Exploring age-related conversational interaction

Jeroen Geertzen
Dept. of Theoretical & Applied Linguistics
University of Cambridge
jg532@cam.ac.uk

Abstract

Does the way we interact in conversation changes as we get older? This paper presents a corpus-based study investigating this question by looking into age-related differences in spontaneous spoken dialogue. Conversations from the Switchboard corpus were analysed using n-gram models, relating the communicative actions as encoded by dialogue acts to speaker age. Results show that older interlocutors generally address the conversation itself (rather than other topics) less than younger adults. Age differences are also reported on feedback strategies, most notably that even though younger interlocutors produce more backchannels overall, older interlocutors use backchannels more often when taking turns.

1 Introduction

Numerous studies on spoken language production have documented language production changes across the life span (see (Mortensen et al., 2006) for an overview). Much of this research is highly controlled in nature and has focused on word level production (see (Burke and Shafto, 2004) for a review). Yet, relatively little research has addressed the question whether normal aging affects the way interlocutors behave and interact with each other in conversation. This, in itself, is not surprising as conversation is difficult to investigate in a controlled manner unless it involves very specific tasks that often make the experimental setting rather artificial.

There is some evidence, however, that suggests that age may have an effect. For instance, (Hupet et al., 1993) showed, by repeatedly asking pairs of interlocutors to discuss how to arrange complex figures in a particular order, that older adults are less likely to take previously shared information into account than younger adults. But many aspects that could be affected by age differences have not been looked at, leaving many questions unanswered.

Do older interlocutors spend more or less effort to address the communication process itself by providing communicative feedback as compared to addressing the topics or tasks under discussion? If there is an aging effect present, what forms of communicative feedback feature such effects? For instance, the use of 'backchannels', short verbal expressions such as 'uh-huh' or 'yeah' are considered to provide implicit positive feedback, whereas expressions such as 'are you listening?' and 'what?' signal communicative problems (Allwood et al., 1992; Bunt, 1994). Age related differences in how and how often feedback is produced could reveal differences in the communicative performance or strategies involved.

The aim of this paper is to investigate whether, and if so how, age affects the way interlocutors interact with each other in spontaneous spoken dialogue by analysing interaction patterns using recorded and annotated telephone conversations from the Switchboard I corpus (Godfrey et al., 1992).

Section 2 provides further background by elaborating on key aspects of spoken dialogue, while Section 3 reports on the Switchboard data, its speaker characteristics, and the annotations used. Section 4 provides the results obtained, which are discussed in Section 5. Conclusions and future work are described in Section 6.

2 Background

The communicative functions of utterances that speakers contribute to a dialogue are usually characterized by dialogue acts, which can be taken to characterise the interaction and constitute sequences such as Question-Answer or Greeting-

Greeting, or more complex patterns. For instance, interlocutor B could respond to a yes-no question posed by interlocutor A by asking A a clarification question and receiving A's reply before finally answering the yes-no question.

Besides communicating about some underlying task or activity that drives the dialogue, interlocutors have to attend to various aspects of the communicative process itself as to keep the communication going in a sufficiently smooth way. This process, often referred to as *dialogue management*, may involve signalling changing time constraints (time management), assigning sender and receiver roles (turn management), or monitoring mutual attention and understanding. Such engagement gives rise to communicative activities such as taking the turn, showing attention, signalling a misunderstanding, and establishing joint attention on the conversational topic under focus. Clark (1996) describes this distinction by means of two communicative tracks: communicative acts in the first track are concerned with the presentation of the information, whereas meta-communicative acts in the second track are concerned with the communication itself. One important aspect is, therefore, how much of the contributions that interlocutors make involve dialogue management, and little dialogue management is arguably an indicator for well going conversation.

At the same time, a conversation without any communication management is not necessarily perceived as more fluent and effective than a conversation with little communication management, which motivates the distinction between communication management that helps to keep a non-problematic interaction continue sufficiently smoothly and dialogue management that is typically used when problems in the communication arise. *Backchannels* are typical of the former category and are short utterances, such as 'uhu' and 'hm-mm', produced by listeners to signal understanding and indicating that they are paying attention, encouraging speakers to continue (Duncan, 1972). They are considered to make the conversation go smoothly. The latter category is mostly represented by *explicit negative feedback*, which concerns utterances conveying meta-communicative acts that indicate a problem in the understanding that may occur for various reasons such as problems in perception (e.g. "I cannot hear you") or interpretation (e.g. "What did you mean?").

For each of the key communicative aspects discussed, there is no clear theory-driven hypothesis that has been proposed. It is generally assumed that healthy aging in (late) adulthood tends to be accompanied by a subtle decline in cognitive and perceptual functioning. Where this affects spoken dialogue, a recent study has established that older adults show more difficulty in following conversation, largely accounted for by perceptual functioning (Murphy et al., 2006). This may cause older adults to engage more often in dialogue management and produce relatively more signals of non-understanding. At the same time, research in human-computer interaction has found that older adults use a wider range of speech acts allowing more flexible interaction patterns (Georgila et al., 2008).

The question whether there are age-related differences in interaction, and the lack of strong hypotheses makes this an exploratory study in which not only the small set of dialogue management functions will be included, but also other less important aspects are considered.

3 Material and method

3.1 Speech data

The speech data for this study come from the Switchboard I Corpus (Godfrey et al., 1992), a collection of about 2,400 dialogues among 543 speakers from all areas of the United States on a wide range of preferred conversational topics. Speakers sharing an interest in the same topic were paired such that no two speakers would converse together more than once, and no one spoke more than once on a given topic. Basic demographic information was collected for each speaker, including age.

A part of the Switchboard corpus was manually annotated with a set of dialogue acts (Jurafsky et al., 1997) and released as the Switchboard Dialog Act Corpus. It covers 1,155 five-minute dialogues comprising over 200K utterances and 1.4 million words. The dialogue act tagset that was used, SWBD-DAMSL, is based on the Discourse Annotation and Markup System of Labeling (DAMSL; (Allen and Core, 1997); (Core and Allen, 1997)) and contains 220 tags that are clustered into 42 larger classes. A dialogue excerpt with labelled utterances is illustrated in Table 1.

Table 1: Sample from dialogue sw_0811_2278

	Utterance	*Dialogue act*
B	Okay,	OTHER (o)
B	um, so, um, do you have any favorite teams?	YES-NO-QUESTION (qy)
A	Well, I kind of like them all.	AFFIRMATIVE NON-YES ANSWERS (na)
A	I played for about eighteen years, all the way through college, and then, uh, kind of hung them up after college,	STATEMENT-NON-OPINION (sd)
A	but, $< laughter >$	ABANDONED OR TURN-EXIT (%)
B	Oh, I was going to say, you played pro ball,	STATEMENT-NON-OPINION (sd)
B	right?	TAG-QUESTION (^g)

3.2 Speakers

The sample for this study consists of speech from 438 different speakers, taken from the 1,155 unique conversations in the corpus. The speakers are between 20 and 68 years old (M=37.6; SD=10.9) and are balanced in gender. Speakers could take part in multiple dialogues: 31% participated only in a single dialogue; 28% participated in two to five dialogues, and 41% participated in more than five dialogues.

3.3 Analysis

The first part of the analysis is based on measuring production rates of individual dialogue acts that target key aspects of the interaction.

In the SWBD-DAMSL and DAMSL annotation scheme, the Information Level layer indicates whether a contribution to the dialogue is about the task, about the management of the task (^t), or about the communication (^c);

As for back-channels, we focus on two specific tags in DAMSL targetting this phenomenon: general backchannels (b) and backchannels in question form (bh), such as really? or yeah?. Explicit negative feedback is indexed by "Signal-non-understanding" which in DAMSL are explicitly marked by (br) and (br^m) tags.

For each of the 438 speakers, production rates were computed for each dialogue act tag by dividing the frequency of the dialogue act tag by the sum of frequencies of all dialogue acts. As the five-minute fragments are rather short, production rates were not first averaged for each dialogue but computed based on all dialogue contributions by a particular speaker. Presence of age effects were then tested by correlating production rates of dialogue acts with interlocutor's age (in years). Pearson correlation coefficients (r) and corresponding p-values (p) are reported.

Correlation analysis has the advantage that the whole age range is modelled as opposed to binning interlocutors in an 'old' and 'young' group according to a more or less arbitrary ranges.

With the same approach as for individual dialogue acts, also interaction patterns of subsequent dialogue acts were analysed to get a better picture of the conversational interaction an interlocutor is involved in. To include turn boundaries as well, turn changes were marked explicitly as turn-beginnings (denoted by $|B|$) and turn-endings (denoted by $|E|$) with respect to a specific interlocutor, e.g. qy+$|B|$+na+sd+%+$|E|$+sd+^g for speaker B. To describe subsequences of dialogue acts, n-gram language models of various orders were estimated: bigrams, trigrams, and 4-grams.[1]

With the individual dialogue acts, we have no clear hypotheses to test. More problematically, the risk of Type I errors (false positives) in significance testing increases with the number of hypotheses being addressed, which are numerous when looking for prominent dialogue act n-grams, making the significance level increasingly meaningless. Multiple comparisons can be taken into account by various adjustments, such as dividing the usual alpha of 0.05 by the number of hypotheses involved (Bonferroni, 1935), but sound hypothesis identification as well as (replication) testing can be achieved by data splitting (Dahl et al., 2008). This technique involves randomly splitting the data into two parts: one for hypothesis formulation and one for hypothesis testing. Only hypotheses that are identified by a p-value below

[1]For instance, a bigram language model would involve pairs {o+qy, qy+na, a+sd, sd+%, %+sd, sd+^g} and is expected to pick up on adjacency pairs.

alpha-level in the first part which then are also tested as significant in the second part are considered to be truly significant as well as replicable, and be reported as estimated over all data.

4 Results

4.1 Individual dialogue acts

It may be argued that the more an interlocutor is engaged in communication management, the less fluent the dialogue tends to become, and age related differences may emerge at Information-level. The correlations of age with the relative frequencies of the labels in Information-level are listed in Table 2, and show that older speakers are generally less involved in communication management.

Table 2: Significant correlations with Information-level: communication (C) and task (T), with task management not replicable

	$r(436)$	p	p_1	p_2
C	-0.199	0.00003	0.0022	0.0045
T	0.186	0.00008	0.0046	0.0067

Table 3: Correlation of age with backchannels (B), which in question form are not replicable

	$r(436)$	p	p_1	p_2
B	-0.187	0.00008	0.0031	0.0073

Pearsons correlations (Table 3) show that relative to all produced dialogue acts, the use of backchannels decreases with age. This is one of the stronger correlation that was found, and is relevant considering that backchannels account for around 19% of all dialogue acts produced (and backchannels in question form for around 1%).

Generally, older interlocutors produced fewer questions and more statements. They tend to signal less non-understanding, even though this turned out not to be replicable (Table 4). Also, older speakers tend to use fewer hedges: expressions that intend to diminish the confidence or certainty of a statement or answers that the speaker made, such as 'I guess' or 'If I am not mistaken'.

4.2 Dialogue act n-grams

Dialogue act sequences involving acts from both interlocutors were described by means of n-grams of which the production rate of a n-gram was calculated as its frequency divided by the cumulative frequency of n-grams of the same length. Production rates were correlated with age, and a selection of the strongest correlating n-grams are listed in Table 5.

Table 4: Correlation with Questions (Q), Hedges (H), Statements-non-opinion (S), and signal-non-understanding (SNU), which cannot be reproduced

	$r(436)$	p	p_1	p_2
Q	-0.173	0.00026	0.0102	0.0078
H	-0.169	0.00038	0.0095	0.0139
S	0.136	0.00436	0.0451	0.0400
SNU	-0.116	0.01539	0.0438	0.2343

Older speakers produce fewer acknowledgements in the form of backchannels following a statement of the dialogue partner (5.1), while receiving more elaborate backchannels or continuers that express appreciation, such as 'That sounds great.' or 'I can imagine' (5.2). Also generally, older speakers produce fewer backchannels and younger speakers produce more (5.7). Furthermore, age shows a negative correlation with hedges followed by an opinion statement (5.6).

Older speakers produce more tag questions, such as 'So you like music, don't you?', which are then followed by a descriptive or narrative statement which acts as a negative answer, such as 'I do not', produced by the other interlocutor (5.3).

5 Discussion

As results suggest, younger adults produce more backchannels in various circumstances, indicating explicitly the active monitoring of the partner's production. This finding is in line with existing work. For instance, (Kemper et al., 1998) report a significant age effect in a referential communication task (a map task as in (Anderson et al., 1991)) in which young adults instructed older adults to reproduce a map. Also in other tasks, such as describing to each other a mutually experienced event such as holidays, younger adults produced more backchanneling than older adults (Gould and Dixon, 1993). This has been explained as an increased "willingness and ability to take on the cognitively demanding task of dividing one's attention between monitoring the social situation and planning one's own speech productions." (Gould and Dixon, 1993). At the same time, turn length

Table 5: Correlation of age with dialogue act sequences

Label	$r(436)$	p	p_1	p_2
1. Statement-opinion (sd) + $\|B\|$ + Backchannel (b)	-0.184	0.00011	0.0081	0.0054
2. Statement-opinion (sd) + $\|E\|$ + Appreciation (ba)	0.182	0.00012	0.0075	0.0056
3. $\|B\|$ + Tag question (statement and YN question) (qy^g)	0.181	0.00014	0.0108	0.0072
.....+ $\|E\|$ + Negative non-no answer (ng)				
4. $\|B\|$ + Backchannel (b) + $\|E\|$ + YN question (qy)	-0.180	0.00015	0.0097	0.0152
5. $\|B\|$ + Action-directive (ad) + $\|E\|$	0.176	0.00022	0.0280	0.0162
6. Hedge (h) + Statement-opinion (sd)	-0.175	0.00023	0.0369	0.0301
7. $\|E\|$ + Backchannel (b) + $\|B\|$	0.175	0.00024	0.0328	0.0294

in number of words and dialogue acts produced by older speakers are also higher on average, as reported by e.g. (James et al., 1998) and confirmed with the Switchboard corpus data. It is then reasonable to expect that long turns elicit more backchannels by the dialogue partner in order to signal continued attention.

Correlation analysis has the advantage that the whole age range is modelled as opposed to binning interlocutors in an 'old' and 'young' group according to a more or less arbitrary ranges. Combined with randomly split data for identifying and testing hypotheses as well as combatting type I errors, prominent hypotheses were highlighted. However, it does not reveal whether interlocutors act differently when they are speaking to addressees of different ages, which is an interesting aspect to further investigate. To this purpose, speakers could be binned into 'young' and 'old' ($>$ 36 yrs) to allow for four groups: young-young, young-old and old-young, and old-old.

6 Conclusions & Future work

Dialogue act production of interlocutors across conversations showed that older interlocutors generally use fewer dialogue acts related to communication management (with more dialogue acts related to task and task management). Even though younger interlocutors produce more backchannels, older interlocutors use backchannels more at the start of a turn. They also tend to start their turn more often by repeating parts of what the speaker said before continuing with an opinion statement.

Current work, about to be completed, is addressing whether interlocutors act differently when they are speaking to addressees of different ages.

Future work in line of this study will investigate whether turn length affects general production rates. Furthermore, temporal aspects will be explored by linking the timing of the speech in the Switchboard I corpus to the dialogue act tags in the Switchboard Dialog Act Corpus. Recent research with the same dataset suggests that various social variables, including age, correlate significantly with turn-taking behavior (Grothendieck et al., 2009) and motivates further analysis.

Additionally, the use of language models as a way to capture relevant and possibly interesting subdialogues will be refined by using grammar induction (by unsupervised machine learning) to capture more complex re-occurring interaction patterns such as clarification or repairs subdialogues (Alexandersson and Reithinger, 1997; Geertzen, 2009).

References

Jan Alexandersson and Norbert Reithinger. 1997. Learning dialogue structures from a corpus. In *Proceedings of Eurospeech 1997*, page 22312234, Rhodes, Greece, September.

James Allen and Mark Core. 1997. Draft of DAMSL: dialog act markup in several layers. Unpublished manuscript.

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, 9(1):1–26.

Anne Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, and Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech*, 34:351366.

Carlo E Bonferroni. 1935. Il calcolo delle assicurazioni su gruppi di teste. In *Studi in Onore del Professore Salvatore Ortu Carboni*, pages 13–60, Rome, Italy. Tipografia del Senato.

Harry Bunt. 1994. Context and dialogue control. *Think Quarterly*, 3(1):19–31.

Deborah M. Burke and Meredith A. Shafto. 2004. Aging and language production. *Current directions in psychological science : a journal of the American Psychological Society*, 13(1):21–24. PMID: 18414600 PMCID: 2293308.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge, UK, May.

Mark G. Core and James F. Allen. 1997. Coding dialogues with the DAMSL annotation scheme. In David R. Traum, editor, *Working Notes: AAAI Fall Symposium on Communicative Action in Humans and Machines*, page 2835, Menlo Park, CA, USA. American Association for Artificial Intelligence.

Fredrik A Dahl, Margreth Grotle, Jūratė Šaltytė Benth, and Bård Natvig. 2008. Data splitting as a countermeasure against hypothesis fishing: with a case study of predictors for low back pain. *European journal of epidemiology*, 23(4):237–242.

Starkey Duncan. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2):283–292.

Jeroen Geertzen. 2009. Dialogue act prediction using stochastic context-free grammar induction. In *Proceedings of the EACL 2009 Workshop on Computational Linguistic Aspects of Grammatical Inference*, CLAGI '09, page 715, Stroudsburg, PA, USA. Association for Computational Linguistics.

Kallirroi Georgila, Maria Klara Wolters, Vasilis Karaiskos, Melissa Kronenthal, Robert H. Logie, Neil Mayo, Johanna D. Moore, and Matthew Watson. 2008. A fully annotated corpus for studying the effect of cognitive ageing on users' interactions with spoken dialogue systems. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco*.

John Godfrey, Edward Holliman, and Jane McDaniel. 1992. SWITCHBOARD: telephone speech corpus for research and development. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-92)*, page 517520, San Francisco, USA.

Odette N. Gould and Roger A. Dixon. 1993. How we spent our vacation: Collaborative storytelling by young and old adults. *Psychology and Aging*, 8(1):10–17, March.

J. Grothendieck, A. Gorin, and M. Borges. 2009. Social correlates of turn-taking behavior. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, pages 4745–4748.

Michel Hupet, Yves Chantraine, and Franois Nef. 1993. References in conversation between young and old normal adults. *Psychology and Aging*, 8(3):339–346, September.

L E James, D M Burke, A Austin, and E Hulme. 1998. Production and perception of "verbosity" in younger and older adults. *Psychology and Aging*, 13(3):355–367, September. PMID: 9793112.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13. Technical Report 97-02, Institute of Cognitive Science, University of Colorado, USA.

Susan Kemper, Andrea FinterUrczyk, Patrice Ferrell, Tamara Harden, and Catherine Billington. 1998. Using elderspeak with older adults. *Discourse Processes*, 25(1):55–73.

Linda Mortensen, Antje S. Meyer, and Glyn W. Humphreys. 2006. Age-related effects on speech production: A review. *Language and Cognitive Processes*, 21(1-3):238–290, January.

Dana R Murphy, Meredyth Daneman, and Bruce A Schneider. 2006. Why do older adults have difficulty following conversations? *Psychology and aging*, 21(1):49.