

Building a chatbot for dialogic teaching

Dialogic teaching and the potential of chatbots

Dialogic teaching involves learning through discussion. It has shown positive results for the **acquisition of content**, and may also help improve students' **cognitive abilities**. However, implementing this pedagogy would require **some conditions that are not always attainable**, such as sufficient teaching staff, curricula open to different approaches, enough time, basic dialogic skills on the part of the students... Some of these issues could at least be partially solved through the use of **chatbots**, which could become a **surrogate teacher** for certain tasks, be it in class or at home. It could also serve as a **training tool for students whose dialogic skills fall behind** those of their peers, so that they can participate in classes where dialogic teaching is already implemented, or in academic cultures where discussions are common. Moreover, it might prove very useful in **language exam preparation**, especially in circumstances like the current pandemic, where students have to prepare at home.

Chatbots in education

Chatbot use in general has increased a lot over the past few years, as the technology has become more effective and accessible, and their benefits for businesses more evident. However, they can also be useful outside the business world. In education, there are quite a few successful experiences with chatbots, though in most cases the chatbots' role was limited to **providing information** about a course's syllabus or an institution's admission policies (similarly to what many companies do when they implement a chatbot for users to access their FAQ). There also exist some chatbots that assist students with language learning by **presenting grammar or vocabulary exercises**, but this does not exploit the full potential of chatbots as conversational tools. Some research has been done in this latter area, showing that **generic chatbots** may not meet the needs of language learners – a **pedagogically motivated chatbot** design is thus needed.

General ideas about chatbots

Chatbot components

Important elements of a chatbot that need to be considered for its design:

- **Intents:** tasks that we want our chatbot to fulfill (e.g. giving the student arguing points, pressing them to reason their answer, etc.).
- **Utterances:** the phrases that the chatbot will need to process and relate to the correct intent (e.g. User: "Do you think that this is a good idea?" → Chatbot needs to express an assessment of the idea and keep the discussion going). For each intent there may be infinite possible utterances, fewer the narrower the intent is; therefore, it is important for the chatbot to be able to process a wide range of utterances related to the same intent.
- **Entities:** objects that are relevant for an intent (e.g. when a student expresses an opinion, the topic and the polarity could be entities).
- **Decision tree:** set of conditional clauses that allow the chatbot to know what has been established in the conversation and what needs to be done depending on the input it's receiving (e.g. if a conclusion has been reached in the argument and reasons have been given to back it up, the chatbot can end the discussion...but only if those conditions are met).
- **Confidence threshold:** we can have our chatbot return the first answer that might fit the input (and risk sometimes giving useless, nonsensical answers), or we can take into account the

confidence score, the probability that the chatbot chose a correct answer, and only give answers that it is confident will be wanted (and perform another action when it's unsure, like asking the user to repeat their question or changing the topic).

Chatbot designs

Chatbots can vary enormously in their design. Looking at popular chatbot design platforms and libraries, I observed differences primarily in these aspects:

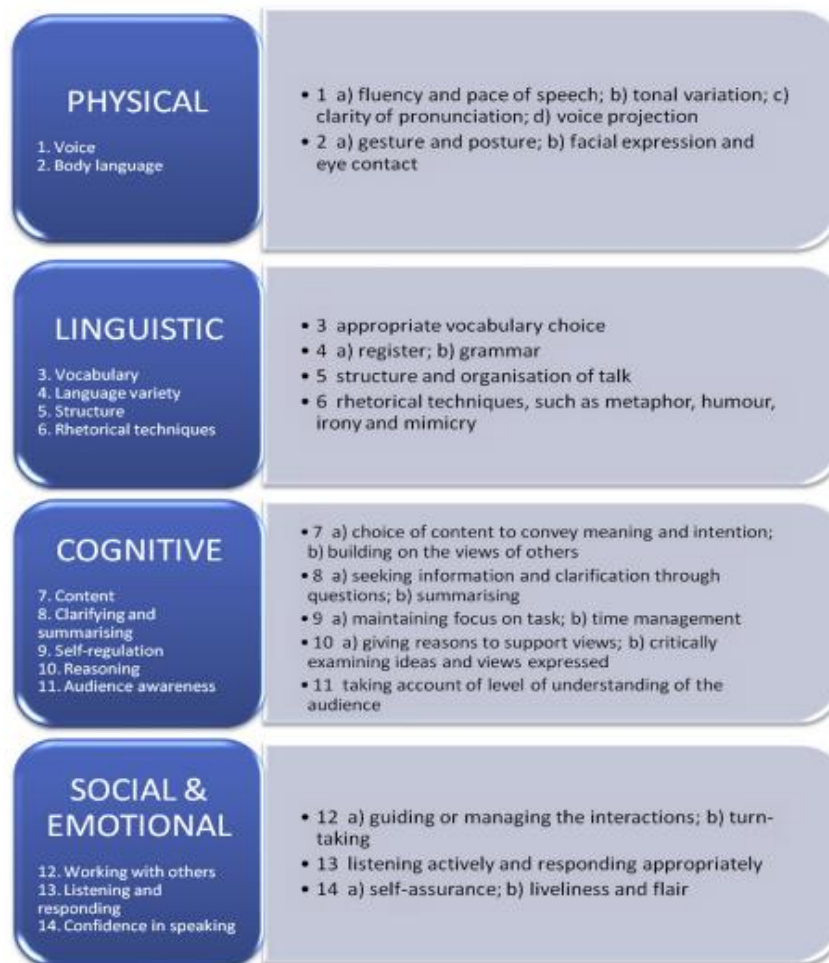
- Type of input: a chatbot may receive written or spoken input, or both.
- Type of output: as with the previous criteria, the output can be given in writing or orally.
- Breadth of topics: some chatbots are designed to carry conversations about as many topics as possible (e.g. just to entertain the user or to really be a universal assistant), whereas others cover more specific areas (e.g. many chatbots simply replace a website's "FAQ" section to present that very narrow content in chat format).
- Approach to NLP: some chatbot design tools depend on keyword/pattern matching (e.g. the chatbots designed with the Pandorabots platform use Artificial Intelligence Markup Language, which requires the developer to write the different linguistic patterns that the input should match and a set of possible answers for each case); other designs rely more on machine learning, drawing on large amounts of data that allow the tool what kinds of answers are most common for each utterance.
- Deployment: some chatbots are web-based tools, others may be developed as phone apps, others may be integrated with pre-existing messaging apps (Whatsapp, Telegram, etc.), and others may even be used with voice-activated assistants (like the chatbots developed with Dialogflow, which can be used with Google Assistant)

My chatbot

With regard to the classification of chatbot design that I outlined in the previous section, this chatbot will be as follows:

- Type of input: written input (to simplify the process, voice recognition is not contemplated here). This will also allow the student to think more carefully and practice returning their best possible answer (the student will be using the "learned system" as a "monitor", following Krashen's influential terminology, to hopefully practice and automatize some processes that will allow them to eventually develop fluency and conduct real-life discussions using the "acquired system").
- Type of output: written output (for simplicity, again)
- Breadth of topics: As a small demo, this chatbot will only cover the topics necessary for a single speaking exercise. To develop a chatbot that could carry out all the FCE model tasks available, a large dataset of such tasks would be needed to either train a system or inform developers of a rule-based classifier.
- Approach to NLP: Given the scarcity of data that is suitable for this very specific task, the best results will be obtained with a pattern/keyword matching approach.
- Deployment: Again, to simplify the development of this demo, only the Python code will be provided, without trying to develop a GUI and deploying the tool for use online or in an app.

As a pedagogically motivated tool, the chatbot's features will be aimed at strengthening specific dialogic subskills, as classified below by Mercer et al.:



Skills 3, 4, 5, 11 and 12 will be practiced through the whole interaction – if the student doesn't speak clearly or respect turns, the chatbot won't be able to respond satisfactorily. Also, if the student sees that the chatbot isn't understanding, they will need to adapt (skill 11).

Intents and decision tree

Greeting the user

Presenting the task

Asking the student to give an opinion: the tool will need to see if the student simply mentions a discussion point without evaluating it.

Asking the student to give reasons to support their views: the tool will need to see when a new opinion is introduced and whether the utterance includes a justification. (Skill 10)

Agreeing with the student: the tool will need to see the logic in the utterance.

Tentatively agreeing with the student: when the tool is not confident about how logical the utterance is, as long as it is backed with reasons, the tool can consider it valid.

Disagreeing with the student: the tool will need to identify the idea expressed in the utterance and return an utterance with arguments that contradict it. The frequency at which the chatbot disagrees with the student could be a customizable parameter to control difficulty.

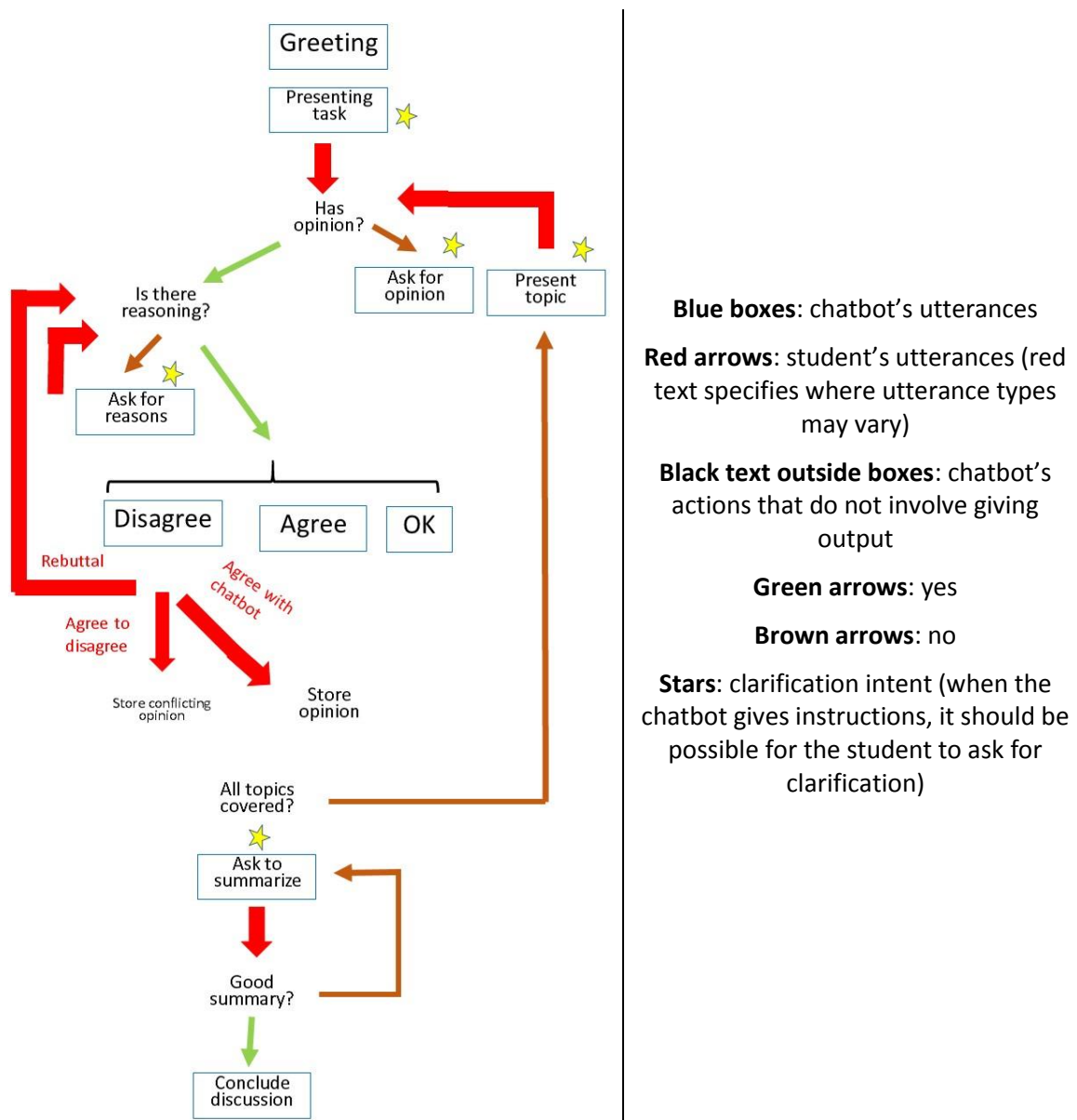
Asking for clarification (Skill 11)

Giving clarification: ideally the tool will be able to rephrase a statement in simpler terms. Also, if the pipeline were to include voice generation, the tool should be able to adapt speed. (Skill 8)

Introducing new points to discuss: the tool will need to keep track of the points that have already been discussed. (Skill 7b if the chatbot also includes an opinion statement)

Asking the student to summarize the conclusion: the tool will need to check whether all the necessary points have been covered for the student to summarize the discussion. (Skill 8b)

Concluding the discussion: the tool will need to check that all the topics have been successfully covered and that a conclusion has been given.



In order to simplify the demo, some intents will not be included:

- Giving clarification
- Disagreeing

- Agreeing (the chatbot will simply accept valid answers as valid, without giving opinions)

Data used

To simplify the task of designing an appropriate discussion prompt, a **Cambridge FCE exam speaking task** will be used for the demo (from the examples available in their website). This ensures that the prompt has been tested and that it is suitable for an intermediate level of English. It must be noted, however, that this also means that the instructions are given in a UK variant of English. The task chosen is the last part of the exam, where exam-takers are asked to interact with each other and discuss several options to pick the ones that best solve a problem introduced by the examiner.

NLP Tasks

Tokenization & lemmatization: the input and the training data will need to be processed to be analyzed.

Sentiment analysis: to detect whether the student has given an opinion or has just uttered a neutral statement, as well as to store as entities the ideas that were considered good and should thus be mentioned in the conclusion.

Text classification (though pattern matching): the inputs will need to be compared with the chatbot's data to see what the student is talking about and also to check whether the statements feature reasoning connectors.

Chatbot features

To better help students develop oracy skills, the chatbot will not only be a chatting tool; it will also give some feedback. In order to do that, it will need to store some information:

- **Usage of expressions for reasoning:** the chatbot will count how many words used for giving reasons were used. For this demo, as there should be reasoned statements about five topics and a conclusion, the threshold has been set to 7, to promote using complex sentences/paragraphs with more than one connector. In addition to that, the chatbot will see how varied the connectors are (if a student uses connectors very often but they are all the same one, that is not a sign of great language skills). As there are many connectors in the English language that the student could employ in the six expositions that constitute the minimum to complete the demo, the threshold has been set to 3. Even when a student uses a wide variety of connectors but happens to repeat one of them a few times, it doesn't hurt to be reminded to avoid repetitions, to develop some awareness of one's own speech.
- **[Seeking agreement:** the chatbot should count how many times the student ask to know the chatbot's opinion. However, as the agreeing and disagreeing intents have been removed for the demo, it would make no sense to expect the student to ask the opinion of a tool that isn't expressing opinions; thus, this part of the feedback has been removed.]
- **Duration of the discussion:** even when a student is able to cover all the topics satisfactorily, the interaction might be too short for a language exam; on the other hand, it could be too long, signaling that the student needs to improve their time-managing skills. Given that, in this chatbot demo, the interaction takes place in writing (this will increase the time needed), as well as the fact that the disagreeing/agreeing intents have been removed (this will reduce the time needed), the threshold of appropriate duration has been established as 200-300 seconds (the spoken task would take 4 minutes in the exam). Using a new, unfamiliar tool would also increase the time needed, hence this broad threshold.

- **[Turn-taking:** as the output of the chatbot has been limited in this demo, the feedback feature that compared how much the student talked with the chatbot's output has been removed.]

Developing the chatbot

Step 1: Coming up with the intents and the decision tree

Step 2: Programming the timing function

Step 3: Copying the task instructions and discussion topics into the code

Step 4: Testing sentiment analysis functions

Results obtained with TextBlob seem promising, so I will use that library.

```

muypositivo = TextBlob("I think that building a nightclub would be a wonderful idea. People love nightclubs!")
positivo = TextBlob("I think that building a nightclub might be a good idea. Many people like nightclubs.")
neutro = TextBlob("Maybe we can discuss the idea of building a nightclub.")
negativo = TextBlob("I think that building a nightclub would be a bad idea. Many people don't like nightclubs.")
muynegativo = TextBlob("I think that building a nightclub would be an awful idea. Most people hate nightclubs!")

muypositivo: Sentiment(polarity=0.8125, subjectivity=0.8)
positivo: Sentiment(polarity=0.6, subjectivity=0.55)
neutro: Sentiment(polarity=0.0, subjectivity=0.0)
negativo: Sentiment(polarity=-0.09999999999999992, subjectivity=0.5833333333333333)
muynegativo: Sentiment(polarity=-0.5, subjectivity=0.7999999999999999)

```

Looking at these figures, and after performing additional steps at later steps, I set the polarity thresholds as:

- Positive opinion: above 0
- Neutral statement: 0
- Negative opinion: below 0

Step 5: Testing tools for the chatbot to know which **topic** is being discussed

- Option 1: TextBlob's classifier based on **Machine Learning**
After 100 iterations, both the Max Entropy and Naïve Bayes classifiers reach 80% accuracy. However, as they were trained with just 25 sentences, the accuracy on the testing set (other 25 sentences) was only slightly above 30%. For this system to be a good option, I would need a lot of training data; for such a specific classification task, no data is publicly available, so I'd have to type thousands of sentences – not feasible for a demo.
- Option 2: Building a **keyword/pattern-based** classifier. From my previous experience building rule-based classifiers, I know that this can be a very effective solution, and I have experience thinking of synonyms, so hopefully my classifier should have good precision and recall (these measures not tested here for the demo, as that would require external data).

Step 6: Building another classifier to detect **reasoning**. For this, I again used regular expressions. This time, they matched a long list of connectors that the students could use to develop their ideas. Both simple and very advanced connectors were included, as either might be used at an intermediate level (even someone who is not very fluent may learn some advanced connectors to make their discourse appear more sophisticated). Other cohesion devices were not considered, partly due to the difficulty of detecting them without reducing the precision of the classifier and also because at an intermediate level student should already know enough connectors and be encouraged to use them.

Step 7: Building yet another classifier to evaluate the **conclusion** of the discussion. Ideally, this should be through aspect-based sentiment analysis, so that the student could summarize all the conclusions reached in the discussion, regarding both ideas they liked and ideas they discarded.

However, here the system was simplified by asking the student to simply mention the ideas that they chose as most suitable to solve the problem presented in the speaking prompt. The classifier then checks those ideas against the opinions stored from the rest of the chat, in search for incoherence or incompleteness.

Step 8: Building a function that analyzes the connectors used throughout the discussion to give feedback

Step 9: Building the main structure of the chatbot and integrating the functions

Step 10: Debugging

To test all the possible scenarios covered by the chatbot, I will simulate a discussion with this plan:

- The student first makes a complete exposition on the topic of clubs, with a positive opinion.
- The chatbot suggests topic 2, and the student gives a negative opinion without backing it.
- The student is asked to give reasons and does so.
- The chatbot suggests topic 3, and the student makes a neutral statement.
- After being asked to express a clear like or dislike, the student expresses dislike and gives a reason.
- The chatbot suggests topic 4, and the student makes a complete exposition with positive opinion.
- Finally, the chatbot suggests topic 5, and the student makes another complete exposition, this one with a negative opinion.
- The chatbot asks the student to summarize the conclusion
- The student says that topics 2 and 3 are good ideas.
- Being asked to give a correct conclusion, the student now says that topics 1 and 2 are the best ideas.
- After being asked again, the student says that topic 1 is the best idea.
- Finally, the student gives a satisfactory conclusion (with topics 1 and 4 as the best ideas).
- The chatbot returns feedback

Conclusions

Limitations

As this is developed as a mere demo only to illustrate the potentiality of this type of tool, numerous aspects have been simplified. The system is completely text-based, though the ideal tool should include **voice** recognition and voice generation. Naturally, for this tool to be appealing and accessible to its target user (language students of an intermediate level), it would need to be deployed via web or some **application** and an attractive **GUI** would need to be developed. The interface would not only make the tool more appealing, but should also make its use easier. For instance, the **task prompt** could remain visible to the student at all times, the fonts could be adjusted to see the text better, the chatbot could be represented with an image for the student to see it more as a partner...

Another limitation resulting from the simplification of the chatbot is that, with the reduced number of intents, the chatbot becomes less like a **partner** and more like a **teacher**. This does not necessarily make it a less effective tool, but means that the task in this case becomes less like the real exam task.

Yet another great limitation has to do with the **topics** covered and how they are recognized. As was mentioned, covering a larger number of topics would require a large amount of data to train a classifier or inform human developers of a rule-based classifier. Moreover, the way that the classifier

in the demo works, it can only discuss one topic at a time. The tool might still be useful to practice simple discussion, though an ideal tool would be able to deal with several topics at once, as the best-performing students in this task are expected to compare ideas instead of simply addressing them one by one.

Even though the TextBlob **sentiment analysis** tool performed reasonably well, further testing has shown that some types of utterances can be problematic. Particularly, when a positive opinion is negated (e.g. “I don’t think that’s a good idea”), the tool gets confused. Also, some sentences that express like or dislike very simply can be misinterpreted as neutral (e.g. “I like the clubs because they rock” is considered neutral, and the verb “rock” is not correctly seen here as an expression of a very positive opinion). A more sophisticated sentiment analysis tool would thus be desirable.

It has also been previously mentioned that the chatbot detects reasoning through the use of connectors, which leaves out some **cohesive devices**. Again, ideally, the classifier should be able to identify subtler devices without losing precision, though the current, limited classifier encourages the use of explicit connectors, which is certainly useful and reasonable at an intermediate level.

Last but definitely not least, as this was only a demo no time was invested in developing **variations of the output**. Even though the student would always know that they are talking to a chatbot, developing a set of equivalent outputs for each intent for the tool to choose from would add variety to the discussion and make it more natural and pleasant.

Concluding remarks

As can be seen in the previous section, this tool has numerous limitations. Still, even in this limited form it can lead a very simple discussion and give feedback that could be useful to a student. Hopefully, then, this demo can illustrate the potential of chatbots for dialogic teaching – with more data and yet more work, a less limited tool could be developed and hopefully successfully tested with its target users.

References

- Cambridge Assessment. (2020, May). *B2 First Preparation*. Retrieved from Cambridge Assessment English: <https://www.cambridgeenglish.org/exams-and-tests/first/preparation/>
- Chopra, S., Gianforte, R., & Scholar, J. (2016). Meet Percy: The CS 221 Teaching Assistant Chatbot. *ACM Transactions on Graphics*, 1-8.
- Jia, J. (2003). The study of the application of a keyword-based chatbot system on the teaching of foreign languages. *ArXiv*, 1-11.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Mercer, N., Warwick, P., & Ahmed, A. (2016). An oracy assessment toolkit: Linking research and development in the assessment of students' spoken language skills at age 11-12. *Learning and Instruction*, 51-60.
- Oxford Royale Academy. (2020, May). *40 Useful Words and Phrases for Top-Notch Essays*. Retrieved from <https://www.oxford-royale.com/articles/words-phrases-good-essays/#ald=9210c857-94ab-43e1-80e8-d51e3b63c006>
- Raj, S. (2019). *Building Chatbots with Python*. Bangalore: Apress.
- TextBlob. (2020, May). *Sentiment Analysis*. Retrieved from TextBlob Quickstart: <https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis>

Appendix: Demonstration of the output

***Note:** The format has been edited for better visualization in this report. Also, even though the input given to the chatbot is not long, this test was done calmly while revising notes on the design of the chatbot, so the discussion time surpassed 7 minutes.

Hello! This is a tool to help you practice discussing in English.

You will be given instructions and then your robot partner will be able to discuss with you.

Press ENTER to continue

Imagine that a town wants to attract more tourists. I will give you some ideas that the Town Council is considering, and you and your robot partner will have to discuss which of those ideas you think are good and why.

These are the ideas:

Building a large nightclub

Putting up security cameras

Having more shops

Providing parks

Building holiday flats

Press ENTER to continue

Hi! I'm your robot partner. You can start saying what you think about one of the ideas

I think that the idea of the clubs is very good, because tourists travel looking for a good time, and nightclubs are a popular option, in my opinion

I think you make a good point

We haven't looked at the topic of cameras

What do you think about that topic?

With regard to the cameras, I think that's a horrible idea, for the reason that they would be spying on people and the citizens would be scared or angry

I think you make a good point

We haven't looked at the topic of parks

What do you think about that topic?

I think the park is a good idea

Uhum, why do you think that?

I believe that parks would be good to attract tourism because, firstly, people like them and can go for walks or picnics, and also because it improves the air of the city

I think you make a good point

We haven't looked at the topic of flats

What do you think about that topic?

I don't like the idea of the flats. It could be bad for the people of the city, due to the fact that their rent could become more expensive

I think you make a good point

We haven't looked at the topic of shops

What do you think about that topic?

I think that shops are ok, it's a good idea. Shops could attract some tourists and also, they are good for the economy because the tourists would spend money in the city

I think you make a good point

Great, it seems that we've looked at all the options.

So, what would you say are the ideas that we liked the most?

I think the clubs were good

Mmm, I think that your conclusion is contradicting our discussion...and you forgot some ideas we thought were good! Try again :)

Think again, what would you say are the ideas that we liked the most?

I like the clubs, for people who want to party, the parks for the environment and picnics, and the shops to bring money

That's right! Good discussion :) I'll give you some feedback now

— ■ ■ ■ ■ ■ YOUR FEEDBACK ■ ■ ■ ■ ■ —

— ■ ■ Time Management ■ ■ —

The discussion took 467.1857213973999 seconds

It's good that you are able to carry a long discussion, but remember that in the FCE exam you have four minutes to discuss all the ideas! Practice summarizing a bit

— ■ ■ Use of Connectors ■ ■ —

I'll give you some feedback on how you linked your ideas

You've been repeating some connectors:

('because', 4)

Using connectors is good, but try to add some variety!

I've made some calculations and I see that you've used 11 connectors.

Good job, those are quite a few! It seems that you know how to structure your ideas, but have a look at some advanced connectors that you might find useful:

.....★.....

LIST OF CONNECTORS THAT YOU MIGHT WANT TO CHECK OUT

.....★.....

EXPRESSING A PURPOSE

in order to
to this/that end
for the purpose of

PARAPHRASING IDEAS

in other words
to put it another way
to put it simply
in other words
that is to say

ADDING IDEAS TO YOUR ARGUMENT

Moreover
Furthermore
What is more
Another key thing to remember
In addition to that
Besides
In fact
Actually
In reality
With respect/regard to
concerning/regarding
As for

COMPARING OR EQUATING

Likewise
Similarly
In the same way
As well as
Not only
Coupled with
Together with

ORDERING

First(ly)
Second(ly)
Third(ly)
Next
To begin with
In the first place

CONTRASTING

However
(Al)though
Nonetheless
Notwithstanding
Despite
In spite of
Having said that
In comparison
That being said
Yet
While
On the one hand... On the other hand
Whereas

CONSIDERING IDEAS

With this in mind
In view of
In light of
Considering
All things considered
At the end of the day
Ultimately
Above all

GIVING EXAMPLES/ESPECIFYING

For instance
For example
To illustrate
As illustration
Namely
Especially/particularly
In particular
Such as

CONSIDERING OTHER PERSPECTIVES

Some (people) might say
Those who disagree
I understand their point of
view/perspective
They have a point
They make a valid point
Another way of seeing this
Quoting
To quote

GIVING REASONS

Therefore
So
As a result/consequence
Due to (the fact that)
Because of
For this reason
This suggests that
Logically
Of course
Naturally

CONCLUDING/SUMMARIZING

In conclusion
To sum up
Finally
Lastly
To conclude
On the whole
In brief

It's been a pleasure talking to you! (* ^ ▽ ^)／