WHAT IS A TEXT?

2.1 TOWARDS A DEFINITION[1]

'What is a text?' This question has occupied text linguistics and discourse analysis since their beginnings,[2] although the two different scientific approaches are currently converging more and more. Both *text* and *discourse* are restricted, in everyday parlance, to written (texts) on the one hand and spoken (discourse) on the other, although this is contested to some extent in the scholarly literature (see Brünner & Graefen 1994a). Apart from this, texts are often considered to be longer pieces of writing. The word evokes the idea of a book, a letter, or a newspaper. The decisive contribution of linguistics in this respect has been to introduce, in the face of such popular opinions, a concept of text that is very broadly and generally accepted and which includes every type of communicative utterance (see below). Clear criteria ultimately decide whether or not something can be viewed as text or discourse (Fairclough 1992a: 3ff.). These criteria are linguistic in nature and relate, most of all, to the syntactic and semantic relations within a text (see sections 2.2 and 2.4 below). A text may equally be an inscription on a tombstone, a form, part of a conversation, or a newspaper article. On the one hand this indicates a very broad concept of communication that regards language and speech as forms of action and derives from Wittgenstein's 'language game' (Wittgenstein 1984: 250, Wodak 1996), while on the other hand it suggests a notion of 'sign', as used in modern semiotics (Kress 1993).[3] The concept of 'semiosis' (meaning-making) relates to any sign (including, for instance, a traffic sign) that according to social conventions is meaningful (Halliday 1978).[4]

In the final analysis, therefore, the answer to our opening question, 'what is a text?', is theory-dependent (see presentation of the individual methods in Part 2) and, as Gruber (1996: 31) claims, in view of its dependency on a particular context and situation, cannot be entirely unambiguous. Let us consider the following poems by Paul Celan (*Einmal*) and Christian Morgenstern (*Fisches Nachtgesang*):
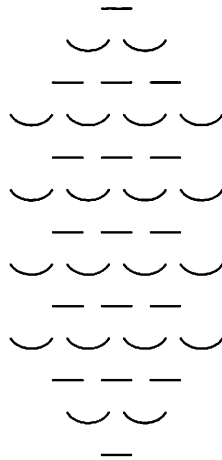
> Einmal,
> Da hörte ich ihn,
> Da wusch er die Welt,
> Ungesehn, nachtlang, wirklich.
> Eins und Unendlich,

Methods of Text and Discourse Analysis

Stefan Titscher, Michael Meyer,
Ruth Wodak and Eva Vetter

Vernichtet,
Ichten.
Licht war. Rettung

Paul Celan (1982) *Atemwende*, Frankfurt: Suhrkamp, 103

```
              —
           ˘     ˘
        —    —    —
     ˘    ˘    ˘    ˘
        —    —    —
     ˘    ˘    ˘    ˘
        —    —    —
     ˘    ˘    ˘    ˘
        —    —    —
     ˘    ˘    ˘    ˘
        —    —    —
           ˘     ˘
              —
```

Christian Morgenstern (1975) *Galgenlieder*, Frankfurt: DTV, 27

It is only when we assume that we are dealing with poems that these are acceptable as texts; otherwise one might suppose that we are dealing with a typing exercise, in the case of Morgenstern, or, in the Celan example, with a random sequence of word-fragments.

One of the most widespread definitions of 'text' comes from de Beaugrande & Dressler (1981: 1ff.) They define a text as a 'communicative event' that must satisfy several conditions, namely the seven text criteria that we consider in detail below. According to this definition a traffic sign, a newspaper article, an argument and a novel are all texts that correspond to the differing rules of particular genres or text types. All the genres mentioned have particular linguistic features, fulfil particular functions and are bound to specific production and reception situations (Swales 1991). There are, therefore, text-internal as well as text-external conditions of meaning, which leads us ultimately to the difficult question of how the extra-linguistic context may be defined and analysed (see below). The examples of the poems demonstrate that we cannot understand these two texts without clear contextual expectations. We bring different expectations, for example, to a political speech than to a news broadcast. The former follows, in part, classical patterns of rhetorical structure (Wodak et al. 1998), while the latter has equally precise conditions of content and form (Lutz & Wodak 1987). These implicit expectations are particularly evident when rules of genre are infringed – as, for example, when a news text is narratively formulated in dialect. Experimental subjects do not accept such texts as official information. They have more trust in news texts that 'sound official', even if they find them perhaps unintelligible (Wodak 1996: 110ff.).

2.2 THE SEVEN TEXT CRITERIA OF DE BEAUGRANDE AND DRESSLER

Here we shall first list the criteria that are invoked by Robert de Beaugrande and Wolfgang Dressler (1981) for the definition of texts. This taxonomy is widely adopted and accepted, and for this reason we wish to use it as the basis for our explanations. Later we shall examine each of the criteria in turn.

Cohesion concerns the components of the textual surface, that is the 'text-syntactic' connectedness. The linear sequence of linguistic elements in a text is in no way accidental, but obeys grammatical rules and dependencies. All the functions that are applied to create relationships between surface elements are categorized as cohesion. In what follows we shall present some of the means by which cohesion is achieved:

- *Recurrence* By means of the repetition of lexical elements, sentence components and other linguistic elements, text structures are formed.

- *Anaphora and Cataphora* Anaphora directs attention to what has previously been said or read (for example, through the use of pro-forms), cataphora points to what is to come through the use of deictic elements.

- *Ellipsis* This element of structure is normally unintelligible without the communicative situation and the shared world knowledge (presuppositions) of participants in a conversation. Textual abbreviations therefore depend particularly on elements of talk-constellations (the reliance on rhetorical devices within text linguistics is no accident since, apart from stylistics, from an historical point of view, rhetoric is one of the most important sources for a supra-sentential grammar).

- *Conjunctions* These signal relations or connections between events and situations. There are conjunctions (linking sentence structures of the same status), disjunctions (linking sentence structures with differing status), contra-junctions (linking sentence structures of the same status that seem to be irreconcilable, such as cause and unexpected effect) and subordinations (used where one sentence structure is dependent on another).

Coherence (or textual semantics) constitutes the meaning of a text. This often refers to elements that do not necessarily also require a linguistic realization. For example, some types of research assume cognitive structures in recipients that are actualized through a text and help to determine interpretations. Similarly, under certain circumstances elements of knowledge that are not expressed in a text may also be implied and may likewise influence reception (see Grice's concept of 'implicature' 1975, Kotthoff 1996, Lutz & Wodak 1987, Kintsch & van Dijk 1983). De Beaugrande & Dressler (1981) suggest that certain 'concepts' (meanings) are bound through 'relationship' and then realized in the textual surface.

For example, causality is a relationship: this affects the manner in which an

event or situation may influence other events or situations; in 'Jack fell down and broke his crown' – *fall* is the cause of the event *break*. A text creates no sense in itself but only in connection with knowledge of the world and of the text. This implies that in the process of language acquisition certain ways of structuring both reality and texts also have to be acquired.[5] This already points to the central concept of intertextuality: every text relates both synchronically and diachronically to other texts, and this is the only way it achieves meaning.

Intentionality relates to the attitude and purpose of text producers. What do they want and intend with the text? Accordingly, talking in one's sleep would not count as a text, whereas a telephone directory would.

Acceptability is the mirror of intentionality. A text must be recognized as such by recipients in a particular situation. This criterion is of course related to conventionality, and does not mean that recipients can simply reject a text 'maliciously'. Acceptability therefore concerns the degree to which hearers and readers are prepared to expect a text that is useful or relevant. Here there can be enormous communicative conflicts. Either the text is not acceptable (unintelligible, incoherent, fragmentary, and so on), or hearers may question its acceptability even though the intentionality is clearly expressed. For example, in some narratives a listener may question a tiny detail that is totally irrelevant to the particular conversation.

Informativity refers to the quantity of new or expected information in a text. This addresses simultaneously not only the quantity but also the quality of what is offered: how is the new material structured and using what cohesive means?[6]

Situationality means that the talk-constellation and speech situation play an important role in text production (Wodak et al. 1989: 120). Only particular varieties or types of text, speech styles or forms of address are both situationally and culturally appropriate. This criterion already leads to the concept of 'discourse, since discourse is very generally defined as "text in context"' (see below).

Intertextuality has two types of meaning. On the one hand it suggests that a text always relates to preceding or simultaneously occurring discourse, and on the other hand it also implies that there are formal criteria that link texts to each other in particular genres or text varieties. In the terminology of text-planning such genres would be described as 'schemas' or 'frames' (Wodak 1986):

- Narrative text varieties (tales, stories, etc.) rely on temporal ordering principles.

- Argumentative text varieties (explanations, scientific articles, etc.) use contrastive devices.

- Descriptive text varieties employ predominantly local (that is, spatial or temporal) elements (as in descriptions, portrayals, etc.).

- Instructive text varieties (such as textbooks) are both argumentative and enumerative.

An additional important feature of all definitions of text is expressed in the seven text criteria: the first two criteria (cohesion and coherence) might be defined as text-internal, whereas the remaining criteria are text-external. In this way a first distinction may be made between traditional text linguistics and discourse analysis. In those approaches which are purely 'text linguistic' in orientation the investigation and modelling of cohesion and coherence are predominant, and all the text-external factors, in the sense of intervening variables, are in the background. In discourse analysis, however, it is precisely these external factors that play an essential role, and texts (that is, cohesion and coherence phenomena) are viewed as a manifestation and result of particular combinations of factors.[7] Modern approaches mostly emphasize the functional aspect (Renkema 1993, Dressler 1989).

2.3 LINGUISTIC TEXT ANALYSIS

Unlike de Beaugrande & Dressler (1981), we believe that these criteria concern different textual dimensions and should therefore not be considered to be of equal importance. We propose that cohesion and coherence should be characterized as constitutive of texts, that is to say, every text (or discourse) must satisfy these two criteria, independently of cotext and context (see below). In this sense *intentionality, informativity, acceptability* and *situationality* are independent of context. We proceed from a complex model of communication that is interactive and dialogic in character, rather than from the sender–hearer type of model used in traditional communication theory (Shannon & Weaver 1949). *Intertextuality* is directly related to this type of complex communication model in its assumption that every text is embedded in a context and is synchronically and diachronically related to many other texts (see Chapter 11). In this way one can, in one's own analysis, take the further step to a critical text or discourse theory: we are not satisfied merely to approach texts according to the linguistic dimensions of cohesion and coherence. On the contrary, we wish to include from the outset the differential relationship to the social, political or other context and to make this the basis of our interpretation.

A linguistic text analysis is therefore defined by its focus on cohesion and coherence, unlike other (sociological) methods of text analysis that select only a few instances of one of these two dimensions. Classical content analysis, therefore, restricts itself to the level of the lexicon (that is, to one dimension of semantics, see Chapter 5); distinction theory text analysis looks for counter-concepts. The focus is therefore on the semantic level. Syntax is used merely to support the selection of units of analysis (sentence **constituents**, see Chapter 13). A linguistic text analysis incorporates syntactic, semantic and pragmatic levels. Most of the sociological methods, on the other hand, are content with only one of these semiotic categories. The particular linguistic theory of grammar that is invoked to connect and specify individual instances of cohesion and coherence is entirely open and theory-dependent (see, for example, the account of

functional pragmatics in Chapter 12, as opposed to Fairclough, in Chapter 11, 11.4.1). In what follows, after outlining the concepts of discourse and content, we shall point to a number of developments since de Beaugrande & Dressler (1981), although in the context of this introductory section we shall not be able to present a complete overview.

2.4 SOME REFLECTIONS ON THE TERM 'DISCOURSE'

> Lastly, instead of gradually reducing the rather fluctuating meaning of the word 'discourse', I believe that I have in fact added to its meanings: treating it sometimes as the general domain of all statements, sometimes as an individualizable group of statements, and sometimes as a regulated practice that accounts for a certain number of statements; and have I not allowed this same word 'discourse', which should have served as a boundary around the term 'statement', to vary as I shifted my analysis or its point of application, as the statement itself faded from view? (Foucault 1972: 80)

Foucault is not the only one to have been confronted by the many meanings of 'discourse'. The notion of discourse, in both the popular and the philosophical use of the term, integrates a whole palette of different meanings that often seem to be contradictory or mutually exclusive. In this chapter we cannot embark on this multi-layered discussion. Instead we can only highlight a number of central modes of use of the term that are also found in the methods we shall discuss (see in Chapters 8(4.2), 11 and 12; see also van Dijk 1985a, 1985b, 1985c, 1990a, 1990b, Schiffrin 1994, Renkema 1993, Vass 1992). First of all we give a brief etymological sketch of the concept.

'Discourse' may be derived etymologically from the Latin *discurrere* (to run to and fro) or from the nominalization *discursus* ('running apart' in the transferred sense of 'indulging in something', or 'giving information about something') (Vass 1992: 7). Medieval Latin *discursus*, in addition to conversation, animated debate, and talkativeness, then also meant *orbit* and *traffic* (Vogt 1987b: 16). Thomas Aquinas (1225 or 1227–1274) was the first to use the term in philosophy. For him it meant something like intellectual reasoning. *Discursive*, 'by reasoning', is contrasted with *simplici intuitu*, 'by means of simple intuition'. Discursive means recognition through concepts and thinking in concepts (Eisler 1927: 286). This bi-polarity is also found in Hobbes, Leibniz and Kant who thought that human thought was quite generally discursive (Kant 1974: 109). Maas (1988) goes on to show that in all Western European languages the popular meaning has developed to refer to 'learned discussion' and then to 'dialogue'. Vass (1992: 9) enumerates the following meanings of 'discourse':

1 (general): speech, conversation, discussion;
2 discursive presentation of a train of thought by means of a series of statements;

3 series of statements or utterances, chain of statements;

4 form of a chain of statements/expressions; the manner in which they came about (archaeology): scientific, poetic, religious discourse;

5 rule-governed behaviour that leads to a chain or similarly interrelated system of statements (=forms of knowledge) (medicine, psychology, etc.) (for instance in the work of Michel Foucault);

6 language as something practised; spoken language (e.g. in the work of Paul Ricoeur);

7 language as a totality, the linguistic universe;

8 discussion and questioning of validity criteria with the aim of producing consensus among discourse participants (e.g. in the work of Jürgen Habermas).

Fairclough (1992a: 3ff.) lists several uses of the term, particularly as they occur in modern discourse analysis: 'samples of spoken dialogue, in contrast with written texts', 'spoken and written language', 'situational context of language use', 'interaction between reader/writer and text', 'notion of genre (for example, newspaper discourse)'. Often such different meanings are used without thinking, and it is frequently unclear whether they refer to a small sequence of text or a complete textual variety, or whether they embrace a very abstract phenomenon. We shall therefore attempt to define the concepts of discourse and discourse analysis very precisely in our own exposition.

First, we shall proceed from van Dijk's (1977) definition that sees discourse quite generally as text in context, and as evidence to be described empirically. Van Dijk points to one decisive aspect, namely that discourse should also be understood as action (see above). Apart from this, its self-contained nature and the act of communication are of central importance. This already leads to a fundamentally more difficult and complex question: how can we set about defining a discourse unit? Where does it begin and end? Is there any relationship between method and unit of investigation? Let us observe that, because of intertextuality, there can in principle be no objective beginning and no clear end, since every discourse is bound up with many others and can only be understood on the basis of others. The determination of the unit of investigation therefore depends on a subjective decision of the investigator and on the research questions that govern the investigation (Kress 1993).

In what follows we wish to emphasize above all else the action aspect. We therefore propose the following definition as a basis for subsequent development (Fairclough & Wodak 1997):

> Critical Discourse Analysis sees discourse – language in use in speech and writing – as a form of 'social practice'. Describing discourse as social practice implies a dialectical relationship between a particular discursive event and situation(s), institution(s) and social structure(s) which frame it: the discursive event is shaped by them, but it also shapes them. That is, discourse is socially constituted, as well as socially conditioned – it constitutes situations, objects of knowledge, and the social identities of and relationships between people and groups of people. It is constitutive both in the sense that it helps sustain and reproduce the social status quo, and in the sense that it contributes to transforming it (Wodak 1996: 15).

The idea of discourse as constitutive of reality is emphasized here. In addition it emerges clearly that questions of power and ideology are closely related to discourse: 'Since discourse is so socially consequential, it gives rise to important issues of power. Discursive practices may have major ideological effects – that is, they can help produce and reproduce unequal power relations . . . through the ways in which they represent things and position people' (ibid.). This introduces the notion of discourse used in critical discourse analysis (CDA) (see Chapter 11).

2.5 CONTEXT

Discourses occur, on the one hand, in macro-contexts, in organizations and institutions ('medical discourse' [Foucault 1993]), but on the other hand they occur at a particular time, in a particular place, with particular participants, and so on (that is, micro-context) (Wodak 1996). Therefore the complete individual discourse must be seen in the macro-context in order to capture the specific meaning of a particular textual or discourse sequence (Lalouschek et al. 1990). In addition there are patterns specific to particular textual varieties, since a political speech, for instance, follows rules that are different from those of a TV talk show or a biographical interview. These genre-specific elements must also be taken into consideration (Wodak et al. 1994: 36ff.). Socioculturally acquired values and norms, as well as psychic predispositions, are in a changing relationship with the process-governed, social production of discourse and can or must be included in the analysis. The inclusion of these factors that influence text production and comprehension goes beyond the limits of traditional investigations that are based on the analysis of content (Matouschek & Wodak 1995/6: 46ff.).[8]

Aaron Cicourel (1992: 295) distinguishes between two types of context: a *broad* and a *local* context. In this way he attempts to bring together the otherwise mutually exclusive starting-points of the ethnographic approach and conversation analysis, where the latter believes that the context is constantly being recreated exclusively through the discourse (Drew & Heritage 1992: 16ff., see Chapter 8). Conversation analysis is closely related to Gumperz's idea of 'contextualization cues' (Gumperz 1982: 162): 'any aspect of linguistic behavior – lexical, prosodic, phonological, and syntactic choices together with the use of particular codes, dialects or styles – may function as such, indicating those aspects of context which are relevant in interpreting what a speaker means'. There is some affinity between this concept and Goffman's 'frame' concept (Goffman 1974, 1981). In this sense the idea of 'frame' means the respective definitions of situation and action that individuals attribute to their communicative actions. This incorporates the subjective experience of individuals. We relate Cicourel's notion of the 'broad context' to the macro-level, and the 'local context' to the micro-level. The goal of sociological investigations is to bring together these two dimensions in all their complexity.

In specific analyses one can pursue the so-called 'discourse-sociolinguistic approach' (see Chapter 11). One the one hand a considerable amount of information is acquired through an ethnographic perspective, and on the other hand the discourse marks particular cases where the context is relevant. There remains a final problem, however: how can one decide how much contextual knowledge is necessary? Where does a context begin and end? This question becomes particularly acute in the analysis of allusions, where it is vital to include the world of the discourse and inttterextuality as factors. Here we would associate ourselves with Cicourel, who says:

> A nagging issue undoubtedly remains for many readers is the familiar one that an infinite regress can occur whereby the observer presumably must describe 'everything' about a context. Such a demand is of course impossible to satisfy because no one could claim to have specified all of the local and larger sociocultural aspects of a context. Observers or analysts, like participants in speech events, must continually face practical circumstances that are an integral part of research of everyday living. (1992: 309)

Finally, the aspects of context that are to be included and excluded must be precisely argued and justified within the concrete analysis of a particular case. And these decisions should take into consideration the theoretical questions posed by the analysis.

2.6 FURTHER DISCUSSIONS ON THE CONCEPTS OF TEXT AND DISCOURSE

De Beaugrande & Dressler (1981) maintain that in the concrete case of a specific text all seven criteria must always apply (see above) if we are to be able to speak of a 'text'. This raises a number of problems because – as Renkema (1993) observes – criteria 3, 4 and 5 (that is, intentionality, acceptability, and informativity) are subjective and dependent on the particular observers. The sequence 'Shakespeare wrote more than 20 plays. Will you have dinner with me tonight?' (Renkema 1993: 36) is probably a 'non-text' for most recipients but might, in a particular situation, be totally acceptable (for instance, if someone has wagered a dinner on the basis that he or she knows roughly how many plays Shakespeare wrote). A consequence of this is that the criteria of de Beaugrande & Dressler (1981) do not admit of predictable and objectivizable distinctions between texts that can be determined in advance. In addition any utterance could ultimately be judged to be a text in a particular context.

Here expectations about genre must also be considered. Renkema (1993) does not deal with these but in Fairclough (1995a), for example, they do indeed play a role as 'orders of discourse'. A particular sequence of signs may therefore be entirely acceptable as a poem even if default expectations (that is, common sense) are infringed.

Van Dijk (1980: 41) sets up a definition of text similar to that of de

Beaugrande & Dressler (1981) which proceeds from the idea that 'only those sequences of sentences which have a macro-structure . . . can be designated as texts'. In simple terms, van Dijk understands 'macro-structure' as an underlying thematic and propositional framework that enables the text to hang together. Van Dijk goes on to introduce many limiting factors (basically similar to those of de Beaugrande & Dressler) which again permit any sequence of sentences, in particular cases, to be characterized as a text.

These problems are avoided in Halliday's purely functional definition of text (Halliday 1978). This assumes that social interaction should be seen as a central unit of investigation. In this way Halliday comes close to the 'critical text analysis' explained above. His approach is linguistically characterized by texts, by the speech situation and the linguistic system – that is to say, by factors that are inseparably bound up with one another. Therefore there can be no 'non-texts' in the sense of Beaugrande & Dressler. Halliday stresses 'the essential indeterminacy of the concept of a "text"'(1978: 136). A text, for him, is everything that is meaningful in a particular situation: 'By text, then, we understand a continuous process of semantic choice' (1978: 137) (see the concept of *semiosis*).

A further important sub-set of 'pure' text-linguistic approaches are the cognitive theories of text that have developed from (and partly as a contradiction of) the model of Kintsch & van Dijk (1983) since the late 1970s. In these, texts are viewed as more or less explicit epi-phenomena of cognitive processes (for example, the pursuance of the principles of causality). Context plays a subordinate role. Text analyses based on these cognitive theories of text also operate in a strictly experimental manner. Many of the models (with the exception of that of Kintsch & van Dijk 1983) can only be used for a very narrow set of specific textual varieties, namely reports and stories.
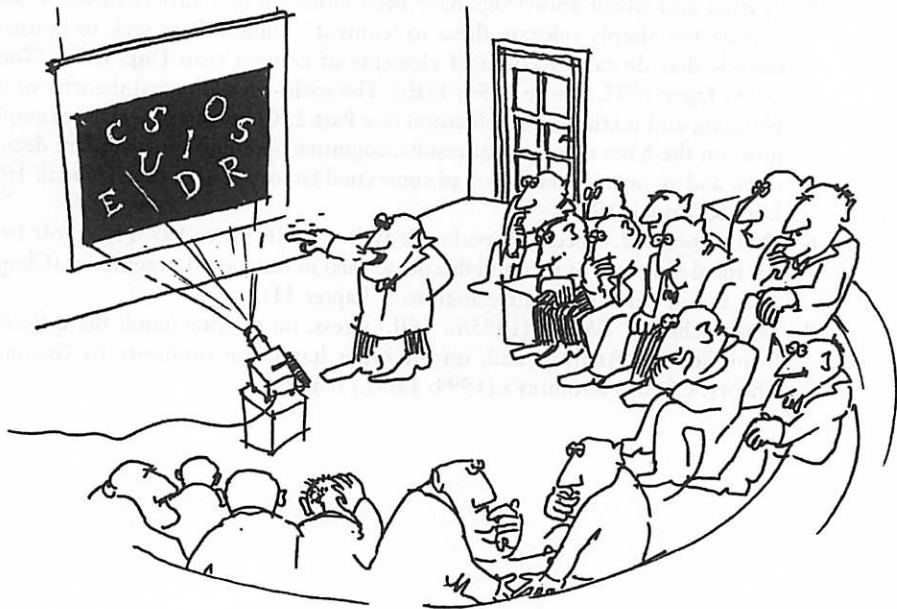
1 We are grateful to Helmut Gruber for some important suggestions in this chapter.

2 Cf. de Beaugrande (1996), van Dijk (1985a, 1985b, 1990a, 1990b), Wodak et al. (1989: 115ff.), Renkema (1993: 36ff.), Brünner & Graefen (1994a: 2), Wodak (1996: 12ff.), Jäger (1993: 138ff.), Shi-xu (1996: 12ff.). Text linguistics and discourse analysis have, in the course of their development, pursued very different goals: text linguistics concerned itself with the isolated text, while discourse analysis has dealt with the text in a context. Latterly, however, the two disciplines have converged more and more, and a clear distinction is often missing. However, many of the authors listed continue to make a clear distinction between text and discourse, as we shall show. This is particularly marked when the notion of discourse is related to the theoretical basis of Michel Foucault (Wodak 1996: 24ff., Jäger 1993: 172ff., Pennycook 1994, Fairclough 1992a: 56ff.) and implies a social construction. In empirical analysis, on the other hand, the two concepts are often used synonymously (cf. Vass 1992).

3 Jäger (1993) refers to Leontjew's speech activity theory that emphasizes a dimension of knowledge in addition to the action aspect. Ultimately the Wittgenstein tradition, that led to critical theory and to Habermas's notion of communication (cf. Wodak 1996: 28ff., Vass 1992), and the Soviet approach, come from different sources: the former from philosophy, the latter from psychology. Both stress the fundamental function of the (speech-)action and speaking as an activity or way of life.

4 Lemke (1995) and Kress & van Leeuwen (1996) give an excellent summary of the different approaches in modern semiotics that derive ultimately from C. Morris and Ch. S. Peirce. In the current context we shall have to dispense with further elaboration.

5 Textual and world knowledge have been modelled in a variety of ways. Some approaches simply relegate these to 'context', while others seek to construct models that do take account of elements of context (van Dijk 1977, Wodak 1996, Jäger 1993, Shi-xu 1996: 17ff.). The socio-psychological theories of text planning and textual comprehension (see Part 2, Chapter 11) seek to exemplify how, on the basis of empirical results, cognitive planning processes are decided upon and an operationalization of contextual factors is attempted (Wodak 1986, Lutz & Wodak 1987).

6 On this point cf. especially Sperber & Wilson 1986, Grice 1975, Kotthoff 1996.

7 Cf. the definitions of text and discourse used in functional pragmatics (Chapter 12), and in critical discourse analysis (Chapter 11).

8 Matouschek & Wodak (1995/6: 46ff.) stress, on the one hand, the difference from Content Analysis and, on the other hand, the similarity to Grounded Theory. Cf. also Kromrey's (1994: 170ff.) remarks.