

## A. Additional Data Analysis

### A.1. Statement Length Distribution

The length distribution for positive and negative statements are presented in Figure 1 and Figure 2, respectively. There is no significant bias in statement lengths for positive and negative statements.

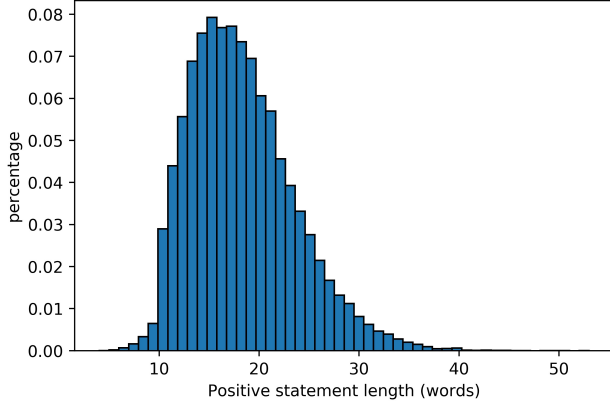


Figure 1. Distribution of positive statement lengths.

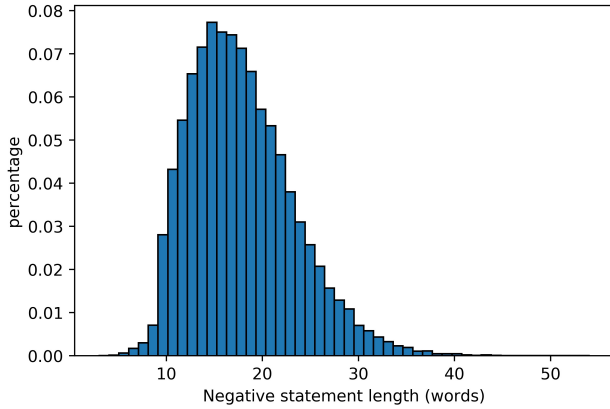


Figure 2. Distribution of negative statement lengths.

### A.2. Statement Content

Table 1 shows the most common nouns, verbs and adjectives in positive statements, respectively.

### A.3. Video Length Distribution

The video clips collected from MovieClips are all 40 seconds long. For video clips collected from TV shows, their lengths vary from 15 to 40 seconds, shown in Figure 3.

Type	Most Common Words
Noun	man, woman, shirt, suit, hair, jacket, girl, lady, boy, dress, sweater, friend, brunette, room, guy, people, tie, glass, table, car, coat, door, hat, phone, hand, top, bed, house, couch, group
Verb	tell, wear, ask, want, sit, try, say, talk, go, explain, walk, get, make, look, see, think, take, give, will, hold, can, stand, know, come, leave, feel, have, find, put, like
Adj	black, blue, blonde, red, white, brown, green, haired, young, dark, grey, old, other, pink, purple, upset, plaid, gray, yellow, long, little blond, happy, good, excited, surprised, striped, light, angry, short

Table 1. Most common words in positive statements.

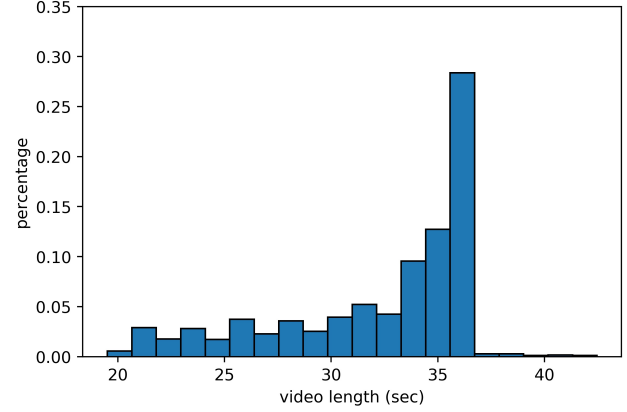


Figure 3. Distribution of video lengths from TV shows.

## B. Instructions for Human Annotators

Figure 4 through 7 show the detailed instructions and user interface for human annotators.

## C. More Examples

Figure 8 and Figure 9 show some more examples of predictions from our model on movie and TV show clips. The model used in these examples is Stmt+Subtt+Vis with BERT and Img features.

## Instructions

Given a movie clip, you are supposed to (1) **drag the start/end slider below the video** to cut out an interval of interest that **contains a single meaningful scene**; (2) **based solely on the interval you cut**, write down a **fact/statement** inferred from the interval, and then slightly **rewrite it to contradict** the interval you cut. The goal of the HIT task is to collect pairs of real/fake statements that are inferred from the video clip.

Task 1: Cut out an interesting and self-contained **interval** from the given video clip:

- The interval you cut should contain a single meaningful scene, use the two handlers in the slider to mark the beginning and end of the scene.
- The **hard requirement** for the length of the interval you cut is **between 15 and 30 seconds**. Do not cut your interval out of this range.
- The interval you cut should be a **single meaningful scene**
  - It **does not have to contain everything** in the scene. You can cut a scene in half to meet the length requirement, as long as the interval you cut is interesting and meaningful so that you can write real/fake statements about it.
  - Your interval should **contain only one scene**. That means, your interval should not contain flashback, change of scenes (locations), opening songs, etc.
  - In this [\[example video\]](#), there are three separate scenes: 0:04->0:40, 0:40->1:10, and 1:10->1:30. Each of these three intervals is a meaningful scene, since each of them contains a coherent interaction among characters. However, the first interval exceeds the time limit, so we need to cut it short. For example, we can cut it into 00:10->00:40, or cut it even shorter to 00:25->00:40. (the interval does not have to contain everything, as long as it is self-contained)
- **In this job you only need to cut one scene and write three pairs of statements for it.**

Task 2: Infer "real" and "fake" statements from the interval you just cut:

### Real statements:

- Should be **inferred** from the interval you cut: e.g. human interaction, dialog, human relation, intention, common sense, reasoning.
- **The statement you write should not directly appear in subtitles or video scenes.** Write something that is implicitly expressed in the video clip.
- Your statements should be only based on the scene (what you see) and the subtitles. **Do not write about the audio information** (e.g. sound in the background, character raising voice, character speaking another language that is not reflected in the subtitles)

### Fake statements:

- Should be contradictory to the interval you cut, by **twisting details** in your real statement to make it fake: e.g. switching characters, changing the cause, describing a wrong activity, changing the mood, intention, facts, etc.
- However, your fake statement should be **confusing with the real one given the interval you cut**. That means, your fake statement should **not be completely out-of-context** with the real one. It should be very relevant to the video and the real statement. (more examples later)
- Your fake statement should look **very similar** to the real statement, in terms of length, structure, writing style, wording, etc.

Figure 4. Overall instructions for human annotators.

## How to write real statements:

The real statement should be inferred from both the video and subtitles. The followings are some examples for accepted real statements.

- **Describing an event with its cause**
  - The people in the living room are trying to make the brunette in the kitchen calm down and get dressed because she is upset about an ex-boyfriend calling. [\[example video\]](#)
  - The woman a black pleads with her friend because she doesn't want to work a humiliating job. [\[example video\]](#)
  - The brunette girl is sitting at the counter when her friends walk in acting all sad when they're actually excited because they won the game. [\[example video\]](#)
  - The woman in the blue top did not want to come downstairs because it was raining and she wasn't dressed. [\[example video\]](#)
- **Describing a complex event**
  - The brunette in the kitchen starts freaking out and yelling at her friends about a possibly old message she heard on the answering machine. [\[example video\]](#)
  - The lady in the kitchen is talking to her friends about when a message was left on their answering machine. [\[example video\]](#)
  - A woman in a black shirt begs her friend to give her money as soon as she opens the apartment door. [\[example video\]](#)
  - The man in the kitchen is breaking up with his roommate's sister via letter and he's running his plan by his other friend first. [\[example video\]](#)
- **Describing human emotions, intentions, relations, etc.**
  - The woman in cream feels sorry for her friend but she can't help her. [\[example video\]](#)
  - The woman seated on the sofa is upset and crying because she did not get the apartment she wanted. [\[example video\]](#)
  - The man wearing the striped shirt is actually relieved he is interrupted from jumping. [\[example video\]](#)
  - The man with the white shirt is willing to give away the entertainment center for free. [\[example video\]](#)

The followings are examples of rejected real statements. Make sure to avoid them when writing your own statements.

- **Ambiguous referring**
  - **[Wrong]** She feels sorry for her friend but she can't help her. [\[example video\]](#)
  - **[Wrong]** The man is breaking up with his roommate's sister via letter and he's running his plan by his other friend first. [\[example video\]](#)
- **Use character names to refer to characters**
  - **[Wrong]** Rachel feels sorry for her friend but she can't help her. [\[example video\]](#)
  - **[Wrong]** Chandler is breaking up with his roommate's sister via letter and he's running his plan by his other friend first. [\[example video\]](#)

**You need to make your referring unambiguous given the interval you cut (by describing the characters' attribute, like the examples in the upper half of this section). Do not use names to refer to characters.**
- **Directly appears in the subtitles, simple copy or rephrase**
  - **[Wrong]** The woman in black tells the woman in cream that she lost all her money and ask her for 100 dollars to get back in the game. [\[example video\]](#)
  - **[Wrong]** The man in black says the letter the man in grey is reading is the hardest letter he has ever written. [\[example video\]](#)

**Your submissions will be immediately rejected if your statements are only based on subtitles (like the ones above). Avoid writing things like "somebody says", "somebody tells somebody", "somebody thinks that", etc.**
- **Unjustifiable statements (cannot be inferred from the interval)**
  - **[Wrong]** The woman in black asks for money to buy a better car. [\[example video\]](#)

Even if you are familiar with the TV shows, you should only write what is presented in the clip, not to use your own knowledge about the plot.
- **Too trivial or simple statements**
  - **[Wrong]** A group of people are hanging out in the living room. [\[example video\]](#)
  - **[Wrong]** The man in suite walk up to the woman in black and talks to her. [\[example video\]](#)
  - **[Wrong]** The group of people in the living room looks at the woman in black when she yells at them. [\[example video\]](#)

Please write statement that is inferred from the clip, combining information from both the video and the dialogue (subtitles).

Figure 5. Instructions for writing real statements.

### How to write fake statements:

When you finish your real statements, please twist a detail in your real statements to make them into fake ones. **Most important requirements** are (1) make fake statements look similar to real ones (length, structure, style); (2) make fake statements confusing with real ones given the clip you cut (they should be in the same context of the video and subtitles).

- **Switch (change) characters**

[Real] The woman in cream feels sorry for the woman in black but she can't help her. [\[example video\]](#)

[Fake] **The woman in black** feels sorry for **the woman in cream** but she can't help her.

[Real] The brunette in the kitchen starts freaking out and yelling at her friends about a possibly old message she heard on the answering machine. [\[example video\]](#)

[Fake] **The man in suite** starts freaking out and yelling at her friends about a possibly old message she heard on the answering machine.

- **Change the cause of the event**

[Real] The woman in the blue top did not want to come downstairs because it was raining and she wasn't dressed.

[\[example video\]](#)

[Fake] The woman in the blue top did not want to come downstairs because **her camping trip was cancelled**.

[Real] The woman a black pleads with her friend because she doesn't want to work a humiliating job. [\[example video\]](#)

[Fake] The woman a black pleads with her friend because she has to **pay for the diner and the costumes**.

- **Change a fact**

[Real] A woman in a black shirt begs her friend to give her money as soon as she opens the apartment door. [\[example video\]](#)

[Fake] A woman in a black shirt begs her friend to give her money as soon as the **woman in cream steps into the apartment**.

**The followings are examples of rejected real statements.** Make sure to avoid them when writing your own statements.

- **Fake statements too different from real ones**

[Real] The brunette in the kitchen starts freaking out and yelling at her friends about a possibly old message she heard on the answering machine. [\[example video\]](#)

[Fake] **[Wrong]** The brunette woman yells at her friends for not leaving her the message. (**Lengths differ too much**. Real and fake statements should roughly have the same length)

[Real] The woman a black pleads with her friend because she doesn't want to work a humiliating job. [\[example video\]](#)

[Fake] **[Wrong]** The woman has to pay for her diner and costume, and so she pleads with her friend. (**Sentence structure and style differ too much**. See good examples in the first half of this section)

- **Fake statements not confusing with real ones IMPORTANT: this is the most common mistake that turkers made**

[Real] The people in the living room are trying to make the brunette in the kitchen calm down and get dressed because she is upset about an ex-boyfriend calling. [\[example video\]](#)

[Fake] **[Wrong]** The people in the living room are trying to make the brunette in the kitchen calm down and get dressed because she was robbed.

[Fake] **[Wrong]** The people in the living room are trying to make the brunette in the kitchen calm down and get dressed because she was fired by her boss.

[Fake] **[Wrong]** The people in the living room are trying to make the brunette in the kitchen calm down and get dressed because her friends did not ask her out for dinner.

[Fake] **[Wrong]** The people in the living room are talking about the baseball game last night because they are huge baseball fans.

**All of the above three fake statements are not acceptable. Even though they are plausible, they are completely out-of-context given the video clip.** A good fake statement is:

[Fake] **[Good]** The people in the living room are trying to make the brunette in the kitchen calm down and get dressed because she is upset about breaking up with the man in suite.

**Suggestion:** If your real statement has some overlap with the subtitles (e.g. 'message', 'calling'), make sure your fake statement also has the same level of overlap with the subtitles (e.g. 'break up'). For example, if your real statement and subtitles has 8 words in common while your fake statement and subtitles only have 2 words in common, then your real and fake statements are not confusing at all since you can tell which one is real by simply comparing the subtitles with the statements without really understanding the scene.

Figure 6. Instructions for writing fake statements.

Watch the given video, (1) cut a single-scene interval using the slider, and then (2) write real and fake statements:

## Your Job



00:17

01:06

check interval

Real-Fake Statement Pair 1:

REAL statement 1 ... (10 words to 40 words)

FAKE statement 1 ... (10 words to 40 words)

Real-Fake Statement Pair 2:

REAL statement 2 ... (10 words to 40 words)

FAKE statement 2 ... (10 words to 40 words)


Real-Fake Statement Pair 3:

REAL statement 3 ... (10 words to 40 words)


FAKE statement 3 ... (10 words to 40 words)

**Before you submit, please check again whether the interval shown in the slider matches the complete scene you have in mind.** You can press the "check interval" button to watch the video interval you cut.


Figure 7. User interface for workers to cut an interval from the video and write three pairs of real and fake statements.

	
00:00:13,640 --> 00:00:20,740 Monsieur Dilek Anthony	00:00:29,789 --> 00:00:33,729 such as I require from your good self I
00:00:20,750 --> 00:00:24,800 Madame Sabine de Barra I'm very happy to	00:00:33,739 --> 00:00:35,720 brought with me a selection of
00:00:24,810 --> 00:00:27,800 meet you I am directed by the office of	00:00:35,730 --> 00:00:40,000 perennials I thought we might exchange
00:00:27,810 --> 00:00:29,779 buildings of Versailles to order shrubs	
<p>(pos) A woman in a leather jacket approaches a man as he is resting on a bench. ✓</p> <p>(neg) A woman in a leather jacket approaches a man on a bench alone crying. ✗</p> <p>(pos) A woman in a leather does a curtsy to the man resting on the bench in front of her. ✓</p> <p>(neg) A woman in a leather jacket gives a hand salute to the man resting on the bench. ✗</p> <p>(pos) The man relaxing on the bench looks annoyed at the woman who has disturbed him. ✗</p> <p>(neg) The man in grey shirt talks with the man in green t-shirt that he hope he didn't throw him off his game. ✗</p>	


(a)

	
00:00:00,000 --> 00:00:03,230 cops never bother iron John something	00:00:18,020 --> 00:00:31,510 you heard the woman
00:00:03,240 --> 00:00:06,680 must be going on it's red	00:00:31,520 --> 00:00:36,460 place creeps me out man hot we think
00:00:06,690 --> 00:00:10,510 we don't know that you have to be sure	00:00:36,470 --> 00:00:38,100 they're gonna do it iron Jonathas boys
00:00:10,520 --> 00:00:18,010 they look weak let's take them all right	00:00:38,110 --> 00:00:40,000 cares hold over sir
<p>(pos) The woman are suspicious of the reasons why a man is being bothered by the cops. ✓</p> <p>(neg) The woman are assured by the presence of the cop in the company of the man. ✓</p> <p>(pos) The woman evaluate the vulnerability of the men and assume they can prevail. ✓</p> <p>(neg) The woman are fearful of the men they watch, and they're scared to act. ✓</p> <p>(pos) One man is apprehensive by the location he is shares in the company of the other man. ✓</p> <p>(neg) The woman in the purple and blue pajamas jumps on-top of the man laying in the bed. ✗</p>	

(c)

	
00:00:03,085 --> 00:00:08,328 Uh, I'll just... fire away then,	00:00:21,897 --> 00:00:24,020 you ever thought of having, um,
00:00:03,085 --> 00:00:08,328 shall I ?	00:00:24,108 --> 00:00:27,856 more, uh, horses in it.
00:00:09,925 --> 00:00:11,965 Right.	00:00:30,991 --> 00:00:34,526 Uh, well, we would have liked to,
00:00:15,431 --> 00:00:18,718 The film's great, and, um,	00:00:34,620 --> 00:00:39,959 but it was, um, difficult,
00:00:18,811 --> 00:00:21,812 I just was wondering whether...	00:00:34,620 --> 00:00:39,959 obviously, being set in space.
<p>(pos) Brunette man sits across a table from a woman in a suit and informs her he is going to ask several questions. ✓</p> <p>(neg) Brunette man sits across a table from a woman in a suit and informs her he is going to sit in silence. ✗</p> <p>(pos) Brunette man asks the woman in a suit if she would have preferred to add more animals to something and she confirms. ✓</p> <p>(neg) Woman in suit asks the brunette man if she would have preferred to add more animals to something and he confirms. ✓</p> <p>(pos) Woman in suit explains to brunette man that having many animals present is difficult. ✓</p> <p>(neg) The man in a black and white top is curious about if the boy in bed has finished what he needed to do for the day. ✗</p>	

(b)

	
00:00:16,210 --> 00:00:18,210 Give me the child	00:00:27,520 --> 00:00:29,520 Then Death will come to you both
00:00:22,570 --> 00:00:24,570 No	00:00:32,549 --> 00:00:34,550 No
<p>(pos) A blonde woman feels really scared and screams when a man comes flying and hits a man in a white shirt. ✓</p> <p>(neg) A blonde woman feels relief and laughs when a man comes flying to rescue her and hits a man in a white shirt. ✗</p> <p>(pos) A blonde woman is running away from a man who came flying trying to protect a baby from him, the baby is upset and cries. ✓</p> <p>(neg) A blonde woman is running away from a man who came flying trying to protect a baby from him, the baby just smiles unaware of everything. ✗</p> <p>(pos) A man wearing a white t-shirt sacrifices himself for a woman and a baby, a blonde woman gets really upset when he falls of the cliff and cries. ✓</p> <p>(neg) The woman and the man come close to fighting over the man's desires. ✓</p>	

(d)

Figure 8. Examples on movie clips. The pos/neg at the beginning of each statement indicates its ground truth. The ✓ or ✗ at the end of each statement indicates the system's prediction. ✓ means the system judges the statement as positive, and ✗ means negative.





00:00:01,169 --> 00:00:02,429 Block me, come on!	00:00:15,899 --> 00:00:18,519 Guys, upstairs now. Let's go.
00:00:02,459 --> 00:00:04,949 Damn, they must have snuck out again.	00:00:18,549 --> 00:00:22,349 Into pj's, into bed. Honey, they're fine.
00:00:05,049 --> 00:00:07,819 Again? How often does this happen?	00:00:22,389 --> 00:00:24,549 You just... you worry too much.
00:00:08,189 --> 00:00:10,469 Boys, get in here now!	00:00:24,579 --> 00:00:26,199 And for good reason.
00:00:10,489 --> 00:00:11,579 Honey, you know how slippery they are.	00:00:26,229 --> 00:00:27,959 Someone could have driven off with them,
00:00:11,629 --> 00:00:13,249 It's like trying to herd cats.	00:00:27,989 --> 00:00:28,949 and you wouldn't have even noticed.
00:00:13,309 --> 00:00:15,869 Tom, it's 9:00 at night.	

(pos) The man in the blue shirt is frustrated because the kids keep running outside when he isn't looking. ✓  
 (neg) The man in the blue shirt is frustrated because the kids keep getting into cookies when he isn't looking. ✗  
 (pos) The woman in the grey suit is upset that the kids weren't in house because it was dark out. ✓  
 (neg) The woman in the grey suit is upset that the kids weren't in the house because they are sick. ✗  
 (pos) The man in the blue shirt thinks the woman in the grey suit worries too much about the kids. ✓  
 (neg) The man in a black shirt and the lady in a brown hair and black dress sit outdoor during their conversation. ✗

(a)



00:00:00,000 --> 00:00:01,009 Me, too.	00:00:17,449 --> 00:00:20,959 It's a non-smoking room!" And I was all, "Hell, no, this is a Cuban!"
00:00:01,500 --> 00:00:02,839 I miss my wife.	00:00:21,699 --> 00:00:24,329 Of course, eventually, I did put it out.
00:00:04,599 --> 00:00:08,129 Hey, let's go around the table and say what our favorite part was.	00:00:25,369 --> 00:00:26,389 Did I put it out?
00:00:08,229 --> 00:00:10,539 Mine was that thing with the typewriter.	00:00:27,129 --> 00:00:28,149 I put it out.
00:00:10,919 --> 00:00:13,869 I mean, she made some spelling mistakes, but still.	00:00:29,209 --> 00:00:30,509 Did I put it out?
00:00:14,799 --> 00:00:17,439 Ooh, and you guys were all, "Barney, put out the cigar!"	00:00:31,750 --> 00:00:33,000 I put it out.
	00:00:34,699 --> 00:00:35,259 Did I put it out?

(pos) The man with the black top and the man with the black hair react with disgust when the blonde man recounts his favorite act. ✓  
 (neg) The man with the black top and the man with the black hair react with amusement when the blonde man recounts his favorite act. ✓  
 (pos) The blonde man is telling the men at the table about his favorite parts from the event they attended. ✓  
 (neg) The man in the green shirt is telling the men at the table about his favorite parts from the event they attended. ✗  
 (pos) The man with black hair reacts with annoyance when the man with blonde hair is trying to remember if he put out his cigar. ✓  
 (neg) The man in the suit and the woman in the blue shirt reminisce about how times have been since they saw each other last. ✗

(b)



00:00:00,000 --> 00:00:00,240 Uh, 9.	00:00:22,341 --> 00:00:26,641 I know, me too. Hey, what if we went away for the whole weekend? No interruptions.
00:00:02,030 --> 00:00:03,410 But it's dark out.	00:00:26,804 --> 00:00:28,933 And we could be naked the entire time.
00:00:03,573 --> 00:00:06,783 Um, well, that's because you always sleep till noon, silly.	00:00:29,515 --> 00:00:30,555 All weekend?
00:00:08,161 --> 00:00:10,750 This is what 9 looks like.	00:00:30,725 --> 00:00:32,424 - That's a whole lot of naked.
00:00:11,831 --> 00:00:14,250 I guess I'll get washed up then.	- Mm-hm.
00:00:14,542 --> 00:00:16,881 Watch that sunrise.	00:00:33,102 --> 00:00:36,610 Yeah, I could say I have a conference and you can say you have a chef thing.
00:00:19,380 --> 00:00:22,170 I'm really getting tired of sneaking around all the time.	

(pos) A man in a grey shirt has a confused conversation about what time of the day it is. ✓  
 (neg) A man in a grey shirt has a confused conversation about where he currently living. ✓  
 (pos) A man and a women in a robe have a conversation about going away for a weekend. ✓  
 (neg) A man and a women in a robe have a conversation about a week long stay-cation. ✗  
 (pos) A confused man in a grey shirt goes into to the bathroom to wash-up. ✓  
 (neg) The woman in the red dress is thrilled that she got divorced and took all of her ex-husband's money in the divorce. ✗

(c)



00:00:00,000 --> 00:00:00,780 by looking through their personal things?	00:00:15,060 --> 00:00:16,390 Why are you so scattered lately?
00:00:00,790 --> 00:00:03,480 Not just now. Since the invention of things.	00:00:16,390 --> 00:00:20,040 Aw, he's just nervous because of his poetry reading tonight.
00:00:03,480 --> 00:00:05,310 Is that one of Claire's brownies?	00:00:20,410 --> 00:00:21,960 But don't worry, papi.
00:00:05,310 --> 00:00:07,130 No, they're delicious. Must be Cam's.	00:00:21,960 --> 00:00:24,320 I am going to be there to support you.
00:00:07,130 --> 00:00:09,000 - Oh, no. - Don't worry. There's more.	00:00:24,990 --> 00:00:26,570 I don't want my mom there.
00:00:09,000 --> 00:00:11,540 No! That's where my backpack is.	00:00:26,930 --> 00:00:28,490 I'm exploring some darker themes
00:00:11,550 --> 00:00:13,520 Cam drove me and Luke home.	00:00:28,490 --> 00:00:28,510 I'm not sure she's ready for.
00:00:13,740 --> 00:00:15,060 I must've left it in his car.	

(pos) The kid in white is worried because he can't find his backpack. ✗  
 (neg) The kid in white is worried because he can't find his homework. ✗  
 (pos) The brunette woman reassures the kid in white that she'll go to his poetry reading. ✓  
 (neg) The brunette woman reassures the man in blue that she'll go to his poetry reading. ✓  
 (pos) The brunette woman is curious who made the brownie that the man in blue is eating. ✓  
 (neg) The woman in the blue shirt reacts with confusion when the man in the red shirt informs her that he didn't order any lemonade. ✗

(d)

Figure 9. Examples on TV show clips. The pos/neg at the beginning of each statement indicates its ground truth. The ✓ or ✗ at the end of each statement indicates the system's prediction. ✓ means the system judges the statement as positive, and ✗ means negative.