

Universität Leipzig
Faculty of Mathematics and Computer Science
Degree Programme Computer Science

How we Argue: A Study of Vocal Argument-Seeking Conversations

Master's Thesis

Simon Kleine

1. Referee: Jun.-Prof. Dr. Martin Potthast
2. Referee: Dr. Christian Kahmann

Submission date: January 29, 2023

Declaration

Unless otherwise indicated in the text or references, this thesis is entirely the product of my own scholarly work.

Hamburg, January 24, 2023

.....
Simon Kleine

Abstract

The goal of this research is to contribute to the understanding of vocal argument-seeking conversations, as we expect them to be part of future speech interfaces and only by understanding them can we evaluate the performance of such interfaces. To this end, we examine whether Kuhlthaus' principle of uncertainty for information seeking is applicable to this type of conversation, and look for patterns and salience peculiar to this type of conversation.

To find answers to these questions we conducted a study containing 10 argument-seeking conversations. Our participants chose a topic that allowed for different opinions, and after we had informed ourselves about the topic, they talked to us until they felt sufficiently informed.

The analysis of these conversations shows that Kuhlthaus' principle is not applicable to argument-seeking conversations. Moreover, we find that these conversations all start with a phase in which the participants' information needs have to be explored in more detail, as the participants' opening questions are not sufficient to explain these needs.

Contents

1	Introduction	1
2	Related Work	4
2.1	Conversation System Evaluation	4
2.2	Investigations and Expectations of Argument-Seeking Conversations	5
2.3	Principle of Uncertainty for Information Seeking	6
3	Conducting a Study to Investigate People’s Behavior in Argument-Seeking Conversations	8
3.1	Study setup	8
3.2	Study conduct	14
4	Investigating the Existence of the Principle of Uncertainty in Argument-Seeking Conversations	18
4.1	Realm of Thoughts	19
4.2	Realm of Actions	22
4.3	Realm of Feelings	25
5	Examination of an Argument-Seeking Conversation	31
5.1	Roles to be Assumed by the Interlocutors	31
5.2	About the Question Introducing the Conversation	33
5.3	Duration of Conversations	35
5.4	Speaking Parts in the Conversation	36
5.5	Conversation Affirming Utterances	37
6	Conclusion	39
A	Lickert Scale	41
B	Analysis of the conversations for subtopics	43
B.1	Visualizations	43

CONTENTS

B.2 Time spent on subtopic	54
C Analysis of the conversations for experienced feelings	58
Bibliography	60

Chapter 1

Introduction

Would it not be great to always have an expert at hand? An expert that you can ask any question and who will not tire of arguing with you. An expert who truly understands your reasoning, who can provide you with facts to support or refute your claims, and who you are comfortable talking to.

Due to ongoing research in the fields of voice interfaces and argument retrieval, such an expert could be realized in the near future. Modern technology giants such as Google with its Google Assistant, Apple with Siri and Amazon with Alexa are showing great interest in the field by introducing speech recognition systems to a wide audience as well as funding research on voice interfaces (Xu et al., 2022) (Chan et al., 2022). Though systems currently available are not able to engage in complex information seeking tasks, recent research will lead to more powerful systems that will be able to carry on conversations that are more than chit-chat (Trippas et al., 2019) (Ni et al., 2021).

The advancement of voice interfaces is still facing the challenge of system evaluation (Ni et al., 2021). This challenge can be further divided into evaluating the system-generated responses and evaluating the conversation itself (see Figure 1.1).

For evaluating the responses, there exist different metrics such as BLEU (Papineni et al., 2002) and BERTScore (Zhang et al., 2020). Evaluating the conversation itself can also be divided into the two parts “how to evaluate” and “what to evaluate” (Ni et al., 2021). If it’s an open-domain conversation, using crowdsourcing for human evaluation is the standard way to go, but it’s subjective. If it is a task-oriented conversation, one way to evaluate it would be to measure the completion rate and its cost (Ni et al., 2021). In an information seeking conversation, the task is to provide the user with the information they need, but it is not easy to automatically assess when this task has been accomplished.

To evaluate any system, it generally is good to know what the system

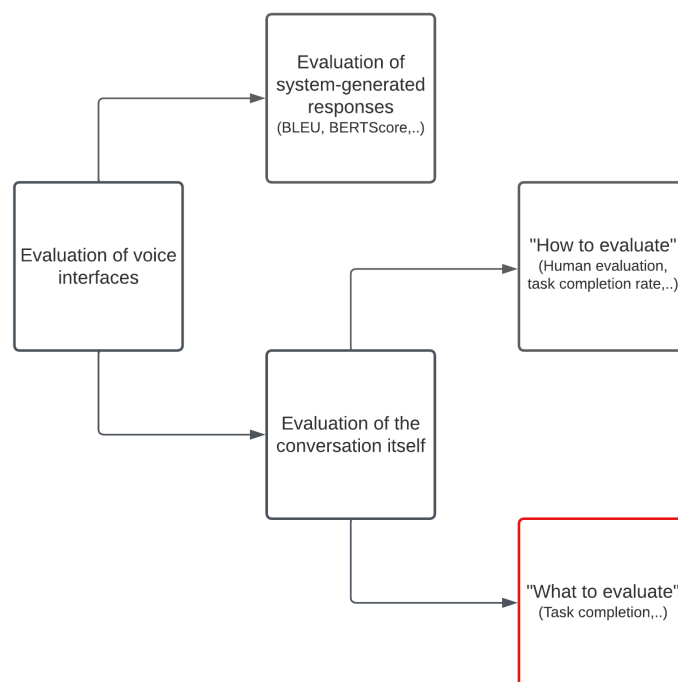


Figure 1.1: Evaluation of voice interfaces.

should look like and how it should behave. Kiesel et al. (2020) did a survey to gain insights in what people expect from voice-based and conversational argument search. They also pointed out the particular challenges of developing a conversational interface for argument search systems - the expert - and expanded our understanding of how such a system should behave (Kiesel et al., 2021).

Since the systems we are interested in should allow users to interact with them as if they were talking to an expert on a particular topic, we need to understand what determines this type of interaction in order to use this knowledge for system evaluation. Between 1988 and 1991 before these systems were within reach Kuhlthau studied the interaction between users and libraries culminating in her *Principle of Uncertainty for Information Seeking* (Kuhlthau, 1993) that describes the thoughts, actions and feelings one experiences in the information seeking process.

In the context of the need to understand a conversation between two interlocutors, one of whom has an information need and the other of whom has the desired information, in order to evaluate a future system for vocal argument retrieval, we derived the following two research questions:

RQ 1: Does the Principle of Uncertainty for Information Seeking proposed by Kuhlthau apply to an individual discussing different stances on a topic with an interlocutor they believe to have the answers to their questions?

RQ 2: What patterns and conspicuities emerge in a conversation in which an individual discusses different stances on a topic with an interlocutor they believe to have the answers to their questions?

To find an answer to these questions, we conducted ten conversations with ten different participants with a total duration of 4.5 hours. Our participants agreed that we would record the conversations and release the translated transcripts (the original conversations were conducted in German). The participants we recruited were asked to look for a topic in their everyday lives that allows for different points of view and on which they would like to form an informed opinion and exchange views with an informed discussion partner. In addition, participants should have no more than a basic knowledge of their chosen topic, otherwise the work required to play the informed interlocutor would have made the study unfeasible. Since the amount of work involved is still considerable, the sample size of the study had to be kept small, as a larger sample would have been beyond the scope of this paper.

This work is divided into the following sections: An overview of the field this thesis is contributing to (Chapter 2), the description of our study setup and how we intend to verify Kuhlthau's principle (Chapter 3), the investigation of research question one (Chapter 4), the investigation of research question two (Chapter 5) and a summary of our findings and contributions (Chapter 6).

Chapter 2

Related Work

The realization of a conversational system for argument-seeking conversations has its roots in two different research areas, argument mining and conversational systems.

The goal of argument mining is to automatically identify the argumentation of a document, i.e., all arguments involved in the argumentation process and the interactions between them (Mochales and Moens, 2011). Although argument mining is a relatively young field of research, it is experiencing rapid growth (Lawrence and Reed, 2019) and gained the interest of big tech companies like IBM (Khatib et al., 2021; Levy et al., 2018).

In the research field of natural language processing, conversational systems are one of the most widely studied topics and receive a lot of attention in industry and daily life (Ni et al., 2021). Industry giants such as Microsoft and Amazon are contributing to ongoing research by developing their own conversational systems (Zhou et al., 2020), contributing financially to the community, or enabling unique evaluation methods (Ram et al., 2018).

The convergence of these two technologies allows users to use a conversational system to find arguments on a particular topic and form an opinion on that topic. Even if such a system does not yet exist in this form, research has already begun to prepare the way for it (Aliannejadi et al., 2020; Kiesel et al., 2020; Spina et al., 2021).

2.1 Conversation System Evaluation

The evaluation of a conversational system is a key component to its success and an ongoing challenge (Ni et al., 2021; Smith et al., 2022; Yeh et al., 2021). The evaluation problem can be divided into two parts: the evaluation of a single system-generated response and the evaluation of the conversation as a whole (Ni et al., 2021). For both parts, human evaluation is still considered

the gold standard today, but it is not ideal either because it is subjective and human labor is expensive (Deriu et al., 2021; Ni et al., 2021; Yeh et al., 2021). In particular, for system-generated responses, there are numerous evaluation metrics such as GRADE (Huang et al., 2020a) and FlowScore (Li et al., 2021) for evaluation (Yeh et al., 2021). In one way or another, all of these metrics measure the extent to which the system is behaving the way we want it to. For evaluating the conversation as a whole, this measurement is not always so straightforward, as it depends on the goals of the conversation. If the conversation is about completing a task, the task completion rate and associated task completion cost is one possible metric for evaluation (Ni et al., 2021). There exists no clear metric for conversations in open domains (Ni et al., 2021). Although conversational argument search systems have a task at hand, e.g., to provide the user with arguments on a particular topic, it is still an open area and unclear when a conversation should be considered good.

2.2 Investigations and Expectations of Argument-Seeking Conversations

A conversation can be classified as good when system-generated responses are consistent in behavior and responses (Huang et al., 2020b). One way of adding consistency to a conversation is to assign some sort of personality to the system that responds (Li et al., 2016; Roller et al., 2020). It is also important that a system does not spam its users with unwanted messages or offend them (Roller et al., 2020).

Kiesel et al. (2020) explored what, why and how users expect to perform argument search including voice based argument search. Their observations provide implications on the design of argument search engines in general, as well as to voice based argument search systems.

Trippas et al. (2019) conducted an exploratory study to examine the exchange between an interlocutor seeking information and an interlocutor providing information. They had two actors playing these parts. The information seeker was given a backstory on an information need that he had to verbalize to the second actor that in turn could use a search engine to provide information. Ren et al. (2021) created a dataset of conversations in which two people have a written conversation, one playing the role of an information seeker and the other playing the role of a knowledgeable expert. They used a dataset containing search logs from a commercial search engine as a database from which the information seeker and the expert had to select the content of their messages in these conversations. The resulting conversations make up the dataset created by Ren et al.. Papenmeier et al. (2022) conducted another study of exchanges

between experts and people with an information need. They wanted to find out what strategies experts and customers use to find the right product. They therefore recorded consultation sessions between an expert and a customer on deciding which laptop to buy or what dinner to cook for a group of friends.

Kuhlthau (1993) developed a Principle of Uncertainty for Information Seeking that provides a theoretical view of people searching for information to deepen their understanding of a topic. To this end, she conducted numerous studies with library users (Kuhlthau, 1988a,b,c,d, 1990). The principle is described in more detail in chapter 2.3.

2.3 Principle of Uncertainty for Information Seeking

In 1993, Kuhlthau developed her *Principle of Uncertainty for Information Seeking*. Her goal was to provide a theoretical view of people using a library to search for information that would deepen their understanding of a topic. Such a view could then have been used to develop a new approach to teaching library skills to use information effectively (Kuhlthau, 1988b).

In her principle, Kuhlthau divides the information seeking process into six phases: Initiation, selection, exploration, formulation, collection and presentation. Furthermore her principle incorporates three realms of human experience, the affective (feelings), the cognitive (thoughts) and the physical (actions) within each stage (Kuhlthau, 1993), as shown in Figure 2.1.

The initiation phase includes the period of the information seeking process when the individual realizes their lack of information or understanding. The individual usually experiences uncertainty and apprehension during this phase. Actions consist only of becoming aware of the lack of information, and thoughts are vague and focused on the general area of the lack of information.

The second phase is the selection phase, which involves identifying and selecting the area or topic to be studied or the approach to be pursued. The feeling of uncertainty is often replaced by a brief feeling of optimism once the area or approach has been selected. The individual feels a readiness to begin the search. Thoughts focus on examining possible topics and weighing them against external circumstances that limit information seeking, such as available time and information, and personal interest. Actions often include obtaining background information on the general area of interest.

In the third phase, the exploration phase, the individual consumes information about the general area of interest to improve his or her personal understanding of it. Feelings of uncertainty usually return and are accompanied by feelings of confusion, frustration and doubt. Thoughts revolve around getting

Tasks	Initiation	Selection	Exploration	Formulation	Collection	Presentation
Feelings (<i>affective</i>)	Uncertainty	Optimism	Confusion, frustration, doubt	Clarity	Sense of direction, confidence	Satisfaction or disappointment
Thoughts (<i>cognitive</i>)	Vague —————→ Focused —————→ Increased interest —→					
Actions (<i>physical</i>)	Seeking relevant information, exploring —————→ Seeking pertinent information, documenting					

Figure 2.1: Principle of Uncertainty for Information Seeking by Kuhlthau (1993)

oriented and sufficiently informed about the topic to form a focus or personal point of view. Actions are determined by finding information relevant to the topic and linking that information to known information to develop understanding.

The fourth phase, the formulation phase, consists of forming a focus from the information found during the exploration. Feelings of confidence increase and feelings of uncertainty decrease. Thoughts become more focused as a focal point of the topic has emerged.

In the fifth phase, the collection phase, information relevant to the previously formulated topic focus is now collected. Feelings of uncertainty continue to subside as confidence keeps increasing and a clear sense of direction is experienced.

The sixth and final phase, the presentation phase, is about completing the search. Individuals experience feelings of satisfaction if the search is successful or disappointment if it is not, often accompanied by a sense of relief. Thoughts focus on completing the search with a personalized understanding of the problem.

Chapter 3

Conducting a Study to Investigate People’s Behavior in Argument-Seeking Conversations

To contribute to the understanding of a conversation between two interlocutors, one of whom has an information need and the other of whom has the desired information, in order to evaluate a future system for vocal argument retrieval, we conducted a study. Although the interlocutor with the desired information will one day be a computer, we cannot directly study the interactions between such a system and its users, because to date there is no voice interface capable of such a conversation. Instead we chose to study the interaction between two human interlocutors. We did not opt for a “Wizard of Oz” setup, in which participants would be under the impression that they were interacting with an advanced computer voice interface, when in fact the voice interface in the study was being simulated by a human (Kelley, 1984). Since people speak differently depending on whether they are talking to another human or to a computer (Hill et al., 2015) such a study design, would have risked examining a biased conversation. To date, no one can predict whether these two forms of communication will one day converge or whether they will always differ. However, since it can be assumed that the current form of a conversation between a human and a machine will change with technological development, we have examined the exchange between two human interlocutors.

3.1 Study setup

The study was conducted in the form of telephone interviews and divided into two parts. All participants agreed on the terms and conditions of the study. This included that we record the conversations, that we are allowed to save

the conversations to analyze it afterwards and that we are allowed to publish the transcripts anonymously. We also informed them that they could withdraw their consent without giving any reason until this work was published. Therefore, they also received the audio recordings and the transcript before the date of publication. Since the preparation effort for the second part of the study was considerable, the study ran for twelve weeks.

In the first part, participants were called and instructed to find an argumentative topic, i.e., a topic that allowed for different points of view in which they were interested. The key points of the instructions mentioned for each participant are listed in Figure 3.1. They could either tell us their results by sending us a message, or we could call them again two days later. Prior to the study conducted, we conducted a pilot study with 5 participants that examined whether our instructions were understandable and whether they resulted in participants naming topics that we could use.

The second part of the study consisted of several one-on-one conversations between myself and one participant at a time about a topic that they had named in the study's first part. These conversations were recorded and then examined for conformance with Kuhlthaus' principle and other patterns and salience.

First part The goal of the first part of the study was to prepare for the second part and consisted of three things.

First, participants were asked to find an argumentative topic that interested them. They were instructed to inspect their everyday life for situations where they would like to have a voice interface to further discuss and elaborate on an argumentative topic that they were interested in. They were also asked to come up with a concretely formulated question that they would ask such an interface. The question did not have to obey any rules, but was only intended as a starting point for the subsequent conversation in the second part of the study. Kuhlthau does not specify a minimum time frame for the observation of her principle. However, it can be assumed that if Kuhlthau's principle applies to the conversations at all, the probability of proving it increases the longer the conversations last. Since people are more willing to engage in conversation about topics that interest them, finding such topics increases the length of conversations.

Second, we asked them to think of keywords and keyphrases related to the feelings that Kuhlthau, in her principle, assigns to the phases of the information seeking process. We intended to examine the transcripts for the occurrence of these keywords and keyphrases in order to draw conclusions about the feelings experienced by our participants. Since, to our knowledge, no such list exists in German for these feelings, we had to create it ourselves. They didn't have

1. Collected data and participant rights
 - We record our conversations and store them securely
 - You can end the study at any time
 - We will publish (possibly translated) anonymized call transcripts
 - You can give us your consent to publish the audio data
 - Before publication, we will send you the transcripts and audio files:
we will exclude each part from publication you want us to exclude
 - Until publication, you can withdraw your consent at any time
2. Goal of the study
 - The study's focus is the information exchange in everyday information-seeking conversations, that is, conversations in which one person tries to acquire knowledge on some topic through talking to someone else
 - We want to analyze the structure of such conversations
 - This could help develop more intuitive search systems in the future (think Alexa).
3. Study procedure
 - After this call:
 - In your daily life, pay attention to topics about that you want to form an opinion (you may have a changeable opinion already)
 - For each such topic think about how you would start a conversation to get information about it (one or two sentences)
 - You can send me these topics or we can arrange another phone call to talk about them in two days
 - We might ask you to change a topic
 - Once we agree on a topic, we will send you a questionnaire to gather your insights on keywords and keyphrases you associate with the feelings described by Kuhlthau.
 - We arrange another call and we prepare for it:
 - We will first ask you whether you are still fine with the study
 - We send you a questionnaire that you fill out immediately
 - Then you should start the conversation as you told me
 - We try to help you form an opinion
 - If we are missing important information, we continue another day
 - We end the conversation naturally and we send you a closing questionnaire

Figure 3.1: Instructions handed out to participants after talking them through. If a participant struggles to find a question, we provide examples “maybe you see a documentary on soccer players and wonder whether they should earn less, or you see some vegan food when shopping and wonder whether a vegan diet is healthy.”

to give us an answer right away, so we provided them with an online survey where they could answer our questions when they had time.

Third, we wanted to get to know each other to some extent through the first phone conversation, so we had direct vocal contact before the second part of the study. The uncertainty people experience when meeting a new person decreases the more time they spend interacting with the new person (Douglas, 1994).

In the first phone call each participant was instructed on what they were supposed to do and what to expect. The call was guided by a manifest (Figure 3.1) that explained all necessary aspects of the study, but left room for the participant to ask further questions, so that each call differed slightly. The opportunity to ask further questions was important, as our pilot study had shown. In this pilot study, all participants were given exactly the same information about what we wanted to study and the same instructions on how to find an argumentative topic before being allowed to ask follow-up questions. These questions revealed that they had a different understanding about what they should do. This confusion seemed to be mainly due to the very different experiences with voice interfaces and the general knowledge about the potential capacities of such systems. We therefore decided against an explanatory video or document because of the risk of greater differences in understanding of the study, despite a possible reduction in variables if exactly the same instructions were used for each participant.

Second part The second part of the study consisted of several one-on-one conversations between us and one participant at a time about a topic the participant was interested in. We prepared the topic by reading up on it to be able to serve as the informed interlocutor. These conversations could consist of multiple phone interviews, depending on how often the complexity of the participants' questions exceeded our knowledge. Before starting that conversation the participants did two things: First, they were asked to give us permission to record the following conversation and to allow us to publish its transcript and audio. Second, they filled out a questionnaire with 9 Likert items shown in Figure 3.2. When they had done both, we initiated the conversation by them asking the question they had named in the study's first part. The length of the conversation depended on two factors: First, participants were free to end the entire conversation at any point. This could be because they felt they had enough information, or because they had lost interest in continuing the conversation for other reasons they did not need to name. Second, if the participant asked a question that we, as the informed interlocutor, could not answer, we ended the telephone interview and we prepared the question that would be the starting point for the next telephone interview. Once the conversation was

over (and not interrupted only due to lack of information from the informed interlocutor) the participant filled out the same Likert scale again and the interview concluded.

As described in Chapter 2.3 Kuhlthaus principle includes three realms of human experiences: thoughts, actions and feelings. The following paragraphs detail how we measured each:

Thoughts In her principle, Kuhlthau describes the thoughts of her participants on two levels. First, on a lower level she describes for each phase of her principle thoughts that are connected to these phases. Second, on a higher level she describes a change from thoughts being vague in the beginning to become more focused throughout the process. We restrict our research to the higher level of these two descriptions. If the changes expected by Kuhlthau occur for this higher level, future work may also investigate the lower level.

To access the participants thoughts we let them self-report their impressions on a Likert scale before and after the conversation of the study's second part to examine if Kuhlthaus expected change from vague to focused thoughts can be found. The Likert scale comprises nine items based on three hypotheses and was created following Nemoto and Beglar (2014). The hypotheses were generated by reviewing the literature on how focusness of thought can be measured and discussing among ourselves under what circumstances focusness of thought can be assumed to exist:

- Participants' thoughts are more focused when their willingness to explain the topic to others is higher.
- Participants' thoughts are more focused when they can divide the topic into subtopics.
- Participants' thoughts are more focused when they feel more comfortable discussing the topic with unfamiliar people.

Based on the hypotheses, we created three times three Likert items each, which were designed to be difficult, moderate and easy for participants to endorse see Figure 3.2 (see Appendix A.1 for the original in German). We tested our Likert scale on four participants to see whether a change in focus of thought was evident.

Actions The next examined realm is the realm of action. Kuhlthau expects a change from seeking relevant information (exploration) to seeking pertinent information (documentation). Measuring such a change is challenging as it depends on the topic and the flow of the conversation what actions can be

	Does not apply at all	Does not apply	Rather does not apply	Rather applies	Applies	Applies fully
I feel comfortable explaining the topic in detail.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable explaining the topic to a stranger.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable talking to strangers about details of the topic.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable explaining standard aspects of the topic.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable explaining the topic to an acquaintance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable discussing the standard aspects of the subject with acquaintances.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable explaining the topic in general terms.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable explaining the topic to a friend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable talking to friends about the topic in general.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 3.2: Questionnaire to determine if participants' thoughts became more focused.

counted to what spectrum. We therefore analyzed each conversation about a specific topic of interest afterwards. We manually split the conversation in different subtopics to see if there is a change in the participants search behavior. We expected our participants to only briefly explore a subtopic in the beginning and change between subtopics. The longer the conversation was the more we expected them to stay on a subtopic or stance and dig deeper.

Feelings The third realm are the participants' feelings. Kuhlthau connects every phase in her principle with certain feelings (see Chapter 2.3). Since the change from one phase to the next and thus the change between the feelings associated with a phase can take place at any time, we needed a way to be able to make continuous statements about the participants' feelings. Therefore, we could not assess the feelings with another questionnaire, as this would have been only a snapshot of their feelings and would have made the conversation unnatural. To our knowledge there exists no text-based feeling recognition system or model for the feelings studied. Following the idea of a word-emotion association lexicon (Mohammad and Turney, 2013), we examined the transcripts for the occurrence of keywords and keyphrases that signal the feelings described by Kuhlthau. To create a list of these keywords and keyphrases we had previously asked our participants what keywords and keyphrases they associated with the feelings studied. We expected that the keywords and keyphrases associated with the feelings of a phase would occur in temporal clusters during the conversation. We further expected that these clusters would split all conversations in the same way.

3.2 Study conduct

Following the study design outlined in Chapter 3.1, we conducted a study of 10 conversations¹. The resulting conversations had a total duration time of approximately 4 hours and 30 minutes each conversation lasting about 27 minutes on average.

Demographic Characteristics We recruited the participants by searching for volunteers at the University of Leipzig as well as among acquaintances. We recruited ten participants consisting of five females and five males. Their age varied between 21 years and 63 years with an average age of 37.3 years. Their educational level was divided as follows: One time "Fachabitur" (vocational baccalaureate diploma), one time "Abitur" (high school diploma), two times a diploma, five times a master's degree and one PhD.

¹The transcripts can be found at: <https://doi.org/10.5281/zenodo.7559232>

The conversations carried out Since the conversations were not structured, the length of each conversation naturally varied. On average, a conversation lasted 26:51 minutes, with the longest conversation lasting 46:12 minutes and the shortest lasting 18:35 minutes. All but one of the participants confirmed that they experienced the conversation to be natural. Participant four felt that the conversation was one-sided. Things were explained to him instead of it being an eye-to-eye conversation, and since the participant was at a distinct knowledge disadvantage, he could not participate in the conversation. In the eighth conversation, the participant made use of the option to interrupt the conversation in order to pick up a topic at a later time on which we lacked information. However, since we were unable to find the necessary information, we did not continue the conversation and declared it finished since he had no further interest in the subject except for the two open questions. The remaining nine conversations all ended with the participants signaling to me that they had gained enough insight for the moment and wanted to end the conversation. In Table 3.1 you will find an overview of the conversations held.

Limitations We need to address some of the problems that arose during the conduct of the study or that were identified in retrospect at the planning stage.

We need to point out two possible sources of bias. First, the small sample size is a problem in itself and bias cannot be ruled out in any of the results we will present below. The second possible source is the high level of education of our participants due to the places where they were recruited. In 2019, 16.9% had a diploma or a higher university degree in Germany²; in our study, 80% had this level of education.

Another source for introducing bias to the conversations lies in the setup itself. Only one person assumed the role of the informed interlocutor and since our study was exploratory, we had not designed any kind of manual for the informed interlocutor. Therefore, all conversations are under the influence of his behavior and decisions. This deficiency was particularly noticeable to us during the conduct of the study in that the informed interlocutor always had to decide how much additional information to give out. If you take a look at Figure 3.3 and Figure 3.4, you can see how the informed interlocutor decides on what information to present. In Figure 3.3, the informed interlocutor offers additional information on his own initiative that he believes might be of interest to the interlocutor with the information need, even though the interlocutor has asked a clear question that has already been clearly answered. In Figure

²<https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Tabellen/bildungsabschluss.html> [accessed: 29.12.22]

Conversation	Conversation Topic	Conversation duration
Conversation 1	Should public transport in Germany be free of charge?	21:15 min
Conversation 2	Should Finland give up its neutrality and join NATO?	23:22 min
Conversation 3	Should we switch to electric cars for the environment?	18:35 min
Conversation 4	Does it make sense for me to get an e-car in three years instead of an internal combustion car?	38:22 min
Conversation 5	Should all pet owners be required to have a certificate of competence?	46:12 min
Conversation 6	Should quotas be introduced for minorities in German leadership positions?	24:25 min
Conversation 7	Does it make more sense to buy a house, rent an apartment or something in between so for the future?	27:41 min
Conversation 8	Should or must the <i>Mohrenbrauerei</i> in Dornbirn in Austria be renamed?	21:35 min
Conversation 9	Does it currently make sense to supply weapons to Ukraine?	21:32 min
Conversation 10	Does it make more sense for my ecological footprint to go shopping in the organic market or to buy only organic products in the supermarket?	25:29 min

Table 3.1: Overview of the conversations held.

3.4, the informed interlocutor shows even more initiative by suggesting a new direction to explore. A different informed interlocutor might not have shared information without clearly being asked. The resulting conversations would most likely differ greatly.

Seeking Interlocutor: Do you have any kind of numbers on how many cases there are, so to speak, um where somehow, where there have been injuries caused by dogs in public spaces or also privately, so?

Informed Interlocutor: Yes, um, that is so annually in the range, so of course always varies a bit, but something like 30 to 40000, um there is but not differentiated, please?

Seeking Interlocutor: How many of them are fatal?

Seeking Interlocutor: Um fatal incidents are very few, um that somewhere in the range of one um depending on the year of course. Most of these reports are also for insurance reasons, because if a dog has bitten you and then something happens later, then you must be able to prove that a dog has bitten you, if something gets infected or so. Um, yes, and most bites or accidents are of course related to the dog and the owner.

Figure 3.3: Conversation excerpt: The informed interlocutor provides additional information even though the interlocutor seeking information has asked an explicit question.

Informed Interlocutor: Than in the private sector. Um exactly. So if we want to take a closer look, I can always differentiate more in that direction. But I couldn't give that here at this point um

Seeking Interlocutor: Yes.

Informed Interlocutor: But maybe also interesting. Um so there are of course also worldwide um other countries that introduce quotas and also for other minorities and

Seeking Interlocutor: Mhm, mhm.

Figure 3.4: Conversation excerpt: The informed interlocutor proposes a new direction of the subject on his own initiative.

Chapter 4

Investigating the Existence of the Principle of Uncertainty in Argument-Seeking Conversations

In this chapter we will address research question one:

RQ 1: Does the Principle of Uncertainty for Information Seeking proposed by Kuhlthau apply to an individual discussing different stances on a topic with an interlocutor they believe to have the answers to their questions?

For answering Research Question One we conducted a study as described in Chapter 3.1, resulting in 10 conversations between an interlocutor seeking information and a second interlocutor who could provide the information sought. We analyzed the conversations transcripts for the presence of Kuhlthaus Principle.

Based on our analysis, we conclude that Kuhlthaus' Principle of Uncertainty for Information Seeking does not apply to these conversations, even though our conversations are a process of Information Seeking for our participants. We have found no evidence that Kuhlthaus' Principle applies to the realm of actions and feelings, but it does to the realm of thoughts.

In the following, we will present our findings on each of the three realms described by Kuhlthau, the realm of Thoughts (Chapter 4.1), the realm of Actions (Chapter 4.2) and the realm of Feelings (Chapter 4.3). Each subchapter is divided into four parts. What we expected to observe (expectations), what we actually observed (observations), the implications of our observations (implications), and the limitations of our research for each of Kuhlthaus' realms.

4.1 Realm of Thoughts

In her principle Kuhlthau describes a change of thoughts in her study participants while searching information. In the beginning their thoughts are vague, focussing on the general problem or area of uncertainty. As the process of information seeking continues, thoughts become more focused and her participants seek to end their search with a personal understanding of the problem.

4.1.1 Expectations

To analyze whether participants' thoughts on a topic were vague or focused, we asked them to self-report their impressions before and after the conversations on a Likert scale. The list of the individual Likert items can be found in Figure 3.2 and were designed to be easier to answer in the affirmative the more focused participants' thoughts are, e.g. "I feel comfortable talking to strangers about details of the subject." The agreement interval from which participants could select responses ranged from "Does not apply at all" to "Applies fully" and did not include a neutral response option.

We expected that our participants' agreement with the Likert items would increase in their second questionnaire compared to the pre-conversation questionnaire.

4.1.2 Observations

Table 4.1 shows participants' responses on our Likert scale before and after the conversations, as well as the mean and median for each Likert item.

Looking at the table, the first thing that stands out is that most answers increase between the first response before the conversations and the second response after. The only participant clearly contradicting this overall rise in agreement is participant ten totally agreeing with every single item before the conversation and only afterwards provides a more differentiated answer. Although this behavior could be treated as an outlier and the reported responses ignored from now on, we decided to include participant 10 in the analysis. It is possible that participant 10 was very convinced of his opinion before the conversation and therefore his or her thoughts were very focused at the beginning of the conversation and this only subsided after the conversation. The impression of an increase in agreement is created by the many upward pointing arrows for the individual answers as well as for the mean and median. In 68 cases the reported value is higher after the conversation, in 15 cases it remains the same and only in 7 cases it decreases. The participants report one time value 1, nine times value 2, twenty-six times value 3, forty-seven times value

Participant	Item 1		Item 2		Item 3		Item 4		Item 5		Item 6		Item 7		Item 8		Item 9	
Participant 1	3	4(↑)	2	5(↑)	2	4(↑)	4	6(↑)	3	6(↑)	4	6(↑)	4	5(↑)	5	6(↑)	5	6(↑)
Participant 2	4	6(↑)	5	6(↑)	5	6(↑)	6	6(○)	5	6(↑)	5	6(↑)	5	6(↑)	6	6(○)	5	6(↑)
Participant 3	3	5(↑)	2	5(↑)	1	4(↑)	4	6(↑)	3	5(↑)	3	6(↑)	4	6(↑)	3	5(↑)	4	6(↑)
Participant 4	4	5(↑)	5	6(↑)	5	6(↑)	3	6(↑)	6	6(○)	6	6(○)	3	6(↑)	6	6(○)	6	6(○)
Participant 5	3	4(↑)	3	4(↑)	4	5(↑)	4	5(↑)	4	5(↑)	4	5(↑)	3	5(↑)	4	6(↑)	4	6(↑)
Participant 6	5	4(↑)	4	3(↓)	4	4(○)	3	4(↑)	3	4(↑)	3	4(↑)	3	4(↑)	3	4(↑)	4	4(○)
Participant 7	3	4(↑)	3	3(○)	2	3(↑)	4	4(○)	3	5(↑)	3	5(↑)	2	4(↑)	5	5(○)	5	6(↑)
Participant 8	2	5(↑)	4	6(↑)	3	5(↑)	3	4(↑)	4	6(↑)	4	5(↑)	2	5(↑)	3	6(↑)	4	6(↑)
Participant 9	2	4(↑)	2	5(↑)	4	5(↑)	4	5(↑)	4	5(↑)	5	6(↑)	4	5(↑)	4	5(↑)	5	6(↑)
Participant 10	6	4(↓)	6	4(↓)	6	5(↓)	6	5(↓)	6	6(○)	6	6(○)	6	5(↓)	6	6(○)	6	6(○)
Mean	3.5	4.5(↑)	3.6	4.7(↑)	3.6	4.7(↑)	4.1	5.1(↑)	4.1	5.4(↑)	4.1	5.5(↑)	3.6	5.1(↑)	4.5	5.5(↑)	4.8	5.8(↑)
Median	3	4(↑)	3.5	5(↑)	4	5(↑)	4	5(↑)	4	5.5(↑)	4	6(↑)	3.5	5(↑)	4.5	6(↑)	5	6(↑)

Table 4.1: The results of the self reported questionnaire. Item 1 to item 9 correspond to the Likert items of the questionnaire. The numbers 1 to 6 correspond to the outcome space, with 1 representing total disagreement and 6 total agreement (see also Figure 3.2). The first number of each item represents the participant’s response before the conversation, and the second number represents the response after the conversation.

4, forty-three times value 5 and fifty-four times value 6. Before the conversations, participants indicate 33 times that they disagree with the statements (corresponding numbers 1-3) and 57 times that they agree with the statements (corresponding numbers 4-6). After the conversations, disagreement is indicated only 3 times and agreement 87 times. This strong shift towards agreement can also be observed in the means and medians of the individual items. The increase in the mean for each element ranges from 1 to 1.5 with an average value of 1.16, and the increase in the median ranges from 1 to 2 with an average value of 1.33. This represents an increase in agreement of 19.3% at the mean and 22.2% at the median. Even though these increases seem very high, we need to keep in mind that our participants only indicated the extent to which they agreed with our Likert-type items, rather than directly indicating how focused their thoughts were. Although our Likert scale measures this value of how focussed once thoughts on a particular topic are, we must not confuse the agreement interval of our Likert scale as a measure of this value, because participants that mostly agree with the option “Applies fully” on their Likert items did not strongly agree to some statement like “my thoughts are focussed on that subject”. Also, we must remember that our agreement interval is ordinal in nature. The above increases of about 20% are the result of assigning the values of the agreement interval to the numbers 1 to 6. However, we do not know the distance between two statements such as “Rather applies” and “applies”. Because of these two features of our measurement method, we should be cautious about making statements about effect size.

To determine if this shift in how focused our participants think about their respective topics really exists, we need to find out if our participants have the

same level of how focused they think about their topic before and after the conversations. We therefore have to examine the likelihood that both pre- and post-conversation questionnaire results are based on the same distribution of how focused our participants think about their respective topics. As we assume our interval to be ordinal as well as how focused participants think about their topics to be our single dependent variable, we did Wilcoxon signed rank sum test¹ yielding $V = 0$ and $p = 0.007971$. Therefore, with a typical alpha of 5%, we reject the hypothesis that the responses to both questionnaires have the same basic distribution. Since the median of the observed variable is higher in the second questionnaire for all nine Likert items we assume that throughout the course of our conversation the participants experience a rise in how focused their thoughts are.

4.1.3 Implications

Given our results and discussion in Chapter 4.1.2, it is likely that the value of how focussed our participants thoughts were increased as the conversations progressed. This complies with the higher level of how participants thoughts change of Kuhlthaus' principle. We therefore assume that Kuhlthaus' principle of uncertainty for information seeking prevails in the realm of thought in argument-seeking conversations.

4.1.4 Limitations

As described in Chapter 3.1, our research on the realm of thought of Kuhlthaus' principle is limited to the change of participants' thoughts from vague to focused.

In addition to the concerns expressed in Chapter 3.2, we need to address another uncertainty regarding our study, which concerns our Likert scale. Because of the lack of research measuring how focused one's thoughts are regarding a particular topic and the small sample size of our pilot study testing our Likert scale, we cannot guarantee that the Likert-type items we use correctly measure the desired variable. As an often used and well-proven measure for evaluating the internal consistency of Likert items Cronbach's alpha was used for this study. The value of Cronbachs alpha varies between 0 and 1 and the higher the value the more consistent are the measurements of individual items. For this study we calculated Cronbach's alpha as $\alpha = 0.92$ for the items before and after the conversation. $\alpha = 0.92$ indicates an excellent internal consistency of the Likert items (Gliem and Gliem, 2003). Due to the internal consistency

¹<https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/> [accessed: 02.01.23]

of the scale shown by this, we can at least assume that the items uniformly measure the same underlying variable, even if, as described above, we cannot guarantee that the variable measured corresponds to the variable sought.

4.2 Realm of Actions

In her principle, Kuhlthau describes a shift in the search behavior of her participants. They shift from exploration, in which they search for relevant information about the topic, to documentation, in which they search for relevant information about the topic.

4.2.1 Expectations

To see if our participants were undergoing the same shift that Kuhlthau expects, we divided the conversations into subtopics (see Chapter 3.1). We expected that participants would exhibit more exploratory behavior at the beginning of the conversation, manifested by only briefly addressing various subtopics. The longer the conversation went on, the more we expected them to concentrate on at least some of the subtopics we had previously addressed.

4.2.2 Observations

Since the conversations differed not only in duration but also in topic, we could not use the same number of subtopics for each conversation. Therefore, we manually divided each conversation into a different number of subtopics².

In Figure 4.1, you can see the conversations divided into their respective subtopics. The number of subtopics we identified varied between 5 and 13. On average we divided a conversation into 8.5 subtopics. Conversation 2 on Finland's accession to NATO was divided into the most subtopics, and Conversation 9 on arms deliveries to Ukraine was divided into the fewest subtopics. It is an interesting coincidence that Conversations 3 and 4, which both address e-cars from different aspects, are divided into the same number of subtopics, however only one of these subtopics is similar.

The behavior we expected from our participant is not present in our conversations, as can be easily seen by looking at Figure 4.1. One would expect the beginning of the conversation, the first quarter or so of each sub-figure, to consist of many subtopics, much like the beginning in 4.1a. The second half of the conversation should consist of a few topics that clearly divide the

²At <https://doi.org/10.5281/zenodo.7559232>, the conversations, divided into their respective subtopics, can be found.

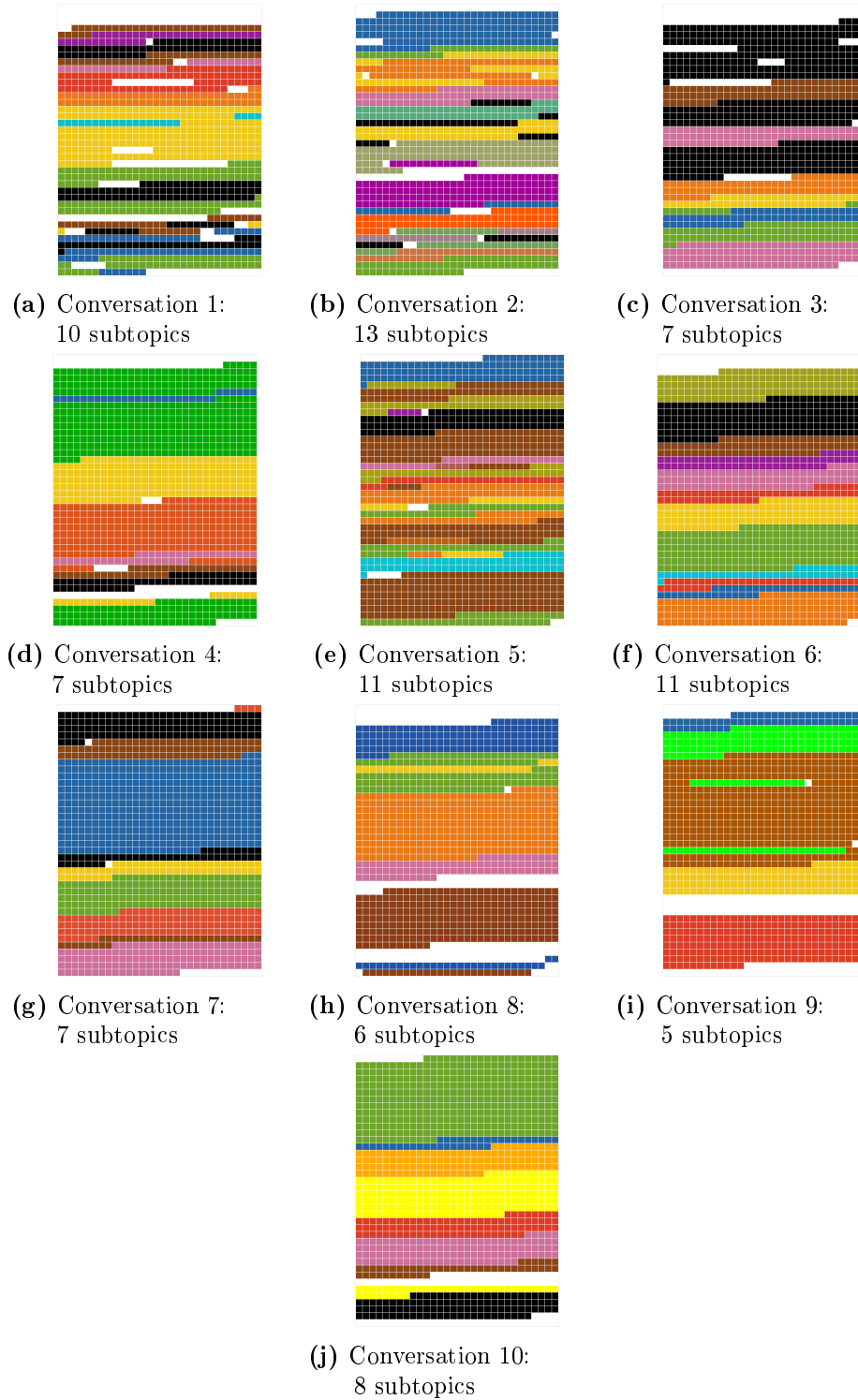


Figure 4.1: Graphical representation of the ten conversations. Each color represents a different topic within a conversation white areas are without topic. For more detailed information see Appendix B.1

Conversation	Part of subtopic dealt with in the beginning	Part of subtopic dealt with in the following
Conversation 1	52.5%	47.5%
Conversation 2	58%	42%
Conversation 3	56.9%	43.1%
Conversation 4	64.5%	35.5%
Conversation 5	44.9%	55.2%
Conversation 6	96.7%	3.3%
Conversation 7	86.9%	13.1%
Conversation 8	85.8%	14.2%
Conversation 9	63.2%	36.9%
Conversation 10	96.6%	3.4%
Average	70.6%	29.4%

Table 4.2: The respective percentages at which, on average, the subtopics of conversations 1 to 10 were discussed in a first attempt or in all subsequent attempts (The subtopics are weighted according to their percentage of the conversation share).

remaining space. However no conversation follows such a pattern. The conversations either deal with one subtopic at a time, as in 4.1f, or alternate between subtopics and take up a subtopic several times, as in 4.1c. To back up these visual impressions we examined the percentages of characters in the transcripts the participants spend on a subtopic. We divided these percentages into two parts. The first part (Part One) includes the characters of the first time a participant touched on the subtopic. This corresponds to the exploring phase of Kuhlthaus' Principle. The second part (Part Two) includes all characters concerning the same subtopic that are not in the first part. This represents the documenting phase of the principle. Between the two parts there has to be at least one subtopic that is touched upon. Thus, while the first part cannot be divided into subparts, the second part may very well consist of numerous subparts that are separated by the participants visiting another subtopic. We calculated these percentages for each subtopic of every conversation (see Appendix B.2). We then examined all conversations to see how much of the conversation belonged to a First Part of some subtopic or to a Second Part of some subtopic. The results can be found in Table 4.2.

There is only one conversation, Conversation 5, in which the participant, on average, did not explore the respective subtopics more thoroughly on its first arising. Conversation 6 and Conversation 10 both deal almost entirely with a subtopic at its first arising. The remaining conversations all too explore subtopics more thoroughly on their first arising.

4.2.3 Implications

The analysis of our results confirm what we already assumed by looking at Figure 4.1. Our expectation that participants would explore a topic by spending little time on subtopics initially and then exploring selected topics thoroughly as the conversation progressed is not supported by the data. In all but one conversation topics are explored more thoroughly on their first arising. Thus, we have no reason to assume that Kuhlthaus' principle applies to the realm of actions in argument-seeking conversations.

4.2.4 Limitations

As mentioned in 3.2, only one person assumed the role of informed interlocutor, leading to a bias in the results. Therefore, the actions of this one informed interlocutor directly affect the search behavior of all searching interlocutors, e.g., by recommending topics not yet discussed, as shown in 3.4. A different approach by the informed interlocutor, e.g., not recommending any subtopics on its own, might have led to different results.

4.3 Realm of Feelings

The last of Kuhlthaus' realms to examine is the realm of Feelings. Kuhlthau divides the information search process into six phases: Initiation, selection, exploration, formulation, collection and presentation. Each of these phases is associated with the information seeker experiencing certain feelings, e.g., in the first phase, the initiation phase, the information seeker is expected to experience uncertainty (see Chapter 2.3).

4.3.1 Expectations

To measure these feelings, we had asked our participants to name keywords and keyphrases that correspond to the feelings described by Kuhlthau. We examined the conversations for the occurrence of these keywords and keyphrases. We expected that keywords and keyphrases belonging to a particular phase would occur together in time in our conversation, so we could divide the conversation into parts corresponding to Kuhlthau's phases.

4.3.2 Observations

Our participants provided us with 198 keywords and keyphrases (see Appendix C.1 for the complete list). For phase two, in which one would experience

optimism, the lowest number of keywords and keyphrases was given: 27, and for phase six, in which one would experience either relief and satisfaction or disappointment, the highest number was given: 50. On average, 33 keywords and keyphrases were reported for each of Kuhlthaus' phases. We examined the conversations for the occurrence of these keywords and keyphrases. Since the informed interlocutor is not seeking for information, Kuhlthau's Principle does not apply to him. Therefore, his feelings are not relevant and we focus only on the parts of the conversations that are spoken by the information seeking interlocutors. We visualized every occurrence of keywords and keyphrases in these parts in all our conversations in Figure 4.2.

Conversation 2 has the fewest occurrences with 13 and conversation 5 the most with 61. On average, we find 35.9 keywords or keyphrases in each conversation. Of the six phases, we find occurrences for only five of them. Keywords or keyphrases for phases one, three and four are present in all 10 conversations. For phase five, we find one keyword in conversation five and for phase six, we find one keyword in conversation six.

The most commonly used keyword is the German equivalent of "um". It occurred 260 times, accounting for more than 72% of all keyword or keyphrase occurrences. Furthermore, the keyword "um" was associated with two phases, phase one and phase three. That is, as long as the keyword "um" is considered, there can be no separation between phases one and three. We have therefore created Figure 4.3 in which this keyword is omitted. This operation reduces the number of keyword or keyphrase occurrences in each conversation to 9.9 on average.

This number is disturbingly low, considering that we have six phases whose existence we need to investigate. Although visual inspection of the Figures 4.2 and 4.3 suggests that there are no sections that can be assigned to any phase, we performed a k-means cluster analysis for each phase. We normalized the conversations by relating the time of occurrence of a keyword or keyphrase to the length of the conversation. We then combined the occurrence of keywords and keyphrases belonging to the same phase from all conversations into a single timeline. We used the keyword list that excludes "um" for the k-means cluster analysis. Since no keyword or keyphrase was found for phase two and only one keyword was found for each of phases five and six, we omitted these three phases for the cluster analysis. We used the elbow method to estimate the most appropriate k and found $k = 3$ for all three phases we considered (Humaira and Rasyidah, 2020). In Figure 4.4 our findings are shown.

As can clearly be seen, we did not find a spatial concentration of keyword or keyphrase occurrence in any phase. The occurrences run throughout the conversation and the clusters identified by the algorithm are not convincing to a human observer, except perhaps for cluster one in phase four. But even if this

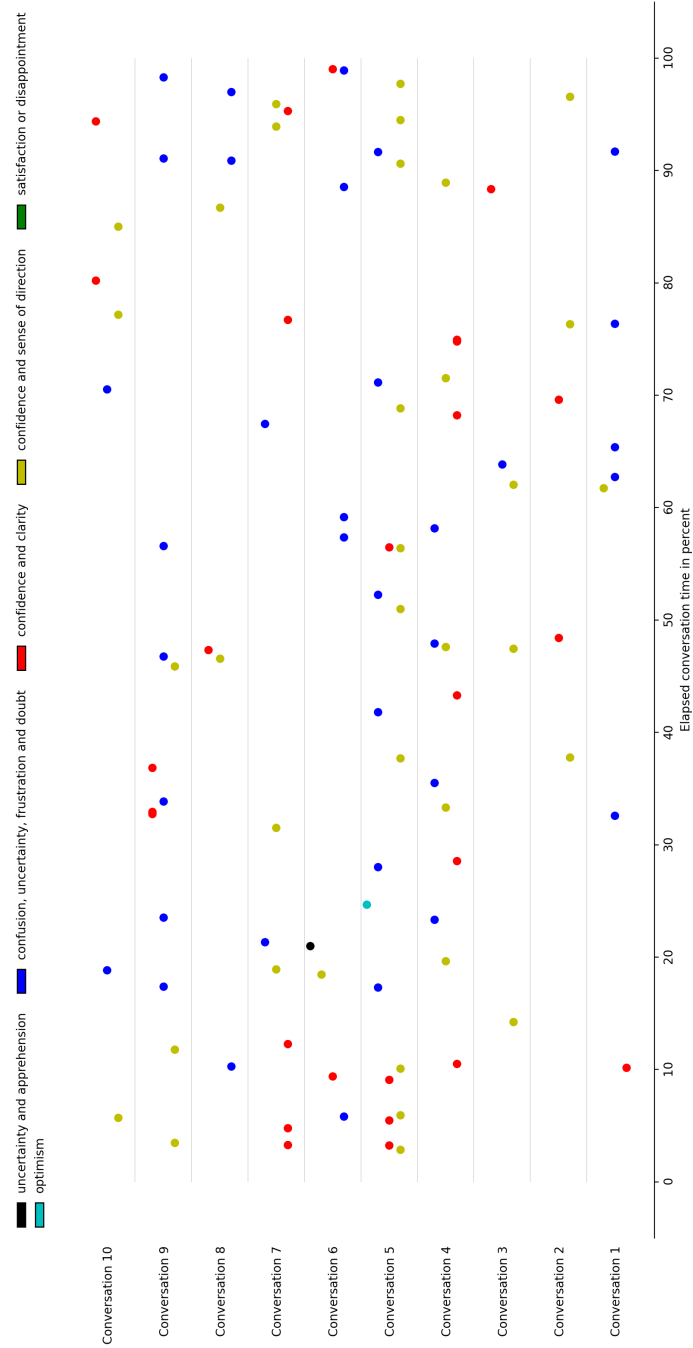
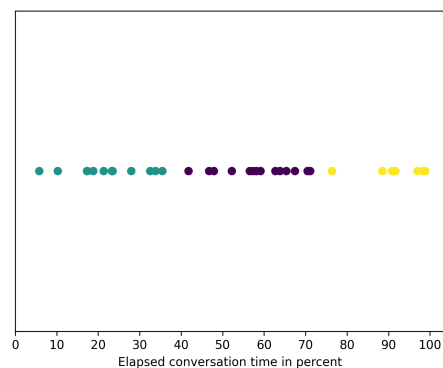
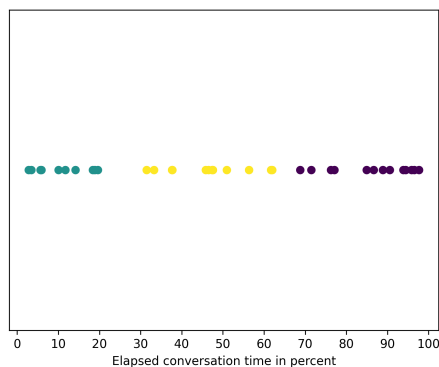


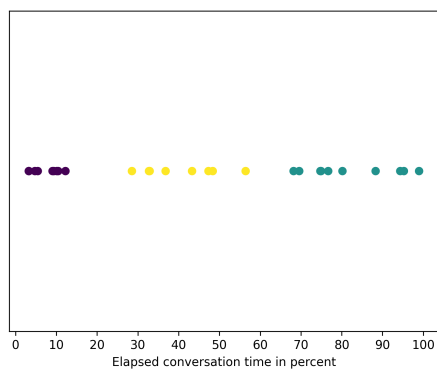
Figure 4.3: Occurrence of keywords or keyphrases of feelings in our conversations without "um". We examined only the parts of speech of the information seeking interlocutor. Each color represents one of the six phases described by Kuhlthau, and each dot in a particular color means that a keyword or keyphrase of a feeling associated with that phase occurred. Unlike in Figure 4.2, we omit the German keyword equivalences of "um".



(a) Phase one



(b) Phase three



(c) Phase four

Figure 4.4: Clustering the occurrences of keywords and keyphrases for the phases one, three and four.

were considered to be a cluster, there are too many occurrences of keywords or key phrases in the temporal course of the conversations associated with this feeling to say that we found one of Kuhlthaus phases. This supports our original assumption that no section of the conversations can be assigned to one of the Kuhlthaus phases.

4.3.3 Implications

The discussion of our results in the previous section showed that the parts of the information seeking interlocutor cannot be divided into subparts by examining the occurrence of keywords and keyphrases from our list. For three of Kuhlthaus six phases we only found one or less occurrences and the conversations could not be divided in subparts of occurrences related to the remaining three phases.

Thus, we have no reason to believe that Kuhlthaus' description of the Realm of Feelings is accurate for argument-seeking conversations.

4.3.4 Limitations

At this point, we should point out that the data, especially the list of keywords and keyphrases on which we base our findings about the realm of feelings, should be taken with a grain of salt. Searching for keywords and keyphrases for a particular feeling or emotion is a difficult task. Since there is no such list for the feelings Kuhlthau describes in German, we had to create one ourselves. We have not verified that the keywords or keyphrases in this list actually show the feelings they are intended to show. Possible future work could be to verify this and if required develop a more accurate list of keywords and key phrases, and to re-examine the presence of Kuhlthaus' principle for the realm of feelings in our conversations.

Chapter 5

Examination of an Argument-Seeking Conversation

As outlined in Chapter 2, the research field of conversational argument retrieval is still in its infancy. To contribute to the development of this field, we conducted a study resulting in 10 argument-seeking conversations between an interlocutor seeking information and a second interlocutor who could provide the information sought. In the previous chapter, we examined whether Kuhlthau’s *Principle of Uncertainty for Information Seeking* applied to these conversations. In this chapter we will address research question two:

RQ 2: What patterns and conspicuities emerge in a conversation in which an individual discusses different stances on a topic with an interlocutor they believe to have the answers to their questions?

To answer this question we present what we noticed and found interesting in and about our conversations, and what might prove informative in designing future systems for conversational argument retrieval.

5.1 Roles to be Assumed by the Interlocutors

In addition to the Likert scale analyzed in Chapter 4.1, we asked all participants two additional questions in the questionnaire after the conversations. First, whether they would agree with the statement “I found the conversation natural.” and second, why they did not find the conversation natural. Since nine of our participants agreed with the statement “I found the conversation natural.” we expected to receive only one response for the second question, but since one participant indicated what he or she liked about the conversation, we received two. One of the participants liked the “good and natural flow of the conversation” as well as it being “very interactive”. He or she also liked the mix

of expertise through the presentation of current findings on the topic, through references to recent studies, and everyday references. The participant that did not agree with the statement “I found the conversation natural.” criticized the one-sidedness of the conversation. He or she felt like things were being explained to him or her instead of it being an eye-to-eye conversation, and he or she was at a distinct knowledge disadvantage and therefore could not really participate in the conversation.

The naturalness of a conversation is, of course, a highly subjective value, and one could argue that all of the conversations in this study were, by definition, not natural because we rarely talk to an expert in a field. Her or his second statement, while probably obvious, is still worth noting: “Things were explained to me rather than it being an eye-to-eye conversation.” We found that in all conversations, the informed interlocutor explained or presented contexts and facts to the interlocutor seeking information see Figure 5.1.

Seeking Interlocutor: What does war of attrition mean?

Informed Interlocutor: Quasi so until it at until one side has no more material or has no more soldiers. So before was

Seeking Interlocutor: Ok.

(a) Excerpt from conversation 9

Informed Interlocutor: Um, in the use of the term Mohr. In general, of course, the term has changed a lot over time in the way it was used.

Seeking Interlocutor: Yeah okay.

Informed Interlocutor: Mohr was originally quite originally quasi the designation simply for the medieval ancient or the inhabitants of the medieval ancient North Africa. Then became so in the 16th century generally [...]

(b) Excerpt from conversation 8

Figure 5.1: Conversation excerpts: The informed interlocutor presents knowledge to the seeking interlocutor.

Similarly, we found that in all conversations, the speech portion of the information seeking interlocutor is characterized by asking questions or opening an area of interest, which are then in turn explained or elaborated on by the informed interlocutor (see Figure 5.2). There are also many conversationally affirmative statements such as “mhm” or “yes” but these represent only a temporally small portion of the total.

We therefore report on the first pattern we have noticed: The information seeking interlocutor asks questions or opens an area of interest, which is then explained or elaborated on by the informed interlocutor.

Seeking Interlocutor: [...] I just ask myself, um why it works cheaper in other countries, for example in Austria. A good friend of mine lives there and she said that they can use the public transport relatively much cheaper than we can in Germany and New Zealand has now also significantly lowered the prices for public transport I'm not so politically involved now, but I find it somehow exciting to look at other countries, because apparently it is somehow possible.

Informed Interlocutor: Yes um I can't tell you exactly how it is regulated in Austria or New Zealand, but if the individual citizen quasi gets the tickets cheaper at the machine [...]

(a) Excerpt from conversation 1

Seeking Interlocutor: So they are part of the EU. Intermediate question, have they also adopted the euro?

(b) Excerpt from conversation 2

Seeking Interlocutor: Um. Yes, and what I um afterwards from this question once again so went through the head. There are in Germany, um I don't know if that is in force, I must honestly admit. Women's quotas

Informed Interlocutor: Mhm.

Seeking Interlocutor: I believe in leadership positions of course then also in my question. Um so succinctly I have so far simply so formulated, but um the women um represent no minority and nevertheless um somehow that is so the only quota, of which I know or that about it is spoken at least.

(c) Excerpt from conversation 6

Figure 5.2: Conversation excerpts: The seeking interlocutor attempts to gain knowledge by either asking explicit questions or opening up an area of interest.

5.2 About the Question Introducing the Conversation

Every conversation started by the information seeking interlocutor asking a question. As described in Chapter 3.1, these questions did not have to obey any rules and served only to open the conversation, yet the questions had some striking similarities. We conducted 10 conversations and therefore received 10 opening questions that could be divided into two groups. The first group contained 6 questions, which were essentially worded as follows: "Should something happen?" (see Figure 5.3). The second group contained 4 questions phrased as either "Does it make sense to do A?" or "Does it make more sense to do A or B?" (see Figure 5.4)

Still both groups of opening questions are quite similar. Both need some

Seeking Interlocutor: Should public transport in Germany be free of charge?

(a) Excerpt from conversation 1

Seeking Interlocutor: Should Finland give up its neutrality and join NATO?

(b) Excerpt from conversation 2

Seeking Interlocutor: Okay. My question is, should all pet owners have to provide proof of suitability or expertise?

(c) Excerpt from conversation 5

Figure 5.3: Conversation excerpts: Opening questions worded like "Should something happen?"

Seeking Interlocutor: Yes. Does it make sense for me to get an e-car instead of a combustion car in three years?

(a) Excerpt from conversation 4

Seeking Interlocutor: So my question to you is, um does it make more sense to buy a house, rent an apartment, or something in between so for the future?

(b) Excerpt from conversation 7

Seeking Interlocutor: So my question is: Does it currently make sense to deliver weapons to Ukraine?

(c) Excerpt from conversation 9

Figure 5.4: Conversation excerpts: Opening questions worded like "Does it make sense to do A?"

kind of goal or value as to why something should be done or make sense. Take for example Figure 5.3a answering this without knowing any objective to why public transport should be free of charge is difficult. Should public transport in Germany be free of charge in order to combat the climate crisis or to increase Germany's gross domestic product? The choice of the goal has a great influence on the further course of the conversation and on which facts are presented. Therefore, in 9 of the 10 conversations, the informed interlocutor tries to find out more about the underlying information need by asking about values or goals that the participant associates with the question, or by introducing some subtopics of the conversation topic to see which ones the participant is interested in, or by combining the two. Only in one conversation, the informed interlocutor assumes to understand the need for information because the initial

question is more precise: “Is it more sensible for my ecological footprint to shop in the organic market or to buy only organic products in the supermarket?” Here, a goal is formulated to measure sense primarily in terms of impact on the ecological footprint. Interesting is also the beginning of the second conversation. Here, Participant 2 asks whether Finland should give up its neutrality and join NATO, and the informed interlocutor presents some possible viewpoints on this issue. Participant 2 then reveals another viewpoint on the issue that has not been presented by the informed interlocutor. Participant 2 is interested in the history of Finland and especially in the alliances the country has made in the past. It seems that this need for information was clear to the participant from the beginning, but he did not include it in his opening question.

It can be noted that 9 of the 10 conversations begin with a phase in which the information seeking interlocutor’s information need is determined in more detail.

5.3 Duration of Conversations

As described in Chapter 3.1, our participants were asked to choose topics they would like to have a conversation about to deepen their understanding. We expected this to have an impact on the duration of the conversations, but since we were not aware of any comparable studies that had been conducted previously, we had no expectations of how long they would actually be. On average, a conversation lasted 26:51 minutes, with the shortest conversation lasting 18:35 minutes and the longest lasting 46:12 minutes (see also Table 3.1). All but one of the participants ended the conversation feeling well informed for the time being and not wanting to continue the conversation. Participant 8 ended the conversation with unanswered questions. We therefore omit conversation 8 for this section, as conversations were supposed to continue as long as the participant had an interest in continuing them. This increases the average conversation duration to 27:26 minutes (the longest and shortest conversation remain the same).

The duration of information-seeking conversations is an interesting value because it limits the maximum amount of information that can be presented, and thus can play a role when the informed interlocutor has to decide whether to elaborate on another topic or just expand on what has been presented so far with another argument.

5.4 Speaking Parts in the Conversation

We did not structure our conversations in advance because we wanted to study conversations that were as natural as possible. We expected that the informed interlocutor would have a larger speaking part in the conversations, since asking a question or expressing interest in a subtopic is usually faster than explaining it. In all but one of the conversations, this assumption proved correct, but in conversation 5, the participant had a 23:24 minute share compared to the informed interlocutor's 22:48 minutes, a 50.65% to 49.35% split. As shown in the table 5.1, in all conversations except the said conversation 5, the informed interlocutor's speech split is larger than that of the seeking interlocutor.

Conversation	Speaking Part	
	Informed Interlocutor	Seeking Interlocutor
Conversation 1	77.8% (83.8%)	22.2% (16.2%)
Conversation 2	81.5% (82.8%)	18.5% (17.2%)
Conversation 3	81.1% (86.3%)	18.9% (13.7%)
Conversation 4	66.2% (56.6%)	33.8% (43.4%)
Conversation 5	49.4% (64.7%)	50.6% (35.3%)
Conversation 6	75.6% (68.7%)	24.4% (31.3%)
Conversation 7	86.2% (92.2%)	13.8% (7.8%)
Conversation 8	56.4% (72.0%)	43.6% (28%)
Conversation 9	64.2% (63.8%)	35.8% (36.2%)
Conversation 10	67.7% (76.0%)	32.3% (24.0%)
Average	70.6% (74.7%)	29.4% (25.4%)

Table 5.1: The conversations were divided by percentage into the informed interlocutor's and the seeking interlocutor's time-share of speech. The percentages for the first ten minutes of each conversation are given in parentheses.

On average, the informed interlocutor contributes 70.6% of the speech to the conversation, this is in line with our expectations.

While we were conducting the conversations, we had the impression that the speech share of the informed interlocutor was larger at the beginning than in the further course of the conversation. Therefore, we examined the speech shares within the first 10 minutes of the conversations. The results are shown in parentheses in table 5.1. Compared to the average values of the whole conversation, we have an increase of 4.1 percentage points in the speaking share of the informed interlocutor. We also like to point out that the speaking share of the informed interlocutor is higher within the first 10 minutes of

conversation 5.

As expected, the informed interlocutor’s share of speech is higher than that of the seeking interlocutor. The even higher proportion of speech by the informed interlocutor at the beginning could be due to the fact that in this phase the information seeker’s knowledge is so low that the questions are shorter and the informed interlocutor takes time to present the basics of the topic. As the conversation progresses, the seeking interlocutor asks more complex questions that correspond to a more differentiated information need and are no longer as easy to formulate as before and must be explained in more detail to the informed interlocutor, which takes more time.

5.5 Conversation Affirming Utterances

Every conversation consists of a sequence of utterances by its participants, and each utterance is made because its maker feels like it at that moment. The overarching goal of the conversations we conducted was to support the information seeking interlocutor in his or her quest for understanding. Therefore, the utterances of the information-seeking interlocutor usually consisted of asking questions, presenting facts that he or she wanted to be put in context, or opening up a subtopic area to explore. The informed interlocutor’s utterances, in turn, usually consisted of presenting facts as answers to questions, contextualizing the facts presented to him, and addressing subtopics. In addition, we have noted another group of utterances that Papenmeier et al. (2022) refer to as “auto” acts. Their purpose is to signal to the other person that one is still listening and can follow what is being said. These utterances consist of a single word or sound such as “Yes” or “Mhm” and add nothing to the previous utterance, and in the case of “Yes”, do not act as a response to a question, see Figure 5.5 for an example.

Although these utterances do not contain task-relevant information, they are numerous in all conversations, see Table 5.2.

On average they make up about 25% of all utterances and on average nearly 75% of this kind of utterances are made by the information seeking interlocutor. This contradicts the findings of Papenmeier et al. (2022). They conducted a study in which they recorded consultation sessions between an expert and a customer on deciding which laptop to buy or what dinner to cook for a group of friends. In contrast to our results, 76% of these “auto” acts were used by the experts in their study.

The right usage of “auto” acts could play a role in how a future voice interface for argument retrieval could be designed to feel more natural to the user, as all participants, as well as the informed interlocutor, exhibited this

Informed Interlocutor: Absolutely. There is also um quasi so um hypotheses so to speak around such quotas. So especially with the women's quota, it's just what I just said with performance and so. There would be quasi the

Seeking Interlocutor: Yes.

Informed Interlocutor: The hypothesis behind this is that in management positions everyone somehow acts in the interests of the company and accordingly um the best qualified candidate or candidate.

Seeking Interlocutor: Mhm.

Informed Interlocutor: Elevate to the vacant position vacant leadership positions. [...]

Figure 5.5: Conversation 6 excerpt: The utterances of the information seeking interlocutor signal that he or she is following what is being said, but do not add further information or indicate a need for information.

behavior. However, the presence of these utterances does not automatically mean that they are desirable, and further research is needed to confirm that a system exhibiting such behavior would indeed feel more natural.

Conversation	Utterances total	Conversation Affirming Utterances
Conversation 1	68	18 (13)
Conversation 2	130	37 (27)
Conversation 3	107	34 (30)
Conversation 4	136	38 (22)
Conversation 5	124	28 (15)
Conversation 6	210	50 (43)
Conversation 7	192	50 (46)
Conversation 8	76	17 (10)
Conversation 9	132	31 (21)
Conversation 10	86	24 (16)
Average	126.1	32.7 (24.3)

Table 5.2: The total number of utterances within a conversation and the number of conversation affirming utterances. The number in parentheses indicates the number of conversation affirming utterances of the information seeking interlocutor.

Chapter 6

Conclusion

With recent advances in argument mining and conversational systems, the possibility of the creation of conversational systems for argument mining is within reach, and soon we could begin to discuss complex topics with voice assistants. With this thesis, we aimed to contribute to this young but growing field of interest by examining the kind of conversations one would have, namely argument-seeking conversations. Understanding what distinguishes these conversations can help design such a system and evaluate whether it is working the way we want it to - whether it is a good system.

In this research, we conducted a study consisting of 10 information-seeking conversations. We asked our participants to find a topic that occurred in their everyday life, about which they knew very little, that is argumentative, i.e. allows for different opinions, and about which they would like to talk to an expert. We then researched the topic ourselves to take on the role of the expert and had an unstructured conversation with our participants, which they used to gather information and form an opinion about. The conversations lasted as long as the participant was interested in continuing them.

We used these conversations to answer our two research questions, first:

RQ 1: Does the Principle of Uncertainty for Information Seeking proposed by Kuhlthau apply to an individual discussing different stances on a topic with an interlocutor they believe to have the answers to their questions?

We found that Kuhlthaus' Uncertainty Principle for Information Seeking does not apply to the conducted kind of information-seeking conversations. However, we were able to confirm that the change in thought she predicted also occurs in information-seeking conversations. Our second research question was more exploratory in nature:

RQ 2: What patterns and conspicuities emerge in a conversation in which an individual discusses different stances on a topic with an interlocutor they believe to have the answers to their questions?

The conversations we conducted had some striking commonalities. All of our participants, as well as ourselves, used utterances that are referred to as “auto” acts. 25% of all utterances made in our conversations are not for the purpose of understanding a topic or formulating an opinion, but to signal to the other person that you understand and are still following what is being said. We found that 75% of these “auto” acts were used by the information seeking interlocutors, which contradicts the results of Papenmeier et al. (2022), which found that 76% of “auto” acts were used by the experts. In addition, we have found that argument-seeking conversations usually begin with a phase in which the information needs of the information seeking interlocutor must be described and explored in more detail.

Future research could address the question of how people explore a topic, as addressed in chapter 4.2. Do people generally prefer to explore a topic by exploring one subtopic at a time, or do they prefer to explore multiple subtopics at once.

Appendix A

Lickert Scale

The original Lickert scale from the questionnaire in German can be found [here](#).

APPENDIX A. LICKERT SCALE

	Trifft über- haupt nicht zu	Trifft nicht zu	Trifft eher nicht zu	Trifft eher zu	Trifft zu	Trifft voll- kommen zu
Ich fühle mich wohl, das Thema detailliert zu erklären.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, das Thema einem Fremden näher zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, wenn ich mit Fremden Details des Themas bespreche.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, Standardaspekte des Themas zu erklären.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, das Thema einem Bekannten näher zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, wenn ich mit Bekannten die Standardaspekte des Themas bespreche.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, das Thema allgemein zu erklären.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, das Thema einem Freund näher zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich wohl, wenn ich mit Freunden allgemein über das Thema spreche.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure A.1: Original questionnaire in german to determine if participants' thoughts became more focused.

Appendix B

Analysis of the conversations for subtopics

B.1 Visualizations

In this section the visualizations of the conversations divided into subtopics can be found.

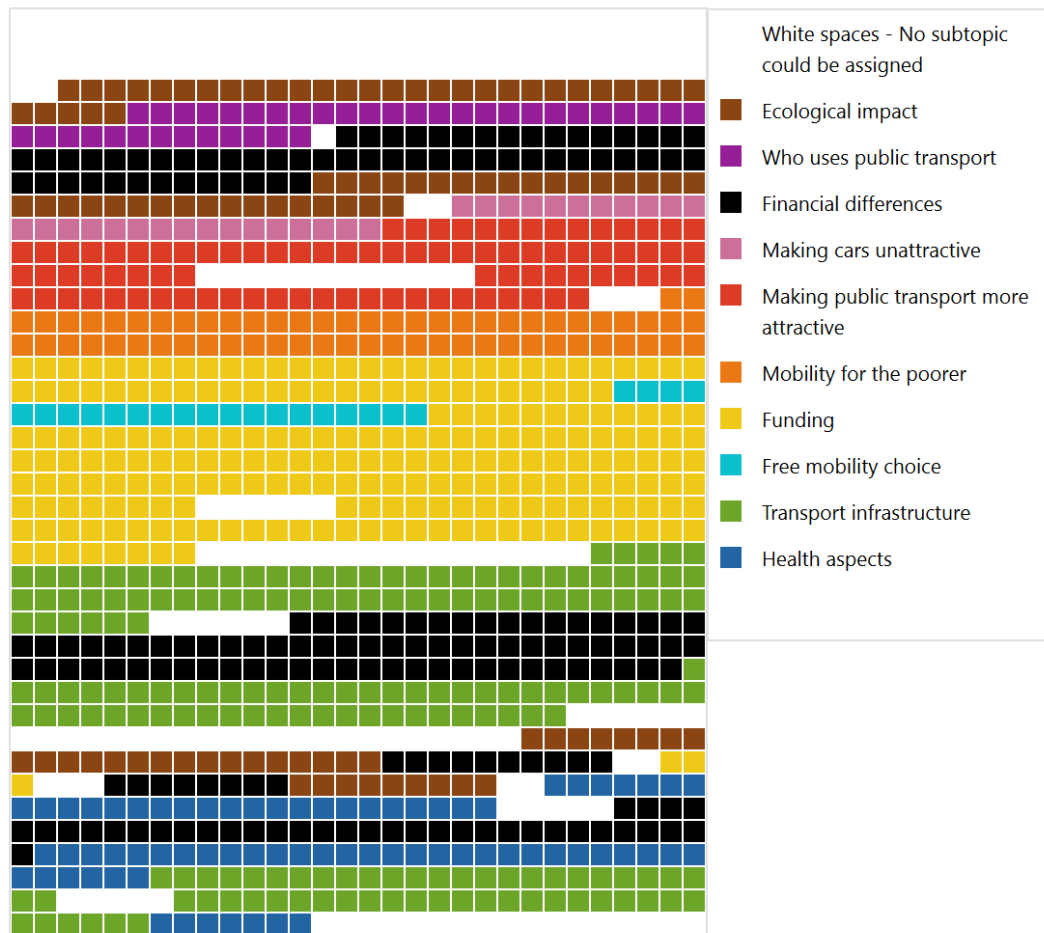


Figure B.1: Conversation 1 addresses a free public transportation system and is broken down into subtopics.

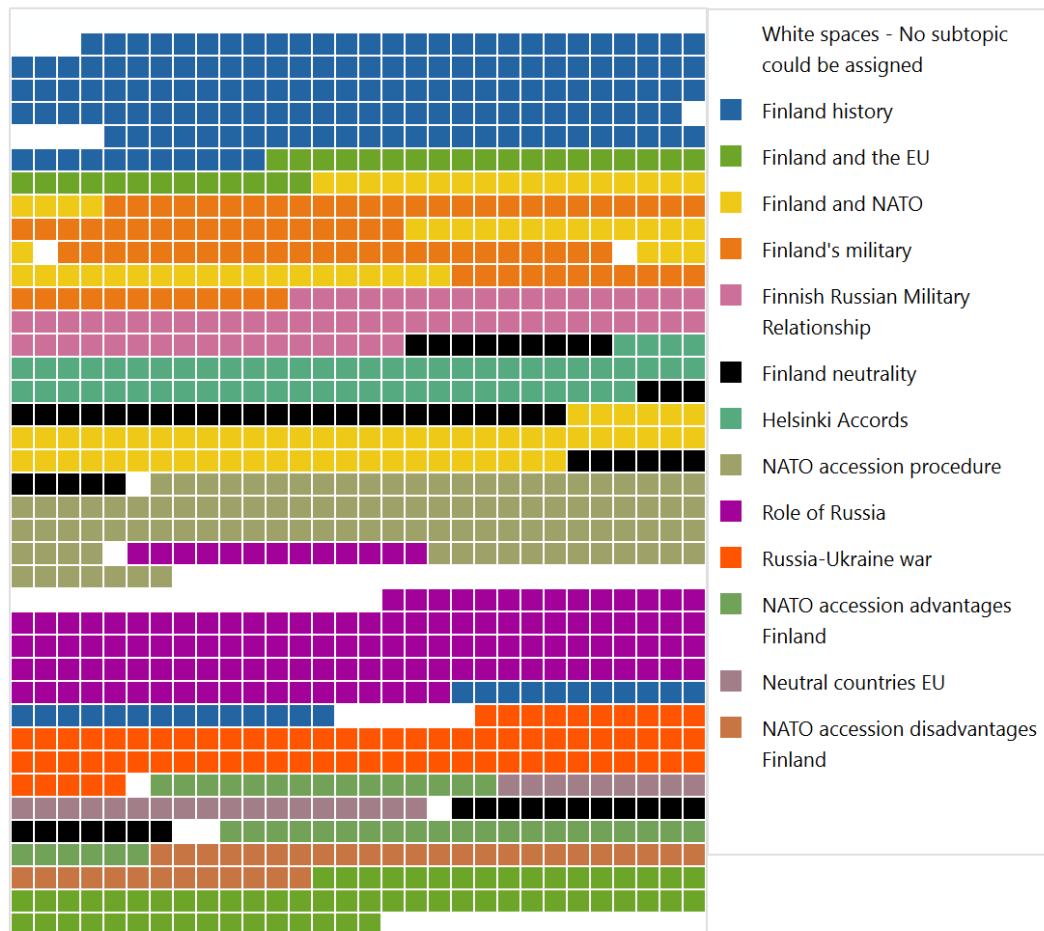


Figure B.2: Conversation 2 addresses the possibility of Finland joining NATO and is broken down into subtopics.

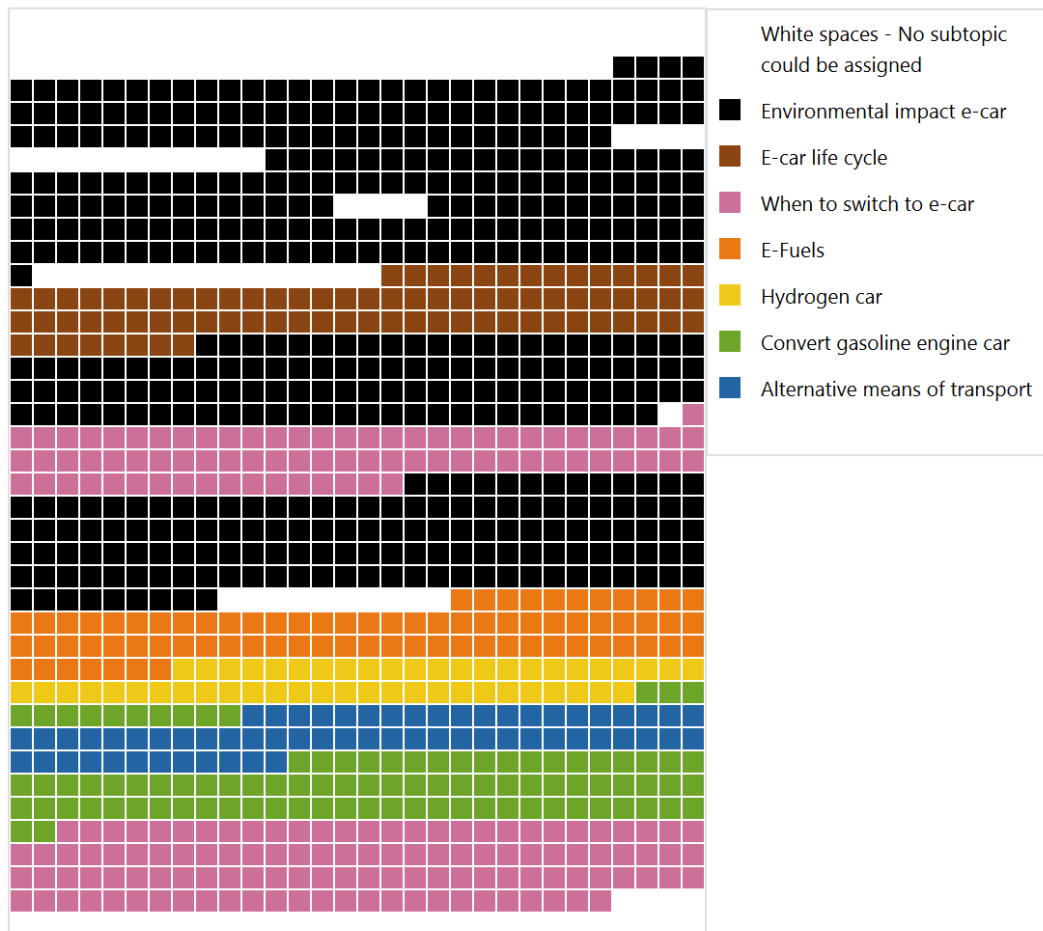


Figure B.3: Conversation 3 addresses the shift to electric cars and its impact on the environment, and is broken down into subtopics.

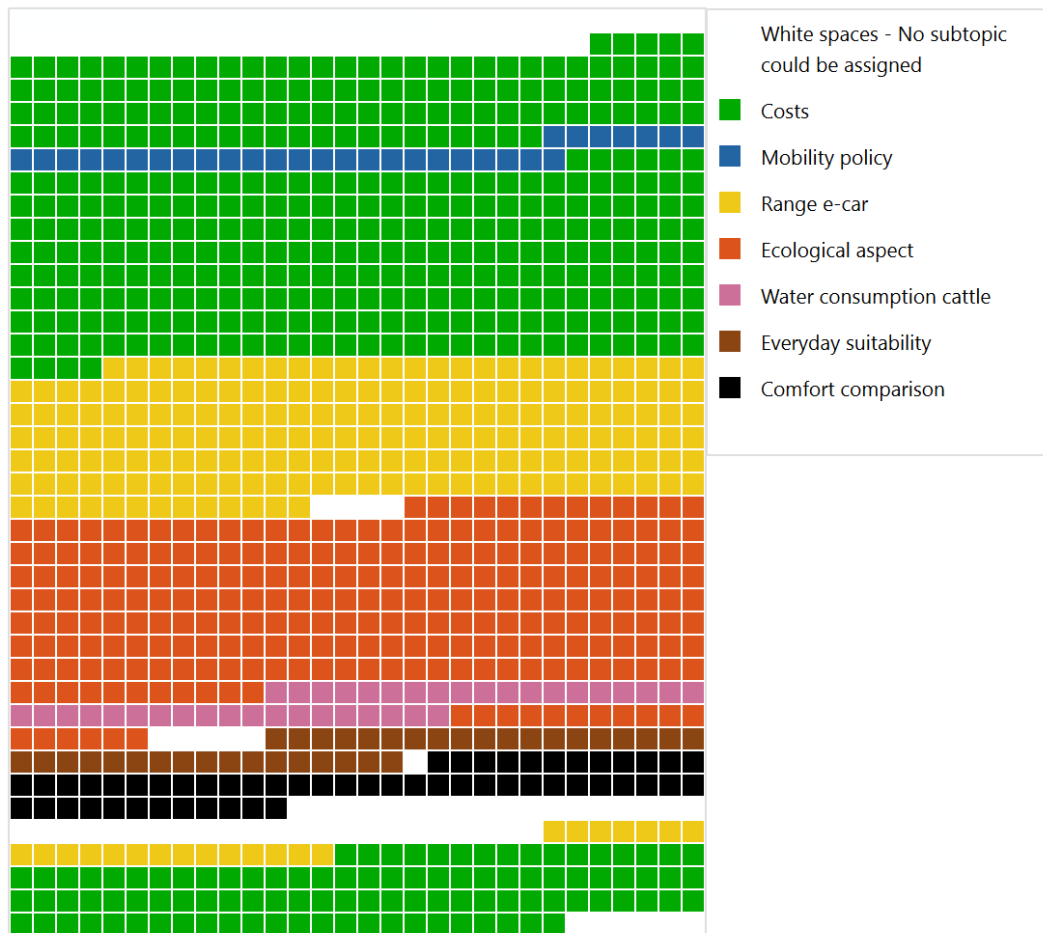


Figure B.4: Conversation 4 addresses a possible individual switch from a combustion engine car to electric car in three years, and is broken down into subtopics.

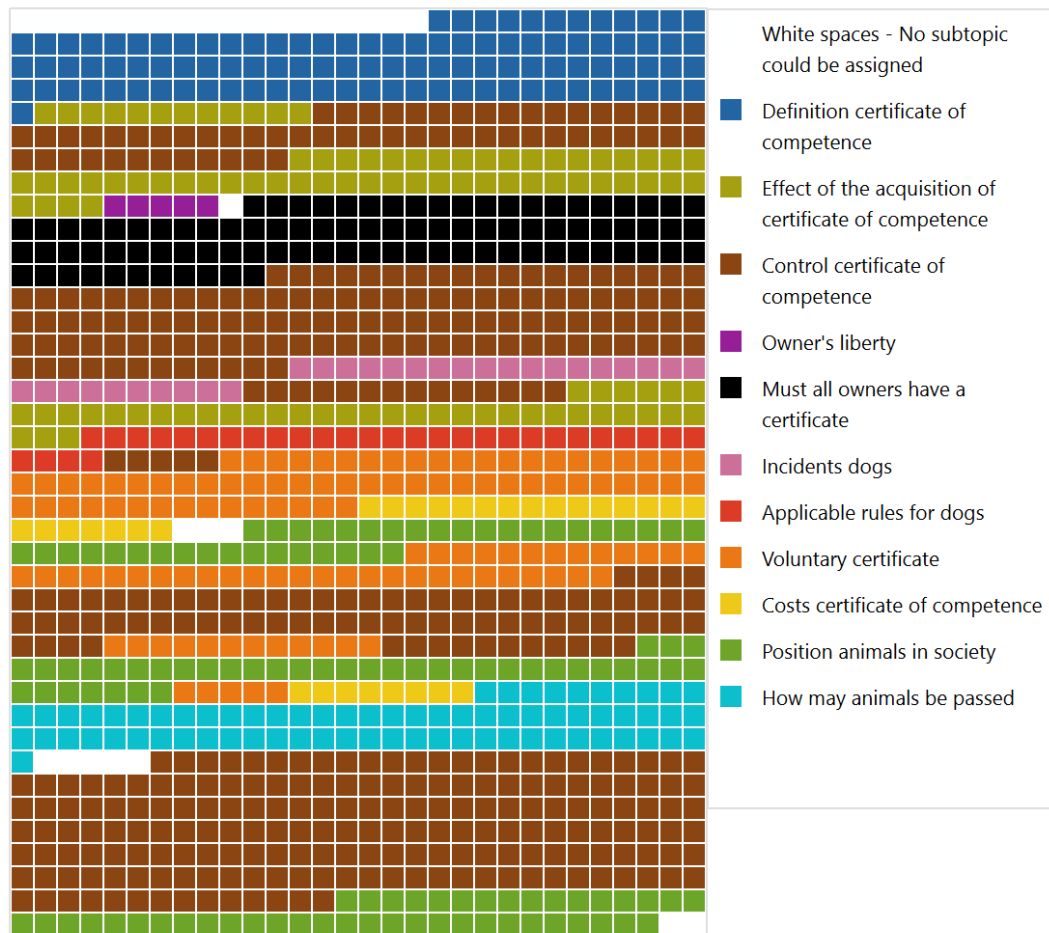


Figure B.5: Conversation 5 addresses certificates of competence for pet owners, and is broken down into subtopics.

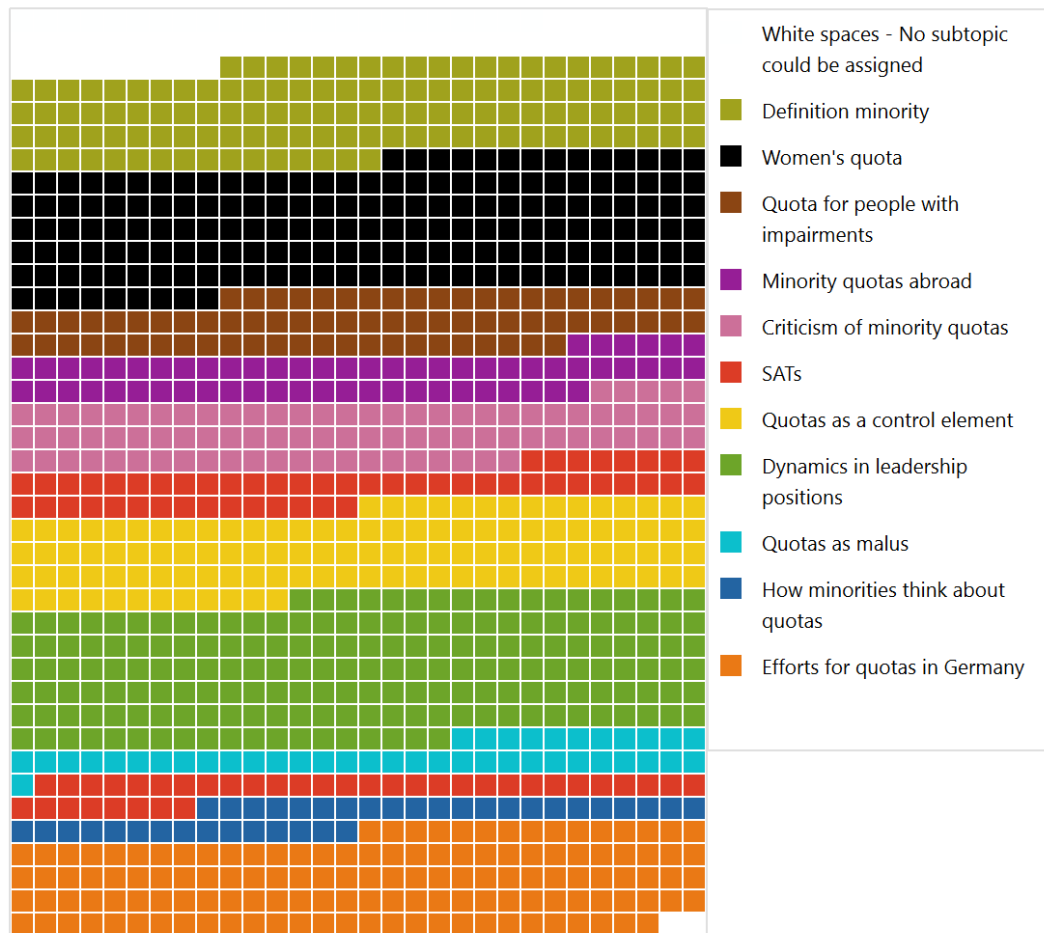


Figure B.6: Conversation 6 addresses quotas for minorities, and is broken down into subtopics.

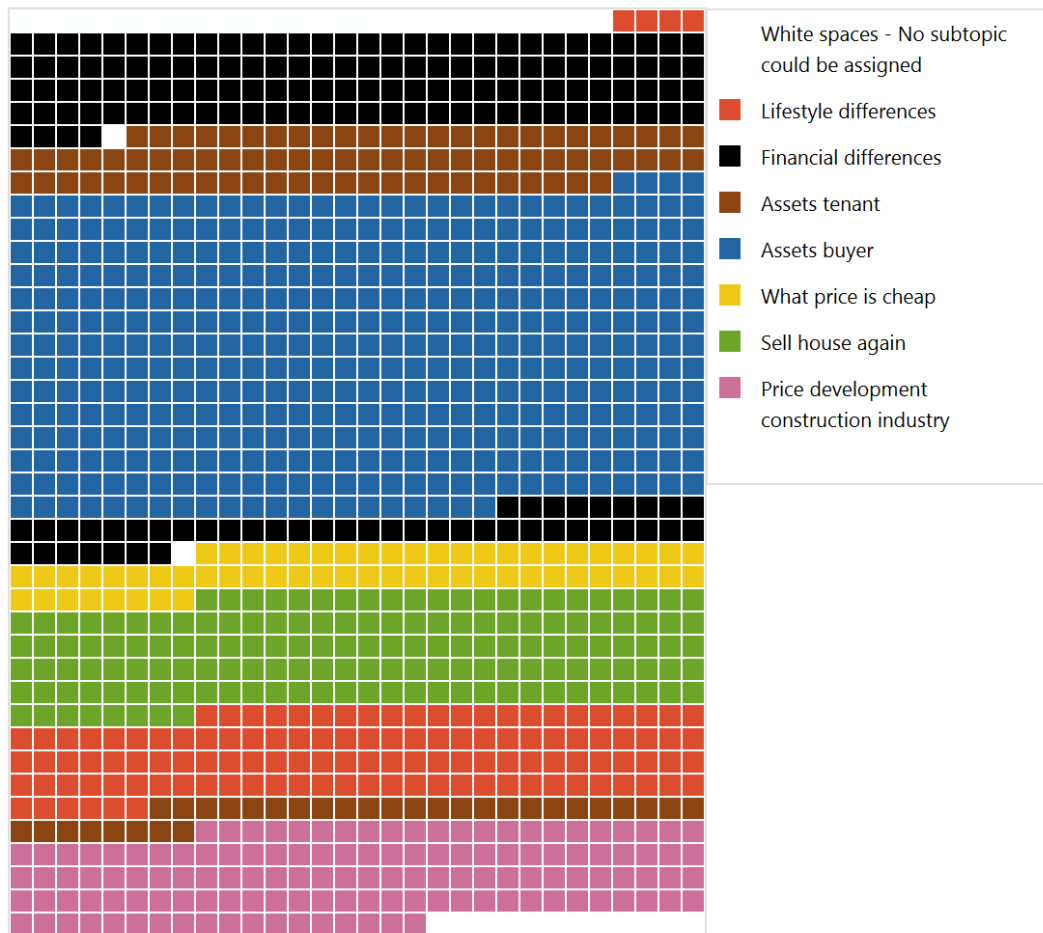


Figure B.7: Conversation 7 explores differences between buying and renting real estate, and is broken down into subtopics.

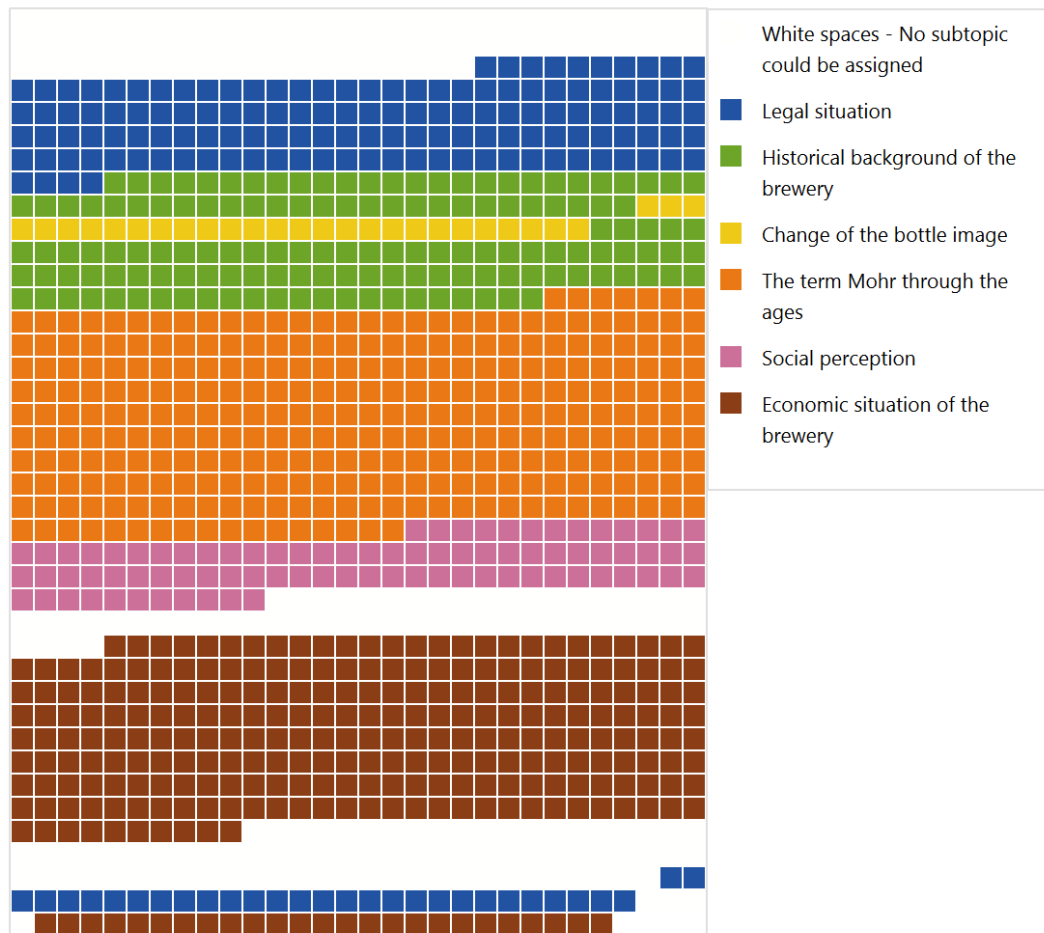


Figure B.8: Conversation 8 addresses the name of the “Mohrenbrauerei”, and is broken down into subtopics.

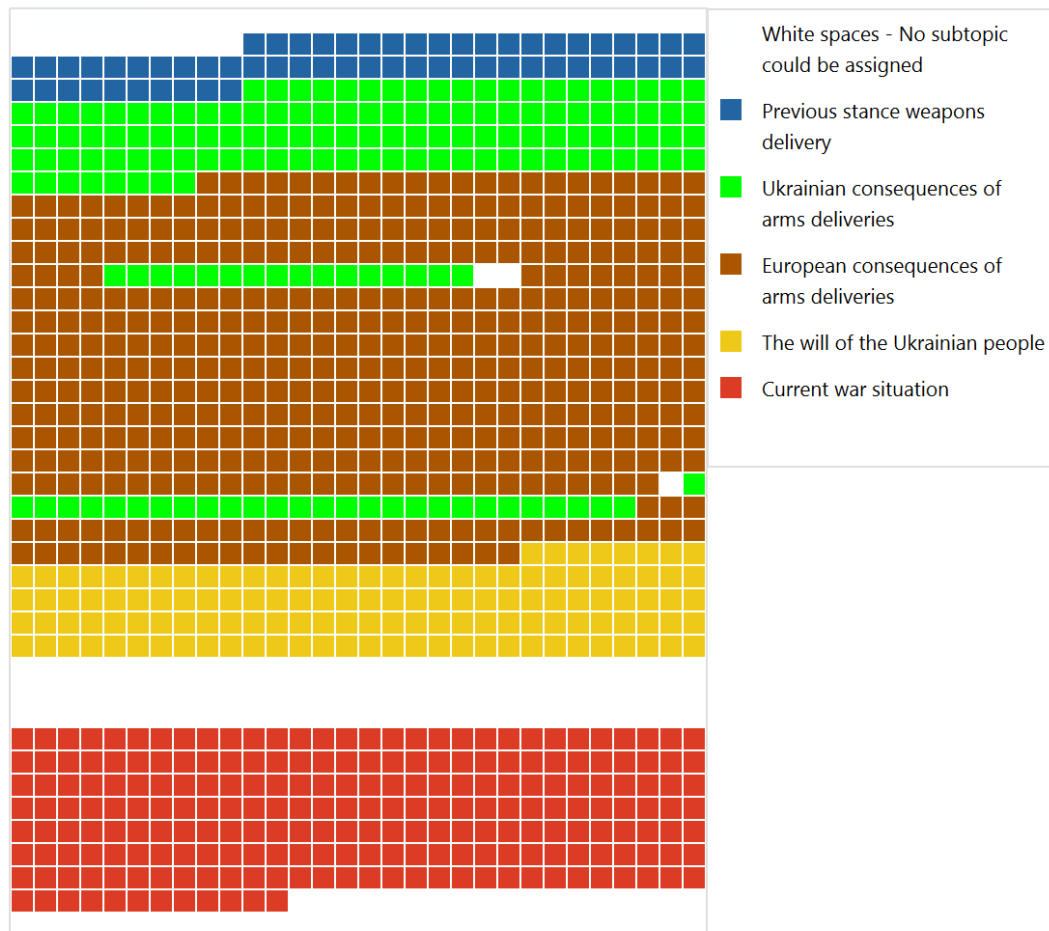


Figure B.9: Conversation 9 addresses arms delivery to Ukraine, and is broken down into subtopics.

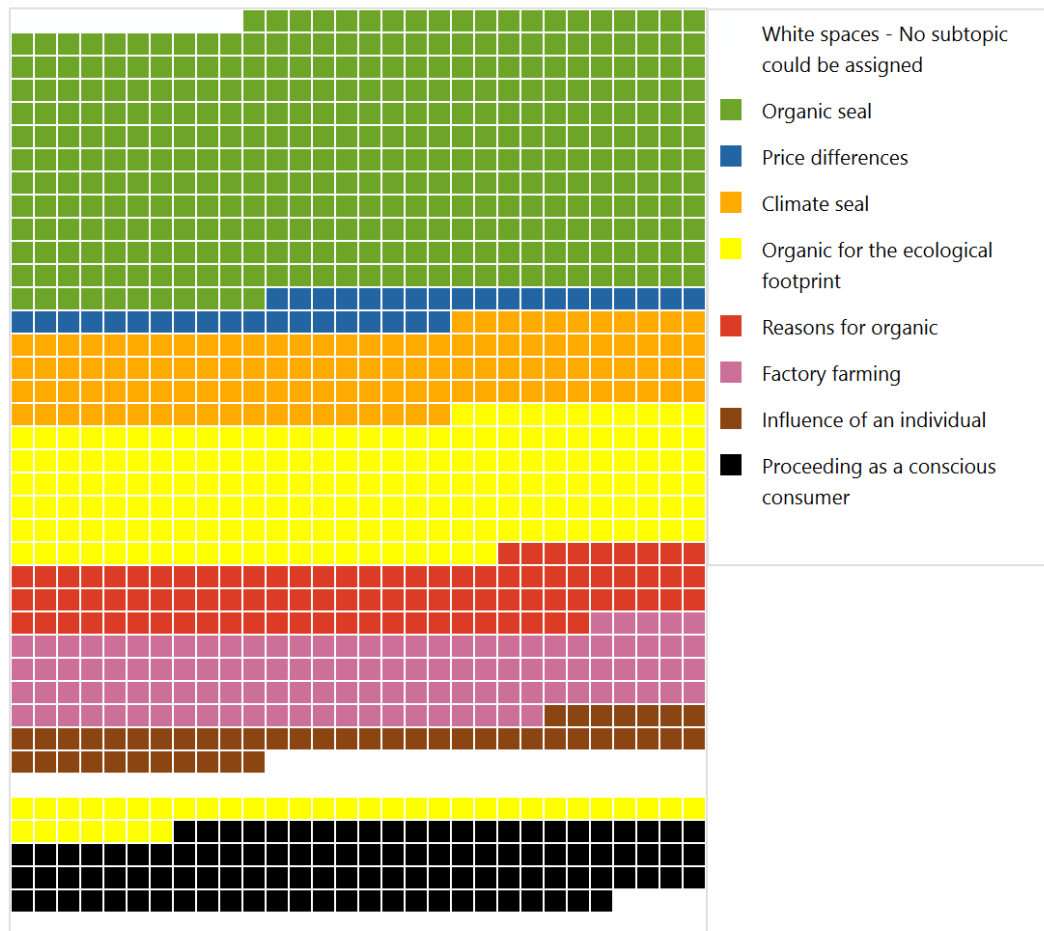


Figure B.10: Conversation 10 addresses the different impacts of supermarkets and organic supermarkets on the environmental footprint and is broken down into subtopics.

B.2 Time spent on subtopic

In this section, the conversations broken down by subtopics in percentages can be found.

Topic	Proportion of the conversation	Percentage breakdown	
Funding	22.7%	25.2%	74.8%
Financial differences	18.4%	33.0%	67.0%
Transport infrastructure	18.3%	39.3%	60.7%
Ecological impact	9.9%	33.7%	66.3%
Making public transport more attractive	8.6%	100.0%	0.0%
Health aspects	7.0%	91.3%	8.7%
Mobility for the poorer	6.4%	100.0%	0.0%
Who uses public transport	3.9%	100.0%	0.0%
Making cars unattractive	2.7%	100.0%	0.0%
Free mobility choice	2.1%	100.0%	0.0%

Table B.1: Conversation 1: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

Topic	Proportion of the conversation	Percentage breakdown	
Finland history	16.5%	86.3%	13.7%
Role of Russia	12.8%	10.1%	89.9%
Finland and NATO	10.8%	18.0%	82.0%
NATO accession procedure	9.8%	82.9%	17.1%
Finland's military	8.3%	49.0%	51.0%
Finland and the EU	7.3%	20.5%	79.5%
Russia-Ukraine war	7.0%	100.0%	0.0%
Finland neutrality	6.1%	14.1%	85.9%
Finnish Russian Military Relationship	6.1%	100.0%	0.0%
Helsinki Accords	5.6%	100.0%	0.0%
NATO accession advantages Finland	3.9%	37.1%	62.9%
NATO accession disadvantages Finland	3.4%	100.0%	0.0%
Neutral countries EU	2.4%	100.0%	0.0%

Table B.2: Conversation 2: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

APPENDIX B. ANALYSIS OF THE CONVERSATIONS FOR SUBTOPICS

Topic	Proportion of the conversation	Percentage breakdown	
Environmental impact e-car	46.5%	47.4%	52.6%
When to switch to e-car	18.4%	40.6%	59.4%
Convert gasoline engine car	8.9%	14.2%	85.8%
E-car life cycle	7.8%	100.0%	0.0%
E-Fuels	7.5%	100.0%	0.0%
Alternative means of transport	6.0%	100.0%	0.0%
Hydrogen car	4.8%	100.0%	0.0%

Table B.3: Conversation 3: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

Topic	Proportion of the conversation	Percentage breakdown	
Costs	43.1%	25.4%	74.6%
Ecological aspect	23.1%	93.6%	6.4%
Range e-car	19.3%	90.1%	9.9%
Comfort comparison	5.0%	100.0%	0.0%
Water consumption cattle	3.5%	100.0%	0.0%
Everyday suitability	3.3%	100.0%	0.0%
Mobility policy	2.8%	100.0%	0.0%

Table B.4: Conversation 4: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

Topic	Proportion of the conversation	Percentage breakdown	
Control certificate of competence	39.8%	12.6%	87.4%
Voluntary certificate	10.5%	54.4%	45.6%
Position animals in society	10.3%	30.6%	69.4%
Effect of the acquisition of certificate of competence	8.8%	12.3%	87.7%
Definition certificate of competence	8.8%	100.0%	0.0%
Must all owners have a certificate	7.8%	100.0%	0.0%
How may animals be passed	6.1%	100.0%	0.0%
Applicable rules for dogs	2.7%	100.0%	0.0%
Costs certificate of competence	2.5%	72.8%	27.2%
Incidents dogs	2.4%	100.0%	0.0%
Owner's liberty	0.4%	100.0%	0.0%

Table B.5: Conversation 5: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

APPENDIX B. ANALYSIS OF THE CONVERSATIONS FOR SUBTOPICS

Topic	Proportion of the conversation	Percentage breakdown	
Dynamics in leadership positions	16.7%	100.0%	0.0%
Women's quota	15.4%	100.0%	0.0%
Efforts for quotas in Germany	11.7%	100.0%	0.0%
Definition minority	11.3%	100.0%	0.0%
Quotas as a control element	10.4%	100.0%	0.0%
SATs	8.0%	58.7%	41.3%
Criticism of minority quotas	7.7%	100.0%	0.0%
Quota for people with impairments	6.6%	100.0%	0.0%
Minority quotas abroad	5.4%	100.0%	0.0%
Quotas as malus	3.6%	100.0%	0.0%
How minorities think about quotas	3.3%	100.0%	0.0%

Table B.6: Conversation 6: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

Topic	Proportion of the conversation	Percentage breakdown	
Assets buyer	37.4%	100.0%	0.0%
Financial differences	15.5%	72.7%	27.3%
Sell house again	13.5%	100.0%	0.0%
Price development construction industry	11.7%	100.0%	0.0%
Assets tenant	10.3%	71.7%	28.3%
Lifestyle differences	6.4%	5.6%	94.4%
What price is cheap	5.3%	100.0%	0.0%

Table B.7: Conversation 7: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

Topic	Proportion of the conversation	Percentage breakdown	
The term Mohr through the ages	30.1%	100.0%	0.0%
Economic situation of the brewery	27.6%	90.8%	9.2%
Legal situation	16.7%	82.3%	17.7%
Historical background of the brewery	14.1%	37.9%	62.1%
Social perception	8.6%	100.0%	0.0%
Change of the bottle image	2.9%	100.0%	0.0%

Table B.8: Conversation 8: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

APPENDIX B. ANALYSIS OF THE CONVERSATIONS FOR SUBTOPICS

Topic	Proportion of the conversation	Percentage breakdown	
European consequences of arms deliveries	43.9%	25.9%	74.1%
Current war situation	21.8%	100.0%	0.0%
Ukrainian consequences of arms deliveries	16.0%	73.1%	26.9%
The will of the Ukrainian people	12.5%	100.0%	0.0%
Previous stance weapons delivery	5.8%	100.0%	0.0%

Table B.9: Conversation 9: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

Topic	Proportion of the conversation	Percentage breakdown	
Organic seal	32.6%	100.0%	0.0%
Organic for the ecological footprint	19.9%	83.1%	16.9%
Climate seal	10.8%	100.0%	0.0%
Factory farming	10.7%	100.0%	0.0%
Proceeding as a conscious consumer	9.9%	100.0%	0.0%
Reasons for organic	8.5%	100.0%	0.0%
Influence of an individual	4.3%	100.0%	0.0%
Price differences	3.4%	100.0%	0.0%

Table B.10: Conversation 10: Column two shows the percentage of the conversation devoted to the subtopic. Column three shows how much of this percentage was covered in the first coverage of the subtopic and column four the subsequent coverage.

Appendix C

Analysis of the conversations for experienced feelings

The list of key words and key phrases used to analyze participants' feelings can be found [here](#).

APPENDIX C. ANALYSIS OF THE CONVERSATIONS FOR EXPERIENCED FEELINGS

[illegible]

Table C.1: The keywords and keyphrases associated with Kuhlthaus’ phases.

Bibliography

- Aliannejadi, M., Chakraborty, M., Ríssola, E. A., and Crestani, F. (2020). Harnessing evolution of multi-turn conversations for effective answer retrieval. In O’Brien, H. L., Freund, L., Arapakis, I., Hoeber, O., and Lopatovska, I., editors, *CHIIR ’20: Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, March 14-18, 2020*, pages 33–42. ACM. 2
- Chan, D. M., Ghosh, S., Chakraborty, D., and Hoffmeister, B. (2022). Multi-modal pre-training for automated speech recognition. In *ICASSP 2022*. 1
- Deriu, J., Rodrigo, Á., Otegi, A., Echegoyen, G., Rosset, S., Agirre, E., and Cieliebak, M. (2021). Survey on evaluation methods for dialogue systems. *Artif. Intell. Rev.*, 54(1):755–810. 2.1
- Douglas, W. (1994). The acquaintanceship process:: An examination of uncertainty, information seeking, and social attraction during initial conversation. *Communication Research*, 21(2):154–176. 3.1
- Gliem, J. A. and Gliem, R. R. (2003). Calculating, interpreting, and reporting cronbach’s alpha reliability coefficient for likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community 4.1.4
- Hill, J., Ford, W. R., and Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Comput. Hum. Behav.*, 49:245–250. 3
- Huang, L., Ye, Z., Qin, J., Lin, L., and Liang, X. (2020a). GRADE: automatic graph-enhanced coherence metric for evaluating open-domain dialogue systems. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 9230–9240. Association for Computational Linguistics. 2.1

- Huang, M., Zhu, X., and Gao, J. (2020b). Challenges in building intelligent open-domain dialog systems. *ACM Trans. Inf. Syst.*, 38(3):21:1–21:32. 2.2
- Humaira, H. and Rasyidah, R. (2020). Determining the appropriate cluster number using elbow method for k-means algorithm. In *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA)*. 4.3.2
- Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.*, 2(1):26–41. 3
- Khatib, K. A., Hou, Y., and Stede, M., editors (2021). *Proceedings of the 8th Workshop on Argument Mining, ArgMining@EMNLP 2021, Punta Cana, Dominican Republic, November 10-11, 2021*. Association for Computational Linguistics. 2
- Kiesel, J., Lang, K., Wachsmuth, H., Hornecker, E., and Stein, B. (2020). Investigating expectations for voice-based and conversational argument search on the web. In O’Brien, H. L., Freund, L., Arapakis, I., Hoeber, O., and Lopatovska, I., editors, *CHIIR ’20: Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, March 14-18, 2020*, pages 53–62. ACM. 1, 2, 2.2
- Kiesel, J., Spina, D., Wachsmuth, H., and Stein, B. (2021). The meant, the said, and the understood: Conversational argument search and cognitive biases. In Schlögl, S., Porcheron, M., and Clark, L., editors, *CUI 2021 - 3rd Conference on Conversational User Interfaces, Virtual Event / Bilbao, Spain, July 27-29, 2021*, pages 20:1–20:5. ACM. 1
- Kuhlthau, C. C. (1988a). Developing a model of the library search process: Cognitive and affective aspects. 2.2
- Kuhlthau, C. C. (1988b). The information search process of high-, middle-, and low-achieving high school seniors. *School Library Media Quarterly*, 17. 2.2, 2.3
- Kuhlthau, C. C. (1988c). Longitudinal case studies of the information search process of users in libraries. *Library & Information Science Research*, 10:257–304. 2.2
- Kuhlthau, C. C. (1988d). Perceptions of the information search process in libraries: a study of changes from high school through college. *Information Processing & Management*, 24(4):419–427. 2.2

- Kuhlthau, C. C. (1990). Validating a model of the search process: A comparison of academic, public and school library users. *Library & Information Science Research*, 12:5–31. 2.2
- Kuhlthau, C. C. (1993). A principle of uncertainty for information seeking. *J. Documentation*, 49(4):339–355. 1, 2.2, 2.3, 2.1, 4, 6
- Lawrence, J. and Reed, C. (2019). Argument mining: A survey. *Comput. Linguistics*, 45(4):765–818. 2
- Levy, R., Bogin, B., Gretz, S., Aharonov, R., and Slonim, N. (2018). Towards an argumentative content search engine using weak supervision. In Bender, E. M., Derczynski, L., and Isabelle, P., editors, *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 2066–2081. Association for Computational Linguistics. 2
- Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., and Dolan, W. B. (2016). A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics. 2.2
- Li, Z., Zhang, J., Fei, Z., Feng, Y., and Zhou, J. (2021). Conversations are not flat: Modeling the dynamic information flow across dialogue utterances. In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 128–138. Association for Computational Linguistics. 2.1
- Mochales, R. and Moens, M.-F. (2011). Argumentation mining. *Artificial Intelligence and Law*, 19. 2
- Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Comput. Intell.*, 29(3):436–465. 3.1
- Nemoto, T. and Beglar, D. (2014). Developing likert-scale questionnaires. In *JALT2013 CONFERENCE PROCEEDINGS*. 3.1
- Ni, J., Young, T., Pandelea, V., Xue, F., Adiga, V., and Cambria, E. (2021). Recent advances in deep learning based dialogue systems: A systematic survey. *CoRR*, abs/2105.04387. 1, 1, 2, 2.1

- Papenmeier, A., Frummet, A., and Kern, D. (2022). "mhm..." - conversational strategies for product search assistants. In Elswailer, D., editor, *CHIIR '22: ACM SIGIR Conference on Human Information Interaction and Retrieval, Regensburg, Germany, March 14 - 18, 2022*, pages 36–46. ACM. 2.2, 5.5, 5.5, 6
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, page 311–318, USA. Association for Computational Linguistics. 1
- Ram, A., Prasad, R., Khatri, C., Venkatesh, A., Gabriel, R., Liu, Q., Nunn, J., Hedayatnia, B., Cheng, M., Nagar, A., King, E., Bland, K., Wartick, A., Pan, Y., Song, H., Jayadevan, S., Hwang, G., and Pettigru, A. (2018). Conversational AI: the science behind the alexa prize. *CoRR*, abs/1801.03604. 2
- Ren, P., Liu, Z., Song, X., Tian, H., Chen, Z., Ren, Z., and de Rijke, M. (2021). Wizard of search engine: Access to information through conversations with search engines. In Diaz, F., Shah, C., Suel, T., Castells, P., Jones, R., and Sakai, T., editors, *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 533–543. ACM. 2.2
- Roller, S., Boureau, Y., Weston, J., Bordes, A., Dinan, E., Fan, A., Gunning, D., Ju, D., Li, M., Poff, S., Ringshia, P., Shuster, K., Smith, E. M., Szlam, A., Urbanek, J., and Williamson, M. (2020). Open-domain conversational agents: Current progress, open problems, and future directions. *CoRR*, abs/2006.12442. 2.2
- Smith, E. M., Hsu, O., Qian, R., Roller, S., Boureau, Y., and Weston, J. (2022). Human evaluation of conversations is an open problem: comparing the sensitivity of various methods for evaluating dialogue agents. In Liu, B., Papangelis, A., Ultes, S., Rastogi, A., Chen, Y., Spithourakis, G., Nouri, E., and Shi, W., editors, *Proceedings of the 4th Workshop on NLP for Conversational AI, ConvAI@ACL 2022, Dublin, Ireland, May 27, 2022*, pages 77–97. Association for Computational Linguistics. 2.1
- Spina, D., Trippas, J. R., Thomas, P., Joho, H., Byström, K., Clark, L., Craswell, N., Czerwinski, M., Elswailer, D., Frummet, A., Ghosh, S., Kiesel, J., Lopatovska, I., McDuff, D., Meyer, S., Mourad, A., Owoicho, P., Cherumanal, S. P., Russell, D., and Sitbon, L. (2021). Report on the future conversations workshop at CHIIR 2021. *SIGIR Forum*, 55(1):6:1–6:22. 2

- Trippas, J. R., Spina, D., Thomas, P., Sanderson, M., Joho, H., and Cavendon, L. (2019). Towards a model for spoken conversational search. *CoRR*, abs/1910.13166. 1, 2.2
- Xu, L., Gu, Y., Kolehmainen, J., Khan, H., Gandhe, A., Rastrow, A., Stolcke, A., and Bulyko, I. (2022). Rescorebert: Discriminative speech recognition rescoring with bert. In *ICASSP 2022*. 1
- Yeh, Y., Eskénazi, M., and Mehri, S. (2021). A comprehensive assessment of dialog evaluation metrics. *CoRR*, abs/2106.03706. 2.1
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2020). Bertscore: Evaluating text generation with BERT. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net. 1
- Zhou, L., Gao, J., Li, D., and Shum, H. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Comput. Linguistics*, 46(1):53–93. 2