Chapter 6

Discours and Dialogue

(Following section is taken from Chapter 6 "Discourse and Dialogue")
of the book: "Survey of the state of the art in human language technology")

6.1 Overview

Barbara Grosz

Harvard University, Cambridge, Massachusetts, USA

The problems addressed in discourse research aim to answer two general kinds of questions:

1. What information is contained in extended sequences of utterances that goes beyond the meaning of the individual utterances themselves?

2. How does the context in which an utterance is used affect the meaning of the individual utterances, or parts of them?

Computational work in discourse has focused on two different types of discourse: extended texts and dialogues, both spoken and written. Although there are clear overlaps between these—dialogues contain text-like sequences spoken by a single individual and texts may contain dialogues—the current state of the art leads research to focus on different questions for each. In addition, application opportunities and needs are different. Work on text is of direct relevance to document analysis and retrieval applications, whereas work on dialogue is of import for human-computer interfaces regardless of the modality of interaction. A good sense of the current state of research in text interpretation can be gained from reading the papers on text interpretation published in a recent special issue of *Artificial Intelligence* (hereafter, *AIJ-SI*), (Hobbs, Stickel, et al., 1994; Jacobs & Rau, 1994; Palmer, Passonneau, et al., 1994).

Text and dialogue have, however, two significant commonalities. First, a major result of early work in discourse was the determination that discourses divide into *discourse segments*, much like sentences divide into phrases. Utterances group into segments, with the meaning of a segment encompassing more than the meaning of the individual parts. Different theories vary on the factors they consider central to explaining this segmentation; a review of the alternatives can be found in a previous survey (Grosz, Pollack, et al., 1989) (hereafter, *Discourse Survey*).[1] However, many of the implications for language processing are shared. For example, segment boundaries need to be detected; recent work suggests there are intonational indicators of these boundaries in spoken language (e.g., Grosz & Hirschberg, 1992 and the references cited in this paper) and can be used to improve speech synthesis (e.g., Davis & Hirschberg, 1988).

Second, discourse research on the interpretation of referring expressions, including pronouns and definite descriptions (e.g., *le petit chat, das grüne Buch*), and the event reference aspect of verb phrase interpretation (e.g., the relationship between the buying and arriving events in the sequence *John went to Mary's house; he had bought flowers at her favorite florist's* is also relevant to both text and dialogue. Work on these problems before 1990 is described in *Discourse Survey*.

6.1.1 Beyond Sentence Interpretation

The major lines of research on determining what information a discourse carries, beyond what is literally expressed in the individual sentences the discourse comprises, fall into two categories which, following Hobbs, we will refer to as *informational* and *intentional*. There are currently efforts to combine these two approaches (e.g., Kehler, 1994; Kehler, 1995; Moore & Pollack, 1992); this is an important area of research.

According to the informational approaches, the coherence of discourse follows from semantic relationships between the information conveyed by successive utterances. As a result, the major computational tools used here are inference and abduction on representations of the propositional content of utterances. *Discourse Survey* describes work in this area under *inference-based approaches*; more recent work in this area is presented in *AIJ-SI*.

According to the intentional approaches the coherence of discourse derives from the intentions of speakers and writers, and understanding depends on recognition of those intentions. Thus, these approaches follow Grice (1969); early work in this area drew on speech act theory (Searle, 1969). A major insight of work in this area was to recognize the usefulness of applying AI planning techniques; this work is described in *Discourse Survey*. Recently, various limitations of this approach have been recognized. In particular, as originally argued by Searle (1990) and Grosz and Sidner (1990), models of individual plans are not adequate for understanding discourse; models of collaborative plans or joint intentions are required. A variety of approaches to developing such models are currently underway (Grosz & Kraus, 1993; Sonenberg, Tidhar, et al., 1994; Cohen & Levesque, 1990) and used for dialogue (Lochbaum, 1993; Lochbaum, 1994; Lochbaum, 1995).

6.1.2 Interpretation and Generation in Context

Research in this area also splits into two approaches, those that examine the interaction of choice or interpretation of expression with focus of attention, and those that are coherence-based.

Focus of attention interacts with the interpretation and generation of pronouns and definite descriptions (Grosz & Sidner, 1986). The coherence-based approaches have been taken with the informational approaches described above. The main new issues in this area concern how to combine these approaches, as it is clear that both kinds of consideration play

[1] Many of the papers cited in this survey may be found in the collection *Readings in Natural Language Processing* (Grosz, Sparck Jones, et al., 1986).

roles both in determining which expressions to use and how to interpret expressions in context. The focus-based approaches have been applied cross-linguistically;

because this is a cognitively-oriented approach, it should have application to multi-media interfaces even when natural language is not being used, or when only a restricted subset can be handled.

6.2 Chapter References

ACL (1994). *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Allen, J. F. (1979). A plan-based approach to speech act recognition. Technical Report 131, Department of Computer Science, University of Toronto, Toronto, Canada.

Allen, J. F. and Perrault, C. R. (1980). Analyzing intention in dialogues. *Artificial Intelligence*, 15(3):143–178.

Andry, F. (1992). Static and dynamic predictions: a method to improve speech understanding in cooperative dialogues. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, volume 1, pages 639–642, Banff, Alberta, Canada. University of Alberta.

Andry, F., Bilange, E., Charpentier, F., Choukri, K., Ponamalé, M., and Soudoplatoff, S. (1990). Computerised simulation tools for the design of an oral dialogue system. In *Selected Publications, 1988-1990, SUNDIAL Project (Esprit P2218)*. Commission of the European Communities.

Appelt, D. E. (1985). *Planning English Sentences*. Cambridge University Press.

ARPA (1993). *Proceedings of the 1993 ARPA Spoken Language Systems Technology Workshop*. Advanced Research Projects Agency, MIT.

Austin, J. L. (1962). *How to do things with words*. Oxford University Press, London.

Bilange, E. (1991). A task independent oral dialogue model. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, Berlin. European Chapter of the Association for Computational Linguistics.

Bilange, E., Guyomard, M., and Siroux, J. (1990). Separating dialogue knowledge and task knowledge from oral dialogue management. In *COGNITIVA'90*, Madrid.

Bobrow, D. and PARC Understander Group, T. (1977). GUS-1, a frame driven dialog system. *Artificial Intelligence*, 8(2):155–173.

Bylander, E. (1991). Complexity results for planning. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 274–279, Sydney, Australia.

Carberry, S. (1985). A pragmatics-based approach to understanding intersentential ellipses. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pages 188–197, University of Chicago. Association for Computational Linguistics.

Carberry, S. (1990). *Plan recognition in natural language dialogue*. ACL-MIT Press Series in Natural Language Processing. Bradford Books, MIT Press, Cambridge, Massachusetts.

Cawsey, A. (1992). *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*. ACL-MIT Press.

Chapman, D. (1987). Planning for conjunctive goals. *Artificial Intelligence*, 32(3):333–377.

Clark, H. H. (1989). Contributing to discourse. *Cognitive Science*, 13:259–294.

Clark, H. H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22:1–39. Reprinted in Cohen, Morgan, et al. (1990).

Cohen, P. R. and Levesque, H. J. (1980). Speech acts and the recognition of shared plans. In *Proceedings of the Third Biennial Conference*, pages 263–271, Victoria, British Columbia. Canadian Society for Computational Studies of Intelligence.

Cohen, P. R. and Levesque, H. J. (1990). Rational interaction as the basis for communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press, Cambridge, Massachusetts.

Cohen, P. R. and Levesque, H. J. (1991a). Confirmations and joint action. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 951–957, Sydney, Australia.

Cohen, P. R. and Levesque, H. J. (1991b). Teamwork. *Noûs*, 25(4):487–512. Also Technical Note 504, Artificial Intelligence Center, SRI International, Menlo Park, California, 1991.

Cohen, P. R., Morgan, J., and Pollack, M. E., editors (1990). *Intentions in Communication*. MIT Press, Cambridge, Massachusetts.

Cohen, P. R. and Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212.

Dahlbäck, N. and Jönsson, A. (1992). An empirically based computationally tractable dialogue model. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society (COGSCI-92)*, Bloomington, Indiana.

Dale, R., Hovy, E. H., Rösner, D., and Stock, O., editors (1992). *Aspects of Automated Natural Language Generation*. Number 587 in Lecture Notes in AI. Springer-Verlag, Heidelberg.

Davis, J. R. and Hirschberg, J. (1988). Assigning intonational features in synthesized spoken directions. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 187–193, SUNY, Buffalo, New York. Association for Computational Linguistics.

Delin, J., Hartley, A., Paris, C., Scott, D., and Vander Linden, K. (1994). Expressing procedural relationships in multilingual instructions. In *Proceedings of the Seventh International Workshop on Natural Language Generation*, pages 61–70, Kennebunkport, Maine. Springer-Verlag, Berlin.

Delin, J., Scott, D., and Hartley, A. (1993). Knowledge, intention, rhetoric: Levels of variation in multilingual instructions. In *Proceedings of the Workshop on Intentionality and Structure in Discourse Relations*, Columbus, Ohio. Association for Computational Linguistics.

Edmonds, P. G. (1993). A computational model of collaboration on reference in direction-giving dialogues. Master's thesis, Computer Systems Research Institute, Department of Computer Science, University of Toronto.

Eurospeech (1993). *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, Berlin. European Speech Communication Association.

Fraser, N. and Gilbert, N. (1989). Simulating speech systems. *Computer, Speech, and Language*, 3.

Fraser, N. M. and Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech and Language*, 5(1):81–99.

Gerbino, E. and Danieli, M. (1993). Managing dialogue in a continuous speech understanding system. In *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, volume 3, pages 1661–1664, Berlin. European Speech Communication Association.

Gilbert, N., Wooffitt, R., and Fraser, N. (1990). Organising computer talk. In Luff, P., Gilbert, N., and Frohlich, D., editors, *Computers and Conversation*, chapter 11, pages 235–258. Academic Press, New York.

Green, N. and Carberry, S. (1994). A hybrid reasoning model for indirect answers. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Grice, H. P. (1969). Utterer's meaning and intentions. *Philosophical Review*, 68(2):147–177.

Grosz, B. and Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, volume 1, pages 429–432, Banff, Alberta, Canada. University of Alberta.

Grosz, B. and Kraus, S. (1993). Collaborative plans for group activities. In *Proceedings of IJCAI-93*, volume 1, pages 367–373, Chambery, France.

Grosz, B., Pollack, M., and Sidner, C. (1989). Discourse. In Posner, M., editor, *Foundations of Cognitive Science*. MIT Press.

Grosz, B. J. and Sidner, C. L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Grosz, B. J. and Sidner, C. L. (1990). Plans for discourse. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 417–444. MIT Press, Cambridge, Massachusetts.

Grosz, B. J., Sparck Jones, K., and Webber, B. L., editors (1986). *Readings in Natural Language Processing*. Morgan Kaufmann Publishers, Inc.

Guyomard, M. and Siroux, J. (1988). Experimentation in the specification of an oral dialogue. In Niemann, H., Lang, M., and Sagerer, G., editors, *Recent Advances in Speech Understanding and Dialog Systems*, volume 46. Springer Verlag, Berlin. NATO ASI Series.

Heeman, P. A. and Hirst, G. (1992). Collaborating on referring expressions. Technical Report 435, Department of Computer Science, University of Rochester, Rochester, New York.

Hirschberg, J. and Grosz, B. J. (1992). Intonational features of local and global discourse structure. In *Proceedings of the Fifth DARPA Speech and Natural Language Workshop*. Defense Advanced Research Projects Agency, Morgan Kaufmann.

Hirschman, L. and Pao, C. (1993). The cost of errors in a spoken language system. In *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, volume 2, pages 1419–1422, Berlin. European Speech Communication Association.

Hobbs, J. R. (1985). On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University.

Hobbs, J. R. (1993). Intention, information, and structure in discourse. In *Proceedings of the NATO Advanced Research Workshop on Burning Issues in Discourse*, pages 41–66, Maratea, Italy.

Hobbs, J. R., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142.

Hobbs, J. R., Stickel, M. E., Appelt, D. E., and Martin, P. (1994). Interpretation as abduction. In Pereira, F. C. N. and Grosz, B. J., editors, *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Hovy, E. H. (1990). Parsimonious and profligate approaches to the question of discourse structure relations. In IWNLG, editor, *Proceedings of the Fifth International Workshop on Natural Language Generation*, Pittsburgh, Pennsylvania. Springer-Verlag.

Hovy, E. H. (1991). Approaches to the planning of coherent text. In Paris, C. L., Swartout, W. R., and Mann, W. C., editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer Academic.

ICSLP (1992). *Proceedings of the 1992 International Conference on Spoken Language Processing*, Banff, Alberta, Canada. University of Alberta.

IJCAI (1991). *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Sydney, Australia.

IWNLG (1994). *Proceedings of the Seventh International Workshop on Natural Language Generation*, Kennebunkport, Maine. Springer-Verlag, Berlin.

Jacobs, P. S. and Rau, L. F. (1994). Innovations in text interpretation. In Pereira, F. C. N. and Grosz, B. J., editors, *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Jefferson, G. (1972). Side sequences. In Sudnow, D., editor, *Studies in Social Interaction*. Free Press, New York.

Jennings, N. R. and Mamdani, E. H. (1992). Using joint responsibility to coordinate collaborative problem solving in dynamic environments. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 269–275, Menlo Park, California. American Association for Artificial Intelligence, AAAI Press/MIT Press.

Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J., Janssen, T., and Stokhof, M., editors, *Formal Methods in the Study of Language*. Mathematisch Centrum, Amsterdam.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer, Dordrecht.

Karis, D. and Dobroth, K. M. (1991). Automating services with speech recognition over the public switched telephone network: Human factors considerations. *IEEE Journal of Selected Areas in Communications*, 9(4):574–585.

Kautz, H. (1990). A circumscriptive theory of plan recognition. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press, Cambridge, Massachusetts.

Kehler, A. (1994). Common topics and coherent situations: Interpreting ellipsis in the context of discourse inference. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Kehler, A. (1995). *Interpreting Cohesive Forms in the Context of Discourse Inference*. PhD thesis, Harvard University.

Lascarides, A. and Oberlander, J. (1992). Abducing temporal discourse. In *Proceedings of the Sixth International Workshop on Natural Language Generation*, pages 167–182, Trento, Italy. Springer-Verlag. Also in Dale, Hovy, et al. (1992).

Levinson, S. (1981). Some pre-observations on the modelling of dialogue. *Discourse Processes*, 4(1):93–116.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Litman, D. and Passoneau, R. (1993). Feasibility of automated discourse segmentation. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Ohio State University. Association for Computational Linguistics.

Litman, D. J. and Allen, J. F. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11:163–200.

Litman, D. J. and Allen, J. F. (1990). Discourse processing and commonsense plans. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 365–388. MIT Press, Cambridge, Massachusetts.

Lochbaum, K. E. (1993). A collaborative planning approach to discourse understanding. Technical Report TR-20-93, Harvard University.

Lochbaum, K. E. (1994). *Using Collaborative Plans to Model the Intentional Structure of Discourse*. PhD thesis, Harvard University.

Lochbaum, K. E. (1995). The use of knowledge preconditions in language processing. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence*, Montreal, Canada. In press.

Mann, W. C. and Thompson, S. A. (1987). Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, Information Sciences Institute, University of Southern California.

Mast, M., Kompe, R., Kummert, F., and Niemann, H. (1992). The dialog module of the speech recognition and dialog system EVAR. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, volume 2, pages 1573–1576, Banff, Alberta, Canada. University of Alberta.

McKeown, K. R. (1985). *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Studies in Natural Language Processing. Cambridge University Press.

Moore, J. D. and Paris, C. L. (1993). Planning texts for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics*, 19(4):651–694.

Moore, J. D. and Pollack, M. E. (1992). A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544.

Palmer, M. S., Passonneau, R. J., Weir, C., and Finin, T. (1994). The kernel text understanding system. In Pereira, F. C. N. and Grosz, B. J., editors, *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Paris, C. and Scott, D. (1994). Stylistic variation in multilingual instructions. In *Proceedings of the Seventh International Workshop on Natural Language Generation*, pages 45–52, Kennebunkport, Maine. Springer-Verlag, Berlin.

Peckham, J. (1993). A new generation of spoken language systems: recent results and lessons from the SUNDIAL project. In *Eurospeech '93, Proceedings of the Third European Conference on Speech Communication and Technology*, volume 1, pages 33–42, Berlin. European Speech Communication Association. Keynote address.

Pereira, F. C. N. and Grosz, B. J., editors (1994). *Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

Perrault, C. R. and Allen, J. F. (1980). A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3):167–182.

Polanyi, R. and Scha, R. (1984). A syntactic approach to discourse semantics. In *Proceedings of the 10th International Conference on Computational Linguistics*, pages 413–419, Stanford University, California. ACL.

Reichman, R. (1981). *Plain-speaking: A theory and grammar of spontaneous discourse*. PhD thesis, Department of Computer Science, Harvard University, Cambridge, Massachusetts.

Sacks, H., Schegloff, E., and Jefferson, G. (1978). A simplest systematics for the organization of turn-taking in conversation. In Schenkein, J., editor, *Studies in the Organization of Conversational Interaction*. Academic Press, New York.

Sadek, D. (1991). Dialogue acts are rational plans. In *Proceedings of the ESCA/ETRW Workshop on the Structure of Multimodal Dialogue*, Maratea, Italy.

Schegloff, E. A. (1981). Discourse as an interactional achievement: Some uses of unh-huh and other things that come between sentences. In Tannen, D., editor, *Analyzing discourse: Text and talk*. Georgetown University Roundtable on Languages and Linguistics, Georgetown University Press, Washington, DC.

Searle, J. R. (1969). *Speech Acts: An essay in the philosophy of language*. Cambridge University Press.

Searle, J. R. (1976). The classification of illocutionary acts. *Language in Society*, 5.

Searle, J. R. (1990). Collective intentionality. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press, Cambridge, Massachusetts.

Seneff, S., Hirschman, L., and Zue, V. W. (1991). Interactive problem solving and dialogue in the ATIS domain. In *Proceedings of the Fourth DARPA Speech and Natural Language Workshop*, Pacific Grove, California. Defense Advanced Research Projects Agency, Morgan Kaufmann.

Shneiderman, B. (1980). Natural vs. precise concise languages for human operation of computers: Research issues and experimental approaches. In *Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics*, pages 139–141, Philadelphia, Pennsylvania. Association for Computational Linguistics.

Sidner, C. and Israel, D. (1981). Recognizing intended meaning and speaker's plans. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 203–208, Vancouver, British Columbia.

Sidner, C. L. (1985). Plan parsing for intended response recognition in discourse. *Computational Intelligence*, 1(1):1–10.

Sinclair, J. M. and Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press, London.

Siroux, J. (1989). Pragmatics in a realization of a dialogue module. In Taylor, M. M., Néel, F., and Bouwhuis, D. G., editors, *The structure of multimodal dialogue*. Elsevier Science, Amsterdam.

Sonenberg, E., Tidhar, G., Werner, E., Kinny, D., Ljungberg, M., and Rao, A. (1994). Planned team activity. Technical Report 26, Australian Artificial Intelligence Institute.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human/machine communication*. Cambridge University Press.

Traum, D. R. and Allen, J. F. (1994). Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico. Association for Computational Linguistics.

Wahlster, W. (1993). Verbmobil, translation of face-to-face dialogs. In *Proceedings of the Fourth Machine Translation Summit*, pages 127–135, Kobe, Japan.

Wahlster, W., André, E., Finkler, W., Profitlich, H.-J., and Rist, T. (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, pages 387–427.

Yamaoka, T. and Iida, H. (1991). Dialogue interpretation model and its application to next utterance prediction for spoken language processing. In *Eurospeech '91, Proceedings of the Second European Conference on Speech Communication and Technology*, pages 849–852, Genova, Italy. European Speech Communication Association.

Young, S. J. and Proctor, C. E. (1989). The design and implementation of dialogue control in voice operated database inquiry systems. *Computer, Speech, and Language*, 3.

Young, S. R., Hauptmann, A. G., Ward, W. H., Smith, E. T., and Werner, P. (1989). High level knowledge sources in usable speech recognition systems. *Communications of the ACM*, 32(2).