

A Corpus Study on Questions, Responses and Misunderstanding Signals in Conversations with Alzheimer’s Patients*

Shamila Nasreen, Matthew Purver, Julian Hough

Cognitive Science Group / Computational Linguistics Lab

School of Electronic Engineering and Computer Science

Queen Mary University of London Mile End Road,

London E1 4NS, UK

{shamila.nasreen,m.purver,j.hough}@qmul.ac.uk

Abstract

This paper describes an initial corpus study of question-answer pairs in the Carolina Conversations Collection corpus of conversational interviews with older people. Our aim is to compare the behaviour of patients with and without Alzheimer’s Disease (AD) on the basis of types of question asked and their responses in dialogue. It has been suggested that questions present an interesting and useful phenomenon for exploring the quality of communication between patients and their interlocutors, and this study confirms this: questions are common, making up almost 14% of utterances from AD and Non-AD patients; and type distributions vary, interviewers asking many Yes-No questions (nearly 6%) from AD patients while more Wh-questions (5.4%) from Non-AD patients. We also find that processes of clarification and coordination (e.g. asking clarification questions, signalling non-understanding) are more common in dialogue with AD patients.

1 Introduction

Alzheimer’s Disease (AD) is an irreversible, progressive deterioration of the brain that slowly destroys memory, language and thinking abilities, and eventually the ability to carry out the simplest tasks in patients’ daily lives. AD is the most prevalent form of dementia, contributing to 60%-70% among all types of dementia (Tsoi et al., 2018). The most common symptoms of AD are memory lapses, difficulty in recalling recent events, struggling to follow a conversation, repeating the

conversation, delayed responses, difficulty finding words for talk, and orientation problems (e.g. confusion and inability to track daily activities).

Diagnosis can be based on clinical interpretation of patients’ history complemented by brain scanning (MRI); but this is time-consuming, stressful, costly and often cannot be offered to all patients complaining about functional memory. Instead, the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and Alzheimer’s Association established criteria for AD diagnosis require the presence of cognitive impairment to be confirmed by neuropsychological testing for a clinical diagnosis of possible or probable AD (McKhann et al., 1984). Suitable neuropsychological tests include the Mini-Mental Status Examination (MMSE; Folstein et al., 1975, one of the most commonly used tests), Mini-Cog (Rosen et al., 1984), Addenbrooke’s Cognitive Examination Revised (ACE-R; Noone, 2015), Hopkins Verbal Learning Test (HVL; Brandt, 1991) and DemTect (Kalbe et al., 2004).

However, these tests require medical experts to interpret the results, and are performed in medical clinics which patients must visit for diagnosis. Currently, researchers are therefore investigating the impact of neurodegenerative impairment on patients’ speech and language, with the hope of deriving tests which are easier to administer and automate via natural language processing techniques (see e.g. Fraser et al., 2016a).

In this paper, we focus on language in conversational interaction. We explore this as a diagnostically relevant resource to differentiate patients with and without Alzheimer’s Disease (AD vs. Non-AD), using the Carolina Conversations Collection data in which patients interact with researchers and community persons on different but not prefixed topics like discussion about breakfast, lunch, special occasions (thanksgiving, Christ-

*This research was partially supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825153, project EMBED-DIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media). The results of this publication reflect only the authors’ views and the Commission is not responsible for any use that may be made of the information it contains.

mas) etc. We particularly focused on the types of questions asked from both groups, how they are responded to, and whether there are any significant patterns that appear to differentiate the groups.

The remainder of this paper is organized as follows. In the next section, we describe earlier work on analyzing conversational profiles of AD and particularly on the types of questions they focused on. In Section 3 we give details about our new corpus study and annotation scheme. We then present and discuss the results in Section 4: in particular, how the distributions of different types of questions, and frequencies of signals of non-understanding, clarification questions and repeat questions, differ between AD patients and Non-AD. We close with a discussion of the overall result, and of possible further work.

2 Related Work

Recent years have seen an increasing amount of research in NLP for dementia diagnosis. Much of this work has looked at properties of an individual’s language in isolation: acoustic and lexical features of speech (Jarrold et al., 2014), or syntactic complexity, fluency and information content (Fraser et al., 2016b,a). However, this is usually studied within particular language tasks, often within specific domains (e.g. the Cookie Theft picture description task of the DementiaBank Pitt Corpus¹); however, conversational dialogue is the primary area of human natural language use, and studying the effects of AD on dialogue and interaction — and particularly more open-domain dialogue — might therefore provide more generally applicable insights.

Recent statistical modelling work shows that AD has characteristic effects on dialogue. Luz et al. (2018) extract features like speech rate, dialogue duration and turn taking measures, using the Carolina Conversations Collection corpus (Pope and Davis, 2011) of patient interview dialogues, and show that this can build a predictive statistical model for the presence of AD.

Work in the conversation analysis (CA) tradition has looked in more detail at what characteristics of dialogue with dementia might be important. Jones et al. (2016) present a CA study of dyadic communication between clinicians and patients during initial specialist clinic visits, while Elsey et al. (2015) highlighted the role of carer,

looking at triadic interactions among a doctor, a patient and a companion. They establish differential conversational profiles which distinguish between non-progressive functional memory disorder (FMD) and progressive neuro-degenerative Disorder (ND), based on the interactional behavior of patients responding to neurologists’ questions about their memory problems. Features include difficulties responding to compound questions, providing specific and elaborated examples and answering questions about personal information, time taken to respond and frequent “I don’t know” responses.

Questions present an interesting testing ground when exploring the quality of communication between caregivers and persons with AD. Question-answer sequences have long been seen as a fundamental building block of conversation; Sacks et al. (1978) formalized these as a type of adjacency pair in which the first utterance represents the question and the second one is an answer. Hamilton (2005) explored the use of questions in conversation with a patient of AD over a period of four years, finding that Yes-No questions are responded to much more frequently than open-ended question i.e. Wh-questions. Gottlieb-Tanaka et al. (2003) used a similar approach, examining Yes-No and open-ended questions in a conversation between family caregivers and their spouse with AD during different activities of daily life. They reported that caregivers used YesNo questions much more frequently than open-ended questions (66% vs. 34%, respectively) and there are fewer communication breakdowns with Yes-No Questions.

Varela Suárez (2018) worked specifically to observe dementia patients’ ability to respond to different types of questions including close-ended questions, open-ended questions, and multiple choice questions. The objective of this study was to verify a) if the ability to answer questions persists until the final stages of dementia b), check if the number of preferred and relevant answers decreases progressively. The interviewers had a list of questions about patients memories, experiences, and daily routine, and were told to talk on the topics introduced by the patients, and only ask the questions from the list when patients are silent. The basic Question-Answer adjacency pair is preserved until the severe stage of the disease; however, the number of answered questions, preferred and relevant answers starts to decrease.

¹<http://talkbank.org/DementiaBank/>

These studies show that the presence of AD affects the production of questions, their use and their responses, but all focus on specific types of question including Yes-No, Wh-questions, and Multiple choice questions. As far as we are aware, none of these studies have extended this approach to look into specific aspects of non-understanding or inability to respond: e.g. non-understanding signals, clarification requests and repetition of questions.

Dialogue Act Models

The ability to model and detect discourse structure is an important step toward working spontaneous dialogue and the first analysis step involves the identification of Dialogue Acts (DAs). DAs represent the meaning of utterances at the level of illocutionary force (Stolcke et al., 2000). Classifying utterances and assigning DAs is very useful in many applications including answering questions in conversational agents, summarizing meeting minutes, and assigning proper DAs in dialogue based games. DAs tagsets classify dialogue utterances based on the syntactic, semantic and pragmatic structure of the utterance.

The most widely used dataset and tagset in DA tagging is the Switchboard corpus, consisting of 1155 annotated conversations containing 205K utterances, 1.4 million words from 5 minute recorded telephonic conversations. The DA types and complete tagset can be seen in (Jurafsky et al., 1997). The corpus is annotated using a variant of the DAMSL tagset (Core and Allen, 1997) with approximately 60 basic tags/classes which combines to produce 220 distinct labels. Jurafsky et al. (1997) then combine these 220 labels into 43 major classes including *Statements*, *Backchannels*, *Questions*, *Agreements*, *Apology* etc.

3 Material and Methods

3.1 Research Questions

This study is a part of a larger project where we analyze what are the significant key indicators in the language and speech of AD patients that can be used as Bio-Markers in the early diagnosis process of Alzheimer’s Disease. The focus of the initial and current study is on the interaction of AD patients and Non-AD patients with interviewers.

Our account suggests these interactions are based on what is being asked from the AD and Non-AD sufferers. We hypothesize that the distri-

bution of questions being asked and the responses generated are not same for both the groups. We hypothesize that the use of different question types such as binary yes-no questions (in interrogative or declarative form), tag questions, and alternative (‘or’) questions will differ between groups; and the signals of non-understanding, back-channels in question form and clarification requests should be more common with AD patients.

In more detail, we are conducting this corpus study to answer the following research questions:

Q1 *Is the distribution of question types asked by the patient and interviewer different when the patient is an AD sufferer?*

Our first interest is in the general statistics regarding what types of questions are asked of the AD and non-AD group. How often does each type occur, and what is the balance between the two groups? What types of questions are more frequently asked from Alzheimer’s patients?

Q2 *How often do signals of non-understanding, clarification requests and back-channel questions occur in dialogues with an AD sufferer compared to those without one?*

We hypothesize that due to the nature of AD, there will be more non-understanding signals and clarification questions in response to questions and statements.

Q3 *Is the distribution of simple-repeat and reformulation questions different for conversations with an AD sufferer compared to those without one?*

We hypothesize that there will be more repeated questions for the AD group from the interviewer, as AD patients find it difficult to follow a conversation.

3.2 Corpus

Our intention was to investigate the behavior of AD patients on the basis of questions and responses observed in a corpus of dialogue. For this purpose, we used the Carolina Conversation Collection (CCC), collected by the Medical University of South Carolina (MUSC)² (Pope and Davis, 2011). This dataset comprises of two cohorts: cohort one contains 125 unimpaired persons of 65

²<https://carolinaconversations.musc.edu/>

years and older with 12 chronic diseases with a total of 200 conversations. Cohort two includes 400 natural conversations of 125 persons having dementia including Alzheimer’s of age 65 and above who spoke at least twice annually with linguistic students. The demographic and clinical variables include: age range, gender, occupation prior to retirement, diseases diagnosed, and level of education (in years) are available. As this dataset includes only older patients with diagnosed dementia, it can only allow us to observe patterns associated with AD at a relatively advanced stage, and not directly tell us whether these extend to early stage diagnosis. However, it has the advantage of containing relatively free conversational interaction, rather than the more formulaic tasks in e.g. DementiaBank. Work in progress is collecting a dataset of conversational language including early-stage and un-diagnosed cases; until then we believe this to be the most relevant corpus for our purposes.

The dataset consists of audio, video and transcripts that are time aligned. The identity of patients and interviewer is anonymized keeping in mind security and privacy concerns. Online access to the dataset was obtained after gaining ethical approval from Queen Mary University of London (hosting the project) and Medical University of South Carolina (MUSC, hosting the dataset), and complying with MUSC’s requirements for data handling and storage.

For our corpus analysis here, we used dialogue data from 10 randomly sampled patients with AD (7 females, 3 males) and 10 patients with other diseases including diabetes, heart problems, arthritis, high cholesterol, cancer, leukemia and breathing problems but not AD (8 females, 2 males). These groups are selected to match age range, to compare the different patterns of interaction and to avoid statistical bias. This portion comprises of 2554 utterances for the AD group and 1439 utterances for the Non-AD group, with a total of 3993 utterances from 20 patients with 23 dialogue conversations.

The CCC transcripts are already segmented at the utterance (turn) level and the word level, and annotated for speaker identity (patient vs. interviewer); however, no DA information is available. We used only the utterance level layers; transcripts were available in ELAN format and we converted them to CSV format. We then manually annotate the transcripts at the utterance level with DA in-

formation.

3.3 Terminology

Throughout this paper, we use specific terms for particular question types and response types, and use these in our annotation procedure. Following Switchboard’s SWBD-DAMSL terminology (Jurafsky et al., 1997), we use **qy** for **Yes-No** questions, and **qy^d** for **Declarative Yes-No** questions. Declarative questions (^d) are utterances which function pragmatically as questions but which do not have “question form” in their syntax. We use **qw** for **Wh-questions** which includes words like *what, how, when, etc.* and **qw^d** for **Declarative Wh-questions**. Yes-No or Wh-questions are questions which do not have only pragmatic force but have a syntactic and prosodic marking of questions or interrogative in nature. We used **g** for **Tag questions**, which are simply confirming questions that have auxiliary inversion at the end of statement e.g. (*But they’re pretty, aren’t they?*). For **Or questions** which are simply choice question and aids in answering the question by giving choices to the patients are represented by **qr** e.g. (*- did he um, keep him or did he throw him back?*).

We used term **Clarification question** for questions that are asked in response to a partial understanding of a question/statement and are specific in nature. These clarification questions are represented by **qc**. **Signal non-understanding** is generated by a person in response to a question that they have not understood and are represented by **br**. **Back-channel Question (bh)** is a continuer which takes the form of question and have question intonation in it. Back-channels are more generic than clarification questions and often occur in many types (*e.g really? Yeah? do you? is that right? etc.*).

When the response to a Yes-No question is just a yes including variations (e.g. *yeah, yes, huh, yes, Yes I do etc.*), it will be represented by **ny** and when there is a yes plus some explanation, it will be represented by **ny^e**.

- (1) A: Do you have children?
B: Yeah, but they’re big children now.
Grown.

[CCC Mason_Davis_001 28-29]

na is an affirmative answer that gives an explanation without the yes or its variation. **nn** is used for

No-answers and **nn^e** is used for an explanation with No answer (see Appendix A for Examples).

3.4 Annotation Scheme

The original SWBD-DAMSL tagset for the Switchboard Corpus contains 43 DA tags (Jurafsky et al., 1997). Our initial manual includes DA tags from SWBD-DAMSL and our own specific new DA tags with a total of 35 tags. For different types of questions and their possible responses, 14 DA tags are taken from SWBD-DAMSL and 2 new tags are introduced. These new tags are for clarification questions (**qc**) and for answers to Wh-Questions (**sd-qw**), and were required to distinguish key response types.³

The ability to tag specific clarification questions is important for our study, as questions asked by the interviewer can be followed by a clarification which indicates partial understanding while requesting specific clarifying information (SWBD-DAMSL only provides the **br** tag for complete non-understanding). The distinction between answers to Wh-Questions and other, unrelated statements is also important (in order to capture whether the response is relevant: a relevant answer should be different from simple general statement), but SWBD-DAMSL provides only a single **sd** tag for statements. Different types of question and their tags are given with examples in Table 1; a list of response types is given in Table 2.

Another new addition is the tagging of *repetition* of questions, with or without reformulation. We marked repeat questions as simple repeats or reformulations, and tagged with the index of the dialogue act (utterance number) they were repeating or reformulating.

Similarly, clarification questions can signal non-understanding with two main distinct CR forms, and this distinction is tagged: pure repeats and reformulated repeated questions that are slightly changed syntactically but the context remains the same – see Table 3 with utterance 144.

3.5 Inter-Annotator Agreement

To check inter-annotator agreement, three annotators annotated one conversation of an AD patient and Non-AD interviewer of 192 utterances. All

³Some other DA tagging schemes provide categories for these and more; however, we chose to begin with SWBD-DAMSL given its prevalence in DA tagging work, and extend it only as necessary. In future work we plan to examine multi-dimensional schemes (e.g. Core and Allen, 1997; Bunt et al., 2010) to see if they provide benefits in this setting.

annotators had a good knowledge of linguistics and were familiar with both the SWBD-DAMSL tagset and the additions as specified above and in the manual. First, all three annotators annotated the dialogue independently by assigning DA tags to all utterances with the 17 tags of interest for this paper as shown in Table 4 (‘other’ means the annotator judged another SWBD-DAMSL act tag could be appropriate apart from the 16 tags in focus). We use a multi-rater version of Cohen’s κ (Cohen, 1960) as described by (Siegel and Castellan, 1988) to establish the agreement of annotators for all tags and also 1-vs-the-rest as shown in Table 4 below.⁴

As can be seen, an overall agreement was good ($\kappa=0.844$) for all tags and the majority of tags which were tagged by any annotator in the dialogue have $\kappa > 0.67$, with only ‘no’ getting beneath $\kappa < 0.5$. We judged this test to be indicative of a reliable annotation scheme for our purposes.

4 Results and Discussion

From the CCC transcripts, we selected 23 conversations, which when annotated yield 3993 utterances. All utterances were tagged with one of the 16 dialogue act tags relating to all question categories and their possible answers as described above, plus an ‘other’ tag. In addition to the dialogue act tag, utterances deemed to be responses (tags in Table 2) were tagged with the index of the utterance being responded to. Repeat questions were also marked as *simple repeats* or *reformulations*, and tagged with the index of the dialogue act they were repeating or reformulating.

Is the distribution of question types asked by the patient and interviewer different when the patient is an AD sufferer?

To investigate the distribution of dialogue acts, we calculated the relative frequency of each question and response type separately for AD and Non-AD group, and for the patient and interviewer within those groups. A comprehensive analysis of particular types and their distribution between AD and Non-AD patient with their interviewer is shown in Table 5. More yes-no questions (qy) are asked by the interviewer from AD Patients than Non-AD patients (6% vs 3.7%) and fewer wh-questions

⁴The annotation results and scripts are available from https://github.com/julianhough/inter_annotator_agreement.

Type	Tag	Example
Yes-No Question	qy	Did you go anywhere today?
Wh-Question	qw	When do you have any time to do your homework?
Declarative Yes-No Question	qy^d	You have two kids?
Declarative Wh-Question	qw^d	Doing what?
Or Question	qr	Did he um, keep him or did he throw him back?
Tag Question	^g	But they're pretty aren't they?
Clarification Question	qc	Next Tuesday?
Signal Non-understanding	br	Pardon?
Backchannel in question form	bh	Really?

Table 1: Question Types for CCC

Type	Tag	Example
Yes answer	ny	Yeah.
Yes- plus expansion	ny^e	Yeah, but they're
Affirmative non-yes answer	na	Oh I think so. [laughs]?
No answer	nn	No
Negative non-no answers	nn^e	No, I belonged to the Methodist church.
Other answer	no	I, I don't know.
Declarative statement wh-answer	sd-qw	Popcorn shrimp and it was leftover from yesterday.

Table 2: Answer Types for CCC

Tag	Speaker:Utterance	Text	Repeat Question?
qw	A:15	-Where's she been?	15
br	B:16	-Pardon?	
qw	A:17	-Where is she been?	
qy	A:142	-Well, are you, are you restricted from certain foods?	142-reformulation
br	B:143	-What?	
qy	A:144	-Like, do they, do they make you eat certain foods because your medication?	

Table 3: Examples of Repeated questions

(qw) are asked in the AD group compared to the non-AD group (4% vs 5.4%). Choice questions (qr) are also asked more from AD patients compared to non-AD patients (2% vs 0.3%). These results suggest there is a systematic difference in question distributions; one plausible explanation for this is that AD patients find it easier to answer a simple Yes-No question or a choice question compared to a wh-question. It is also obvious from the results that AD patients are also asking more questions than Non-AD patient during their conversation with the interviewer (*qy*: 1% vs 0.3%), (*qw*: 1% vs 0.3%), (*^g*: 0.2% vs 0.1%), (*br*: 3% vs 0.4%), and (*qc*: 2% vs 0.1%).

We also compared the distribution of these tags with the Switchboard SWDA corpus, as shown in Table 6. As the CCC is a set of clinical interviews, the percentage of tags which are questions is higher in this corpus compared to Switchboard. Although simple yes-no questions have almost identical frequencies in both corpora, declarative yes-no, wh-questions, declarative wh-questions, tag questions, and signals of non-understanding are higher in the CCC than Switchboard. Our new clarification question (*qc*) tag accounts for 1% for both AD group and Non-AD group tags but is not annotated in SWDA.

Tag	# times annotated	κ
qy	26	0.758
qw	30	0.895
qy ^d	12	0.660
qw ^d	3	1.000
^g	2	0.498
br	22	0.953
bh	0	0
qc	15	0.795
qr	0	0
ny	12	1.000
ny ^e	11	0.907
na	8	0.873
nn	1	0
nn ^e	6	0.663
no	4	0.497
sd-qw	26	0.637
other	398	0.902
all tags	576	0.844

Table 4: Multi-rater Cohen’s κ statistics for one-vs-rest and overall agreement score for one dialogue.

DA tag	AD		Non-AD	
	Pat	Int	Pat	Int
qy	1%	6%	0.3%	3.7%
qy ^d	1%	6%	0.1%	5%
qw	1%	4%	0.1%	5.4%
qw ^d	0.4%	1%	0.5%	0
^g	0.2%	2%	0.1%	0.7%
qr	0.1%	2%	0	0.3%
br	3%	0.1%	0.4%	0
bh	1%	1%	1%	1%
qc	2%	1%	0.1	1%
simple-Repeat	0	1%	0	0
reformulation	0	2%	0	0

Table 5: Distribution of DA question tags among the AD group and Non-AD group

How often do signals of non-understanding, clarification requests and back-channel questions occur in dialogues with an AD sufferer compared to those without one?

An examination of signals of non-understanding, clarification requests and back-channel requests reveals that the ability to follow and understand questions decrease for AD patients so they produce more signals of non-understanding (e.g. *sorry*, *Maam?*, *Pardon?*, *huh?*, *eh?*), when questions are posed to them. On the other hand, signals of non-

DA Tag	CCC-AD	CCC-Non-AD	SWDA
qy	3%	2%	2%
qy ^d	4%	2%	1%
qw	3%	3%	1%
qw ^d	1%	0.3%	<.1%
^g	1%	0.5%	<.1%
br	1%	0.2%	0.1%
bh	1%	1%	1%
qc	1%	1%	-
qr	1%	0.2%	0.1%
ny	3%	1%	1%
ny ^e	2%	2%	0.4%
na	3%	3%	1%
nn	0.4%	0.4%	1%
nn ^e	1%	1%	0.1%
no	0.4%	0.3%	1%
sd-qw	4%	6%	-

Table 6: Comparison of relative frequency of DA tags in the AD group, Non-AD group of the CCC and SWDA corpora

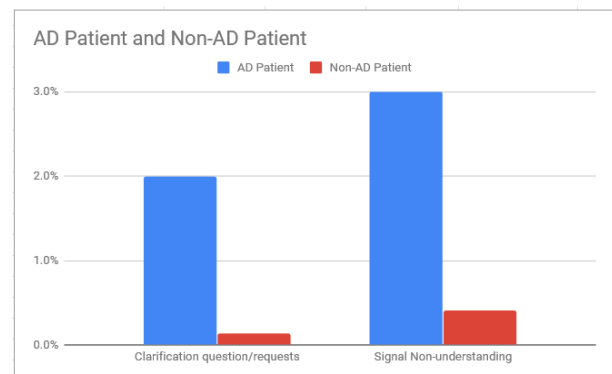


Figure 1: Clarification questions and Signal Non-understanding

understanding from Non-AD patients are much less frequent as shown in Figure 1. The overall frequency of clarification questions (qc) between the two conversation groups was not systematically different as shown in Table 6 when utterances from both patient and interviewer are combined, but dealing with them separately, AD patients produce more clarification requests than non-AD patients (2% vs 0.1%) – see Table 5 and Fig. 1.

We further examine how often signals of non-understanding and clarification requests are issued in response to questions rather than statements/answers. Examination of the data shows that clarification requests are more often gener-

	AD Group	Non-AD Group
Question followed by Signal of Non-understanding	24 (35)	2 (3)
Statements followed by Signal of Non-understanding	11 (35)	1 (3)
Question followed by Clarification Question	8 (34)	1 (11)
Statement followed by Clarification Question	26 (34)	10 (11)

Table 7: Occurrences of signal non-understanding and clarification question followed by question/statements

ated in response to statements, and less often after questions are raised; but signal non-understanding happen more often after questions. Out of total 35 signal non-understanding, 24 are generated in response to a question of AD Group as shown in Table 7. However, only 8 clarification questions are asked in response to questions, with 26 asked in response to declarative statements – (see Appendix A for more examples and context).

Is the distribution of simple-repeat and reformulation questions different for conversations with an AD sufferer compared to those without one?

Many questions are followed by clarification questions or signal non-understanding, so there will be more repetition of a similar type of question in case of the AD patients. Repeated questions are asked in two variations; either repeated simply or reformulated so that the patient can understand the question properly as in (4). In the AD group 4.7% questions are simple-repeat questions and 6.7% are reformulated as shown in Table 8 while for the non-AD group only 2.4% are reformulated questions and there were no repeated questions.

- (4) A: Your dad worked for who was it? Swisten
A: and that's why you went up to Baltimore?.
B: Huh?
A: Your dad went to –worked at – worked for Swisten?
B: My Father?
A: Yeah. Is that why you guys went to Baltimore?

[CCC Tappan_Patte_001 37-43]

5 Conclusion and Future work

Our study provides the first statistical analysis of different types of question asked in conversations

Repeat Type	AD Group	Non-AD Group
Total Question	313	127
Simple-Repeat Question	15 (4.7%)	0
Reformulated Question	21 (6.7%)	3 (2.4%)

Table 8: Repetition and reformulation of questions for AD group and Non-AD group

with AD patients in the Carolina Conversation Collection (CCC) Corpus. We found that yes-no questions were asked more frequently in the AD sufferer conversations than the Non-AD conversations (6% vs 3.7% of all dialogue acts) and less Wh-questions were asked in AD sufferer conversations compared to Non-AD ones (4% vs 5.4%). While our newly introduced tags were not frequent, they are significant in AD sufferer conversations, with 2% of all dialogue acts by AD sufferers being clarification questions and 3% being signals of non-understanding.

In future work, we plan to work on the CCC corpus conversations of both AD and Non-AD conversations to build an automatic dialogue act tagger for the tagset we used in this study. We will also explore more complex questions including compound questions and questions that relate to semantic memory and episodic memory. We also plan to look into disfluency and repairs in this data collection which could further aid interpretation and automatic diagnosis.

References

- Jason Brandt. 1991. The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The Clinical Neuropsychologist*, 5(2):125–142.
- Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010. Towards an ISO standard for dialogue act annotation. In *Proceedings of LREC 2010, the Seventh International Conference on Language Resources and Evaluation*.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Mark G Core and James Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*, volume 56. Boston, MA.
- Christopher Elsey, Paul Drew, Danielle Jones, Daniel Blackburn, Sarah Wakefield, Kirsty Harkness, Annalena Venneri, and Markus Reuber. 2015. Towards diagnostic conversational profiles of patients presenting with dementia or functional memory disorders to memory clinics. *Patient Education and Counseling*, 98(9):1071–1077.
- M F Folstein, S E Folstein, and P R McHugh. 1975. Mini-mental status. a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198.
- Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. 2016a. Linguistic features identify Alzheimer’s disease in narrative speech. *Journal of Alzheimer’s Disease*, 49(2):407–422.
- Kathleen C. Fraser, Frank Rudzicz, and Graeme Hirst. 2016b. [Detecting late-life depression in Alzheimer’s disease through analysis of speech and language](#). In *Proc. CLPsych*, pages 1–11, San Diego, CA, USA. Association for Computational Linguistics.
- Dalia Gottlieb-Tanaka, Jeff Small, and Annalee Yassi. 2003. A programme of creative expression activities for seniors with dementia. *Dementia*, 2(1):127–133.
- Heidi Ehernberger Hamilton. 2005. *Conversations with an Alzheimer’s patient: An interactional sociolinguistic study*. Cambridge University Press.
- William Jarrold, Bart Peintner, David Wilkins, Dimitra Vergryi, Colleen Richey, Maria Luisa Gorno-Tempini, and Jennifer Ogar. 2014. [Aided diagnosis of dementia type through computer-based analysis of spontaneous speech](#). In *Proc. CLPsych*, pages 27–37, Baltimore, Maryland, USA. Association for Computational Linguistics.
- Danielle Jones, Paul Drew, Christopher Elsey, Daniel Blackburn, Sarah Wakefield, Kirsty Harkness, and Markus Reuber. 2016. Conversational assessment in memory clinic encounters: interactional profiling for differentiating dementia from functional memory disorders. *Aging & Mental Health*, 20(5):500–509.
- Daniel Jurafsky, Elizabeth Shriberg, and Debra Bisasca. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual.
- Elke Kalbe, Josef Kessler, Pasquale Calabrese, R Smith, AP Passmore, Met al Brand, and R Bullock. 2004. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International journal of geriatric psychiatry*, 19(2):136–143.
- Saturnino Luz, Sofia de la Fuente, and Pierre Albert. 2018. A method for analysis of patient speech in dialogue for dementia detection. In *Proceedings of the LREC 2018 Workshop Resources and Processing of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric impairments (RaPID-2)*.
- Guy McKhann, David Drachman, and Marshall Folstein. 1984. [Clinical diagnosis of Alzheimer’s disease](#). *Neurology*, 34(7):939—944. Views & Reviews.
- Peter Noone. 2015. Addenbrooke’s Cognitive Examination-III. *Occupational Medicine*, 65:418–420.
- Charlene Pope and Boyd H Davis. 2011. Finding a balance: The Carolinas Conversation Collection. *Corpus Linguistics and Linguistic Theory*, 7(1):143–161.
- W G Rosen, R C Mohs, and K L Davis. 1984. A new rating scale for Alzheimer’s disease. *American Journal of Psychiatry*, 141(11):1356–1364.
- Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. 1978. A simplest systematics for the organization of turn taking for conversation. In *Studies in the Organization of Conversational Interaction*, pages 7–55. Elsevier.
- Sidney Siegel and NJ Castellan. 1988. Measures of association and their tests of significance. *Nonparametric Statistics for the Behavioral Sciences*, pages 224–312.
- Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–373.
- Kelvin K F Tsoi, Lingling Zhang, Nicholas B Chan, Felix C H Chan, Hoyee W Hirai, and Helen M L Meng. 2018. Social Media as a Tool to Look for People with Dementia Who Become Lost : Factors

That Matter. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 9:3355–3364.

Ana Varela Suárez. 2018. The question-answer adjacency pair in dementia discourse. *International Journal of Applied Linguistics*, 28(1):86–101.

A Examples from Carolinas Conversation Collection

Yes-No Question followed by no plus expansion answer:

Tag	Text
qy nn ^e	A: were you Primitive Baptist? B: — no, I belonged to the Methodist church.

[CCC Mason_Davis_001 92-93]

Yes-No Question followed by other answer:

Tag	Text
qy no	A: are you going to go with them to see the Christmas Lights? B: Oh, I, I dont know.

[CCC Wakefield_Brock_001 51-52]

Two Wh-Questions followed by declarative statements wh-answer:

Tag	Text
qw sd-qw	A: - what does he preach about? B: – hell hot and heaven beautiful.
qw sd-qw +	C: what types of food do you like the best? D – vegetables, meat, - and desserts.

[CCC Mason_Davis_001 31-32]

[CCC Wakeman_Rhyne_001 6-7]

Wh-question followed by a clarification question(*qc*) and a wh-question followed by a statement and then a clarification(*qc*):

Tag	Text
qw qy br qw qc	A: where is Jerusalem Primitive Baptist Church? - is that near Fountain Hill? B: - m'am? A: where is that church? B: Fountain Hill?
qw sd-qw qc	A: what do you do? B - I'm a teacher. A: Preacher?-

[CCC Mason_Davis_001 83-86,64-66]

Declarative wh-question followed by signal non-understanding(*br*) and then by reformulated-repeat wh-question:

Tag	Text
qw ^d br qw	A: You were married for– B: Huh? A: How long– have you been married? (reformulated-repeat)

[CCC Tappan_Patte_001 7-9]

Declarative statement followed by back-channel question(*bh*) and then by yes answer:

Tag	Text
sd bh ny	A: huh, it used to be something special. it used to be my Mother's birthday. B: Really ? A: Yeah

[Wheaden_Lee_001 52-54]