Dialogue Acts in the AusTalk Map Tasks

Dominique Estival, Valeria Peretokina

MARCS Institute, Western Sydney University

[d.estival/v.peretokina]@westernsydney.edu.au

Abstract

This paper reports on the analysis of 12 Map Task dialogues from the AusTalk corpus, which were annotated for Dialogue Acts. We describe the annotation process, including the modifications to the SWBD-DAMSL coding scheme, and present the results of an analysis of the number, complexity, and function of turns across conversation role (Information Giver or Follower), gender, and gender dyad. Significant differences are found depending on the gender and role of the speakers and on the gender dyad. Overall, female speakers produce more turns than male speakers, and female-female dyads produce more multiple-label and mixed function turns.

Index Terms: speech corpus, Australian English, dialogue acts

1. Introduction

One of the components of the AusTalk corpus [1, 2] is the Map Task, a data gathering game aimed at collecting naturalistic conversational speech [3]. Each Map Task participant is given a map of the same environment. The map given to the information giver (IG) also shows a route that winds between landmarks and the IG must describe this route to the information follower (IF) who draws it on the other map. Landmark discrepancies between the maps are included to encourage IF-IG negotiation. Thus, the Map Task allows us to study discourse phenomena in spontaneous speech.

The AusTalk Map Task was designed for Australian English by Jette Viethen, Robert Dale and Felicity Cox, with landmark names chosen to elicit specific phonetic combinations of interest to researchers studying that language variety. In the AusTalk corpus, each of the 853 participants participated in two Map Tasks, once as the IG and once as the IF. This paper reports on the preliminary results of a project aiming to identify the Dialogue Acts (DAs) present in AusTalk Map Tasks and their linguistic features, building on [4, 5], and to investigate whether any gender, age, and dialect differences are observable in the participants' choice of DAs. Specifically, here we discuss the combined effect of conversation role and gender on the number, complexity, and function of the participants' turns.

2. Data and data processing

2.1. Data

Speech from 60 AusTalk Map Task recordings (up to 20 minutes each) had already been transcribed in the AusTalk Annotation project [6]. The transcriptions, as well as the audio data, are available from the Alveo Virtual Lab [7]. Twelve (12) of these 60 transcribed Map Tasks were selected for the current analysis to counterbalance speaker gender and recording location, and were then fully annotated for DAs. Table 1 presents the details of the 12 Map Tasks selected. More detailed information about the speakers is retrievable from the AusTalk demographic information collected at the time of recording in the speaker questionnaire (www. alveo.edu.au).

Table 1. *Transcribed AusTalk Map Tasks selected for analysis.*

Dyad	Site	Information Giver	Information Follower
female-female	USYD	4_1145	3_306
	UQ	2_534	3_377
	ANU	4_232	4_644
female-male	USYD	3_794	3_1035
	UQ	4_1277	1_738
	ANU	1_1119	3_1274
male-female	USYD	3_627	3_705
	UQ	4_68	4_1151
	USYD	3_42	4_923
male-male	USYD	3_112	1_322
	UQ	3_81	3_644
	ANU	1_178	2_113

2.2. Data processing

The AusTalk speech data had been originally transcribed with Transcriber [8] and the transcription files were in .trs format. There were at least two audio files per Map Task, one for the main speaker, who was the IG (Channel 6), and one for the second speaker, who was the IF (Channel 1). The audio files for the two channels were converted into one audio file readable by Transcriber, which was then opened with corresponding transcription file(s) and saved in Transcriber. Some transcription errors were corrected at that stage. The resulting .trs files were converted to Praat textgrids [9]. Two tiers were added to the textgrid files: one Dialogue Act label tier for the IG and one for the IF, as shown in Figure 1.

3. Coding the AusTalk Map Task dialogues

In the context of dialogues such as the Map Tasks, the coding is considered in terms of DAs rather than speech acts (SAs). A DA refers to a move within a dialogue, so it can link back to a previous utterance and it assumes there is a conversation. By contrast, a SA refers to the intention of the speaker, and does not assume there is a conversation or dialogue or even the existence of a listener (e.g., praying or swearing). DAs include questions, answers, directives, statements, and replies. The taxonomy of DAs is not based on the form of the utterances, but on their interpretation in context. For instance, an utterance with the syntactic form of a question is not necessarily a query (e.g., it can be a rhetorical question) and a query can be realised as a declarative as well as by a question (e.g., *"you want me to go right?"* or *"do you want me to go right?"*).

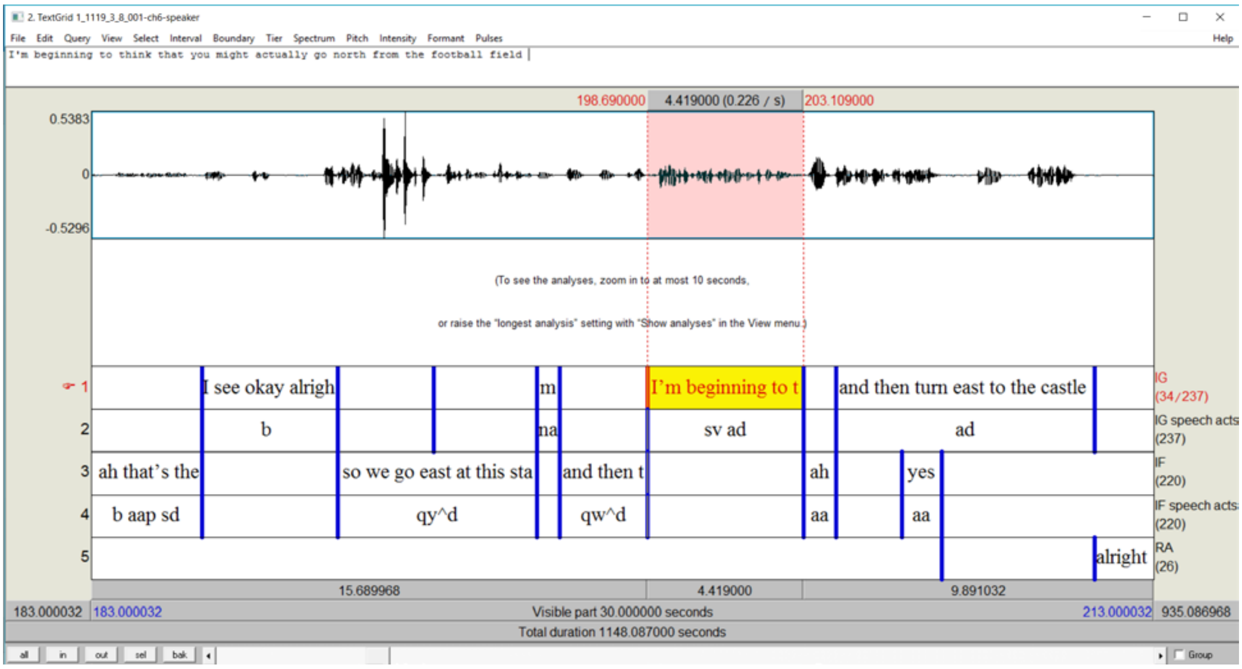

Figure 1. *Textgrid of AusTalk Map Task with orthographic transcription and Dialogue Acts labels for Information Giver (Tiers 1 and 2) and for Information Follower (Tiers 2 and 3). Tier 5 shows the transcription of the AusTalk Recording Assistant.*

3.1. Dialogue coding scheme

The coding scheme used in this project was adapted from the SWBD-DAMSL coding scheme [10], using insights from [5] and [4]. A useful aspect of DAMSL (Dialog Act Markup in Several Layers) is that it allows coding of dialogue moves at different levels at the same time, thus permitting the distinction between Task and Dialogue Management, as well as between backward- and forward-looking moves as follows:

- Task Management (e.g., request for clarification)
- Communication Management (e.g., clarification or apology)
- Backward-Communicative Function (e.g., clarification, confirmation or explanation)
- Forward-Communicative Function (e.g., follow-up or additional question).

In addition to the DA labels listed in [5], we adopted the following conventions for coding AusTalk Map Tasks DAs.

1) Questions

We specify whether a declarative question (*^d*) is a WH- or a Yes-No question, parallel to full interrogatives:

qy	yes-no-question	*Do you want me to turn left?*
qw	wh-question	*How far do I go?*
qy^d	declarative yes-no question	*We're going straight?*
qw^d	declarative wh-question	*You can find it where?*

2) Short responses (e.g., "*Yes.*", "*Right.*")

If the preceding DA was a question, the response is coded as *ny* for yes-answer, or *na* for affirmative non-yes answer. If the speaker is acknowledging a statement, the response is coded *aa* (accept/agreement).

3) Signal non-understanding (*br*) must be explicit (e.g., *"Excuse me?"),* rather than implied (e.g., *"mmm").*

4) *^e* (expansion) in *ny^e* (answer to yes-no question)

A simple expansion of a yes-answer (e.g., *"Yes, that's right*") remains *^e*. However, if a particular function can be identified, it is specified instead (e.g., *"Yep, and then you turn right"* is coded as *ny ad*: yes-answer + action directive).

5) *fo* (other forward function)

Anything used to propel the conversation forward which cannot be identified as any other label is coded as *fo*. An example of *fo* is *"So"* produced with long fricative and vowel.

6) *aa* (accept) versus *ny* (yes-answer)

IFs often produce *"yep", "yes", "Okay", "alright"* in response to action directives from IGs. If we identified the DA based on the lexical item, these responses would always be coded as *ny* or *na*. However, we view the DA function in combination with the previous utterance, that is, whether the interlocutor had asked a yes-no question or had been giving directions. Thus, in response to a question, *"yes"* is coded as *ny*. If it indicates that the IF is keeping up with the directions given to them (*"After the church you turn right - yep - and then go southerly"),* it is coded as *b* (acknowledge). If it indicates that the IF is agreeing with those directions (*"After the church you turn right - yep, got it"*), it is coded as *aa* (accept).

3.2. Inter-rater agreement

Three Map Tasks were coded independently by two research assistants. There was an average of 400 DAs per Map Task, 200 each for the IG and the IF, with an average of 36 DAs (16%) with different labels per Map Task. Manual examination showed most differences were systematic, with only 15 coding errors per Map Task (i.e., <4%). The systematic differences concerned the coding of:

- *aa* (accept) versus *b* (acknowledge) for utterances such as *"yes"* (see above);
- *nn^e* (no answer w/expansion) versus *nn* (no answer) for *"No, I don't"*;
- *oo* (open option) versus *ad* (action directives) for *"and we see a yellow church on the west side of the road"*.

4. Analysing Dialogue Acts

The 12 Map Tasks were analysed for differences in number of: turns; number of DAs per turn; relative frequency of backward and forward functions; relative frequency of complex and simple dialogue turns, according to the speakers' role in the conversation (IG vs. IF), gender (female vs. male), and the gender dyad completing the task (female-female, female-male, male-female, male-male). As Table 1 demonstrates, the chosen Map Tasks participant pairs consisted of three female-female dyads, three male-male dyads, and six mixed gender dyads. Of the mixed gender dyads, three had a female speaker as IG and a male speaker as IF, with the roles reversed (i.e., male IG and female IF) for the other three. For each participant, the total number of turns produced during the whole Map Task was calculated. Then, each turn was manually coded using the DA scheme described in Section 3.

Specific DAs and their variants were extracted from the textgrid by Praat scripts. Turns were grouped into two categories depending on the number of DAs identified in the turn. Turns coded with a single DA label comprised the single-label category, while turns containing multiple different DA labels were identified as multiple-label. In addition, the turns were categorised according to the function of their DAs: forward, backward, or mixed. Examples of different turn categories are presented in Table 2.

Table 2. *Turn categories.*

Category	Example	Labels assigned
Single-label	I want you to go past the orange castle should be on your right towards a yellow church with a tower but no steeple (3_42_4_10_IG)	*ad* (action directive)
Multiple-label	have we already been past this church yes okay (3_112_3_8_IF)	*qy* (yes-no question) *b* (acknowledge) *aa* (accept)
Forward	then we're going to head west till we get to New East Tunnel now that might be called something else (4_68_4_10_IG)	*ad* (action directive) *co* (offer)
Backward	okay just after the house (4_232_3_8_IF)	*b* (acknowledge) *^m* (repetition)
Mixed	oh okay so I turn right at the factory (1_178_4_10_IF)	*b* (acknowledge) *aa* (accept) *qy^d* (declarative yes-no question)

The aim of the present study was to determine whether the total number of turns, the frequency of occurrence of different types of turns (single- vs. multiple-label), and the dialogue function (forward vs. backward vs. mixed) varied depending on the participant's role in the Map Task, their gender, and the dyad type. There is no theoretical reason to assume other than the null hypothesis, except for the participant role (IG vs. IF) and the discourse function: we would expect that IGs would produce more forward function turns because they are giving directions to the IF. Conversely, IFs are expected to produce more backward function turns, either to acknowledge or question the information provided by the IG.

4.1. Results

One-sample chi-square tests were conducted to compare raw counts of: total turns, single-label turns, multiple-label turns, forward turns, backward turns, and mixed turns produced by the IG and IF participants. Similar tests compared the number of each type of turns produced by females and males, and by each dyad type. The alpha level for all tests was set at 0.05. Chi-square statistics are presented in Table 3.

Table 3. *One-sample chi-square statistics (cells containing significant results are highlighted in grey).*

	Role (df = 1)	Gender (df = 1)	Dyad (df = 3)
Total number of turns $N = 2{,}815$	$\chi^2 = 0.92$ $p = .336$	$\chi^2 = 74.84$ $p < .001$	$\chi^2 = 217.42$ $p < .001$
Complexity			
Single-label $N = 1{,}898$	$\chi^2 = 12.17$ $p < .001$	$\chi^2 = 35.07$ $p < .001$	$\chi^2 = 147.21$ $p < .001$
Multiple-label $N = 917$	$\chi^2 = 11.12$ $p = .001$	$\chi^2 = 44.06$ $p < .001$	$\chi^2 = 112.14$ $p < .001$
Function			
Forward $N = 957$	$\chi^2 = 260.19$ $p < .001$	$\chi^2 = 53.84$ $p < .001$	$\chi^2 = 113.17$ $p < .001$
Backward $N = 1{,}137$	$\chi^2 = 361.37$ $p < .001$	$\chi^2 = 8.97$ $p = .003$	$\chi^2 = 48.01$ $p < .001$
Mixed $N = 641$	$\chi^2 = 11.27$ $p = .001$	$\chi^2 = 14.08$ $p < .001$	$\chi^2 = 76.39$ $p < .001$

4.1.1. *Total number of turns*

According to the analyses, there was no statistical difference between the two roles (IG and IF) in the raw count of total number of turns used ($N_{IG} = 1{,}382$; $N_{IF} = 1{,}433$), confirming our hypothesis that total number of turns should not vary depending on role. However, the total number of turns was significantly different between genders as well as across the four dyad types. Female participants were found to produce more turns than male participants ($N_f = 1{,}637$; $N_m = 1{,}178$) and female-led dyads produced more turns than male-led dyads, with the highest number of total turns observed in female-female dyads ($N_{f\text{-}f} = 976$; $N_{f\text{-}m} = 803$; $N_{m\text{-}f} = 514$; $N_{m\text{-}m} = 522$).

4.1.2. *Turn complexity*

IGs were found to use significantly more multiple-label turns ($N_{IG} = 509$; $N_{IF} = 408$), forward turns ($N_{IG} = 728$; $N_{IF} = 229$), and mixed turns ($N_{IG} = 363$; $N_{IF} = 278$) than IFs. Conversely, IFs used single-label turns ($N_{IG} = 873$; $N_{IF} = 1{,}025$) and backward turns ($N_{IG} = 248$; $N_{IF} = 889$) more frequently than IGs. As female speakers produced more turns than male speakers overall, this difference remained statistically significant for both the number of single-label turns ($N_f = 1{,}078$; $N_m = 820$) and the number of multiple-label turns ($N_f = 559$; $N_m = 358$). Female-led dyads produced significantly more single-label turns ($N_{f\text{-}f} = 609$; $N_{f\text{-}m} = 604$; $N_{m\text{-}f} = 334$; $N_{m\text{-}m} = 351$) than male-led dyads. Interestingly, female-female dyads produced multiple-label turns more frequently than any other dyad type ($N_{f\text{-}f} = 367$; $N_{f\text{-}m} = 199$; $N_{m\text{-}f} = 180$; $N_{m\text{-}m} = 171$).

4.1.3. *Turn function*

As expected, IGs differed from IFs in their frequency of producing forward and backward function turns. The use of forward turns was more prevalent in the speech of IGs than IFs ($N_{IG} = 728$; $N_{IF} = 229$), with the opposite pattern found for the use of backward turns ($N_{IG} = 248$; $N_{IF} = 889$). It is interesting to note that the IG speakers also produced more mixed function turns than the IF speakers ($N_{IG} = 363$; $N_{IF} = 278$). Consistent with gender differences observed for the total number of turns and the number of turns varying in complexity, females produced a higher number of turns of each function than males (forward turns: $N_f = 592$; $N_m = 358$; backward turns: $N_f = 619$; $N_m = 518$; mixed turns: $N_f = 368$; $N_m = 273$). Female-led dyads produced significantly more forward turns ($N_{f\text{-}f} = 342$; $N_{f\text{-}m} = 298$; $N_{m\text{-}f} = 155$; $N_{m\text{-}m} = 162$), and backward turns ($N_{f\text{-}f} = 346$; $N_{f\text{-}m} = 339$; $N_{m\text{-}f} = 221$; $N_{m\text{-}m} = 231$) than male-led dyads. Female-female dyads produced mixed function turns more frequently than any other dyad type ($N_{f\text{-}f} = 256$; $N_{f\text{-}m} = 130$; $N_{m\text{-}f} = 130$; $N_{m\text{-}m} = 125$).

4.1.4. *Combined influence of Role and Gender*

To disentangle the influence of conversation role from gender on the number and type of turns produced, a two-factor chi-square analysis was carried out. A significant relationship between role and gender was found for the total number of turns used in map tasks, $\chi^2(1, N = 2{,}815) = 31.41$, $p < .001$, for the number of single-label turns, $\chi^2(1, N = 1{,}898) = 48.84$, $p < .001$, and for the number of backward turns, $\chi^2(1, N = 1{,}137) = 47.88$, $p < .001$.

Females produced significantly more turns than males as IGs ($N_f = 877$; $N_m = 505$), but those gender differences were not as pronounced when males and females assumed the role of IFs ($N_f = 760$; $N_m = 673$). IG males produced significantly fewer single-label turns ($N = 302$) than IF males ($N = 518$), IG females ($N = 571$), and IF females ($N = 507$). Finally, IG females produced backward turns ($N = 183$) more frequently than IG males ($N = 65$), while the raw counts of this turn category was comparable (and, as expected, significantly higher) for IF males and females IFs ($N_f = 436$; $N_m = 453$).

5. Conclusions

This project contributes to the field of Dialogue Analysis by building upon the work of [5] and [4] and by further extending the SWBD-DAMSL Dialogue Coding Scheme [10] for Map Tasks. We refined the coding of questions by adding two labels for declarative questions, and specified the coding of short answers and expansions according to dialogue context.

The analysis of the distribution of DAs according to gender and Map Task role presented in Section 5 revealed interesting patterns which contribute to a better understanding of gender and conversation. No single independent variable (role, gender, and gender dyad) accounted for the communicative pattern variation observed in the data. Females did not conform to what we might expect from the conversational role: while producing more forward-function turns as IG than IF, they produced more backward-function turns than male IGs. Females also produced more turns and more complex turns than males, with this being even more pronounced in female-female dyads. Finally, IFs produced more single-label turns while IGs produced more multiple-label and mixed turns. Further investigation needs to be undertaken to explain these varied and complex data patterns.

Future work on this subset of AusTalk Map Tasks will examine the differences between speakers using total speaking time as a proxy for task success (all AusTalk Map Tasks were successful but they differed in length, with some pairs completing the task much more quickly than others) and turn complexity as a potential marker of difficulty in negotiating the map discrepancies. Another direction would be to analyse the differences between turns dedicated to task/conversation management and whether they also vary depending on the role, gender, and dyad.

In addition to providing preliminary observations on the influence of Map Task role and gender on the complexity and function of DAs in conversational speech, this project is a good example of what can be done with the AusTalk corpus, the audio-visual corpus of Australian English created by the Big ASC project [1]. This study built upon the Annotation Task of the Big ASC, which had produced transcriptions for selected components (both spontaneous and read speech) for a subset of the 853 AusTalk speakers; it also contributed to the data repository of the Alveo Virtual Lab, by adding specific annotations to the AusTalk corpus. Thus, an important outcome of the project consists of a revised set of transcriptions for 16 AusTalk Map Tasks, combining both IG and IF, and of DA annotations for 12 of those Map Tasks. As an early user of the new Alveo functionality allowing adding annotations to data held in Alveo, this project helped with the development and testing of the Alveo Annotation Contribution tool and provided feedback on the usability of the tool and on the interface.

6. Acknowledgements

The project was funded by a Transdisciplinary Grant from the Centre of Excellence for the Dynamics of Language (COEDL).

7. References

[1] Burnham, D., et al., *Building an audio-visual corpus of Australian English: large corpus collection with an economical portable and replicable Black Box*, in *Interspeech 2011*. 2011: Florence, Italy.

[2] Estival, D., et al., *AusTalk: an audio-visual corpus of Australian English*, in *9th Language Resources and Evaluation Conference (LREC 2014)*. 2014: Reykjavik, Iceland.

[3] Anderson, A.H., et al., *The HCRC Map Task Corpus.* Language and Speech, 1991. **34**(4): p. 351-366.

[4] Fletcher, J. and L. Stirling, *Prosody And Discourse In The Australian Map Task Corpus*, in *The Oxford Handbook of Corpus Phonology*. 2014, Oxford University Press.

[5] Fletcher, J., et al., *Intonational rises and dialog acts in the Australian English map task.* Language and speech, 2002. **45**(3): p. 229-253.

[6] Cassidy, S., D. Estival, and F. Cox, *Case Study: The AusTalk Corpus*, in *Handbook of Linguistic Annotation*, J. Pustejovski and N. Ide, Editors. 2017, Springer Verlag. p. 1287-1302.

[7] Cassidy, S. and D. Estival, *Supporting Accessibility and Reproducibility in Language Research in the Alveo Virtual Laboratory.* Computer Speech & Language, 2017. **45**: p. 375-391.

[8] Barras, C., et al., *Transcriber: development and use of a tool for assisting speech corpora production.* Speech Communication, 2000. **33**(1-2).

[9] Boersma, P. and D. Weenink. *Praat: doing phonetics by computer (Version 5.3.51)*. 2014; Available from: http://www.praat.org/.

[10] Jurafsky, D., L. Shriberg, and D. Biasca. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation, Coders Manual, Draft 13* 1997; Available from: https://web.stanford.edu/~jurafsky/ws97/manual.august1.html.