Perspectives on Dialogue: Introduction to this Special Issue

Paul Piwek *

Abstract

We briefly summarize the papers in this volume, draw attention to the variety of perspectives that they bring to the subject of dialogue, identify a number of common themes, and conclude with a discussion of directions for further research.

1   Introduction

In many ways, dialogue is the primary setting of language use: it requires no formal training, serves as a means of communication in all cultures, and plays an important role during language acquisition. This has led a number of researchers to propose that a general understanding of language processing needs to start with the study of dialogue (Clark, 1996; Pickering and Garrod, 2004; this volume).

At the same time, dialogue is one of the most challenging topics in the study of language. In particular, spoken dialogue lacks the tidiness of letters, essays and other forms of written discourse. It more often than not involves non-sentential utterances, hesitations and overlapping speech events. Perhaps even more importantly, the ability to engage in a dialogue has been trumpeted as a good indicator of general intelligent behaviour (Turing, 1950). In other words, the capacity to engage in a dialogue could very well be *AI-complete*, i.e., employ all the skills and abilities that constitute human-level intelligence.

The difficulty of isolating dialogue from other human skills and abilities suggests that we cannot afford to study it from one single perspective. To do justice to this insight, this special issue brings together recent work on dialogue which approaches the subject from a variety of perspectives. The papers illustrate that work under these different perspectives has a number of common underlying themes and that connections are being forged that lead to fresh results.

---

* Centre for Research in Computing, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom (E-mail: p.piwek@open.ac.uk)

We hope that this issue will further stimulate cross-disciplinary research into dialogue. In this respect, it is in the spirit of the annual workshops on the Semantics and Pragmatics of Dialogue (SEMDIAL) and the SIGdial workshops on Discourse and Dialogue, which bring together researchers from a variety of backgrounds to discuss recent development in dialogue research.[1]

2 The papers in this volume

Branigan's paper investigates in what ways *two-party dialogue* differs from *multi-party dialogue*. For this purpose, she provides a thorough overview of empirical findings for both two and multi-party dialogue. The overview includes findings on the construction of common ground (covering both textual and situational representations) and audience design. Branigan reaches the conclusion that any differences between two and multi-party dialogue are quantitative rather than qualitative. In both varieties of dialogue, participants actively try to bring about a situation in which they "mutually believe that the parties have understood what the contributor meant to a criterion sufficient for the current purposes" (*The Principle of Responsibility* from Clark & Schaefer, 1987, p. 262).[2] Branigan points out that in two-party dialogues the criterion for understanding has typically been higher than in multi-party dialogues for reasons that are independent of the number of dialogue participants. In particular, most work on two-party dialogue concerns task-oriented dialogues that require a high level coordination between the interlocutors.

Kibble opens with a critique of mentalistic belief and intention-based accounts of dialogue phenomena. Amongst other things, he identifies problems with formalizations of the notion of belief in terms of normal modal logics. He also notes that communicative actions can be ascribed to entities (e.g., corporations) that would normally not be considered capable of mental states. A further problem for any intersubjective account of dialogue based on belief and intention is that in communication agents are not necessarily sincere about their beliefs and intentions. Thus it is difficult to provide an intersubjective report of the beliefs and intentions that are entertained by dialogue participants. Kibble proposes to discard the notions of belief and intention altogether in favour of an account in terms of the normative framework that has been developed over two decades by the philosopher Robert Brandom (Brandom, 1983; Brandom, 1994; Brandom, 2000). Kibble outlines such an account in which *commitment* and *entitlement* are centre stage. Commitment concerns what an agent is compelled to do whereas entitlement pertains to what it is reasonable for an agent to maintain. Whereas an agent can be committed to incompatible propositions, it is not possible to be entitled to such commitments. Using this normative vocabulary, the outline of a characterization of a various dialogue acts (such as

---

[1] For further details on these workshop series, see http://cswww.essex.ac.uk/semdial/ and http://www.sigdial.org/, respectively.

[2] The activity of bringing about such a mutual belief is called *grounding*. An early computational model of grounding can be found in Traum (1994).

assert, endorse and *challenge*) is given.

Pickering and Garrod present a more detailed account of the notion of alignment in dialogue that they have extensively argued for in Pickering and Garrod (2004). They contrast their approach with the information transfer model of language use which emphasizes encoding of meaning by a speaker and subsequent decoding by a hearer. In Pickering and Garrod's model, alignment is not the result of such packing and unpacking of meaning, but the consequence of a primitive priming mechanism. This mechanism is a form of automatic, non-inferential imitation and operates on various levels of representation (e.g., lexical and syntactic). Furthermore, alignment on such lower levels gives rise to alignment on the level of situational representations. During normal processing, there is no appeal to higher level processes that involve negotiation between interlocutors or explicit representation of each other's mental states. Even repair is thought of in mechanistic terms: if an utterance does not fit the current situational model, the first course of action for a hearer is to attempt a reformulation compatible with their current situation model, and to see whether the speaker agrees with it. Solutions that require explicit modeling of other participants' beliefs are to be more costly (in terms of processing effort). Such alternative solutions are normally only used if shallow mechanistic repair strategies have failed. Pickering and Garrod illustrate various aspects of their model using an example from a corpus of maze conversations.

Poesio, Patel and Di Eugenio use a corpus of tutorial dialogues to compare two models of *global* focus in discourse: Grosz and Sidner's *stack model* (Grosz and Sidner, 1986) and Walker's *cache model* (Walker, 1998). The corpus is annotated in terms of Relational Discourse Analysis (RDA; Moser and Moore, 1996). They show how the theoretical constructs of both the stack and cache model can be interpreted in terms of RDA. This then gives them a basis for comparing the two models. The comparison is in terms of *accessibility* of antecedents and *ambiguity* for pronouns and definite descriptions. For pronouns, they find that the accessibility of the antecedent is not affected by varying the model. This resonates with previous work which has also suggested that the global focus has little to no effect on the interpretation of pronouns. In contrast, they do find that both accessibility and ambiguity are affected for definite descriptions: changing the model does mean that an antecedent might be ruled out (accessibility) or that a larger number of non-antecedents is put forward by the model as candidates (ambiguity). The best cache model they tested does better than the best stack model. They suggest that this might, however, be an artefact of the annotation scheme (in particular, the absence of intentional relations across speaker turns). For cache models, it turned out that cache size matters: in particular, the standard size of 7 that was originally proposed does not seem to be the most optimal one.

Purver presents a computational model of clarificatory exchanges in dialogue. A prototype information state-based implementation (CLARIE) is described. It deals with both user and system clarification and follows a classification scheme for *clarification requests* (CRs) that was derived from a 150,000 word corpus of general dialogue extracted from the British National Corpus (BNC) (Burnard,

2000). The proposed model has a number of appealing aspects. Firstly, it deals with a wider range of CR forms than is common in existing formal and computational treatments of CRs. Secondly, CRs and their answers are treated as regular *ask* and *assertion* moves, respectively (rather than, for example, specialized *requests-for-clarifications*). What distinguishes CRs from other ask moves is not the type of dialogue act that is being performed, but rather the fact that this act happens to concern a previous utterance. The paper should be of general interest to anyone interested in building practical dialogue systems, also, because the underlying idea that utterances can be modelled as contextual abstracts that require a grounding process to become fully specified remains largely neutral with respect to the underlying grammar formalism.

Purver, Cann and Kempson describe how their grammar formalism, *Dynamic Syntax* (Kempson et al., 2001), addresses the challenge posed by Pickering and Garrod (2004; this volume) to provide an account for the basic processing mechanisms in dialogue. The formalism is procedural in nature: parsing is at the centre and no parsing-independent grammar is recognized. Rather, the grammar consists of tree update actions that allow for word-by-word incremental construction of parse trees. A parse tree represents the predicate-argument structure of an input sentence. Generation operates on the same tree structures as parsing, but now the input is a goal tree. Generation consists of finding lexical actions –associated with words– that incrementally create a structure which subsumes the goal tree. An important aspect of the approach is the introduction of a *context* relative to which both parsing and generation operate. This context makes it possible to re-use both preceding *tree structure* and *tree update actions*. Such re-use is shown to account for a range of linguistic phenomena including bare answers, shared utterances and ellipsis. Alignment is shown to be a natural consequence of context re-use. Crucially, re-use is shown to reduce processing costs. This provides a further motivation for lexical and syntactic alignment in dialogue in addition to Pickering and Garrod's claim that alignment on lower levels results in alignment on the level of situational representations.

3 Emergent Perspectives on Dialogue

The natural starting point for identifying different perspectives on dialogue is the traditional organization of research into different fields. From this point of view, the papers by Purver, Purver et al. and Poesio et al. can be seen as belonging to Computional Linguistics; Branigan, and Pickering and Garrod, are foremost studies in Psycholinguistics; and Kibble's paper is on the borderline between Analytical Philosophy/Logic and Computational Linguistics.

Such a division into fields of research gives us some indication of the specific angles from which dialogue can be examined. For example, whereas in (computational) linguistics the focus is on specific linguistic constructions, from an AI perspective the main concern is typically shifted towards underlying language independent mechanisms for communication, abstracting away from the connection with the linguistic inventory of specific languages (see Purver, this

volume for further discussion). However, on the whole, the existing divisions are too crude for a systematic inventory of perspectives; they cut across various dimensions and consequently bring with them the danger of ending up with a classification of research into dialogue a bit like the classification of animals in the infamous Chinese encyclopedia alluded to in Borges (1984): "(a) belonging to the emperor, (b) embalmed, (c) tame [...] (k) drawn with a very fine camel hair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies.".

In the following sections, we add to the common division into research fields a number of mainly orthogonal dimensions of classification that should help us to make the relations between the papers in this volume explicit.

3.1   Is versus Ought

A dialogue involves participants, and the behaviour of these participants constitutes the dialogue. We can approach the study of such behaviour from two different angles. From an empirical point of view, one can ask what the behaviour of human dialogue participants *is* like. Alternatively, one can look for the norms that are implicit in this behaviour and shift the emphasis to the question what the behaviour of dialogue participants *ought to* be like. In this special issue, most papers tend to toward the point of view that asks what dialogue behaviour *is* like rather than what it *ought* to be like. Kibble's paper is an exception in that it does employ a normative vocabulary to specify dialogue behaviour. The characterization of dialogue in normative terms, i.e., norms, commitments, obligations, etc., can be traced back to the view of dialogue as a type of game that is governed by rules that the interlocutors are expected to follow.[3] Such rules describe what participants ought to do, but do not necessarily also specify how participants in fact act and even on occasion break the rules. The relation between such a normative approach and the empiricist point view is contentious. For example, the philosopher Robert Brandom proposes that: "Cultural products and activities [including communicative acts in dialogue] become explicit as such only by the use of normative vocabulary that is in principle not reducible to the vocabulary of the natural sciences (though of course the same phenomena under other descriptions are available in that vocabulary)." (Brandom, 2000, p.33).

3.2   What versus How, and Who

Independent of the question whether certain behaviour *is* or *ought* to be produced in dialogue, one can ask the question *how* such behaviour is or can be

---

[3]The treatment of dialogue, and language use in general, as a type of game is usually attributed to Wittgenstein (1953). Formalizations have been proposed by, amongst others, Hamblin (1970), Bunt and Van Katwijk (1979) and Carlson (1983). More recent work on obligations in dialogue by, for example, Matheson et al. (2000) introduces a normative vocabulary into computational models of dialogue.

produced. This move shifts the focus from describing observed or desirable patterns of behaviour to uncovering underlying mechanisms that can generate such behaviour. The question about the *how* allows for two alternative inpretations: one can conceive of it as being focused on how humans produce dialogue behaviour or, alternatively, take a wider perspective and ask how agents (whether human or not) can produce dialogue behaviour.

Of course, the distinction between research into the *what* of behaviour and the *how* on the other side is not absolute; rather one can see them as opposite extremes on a scale. Branigan, and Pickering and Garrod, put somewhat more emphasis on the quest for empirical data on *what* behaviour humans produce, whereas Purver and Purver et al. tend towards addressing specific *how* questions by providing detailed algorithmic specifications. The work of Poesio et al. addresses both sides: the collection of empirical data and the use of these to throw light on different theories of how language is processed. With regards to the question of whose behaviour is being examined, Branigan, Pickering and Garrod, and Purver et al., are primarily interested in the *what* and *how* of human behaviour, whereas Poesio et al., Purver, and Kibble also concern themselves with behaviour by non-human agents (computer programs).

3.3   Pure versus Applied

One can ask the aforementioned questions for their own sake, but also in the context of specific applications. The distinction between pure and applied research cuts across the questions regarding *what*, *how*, *who* and *is* versus *ought*. From the point of view of applications, research into dialogue is relevant to human–human dialogue, human–machine dialogue and machine–machine dialogue. In this volume, the issue of applications for human–human dialogue is not addressed, i.e., the research on dialogue for helping humans communicate better with each other, whether it is with or without the help of technology. The technology-based enhancement of human-human dialogue is the subject of the field of computer-mediated communication (see, e.g., Herring, in press).

A number of papers in this volume are concerned with issues resulting from applications in the area of human–computer dialogue. Purver explicitly motivates his research into clarification requests by pointing out that as dialogue systems "start dealing with more complex tasks and domains, [...] the scope for various misunderstandings increases" and as result there is "an advantage to being able to clarify at the individual word or phrase level, making the user aware of the specific source of the problem".

Interestingly, Purver's work also illustrates the tension that sometimes exist between pure and applied research into dialogue. From the point of view of current applications, Purver's work on addressing clarification requests regarding unknown words is inspired by human behaviour, but difficult to interface with existing speech recognizers which given an input guess which known words it matches, ignoring the possibility that the input is an unknown word.

Although Poesio et al. are primarily concerned with comparing two theoretical accounts of discourse structure and anaphora, they also bring into consider-

6

ation practical differences between these accounts. In particular, they point out that one of the models under investigation (the cache model) is much easier to implement than the other (the stack model), because the former does not depend the recognition of intentions (which are notoriously difficult to recognize).

Finally, Kibble draws attention to the relevance of foundational work on protocols and update rules for dialogue to the semantics of practical languages for inter-agent communication languages (e.g., KQML; Finin et al., 1997).

3.4 Differences in Method

The methods that are used in each of the papers in this volume range from empirical (experimental and corpus-based) to computational and analytical. The work of Branigan, and Pickering and Garrod, is primarily a discussion of empirical work, whereas some aspects of this work are formalized and framed in computational terms in Purver and Purver et al. Poesio et al. use a corpus and automated processing of the corpus to compare two alternative theoretical accounts of global focus. Kibble approaches dialogue through analytical means, examining different foundational assumptions for modelling dialogue.

From the point of view of this journal, the computational approach is of particular importance. We want to emphasize that computational approaches have merits beyond their obvious potential for applications in dialogue systems. One of the pioneers of research on dialogue using computers, Richard Power, has put it as follows:

> "There are two main reasons why computer models are used [in theoretical psychology]. First, they give quick accurate feedback on the coherence, completeness, and detailed consequences of a theory. Theories of cognitive processes are necessarily complicated, and are likely to be too complicated for the unaided human intelligence to cope with. Second, computer models are an appropriate medium of expression for a theorist who believes that the mind is a computational system. Even theorists who don't write programs have found computational concepts useful (see Neisser, 1967)." (Power, 1979, p.109)

4 Trends and Outlook

The papers in this volume suggest a number of trends in current research on dialogue. A common theme in Pickering and Garrod, Purver, and Purver et al., is the emphasis on low level mechanisms that are at play in dialogue and which require no explicit modelling of higher level beliefs and intentions.[4] Kibble questions the use of belief and intentions for modelling dialogue from a

---

[4]Others have suggested that although some notion of belief, intention and desire are required, there is no need for computationally intractable planning to arrive at complex dialogue structure. See, e.g., Beun (2001).

different angle, arguing that Brandomian commitments and entitlement should replace them as a more solid, intersubjective basis. Recently, Robert Brandom's work seems to have gained some influence in the community of researchers concerned with modelling language use, as witnessed by a special issue of Pragmatics and Cognition dedicated to Brandom's Inferentialism (Stekeler-Weithofer, 2005). Proof and model-theoretic commitment-based analyses of dialogue have also been proposed by, for example, Piwek (1998; 2000) to deal with various phenomena in dialogue.

There is also a clear trend away from armchair examples to the use of corpora of naturally occurring dialogue. Furthermore, the availability of large quantities of corpus-based data suggests that there is potential for the automated acquisition of regularities in dialogue (e.g., by means of machine learning). Purver points out that automation is necessary for extending coverage to new domains and languages.

Let us conclude this introduction by highlighting a few directions for further research. Firstly, in the future, empirical research into the *how* of human dialogue behaviour will hopefully get new data from neurophysiological studies. These are currently still limited to isolated production tasks (see Roelofs, 2004 for an overview), but further technological developments should, in the future, make it possible to obtain information on how the brain processes information during dialogue.

The work by Pickering and Garrod on alignment as grounded in imitation strongly suggests that low-level non-verbal signals also play an important role in achieving alignment of situational representations. This requires further research into the integrated use in dialogue of verbal and non-verbal means (see, e.g., Nakano et al., 2003). From the point of view of applications, this is an area which has recently gained interest in the context of the construction of *Embodied Conversational Agents* that is, computer-animated characters that can engage in a dialogue with a user or other computer-animated characters (e.g., Cassell et al., 2000; Prendinger and Ishizuka, 2004). Others have taken up the even more challenging task of constructing robots that can engage in dialogue (e.g., Bos et al., 2003).

Finally, the emphasis on low-level processes in dialogue behaviour leads to models that are more concerned with the central place of emotion and feelings in dialogue behaviour (See Damasio, 1999 regarding the revival of emotions in psychology and André et al, 2004 for work on affective dialogue systems).

References

[1] E. André, L. Dybkjær, W. Minker, and P. Heisterkamp, editors. *Affective Dialogue Systems (ADS 2004)*, volume 3086 of *Lecture Notes in Computer Science*, Berlin, 2004. Springer.

[2] R.J. Beun. On the Generation of Coherent Dialogue: A Computational Approach. *Pragmatics and Cognition*, 9(1):37–68, 2001.

[3] J.L. Borges. The Analytical Language of John Wilkins. In *Other Inquisitions 1937–1952*. University of Texas Press, 1984.

[4] J. Bos, E. Klein, and T. Oka. Meaningful Conversation with a Mobile Robot. In *Proceedings of the tenth conference of the European chapter of the Association for Computational Linguistics*, pages 71–74, 2003.

[5] R. Brandom. Asserting. *Nous*, 17:637–650, 1983.

[6] R. Brandom. *Making it explicit*. Harvard University Press, Cambridge, Massachusetts, 1994.

[7] R. Brandom. *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press, Cambridge, Massachusetts, 2000.

[8] H. Bunt and H. van Katwijk. Dialogue acts as elements of a language game. In *Linguistics in the Netherlands 1977–1979*. Foris, Dordrecht, 1979.

[9] L. Burnard. Reference Guide for the British National Corpus. Technical report, Oxford University Computing Services, 2000.

[10] L. Carlson. *Dialogue Games: An Approach to Discourse Analysis*. D. Reidel, Dordrecht, 1983.

[11] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. *Embodied Conversational Agents*. The MIT Press, Cambridge, MA, 2000.

[12] H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.

[13] H. Clark and E. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1987.

[14] A. Damasio. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Hartcourt Brace, New York, 1999.

[15] T. Finin, Y. Labrou, and J. Mayfield. Kqml as an agent communication language. In J. Bradshaw, editor, *Software Agents*. MIT Press, Cambridge, MA, 1997.

[16] B. Grosz and C. Sidner. Attention, Intention, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[17] C. Hamblin. *Fallacies*. Methuen and Co. Ltd., 1970.

[18] S. Herring, editor. *Computer-mediated Conversation*. Hampton Press, Cresskill, NJ, in press.

[19] R. Kempson, W. Meyer-Viol, and D. Gabbay. *Dynamic Syntax: the Flow of Language Understanding*. Blackwell, Oxford, 2001.

[20] C. Matheson, M Poesio, and D. Traum. Modeling grounding and discourse obligations using update rules. In *Proceedings of the First Annual Meeting of the North American Chapter of the ACL*, pages 1–8, Seattle, April 2000.

[21] M. Moser and J. Moore. Toward a synthesis of two accounts of discourse structure. *Computational Linguistics*, 22(3):409–419, 1996.

[22] Y. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a model of face-to-face grounding. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 03)*, pages 553–561, 2003.

[23] U. Neisser. *Cognitive Psychology*. Appleton-Century-Crofts, New York, 1967.

[24] M. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioural and Brian Sciences*, 27(2):169–226, 2004.

[25] P. Piwek. *Logic, Information and Conversation*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 1998.

[26] P. Piwek. Imperatives, Commitment and Action: Towards a Constraint-based Model. *LDV Forum: GLDV-Journal for Computational Linguistics and Language Technology*, 17(1,2):92–115, 2000.

[27] R. Power. The organisation of purposeful dialogues. *Linguistics*, 17:107–152, 1979.

[28] H. Prendinger and M. Ishizuka, editors. *Life-Like Characters: Tools, Affective Functions, and Applications*. Springer, Berlin, 2004.

[29] A. Roelofs. The seduced speaker: Modeling of cognitive control. In A. Belz, R. Evans, and P. Piwek, editors, *Natural Language Generation*, volume 3123 of *Lecture Notes in Computer Science*, pages 1–10, Berlin, 2004. Springer.

[30] P. Stekeler-Weithofer, editor. *Special Issue on The Pragmatics of Making it Explicit*, volume 13 of *Pragmatics and Cognition*, 2005.

[31] D. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Dept. of Computing, University of Rochester, 1994.

[32] A. Turing. Computing Machinery and Intelligence. *Mind*, LIX(2236):433–460, 1950.

[33] M. Walker. Centering, anaphora resolution, and discourse structure. In M. Walker, A. Joshi, and E. Prince, editors, *Centering in Discourse*, pages 401–435. Oxford University Press, 1998.

[34] L. Wittgenstein. *Philosophical Investigations*. Blackwell, Oxford, 1953.