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.](image)

![Figure 2. Distribution of negative statement lengths.](image)

A.2. Statement Content

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

<table>
<thead>
<tr>
<th>Type</th>
<th>Most Common Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>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</td>
</tr>
<tr>
<td>Verb</td>
<td>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</td>
</tr>
<tr>
<td>Adj</td>
<td>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</td>
</tr>
</tbody>
</table>

![Table 1. Most common words in positive statements.](image)

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.

![Figure 3. Distribution of video lengths from TV shows.](image)

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.

<table>
<thead>
<tr>
<th>Task 1: Cut out an interesting and self-contained interval from the given video clip:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 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.</td>
</tr>
<tr>
<td>• The <strong>hard requirement</strong> for the length of the interval you cut is <strong>between 15 and 30 seconds</strong>. Do not cut your interval out of this range.</td>
</tr>
<tr>
<td>• The interval you cut should be a <strong>single meaningful scene</strong></td>
</tr>
<tr>
<td>• 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.</td>
</tr>
<tr>
<td>• Your interval should <strong>contain only one scene</strong>. That means, your interval should not contain flashback, change of scenes (locations), opening songs, etc.</td>
</tr>
<tr>
<td>• 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 0:04–0:40, or cut it even shorter to 0:25–0:40. (the interval does not have to contain everything, as long as it is self-contained)</td>
</tr>
<tr>
<td>• In this job you only need to cut one scene and write three pairs of statements for it.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task 2: Infer “real” and “fake” statements from the interval you just cut:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Real statements:</strong></td>
</tr>
<tr>
<td>• Should be <strong>inferred</strong> from the interval you cut: e.g. human interaction, dialog, human relation, intention, common sense, reasoning.</td>
</tr>
<tr>
<td>• <strong>The statement you write should not directly appear in subtitles or video scenes.</strong> Write something that is implicitly expressed in the video clip.</td>
</tr>
<tr>
<td>• Your statements should be only based on the scene (what you see) and the subtitles. <strong>Do not write about the audio information</strong> (e.g. sound in the background, character raising voice, character speaking another language that is not reflected in the subtitles)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fake statements:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Should be <strong>contradictory</strong> to the interval you cut, by <strong>twisting details</strong> 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.</td>
</tr>
<tr>
<td>• However, your fake statement should be <strong>confusing with the real one given the interval you cut</strong>. That means, your fake statement should not be <strong>completely out-of-context</strong> with the real one. It should be very relevant to the video and the real statement. (more examples later)</td>
</tr>
<tr>
<td>• Your fake statement should look <strong>very similar</strong> to the real statement, in terms of length, structure, writing style, wording, etc.</td>
</tr>
</tbody>
</table>

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 in 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 are out (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 asks 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 her 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**

![Video Interface](image)

**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.
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.