Where's The Answer: Dialogue Annotation

UE 805 - Supervised Project 2018-2019

Edited By

Marta Carletti
Lea Dieudonat
Yiting Tsai

University of Lorraine
Nancy

May 28, 2019

supervised by :
Maxime Amblard
Maria Boritchev

Contents

Introduction

The study of dialogue is of major interest in linguistics as human interactions compose the basis of our communication. Dialogue takes place between two or more interlocutors, therefore it involves a continuous exchange of information. The most efficient way to learn about the others and to exchange information is through the act of questioning. Hence, it comes with no surprise that researchers have paid a great deal of attention to questions and answers as a core topic in dialogue. During the last years the subject raised a lot of interest also in the field of Natural Language Processing (NLP). In NLP, dialogue still constitutes a big challenge for automated systems, especially for spontaneous and non cooperative communication. The challenge not only concerns the extraction and analysis of data, but also the availability of the data itself. Firstly, questions and answers in spontaneous dialogues are puzzling: phenomena such as incomprehension, discontinuity and misunderstanding are extremely difficult to analyze. Secondly, it is still hard to find available annotated corpora of spontaneous conversations for languages other than English.

Motivation

This paper aims at summarizing and presenting the work carried out during our supervised project of master 1 NLP. Our research takes root in the inquiries currently pursued in the SLAM project (Schizophrénie et Langage, Analyse et Modélisation) developed by the Sémagramme group at LORIA whose main goal is to analyze the conversations between schizophrenic patients and identify the inconsistencies in dialogue. Our work can also be seen as a continuation of last year's supervised project conducted by a team of NLP students and supervised by Maxime Amblard and Maria Boritchev. As linguists, the main objective of our work was to contribute to a better understanding of the nature of questions and answers and to propose a fine-grained annotation schema that could account for their complex structure.

The structure of the report

Section 1 is dedicated to the classification of questions and answers. We will introduce the need of analyzing them both in isolation and as a pair.
Section 2 will constitute an attempt at formalizing our considerations into an annotation schema. Section 3 will be dedicated to the segmentation of the text for both simple exchanges and complex exchanges containing multiple questions and answers.
In the last two sections we will describe the results obtained by the application of our annotation schema in English, Italian, Chinese and French.

Chapter 1

Questions and Answers Classification

In this paragraph we want to focus on the classification of questions and answers. Classification is, indeed, the first step needed before the conception of an annotation schema. In last year's project, questions were classified according to classes where a tag was associated to a function conveyed by different possible forms. See table 1.

Tag	**Name**	**Form**	**Function**
YN	Yes/no question	Inversion, do-support	Asks for the truth value of a proposition
WH	Wh-question	Wh-constituent	Asks for a feature (see feature table)
DQ	Disjunctive question	Contains a disjunction ("or")	Asks for a feature (see feature table)
PQ	Phatic question	Any form	Phatic function
CS	Completion suggestion	Any form	Complete the previous speaker's utterance

Table 1: Previous classification

In our opinion it is important to annotate forms and functions separately. Indeed, one question/answer form can have different functions and one function can be conveyed by more question/answer forms. In the annotation schema proposed by last year's students, for example, a phatic question received the tag PQ. In our proposition a phatic question received both the tag FORM and FUNCTION and the tag PHA was associated only to the function layer. The same logic has been applied to answers. The main objective of this distinction was to understand what questions and answers could be paired without giving raise to incomprehension or absurd exchanges through the analysis of the interaction between form and function.

1.1 The forms of Questions and Answers

In this section we will see how we can exploit the tags FORM and FUNCTION to untangle the complex nature of questions and answers. We will first detail the forms and functions that deserve special attention for our analysis.

1.1.1 The form of questions

We used two criteria to classify question forms:

Criterion 1:
A question is classified according to the span of answers it accepts.

According to Criterion 1 we classified:

(i) yes/no questions.

(ii) Auxiliary questions against their traits [+/-Deontic],[+/-Epistemic].

(iii)Disjunctive questions against their trait [+/-Inclusive].

Criterion 2:
A question is classified by the specific lexical item it contains.

According to Criterion 2 we classified:

(i) WH questions (presence of a WH pronoun)

(ii) Disjunctive (presence of "or")

(iii) Auxiliary questions (presence of an auxiliary verb).

Definition 1 (Deontic auxiliary) *The deontic function of auxiliary verbs expresses the possibility and necessity in terms of freedom to act (including permission and duty).*

Definition 2 (Epistemic auxiliary) *The epistemic function of auxiliary verbs expresses the possibility of propositions being true or false.*

Definition 3 (Disjunctive inclusive) *A disjunctive "or" is inclusive if the truth or the falsity of one disjunctive term implicates the truth/falsity of both.*

Definition 4 (Disjunctive exclusive) *A disjunctive "or" is exclusive when the truth of one option implies the falsity of the second one and vice versa.*

Example 1.
A: Can you tell me more about it?(deontic reading)
(The speaker starts telling something)

Example 2.
A: can you make it through the year?[1]. (epistemic reading)

Example 3.
A: will you pay with a credit or a debit card? (inclusive "or")
B: Yes / debit

Example 4.
A: Do you want tea or coffee? (exclusive "or")
B: *Yes / Tea

This distinction is important for us because the traits described above determine the span of acceptable answers for a given question. For instance, both questions in Examples 3 and 4 above contain a disjunctive "or" . However, (A) in 4 accepts only a WH response, while (A) in 3 accepts both a WH response and a YN response. This is due to their distinct traits [+ Exclusive] vs. [+Inclusive].

[1]SCoSE/Addie and Brianne, 570

According to the two criteria mentioned above, we built the following classification:

Name	Tag	Description	Examples
Yes no	YN	Contains do support, inversion	Are you fine?
Wh	WH	Contains WH Require: FEATURE	What time is it?
Disjunctive inclusive	DQ_I	Contains "or", inclusive	Are you a citizen of European Union or Switzerland? If yes, click here
Disjunctive exclusive	DQ_E	Contains "or", exclusive	Do you want tea or coffee?
Auxiliary deontic	AUX_D	Contains an auxiliary deontic	Can you open the window for me?
Auxiliary epistemic	AUX_E	Contains an auxiliary epistemic	Can you survive all this?

Table 2: Questions

Concerning the FEATURES required by WH questions we decided to keep unchanged last year annotation schema.[2]

Tag	Name
TMP	Temporality
LOC	Location
AG	Agent
TH	Theme
OW	Owner
RE	Reason
CH	Characteristic

Table 3: Features

1.1.2 The form of answers

Answers have their own form too. It is, of course, difficult to list all forms that answers can take for each tag. However, we can intuitively list some usual forms used to convey: (i) a positive/negative answers (ii) feature answers (iii) uncertain answers (iv) unknown answers.
Some examples with their corresponding tags are shown in the table below.

Name	Tag	Examples
Yes no	YN	Yes, yeah, yep, jeez, sure, of course, absolutely../ No, nope, no way, not at all, nah..
Wh	WH	I go home tomorrow, When I.., Because I..
Uncertain	UNC	I'm not sure, maybe, still don't know, could be..
Unknown	UNK	I don't know, dunno, have no clue..

Table 4: Answers

[2]Maria Andrea Cruz Blandon, Gosse Minnema, Aria Nourbakhsh, What's The Answer: Dialogue Annotation, 2017-2018

Answers stay puzzling. Some forms don't correspond to any form that can be categorized according to our classification. Such forms were annotated as NONE in our schema.
One example from our corpus[3]:

Example 5.
Speaker B: oh just go
Speaker A: go?
Speaker B: go.

If we take the answer of B in isolation we can notice that its form doesn't correspond to any of the forms listed in Table 3. The answer doesn't contain an explicit "yes", which would make it fall under the YN category, nor a WH pronoun or a lexical item expressing uncertainty/unknown. However, its function is to convey a confirmation about the question asked. For these kind of answers, we kept the form as NONE and attributed the conveyed function. Their analysis will be carried out in a second moment at the level of the pair.

1.2 The functions of Questions and Answers

When we talk about functions, what we have in mind is close to the concept of Austin's illocutionary force (Austin, 1962). Austin draws a distinction between what it is said, how and the intention behind it. For example, a question such as "Is there any salt?" is used to ask whether some salt is present. However, the intention of the speaker using this particular form is to request some salt. This idea is included in the Speech Act theory. We could not include the Speech Act Theory itself in our study because it is not detailed enough for the questions. In other words, asking a question would be a Speech Act itself, while what we need is to specify more functions attributed to the act of asking. However, starting from this idea we built a list of functions. See the tables below:

Name	**Tag**	**Description**
Completion suggestion	CS	The speaker completes the turn of another speaker
Phatic	PHA	Phatic function
Ask confirmation	ASK_CONF	The speaker asks the truth value of a proposition or the hearer's engagement to an action
Ask feature	ASK_FEAT	The speaker asks for a feature
Ask to perform	ASK_PERF	The speaker asks to perform an act
Reported speech	RS	The speaker report someone's else question

Table 5: Functions of questions

[3]SCoSE/Addie and Brianne, 216

Name	Tag	Description
Refuse	REFUSE	The speaker refuses to engage in an action
Accept	ACCEPT	The speaker accepts to engage in an action
Phatic	PHA	Phatic function
Give confirmation	GIVE_CONF	The speaker conveys the truth values of a proposition
Give uncertainty	GIVE_UNC	The speaker conveys uncertainty about something
Give unknown	GIVE_UNK	The speaker conveys an unknown response
Report speech	RS	The speaker reports someone else speech
Give feature	GIVE_FEAT	The speaker gives a feature
Perform	PERF	The speaker performs the act requested
NONE	NONE	No answer is given

Table 6: Functions of answers

In Tables 5 and 6 each intention of the speaker by uttering a question/answer is linked to a specific function.

1.3 Questions and Answers as a pair

How do the questions and answers interact with each other? After an analysis of them in isolation we tried to understand how their association works and why it can result in comprehension or incomprehension. To do this, we played with the notions of symmetry and mismatch.

Definition 5 (Symmetry) *A question is symmetric to its answer when the semantic or syntactic requirements imposed by the question are fulfilled by the answer.*

Definition 6 (Asymmetry) *A question is asymmetric to its answer when the semantic or syntactic requirements imposed by the question are not fulfilled by the answer.*

When we talk about syntactic requirements we refer to the form. In English, for example, a yes/no question is characterized by subject-auxiliary inversion and do-support (form) and its answer is expected to have the same yes/no form of Table 4. When we talk about semantic requirements we refer to the function. A yes/no question has prototipically the function of asking for the truth-values of a proposition (function) (see the tag ASK_CONF in Table 5) and its answer is expected to convey the veridicity or falsity of the proposition asked (see the tag GIVE_CONF in Table 6). Let's consider some examples of asymmetry (Tables 8 and 9) from our corpus SCoSE.

Example 1.

Symmetry
of form and function
A: Why are you crying?
B: Because I hurt myself.

Q_form	**Q_function**
WH	ASK_FEAT
A_form	**A_function**
WH	GIVE_FEAT

Table 7: Symmetry

Example 2.

Asymmetry
of form
B: it includes heat and uhm I think..
B: Water?
A: Water.

Q_form	**Q_function**
YN	ASK_CONF
A_form	**A_function**
WH	GIVE_CONF

Table 8: Asymmetry

Example 3.

Asymmetry
of form and function
A: so- wh- where can you move to?
B: Well..you know..I don't even know where I'm living next year.

Q_form	**Q_function**
WH	ASK_FEAT
A_form	**A_function**
UNC	GIVE_UNC

Table 9: Asymmetry

We define the notion of mismatch as follows:

Definition 7 (Mismatch of form) *In case of asymmetry of form between a question and its answer (see definition 6) a mismatch of form occurs if and only if the form of the given answer doesn't fall under one of the forms prototypically accepted by the question.*

Definition 8 (Mismatch of function) *In case of asymmetry of function between a question and its answer (see definition 6), a mismatch of function occurs if and only if the given answer doesn't convey an intention that falls under one of the intention prototypically required by the question.*

To identify what prototypical answer forms and answer functions a question requires, we made a table of compatibility. The idea of the compatibility is to map the forms and functions that in both cases of symmetry and asymmetry can combine with each other. In Table 10 each question tag in the first column is associated with a group in the second column (`Fo`) containing the answer forms that do not trigger a mismatch when combined with that specific question form.

Question Forms	**Expected answer forms**
YN	$\texttt{Fo}_1$ ⟨ YN,UNC,UNK ⟩
WH	$\texttt{Fo}_2$ ⟨ WH,UNC,UNK ⟩
DQ_I	$\texttt{Fo}_3$ ⟨ YN,UNC,UNK ⟩
DQ_E	$\texttt{Fo}_4$ ⟨ WH,UNC,UNK ⟩
AUX_D	$\texttt{Fo}_5$ ⟨ YN, NONE, PERF ⟩
AUX_E	$\texttt{Fo}_6$ ⟨ YN,UNC,UNK ⟩

Table 10: Compatibility form

In Table 11 each question tag in the first column is associated with a group (Fu) in the second column containing the answer functions that do not trigger a mismatch when combined with that specific question function.

Question Function	**Expected answer function**
CS	$\texttt{Fu}_1$ ⟨ REFUSE, ACCEPT, PHA, GIVE_CONF, REPORT ⟩
PHA	$\texttt{Fu}_2$ ⟨ REFUSE, PHA, GIVE_CONF, REPORT, NONE ⟩
ASK_CONF	$\texttt{Fu}_3$ ⟨ REFUSE, ACCEPT, GIVE_UNC, GIVE_UNK, GIVE_CONF ⟩
ASK_FEAT	$\texttt{Fu}_4$ ⟨ GIVE_FEAT, GIVE_UNC, GIVE_UNK ⟩
ASK_PERF	$\texttt{Fu}_5$ ⟨ GIVE_PERF, NONE, GIVE_UNK, GIVE_UNC, YN ⟩
RS	$\texttt{Fu}_6$ ⟨ PHA,REPORT,NONE ⟩

Table 11: Compatibility function

We want to clarify the concept of compatibility with one example:
A phatic question such as:
"Really?"
is categorized as follows:

Form: YN (see Table 2)
Function: PHA (see Table 5)

If no answer is given we can say that there's an asymmetry between the question and the answer (lack of answer). However, the lack of answer is part of the possible answer span for a phatic function (see Fu2 in Table 11). Therefore, a mismatch of form is triggered but no mismatch of function.

On the contrary this doesn't apply to a WH question such as:
"What colour is it?", which is categorized as follows:

Form: WH (see Table 2)
Function: ASK_FEAT (see Table 5)
Feature: CH (see Table 3)

In this case, if no answer is given, we are in the presence of a mismatch between the question form and its answer form (lack of answer) as well as a mismatch between the two functions. Indeed, if we look at our Tables 10 and 11, a WH question doesn't accept a NONE response.

This distinction is vital to distinguish the cases where someone asks a question and:
(i) An answer is not given on purpose: the speaker doesn't engage in a response (our WH example above).
(ii) An answer is not given because it's not necessary. (Mostly the case of phatic questions).

Chapter 2

Annotation in practice

In this chapter, we will see how to put our previous considerations into practice. Therefore, we will describe the steps of our annotation that include: (i) annotation of questions and answers alone (ii) addition of a second layer to detect whether the annotated answer is compatible with its question (iii) in case of mismatch a possible resolution of it at the level of the question-answer pair.

2.1 The semantics of questions and answers: tags in isolation

In tables 2, 4, 5, 6 we listed the possible forms and functions for questions and answers. With these tags available, we first proceeded by detecting a question through the presence of a question mark or inversion. Successively, our annotation follows some precise steps (the notation is in Xml format):

Questions:

1. We tag the question as <Question>
   2. We open a tag <question_type> containing four elements:
      a. <question_form>: contains one of the tag available in Table 2
      b. <question_function>: contains one of the tag available in Table 5
      c. <expected_answer_form>: contains one of the groups available in Table 10
      d. <expected_answer_function>: contains one of the groups available in Table 11
   3. We close the tag <question_type>

Answers:

   4. We tag the answer as <Answer>
   5. We open a tag <answer_type> containing two elements:
      a. <given_answer_form>: contains one of the tag available in Table 4
      b. <given_answer_function>: contains one of the tag available in Table 6
   6. We close the tag <answer_type>

The logic behind our annotation is the following: a question, as we said, has a specific form and function. However, it is not possible to state precisely what answer will be given for a question. For example, a WH question such as "What time are you coming?" has a WH form but it cannot be said to expect a WH answer only. One can perfectly reply "I don't know", which would fall into the category UNCERTAIN. This is why we annotated the expected answers with potentially compatible groups (see table 10 and 11) within the annotation of the question. There's no need to apply the same grouping to answers, but only to check if the answer given is compatible with the expectation of the question asked.

The way we managed to find the answer to a question is a complex topic and we will treat it in section 3.

2.2 From Semantics to Pragmatics: mismatch analysis

In the previous sections we stated the need of keeping the analysis of questions and answers in isolation and introduced the notions of symmetry and mismatch. We can now put all these elements together and proceed to analyze our QA pairs. We decided to proceed in an algorithmic way. For convenience, we will restart our annotation after the end of step 6 in the previous section 2.1. After step 6 we start the analysis of the pair.
The first thing we do is to check the compatibility of the <expected_answer_form> and the <given_answer_form>. If the tag of the <given_answer_form> is present in the group corresponding to the <expected_answer_form> (see Table 10) we signal NO in the <mismatch_form> section.
The second thing we do is to check the compatibility of <expected_answer_function> and the <given_answer_function>. If the tag of the <given_answer_function> is present in the group corresponding to the <expected_answer_function> (see Table 11) we signal NO in the <mismatch_function> section. Finally, we go on according to this algorithm:

7. <mismatch_form>
8. <mismatch_function>
   if NO go to 9, tag YES and STOP
   if YES go to 9, tag NO and continue to 10
9. <direct_answer>
10. <indirect_answer>

Intuitively, if we signal NO in <direct_answer>, we should signal YES in <indirect_answer>. However, there's a difference between the two:

Definition 9 (direct answer) *The direct answer is an answer fulfilling all the semantic requirements of the questions.*

Definition 10 (indirect answer) *The indirect answer is an answer that doesn't fulfill all the semantic requirements of the questions.*

Indeed, both <direct_answer> and <indirect_answer> are actually answers. If we signal NO to both, it means that no answer is present and we should end the query (we go to step 13 directly). If we signal YES only to indirect answer we are allowed to continue with step 11 (see below) and detect where the indirectness lies. For convenience we retake the previous steps 9 and 10 and continue as follows:

9. <direct_answer>
10. <indirect_answer>
    if NO go to 13
    if YES go to 11
11. <has_dialogic_function>
    if YES tag <_dialogic_function> <id>
    if NO go to 12
12. <has_implication>
    tag <_has_implication> <id>
13. <not_answer>
    tag YES and STOP

An example of an output from our corpus SCoSE[1]:

```
<turn>
    <id>Speaker A</id>
    <text>does that include everything? </text>
            <type>
            Question
            </type>
            <question_type>
                <question_form>
                YN
                </question_form>
                <expected_answer_form>
                Fo1
                </expected_answer_form>
                <question_function>
                ASK_CONF
                </question_function>
                <expected_answer_function>
                Fu3
                </expected_answer_function>
            </question_type>
</turn>
<turn>
    <id>Speaker B</id>
    <text> it includes heat and uhm I think uhm..</text>
            <type>
            Answer
            </type>
            <answer_type>
                <given_answer_form>
                WH
                </given_answer_form>
                <given_answer_function>
                GIVE_FEAT
                </given_answer_function>
                <feature>
                TH
                </feature>
            </answer_type>
</turn>

<mismatch_form>YES</mismatch_form>
<mismatch_function>YES</mismatch_function>
<direct_answer>NO</direct_answer>
<indirect_answer>YES</indirect_answer>
<has_dialogic_feature>NO</has_dialogic_feature>
<has_implication>YES</has_implication>
<id>Indirect, infer no</id>
```

[1]SCoSE/Addie and Brianne, 534-535

In the example above, we have an asymmetry of form and function between the question (YN)(ASK_CONF) and the answer (WH)(GIVE_FEAT). However, the answer seems to be pertinent and the dialogue goes on. The mismatch analysis allows us to justify it through the presence of an implicature[2]: the Speaker A is able to infer a negative answer from Speaker B who adds more information about the subject.

2.2.1 Dialogic functions and Implicatures

In our analysis we treated only the simple cases of indirectness and grouped them into two big categories: Dialogic functions and Conversational implicatures.

a) Dialogic functions:

Other functions/intentions not belonging to the QA pair functions. They mostly include speakers making comments about what is said before. The statement is usually related to a question previously asked but it cannot be considered as its direct answer. Examples from our corpus:[3]

Example 1.
A: They just decided about 6 o'clock.
B: Oh really?
A: It's so late to be home (comment)

Example 2.
B: which is pretty small but still bigger than a dorm room you know?
A: uh-huh and you have your own bathroom (comment)

b) Conversational implicatures:

We remind quickly of the definition of an implicature and one of its particular type: the conversational implicatures (Grice, 1975).

Definition 11 (Implicature) *We are in the presence of a case of implicature when what a speaker suggests with his/her utterance differs from the literal meaning of the utterance used to express it.*

Definition 12 (Conversational implicature) *An implicature is said to be conversational when it is not part of the conventional meaning of the sentence uttered, but depends on the conversational context.*

In our analysis we only considered the cases where the interlocutor is able to infer a positive, negative or uncertain answer from an indirect statement. The implicatures are almost all conversational. Examples from our corpus:

Example 3.
A: can you get out of your contracts anyway?
B: I talked to a couple of different people who have done it.
(Infer yes)

Example 4.
B: when will you guys get off?
A: my last exam is like I don't know. I think Wednesday. Tuesday or Wednesday.
(Infer uncertainty)

The detection of indirectness is a wide topic to treat and it's worth to be a subject of future research.

[2]Conversational and conventional implicatures, Jacques Moeschler. Retrieved from `https://pdfs.semanticscholar.org/b548/c00d985d61c03b20b1c424a1a46becdc5587.pdf`

[3]SCoSE/Addie and Brianne, 16 nad 476

Chapter 3

The pair Q-A: the NP algorithm

An issue pointed out in "Coffee or Tea?" (Amblard, Boritchev, 2018) is the way dialogue can be divided into smaller sub-parts in order to make the analysis more approachable while taking into account the communicative context. These smaller parts of dialogue have been named Negotiation Phases (NP). NPs intuitively correspond to self-contained sub-dialogues following an ongoing topic. We could not rely on the notion of topic so our approach slightly deviates from the definition of NP. Our main objective was to propose a systematic way to group multiple questions and answers and segment the text accordingly. We will describe in the next paragraphs how we decided to treat simple and complex exchanges between speakers during our annotation of the corpora.

3.1 Simple NP

A simple NP is an exchange containing only one question and one answer. Now, it happens many times that an answer doesn't follow its question directly. Our idea is that in such cases we continue our analysis trying to find an answer as far as we can according to two cases:

Case 1:
A Speaker A asks a question and the Speaker B makes statements before answering the question.

Case 2:
The Speaker A asks a question and the Speaker B makes completely unrelated statements without actually ever responding.

To analyze such exchanges we decided to proceed as follows:

1. We analyze the question of one speaker (see section 2.1).
2. When the speaker changes, we analyze his/her first utterance and look for an answer.
   if it's an answer:
      we close the NP
   if it's a statement (not answer):
      we move to the next segment
3. We keep doing this until the speaker changes again.
   if the speaker has changed (order Speaker A, Speaker B, Speaker A) and no answer is found
      we tag answer_not_found unless a new question is introduced (see next section)

3.2 Complex NPs

The NP method we have presented above doesn't take into consideration the case where more than one questions and answers are embedded. Let's consider the following example from our English corpus :

Example 1.
A: no I don't suppose I got invited
B: did you? good ha ha ha good
A: do you want me to go?
B: yes

Such cases are complex to treat, so we tried to think about the different combinations between speakers in an embedded exchange and build a graph that represents the paths of their interaction.

Legend
1 = Speaker 1
2 = Speaker 2
Qrel = Question related
Qunrel = Question unrelated
R = Response (It can be S or A)
S = Statement
A = Answer
EB = End of NP before the segment
EA = End of NP after the segment
Numbers from 1 to 20: paths for the reader(references to smooth the explanation of the examples)

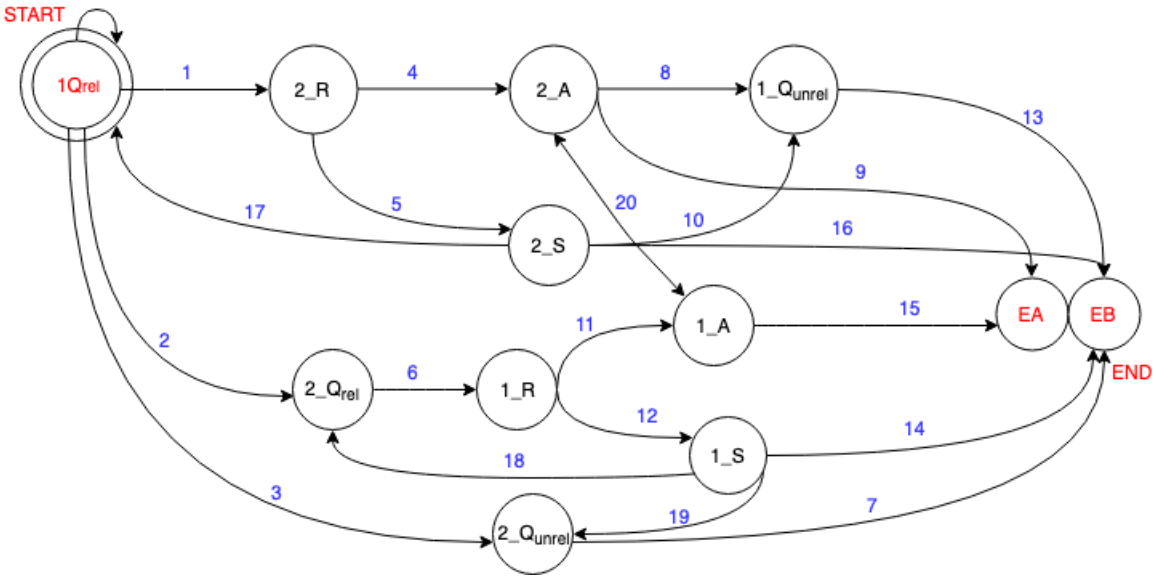

Figure 3.1: NP Algorithm

The reading of the graph is the following:
We suppose that the starting point is Speaker 1 asking a question (1_Qrel in the starting node). The end points are the two nodes in red colour representing (i) the end of the NP before the segment analyzed (EB) and (ii) the end of the NP after the segment analyzed (EA). Starting from the node Q_rel we have two main possibilities:
(i) The path of the answer.
(ii) The path of the question.

(i) The path of the answer

a. We find a response from Speaker 2 (2_R). We take the path 1.
b. The response can be an answer (path 4 to 2_A) or an unrelated statement (path 5 to 2_S) . In the first case: we can take the path 9 and close the NP after the answer or find an unrelated question (path 10) and close the NP before it (path 13). In the second case we take the path 9.

If we take the path 9 we go on only if we encounter a new question, otherwise we take path 16 and close the NP.

If we encounter a question:

b.1 the question is unrelated. We take the path 10,arrive at node (Q_unrel) and close the NP through path 13 (EB).

b.2 the question is related, we take the path 17 and loop.

(ii) The path of the question:

a. We find a question from Speaker 2

a.1 The question is related.We take the path 2.

If the speaker A gives a response we have here the same choices that we had for the node 2_R but for Speaker 1 (1_R).

a.2. The question is unrelated. We take the path 3 to 2_Qunrel. We close the NP through path 7.

We can test the algorithm on the example introduced at the beginning of the section and other examples from our French and Italian corpora. We first provide the translations below:

Example 2:
A: tu me feras penser à appeler mon frère
tout à l' heure?
B: ton frère?
A: ouais mais je veux pas que tu
écoutes quand je l' appelle.

Example 2: translation
A: will you remind me of calling
my brother earlier?
B: your brother?
A: yes, but I don't want you to
listen when I call him.

Example 3:
A: secondo voi macchia?
B: sui vestiti?
A: eh
B: non dovrebbe dovrebbe evaporare
tutto praticamente, è zucchero

Example 3: translation
A: Do you think it stains?
B: On the clothes?
A: Yeah
B: It shouldn't, it's all sugar,
it should evaporate

NP testing:

Example 1:
A: did you? → (1_Qrel)
B: do you want me to go? → (2_Qunrel)
Path followed:
1_Qrel → 3 → 7 → EB

Indeed, the conversation goes on like this:
B: do you want me to go? → (1_Qrel)
A: yes → (2_R) → (2_A)
Path followed:
1_Qrel → 1 → 4 → 9 → EA

Final division of the text:

1st segmentation
A: no I don't suppose I got invited
B: did you? good ha ha ha good
[answer_not_given]

2nd segmentation
A: do you want me to go?
B: yes *[answer_given]*

Example 2:
A: tu me feras penser à appeler mon frère tout à l' heure? → (1_Qrel)
B: ton frère? → (2_Qrel).
A: ouais mais je veux pas que tu écoutes quand je l' appelle. → (1_R)(1_A).
Path followed:
(1_Qrel) → 2 → 6 → 11 → 15 → EA

Example 3:
A: Do you think it stains? → (1_Qrel)
B: On the clothes? → (2_Qrel)
A: Yeah → (1_R)(1_A)
B: It shouldn't, it's all sugar, it should evaporate → (2_A)
Path followed:
1_Qrel → 2 → 6 → 11 → 20 → 9 →EA

After this analysis the two questions and the answer in Examples 2 and 3 will be considered within the same NP and we can proceed with their analysis.

While annotating our corpus we realized that the main issue we had was to decide whether we should consider multiple questions in the same NP or not. This algorithm constituted a very efficient way to solve the problem and improve the agreement between annotators.

Chapter 4

Annotations on English

After building our annotation schema as described in the previous sections, we put it into practice. This section will present the results obtained by comparing three annotations on the same data. The last section will be dedicated to the comparison between our results and last year's results on the Saarbrucken Corpus of Spoken English (SCoSE).

4.1 The corpus

We annotated 90 pairs of Q-As from a dialogue in the Saarbrucken Corpus of Spoken English (SCoSE). The dialogue is a conversation between two friends, Addie and Brianne, meeting again after their departure to different universities. Both women are close to twenty-one years old, and both come from northern Illinois. The annotation was carried out directly on the transcript without relying on the related MP3 audio. The results of three annotations were compared. Afterwards, a gold corpus was created to enable a better analysis of the data. The dialogue, its audio file as well as the outputs of our annotations can be found in the folder annexes at this link `https://drive.google.com/drive/folders/1jn9eD4mbadNfRdZSAy2FySmIS3dGqzpM`

4.2 The output: agreement and disagreement

The output of our annotation is in the form of a Json file. Each question and answer were annotated following the directives given in our annotation guideline (see Annexes). Successively, we put our data in an Excel table where each column represents the tag assigned and each row points to the annotator's name. The table can be found in the annexes. The first comparison of the data included the agreement between pairs of annotators A, B and C.
It was calculated for each tag and for the total of the tags with K-score. The results of the agreement are summarized in the table below and the overall score for each tag is plotted in the bar graph 4.1 for 7 types of tags. For other results see the annexes.

	Question Form	Expected Answer Form	Question Function	Expected Answer Function	Feature	Answer Form	Answer Function
Annotator A_B	0.94	0.91	0.91	0.91	0.95	0.90	0.88
Annotator A_C	0.94	0.94	0.88	0.88	1	0.85	0.85
Annotator B_C	0.89	0.89	0.88	0.88	0.9	0.84	0.79

Table 12: Agreement between A, B and C

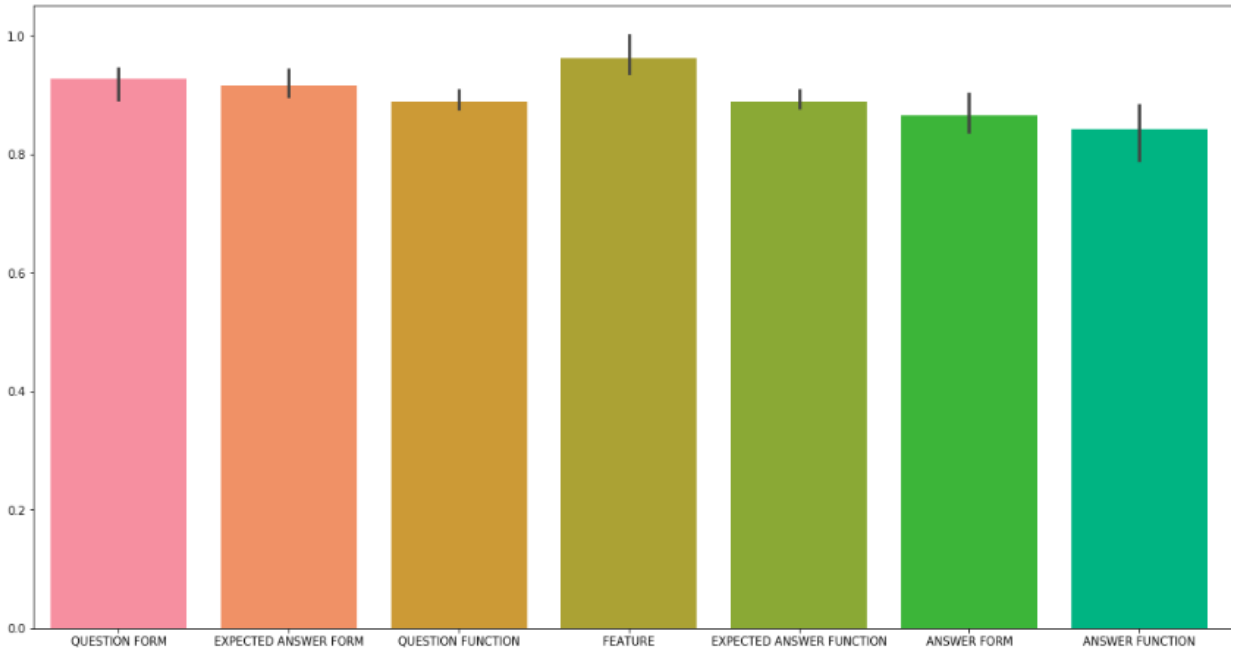

Figure 4.1: Overall agreement on each tag

The agreement between the three annotators taken as pairs is high after the whole annotation. The first results we achieved were lower (around 0.65). The more we annotated and the more we got familiar with the complexity of the annotation schema. The current results show agreements between 0.79 and 0.96.

4.3 Comparing two groups of annotations

After annotating our corpus we found it interesting to make some comparisons with last year's results. In our annotation we had a total of 5 tags for questions: Question Form, Question Function, Expected Answer Form, Expected Answer Function, Feature.
In last year annotation they had a total of 3 tags: Question Type, Is quoted, Complexity, Feature.
Due to the different amount of labels, we grouped our annotations into wider classes to compare the data. The annotation of this year has been named GOLD1 and the annotation of last year has been named GOLD2. The comparable classes were grouped as follows :

1. The two classes “Feature” were directly compared for GOLD1 and GOLD2.
2. The classes “Question Form” and “Question Function” for GOLD1 were put together according to figure 4.2:

	YN	CS	WH	PHA	DQ	FEATURE
GOLD_2	YN	CS	WH	PHA	DQ	FEATURE
GOLD_1	AUX_E AUX_D RS YN	CS	WH	PHA	DQ_E DQ_I	FEATURE

Figure 4.2: Grouping for comparison

The tags were compared for each corresponding class of questions. However, some questions were not annotated in last year's output and we had to omit these data. We calculated the inter-annotator agreement between the two GOLDS. The results are shown below:

	question_type	**features**
GOLD12	0,75	0,89

Table 13: Agreement between GOLDS

The overall agreement is more than 0.7, which means that a more fine-grained annotation didn't affect the choice of the main tags on questions and answers. We want now to compare the inter-annotator agreement between the three annotators of last year and us.

	A B	**B C**	**A C**
question_type	0.62	0.59	0.62
answer_type	0.6	0.42	0.35
features	0,74	0.64	0.61

Table 14: Agreement 2017/2018

	A B	**B C**	**A C**
question_type	0.92	0.92	0.88
answer_type	0.88	0.85	0.82
features	0.94	1	0.87

Table 15: Agreement 2018/2019

The difference in agreement makes us think that a simpler and plane annotation leaves too much space to the interpretation from the annotator's side. Indeed, a more structured annotation schema gives a score that is no lower than 0.8, while a less structured one gives a results of around 0.6. On the other hand the more labels we have, the easier is to disagree on one single label. Indeed, even though the agreement between annotators this year is higher compared to the one of the previous year, the score of the agreement on all labels results to be quite low:

	AB	**BC**	**BC**
total_score	0.5	0.6	0.6

Table: Agreement on all labels

Chapter 5

Annotations on other languages

5.1 Italian

Before introducing the annotations on Italian we need a short historical parenthesis. At the time where Roman Empire started falling down, and after the decentralization of political power, Latin was only spoken in the Roman empire. Its variations were absorbed by Italian regions and gave raise to our dialects. The dialects were spoken for a long time in Italy and language unification came very late. The first attempts to normalize Italian language is in 1525 with Pietro Bembo who proposed to take as language models Dante, Petrarca and Boccaccio (Tuscan variation of Italian). Even though these models were accepted, Italy officially unified only in 1861 and the problem of language disharmony was never really solved due to the lack of central political power. The consequence of this is that Italian language is still extremely fragmented.
The historical introduction aims at explaining a curious fact: all corpora currently available in Italian are grouped into regions. This is, on one hand, a rational wise choice. Indeed, a dialogue between two interlocutors from different regions would probably affect its spontaneity. On the other hand, grouping regions corresponds to a big loss of data for our purpose. It would've been extremely interesting to have mixed corpora where incomprehension can be linguistically investigated and explained through language variation.

5.1.1 The corpora

The first annotated corpus comes from the Corpus del parlato italiano (API). All the material in the corpus totally reflects our previous considerations on the Italian language. The dialogues are divided by city (mostly Naples and Pisa) and annotated to detect particular dialectal phenomena. For example, all corpora in Tuscan are annotated with the phonetic trait [+GORGIA], which affects the voiceless stops /k/ /t/ and /p/ pronounced as fricative consonants in post-vocalic position. However, scraping a bit the data in the corpus we were able to find a spontaneous dialogue containing a sufficient number of questions and answers.
The dialogue contains 78 pairs of QA and it's a spontaneous conversation between two interlocutors from Naples of the duration of 14 minutes and 17 seconds. The annotation was carried out directly on the transcript without relying on the related MP3 audio. The .pdf file and the output of the annotation can be found in the the folder Annexes mentioned in section 4.1.
Because of the lack of data we decided together with our supervisors to record a second corpus that could serve to our analysis. The corpus was recorded in Italy in April and it involves four participants playing a game called Catan. The objective of the game is to build colonies by piling up resources. Each gamer

can exchange their resources with the adversary and to do this he/she must negotiate through questions and answers. The recording lasts 1 h and 10 minutes in total, 14 of which were transcribed according to the transcription guideline of the API corpus. They were successively annotated.
In the next section I will present the issues found in both corpora while trying to adapt our annotation schema designed for English to Italian.

5.1.2 Language-specific issues

1. Prosody: questions in Italian do not involve any inversion or morphological insertions. The only way to distinguish a question from a statement in a transcription is the presence or absence of the interrogation mark. The trait [+inversion/do-support] is, therefore, not pertinent for Italian.

2. Ellipses: In Italian it very common to omit parts of the questions, which makes it complicated to classify them according to the form. Two example from the corpora:

Example 1.
(context: Speaker 1 wants to know something about the rear-view mirror.)
A: lo specchietto retrovisore in alto
the rear-view mirror on the top
quello all'interno della macchina ?
that one inside the car?
B: non c'è!
there isn't.

Example 2.
A: ma io come è possibile che ho tutti sti
but how is it possible that I have all these
pascoli e non ho mai preso una pecora?
grazing lands and I didn't get one single sheep?
B: no ma perché le rocce scusami ?
no but the rocks sorry?
C: eh ma le rocce non ce l'ha nessuno
yeah but nobody has rocks

In the example 1 above it's almost impossible to know whether the reconstruction of the ellipsis takes one of the following forms:
(a) [Is there] the rear-view mirror, that one at the top inside the car? (YN, no mismatch)
(b) [Do you see] the rear-view mirror, that one at the top inside the car? (YN, denial of expectation)
(c) [What about] the rear-view mirror, that one at the top inside the car? (WH, mismatch)
Due to the difficulties in the reconstruction of the ellipses, even questions involving WH pronouns were classified according to Criterion1 (see section 1.1.1) and not according to Criterion2. The same thing can be said for the question asked from B in example 2:
B: no ma perché le rocce scusami ?
Not but the rocks sorry?
The form above is used to approximately convey the meaning "You are complaining but do we want to talk about the rocks and the fact that there's never the possibility to take them?". This form is very hard to categorize and analyze.

3. Interjections and prosody: In Italian interjections are widely used to respond to questions. The issue is that the tone usually changes the meaning of the same interjection. Let's consider the example below:

Example 3.
Speaker 1: quello che sta in alto diciamo , no ?
Speaker 2: eh
Speaker 1: eh e invece io sì

Now, the exchange is minimal in terms of words but very rich in terms of meaning. The paraphrase in English would be:
Speaker 1: that one at the top, no?

Speaker 2: yeah that one and no I don't have it
Speaker 1: I have it instead
We have again one ellipsis in question 1. The form "no?" means: "That one at the top, [you do] not [have it]?" The answer is only "eh". In Italian "eh" has 9 meanings according to the intonation. It can, indeed, express doubts, confirmation, surprise, disappointment, exaggeration and so on. The issue is that such forms are commonly use to respond to questions.

5.2 French

We chose to carry out our annotations on three different corpora selected from the same environment: **O**pen **R**esources and **TO**ols for **LANG**uage (Ortolang). Ortolang is a platform that groups various types of French data, notably corpora, lexicon, dictionaries to provide efficient tools for the study of French in multiple aspects. The corpora consist of three spontaneous dialogues between French adults with their transcriptions, recordings and syntactic trees.

1. TCOF (Traitement de Corpus Oraux en Français)
TCOF is a corpus of spoken French collected by the ATILF team divided in two parts: recordings of interactions between children and adults, and recordings of interaction between adults in multiple communicative situations. The second part is composed of twenty-three hours of speech.
The first recording is a dialogue of ten minutes between a couple of students talking about different topics (Christmas gifts, friendships) resulting in 41 pairs of QAs.
The second recording is a dialogue of ten minutes, inside of which we analyzed 25 pairs of QAs. The conversation is held between two young adults discussing about nightclubs, their friends and role games.

2. Valibel
The Valibel center - Discourse and Variation gives access to a data base of twenty-two corpora of spoken dialogues, registered between 1987 and 1995. 43 hours of speech are available on the Orfeo project.
This dialogue contains 13 pairs of QAs and present a conversation of nine minutes between two former roommates playing a game.

5.2.1 Language-specific issues

1. Punctuation:
In the transcription of the corpus, utterances corresponding to questions lack a question mark. This issue can be resolved thank to the presence of wh-words. When no syntax mark of question was found, the only way to identify questions was through the intonation in the audio file.

2. Overlapping:
In the french corpus, overlaps between speech turns (i.e., both speakers are talking at the same time) are much more frequent than in the English corpus. This causes a problem when both speakers ask a question at the same time.
We solved this issue by defining new rules that decide to close or continue the negotiation phase whether the question is related or not and whether the speaker change or not.

Example 1.
L1: c' est pour le cadeau de Noël?
It's for the christmas gift? L2: tu peux aller chercher ma pommade dans le...?
Can you get my ointment from the...?

L1: je te rapporte ton sac si tu veux.
I bring you your bag if you want.

3. Transcription:
Many 'utterances' in the transcription were associated to only one speaker. However, this didn't match with the audio files where multiple speakers would utter the same segment of speech. For example, the utterance below in Example 2 has been transcribed as uttered by a single speaker. However the audio file reveals a different segmentation (Example 3).

Example 2.
L2: non ben non c' est le gros ah
c' est dans la trousse de toilette verte
dedans il faut que tu montes sur une chaise

Example 3.
L2: non ben non c'est le gros
No but no, it's the big one.
L1: ah c'est dans la trousse de toilette
Ah it's in the green toiletry bag.
L2: verte dedans il faut que
inside you should
tu montes sur une chaise.
get on a chair.

5.3 Chinese

One point to specify is that the Chinese language here refers exclusively to the standard language of Mandarin. Other Chinese dialects, such as Cantonese, Hakka or Min, are not in our discussion.
Two corpora used in this annotation include **PolyU Corpus of Spoken Chinese**, conducted by Hong Kong Polytechnic University and released in 2015, and **NCCU Corpus of Spoken Taiwan Mandarin**, data of daily face-to-face conversations collected since 2006 by National Chengchi University in Taiwan.
Four free conversations from **PolyU Corpus of Spoken Chinese**, accompanied by transcription, in average of 5 minutes of conversational exchanges, resulted in 22 pairs of QA.
On the other hand, **NCCU Corpus of Spoken Taiwan Mandarin** consists of 43 free-subject dialogues, varying from 15 to 30 minutes in terms of dialogue duration, with the precise transcription convention (speech overlap, pause of different lengths, etc).
However, lacking of question marks in the transcription of **NCCU Corpus of Spoken Taiwan Mandarin** poses a big problem of laborious annotation, which requires listening to the recordings line-by-line and annotating simultaneously to obtain the accurate NPs and QA pairs. 7 recordings are annotated in this project, resulted in 287 pairs of QA.
In total of 309 QA pairs are annotated in two Chinese corpora.

5.3.1 Language-specific issues

1. English tagset: It is noticeable that the annotator must have basic knowledge of English and Chinese linguistics, otherwise it would be onerous to apply the tagset on the Chinese corpora. Generally, the tagset can be implemented effortlessly, despite the linguistic distance between English and Mandarin , yet the analysis of the form and the function of both questions and answers demand more attentions to bring to.

2. Languages blending: We can observe the mix of Mandarin, Southern Min, Hakka, Japanese and English in various conversations from the NCCU Corpus of Spoken Taiwan Mandarin. Southern Min and Hakka are two major Chinese dialects spoken in Taiwan, due to the history of immigrants from Southeast China since 17^{th} century. During the implementation of Southern Expansion Doctrine of Japan since late 19^{th} century, official language in Taiwan was set to Japanese under the colonization of Japan. Elder people (whose age $\geq$60) use frequently Japanese vocabulary in the dialogues. A frequent phenomenon of intertwining Chinese Pidgin English or standard English in dialogues among youths can also be seen in

this corpus. Meanwhile, the PolyU Corpus of Spoken Chinese doesn't have this issue, the speakers were requested to discuss only in standard Mandarin beforehand.

3. Absence of interrogation mark: Due to the lacking of question marks in transcriptions, the annotation of NCCU Corpus of Spoken Taiwan Mandarin necessitates listening attentively to the recordings in order to detect the questions, although interrogation marks belong to the standard Mandarin writing system. Moreover, interrogations in Chinese don't always appear in specific grammatical structures (inversion or auxiliary verb support), nor always contain a particle at the end of a certain type of questions, especially in daily conversational exchanges, they could be omitted. Besides, the prosody isn't always the benchmark to identify questions in tonal language, such that normally a phatic question can be articulated in descending tone.

4. Embedding of QA: In both Chinese corpora, complexity of embedding of questions and answers during a negotiation phase could be arduous for the annotation task. Different from the overlapping issue found in French corpora, a WH question followed by a phatic question, or a YN question came after a disjunctive question, the answer to the questions may not even be given or even the question following itself could be the answer to the previous question, relations of associated questions rely on the contexts. The solution of NP annotation of complicated embedding of QA is proposed at chapter 3.2.

5. Implicatures: Real-life daily conversational exchanges in Chinese involves some implicatures. One example here :

Example 1.
Speaker 1: 中午要吃什麼？
What do you want for lunch?
Speaker 2: 這邊附近好像新開了一間麵店。
It seems that there is a newly opened noodle restaurant right around this corner.

From the example above, the form of the given answer is proposing the location feature to the speaker 1, but the function is giving feature of the initial WH question raised by speaker 1 exactly asking for. In Chinese, the contextualization of the response, the relationship between speakers, the politeness norm of responding indirectly and the intentions of the speaker play a vital role in the answers containing implicatures. This entails a need of delicate attention-demanding annotation task.

5.4 Comparison between languages

In this section we will compare English, Italian, French and Chinese in different ways. Firstly we will identify the pertinence of the elements used to detect the questions in the corpora. Secondly we will compare the statistics of the annotations in all languages and show some visualizations.

5.4.1 Pertinence of the interrogation tags

Not all languages express linguistically the act of questioning in the same way. We relied on the traits inversion and do-support to detect questions in the English corpus. However, when we started working on different languages we realized that the same traits were not pertinent or sufficient.

Below is the table showing the features for detecting questions and their pertinence in English, Italian, French and Chinese:

	English	Italian	French	Chinese
Interrogative particles	x		x	x
Prosody	x	x	x	x
Inversion	x		x	

Table 16: Presence of interrogation tags

5.4.2 Statistics on four languages

This section is dedicated to some statistics from our annotations in four languages. It will include visualizations for:

- Number of instances for each label (Table 17)
- Frequency of the tags on questions (Bar graphs 5.1,5.2)
- Frequency of the tags on answers (Bar graphs 5.3,5.4)
- Number of answer type for direct and indirect answers (Bar graph 5.5)

	number questions	number direct answers	number mismatches	number indirect	number not answered
English	93	82	55	20	11
Italian	90	75	19	9	15
French	90	72	32	16	18
Chinese	90	87	32	9	3

Table 17: Number of instances for each label

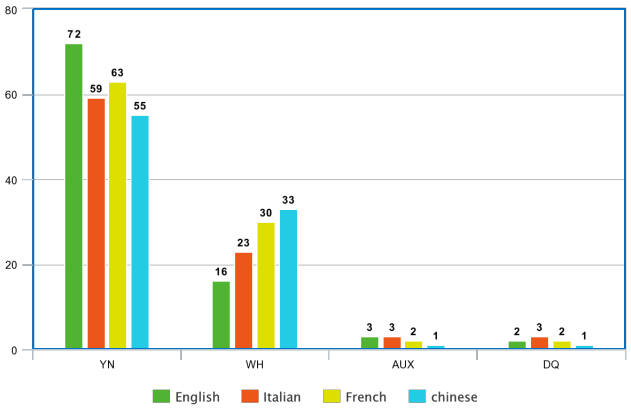

Figure 5.1: Question form frequency

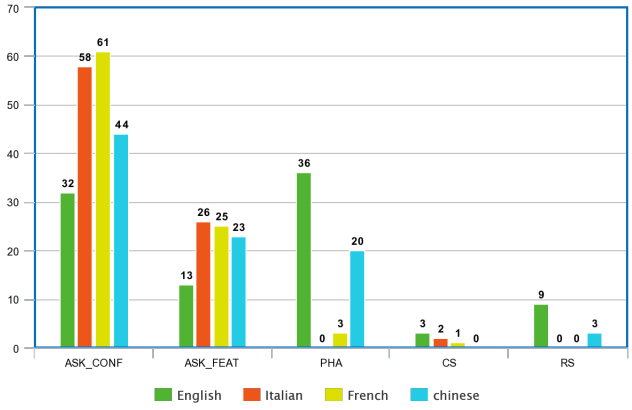

Figure 5.2: Question function frequency

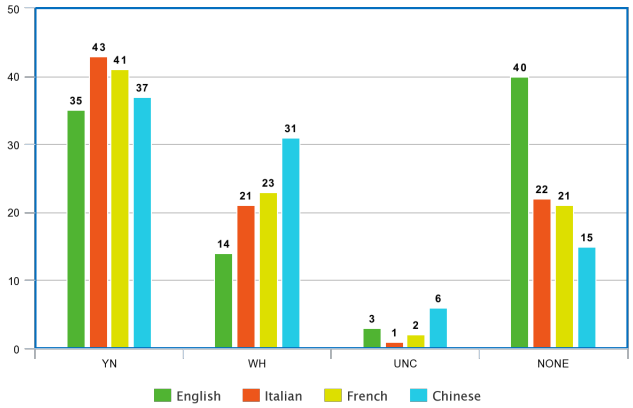

Figure 5.3: Answer form frequency

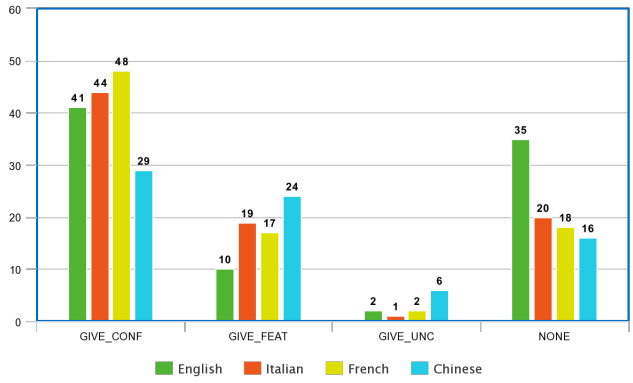

Figure 5.4: Answer function frequency

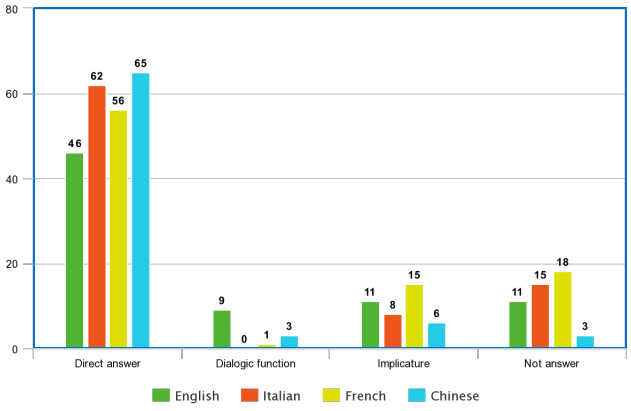

Figure 5.5: Type of answer frequency

Chapter 6

Conclusions

In this report, we presented the results carried out during our supervised project. The main objective was to build a fine-grained annotation schema that could easily adapt to multi-language annotations. In particular our goals were:

- To enrich the number of tags for questions to detect the different span of answers they can accept.
- Work on the possible forms that answers can take and on the possible functions they can convey.
- To investigate the reasons why the questions forms and functions are or aren't compatible with their respective answers.
- To detect some phenomena of indirectness.
- To propose a way to resolve complex exchanges containing multiple questions and answers within the same negotiation phase.

In what follow we will present the main challenges we encountered during our work and list all those that we think should and could be improved in further research.

- **Form and Function** While almost all the asymmetries between the expected answer function and the given answer function trigger a mismatch, only some of the asymmetry of forms do. In a further research we should be able to analyze the fact more closely.
- **Text segmentation** Sometimes a speaker asks a question and answers to him/herself. We didn't take into account the eventuality of a self-response because of the complexity of the already existing NP algorithm. The algorithm can be improved and segmented into different algorithms that account for the particular situation we are dealing with in a complex NP.
- **Questions within dialogue** In our proposal we tried to cover a wide variety of types of questions and answers. However, we think that two phenomena we encountered should deserve a special attention because of their frequent occurrence.
  - Rhetoric questions (see Example 1)
  - Ironic questions (see Example 2)

Example 1.
A: ma cosa devo fà?
what should I do?
(Speaker after throwing the dice and getting always the same number)

Example 2.
A: Le ruote sono rotonde?
And the wheels are round?
B: <risata>
((laugh))
A: Peccato era interessante le ruote quadrate <risata>
Pity, it squared wheels would've been interesting ((laugh))

- **Implicatures**

  In our work we treated only the cases where a speaker manages to infer a positive, negative or uncertain response. However, our corpora contained other cases of indirectness that are much more complex to analyze and resolve. Two examples from the Italian Catan corpus:

  Example 3.
  A: senti qualcuno ha un mattone?
  does anyone have rocks?
  B: che strana richiesta Mattew
  What a weird request Matthew.

  Example 4.
  A: ma noi siamo sicura che non ho fatto
  are you sure that I didn't build
  la strada più lunga in questo modo
  the longest route in this way?
  B: consecutive!
  consecutive!

  In Example 3 the speaker A must have asked for rocks many times during the game and he doesn't manage to gather this precise resource. Speaker B makes him infer that nobody wants to give him rocks and he should stop asking insistently.
  In Example 4 the speaker A asks whether the others are sure that he didn't build the longest route with his colonies. The speaker B responds "consecutive", which means that he has the number of routes needed to build the longest route but as they are not consecutive, he didn't build the longest route.

- **Formalization** On one hand our annotation schema is thought to cover many possibilities and be easily adaptable to every language. On the other hand its complexity makes its formalization hard. To improve the formalization we think that we must rethink about the Xml/Json tags used so that our annotations can be encoded in more proper classes.

It comes with no surprise that there is still much room for improvement and implementation. Indeed our work is conceived for the analysis of spontaneous dialogue, which is extremely complex. Nowadays, most of the spoken dialogue systems focus on task-based communication (making reservations, airlines and train information). The consequence of this is to limit the annotations to some highly specific purposes required by the system to develop. This is why during our work we found many corpora whose annotations were focusing on oriented-tasks but that could not easily be generalized. This is the case for example, of the SPAADIA[1] corpus, consisting of interactions between a call centre agent and her callers trying to get information about timetables and book trains. The corpus is annotated with Speech acts, which are a powerful tool to detect questions and answers functions and speakers' intentions in a dialogue. However, the Speech Acts in the corpus are limited to the specific task of booking and asking for information. Our annotation is thought for unplanned conversations, which makes the set of phenomena to annotate way wider than those in oriented-task dialogues. We believe that it constitutes a good (but, of course, not complete) generalization about what phenomena spontaneous human interaction entails. There is still much work to do on the side of formalization, but we hope that the study behind our annotation schema can help in the realization of new ways to formalize spontaneous dialogue.

[1]SPAADIA, retrieved from `http://martinweisser.org/`

Bibliography

[1] Stergos Afantenos, Eric Kow, Nicholas Asher and Jéremy Perret. "Discourse parsing for multi-party chat dialogues". *Conference on Empirical Methods in Natural Language Processing*, 2015.

[2] Stergos Afantenos, Nicholas Asher, Farah Benamara, Anaïs Cadilhac, Cédric Dégremont, et al. "Developing a corpus of strategic conversation in The Settlers of Catan". SeineDial 2012 - The 16th WORKSHOP ON THE SEMANTICS AND PRAGMATICS OF DIALOGUE, Sep 2012, Paris, France. 2012.

[3] Maria Aloni and Paul Dekker. "The Cambridge Handbook of Formal Semantics". *Cambridge University Press*, 2016.

[4] Patrick Amsili and Myriam Bras. "DRT et compositionnalité, T.A.L.". vol 39, no 1, pp.131-160.

[5] Maria Boritchev, Maxime Amblard, Laurent Prévot, Magalie Ochs and Benoît Favre. "Coffee or tea? Yes." *The 22nd workshop on the Semantics and Pragmatics of Dialogue*, Nov 2018, Aix-en-Provence, France, 2018.

[6] Nicholas Asher and Alex Lascarides. "Logics of conversation". *Cambridge University Press*, 2003.

[7] John Langshaw Austin. "How to do things with words". *Oxford University Press*, London, 1962.

[8] Maria Boritchev. "Approaching dialog modeling in a dynamic framework". MA thesis, Université de Lorraine, 2017.

[9] Joan Busquets, Laure Vieu and Nicholas Asher. "La SDRT : une approche de la cohérence du discours dans la tradition de la sémantique dynamique". Verbum, vol. 22, no 1, 2001, p. 73-102.

[10] Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. "Inquisitive semantics". *NASSLLI lecture notes*, volume 187, 2012.

[11] Corblin F. "Représentations du discours et sémantique formelle. Introduction et application au français". Paris, 1992.

[12] Helen Wright Hastie, Rashmi Prasad, Marilyn Walker. "What's the Trouble: Automatically Identifying Problematic Dialogues in DARPA Communicator Dialogue Systems". *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, U.S.A, July 2002, pp. 384-391.

[13] Martin Weisser. "DART – the Dialogue Annotation and Research Tool". *Corpus Linguistics and Linguistic Theory*, 12(2), 355-388, 2016.

[14] Philippe De Groote. "Towards a montagovian account of dynamics". *Proceedings of semantics and linguistic theory XVI*, M. Gibson and J. Howell (eds), SALT XVI 1-16, Ithaca, NY: Cornell University, 2006.

[15] "SCoSE Part 1: Complete Conversations". English Linguistics, Department of English at Saarland University, https://ca.talkbank.org/access/SCoSE.html, 2017.

[16] Jonathan Ginzburg. "Semantics of dialogue". *The Cambridge Handbook of Formal Semantics*. 2016.

[17] Jonathan Ginzburg and Ivan A. Sag. "Interrogative investigations". *Stanford: CSLI publications*, 2000.

[18] Herbert Paul Grice. "Logic and conversation". *Syntax and Semantics*, 1975.

[19] Jeroen Groenendijk and Martin Stokhof. "Studies on the Semantics of Questions and the Pragmatics of Answers". dissertation, Amsterdam, 1984.

[20] Jeroen Groenendijk and Martin Stokhof. "On the Semantic of Questions and the Pragmatics of Answers". *Varieties of Formal Semantics*, 1984.

[21] Hans Kamp and Uwe Reyle. "From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory". volume 42, Springer Science and Business Media, 2013.

[22] Jacqueline Leon. "Approche séquentielle d'un objet sémantico-pragmatique: le couple Q-R. Questions alternatives et questions rhétoriques". *Revue de Sémantique et de Pragmatique*, pp.23-50, Presses de l'Université d'Orléans, 1997,

[23] Laurent Prevot and Muller, P. and Denis, P. and Vieu, L. "Une approche sémantique et rhétorique du dialogue. Un cas d'étude: l'explication d'un itinéraire". https://hal.archives-ouvertes.fr/hal-01231956, Traitement Automatique des Langues, Association pour le traitement automatique des langues, 2002.

[24] Laurent Prevot. "Topic Structure in Route Explanation Dialogues". 13th EUROPEAN SUMMER SCHOOL IN LOGIC, LANGUAGE AND INFORMATION (ESSLLI2001), Helinski, Finland, *Workshop on Information Structure, Discourse Structure and Discourse Semantics*, pp.145, 2001.

[25] Van Rooy R., Safarova M. "On Polar Questions". ILLC University of Amsterdam.

Appendices

Appendix A

Deliverable

A.1 Annexes

Our annexes can be found on Google Drive at the following adress:
`https://drive.google.com/drive/folders/1jn9eD4mbadNfRdZSAy2FySmIS3dGqzpM`:

Folder**Annexes** containing for each language:

- Corpora
- Python_programs
- Outputs_Json
- Tables_comparison_languages

For access, please contact:
Marta Carletti at `martacarletti1993@gmail.com`
Léa Dieudonat at `leadieudonat@gmail.com`
Yi Ting Tsaï at `yi-ting.tsai5@etu.univ-lorraine.fr`

Appendix B

Annotation Guideline

B.1 Introduction

This guideline aims to explain the procedures step-by-step, clarify the different question and answer tags and understand the analysis of the interaction (form and function) of the QA pairs for facilitating the corpus annotation task.
The annotation is developed to be extremely fine-grained and it requires good linguistic analysis skills. Yet, the tag set is still easily adaptable to annotators of different background and can be effortlessly applied on various languages other than English. All examples used in the following sections are part of the Saarbrucken Corpus of Spoken English (SCoSE).

B.2 Setting Environment

Before jumping into the annotation task, it is primordial to set up the working environment. This section gives an explanation of prerequisite for applying the annotation task.

- **Obtaining corpus files**
  The SCoSE corpus are downloadable from https://ca.talkbank.org/access/SCoSE.html. Although it will be enough to download the transcriptions, it is recommended to also download the corresponding audio files. Listening to the audio files makes the annotation task easier, and, more importantly, there are some cases in which listening to the intonation pattern will help resolve ambiguities in the transcription. All the transcriptions of the corpora used in this project use the pdf format for the dialogues files.

- **XML/JSON Editing Tools**
  Our annotation output format is required to be in *.json* file. This can be edited on a text editor for coding, such as Sublime Text or Atom, in a more human-readable XML format at first hand, then converted into JSON format via online XML/JSON converter tool easily.

- **Standard Layer of Annotation Element Scheme**
  The standard layer of annotation element scheme should be like this:

```
<NPs>
    <NP1>
        <turn>
            <id>Speaker A</id>
            <text> </text>
                    <type>
                    Question
                    </type>
                    <question_type>
                        <question_form>

                        </question_form>
                        <expected_answer_form>

                        </expected_answer_form>
                        <question_function>

                        </question_function>
                        <expected_answer_function>
                        </expected_answer_function>
                    </question_type>
        </turn>
        <turn>
            <id>Speaker B</id>
            <text> </text>
                    <type>
                    Answer
                    </type>
                    <answer_type>
                        <given_answer_form>

                        </given_answer_form>
                        <given_answer_function>

                        </given_answer_function>
                        <feature>

                        </feature>
                    </answer_type>
        </turn>

        <mismatch_form> </mismatch_form>
        <mismatch_function> </mismatch_function>
        <direct_answer> </direct_answer>
        <indirect_answer> </indirect_answer>
        <has_dialogic_feature> </has_dialogic_feature>
        <has_implication> </has_implication>
    </NP1>
</NPs>
```

B.3 Tagging Questions

The process of question tagging will be explained here in this section. We explain our tag set in section B.3.1, and the detailed annotating steps in section B.3.3.

B.3.1 Tags for Questions

Our annotation scheme annotates the form and the function of a question separately, and the interaction of QA pair will be annotated after tagging the answer. The different formal and functional definitions of the questions are explained here.
We see each question as a combination of a **form** (the syntactic outlook and characteristics) and a **function** (the role that a question plays in the discourse). In this section, we define these tag sets, type by type.

- **Question Form and Expected Answer Form**
  The question form is the syntactic outlook and characteristics of a question, and the expected answer form predicts the form of a following plausible answer. They are treated as a pair and tagged as <question_form> and <expected_answer_form> elements respectively under the <question_type> layer. In this subsection, we define the tags for the question form and the expected answer form.

  – **Yes/No question YN** : In these types of question, a proposition is expressed and the expected answer will confirm or deny this proposition. **YN** questions look for a confirmation from the other participant of the conversation. The form of a **YN** question can contain do- support or inversion in English.
  Expected answer form associated with **YN** questions is $\texttt{Fo}_1$. Remember here we only examine a question and predict the form of its answer.

  Prototypical Examples 1

  (1) So- so you're gonna be HOME then? (SCoSE/Addie and Brianne a, 24)

  (2) Does it ? (SCoSE/Addie and Brianne a, 62)

  (3) Isn't that awful? (SCoSE/Addie and Brianne d, 8)

  – **Wh-questions WH** : This type of question is easy to spot as the presence of wh- constituent. Usually, this constituent, introduced by a wh-word (like what, who, where, when, which, how, etc), appears at the beginning of the question.
  Expected answer form associated with **WH** questions is $\texttt{Fo}_2$. Remember here we only examine a question and predict the form of its answer.

  Prototypical Examples 2

  (1) What color is it? (SCoSE/Addie and Brianne a, 67)

  (2) When will you guys get off? (SCoSE/Addie and Brianne a, 253)

– **Disjunctive inclusive question DQ_I** : Questions of this type can be spotted by the presence of *inclusive "or"*, to distinguish between different options which are proposed to the hearer. According to Definition 1 below, the intention of this type of questions is to obtain true or false value of one or two options proposed. **DQ_I** question accepts both **WH** and **YN** responses.

Definition 1. (Disjunctive Inclusive) *A disjunctive "or" is inclusive if the truth or the falsity of one disjunctive term implicates the truth/falsity of both.* Expected answer form associated with **DQ_I** questions is $\texttt{Fo}_3$. Remember here we only examine a question and predict the form of its answer.

Prototypical Examples 3

Are you a citizen of European Union or Switzerland? If yes, click here.

A: will you pay with a credit or a debit card? (inclusive "or")
B: Yes

– **Disjunctive exclusive question DQ_E** : Questions of this type can be spotted by the presence of *exclusive "or"*, to distinguish between different options which are proposed to the hearer. The hearer can choose one of the options or state a third option of his own, and **DQ_E** question accepts only a **WH** response.

Definition 2. (Disjunctive Exclusive) *A disjunctive "or" is exclusive when the truth of one option implies the falsity of the second one and vice versa.*

Expected answer form associated with **DQ_E** questions is $\texttt{Fo}_4$. Remember here we only examine a question and predict the form of its answer.

Prototypical Examples 4

Should I go to art school or should I stay in a university? (SCoSE/Addie and Brianne c, 92)

A: Do you want tea or coffee? (exclusive "or")
B: Tea

– **Auxiliary deontic question AUX_D** : Questions of this type can be detected by the presence of the *deontic auxiliary verbs*, such as "can" or "could". According to Definition 3 below, **AUX_D** question expresses the intention of the speaker to ask for a favor or a permission.

Definition 3. (Deontic Auxiliary) *The deontic function of auxiliary verbs expresses the possibility and necessity in terms of freedom to act (including permission and duty).*

Expected answer form associated with **AUX_D** questions is $\texttt{Fo}_5$. Remember here we only examine a question and predict the form of its answer.

Prototypical Examples 5

(1) Can you open the window?

(2) Could you pass me the salt?

– **Auxiliary epistemic question AUX_E** : Questions of this type can be detected by the presence of the *epistemic auxiliary verbs*, such as "can". According to Definition 4 below, the auxiliary verbs in **AUX_E** questions express the ability, capability or possibility.

Definition 4. (Epistemic Auxiliary) *The epistemic function of auxiliary verbs expresses the possibility of propositions being true or false.*

Expected answer form associated with **AUX_E** questions is $\texttt{Fo}_6$. Remember here we only examine a question and predict the form of its answer.

Prototypical Examples 6

(1) Can you survive all this?

(2) Can you get out of your contracts anyway? (SCoSE/Addie and Brianne a, 572)

- **Question Function and Expected Answer Function**

  The question function is the role that a question plays in the discourse, and the expected answer function predicts the function of a following plausible answer. They are treated as a pair and tagged as <question_function> and <expected_answer_function> elements respectively under <question_type> layer. In this subsection, we define the tags for the question function and the expected answer function.

  – **Completion suggestion CS** : The speaker completes the turn of another speaker. Sometimes, in the middle or at the end of an utterance, when the speaker hesitates to say what he/she means and the one who asks the question tries to help him/her to finish his utterance. The form of the CS-function questions can appear in different forms, such as **YN** or **WH**. In consequence, different from annotation guide of last year, **CS** is defined to be a question function.
  Expected answer function of **CS** is $\texttt{Fu}_1$. Remember here we only examine a question and predict the function of its answer.

  Prototypical Examples 7

  A: They're in Dubuque .. Christmas shopping and Amanda-
  B: Making an evening of it? (SCoSE/Addie and Brianne a, 14)

  A: It includes heat and uhm I think-
  B: Water? (SCoSE/Addie and Brianne b, 535)

  – **Phatic PHA** : Phatic questions are uttered to continue participation in communication or to indicate that the listener follows the conversation or to express surprise, rather than asking for information. Phatic questions often contain clichés like 'you know?' or 'right?'. In principle they can have vary forms and they must be distinguished by question functions.
  Expected answer function of **PHA** is $\texttt{Fu}_2$. Remember here we only examine a question and predict the function of its answer.

  Prototypical Examples 8

  I guess I could come back but.. you know? (SCoSE/Addie and Brianne c, 273)

  (2) A: They hire people at the Chamber of Commerce too.
  B: Really? (SCoSE/Addie and Brianne c, 281)

– **Ask confirmation ASK_CONF** : The function of this type of questions is uttered to seek for the confirmation from the hearer. The speaker asks the truth value of a proposition or the hearer's engagement to an action.
Expected answer function of **ASK_CONF** is $\texttt{Fu}_3$. Remember here we only examine a question and predict the function of its answer.

Prototypical Examples 9

(1) You have to PAY for all this? (SCoSE/Addie and Brianne a, 109)

(2) And everybody had to get it? (SCoSE/Addie and Brianne a, 140)

– **Ask feature ASK_FEAT** : The function of this type of questions is uttered to ask for the feature from the hearer. The question of function **ASK_FEAT** can come in different questions forms, like **WH** or **DQ_E**. The further feature tags of function **ASK_FEAT**, which are introduced in subsection B.3.2, must be annotated inside <feature> element afterwards.
Expected answer function of **ASK_FEAT** is $\texttt{Fu}_4$. Remember here we only examine a question and predict the function of its answer.

Prototypical Examples 10

(1) Is the wedding on Saturday or Sunday?(SCoSE/Addie and Brianne a, 237)

(2) What are your plans now? (SCoSE/Addie and Brianne a, 251)

– **Ask to perform ASK_PERF** : The function of this type of questions is uttered to ask to the hearer to perform an act or a task.
Expected answer function of **ASK_PERF** is $\texttt{Fu}_5$. Remember here we only examine a question and predict the function of its answer.

Prototypical Examples 11

(1) Can you open the door for the old lady?

(2) Could you please hand me my bag?

– **Reported speech RS** : Sometimes in dialogues, the speaker report someone's else question. In SCoSE transcription, this type of questions are quoted with quotation marks "".
Expected answer function of **RS** is $\texttt{Fu}_6$. Remember here we only examine a question and predict the function of its answer.

Prototypical Examples 12

(1) "Hi guys how are you doing?... bye now" (SCoSE/Addie and Brianne c, 294)

(2) "Will you get a job any time soon?" (SCoSE/Addie and Brianne c, 411)

B.3.2 Feature Tags

Tags are inherited from last year's work, but two main changes this year : firstly, both questions and answers should be tagged with their respective feature tag set if they have **GIVE_FEAT** or **ASK_FEAT** function ; secondly, the feature tags should be inserted at <feature> element layer right after the layer **GIVE_FEAT** or **ASK_FEAT** function. There are eight tags for features:

1. Temporality (**TMP**): the constituent refers to a moment or a period related to the event described in the question;
2. Location (**LOC**): the constituent refers to a location related to the event described in the question;
3. Agent (**AG**): the constituent refers to the person (or, less commonly, the entity) that performed the action described in the question;
4. Theme (**TH**): the constituent refers to the person or entity that underwent the action described in the question;
5. Owner (**OW**): the constituent refers to the person (or other entity capable of ownership, such as an organization) who owns (in a broad sense that also includes, for example, family relations) the entity that is described in it;
6. Reason (**RE**): the constituent refers to the the reason or motive behind the event described in the question;
7. Characteristic (**CH**): the constituent refers to a characteristic of the event described in the question.

Tag	**Name**
TMP	Temporality
LOC	Location
AG	Agent
TH	Theme
OW	Owner
RE	Reason
CH	Characteristic

Table A1: Features

B.3.3 Question Tagging Procedure

While annotating the corpus, it requires the line-by-line reading to include the also utterance in text for an accurate negotiation phase NP. For each question, please follow the steps beneath:

Step 1 : Identify the questions
An important criterion for a question to be question in our English corpus SCoSE is that it must contain an interrogation mark. Although, a question can be presented sometimes without a question mark, but rather in a specific syntactic structure, such as inversion. However, it would be hyper time-consuming and complex to apply such analysis on every utterance of the corpus. We tag only the questions with proper question marks, other types of presence of questions without interrogation marks would be ignored. Once identifying a question, we can proceed to step 2.

While applying the tag set on other languages, which may not have question mark in their standard writing system, we could determine a question by other specific syntactic structures or lexical characteristics in that language. Such as Japanese, in the formal writing system, a question is usually presented without a question mark, but with a sentence-final particle か ka.

Step 2 : Single or multiple questions
In a corpus, we might encounter several questions uttered at once by the same speaker, which means at the same single line of the corpus transcription, we could see one question followed by another question immediately. Or even in a same negotiation phase. To know how to analyze these exchanges we follow the following algorithm:

The reading of the graph is the following:
We suppose that the starting point is Speaker 1 asking a question (1_Qrel in the starting node). The end points are the two nodes in red colour representing (i) the end of the NP before the segment analyzed (EB) and (ii) the end of the NP after the segment analyzed (EA). Starting from the node Q_rel we have two main possibilities:
(i) The path of the answer.
(ii) The path of the question.

(i) The path of the answer

a. We find a response from Speaker 2 (2_R). We take the path 1.
b. The response can be an answer (path 4 to 2_A) or an unrelated statement (path 5 to 2_S) . In the first case we can take the path 9 and close the NP after the answer or find an unrelated question (path 10) and close the NP before it (path 13), in the second case we take the path 9.

If we take the path 9 we go on only if we encounter a new question otherwise we take path 16 and close the NP.

If we encounter a question:

b.1 the question is unrelated. We take the path 10,arrive at node (Q_unrel) and close the NP through path 13 (EB).

b.2 the question is related, we take the path 17 and loop.

Legend
1 = Speaker 1
2 = Speaker 2
Qrel = Question related
Qunrel = Question unrelated
R = Response (It can be S or A)
S = Statement
A = Answer
EB = End of NP before the segment
EA = End of NP after the segment
Numbers from 1 to 20: paths for the reader(references to smooth the explanation of the examples)

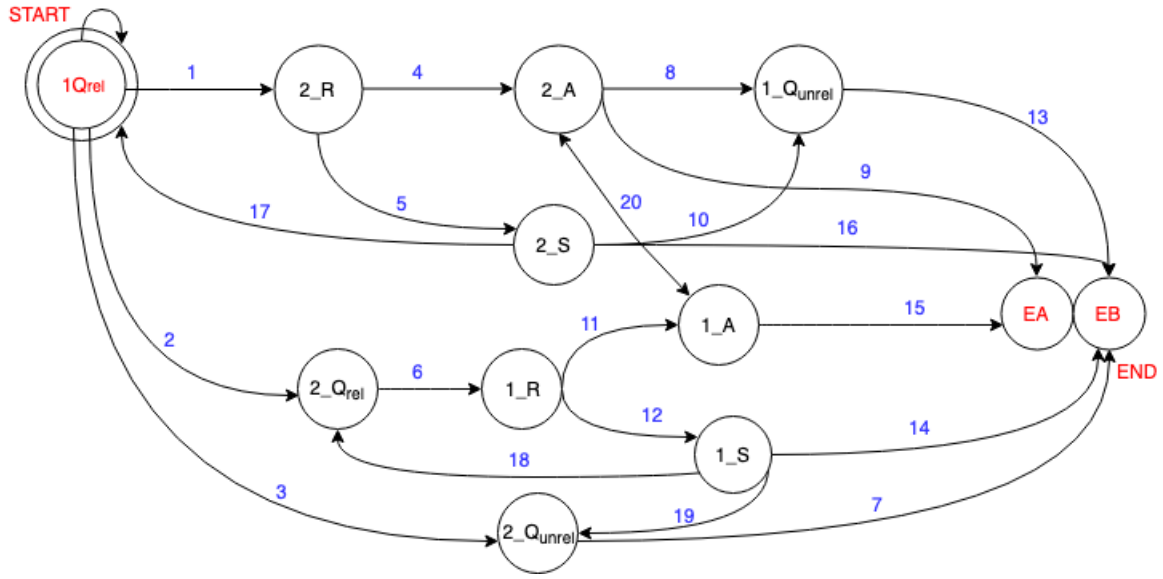

Figure A1: NP Algorithm

(ii) The path of the question:

a. We find a question from Speaker 2
   a.1 The question is related.We take the path 2.
   If the speaker A gives a response we have here the same choices that we had for the node 2_R but for Speaker 1 (1_R).
   a.2. The question is unrelated. We take the path 3 to 2_Qunrel. We close the NP through path 7.

Step 3 : Assign the question form and expected answer form
Go through all of the definitions of the question forms and expected answer form tags presented in previous section, and match both defined question form and expected answer form tags to your question. The question form and expected answer form should be tagged inside <question_form> element and <expected_answer_form> element under <question_type> layer.

Name	Tag	Description	Examples
"yes no"	YN	Contains do support, inversion	Are you fine?
"wh"	WH	Contains WH	What time is it?
"disjunctive inclusive"	DQ_I	Contains "or", inclusive	Are you a citizen of European Union or Switzerland? If yes, click here
"disjunctive exclusive"	DQ_E	Contains "or", exclusive	Do you want tea or coffee?
"auxiliary deontic"	AUX_D	Contains an auxiliary deontic	Can you open the window for me?
"auxiliary epistemic"	AUX_E	Contains an auxiliary epistemic	Can you survive all this?

Table A2: Forms of questions

Step 4 : Assign the question function and expected answer function
Go through all of the definition of the question function and expected answer function tags presented in previous section, and match both defined question function and expected answer function tags to your question. The question function and expected answer function should be tagged inside <quesiton_function> element and <expected_answer_function> element under <question_type> layer.

- If you assign **ASK_FEAT** function to that question, please go to step 5;
- Otherwise, go to step 6.

Name	Tag	Description
"Completion suggestion"	CS	The speaker completes the turn of another speaker
Phatic	PHA	Phatic function
"Ask confirmation"	ASK_CONF	The speaker asks the truth value of a proposition or the hearer's engagement to an action
Ask feature	ASK_FEAT	The speaker asks for a feature
Ask to perform	ASK_PERF	The speaker asks to perform an act
Reported speech	RS	The speaker report someone's else question

Table A3: Functions of questions

Step 5 : Assign feature tag
If a question's function contains **ASK_FEAT** tag at <feature> element after <question_function> under <question_type> layer, please further apply the analysis of feature tags, which are presented in B.3.2. Once match the proper feature tag accordingly to the question, we can proceed to step 6.

Step 6 : Check your annotation
Check your own annotations for current question. Please pay attention to the following:

- Check your question form/function and expected answer form/function are properly associated;
- Check that you included a tag in each layer that you needed to tag (e.g., if you annotated a question function as **ASK_FEAT** , make sure that you also include a feature tag inside <feature> element afterwards;
- Check that every tag is placed in the corresponding layer;
- Check that there are no typos in the tags that you used;
- Check that every element layer is closed properly.

B.4 Tagging Answers

The process of question tagging will be explained here in this section. We explain our tag set in section B.4.1, and the detailed annotating steps in section B.4.2.

B.4.1 Tags for Answers

Our annotation scheme annotates the form and the function of an answer separately, and the interaction of QA pair will be annotated after tagging the answer. The different formal and functional definitions of the answers are explained here.
We see each answer as a combination of a **form** (the syntactic outlook and characteristics) and a **function** (the role that a question plays in the discourse). We define the tag set of answer form and answer function, individually type by type, below.

- **Answer Form**
  The answer form is the syntactic outlook and characteristics of an answer, and it is tagged as <given_answer_form> element under <answer_type> layer. In this subsection, we define the tags for the answer form.

  - **Yes/No YN** : The form of this type of answers has either positive or negative forms, like in confirmation "yeah", "of course", "sure", or in refusal "no", "never", or including other lexical token, such as "not".
    The **YN** answer form tag is associated with $\texttt{Fo}_1$, $\texttt{Fo}_3$, $\texttt{Fo}_5$ or $\texttt{Fo}_6$ group, and can further correspond to **YN**, **DQ_I**, **AUX_E** or **AUX_D** question form tag.

    Prototypical Examples 13

    (1) A: What are your plans now?
    B: I don't have any plans. (SCoSE/Addie and Brianne a, 252)

    (2)A: Just for the hell of it?
    B: Yeah (SCoSE/Addie and Brianne b, 417)

    (3) A: Going looking for apartments?
    B: NO (SCoSE/Addie and Brianne b, 420)

  - **Uncertain UNC** : The form of this type of answers often contains expressions, like "maybe", "I am not sure", "could be", or other lexical tokens which demonstrate uncertainty or reservation in his/her utterance.
    The **UNC** answer form tag is associated with $\texttt{Fo}_1$, $\texttt{Fo}_2$, $\texttt{Fo}_3$, $\texttt{Fo}_4$ or $\texttt{Fo}_6$ group, and can further correspond to **YN**, **WH**, **DQ_I**, **DQ_E**,or **AUX_E** question form tag.

    Prototypical Examples 14

    (A: Are they?
    B: They probably plan to go. (SCoSE/Addie and Brianne a, 208)

    (2) A: Is the wedding on Sunday or is it Saturday?

B: It's- I don't even know (SCoSE/Addie and Brianne a, 238)

– **Unknown UNK** : The form of this type of answers often contains expressions, like "I don't know", "dunno", "have no clue",or other lexical tokens which show his/her lacking of knowledge or idea to the proposition in the question.
The **UNK** answer form tag is associated with $\texttt{Fo}_1$, $\texttt{Fo}_2$, $\texttt{Fo}_3$, $\texttt{Fo}_4$ or $\texttt{Fo}_6$ group, and can further correspond to **YN**, **WH**, **DQ_I**, **DQ_E**,or **AUX_E** question form tag.

– **Wh-constituent WH** : The form of this type of answers contains wh- constituent, plus it doesn't contain an explicit positive or negative form, nor expressions showing uncertainty/unknown.
The **WH** answer form tag is associated with $\texttt{Fo}_2$ group, and can further correspond to **WH** question form tag.

Prototypical Examples 15

(1) What color is it?
B: It's midnight blue. (SCoSE/Addie and Brianne a, 68)

(2) A: What day did you come back?
B: I came on Wednesday night. (SCoSE/Addie and Brianne a, 91)

– **None NONE** : Some forms don't correspond to any form that can be categorized according to our classification. The form of this type of answer doesn't contain an explicit positive or negative form, nor a WH constituent or a lexical tokens expressing uncertainty/unknown. Such forms were annotated as **NONE** in our schema.
The **NONE** answer form tag is associated with $\texttt{Fo}_5$ group, and can further correspond to **AUX_D** question form tag.

Prototypical Examples 16

(1) A: If you could just be home, you know?
B: Mhm-mhm (SCoSE/Addie and Brianne a, 167)

(2) A: Oh just go
B: Go?
A: Go. (SCoSE/Addie and Brianne a, 217)

NOTE : The second example here, its function is to convey a confirmation about the question asked. For these types of answers, we kept the form as NONE and attributed the conveyed function. Their analysis will be carried out in a second moment at the level of the pair.

– **Performance PERF** : The form of this type of answers expresses a specific performance action, which is responding to the question of **ASK_PERF** function. The answer form does not contain an explicit positive or negative form, nor a WH constituent or a lexical tokens expressing uncertainty/unknown.
The **NONE** answer form tag is associated with $\texttt{Fo}_5$ group, and can further correspond to **AUX_D** question form tag.

Prototypical Examples 17

A: Can you open the window?
B: I can open the window.

- **Answer Function**
  The answer function is the role that an answer plays in the discourse It is tagged as <given_answer_function> element under <answer_type> layer. In this subsection, we define the tags for the answer function.

  – **Give confirmation GIVE_CONF** : The function of this type of answers is to express the confirmation to the proposition in the question, in other words, to convey the truth values of a proposition.
  The **GIVE_CONF** answer function tag is associated with $\texttt{Fu}_1$, $\texttt{Fu}_2$ or $\texttt{Fu}_3$ group, and can further correspond to **CS** , **PHA** or **ASK_CONF** question function tag.

  Prototypical Examples 18

  (1) A: You have to PAY for all this?
  B: Oh yeah (SCoSE/Addie and Brianne a, 110)

  (2) A: You mean your town is kind of spread out like that?
  B: Yeah (SCoSE/Addie and Brianne a, 208)

  – **Give uncertainty GIVE_UNC** : The function of this type of answers is to convey the uncertainty about the proposition in the question. It's usually used when the speaker does not know the answer or cannot provide the information that is asked in the question.
  The **GIVE_UNC** answer function tag is associated with $\texttt{Fu}_3$, $\texttt{Fu}_4$ or $\texttt{Fu}_5$ group, and can further correspond to **ASK_CONF**, **ASK_FEAT** or **ASK_PERF** question function tag.

  Prototypical Examples 19

  A: Are they?
  B: They probably plan to go (SCoSE/Addie and Brianne a, 209)

  – **Give unknown GIVE_UNK** : The function of this type of answers is to express the unknown to the proposition in the question, in other words, to convey an unknown.

  The **GIVE_UNK** answer function tag is associated with $\texttt{Fu}_3$, $\texttt{Fu}_4$ or $\texttt{Fu}_5$ group, and can further correspond to **ASK_CONF**, **ASK_FEAT** or **ASK_PERF** question function tag.

  Prototypical Examples 20

  A: W- is the wedding on Sunday or is it Saturday?
  B: Its-I don't even know (SCoSE/Addie and Brianne a, 167)

  – **Accept ACCEPT** : The function of this type of answers is to accept to engage in an action.
  The **ACCEPT** answer function tag is associated with $\texttt{Fu}_1$, $\texttt{Fu}_2$ or $\texttt{Fu}_3$ group, and can further correspond to **CS**, **PHA** or **ASK_CONF**question function tag.

– **Refuse REFUSE** : The function of this type of answers is to refuse to engage in an action. It's usually used to express denial to the proposition in the question.
The **REFUSE** answer function tag is associated with $\texttt{Fu}_1$, $\texttt{Fu}_2$ or $\texttt{Fu}_3$ group, and can further correspond to **CS**, **PHA** or **ASK_CONF**question function tag.

Prototypical Examples 21

A: Just for the- did I tell you about that?
B: No (SCoSE/Addie and Brianne b, 419)

– **Give feature GIVE_FEAT** : The function of this type of answers is to give a feature to the proposition in the question. Remember, when you have **GIVE_FEAT** as answer function, please include feature tags, which is introduced in subsection B.3.2, inside <feature> element afterwards.
The **GIVE_FEAT** answer function tag is associated with $\texttt{Fu}_4$ group, and can further correspond to **ASK_FEAT** question function tag.

Prototypical Examples 22

(1) A: What day did you come back?
B: I came on Wednesday night (SCoSE/Addie and Brianne a, 91)

(2) A: Ahat for?
B: To get ready for school (SCoSE/Addie and Brianne a, 246)

– **Give performance GIVE_PERF** : The function of this type of answers is to perform the action requested from the proposition in the question.
The **GIVE_PERF** answer function tag is associated with $\texttt{Fu}_5$ group, and can further correspond to **ASK_PERF** question function tag.

Prototypical Examples 23

A: Can you please give me that book?
B: I can give you the book.

– **Phatic PHA** : Phatic answers are uttered to continue participation in communication or to indicate that the speaker follows the conversation or to express surprise. Phatic questions often contain clichés like 'right'.
The **PHA** answer function tag is associated with $\texttt{Fu}_1$, $\texttt{Fu}_2$ or $\texttt{Fu}_6$ group, and can further correspond to **CS**, **PHA** or **RS** question function tag.

Prototypical Examples 24

A: "Should I go to art school or should I stay in a university?"

B: Right (SCoSE/Addie and Brianne c, 95)

– **Report REPORT** : The function of this type of answers is that the answer is reported from someone else's speech. In SCoSE transcription, this type of answers are quoted with quotation marks "".
The **REPORT** answer function tag is associated with $\texttt{Fu}_1$, $\texttt{Fu}_2$ or $\texttt{Fu}_6$ group, and can further correspond to **CS**, **PHA** or **RS** question function tag.

Prototypical Examples 25

A: 'Profesor, can you drop our lowest quiz?'
B: You have- you know like 'oh sure' (SCoSE/Amy Briget B a, 25)

– **None NONE** : We tag the answer function as **NONE** when no answer is given. Even though an utterance could be given by the speaker, but not necessarily being the answer to the proposition in the question.
The **NONE** answer function tag is associated with $\texttt{Fu}_2$, $\texttt{Fu}_5$ or $\texttt{Fu}_6$ group, and can further correspond to **PHA**,**ASK_PERF** or **RS** question function tag.

Prototypical Examples 26

(1) A: If you could just be home, you know?
B: Mhm-mhm (SCoSE/Addie and Brianne a, 167)

(2) A: So- wh- where can you move to?
B: Well .. you see (SCoSE/Addie and Brianne a, 398)

B.4.2 Answer Tagging Procedure

Same as question tagging procedure B.3.3, it requires the line-by-line reading to include the also utterance in text for an accurate negotiation phase NP. For each question, please follow the steps beneath:

Step 1 : Identify the answers
The hardest part of answer tagging is identifying what and where exactly the answer is. What you should do is try to find a single, complete utterance that satisfies the following conditions:

- The utterance should be uttered shortly after the question.
  - Unfortunately, 'shortly' cannot be defined objectively. In many cases, the answer will be the utterance directly after the question, but in some cases, it can be one or a few utterances after that.
  - In some cases, the answer will interrupt the question, i.e., the answer will be given while the question is still being formulated.
- The utterance should be pragmatically and/or semantically related to the question that it answers.
  - Prototypically, the answer will be uttered in response to the question (pragmatics) and contain the information that was requested in the question (semantics).
  - In some cases, the utterance can be indirect, or be semantically unrelated to the question. However, if it is uttered as a response to the question, it should still be considered to be an answer.
  - If the utterance is uttered in response to either an external stimulus or to an utterance other than the question, it should not be tagged as an answer.
- The utterance is not itself a question.
  - If a question is directly followed by another question, this second question should always be tagged as a question rather than as an answer, even if it is asked in response to the earlier question.
- Interjections (e.g. 'uhm', 'mhm mhm') can also be answers.

Step 2 : Assign answer form
Go through all of the definitions of the answer form tags presented in previous section. The answer form should be tagged as <given_answer_form> element under <answer_type> layer.

Name	Tag	Examples
"yes no"	YN	Yes, yeah, yep, jeez, sure, of course, absolutely.. No, nope, no way, not at all, nah..
"wh"	WH	I go home tomorrow, When I.., Because I..
"uncertain"	UNC	I'm not sure, maybe, still don't know, could be..
"unknown"	UNK	I don't know, dunno, have no clue..
"performance"	PERF	I can open the window, I can't help to turn off the light..
"none"	NONE	Mhm-mhm, Ugh, or other utterance doesn't fit into the categories above..

Table A4: Answer Forms

Step 3 : Assign answer function
Go through all of the definitions of the answer function tags presented in previous section and tag the answer function as <given_answer_function> element under <answer_type> layer.

- If you assign **GIVE_FEAT** function to that answer, please go to step 4;
- Otherwise, go to step 5.

Name	Tag	Description
Refuse	REFUSE	The speaker refuses to engage in an action
Accept	ACCEPT	The speaker accepts to engage in an action
Phatic	PHA	Phatic function
Give confirmation	GIVE_CONF	The speaker conveys the truth values of a proposition
Give uncertainty	GIVE_UNC	The speaker conveys uncertainty about something
Give unknown	GIVE_UNK	The speaker conveys an unknown response
Report speech	REPORT	The speaker reports someone else speech
Give feature	GIVE_FEAT	The speaker gives a feature
Perform	PERF	The speaker performs the act requested
NONE	NONE	No answer is given

Table A5: Functions of answers

Step 4 : Assign feature tag
If a answer's function contains **GIVE_FEAT** tag, please further apply the analysis of feature tags at <feature> element, which are presented in B.3.2. Once match the proper feature tag accordingly to the question, we can proceed to step 5.

Tag	**Name**
TMP	Temporality
LOC	Location
AG	Agent
TH	Theme
OW	Owner
RE	Reason
CH	Characteristic

Table A6: Features

Step 5 : Check your annotation
Check your own annotations for current question. Please pay attention to the following:

- Check that you included a tag in each layer that you needed to tag (e.g., if you annotated a question function as **GIVE_FEAT** , make sure that you also include a feature tag inside <feature> element afterwards);
- Check that every tag is placed in the corresponding layer;
- Check that there are no typos in the tags that you used;
- Check that every element layer is closed properly.

B.5 Analysis of QA interaction

The first thing we do is to check the compatibility of the <expected_answer_form> and the <given_answer_form>. If the tag of the <given_answer_form> is present in the group corresponding to the <expected_answer_form> (see Table below) we signal NO in the <mismatch_form> section.

Question Forms	Expected answer forms
YN	$\texttt{Fo}_1$ ⟨ YN,UNC,UNK ⟩
WH	$\texttt{Fo}_2$ ⟨ WH,UNC,UNK ⟩
DQ_I	$\texttt{Fo}_3$ ⟨ YN,UNC,UNK ⟩
DQ_E	$\texttt{Fo}_4$ ⟨ WH,UNC,UNK ⟩
AUX_D	$\texttt{Fo}_5$ ⟨ YN, NONE, PERF ⟩
AUX_E	$\texttt{Fo}_6$ ⟨ YN,UNC,UNK ⟩

Table A7: Compatibility form

The second thing we do is to check the compatibility of <expected_answer_function> and the <given_answer_function>. If the tag of the <given_answer_function> is present in the group corresponding to the <expected_answer_function> (see Table below) we signal NO in the <mismatch_function> section.

Question Function	Expected answer function
CS	$\texttt{Fu}_1$ ⟨ REFUSE, ACCEPT, PHA, GIVE_CONF, REPORT ⟩
PHA	$\texttt{Fu}_2$ ⟨ REFUSE, PHA, GIVE_CONF, REPORT, NONE ⟩
ASK_CONF	$\texttt{Fu}_3$ ⟨ REFUSE, ACCEPT, GIVE_UNC, GIVE_UNK, GIVE_CONF ⟩
ASK_FEAT	$\texttt{Fu}_4$ ⟨ GIVE_FEAT, GIVE_UNC, GIVE_UNK ⟩
ASK_PERF	$\texttt{Fu}_5$ ⟨ GIVE_PERF, NONE, GIVE_UNK, GIVE_UNC, YN ⟩
RS	$\texttt{Fu}_6$ ⟨ PHA,REPORT,NONE ⟩

Table A8: Compatibility function

Finally, we go on according to this algorithm:

Step 1 : Check <mismatch_form>
According to the table A7, we check if the tags inside <question_form> and <given_answer_form> are matchable or not. At the element of <mismatch_form>, if the tags entail a mismatch, tag YES ; else, tag NO. Then, proceed to Step 2.

Step 2 : Check <mismatch_function>
According to the table A8, we check if the tags inside <question_function> and <given_answer_function> are matchable or not. At the element of <mismatch_function>, if the tags entail a mismatch, tag YES ; else, tag NO. Then, proceed to Step 3.

Step 3 : Check <direct_answer>
If the answer fulfills the semantic requirements of the question, we tag YES at <direct_answer> and STOP.

Definition 1. Direct Answer *The direct answer is an answer fulfilling all the semantic requirements of the questions.*

Step 4 : Check <indirect_answer>
If the answer doesn't fulfill the semantic requirements of the question, at <indirect_answer> element, tag YES and continue to Step 5 ; else, tag No and proceed to Step 7.

Definition 2. Indirect Answer *The indirect answer is an answer that doesn't fulfill all the semantic requirements of the questions.*

Intuitively, if we signal NO in <direct_answer>, we should signal YES in <indirect_answer>. However, we have to further examine this session of negotiation phase, question's dialogic function and answer's implicature.

- If we signal NO at both <direct_answer> and <indirect_answer>, it means that no answer is present here, we should end the query and proceed to Step 7 directly.
- If we signal YES at <indirect_answer>, we proceed to Step 5 and detect where the indirectness lies.

Step 5 : Check <has_dialogic_function>
Please check next page for more details on dialogic features. Dialogic functions mean the other functions and intentions outside the scope of QA pair functions. If the pair has dialogic function, at <has_dialogic_function> element, tag YES, also give the dialogic feature tag at <id> element and STOP ; else, tag No and proceed to Step 6.

Step 6 : Check <has_implication>
Please check page afterwards for more details on implications. If the answer contains a linguistic implication, at <has_implication> element, tag YES, also give the inference of the indirect answer at <id> element and STOP ; else, tag No and proceed to Step 7.

Step 7 : Check <not_answer>
If both tags at <direct_answer> and <indirect_answer>, we tag YES at <not_answer> element and STOP.

NOTE:

- A mismatch of form and function together always triggers a mismatch.
- A mismatch of only form and no function triggers a mismatch but not frequently (must be checked).
- A mismatch of only function often triggers a mismatch.
- Symmetry of form and function should never trigger a mismatch.

B.5.1 Dialogic Features

Once at <has_dialogic_function> has a YES tag, we need to tag the dialogic feature of that question at <id> element afterwards. Below are the dialogic feature tags available for the moment, with different contexts encountered, tags could be further developed if needed.

- **Comment statement**: comment on previous statement.

Prototypical Examples 27

A: oh really?
B: It's so late to be home.(SCoSE/Addie and Brianne a, 19)

A: what kind of a health standard is that?
B: he's so goofy (SCoSE/Amy bridget A e, 62)

- **Irony**: the speaker is ironic and doesn't expect a real answer from another speaker.

Prototypical Examples 28

A: oh God
A: I mean school is great
A: **who wants to work?**
B: so, are you gonna be done after four years or? (SCoSE/Amy bridget B d, 408)

A: you know
A: 'oh Brianne do you have your cherry berry?'
B: [((laughs))]
A: **he has all kinds of names you know?**
B: uh-huh (SCoSE/Amy bridget A e, 32)

- **Rhetoric**: speaker asking questions and answering to him/herself.

Prototypical Examples 29

B: and then uhm
B: what else do I have?
B: oh I have sculpture (SCoSE/Amy bridget B a, 334)

B.5.2 Implications

Once at <has_implications> has a YES tag, we need to tag the implication of that answer at <id> element afterwards. Below are the implication tags available for the moment, with different contexts encountered, tags could be further developed if needed.

- **Infer no**

Prototypical Examples 30

A: if you could just be home, you know?
B: mhm-mhm
A: if i had car (SCoSE/Addie and Brianne a, 165)

- **Infer uncertainty**

Prototypical Examples 31

A: can you get out of your contracts anyway?
B: uhm
B: I talked to a couple of different people who have done it (SCoSE/Addie and Brianne b, 572)