“Mhm...” – Conversational Strategies For Product Search Assistants

ANDREA PAPENMEIER, GESIS - Leibniz-Institute for the Social Sciences, Germany
ALEXANDER FRUMMET, University of Regensburg, Germany
DAGMAR KERN, GESIS - Leibniz-Institute for the Social Sciences, Germany

Online retail has become a popular alternative to in-store shopping. However, unlike in traditional stores, users of online shops need to find the right product on their own without support from expert salespersons. Conversational search could provide a means to compensate for the shortcomings of traditional product search engines. To establish design guidelines for such virtual product search assistants, we studied conversations in a user study (N = 24) where experts supported users in finding the right product for their needs. We annotated the conversations concerning their content and conversational structure and identified recurring conversational strategies. Our findings show that experts actively elicit the users' information needs using funneling techniques. They also use dialogue-structuring elements and frequently confirm having understood what the client was saying by using discourse markers, e.g., “mhm”. With this work, we contribute insights and design implications for conversational product search assistants.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Natural language interfaces**; *Empirical studies in interaction design.*

Additional Key Words and Phrases: intelligent assistant, conversational search, natural language interface, conversation analysis, product search

ACM Reference Format:
Andrea Papenmeier, Alexander Frummet, and Dagmar Kern. 2022. “Mhm...” – Conversational Strategies For Product Search Assistants. In *Proceedings of the 2022 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22), March 14–18, 2022, Regensburg, Germany*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3498366.3505809

1 INTRODUCTION

Online stores are an important channel for selling products. Many retailers offer both online and in-store shopping and consumers show flexible buying behavior, switching between online and offline stores [14]. Online product search systems, however, offer limited customer support compared to on-site stores, where expert salespersons consult customers about individually suitable products. In an e-commerce setting, customers are left alone with the challenge of verbalizing their information need and mapping it to the information displayed on the search interface [4, 16, 33]. For example, a user searching for a new laptop on *Amazon* needs to formulate a search query that matches the vocabulary of *Amazon*'s database. Additionally, *Amazon* offers facets with which the user can tweak and modify many parameters, such as the price range, processor generation, or screen width. However, many users might not know which technical features are required for their individual use case, e.g., which processor type is suitable for their needs. The literature on product search interfaces has highlighted several misalignments between users' information needs and information and facets offered by search systems [28].

Recent studies found that creating conversation-like interfaces either by adapting the graphical user interface [27] or by asking investigative questions [25] help users to express their information need. Furthermore, conversational search is known to be a suitable framework to elicit user information needs [30]. Conversational systems might, therefore, serve as means to align search systems with the users' information needs and take on a consulting role similar to a salesperson in in-store settings. The literature on conversational systems suggests some guidelines for developing conversational systems, e.g., designing proactive agents [11, 34], context-aware agents [9, 12, 18], and asking questions to elicit and clarify the users' complex information needs [25, 39, 40]. However, high-level guidelines for conversations exist mainly for factual web search [36, 39, 40] or on the conversational aspects of voice assistants and chatbots [9, 18, 22], while works on specific use cases of product search (e.g., flight booking [11, 12], cooking [37]) do not propose explicit design guidelines for product search in general.

With the present work, we extend the current literature on design guidelines for conversational product search systems with an expert role by observing dialogues (N = 24) in which product experts help customers to find suitable products, similar to in-store situations. In our work, we define "product" as an object with multiple attributes, for which users have complex, i.e., multi-faceted, information needs and aim to eventually decide for a single object. We observe conversations for two types of products, laptops and recipes. Based on the dialogues, we analyzed which strategies experts and customers employ to communicate about products. Our findings show five conversational strategies in product search conversations: (1) Experts take a proactive role in the beginning and a reactive role towards the end of a product consultation session, (2) experts use funneling strategies to elicit the user's information needs, going from broad, open questions to nuanced, closed questions, (3) product search conversations contain many interaction-structuring utterances, and (4) experts frequently reflect to the user that they understood what was said. In the recipe domain, (5) experts also ask for exclusion criteria in addition to inclusion criteria, whereas common search engines focus on inclusion. Overall, the present study delivers insights for the design of future conversational product search assistants.

2 RELATED WORK

Research in the area of conversational search has gained momentum in recent years which provides a starting point for developing product search systems that address users' information needs. In this section, we discuss prior work in the field of product search, conversational search, and methods offered to analyze the structure of conversations.

2.1 User Information Needs in Product Search

In online product search, users are required to communicate what they are looking for to the search system (express their information need). Information needs, however, can be complex and multifaceted [1, 30]. Users do not always have a specific product in mind when they start their search [4], which leads to an "exploratory search" [23]. Moreover, even when they have an idea of what they need, they lack the knowledge of how to formulate what they are looking for in a machine-interpretable manner [19, 33]. Product search systems must, therefore, provide guidance and assistance so that users find the most suitable product for them. In the case of complex information needs, users often describe their information need with terms different to those that the machine can understand, e.g., by using synonyms [16, 21, 31] or using vague expressions [28]. When interacting with search systems, users adapt their query formulation to what they believe the search system can understand [19]. In a study on laptop search, Papenmeier et al. [28] found clear differences between how users describe their information need in natural language (e.g., when talking to a peer) and how they formulate a search query for a search system: Natural language descriptions were longer, more vague, and contained more information about the target product than the actual search queries. To help users disclose more information about their genuine information need, the graphical search interfaces can be adapted to a more conversational style [27] and the system can proactively ask investigative questions [25].

2.2 Conversational Search Systems

Radlinski and Craswell [30] defined a conversational search system as "a system for retrieving information that permits a mixed-initiative back and forth between a user and agent, where the agent's actions are chosen in response to a model of current user needs within the current conversation, using both short- and long-term knowledge of the user". They introduced a theoretical framework that illustrates the action space of a conversational search agent and considered this framework to be especially sensible in multi-item faceted search. Complementing this theoretical basis, other work provides practical contributions by answering the question of how people use such systems. Kiesel et al. [20], for example, studied the potential of various real-life situations for conversational argument search. In an online user study (N = 500), they found that users are more likely to use conversational search at home than in public settings. Trippas et al. [35], on the other hand, focused more on behavioral aspects in a conversational search situation. In user studies, they observed two participants at a time, where one acted as the user and the other one mimicked the search system. Their results show that the more complex the information need, the longer it took the pairs to finish the search [35]. Furthermore, they observed that some users issued their query as a thought-out phrase, whereas others used several sentences to paraphrase their information need. Using a similar setup of observing human-human dialogues, Trippas et al. [36] derived four high-level design guidelines for conversational search systems in general web search, concluding that such systems should be transparent, include active search assistance, engage in grounding, and facilitate easy navigation through the search space. Luger and Sellen [22] performed in-depth interviews with 14 regular users of conversational agents (e.g., Siri or Alexa) to identify barriers and opportunities for conversational agents. Overall, their participants' expectations were not aligned with the systems' capabilities, leading both to frustration and hindering long-term usage. Similar observations have been made in long-term user studies [9] and in studies with chatbots [18]: Users expected the system to retain information throughout a conversation or even beyond and to understand a broad vocabulary.

As early as 1968, in studies of library patrons, Taylor [32] found that library visitors relied on librarians to resolve their vague information needs if they had trouble expressing their needs. Librarians then engaged in dialogues with the visitors to identify the exact information needs. Clarifying questions can help to resolve complex information needs and lead to better search results (see [39, 40]). Clarifying questions can improve the search performance by asking questions about a product or about product attributes [1, 5].

Other research has focused on the usability of conversational systems. Fergencs and Meier [13], for example, conducted a controlled interactive information retrieval experiment to compare chatbots and graphical search user interfaces in terms of engagement and usability. They found that current conversational systems are still prone to errors and require flawless usability to provide additional value in information retrieval tasks. Trippas et al. [34] pointed out that conversational systems that rely mostly on the auditory channel have to become proactive collaborators in the conversation to reduce complexity and cognitive load. Furthermore, conversational search systems lead to faster task completion times and better usability when they are able to preserve contextual information [12] and should take an active role by suggesting query reformulations and providing summaries of available options [11]. These studies illustrate that conversations are a means to cope with the complexity and vagueness of information needs and might, thus, help to better understand users' information needs.

2.3 Conversational Analysis

Conversational analysis is a methodology that facilitates an in-depth understanding of how conversations work. Conversations consist of intervowen utterances that form the dialogue [24]. Since 2012, the ISO standard 24617-2 for dialogue annotation exists to analyze the structure and semantic of conversations. Bunt et al. [8] described the 26 general-purpose dialogue acts (e.g., *inform*, *question*, *answer*) and 30 communication dimension dialogue acts (e.g., *pausing*, *insteraction structuring*, *init-greeting*) of the ISO standard. Novielli and Strapparava [26] compiled a comprehensive overview of works that used dialogue acts as conversational analysis method, demonstrating, amongst other things, how to utilize these in automatic conversation systems. In the field of conversational search, dialogue acts are frequently used and adapted to particular settings. Azzopardi et al. [3] proposed an experimental framework for analysing search conversations, containing agent-dependent acts (e.g., *interrogate* for the user, or *explain* for the search agent) and agent-agnostic acts (e.g., *reveal*, *inquire*). Kiesel et al. [20] built on this framework to analyze action frequencies in search conversations with a wizard-of-oz setup. In their study, they used four generic action categories that apply to both the agent and the user: generic actions, navigating, inquiring, and revealing. While the latter three are based on Azzopardi et al.'s framework [3], Kiesel et al. [20] added "generic actions" to cover conversational conventions such as greetings. In the same line of research, Qu et al. [29] developed a more detailed list of labels for search conversation utterances. Their schema contains 12 user intent categories that apply to generic actions, such as greetings, as well as to search-related utterances, like original questions and follow-up questions.

In recent years, virtual shopping assistants have received attention. Yan et al. [38] defined six generic action states that can arise during conversation with a shopping assistance bot: Question-answering, proactive questioning, comparison, recommendation, opinion summary, and chit-chat. Ivanovic [17] developed dialogue acts for written dialogues in a chat with shopping assistants. Similar to the ISO standard 24617-2, their acts are separated into content-related acts (e.g., *open question*) and conversation-related acts (e.g., *thanking*). They use their dialogue act schema to annotate and predict turns in a chat conversation.

2.4 Summary

Searching for products online can be cumbersome for users who do not know what exactly they are looking for [4, 23] or how to accurately describe what they are looking for [16, 19, 21, 33]. The literature proposes natural language interactions to overcome these challenges [1, 25, 28, 40]. However, voice assistants and conversational search systems are not yet sufficiently refined to be accepted by the users [9, 13, 18, 37]. Therefore, we need to improve the design guidelines for product search systems to better align users' expectations with the system's behavior. Previously, research has studied conversational settings by simulating conversational search systems with humans (e.g., in general search [35]) and by recording and analyzing conversations (e.g., in argument search [20]). Although existing research provides valuable insights into search dialogues, explicit high-level design guidelines for conversational product search systems that function as an expert consultant do not yet exist.

3 STUDY DESIGN

Our research is driven by the following question: *Which strategies do experts and customers in product search sessions employ to define and find suitable products?* In an exploratory user study, we collected recordings of consultation sessions in laptop and recipe search and analyzed the dialogues concerning their conversational structure and content. Similar to in-store consultations, an expert supported a participant in finding a suitable product online.

3.1 Scenario and Task

Before starting the consultation sessions, participants in the recipe domain were prompted with the following task:

> *Imagine two good friends will be visiting next weekend. You decide to cook dinner for the group. Please think about the dinner you want to cook.*

The scenario for the laptop domain was formulated as follows:

> *Imagine your laptop broke down yesterday. Please think about the laptop with which you want to replace your broken one.*

After reading the scenario prompts, participants were asked to find a suitable recipe or laptop with the help of an expert. Experts were aware of the task.

3.2 Apparatus and Procedure

The consultation sessions were conducted in German via the video communication program Zoom[1]. Due to travel and meeting restrictions imposed by the COVID-19 pandemic, we were unable to conduct the study in an in-person setting. Participants started the study by completing an online questionnaire collecting details on demographics and domain knowledge. After reading the scenario, participants entered the consultation sessions. The study leader introduced participant and expert and repeated the task instructions. The video feed was turned off during the consultation dialogue and only the audio was recorded. Experts were instructed to find a suitable recipe or laptop on the internet (they were free in their choice of search engine and product websites) and send a link to the product via the chat. Similar to in-store settings, the experts were allowed to send multiple links, but were asked to keep the links sent to a minimum. The consultation sessions were finished once the participants were satisfied with a recipe or laptop suggested by the expert. Each expert conducted three consultation sessions. The questionnaire and the task descriptions are available in their original (German) and translated form (English) online[2].

3.3 Experts

Overall, five recipe experts and three laptop experts took part in the study. The **recipe experts** (female (f) = 3, male (m) = 2, non-binary (nb) = 0), recruited via the author's personal contact network, had diverse backgrounds in cooking and recommending food. Two experts worked in the restaurant industry where they regularly recommend dishes to customers. One expert had completed training for educational cooking, and the other were experienced hobby cooks. All experts had previously been asked to recommend recipes by their families and friends. We recruited the *laptop experts* through a manually compiled emailing list that contained a total of 50 companies selling laptops in-store in Germany. The three laptop experts (f = 0, m = 3, nb = 0) had several years experience (between 5 and 12 years) working as laptop salespersons and usually spend around 20-30% of their working time on consultations. On average, a session with an expert took 1.5 hours. All experts received financial compensation of 45 Euros.

3.4 Participants

We recruited a convenience sample of 24 volunteers who responded to the word-of-mouth advertising in the authors' social and professional networks. 15 volunteers (female (f) = 7, male (m) = 8, non-binary (nb) = 0) took part in the recipe condition and nine (f = 4, m = 5, nb = 0) were consulted on finding a new laptop. A single consultation session

[1]https://zoom.com,lastaccessed06-10-2021
[2]https://git.gesis.org/papenmaa/chiir22_conversationstrategies

took between 10 to 15 minutes. Participants were on average M = 33 years old (SD = 6 years) and self-assessed their knowledge of the product (laptop or recipe, respectively) on a scale from 1 = "do not agree at all" to 5 = "completely agree". In the recipe domain, participants indicated to have slightly more domain knowledge (M = 2.9, SD = 0.9) than in the laptop domain (M = 2.7, SD = 1.6). Most participants in the recipe domain had recently searched for a recipe (14 within 6 months, 1 within last year). In the laptop domain, the last search for a laptop was a longer time ago (3 within 6 months, 1 within last year, 5 longer than a year ago). Participants did not receive any type of compensation for participating in the study.

3.5 Transcription

Two experienced transcribers interpreted the audio files of the consultation sessions using the transcription guidelines for German conversations of Dresing and Pehl [10]. The transcribers were instructed to note down the turns and actors (anonymized), the spoken words, interjections (e.g., "mhm (affirmative)"), intonation at sentence endings (high (↑) and low (↓) pitch), overlapping utterances ("//(...)//"), and pauses.

3.6 Annotation

First, to investigate the topics of the conversations, we developed schemata of themes and topic categories in a bottom-up coding manner (see [2, 6]) for each domain (recipe, laptop). Two annotators read and coded the recipe and laptop conversations. In two discussion workshops, the annotators unified common categories and clustered categories into groups.

Second, to derive insights about the conversational structure of the conversations, we adapted the ISO standard 24617 as described by Bunt et al. [8]. The ISO standard 24617 is a generic schema that is domain independent. Since we did not know which dialogue acts will occurr in our experimental user study, we chose this generic standard. Besides the 26 general-purpose acts, we included the four feedback acts (e.g., *auto-positive*, *allo-positive*), and one discourse structuring act. Additionally, we added "opening" as a representation of social obligations acts that are connected to the opening and closing of a conversation. We further added the "think-aloud" act for utterances that were caused by the setup of our experiment, i.e., verbalizations of actions that the other could not see, such as searching for the chat window to paste the link to a product. We report the final schemata in Section 4.1.

When annotating the conversation transcripts with content codes and dialogue acts, we recruited two experienced annotators. Each transcript was annotated by one annotator after ensuring that both annotators had a similar understanding of the schemata. The annotators started with a training phase with a subsequent workshop in which difficult cases were discussed. We then verified the agreement between annotators on one recipe transcript and one laptop transcript.

For annotations with dialogue acts, we found a substantial agreement between both annotators (Cohen's kappa of 0.674) with 75% of words being ascribed to the same dialogue act by both annotators. The agreement on topic labels was similarly substantial (Cohen's kappa of 0.731 for the recipe conversation and 0.764 for the laptop conversation), with a percentage overlap of 78% in the recipe domain and 80% in the laptop domain. As the inter-annotator agreement on the first laptop and recipe conversation was sufficiently high, the remaining conversations were annotated by one annotator each.

4 RESULTS

With 15 conversations about recipes and nine conversations about laptops, we analyzed a total of 24 consultation sessions. In both domains, experts took a prominent role during the conversations, being the active speaker most of the time (68% of words in the recipe domain, 78% in the laptop domain). Conversations about recipes were on average shorter (M = 1005 words, SD = 392 words) than those in the laptop domain (M = 1856 words, SD = 959 words). The information needs were diverse in both domains. Participants were asking for both simple recipes (e.g., pasta with sauce) and advanced recipes (e.g., a two-course menu with wine suggestions). In the laptop domain, five participants needed a mid-range computer for text editing and surfing, while four needed a high-end computer for software engineering, gaming, or business purposes. After discussing the participants' information needs, experts pasted a link to a product in the chat. After the link was sent, the conversations continued: Experts justified their choice and participants asked questions about the product. Experts could send another link if needed. In nine of 15 recipe conversations, a second link was sent (six contained an alternative to the first recipe, while three provided an addition to the first recipe, e.g., a side dish or dessert). Around 59% of the recipe conversations took place before the first link was sent. In all conversations in the laptop domain, only a single link was sent, and 70% of the conversations took place before exchanging the product link. The conversations were finished once the participants were satisfied with the proposed product. Despite the diverse information needs, all experts succeeded in choosing suitable products for the participants. One recipe conversation seemed to be an outlier with the expert having a specific recipe in mind but being unable to find the link. In that case, the expert and participant agreed that the link will be sent later and finished the conversation at that point, but the participant confirmed that the recipe described by the expert would fit his needs.

In the following subsections, we discuss the findings of analysing the content and the conversational structures of the conversations in more detail.

4.1 Schemata

The bottom-up coding of the conversations with regard to content led to a categorical schema with two hierarchical levels for each domain. Table 1(a) presents the seven main categories of the recipe conversations. Table 1(b) describes the eight main categories of the laptop conversation content schema. For conciseness, we show only the main categories, for the full schemata, the reader is referred to the accompanying online repository[4].

Building on the work of Bunt et al. [8], Table 1(c) shows the main dialogue acts annotated at least once. For a better overview, we collapsed dialogue acts of the same type (e.g., all question types into "Question", all request types into "Request"). Additionally, as explained in Section 3.6, the "Think-aloud" act was introduced to label utterances related to the technical setup of the experiment.

4.2 Conversation Analysis: Content

Figure 1 shows the **distribution of topic labels** over all conversations. In the recipe domain, conversations mostly revolved around the type of dish (e.g., from which culture the dish should be, or whether it is a stew, salad, or composed dish), the ingredients to include or exclude, and the preparation effort, as well as general instructions. In laptop conversations, the predominant topics are the hardware (e.g., the processor, display, performance) and the purpose (i.e., what the laptop will be used for).

[4]https://git.gesis.org/papenmaa/chiir22_conversationstrategies

(a) Category	**Explanation**
Premises	Information on the cook (e.g., preferences) and the physical setup (e.g., available kitchen tools)
Dish-type	Information on the composition, culture, level of satiation, and main flavor of the dish
Ingredients	Information on ingredients to exclude (e.g., diets and allergies) and include
Preparation	Information concerning the required effort, mode of serving, and preparation instructions
Scale	Information on the serving size and number of courses
Recipe	Information on the presentation of the recipe itself (e.g., pictures or background stories)
Assessment	Statements and comments of speaker's opinion on a recipe (e.g. *"I like this recipe."*)
(b) Category	**Explanation**
Meta	Information on the brand, series, model, and general type of device (e.g., convertible)
Purpose	Information on the purpose or desired applicability of a laptop (e.g., for work, for gaming, for writing emails)
Outer	Information on visual aspects (e.g., color, size, weight, and the robustness)
Software	Information on the operating system and software that the laptop needs to be able to run
Hardware	Information that concern the (functional) hardware of a laptop (e.g., performance, CPU, display, webcam)
Experience	Information that can only be acquired through interaction with the laptop (e.g., usability, lifespan, quality)
Purchase	Information on the process of buying a laptop (e.g., price, discounts, seller, warranty)
Additional	Information that concern additional products with the purchase (e.g., screen protector, laptop bag)
(c) Dialogue Act	**Explanation**
Interaction Structuring	Utterances to structure the conversation
Opening	Utterances that indicate the beginning or end of the conversation
Question	Utterances that aim to acquire information from the conversation partner
Inform	Utterances that serve to inform, explain, answer, justify, elaborate or make statements
Agreement	Utterances that agree and disagree with a previously made statement or plan
Confirmation	Utterances that explicitly answer a closed question
Suggest	Signifying that one wants the other to consider a proposition that concerns both conversation partners
Offer	Signifying that one wants the other to consider an offer that most of the times concerns only the speaker
Request	Signifying that one uses to ask the other to do something
Auto	The speaker believes to have (in)correctly understood what has been said before, including literal repetition
Allo	The speaker (dis)confirms that the other one has correctly understood what has been said before
Think-aloud[3]	Utterances about actions related to the task (category added by the authors, not present in the original scheme of [7])

Table 1. Coding schemata used in the user study to analyze the content in the recipe (a) and laptop (b) domain and the conversational structure (c) in both domains.

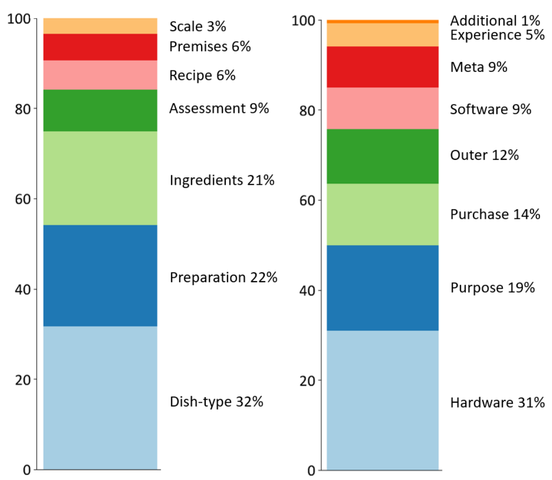

Fig. 1. Distribution of topic labels in recipe (left) and laptop (right) conversations.

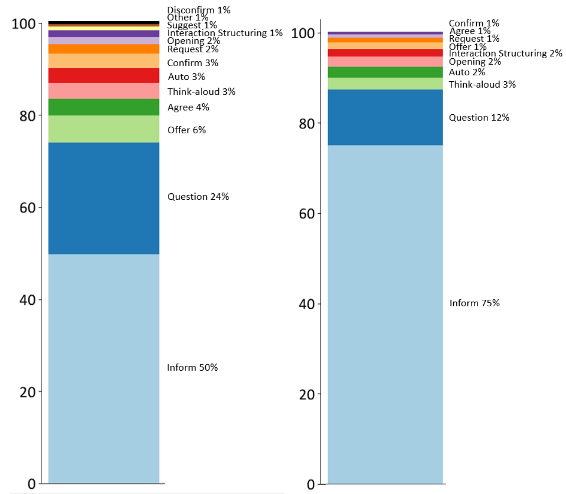

Fig. 2. Distribution of dialogue acts in recipe (left) and laptop (right) conversations.

4.3 Conversation Analysis: Dialogue Acts

Besides the topics appearing in the conversations, we analyzed the **structure of the conversation with dialogue acts**. Around 20% of the words (20% in the recipe domain, 18% in the laptop domain) did not carry information about the product but were essential to the conversation, e.g., opening the conversation, showing understanding, asking for a short break. Figure 2 depicts the shares of dialogue acts in the conversations in both domains. We used single words as base unit and present the highest categories of dialogue acts (e.g., “request”, “request_address”, “request_accept” and “request_decline” are all summarized as “request”). In both domains, the “inform” act takes the greatest share of the conversations, followed by “question”.

A dialogue act that appears frequently throughout all conversations is the **“auto” act** [8], a discourse marker that communicates feedback of a speaker’s own (auto) understanding of what was said before. We identified seven types of auto acts, which are summarized in Table 2. The auto feedback can be expressed either with words (e.g., *“yes”*, *“I see”*), with interjections (e.g., *“mhm”*), or literal repetition of what the other one said (e.g., participant: *“More simple things. (...)”*, expert: *“Simple things, okay. (...)”*). It should be noted that auto acts were mostly used by experts in both domains (82% in the recipe domain and 70% in the laptop domain), with *“mhm”* being the only auto act that was uttered equally by experts and participants. Auto acts comprised 2% of the words in the recipe conversations, and 3% of the words in the laptop conversations (see Figure 2). Contrary to other utterances, auto acts often comprised only a single word and therefore underrepresented in the word count. Taking utterances as a basis, auto acts made up 25% of the utterances in the laptop conversations, and 22% in the recipe conversations. On average, a conversation contained eleven auto utterances (same in both domains).

To investigate the roles of the expert and the participant in the conversations, we analyzed the **distribution of dialogue acts per actor**, using utterances (not words) as basis. Comparing the speech acts over the complete course of a conversation did not show notable differences between experts and participants. Looking at the distribution at parts of the conversation, however, shows a clear change in roles. In both domains, experts ask more questions before

Table 2. Types of auto-positive discourse markers and absolute occurrences in the 15 recipe and 9 laptop conversations.

Discourse marker	Recipe	Laptop
"okay"	72	27
"yes"	25	36
"mhm" (affirmative)	24	14
"all right"	17	10
"great"	9	4
"I see"	2	0
literal repetition	12	7
total	161	98

Table 3. Types, actors and topics of first three questions in the conversations.

	Recipe	**Q1**	**Q2**	**Q3**	**Laptop**	**Q1**	**Q2**	**Q3**
Types	Set	13	9	7	Set	9	6	3
	Choice	1	3	5	Choice	0	1	0
	Propositional	1	2	2	Propositional	0	0	3
	Check	0	1	1	Check	0	2	3
Actors	Expert	15	14	15	Expert	9	9	9
	Participant	0	1	0	Participant	0	0	0
Topic	Ingredients	4	8	6	Purpose	7	6	5
	Dish-type	6	3	7	Hardware	0	2	2
	Premises	4	1	1	Experience	2	0	0
	Preparation	1	3	1	Purchase	0	1	1
					Outer	0	0	1

recommending a product than they did afterward (recipe domain: 27% of utterances before, 13% of utterances after sending a link; laptop domain: 17% before, 12% after). Contrarily, experts make more informing utterances after they chose a product than they did before (recipe domain: 17% before, 24% after; laptop domain: 38% before, 49% after). Contrarily, participants make more informing statements before receiving the product recommendation than afterward (recipe domain: 39% before, 19% after; laptop domain: 45% before, 17% after) and, at least in the laptop domain, ask more questions after receiving the link than before (recipe domain: 7% before, 6% after; laptop domain: 3% before, 6% after). In summary, the roles of the actors changed over the course of the conversation, from a proactive expert that elicits information about the information need to an expert that informs and reacts to the questions of the participants. Conversely, by trend, participants became more proactive after they received the link. As the effects cancel out on the level of a whole conversations, the roles of the actors appeared to be similar at first.

A detailed analysis of the **questions in the beginning of the conversation** (see Table 3) revealed that the majority of questions in the beginning are formulated as open questions. Focusing only on the first question in a conversation confirms this observation. 13 of 15 recipe conversations start with a *set* question (open questions, e.g., *"What are your food preferences?"*), one with a *propositional* question (yes or no questions, e.g., *"Do you cook often?"*), and one with a *choice* question (A or B questions, e.g., *"Do you usually cook often or rather less frequently?"*). In the laptop domain, all nine consultation conversations initially start with a *set* question from the expert (e.g., *"What is most important to you* *when it comes to laptops?”*). All first questions were asked by the experts. In one recipe consultation conversation, a participant asked a question within the first three questions of the conversation. After explaining that he liked couscous and how he usually cooks with couscous, the participant asked the expert: *“Are there alternatives to how I usually prepare couscous?”*. However, in all other cases, the experts took the proactive role of steering the conversation with their questions.

Regarding the content, the experts asked the first question to gather information about the context and general requirements. In the recipe domain, the first question asked about the general direction of the dish, intolerances and other constraints, or the main ingredient. Subsequently, more detailed questions were asked about specific ingredients or the mode of preparation. Interestingly, most of the ingredients questions in the beginning (16 out of 18 questions) were asking for which ingredients to exclude, rather than which to include. The laptop experts mostly asked about the purpose of the laptop (*“What do you want to use the laptop for?”*) and prior experiences (*“For how long did you use your last laptop?”*) in their initial question. Afterward, they continued to explore specific purposes, such as portability, or asked about hardware requirements. We also noted that the expert questions in the beginning served a different role than the participant questions in the end. While experts asked questions to find out about participants’ information needs (e.g., *“What do you need your laptop for?”*), participants asked questions to judge whether the given product is a good fit for their needs (e.g., *“Are the quantities [mentioned in the recipe] enough for three people?”*).

5 DISCUSSION

In this final section, we highlight important findings and discuss five design implications for product search assistants that emerge from our analyses of two product domains (recipe and laptop). Despite the diversity of domains and information needs, we observed common strategies in the conversations. We start with our findings on how experts elicited participant information needs before discussing conversational features that we observed in the consultation sessions.

5.1 Eliciting the Genuine Information Need

We observed two domain-independent strategies and one domain-dependent strategy, which experts employed to elicit the users’ genuine information needs. First, experts took a proactive role in the beginning and a reactive role towards the end of conversations in both domains. The large majority of the first three questions in a conversation was asked by the experts (71 of 72 questions). Experts asked more questions before recommending a product than they did afterward, but informed more after recommending a product than they did before. On the contrary, participants made more informing statements before receiving the link to a product than afterward. At the beginning of the conversations, experts proactively ask about the participants’ information needs, which confirm the consensus in the literature that conversational search systems should be more proactive [11, 12, 34, 36]. Trippas et al. [36], for example, state that being proactive and providing search assistance should be considered in the design of conversational assistants for general web search. Whereas prior research found that a proactive conversational search system can improve the search outcome, our findings indicate that proactive interaction would also establish a more natural form of interactivity. The proactive role of the expert also shows that a system can actively support the user in clarifying the information need. When designing a search system, we need to think about questioning strategies of the system to elicit these needs, as proposed by Zhang et al. [41], e.g., with clarifying questions [1, 39, 40]. Our findings also indicate that an agent’s level of proactivity can change in the course of a conversation, which supports Radlinski and Craswell’s [30] suggestion of mixed-initiative dialogues, that is, dialogues in which both interaction partners can take an active role. From this discussion, we derive the first design implication for search assistants:

DI1: Product search assistants should be able to take a "proactive" and "reactive" role and should transition from "proactive" to "reactive" throughout a search session.

Our findings show that experts use funneling strategies to elicit the participants' information needs and narrow down the search space step by step throughout the conversation. In the recipe domain, the experts initially asked about general information (general composition, level of knowledge, preferences), before continuing with questions more tailored to the participant (e.g., specific ingredients and preparation modes). Similarly, in the laptop domain, initial questions concerned the general purpose of the machine, while later questions asked about specific details such as individual hardware components. The very general questions in the beginning served to understand the context and get initial information before asking more detailed questions. Interestingly, in the laptop domain, current search systems do not support filtering for the purpose [28], although this seems to be the most important information for experts to reduce the search space. The funneling strategy also showed in the type of questions that experts asked in the beginning. The first question, asking about general information, was formulated as a set question that elicits an open answer. Participants were free to answer in as much detail as they seemed suitable. Later, questions were more and more often formulated in a closed manner (choice, propositional, or check questions). With closed questions, experts strategically reduced the search space and addressed aspects that the participants did not yet mention in their answers to the open questions in the beginning. Current search systems often offer a search bar to enter keywords and facets to parametrize the search. However, the facets concern usually the "hard facts" [28], equivalent to specific, closed questions asked by the experts. Facets that correspond to the general, open questions, are missing. This search flow from broad to detailed could be solved by conversational search, as proposed by Radlinski and Craswell [30]. We conclude with a second design implication:

DI2: Product search assistants should make use of funneling strategies, going from broad to specific, to help the users narrow down their information need.

Besides those two findings that concern search assistants in general, we observed a domain-specific strategy for eliciting the genuine information needs of participants. In the recipe domain, experts used a technique that differs from typical filtering mechanisms. Instead of asking which ingredients to include, they asked which ones to exclude (e.g., due to intolerances or personal dislike). Although these questions did not notably reduce the search space, all recipe experts asked this question at some point. A possible explanation could be the potential consequences of a recipe that includes a deal-breaker. If someone is allergic to one of the ingredients, the proposed recipe will be rejected, regardless of how well other factors match the requirements. We therefore suggest a domain-dependent design implication:

DI3: Recipe search assistants should account for potential deal-breakers by offering to explicitly exclude ingredients.

This design implication might be applicable in other domains as well, if they contain similar deal-breaking aspects. We did not see a comparable strategy in the laptop domain, which might be due to less severe consequences in case requirements are not met.

5.2 On the Relevance of Conversational Features

Besides analysing how experts elicit information needs, we observed repeated occurrences of linguistic features that led to two additional design guidelines. First, a substantial part of the conversations (18-20%) was conversational structure and did not convey explicit information about the search target. Even though those utterances did not address the product search, they were an important part of the search process, e.g., by signaling the beginning or ending of a conversation, exchange opinions, or introduce an action:

expert: *"I think this fits well with your cooking habits (↓)"*
participant: *"Yes (↓) This is a recipe I would like to cook (↓)"*
expert: *"Okay, so we have reached our goal (↓)"*
participant: *"Perfect (↓)"*
expert: *"Alright (↓)"*[5]

Besides the opening and closing of conversations, both the experts and the participants used dialogue-structuring utterances such as *"wait a second"* or *"another question"*. These elements inform the interaction partner about the intended progress of the interaction, which can help to reduce misunderstandings and overlapping turns: expert: *"I'll be taking a minute to search for a laptop for you (↓)*
I'll be back in a minute (↓)"
participant: *"Okay, yes (↓)"*
(silence of 01:01 minutes)
expert: *"Alright (↓)"*

If users interact with a search assistant, they might be unsure about how the system functions and hence how the dialogue will proceed in the future. To increase transparency and feedback on the system's planning of the dialogue, we suggest to include interaction-structuring elements in product search assistants. We conclude with a forth design guideline:

DI4: Product search assistants should make use of dialogue-structuring acts.

Secondly, we observed a repeated usage of "auto" acts such as *"mhm"*, *"yes"*, and *"okay"*. Although such utterances did not make up a great percentage of the spoken words, they did occur often throughout the conversations, accounting for roughly one forth of the utterances. In our experiment, the majority of auto acts carried a confirmative meaning to signalize that one thinks to have understood what the other one was saying. Those discourse markers took a small but vital role in consultation sessions. On the one hand, experts used them to signify that they understood participants' descriptions of requirements for the products. On the other hand, participants reflected with auto-positives that they understood the explanations of the experts. Previous research has shown that people with different levels of domain knowledge (i.e., experts and laypersons, or retailers and customers) use different vocabularies to describe a product [21, 31]. Especially in those cases, it is vital to reflect whether something has been understood. In our experiment, auto acts were mainly used by experts, for example:

participant: *"I am mainly using it for university (↑)"*
expert: *"//mhm (↑)//"*
participant: *"//So,// writing essays (↑)"*
expert: *"mhm (↑)"*

[5] "(↓)" marks a low intonation and "(↑)" marks a high intonation.

participant: *"Everything that has to do with Word, Excel (↑)"*
expert: *"mhm (↑)"*

Experts, hence, seemed to believe that participants needed this type of confirmation and reflection of their state of attentiveness.

DI5: Product search assistants should use auto acts to reflect their understanding of the user's input and recognize whether the user has understood the system.

Both observations above (DI4 and DI5) match with findings made by Vtyurina et al. [37] and Frummet et al. [15] in the cooking domain: Conversational cues like (*"okay"*, *"mhm"*) are implicitly used as linguistic feedback, i.e., grounding, clarification, and confirmation. Trippas et al. [36] likewise considered grounding as means of disambiguating information needs. Product search assistants should, thus, be capable of recognizing and inserting these cues which take a relevant role in user-system interaction. Auto acts could also be utilized to adapt the assistant's behavior, for example to adapt its vocabulary to include less technical terms when the user does not show auto positive cues during an explanation.

5.3 Summary and Outlook

We analyzed which strategies experts and customers employ to define and find suitable products that satisfy the customers product needs (**RQ**). We first analyzed what constitutes a product consultation session. In both domains, consultation conversations show a generic structure that includes an opening phase with greetings and personal introduction as well as an ending phase with thanking and greetings. In between, the search task was executed. Experts first elicited the information needs before participants asked questions about the recipe or laptop that the expert recommended.

Moreover, we identified strategies for information need elicitation that experts showed in our experiment. We identified the proactive role of the expert (DI1) and the funneling strategy using broad, open questions in the beginning and more nuanced, closed questions in the course of the conversations (DI2). For the recipe domain, we observed filtering by exclusion (DI3) that applies for deal-breaking aspects such as ingredients that participants were allergic to. Furthermore, we found that the consultation sessions profited from conversational features such as dialogue-structuring acts (DI4) and auto acts to signalize confirmation of understanding (DI5). We found that the large majority of strategies and conversational features were constant across domains. Only one out of five design guidelines (DI3) was only observed in one of the two domains. Our experiment included two products that are quite different. However, even though the recipe and the laptop domain do not share many common attributes, we still observed conversational strategies and conversational features that were common in both domains. We are therefore optimistic that our guidelines can be generalized to other product domains as well.

Our work is subject to several limitations. First, the experts in the recipe domain did not have a homogeneous background and, unlike the laptop experts, did not all consult customers about recipes as their profession. However, for our experiment, we recruited persons whose level of domain knowledge is above average and likely to be above that of our participants. Furthermore, all consulting sessions were successful and all participants were satisfied with the product they were recommended in the end. We therefore expect that the influence of the heterogeneous background of recipe experts did not have a negative influence on the consulting sessions.

Second, the sample size of our experiment is rather small (N = 24) and not representative. As we evaluated the conversations in a qualitative manner, we assume that the sample size was sufficient to support our findings. Nevertheless, the design guidelines should be evaluated in a quantitative study with a bigger sample group. In future work, we plan to implement a conversational product search system that explicitly models our four domain-independent design guidelines (DI1, DI2, DI4, DI5) and investigate whether they improve usability (efficiency, effectiveness, satisfaction) as compared to a system that does not make use of these guidelines.

6 CONCLUSION

In the present work, we investigated how experts consult customers about products in product search. In 24 online conversations similar to a sales session in on-site stores, we recorded the dialogues between an expert and a client to leverage their conversational strategies for the design of future conversational product search assistants. We analyzed the conversations with respect to their content and their conversational form. We derived five design guidelines for conversational product search systems: (1) Systems should be able to proactively ask questions about the user's information need as well as reactively answer questions about suggested products. (2) Systems should narrow down the user's information need by employing funneling strategies. (3) Systems should support interaction structuring, e.g., greetings, thanking, and introducing future actions. (4) Systems should communicate their understanding of what the user communicates and recognize signifiers of understanding in the user's reactions. Additionally, (5) in the recipe domain, systems should be able to explicitly exclude ingredients. With the derived design guidelines, we extend the insights into conversational product search and offer user-centered guidelines for the design of conversational systems.

ACKNOWLEDGMENTS

We would like to thank our anonymous reviewers for their encouraging and helpful feedback on our manuscript. This work was partly funded by the DFG, grant no. 388815326; the VACOS project at GESIS.

REFERENCES

[1] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 2019. Asking Clarifying Questions in Open-Domain Information-Seeking Conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 475–484. https://doi.org/10.1145/3331184.3331265

[2] Carl Auerbach and Louise B Silverstein. 2003. *Qualitative data: An introduction to coding and analysis*. Vol. 21. NYU press.

[3] Leif Azzopardi, Mateusz Dubiel, Martin Halvey, and Jeffery Dalton. 2018. Conceptualizing agent-human interactions during the conversational search process. In *The Second International Workshop on Conversational Approaches to Information Retrieval (CAIR'18)*.

[4] Marcia Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online Review* 13(5) (1989), 407–424. https://doi.org/10.1108/eb024320

[5] Keping Bi, Qingyao Ai, Yongfeng Zhang, and W. Bruce Croft. 2019. Conversational Product Search Based on Negative Feedback. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (Beijing, China) *(CIKM '19)*. Association for Computing Machinery, New York, NY, USA, 359–368. https://doi.org/10.1145/3357384.3357939

[6] Virginia Braun and Victoria Clarke. 2013. *Successful qualitative research: A practical guide for beginners*. sage. https://books.google.de/books?id=nYMQAgAAQBAJ

[7] Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010. Towards an ISO Standard for Dialogue Act Annotation. In *Seventh conference on International Language Resources and Evaluation (LREC'10)*. La Valette, Malta. https://hal.inria.fr/inria-00544997

[8] Harry Bunt, Volha Petukhova, David Traum, and Jan Alexandersson. 2017. *Dialogue Act Annotation with the ISO 24617-2 Standard*. Springer International Publishing, Cham, 109–135. https://doi.org/10.1007/978-3-319-42816-1_6

[9] Minji Cho, Sang-su Lee, and Kun-Pyo Lee. 2019. Once a Kind Friend is Now a Thing: Understanding How Conversational Agents at Home Are Forgotten. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS'19)*. Association for Computing Machinery, New York, NY, USA, 1557–1569. https://doi.org/10.1145/3322276.3322332

[10] Thorsten Dresing and Thorsten Pehl. 2015. *Praxisbuch Interview, Transkription & Analyse*. Dr. Dresing und Pehl GmbH, Marburg. 72 pages.

[11] Mateusz Dubiel, Martin Halvey, Leif Azzopardi, Damien Anderson, and Sylvain Daronnat. 2020. Conversational strategies: impact on search performance in a goal-oriented task. In *The Second International Workshop on Conversational Approaches to Information Retrieval (CAIR'18)*.

[12] Mateusz Dubiel, Martin Halvey, Leif Azzopardi, and Sylvain Daronnat. 2018. Investigating how conversational search agents affect user's behaviour, performance and search experience. In *The Second International Workshop on Conversational Approaches to Information Retrieval (CAIR'18)*.

[13] Tamás Fergencs and Florian Maximilian Meier. 2021. Engagement and Usability of Conversational Search – A Study of a Medical Resource Center Chatbot. In *Proceedings of iConference 2021*.

[14] Weixing Ford, Yixiu Li, and Jie Zheng. 2021. Numbers of bricks and clicks: Price competition between online and offline stores. *International Review of Economics & Finance* 75 (2021), 420–440. https://doi.org/10.1016/j.iref.2021.04.027

[15] Alexander Frummet, David Elsweiler, and Bernd Ludwig. 2019. Detecting domain-specific information needs in conversational search dialogues. In *Proceedings of the 3rd Workshop on Natural Language for Artificial Intelligence*.

[16] George Furnas, Thomas Landauer, Louis Gomez, and Susan Dumais. 1987. The Vocabulary Problem in Human-System Communication. *Commun. ACM* 30, 11 (Nov. 1987), 964–971. https://doi.org/10.1145/32206.32212

[17] Edward Ivanovic. 2005. Dialogue Act Tagging for Instant Messaging Chat Sessions. In *Proceedings of the ACL Student Research Workshop* (Ann Arbor, Michigan) *(ACLstudent'05)*. Association for Computational Linguistics, USA, 79–84.

[18] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS'18)*. Association for Computing Machinery, New York, NY, USA, 895–906. https://doi.org/10.1145/3196709.3196735

[19] Makoto P. Kato, Takehiro Yamamoto, Hiroaki Ohshima, and Katsumi Tanaka. 2014. Cognitive Search Intents Hidden behind Queries: A User Study on Query Formulations. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW'14 Companion)*. Association for Computing Machinery, New York, NY, USA, 313–314. https://doi.org/10.1145/2567948.2577279

[20] Johannes Kiesel, Kevin Lang, Henning Wachsmuth, Eva Hornecker, and Benno Stein. 2020. Investigating Expectations for Voice-Based and Conversational Argument Search on the Web. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval* (Vancouver BC, Canada) *(CHIIR'20)*. Association for Computing Machinery, New York, NY, USA, 53–62. https://doi.org/10.1145/3343413.3377978

[21] Aarno Lehtola, Johannes Heinecke, and Catherine Bounsaythip. 2003. Intelligent human language query processing in MKBEEM. In *Proceedings of the International conference on Universal Access in Human-Computer Interaction (UAHCI'03)*. 22–27.

[22] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI'16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[23] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (2006), 41–46. https://doi.org/10.1145/1121949.1121979

[24] Jacques Moeschler. 2001. Chapter 12. Speech act theory and the analysis of conversation. In *Essays in Speech Act Theory*. John Benjamins, 239–261. https://www.jbe-platform.com/content/books/9789027298157-pbns.77.15moe

[25] Magnus Niemann, Malgorzata Mochol, and Robert Tolksdorf. 2009. Smart Shop Assistant: Using Semantic Technologies to Improve Online Shopping. In *Proceedings of the Second Future Internet Conference on Future Internet* (Berlin, Germany) *(FIS'2009)*. Springer-Verlag, Berlin, Heidelberg, 106–115.

[26] Nicole Novielli and Carlo Strapparava. 2011. Dialogue act classification exploiting lexical semantics. In *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*. IGI Global, 80–106.

[27] Andrea Papenmeier, Dagmar Kern, Daniel Hienert, Alfred Sliwa, Ahmet Aker, and Norbert Fuhr. 2021. Starting Conversations with Search Engines - Interfaces That Elicit Natural Language Queries. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) *(CHIIR'21)*. Association for Computing Machinery, New York, NY, USA, 261–265. https://doi.org/10.1145/3406522.3446035

[28] Andrea Papenmeier, Alfred Sliwa, Dagmar Kern, Daniel Hienert, Ahmet Aker, and Norbert Fuhr. 2020. "A Modern Up-To-Date Laptop" - Vagueness in Natural Language Queries for Product Search. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS'20)*. Association for Computing Machinery, New York, NY, USA, 2077–2089. https://doi.org/10.1145/3357236.3395489

[29] Chen Qu, Liu Yang, W. Bruce Croft, Johanne R. Trippas, Yongfeng Zhang, and Minghui Qiu. 2018. Analyzing and Characterizing User Intent in Information-Seeking Conversations. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (Ann Arbor, MI, USA) *(SIGIR'18)*. Association for Computing Machinery, New York, NY, USA, 989–992. https://doi.org/10.1145/3209978.3210124

[30] Filip Radlinski and Nick Craswell. 2017. A Theoretical Framework for Conversational Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR'17)*. ACM, New York, NY, USA, 117–126. https://doi.org/10.1145/3020165.3020183

[31] Napawan Sawasdichai and Sharon Poggenpohl. 2002. User Purposes and Information-Seeking Behaviors in Web-Based Media: A User-Centered Approach to Information Design on Websites. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (London, England) *(DIS'02)*. Association for Computing Machinery, New York, NY, USA, 201–212. https://doi.org/10.1145/778712.778742

[32] Robert S. Taylor. 1968. Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries* 29, 3 (1968), 178–194. https://doi.org/10.5860/crl_29_03_178

[33] Arthur ter Hofstede, Henderik Alex Proper, and Theo van der Weide. 1996. Query formulation as an information retrieval problem. *Comput. J.* 39, 4 (1996), 255–274.

[34] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, Hideo Joho, and Mark Sanderson. 2018. Informing the Design of Spoken Conversational Search: Perspective Paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) *(CHIIR'18)*. Association for Computing Machinery, New York, NY, USA, 32–41. https://doi.org/10.1145/3176349.3176387

[35] Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. 2017. How Do People Interact in Conversational Speech-Only Search Tasks: A Preliminary Analysis. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Oslo, Norway) *(CHIIR'17)*. Association for Computing Machinery, New York, NY, USA, 325–328. https://doi.org/10.1145/3020165.3022144

[36] Johanne R. Trippas, Damiano Spina, Paul Thomas, Mark Sanderson, Hideo Joho, and Lawrence Cavedon. 2020. Towards a model for spoken conversational search. *Information Processing & Management* 57, 2 (2020), 102162. https://doi.org/10.1016/j.ipm.2019.102162

[37] Alexandra Vtyurina and Adam Fourney. 2018. Exploring the Role of Conversational Cues in Guided Task Support with Virtual Assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI'18)*. Association for Computing Machinery, New York, NY, USA, Article 208, 7 pages. https://doi.org/10.1145/3173574.3173782

[38] Zhao Yan, Nan Duan, Peng Chen, Ming Zhou, Jianshe Zhou, and Zhoujun Li. 2017. Building Task-Oriented Dialogue Systems for Online Shopping. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, California, USA) *(AAAI'17)*. AAAI Press, 4618–4625.

[39] Hamed Zamani, Susan Dumais, Nick Craswell, Paul Bennett, and Gord Lueck. 2020. Generating Clarifying Questions for Information Retrieval. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW'20)*. Association for Computing Machinery, New York, NY, USA, 418–428. https://doi.org/10.1145/3366423.3380126

[40] Hamed Zamani, Bhaskar Mitra, Everest Chen, Gord Lueck, Fernando Diaz, Paul N. Bennett, Nick Craswell, and Susan T. Dumais. 2020. Analyzing and Learning from User Interactions for Search Clarification. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR'20)*. Association for Computing Machinery, New York, NY, USA, 1181–1190. https://doi.org/10.1145/3397271.3401160

[41] Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. 2018. Towards Conversational Search and Recommendation: System Ask, User Respond. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) *(CIKM'18)*. Association for Computing Machinery, New York, NY, USA, 177–186. https://doi.org/10.1145/3269206.3271776