CHAPTER 5

Appraisal Framework for Clinical Empathy: A Novel Application to *Breaking Bad News* Conversations

In Part I, we scrutinized the research on empathy across the NLP field and identified limitations in empathy construct definitions and operationalizations. We observed in the previous chapter that the more specific the construct operationalizations the better, but there is a lack of data and models reflecting them. In Part II, we shift our focus to certain applied settings, focusing on a broader challenge of shedding light on support-oriented interactions and understanding the dynamics of various forms of communication. We aim to adopt theory-driven operationalizations and computational models to gain insight into human-to-human supportive interactions.

We begin Part II with a focus on doctor-patient conversations and clinical empathy. Empathy is essential in healthcare communication. Studying conversational empathy from a computational perspective may provide insight into the complex dynamics that are difficult to identify manually. Computational models could contribute a better understanding of these dynamics, which can be leveraged in pedagogical tools for communication skills training [16]. To this end, there is a need for systematic representations of these conversational dynamics, which involves a model of discourse elements that constitute empathy and support-oriented interactions. The work in this chapter aims to address the lack of such representations with a conceptual model of empathic discourse elements for the clinical domain, applied to *breaking bad news* (BBN) conversational scenarios. Effective communication of empathy in such situations not only fosters understanding between patients and clinicians but also influences treatment outcomes and patient satisfaction. Ultimately, this work aligns with objectives of improving empathy communication in medical encounters and enhancing the training of medical students in breaking bad news scenarios.

This chapter focuses on the following research question:

RQ3

How can we integrate theory and linguistics-based frameworks into the conversational setting of *breaking bad news* dialogues?

To address RQ3, we introduce a novel annotation scheme for clinical empathy communication in

BBN conversations drawing on established theoretical and pedagogical frameworks. This scheme aims to identify key discourse elements and their dynamic interactions that constitute empathic successes and failures during these exchanges. Additionally, it provides a structural representation of empathic conversations, complementing established pedagogical methods for training communication skills in medical education.

We construct Empathy in BBNs, a span-relation task dataset of simulated BBN conversations in German, using our annotation scheme, in collaboration with a large medical school to support research on educational tools for medical didactics. The annotation is based on 1) Pounds [14]'s appraisal framework for clinical empathy, which is grounded in systemic functional linguistics, and 2) the SPIKES protocol for breaking bad news [260], commonly taught in medical didactics training. This approach presents novel opportunities to study clinical empathic behavior and enables the training of models to detect causal relations involving empathy, a highly desirable feature of systems that can provide feedback to medical professionals in training. We present illustrative examples, discuss applications of the annotation scheme, and insights we can draw from the framework.

The key contributions of this chapter are:

- We develop a novel annotation framework for capturing discourse elements and relations of empathic interactions in clinical scenarios based on an appraisal framework for clinical empathy grounded in systemic functional linguistics. Furthermore, we integrate this framework with an established protocol of clinical strategies and structural stages of *breaking bad news* conversations.

- We produce a dataset of German breaking bad news conversations annotated with the framework, which contributes to a gap in empathy language resources in non-English languages and enables investigations into empathic conversational strategies for clinical settings.

This chapter is based on the following publication [43]:

- **Allison Lahnala**, Béla Neuendorf, Alexander Thomin, Charles Welch, Tina Stibane, Lucie Flek. 2024. "Appraisal Framework for Clinical Empathy: A Novel Application to Breaking Bad News Conversations." In Proceedings of The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. European Language Resources Association.

The chapter is structured as follows. Section 5.1 introduces the work, providing an overview of the objectives and methods. Section 5.2 reviews related work concerning digital tools for empathic communication training and development datasets. Section 5.3 describes the annotation scheme, including the background on the theoretical frameworks and motivations for applying them for an NLP dataset. Section 5.4 describes the development of the dataset, and data statistics. Section 5.5 offers a discussion of the work, describing the information the scheme would provide for an example BBN scenario, and presenting future work. Section 5.6 summarizes the chapter.

5.1 Introduction

Empathy in medical encounters is considered a core element to high-quality patient care and an important skill to develop in medical training [15, 17]. Theoretical models of *clinical empathy* suggest it fosters more open patient-clinician communication for more deeply understanding patients and their conditions, providing valuable information for diagnosis and addressing therapeutic needs. In turn, this leads to better treatment strategies and adherence, therapeutic outcomes, and higher patient satisfaction [12, 13, 14], and proper empathy can be intrinsically therapeutic [11].

Empathy is crucial to *breaking bad news* conversations (BBNs), scenarios where a clinician must inform the patient about life-altering circumstances. Clinicians frequently must deliver bad news to patients, a high-stress and complex communication task [260]. Many medical students must pass formal BBN training, an area where digital tools have the potential to support students via automated feedback and practice with virtual standardized patients [23, 24, 25]. Models informed by clinical empathy could be made more transparent and explainable, leading to higher quality feedback on empathic responses or suggestions via an interface that suits their learning needs [16, 26]. Natural language processing (NLP) researchers are currently exploring models for automatic empathy detection and generation, which are essential for such tools.

Objectives

We introduce a new annotation scheme for clinical empathy communication in patient-clinician conversations, guided by three key objectives. First, we aim to identify precise discourse elements and their dynamic interactions that constitute empathic successes and failures during these exchanges. Second, we aim to provide a novel structural representation of empathic conversations that can support addressing existing shortcomings of NLP models for empathy, which struggle in identifying finer-level empathy components and utilizing the broader conversational context necessary for accurate evaluation [252]. Third, we aim to complement established pedagogical methods for training communication skills for breaking bad news conversations by integrating SPIKES, a protocol for breaking bad news commonly taught in medical didactics training [260].

Method Overview

We address these goals through a *span-relation* labeling method. Central to the scheme is identifying interactional sequences formally defined by Suchman, Markakis, Beckman, and Frankel [11]'s *model of empathic communication in medical conversations*. These sequences encompass three elements: 1) *empathic opportunities* (EO) in patient turns, 2) *elicitations* of such opportunities, and 3) *empathic responses* within clinician turns. We label specific spans of patient and clinician turns containing one of these three elements according to Pounds [14]'s *appraisal framework for clinical empathy* (AF) which defines types of EOs and responses based on linguistic aspects (see §5.3.1).

The scheme provides two further innovations. After labeling the spans, we create relations between identified empathic opportunities and the identified elicitations and empathic responses that correspond to them. Figure 5.1 illustrates spans and relations in an example portion of a BBN dialogue. Furthermore, we incorporate the structure of BBN conversations by labeling the six stages of BBN conversations defined by the SPIKES protocol within the clinician turns (see §5.3.2). This integration aligns the appraisal framework with the stages of BBN conversations, facilitating the examination of communicative strategies employed at each stage and informing the development of NLP models for digital training tools.

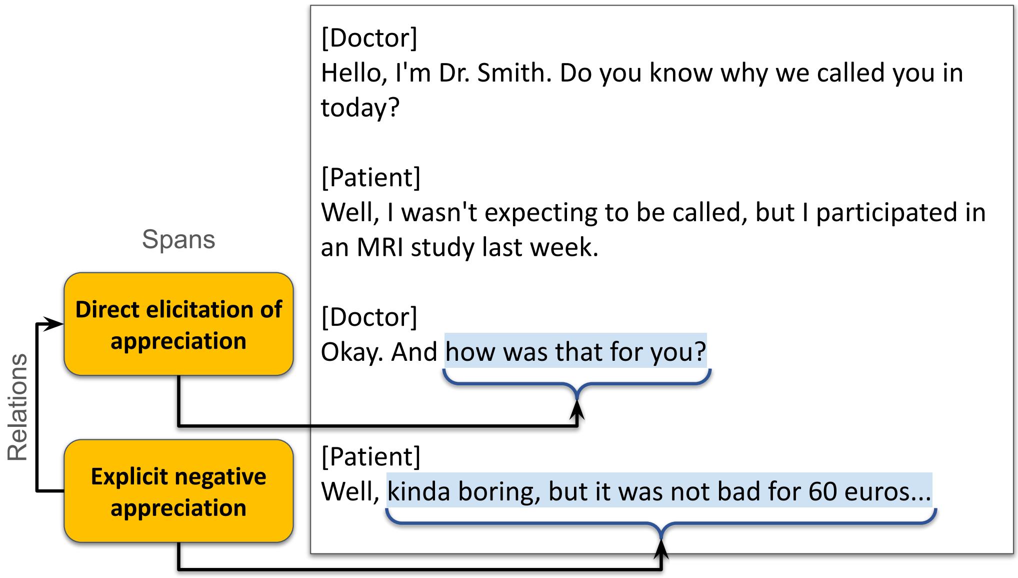

Spans
Relations
Direct elicitation of appreciation
Explicit negative appreciation
[Doctor]
Hello, I'm Dr. Smith. Do you know why we called you in today?
[Patient]
Well, I wasn't expecting to be called, but I participated in an MRI study last week.
[Doctor]
Okay. And how was that for you?
[Patient]
Well, kinda boring, but it was not bad for 60 euros...

Figure 5.1: Example showing spans highlighting an elicitation of an empathic opportunity, followed by a highlighted empathic opportunity span, and a relation connecting the empathic opportunity to the elicitation that prompted it.

The BBN Empathy Dataset

Finally, as part of a collaboration between NLP researchers and medical didactics experts at a large medical school to support research on educational tools for medical didactics, we construct a dataset of BBN conversations annotated with the new scheme. The dataset contains practice BBN conversations between medical students and standardized patients and fine-grained annotations of the components of empathic interactions. The BBN Empathy Dataset is the first dataset to contain discourse labels and relations for clinical empathy, which we make available for other researchers.

Summary of Contributions

We contribute 1) an innovative annotation scheme for clinical empathy 2) made available on an open-source platform for other researchers (§5.3), and 3) a public dataset of annotated BBN dialogues (§5.4).[1]

5.2 Related work

Digital tools have the potential to support training medical professionals in empathetic communication in ways such as offering communication practice with virtual standardized patients [23, 24, 25], assessing the quality of empathetic responses, and providing feedback on empathic responses or suggestions via an interface that suits their learning needs [16, 26]. These tools require effective NLP models of empathic language and conversational behaviors. In NLP, there is current momentum toward models for empathy detection and generation. Recognition is the task of determining the presence [5,

[1] Dataset and code: https://github.com/caisa-lab/BBN-Empathy

Patient Role: Empathic Opportunities	
Explicit	**Implicit**
Feelings	
Describing or exhibiting • emotion quality: "I'm sad" • emotive behavior "I cried" or "I laughed" • mental state: "I'm in pain" or "I feel alone"	May occur through expression of judgement or appreciation. Implicitly expresses feeling or perception by referring to • negative experiences: "My aunt had the same condition. She was in a lot of pain, and she didn't make it" (fear) • critical life stages and experiences: "Everything was going well...I just started my master's thesis" (surprise/disbelief)
Appreciation	
Attitude or perception toward things, events, actions, and behaviors, e.g. • event: "The MRI was boring," • thing: "The medication is not helping me."	Indirectly conveys attitude or perception toward things, events, actions, and behaviors that a clinician may infer and explore • thing: "My symptoms don't seem to improve with the medication."
Judgement	
Attitude or perception toward • self: "I'm not good at medication adherence" • others: "The nurses weren't helpful"	Indirectly conveys attitude or perception toward • themselves: "I'm not very consistent about taking my medication" • others: "The nurse had to poke me several times to withdraw blood"

Table 5.1: Description and examples of explicit and implicit functions (feeling, appreciation, judgement) in patients, representing empathic opportunities.

144] or degree of empathy [27] or subtypes of empathic behaviors [122, 123]. Detection models can be designed to evaluate empathic language and identify improvement areas to provide feedback [10]. Generative models attempt to generate a text response that is empathetic to a conversational partner. Research on generative models has focused on applications to open-domain dialogue [1], customer support [92], and counseling [28]. Generative models can be designed to deliver feedback and provide suggestions for empathetic responses. For example, Sharma et al. (2021) [204] developed a model for "empathic rewriting" to provide suggestions that increase the level of empathy in a given text, an approach that could support students in reflecting on ways to improve their empathic communication.

Despite the progress, current shortcomings in this research include poor operationalization of empathy, tending to employ only abstract notions focused on emotional aspects and overlooking cognitive and behavioral aspects, and a lack of empathic language resources that incorporate these dimensions. These issues are exasperated by lacking measurements with construct validity [39]. In turn, models trained on widely available datasets, such as the empathic dialogues dataset which grounds empathetic engagement in specific emotional situations, could help reveal patterns of emotional understanding, but this is only one facet of the empathy concept [255]. Thus, such models are limited in providing detailed and reliable assessments, explanations of the relation between EOs and empathic responses, or validated guidance for developing clinical empathy skills. However, NLP can draw from

Clinician Role: Elicitations	
Direct	**Indirect**
Feelings	
Inquiring directly about the patient's • emotions or mental state: "How do you feel about that?" • emotive behaviors: "What was your reaction?"	Asking about experiences or emotive behaviors, where the clinician may convey interpretation which invites the patient to confirm, reject, or clarify • "So you're worried that the treatment won't work."
Appreciation	
Directly asking the patient about appreciation of things, events, actions, or behaviors • event: "How was the MRI for you?" • thing: "Do you find the medication helpful?"	Explores preferences and statements that convey clinician's interpretation which invites the patient to confirm, reject, or clarify • "It sounds like the medication isn't helping."
Judgement	
Asking the patient about judgement of • self: "Do you think you are a good father?" • others: "Was the nurse helpful?"	Inquire about behaviors or make statements that convey clinician's interpretation, which invites the patient to confirm, reject, or clarify • "So the nurse was not very helpful then?"

Table 5.2: Description and examples of explicit and implicit functions (feeling, appreciation, judgement) in clinicians in how they elicit empathic opportunities.

extensive research in psychology and linguistics, which has empirically studied theoretical models of clinical empathy and measurement approaches.

Though they are still few, NLP studies exploring tools for training and education in empathetic communication have generally integrated insights from these fields, often in collaborations with psychologists. For example, Wambsganss et al. [10] and Wambsganss et al. [202] investigate the effectiveness of digital empathy training tools in enhancing students' empathetic communication skills when writing peer feedback. Other work focuses on applications for psychotherapy [18], examining for instance, linguistic behaviors that signal empathy and session quality in Motivational Interviewing conversations [8, 19, 20], and crisis counseling conversations [21, 22].

Some recent NLP studies on clinical empathy started integrating linguistic theories and discourse analysis approaches. Dey and Girju (2022) created a corpus of medical students' essays about breaking bad news to patients with sentence-level labels of cognitive, affective, and prosocial empathy [17]. They built a novel system architecture informed by frame semantics [261] that outperformed state-of-the-art empathy classification models. On the same dataset, Dey and Girju (2023) showed that incorporating features of Construction Grammar [262] and Systemic Functional Grammar [263] also improves deep learning models for empathy classification [264]. Shi et al. (2021) also demonstrated the potential of the discourse annotation resources to improve empathy NLP models [265]. Nevertheless, creating clinical dialogue datasets with quality clinical empathy annotations suitable for training NLP models requires expertise, labor, and ethics and privacy protections.

To address the described shortcomings, we contribute a novel dataset of annotated clinical empathy

	Clinician Role: Responding to Patient Cues	
Acceptance	**Unconditional Positive Regard**	
	Explicit Positive Judgement Expression of positive judgment of the patient as a person • “You are a reasonable parent” • “You’re very responsible”	*Implicit Positive Judgement* Expression of a judgement of the patients thoughts or feelings • “It’s really great that you’ve been taking care of your parents” • “It sounds like you’ve been working hard to improve your health”
	Repetition Repeating or paraphrasing the patients words without countering statements or premature reassurance • (patient says they’re worried about cancer spread) “I understand you’re worried about the cancer spreading”	*Allowing Full Expression* Allowing patients to express feelings and views fully through minimal responses, nodding, and avoidance of interruption
	Neutral Support	
	Explicit Appreciation Appreciation of ideas, feelings, or behaviors regarding the patients normality or acceptability • “It’s completely normal to be upset about this” • “It wouldn’t be surprising to feel that way”	*Explicit Judgement* Denying negative self-assessment by the patient • “You’re not crazy for being worried about that” • “It’s not bad to be thinking about these things”
Sharing	Sharing patient views or feelings through expressed agreement • “I’m sure I would also feel anxious in this situation if I were going through everything you have going on right now” • “Yes, this medication really does not taste good” • “Oh, no!”	
Understanding	Understanding and acknowledgement of the patients views and feelings can be expressed directly as acknowledgement or interpretations and may attempt to elicit explicit expressions from the patient • “I get the sense that you found the other doctor unhelpful” • “I see that the medication isn’t working for you” • “I understand that you’re worried about infections and perhaps that makes you anxious”	

Table 5.3: Descriptions and examples of categories of clinician responses to patient cues/empathic opportunities.

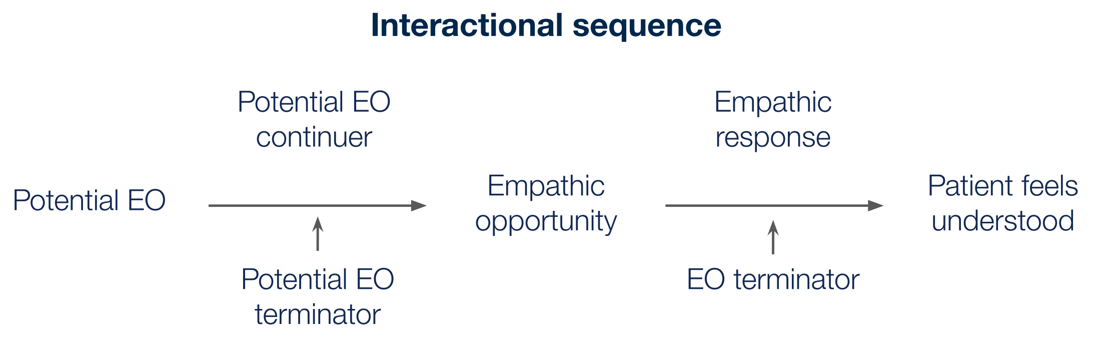


Figure 5.2: An interactional sequence of EOs and EO responses in the clinical interview. Figure is adapted from Suchman et al. (1997).

in breaking bad news conversations, an annotation method based on a linguistic framework for structured analysis of clinical empathy, and a framework of communication strategies for BBNs developed by medical experts. We tailor these resources to support the development of NLP models that can be integrated with digital training tools. Related work outside of NLP has also used discourse analysis approaches to identify strategies in clinical BBN conversations [266]. Recently, Rey Velasco et al. (2022) applied SFL approaches and Pounds [14]'s clinical empathy framework to asynchronous health interactions, revealing insights into implicit/explicit EOs and their relationship to trust between healthcare providers and patients [216]. To our knowledge, this work is the first to apply Pounds's (2011) framework [14] to live conversations and BBN scenarios, which we make public to support NLP research.

5.3 Annotation Scheme

5.3.1 Appraisal Framework

Empathic opportunities (EOs) are expressions or behaviors that reveal a patient's feelings or views, which can be either *explicit* or *implicit* [11]. *Explicit empathic opportunities* directly reveal a patient's feelings or attitude via explicit expressions of behavior or emotive behaviors (e.g., crying). *Implicit empathic opportunities* are defined as "patient statements from which a clinician might infer an underlying emotion that has not been explicitly expressed" [11].

The appraisal framework for clinical empathy extends Suchman et al.'s (1997) model of clinical empathic communication [11], by integrating a finer-grained taxonomy that categorizes EOs and doctor responses based on the linguistic functions of attitudinal expression [14]. It draws on insights from [267]'s linguistic research on interactional sequences that build empathy in psychotherapy settings, and Martin and White's (2005) appraisal framework [268], a systemic functional linguistics (SFL) approach to discourse analysis [263] that concerns the interpersonal, interactive functions of language in specific social settings. The framework aligns various linguistic functions with components of attitudinal expression within empathic communication in order to gain insights that support the teaching of empathic communication skills. The three dimensions of attitudinal expression are Feeling (i.e., affect), Judgment of oneself or other people, and Appreciation; an attitude or perception toward things, events, actions, and behaviors.

Previous research observed that clinicians often overlook empathic opportunities, hindering effective, satisfactory communication with patients [269, 270]. Moreover, Suchman, Markakis, Beckman, and Frankel [11] finds implicit EOs are more common in a medical interview than explicit empathic opportunities. As implicit EOs are hidden in patients' expressions, they are particularly challenging to identify and infer. Thus, the ability to model explicit and implicit EOs is an important step toward digital tools that support communicative skills training for BBNs. By developing a dataset of EOs and their relations to clinician expressions, not only do we enable training such NLP models, but we also provide a resource for investigating linguistic aspects that could add to the body of knowledge about BBN conversation strategies. This resource contributes to ongoing efforts in broader NLP empathy research seeking multidimensional representations of empathy, including affective and cognitive empathy and empathic behaviors [39]. The segments of inferred implicit EOs provide opportunities, especially to better understand cognitive empathy, and can be leveraged in research on abductive social reasoning [199, 271].

As we present the framework, we provide descriptions and examples inspired by Pounds [14].

Categories of Empathic Opportunities

Table 5.1 contains examples and descriptions of the Feelings, Judgments, and Appreciations categories of explicit and implicit EOs, yielding six possible labels to apply to EO spans. Explicit EOs are directly observable in the patient's expressions and behaviors. Implicit EOs can be explored by the clinician to more deeply understand patient views and can lead to explicit EOs and a better consensus on empathic accuracy, which informs the clinician's empathetic response. Implicit feelings can occur with the expression of judgments or appreciation.

Categories of EO Elicitations

Table 5.2 provides descriptions and examples of EO elicitations along the three attitudinal dimensions. Elicitations are parts of the empathic interaction in which the clinician elicits the patient's feelings and perspectives. Direct elicitations typically are questions, whereas indirect elicitations may be carried out in various ways and help clinicians avoid imposing their ideas (and potential misunderstandings) on the patient. They may soften or hedge (e.g., *perhaps, it sounds like, I have a feeling*).

Categories of Empathic Responses

The third family of spans, shown in Table 5.3, are the clinician's empathic response to patient EOs. The framework describes three broad types of responses: 1) Expressing explicit Understanding or Acknowledgement of the patient's feelings/views, 2) Sharing the patient's feelings/views through expressions of agreement, and 3) expressing Acceptance in response to patients' explicit, implicit, or potential negative or positive self-judgment.

There are a variety of ways clinicians express their understanding or acknowledgement. This can overlap with indirect elicitations, as one approach is to express an interpretation/inference about the patient's views. We label these as understanding rather than an indirect elicitation because the former is more specific, and we can label the parts of the turn that show the invitation to the patient to confirm, reject, or clarify the interpretation. In the case of an implicit EO, we may observe first an understanding response followed by elicitation of more explicit patient cues.

Clinicians may express shared feelings/views through expressions that agree with the patient's attitudes (e.g., "I would also feel...", "Yes, [agree with judgment/appreciation]"). Acceptance is exhibited through principles of patient-centered care. Pounds [14] describes two forms of acceptance: unconditional positive regard and neutral support. The former are expressions of positive judgment of the patient. Opportunities for such praise are possible when patients similarly show positive self-judgment. Neutral support can take the form of explicit appreciation or judgement, helping the patient understand their feelings are justified.

Relations form Interactional Sequences

Here, we describe different types of interactional sequences, observable via the relations drawn between empathic opportunities and elicitations or empathic responses. The general sequence is shown in Figure 5.2. *Empathic response sequences* are cases when an EO is directly linked to an empathic response that explicitly recognizes the attitudes conveyed in the EO. *EO continuer sequences* involve a "potential EO continuer" that facilitates further exploration, which can lead to more explicit empathic opportunities, help the clinician gain more insight, increasing empathic understanding [11]. These are identifiable via span relationships that form a sequence of implicit EOs, elicitations, and explicit EOs. *EO terminating sequences* involve an empathic opportunity terminator, when a clinician directs the conversation away from an explicit EO in the patient's prior turn, or a *potential* EO terminator, which occurs after implicit empathic opportunities when the clinician does not explore the implied feeling via further elicitations, instead directing the conversation away from implied cues. The scenario in §5.5.1 demonstrates examples of missed EOs.

5.3.2 SPIKES Protocol

SPIKES is a pedagogical tool commonly employed for training medical students' communication skills for BBN scenarios. It provides a high-level conversation structure and communication strategies to help manage a BBN conversation compassionately while fulfilling four objectives: 1) gathering information from the patient; 2) transmitting the medical information; 3) providing support to the patient; and 4) eliciting the patient's collaboration in developing a strategy or treatment plan for the future [260]. Previous work observed significant increases in confidence in handling aspects of BBN conversations for medical students and faculty trained with the protocol. Mahendiran, Yeung, Rossi, Khosravani, and Perri [272], similarly, found that it improved learner satisfaction, performance, and knowledge.

Coding SPIKES stages

For the medical conversations we collect, the students practice the SPIKES Protocol for Delivering Bad News [260]. This protocol contains six steps which we summarize from [260]:

1. **S**etting: This step involves planning and establishing a time and place for the conversation to take place. This includes determining an appropriate time to have the conversation so that there are no interruptions or need to rush, creating a private space, deciding who needs to be there, such as significant others or other supportive persons for the patient, establishing comfort by sitting down and rapport by other body language.

2. **P**erception: In this stage, the medical professional seeks to learn about what the patient already knows and their current perception of their situation. As patients' knowledge and perception can vary from feelings like denial or wishful thinking, it enables the doctor to tailor the conversation to the patient's understanding.

3. **I**nvitation: The medical professional seeks to understand how much information the patient is ready for at the time of the conversation. Patients can range from wanting the full medical information about their situation to not being ready to process it. The doctor asks questions to assess this and makes plans about what to discuss in the present conversation, in the next meeting, or whether to offer the details to a support person of the patient, such as a relative or friend.

4. **K**nowledge: This is the stage where the medical professional gives knowledge and information to the patient. The doctor first tries to introduce or warn that bad news is coming with a phrase such as "I'm sorry to tell you that I have some bad news." The protocol guides during this step encourage the doctor to match the comprehension and vocabulary of the patient, use nontechnical words, and avoid excessive bluntness. They try to deliver small bits of information broken up with check-ins of the patient's understanding and, furthermore, not to convey that there are no therapeutic options for the patient.

5. **E**mpathy/Emotion: Delivering bad news can be met with various reactions like shock, anger, disbelief, and sadness. Perhaps the most difficult challenge in delivering bad news is responding empathetically in order to support the patient. The protocol details four steps of an empathetic response: (i) Observe the patient's reaction; (ii) Identify the type of reaction or emotion the patient is experiencing, asking open questions about what they are thinking and feeling if necessary; (iii) Identify the emotion reason; (iv) Allow the patient time to express their feelings, and connect to their emotion through a response statement that reflects their understanding and legitimize how the patient feels, and the physician may use other body language like moving closer to the patient or touching their arm. The physician continues empathic responses until the patient becomes calm.

6. **S**trategy and **S**ummary: Here, the physician discusses treatment plans with the patient if the patient is ready.

Except for **Empathy**, which is covered by the appraisal framework labels, each of the SPIKES stages is labeled on clinician turns or segments of the clinician turns when identifiable. We note that Setting is rarely applied given that much of the behaviors involved in this stage occur prior to the BBN conversation.

5.4 The BBN Empathy Dataset

The BBN Empathy Dataset is constructed via a collaboration with medical didactics experts at a large medical school. The dialogues are practice BBN conversations between medical students taking part in a medical didactics seminar and trained standardized patient actors. These simulate BBN scenarios as realistically as possible for student practice. They take place in rooms modeling medical environments, such as a doctor's office or a hospital bed. During the seminar, the students are trained in BBN communication skills with the SPIKES Protocol (§5.3.2). The standardized patients are provided a role description and background for the scenario, and the students are provided a full scenario description and patient history. The scenarios include delivering cancer diagnoses, failures of treatments for serious diseases, and informing a family member of a severe accident, among others. We curated a total of 63 conversations in German over two semesters of the seminar.

5.4.1 Transcription and Annotation Procedure

Transcriptions

Four trained L1 German speakers transcribe them. Though these are practice conversations, we anonymize the participants' names. The time it takes to transcribe a full conversation is very dependent on the audio quality of the file. With prioritizing simulating a realistic setting for the medical students' practice, the microphone position reduces the audio quality. Some transcribers report that in especially low-quality circumstances in which the voices are muffled and unclear, the time it takes to complete transcribing the dialogue can amount to 12 hours since the writing, rewinding, correcting, rewinding, repeating, can be cumbersome.

To aid the transcription labor, we searched for effective ASR tools for German that can be run offline on a private server (to protect the data) and handle low quality. We found setting up Whisper[2] [273] offline was the most effective, and integrated it as a starting point for the transcribers. In the best-case scenario, with a quality Whisper base script, the task consists mostly of setting the correct timestamps in the right places, correcting the occasional misinterpretation, and filling in the parts that Whisper failed to recognize at all. The task of transcribing the dialogue can be done in two to three hours. Even so, it still depends on the audio quality, which impacts the quality of Whisper transcriptions.

Platform Integration

Our scheme is integrated in INCEpTION [274], an open-source tool for semantic annotations that supports labeling text spans and relations.[3] We setup custom layers for each. The annotator highlights a text segment and selects a coarse-grained span category (*EOs*, *EO elicitations*, *EO responses*, and *SPIKES*) for the span. This opens tags representing the fine-grained labels for the selected category which the annotator then applies as the span label.

Annotator Training

The annotation task is complex and non-trivial, requiring dedicated time and work effort to properly reason through the types of spans. This requires expert annotators who have a comprehensive understanding of the theoretical frameworks, the dialogue setting, and experience applying the framework. Two L1 German speakers were trained to perform the annotations in three 1-hour sessions. The training included background on SFL, a tutorial on the full coding manual, additional training materials involving real samples for demonstration and practice applying the coding scheme, and a tutorial on the annotation tool. We explain how to approach the analysis and labeling throughout example scenarios in §5.5.1.

[2] https://github.com/openai/whisper

[3] https://inception-project.github.io/

After the training sessions, the annotators performed three calibration rounds on dialogues that they coded independently. We met as a group to discuss the source of disagreements between annotators. The primary source in the first round was distinguishing between Judgments and Appreciations, which we clarified and added further examples with explanations to the coding manual. After this, the agreement improved in the subsequent calibration rounds. In those rounds, the main challenge related to identifying the implicit EOs. Some disagreements on this aspect are reasonable, as it requires the annotator to make their own inferences, which can vary subjectively. However, during our discussions, we reached a consensus for most implicit EOs. We met weekly to review coding conventions, clarify questions, and discuss specific cases throughout the months of the annotations. We present an analysis of interannotator agreement post-training and calibration in the next section.

5.4.2 Agreement Study

After training and the calibration rounds, the annotators independently coded eight dialogues. For these, we study the interannotator agreement on the text spans and span labels.

Text span agreement

We measure the interannotator agreement on the text spans labeled by the annotators at the string level based on Wiebe, Wilson, and Cardie [275]'s approach shown in Equation 5.1.

$$agr(a||b) = \frac{|A \text{ matching } B|}{|A|} \tag{5.1}$$

A and B are the sets of spans highlighted by annotator a and b respectively. $agr(a||b)$ is the proportion of text marked by annotator a that b marked, and $agr(b||a)$ is the proportion of text marked by b that a marked. agr is the mean of both measures.

As with Wiebe, Wilson, and Cardie [275]'s span labeling task, agr is a suitable metric for text span agreement for our task because we do not employ nor instruct strict rules about the precision of the text boundaries. The main concern, rather, is whether the annotators mark the same general expression. Example (1) shows a case where for turn t, annotator A marked an additional clause not marked by B, but the expression is generally the same in terms of the appraisals they convey.

(1) A_t: Ähm, okay, Operation? **Okay, also das, das muss raus, oder was?**
B_t: **Okay, also das, das muss raus, oder was?**
English: Um, okay, surgery? Okay, so that, that has to come out, or what?

Here, $agr(a||b) = 0.65$ and $agr(b||a) = 1.0$. In example (2), $agr(a||b) = 0.94$ and $agr(b||a) = 0.86$. Bold indicates the text was marked by both annotators and bold and italic are spans marked by only one annotator.

(2) A_t: ***Ähm, okay,*** **kann ich, vielleicht Wasser, oder irgendwie... €** (inaudible) [**Patient crying]** Oh fuck. (inaudible) Also kann ich dann wieder arbeiten, oder was? (inaudible) [**laughing desperately**] **Ich weiß nicht, ob Sie das verstehen, aber wenn ich nicht arbeiten kann, dann...**
B_t: Ähm, okay, **kann ich, vielleicht Wasser, oder irgendwie... €** ***(inaudible) [*****Patient crying] Oh fuck.** (inaudible) Also kann ich dann wieder arbeiten, oder was? (inaudible)

> **[laughing desperately] Ich weiß nicht, ob Sie das verstehen, aber wenn ich nicht arbeiten kann, dann...**
> *English*: Um, okay, can I have, maybe water, or something... Oh fuck. So can I then go back to work, or what? I don't know if you understand but if I can't work, then...

First, we compute *agr* for each turn. Then, we take the average across all turns to get the *agr* for a dialogue. Table 5.4 shows the *agr* metrics for each dialogue and the means across all eight. The agreement on all spans improved notably between dialogue 0 and 1, which is due to our continued discussions after each dialogue annotation which focused on general clarifications about the scheme and coding conventions rather than resolving disagreements. In our analysis, we observed small differences between the annotator's choice to include punctuations and short subclauses in the annotations. Overall, the annotators match the same general expression.

Dialogue	agr	$agr(a\|\|b)$	$agr(b\|\|a)$	α
		Calibration Dialogues		
1	.789	.738	.840	.30
2	.910	.850	.970	.43
3	.902	.844	.960	.47
4	.952	.923	.981	.50
5	.931	.865	.997	.58
6	.912	.935	.890	.63
7	.905	.948	.862	.85
8	.948	.933	.963	.89
mean	.92 ± .02	.90 ± .04	.95 ± .04	
		All Dialogues		
mean	.97 ± .03	.96 ± .05	.98 ± .03	

Table 5.4: Span text agreement for each of the eight calibration dialogues and the mean agreement with standard deviations across all 63 dialogues. The right column shows the span label agreement measured by Krippendorff's α.

Span label agreement

As a first point of reference, we computed Krippendorff's α for all labels (strict) and report them in Table 5.4. We also study agreement on the span labels by computing *agr* for each type of label; 1) All labels include all fine-grained AF labels and SPIKES labels; 2) Coarse-grained labels are the three AF categories (EO, EO elicitation, EO response); 3) Fine-grained labels include the attitudes and explicit/implicit for EOs, direct/indirect for EO elicitations, and each fine-grained category of EO responses; 4) Attitudes only include the *feelings*, *judgement*, and *appreciation* labels (i.e., combining explicit and implicit, and direct and indirect); and 5) only SPIKES labels. $agr(a||b)$ represents annotator b's precision evaluated against annotator a's labels and $agr(b||a)$ is a's with respect to b. We compute these metrics by a *strict* evaluation denoted by ∪, which includes spans where there was no overlap between annotators, and by a *matched* evaluation denoted by ∩, which includes only spans with some overlap.

As noted, the annotators improved their text span agreement after dialogue 1 following further clarifications. We observed a stark contrast in the agreements between dialogue 1 and the other

	agr	$agr(a\|\|b)$	$agr(b\|\|a)$
All labels			
∪	0.71	0.72	0.69
∩	0.73	0.74	0.73
Coarse-grained labels			
∪	0.87	0.90	0.85
∩	0.90	0.91	0.90
Fine-grained labels			
∪	0.59	0.59	0.59
∩	0.61	0.60	0.62
Attitudes			
∪	0.69	0.70	0.69
∩	0.72	0.71	0.72
SPIKES	0.63	0.67	0.59

Table 5.5: Mean interannotator agreements on span labels for the eight calibration dialogues. ∪ indicates *agr* over all spans, whereas ∩ indicates *agr* only on spans that have overlap.

	agr	$agr(a\|\|b)$	$agr(b\|\|a)$
All labels			
∪	0.88	0.88	0.87
∩	0.89	0.89	0.90
Coarse-grained labels			
∪	0.95	0.96	0.94
∩	0.97	0.96	0.97
Fine-grained labels			
∪	0.81	0.81	0.81
∩	0.82	0.81	0.83
Attitudes			
∪	0.85	0.85	0.85
∩	0.87	0.86	0.88
SPIKES	0.90	0.91	0.89

Table 5.6: Mean interannotator agreements on span labels across all 63 dialogues. ∪ indicates a strict evaluation over all spans, including those where the annotators had no overlap, whereas ∩ indicates evaluation only on spans that have overlap.

dialogues. Dialogue 1 had very low *agr* on the span labels: ∩ *agr* for *all labels* was 0.5; AF coarse-grained and SPIKES ∩ agreements were 0.85 and 0.60 respectively; the rest ranged from 0.22 (AF fine ∩) to 0.38 (AF attitude ∩). Meanwhile, the label agreements across the rest improved; Table 5.5 shows the mean values. We find that the annotators generally perform well at matching each other. Disagreement was mostly between implicit feelings and other implicit labels or explicit judgement. Implicit responses are more subjective or difficult to interpret. For SPIKES, the most common disagreement was between the invitation and knowledge steps, as determining how much the patient is ready to hear may be part of the step of delivering that information. After we reached substantial agreement across all categories, the remaining dialogues were annotated by one annotator and revised for quality by the other. Table 5.6 shows the agreements over all dialogues. Note that only the eight dialogues discussed here were annotated independently; otherwise, they were annotated first by one annotator, and reviewed by the other.

	Patient Role: Empathic Opportunities			
	Explicit		**Implicit**	
	A	*B*	*A*	*B*
Feelings	248	203	1043	1146
Appreciation	608	618	180	178
Judgement	874	991	135	56
	Clinician Role: Empathic Elicitations			
	Direct		**Indirect**	
	A	*B*	*A*	*B*
Feelings	109	113	55	107
Appreciation	203	172	96	89
Judgement	151	149	68	40

Table 5.7: Counts of each type of Empathic Opportunity (EO) label and each type of EO Elicitation label from each annotator (*A* and *B*). For EOs, both annotators identified significantly more *implicit* rather than *explicit* feelings, whereas they identified more *explicit* than *implicit* appreciations and judgements. *Implicit feeling* is the most common type of EO. Both annotators identify more *direct* rather than *indirect* EO elicitations.

5.4.3 Dataset statistics

The label distributions for EOs and EO Elicitations are presented in Table 5.7 and for EO responses in Table 5.8. Table 5.10 shows the average and standard deviation of the label counts per dialogue. For the Patient EOs, we observe that implicit feelings are much more common than explicit feelings in line with Suchman, Markakis, Beckman, and Frankel [11]'s findings, whereas the opposite is the case for judgment and appreciation. However, this is consistent with Pounds [14]'s observation that implicit feelings are often identified within explicit judgments and appreciations.

We quantify the relations between patient EOs and clinician responses, showing the percentage of EOs that are linked to responses in Table 5.9. Interestingly, we observe that the *implicit EOs* have a higher rate of clinician responses than *explicit EOs*. *Explicit appreciations* and *judgements* are more frequently identified than *implicit* ones. As the annotators remarked on the difficulty with identifying appreciations and judgements compared to feelings, it may be that observing the clinicians' responses aids the annotator in observing the implicit EOs, thus biasing the relation rates. However, the higher response rate to *implicit* EOs is also the case for *feelings*, for which *implicit* cases are more frequently marked. Future work could investigate the possible effects further by testing the annotation scheme in a setup in which the identified EOs are locked before observing the rest of the dialogue. Figure 5.3

	Clinician Role: Empathic Responses	
	ACCEPTANCE	
UNCOND. POSITIVE REGARD	*A*	*B*
Explicit Positive Judgement	374	454
Implicit Positive Judgement	143	109
Repetition - no counter	31	27
Allowing Full Expression	156	184
NEUTRAL SUPPORT	A	B
Explicit Appreciation	373	405
Explicit Judgement	185	141
	SHARING FEELINGS AND VIEWS	
	A	*B*
Feelings	40	42
Appreciation	50	59
Judgement	44	60
	UNDERSTANDING FEELINGS AND VIEWS	
	A	*B*
Feelings	427	462
Appreciation	75	56
Judgement	167	170

Table 5.8: Counts of each type of Empathic Response label from each annotator (*A* and *B*). With all sublabels, *unconditional positive regard* is the most common response type, the most frequent among them being *explicit positive judgement* and *explicit appreciation*. *Understanding* rather than *sharing* feelings and views is more common; here, understanding *feelings* is most frequently observed.

shows the percentages broken down by each EO response type. We observe higher proportions of relations between EOs and EO responses of the same attitude type.

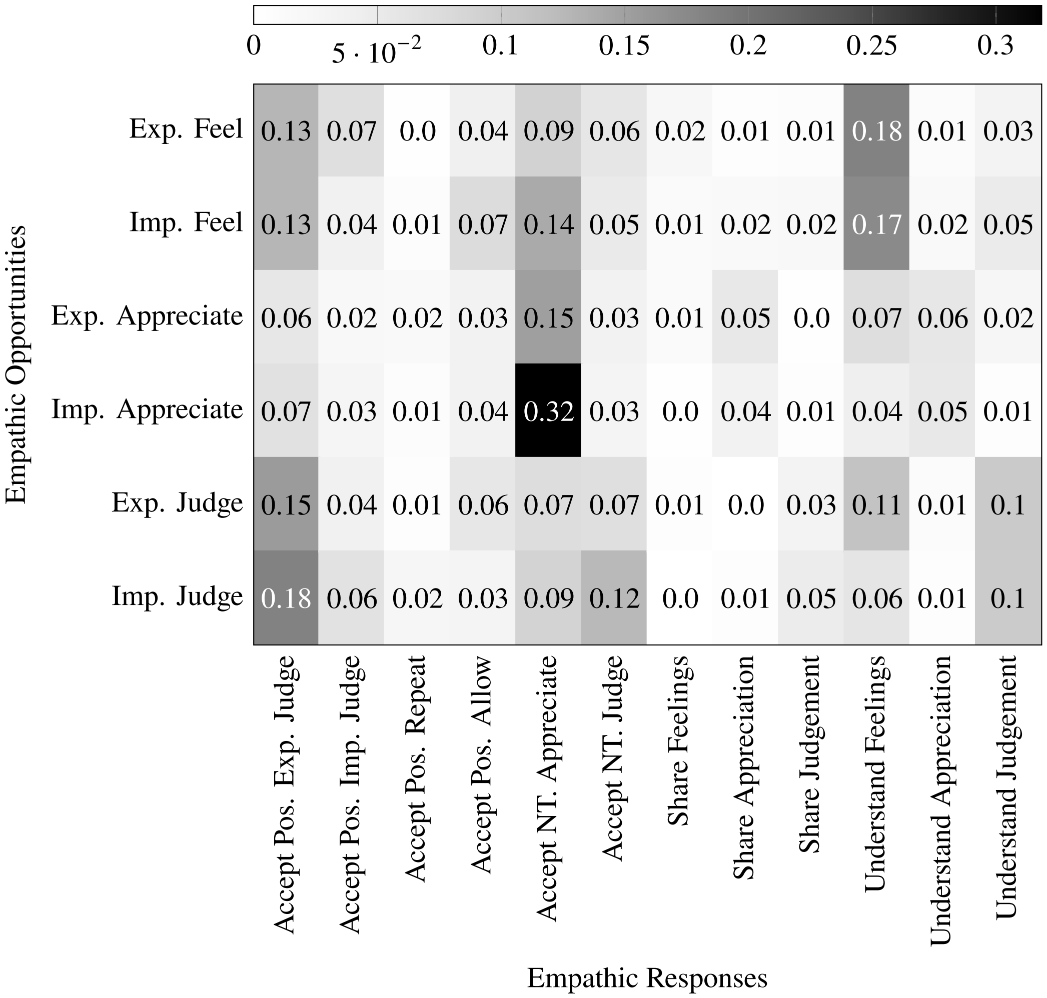
5.5 Discussion

In this section, we first illustrate an example BBN scenario to discuss the application of the annotation scheme. Then, we discuss opportunities for further investigation.

5.5.1 How do I Tell my Family? Applying the Framework

A physician informs a patient that an MRI, performed as part of an anonymous research study, detected a mass in their brain. The patient is in disbelief, as they did not expect a follow-up and had no previous cause for concern, suggesting that the physician had mistaken the results for someone else's.

The physician allows the patient to experience shock and express disbelief implicitly by denying that the results could indeed be for them (IMPLICIT FEELING EO). The physician identifies the emotion of disbelief and responds, "You can't believe it quite yet, I have a feeling." The first clause is the



	Accept Pos. Exp. Judge	Accept Pos. Imp. Judge	Accept Pos. Repeat	Accept Pos. Allow	Accept NT. Appreciate	Accept NT. Judge	Share Feelings	Share Appreciation	Share Judgement	Understand Feelings	Understand Appreciation	Understand Judgement
Exp. Feel	0.13	0.07	0.0	0.04	0.09	0.06	0.02	0.01	0.01	0.18	0.01	0.03
Imp. Feel	0.13	0.04	0.01	0.07	0.14	0.05	0.01	0.02	0.02	0.17	0.02	0.05
Exp. Appreciate	0.06	0.02	0.02	0.03	0.15	0.03	0.01	0.05	0.0	0.07	0.06	0.02
Imp. Appreciate	0.07	0.03	0.01	0.04	0.32	0.03	0.0	0.04	0.01	0.04	0.05	0.01
Exp. Judge	0.15	0.04	0.01	0.06	0.07	0.07	0.01	0.0	0.03	0.11	0.01	0.1
Imp. Judge	0.18	0.06	0.02	0.03	0.09	0.12	0.0	0.01	0.05	0.06	0.01	0.1

Figure 5.3: **Relation heatmap.** This heatmap reflects relations between patient EOs (rows) and subsequent EO responses (columns). The cell values reflect the proportion of the EOs of the type specified by the row that was responded to with the EO response type specified by the column.

physician's interpretation of how the patient feels (Understanding/Acknowledgement). The second serves to soften the delivery of the interpretation and to formulate the statement as an *indirect elicitation:* Feeling and/or Appreciation.

The patient speaks more openly, sharing that everything was going great for them. They state, "It is probably not easy. If there's really something there, it won't just go away on its own." This could signal that the patient is accepting the news. One may interpret this as an *implicit EO*, inferring a *negative Appreciation* that treatment will be difficult and a *negative Feeling* (anxiety/fear). The physician acknowledges the patient's view, confirming treatment probably will not be easy (*empathic response: Acceptance, neutral support*). The patient asks if, theoretically, one can die from such a mass; an *implicit EO: Feeling* (fear of severity/uncertainty). The physician confirms this but says further analysis must clarify the type and severity. This balances the SPIKES/BBN principles of *honesty* and *lending hope*, delivering Knowledge clearly and compassionately.

EO type	No Response	EO Response
Explicit Feeling	35.4	64.6
Implicit Feeling	27.8	72.2
Explicit Appreciation	46.9	53.1
Implicit Appreciation	36.5	63.5
Explicit Judgement	34.3	65.7
Implicit Judgement	27.3	72.7

Table 5.9: The percentage of EOs by EO type that had *no response* or an *EO response* linked by a relation. The rate of responses is higher for *implicit EOs* than *explicit EOs*. The percentages broken down by response type are shown in Figure 5.3.

Later, the patient shares that their aunt had had a brain tumor a few years earlier, describing a quick escalation that was painful and fatal. The loss weighs heavily on the family. We consider this turn to have *implicit EOs: negative* Feeling (signaling fear/worry) and Appreciation (of this significant event in the family). The physician responds, saying it does not mean the patient's case will be the same. While aspects of this response reflect SPIKES/BBN principles (e.g., *lending hope*, attempting to *reduce a sense of isolation*), it is also an *EO terminator* since it directs the conversation away from the implicit EOs. The patient's underlying perspectives that this story communicates become clearer as the dialogue continues, suggesting there were indeed missed EOs.

After the physician's response, the patient immediately expresses anxiety about telling their family, asking how to break the news, implying concern for how the family will react. The physician misses this EO, responding "you can figure that out for yourself." The patient asks again what to do, saying that they cannot simply go home and tell their family they might have cancer (*implicit EOs: negative* Feeling *(anxiety) and* Appreciation *(anticipating significant difficulty in informing their family)*). The physician says nothing, perhaps allowing space for the patient's emotive behaviors. As the conversation continues, the patient continues to express the same sentiment about their aunt and not knowing how to tell their family, EOs that are repeatedly missed by the physician.

5.5.2 Future Work

We have developed a span-relation labeling method based on models of interactional sequences and semantic exchanges in clinical empathy conversations. This establishes links between empathic opportunities and empathic responses, enabling analysis of the types of interactional sequences that achieve empathy. Future work can explore the linguistic components of the discourse elements through computational linguistic analyses, and dynamic models of the interactional sequences to study their impacts on empathic understanding. In addition, future work can investigate NLP models trained on the BBN Empathy dataset for supporting scaling up such resources in collaboration with trained human annotators. The dataset enables establishing novel NLP tasks of automatically labeling spans of empathic opportunities, EO elicitations, and EO responses, and relations between them. Effective models developed for these tasks could be leveraged in pedagogical tools. They could also have a broader impact in NLP research on empathy by providing insight into other conversational empathy datasets, informing empathy generation models, and evaluating them.

There may be unknown effects that the simulated scenarios have on the dialogues. Further research could investigate artifacts of the simulated scenarios and how this might affect future approaches. Although we see progress in applying the SPIKES protocol and its pedagogical benefits, there is still room for improvement, especially as SPIKES does not specify higher-level aspects of the interaction or the patient's role in the conversation, which appraisal framework for clinical empathy helps address.

Other research has brought to light insights about the patient perspectives on SPIKES. Assessing patients' preferences for BBN communication and their perception and satisfaction of actual BBN disclosures, Seifart et al. [276] administered the Marburg Breaking Bad News Scale (MABBAN), a questionnaire based on the SPIKES protocol, to 350 cancer patients. They observe that only 46.2% of the patients are fully satisfied by how the bad news was broken to them and that there is a highly significant discrepancy between the patients' preferences for receiving bad news and the actual disclosure. Furthermore, they find that the overall patient's satisfaction with the first BBN disclosure significantly correlates with their emotional state, including depression, anxiety, and sleeplessness, after receiving the bad news. Von Blanckenburg et al. [277] later administered MABBAN to 336 cancer patients. Analyzing its psychometric properties, they observed an accordance between the SPIKES protocol and the MABBAN scale, suggesting that SPIKES meets the preferences of German cancer patients. The study emphasizes that differentiated communication of BBN is highly important due to discrepancies in patient preferences.

5.6 Chapter Summary

In this chapter, we pursued three core objectives toward modeling clinical empathy in patient-clinician conversations: 1) develop an annotation scheme for labeling precise discourse elements and their relations in clinical empathy encounters; 2) produce a structural representation of the dynamics of these elements over a full conversation toward addressing current shortcomings of NLP models for empathic interactions; and 3) tailor the approach for medical didactics research for training empathic communication skills in BBN conversations. These objectives target RQ3, in which we aim to integrate theory and linguistics-based frameworks into clinical conversational settings and, more specifically, breaking bad news scenarios. In Section 5.3, we presented a span-relation labeling method based on models of interactional sequences and semantic exchanges in clinical empathy conversations. This establishes links between empathic opportunities and empathic responses, enabling analysis of types of interactional sequences that achieve empathy. In addition, we produced the BBN Empathy dataset described in Section 5.4, the first of its kind curating discourse-level annotations tailored to clinical empathy, which is available for fellow researchers. In Section 5.5, we illustrated the application of the annotation scheme by narrating an example BBN and discussed future opportunities with the dataset. These contributions enable open research and interdisciplinary collaboration addressing critical aspects of empathic communication in healthcare contexts.

In the next chapter, we explore another setting of support-oriented interactions: exchanges in online mental health forums. While this chapter followed a theory-first perspective (refer to Figure 1.1 in Chapter 1), the next uses computational models first, to provide empirical insight into existing theories of relevant conversational settings.

Label	**Avg ±std**
Patient EO	33.9 ±14.6
Explicit Appreciation	5.9 ±5.0
Implicit Appreciation	2.4 ±2.6
Explicit Feeling	2.4 ±2.3
Implicit Feeling	12.5 ±5.6
Explicit Judgement	9.5 ±6.0
Implicit Judgement	1.2 ±1.7
EO elicitation	6.8 ±6.4
Direct Appreciation	1.8 ±3.0
Indirect Appreciation	0.9 ±1.6
Direct Feeling	1.2 ±1.2
Indirect Feeling	0.8 ±1.7
Direct Judgement	1.6 ±1.4
Indirect Judgement	0.5 ±1.0
EO Response	19.1 ±9.3
Acceptance: Neutral Support–Explicit Appreciation	3.8 ±2.5
Acceptance: Neutral Support–Explicit Judgement	1.7 ±1.8
Acceptance: Unconditional Positive Regard–Allowing Full Expression	1.7 ±1.5
Acceptance: Unconditional Positive Regard–Explicit Positive Judgement	3.5 ±2.8
Acceptance: Unconditional Positive Regard–Implicit Positive Judgement	1.2 ±1.7
Acceptance: Unconditional Positive Regard–Repetition	0.4 ±1.0
Sharing Attitude: Appreciation	0.5 ±1.0
Sharing Attitude: Feeling	0.3 ±0.7
Sharing Attitude: Judgement	0.5 ±1.1
Understanding Attitude: Appreciation	0.6 ±1.0
Understanding Attitude: Feeling	3.4 ±2.4
Understanding Attitude: Judgement	1.4 ±1.6
SPIKES	15.1 ±5.6
Setting	0.7 ±0.6
Invitation	1.8 ±1.5
Knowledge	5.0 ±3.0
Perception	2.3 ±2.2
Strategy/summary	5.2 ±3.3

Table 5.10: Averages and standard deviations of label counts per dialogue.