Formal Method in Discourse Analysis

John Frow

Developing a workable method of discourse analysis matters, presumably, and if it does, because it would allow us to demonstrate the play of power in language, the play of ideology. To demonstrate means being able to substantiate arguments in the public arena – to politicians, in courts of law, in hospitals or prisons or museums. What we envisage is something like the possibility of proof. We envisage, not a political process of judgement, but procedures of demonstration which could be mechanically repeated to produce identical results, and so bear witness to the underlying structure which causes them. Discourse analysis, like other uncertain knowledges, aspires to a state of hardness.

My tone of irony covers a real ambivalence about this project, these projects, and certainly I intend no easy dismissal of them. Here I want to raise a series of questions about some of the work being done in and around the Dutch *Konteksten* series, especially by Tony Hak and Brian Torode;¹ and to see whether it might be possible to define some of the limits touched by this raw, exploratory, interesting work.

This is not, even in the case of individual writers, a unified body of research, but it has certain common points of reference. One of them is the ethnomethodological conversation analysis of Sacks and Schegloff; the other, more properly linguistic, is the work of Michel Pêcheux, and to a certain extent the analytic methods of Zellig Harris. The attraction of Pêcheux's work is, I think, that it promises a rigorous and quasi-grammatical account of presupposition: that is, the concept of the preconstructed promises direct access to ideology insofar as it appears to be equivalent to the Gramscian categories of 'common sense' – the obvious, the unthought, the taken-for-granted; and insofar as it both embodies these in a particular grammatical construction and relates them to a particular process of subject formation.

The development of Pêcheux's argument in *Les vérités de la Palice* goes something like this: distinguishing between two types of sentence information – preconstructed information, which is intertextually established in relation to another, absent or virtual, discourse; and new information constructed in the chain of the utterance – he posits an asymmetrical split between two domains of thought, especially in embedded phrases, and in such a way 'that the subject encounters one of these domains as the unthought of his thought, necessarily preexisting it.'²

The privileged example of this split is the relative clause, which is further divided into explicative and determinative relatives, each producing different effects of preconstruction. Determinative relatives (**Paul, who came to lunch*), specifying a particular feature of a known entity, articulate a particular knowledge; explicatives (**the man who came to lunch*) identify what is already known about an otherwise empty category, and thus *preconstruct* a knowledge, project it as enunciated elsewhere; they intervene, through a relation of implication, to sustain the thought contained in another proposition, and thereby constitute 'a *lateral reminder* of what we already know.'³

Explicatives are exemplary of the workings of ideology in that their 'empty' grammatical form ('the man who,' 'whoever') functions, through a discursive binding, as the site of a process of production of the subject form: juridical discourse, for example, constructs legal subjects as such through the appeal of an 'empty' future anterior ('whoever shall have ...').⁴ This is to say that the constitutive splitting of the subject takes place in this gap between disparate modes of information within the process of the signifier, where the subject finds itself where it always already was. The grammatical account is fitted to the Lacanian and Althusserian accounts of the formation of the subject's imaginary self-identity through and by the signifying chain.

My own use of Pêcheux has drawn largely on his theorization of discourse as a systemic process (this overturns the whole basis of the *langue/parole* opposition) and on his account of the discursive constitution of meaning; I am less convinced by his strict Althusserian conception of ideology and its formulation of the subject-effect (there is a good critique of this by Colin MacCabe).⁵ By contrast, Hak and Torode seem to value most highly the opportunity Pêcheux opens for the development of a methodology which is at once empirical and capable of extensive formalization. Unlike, for example, Schutz's concept of finite 'provinces of meaning,' which 'refers to self-contained domains of consciousness' and for which Schutz 'suggests no way in which these could be empirically investigated,' Pêcheux's develops a model in which 'a discourse which is *preconstructed* in one instance will be *enunciated* in another,' and thus in which 'every discourse is subject to the intrusion of other discourses into itself' (p. 14).

The simplest model of formalization used by Hak and Torode – one that Pêcheux draws on heavily in developing a computer analysis of discourse – is Harris's distributional method. Torode gives a schematic account of the relevant rewriting rules in 'Discourse Analysis Between Harris and Pêcheux.' They involve, first, finding repetitions of strings of words within sentences; second, constructing groups of equivalence classes of the environments of each repeated string; third, if these environments themselves constitute repeated strings, then grouping together *their* environments; and finally, establishing *approximate* equivalence classes by setting up chains of equivalences between *almost* identical environments (pp. 7–8). Torode sees anaphoric transformation as an example of the constitution of an approximate equivalence class, and he superimposes the distinction between automatic and approximate equivalence classes on Pêcheux's distinction between the two modes of information – preconstruction and enunciation – organizing the sentence. Thus 'a "proper" discourse analysis should search first and foremost for automatic equivalences (=preconstructions) within the texts (including transcripts) which, hypothetically, constitute the discourse ... Subsequently, the approximate equivalences (=enunciations) produced by grammatical transformations can be examined, noting the

ways in which these unavoidably introduce ideological shifts into the interpretation of the discourse' (p. 18).

In a later paper these categories are reversed. Anaphora now becomes equivalent to preconstruction (or 'discontinuity in discourse'), and lexical cohesion (or thematic repetition) 'identifies what Pêcheux calls *enunciation*, i.e., continuity in discourse.'⁶ In both cases, however, there is an assumption that discursive struggle can be read off from the relation between constructed thematic paradigms – between 'automatic' and 'approximate' equivalence classes, or between discourse 'trains.' There are some fairly simple things to say about this. All of them turn on the question of how relations of identity are constructed across discursive chains, and on the possibility of deriving semantic sameness and difference from structural sameness and difference.

There is a problem in the first place simply in postulating the identity of a discursive string in different locations in a text, since it is clearly possible that an element may have a quite different valency in relation to different enunciations (in the traditional sense of the word). Content analysis of this type can certainly locate explicit acts of enunciation described in the text (the fable Torode analyses contains definite acts of speaking) but it cannot identify non-explicit speech acts and the discursive and semantic shifts they may entail: this is to say that it cannot relativize statements to the position from which they are spoken. It might seem that conversation analysis, with its concern with turns of talk, might provide a relevant solution to this problem, but a turn is of course not necessarily identical to a single act of enunciation; a stretch of talk may incorporate more than one 'voice,' more than one discursive register. In the same way, it would be wrong to expect that what Torode calls discursive trains, and the play between trains, would allow us to assess the conflict of textual voices, since trains are not enunciative but lexical structures (and more than one voice may contribute to producing a single lexically cohesive sequence).

The second problem is that what is grouped in equivalence classes is linguistic surface structures which may have a quite different form or forms in the deep structure. Torode himself points to this problem when he indicates the inability of Harris's analysis to catch the word play on 'propaganda' and 'proper gander,' their reduction by Harris to an equivalence. His solution to this problem is to argue that the fable can be read in terms of an interdiscursive relation between the discourses of the fairy tale and the political fable. This solution seems to me quite correct, but note that it involves a retreat from any strict empiricism – involves 'mak[ing] assumptions about discourses *in the absence of specific texts which exemplify them*' (p. 13) (the fact that these discourse *could* be empirically investigated is a different question). I shall return to this point later.

The third problem I want to identify involves the relation between structural and semantic equivalence. All that a distributional analysis can identify is the former, and Harris specifically warns against any inference of semantic categories from structural equivalences. Nevertheless, this is surely the whole point of doing such an analysis. There are in fact two separate difficulties here. The first concerns the construction of equivalence classes out of material which is supposedly semantically neutral; Torode points out, for example, the problems that arise from Harris's assumption of the ideological neutrality of an active/passive transformation. The second is the problem of knowing when and how to interpret equivalence chains.

This problem is broached at some length in Tony Hak's 'Constructing a Psychiatric case.' Citing Harris's description of 'the principle of a formal method with which relations between statements can be described without having to presuppose an (arbitrary) interpretation by the researcher,'⁷ he notes that '(groups of) words that are in an identical environment are called "equivalent," which means no more than that they are in an "equivalent position" (in relation to an identical "environment").'⁸ This is a relation of *substitution*; but 'substitutability of two sentence parts still does not mean that the

substituents also have a similar "meaning."⁹ What then is the point in establishing relations of identity and equivalence? Hak answers this by distinguishing between strict synonymy and a more general 'connectedness' between statements. What Harris's method yields is the structure (the 'matrix of substitution') within which meaning is realized: a connection which must still be completed in a particular way by a reader.

In Hak's own analysis the key interpretive act seems to be the decision to read equivalent elements in a chain as mutual transformations (transpositions). The assumption seems to be that these are *real* transformations (not merely effects of the reading); and this means that the structurally equivalent elements around the 'sex(ual) life' group are read in their rewritten form as semantically equivalent.

In an extension of the analysis Hak constructs a further series of equivalences from his six texts, as follows:

- (Anna-Lize:) I've been used all my *life*.
- (Assistant:) *Is it broadcast* that you're being used?
- (Anna-Lize:) *It is broadcast* with whom I went to bed.
- (Father:) She reels off that she's being used for *sexual life*.
- (Assistant:) Her *sex life* is broadcast.
- (Report:) She belches out descriptions of her *sex life*.¹⁰

Here again the problem of what constitutes identity between textual strings is raised: 'all my life' and 'her sex life,' for example, are not 'identical environments' (and the lexeme 'life' comes close to being a pun here); and again it should be clear that positional identity (for example, between 'that you're being used' and 'with whom I went to bed')¹¹ can't be taken as justifying an inference of semantic identity; but this is precisely what happens in the move from Harris's purely formal method to a method concerned with semantic constants and transformations. At the same time, the overt eschewal of 'interpretation' means that Hak declines to consider the implications of the metaphor of Anna-Lize's sex-

life being 'broadcast on the radio' or the relation between this broadcasting and the 'voices' – apparently of the father – mentioned in Appendix I. In limiting himself to *describing* the constitution of a psychiatric 'knowledge' (and so a 'reality') he refuses to answer the question of the *adequacy* of the relations between discourses; the question of the specificity of Anna-Lize's language can't be asked.

Another way of posing the problem of how 'sameness' is constituted is to ask whether the equivalences established aren't perhaps merely the effect of the method used to identify and group them. This is to suggest that what interpretation finally interprets is the analytic method employed rather than a structure inherent in the material (there is a classic demonstration of this in Lévi-Strauss's structural analysis of the Oedipus myth, which groups narrative relations according to a common feature and then blithely proposes that 'all the relations belonging to the same column exhibit one common feature which it is our task to discover').¹² Torode's use of Prolog grammar rules for syntactic parsing is based on the supposition that 'the application of an axiomatic approach can reveal complex narrative structure in the everyday utterance,' and specifically that 'the discourse consistently distinguishes two subject positions,' described as here-reflective and there-transitive.¹³ In fact neither the reflective/transitive nor the here/there distinctions are *discovered* by the automatic analysis, since they're built into the parsing procedure. There is no 'recognition' involved in the parsing, only a transcription and a reorganization. This is not to condemn the 'arbitrariness' of interpretation, nor to deny any heuristic value to tautology; it is merely to insist that any description necessarily involves interpretation.

Or rather: any description except one based entirely and strictly on formal and positional identity (and even here, 'positional identity' can not be exact between two separate moments of a text). But such a formal description would suffer the crippling disadvantage that it can analyse only surface structures; and this would only *not* be arbitrary if there were a direct and invariant relation between surface lexical and syntactic structures and

logical and semantic deep structures. An example to the contrary, in passing, is Torode's distinction between two forms of the 'I' in an interview text – the I as teller and as actor.¹⁴ Recognising this functional distinction involves the need to rewrite the pronoun as two semantically distinct forms; and such a recognition is not available to a purely formal methodology.

The Proppian thesis that 'the formal analysis does not involve any interpretation of the discourse. Interpretation should follow the formal analysis, upon which it should be based'¹⁵ is therefore not tenable for any serious purposes. More generally, the model of formal analysis derived from Propp manifests the characteristic limits of formal narrative analysis in its inability to specify most of the dimensions of *generic* structure:¹⁶ modal relations (as some of Torode's other work on the distribution of pronominal structures does); shifts in rhetorical level; conditions of occurrence and use; and interdiscursive relations.

The central problem in the work of Hak and Torode concerns the relation between actual and virtual textual structures. In 'Discourse Analysis Between Harris and Pêcheux' Torode argues against Harris that discursive conflict depends on the possibility of one discourse referring to another, and that this means – in a phrase I have already quoted – that 'we must make assumptions about discourses *in the absence of specific texts which exemplify them*' (p. 13). Exemplification is methodologically desirable, and one of the tasks of a discourse analysis is to discover exemplifying texts; this is the advantage of Pêcheux's concept of the preconstructed, that it seems to offer a positive enunciation of an intertextual structure. In other words, although 'often, preconstruction is hidden' (p.14), it is made explicit in relative clauses. Explicitness, positivity, is possible but not by any means inevitable. Later in the article, however (p.15), he seems to take the preconstructed/enunciated relation not as a relation between the virtual and the manifest but *solely* as a relation of levels of enunciation within the enunciated - a relation between, for example, narrator's discourse and characters' discourse, that is, a relation between

two actualized bits of text.

The point is that textually presupposed discourse – the preconstructed – may or may not find linguistic manifestation in a particular text. There is no guarantee of its visibility, and no certainty that the ideological can be read off from the linguistic. If we can't discover preconstruction in a given text, we must look for it elsewhere. But in his 'Response to Tony Hak' Torode explicitly refuses this alternative: he follows MacCabe in rejecting the location of preconstruction in 'some external source (some other discourse)' and opts instead for restricting the preconstructed/enunciated distinction 'to occasions when it appears within the discourse itself' (p. 200).

The issue here is whether what is of analytic interest is a *text* or *textual structure*. Some of the ambiguity surrounding this issue is evident in Hak's 'Why Voloshinov Needs Formal Method.' Here he uses the concept of a substitution matrix as a way of restricting the infinity of possible alternatives to a given utterance. Taken from Pêcheux, the concept promises to make interdiscursive relations visible 'by the systematic confrontation of series of sentences which are generated from out of the whole of utterances that belong to a "sociolect" or "discourse"' (p.32). The substitution matrix, that is to say, is a phrase-generating structure; but Hak seems not to understand it in a generative way, since he analyses it not structurally but through an enumeration of the totality of its contents. Paradoxically, then, this list, this 'series of sentences,' would have to be almost infinitely long in order to account for all of the phrases that could be generated, and the purpose of an 'empirical' analysis is thereby defeated.

A better solution to this problem – and one of which Torode apparently, if inconsistently, approves – is the one Hak proposes in 'Constructing a Psychiatric case.' The text of a psychiatric report, which 'forms the set of present statements to be analysed,' is related to 'two corpora of absent statements, i.e., statements that may be regarded to play a part in the report even if they are absent there.' The first of these consists of 'the set of statements

which were made about the patient *previously and elsewhere*,' and the second consists of 'statements which represent the implicitly assumed "psychiatric knowledge."' The analysis is thus characterized by the fact 'that "present" and "absent" (but elsewhere present) statements are confronted with each other,'¹⁷ in such a way that the text of the report can be read as a production of 'statements that represent a "reality" and a "knowledge" all in one.'¹⁸ Nevertheless, there remains a problem about how the relation between 'present' and 'absent' statements can be thought.

It is to Foucault's 'positive' understanding of the mode of existence of statements that Hak turns for clarification of the methodological bases for a 'literal description'¹⁹ of discourse. Foucault's genealogical method – which is then operationalized by Pêcheux's computer analysis – establishes 'the *return* or the *failure to return* of statements':²⁰ regularities of occurrence and non-occurrence. Now, the problem with turning to Foucault, and precisely to the notion of positivity, is that this marks the site of a fundamental ambiguity in *The Archaeology of Knowledge* as to whether the concept of statement (*énoncé*) refers to a *function* which is the precondition of discursive events, or to an *actual* discursive event. The ambiguity derives, I think, from the fact that Foucault needs to fit the concept of statement both into the 'structuralist' task of an analysis of the conditions of possibility of discourse, and into the positivist task of a 'pure description of discursive events,'²¹ the definition of 'a limited system of presences.'²²

What governs the limitation of the realm of discourse is a law of rarity: discourse is a closed realm of necessary events (in the sense that only the statements that did appear could have appeared). Because Foucault understands discourse as an *aggregation* of events (as does Torode when he defines it as a 'succession of sentences'),²³ it thus becomes difficult for him to explain the conditions of production of new discourse. A 'positive' conception of discourse cannot found an analytic method because it defines 'absence' only as non-presence, rather than as a *structured* relationship; and because it therefore cannot adequately theorize the complex layering of discursive information.

Let me conclude by trying to extrapolate from my criticism some of the alternatives it would entail. Very briefly, it would involve positing a level of generative structure situated between the levels of the language system and the utterance; following Volosinov I call this level that of the discursive *genres*. Working at this level makes it possible to understand the infinity of utterances by recourse to a limited set of codes; discourse is thus understood as a systemic, rule-governed process rather than as a random, 'subjective' or aggregative process. The relation between code and message is the model for thinking the relation between the virtual and the actual (interdiscursive relations are thus in the first place a relation to codes and only secondarily to messages: in Hak's example, to the genre of the psychiatric textbook and its knowledges rather than to a particular knowledge), and preconstruction will be thought in terms of a layered distribution of information between a background and a textual foreground. Following Halliday, finally, the concept of discourse genre will be thought as a multideterminate structure which specifies a range of semantic dimensions – thematic, modal, rhetorical – for a particular type of speech situation, as well as specifying conditions of use and availability. Discourse will be understood as the play and conflict of genres and the ideological presuppositions they carry with them at any point in time. Thinking discourse in this way doesn't by any means preclude formalization: but it does mean opting for a 'softer' and interpretively oriented mode of analysis.

Notes

¹ Cf. *Working Papers on Discourse and Conversational Analysis*, ed. Tony Hak, Joke Haafkens and Gerhard Nijhof, *Konteksten*, No. 6, 1985; page references to this volume will be given in the text.

² Michel Pêcheux, *Les Vérités de la Palice*, Coll. Théorie (Paris: Maspéro, 1975), p. 92.

³ Ibid., p. 99.

⁴ Ibid., p. 143.

⁵ Colin MacCabe. 'On Discourse,' *Economy and Society*, 8:4 (August, 1979), 279-307.

⁶ Brian Torode, 'Is There an Empiricist Discourse Analysis?' Unpublished paper, p. 9.

⁷ Tony Hak, 'Constructing a Psychiatric case.' Unpublished paper, p. 19.

⁸ Ibid.

⁹ Ibid., p. 20.

¹⁰ Ibid., p. 27.

¹¹ Ibid.

¹² Claude Lévi-Strauss, 'The Structural Study of Myth,' *Structural Anthropology*, trans. Claire Jacobson and Brooke Grundfest Schoepf (London: Allen Lane, 1968), p. 215.

¹³ Brian Torode, 'Computational Sociolinguistics and Communication Failure: On the Resolution of Incompleteness in Automatic Discourse Parsing,' forthcoming in R. Reilly (ed.), *Communication Failure in Dialogue* (Amsterdam: North Holland), p. 20.

¹⁴ Ibid., p. 5.

¹⁵ Ibid.

¹⁶ Cf. John Frow, *Marxism and Literary History* (London: Blackwell, 1986), pp. 67–82.

¹⁷ Hak, 'Constructing a Psychiatric Case,' p. 14.

¹⁸ Ibid., p. 36.

¹⁹ Ibid., p. 12.

²⁰ Ibid., p. 15

²¹ Michel Foucault, *The Archaeology of Knowledge*, trans. A.M. Sheridan Smith (London: Tavistock, 1972), p. 27.

²² Ibid., p. 119

²³ Torode, 'Is There an Empiricist Discourse Analysis?' p. 1.

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

FROW, JOHN

Title:

Formal Method in Discourse Analysis

Date:

1989

Citation:

Frow, J. A. (1989). Formal method in discourse analysis. Journal of Pragmatics, 13, 333-341.

Publication Status:

Published

Persistent Link:

<http://hdl.handle.net/11343/34361>