Towards a Model for Spoken Conversational Search

Johanne R. Trippas[a,*], Damiano Spina[a], Paul Thomas[b], Mark Sanderson[a],
Hideo Joho[c], Lawrence Cavedon[a]

[a]*RMIT University, Melbourne, Australia*
[b]*Microsoft, Canberra, Australia*
[c]*University of Tsukuba, Tsukuba, Japan*

Abstract

Conversation is the natural mode for information exchange in daily life, a spoken conversational interaction for search input and output is a logical format for information seeking. However, the conceptualisation of user–system interactions or information exchange in spoken conversational search (SCS) has not been explored. The first step in conceptualising SCS is to understand the conversational moves used in an audio-only communication channel for search. This paper explores conversational actions for the task of search. We define a qualitative methodology for creating conversational datasets, propose analysis protocols, and develop the *SCSdata*. Furthermore, we use the SCSdata to create the first annotation schema for SCS: the *SCoSAS*, enabling us to investigate interactivity in SCS. We further establish that SCS needs to incorporate interactivity and pro-activity to overcome the complexity that the information seeking process in an audio-only channel poses. In summary, this exploratory study unpacks the breadth of SCS. Our results highlight the need for integrating discourse in future SCS models and contributes the advancement in the formalisation of SCS models and the design of SCS systems.

Keywords: Spoken Conversational Search, Information Seeking

*Corresponding author.
  Email addresses: johanne.trippas@rmit.edu.au (Johanne R. Trippas), damiano.spina@rmit.edu.au (Damiano Spina), pathom@microsoft.com (Paul Thomas), mark.sanderson@rmit.edu.au (Mark Sanderson), hideo@slis.tsukuba.ac.jp (Hideo Joho), lawrence.cavedon@rmit.edu.au (Lawrence Cavedon)

1. Introduction

Voice-only, spoken conversational systems such as Google Home, Amazon Echo, or Apple Homepod, are becoming widely used. These systems can answer factoid questions. However, they are yet not able to engage in complex information seeking tasks where multiple turns are needed to exchange information, reformulate queries, or proactively recommend different search strategies.

Trippas et al. [53] suggested that existing information seeking models do not suffice for the increased interactivity, complexity, or the agency of SCS systems. To the best of our knowledge, only a few models include the system as an integral part of the search process [3, 33, 39, 54]. Recently, Azzopardi et al. [3] created a conceptual framework of the probable action and interaction space for conversational agents as a first step, acknowledging that their initial framework would need expansion and empirical evidence.

Understanding the communication behaviours of dialogue is crucial to SCS and many different annotation schemas have been developed [11]. These schemas are classifications of dialogues and consider an utterance as an action inside the information exchange. Even though many different domain independent annotation schemas exist, such as DAMSL [1], these were mainly used for the creation of spoken dialogue systems (SDS) or the general understanding of dialogue. Information seeking actions were not covered in depth but instead broad categories such as "answer" or "info-request" are presented. Recent work has made initial steps towards understanding such actions [32, 51, 49], however, no complete set has been developed so far.

We create a first annotation schema for SCS: the spoken conversational search annotation schema (SCoSAS). The schema reveals the different atomic actions or utterance functions and interactions taken by prospective user and system in the stages of an information seeking process. Thematic analysis was used to identify and summarise communicative activities, strategies, and challenges [8]. We provide a schematic overview of the possible interactions by either actor, which allows us to understand a seeker's conversational patterns [22].

The analysis is based on an experimental lab study simulating a natural dialogue to understand how a user and system may interact. Since no existing SCS systems can reliably manage multiple turns, we used conversations between two people, the Seeker, the other an Intermediary. The information seeking dialogues between the two actors were filmed, transcribed, and annotated. The annotated data is publicly available.[1]

The contributions of this work include:

1. We create the first fully labelled dataset for SCS, the SCSdata;
2. We define a multi-level annotation schema for SCS, SCoSAS, to identify the interaction choices for SCS;
3. We create a new model based on multi-turn activities and multi-move utterances;
4. We validate the proposed annotation schema by re-annotating the SCS-data and with annotating another conversational dataset;
5. We suggest new design recommendations and hypotheses for SCS.

The aim of this qualitative research is to explore SCS as a new search paradigm and seeks to understand the exhibited behaviours demonstrated in an ideal scenario for SCS. Thus, we aim to gain a deeper understanding and overview of the interaction behaviours of a group of participants through qualitative analysis. We first create a rich and detailed dataset through a natural dialogue study (NDS) [58], which we refer to as our observational study, to explore SCS interaction behaviours and to seek patterns within this data through an inductive method. In particular, we use our observational study dataset to better understand the communication behaviours at first-hand instead of relying on questionnaires or self-report. The strengths of our qualitative analysis are that it provides an in-depth and detailed analysis to explain complex interactions during the information seeking process.

---

[1]Transcripts together with codes/labels of the experiment are available from the corresponding author upon request.

The remainder of the paper is organised as follows. In Section 2, we present related work. In Section 3, we describe our methodological approach to create the SCSdata and analysis. In Section 4, we present all the identified themes and sub-themes, and we then validate the coding consistency within the SCSdata. We further validate our coding schema in Section 5. Then, discussion, implications, and design recommendations for SCS are presented in Section 6, before concluding in Section 7.

2. Related Work

We organise previous work into four sections: Spoken Dialogue Systems (Section 2.1), annotating dialogues (Section 2.2), interaction space in conversational search (Section 2.3), and natural dialogue studies (Section 2.4).

2.1. Spoken Dialogue Systems

SDS provide a platform for people to interact with computer applications such as databases with the use of spoken natural language. These systems exchange information on a turn-by-turn basis providing an interface between the user and the computer [17].

In recent years, interest in SCS has grown, as speech technology [56] and machine learning for spoken systems [57] have developed. A range of SDS are available, from question answering to semi-conversational systems [28]. Research has been devoted to task-oriented SDS which has defined search boundaries, such as travel planning or route planning, and can be developed with slot filling approaches [55]. Thus, task-oriented dialogue systems are created on a particular closed domain. However, non-task-oriented dialogue systems or open-domain conversations such as search for SCS systems may not benefit from a rigid plan-based dialogue approach and introduce many new challenges [20]. These challenges include how to deal with the variety of user utterances and how answers or replies could be simplified or abstracted to generate appropriate system responses [43].

4

2.2. Annotating Dialogues

Research interest in SCS has increased the recording of spoken search interactions [45, 54]. Such records are a valuable source of data to understand how users interact and which tactics are used for driving effective search performance in this new search paradigm. Thus, this data is useful to understanding the characteristics of a search conversation to build SCS systems acting as a dialogue participant [17]. The spoken data recordings themselves need to be appropriately transcribed and "annotated" [24]. Thus, exposing the structure of the conversations by annotating the actions taken is one of the first steps towards analysing these spoken interactions [59].

Previously, much research has been devoted to creating annotation schemas and classifying taxonomies for dialogue and SDS [1, 11, 38]. Annotating these dialogues has been based on the understanding that classifying utterances provides insight into the dialogue behaviour [34] additionally research on dialogue is often based on the assumption that dialogue acts provide a useful way of characterising dialogue behaviors in human–human dialogue, and potentially in human–computer dialogue as well [1, 7, 10]. For example, annotated conversations can help to identify answers in texts or characterise user intents [31].

Annotating dialogue transcriptions has been explored by sociologists (via *conversation analysis*, e.g., [36]) and socio-psychologists (e.g., [13]) for the purpose of understanding the organisation and communicative purpose of dialogue contributions. Within Computational Linguistics, annotation via *dialogue acts* – which extend Searle's *speech acts* [38] by adding the social-communicative purpose – has been used to analyse dialogue transcriptions for the purpose of designing computational models of dialogue management. While Allen and Core's original DAMSL framework was designed for task-based dialogue, subsequent formulations have been designed for specific types of dialogues [1].

Thus, several different annotation schemas have been proposed which cover the general speech interactions. Such schemas emphasised information seeking, such as the dynamic interpretation theory (DIT) by Bunt [10]. The DIT was based on the empirical investigation of spoken human–human information dia-

5

logues. Bunt suggested that these information dialogues have two motivational sources, namely, to proceed in the task and to exchange communicative functions to drive the conversation [10]. He noticed that an information dialogue consisted of the expected greetings, apologies, and acknowledgements but also included information-exchange utterances such as questions, answers, checks, and confirmations. Later, Bunt developed an annotation schema called DIT++ for these information dialogues [11]. Nevertheless, DIT++ lacks the detailed distinctions made when a user interacts with a search system while satisfying their information need, for example the techniques used to represent documents or information units.

2.3. Interaction Space in Conversational Search

Different schemas have been proposed for information-seeking dialogues based on dialogue acts (DAs) [38] which try to capture the role of an utterance. In particular, schemas such as the COnversational Roles (COR) [40] and Query Request Feedback Answer (QRFA) [54] aim to provide the structure of a single dialogue contribution or move. In our study, we are interested in interactions between a user and SCS system in a more exhaustive manner: for example, utterances such as relevance feedback statements or physical actions (i.e., a mouse click to open a document). These are not covered by general-purpose DA models.

A more relevant conceptual framework was recently created by Azzopardi et al. [3]. This framework combined the action and interaction space discussed in Radlinski and Craswell [32] and Trippas et al. [53]. The conceptual framework, therefore, is not restricted to DAs but provides an overview of the possible actions taken by either actor. We develop the action and interaction space while enriching the current frameworks.

2.4. Natural Dialogue Study

A first step to conceptualising SCS is to explore how people interact or speak in the SCS task they are trying to accomplish [22]. In the case of a SCS system,

one could investigate the reference interview techniques or record elicitation processes librarians undertake with information seekers [15, 6]. However, a more direct approach is to record a situation where people are acting as closely as possible to the task of interest [22]. A natural setting will encourage participants to converse more intuitively and thus provide insights into the language or vocabulary people use, their turn-taking behaviours, and the information flow [10, 58].

A natural dialogue study (NDS) supports an understanding of the accepted conversational patterns in human dialogue. Thus, more natural and usable conversational systems can be created by studying human dialogue [58]. In other words, NDS helps to explore the behavioural patterns and provides insights to improve the design of the system while creating a conceptual understanding of human dialogue behaviour [10].

NDS is not a Wizard of Oz (WOZ) technique. In a WOZ setting, a human acts as a system while the user thinks they are interacting with a live system [18].

3. Methodology

We conducted a laboratory study to collect utterances and search interactions to develop the SCSdata. This dataset captures the utterances of two participants or actors communicating to fulfil an information need. In particular, the purpose of the SCSdata is to understand how users communicate in an audio-only search setting where no screens are available to exchange information and focuses on the issues one could encounter when using such a search system. Thus, observing how people search in this setting provides initial insight into the interactions taken [53]. To this end we conducted a study to collect a set of utterances and search interactions from two actors communicating to fulfil an information need: *SCSdata* [53]. We developed an annotation scheme for SCS, the *SCoSAS*, analysed this and validated it with inter-rater reliability; further tested it with an independent data set, Microsoft Information-Seeking

Conversation data (MISC) [45, 46, 50]. Our analysis provides insight into the interaction space and design recommendations for further research into SCS.

3.1. Approach

The development of spoken language datasets is a work-intensive and time-consuming process. Nevertheless, these datasets are invaluable for conversational modelling, as a resource for system development, or defining of vocabulary coverage [17]. The development and evaluation of SDS is a well-studied problem and has shown that iterative analysis and assessment is needed.

To enhance our understanding of SCS, we adopt NDS as a well-established technique used in SDS to develop a spoken language dataset and utilise qualitative analysis to identify meaningful patterns in our dataset [17, 8]. The purpose of our experimental setup is to specify the interaction possibilities in SCS. By outlining these different interactions, we provide the first step towards uncovering the details of the SCS process [17].

Our observational study consisted of a number of sessions with two participants, where one participant acted as the *Seeker* and the other participant as the *Intermediary* as illustrated in Figure 1.

Information Need   Seeker   Spoken Conversational Channel   Intermediary   Search Engine

Figure 1: Experimental setup.

The Seeker received a *backstory*: a short information need statement, to motivate and contextualise the search need.[2] The Seeker had to read the backstory and verbalise the information need without reading out the backstory to the Intermediary. Instead, the Seeker had to personally formulate their information need problem to convey it to the Intermediary. The Intermediary had access

---

[2]Information needs and backstories used in our experiments are listed in Appendix B.

to a search engine through a desktop computer. In effect, the Seeker acted as the searcher and the Intermediary simulated the audio-only interface and search system. Participants could not access each others' tasks or search engine, were not able to see each others' facial expressions, and could only verbally communicate. All backstories were randomised and the participant roles were randomly assigned.

Participants completed pre-test, pre-task, post-task, and exit questionnaires, as well as a semi-structured interview. Sessions took around 90 minutes.

3.1.1. Thematic Analysis

Thematic analysis involves identifying, analysing, and reporting patterns (themes) within qualitative data [8]. This method allows analysing qualitative data in an accessible and theoretically flexible manner, and it is often seen as a fundamental way of examining this kind of data [8]. We adopted the six-step process as outlined by Braun and Clarke [8]: (Step 1) familiarising self with data, (Step 2) generating initial codes, (Step 3) searching for themes, (Step 4) reviewing themes, (Step 5) defining and naming themes, and (Step 6) producing the report. All steps were completed by the first author with continuous systematic feedback sessions with two other authors from Step 2 onward and a second independent annotator as validation of the full schema.

We illustrate the two-tiered coding process in the following example from Participant 8 (Seeker). The example utterance was coded as "Intent clarification" as it describes the Seeker further explaining their query intent. This code is then grouped with similar codes into the "Information Request" sub-theme in the Task Level theme. We describe all themes and sub-themes in Section 4.

P8 -SEEKER: Yeah, so I just want to know where it comes from

[Intent clarification]

To the best of our knowledge, we are the first to use thematic analysis to create of an annotation schema for SCS.

3.1.2. Validation of the SCoSAS schema

To reduce the possibility of missing important data points, we validate our coding schema in two ways. We computed (1) inter-rater reliability and code overlap, and (2) overlap and coverage based on the coding of a different dataset, the MISC[3], with our predefined codes [45].

A second independent annotator, who is familiar with information seeking and information retrieval research, recoded all utterances in the SCSdata to obtain the inter-rater reliability with Cohen's Kappa and code overlap [23]. The second annotator used the codebook for closed coding (i.e., the categories were already determined).

Identifying useful actions for SCS which have not been covered in the SCoSAS provides an understanding of the scope of our coding schema. Therefore we apply the SCoSAS to a second and similar dataset, the MISC, to calculate the overlap and coverage [45]. We took a random sample from the MISC and coded the utterances according to our dataset. Nevertheless, it may not be possible to achieve complete coverage with our annotations given the complexity and unexplored interactivity of a SCS information seeking dialogue [41]. In addition, achieving full coverage is difficult and often not possible to achieve [42]. Hence, declarations which were not covered in SCSdata received new codes according to the steps of thematic analysis.

3.2. Data Collection Setup

This section introduces the experimental setup by describing the tasks used in the experiment, an overview of the participants, and the annotation steps.

3.2.1. Task Design

We used nine search tasks and backstories from Bailey, Moffat, Scholer, and Thomas [4] (Appendix B). These tasks covered three levels of the Taxonomy of Learning [2]: *Remember, Understand,* and *Analyse.*

---

[3]The MISC data was accessed at `http://aka.ms/MISCv1`.

3.2.2. Participants

The study involved 26 participants recruited through a mailing list.[4] Fifteen participants were female and 11 were male with a mean age of 30 years (SD=11, range 18–54). Twenty-two participants reported being a native English speaker, and four participants said they had a high level of English proficiency. The highest level of degree held was a Master's degree. Eighteen participants reported that they were awarded a Bachelor's degree or higher and eight participants said their highest level of degree awarded was High School graduation. The majority of participants were students (73%), 19% was employed, and 7% were unemployed. The most common fields of education were Science and Engineering (both 19% respectively) and Law (11%). Participants reported that they had been using a computer for more than ten years (85%) and 15% reported using a computer for 5–10 years. All participants said that they used search engines daily with the majority of participants reporting that they used a search engine more than eight times per day (54%).

Participants rated their search skills on a 5-point scale, where 1=novice and 5=expert. Participants' mean search skills were 3.9 (SD=0.5), with a minimum score of 3 and a maximum of 5.

Participants' search self-efficacy was measured with the Search Self-Efficacy scale [9], which contains 14 items describing different search activities. Participants indicated their confidence in completing each activity using a 10-point scale, where 1=totally unconfident and 10=totally confident. Participants' average Search Self-Efficacy was 7.3 (SD=1.51 and Cronbach's alpha=0.93).

Participants reported their usage of intelligent personal assistants, such as Google Now, Apple's Siri, Amazon Alexa or Microsoft Cortana. Four participants had never used an intelligent assistant and eight had used one a couple of times but did not use them anymore. The majority (54%) of the participants

---

[4] The protocol was reviewed and approved by RMIT University's Ethics Board (ASEHAPP 08-16). The mailing list is created and maintained by the Behavioural Business Lab at RMIT University: `https://orsee.bf.rmit.edu.au/public/index.php`.

said they used an assistant, consisting of five participants using one at least once a month and nine participants using one at least weekly.

3.3. Data Analysis and Annotation Schema Creation

3.3.1. SCS Dataset

The SCSdata consists of 1044 turns between the 13 pairs of actors. Seekers took a total of 528 turns and Intermediaries 516. (Seekers instigated and could conclude the search, so they took 12 turns more than Intermediaries.) We recorded an average of 80 turns per pair and 26.76 turns per task. Participants exchanged 15.82 words per utterance on average with a minimum of one word per turn and a maximum of 359 words per turn. This maximum involved an Intermediary reading out a document, an action which is unusual for the dataset where the median words per turn was 9. The SCSdata which was manually transcribed and subjected to the three-pass-per-tape policy [27]. An editor then proofread the SCSdata transcription [52]. To mitigate automatic speech recognition (ASR) problems such as out-of-vocabulary utterances, we transcribed the SCSdata manually allowing us to conceptualise the user-system interactions. For future systems, investigation will be necessary to understand the impact of ASR transcriptions on the user-system interactions.

3.3.2. Coding Transcriptions With Thematic Analysis to Develop SCoSAS

We coded (i.e., labelled) our transcriptions using thematic analysis as described previously in Section 3.1.1. The labels of the SCSdata form the annotation schema, SCoSAS. We recorded both participants, and the Intermediary's screen. The recordings were synchronised and merged for transcription. We adopted the following steps:

Step 1: Identifying when each participant spoke, i.e., identifying turns. We used the approach of *taking the initiative equals taking the turn*, as described by Hagen [19]. This means that one turn can consist of multiple moves, actions, or communication goals [47].

Step 2: Transcribing each turn of the full dataset. However, we deliberately did not eliminate any errors, false starts, or confirmations since these occur in real case voice search scenarios and to preserve the morphological naturalness of the transcription and the naturalness of the transcription structure [27, 52]. Instances where either Seeker or Intermediary was unintelligible were not transcribed but were coded as *[inaudible segment]*. We assumed that if the audio recording was not clear, it was probably not clear to the other participant either.

Step 3: Designing and assigning codes to each turn with ELAN [52, 25]. Observational notes were added. The full dataset was coded with each utterance receiving equal attention. We classified concepts from the recordings and devised a coding scheme according to the similarities across different actors. The codes were designed to identify the action(s) of that particular turn, describing features of the data and defining the function of the turn (i.e., one turn/utterance could consist of multiple codes). Thus, turns were annotated with the actions taking place. Consequently, meaningful labels were developed from the original annotations. *Controlled Vocabulary* was added to a *dictionary* which was created during coding. This dictionary was then developed into a *codebook*.

Step 4: Combining codes to themes for further analysis. Themes may consist of *sub-themes* which capture specific concepts as illustrated in Figure 2.

Step 5: Checking quality assurance. Transcriptions and codes were exported from ELAN to a text file. Spelling and codes were checked.

Step 6: Importing files into R and aggregating codes to check whether codes within a theme conceptually belonged to that theme.

Note: Steps 3–6 were conducted iteratively. This process reduced the initial 100 codes to 84 through the identification of overlapping codes. To preserve the nuanced action described in the codes for future information seeking research, distinctions between closely defined codes were retained. For example, the codes "Information request" or "Information request within document" were retained to identify in which section of the interaction
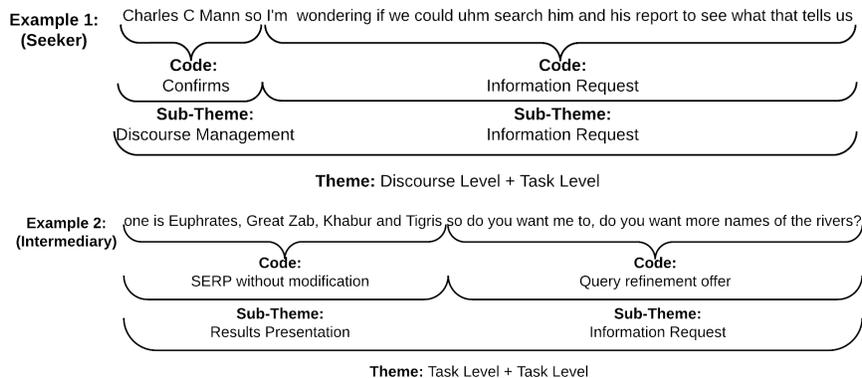
Figure 2: Example of coding utterances.

particular information was requested.

Steps 3–4 were conducted iteratively by Trippas with feedback sessions with two other authors. Random samples were investigated and compared against the coding schema, and feedback was incorporated in the next coding iteration.

4. Results

In this section, we present the themes derived from the thematic analysis together with the sub-themes which are based on the constructed codes/labels (Sections 4.1–4.4). These themes provide the characteristics of information seeking dialogues in a conversational setting, the actor's role, and the actor's relationship with the conversation. Then, we focus on the inter-rater reliability and code overlap calculations addressing the consistency of our coding schema (Section 4.5).

4.1. Themes for Spoken Conversational Search

Every utterance received one or more codes, based on the action taken in that utterance. For example, when an Intermediary read out a document to the Seeker, this utterance was coded with "Scanning document". However, when

there were two actions present in one utterance – such as when an Intermediary read out a document and then asked the Seeker whether that was useful for them – then the utterance received two codes: "Scanning document" and "Asking about usefulness". Other examples of multiple actions and the coding of these actions are provided in Figure 2.

To understand which actions are taken, we split all codes where more than one code was attached to an utterance — thus creating atomic actions per utterance for a more natural grouping of these actions into themes and sub-themes. We present the three themes and their corresponding sub-themes and codes as follows. The first theme, *Task Level*, is related to search interactions and the topical investigation. The second theme, *Discourse Level*, is associated with communicative functions between the Intermediary and Seeker for smooth collaboration. The third theme, *Other*, consists of utterances that belong to neither the Task nor the Discourse levels. Example utterances are provided for each sub-theme. Tables of all the themes, corresponding sub-themes, participants (or actors), and codes are included in Appendix A.[5]

4.2. Theme 1: The Task Level

The Task Level theme covers search actions such as queries and search results presentation. In other words, this theme is related to the performed search task. The theme includes four sub-themes:

Information Request. This sub-theme covers utterances which are associated with topical information requests. It includes all utterances with codes which are related to forming, suggesting, refining, confirming, repeating, spelling, or embellishing information requests. The following example is of two information request sub-theme utterances:

> P13 -SEEKER: So which state in Australia consumes the most alcohol

---

[5]The data provided with this paper supplies all the transcripts and corresponding sub-themes and codes.

per person?

[Information request]

P14 -Intermediary: Again 2016 or the most recent information?

[Information request]

Information requests from Seekers could be expressed at any time, and they often asked for information *from* a document itself, asked for meta-information *about* a document or search engine results page (SERP), or provided clarification about their search intent. Intermediaries were more likely to provide support in (re)forming the information request, for example by providing information request refinements, suggesting query expansions (i.e., whereby the initial query is augmented with query terms), or eliciting extra information.

Results Presentation. These sub-theme utterances convey the results from the search engine or documents: reading, interpreting, or providing an overview of a SERP or document. Only Intermediaries use this sub-theme, and the majority of Intermediary actions are linked to this sub-theme. In the next example the Intermediary reads out the results exactly as they were displayed in a document:

P6 -Intermediary: The history of valuable cinnamon. The first mention of cinnamon is in Chinese documents dating from 2800 BC. The ancient Egyptians logged cinnamon as a spice used in the embalming process...

[Results presentation]

Other categories of utterances where Intermediaries conveyed the documents or search engine results but modified them (i.e., interpreting the results so that they would be most beneficial for the user) are also sorted in this theme. Intermediaries modified SERPs or documents via synthesis; interpretation; paraphrase; summarisation; clarification; and comparison.

Search Assistance. This sub-theme captures interactions where the Intermediary assisted the search process by providing explicit search suggestions, advice,

16

or relevance judgements:

> P2 -INTERMEDIARY: there is a lot on health benefits conversation uhm [long pause] I don't see how some of these are relevant
>
> *[Results presentation + search assistance]*

In contrast to directly providing assistance, Intermediaries also asked how to help the Seekers in their search process. This was seen by asking about the usefulness of a result, requesting spelling, or suggesting a different search engine.

Additionally, this sub-theme captures the Seeker explicitly asking for assistance during their search session: for example by asking for recommendations or judgements on whether they covered enough of the information space.

Search Progression. This sub-theme is only used by the Seeker to provide feedback on progress: for example by giving performance feedback, rejecting search results, or informing whether they found enough information for a topic:

> P15 -SEEKER: OK that's probably enough information
>
> *[Search progression]*

In summary of this first theme, the Seeker evoked all sub-themes in the Task Level theme except the Results Presentation sub-theme. The Results Presentation sub-theme was only used by the Intermediary allowing them to present found information from the search engine to the Seeker. The Intermediary also evoked all sub-themes except the Search Progression sub-theme, which was used by the Seeker to provide feedback to the Intermediary.

4.3. Theme 2: The Discourse Level

The Discourse Level theme covers aspects which are not linked to performing a topical (search) task but instead are concerned about the audio channel between participants. The Discourse Level theme consists of four sub-themes: *Discourse Management* which allows the conversation to take place between the actors; *Grounding* captures interactions for creating mutual knowledge, beliefs, and assumptions between the two actors [48, 13]; *Navigation* which covers the

communications of moving around web pages, documents, and browser tabs; and *Visibility of System Status* which allows actors to provide feedback on what is happening throughout interactions.

Discourse Management*.* This sub-theme includes conversational coherence and cohesion between the actors [37]. In other words, the utterances in this sub-theme are part of the communication between the actors to check whether the other actor has understood a message. In our dataset, these discourse building utterances are independent of the participant role. For example, both Seeker and Intermediary confirmed, checked, or asked to repeat and repeated utterances as illustrated in the snippet below.

> P1 -SEEKER: So uhm can you go and change the search question to effectiveness of uhm... passenger and baggage screenings at airport
>
> P2 -INTERMEDIARY: Passenger and
> *[Discourse management]*
>
> P1 -SEEKER: Baggage
> *[Discourse management]*

Often an information request was echoed or either actor confirmed a command. These discourse actions are crucial to have a meaningful conversation, for example indicating that one actor has understood the other.

Grounding*.* *Grounding in communication* as described by Clark and Brennan is "sharing and synchronising mutual beliefs and assumptions" and is fundamental for communication between actors [13]. The two actors' mental model of each others' beliefs needs to be continuously updated to coordinate the build of a mutual understanding.

We observed utterances belonging to this particular sub-theme which was used by Seekers to coordinate the shared information or common ground [13]. Seekers summarised or paraphrased the information given to them and created a bigger picture of the search results as a way of synchronisation.

Through this dynamic process, Seekers provided insight on what they understood from the information provided; Intermediaries then knew whether the information was correctly conveyed.

> P14 -INTERMEDIARY: [...] yeah 20 to 29 is the most high risk drinking people in Australia for alcohol related harm... I don't know what that means about consumption

> P13 -SEEKER: Yeah so they consume a lot
> *[Grounding]*

Grounding differs from Search Progression and Discourse Management. While Grounding involves sharing the beliefs and values of the information, Search Progression is concerned with the feedback on the search task progress and Discourse Management is related to effective information transfer.

The Grounding sub-theme was only seen in Seekers' utterances. This is because Intermediaries, by having the information to hand, summarised results presented and they did not need to confirm or share their beliefs or meaning of the content. As such, their utterances are captured by Results Presentation.

Navigation. Navigational utterances allow actors to progress the task by manoeuvring around the online information space. We observed Seekers navigate the search results by instructing the Intermediaries. Seekers asked to access specific sources, navigated between documents, singled out particular documents, and read more from a document or the next document:

> P9 -SEEKER: Uhm maybe uhm can you go into the result [...] that mentions how uhm outsourcing damages the industry
> *[Navigation]*

Visibility of System Status. Seekers asked the Intermediaries to provide information on what was occurring throughout the interactions: for example, whether what they asked for was fulfilled, or what the results were. Intermediaries provided feedback on what was taking place on their side of the

conversation by giving input on what was happening (i.e., keeping each other informed [30]) if they had seen certain items before, or by way-finding (i.e., orienting where they were positioned). For example:

> P25 -SEEKER: Oh TIBER sorry Tiber yeah
> *[Discourse management]*
>
> P26 -INTERMEDIARY: Yeah uhm just searching just one second
> *[Visibility of system status]*
>
> P25 -SEEKER: Any luck?
> *[Visibility of system status]*

4.4. Theme 3: Other

Five utterances from the Seeker were not classified in any of the above (sub-)themes. Two of these utterances were disfluencies from the Seeker, one utterance was where the Seeker provided information about the search engine, one utterance was asking if the Seeker was allowed to embellish a query, and the last unclassified utterance involved the Seeker offering to spell a word. These five categories were not classified after much deliberation and given the theme "Other" instead.

To finalise the themes examination, an overview of the themes and sub-themes used by each actor is presented in Table 1. The development of the classifications in themes, sub-themes, and codes form the basis of the Spoken Conversational Search Annotation Schema (SCoSAS).

4.5. Inter-rater Reliability and Code Overlap

As part of the validation of the SCoSAS, we calculate the inter-rater reliability and code overlap (i.e., the utterance code overlap and code usage overlap).

The first author (Assessor 1) created the codes as described above. A second independent researcher (Assessor 2) used the codebook for closed coding of all utterances in the SCSdata. The inter-rater reliability on code level (i.e., atomic action identified on an utterance) was moderate (Cohen's $\kappa = 0.59$) [23],

Table 1: Themes and sub-themes used by different actors

Theme	Sub-theme	Seeker	Intermediary
Task Level	Information Request	✓	✓
	Results Presentation		✓
	Search Assistance	✓	✓
	Search Progression	✓	
Discourse Level	Discourse Management	✓	✓
	Grounding	✓	
	Navigation	✓	
	Visibility of System Status	✓	✓
Other		✓	

and substantial at the sub-theme level (i.e., classification based on the code) ($\kappa = 0.71$).

The overlap of codes was high with 90% of the predefined codes being used by both assessors. More precisely, Assessor 1 applied 84 different codes consisting of 41 codes for the Seeker and 43 for the Intermediary. Assessor 2 used 76 codes, 38 codes for the Seeker and 38 for the Intermediary as seen in Table 2.

Table 2: Independent Assessors' Code Overlap

	Assessor 1	Assessor 2
Total number of utterances	1,044	1,044
Total number of codes used	84	76
Total number of codes for Seeker	41	38
Total number of codes for Intermediary	43	38
Unused codes	0	8 (10%)

The eight codes used by Assessor 1 but not Assessor 2 could potentially be consolidated in a future refinement.

5. Validation of SCoSAS

To explore the extent to which SCoSAS covers SCS interactions, we applied the SCoSAS to a subset of interactions from a second dataset, the MISC [45].

5.1. Using the MISC dataset to validate SCoSAS

As with the SCSdata, the MISC dataset [45] is a collection of recorded information-seeking conversations between a Seeker and an Intermediary. MISC contains audio and video recordings with automatic speech recognition transcriptions of these recordings. We coded the MISC dataset according to our predefined codes to investigate which actions were covered or not covered by our coding schema. Thus, by using our predefined codes, we validate the coverage (i.e., is there an action applicable for every situation) and overlap (i.e., is there a situation where more than one action could be relevant).

We selected a random set of four participant pairs for the labelling: participants 1–2, 7–8, 19–20, and 27–28. The MISC setup has five tasks for each pair, of which one is a practice. We labelled the four remaining tasks per participant pair for a total of 16 task-instances. The labelling was completed by Trippas.

Among the four pairs, we have a total of 701 turns with an average of 175.25 turns per pair and an average of 43.81 turns per task. However, 5% of the total turns in the MISC transcriptions were inserted by the ASR and were not present in the audio. These turns were ignored which means that a total of 666 turns were labelled on code-level with an average of 166.5 turns per pair and an average of 41.62 turns per task.

5.2. Differences Between the SCSdata and MISC

The setup and instructions between the SCSdata and MISC protocols were marginally different, which led to differences in the data. We provide an overview of the differences in this section; a fuller account is in Trippas and Thomas [50].

Setup of SCSdata and MISC. As with SCSdata, MISC Seekers did not have access to any information source, but received an information need which they relayed to an Intermediary over an audio connection. Unlike SCSdata, MISC Seekers were allowed to read out the need as given, but were also asked to record an answer.

Transcription Differences. The MISC dataset was transcribed using ASR in contrast to the SCSdata which was manually transcribed, subjected to the three-pass-per-tape policy, and proofread by a professional editor.

The ASR was prone to error, in particular "recognising" utterances such as "thank you" that were not in the audio. The following snippet of a conversation is an illustration: speakers appear more polite than they were.

> P20 -INTERMEDIARY: [...] She wanted them to donate to charity
>
> P19 -SEEKER: Thanks
>
> *[Utterance not present in audio]*
>
> P20 -INTERMEDIARY: To provide clean water // and she um
>
> P19 -SEEKER: Thank you
>
> *[Utterance not present in audio]*

We encountered an occasion where the researcher interfered due to a technical issue and sections where the ASR created many unnecessary turns between the actors because it falsely believed that someone was talking. We excluded these utterances from this analysis.

5.3. Creating Comparable Datasets

The MISC dataset contains ASR errors and the subset we used did not include screen capture video. We labelled MISC at the code level. However, subtleties such as whether an Intermediary was reading from a SERP or document could not be distinguished without screen captures and all Results Presentation utterances were labelled just with that sub-theme.

5.4. Results: Overlap and Coverage Between SCSdata and MISC

We are interested in the number of actions shared between the SCSdata and MISC, and where actions are different. After collapsing all Results Presentation utterances to the sub-theme level, for compatibility with MISC, SCSdata used 66 distinct codes: 41 from Seekers and 25 from Intermediaries. MISC used 31 for Seekers and 18 for Intermediaries (Table 3).

Table 3: SCSdata and MISC Descriptives

	SCSdata	MISC subset
Total number of utterances	1044	666
Total number of unique codes*	66*	49*
Unique codes Seeker	41	31
Unique codes Intermediary	25	18

*NOTE: Due to insufficient details, utterances which were related to presenting results were aggregated to the Results Presentation sub-theme level. The SCSdata's unique number of codes without aggregation of the Results Presentation is 84.

To label MISC, we needed 49 codes, of which 35 were used in both sets; 14 additional codes were needed to cover actions not seen in SCSdata. These additional codes however were infrequently used and 94% of MISC utterances could be coded with SCoSAS. The 14 additional codes, covering 6% of utterances, are summarised in Table 4, and we discuss these below.

Table 4: Set difference between SCSdata and MISC.

Code	Actor	Nr used
Chitchat		1
Communication about the task		2
Decision offloading		1
Feedback on writing down the answer for the given task		3
Negotiation	Seeker	7
Rejects spelling offer		1
Requests spelling		1
Uncertainty expression of what to search		2
Chitchat		5
Enough information?		9
Negotiation		6
Offers to spell	Intermediary	1
Spells		5
Too many results to sum up		1
Total number of instances of code used by MISC and not by SCSdata		45 (6%)

Chitchat or Negotiation. We encountered new types of utterances in the MISC where the actors were negotiating or chitchatting. The negotiation utterances were used to bridge differences and reach agreements [60]. Examples include instances where actors share their own experiences about particular topics or subjects. However, this is not to be confused with the already defined Grounding sub-theme which covers utterances from the Seeker expressing their beliefs and values of information provided by the Intermediary.

Chitchat and negotiation utterances have greater overlap between speakers, meaning that more than one actor at a time is speaking [36]. For example, the following utterances overlapped while the Seekers and Intermediary negotiated their shared understanding of non-traditional medicine:

> P1 -Seeker: I think herb sounds more like // not
> *[Negotiation]*
>
> P2 -Intermediary: More like medicine
> *[Negotiation]*
>
> P1 -Seeker: I think it sounds more like naturopathic but that fits it
> *[Negotiation]*

Participants seemed forthcoming in sharing their own opinions and experiences. The following example is from an Intermediary who shares her own travel experiences which are related to the task:

> P8 -Intermediary: That's what I love to do actually when I traveled all the public transportation and all sorts of continents
> *[Chitchat]*

Communication about the task. SCSdata participants were instructed not to share the given search task but instead rephrase request. However, for MISC, participants were allowed to read out their search task. This led to Seekers talking informally about the task itself. For example,

> P1 -Seeker: Yeah the task is a bit // um very generalised so um

Agency and Decision Offloading or Taking Control*.* In MISC, both Seeker and Intermediary share the same information and underlying ideas of what they need to search for. This created an equal level of collaboration between the two actors. However, it also allowed the Intermediary to instantiate more agency. In contrast, Intermediaries in the SCSdata acted more as the interface between the Seeker and the found information.

We noticed this idea of agency throughout the subset of the MISC in actions resulting in the following codes "Enough information?" (Intermediary), "Too many results to sum up" (Intermediary), and "Decision offloading" (Seeker). For example, the Intermediaries suggested that a search task has been finished *"excellent, so we are finished..."* (P8) or that they are not going to sum up all the results. The Seekers also handed over the decision making to Intermediaries: e.g. *"it's up to you [ed. if we look at the other site or not]"* (P20).

Trippas et al. [53] suggested decision offloading and taking control may be an artefact of the linear audio channel. The system thus creates a cost-benefit estimation of whether further information from the Seeker is required and therefore receives more autonomy. Simultaneously the Seeker can suggest the system to make the discussion due to the limited knowledge which can be transferred in the audio channel.

Writing Down the Answer*.* Seekers in the MISC setup were asked to write down an answer to their given information seeking task. Seekers' utterances therefore include how they were progressing with the writing. We also encountered several different instances of spelling actions in MISC which we had not encountered in the SCSdata.

Uncertainty in What to Search*.* As mentioned, MISC Seekers were allowed to read their search task aloud. Here, the Seeker is expressing their confusion with the task:

> P19 -SEEKER: I am not sure what you're supposed search

This could be interpreted as identifying a gap in the Seeker's knowledge. However, the information need expression is not formalised [44]. Recently, Trippas et al. [53] suggested that formulations of needs in SCS do not conform to the typical textual query. In a voice environment, users can use natural language to describe their search, and the information request may not go through Taylor's four stages of information need [44]. In the above example, we could even argue that users may now have the freedom to tell the system that they have identified a gap in their knowledge before yet formalising their problem.

5.5. Discussion of SCoSAS Validation

The majority of the codes (71%) seen in MISC overlapped with the SCoSAS, and the novel codes only covered 6% of utterances. Some of the new codes were not encountered in the SCSdata due to the difference in experimental setups, such the array of possible spelling requests, suggestions, or declines. These would be valuable expansions to the SCoSAS.

6. Discussion

In this work, we used a qualitative analysis approach to uncover the range of possible interactions of information moves for Seekers and Intermediaries in a SCS setting. We constructed insight into the conversational structure of information seeking processes. To do so, we first created a spoken dataset, the SCSdata, and then derived an annotation schema for conversational search via a thematic analysis approach. Finally, we evaluated these actions against a similar dataset as evaluation.

6.1. Schematic SCS Themes Model

The SCoSAS presents the Task Level at the centre of conversations. The Discourse Level surrounds this, representing the statements which are about the mechanism, not the task (see Figure 3). The Discourse Level would still exist if the search task is changed to a different, unrelated, task. Previous research in
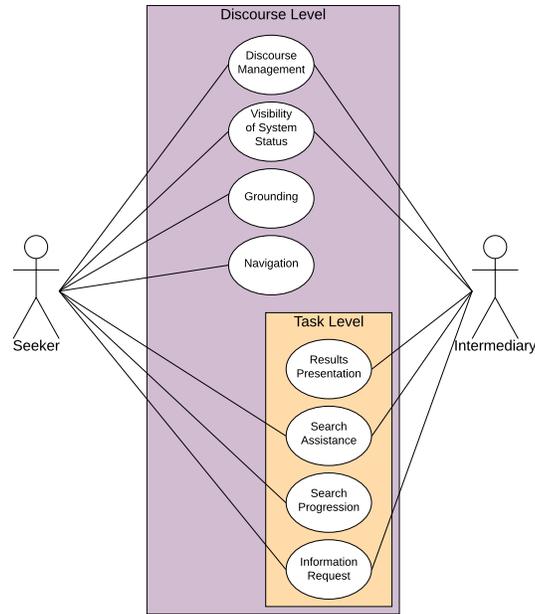
Figure 3: Schematic Model of Themes and Sub-themes used by each actor.

communication goal studies suggests a similar two-tiered model [47, 10]. Furthermore, the goal studies community argues that ordinary discourse is segmented in different types of goals such as communicative functions or interaction outcomes which is similar to our two themes of Task and Discourse. Bunt [10] provided a two-tiered model where general information dialogues consist of two motivations, that is, one tier was concerned about the task communication and the second.

Our results highlight the importance and need for integrating discourse in SCS systems and to the best of our knowledge, discourse functions are yet to be integrated in information seeking models [40, 54]. Furthermore, including these discourse utterance inherently creates a system which interacts in a mixed-initiative information seeking communication (the system can ask for clarification and thus takes initiative). Such mixed-initiative dialogue is a requirement of what makes a SCS system truly conversational [14, 53].

This first attempt at creating an interaction model of two actors in a SCS setting may not have included all possible future actions. One action we believe may be observed in a real system is for the user to test the abilities of the system or access the settings of the system itself. This might be coded as a System Level theme, overlapping both Discourse and Task Levels.

6.2. Design Recommendations for SCS Systems

Our analysis leads to some design recommendations for SCS systems.

Integrating Search Assistance*.* Search assistance is integrated in many different ways in browser-based search, for example by query or spelling suggestions. We could extend these assistance functions to include the system providing relevance feedback to the user about a given document, suggesting to move on, or even asking about the usefulness of a given result. These pro-active features can become part of the model of the user preferences given the interaction history [c.f. 32].

Grounding as Relevance Feedback*.* Grounding (i.e., discourse for the creation of mutual knowledge and beliefs) is when participants in a conversation engage in a specific discourse activity to share their mutually understood utterances [13]. We observed grounding actions in the SCSdata. For example, Seekers provided indirect feedback by reciting their interpretation of the found results. This grounding process could enable a future SCS system to better understand a users awareness of the results or information space, including helping the SCS system to disambiguate a users information need.

Visibility of System Status*.* Visibility of system status enables greater control, explainability, and transparency of the system processes and outputs [14]. However, providing constant feedback on what is happening in a system is not convenient in a spoken environment and will overwhelm the user with too much unnecessary information. It will be essential to understand which aspects should be given to the user. At any point in time, the system should be able to disclose how it retrieved or computed specific information.

Navigation. Navigational interactions often contain a mix of selecting links on a web page or using backtracking techniques such as the back button or history list [16]. These navigation actions have been extensively studied in a text environment [12, 21]. Recently Azzopardi et al. [3] and Trippas et al. [53] recognised the importance of these actions in a conversational search setting. Instead of interacting with lists in a spoken environment as often is done in a SDS, users can freely navigate in a multi-dimensional information space.

Navigational interactions in this study may have been influenced by how a system works through back buttons and links. In the future we expect the navigation space to expand together with the adoption and creation of conversational systems. Being able to present a traceable history also provides further transparency for the user and supports the explainability of the system. For example, breadcrumbs could refer to previous information spaces or provide summaries of information the user visited instead of titles of documents as in a browser-based back-button action.

6.3. Evaluating Existing Search Behaviour Models with SCoSAS

To our knowledge many well-known models such as Belkin's ASK [5] or Marchionini's ISP [26] do not include the system's "responsibility" of interacting with the user and thus do not capture all SCS behaviours.

Other models, such as Sitter and Stein's COR model [39], Belkin et al.'s scripts [7], or the recently proposed QRFA model by Vakulenko et al. [54] encompass the interaction between two actors. However, these models either lack the flexibility of the speech aspect, such as multiple moves in one turn, or are based on broad DA categorisations. Additionally, the broad DA categorisation only provides a high level insight of the actions users take while the SCoSAS discloses more refined details of the users' and systems' state in each turn.

Finally, Saracevic's stratified model includes the system as an active participant in the information seeking process [35]. Furthermore, Saracevic specifies that the process consists of a dialogue between the two actors. He also mentions that the dialogue can be used for not only "searching" utterances but

also for a number of "other engagements" beyond the searching, for example, obtaining and providing different types of feedback, judgements, or states. In the SCS model, we also identify the system as an active participant throughout the search process, which is in itself a conversation. In addition, the "other engagements" Saracevic mentions could be interpreted as our Discourse Level interactions, such as our identified grounding utterances. Furthermore, the stratified model could be used to illustrate the effect of the audio-only interaction channel limitation. That is, Saracevic says that a weak point in the system could hamper the desirable outcome for the search process [35]. The stratified model and the schematic SCS themes model may be complementary for the abstraction of a SCS process.

6.4. Future Extensions

Human to human interaction: Human–human interaction may differ from the human–machine interactions we really want to model. We plan to conduct further studies to test our hypotheses in a human–machine interaction setting. Thus, further research will investigate if the mindlessness projection of users' own believes and expectations of computers transfers to SCS [29].

Laboratory setting: Participating in a laboratory setting influences the participants' behaviour. Even though this study was conducted in a laboratory setting, we believe findings will apply to a day-to-day environment.

Investigating the information needs for SCS which arise in a natural setting will be crucial to develop natural systems. This will include understanding the different information needs and creating new taxonomies for these needs.

Taking initiative equals one turn: Our coding schema allows for coding per turn since we segmented the users' utterances with the idea that taking the initiative equals one turn. This means that slight subtleties inside a turn may be lost such as long pauses. However, we believe this was necessary to understand the broader context of SCS.

This study should be interpreted within the context of its limitations. Firstly, it is possible that cross-coding a larger dataset from the MISC may have added

31

further sub-themes. However, we feel that the discrepancies identified through the current cross-coding are due to differences in experimental set-up rather than to substantial content differences.

6.5. Informing Wider Research Agendas

Existing systems and models have difficulties with multi-turn actions, utterances which consist of multiple moves, or intent extraction. In this paper, we attempted to better understand these unique features of SCS by creating a labelling schema and schematic model of these labels. Our model and annotation schema are a more in-depth study than prior preliminary models [51] or conceptual framework [3]. While the labelling schema developed in this paper focuses on the interaction space of conversational search, we expect it will be useful for non-search related or discourse actions. The implications of this analysis are many. Firstly, this analysis can support the feature extraction of particular utterance-types, or assist with the engineering and evaluation of conversational retrieval. The analysis can also be used for language modelling of information seeking conversations and the development of results presentation strategies.

7. Conclusions

In this paper, we address the challenge of spoken conversational search (SCS), where no screens are available and user–system interactions are entirely voice-based. After identifying the limitations of existing information seeking models, we used a qualitative analysis approach to explore how people interact in an audio-only communication search setting. We created the first dataset for SCS (SCSdata), defined a labelling set identifying the interaction choices for this dataset (i.e., the SCoSAS annotation schema), and translated these interactions in a schematic model. This approaches both actors in the seeking process, the Seeker and the Intermediary, as equal, leveraging multi-turn activities and multi-move utterances. The validation of SCoSAS using an independent dataset demonstrated high overlap and coverage.

Furthermore, our transparent annotation process contributes by strengthening the analysis and the methodological foundations of annotation schema development.

The significance of this paper is twofold: we (i) develop a classification schema, (ii) test and validate this schema, and (iii) provide a transparent annotation schema process. Furthermore, our contributions highlight the need of new models for SCS, especially integrating discourse. The resources described and validated in this paper – including the SCoSAS annotation schema – also allow us to suggest possible extensions of the schematic model and inform the design of SCS systems in the future.

8. Acknowledgements

References

[1] J. Allen and M. Core. DAMSL: Dialogue act markup in several layers (draft 2.1). Technical report, Multiparty Discourse Group, Discourse Resource Initiative, 1997.

[2] L. W. Anderson, D. R. Krathwohl, and B. S. Bloom. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, 2001.

[3] L. Azzopardi, M. Dubiel, M. Halvey, and J. Dalton. Conceptualizing agent-human interactions during the conversational search process. In *Proc. of the Int. Workshop on Conversational Approaches to Information Retrieval*, 2018.

[4] P. Bailey, A. Moffat, F. Scholer, and P. Thomas. User variability and IR system evaluation. In *Proc. of the Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 625–634. ACM, 2015.

[5] N. J. Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5(1):133–143, 1980.

[6] N. J. Belkin, H. M. Brooks, and P. J. Daniels. Knowledge elicitation using discourse analysis. *International Journal of Man-Machine Studies*, 27(2):127–144, 1987.

[7] N. J. Belkin, C. Cool, A. Stein, and U. Thiel. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Syst. Appl.*, 9(3):379–395, 1995.

[8] V. Braun and V. Clarke. *Successful qualitative research: A practical guide for beginners*. Sage, 2013.

[9] K. Brennan, D. Kelly, and Y. Zhang. Factor Analysis of a Search Self-Efficacy Scale. In *Proc. of the ACM SIGIR Conf. on Human Information Interaction and Retrieval*, pages 241–244. ACM, 2016.

[10] H. Bunt. Dynamic interpretation and dialogue theory. *The Structure of Multimodal Dialogue*, 2:1–8, 1999.

[11] H. Bunt. The DIT++ taxonomy for functional dialogue markup. In *AA-MAS 2009 Workshop, Towards a Standard Markup Language for Embodied Dialogue Acts*, pages 13–24, 2009.

[12] L. D. Catledge and J. E. Pitkow. Characterizing browsing behaviors on the world-wide web. Technical report, Georgia Institute of Technology, 1995.

[13] H. H. Clark and S. E. Brennan. Grounding in communication. In *Perspectives on Socially Shared Cognition*, pages 222–233, 1991.

[14] J. S. Culpepper, F. Diaz, and M. D. Smucker. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in Lorne (SWIRL 2018). *ACM SIGIR Forum*, 52(1):34–90, 2018.

[15] B. Dervin and P. Dewdney. Neutral questioning: A new approach to the reference interview. *RQ*, 25(4):506–513, 1986.

[16] W. Fu and P. Pirolli. Snif-act: A cognitive model of user navigation on the world wide web. *Human–Computer Interaction*, 22(4):355–412, 2007.

[17] D. Gibbon, R. Moore, and R. Winski. *Handbook of standards and resources for spoken language systems.* de Gruyter, 1997.

[18] J. D. Gould, J. Conti, and T. Hovanyecz. Composing letters with a simulated listening typewriter. *Communications of the ACM*, 26(4):295–308, 1983.

[19] E. Hagen. An approach to mixed initiative spoken information retrieval dialogue. *User modeling and User-Adapted Interaction*, 9(1):167–213, 1999.

[20] R. Higashinaka, K. Imamura, T. Meguro, C. Miyazaki, N. Kobayashi, H. Sugiyama, T. Hirano, T. Makino, and Y. Matsuo. Towards an open-domain conversational system fully based on natural language processing. In *Proc. of the Int. Conf. on Computational Linguistics*, COLING'14, pages 928–939, 2014.

[21] M. Kellar, C. Watters, and M. Shepherd. A field study characterizing web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology*, 58(7):999–1018, 2007.

[22] J. Lai, C. Karat, and N. Yankelovich. Conversational speech interfaces and technologies. *Human-Computer Interact. Des. Issues, Solut. Appl.*, pages 53–63, 2009.

[23] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[24] M. Larson and G. J. F. Jones. Spoken content retrieval: A survey of techniques and technologies. *Foundations and Trends in Information Retrieval*, 5(4–5):235–422, 2012.

[25] H. Lausberg and H. Sloetjes. Coding gestural behavior with the neurogeselan system. *Behavior Research Methods*, 41(3):841–849, 2009.

[26] G. Marchionini. *Information seeking in electronic environments.* Number 9. Cambridge university press, 1997.

[27] E. McLellan, K. M. MacQueen, and J. L. Neidig. Beyond the Qualitative

Interview: Data Preparation and Transcription. *Field Methods*, 15(1):63–84, 2003.

[28] M. McTear, Z. Callejas, and D. Griol. *The Conversational Interface*. Springer, 2016.

[29] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000.

[30] J. Nielsen. Ten usability heuristics for user interface design. http://www.nngroup.com/articles/ten-usability-heuristics/, 1995. (Last accessed January 24, 2019).

[31] C. Qu, L. Yang, W. B. Croft, J. R. Trippas, Y. Zhang, and M. Qiu. Analyzing and characterizing user intent in information-seeking conversations. In *Proc. of the Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, SIGIR '18, pages 989–992. ACM, 2018.

[32] F. Radlinski and N. Craswell. A theoretical framework for conversational search. In *Proc. of the ACM SIGIR Conf. on Human Information Interaction and Retrieval*, CHIIR '17, pages 117–126. ACM, 2017.

[33] R. Reichman. *Getting Computers to Talk Like You and Me: Discourse Context, Focus, and Semantics:(an ATN Model)*. MIT press, 1985.

[34] N. Reithinger and E. Maier. Utilizing statistical dialogue act processing in verbmobil. In *Proc. of the 33rd annual meeting on Association for Computational Linguistics*, pages 116–121. ACL, 1995.

[35] T. Saracevic. The stratified model of information retrieval interaction: Extension and applications. In *Proc. of the Annual Meeting-American Society for Information Science*, volume 34, pages 313–327, 1997.

[36] E. A. Schegloff. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1):1–63, 2000.

[37] D. Schiffrin. Conversational coherence: the role of well. *Language*, pages 640–667, 1985.

[38] J. R. Searle. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press, 1969.

[39] S. Sitter and A. Stein. Modeling the illocutionary aspects of information-seeking dialogues. *Information Processing & Management*, 28(2):165–180, 1992.

[40] A. Stein and E. Maier. Structuring collaborative information-seeking dialogues. *Knowledge-Based Syst.*, 8(2-3):82–93, 1995.

[41] A. Stent. Rhetorical structure in dialog. In *Proc. of the first international conference on Natural language generation-Volume 14*, pages 247–252. ACL, 2000.

[42] A. Stent and J. Allen. Annotating argumentation acts in spoken dialog. *University of Rochester, Rochester, NY*, 2000.

[43] H. Sugiyama, T. Meguro, R. Higashinaka, and Y. Minami. Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures. In *Proc. of the Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL'13, pages 334–338. ACL/ISCA, 2013.

[44] R. S. Taylor. The process of asking questions. *Journal of the Association for Information Science and Technology*, 13(4):391–396, 1962.

[45] P. Thomas, D. McDuff, M. Czerwinski, and N. Craswell. MISC: A data set of information-seeking conversations. In *Proc. of the Int. Workshop on Conversational Approaches to Information Retrieval*, 2017.

[46] P. Thomas, M. Czerwinski, D. McDuff, N. Craswell, and G. Mark. Style and alignment in information-seeking conversation. In *Proc. of the ACM SIGIR Conf. on Human Information Interaction and Retrieval*, CHIIR '18, pages 42–51. ACM, 2018.

[47] K. Tracy and N. Coupland. Multiple goals in discourse: An overview of issues. *Journal of Language and Social Psychology*, 9(1-2):1–13, 1990.

[48] D. R. Traum. Speech acts for dialogue agents. In *Foundations of Rational Agency*, pages 169–201. Springer, 1999.

[49] J. R. Trippas. *Spoken Conversational Search: Audio-only Interactive Information Retrieval*. PhD thesis, RMIT, Melbourne, 2019.

[50] J. R. Trippas and P. Thomas. Data sets for spoken conversational search. In *Proc. of the CHIIR 2019 Workshop on Barriers to Interactive IR Resources Re-use (BIIRRR 2019)*, 2019.

[51] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. How do people interact in conversational speech-only search tasks: A preliminary analysis. In *Proc. of the ACM SIGIR Conf. on Human Information Interaction and Retrieval*, CHIIR '17, pages 325–328. ACM, 2017.

[52] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. A Conversational Search Transcription Protocol and Analysis. In *Proc. of the Int. Workshop on Conversational Approaches to Information Retrieval*, 2017.

[53] J. R. Trippas, D. Spina, L. Cavedon, H. Joho, and M. Sanderson. Informing the design of spoken conversational search: Perspective paper. In *Proc. of the ACM SIGIR Conf. on Human Information Interaction and Retrieval*, CHIIR '18, pages 32–41. ACM, 2018.

[54] S. Vakulenko, K. Revoredo, C. Di Ciccio, and M. de Rijke. QRFA: A data-driven model of information-seeking dialogues. In *Proc. of the European Conf. on Information Retrieval*, ECIR'19, pages 541–557, 2019.

[55] M. A. Walker, R. Passonneau, and J. E. Boland. Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems. In *Proc. of the Annual Meeting on Association for Computational Linguistics*, ACL '01, pages 515–522. ACL, 2001.

[56] W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, and A. Stolcke. The microsoft 2017 conversational speech recognition system. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, ICASSP '18, pages 5934–5938. IEEE, 2018.

[57] L. Yang, M. Qiu, C. Qu, J. Guo, Y. Zhang, W. B. Croft, J. Huang, and H. Chen. Response ranking with deep matching networks and external knowledge in information-seeking conversation systems. In *Proc. of the Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, SIGIR '18, pages 245–254. ACM, 2018.

[58] N. Yankelovich. Using natural dialogs as the basis for speech interface design. In *Human Factors and Voice Interactive Systems*, pages 255–290. Springer, 2008.

[59] E. Zarisheva and T. Scheffler. Dialog act annotation for twitter conversations. In *Proc. of the Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL'15, pages 114–123. ACL/ISCA, 2015.

[60] I. Zuckerman, E. Segal-Halevi, A. Rosenfeld, and S. Kraus. First steps in chat-based negotiating agents. In *Next Frontier in Agent-based Complex Automated Negotiation*, pages 89–109. Springer, 2015.

Appendix A. Themes, Sub-themes, and Codes in SCoSAS

Appendix A.1. Theme 1: Task Level

Table A.5: Information Request Codes (Seeker)

Theme	Sub-theme	Actor	Code	Frequency
Task Level	Information Request	Seeker	Automated repetitive search [53]	3
			Definition explanation (or clarification, i.e., intent clarification)	1
			(Information request for a) Definition lookup or person	1
			(Requesting) Information about document [53]	6
			Information about SERP overview	2
			Information request	67
			Information request within document	80
			Information request within SERP	15
			Initial information request	39
			Intent clarification	52
			Query embellishment [51]	20
			Spells (query or query word)	2
		Intermediary	(Requests) Definition clarification (i.e., requests more details about the information request)	1
			Enquiry for further information	11
			Google query expansion suggestion	3
			Query refinement offer	57
			Query rephrase	12
			Requests more details about information request	5
			Query formulation for information found in document	1
			Asking what they (i.e., the Seeker) are looking for	2
			Within-Document search result entity lookup request	1

Table A.6: Result Presentation Codes (Intermediary)

Theme	Sub-theme	Actor	Code	Frequency
Task Level	Results Presentation	Intermediary	Source information	8
			Image overview on SERP	2
			Interpretation of photos	1
			Multi-document summary	3
			Paraphrasing from document which is not in front of them	1
			Scanning document with modification	51
			Scanning document without modification	79
			Scanning document without modification but with interpretation of photos	1
			SERP Card	16
			SERP overview without modification	1
			SERP with modification	19
			SERP without modification	72
			Within SERP search result	4
			Within-Document command response	1
			Within-Document search result	60
			Interpretation biased towards information request or clarification given by the User	1
			Comparing results against each other	1
			Interpretation	22

Table A.7: Search Assistance (Seeker and Intermediary)

Theme	Sub-theme	Actor	Code	Frequency
Task Level	Search Assistance	Seeker	(Requests further search) Recommendations	1
			Requests "enough information" judgement	1
		Intermediary	Asking about usefulness (of presented result)	4
			Requests spelling	2
			Suggestion to move on	2
			Relevance judgement	6
			Suggestion to search more	1
			Requests to access search engine	1
			Search suggestion based on info encountered in document	1

Table A.8: Search Progression (Seeker)

Theme	Sub-theme	Actor	Code	Frequency
Task Level	Search Progression	Seeker	Enough information	6
			Performance feedback	18
			Rejects (suggestion from Intermediary)	9

Table A.9: Discourse Management (Seeker and Intermediary)

Theme	Sub-theme	Actor	Code	Frequency
Discourse Level	Discourse Management	Seeker	Asks to repeat	31
			Asks to repeat first search result	6
			Asks to repeat Nth search result	1
			Confirms	114
			Query repeat	14
		Intermediary	Asks to repeat	38
			Checks navigational command	13
			Confirms	46
			Repeats	12
			Repeats the query back	9

Table A.10: Grounding (Seeker)

Theme	Sub-theme	Actor	Code	Frequency
Discourse Level	Grounding	Seeker	Creating bigger picture	1
			Interpretation	12

Table A.11: Navigation (Seeker)

Theme	Sub-theme	Actor	Code	Frequency
Discourse Level	Navigation	Seeker	Access link within document	1
			Access search engine	2
			Access source	29
			Access source (implicit)	2
			Between-document navigation	1
			Is there more information	6
			Leave document	1
			Next	3
			Read more from the document	1
			Within-document command	3

Table A.12: Visibility of System Status (Seeker and Intermediary)

Theme	Sub-theme	Actor	Code	Frequency
Discourse Level	Visibility of system status	Seeker	Access source feedback-request	3
			Feedback on what is happening	1
			Results?	10
		Intermediary	Feedback on what is happening	13
			Misheard	1
			Previously seen results	2
			Wayfinding	3

Appendix A.3. Theme 4: Other Level

Table A.13: Other Level (Seeker)

Theme	Actor	Code	Frequency
Other Level	Seeker	Utter ("So I'm" and "Well so they are saying")	2
		Provides information about the Search Engine ("So it's [a] search engine")	1
		Asks if allowed to query embellish ("Actually can I add something else to that?")	1
		Offers to spell ("[...] would you like me to spell it?")	1

Appendix B. Tasks and Backstories

Table B.14: Example Search Tasks

Dimension	Query and Example Backstory
Remember	**What river runs through Rome, Italy?** Many great cities have rivers running through them, as rivers facilitated trade and commerce as well as supplying fresh water to drink. You remember that Paris has the Seine, London has the Thames, but what does Rome have? **What language do they speak in New Caledonia?** You and your partner are thinking of places to go on holiday. New Caledonia is an option, but you realize you don't know what language is spoken there and you decide to find out. **Where does cinnamon come from?** The other day you were eating some spiced biscuits from Europe, when it occurred to you that cinnamon probably isn't native to that part of the world. You would like to know where it comes from.
Understand	**recycle, automobile tires** You need to buy new tires for your car, and the local dealer has offered to take the old ones for recycling. You didn't know tires could be recycled and you wonder what new uses they are being put to. **Outsource job India** A recent report on the radio quoted a politician as saying that one of the causes of rising unemployment in the U.S. was the outsourcing of jobs to India. This has made you interested in finding out what jobs that used to be in the U.S. have been outsourced to India. **Marine Vegetation** You recently heard a commercial about the health benefits of eating algae, seaweed and kelp. This made you interested in finding out about the positive uses of marine vegetation, both as a source of food, and as a potentially useful drug.
Analyse	**Turkey Iraq Water** Looking at a map, you realize that there are several rivers that commence in Turkey and then flow over the border into Iraq. You wonder if Turkish river control projects, including dams and irrigation schemes, have affected Iraqi water resources. **Airport Security** Every time you go through the security screening at an airport, you wonder whether it is making any difference. Find out how effective the many new measures (beyond just standard screening) at airports actually are, both for scrutinizing of passengers and their checked and carry-on baggage. **per capita alcohol consumption** You recently attended a big party and woke up with a hangover, and have decided to learn more about the average consumption of alcohol. You are particularly interested in any information that reports per capita consumption, and want to compare across groups, for example at the country, state, or province Level.