Question Avoidance Study

Polina Gusenkova
University of Potsdam
polinagpa@gmail.com

Karina Hensel
University of Potsdam
henselkarina@gmail.com

Iana Palacheva
University of Potsdam
palachevayana@gmail.com

Abstract

The unwillingness to answer a question is a common situation in everyday conversations. In this paper, we focus on the phenomenon of question avoidance. We pursue two goals: the first one is to create a manually annotated dataset with question avoidance examples, the second is to identify some linguistic patterns which are specific to avoidance responses. As a result, we developed a novel classification of question avoidance situations (including two avoidance strategies), constructed a custom dataset using question-response pairs from datasets of different domains, annotated it manually, and conducted several experiments aimed at discovering underlying lexico-syntactic features of question avoidance. Some differences between the two main avoidance strategies were found as a result. Finally, we transformed our annotated dataset into an easily-processible corpus and augmented it with our findings.

1 Introduction

In the social sciences and psychology, the unwillingness to answer a question has been a widely discussed topic. Also in politics not providing an exact answer to a question is a common rhetorical strategy. There are various studies (e.g. (Clementson, 2018), (Bull, 2003)) which examine general methods and patterns of answer avoidance and provide a theoretical basis for the numerous practical guides which can be found on the internet.

Therefore, it is quite surprising that question avoidance does not seem to be a field of current research in linguistics and computational linguistics. There have been some recent studies that have developed a classification system for answer types (Lupkowski and Ginzburg, 2013) and tested it on corpora in multiple languages (Ginzburg et al., 2019). However, none of them addressed the topic of avoidance.

In this paper, we focus on responses that avoid answering a question. We pursue two goals: the first one is to create a manually annotated dataset with question avoidance examples, and the second is to identify some linguistic patterns which are specific to avoidance responses.

This line of research could potentially be useful for improving dialogue systems. The extracted linguistic indicators could be used by an intelligent system to detect whether a user is unwilling to answer its questions and adapt its behavior strategy accordingly. Another possible application might be an automated trustworthiness level detection: the tendency to avoid a question (i.e. during a job interview) can indicate that a speaker is trying to confuse a questioner or hide some vital information.

With this project, we would like to contribute to a relatively unexplored but yet highly interesting area in (computational) linguistics by providing some basis for further research. On the one hand, we hope to find linguistic indicators for answer avoidance which can be used as features by existing computational applications. The second aim is to create a new corpus of question-response pairs annotated with respect to their level of avoidance.

This paper is structured as follows: Section 2 describes the general approach taken in this project. Section 3 presents details on the data collection process, annotation guidelines and the annotation itself. Section 4 covers the experiments aimed at exploring the linguistic features of the data. Afterward, the results and the new corpus are described in Section 5. Finally, we conclude with some final remarks on open research questions that remain for the future in Section 6. Appendix A reports on our collaboration experience.

Overall, our contributions can be summarized as follows:

1. We developed a novel classification of question avoidance situations

2. We constructed a custom dataset using question-response pairs from datasets of different domains and manually annotated it

3. We conducted several experiments aimed at discovering underlying lexico-syntactic features of question avoidance

4. We transformed our annotated dataset into an easily-processible corpus and augmented it with our findings.

2 Approach

For our project, we have developed an iterative approach. It consists of two iterations: the first iteration is focused on data collection and its annotation when the second iteration is dedicated to exploring the data and finding linguistic indicators of question avoidance. The first iteration is outlined in Section 3: we explain how we constructed the question-response dataset and describe the developed annotation guidelines and the annotation procedure. Section 4 covers the second iteration: we present the experiments aimed at uncovering some underlying linguistic patterns of question avoidance.

3 Data Collection and Processing

3.1 Data

Since one of our goals was to analyze utterances from different domains, we decided to use data from Cornell Conversation Analysis Toolkit - ConvoKit (Chang et al., 2020). The toolkit contains tools to extract conversational features and analyze social phenomena in conversations. The corpora from the toolkit are similarly structured, which allowed us to sample question-response pairs from different domains in a similar format.

We chose three corpora from ConvoKit relying on our intuition and the corpora structures. The criteria were the following: domain relevancy, whether the structure allows us to extract question-response pairs, whether there is additional metadata that can be relevant to the task, and the number of utterances we could use (Figure 1).

Domains: In line with our intuition and common sense, we picked out the corpora where we would expect to find cases of avoidance. Among them were parliament question periods (Parliament Question Time Corpus[1]), court proceedings (Supreme Court Corpus[2]), press-conferences (Tennis Interviews[3]), phone conversations (Switchboard Dialog Act Corpus[4]) and discussion threads on Reddit (Coarse Discourse Sequence Corpus[5] and Wikipedia (Conversations Gone Awry Dataset[6]).

Structure and additional metadata: According to the corpora descriptions, only three corpora contain labels for questions and answers: Parliament Question Time Corpus, Tennis Interviews, and Coarse Discourse Sequence Corpus. We found the three domains different enough to match our goals: political discussions (the obvious domain for question avoidance research), sport interviews (more personal questions) and social media interaction (informal environment).

The Conversation Gone Awry Dataset is also interesting because it contains an additional feature on a conversational level which indicates whether the comment contains a personal attack. However, we did not use this corpus because the number of utterances with a personal attack is relatively small: 1,493 utterances out of 30,021 in the corpus.

Balancing the data: From each corpus, we took a random sample of question-response pairs to create a corpus of 500 pairs to annotate manually. To balance the data, we sampled 200 pairs from the Parliament Question Time Corpus (216,893 utterances) and 150 from each the Tennis Interviews dataset (81,974 utterances) and the Coarse Discourse Sequence Corpus (39,749 utterances). This decision was partially based on our

[1]https://convokit.cornell.edu/documentation/parliament.html
[2]https://convokit.cornell.edu/documentation/supreme.html
[3]https://convokit.cornell.edu/documentation/tennis.html
[4]https://convokit.cornell.edu/documentation/switchboard.html
[5]https://convokit.cornell.edu/documentation/coarseDiscourse.html
[6]https://convokit.cornell.edu/documentation/awry.html

Dataset	Domain	QA pairs	Relevant features	Num of utt.	Num of utt. in dataset
Parliament Question Time Corpus	Parliamentary question periods		Q, A, QA-pair	216893	200
Tennis Interviews	Press-conferences		Q, A, QA-pair	81974	150
Coarse Discourse Corpus	Reddit conversations		majority_type: Q, A, ...	39749	150
Switchboard Dialog Act Corpus	Phone conversations		tag: speech act	0 (bad format)	
Conversations Gone Awry Dataset	Personal attacks: Wiki, Reddit		comment_has_personal_attack	1493	
Supreme Court Corpus	Court oral arguments				

Figure 1: Summary of relevant ConvoKit datasets

expectations: it seemed more likely to find avoidance cases in the Parliament proceedings.

3.2 Guidelines

In parallel with data collection, we were developing annotation guidelines. This document served two main purposes: firstly, it describes the annotation scheme, secondly, it comprises a simple classification of questions and responses (with examples) with respect to the question avoidance phenomenon.

Theoretical background: As the first step, we did theoretical research on the topic of question avoidance. By doing that, we were trying to give some finite definition to the question avoidance phenomenon, to determine conversational situations when the avoidance can happen more likely and to compile a list of avoidance techniques.

Firstly, it is worth noting that this phenomenon doesn't have an established name. One can look for the related information using the following keywords: *question avoidance, question evasion, question dodging, equivocation, obfuscation*. However, according to Judge (Judge, 2006), avoidance is "orienting the dialogue to other topics, or simply by avoidance of situations in which questions can be asked", while evasion is a "dishonest misrepresentation or dissimulation".

Based on different sources, we agreed on the following list of situations which can potentially trigger question avoidance:

- An answerer does not know the answer and does not want to show it.
- Interrogation, debate: an answerer wants to avoid a direct response to an accusatory question.
- The questioner asks an uncomfortable personal question.

Peter Bull (Bull, 2003) presented a comprehensive list of question evasion techniques. Below is the generalized version of this list with examples:

- Ignoring the question (continuing the conversation like the question wasn't asked)
- Acknowledging the question without answering it (responds like 'Mhm', 'Good question!')
- Echoing the question back to the questioner (for example, saying "you tell me")
- Attacking the question
- Attacking the questioner
- Being unwilling to answer (responds like "I don't want to talk about it", "I can't speak for someone else")
- Deferring answer("It is not possible to answer the question for the time being")
- Pleading ignorance
- Joking

Overall, the described types of avoidance fall into two high-level strategies: attacking and defending. It reminds of the well-known fight-or-flight response - "a physiological reaction that occurs in response to a perceived harmful event, attack, or threat to survival" (Walter, 1932). Inspired by this parallel, we adapted this idea in our annotation scheme as *Fight* and *Flight* avoidance strategies.

Annotation Scheme: The annotation scheme consists of three sequential levels:

- Level 1: (information-seeking) question vs. non-question

  Despite of the fact we have collected the question-response pairs with the corresponding labels for questions and answers, we could not be sure that all the pairs consisted of actual questions and answers in a linguistic sense. Some data could be corrupted, some pairs could be assessed as related to each other only in a broader context, some utterances could be rhetorical. In the guidelines, we suggest to label all corrupted utterances, simple statements, rhetorical questions and not-making-sense without context queries as non-questions.

- Level 2: signs of avoidance vs. no signs of avoidance

  Signs of avoidance are described in more detail in the next point. We suppose that the response does not express avoidance if it is: 1) a direct answer, 2) an indirect yet implicit and inferable answer, 3) a clarification request (depending on context), 4) ”I don't know” response.

  We suggest to rate each question-response pair on a 5-point avoidance scale based on certainty: 0 - no avoidance, 1 - rather not avoidance, 2 - uncertain, 3 - rather an avoidance, 4 - definitely avoidance. Also, we explicitly advise using score ”2 (uncertain)” only when it is necessary.

- Level 3: *Fight* strategy vs. *Flight* strategy

  Such response strategies as attacking the question or the questioner, answering with a vague or confrontational question, or reflecting the original question we are referring to *Fight* strategy. In contrast, *Flight* strategy is associated with the following response types: 1) ignoring the question or acknowledging it without answering, 2) declining to answer, 3) changing the topic, 4) respond with a joke. For the extended classification with examples, refer to Appendix B.

We suggest to assess paired utterances by answering the following questions representing the described levels:

- Is the first utterance (u1) a question? If yes,

- What is the level of avoidance in the second utterance (0 - no avoidance, 4 - definitely avoidance)? If the level is greater than 2,

- What is the avoidance strategy (*Fight* or *Flight*)?

For an annotation setting example, refer to Appendix C.

3.3 Annotation

The annotation was done by the authors (Master's students of the University of Potsdam), two of them are native Russian speakers and one is a native German speaker. Each annotator rated all 500 question-response pairs according to the guidelines. This was done for cross-evaluation and bias reduction.

After the annotation was completed, all the ratings were added to our new corpus and additionally averaged. If the first utterance was marked as a non-question by at least two annotators, the corresponding pair was deleted from the final set of samples.

The final avoidance rating for each question-response pair was obtained by averaging over the scores given by all annotators. We consider the following cases: if the average rating is below 2, the response is treated as non-avoidance, for a score above 2 - as avoidance. Cases with an average rating of 2 are treated as ‘undetermined’ and were treated separately as a non-informative part of the corpus.

For the *Fight* vs. *Flight* classifications, the majority label was chosen as the correct class. Unclear cases (one *Fight*, one *Flight*, one *undetermined* or not labeled) were reviewed separately and also treated as non-informative.

With the resulting set of annotated and classified question-response pairs, we eventually compiled the final corpus. Table 1 presents some more fine-grained statistics about the text collection.

It can be seen that there are some issues with the data. For instance, the data set is not balanced. The proportion of samples from the Parliament corpus is much larger than those from the other two sources. Moreover, the corpus does not comprise many examples in general (173 avoidance cases).

Finally, it must be noted that, although the entire corpus is in English, none of the annotators is a native English speaker. It might be interesting to have the corpus re-annotated by native English speakers who possibly have a different intuition about the ranking.

	PQTC	**TI**	**CDC**	**Total**
Non avoidance	36	108	84	228
Avoidance	142	21	10	173
Fight	16	1	2	19
Flight	105	18	8	131

Table 1: Corpus statistics

4 Experiments

In this section, we describe the conducted experiments. In these experiments, we were looking for correlations between avoidance score or strategies and scores such as negation (Section 4.1), number of queries in answers (Section 4.2), sentiment (Section 4.3), politeness (Section 4.4), and syntactic similarity (Section 4.5). In most experiments, we measured the difference in scores between avoidance and non-avoidance pairs and between the two avoidance strategies (*fight* and *flight*). Additionally, the question-response pairs from the Parliament Corpus were checked for a correlation between who asks who We also looked for syntactic patterns and co-occurrences in answers (Section 4.6).

4.1 Negation Score

The negation score was calculated using a pre-trained model for predicting syntactic dependencies from spaCy[7] (a Python-based NLP tool). The dependency parsing contains labels for negation.

We counted the number of occurrences of negation labels in the whole utterance as well as in the first two sentences. The latter was done under a hypothesis that if a question is avoided, the avoidance markers appear at the beginning of the answer. This hypothesis was proposed during the annotation phase.

No correlation was found between the average number of negation and avoidance rate (Figure 2). The low point at the avoidance rate 1.5 can be explained by a small number of examples (only 3 utterances with this score) in the data.

[7]https://spacy.io/

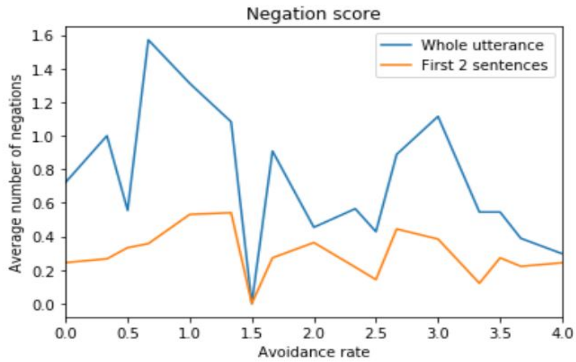


Figure 2: Negation score over the avoidance rate for the whole response (blue) and the first two sentences of the response (orange)

However, there is a possible correlation between the number of negations and the avoidance strategy (Figure 3). Responses labeled with the *fight* strategy tend to contain more negation on average than the responses labeled with *flight*. The low point for the 'fight' type at avoidance rate 3.5 can once again be explained by a small number of examples (only one utterance) in the data. This result corresponds to our intuition towards the bigger level of aggression and, therefore, negations in the case of the *fight* strategy.

4.2 Queries in Answers

This experiment was based on the idea that a query, or a question, can be answered by another query. For example, (Łupkowski and Ginzburg, 2013) presents a taxonomy for such responses which includes questions aimed at avoiding the initial question and questions ignoring it.

The number of questions appearing in a response was counted in the simplest possible way: by the number of question marks. As a result, no correlation was found between the average number of question marks and the avoidance rate (Figure 4). Since there are only 19 utterances labeled with *fight*, too few question marks were found for this part of the dataset to compare the two strategies.

4.3 Sentiment Score

To calculate the sentiment score, we used the pre-trained VADER Sentiment Lexicon model (Gilbert and Hutto, 2014) available in the NLTK[8] package. VADER (Valence Aware Dictionary for

[8]https://www.nltk.org/

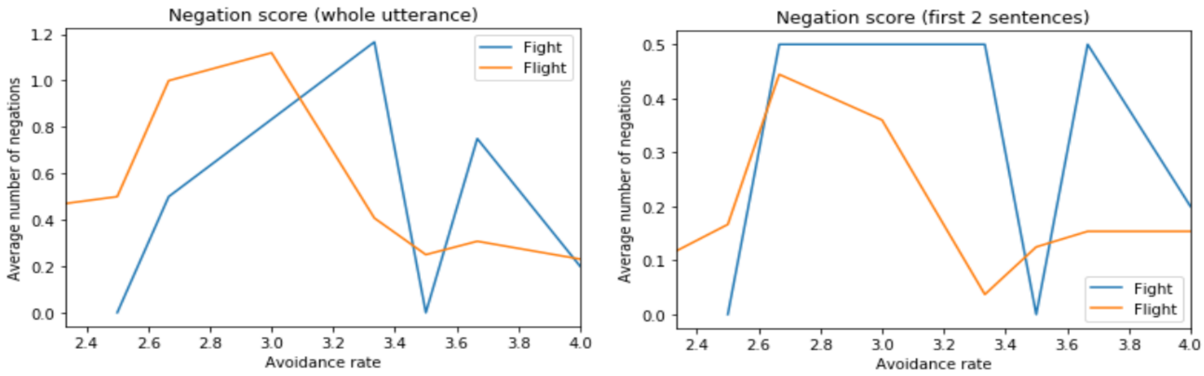


Figure 3: Negation score for *fight* (blue) and *flight* (orange) strategies over the avoidance rate for the whole response (left) and the first two sentences of the response (right)

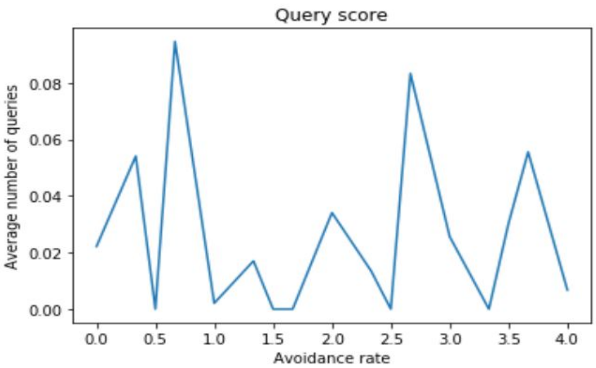


Figure 4: Number of questions (question marks) over the avoidance rate.

Sentiment Reasoning) takes into account emotional polarity and intensity. It predicts sentiment scores in relation to four aspects: negative, neutral, positive, and compound (the three previous scores normalized). The scores lie in the interval from -1 (negative) to 1 (positive).

The pre-trained VADER model is rule-based and is tuned to perform sentiment analysis in the context of social media. Our annotated dataset consists mostly of political discussions and interview questions. On the one hand, this is a ground for possible improvement. On the other hand, the results are still worthy of attention.

There is evidence of a positive correlation between the avoidance score and the sentiment score (Figure 5). This is a rather counter-intuitive result since we expected more negative scores for avoidance in comparison to non-avoidance. However, this could be explained by the fact that our dataset contains a lot of *flight* examples and just a few *fight* examples. This could mean that responses using the *flight* strategy are more positive and polite because the point of this strategy is to avoid conflict.

This hypothesis is supported by the results of the comparison of the average sentiment scores between *fight* and *flight* strategies (Figure 6). The utterances labeled as *fight* have more negative sentiment score than the utterances labeled as *flight*.

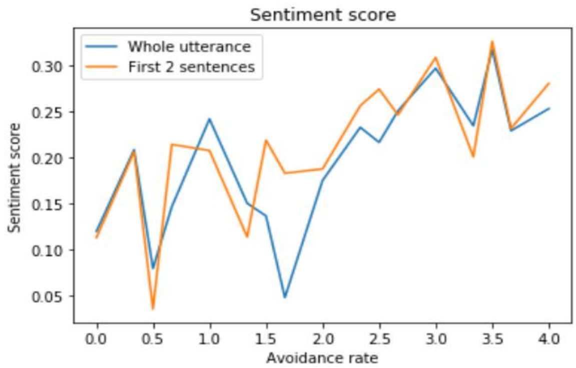


Figure 5: Sentiment score over the avoidance rate for the whole response (blue) and the first two sentences of the response (orange)

4.4 Politeness Score

The above results suggest that there could be some evidence of the correlation between the avoidance score and politeness. The politeness score was calculated using the Politeness Strategies (Danescu-Niculescu-Mizil et al., 2013) from ConvoKit. A ConvoKit classifier was trained on Stanford Politeness Corpus[9] (based on Wikipedia), which is annotated with politeness scores. Then the model predicted the politeness scores for our dataset.

Once again, there is no correlation found between avoidance and non-avoidance cases (Fig-

[9] https://convokit.cornell.edu/documentation/wiki_politeness.html

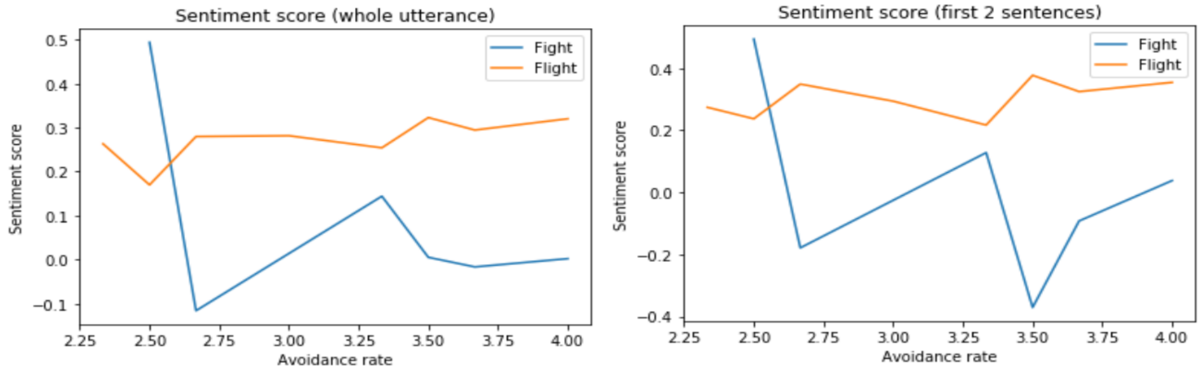


Figure 6: Sentiment score for *fight* (blue) and *flight* (orange) strategies over the avoidance rate for the whole response (left) and the first two sentences of the response (right)

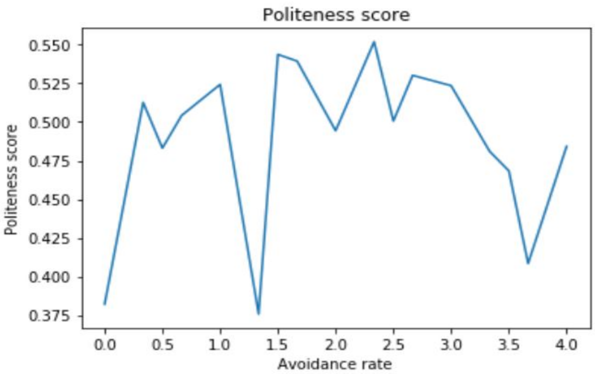


Figure 7: Politeness score over the avoidance rate for the responses.

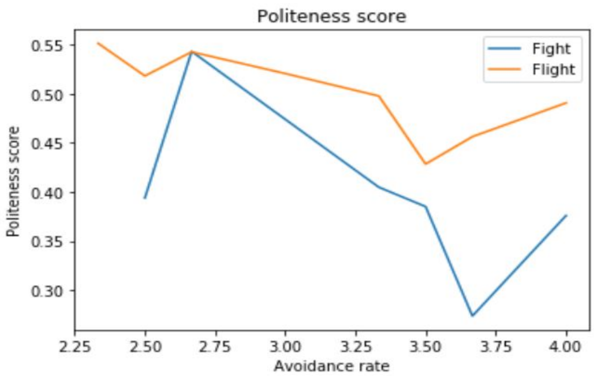


Figure 8: Politeness score for *fight* (blue) and *flight* (orange) strategies over the avoidance rate for the responses.

ure 7). However, as expected, the utterances labeled *fight* are less polite than the utterances labeled *flight* (Figure 8). This trend is consistent for all avoidance scores and is in agreement with the previous results regarding sentiment score.

4.5 Syntactic Similarity

A widely used metric to quantify the closeness of two strings is the Levenshtein distance (Levenshtein, 1966). It measures syntactic similarity in terms of the number of character insertions, deletions, and substitutions which are necessary to convert one phrase to exactly match another. In this project, we use the implementation from the NLTK package[10] for the measurements.

In this experiment, we calculated the edit distance between the utterances in a question-response pair to assess whether the distance correlates in any way to avoidance. To this end, we compared the distances between avoidance and non-avoidance pairs and between the two avoidance strategies.

The utterances were preprocessed beforehand. Stop words were removed to eliminate the most common and repetitive patterns. For that, we used the list of stop words for the English language from NLTK.

We also hypothesized that a question is more often posed at the end of an utterance, and whenever it is answered, the answer is found in the first part of the response. Therefore, only the last two phrases of a question and the first two phrases of a response were included in the calculation. This was necessary because significant differences in utterances' length would incorrectly increase the distance.

The non-avoidance pairs have a bigger similarity score than avoidance pairs, even though the percentile intervals and measured distances are very large (Table 2). The numbers do not demon-

[10] `https://www.nltk.org/api/nltk.metrics.html`

strate a significant difference but generally correspond to our intuition.

For pairs containing avoidance, *fight* pairs have a smaller similarity scores than *flight* pairs. It can be explained by the prevalence of entries from Parliament Question Time Corpus among all marked as avoidance: politicians often use a 'slight topic change' strategy to avoid unwanted questions, which means that they use the same lexicon and stick to the topic from the question, however they do not actually *answer* the question.

Class	**Distance**	**10% percentile**	**90% percentile**
Non avoidance	314.91	138.4	491.8
Avoidance	372.79	127.7	466.8
Fight	375.96	152.4	502.4
Flight	303.16	127.7	466.8

Table 2: Edit distances

4.6 Syntactic Patterns

One research goal of this project is to investigate whether there are certain phrases and syntactic constructions that are either typical for avoidance or non-avoidance responses. If this is indeed the case these formulations could be used as features for the prediction of response types.

Just like for the calculation of the string distance in Section 4.5 the responses were preprocessed by removing stop words to obtain more generalized and less domain-specific results. Afterwards bigrams and trigrams were extracted by using the collocations module[11] provided by NLTK (Bird et al., 2009).

However, the extracted n-grams still reflect the unbalanced sample selection of the data set. Figures 9 and 10 present the ten most frequent bigrams in avoidance and non-avoidance responses. It can be seen that instances from the Parliament corpus outnumber question-response pairs from the other two corpora. Hence, it is not possible to find any characteristic phrasings for avoidance or non-avoidance respectively by using automated methods for n-gram extraction on the data at hand. Future studies might consider obtaining n-grams from more balanced or less domain-specific corpora.

[11] http://www.nltk.org/api/nltk.html?highlight=freqdist

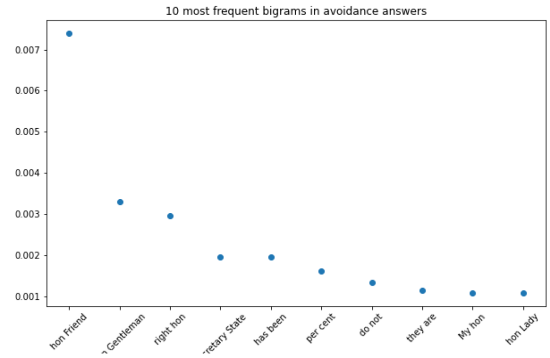


Figure 9: Avoidance bigrams

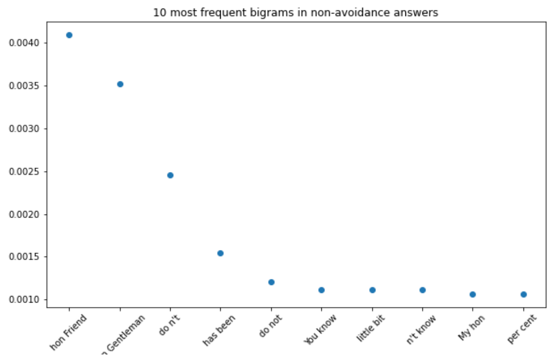


Figure 10: Non-avoidance bigrams

5 Results

To conclude the research, we collected, annotated, and experimented on data in a form of question-response pairs, aiming towards the detection of avoidance in such data. Overall, the results of this research include the following:

1. Data collection: the question-response pairs were collected from three different domains.

2. A novel classification of question avoidance situations: *fight* and *flight* strategies were introduced. As a result, we have explicit annotation guidelines.

3. A manually annotated custom dataset of question-response pairs from different domains.

4. Experiments showed that responses with high avoidance scores have a higher sentiment score. Other measurements did not demonstrate any tendencies in response to the avoidance score.

5. Experiments on *fight* and *flight* strategies provide evidence of distinction between these two types. For instance, fight responses generally contain more negations, are less positive in terms of sentiment, and less polite.

6. Experiments on n-grams demonstrated the fact that the common patterns are domain-specific due to the unbalanced corpora.

7. The annotated dataset was transformed into an easily-processible corpus and augmented with our findings.

All the data and code are available on our GitHub repository[12].

5.0.1 New corpus

To summarize the results, we used ConvoKit's API to create a new custom question avoidance corpus. Each corpus contains *utterances* which are grouped into *conversations*. We augment the question-response pairs with our annotation labels as well as with the linguistic features described in Section 4.

Utterances are both questions and responses, the utterance-level information includes:

1. Utterance text

2. ID from the original corpus (Parliament Question Time Corpus, Tennis Interviews or Coarse Discourse Corpus)

3. ID of the question utterance (only for responses)

4. Negation scores for the whole utterance and two first sentences of the utterance (only for responses)

5. Sentiment scores for the whole utterance and two first sentences of the utterance (only for responses)

6. Politeness scores (only for responses)

Conversations are pairs of related questions and responses linked by the ID of the question utterance, the conversation-level information includes:

1. IDs of question and response utterances

2. Annotation data: rates from all three annotators, along with the averaged avoidance score and averaged strategy

3. Levenstein distance between whole utterances and between the last two sentences of the question and the first two sentences of the response

[12] https://github.com/YanaPalacheva/avoidance_study

6 Future work

We hope that this project will inspire other researchers to explore the topic of question avoidance. Below we suggest several ideas for future research:

1. The more data the better! For such corpus-based research, the amount of data is the crucial factor. We suggest to follow the developed annotation guidelines to annotate more data from different domains: we believe it is the only way to discover some common patterns.

2. It can be very interesting to get this dataset annotated by native English speakers and to compare the results: in such social-linguistic cases as question avoidance, linguistic intuition plays a huge part.

3. Use a more domain-related sentiment classifier: the one we use in the current project is trained on social media.

In our opinion, the most interesting research idea is to implement our approach on a dataset of 1,256 questions asked to various Prime Ministers labeled by political scientists (Bates et al., 2012). They label questions as *answered, not answered, deferred*. We have checked if we have any intersection with this dataset, but it was not the case, unfortunately, so we could not compare our annotation to these labels. The creators of Parliament Question Time Corpus we used utilized this dataset in their work on rhetorical roles of question (Justine Zhang, 2017) as well.

References

Stephen Bates, Peter Kerr, Christopher Byrne, and Liam Stanley. 2012. Questions to the prime minister: A comparative study of pmqs from thatcher to cameron. *Parliamentary Affairs*, 67:253–280.

Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

P. Bull. 2003. *The Microanalysis of Political Communication: Claptrap and Ambiguity*. Routledge research international series in social psychology. Routledge.

Jonathan P Chang, Caleb Chiam, Liye Fu, Andrew Z Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. Convokit: A toolkit for the analysis of conversations. *arXiv preprint arXiv:2005.04246*.

David Clementson. 2018. Deceptively dodging questions: A theoretical note on issues of perception and detection. *Discourse Communication*, 12:478–496.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*.

CHE Gilbert and Erric Hutto. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*, volume 81, page 82.

Jonathan Ginzburg, Zulipiye Yusupujiang, Chuyuan Li, Kexin Ren, and Paweł Łupkowski. 2019. Characterizing the response space of questions: a corpus study for english and polish. pages 320–330.

Anthony Judge. 2006. Question avoidance, evasion, aversion and phobia: Why we are unable to escape from traps.

Cristian Danescu-Niculescu-Mizil Justine Zhang, Arthur Spirling. 2017. Asking too much? the rhetorical role of questions in political discourse. In *Proceedings of EMNLP*.

Vladimir I. Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics Doklady*, volume 10, pages 707–710.

Pawel Lupkowski and Jonathan Ginzburg. 2013. A corpus-based taxonomy of question responses. In *IWCS*.

Paweł Łupkowski and Jonathan Ginzburg. 2013. A corpus-based taxonomy of question responses.

Cannon Walter. 1932. The wisdom of the body.

A Teamwork

During the research process, we were determining milestones and for each interval, we distributed the tasks evenly among the three of us and did not have any problems to organize our work and to resolve misunderstandings.

As for the distribution of the work, Iana Palacheva was mainly concerned with developing the annotation guidelines, creation and structuring of the final corpus and maintaining the code. Polina Gusenkova was in charge of the data (she did research on data, compiled the dataset for annotation, merged the results) and moreover she conducted several experiments on the data. Karina Hensel did research on the theoretical and linguistic basis for avoidance annotation and also analyzed the annotated corpus with respect to syntactic structure and linguistic features, such as n-grams.

All group members worked jointly on the avoidance rating and assessed the final results together.

We mainly organized our project online. Although we only met twice in person, we had regular Skype calls and maintained active communication via Slack. To make the collaboration comfortable, we used Google Drive[13] for data and documents and GitHub[14] for the code.

In general, our experience of working as a team was mainly positive and we did not have any major problems in spite of being in different places physically.

B Avoidance strategies with Examples

See Figure 11.

C Annotation Setting

See Figure 12.

[13] https://drive.google.com/

[14] https://github.com/

u2 is ‘fight’ avoidance

- Attacking the question:

u1: Who you think killed Mary?
u2: What kind of question is that?!

u1: When is the wedding?
u2: This is an inappropriate thing to ask!

- Attacking the questioner:

u1: Tell me about your deepest fears.
u2: And who are you to ask such questions?

u1: Tell me about your deepest fears.
u2: We are not that close.

- Respond with a vague question:

u1: What has happened here?
u2: How would I know?

It can be the case that the respondee is truly not aware of what has happened, but the form of saying that can indicate evasion

- Reflecting the question:

u1: What has happened here?
u2: You tell me!

u1: What has happened here?
u2: What do you think?

u1: Were you up late again?
u2: And what about you?

- Respond with a confrontational question:

u1: Are you going to marry him?
u2: Are you so concerned about my marital status?

u2 is ‘flight’ avoidance

- Ignoring the question:

u1: What time is it?
u2: Vladimir Putin is the president of Russia.

u2 has no semantic connection to u1, so we assume that the question u1 was ignored

- Acknowledging the question without answering:

u1: What shall we eat for lunch?
u2: Good question!

u2 expresses that the respondee heard the question, but it's not an answer

- Circular reasoning:

u1: Why did your boss lose?
u2: He lost because he didn't get enough votes.

u2 just defines what ‘lose’ means in this context, but doesn't give any reasons.

- Declining to answer:

  - Unwilling to answer:

u1: What is your greatest fear?
u2: I don't want to talk about it.

  - Referring to another person:

u1: Tell me how your friend is today.
u2: I can't speak for someone else.

- Joking:

u1: How much money do you make?
u2: Not enough!

- Changing the topic:

  - Bridge response

u1: What is the capital of Great Britain?
u2: I can't tell you that for sure, but here's something I do know: the UK is a great country!

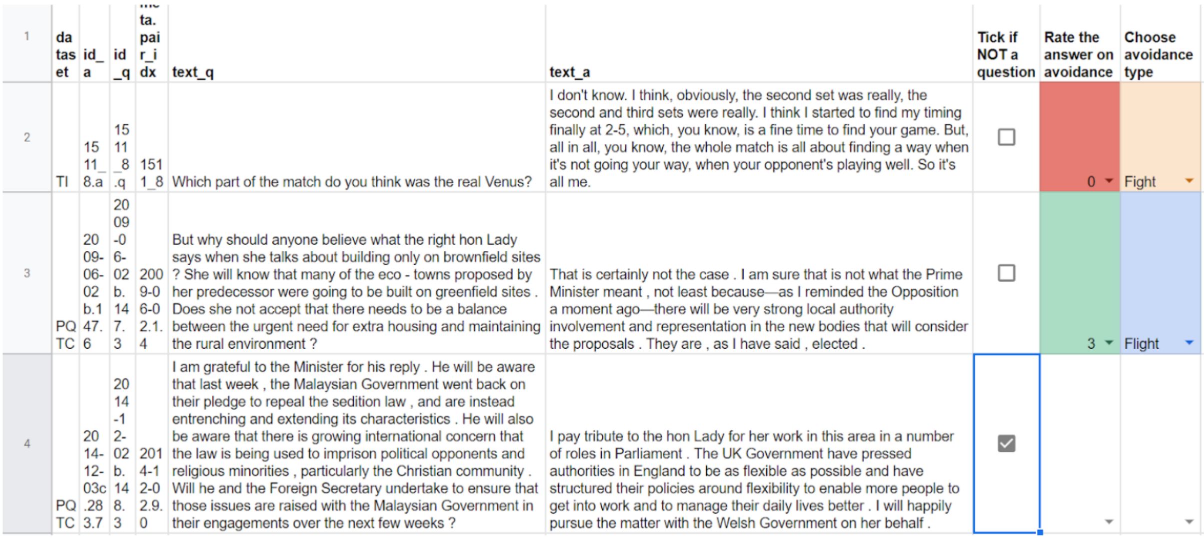
  - *Slight change*

u1: When will you pay off the loan?
u2: I'm looking for a job day and night!

  - *Forced change*

u1: Are you pregnant?!
u2: Look! Isn't it your favorite coffeeshop?

Figure 11: Classification of question avoidance situations with examples with respect to ’Fight’ and ’Flight’ strategies



1	dataset	id_a	id_q	meta.pair_idx	text_q	text_a	Tick if NOT a question	Rate the answer on avoidance	Choose avoidance type
2	TI	1511_8.a	1511_8.q	1511_8	Which part of the match do you think was the real Venus?	I don't know. I think, obviously, the second set was really, the second and third sets were really. I think I started to find my timing finally at 2-5, which, you know, is a fine time to find your game. But, all in all, you know, the whole match is all about finding a way when it's not going your way, when your opponent's playing well. So it's all me.	☐	0	Fight
3	PQTC	2009-06-02b.147.6	2009-06-02b.147.3	2009-06-02.1.4	But why should anyone believe what the right hon Lady says when she talks about building only on brownfield sites ? She will know that many of the eco - towns proposed by her predecessor were going to be built on greenfield sites . Does she not accept that there needs to be a balance between the urgent need for extra housing and maintaining the rural environment ?	That is certainly not the case . I am sure that is not what the Prime Minister meant , not least because—as I reminded the Opposition a moment ago—there will be very strong local authority involvement and representation in the new bodies that will consider the proposals . They are , as I have said , elected .	☐	3	Flight
4	PQTC	2014-12-03c.283.7	2014-12-02b.148.3	2014-12-02.9.0	I am grateful to the Minister for his reply . He will be aware that last week , the Malaysian Government went back on their pledge to repeal the sedition law , and are instead entrenching and extending its characteristics . He will also be aware that there is growing international concern that the law is being used to imprison political opponents and religious minorities , particularly the Christian community . Will he and the Foreign Secretary undertake to ensure that those issues are raised with the Malaysian Government in their engagements over the next few weeks ?	I pay tribute to the hon Lady for her work in this area in a number of roles in Parliament . The UK Government have pressed authorities in England to be as flexible as possible and have structured their policies around flexibility to enable more people to get into work and to manage their daily lives better . I will happily pursue the matter with the Welsh Government on her behalf .	☑		

Figure 12: Annotation setting