Guidelines for Annotating Fine-grained Emotion Trajectories in Customer Service Dialogues

Version 1.0

LT3 Technical Report – LT3 20-01

Sofie Labat, Thomas Demeester and Veronique Hoste

LT3 – Language and Translation Technology Team
Department of Translation, Interpreting and Communication – Ghent University
firstname.lastname@ugent.be

December 16, 2020

Contents

Chapter 1

Introduction

The field of conversational AI has achieved immense progress in recent years, see e.g., Google's Meena (Adiwardana et al., 2020); Facebook's Blender (Roller et al., 2020); Emora, the socialbot winning the 2020 Alexa Prize (Finch et al., 2020); and OpenAI's GTP-3 (Brown et al., 2020). Nevertheless, an important challenge that remains to be solved is the lack of connotative commonsense knowledge in such conversational systems, more precisely their inability to correctly detect, handle and respond to emotions during a conversation. Most studies on emotion recognition in conversations and affective dialogue generation focus on open-domain conversations (see e.g., Rashkin et al. (2019); Lin et al. (2020)). For the purpose of our research, we investigate emotion dynamics in domain-specific customer service dialogues, as these dialogues contain a clear role division between the customer who addresses the company and the company's handler who attempts to help the customer.

Customer service encompasses all types of interactions a company has with its current and potential clients. It is an essential marketing tool for every business in that such interactions not only enhance customer experience, but they also have a positive effect on customer satisfaction and, subsequently, on customer retention (Goffin, 1999; Gustafsson et al., 2005; Mithas et al., 2005). Due to the digital transformation and the arrival of the Web 2.0, many companies have discovered the possibilities of performing online and personalized customer relationship management through social media (Trainor et al., 2014). This evolution is also visible on Twitter, which has emerged as a popular communication channel for companies to engage in customer service dialogues (Culnan et al., 2010). While the popularity of online customer service conversations increases, a need arises to (i) automatically monitor customer satisfaction on these platforms and to (ii) create automated customer service agents that assist human agents in giving live, affective and personalized responses.

In these guidelines, we propose a novel framework to annotate fine-grained emotion trajectories in Dutch customer service dialogues on Twitter. We build on insights from various research fields such as psychology (Mehrabian and Russell, 1974), marketing (Simon, 2013), and emotion analysis in natural language processing (NLP) (Xia and Ding, 2019; Demszky et al., 2020) to propose an all-encompassing, interdisciplinary framework. The resulting guidelines consist out of four annotation layers, which can easily be transferred to other languages or to other communication tools (e.g., chat or e-mail correspondences). In the first and most global layer (see Chapter 2), we indicate whether the conversation is subjective or objective and whether an explicit description of the event (potentially) triggering emotions is present in the conversation. We also flag conversations that contain turns in other languages than Dutch or that do not make any sense to the annotator. If in the first annotation layer an annotator decided that the conversation contains an explicit cause, then this cause is extracted and labelled along the procedures described in Chap-

ter 3. Similarly, if an annotator marks the conversation as subjective in the first layer, then the emotions in the customer turns and the response strategies in the employee turns are extracted and labelled along the guidelines introduced in Chapter 4 and Chapter 5 respectively.

All annotations are made in INCEpTION (Klie et al., 2018), a publicly available and open-source platform designed for semantic annotation tasks. In accordance with Twitter's Developer Agreement and Policy[1], we choose to not display any actual Twitter conversations. Instead, the examples in the following chapters are reconstructed (fragments of) dialogues, based on the real-life Twitter dialogues in our corpus. Each example first displays a Dutch version in item (a), while its translation is provided in item (b). It is important to remark that examples are only annotated along one specific annotation layer, namely the layer that is introduced in that respective chapter. We refer to Chapter 6 for a complete overview of the annotation procedure in which the different layers are combined, and to Chapter 7 for some examples of fully-annotated dialogues in the INCEpTION environment.

---

[1]see https://developer.twitter.com/en/developer-terms/agreement-and-policy

Chapter 2

Conversation characteristics

For this first layer, annotators have to inspect the entire dialogue to determine three different aspects, namely whether:

   (i) the dialogue pertains to the scope of this research on Dutch customer support dialogues;

  (ii) the customer turns contain emotions;

 (iii) the dialogue contains an explicit description of an event that caused a customer to initiate the dialogue and that has the potential to trigger emotions in the customer.

All annotations of this layer are made on the conversation level by creating a "zero-width span annotation" at the beginning of each conversation. This means that the span on which annotations are made has a length of zero. To make such annotations in INCEpTION, one should hold SHIFT and click on the position where one wishes to place the annotation. The examples in this chapter are dialogues between customers and employees. To clarify whether a turn is made by a "Customer" or an "Employee" representing the company, each turn is preceded by either a **[C]** or an **[E]**.

2.1 Out of scope annotation

We flag the dialogue with the "out of scope" label if one of the following two conditions applies:

  (i) **The dialogue is not in Dutch.**
     → A dialogue can contain some words or slang in other languages. Nevertheless, if entire sentences are in another language than Dutch, the dialogue should be labelled with "out of scope".

  (ii) **The dialogue does not make sense or is not understandable.**
     → If it is impossible for the annotator to understand the dialogue, one cannot possibly make fine-grained annotations on it. In such cases, we flag the dialogue with "out of scope".

The examples below illustrate when an "out of scope" label should be given to a dialogue. Example 1 consists of an English dialogue. As this dialogue is in another language than Dutch, annotators have to flag it with "out of scope". Conversations 2 and 3 are also tagged with the same

label, since they do not make a lot of sense and can therefore not be annotated. In dialogue 2, the customer refers to the non-existing company's mother without any additional explanations. Dialogue 3 contains a number of unfamiliar letter combinations/self-invented words which make the conversation hard to understand.

(1) **[C]** Hey @Company! Due to the recent changes to the subscription plans, my current plan is no longer available... Can you help me in finding another plan?

   [E] @Customer of course we can! Can you please send me a message in DM with your phone number?

   → Out of scope

(2) a. **[C]** @Company uw moeder!

     [E] @Customer Bedankt om Genesis te lezen en dit met ons allen te delen 🙂 Kan ik je misschien met iets verderhelpen?

   b. **[C]** @Company your mother!

     [E] @Customer Thanks for reading Genesis and sharing this experience with us 🙂 Can I maybe help you with something?

   → Out of scope

(3) a. **[C]** Dees internet geeft mij klnoen wslh stress kma @Company 😡

     [E] Oei @Customer dat klinkt niet goed! Kan je mij eens vertellen wat je aan de hand hebt?

   b. **[C]** This internet gives me *klnoen wslh* stress *kma* @Company 😡

     [E] Oh @Customer that doesn't seem right! Can you tell what's going on?

   → Out of scope

2.2 Subjectivity identification

A Twitter dialogue contains emotions if there is at least one customer turn that is subjective. A turn is characterized as subjective when an interlocutor either explicitly or implicitly expresses his or her emotions in it. Once one customer turn is identified as subjective, the entire conversation has to be labelled as "subjective". If the conversation contains no subjective customer turns, the conversation has to be labelled as "objective".

In some cases, emotions are expressed in an explicit manner. Upon deciding if a turn is subjective or not, it helps to focus on the presence of frequently re-occurring and prototypical constructions. Examples of such Dutch constructions are *dank u* ("thank you"), *ik hoop dat* ("I hope that"), *ik wil graag dat* ("I'd like that"), *ik haat dit* ("I hate this")... In addition, some adjectives and adverbs explicitly signal the presence of emotions, e.g., *vervelend* ("annoying"), *super* ("great"), and *slecht* ("bad")... Finally, emojis, excessive use of punctuation, and capitalized words also indicate the presence of emotions.

Nevertheless, it is often difficult to determine whether a conversation is subjective or objective, because many emotions are implicitly conveyed. Annoyance is, for example, an emotion that can

be implicitly communicated through irony, by mentioning recurring events (e.g., *al de 5de keer* ("already the 5th time")), or by referring to issues that started in the past and are still ongoing (e.g., *al een maand geleden* ("already a month ago")). In many cases, one can only detect implicit emotions by taking the underlying meaning and surrounding context of the other turns into account.

Dialogue 4 contains an example of an objective conversation, as the customer only poses a neutral question to the company. In dialogue 5 on the other hand, the customer clearly expresses his or her negative feelings towards the company. Hence, this conversation is labelled as "subjective".

(4) a. **[C]** @Company Goedemorgen! Wat is er precies mis met de trein naar Sint-Truiden van 8u34?

   [E] @Customer Goeiemiddag, door een technische storing met een locamotief is deze trein met vertraging vertrokken in Gent-St-Pieters. We proberen deze vertraging doorheen de dag in te halen.

   b. **[C]** @Company Good morning! What exactly went wrong with the train to Sint-Truiden from 8h34?

   [E] @Customer Good afternoon, due to a technical failure with a locomotive, this train left Ghent-St-Pieters with a delay. We'll try to make up for this delay throughout the day.

   → Objective conversation

(5) a. **[C]** @Company er is nog steeds geen informatie over mijn vlucht! Ik zou wel binnen een half uur moeten vertrekken!!! Wat flutservice zeg, echt de laatste keer EVER 😡

   [E] @Customer We vinden het jammer om dit te horen 🙁 Onze excuses hiervoor. Hebben ze je ondertussen kunnen verderhelpen?

   b. **[C]** @Company there is still no information about my flight! I'm supposed to leave in half an hour!!! What a sh*tty service, really the last time EVER 😡

   [E] @Customer We're sorry to hear this 🙁 We apologize for this. Have they been able to help you in the meantime?

   → Subjective conversation

2.3 Presence of explicit cause

Finally, we also have to indicate whether the customer explicitly describes the event that led up to a conversation and that caused a customer to contact the company. Events often trigger or have the potential to trigger emotions in the customer. We include past events, events that started in the past and are still ongoing at the point the tweet was published, and repeated events/habits. We leave out hypothetical or in the future situated events, since these have not yet happened and will thus, in most cases, not have triggered any emotions. When the conversation contains an explicit event description in one of the customer turns, the conversation receives the "cause present" label. If this is not the case, the conversation is annotated with "cause absent".

There are a couple of remarks to be made regarding this annotation procedure. First, a preliminary condition that we imposed is that each conversation can at most have one cause. If in some

rare cases more than one cause is present, the conversation is also annotated with "cause present" (for details on how to handle the different cause descriptions we refer to Chapter3). Second, one should only inspect customer turns for event descriptions, not employee turns. By making a single, general statement that is directed towards all customers, a company representative might start a number of parallel dialogues with different clients. When a client responds to the general message, the event that caused this client to initiate the conversation is actually the publication of that first message. In these particular cases, we label the conversation with "cause absent", since the conversation does not contain an explicit description of the event. Third, annotators should be looking for an event, not an intent. For example, if a customer asks when the next bus will arrive, his or her intent is to require information about a certain service, but the same request does not contain an explicit event description. One could infer from the context that the bus is delayed, but such implicit events are annotated as "cause absent" according to these guidelines.

The following examples illustrate this annotation procedure. A conversation containing turn (6) is labelled as "cause present", because the first sentence of this turn contains a clear description of an event that occurred and that has triggered certain emotions in the customer. Turn (7) contains no explicit event description. In most cases, questions do not carry explicit event descriptions. One exception to this guideline is turn (8). Questions that do contain event descriptions are either rhetorical questions or questions that carry constructions such as *is het normaal dat + evenement?* ("is it normal that + event?"), *dat is nu toch niet oké dat + evenement?* ("that's not okay now that + event?"), *ligt het aan mij of het maar raar dat + evenement?* ("is it me or is it strange that + event?"), etc. An explicit description of an event that is situated in the future can be found in turn (9). As described above, we exclude such events, because they will most likely not trigger any emotions. Turn (10) does not carry an explicit event description, but we can implicitly infer from the context that the bus driver passed a bus stop without actually picking up travellers. Nevertheless, according to these guidelines we do not consider implicit events. In dialogue (11), the publication of that first tweet by the company representative is the event that caused the conversation to start and that has the potential to trigger emotions in the customer. As dialogue (11) does not carry an explicit description of that event (i.e., it contains the first tweet, but not a description of the publication of that tweet in one of the customer turns), the conversation should be annotated with "cause absent".

(6) a. **[C]** Dat is de vijfde keer vandaag dat mijn internet wegvalt! Hoe kan ik zo telewerken?!

   b. **[C]** That's the fifth time today that my internet drops! How can I telework like this?!

   → Cause present

(7) a. **[C]** Waar blijft de 651 richting Zaventem?

   b. **[C]** Where is the 651 in the direction of Zaventem?

   → Cause absent

(8) a. **[C]** Is dat normaal dat ik al twee uur op mijn bagage aan het wachten ben?! 😡

   b. **[C]** Is it normal that I have been waiting for my luggage for two hours?! 😡

   → Cause present

(9) a. **[C]** Ik vertrek woensdag naar Spanje. Om hoe laat moet ik aanwezig zijn op de luchthaven?

b. **[C]** I'm leaving for Spain on Wednesday. What time do I have to be at the airport?

→ Cause absent

(10) a. **[C]** Volgende keer spring ik voor de bus. Misschien wil de chauffeur dan eindelijk wel eens stoppen... 🙃

b. **[C]** Next time I'll jump in front of the bus. Maybe then the driver would finally like to stop... 🙃

→ Cause absent

(11) a. **[E]** Vandaag: kortingsactie bij @Company!! Wees niet te laat ;-)

[C] @Company kan ik jullie korting ook krijgen als ik geen package 1, maar een package 2 heb (enkel mobiele data)?

b. **[E]** Today: discount promotion at @Company!! Don't be too late ;-)

[C] @Company can I also receive your discount if I don't have a package 1 subscription, but I do have a package 2 subscription (only mobile data)?

→ Cause absent

Chapter 3

Cause

While most research on text-based emotion processing focuses on the detection and classification of emotions, in recent years the field has also become interested in the events that trigger emotions. According to the psychological theory of appraisal, our own evaluation or interpretation of an event causes an emotional response (Scherer et al., 2001). Following this theory, Hofmann et al. (2020) released the first event-centered text corpus that contains dimensional appraisal annotations. They demonstrated that appraisal-based emotion classification performs better than text-based classification, under the condition that the model has access to perfect appraisal predictions and complementary correct predictions are also available. Although their findings demonstrate the potential of appraisal information to boost the performance of emotion detection, they were not able to reach a substantial improvement in a real-world setting, which suggests that more data is needed to train a model (Hofmann et al., 2020).

Another recent development is the task of emotion cause detection and extraction which was first proposed by Lee et al. (2010) who approached the task as a word-level sequence labelling problem. Gui et al. (2016) were the firsts to formally define the task of emotion cause extraction (ECE) as a clause-level binary classification problem in which the goal is the determine whether or not a clause is the cause of a certain emotion, given that emotion. In order to solve this task, Gui et al. (2016) built the first open-source corpus using SINA city news data on which they trained a multi-kernel classifier (SVMs) to identify emotion cause events. Recently, Xia and Ding (2019) introduced the new task of emotion-cause pair extraction (ECPE) which aims at extracting both the potential causes and their corresponding emotions from a document. The ECPE task addresses two shortcomings from the ECE task in that (i) emotion annotations no longer have to be available but will be extracted instead, and (ii) by first annotating emotions and then extracting the corresponding causes the original ECE task ignores the fact that causes and emotions are mutually indicative.

Building on these insights, we decided to extract and label events happening prior to the conversations that caused or have the potential to cause emotions (cf. Section 2.3). This implies that we include explicit descriptions of events in conversations that are objective, as these events have the potential to trigger emotions and they might have actually accomplished this with other interlocutors. As a prior condition, we impose that each conversation has a maximum of one cause. In the rare cases where more than one event description is present, one should only annotate the event that is the most direct cause of an emotion. If this rule of thumb cannot be applied, annotators should annotate the cause that is most explicitly described in the conversation.

Annotations should be made at the sentence level. The sentence with the cause description resides in most cases in the first customer turn. Nevertheless, the event description can also be

contained in other turns or can even span several turns (e.g., a sentence which is interrupted in turn $n-1$ because each tweet can only hold a predefined number of characters; the same sentence continues in turn n). Since we are working on rather noisy social media data, annotators are allowed to deviate from sentence-level spans if one of the following conditions applies:

(i) Punctuation is not properly used by the customer. This means that sentences either end too early or that several sentences are combined together without any proper end of sentence markers.

(ii) One sentence alone has too little context to have a proper understanding of the event (potentially) triggering emotions.

Once the span of an event description is selected, one should assign an adequate label to that event. For our customer service data, we decided to base these category labels on insights from research in the field of marketing, management and customer services. Therefore, we used the findings from Simon (2013) who let her students participate in a survey on customer complaint handling to investigate the influence of empathy on customer loyalty. Simon (2013) manually divided the submitted complaints in six different categories (viz. (i) employees' difficulty in resolving problems and attending to consumers; (ii) lack of product quality; (iii) delays and service breakdowns; (iv) product information and website design inadequacies; (v) environmental or consumer health; (vi) other causes). On our turn, we adapted her labels to the following eight categories:

– **Employee service**
  Events related to employees' or company's difficulty or ease in resolving problems and attending to consumers.

  (12) a. Dit is SERIEUS de 10e keer dat ik naar @Company bel, maar jullie willen gewoon niet opnemen of wat??!
       b. This is SERIOUSLY the 10th time I've called @Company, but you guys just don't want to answer or what??!

  (13) a. @Company merci te blijven rijden in deze gekke tijden, jullie zijn echte helden 🙏
       b. @Company thanks to keep driving in these crazy corona times, you are real heroes 🙏

  (14) a. WOOOW die buschauffeur hier discrimineert juist twee zwarte schoolkinders! 🤯
       b. WOOOW that bus driver here just discriminated against two black school kids! 🤯

– **Product quality**
  Events related to the apparent lack or presence of product quality.

  (15) a. HOERA 🥳🎉 mijn wifi werkt weer!
       b. HURRAY 🥳🎉 wifi is working again!

  (16) a. Surfen aan 512kbps... opnieuw 😒 is niet het "unlimited" 4G abonnement waar ik voor betaal.
       b. Surfing at 512kbps ... again 😒 does not meet the expected quality of the "unlimited" 4G plan I pay for.

  (17) a. De trein om 6:02u van Leuven naar Gent-St-Piet werd goed bevonden door reizigers: stiller en minder schokken :)

b. The train at 6:02 am from Leuven to Gent-St-Piet was approved by travelers: quieter and less shocks :)

– **Delays and cancellations**

All events characterized by some discrepancy between the actual time a service is performed or a product arrives and the time the customer expected. Most such cases are delays and cancellations, but we also decided to include services/product that are performed/arrive earlier then the time expected by the customer.

(18) a. De 1 van Hasselt naar Genk rijdt weer niet vandaag...

b. The 1 from Hasselt to Genk doesn't drive again today...

(19) a. Na 3 uur wachten in vliegtuig A, beslissen ze opeens om ons in een busje te steken naar vliegtuig B, waarna het wacht opnieuw begon #efficiëntenzo

b. After waiting 3 hours in plane A, they suddenly decide to put us in a van to plane B, after which the wait started again #efficientandall

(20) a. Echt waar, sta ik eens op tijd op het perron en dan beslist die trein er eerder van door te gaan 😡

b. Really, I'm on the platform on time and then the train decides to leave earlier 😡

– **Breakdowns**

Events that describe breakdowns in the main products or services provided by a certain company.

(21) a. Ja, het is weer zo ver: wifi ligt weeral uit.

b. Yep, the time has come again: WiFi is down again.

(22) a. @Company niet om te lachen, maar er zit een gat in het plafond van jullie tram en het regent hier dus binnen 🙃

b. @Company not to laugh, but there's a hole in the roof of the tram's ceiling and so it's raining inside 🙃

(23) a. De wifi is extreem traag vandaag en herstarten lijkt niet te werken... 😭

b. The wifi is extremely slow today and rebooting doesn't seem to work... 😭

– **Product information**

Events that are related to information about a company's product or changes to this information made by the company or caused by external factors over which the customer has no control (e.g., pandemic, bad weather). In addition, event descriptions in which a customer explicitly chooses to opt for another service/product are also included. Events falling under the "product information" category often have the potential to cause a customer to pose questions to a company. In this respect, it is very important to underscore the difference between this category (i.e., a mere label for an event description) and the "request information" intent of a customer. While an explicit question signals the intent of requesting information, that same question will probably not contain an explicit event description dealing with a shift in product information.

More precisely, events that fall under one of the following conditions should be labelled as "product information":

– Changes to the product/service provided by a company, or changes to similar products/services of competitors. This condition is illustrated in example sentences (24) and (25). Note that both sentences are part of a turn, and most turns contain more than one sentence. Hence, such event descriptions can be followed by questions such as *Geldt dat ook voor België?* ("Does this also apply to Belgium?"), *@Company hebben jullie jullie service aangepast door corona?* ("@Company have you adjusted your service due to corona?"), etc.

– Constructions such as *ik zie hier staan dat + evenement* ("I see here that you + event"), *ik heb vernomen dat + evenement* ("I heard that + event"), etc. An example of a sentence to which the following condition applies can be found in (26).

– Customer initiated change in which a customer opts for a different products/service (see example (27)).

– Events that trigger a customer to consider changing his/her service (see example (28)).

– Missing information or a discrepancy in the information provided by the company. The sentences in (29) and (30) are examples that fall under this final condition.

(24) a. In de UK worden alle vluchten naar Spanje geschrapt.

b. In the UK, all flights to Spain will be canceled.

(25) a. Ik heb al de hele week de trein genomen, en dit keer op keer zonder gecontroleerd te worden 🤔

b. I've been taking the train all week, and this time and time again without being checked 🤔

(26) a. Ik zie hier staan dat jullie tarieven recent veranderd zijn.

b. I see here that your rates have recently changed.

(27) a. Ik ben juist van abonnement X naar jullie nieuw datapakket Y overgeschakeld.

b. I just switched from subscription X to your new data package Y.

(28) a. Met de coronacrisis gebruik ik mijn NMBS abonnement niet meer.

b. Due to the corona crisis, I no longer use my NMBS subscription.

(29) a. Ik heb meermaals gezocht, maar ik vind nergens terug of de @Company abonnementen automatische verlengd worden en wanneer men deze mogelijks zou kunnen opzeggen.

b. I've searched several times, but I can't find any information about whether the @Company subscriptions are automatically renewed and when they could possibly be canceled.

(30) a. Ik heb recent een brief van jullie gekregen met de mededeling dat ik nog een openstaande factuur heb van 10, maar in de app staat dat deze reeds betaald is.

b. I recently received a letter from you informing me that I still have an outstanding invoice of 10, but the app says it has already been paid.

– **Digital design inadequacies**

Events related to flaws on a company's website, bugs in a certain update, or ill-working applications.

(31) a. Ik kan om de een of andere reden mijn openstaande rekening niet betalen via de app, maar mijn rekening moet wel morgen betaald zijn! 😰

    b. For some reason I cannot pay my outstanding bill via the app, but my bill must be paid by tomorrow! 😰

(32) a. Als ik mij wil aanmelden op jullie site, krijg ik constant een error message??

    b. If I want to register on your site, I constantly get an error message??

– **Environmental and consumer health**

Event description in which the customer addresses certain environmental or consumer health issues which seem insufficiently addressed by the company policies.

(33) a. Terwijl ik ga wandelen met mijn hond, moet ik hem beschermen van al het zwerfvuil dat in de buurt van jullie bushokjes ligt?! 😡

    b. While I go for a walk with my dog, I have to protect him from all the litter that lies near your bus shelters?! 😡

(34) a. Wat ik niet begrijp, is hoe men de coronamaatregelen kan respecteren op overvolle treinen en platformen #onverantwoord #datwasgeen1,5meter

    b. What I do not understand is how to respect the corona measures on overcrowded trains and platforms #unresponsible #thatwasnot1.5meters

– **Other**

All events that do not fall under one of the previously introduced cause categories, such as the loss of items, spam and fake phone calls, etc.

(35) a. Ik ben mijn rugzak kwijtgespeeld op de IC2556 naar Brugge gisteren...

    b. I lost my backpack on the IC2556 to Bruges yesterd...

(36) a. Ik heb zonet spam ontvangen van een e-mailadres dat verdacht veel op dat van jullie lijkt 🤔

    b. I just received spam from an email address that looks suspiciously like yours 🤔

Chapter 4

Emotions

Emotion annotations should only be made if the conversation is labelled as "subjective" (see Section 2.2). In case the conversation is "objective", annotators can skip this annotation layer. This chapter introduces two different types of annotation methods that complement each other. In Section 4.1, we will present a set of distinct emotion labels that have to be assigned to customer turns. These categorical annotations are accompanied with additional dimensional scores along valence-arousal-dominance (VAD) axes in Section 4.2.

As explained in Section 2.2, emotion can be expressed both explicitly and implicitly. If a tweet contains some explicit words that signal the presence of emotions, annotators should annotate these specific words. If this is not the case and it is rather the context itself that conveys emotions in an implicit manner, annotators should create an emotion annotation over the span of text containing that context. This can either be the entire customer turn or a (combination of) sentence(s) in the customer turn. The motivation behind the possibility to annotate at sentence level is that we want to acquire a better understanding of which parts of a turn contain which types of emotions. Such information is important as in customer turns one part of a turn might carry a particular emotion, but another part of the same turn might hold a different emotion.

4.1 Emotion categories

There exists a wide variety of different categorical frameworks that try to capture the spectrum of emotions. In his influential work on emotions and their relation to facial expressions, Ekman (1992) reports that there are six universal basic emotions: joy, surprise, anger, fear, disgust and sadness. Plutchik (1980) extends Ekman's schema with two additional emotions, namely trust and anticipation. While both frameworks are extensively adopted in NLP research, Skerry and Saxe (2015) remark that our ability to recognize emotional states is not limited to basic emotions, which causes them to introduce a larger set of 20 emotions that are inferred from contextual situations. Rashkin et al. (2019) notice that such situationally-inferred emotions are important in dialogue scenarios. For their research on empathetic dialogue generation, they released a novel dataset (viz. EmpatheticDialogues) containing 25k conversations that are grounded in emotional situations. Rashkin et al. (2019) employed 32 different emotion labels aggregated from different emotional taxonomies to create situation descriptions that are associated with a particular emotion, remarking that some of these emotions (e.g., *afraid* and *terrified*) are very closely related and can therefore be merged.

In our annotation scheme, we adopt the emotion labels that Demszky et al. (2020) used in their

creation of the GoEmotions dataset. This novel resource consists of 58k English Reddit comments, labelled with 27 emotion categories and an additional Neutral label. We opt for this framework, because Demszky et al. (2020) implemented a multi-step taxonomy selection procedure to maximize the coverage of emotions, while simultaneously limiting the overlap among emotions. Moreover, the researchers demonstrated that their dataset generalizes well to other domains and different sets of emotion categories Demszky et al. (2020).

This implies that our annotators can choose from 27 emotion labels and an additional Neutral label to annotate emotions in customer turns. The annotation task involves selecting all the different types of emotions the customer was experiencing at the time the tweet was written. Since we are working with conversations that often consist of more than one customer turn, it is perfectly possible that one customer turn contains emotions, while another customer turn in the same conversation does not contain any emotions. In that latter case, annotators should annotate the entire span of that turn with the Neutral label. It is important to remark that this Neutral label cannot be combined with any emotion label: a turn contains either emotions (selection from 27 emotion labels) or it is objective (assignment of the Neutral label).

Annotators can select emotions from the following taxonomy (qtd. from Demszky et al. (2020)):

- **Admiration** 👏 Finding something impressive or worthy of respect.
- **Amusement** 😂 Finding something funny or being entertained.
- **Anger** 😡 A strong feeling of displeasure or antagonism.
- **Annoyance** 😒 Mild anger, irritation.
- **Approval** 👍 Having or expressing a favorable opinion.
- **Caring** Displaying kindness and concern for others.
- **Confusion** 😕 Lack of understanding, uncertainty.
- **Curiosity** A strong desire to know or learn something.
- **Desire** A strong feeling of wanting something or wishing for something to happen.
- **Disappointment** Sadness or displeasure caused by the nonfulfillment of one's hopes or expectations.
- **Disapproval** 👎 Having or expressing an unfavorable opinion.
- **Disgust** 🤢 Revulsion or strong disapproval aroused by something unpleasant or offensive.
- **Embarrassment** 😳 Self-consciousness, shame, or awkwardness.
- **Excitement** 🤩 Feeling of great enthusiasm and eagerness.
- **Fear** 😨 Being afraid or worried.
- **Gratitude** 🙏 A feeling of thankfulness and appreciation.
- **Grief** Intense sorrow, especially caused by someone's death.
- **Joy** 😃 A feeling of pleasure and happiness.
- **Love** ❤️ A strong positive emotion of regard and affection.
- **Nervousness** Apprehension, worry, anxiety.
- **Optimism** 🤞 Hopefulness and confidence about the future or the success of something.

- **Pride** Pleasure or satisfaction due to ones own achievements or the achievements of those with whom one is closely associated.

- **Realization** Becoming aware of something.

- **Relief** Reassurance and relaxation following release from anxiety or distress.

- **Remorse** Regret or guilty feeling.

- **Sadness** 😢 Emotional pain, sorrow.

- **Surprise** 😲 Feeling astonished, startled by something unexpected.

Moreover, we added Table 4.1 (qtd. from Demszky et al. (2020)) to assist annotators in picking emotions. The proposed division of emotions in positive, negative and ambiguous sentiment can help annotators, as it reduces the number of emotion categories between which he or she can choose.[1] It is, however, important to emphasize that the proposed sentiment division is by no means fixed. For example, the emotion *desire* can also be negative, especially when someone repeatedly insists on a certain change that did not yet happen.

admiration	amusement	approval	caring	anger	annoyance	disappointment	disapproval	confusion
great (42)	lol (66)	agree (24)	you (12)	fuck (24)	annoying (14)	disappointing (11)	not (16)	confused (18)
awesome (32)	haha (32)	not (13)	worry (11)	hate (18)	stupid (13)	disappointed (10)	don't (14)	why (11)
amazing (30)	funny (27)	don't (12)	careful (9)	fucking (18)	fucking (12)	bad (9)	disagree (9)	sure (10)
good (28)	lmao (21)	yes (12)	stay (9)	angry (11)	shit (10)	disappointment (7)	nope (8)	what (10)
beautiful (23)	hilarious (18)	agreed (11)	your (8)	dare (10)	dumb (9)	unfortunately (7)	doesn't (7)	understand (8)
desire	excitement	gratitude	joy	disgust	embarrassment	fear	grief	curiosity
wish (29)	excited (21)	thanks (75)	happy (32)	disgusting (22)	embarrassing (12)	scared (16)	died (6)	curious (22)
want (8)	happy (8)	thank (69)	glad (27)	awful (14)	shame (11)	afraid (16)	rip (4)	what (18)
wanted (6)	cake (8)	for (24)	enjoy (20)	worst (13)	awkward (10)	scary (15)		why (13)
could (6)	wow (8)	you (18)	enjoyed (12)	worse (12)	embarrassment (8)	terrible (12)		how (11)
ambitious (4)	interesting (7)	sharing (17)	fun (12)	weird (9)	embarrassed (7)	terrifying (11)		did (10)
love	optimism	pride	relief	nervousness	remorse	sadness	realization	surprise
love (76)	hope (45)	proud (14)	glad (5)	nervous (8)	sorry (39)	sad (31)	realize (14)	wow (23)
loved (21)	hopefully (19)	pride (4)	relieved (4)	worried (8)	regret (9)	sadly (16)	realized (12)	surprised (21)
favorite (13)	luck (18)	accomplishment	relieving (4)	anxiety (6)	apologies (7)	sorry (15)	realised (7)	wonder (15)
loves (12)	hoping (16)	(4)	relief (4)	anxious (4)	apologize (6)	painful (10)	realization (6)	shocked (12)
like (9)	will (8)			worrying (4)	guilt (5)	crying (9)	thought (6)	omg (11)

Table 4.1: Top 5 words associated with each emotion (positive, negative, ambiguous). The rounded z-scored log odds ratios in the parentheses, with the threshold set at 3, indicate significance of association (qtd. from Demszky et al. (2020)).

4.2   Emotion dimensions

Alongside the categorical frameworks, dimensional models are also used to capture emotions. These latter models portray emotions as vectors in a multidimensional space. Hence, the task of emotion analysis becomes redefined as a regression problem in which a system should predict a real-valued score for each dimension. One of the most well-known dimensional models consists of three different orthogonal axes, namely Valence (pleasure - displeasure), Arousal (excited - calm), and Dominance (dominant - submissive) (Osgood et al., 1957; Mehrabian and Russell, 1974).[2] An even more widely used version of this framework consists only of Valence and Arousal dimensions (Russell, 1980), thus removing the Dominance dimension. Although categorical models are still predominantly employed for the task of emotion analysis (cf. Section 4.1), dimensional models have recently gained more recognition in the NLP community (Preoţiuc-Pietro et al., 2016; Buechel and Hahn, 2017; Wood et al., 2018). Buechel and Hahn (2017) underscore the potential of the VAD-model compared to categorical approaches, as these categorical

---

[1]In these guidelines, we did not explicitly translate the associated English words to Dutch.
[2]Henceforth, the Valence-Arousal-Dominance model will be abbreviated to VAD-model.

approaches suffer from a lack of consensus on the set of incorporated emotions to be used and the emotion labels themselves are often not equally distributed in the Valence-Arousal space.

We implement this VAD-model (Mehrabian and Russell, 1974) into our annotation scheme. Each selected text span in a customer turn receives therefore not only a categorical label, but it also gets three additional scores along five-point scales (for Valence, Arousal, and Dominance respectively). Valence indicates the positivity or negativity associated with the expressed emotion; Arousal denotes the energy or activity that is involved in expressing an emotion; and Dominance relates to the feeling of control and influence that one has Mehrabian (1996). We attached a pictographic aid to these guidelines to further assist annotators (see Figure 4.1). The tool is called the Self-Assessment Manikin (SAM) and its purpose is to visually assist annotators in selecting a score for each emotion dimension Bradley and Lang (1994). The three rows illustrate the Valence, Arousal, and Dominance axis respectively. Each emotion dimension and the corresponding scores for that dimension are thoroughly discussed below. We employed the same definitions and descriptions as the ones De Bruyne et al. (forthcoming) introduced in their annotation scheme.

- **Valence** (also often referred to as Pleasure or Evaluation)

  A continuum ranging from extreme pain or unhappiness to extreme happiness or ecstasy (e.g., cruelty, humiliation, disinterest and boredom versus excitement, relaxation, love and tranquility).

  TIP: try to capture the overall feeling of the customer.

  1 – very negative
  2 – rather negative
  3 – neutral
  4 – rather positive
  5 – very positive

- **Arousal**

  Denotes the level of mental alertness and physical activity or energy (e.g., sleep, inactivity, boredom and relaxation at the lower end versus wakefulness, bodily tension, strenuous exercise, and concentration at the higher end). It is important to remark that the arousal dimension is not equal to emotion intensity.

  TIP: try to imagine how the customer would say what he or she wrote.

  1 – very calm, monotonous, bored
  2 – rather calm
  3 – neutral
  4 – faster, a bit excited
  5 – fast, loud, dynamic voice use, gestures

- **Dominance**

  Relates to social position; the feeling of control and influence over one's surroundings versus the feeling of being controlled or influenced by situations and others (e.g., relaxation, power, and boldness versus anxiety, infatuation, fear, and loneliness). Note that this applies to the feeling one has after a situation emerged. After all, not that many situations occur in which one is in complete control. However, it is possible to have a feeling of dominance after something happened that was out of a someone's control (e.g., yelling out of anger about a stolen phone). Likewise, it is possible to feel submissive after a similar event (e.g., crying because the phone really meant a lot to you).

  TIP: try to visualize the pose of the customer (if he or she sits (not) just behind a computer).

1 – cringed, small, feeling of being overpowered, very insecure

2 – introvert posture, rather insecure

3 – neutral

4 – more open posture, rather self-confident
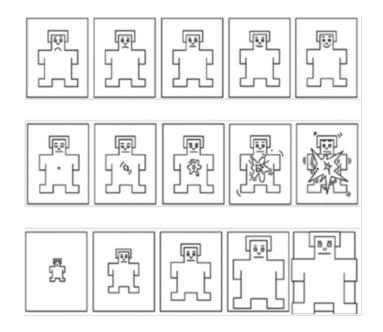
5 – standing up, feeling of self-confidence and power

Figure 4.1: The Self-Assessment Manikin (SAM) to assist annotators in assessing emotions along three affective spaces.

To conclude this section, a couple of important remarks should be made:

- A dialogue that is annotated as subjective on the conversation level (cf. Section 2.2) can still contain certain turns that are Neutral, i.e., turns do not contain any emotions. If a customer turn is categorically labelled as Neutral, annotators should not assign any VAD scores to this turn. The blank fields will in this particular case automatically be converted to a score of 3 (viz., a neutral score) once the annotations are processed.

- While Valence changes quite easily during a conversation, Arousal and especially Dominance seem to be somewhat harder to change with a small number of conversational turns. This depends obviously on the conversational context and setting. For example, a customer can become less aroused when he or she is helped immediately. A customer can also feel helpless at the start of a conversation (e.g., recurrent problem and his/her problem is consistently ignored by the company), but become very dominant and take matters into his/her own hands by deciding to leave the company. Nevertheless, when annotators are in doubt about the exact VAD score for a certain text span, they can check surrounding turns and their annotations for VAD scores.

- All explicit formulations of the emotion gratitude (e.g., *dank u, dankuwel, thx, bedankt* ("thank you")) are labelled with the corresponding emotion category Gratitude. In most cases, this gratitude is either very sincere or out of courtesy. In some other cases, gratitude is expressed to be ironic (e.g., *Amai, bedankt om mij tweemaal te veel aan te rekenen! Afzetters!!!* ("Thanks for overcharging me twice! Ripoffs!!!")). These differences in meaning should be made explicit through VAD scores. For example, neutral gratitude is indicated as V=3, ironic gratitude can be expressed as V=2 or V=1, and sincere gratitude is annotated with a V=4 or V=5.

4.3 Annotation examples

The following examples are meant to illustrate (i) how annotators should select text spans for annotating explicit and implicit emotions, and (ii) how emotions can be annotated along the proposed categorical and dimensional framework. All examples consist of customer turns. We refer to Chapter 7 for examples of fully-annotated dialogues. Finally, we would like to emphasize that annotating emotions is a highly subjective task in which there are no absolute right or false answers. The annotations below are one possible solution for this annotation task, but other similar annotations are also possible.

(37) a. Hallo 🙂 Ik heb jullie al een aantal keer proberen bereiken (zowel via mail als telefonisch) i.v.m. de terugbetaling van mijn geannuleerde vlucht. Zouden jullie mij hier kunnen verderhelpen?

    b. Hello 🙂 I have already tried to reach you a number of times (both by mail and by telephone) concerning the refund of my canceled flight. Could you help me here?

    → Although smileys often signal the presence of emotions, annotators should be aware that this is not always the case. The smiley in this example is used for punctuation, not to convey the emotion Joy. The span of the entire turn has therefore to be labelled with the category Neutral, as it is an objective turn that does not carry any emotions. Hence, no additional annotation for VAD dimensions should be made.

(38) a. Mijn aankoop is op mijn rekening in mindering gebracht, maar op jullie site zie ik geen verandering en ik heb ook geen bevestiging gekregen... Hoe kan dit nu??! Moet ik opnieuw betalen, heb ik zelf iets fout gedaan of ligt het probleem bij jullie??

    b. My purchase has been deducted from my bank account, but I don't see any change on your site and I have not received any confirmation... How is this possible??! Do I have to pay again, did I do something wrong myself or does the problem lie with you?

    → This turn implicitly conveys the emotion Confusion. The text span to annotate is *Hoe kan dit nu??! Moet ik opnieuw betalen, heb ik zelf iets fout gedaan of ligt het probleem bij jullie??*. The Valence score for the emotion is either 3 or 2 (neutral to slightly negative sentiment about what happened), the Arousal score is 4 (additional question marks, agitation), and the Dominance score is 2 (client does not understand what is wrong, has no control over the situation).

(39) a. Dit is echt DE LAATSTE KEER OOIT! Ik heb genoeg van dat veel beloven maar niets presteren, BENDE LUIE ZAKKEN!! Awel, maar goed dat ik volgende maand overstap naar @Company! Tot nooit meer xoxo

    b. This really is THE LAST TIME EVER! I'm tired of that promising a lot but not achieving anything, GANG OF SLACKERS!! Good thing I am switching to @Company next month! Until never again xoxo

    → This turn displays rather explicit emotions. *DE LAATSTE KEER OOIT!* and *BENDE LUIE ZAKKEN!!* are expressions of Anger, with Valence equal to 1 (very negative), while Arousal is equal to 5 (capitalized words, exclamation mark), and Dominance is also equal to 5 (remark that the customer is in control and decided to switch to another company). In addition, *BENDE LUIE ZAKKEN!!* is also an expression of Disgust (same VAD scores as the Anger case). *Ik heb genoeg van* is an explicit expression of Disapproval (V=1, A=4, D=5), while *maar goed dat* contains Approval/Optimism (V=3 or 4, A=4, D=5).

Chapter 5

Response strategies

Similar to the previous chapter, annotations introduced in this annotation layer should only be made if the conversation is labelled as "subjective" (see Section 2.2). In case the conversation is "objective", the annotation layer can be skipped. Since customer turns receive emotion annotations, we also want to annotate emotions in the agent/employee turns. Rafaeli and Sutton (1987) remark that emotions expressed by company agents are typically governed by that company's policies. We will therefore work with "response strategies" instead of emotions. To our knowledge, Herzig et al. (2016) were the firsts to propose four emotional techniques to label the techniques an employee utilize to respond to a customer, namely apology, cheerfulness, empathy, and gratitude. We decided to add four additional labels that are less related to emotions, but that also have to potential to mitigate customer emotions (e.g., an explanation might be sufficient to reassure a customer). Concretely, we added explanation, help offline, request information, and other to the framework.

Response strategy annotations are made at the sentence level. In most cases, one label per sentence is sufficient to annotate all response strategies in a turn. It is, however, possible that a sentence contains two response strategies (e.g., *Dankje, kan ik je nog met iets verderhelpen?* ("Thank you, can I still help you with something else?") = gratitude; request information). If this occurs, annotators are allowed to deviate from the sentence level and annotate at clause level. The following part contains a definition of the different response strategies and some examples:

- **Apology**
  Denotes all types of apologies and regrets an employee makes towards a customer.

  (40) a. Onze excuses voor dit ongemak.
        b. We apologize for this inconvenience.

  (41) a. Het spijt me van dit te horen... 😔
        b. I'm sorry to hear about this... 😔

- **Cheerfulness**
  Describes all cheerful, happy and funny responses made by the employee. These sentences often contain emojis.

  (42) a. Veel succes nog! Wij duimen alvast! 🤞
        b. Good luck! We keep our fingers crossed! 🤞

(43) a. Graag gedaan hoor 😃

  b. You're welcome 😃

(44) a. Super van dit te horen! Ik wens je nog een hele fijne dag toe met de goede wifi!!

  b. Great to hear about this! I wish you a very nice day with the good WiFi!

– **Empathy**

All sentences in which an employee empathizes with the customer or gives an empathic response by understanding, acknowledging and mirroring the emotion the customer is experiencing.

(45) a. Jammer van dit de horen, dat was zeker niet de bedoeling 😌

  b. Too bad to hear this, that was certainly not the intention 😌

(46) a. Dat is inderdaad niet fijn!

  b. That is indeed not nice!

(47) a. Dat klinkt niet goed.

  b. That doesn't sound well.

(48) a. We begrijpen dat dit vervelend is.

  b. We understand this is annoying.

– **Explanation**

All explanations and motivations behind certain actions, events, company policies, etc. fall.

(49) a. Dat kan zeker. Je mag meerdere abonnementen of Campuskaarten hebben, zo lang je school of universiteit maar de bestemming is.

  b. You certainly can. You may have multiple subscriptions or Campus cards, as long as your school or university is the destination.

(50) a. Binnen 6 minuten zal de tram er zijn.

  b. The tram will be there within 6 minutes.

(51) a. Ik geef de boodschap door aan de juiste werknemer.

  b. I pass the message on to the right employee.

(52) a. Indien er nog iets niet werkt, kan je best eens je gsm heropstarten.

  b. If something still doesn't work, you can restart your mobile phone.

(53) a. Meer informatie over deze actie kan je terugvinden op onze webiste: @website.

  b. More information about this promotion can be found on our website: @website.

– **Gratitude**

Relates to every sentence in which an employee expresses his or her gratitude to the customer. Responses to gratitude of a customer (e.g., Graag gedaan!) fall under the "cheerfulness" category.

(54) a. Dankjewel om ons dit te laten weten 🙂

  b. Thank you for letting us know 🙂

(55) a. Dank u voor uw feedback.

b. Thank you for your feedback.

- **Help offline**

  This label should be used when one of the following conditions applies:

  (i) an employee asks the customer to send him a private message/DM to solve the issue;

  (ii) an employee confirms that he or she sent a private message/DM to the customer;

  (iii) an employee makes the offer that the customer can always contact him or her again if any questions remain.

  The following sentences are some examples that fall under the help offline category.

  (56) a. Stuur je je gsmnummer in DM door? Dan kijk ik dit even na voor jou.

  b. Do you forward your mobile number in DM? Then I'll check this for you

  (57) a. Zou je ons je gegevens kunnen doorsturen via @website?

  b. Could you send us your details via @website?

  (58) a. Ik heb je een privébericht gestuurd.

  b. I sent you a private message.

  (59) a. Mocht je nog vragen hebben, aarzel dan zeker niet om ons opnieuw te contacteren!

  b. If you have any questions, don't hesitate to contact us again!

  (60) a. Laat ons maar iets weten, we volgen het samen verder op!

  b. Just let us know, we will follow it up together!

  (61) a. Je kan dit best melden via onze app. Onze collega help jou dan verder!

  b. You can best report this via our app. Our colleague will help you further!

  (62) a. Op onze website kan je onze klacht achterlaten: @website... We onderzoeken dit dan verder voor jou.

  b. You can leave our complaint on our website: @website... We will then investigate this further for you.

- **Request information** All sentences in which an employee asks the customer for more information in order to better understand or solve an issue. Remark that not all questions fall under this category, as questions related to offering offline help should be labelled as "help offline".

  (63) a. Kan je ons de link doorgeven waar je dit hebt gevonden?

  b. Can you give us the link where you found this?

  (64) a. Heb je het ongemak ook op andere browsers?

  b. Do you have the inconvenience on other browsers as well?

  (65) a. Kan je je gsm eens aan en uitzetten?

  b. Can you turn your mobile on and off?

- **Other**

  Denotes all sentences that do not fall under one of other categories described above.

  (66) a. https://...

b. https://...

(67) a. Groetjes, Bob

b. Kind regards, Bob

We label these two examples with "other" when no additional text is written in a turn. If, however, *Groetjes Bob* follows some other sentence(s) in the turn, then *Groetjes Bob* can simply be added to the text span of the previous sentence and thus receive the same label as the previous sentence.

Chapter 6

Annotation procedure

This chapter serves as an overview of the annotation procedure, as it combines the different isolated annotation layers that were described in the previous chapters. Moreover, a flowchart (see Figure 6) was created to assist the annotators by visually illustrating how the different layers should be combined. Each document in the INCEpTION environment contains one dialogue, while each line in that document corresponds to one tweet/turn.

At the start of the annotation procedure, one should first apply the conversation characteristics layer to the dialogue (see Chapter 2), as this is the most global annotation layer. Decisions made at this layer will determine the course of the procedure. In the conversation characteristics layer, one first has to inspect if the conversation is in Dutch and if the conversation has enough context to make sense of it. If either of these two conditions is not met, the conversation receives an "out of scope" label and the rest of the annotation procedure is aborted. Otherwise, the procedure continues and annotators have to indicate whether the conversation is "subjective"/"objective" and whether an explicit event description (potentially) causing emotions is present.

In a second step we continue with the next layer, namely the cause layer (see Chapter 3). If annotations in the first layer indicate that the dialogue contains an explicit event description, this layer has to be applied. Otherwise this layer can be skipped. In the former case, one should search for a sentence describing the cause. That sentence has to be selected and a corresponding cause label has to be assigned to it. Each conversation can at most have one cause annotation.

The third and fourth layer in this procedure also build on the first layer. If a conversation is "objective", both layers can be omitted and the annotation procedure is therefore finished. If a conversation is "subjective", we will continue with the third layer, namely the emotion annotations (see Chapter 4). For this layer, one has to annotate emotions in customer turns along a categorical and a dimensional framework. Each selected text span receives four annotations: one category for the emotion class, and three scores along a five-point scale representing Arousal, Dominance, and Valence respectively.[1] If a text span contains several emotions, the same text should be annotated several times with each of the emotions. Finally, the fourth annotation layer can be executed. This layer is designed to extract and label the different response strategies in employee turns (see Chapter 5). Each sentence in an employee turn should be assigned to a certain response strategy. In case a sentence has more than one response strategy, annotators are allowed to annotate these strategies at clause level instead of sentence level. Once all sentences in the employee turns are labelled, the annotation procedure is finished and one can proceed to the next dialogue.

---

[1]Note that the order is not VAD, but ADV. The reason for this is that in the INCEpTION interface, the different dimensions are placed in alphabetical order. The same applies for the examples in Chapter 7.
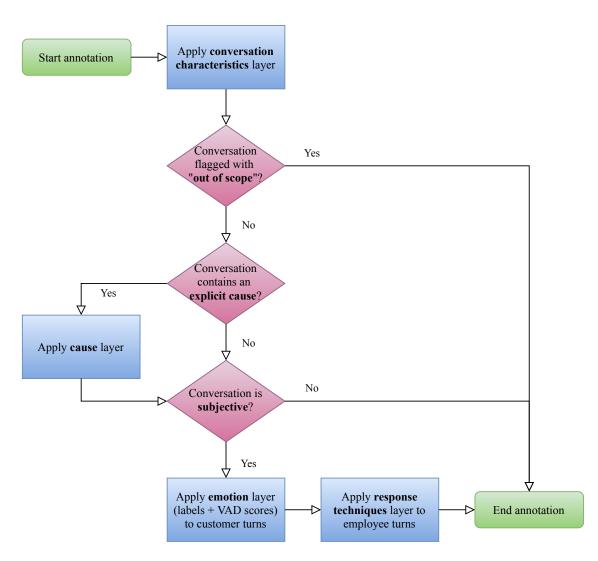
Figure 6.1: Flowchart illustrating the procedure to annotate fine-grained emotion trajectories in customer service dialogues.

Chapter 7

INCEpTION annotation examples

In this final chapter, we briefly discuss four dialogues that were annotated in the INCEpTION environment. Each Dutch dialogue is accompanied with a English translation that is annotated in the same manner.

Figure 7.1: Example dialogue 1 in Dutch.

Figure 7.2: Example dialogue 1 translated to English.

– Dialogue 1 is **subjective** and contains an **explicit cause** description

– *@Company De wifi laat het hier soms echt afweten...* 😓 = **Breakdowns**

– 😓 = **Annoyance, A=4, D=3, V=2**

– *Super* = **Approval, A=4, D=3, V=5**

– *Echt heel erg bedankt* = **Gratitude, A=4, D=3, V=5**

– *hebben mijn dag gemaakt!* 😊 = **Joy, A=4, D=3, V=5**

– *@Customer Goedemorgen, we hebben je net een DM gestuurd; Indien je in de toekomst nog storingen hebt, mag je ons altijd contacteren* = **Help offline**

– *@Customer Fijn te horen dat je probleem is opgelost.* = **Empathy**

Figure 7.3: Example dialogue 2 in Dutch.

Figure 7.4: Example dialogue 2 translated to English.

– Dialogue 2 is **subjective** and contains an **explicit cause** description

– *@Company nog steeds geen information over mijn vluch SAL-BRU* = lack of **product information**

– *nog steeds geen* = **Annoyance, A=4, D=2, V=2**

– *flut* = **Disapproval, A=4, D=2, V=2**

– *Dat is een leugen! ; Laatste keer ever! En wees maar zeker dat ik dit vertel aan mijn omgeving* 😡 = **Anger, A=5, D=2, V=1**

– *creperen* = **Disgust, A=5, D=2, V=1**

– *Onze excuses voor de vertraging ; en we willen ons oprecht excuseren voor de bijkomende ongemakken.* = **Apology**

– *Jullie vlucht staat gepland te vertrekken om 11u30 vanuit Sal naar Brussels Airport ; @Customer Geen leugens, maar helaas zijn er onverwachte technische problemen opgedoken ; We hadden jullie ook al liever onderweg gezien, maar op vlak van veiligheid nemen we geen risico's, vandaar de vertraging ; Voorlopig hebben we ook nog geen exact vertrekuur doorgekregen ; Eens we hier meer over weten, houden we jullie zeker op de hoogte. 2/2 =* **Explanation**

– *We begrijpen dat een dergelijke lange vertraging geen pretje is =* **Empathy**

Figure 7.5: Example dialogue 3 in Dutch.

Figure 7.6: Example dialogue 3 translated to English.

– Dialogue 3 is **subjective** and contains an **explicit cause** description

– *@Company Wanneer de conducteur zo fun en enthusiast is dat ik instant een lach op mn gezicht krijg* 😍 = **Employee service**

– *fun en enthusiast =* **Joy, A=5, D=4, V=5**

– 😍 =**Joy, A=5, D=4, V=5**

– *Steek hem maar een hoop pluimen op zijn hoed! =* **Admiration, Caring and Gratitude with A=5, D=4, V=5**

- *WOOOW* = **Surprise, A=5, D=4, V=5**

- *chill* = **Approval, A=5, D=4, V=5**

- *Kunnen jullie anders ook regelen dat hij vanaf nu iedere ochtend die trein begeleidt?* = **Optimism, A=5, D=4, V=4**

- *Jammeeeer!* = **Disappointment, A=5, D=4, V=2**

- *Nogmaals een dikke merci!* = **Gratitude, A=5, D=4, V=5**

- *Zou je ons misschien kunnen meedelen op welke treinverbinding dit was?* = **Request information**

- *Dan zorgen wij ervoor dat je wensen persoonlijk aan hem worden overgemaakt 😉 ; Dat eerste kunnen we zeker, over dat laatste hebben wij helaas zelf geen controle. Dat wordt door de planning geregeld* = **Explanation**

- *@Customer Dat is zeker graag gedaan! Geniet jij nog van dat zonnig dagje vandaag 😎* = **Cheerfulness**

Figure 7.7: Example dialogue 4 in Dutch.

Figure 7.8: Example dialogue 4 translated to English.

- Dialogue 4 is **subjective** and contains an **explicit cause** description

- *Rechtzetting: @Company heeft het bedrag nu teruggezet na mijn klacht (en zelfs naar boven afgerond!)* = **Employee service**

- *en zelfs naar boven afgerond!* = **Excitement, A=4, D=4, V=5**

- *Als het goed is mag het ook gezegd worden! ; 👍* = **Approval, A=4, D=4, V=5**

- *Bedankt* = **Gratitude, A=4, D=4, V=5**

- *@Customer Graag gedaan!* = **Cheerfulness** *Foutjes gebeuren soms in het systeem, maar we doen er altijd ons best aan om deze zo snel mogelijk op te lossen* = **Explanation**

- *Laat zeker weten als je nog eens problemen zou hebben! Groetjes* = **Help offline**

Bibliography

Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. Towards a Human-like Open-Domain Chatbot, 2020.

Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59, 1994.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners, 2020.

Sven Buechel and Udo Hahn. Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation. In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 1–12. Association for Computational Linguistics, 2017.

Mary Culnan, Patrick McHugh, and Jesus Zubillaga. How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value. *MIS Quarterly Executive*, 9(4):243–259, 2010.

Luna De Bruyne, Orphée De Clercq, and Hoste Veronique. Annotating affective dimensions in user-generated content. *Language Resources and Evaluation*, forthcoming.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. GoEmotions: A Dataset of Fine-Grained Emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020.

Paul Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200, 1992.

Sarah E. Finch, James D. Finch, Ali Ahmadvand, Ingyu, Choi, Xiangjue Dong, Ruixiang Qi, Harshita Sahijwani, Sergey Volokhin, Zihan Wang, Zihao Wang, and Jinho D. Choi. Emora: An Inquisitive Social Chatbot Who Cares For You, 2020.

Keith Goffin. Customer support: A cross-industry study of distribution channels and strategies. *International Journal of Physical Distribution & Logistics Management*, 29(6):374–397, 1999.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. Event-Driven Emotion Cause Extraction with Corpus Construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1639–1649. Association for Computational Linguistics, 2016.

Anders Gustafsson, Michael D. Johnson, and Inger Roos. The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention. *Journal of Marketing*, 69(4):210–218, 2005.

Jonathan Herzig, Guy Feigenblat, Michal Shmueli-Scheuer, David Konopnicki, Anat Rafaeli, Daniel Altman, and David Spivak. Classifying emotions in customer support dialogues in social media. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 64–73. Association for Computational Linguistics, 2016.

Jan Hofmann, Enrica Troiano, Kai Sassenberg, and Roman Klinger. Appraisal Theories for Emotion Classification in Text. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 125–138, Barcelona, Spain (Online), 2020. International Committee on Computational Linguistics.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics, 2018.

Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. A Text-driven Rule-based System for Emotion Cause Detection. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 45–53. Association for Computational Linguistics, 2010.

Zhaojiang Lin, Peng Xu, Genta Indra Winata, Farhad Bin Siddique, Zihan Liu, Jamin Shin, and Pascale Fung. CAiRE: An Empathetic Neural Chatbot, 2020.

Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, 14:261–292, 1996.

Albert Mehrabian and James A. Russell. *An Approach to Environmental Psychology*. M.I.T. Press, 1974.

Sunil Mithas, Mayuram S. Krishnan, and Claes Fornell. Why Do Customer Relationship Management Applications Affect Customer Satisfaction? *Journal of Marketing*, 69(4):201–209, 2005.

Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum. *The measurement of meaning*. University of Illinois Press, 1957.

Robert Plutchik. A general psychoevolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, *Theories of Emotion*, pages 3 – 33. Academic Press, 1980.

Daniel Preoţiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. Modelling valence and arousal in Facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 9–15. Association for Computational Linguistics, 2016.

Anat Rafaeli and Robert I. Sutton. Expression of emotion as part of the work role. *The Academy of Management Review*, 12(1):23–37, 1987.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy, 2019. Association for Computational Linguistics.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M. Smith, Y-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot, 2020.

James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39: 1161–1178, 1980.

Klaus R. Scherer, Angela Schorr, and Tom Johnstone. *Appraisal processes in emotion: theory, methods, research*. Series in Affective Science. Oxford University Press, USA, 2001.

Françoise Simon. The influence of empathy in complaint handling: Evidence of gratitudinal and transactional routes to loyalty. *Journal of Retailing and Consumer Services*, 20(6):599–608, 2013.

Amy E. Skerry and Rebecca Saxe. Neural representations of emotion are organized around abstract event features. *Current Biology*, 2015.

Kevin J. Trainor, James (Mick) Andzulis, Adam Rapp, and Raj Agnihotri. Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. *Journal of Business Research*, 67(6):1201–1208, 2014.

Ian Wood, John P. McCrae, Vladimir Andryushechkin, and Paul Buitelaar. A comparison of emotion annotation schemes and a new annotated data set. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 1197–1202. European Language Resources Association (ELRA), 2018.

Rui Xia and Zixiang Ding. Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1003–1012. Association for Computational Linguistics, 2019.