Mohammad, 2018b) contains real-valued scores of association for anger, fear, joy, and sadness; Fig. 21.5 shows examples.

<table>
<thead>
<tr>
<th>Word</th>
<th>anger</th>
<th>anticipation</th>
<th>disgust</th>
<th>fear</th>
<th>joy</th>
<th>sadness</th>
<th>surprise</th>
<th>trust</th>
<th>positive</th>
<th>negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>reward</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>worry</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>tenderness</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>sweetheart</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>suddenly</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>thirst</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>garbage</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 21.5** Sample emotional intensities for words for anger, fear, joy, and sadness from Mohammad (2018b).

**LIWC**, Linguistic Inquiry and Word Count, is a widely used set of 73 lexicons containing over 2300 words (Pennebaker et al., 2007), designed to capture aspects of lexical meaning relevant for social psychological tasks. In addition to sentiment-related lexicons like ones for negative emotion (bad, weird, hate, problem, tough) and positive emotion (love, nice, sweet), LIWC includes lexicons for categories like anger, sadness, cognitive mechanisms, perception, tentative, and inhibition, shown in Fig. 21.6.

There are various other hand-built affective lexicons. The General Inquirer includes additional lexicons for dimensions like strong vs. weak, active vs. passive, overstated vs. understated, as well as lexicons for categories like pleasure, pain, virtue, vice, motivation, and cognitive orientation.

Another useful feature for various tasks is the distinction between concrete words like banana or bathrobe and abstract words like belief and although. The lexicon in Brysbaert et al. (2014) used crowd-sourcing to assign a rating from 1 to 5 of the concreteness of 40,000 words, thus assigning banana, bathrobe, and bagel 5, belief 1.19, although 1.07, and in between words like brisk a 2.5.

### 21.3 Creating Affect Lexicons by Human Labeling

The earliest method used to build affect lexicons, and still in common use, is to have humans label each word. This is now most commonly done via crowd-sourcing: breaking the task into small pieces and distributing them to a large number of anno-
tators. Let’s take a look at some of the methodological choices for two crowdsourced emotion lexicons.

The NRC Emotion Lexicon (EmoLex) (Mohammad and Turney, 2013), labeled emotions in two steps. To ensure that the annotators were judging the correct sense of the word, they first answered a multiple-choice synonym question that primed the correct sense of the word (without requiring the annotator to read a potentially confusing sense definition). These were created automatically using the headwords associated with the thesaurus category of the sense in question in the Macquarie dictionary and the headwords of 3 random distractor categories. An example:

Which word is closest in meaning (most related) to startle?

- automobile
- shake
- honesty
- entertain

For each word (e.g. startle), the annotator was then asked to rate how associated that word is with each of the 8 emotions (joy, fear, anger, etc.). The associations were rated on a scale of not, weakly, moderately, and strongly associated. Outlier ratings were removed, and then each term was assigned the class chosen by the majority of the annotators, with ties broken by choosing the stronger intensity, and then the 4 levels were mapped into a binary label for each word (no and weak mapped to 0, moderate and strong mapped to 1).

The NRC VAD Lexicon (Mohammad, 2018a) was built by selecting words and emoticons from prior lexicons and annotating them with crowd-sourcing using best-worst scaling (Louviere et al. 2015, Kiritchenko and Mohammad 2017). In best-worst scaling, annotators are given N items (usually 4) and are asked which item is the best (highest) and which is the worst (lowest) in terms of some property. The set of words used to describe the ends of the scales are taken from prior literature. For valence, for example, the raters were asked:

Q1. Which of the four words below is associated with the MOST happiness / pleasure / positiveness / satisfaction / contentedness / hopefulness OR LEAST unhappiness / annoyance / negativeness / dissatisfaction /
Q2. Which of the four words below is associated with the LEAST happiness / pleasure / positiveness / satisfaction / contentedness / hopefulness OR MOST unhappiness / annoyance / negativeness / dissatisfaction / melancholy / despair? (Four words listed as options.)

The score for each word in the lexicon is the proportion of times the item was chosen as the best (highest V/A/D) minus the proportion of times the item was chosen as the worst (lowest V/A/D). The agreement between annotations are evaluated by split-half reliability: split the corpus in half and compute the correlations between the annotations in the two halves.
21.6 Using Lexicons for Sentiment Recognition

In Chapter 4 we introduced the naive Bayes algorithm for sentiment analysis. The lexicons we have focused on throughout the chapter so far can be used in a number of ways to improve sentiment detection.

In the simplest case, lexicons can be used when we don’t have sufficient training data to build a supervised sentiment analyzer; it can often be expensive to have a human assign sentiment to each document to train the supervised classifier.

In such situations, lexicons can be used in a rule-based algorithm for classification. The simplest version is just to use the ratio of positive to negative words: if a document has more positive than negative words (using the lexicon to decide the polarity of each word in the document), it is classified as positive. Often a threshold $\lambda$ is used, in which a document is classified as positive only if the ratio is greater than $\lambda$. If the sentiment lexicon includes positive and negative weights for each word, $\theta^+_w$ and $\theta^-_w$, these can be used as well. Here’s a simple such sentiment algorithm:

$$f^+ = \sum_{w \text{ s.t. } w \in \text{ positive lexicon}} \theta^+_w \text{count}(w)$$
$$f^- = \sum_{w \text{ s.t. } w \in \text{ negative lexicon}} \theta^-_w \text{count}(w)$$

$$\text{sentiment} = \begin{cases} + & \text{if } \frac{f^+}{f^-} > \lambda \\ - & \text{if } \frac{f^-}{f^+} > \lambda \\ 0 & \text{otherwise.} \end{cases}$$

(21.12)

If supervised training data is available, these counts computed from sentiment lexicons, sometimes weighted or normalized in various ways, can also be used as features in a classifier along with other lexical or non-lexical features. We return to such algorithms in Section 21.8.
Many other kinds of affective meaning can be extracted from text and speech. For example detecting a person’s **personality** from their language can be useful for dialog systems (users tend to prefer agents that match their personality), and can play a useful role in computational social science questions like understanding how personality is related to other kinds of behavior.

Many theories of human personality are based around a small number of dimensions, such as various versions of the “Big Five” dimensions (Digman, 1990):

- **Extroversion vs. Introversion**: sociable, assertive, playful vs. aloof, reserved, shy
- **Emotional stability vs. Neuroticism**: calm, unemotional vs. insecure, anxious
- **Agreeableness vs. Disagreeableness**: friendly, cooperative vs. antagonistic, fault-finding
- **Conscientiousness vs. Unconscientiousness**: self-disciplined, organized vs. inefficient, careless
- **Openness to experience**: intellectual, insightful vs. shallow, unimaginative

A few corpora of text and speech have been labeled for the personality of their author by having the authors take a standard personality test. The essay corpus of Pennebaker and King (1999) consists of 2,479 essays (1.9 million words) from psychology students who were asked to “write whatever comes into your mind” for 20 minutes. The EAR (Electronically Activated Recorder) corpus of Mehl et al. (2006) was created by having volunteers wear a recorder throughout the day, which randomly recorded short snippets of conversation throughout the day, which were then transcribed. The Facebook corpus of (Schwartz et al., 2013) includes 309 million words of Facebook posts from 75,000 volunteers.

For example, here are samples from Pennebaker and King (1999) from an essay written by someone on the neurotic end of the neurotic/emotionally stable scale:

> One of my friends just barged in, and I jumped in my seat. This is crazy. I should tell him not to do that again. I’m not that fastidious actually. But certain things annoy me. The things that would annoy me would actually annoy any normal human being, so I know I’m not a freak.

and someone on the emotionally stable end of the scale:

> I should excel in this sport because I know how to push my body harder than anyone I know, no matter what the test I always push my body harder than everyone else. I want to be the best no matter what the sport or event. I should also be good at this because I love to ride my bike.

Another kind of affective meaning is what Scherer (2000) calls **interpersonal stance**, the ‘affective stance taken toward another person in a specific interaction coloring the interpersonal exchange’. Extracting this kind of meaning means automatically labeling participants for whether they are friendly, supportive, distant. For example Ranganath et al. (2013) studied a corpus of speed-dates, in which participants went on a series of 4-minute romantic dates, wearing microphones. Each participant labeled each other for how flirtatious, friendly, awkward, or assertive they were. Ranganath et al. (2013) then used a combination of lexicons and other features to detect these interpersonal stances from text.


Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *ACL-02*.


