Good call!
Grounding in a Directory Enquiries Corpus

Christine Howes **Anastasia Bondarenko** **Staffan Larsson**
Centre for Linguistic Theory and Studies in Probability (CLASP)
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg, Sweden
christine.howes@gu.se

Abstract

This paper describes the collection of a corpus of telephone directory enquiries conversations. We analyse the feedback used in the corpus and discuss implications for dialogue systems.

1 Introduction

Effective communication requires collaboration between all participants, with dialogue co-constructed by speakers and hearers. Even in contexts such as lectures or storytelling, which are largely monological (Rühlemann, 2007), listeners provide frequent feedback. This feedback demonstrates whether or not they have *grounded* the conversation thus far (Clark, 1996), i.e. whether something said can be taken to be understood, and comes in the form of relevant next turns, or backchannels (e.g. 'yes', 'yeah', Example 1; lines 6 and 8[1] or 'mm').[2] Other responses, such as clarification requests (e.g. Example 1; lines 10 and 17) indicate processing difficulties or lack of coordination and signal a need for repair (Purver, 2004; Bavelas et al., 2012).

These communicative grounding strategies (Clark and Brennan, 1991; Traum, 1994) enable dialogue participants to manage the characteristic divergence and convergence that is key to moving dialogue forward (Clark and Schaefer, 1987, 1989), and are therefore crucial for dialogue agents. Importantly, feedback is known to occur subsententially (Howes and Eshghi, 2017), but most dialogue models do not operate in an incremental fashion that would allow them to produce or interpret feedback in a timely fashion.

(1) DEC07:1–32

1	Caller	hello
2	Operator	hello
3	Caller	hello
4	Operator	how may i help you?
5	Caller	oh hi i'm uh looking for some phone numbers
6	Operator	yes
7	Caller	er here in london
8	Operator	yeah
9	Caller	and the first
10		one is rowans tenpin bowl
11	Operator	can you repeat that for me?
12	Caller	rowans tenpin bowl
13		so it's rowan
14		R O W A N S
15	Operator	yes
16	Caller	tenpin
17	Operator	tenpin?
18	Caller	yeah
19	Operator	the number ten
20	Operator	and pin?
21	Caller	yes
22	Caller	yes
23	Operator	tenpin
24	Operator	road?
25	Caller	bowl
26	Operator	th- like the bird?
27	Caller	uh like bowling
28	Operator	uh bowling
29	Caller	bowl
30	Operator	yes
31		the thing you eat from right?
32		okay here we go

While it is difficult to compare corpus studies of feedback, as terms such as backchannels and repair have not been used consistently in the literature (see Fujimoto, 2007, for review), there are a number of quantificational studies of feedback that bear mentioning. One of the earliest is that described in Duncan (1972, 1974), which presents a detailed multimodal annotation of backchannel responses, and finds that in 885 'units' (roughly corresponding to utterances) there are a total of 71 instances of feedback (8%). Corpus studies that cover aspects of feedback include (Fernández, 2006), whose annotations of non-sentential utterances (NSUs) in a subcorpus of the British

[1] Examples are all taken from our Directory Enquiries Corpus (DEC), described below.

[2] In face-to-face dialogue this includes non-linguistic cues (e.g. nods), but as our corpus is telephone conversations, we do not consider these here.

National Corpus (BNC; Burnard, 2000) include the classes 'acknowledgements' (5% of all utterances), and 'clarification ellipsis' (1%). However, as her focus is on NSUs, Fernández (2006) deliberately excludes cases in overlap, which means many genuine feedback utterances will be missed (Rühlemann, 2007). For clarification requests, the numbers reported in (Fernández, 2006) are also an underestimate, as she is not concerned with sentential cases (e.g. "what do you mean?"). In another BNC study, Purver (2004) found that CRs made up just under 3% of utterances, whilst Colman and Healey (2011) found different levels of CRs in different dialogue domains, with more in the task oriented Map Task (Anderson et al., 1991). Interestingly, this varied significantly depending on role; route followers produced significantly more CRs than route givers. Additionally, and importantly for phone conversations, participants in the Map Task also produce more backchannels when they are not visible to one another (Boyle et al., 1994)

Although using low-level features (Cathcart et al., 2003; Gravano and Hirschberg, 2009) may allow a dialogue model to sound 'more human', it can't provide any insight into why feedback occurs where it does, or whether there are different appropriate responses to feedback dependent on its positioning and other characteristics. It is also unclear whether models in which feedback incorporates reasoning about the intentions or goals of one's interlocutor (Visser et al., 2014; Buschmeier and Kopp, 2013; Wang et al., 2011) presuppose a level of complexity that is unnecessary in natural conversation (Gregoromichelaki et al., 2011).

Here, we focus on feedback in an extremely restricted domain – that of telephone directory enquiries (see also Clark and Schaefer, 1987; Bangerter et al., 2004), which can be seen as a good test case for dialogue systems. Directory enquiries is a real world application for dialogue systems (e.g. Chang, 2007) that has particular features that can be problematic for a speech recogniser, such as understanding names which are not present in an existing lexicon over a noisy channel. As we argue below, this is a particularly good domain for studying feedback, as feedback should be more frequent and necessary than in less restricted domains. The reasons for this are two-fold. Firstly, in task-oriented dialogue, where information transfer is crucial for success, and avoiding miscommunication is vital, feedback is more common than in less goal-directed conversations (Colman and Healey, 2011). Secondly, verbal feedback is more frequent in dialogues where participants cannot see each other, and therefore do not have the ability to employ non-verbal feedback (Boyle et al., 1994), such as telephone conversations. In addition, the specific task of a directory enquiries call is less asymmetric than many tasks used to study dialogue, such as the Map Task (Anderson et al., 1991), because both participants act as 'information giver' (caller for name to be looked up; operator for phone number) and 'information receiver' (the reverse) at different stages in the dialogue. Additionally, in contrast to corpora which have similar features (such as SRI's Amex Travel Agent Data, Kowtko and Price, 1989), relevant parts of the dialogue (names and numbers, see below) do not require anonymisation.

In this paper, we present a new corpus of human-human telephone directory enquiries dialogues, and explore the strategies for feedback that human participants use, especially in cases where misunderstandings arise. We suggest that dialogue models need to be able to perform incremental grounding, particularly in the context of spelling out words and dictating number sequences, with a number of increasingly specific strategies available for both acknowledgements and clarifications. The complete corpus (transcriptions, audio and annotations) is freely available on the Open Science Framework (`osf.io/2vjkh`; Bondarenko et al., 2019) thus aiding in the development of spoken dialogue systems that need to both acquire and offer accurate information to the user (e.g. directory enquiries, travel agents etc).

2 Method

2.1 Data collection

The data was collected with the help of 14 volunteers who were paired up for each recording session. Eight of the volunteers were male and six were female. The participants were native speakers of a number of different languages and had various levels of English proficiency.

Each pair of participants was instructed that they were to take turns playing the roles of a directory service enquiries caller and operator. Each caller was provided with a list of three businesses located in London, and told that their task was to find out the phone numbers of the businesses on

their list through a telephone conversation with the operator. The operators task in turn was to provide the caller with the phone numbers using the online Phone Book service (thephonebook.bt.com). Each caller made two calls to the operator who was situated in the studio. The recording sessions resulted in 4 dialogues per pair (28 in total) with the shortest dialogue duration being 2 minutes 31 seconds and the longest one being 10 minutes 46 seconds.

2.2 Transcription

The audio recordings were transcribed using ELAN (Brugman and Russel, 2004).

2.3 Annotation

All of the transcripts were manually annotated, with the overview of annotations used shown in Table 1. Two dialogues (281 utterances) were annotated by two coders to ensure inter-rater reliability. Cohen's kappa tests showed good agreement for all tags: `turn-type (ack/CR/C)` $\kappa = 0.635$; `AckType` $\kappa = 0.625$; `CRType` $\kappa = 0.689$.

2.4 Feedback subtypes annotation

Following observations of the data, we further annotated our feedback utterances into subtype. For acknowledgements these are:

`Continuer` acknowledgement/backchannel words like "okay", "yeah", "yes, "mmhm" (e.g. Example 1; line 8).

`Verbatim` verbatim repetitions of (parts of) previous utterances (e.g. Example 1; line 27)

`Paraphrase` paraphrased repetitions of (parts of) previous utterances

`Confirm` confirmation phrases like "correct", "exactly", "thats correct"

`Appreciate` appreciative response to the previous utterance: "great", "good", "perfect".

For clarification requests these are:[3]

`General request` indicates a non-specific lack of perception/understanding of other speaker's previous utterance (e.g. "sorry?", "what?")

`Repeat request` asks other speaker to repeat a previous utterance (e.g. Example 1; line 11)

`Confirmation request` asks other speaker to provide a confirmation (e.g. Example 1; line 17)

`Spelling request` asks other speaker to spell out the name of the queried business or its address (e.g. "could you spell that for me please?", "is that a W?")

2.5 Content annotation

Since the main purpose of the data collection was to investigate the domain of telephone directory enquiries each of the the utterances was also labelled according to its content: namely, whether it includes any information about the names, addresses and phone numbers of businesses. Each utterance labelled with any of these was then labelled according to the form such information was conveyed in:

`Word (part)` speaker mentions the name of a business or its address in full or in part

`Spelling installment (part)` speaker provides a spelling for the name or the address of a business in full or in part, usually in installments of one or more letters

`Dictation installment (part)` speaker dictates a phone number in full or in part, usually in installments of one or more digits

`PreviousWord/spelling/dictation, PreviousContent` each utterance is also annotated with the content and form labels of the previous utterance.

3 Results

In our 28 dialogues, there were a total of 4165 utterances, or 3002 speaker turns (for our purposes a turn constitutes multiple consecutive utterances by the same speaker with no intervening material from the other participant). The shortest dialogue consists of 64 utterances (48 turns) and the longest consists of 246 utterances (190 turns). 1285 of these utterances are acknowledgements, which constitutes 31% of utterances or 43% of turns. There are also 277 clarification requests, i.e. 7% of utterances and 9% of turns.[4] This is higher than found in previous studies (Purver,

[3] As pointed out by an anonymous reviewer, the categories for acknowledgements may conflate form and function, whilst those for CRs do not consider the form. This may mean that we miss important parallels or differences between acknowledgements and clarification requests and we intend to address this in future work.

[4] As the pattern of results is consistent over turns or utterances, for the remainder of this paper we focus on the by utterance numbers.

Tag	Value	Explanation
`acknowledge (Ack)`	y/n	For all utterances: does this sentence contain a backchannel (e.g. 'yeah', 'mhm', 'right') or a repeated word or phrase acknowledging the proposition or speech act of a previous utterance? (Note this category does not include direct answers to yes/no questions)
`clarification request (CR)`	y/n	For all utterances: does this utterance contain a clarification request, indicating misunderstanding of the proposition or speech act of a previous utterance
`clarify (C)`	y/n	For utterances following a clarification request: does this utterance contain a response to a clarification request, clarifying the proposition or speech act of a previous utterance?

Table 1: Annotation Tags

2004; Fernández, 2006; Boyle et al., 1994, a.o.), and, as discussed in the introduction, is probably due to the nature of the task.

As shown in Table 2, operators produce more acknowledgements and clarification requests than callers (Acks: 36% vs 26% $\chi^2_1 = 48.466, p < 0.001$; CRs: 9% vs 4% $\chi^2_1 = 36.961, p < 0.001$). This result stems from the greater possibility for error in the understanding of names compared to numbers (see section 3.1 below).

	Role					
	Caller		**Operator**		**Total**	
Ack	559	*26%*	726	*36%*	1285	*31%*
C	189	*9%*	64	*3%*	253	*6%*
CR	94	*4%*	183	*9%*	277	*7%*
(blank)	1306	*61%*	1044	*52%*	2350	*56%*
Total	**2148**	***100%***	**2017**	***100%***	**4165**	***100%***

Table 2: Summary of results by speaker role

3.1 Asymmetry of information

As shown in Tables 3 and 4, as in Colman and Healey (2011), the pattern of feedback mirrors the asymmetry of roles, with information receiver (i.e. operator for the business name, and the caller for the phone number) providing the majority of acknowledgements and clarification requests.

	Role					
	Caller		**Operator**		**Total**	
Ack	50	*11%*	441	*68%*	491	*44%*
C	78	*16%*	1	*0%*	79	*7%*
CR	3	*1%*	100	*15%*	103	*9%*
(blank)	342	*72%*	105	*16%*	447	*40%*
Total	**473**	***100%***	**647**	***100%***	**1120**	***100%***

Table 3: Results by speaker role where the previous utterance is about a business name

	Role					
	Caller		**Operator**		**Total**	
Ack	364	*73%*	92	*28%*	456	*55%*
C	0	*0%*	30	*9%*	30	*4%*
CR	60	*12%*	0	*0%*	60	*7%*
(blank)	75	*15%*	210	*63%*	285	*34%*
Total	**499**	***100%***	**332**	***100%***	**831**	***100%***

Table 4: Results by speaker role where the previous utterance is about a business phone number

3.2 Feedback subtypes

As shown in Table 5, most of the acknowledgements in our corpus consist of continuers, with 772 (60%) acknowledgements containing at least one continuer. The next most common type of acknowledgement is a verbatim repeat of material from a prior utterance, with 492 (38%) acknowledgements. For a dialogue system, this is good news: simple utterances of just a continuer or repeated material accounts for 91% of all acknowledgements, suggesting that these may be the only two strategies that need to be implemented for both production and comprehension.

For clarification requests (Table 6), the majority (48%) are confirmation requests – checking that something has been understood by offering a provisional interpretation. These serve to pinpoint the (potential) source of miscommunication in a way that the more general types do not (see also Ginzburg, 2012). In practice, they are very similar to the verbatim acknowledgements, as in example 1 line 17, but with questioning intonation suggesting that they are more tentative. These ought to therefore be generatable in the same way as verbatim acknowledgements. The data suggest a scale of feedback, analogous to Clark and colleagues' levels of evidence of understanding

(Clark and Brennan, 1991; Clark and Schaefer, 1989; Clark, 1996), with listener confidence being a key component of which type of feedback is appropriate.

Type(s)	Number	%
Appreciate	5	0.4%
Confirm	21	1.6%
Confirm, Continuer	1	0.1%
Continuer	718	55.9%
Continuer, Appreciate	9	0.7%
Continuer, Appreciate, Continuer	1	0.1%
Continuer, Confirm	9	0.7%
Continuer, Paraphrase	2	0.2%
Continuer, Verbatim	3	0.2%
Paraphrase	25	1.9%
Paraphrase, Continuer	2	0.2%
Verbatim	456	35.5%
Verbatim, Appreciate	1	0.1%
Verbatim, Continuer	25	1.9%
Verbatim, Continuer, Appreciate	2	0.2%
Verbatim, Paraphrase	1	0.1%
Verbatim, Verbatim	4	0.3%
Total	**1285**	**100%**

Table 5: Types of acknowledgement

Type	Number	%
Confirmation request	134	48.4%
General request	28	10.1%
Repeat request	64	23.1%
Spelling request	51	18.4%
Total	**277**	**100%**

Table 6: Types of clarification request

3.3 Strategies

As there is greater scope for miscommunication in the transmission of names than numbers, we now focus on the examples where the feedback follows an utterance whose content is about a name.[5] For these cases, there is large variability in how easily the names are conveyed, with the number of turns taken from the first mention of any part of the name to the operator confirming that they have found the number ranging from 2 utterances to 82 utterances, with 3 (of 84) cases unresolved.

Table 7 shows that of the turns following an utterance about a business name, 45% contain a spelling installment, or part of one, with similar proportions for acknowledgements (36%) and clarification requests (41%), with only 15% (acks 12%, CRs 21%) relating to the word level. This shows that models of dialogue need to be able to produce and interpret increments of different sizes – potentially of a single letter, as people do when they are pinpointing sources of (potential) trouble within an unfamiliar name.

Tables 8 and 9 demonstrate that feedback strategies are highly dependent on the information giving strategy employed in the preceding utterance. While generic strategies (continuers or non-specific repairs such as "what?") are common and always available, participants are also likely to match the prior strategy used in their feedback – it is, for example, rare to acknowledge or clarify a spelling installment with a word, and vice versa.

3.4 Qualitative results

Examples 2–9 show a variety of these strategies in action. In Example 2, the Operator relies on continuer acknowledgements, which, according to Clark and colleagues' model of levels of evidence of understanding, are weaker signals of understanding than e.g. verbatim repeats and might be therefore more likely to allow misunderstandings to occur. Example 3 from another pair shows the same business name split into different increments (with the first half of the name "bistro" treated as an independent word and the rest spelled out in increments of 3 letters; see also section 3.5, below), with different feedback techniques for different subparts of the utterance – a continuer at line 126, a verbatim acknowledgement at line 128.

(2) DEC11:88–98

88	Operator	er can you spell bistrotheque for me?
89	Caller	abs-
90	Caller	sure er it's
91	Caller	B I S
92	Operator	yes
93	Caller	T R O
94	Operator	mmhm
95	Caller	T H E
96	Operator	okay
97	Caller	Q U E
98	Operator	er yes i have it here for you

(3) DEC3:123–128

123	Caller	so bistro
124	Caller	T
125	Caller	H E
126	Operator	yeah
127	Caller	Q U E
128	Operator	Q U E

Example 4 splits the business name into two increments of 3 and 4 letters respectively, and is acknowledged by verbatim repeats in each case.

[5] Note that row totals in Tables 7, 8 and 9 do not add up to 100% as some turns contain more than one strategy.

	Ack		CR		Total	
Spelling installment	137	*28%*	31	*30%*	394	*35%*
Spelling installment part	41	*8%*	11	*11%*	107	*10%*
Word	21	*4%*	5	*5%*	47	*4%*
Word part	40	*8%*	16	*16%*	127	*11%*
Other	253	*52%*	42	*41%*	452	*40%*
Total	**491**	***100%***	**103**	***100%***	**1120**	***100%***

Table 7: Strategies for feedback following an utterance about a business name

	Previous utterance content type								
	Spelling installment		Spelling instmt part		Word		Word part		Total
Spelling installment	127	*40%*	9	*20%*	0	*0%*	1	*1%*	137
Spelling installment part	23	*7%*	18	*39%*	0	*0%*	4	*6%*	41
Word	3	*1%*	2	*4%*	10	*20%*	6	*9%*	21
Word part	3	*1%*	0	*0%*	15	*30%*	22	*32%*	40
(continuer/confirm/appreciate)	171	*54%*	18	*39%*	25	*50%*	42	*62%*	253
Total	**319**	***100%***	**46**	***100%***	**50**	***100%***	**68**	***100%***	**491**

Table 8: Strategies for acknowledgements about a business name by previous utterance content type

A common strategy for avoiding miscommunications in spellings is developed in Example 5: namely using unambiguous words which start with the same letter. This strategy is prompted by the operator's clarification request in line 19. Note that the acknowledgements provided by the operator here are sometimes only the word (e.g. line 23 "america") but sometimes include the letter in a direct repeat of the whole utterance (e.g. line 35 "R for Russia"). In our corpus, different pairs come up with different sets of words for spelling out the letters (e.g. country/city names, as here, or people's first names – note that this choice can also be the source of miscommunication, as in Example 12). This strategy can be initiated by either participant, or in co-constructions (as in Example 7), and, after repeated interactions, participants may use this strategy productively – even dropping the letter with the country name standing in for the whole, as in Example 6 (this mirrors the way participants strategically align in tasks such as the Maze Game; Mills and Healey, 2006).

(4) DEC16:54–61

54	Caller	the next place i'm looking for is called
55	Caller	er tayyabs which is spelled
56	Caller	T A Y
57	Operator	T A Y
58	Caller	Y A B S
59	Operator	Y A B S
60	Caller	it's a restaurant
61	Operator	okay

(5) DEC28:17–35

17	Caller	okay so it starts with a
18	Caller	L
19	Operator	L?
20	Caller	as in london
21	Operator	yes
22	Caller	A as in america
23	Operator	america
24	Caller	er U
25	Caller	as in er
26	Caller	er under
27	Caller	<laugh>
28	Operator	under yes
29	Caller	er D as in denmark
30	Operator	denmark
31	Caller	E as in england
32	Operator	england
33	Caller	and R
34	Caller	for russia
35	Operator	R for russia

(6) DEC26:61–69

61	Caller	it's it's a restaurant by name tayyabs
62	Operator	okay can you spell that for me please?
63	Caller	should i
64	Caller	yes it's a thailand
65	Operator	yes
66	Caller	america
67	Operator	yes
68	Caller	yugoslavia
69	Operator	yes
:	:	:

(7) DEC28:138–141 Co-construction

138	Caller	and K for er
139	Caller	<laugh>
140	Operator	as in king?
141	Caller	k- king <laugh> yeah

3.5 Increments

People often break the names into increments to aid understanding, but what counts as an incre-

	Previous utterance content type								
	Spelling installment		**Spelling instmt part**		**Word**		**Word part**		**Total**
Spelling installment	24	*52%*	3	*43%*	1	*4%*	4	*19%*	31
Spelling installment part	8	*17%*	3	*43%*		*0%*		*0%*	11
Word		*0%*		*0%*	4	*16%*		*0%*	5
Word part	2	*4%*	1	*14%*	5	*20%*	10	*48%*	16
(generic repair)	17	*37%*		*0%*	16	*64%*	12	*57%*	42
Total	**46**	***100%***	**7**	***100%***	**25**	***100%***	**21**	***100%***	**103**

Table 9: Strategies for clarification requests about a business name by previous utterance content type

ment is not fixed, and may be further subdivided in case of failure. Examples 8 and 9 show two different ways in which the same name was divided into increments, with Example 9 having many more utterances, including several verbatim acknowledgements to convey the same information.

(8) DEC7:89–98

89	Caller	phoenicia mediterranean food
90	Operator	can you repeat that for me?
91	Operator	tenicia?
92	Caller	yeah
93	Caller	it's P H
94	Caller	O E N
95	Operator	mmhm
96	Operator	co- continue please
97	Caller	I C I A
98	Operator	I C I A

(9) DEC23:101–117

101	Caller	yeah it's phoenicia
102	Operator	clomissia?
103	Caller	mediterranean food
104	Caller	yes you spell it with a P
105	Operator	P
106	Caller	H
107	Caller	O
108	Operator	H O
109	Caller	E
110	Operator	yes P H O E
111	Caller	E N
112	Operator	N
113	Caller	A C
114	Operator	A C
115	Caller	A-
116	Caller	I A
117	Operator	I A

3.6 Repair Strategies

In our data there is some indication that participants are generally good at predicting potentially problematic elements and further specifying those before they lead to miscommunication, such as non-conventional spellings of words as in Examples 10 and 11.

(10) DEC20:4–9

4	Caller	the first one being first one being one called cittie of yorke which is C I T T I E of
5	Caller	yorke spelled with an E at the end
6	Operator	cittie of yorke with two Ts?
7	Caller	cittie of yorke where cittie isn't
8	Caller	C I T Y it's C I T T I E
9	Operator	yeah

(11) DEC10:59–9

59	Caller	it's called lyle's
60	Caller	with a Y
61	Operator	lyle's

In general, misunderstandings are resolved quickly and locally, however, there are also interesting cases where misunderstandings persist, such as Example 12, with the specific problematic letter in the name taking 57 utterances to resolve. In this case, as in 13, the participants started by trying to just spell out the names (which can be ambiguous, especially in noisy settings) and then switch strategy to a more specific method (here using the initial letter of a name or place) when the initial strategy fails.

(12) DEC22:82–139

82	Caller	with a - filip with an F
83	Operator	filip
84	Operator	yeah
:	:	:
107	Caller	er
108	Operator	pilip
109	Caller	fanny
110	Operator	mmhm
111	Caller	fanny
:	:	:
113	Operator	P
114	Operator	P as in panda
115	Operator	right?
116	Caller	sorry i didn't hear you
117	Operator	P
118	Operator	the next one is a P
119	Operator	as in panda
120	Caller	P?

121	Operator	or okay
122	Operator	then
123	Caller	no
124	Caller	it's er
:	:	:
133	Caller	uh fanny
134	Operator	<unclear> I don't know that name funny?
135	Caller	yeah or like filip but with an F
136	Caller	or if you say fruits
137	Operator	with an F?
138	Operator	okay
139	Caller	F yeah

(13) DEC25:67–112 Change of strategy

67	Caller	yes and the business i was looking for hot- it's a hotel it's called hotel wardonia
:	:	<lines 68–94 spell out the name >
95	Operator	er i'm sorry i couldn't find any result for
96	Operator	hotel swarbonia maybe i spelled
97	Operator	wrong
98	Caller	yes i can spell that once again
99	Operator	yes please
100	Caller	it's er W for wales
101	Operator	er so it's hotel first?
102	Caller	yes it's hotel and W for washington yeah
103	Operator	W for washington
104	Caller	yeah then A for er
105	Caller	atlanta
106	Operator	yeah

In Example 14, one of the few cases where misunderstandings did not get resolved, it is clear that the participants are unable to align due to the similarity in sound of a 'B' and a 'V' (especially for the native Spanish caller). Note that this pair did not manage to ascertain the source of the trouble, which a letter + name using the initial letter strategy may have resolved. A dialogue model should therefore be able to generate this type of strategy for disambiguating letter sounds, even where the human user does not do so.

(14) DEC14:4–112 Complete failure

4	Caller	er one is a pub
5	Caller	it's called the star tavern
6	Operator	can you repeat please?
7	Caller	the star
8	Caller	tavern
:	:	:
16	Caller	yeah the well the place is called the star tavern
17	Operator	the star
18	Caller	tavern
19	Caller	yeah
:	:	:
29	Operator	i'm not sure if i heard the name of the place correctly
30	Operator	can you repeat?
31	Caller	yeah the the name of the place the
32	Operator	yes
33	Caller	the tavern it's the star
34	Caller	star like a star in the sky you know <laugh>
35	Operator	yes
36	Caller	the night
37	Operator	mmhm
38	Caller	er tavern
39	Operator	can you spell it er please ta-?
40	Caller	the address you say?
41	Operator	er the star ta- what?
42	Caller	the star tavern
:	:	:
58	Caller	and it's tavern it's T A
59	Operator	and then
60	Caller	V E er <R> un <N>
61	Caller	N
62	Caller	sorry
:	:	:
72	Operator	T A B E R N
73	Operator	is that correct?
74	Caller	yeah
:	:	:
94	Caller	okay you have the name of the place correct?
95	Caller	right?
96	Operator	star tabern right?
97	Caller	yeah
:	:	:
112	Operator	website still says we're sorry we co- couldn't find any results

4 Discussion and future work

We have presented a new corpus of telephone directory enquiries that is freely available, and a preliminary exploration of the feedback used in these dialogues.

In future work, we hope to provide a formal model of incremental grounding incorporating the phenomena observed in our corpus including spelling and dictation installments, as well as a comparison with previous work (e.g. Purver, 2004; Fernández, 2006; Rieser and Moore, 2005). Work on formal modelling of grounding (e.g. Traum, 1994; Larsson, 2002; Visser et al., 2014) has often assumed that the minimal units being grounded are words. In a complete model, this needs to be complemented by the grounding of subparts of words, including single letters. Work in this direction includes Skantze and Schlangen (2009), where dictation of number sequences is used as a test case "micro-domain" for an implemented model of incremental grounding. However, this system works exclusively on the level of single digits (or sequences thereof). A challenge for a general model of grounding is to combine grounding of whole words/utterances with grounding of sub-parts of words, using the many strategies that people do.

Acknowledgements

This work was supported by two grants from the Swedish Research Council (VR): 2016-0116 – Incremental Reasoning in Dialogue (IncReD) and 2014-39 for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg. We are also grateful to the three anonymous reviewers for their helpful comments.

References

Anne Anderson, Miles Bader, Ellen 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 data. *Language and Speech*, 34(4):351–366.

Adrian Bangerter, Herbert H Clark, and Anna R Katz. 2004. Navigating joint projects in telephone conversations. *Discourse Processes*, 37(1):1–23.

Janet Beavin Bavelas, Peter De Jong, Harry Korman, and Sara Smock Jordan. 2012. Beyond backchannels: A three-step model of grounding in face-to-face dialogue. In *Proceedings of Interdisciplinary Workshop on Feedback Behaviors in Dialog*.

Anastasia Bondarenko, Christine Howes, and Staffan Larsson. 2019. Directory enquiries corpus. Available at osf.io/2vjkh.

Elizabeth A Boyle, Anne H Anderson, and Alison Newlands. 1994. The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and speech*, 37(1):1–20.

Hennie Brugman and Albert Russel. 2004. Annotating multi-media/multi-modal resources with ELAN. In *4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 2065–2068. European Language Resources Association.

Lou Burnard. 2000. *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services.

Hendrik Buschmeier and Stefan Kopp. 2013. Co-constructing grounded symbols–feedback and incremental adaptation in human-agent dialogue. *KI-Künstliche Intelligenz*, 27(2):137–143.

Nicola Cathcart, Jean Carletta, and Ewan Klein. 2003. A shallow model of backchannel continuers in spoken dialogue. In *Proceedings of the tenth EACL conference*, pages 51–58. Association for Computational Linguistics.

Harry M Chang. 2007. Comparing machine and human performance for callers directory assistance requests. *International Journal of Speech Technology*, 10(2-3):75–87.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press.

Herbert H. Clark and Susan A. Brennan. 1991. *Grounding in communication*, pages 127–149. Washington: APA Books.

Herbert H. Clark and Edward A. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

Herbert H Clark and Edward F Schaefer. 1987. Collaborating on contributions to conversations. *Language and cognitive processes*, 2(1):19–41.

Marcus Colman and Patrick G. T. Healey. 2011. The distribution of repair in dialogue. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, pages 1563–1568, Boston, MA.

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

Starkey Duncan. 1974. On the structure of speaker–auditor interaction during speaking turns. *Language in society*, 3(2):161–180.

Raquel Fernández. 2006. *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. Ph.D. thesis, King's College London, University of London.

Donna T Fujimoto. 2007. Listener responses in interaction: A case for abandoning the term, backchannel. *Journal of Osaka Jogakuin College*, 37:35–54.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Agustín Gravano and Julia Hirschberg. 2009. Backchannel-inviting cues in task-oriented dialogue. In *INTERSPEECH*, pages 1019–22.

Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, Greg J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Patrick G. T. Healey. 2011. Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, 2(1):199–233.

Christine Howes and Arash Eshghi. 2017. Feedback relevance spaces: The organisation of increments in conversation. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS 2017)*. Association for Computational Linguisitics.

Jacqueline C Kowtko and Patti J Price. 1989. Data collection and analysis in the air travel planning domain. In *Proceedings of the workshop on Speech and Natural Language*, pages 119–125. Association for Computational Linguistics.

Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University. Also published as Gothenburg Monographs in Linguistics 21.

Gregory Mills and Patrick G. T. Healey. 2006. Clarifying spatial descriptions: Local and global effects on semantic co-ordination. In *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, Potsdam, Germany.

Matthew Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, University of London.

Verena Rieser and Johanna Moore. 2005. Implications for generating clarification requests in task-oriented dialogues. In *Proceedings of the 43rd Annual Meeting of the ACL*, pages 239–246, Ann Arbor. Association for Computational Linguistics.

Christoph Rühlemann. 2007. *Conversation in Context: A Corpus-Driven Approach*. Continuum.

Gabriel Skantze and David Schlangen. 2009. Incremental dialogue processing in a micro-domain. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, pages 745–753, Stroudsburg, PA, USA. Association for Computational Linguistics.

David Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester.

Thomas Visser, David Traum, David DeVault, and Rieks op den Akker. 2014. A model for incremental grounding in spoken dialogue systems. *Journal on Multimodal User Interfaces*, 8(1):61–73.

Zhiyang Wang, Jina Lee, and Stacy Marsella. 2011. Towards more comprehensive listening behavior: beyond the bobble head. In *Intelligent Virtual Agents*, pages 216–227. Springer.