Quantifying Ellipsis in Dialogue: an index of mutual understanding

Marcus Colman, Arash Eshghi and Patrick G. T. Healey
Interaction, Media and Communication Research Group
Queen Mary, University of London
E1 4NS UK
{marcus, arash, ph}@dcs.qmul.ac.uk

Abstract

This paper presents a coding protocol that allows naïve users to annotate dialogue transcripts for anaphora and ellipsis. Cohen's kappa statistic demonstrates that the protocol is sufficiently robust in terms of reliability. It is proposed that quantitative ellipsis data may be used as an index of mutual-engagement. Current and potential uses of ellipsis coding are described.

1. Introduction

Spontaneously generated dialogue, whether naturally occurring or task-oriented, rarely sticks to accepted rules of grammar or even politeness. Interruptions, ungrammatical utterances and grunts or other noises are found in the majority of contributions in dialogue corpora. One reason for this is the ubiquitous use of ellipsis; the omission of words or phrases from a contribution which can be inferred or extracted from previous contributions. Ellipsis is optional; the full constituent could serve communication as well as the elliptical version. Where ellipsis occurs across speakers i.e., one participant makes (elliptical) use of another's contribution, it provides a direct index of the mutual-accessibility of the current conversational context (cf. Healey et. al. 2007; Eshghi and Healey, 2007).

In some cases elliptical contributions are obvious, as in the polar response 'yeah', signifying that a question has been heard, understood and considered; however, there are degrees of complexity that would seem to require a close understanding of what another participant is referring to. It is this issue of mutual-accessibility or 'grounding' that we propose can be investigated through the quantification of elliptical phenomena. These phenomena are, we propose, also related to the way referring expressions can contract over repeated use. (e.g. Schober and Clark, 1989; Wilkes-Gibbs and Clark, 1992). The approach taken in Clark et al.'s 'collaborative theory' is that as mutual understanding increases, dialogue contributions become shorter as referring terms become part of the common ground. Clark and Krych (2004) note that various elliptical phrases can be used to establish common ground, from continuers ('uh-huh', 'yeah') or assessments ('gosh') to establishing shared attention through deictic expressions such as 'this', 'that', 'here' and 'there'.

Healey et al. (2007) demonstrated the basic concept and viability of quantifying ellipsis phenomena as a quantitative index of mutual-accessibility of context. They showed that the frequency of use of cross-speaker elliptical expressions in online chat varies systematically depending on whether communication is 'local' i.e. within a single chat room or 'remote'. However, the coding of ellipsis in this study did not follow an explicit protocol. It relied mainly on the distinctions made by Fernandez et al. (2004) but specific measures of reliability and validity were not calculated.

Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue, pages 96–99,
Columbus, June 2008.

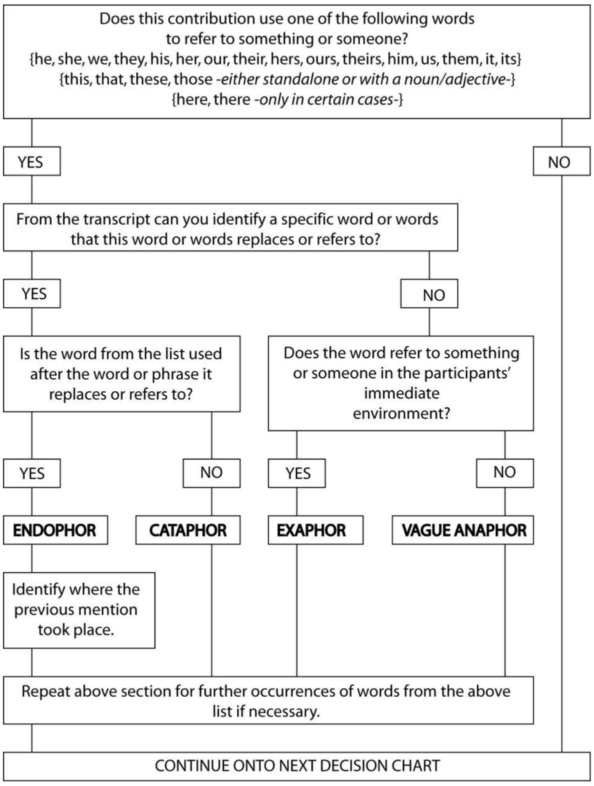

Figure 1. 'Anaphora' decision chart

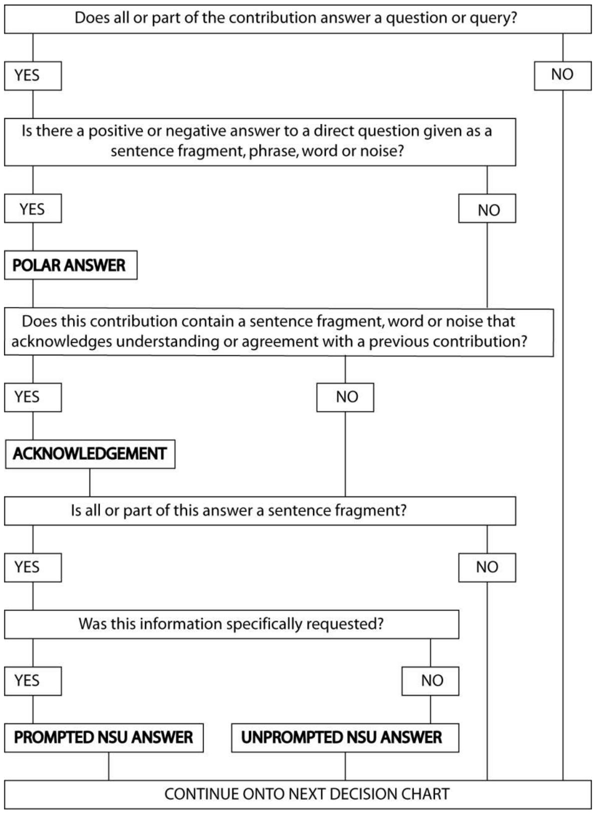

Figure 2. 'Answers' decision chart

In this paper we present an ellipsis coding protocol that provides a set of coding categories and we report the inter-rater reliability scores that have been obtained with it. In order to simplify coding and increase reliability, categories suggested by Fernandez et al. have been collapsed into broader ones. It should be pointed out that we are not, in general, trying to produce an accurate or definitive analysis of ellipsis. The protocol is rather the product of contending with the compromise between robust coding categories and linguistic elegance. The categories presented here are generally ordered in terms of occurrence in order to assist the coder. A contribution to dialogue may contain more than one type of elliptical utterance; contributions are not assigned to one mutually exclusive category. Rather, coders are able to use the protocol to label any part of a dialogue that is elliptical.

2. The Ellipsis Protocol

The protocol is designed as a tool for coding one aspect of dialogue, developed with the intention that users with no specific knowledge of linguistics can use it. As can be seen from Figures 1-4, it consists of four binary branching decision trees that are applied to each contribution in an interaction. Full instructions for use of the protocol have also been written and are available from the authors.

3. Inter-rater reliability

In order to demonstrate reliability between coders, two coders (one computer scientist, one psychologist) applied the ellipsis protocol to a sample of task oriented dialogue. This was taken from the HCRC Map Task corpus (Anderson et al, 1991); a series of dialogues in which one participant attempts to describe a route on a fictional map to another. The longest of these dialogues was chosen to be coded (transcript Q1NC1) which consisted of 446 turns and 5533 words. Cohen's kappa was calculated using the procedure outlined in Howell (1994); see Carletta (1996) for a discussion of the use of kappa in dialogue coding. Kappa in this instance was .81, which shows very high reliability, even by conservative standards (Krippendorff,

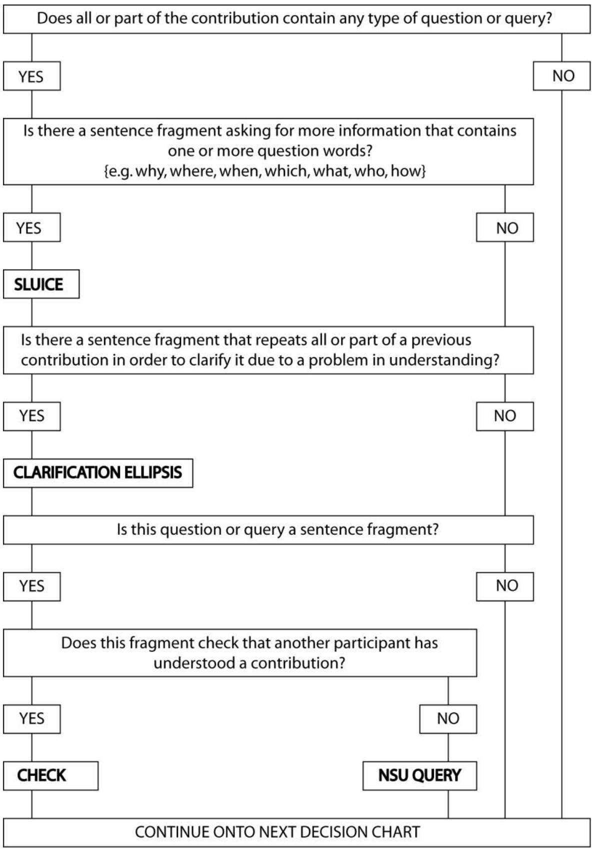

Figure 3. 'Questions' decision chart

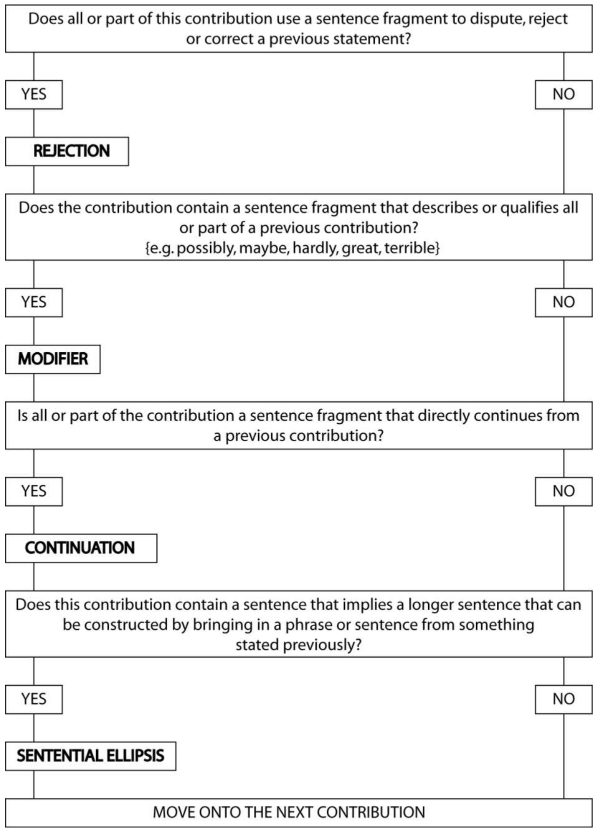

Figure 4. 'Statements' decision chart

1980). Table 1 below presents a breakdown of the instances of categories that were agreed upon.

Table 1 shows the total number and approximate percentage of agreements. Also given, '1.dis' and '2.dis' are the number of observed instances by coders one and two respectively identified but disputed for that particular category. The total number of elliptical or non-elliptical instances coded, from single words or phrases to entire turns was 624; of these, 100 (16%) were disagreed upon and 78 instances (12.5%) were agreed to contain no elliptical phenomena (no ellipsis disagreements = 50). Some categories have very low frequencies; however, previous work suggests that these categories are necessary. To some extent this table shows the limitations of the kappa statistic; coder agreement varies considerably across these categories.

	Endophor	Cataphor	Exaphor	Vague Anaphor
Total	**119**	**2**	**8**	**33**
%	**19**	**.03**	**1.3**	**5.3**
1.dis	12	1	1	20
2.dis	10	3	17	6
	Polar Answer	Acknowledge	Prompted NSU Ans.	Un-prompted NSU Ans.
Total	**113**	**78**	**1**	**7**
%	**18.1**	**12.5**	**0.2**	**1.1**
1.dis	7	15	0	1
2.dis	5	9	1	5
	Sluice	Clarification Ellipsis	Check	NSU Query
Total	**2**	**7**	**20**	**27**
%	**.03**	**1.1**	**3.2**	**4.3**
1.dis	0	0	2	5
2.dis	2	2	0	2
	Rejection	Modification	Continua-tion	Sentential Ellipsis
Total	**2**	**1**	**13**	**13**
%	**.03**	**.002**	**2.1**	**2.1**
1.dis	1	0	3	10
2.dis	4	0	3	3

Table 1. Total agreements by category

4. Discussion

Although mutual-accessibility of context is fundamental to communication, there has not been a reliable method for observing or measuring it. The ellipsis protocol presented here thus provides a useful step in this direction. It gives a standardised coding scheme that can quantify the extent to which speakers can directly access the constituents of each other's turns.

In previous work there have been several different attempts to define taxonomies of elliptical or context dependent utterances. For example, non-sentential utterances (NSUs), e.g. Schlangen and Lascarides (2003); Fernandez and Ginzburg (2002); Fernandez, Ginzburg and Lappin (2007). One issue with these previous approaches is the lack of reliability data; a statistic such as Cohen's kappa is needed in order to demonstrate that a taxonomy or coding scheme can be reliably applied between independent coders. Carletta et al. (1997) presented a reliable coding scheme for the classification of dialogue moves; although there are overlaps between their categories and ours, the questions used in the scheme are intended to establish solely the function of an utterance and importantly, not whether the utterance is elliptical. The protocol presented here achieves a high level of reliability for some of these context dependent phenomena without requiring specific prior knowledge of the relevant linguistic theory.

Further work will code a sample from the BNC (Burnard, 2000) in order to allow comparisons with previous taxonomies. The HCRC map task corpus has previously been examined in terms of various features of dialogue, e.g. Dialogue Games Analysis (Kowtko et al, 1991) and disfluencies (Lickley and Bard, 1998). Ongoing work will develop this through coding the entire HCRC map task corpus; providing data on how ellipsis varies over different conditions such as medium, familiarity and task role.

Acknowledgments

Thanks go to the HCRC group for providing the map task data. Thanks also to Jackie Folkes and Greg Mills for help and advice.

References

Anderson, A., Bader. M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34, 351-366.

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

Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2): 249-254.

Clark, H. H. and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50, 62-81.

Eshghi, A. and Healey, P. G. T. (2007). Collective states of understanding. *Proceedings of the 8th SIGdial workshop on discourse and dialogue*. pp 2-9.

Fernandez, R. and Ginzburg, J. (2002). Non-sentential utterances: a corpus study. *Traitement automatique des languages: dialogue*, 43(2), 13-42.

Fernandez, R., Ginzburg J and Lappin, S, (2004). Classifying ellipsis in dialogue: a machine learning approach. *Proceedings of the 20th international conference on computational linguistics*. pp 240-246.

Fernandez, R. , Ginzburg J and Lappin, S. (2007). Classifying non-sentential utterances in dialogue: a machine learning approach. *Computational Linguistics* 33(3), 397-427.

Healey, P. G. T., White, G., Eshghi, A. and Light, A. (2007). Communication Spaces. *Computer Supported Co-operative Work*, 2007.

Howell, D. C. (1997). *Statistical Methods for Psychology*. Duxbury Press.

Kowtko, J. C., Isard, S. D. and Doherty, G. M. (1991). Conversational games within dialogue. *Proceedings of the espirit workshop on discourse coherence*, 1991.

Krippendorff, K. (1980). *Content Analysis: an introduction to its methodology.* Beverly Hills: Sage Publications.

Lickley, R. and Bard, E. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech,* 41.

Schlangen, D. and Lascarides, A. (2003). The interpretation of non-sentential utterances in dialogue. *Proceedings of the 4th SIGdial workshop on discourse and dialogue,* 2003.

Schober, M. F. and Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211-232.

Wilkes-Gibbs, D. and Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183-194.