

# Aspect-Based Sentiment Analysis as Fine-Grained Opinion Mining

Gerardo Ocampo Diaz<sup>1</sup>, Xuanming Zhang<sup>2</sup>, Vincent Ng<sup>1</sup>

<sup>1</sup>Human Language Technology Research Institute, University of Texas at Dallas

<sup>2</sup>University of Nottingham, Ningbo China

godiaz@hlt.utdallas.edu, zy21855@nottingham.edu.cn, vince@hlt.utdallas.edu

## Abstract

We show how the general fine-grained opinion mining concepts of opinion target and opinion expression are related to aspect-based sentiment analysis (ABSA) and discuss their benefits for resource creation over popular ABSA annotation schemes. Specifically, we first discuss why opinions modeled solely in terms of (entity, aspect) pairs inadequately captures the meaning of the sentiment originally expressed by authors and how opinion expressions and opinion targets can be used to avoid the loss of information. We then design a meaning-preserving annotation scheme and apply it to two popular ABSA datasets, the 2016 SemEval ABSA Restaurant and Laptop datasets. Finally, we discuss the importance of opinion expressions and opinion targets for next-generation ABSA systems. We make our datasets publicly available for download.

**Keywords:** opinion mining, sentiment analysis, text mining

## 1. Introduction

For almost two decades, researchers in the text mining and natural language processing (NLP) communities have worked to improve the state of aspect-based sentiment analysis (ABSA); a task that, roughly speaking, involves extracting sentiment/opinions from text in terms of targets they address. For example, given the sentence:

The soup here is very expensive.

The goal of ABSA is to extract the subject matter of the *opinion*, which is typically represented by an (*entity, aspect*) pair along with the sentiment expressed towards it (frequently in terms of polarity). For example, the previous sentence expresses a *positive* polarity towards (food, price). The entity and the aspect are both chosen from predefined sets. So, if an opinion is expressed on an entity that does not appear in the predefined set of entities, then an ABSA system is not expected to output any opinion for that entity. The important role this task can play in decision making for governments, companies, and individuals has resulted in a large body of research on the subject being developed over the years, along with the organization of multiple shared tasks in SemEval (Pontiki et al., 2014; Pontiki et al., 2015; Pontiki et al., 2016), TASS (Martínez Cámara et al., 2018), and GermanEval (Wojatzki et al., 2017). The datasets published in the SemEval ABSA tasks (Pontiki et al., 2016), which are composed of user-generated reviews of products and services such as those found on Amazon and Yelp, have heavily influenced work in the area, as researchers frequently rely on them for system training and testing. However, these datasets follow a task-specific annotation scheme, under which opinion annotations do not always capture the full meaning of the sentiment expressed in text. Consider the following pair of sentences taken from the SemEval 2016 Restaurants dataset:

(1) I tend to judge a sushi restaurant by its sea urchin, which was heavenly at sushi rose. (2) It melted in my little mouth and the perfect consistency-not too fishy, creamy, and slightly buttery.

According to the annotation scheme, one opinion is listed for each sentence:

```
(1) <OTE="sea urchin"
    category="FOOD#QUALITY"
    polarity="positive"
    from="42" to="52"/>
(2) <OTE="NULL"
    category="FOOD#QUALITY"
    polarity="positive"
    from="0" to="0"/>
```

Here, each opinion’s subject matter is labeled under the *category* field; each category is an (entity, aspect) pair, while the *opinion target expression* (OTE) field labels the explicit reference in the sentence to the entity in the category field. In sentence (1), the author describes the *sea urchin* as *heavenly*. The value of the category field in the annotation is as food/quality, while *sea urchin* is marked as a reference to the entity *food*. This roughly seems to encode the original meaning of the sentiment expressed by the author.

In sentence (2), however, the author expresses an opinion directly towards the *consistency* of the *sea urchin*, but the category *FOOD#QUALITY* does not reflect this. More specifically, this annotation is by no means wrong: *sea urchin* is a dish in the restaurant (which leads to *FOOD*), and consistency is an attribute closely related to *QUALITY*. However, by relying on predefined entities and aspects, *consistency* is being abstracted to *QUALITY*, resulting in *loss of information*. Moreover, the OTE field is *null* because by definition it cannot be filled with a pronoun. Overall, the annotation for this sentence does not precisely encode the author’s original intended meaning.

In this paper, we make the argument for a different annotation scheme for ABSA; one that aims to more closely preserve the semantic information of the opinions originally expressed in text.

Our contributions in this paper are three-fold. First, we make a connection between key ideas in the general opinion mining community (e.g., *opinion target*<sup>1</sup> and *opinion*

<sup>1</sup>As we will explain later, the notion of *opinion targets* in the

expression) and ABSA, which uses the notion of *aspects*, and explore how opinion targets and opinion expressions are used by authors in ABSA datasets to express sentiment. Second, we present our annotation scheme and reannotate two popular ABSA datasets from the SemEval 2016 ABSA shared task (laptop and restaurants), which we make publicly available. Finally, we discuss how our datasets represent a first step towards building next-generation ABSA systems.

## 2. Opinion Mining and ABSA: A Critical Overview

Before discussing an annotation scheme for ABSA, it is sensible to try to understand the core ideas behind it. Here, we aim to cover key concepts such as opinion models, opinion targets, and opinion expressions, as well as provide readers with an intuitive understanding of the breadth of ways in which authors express sentiment in text and the information that might be useful for understanding it. We also introduce our own concepts, examples, and explanations to clear up ambiguity and enable a deeper discussion on key ideas, as the well-established literature is sometimes lacking in this sense (Liu, 2012; Liu, 2015; Liu, 2017).

### 2.1. Opinion Models and ABSA

In this section, we examine two well-known opinion models. The first one, which we refer to as *M1*, models an opinion conceptually as a quadruple  $(g, s, h, t)$  (Liu, 2015), where  $g$  represents the target of the opinion (i.e., the topic, event, object, etc... towards which sentiment is expressed),  $s$  represents whatever sentiment is expressed,  $h$  represents the holder of the opinion/sentiment, and  $t$  represents the time when the opinion/sentiment is expressed/held. Although the sentiment  $s$  can be represented in terms of different *attitudes* (Wilson, 2008), researchers in ABSA frequently extract only *sentiment polarity*. We say that an opinion can be modeled as a quadruple *conceptually*, because coming up with a string that fully encodes the target  $g$  of an opinion is not always straightforward. For example, in the sentence *I love ice cream*, it is easy to identify the target *ice cream*, but this is not the case for *These books are inappropriate for children*; this is not a general opinion on *these books*, but on books in relation to children reading them (hence, *inappropriate*). Further, even when a string can fully describe the target  $g$ , it may not be fully contained in the sentence that expresses sentiment (Liu, 2012). Consider the following interaction between two speakers:

- S1: (1) I really like the new AI-centric cars companies are coming up with.  
 S2: (2) Me too, I love the new Tesla Model 3.  
 (3) The self-driving features are great!

(1) expresses a general opinion on *the new AI-centric cars companies are coming up with*. Here the target of the opinion is completely contained in the sentence. (3), on the

---

general fine-grained opinion mining community is not the same as *opinion target expression* in ABSA datasets.

other hand, does not express an opinion on *self-driving features* in general, but specifically the *self-driving features of the new Tesla Model 3*. The fact that the target is not fully specified in the sentence is considered particularly important in the domain of online user-generated reviews, where most of the initial work on ABSA was originally carried out (Hu and Liu, 2004a; Hu and Liu, 2004b; Liu et al., 2005). In this domain, the objects of interest (e.g., a laptop, restaurant, etc) can be modeled as a hierarchy of “parts” where every part can have *attributes* or *dimensions* and *sub-parts*, and opinion targets can be mapped to a specific leaf node in the hierarchy. For example (Liu, 2012):

A particular model of camera is an entity, e.g., Canon G12. It has a set of attributes, e.g., picture quality, size, and weight, and a set of parts, e.g., lens, viewfinder, and battery. Battery also has its own set of attributes, e.g., battery life and battery weight. A topic can be an entity too, e.g., tax increase, with its parts “tax increase for the poor,” “tax increase for the middle class,” and “tax increase for the rich.”

From this observation, work on ABSA frequently uses a simplified notion of *flat* targets. Here, target  $g$  from *M1* is now represented by a pair  $(e, a)$ , where  $e$  represents the entity or topic of interest, and  $a$  represents an *aspect*: any *attribute/dimension* or *part/component* of  $e$ . This representation of targets corresponds to another opinion model, which we refer to as *M2*. Specifically, *M2* is a quintuplet model of opinions  $(e, a, s, h, t)$ . Consider the sentence *The ink of the printer is expensive* in the context of a printer review. Since the main entity being described is the printer and ink is an aspect, we could represent this opinion’s target  $(e, a)$  as (printer, ink); alternatively, if we choose to model the ink as a separate entity, the target could be (ink, price), but then one needs to separately store the relation between ink and printer.

It is important to point out that the targets under *M2* are in a sense open to interpretation. Consider again the sentence *The ink of the printer is expensive*. Here, the target the author has in mind is clearly *the price of the ink of the printer*, but under *M2*, the valid targets for this opinion include (printer, ink), (printer, cost of operation), or (ink, price). Intuitively, only (ink, price) matches the meaning intended by the author, but all of the targets correspond to valid interpretations from a reader’s perspective, each with a different focus. In other words, modeling opinion targets as (entity, aspect) pairs involves an arbitrary level of granularity, frequently defined by the scope of the task at hand, and does not necessarily preserve authors’ originally intended meaning.

### 2.2. Opinion Types and General Opinion Mining

Opinion models describe desired system output, not how opinions are expressed. Unfortunately, work on ABSA rarely touches on this subject, and even when it does it is frequently lacking in depth. In this subsection, we aim to share our perspectives on how opinions are expressed, basing our discussion on already-established ABSA and opinion mining concepts but using our own examples to illus-

trate how opinions are expressed within the scope of the SemEval ABSA Laptop and Restaurant 2014-2016 datasets (Pontiki et al., 2014)<sup>2</sup>.

One of the more common ways of explaining how opinions are expressed is to distinguish between *explicit* and *implicit opinions* (Liu, 2012). Intuitively, an **explicit opinion** is expressed through a *subjective statement*: one that a) explicitly states an opinion holder’s attitude or feelings towards a target, such as *I love ice cream*, or b) uses language that inherently describes an opinion holder’s attitude or feelings, such as *The ice cream is delicious*. On the other hand, **implicit opinions** are expressed through *objective or factual statements*, such as *I bought this phone a month ago and it died today* or *Repairing an iPhone costs almost as much as buying a new one*.

Conceptually, extracting implicit and explicit opinions involves very different tasks: we can say that explicit opinions are **extracted**, as they are already indicated in the text, while implicit opinions are **inferred** due to the fact that they require accounting in some way for context and/or domain-knowledge.

Although this distinction is fairly important, researchers frequently fail to explain opinions further. There are many different ways in which one can use subjective or objective statements to express opinions. Below we attempt to alleviate this problem by providing a quick overview of the types of opinions that we have found during our survey, as well as a discussion of the importance of different lexical elements from which opinions can be inferred.

### 2.2.1. Explicit Opinions and Opinion Expressions

Explicit opinions are characterized by containing an **explicit opinion expression**: a word or phrase that describes an opinion holder’s sentiment or attitude towards a particular target. Adjectival phrases are the most common type of explicit opinion expression, but adverbial and noun phrases can be used too, along with verbs. Examples include adjectives such as *beautiful*, *great*, nouns such as *hero*, *idiot*, and adverbs such as *beautifully*, *effectively*. Although there are words or phrases that can be considered opinion expressions related to the same attitude or polarity in multiple domains (e.g., *fantastic*, *good*), there might be others that are domain or even target-specific. For example, describing a pillow as hard might imply negative sentiment, while describing a hammer as hard might not imply any sentiment at all. In practice, explicit opinion expressions frequently serve one of the following functions:

- Explicitly describing an opinion holder’s attitude/sentiment
  - Transitive verbs like *love*, *hate*, *agree*, *enjoy* can be used to explicitly describe sentiment towards a target such as in *I love the dress you bought me*, or *I agree that we should have stricter immigration laws*.

- Adjectives that are used to describe internal states, such as *happy*, *sad*, *angry*, *joyous*, can also be used like this, although the syntactic relations between them and targets can be more complicated than examples using transitive verbs (above). Examples include *I am happy that I came here* and *I am angry because you did not buy me ice cream*.

- Using language that implies sentiment on a target

- Simple adjectives like *good*, *bad*, *great*, *tasty* are frequently found in opinionated text, such as *This restaurant is awful*, but they can also appear in slightly more complex statements like *You are an amazing person*, where syntactically speaking, *amazing* describes *person*, but because of the copula relation between *You* and *person*, the entire phrase *amazing person* can be considered an opinion expression on the target *You*.
- Nouns like *genius*, *hero*, *champion* can be used in simple copula relations to describe targets, such as in *I finally realized, you are a genius*.
- Adverbs, like adjectives can be used in simple phrases like *The steak tasted delicious* or in more complex ones like *The fried shrimp was too spicy*, where *too spicy* describes *The fried shrimp*: even though *spicy* might not necessarily imply an opinion, the adverb *too* does.

### 2.2.2. Implicit Opinions

Implicit opinions are inferred from objective statements. One of the explanations provided in existing ABSA literature is that implicit opinions are derived from *desirable facts* (Liu, 2012): common-sense or domain-specific knowledge. For example, the sentence *We initially ordered the kung pao chicken and wanted to cancel it so we’d have room for dessert, but the waiter ignored us when we called him* implies a negative opinion on *The waiter*. It is commonsense knowledge that restaurant staff should be attentive to customer’s requests and therefore ignoring a customer is considered inappropriate. However, explaining implicit opinions simply as *desirable facts* is a significant step back from explicit opinion expressions discussed in the previous subsection. From the point of view of general fine-grained opinion mining, however, one can still identify *implicit opinion expressions*: specific words, phrases, or clauses in an objective statement from which the opinion holder’s sentiment can be inferred<sup>3</sup>. In practice, implicit opinion expressions describe the target itself (through noun phrases) or the target’s behavior (what the target does/ is done to the target):

- Describing the target
  - Nouns or noun phrases may be used in copula relations with targets to describe them. No single part of the phrase needs to imply sentiment,

<sup>2</sup>The concepts and explanations provided here (both from the literature and our own) are not meant to be exhaustive, but roughly cover the breadth of opinions contained in the SemEval datasets. For further reading on traditional explanations of the task, we refer the reader to Liu (2015).

<sup>3</sup>In general opinion mining literature, the term *opinion expression* is used to refer to both implicit and explicit opinion expressions (Wiebe et al., 2005).

but the phrase should imply some sentiment as a whole. For example, in the sentence *The new president is truly someone who has no idea of what he's doing*, the noun phrase *someone who has no idea of what he's doing* describes the new president.

- Describing target behaviour

- Verb phrases can imply sentiment on their subjects or objects. Consider the sentence *The waiter took our order while holding a garbage bag* in a restaurant review. Here, *took our order while holding a garbage bag* implies negative sentiment on the waiter, who performs the action. Similarly, in *I could fit the carry on bag in the overhead bin easily*, the verb phrase *could fit in the overhead bin easily* implies a positive sentiment on the carry on bag.

Nevertheless, less common, implicit opinions can also sometimes be inferred from events, such as *I took the pill and then I felt much better*. Here, the entire sentence can be viewed as the opinion expression. Conceptually, extracting the opinion from this example involves inferring first that *The pill caused me to feel better* and then that *cause me to feel better* implies positive sentiment (like in explicit opinions).

## 2.3. Lexical Targets, Opinion Expressions, and Semantic Targets

To facilitate our discussion, in this subsection we introduce the concept of lexical and semantic targets. **Lexical targets** are phrases found in the text which the author uses to refer to the opinion target, while **semantic targets** are the “real”, fully-specified targets in the opinion holder’s mind<sup>4</sup>. For example, in *I went to Feng Cha. The tea was so cheap!*, the lexical target is *The tea*, while the semantic target is *the price of Feng Cha’s tea*. Note that semantic targets are the same as  $g$  in  $M1$ .

In practice, what really matters is the semantic target, so why are lexical targets important? To answer this question, we make the following observation: the combination of lexical target + opinion expression can be used to infer an opinion’s semantic target. In the previous example, *the price of Feng Cha’s tea* can be inferred from the *The tea* (lexical target) and *expensive* (opinion expression). This is also true even when the semantic target can not be described in a straightforward manner, such as in *The way the waiter looked at me was totally inappropriate*, where the lexical target is *The way the waiter looked at me* and the opinion expression is *totally inappropriate*.

To better understand the significance of our observation, consider its implication for opinion model  $M2$ . Recall that *The ink of the printer is too expensive*. has three valid targets that correspond to different levels of granularity under  $M2$ : (ink, price), (printer, cost of operation),

and (printer, ink). Simply put, the combination of lexical target and opinion expression sufficiently encodes the original meaning of the sentiment intended by the author, and can therefore be used to infer all valid (entity, attribute) pairs in  $M2$ . Specifically, from the lexical target *ink of the printer* and opinion expression *too expensive*, it is easy to see that among the three valid targets, (ink, price) matches most closely the meaning of the original text. In addition, it can be used to infer the other two valid targets, which are less fine-grained. In other words, the combination of lexical target and opinion expression encodes the sentiment expressed in the text without any loss of information, so it can be useful regardless of the level of target granularity one is interested in.

## 3. Dataset

In this section we provide an overview of related fine-grained sentiment analysis datasets, detail our meaning-preserving annotation scheme and procedure, and detail inter-annotator agreement and dataset statistics.

### 3.1. Related Resources

Over the years, researchers have produced datasets for ABSA of different sizes and domains. However, not all datasets have been made publicly available and few have received as much attention as the SemEval ABSA Restaurant and Laptop datasets (Pontiki et al., 2014; Pontiki et al., 2015; Pontiki et al., 2016), which consist of 440 and 530 user-generated reviews mined from online websites respectively. The datasets aim to capture sentiment in terms of pre-defined opinion targets, where a target is defined as a specific (entity, attribute) pair, such as FOOD#QUALITY or LAPTOP#PRICE. These categories are assigned if an opinion directly references an ENTITY#ATTRIBUTE pair, or if the opinion itself can be generalized to describe one. For each sentence, every opinion in the sentence is annotated, with each set of annotations consisting of 1) the opinion target as an entity,attribute pair, 2) the opinion polarity (positive or negative), and 3) the opinion target expression (OTE), which is defined as the explicit reference to the entity in the opinion target (if any). For example, in the context of a restaurant review, there would be one opinion annotation for *The shrimp was delicious!*, with the opinion target (FOOD,QUALITY), OTE *shrimp*, and polarity *positive*. Opinion holders and the time at which opinions are held (from  $M1$ ) are not annotated, as it is assumed that the opinions expressed in a review are in line with the author’s own.

It is important to point out that OTEs are not equivalent to our notion of lexical targets. To see why, consider the example *The way the waiter looked at me was totally inappropriate*. Under the SemEval annotation scheme, the corresponding target is (SERVICE,GENERAL), since the author is expressing an opinion towards the customer service (the aspect is GENERAL because SERVICE does not have any specific attributes under the SemEval annotation scheme), and the corresponding OTE is *waiter*, with opinion polarity negative. However, in reality, this statement directly expresses an opinion on *the waiter’s behavior*, specifically, its *appropriateness*. Because of this, we can interpret this

<sup>4</sup>In the general fine-grained opinion mining literature (Yang and Cardie, 2013), lexical targets are simply referred to as *opinion targets*, while the notion of semantic targets is absent.

as an opinion on *the waiter*, and as a consequence, as an opinion on *customer service*. This information is lost under the SemEval annotation scheme due to the fact that the lexical target and opinion expression are not annotated.

Another relevant well-known resource is the Multi-Perspective Question Answering (MPQA) dataset (Wiebe et al., 2005; Wilson, 2008; Deng and Wiebe, 2015), which is composed of 535 political news articles spanning 10 different topics and annotated with opinion expressions, opinion holders, and lexical opinion targets. The MPQA dataset introduces a general framework for fine-grained sentiment analysis, and contains annotations for opinions in terms of lexical targets, opinion expressions, and opinion holders. Given that it is composed of political news articles, there is no notion of targets as *entity*, *aspect* pairs, and the language, topics, and opinion holders can be quite different from those found in user-generated reviews.

### 3.2. Annotation Scheme and Procedure

Recall that the SemEval datasets a) annotate targets as (entity, aspect) pairs, leading to annotations that lose information from authors' originally intended sentiment, and b) ignore opinion expressions that are useful for inferring the semantic target and sentiment polarity of the opinion. Therefore, we reannotate both the Laptop and Restaurant datasets from the SemEval 2016 ABSA shared task. For each opinion in the dataset, we annotate the lexical opinion target, opinion expression and opinion polarity. By annotating lexical opinion targets and opinion expressions, 1) authors' originally intended sentiment is preserved, and 2) we provide useful information to infer semantic targets, valid (entity, aspect) pairs at different levels of granularity, and the sentiment of the opinions.

In the following subsections, we describe the different types of annotations found in the dataset. The different types of opinion targets and expressions are exhaustive; no opinions found in the dataset are considered out of scope.

#### 3.2.1. Opinion Targets

We annotate three different types of opinion targets:

- *Lexical Targets*: Targets appearing as phrases in the original text
  - The screen of the laptop is great → *The screen of the laptop*
- *Implicit Targets*: Opinion targets that do not appear in the original text but can be inferred from context/opinion expressions.
  - Not a great place for family dining → *restaurant*. The opinion is given towards the entity itself - the restaurant. Implicit targets are annotated based on the corresponding SemEval entity catalog. Here, the entity is annotated in place of the lexical target, and the semantic target is encoded by the combination of the implicit target and the opinion expression.
- *Resolved Targets*: Pronominal targets that can be resolved to noun phrases based on neighboring sentences. For example, in *I ate the crab cake. It was*

*the best in town!*, it is annotated as the lexical target, but *crab cake* is also labeled as the resolution of *it*.

As far as we know, none of the existing ABSA datasets has pronominal targets resolved.

In addition, for each lexical target we annotate its semantic head:

- The views around the mountain are beautiful → The opinion target is *The views around the mountain* and its head is *views*.
- I love swimming in the ocean → The opinion target is *swimming in the ocean* and its head is *swimming*

Our annotation of semantic heads is motivated by the scheme used for annotating the MUC-6 and MUC-7 datasets (Grishman and Sundheim, 1995; Chinchor, 1998), which were designed for information extraction tasks such as named entity recognition and entity coreference resolution; Complex phrases are typically hard for a system to extract, so in order not to penalize a system for its failure to extract a phrase because of its complexity, one can consider that it extracts the phrase correctly as long as it extracts the simpler (and thus arguably easier to extract) phrase denoted by the semantic head. As far as we know, none of the existing fine-grained sentiment analysis corpora has semantic heads annotated.

#### 3.2.2. Opinion Expressions

We annotate both *explicit* and *implicit* opinion expressions, which can be broadly divided into five categories:

- *Type 1: Noun or adjectival phrases inherently associated with sentiment*
  - The fried chicken was far too oily → *far too oily*
  - This place is the best → *the best*
- *Type 2: Verbs or verb phrases that imply sentiment on their subject*
  - These pills got me to stop throwing up! → *got me to stop throwing up*
  - The waiter took our order while holding a garbage bag → *took our order while holding the garbage bag*
- *Type 3: Verb or verb phrases that imply sentiment on their object*
  - My dog ripped it [this leash] apart in less than a month → *ripped it apart in less than a month*
  - I had to ask the waiter because she was not paying attention → *had to ask the waiter three times because she was not paying attention*
- *Type 4: Explicit expressions of sentiment*
  - I am so happy I came here. → *happy*
- *Type 5: Miscellaneous*

- Since I started eating here, my life changed. → The whole phrase is marked as an opinion expression.
- I bought this laptop and now I can play all the games I want! → The whole phrase is marked as an opinion expression.

For each opinion expression, we also annotate its polarity (positive/negative):

- The laptop is nice but kind of expensive. → Two opinion expressions are labeled: *nice* with polarity *positive* and *kind of expensive* with polarity *negative*

Further, each opinion expression is labeled as either *subjective* or *objective*:

- *I had an amazing time here!* → *had an amazing time* is labeled as *subjective*
- *We had to wait 45 minutes for our food* → *had to wait 45 minutes for my food* is labeled as *objective*

Finally, just like opinion targets, subjective opinion expressions have their semantic heads labeled:

- *The fish was far too oily!* → the opinion expression is *far too oily* and semantic head is *oily*
- *Applebee's is my favorite restaurant in the whole world!* → the opinion expression is *my favorite restaurant in the whole world* and the semantic head is *favorite*
- *My dog broke this leash in less than a week* → the opinion expression is *my dog broke this leash in less than a week* and the semantic head is *broke* (the lexical target is *this leash*).

Note that objective opinion expressions do not have their semantic heads annotated since the entirety of the expression is needed to infer sentiment polarity and the semantic target.

### 3.3. Annotation Procedure and Inter-annotator Agreement

The dataset is labeled by two human annotators and a meta-annotator. First, annotators are given a small set of 20 reviews to familiarize themselves with the annotation scheme and procedure. After discussion with the meta-annotator, both annotators separately label opinion target and opinion expression spans for 30% of laptop and restaurant reviews (159 and 132 reviews, respectively). This includes pronominal target resolution, implicit targets, and semantic head annotations for subjective opinion expressions and lexical targets. Inter-annotator agreement is calculated and disagreements are resolved with the meta-annotator. Due to the high agreement between annotators for pronominal and implicit target resolution, as well as semantic head annotations (.98–1.0). It is agreed to let the meta-annotator label these fields over the final lexical target and opinion expression annotations. The two annotators are then each assigned 50% of the remaining laptop and restaurant reviews and annotate opinion target and expression spans.

Finally, pronominal target resolution, implicit targets, semantic heads of subjective opinion expressions and lexical targets, and opinion expression types are annotated by the meta-annotator.

To measure agreement we use both Cohen's kappa (Cohen, 1960) and recall-based agreement. Recall-based agreement is calculated as follows

$$agr(A, B) = \frac{\# \text{ of spans annotated by A and B}}{\# \text{ of spans annotated by A}}$$

agr(A,B)	agr(b,a)	avg.
0.92	0.90	0.91

Table 1: Recall-based agreement for opinion target spans

agr(A,B)	agr(B,A)	avg.
0.99	0.96	0.975

Table 2: Recall-based agreement for opinion expressions

Cohen's kappa (Cohen, 1960) is calculated over every word in the corpus. Each word is tagged as *no label*, *opinion expression word* or *opinion target word*.

Span	Cohen's Kappa
Opinion Target	0.90
Opinion Expression	0.97

Table 3: Cohen's K for opinion targets and opinion expressions annotation

### 3.4. Dataset Statistics

We reannotate the restaurant and laptop datasets from SemEval 2016 (Pontiki et al., 2016), consisting of 440 restaurant reviews and 530 laptop reviews. The restaurant dataset contains 2,687 sentences in the dataset, with 2,420 opinion annotations. The laptop dataset contains 3,308 sentences with 1,735 opinion annotations<sup>5</sup>.

	Restaurants	Laptops
Opinions	2,420	1,735
Subjective Opinion Expressions	2,089	1,456
Objective Opinion Expressions	339	279

Table 4: Dataset statistics

## 4. Implications for New Applications

The original motivation behind ABSA was to summarize user-generated online reviews (Hu and Liu, 2004a; Hu and Liu, 2004b). The idea is that if all opinion targets can be modeled in terms of *entity*, *aspect* pairs, then all of the opinions in a given review can be reported in a structured manner. Current work on ABSA is based on *pre-defining* aspects, but there are inherent limitations to handling opinion targets in this way. First, targets defined in this way

<sup>5</sup>The datasets are publicly available for download at <https://godiaz01.github.io/resources/>

are frequently highly dependent on domains (e.g., laptops and restaurants do not have many aspects in common under the SemEval annotation schemes), making it complicated to apply ABSA to a domain for which one does not have annotated data. Second, working at the aspect level can result in the loss of important information. Consider *The charge capacity is excellent!* and *The battery gets really hot when charging* in the context of a laptop review. Both sentences can be reported under the target (battery, operating performance), but they do not address the same dimension of the operating performance. Attempting to solve this problem by pre-specifying more aspects increases the number of needed annotations and exacerbates the previous problem. Finally, a system can only report opinions in terms of the fixed aspects and aspect hierarchy that are defined by its training data; in particular, there is no way for a system to report opinions in terms of dimensions that are of interest to a specific user.

We envision that more advanced fine-grained sentiment analysis systems should be able to learn meaningful ways to report opinions without requiring humans to specify aspect hierarchies, prevent information loss that may make two different opinions be reported under the same target when they in fact do not talk about the same thing, and report opinions under different aspects to satisfy the needs of different users. Such systems would be able to provide shopping assistance for example, providing side by side comparisons of two unseen products based entirely on their reviews, identifying meaningful axes around which to categorize and order opinions independently, or reporting reviewers' opinions in terms of dimensions specified by the user. Given that the combination of lexical targets and opinion expressions encodes sentiment as expressed by writes without loss of information, we believe we will need resources based around them to work towards more advance sentiment analysis systems in this domain.

## 5. Conclusion

In this paper we attempted to bridge the gap between general opinion mining and aspect-based sentiment analysis by showing the relationship between ABSA concepts such as opinion models and aspects and those used in opinion mining, opinion expressions and opinion targets. We showed how traditional ABSA annotation schemes based on representing targets as (entity, aspect) pairs can result in losing the original opinion target intended by the author and how the combination of lexical opinion targets and opinion expressions solves this problem while at the same time enabling the inference of valid (entity, aspect) pairs at different levels of granularity. We reannotated the popular SemEval 2016 ABSA restaurant and laptop datasets based on this notion, annotating opinion expressions, opinion targets, and opinion polarities, as well as the semantic heads of opinion expressions and opinion targets. Finally, we discussed how these datasets could be useful for building next-generation ABSA systems. To stimulate work on this problem, we make our datasets publicly available.

## 6. Acknowledgments

We thank the three anonymous reviewers for their detailed and insightful comments on an earlier draft of the paper. This work was supported in part by NSF Grants IIS-1528037 and CCF-1848608. Any opinions, findings, conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views or official policies, either expressed or implied, of NSF.

## 7. Bibliographical References

- Chinchor, N. A. (1998). Overview of MUC-7/MET-2. In *Proceedings of the Seventh Message Understanding Conference*.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Deng, L. and Wiebe, J. (2015). Mpqa 3.0: An entity/event-level sentiment corpus. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1323–1328.
- Grishman, R. and Sundheim, B. (1995). Design of the MUC-6 evaluation. In *Proceedings of the Sixth Message Understanding Conference*, pages 1–11.
- Hu, M. and Liu, B. (2004a). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177.
- Hu, M. and Liu, B. (2004b). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artificial Intelligence*, pages 755–760.
- Liu, B., Hu, M., and Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web*, pages 342–351.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
- Liu, B. (2017). Many facets of sentiment analysis. In *A Practical Guide to Sentiment Analysis*, pages 11–39. Springer.
- Martínez Cámara, E., Almeida Cruz, Y., Díaz Galiano, M. C., Estévez-Velarde, S., García Cumbreiras, M. Á., García Vega, M., Gutiérrez, Y., Montejó Ráez, A., Montoyo, A., Muñoz, R., et al. (2018). Overview of TASS 2018: Opinions, health and emotions. *Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN*.
- Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35.
- Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495.

- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryigit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30.
- Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210.
- Wilson, T. A. (2008). *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. Ph.D. thesis, University of Pittsburgh.
- Wojatzki, M., Ruppert, E., Holschneider, S., Zesch, T., and Biemann, C. (2017). GermEval 2017: Shared task on aspect-based sentiment in social media customer feedback. *Proceedings of the GermEval*, pages 1–12.
- Yang, B. and Cardie, C. (2013). Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1640–1649.