A Conversational Search Transcription Protocol and Analysis

Johanne R. Trippas
RMIT University
johanne.trippas@rmit.edu.au

Damiano Spina
RMIT University
damiano.spina@rmit.edu.au

Lawrence Cavedon
RMIT University
lawrence.cavedon@rmit.edu.au

Mark Sanderson
RMIT University
mark.sanderson@rmit.edu.au

ABSTRACT

The analysis of spoken interactions between human participants, for the purpose of understanding tactics and strategies for performing a task such as collaborative search, involves a number of steps, completing with the identification and classification of operations designed to interpret search results or to progress the interaction to a successful outcome. This paper provides guidelines for transcription and analysis of conversational search recordings which are critical steps for the overall goal in identifying and classifying operations in conversational search. Many decisions need to be made when transcribing utterances, such as determining pauses, punctuation, or when a speaker stops or starts uttering. To our knowledge, no publicly available guidelines are available for conversational search transcription and analysis, thus we adapted several different guidelines from other fields. In this paper we provide fundamental principles for transcriptions and a protocol based on those principles. We also introduce annotation tools, transcription quality assurance guidance, analysis and coding instructions.

CCS CONCEPTS

•**Information systems** →**Information retrieval;** *Collaborative search; Speech / audio search;*

KEYWORDS

Conversational Search; Transcription Protocol; Qualitative Analysis

ACM Reference format:
Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson. 2017. A Conversational Search Transcription Protocol and Analysis. In *Proceedings of ACM SIGIR Workshop on Conversational Approaches for Information Retrieval, Tokyo, Japan, August 2017 (CAIR'17),* 5 pages.

1 INTRODUCTION

The increasing research interests in conversational search has resulted in a consequent growth in recordings of spoken search interactions. Such recordings are a valuable source of data for understanding how users interact in this particular search setting and tactics used for driving effective search performance. Yet the transient nature of audio necessitates a more permanent form of the data to facilitate analysis. This is achieved through transcripts (i.e., creating a written version of audio). Thus, generating transcripts is a critical first step in the process of understanding conversational search. However, transcribing throws up a number of challenges in an interactive scenario, particularly when two human participants are involved. These challenges include understanding how to represent turn-taking (i.e., utterances where a participant speaks for a certain period of time) or punctuating utterances.

Analysis of such interactions or transcriptions requires a number of important steps, culminating in the identification of actual dialogue operations that interpret results or drive further interaction. In general spoken dialogue research, this involves the definition and classification of utterances using *dialogue acts* [5]. These acts may be generic or tailored to domain- or task-specific circumstances [1, 2]. Thus, thematic analysis can be used in order to systematically identify, analyze and report patterns in the transcripts [3].

Other fields have guidelines for both transcriptions and analysis (e.g., social sciences [4], automatic speech recognition [17]), however, to our knowledge, there are no publicly available guidelines for conversational search. Given the importance of consistent research techniques in order to establish a body of comparable work, we propose in this paper a protocol for spoken search interactions which includes data preparation, quality assurance, and analysis to assist future researchers in the field.

The methodology in this paper was used in our previous work which presented a study designed to understand how users conduct searches over voice where a screen is absent but where users can converse interactively with the search system [18]. We recorded spoken interactions between two participants in a search setting where one participant received a search task (*User*) and the other had access to a search engine (*Retriever*). The communication between participants was analyzed for interaction patterns used in the search process. We used these patterns to create an initial annotation schema of spoken conversational search interactions. There are no factors that make the protocol task dependent; however, the generalizability of the protocol has not yet been explored –it has been applied only in the settings described in [18].

We firstly present the transcription methodology. Then, we describe the qualitative analysis for those conversational search transcriptions.

2 TRANSCRIPTION METHODOLOGY

In this section, we present general transcription principles followed by more detailed examples of how these principles can be translated into a protocol. We then suggest tools for transcription and present quality assurance processes.

CAIR'17, Tokyo, Japan

While various transcribing techniques are available such as Automatic Speech Recognition (ASR) [13], crowdsourcing [16], or manual transcriptions [14], none of these techniques are perfect and therefore we strongly encourage incorporating these quality assurance processes. The recent progress in ASR allows for improved automatic conversational transcriptions [20]. Nevertheless, manual checking of ASR transcriptions is still necessary. High quality transcriptions could then also be suitable both for indexing purposes and for result representations of these verbalised documents [9].

2.1 Transcription Principles

This section discusses the transcription principles used in preparing the recordings for analysis. Although different researchers have tried to form guidelines allowing for consistent transcriptions, no universal rules are available for transcriptions of video or audio recordings [11]. Therefore we propose guidelines for transcriptions in conversational search which have been obtained from McLellan, MacQueen, and Neiding [11].

Deciding what to transcribe is left to the researcher and the purpose of the research questions, and the level of transcription should complement the level of the analysis [6]. For example, researchers exploring general themes and patterns may not need a detailed transcription [11]. However, if the transcription is used for a test collection for conversational search, it is necessary to include other information such as relevance judgments.

We followed these principles allowing for high quality transcriptions. In the transcription process we wrote what was said, thus we did not include non-linguistic observations such as facial expressions and body language, or intonations. Our transcription is therefore verbatim and often referred to as orthographic transcription. Including non-linguistic observations adds a level of complexity to the transcriptions and results in a more costly process.

For our transcriptions we wanted to capture how people expressed themselves in a search situation and therefore transcribed all recorded utterances with the following transcription principles [11]:

(1) *Preserve the morphological naturalness of transcription.* Keep word forms, the form of commentaries, and the use of punctuation as close as possible to speech presentation and consistent with what is typically acceptable in written text.
(2) *Preserve the naturalness of the transcript structure.* Keep text clearly structured by speech markers (i.e., like printed versions of plays or movie scripts).
(3) *The transcript should be an exact reproduction.* Generate a verbatim account. Do not prematurely reduce text.
(4) *The transcription rules should be universal.* Make transcripts suitable for both human/researcher and computer use.
(5) *The transcription rules should be complete.* Transcribers should require only these rules to prepare transcripts. Everyday language competence rather than specific knowledge (e.g., linguistic theories) should be required.
(6) *The transcription rules should be independent.* Transcription standards should be independent of transcribers as well as understandable and applicable by researchers or third parties.
(7) *The transcription rules should be intellectually elegant.* Keep rules limited in number, simple, and easy to learn.

We chose ELAN[1] as a transcription tool but the principles can be used with other tools [10]. ELAN accommodated the use of the above principles and precise transcription protocol (See Section 2.2). These principles and rules allowed us to create high quality transcripts with an iterative manner which was systematic and consistent. However, it is important to note that transcriptions of videos never completely embody all details which take place in the recording [8]. Researchers have to determine the scope of their transcription needs in order to answer their research questions.

2.2 Transcription Protocol

Transcription protocols have two main goals: minimizing the probability that the transcripts produced are inconsistent; and reducing the likelihood that the data analysis will be weakened or delayed [11]. To avoid these two main problems we used the following transcription protocol adapted from [4, 11].

- Turns were identified and every first word of each new turn was capitalized.
- Audiotapes were transcribed verbatim (i.e., recorded word for word, exactly as said), non-complete words or sentences were transcribed to the best of the transcriber's ability. Nonverbal or background sounds were not included (e.g., laughter, sighs, or coughs).
- If participants mispronounced words, these words were transcribed as the individual said them. The transcript was not "cleaned up" by removing slang, grammatical errors, or misuse of words.
- While "aha", "hmm" or "uhm" were included, linguistic- or phonetic-type transcripts were not produced making the transcripts more accessible for other researchers.
- Abbreviations were written as said, such as "TV" for "television". No abbreviations were written if they were not used by the participants.
- Numbers were all spelled out (e.g., "90" is written as "ninety").
- Spelled out words were capitalized (e.g., User spells the country "New Caledonia" which is transcribed as "NEW CALEDONIA").
- URLs were written as pronounced (e.g., "drive dot com dot AU").
- Place names and brand names were written with an initial capital.
- Portions of the audiotape that were inaudible or difficult to decipher were identified as "inaudible segment". This information was transcribed as *[inaudible segment]*.
- Pauses in speech were indicated with an ellipsis. A brief pause was defined as a two- to five- second break in speech. Pauses longer than five seconds were transcribed as *[long pause]*.
- A style guide with vocabulary was kept throughout the project.

We did not note overlapping speech since this was not in the scope of our analysis [11].

[1] http://tla.mpi.nl/tools/tla-tools/elan/

2.3 Transcription Tools

Qualitative Data Analysis (QDA) tools such as ELAN help organize transcriptions. Even though our protocol called for verbatim transcribing, the transcript is the product of interactions between the recordings and transcriber, who will make choices about what to preserve and how to transcribe [4]. Thus different QDA tools may lead to different transcriptions. As a result, the QDA tool and analysis should be decided before the transcription process [15]. QDA tools dictate how a transcription is formatted. For example, in some tools only plain text is allowed while in other tools one can use formatted text. The output of the transcription is also dictated by the QDA tool. In our example, we exported the transcription file with labels and converted the exported file into a CSV file for further analysis in R.

Our transcriptions principles helped us to organize and analyze all the data independently of the QDA tool used [10]. ELAN did not impose constraints on the data collected and accommodated our iterative process in relationship to our research questions. Other projects may require a different kind of transcription and may not need QDA tools or different tools such as Nvivo,[2] Atlas.ti,[3] or MAXQDA.[4] Note, many QDA tools do not support video analysis.

2.4 Transcription Quality Assurance

In a quantitative database, data cleanup and recording are performed before analysis is undertaken. With qualitative data, these processes generally proceed simultaneously [11]. With smaller scale research projects, transcription is usually handled by the researcher, and a continuous or iterative process between transcription and data interpretation occurs [12].

In order to provide highly accurate transcripts we developed our process for reading and reviewing the text whereby each recording was listened to three times against the transcript before it was submitted, also referred to as the three-pass-per-tape policy [11]. All transcripts were audited for accuracy by a professional editor.

Even the most proficient transcriber misses a word or two or transcribes some phrases that are slightly different from what was actually said [19]. Therefore, it is recommended to proofread all or a random selection of transcripts by checking the final transcript against the audio [11].

3 QUALITATIVE ANALYSIS

This section introduces a qualitative analysis method thematic analysis and presents an applied example of coding and analysis of themes.

3.1 Thematic Analysis

Thematic analysis involves identifying, analyzing, and reporting patterns (themes) within the qualitative data [3, 4]. This method allows for analyzing qualitative data in an accessible and theoretically flexible manner and it is often seen as a fundamental way for analysing qualitative data [3]. We adopted the six-step process as outlined by Braun and Clarke [4]: (step 1) familiarizing 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. The first author generated initial codes and categorized them into themes and sub-themes.

Thematic analysis in conversational search can be used for creating new information seeking models or identifying issues in particular search stages. For example, a map of the identified search stages created by the six-step process can lead to a formal information seeking model. Simultaneously, one stage (e.g., examine results) could be investigated for particular interests (e.g., sensemaking). Thus thematic analysis allows for systematical investigations of a non-functional system. Those investigations would inform the design of new presentation strategies for conversational search.

3.2 Coding of Transcriptions

The conversational search task we transcribed were pairs of participants in a spoken conversational search experiment [18]. The pairs conducted searches where one participant acted as the User (participant with the search task) and the other acted as the Retriever (participant with the search engine). Users and Retrievers did not have access to each others' search task or search engine interface, could not see each other, and could communicate only verbally. This setup can be seen in Figure 1 ((a) User, (b) Retriever).

Both participants were recorded during the session as well as the Retriever's screen. The recordings were synchronized and merged for transcription. Recordings were transcribed and coded in order of their chronological occurrence. The codes were created on the basis of the video and transcriptions in ELAN. We adopted the following steps:

Step 1: Identifying when each participant spoke, i.e. identifying turns.
Step 2: Transcribing turn.
Step 3: Assigning codes to each turn with ELAN. Observational notes were added.
- Adding *Controlled Vocabulary* and descriptions of the Controlled Vocabulary to ELAN and spreadsheet for crosschecking. The Controlled Vocabulary can be seen as a *dictionary* which is created during coding. This dictionary is then developed in to a full *codebook*.

Step 4: Combining codes to themes for further analysis.
Step 5: Quality assurance: Spelling check, checking if all turns received codes done through exporting of transcriptions and codes with ELAN to text file.
Step 6: Converting transcription and code text file to CSV file.
Step 7: Importing CSV files into R and aggregating codes to check whether they have been coded correctly. The aggregation of codes allows for checking whether all codes belong to the same category.
Note: Steps 3–7 were conducted iteratively and the codebook was updated.

Instances where either User or Retriever was unintelligible while reading were not transcribed. Instead these instances were coded as *[inaudible segment]* (See Section 2.2). We assumed that if the audio recording was not clear, it was probably not clear to the other participant either.

[2] http://www.qsrinternational.com/nvivo-product
[3] http://atlasti.com/
[4] http://www.maxqda.com/

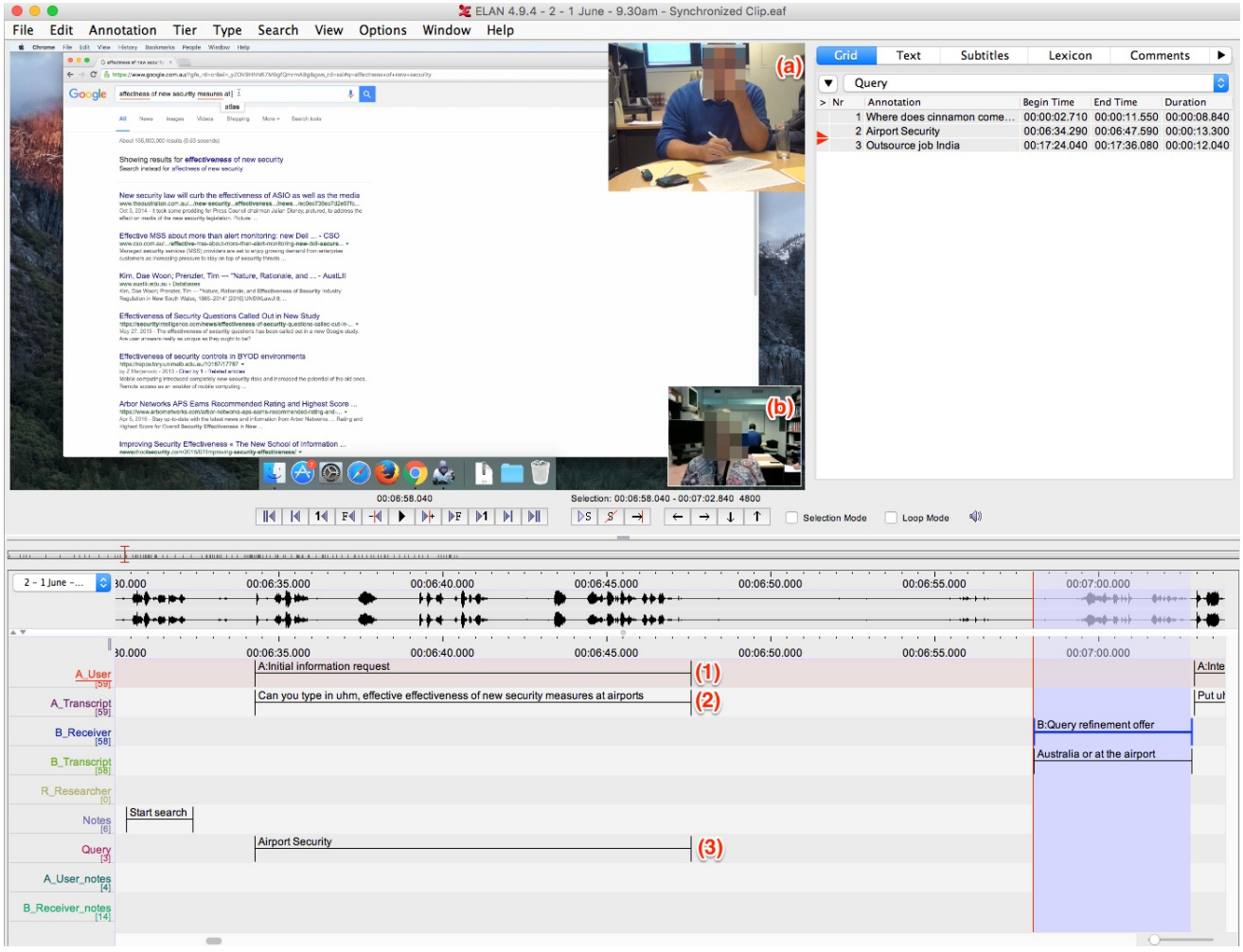

Figure 1: Sample screenshot of ELAN transcription and analysis tool (anonymized). Annotations indicate (a) User, (b) Retriever, (1) Controlled vocabulary User, (2) Transcription, (3) Query.

3.3 Analysis of Coding

The ELAN transcriptions with codes and observational notes were transferred to R for further analysis. The transcriptions were modified to lower case, and punctuation and extra spacing were removed. The fill-word "uhm" was removed for analysis purposes. However, we deliberately did not remove any errors, false starts or confirmations since these occur in real case voice search scenarios.

In the context of mixed initiative information retrieval dialogues, the terms *control* and *initiative* are used interchangeably. However, we used the approach of *taking the initiative equals taking the turn*, as described by [7]. This means that one turn can consist of multiple moves or communication goals. We coded the complete dataset identifying anything of relevance to our research aim. Thus, codes were applied to each turn taken by either User or Retriever and these codes were collated and given *themes*. Themes may consist of *sub-themes* which capture specific concepts of that theme. Themes were created independently of the previous turn meaning that each turn may consist of similar themes or sub-themes.

4 SUMMARY

In this paper we proposed a protocol for transcribing and annotating conversational search observations. We summed up broad transcription principles and detailed protocols. Transcription and annotation tools were introduced as well as the notion of quality assurance policies. We provided a step by step approach of a qualitative analysis of conversational search transcriptions.

The proposed protocol has been used to understand how users conduct searches over voice where a screen is absent but where users can converse interactively with the search system [18].[5] We envisage that this protocol could be used in further research studies on conversational and spoken collaborative search, fostering the reproducibility of studies and comparability of the results.

[5]Transcripts and annotations are available at https://jtrippas.github.io/Spoken-Conversational-Search.

ACKNOWLEDGMENTS

This research was partially supported by Australian Research Council Project LP130100563 and Real Thing Entertainment Pty Ltd.

REFERENCES

[1] J. Allen and M. Core. Draft of DAMSL: Dialog act markup in several layers. Technical report, University of Rochester, 1997. Retrieved July 14, 2017 from http://www.cs.rochester.edu/research/cisd/resources/damsl/.

[2] S. Bangalore, G. Di Fabbrizio, and A. Stent. Learning the structure of task-driven human–human dialogs. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(7):1249–1259, 2008.

[3] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

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

[5] H. Bunt. Conversational principles in question-answer dialogues. *Zur Theory der Frage*, pages 119–141, 1981.

[6] J. W. Drisko. Strengthening qualitative studies and reports: Standards to promote academic integrity. *Journal of social work education*, 33(1):185–197, 1997.

[7] E. Hagen. An approach to mixed initiative spoken information retrieval dialogue. In *Computational Models of Mixed-Initiative Interaction*, pages 351–397. Springer, 1999.

[8] S. Kvale. Interviews: An introduction to qualitative research interviewing. *Lund: Studentlitteratur*, 1996.

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

[10] H. Lausberg and H. Sloetjes. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior research methods*, 41(3):841–849, 2009.

[11] E. McLellan, K. MacQueen, and J. Neidig. Beyond the Qualitative Interview: Data Preparation and Transcription. *Field methods*, 15(1):63–84, 2003.

[12] M. B. Miles and A. M. Huberman. *Qualitative data analysis: An expanded sourcebook*. Sage, 1994.

[13] M. Mulholland, M. Lopez, K. Evanini, A. Loukina, and Y. Qian. A comparison of ASR4 and human errors for transcription of non-native spontaneous speech. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 5855–5859. IEEE, 2016.

[14] S. Shiga, H. Joho, R. Blanco, J. R. Trippas, and M. Sanderson. Modelling information needs in collaborative search conversations. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017.

[15] C. Silver and A. Lewins. *Using software in qualitative research: A step-by-step guide*. Sage, 2014.

[16] R. Sprugnoli, G. Moretti, L. Bentivogli, and D. Giuliani. Creating a ground truth multilingual dataset of news and talk show transcriptions through crowdsourcing. *Language Resources and Evaluation*, pages 1–35, 2016.

[17] Telephone Speech Collection Group. Transcription guidelines (NQTR) - draft, 2006. Retrieved July 14, 2017 from https://catalog.ldc.upenn.edu/docs/LDC2010S01/trans_guide_nqrt_span.doc.

[18] J. R. Trippas, D. Spina, L. Cavedon, and M. Sanderson. 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 (CHIIR)*, pages 325–328. ACM, 2017.

[19] R. S. Weiss. *Learning from strangers: The art and method of qualitative interview studies*. Simon and Schuster, 1995.

[20] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, and G. Zweig. Achieving human parity in conversational speech recognition. *arXiv preprint arXiv:1610.05256*, 2016.