TWIST Dialogue Annotation Tool User Guide

Brian Plüss b.pluss@open.ac.uk

July 2012

Contents

1	Ove	erview	2
2	Get	ting Started	3
	2.1	Running the Annotation Tool	3
	2.2	First Session	3
	2.3	Annotation Procedure Overview	4
	2.4	Support for the Annotation Workflow	6
3	Am	notating Dialogues	7
	3.1	First Stage: Segmenting Turns	8
	3.2	Second Stage: Annotating Content Features	10
4	Ado	litional Features	12
	4.1	Converting Dialogues in Text Format to XML	12
	4.2	Changing the Annotation Stage	13
	4.3	Modifying the Annotator Profile	14
	4.4	Modifying the Annotator Familiarity	
5	Get	ting Help	15

1 Overview

The annotation tool consists of two main windows:

- The Dialogue window (Figure 1) is the main window in the application. It shows the dialogue being annotated, a description of the the context in which it took place and the annotations for each turn.
- The Turn Annotator window (Figure 2) opens from the main window by clicking on one of the annotation cells. It shows a single dialogue turn and allows adding or removing segments, and creating or modifying annotations for each segment in the turn.

ile	Annot	tation Help	
Cont	ext		
riso	n Servic	Jeremy Paxman questions former UK Home Secretary Michael Howard e, Derek Lewis, about the dismissal of the governor of Parkhurst Prisor attention in the media, as a result of accusations by Lewis that Howard	
¥	Speaker	Turn	Annotated Turn
)		Right, uh can you help us with this then? You stated in your statement that the Leader of the Opposition had said that I (that is, you) personally told Mr Lewis that the governor of Parkhurst should be suspended immediately, and that when Mr Lewis objected as it was an operational matter, " I threatened to instruct him to do it ". Derek Lewis says" Howard had certainly told me that the Governor of Parkhurst should be suspended, and had threatened to overrule me ". Are you saying Mr Lewis is lying ?	0 :: Init-Inform :: On-Topic :: Objective :: Accurate :: You stated in your statement that the Leader of the Opposition had said that ! (that is, you) personally told Mr Lewis that the governor of Parkhurst should be suspended immediately, and that when Mr Lewis objected as it was an operational matter, " I threatened to instruct him to do it ". 1 :: Init-Inform :: On-Topic :: Objective :: Accurate :: Derek Lewis says " Howard had certainly told me that the Governor of Parkhurst should be suspended , and had threatened to overrule me ".
			2 :: Init-InfoReq :: On-Topic :: Loaded :: Reasonable :: Are you saying Mr Lewis is lying ?
L	IE	I have given a full account of this , and the position is what I told the House of Commons , and let me tell you what the position is –	3 :: Resp-Inform :: Irrelevant :: Objective :: Accurate :: Incomplete :: @2 :: I have given a full account of this , and the position is what I told the House of Commons
2	IR	(Interrupting) So you are saying that Mr Lewis lied ?	4 :: Init-InfoReq :: On-Topic :: Loaded :: Reasonable :: So you are saying that Mr Lewis lied ?
3		Let me tell you exactly what the position is . I was entitled to be consulted and I was consulted , I was entitled to express an opinion and I did express an opinion . I was not entitled to instruct Derek Lewis what to do , and I did not instruct him what to do and you will understand and recail that Mr Marriot was not suspended , he was moved , and Derek Lewis told the Select Committee of the House of Commons that it was his opinion , Derek Lewis 's opinion , that he should be moved immediately . That is what happened .	5 :: Resp-Accept :: @4 :: Let me tell you exactly what the position is . 6 :: Resp-Inform :: Relevant :: Objective :: Inaccurate :: Incomplete :: @4 :: I was entitled to be consulted and I was consulted , I was entitled to express an opinion and I did express an opinion . I was not entitled to instruct Derek Lewis what to do , and I did not instruct him what to do and you will understand and recall that Mr Marrior was not suspended , he was moved , and Derek Lewis told the Select Committee of the House of Commons that it was his opinion , Derek Lewis 's opinion , that he should be moved immediately . That is what happened .
1		Mr Lewis says " I (that is , Mr Lewis) , told him what we had decided about Marriot , and why . He , (that is , you) , exploded . Simply moving the governor was politically unpalatable , it sounded indecisive , it would be	

Figure 1: Main Dialogue Window

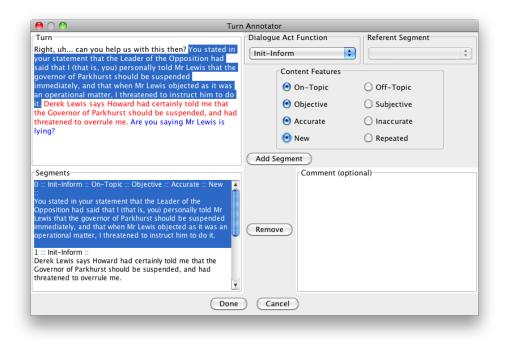


Figure 2: Turn Annotator Window

2 Getting Started

2.1 Running the Annotation Tool

There are three scripts to lunch the tool, depending on your operating system:

- In Windows: double click on runTool-Windows.bat
- In MacOS: double click on runTool-MacOS.command
- In Linux: in a terminal run ./runTool-Linux.sh

2.2 First Session

The first time you use the annotation tool, you will be asked to provide your name and to answer a few questions about your background (Figure 3). This information is important for the analysis of the your annotations, so please answer every question¹. It will take just a few seconds!

¹The background information can be changed at any time by using the menu option **Annotation** | **Annotator Profile...** (see details in Section 4.3).

00	Annotator Profil	e		
	Please answer the questions below:			
3	1. What is your name?)	
~	2. What is your first language?	🔘 English	Other	
	3. What is your nationality?	🔘 British	O Other	
	4. How many years have you lived in the UK?	0		
	5. Previous experience in dialogue annotation	? 🔘 Yes	🔘 No	
		ОК		Cancel

Figure 3: Annotator Profile

After filling in this information, you can open a dialogue file and start your annotation.

2.3 Annotation Procedure Overview

The annotation is carried out in two stages. Most likely, you will be asked to carry out only one of them. The general steps to annotate a dialogue file are as follows:

• First Stage: Turn Segmentation

- 1. Open a dialogue file
- 2. Click on a cell under the Annotated Turn column
 - a. Highlight a segment (part of a turn)
 - b. Select the Function of the segment
 - c. If it is a responsive segment, select a $\mathsf{Referent}\ \mathsf{Segment}$
 - d. Click Add $\mathsf{Segment}$ to add the new segment
 - e. Repeat the steps above until the turn is fully segmented
 - f. Click **Done** to complete the segmentation of the turn
- 3. Repeat step 2. until all the turns are segmented
- 4. Save the annotation: File | Save Dialogue
- Second Stage: Content Feature Annotation
 - 1. Open a dialogue file
 - 2. Click on a cell under the Annotated Turn column

- a. Click on a segment from the Segments list
- b. For each row in Content Features, select the option that applies
- c. Repeat the steps above until every segment is annotated
- d. Click **Done** to complete the annotation of the turn
- 3. Repeat step 2. until all the turns have been annotated
- 4. Save the annotation: File | Save Dialogue

Regardless of the stage, the first time you open a dialogue file, you will be asked a few questions about your familiarity with the dialogue, its context and the participants (Figure 4). This will happen only once for each file you annotate².

3	You have not worked on this file before. Please read the given in Context and answer the questions below.	informati	ion				
~*	BBC presenter Jeremy Paxman questions former UK Home Secretary Michael Howard with respect to a meeting in 1995 between Howard and the head of the Prison Service, Derek Lewis, about the dismissal of the governor of Parkhurst Prison, John Marriott, due to repeated security failures. The case was given considerable attention in the media, as a result of accusations by Lewis that Howard had instructed him, thus exceeding the powers of his office.						
	1. Have you watched or read this interview before?	() Yes	O No	O Not Sure			
	2. Are you familiar with the political/historical context?	🔘 Yes	🔘 No	🔘 Not Sure			
	3. Have you heard of the interviewer?	🔘 Yes	🔘 No	🔘 Not Sure			

Figure 4: Familiarity of the annotator with the dialogue and its context

Dialogue files are found in the folder called "data". They are named following the convention <interviewer>-<interviewee>.xml (e.g. paxmanhoward.xml). The first time you save an annotated dialogue, you will be asked for a new filename. The annotation tool automatically suggests appending the suffix "-annot" to the original name (e.g. paxman-howardannot.xml). It is strongly recommended that you follow this convention

²This information can be changed at any time by using the menu option File | Annotator Familiarity... (see details in Section 4.4).

when first saving annotated files. This preserves the original file in case you need to start the annotation from scratch.

When annotating longer dialogues, you should save your work often. Once you have saved the annotated file with a new name, you can quickly save changes by using the **File** | **Save Dialogue** menu option or the keyboard shortcut **Ctrl+S** (**Command+S** in MacOS).

2.4 Support for the Annotation Workflow

The annotation tool is configured to open one of the files in the "data" folder after being launched for the first time, and to operate according to the annotation stage you have been asked to carry out. Once you have annotated the first dialogue, the menu option **File** | **Open Next Dialogue** will automatically save the current annotated file and open the next one in the dataset (see Figure 5).

9 🔿 🔿		TWIST Dialogue Annotation Tool	
File Annotation Hel	р		
Open Next Dialogue	ЖN		
Open Dialogue Save Dialogue Save Dialogue As	жо жs	estions former UK Home Secretary Michael H ut the dismissal of the governor of Parkhurs dia, as a result of accusations by Lewis that I	
Annotator Familiarity.	can yo ar of t	help us with this then ? You stated in your stater e Opposition had said that I (that is , you) perso	
Convert to XML		e governor of Parkhurst should be suspended at when Mr Lewis objected as it was an operation	
Exit	ЖE	d to instruct him to do it " . Derek Lewis says " H e that the Governor of Parkhurst should be suspe	
and had thre	ateneo	to overrule me " . Are you saying Mr Lewis is lyin	

Figure 5: Opening the next dialogue in the dataset

Once all the dialogues have been annotated, this option will invite you to submit the annotated data (see Figure 6). Your computer needs to be connected to the internet in order to send the data. If that is not the case, you can submit the data later by using the menu option **Annotation** | **Submit Data...**, shown in Figure 7.

If you have to go back to one of the files you have annotated earlier in the workflow, you will need to use the **File** | **Open Dialogue...** menu option and look for the annotated file in the "data". Note that this will move

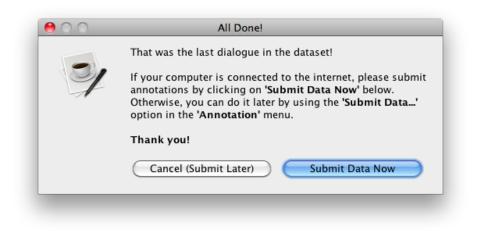


Figure 6: Completing the annotation workflow

0) 🔘		TWIST Dialogue Annotation Tool
File	Annotation	Help	
Cont	Submit Dat	a	
BBC p Prisor consi	✓ First Stage	ge	stions former UK Home Secretary Michael H t the dismissal of the governor of Parkhurs ia, as a result of accusations by Lewis that
# 0	that the	uh can you Leader of the	help us with this then ? You stated in your state Opposition had said that I (that is , you) pers governor of Parkhurst should be suspended

Figure 7: Submitting the annotated data

you backwards in the annotation workflow, so you might need to use the menu option **File** | **Open Next Dialogue** several times to go back to your last annotated dialogue file.

3 Annotating Dialogues

After opening a file, the context of the dialogue will be shown at the top of the main window and the transcript will appear on the table, as shown in Figure 1. Dialogues are divided in turns. The number of each turn is on the first column and the speaker on the second (IR for the interviewer and IE for the interviewee).

The third column shows the annotations for each turn. Depending on the

stage of annotation you were asked to perform, this column can be initially empty or contain information. The annotations of a turn are organised in segments and have the following structure:

```
<Segment ID> :: <Function Label> :: @<Referent Segment ID> ::
<Content Feature 1> :: ... :: <Content Feature N> ::
<Segment Transcript>
```

If you were asked to do the first stage of the annotation, the annotations column will be initially empty. The annotations you produce will only have a segment ID, a function label, an optional referent segment ID (only for responsive segments) and the segment transcript.

If you were asked to do the second stage of the annotation, the annotations column will already have segments annotated with dialogue act functions, referent segment IDs and the segment transcript. Your job will be to annotate the content features for each segment.

In either case, clicking on the cell under Annotated Turn for a turn's row opens the Turn Annotator window (Figure 2) which allows you to add or edit annotations for that turn.

3.1 First Stage: Segmenting Turns

To segment a turn, click on the cell under Annotated Turn on the row of the turn. This opens the Turn Annotator window (Figure 8):

• To create a new segment, in the Turn text area, select the stretch of the turn that constitutes a segment, choose the corresponding value in Function and, if it is a responsive dialogue act, choose a previously annotated segment in Referent Segment³.

Then click on the Add Segment button. This will add a new entry under Segments with ID "-1" and the values you selected for function and referent segment, separated with "::" (unique segment IDs are generated once you save the annotation for this turn). The stretch of text for the new segment will be painted in blue or red in the Turn text area, as shown in Figure 2. This will help in segmenting the rest of the turn, as segments can not overlap.

• To modify an existing segment, select it from the list of segments. The corresponding stretch will be highlighted in the Turn text area and

 $^{^{3}}$ For details on how to select segments and adequate values for dialogue act function and referent segments, see the *Annotation Guidelines*.

\varTheta 🔿 🔿 Turr	Annotator
Turn	Dialogue Act Function Referent Segment
Mr Lewis says I (that is, Mr Lewis), told him what we had decided about Marriot, and why. He, (that is, you), exploded. Simply moving the governor was politically unpalatable, it sounded indecisive, it would be seen as a fudge. If I did not change my mind and suspend Marriot he would have to consider overruling me. You can't both be right.	(Init-Inform 🗘
	Add Segment
Segments	Comment (optional)
	Remove
Done	Cancel

Figure 8: The turn annotator window for segmenting turns (First Stage)

the annotation values will appear in Function and, when applicable, in Referent Segment (see Figure 2). You can change these values to update the annotation. To modify the stretch of the segment within the turn, you will have to remove the segment and create a new one.

• To remove an existing segment, select it from the list and click on Remove. When you remove a segment that is referenced by other segments, a warning will pop up and you will be asked to confirm the operation (Figure 9). If you confirm the removal, the lost references will be shown as "@??". As all responsive segments must have a referent, you will have to edit any segment with missing referents after the removal.

Adding or removing segments in the middle of an annotated dialogue automatically adjusts the numbering of subsequent segments and of any references.

Adding a segment fails when the new segment overlaps with an existing one and an error message is displayed. Also, adding and changing segments fail when none of the values in **Function** have been selected, or when the

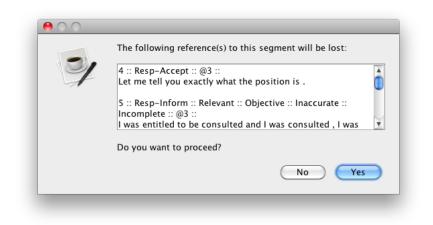


Figure 9: Confirmation request when removing a segment with references.

segment was classified as responsive, but none of the options in Referent Segment have been chosen.

If you have any observations regarding the segmentation of a turn, you can write a comment in the Comment (optional) text area.

Once you have identified all the segments in the turn, confirm the annotation by clicking on Done. This closes the turn annotation window and the segments will appear next to the turn, under the column Annotated Turns in the main window.

If you wish to discard the annotation (of if you have not made any changes on an existing annotation), click on Cancel. This will close the Turn Annotator window. Note that any changes or new annotations made since last opening the annotator window will be lost if you close it with the Cancel button.

3.2 Second Stage: Annotating Content Features

In this stage, turns in the dialogue are already segmented and each segment is annotated with a function label and, if applicable, a referent segment. To annotate the content features in a turn, click on the cell under Annotated Turn on the row for the turn. This opens the Turn Annotator window (Figure 10):

• To annotate the content features in a segment, select the segment from the list. The corresponding stretch will be highlighted in the Turn text area and the annotation values will appear in Function and, when applicable, in Referent Segment. These fields will be disabled and their values can not be modified.

You must select one option for each row under **Content Features**. Selecting these values autimatically updated the annotation of the segment. Note that some segments might not have associated content features, in which case no annotation is required⁴.

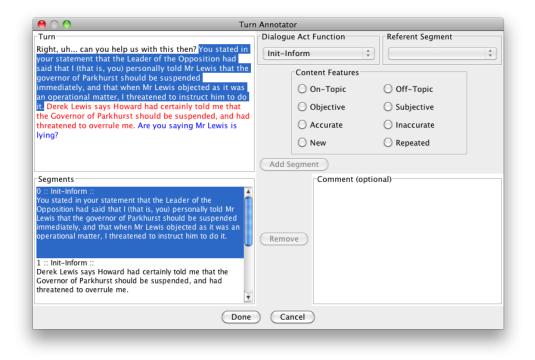


Figure 10: The turn annotator window for content features (Second Stage)

The annotation of content features for a segment fails if a choice has not been made for any of the rows under **Content Features** has not been .

If you have any observations regarding the annotation, you can write a comment in the **Comment (optional)** text area.

Once you have annotated the content features for all the segments in the turn, confirm the annotation by clicking on **Done**. This will close the turn annotation window and the updated segments will appear next to the turn, under the column **Annotated Turn** in the main window.

If you wish to discard the annotation (of if you have not made any changes on an existing annotation), click on Cancel. This will close the Turn Annotator window. Note that any changes or new annotations made since last opening the annotator window will be lost if you close it with the Cancel button.

 $^{{}^{4}}$ For details on how to select adequate values for content features, see the Annotation Guidelines.

4 Additional Features

4.1 Converting Dialogues in Text Format to XML

The annotation tool includes a function for converting dialogues in text format to the XML files used as input to the annotation. To access this feature, use the menu option **File** | **Convert to XML...** (Figure 11). This will open a window in which you can select one or more files for conversion. The files must have extension ".txt" and converted files will have the same name and extension ".xml".

00	TWIST Dialogue Annotation Tool
File Annotation Help	
Open Next Dialogue	
Open Dialogue #O Save Dialogue #S Save Dialogue As	estions former UK Home Secretary Michael How ut the dismissal of the governor of Parkhurst Pri dia, as a result of accusations by Lewis that Hov
Annotator Familiarity	help us with this then? You stated in your statement position had said that I (that is, you) personally told por of Parkhurst should be suspended immediately,
Convert to XML	
Exit #E	wis objected as it was an operational matter, I t him to do it. Derek Lewis says Howard had certain ernor of Parkhurst should be suspended, and had
threatened to over	ule me. Are you saying Mr Lewis is lying?

Figure 11: Converting dialogue in text files to XML format

The text files should follow the structure below, with the speakers being either IR or IE:

```
CONTEXT<Blank Space><Text>
<Empty Line>
<Speaker 1><Blank Space><Text>
<Empty Line>
<Speaker 2><Blank Space><Text>
<Empty Line>
<Speaker 1><Blank Space><Text>
<Empty Line>
<Speaker 2><Blank Space><Text>
...
<End of File>
```

The text in CONTEXT and in each turn can span several lines, with an empty line marking the change of turn (and speaker). An example follows:

CONTEXT During the American Presidential campaign in January 2008, Fox News host Bill O'Reilly interviews Hermene Hartman, the editor of an African-American newspaper in Chicago, about Obama's pastor Jeremiah Wright and his connections with Nation of Islam's leader Louis Farrakhan.

IR How would you describe Dr Wright's church?

IE It's a middle-class church. It is a superb church. Reverend Wright started a church with 87 people; today, has 8,000 in that particular congregation. United Church of Christ is basically a white denomination. And I think there's been just a lot of miscasting here. Seventy ministries within the church, to include Girl Scouts, prison outreach, marital counselling, education, children's counselling, a lot of Adopt-A-School. They have done a lot to empower that community and to improve that community.

IR OK. But you could make the same argument about Louis Farrakhan, that he's done, you know, some good things, yet you know, he's anti-Semitic in his rhetoric and sometimes anti-white or whatever. And-

IE (Interrupting) But that is, that is not Jeremiah Wright.

IR No, but it is association there. And the association, you can draw your own conclusion.

The result of the converting this dialogue is an XML file with the following content:

```
<DIALOGUE>
   <!--automatically generated by the TWIST Annotation Tool-->
   <CONTEXT DESCRIPTION="During the American Presidential campaign in January
        2008, Fox News host Bill O'Reilly interviews Hermene Hartman, the
        editor of an African-American newspaper in Chicago, about Obama's
         pastor Jeremiah Wright and his connections with Nation of Islam's
        leader Louis Farrakhan. "/>
      <TURN SPEAKER="IR" SPEECH="How would you describe Dr Wright 's church ? "/>
      <TURN SPEAKER="IE" SPEECH="It 's a middle - class church . It is a superb
           church . Reverend Wright started a church with 87 people ; today ,
           has 8,000 in that particular congregation . United Church of
           Christ is basically a white denomination . And I think there
           been just a lot of miscasting here . Seventy ministries within the
           church , to include Girl Scouts , prison outreach , marital
           counselling , education , children 's counselling , a lot of
           Adopt - A - School . They have done a lot to empower that
           community and to improve that community . "/>
      <TURN SPEAKER="IR" SPEECH="OK . But you could make the same argument
           about Louis Farrakhan , that he 's done , you know , some good
           things , yet you know , he 's anti - Semitic in his rhetoric and
           sometimes anti - white or whatever . And - "/>
      <TURN SPEAKER="IE" SPEECH="( Interrupting ) But that is , that is not
           Jeremiah Wright . "/>
      <TURN SPEAKER="IR" SPEECH="No , but it is association there . And the
           association , you can draw your own conclusion . "/>
</DIALOGUE>
```

4.2 Changing the Annotation Stage

Your version of the annotation tool will come configured by default for the annotation stage you were asked to perform. However, this can be changed by using the options in the **Annotation** menu (Figure 12):

• Annotation | First Stage allows enabling and disabling turn segmentation and the annotation of dialogue act functions and referents.

• Annotation | Second Stage allows enabling and disabling the selection of content features for segments.

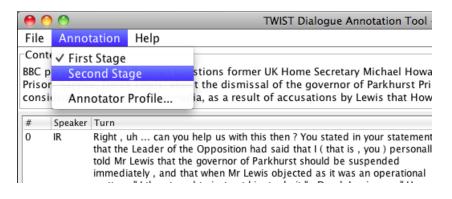


Figure 12: Enabling and disabling annotation stages

4.3 Modifying the Annotator Profile

You can modify the information you provided for the annotator profile the first time you run the annotation tool. This is useful in case you have made a mistake or skipped one of the answers. To re-enter the information, use the option **Annotation** | **Annotator Profile...** from the menu bar (Figure 13).

0) 🔘 👘			TWIST Dialogue Annotation Tool
File	Anno	tation	Help	
BBC p Prisor consi	Jee	ond Sta	ge Profile	stions former UK Home Secretary Michael Howa t the dismissal of the governor of Parkhurst Pri ia, as a result of accusations by Lewis that How
#	Speaker	Turn		
0	IR	that the told Mr	Leader of th Lewis that the ately , and th	help us with this then ? You stated in your statement e Opposition had said that I (that is , you) personall e governor of Parkhurst should be suspended at when Mr Lewis objected as it was an operational d to instruct him to do it " Derek Lewis says " Howa

Figure 13: Re-entering the annotator profile information

4.4 Modifying the Annotator Familiarity

You can modify the information you provided regarding your familiarity with the context of the dialogue the first time you opened a file. This is useful, for instance, in case you have made a mistake. To re-enter the information, use the option **File** | **Annotator Familiarity...** from the menu bar (Figure 14).

ile Annotation Help		
Open Next Dialogue		
Open Dialogue #O Save Dialogue #S Save Dialogue As	estions former UK Home Secretary Michael How ut the dismissal of the governor of Parkhurst Pr dia, as a result of accusations by Lewis that Hov	
Annotator Familiarity	help us with this then? You stated in your statement to position had said that I (that is, you) personally told	
Convert to XML	nor of Parkhurst should be suspended immediately, wis objected as it was an operational matter, I	
Exit #E	t him to do it. Derek Lewis says Howard had certain renor of Parkhurst should be suspended, and had	

Figure 14: Re-entering the annotator familiarity information

5 Getting Help

The **Help** menu gives you access to a quick set of instructions via **Help** | **Instructions** (Figure 15).

The menu also has options for opening this guide (**Help** | **User Guide**), the annotation guidelines (**Help** | **Annotation Guidelines**), and for showing the credits and contact information (**Help** | **About**).

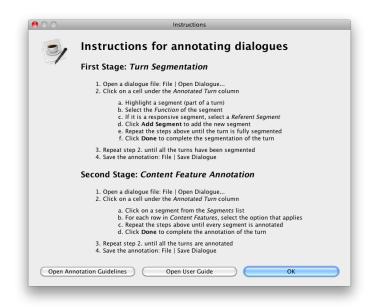


Figure 15: Overview of instructions for annotating dialogues

To report any problems and bugs, or if you have questions or suggestions, please send an email to b.pluss@open.ac.uk.