

Sarrasine Revisited: A Perspective in Text-Analysis

Jacqueline Léon and Jean-Marie Marandin

To Describe, to Read a Text

A text, any text, is a complex organization that cannot be reduced to simple patterns or described in a unidimensional way. The procedures we describe here aim at capturing two different aspects of a text: SEGTHEM (THEMatic SEGmentation) in Section 1 and DENSITE (Density) in Section 2. They are currently under development and we see them as heuristic tools in text analysis. We apply them to texts to remodel the hypotheses or assumptions which we accept when constructing them; furthermore, they allow us to "read" these texts. We have applied our procedures to Balzac's short story *Sarrasine*. "To read is a labour of language" (Barthes 1970: 17, [1975]: 11). For us, to read is the labour of making sense out of the confrontation of three states of the same text: the text as it reached us, laden with literary history and, in particular, impregnated with the description Barthes gave of it in *S/Z*, and the two texts yielded by our procedures. In Section 3 we shall read a fragment from *Sarrasine*: the portrait of the weird old man in Part I (given in annex).

We use P. Plante's DEREDEC programming system. DEREDEC, written in LISP, is a "workbench" in Linguistics and Text Analysis. Its remarkable flexibility allows various types of syntactic or textual parsings. Our procedures use the output of the syntactic parser GDSF (Grammaire de Surface du Français).¹

Jacqueline Léon and Jean-Marie Marandin are both full time CNRS researchers and belong respectively to LISH and INaLF working in the SLID team (Syntaxe, Lexique, Informatique, Discours).

1. SEGTHEM

The SEGTHEM procedure is intended to provide a representation of the thematic organization of a series of sentences; it consists of a parenthesizing procedure which groups sentences into thematic segments and thematic sequences. Rather than describe it in full detail here, we shall limit ourselves to a description of the rules which involve a definite or demonstrative thematized Noun Phrase (NP). This prototype procedure in fact produces a first fairly good approximation of the thematic organization of our sample of text.

1.1 *The notion of theme*

The notion of theme which we postulate and attempt to build with the SEGTHEM procedure is different from the notion inherited from the Prague School. We distinguish between the thematic position of a Noun Phrase in a clause, and the theme of a discourse. The thematic position of a NP is a quasi-structural notion, and we postulate that in French, it is occupied by the leftmost (non-prepositional) NP (within the exception of NPs which are in apposition); we call this NP "a thematized NP." The notion of theme of discourse is redefined in a non-referential sense of lexical meaning. Within this paradigm, the NP which we call theme of discourse is the name of an object of discourse. An object of discourse is taken as a set (or more precisely a mereological class) of "quasi-propositions"; we call the constituents of these propositions the ingredients of the object.²

A theme of discourse is the NP which groups one or more sentences developing the ingredients of an object and which names it. These sentences are the thematic segments; the segments which can be grouped form a thematic sequence. It is possible to

identify different types of thematization; one can, for example, contrast an accumulating process which is typically made up of a definite NP head of a chain of pronominal anaphoras, and a compacting process typically made up of a demonstrative NP interpreted anaphorically.

For example, in our text:

a) accumulating process based on a definite NP (here *le petit vieillard*):

- (1) “... Quoique le petit vieillard eût le dos courbé... *Son* excessive maigreur... *Il* portait une culotte...”

Thematization operates from left to right. The text accumulates ingredients of the named object *le petit vieillard*.

b) Compacting process by a demonstrative NP (here *ces gibbosités*):

- (2) “Le menton était creux; les tempes étaient creuses; les yeux étaient perdus... Les os maxillaires... dessinaient des cavités... *Ces gibbosités*... produisaient des ombres et des reflets curieux...”

Thematization operates from right to left. The demonstrative NP *ces gibbosités* here closes an enumeration and names the enumerated ingredients.

1.2 The procedure

SEGTHEM, made up of four DEREDEC-automata (in its current version), inserts segment and sequence opening and closing marks in a series of sentences. The steps of the procedure which we describe under a simplified form, are as follows:

Step 1: The text is syntactically parsed by GDSF. From this parsing we obtain NPs in a sentential thematic position (as given by the topic-comment relation yielded by GDSF).

Step 2: The NPs in thematic position are recategorized according to their determiner (definite, indefinite, possessive, demonstrative). For example:

- (le petit vieillard) (le petit vieillard)

GN → SNLTH

where GN = Noun Phrase.

SNLTH = NP definite (L) thematized (TH).

Step 3: The occurrence of recategorized NPs triggers the insertion of thematic markers at the beginning or end of a sentence:

a) the occurrence of a SNLTH is interpreted as the

opening of a segment: a left round bracket is inserted at the beginning of the sentence. For example:

- (3) ... (Quoique <SNLTH: le petit vieillard> ...
... (<SNSTH: son excessive maigreur> ...

In the current state of the procedure, the occurrence of a thematic possessive NP (SN S[son] TH) triggers the same operation as a SNLTH.

b) The occurrence of a SNCTH (a thematized demonstrative NP) is interpreted as the end of a sequence and the beginning of a segment; a sequence closing mark (a right square bracket) is inserted at the end of the sentence, and a segment opening mark (a left round bracket) is inserted at the beginning of the sentence. For example:

- (4) ... (<SNCTH: ces gibbosités> ... achevaient d'ôter à ce visage les caractères de la face humaine.] Puis, les années... .

A demonstrative NP names a sequence on its left and can, in turn, be processed as an accumulative NP (in the same way as a SNLTH) by an anaphorical chain. The occurrence of a SNCTH thus triggers the insertion of a segment opening mark at the beginning of the sentence and the insertion of a sequence closing mark. In this case, we insert the sequence closing mark at the end of the sentence containing the last occurrence of a thematized anaphorical pronoun. For example:

- (5) ... (<SNCTH: Cette espèce d'idole japonaise> ... Silencieuse, ... elle exhalait l'odeur... pendant un inventaire].

We are still investigating a particular case exemplified in (6) below: the head noun of the SNCTH is repeated from the immediate left context. In the current version, such an occurrence is interpreted as a closing mark of a complex segment (which is opened at the beginning of the sentence where the repeated noun occurs in the first place). In this case, we have at the same time a promotion of an ingredient to the status of sentential theme and a restriction of the naming effect to this ingredient. For example:

- (6) ... (Un jabot de *dentelle* d'Angleterre assez roux... ; mais sur lui, cette *dentelle* était plutôt un haillon qu'un ornement).

Step 4: We construct the sequences whenever it is possible: a sequence is made up of segments which are named by a same NP. So we have to determine whether there is a NP which can fulfill this function and which NP is such. Or to use Reinhart's

definition (see note 2 above), we have to determine which NP works as “an entry under which” to file thematic segments. Here we shall restrict ourselves to examples where such a NP is at the beginning or end of a sequence, and limit ourselves to providing some procedural rules for the construction of a sequence, triggered by the occurrence of a SNLTH or a SNCTH.

In the case of a configuration such as:

...(<SNLTH: x> . . . (<SNLTH: y> . . . (<SNLTH: z>)...
we have to decide if the first SNLTH can function as the name of an object of discourse, developed in the series of segments. This decision is easy in two cases: the occurrence of a proper noun or of a NP with the status of regular designator (this concept we owe to Corblin 1983). A clue to identifying a regular designator is provided by the frequency of the substantive head of the NP in sentential thematic position, which is higher than that of the other NPs. Such NPs trigger the opening of a sequence: we call them “accumulative NPs.” It is the case in (1) above where *le petit vieillard* (occurring 14 times in the text) is the regular designator of the mysterious character introduced in the first part of the story whereas in (2) above *le menton* has no special status compared to *les tempes*, *les yeux*, *les os*.

In the case of an occurrence of a SNCTH (remember that this occurrence has been interpreted as marking the end of a sequence in the third step), we construct the sequence when a sequence opening mark (such as those triggered by accumulative NPs) is met on the left; if no such mark is met, the sequence is opened when a sequence closing mark is met. Such a rule is used to process (2) above. This “default” opening will be represented by the sign in (7) below.

1.3 Results

When applied to our text, SEGTHEM (reduced to the rules involving thematized definite and demonstrative NPs) gives a representation of the text under the form of sequences which is shown in a simplified format in (7); a complete parenthesized text appears in the annex. The sequences are numbered from 1 to 6.

(7): [₁(Le petit vieillard . . . [₂(L'inconnu . . . (. . . (Ce luxe . . . , ce trésor . . . cet être bizarre)₂)
<₃(Le cadre . . . (Ce visage noir . . . dans tous les sens)₃) <₄(Le menton . . . (. . . (Ces gibbo-

sités . . . les caractères de la face humaine)₄)
<₅(Puis les années . . . (Cette espèce d'idole japonaise . . . Silencieuse . . . pendant un inventaire)₅) [₆(Si le vieillard . . . Ah, c'était bien la mort³ . . . par le corsage)₆].

2. DENSITE

The DENSITE procedure is of a different type: not based on sharply defined hypotheses as is SEGTHEM but devised to explore a fuzzy phenomenon: the textual functions of lexical repetitions. At the first blush, lexical repetition is the recurrence of lexical items in a text.

2.1 Lexical repetition

Lexical repetition has been the object of many sophisticated quantitative treatments since the use of computers in the description of texts. Usually, these methods postulate that the recurrence of lexical items or the recurrence of cooccurrences of lexical items may characterize the content of a text. The lexical forms are considered independently of the phrase in which they occur; if, for example, a substantive Ni is repeated, they do not take into account the features of the Noun Phrase (NP) of which Ni is a constituent, such as their determiners (definite, indefinite, etc.), their syntactic status (subject, object, etc.), their clausal locus (extraposed on the left, etc.)⁴. We take up the problem from a standpoint where we do not discriminate between Form and Content and which enables us to take into account the properties of the occurrence of a given item in its contexts.

Lexical repetition is not a simple phenomenon. Bolinger (1979) has shown that repetition of a noun and use of a pronoun fulfill different functions when they are considered in actual chains of utterances. Bolinger ends his description of the pragmatic or textual functions of the repetition of a substantive with these words: “The decision to repeat a noun, or, instead, to use a pronoun depends on how necessary or desirable it is to re-identify the referent at a given point.(.) Other re-identifications respond to some implied or underlying assertion about the referent. It may be asserted as topic (“we are talking about Tom”) or asserted in terms of its nature (“Tom qua Tom”) or the assertion may involve an extraneous viewpoint whereby the speaker attributes to the referent some expression that is not (or not entirely) the referent’s

own at the time (. . .)' (1979:308). Two features are of considerable importance in our framework:

- the repetition of a Ni in a definite thematized NP is usually a good clue to its function as a regular designator of an object. This “quantitative” information is exploited in SEGTHEM when it is necessary to choose, among a number of definite NPs which open thematic segments, the one which has the property of accumulating these different segments.
- the second occurrence of a lexical item Xi in a series of utterances is not the mere repetition of a material segment of language. By its very repetition, a lexical item gains in “semantic opacity”: its semantic value or its connotative potentiality seems to be enhanced (“the X qua X effect” in Bolinger’s description). This effect which is noticeable in a chain of utterances, is more directly approachable on the macro-level of textual organization.

2.2 DENSITE

So far, we have considered lexical repetition on the micro-level of a chaining of utterances. DENSITE tackles it on the macro-level of textual organization. We are interested here in giving some substance to the distinction drawn by Barthes between “a sequential space” and “an agglomerative space” in his discussion of the notion of connotative anaphora (1970: 14 et sq, [1975]: 11). He defines “agglomerative space” as “a nebula of signified” or, to use another metaphor, a web of links related to other texts or other areas of the text itself. These characterizations do not lend themselves to direct processing; hence our redefinition of them in terms of lexical connotation via lexical repetition. Certainly, our redefinition does not exhaust what Barthes had in mind; our goal is simply to implement procedures to capture his intuition. We assume that the occurrence of an item Xi in a text “has the power to relate itself to anterior, ulterior, or exterior mentions [of that Xi], to other sites of the text (or of another text”’; (ibid: 14, [8]). On such an assumption, a unit built up of recurrent items can be viewed as potentially having a higher density of connotative links than a unit fashioned from hapax. In the present version this unit is defined as a clause (a “Groupe Propositionnel” in GDSF). Such an assumption is typically a proposition which cannot be verified or falsified; this

should not prevent us from tackling the issue if it is to be pertinent in describing a text. Wittgenstein addresses a similar problem in his *Philosophical Grammar*: “It looks as if a sentence with eg. the word ‘ball’ in it already contained the shadow of other uses of this word. That is to say the possibility of forming those sentences. To whom does it look like that? And under what circumstances?” (:54).

2.3 Procedure

The DENSITE procedure traces within a text those clauses that contain at least two recurring full lexical items. We call these clauses dense: a clause Pi is denser than Pj if Pi contains more recurring items than Pj. The different steps of the procedure, coded in DEREDEC⁵ are as follows:

Step 1: The text is parsed by GDSF (Grammaire de Surface du Français).

Step 2: Full lexical items are searched;⁶ these items are indexed according to their addresses in terms of sentence and clause.

Step 3: A lexicon of recurring (frequency ≥ 2) full items with their addresses is created. Note that at this level, an interactive morphological grouping procedure provides a lexicon of lemmatized lexical items.

Step 4: The addresses of clauses containing n full recurring lexical items are stored; the list of the addresses of the dense clauses with the number of their full recurring lexical items is produced.

The number n (2 by default) is a parameter; a higher number can be chosen to provide speedier retrieval of the results.

At this point it is possible to calculate the ratio between the number of full recurring items, and the total number of items in the clause.⁷

Step 5: A list of dense clauses is obtained. Two representations of the results are provided: clauses under their surface representation or as generated by the parser.

2.4 Results

DENSITE was applied to part I of *Sarrasine*; lexical items are not lemmatized. We give below the densest clauses of our fragment of text (i.e., the portrait); they are dense relative to the set of clauses of the first part of the novel.

1) The pyramid of dense clauses is the output of Step 4; it comprises the list of dense clauses

ordered hierarchically according to the number *n* of full recurring lexical items.

("160-1" 6)	("166-1" 3)
("172-1" 6)	("167-2" 3)
("186-1" 5)	("168-1" 3)
("193-1" 5)	("171-1" 3)
("161-1" 4)	("173-1" 3)
("180-3" 4)	("185-1" 3)
("190-1" 4)	("185-4" 3)
("161-2" 3)	("162-2" 2)
("162-1" 3)	("163-2" 2)

to be read as follows:

("number of the sentence"—"number of the clause" number of full recurring items)

2) the densest clauses, i.e., the output of Step 5, under their surface representation.

(8) ("160-1" 6) elle s'enhardit alors assez pour examiner pendant un moment cette créature sans nom dans le langage humain, forme sans substance, être sans vie, ou vie sans action.

(9) ("172-1" 6) ce luxe suranné, ce trésor intrinsèque et sans goût, faisaient encore mieux ressortir la figure de cet être bizarre.

(10) ("186-1" 5) enfin, cette espèce d'idole japonaise conservait sur ses lèvres bleuâtres un rire fixe et arrêté, un rire implacable et goguenard, comme celui d'une tête de mort.

(11) ("193-1" 5) c'était bien la mort et la vie, ma pensée, une arabesque imaginaire, une chimère hideuse à moitié, divinement femelle par le corsage.

3. Reading

We presented in Sections 1 and 2 the results of our two description procedures. We shall now "read" our extract in the light of these results. First a number of observations are called for:

Observation 1: The locus of the occurrence of the four densest clauses is not insignificant: they appear at the hinges of the rhetorical organization which characterizes the conventional form of a portrait:

—(8) appears at the beginning of the portrait (this is the classical introduction: a portrait is what a character sees) and introduces the passage which is dedicated to the full-length description.

—(9) appears at the end of the passage dedicated to the description of the full-length portrait.

—(10) appears at the end of the passage dedicated

to the impression conveyed by the face.

—(11) appears at the end of the portrait and at the end of the passage dedicated to the general impression that the character who is portrayed gives to another (this is the conventional way of ending a portrait).

Observation 2: The thematic organization concurs with the rhetorical organization: the compacting NPs (*ce luxe suranné, ce trésor intrinsèque, cette espèce d'idole japonaise, c'était bien la vie et la mort*) appear in the dense clauses (9), (10), (11), and in the case of *ces gibbosités* at the joint between the passage concerning the impression left by the face.

Observation 3: The thematic sequences [2] and [6] are remarkable precisely because they bring two thematization processes into play: accumulation from a regular designator and compacting by a demonstrative NP. Considering the NP which names the object (here the character), it is notable that the double movement of thematization is accompanied by a transformation of the object itself; first a person (*l'inconnu, le vieillard*) then it becomes a collection of properties (*ce luxe suranné, ce trésor intrinsèque et sans goût, la mort, la vie . . .*).

Observation 4: Isolating the four dense clauses, it becomes obvious that the character is first *une créature sans nom* (a creature with no name); he takes on an outline (*une figure, cette espèce d'idole japonaise*), an expression (*un rire implacable et goguenard*) and a sex (*chimère divinement femelle*).

Observation 5: Accepting the hypothesis of connotation by lexical repetition and searching for the areas of the text where recurring items appear (namely *mort, vie, être, rire, corsage, créature*, it is clear that these items are found in the passages concerning the narrator's visions (the garden of the Lanty's hotel, the young women, the weird old man).

These few descriptive elements seem to us to seize on two aspects of the passage which Barthes underlines in *S/Z* (:67, [:61]):

—"The portrait (in the present text) is not a realis-

tic, a related copy, an idea such as we might get from figurative painting; it is a scene made up by blocks of meaning, at once varied, repeated and discontinuous"; this portrait is "cubist."

—"the old man's body is not 'detached' like a real referent from the background of the words or of the salon".

It can be called "cubist" because the identity of the character is "exploded," smashed to a collection of properties (observation 3). This fragmentation is perpetrated in the classical form of portraiture (Observations 1 and 2) in the same way that Picasso or Braque fragmented the representation of objects within the classical framework of the canvas. Besides, this fragmentation goes alongside the "feminization" of the character (Observation 4); Barthes develops this theme lengthily enough to make it central to the story: the problem of the castrato and castration.

Finally, the character's body, even the character himself is not detached from the text as an object belonging to an extra-linguistic world: he is dissolved in his materials and these materials are those of which the narrator's visions are made (Observation 5).

Our elements of description thus tend to support Barthes' reading (this is not always the case). The work which underlies Barthes' reading is implicit, hidden; our procedures are explicit and use hypotheses and assumptions the scope and value of which can be gauged by applying them to other texts. Our reading is done systematically and we show how we read. In fact we put into practice what Barthes only postulates: "There is no other proof of a reading than the quality and endurance of its systematics" (:17, [:11]).

4. Conclusion

This description of the portrait in part I of *Sarrasine* is a sample of the approach we advocate in Text-Analysis (Lecomte-Marandin, 1984). Among the texts we are currently analyzing, we consider literary texts a touchstone for our procedures, since they are most likely to provide us with the most sophisticated levels of textual organization. One aim of Text-Analysis is to yield representing models of forms of textual organizations; in metaphorical terms, they yield different maps of a given text. SEGTHEM, for example, maps out the thematic organization of a text: the parenthesized

text (1.3 supra) is one map of our text. It gives us a description of the thematic links supported by sentential anaphoric elements and a view to describe how information about the objects of discourse (the character in our description) are structured in the referential (the context set, in Reinhart's words) built by a text (cf. Léon-Marandin, to be published). These maps can be the basis of high level interpretative processes such as Barthes' reading or the types of readings of those who analyze texts in the Social Sciences (Sociology, Discourse Analysis. . .). Our procedures are then simulations of the labour of language involved by these readings. By computing our procedures, we give an explicit formulation of the concepts of Text-Analysis and we meet the requirements of a genuinely experimental description of texts.

ANNEX: the portrait of the old man in *Sarrasine*, Part I.

(. . .) : segment opening and closing marks.

[. . .] : sequence opening and closing marks

< default sequence opening mark.

the densest clauses are underlined.

the sequences are numbered from 1 to 6.

Elle s'enhardit alors assez pour examiner pendant un moment cette créature sans nom dans le langage humain, forme sans substance, être sans vie, ou vie sans action. Elle était sous le charme de cette craintive curiosité qui pousse les femmes à se procurer des émotions dangereuses, à voir des tigres enchaînés, à regarder des boas, en s'effrayant de n'en être séparées que par de faibles barrières. [1 (Quoique le petit vieillard eût le dos courbé comme celui d' un journalier, on s'apercevait facilement (que sa taille avait dû être ordinaire.(Son excessive maigreur, la délicatesse de ses membres, prouvaient que (ses proportions étaient toujours restées sveltes. Il portait une culotte de soie noire, qui flottait autour de ses cuisses décharnées en décrivant des plis, comme une voile abattue. Un anatomiste eût reconnu soudain les symptômes d'une affreuse étiologie en voyant les petites jambes qui servaient à soutenir ce corps étrange. Vous eussiez dit de deux os mis en croix sur une tombe. Un sentiment de profonde horreur pour l' homme saisissait le coeur quand une fatale attention nous dévoilait les marques imprimées par la décrépitude à cette casuelle machine. [2 (L' in-

connu portait un gilet blanc, brodé d'or, à l'ancienne mode, et (son linge était d'une blancheur éclatante. (Un jabot de dentelle d'Angleterre assez roux, (dont la richesse eût été enviée par une reine, formait des ruches jaunes sur sa poitrine; mais sur lui cette dentelle était plutôt un haillon qu'un ornement). Au-milieu-de ce jabot, un diamant d'une valeur incalculable scintillait comme le soleil. (Ce luxe suranné, ce trésor intrinsèque et sans goût, faisaient encore mieux ressortir la figure de cet être bizarre₂). ζ_3 (Le cadre était digne du portrait. (Ce visage noir était anguleux et creusé dans tous les sens₃). ζ_4 (Le menton était creux; (les tempes étaient creuses; (les yeux étaient perdus en de jaunâtres orbites. (Les os maxillaires, rendus saillants par une maigreur indescriptible, dessinaient des cavités au-milieu-de chaque joue. (Ces gibbosités, plus ou moins éclairées par les lumières, produisaient des ombres et des reflets curieux qui achevaient d'ôter à ce visage les caractères de la face humaine₄). ζ_5 (Puis les années avaient si fortement collé sur les os la peau jaune et fine de ce visage qu'elle y décrivait partout une multitude de rides, ou circulaires comme les replis de l'eau troublée par un caillou que jette un enfant, ou étoilées comme une fêlure de verre, mais toujours profondes et aussi pressées que les feuilletés dans la tranche d'un livre. Quelques vieillards nous présentent souvent des portraits plus hideux; mais ce qui contribuait le plus à donner l'apparence d'une création artificielle au spectre survenu devant nous était le rouge et le blanc dont il reluisait. (Les sourcils de son masque recevaient de la lumière un lustre qui révélait une peinture très bien exécutée. (Heureusement pour la vue attristée de tant de ruines, (son crâne cadavérique était caché sous une perruque blonde dont les boucles innombrables trahissaient une prétention extraordinaire. (Du reste, la coquetterie féminine de ce personnage fantasmagorique était assez énergiquement annoncée par les boucles d'or qui pendaient à ses oreilles, par des anneaux dont (les admirables pierrieres brillaient à ses doigts ossifiés, et par une chaîne de montre qui scintillait comme les chatons d'une rivière au cou d'une femme. (Enfin, cette espèce d'idole japonaise conservait sur ses lèvres bleuâtres un rire fixe et arrêté, un rire implacable et goguenard, comme celui d'une tête de mort. Silencieuse, immobile autant qu'une statue, elle exhalait l'odeur musquée des vieilles robes que les

héritiers d'une duchesse exhument de ses tiroirs pendant un inventaire₅). [Si le vieillard tournait les yeux vers l'assemblée, il semblait que (les mouvements de ces globes incapables de réfléchir une lueur se fussent accomplis par un artifice imperceptible ;et quand (les yeux s'arrêtaient, celui qui les examinait finissait par douter qu'ils eussent remué. Voir, auprès de ces débris humains, une jeune femme, dont (le cou, les bras et le corsage étaient nus et blancs; dont (les formes pleines et verdoyantes de beauté, dont (les cheveux bien plantés sur un front d'albâtre inspiraient l'amour, dont les yeux ne recevaient pas, mais répandaient la lumière, qui était suave, fraîche, et dont (les boucles vaporeuses, dont (l'haleine embaumée, semblaient trop lourdes, trop dures, trop puissantes pour cette ombre, pour cet homme en poussière; (ah! c'était bien la mort et la vie, ma pensée, une arabesque imaginaire, une chimère hideuse à moitié, divinement femelle par le corsage₆].

NOTES

1. A complete description of DEREDEC is provided in Plante (1985). Our programming philosophy is exposed in Lecomte et al. (1984). GDSF, described in Plante (1983), is a modular bottom-up parser. It produces an hierarchical analysis of the phrases of sentences and links those phrases by contextual dependency relations: among others, the topic-comment relation between the leftmost (non-prepositional NP) and the Verb Phrase.
2. We cannot fully describe the concepts we introduce here. See: —Reinhart (1982), for a critical description of the Prague School notion of theme. —Reinhart (1982), for a description of a concept of theme which is closer to our own: "Sentence topics(. . .) are one of the means available in the language to organize, or classify the information exchanged in linguistic communication. They are signals for how to construct the context set, or under which entries to classify the new proposition" (:24). For us, this definition can be applied to the notion of theme of discourse. —Lecomte (1981), for the concept of an object of discourse as a mereological class, and Marandin (1986), for the links between lexical meaning and object of discourse. our concept of object of discourse is close to the concept of notion (Culioli 1980) or of stereotype (Putnam 1975). —Lecomte (1985 and 1986) for a modelling of these concepts. The procedure SEGTHEM is fully described in Léon-Marandin (to be published).
3. "ce + être [to be] + NP" is treated as a compacting formula. See Léon-Marandin for a justification.
4. We are currently working on a procedure VSIL (Syntactic Variation of a Lexical Item) which provides these properties, for all the items of a text, or a subset of them, in the form of a table.
5. We are grateful to P. Plante for most of the coding of DENSITE.
6. Substantives, verbs other than auxiliaries and support-verbs, adjectives, adverbs ending in "-ment" are considered full lexical items. A special module in GDSF performs the context sensitive disambiguation of such verbs: for example: Pierre a un voiture (full verb), Pierre a parlé (auxiliary), Pierre a faim (support-verb).
7. DENSITE does not necessarily pick out the longest sentences in a text since it works on a clausal level. Besides, DENSITE does not necessarily pick out the longest clauses in a text.

REFERENCES

- Barthes, R. 1970, *S/Z*, Seuil, Paris.
English transl., R. Miller (transl.), 1975, Jonathan Cape Ltd, London.
- Bolinger, D., 1979, "Pronouns in Discourse," [Gvon T. ed.] *Discourse and Syntax*, Academic Press, New-York.
- Corblin, F. 1983, "Les désignateurs dans le roman," *Poétique* 54: 199-211.
- Culioli, A., 1981, "Sur le concept de notion," *BULAG* 8, :62-77.
- Lecomte, A., 1981 "Comment Einstein raconte comment Newton expliquait la lumière," *Revue Européenne des Sciences Sociales* 56 :69-93.
- Lecomte, A., 1986, "Espace des séquences," *Langages* 81:91-10g.
- Lecomte, A., Marandin, J.M., 1984 (to be published) "Analyse du Discours et Morphologie Discursive," [Charolles M. and Petöfi J. eds.], *Buske Verlag*, Hamburg.
- Lecomte, A., Léon, J. Marandin, J.M., 1984, "Analyse du discours, stratégies de description textuelle," in *MOTS* 9 :142-165.
- Léon, J., Marandin, J.M., 1986 (to be published) "Détermination, Anaphore et Thématisation: la procédure SEGTHEM."
- Marandin, J.M., 1979, "Problèmes d'Analyse du Discours. Essai de description du discours français sur la Chine," in *Langages* 55 :17-88.
- Marandin, J.M., 1986, "Ce est un autre. L'interprétation anaphorique du syntagme démonstratif," in *Langages* 81: 75-83.
- Plante, P., 1983, *GDSF, une Grammaire de Surface du Français*, Centre d'Analyse des Textes par Ordinateur, Université du Québec à Montréal (to be published).
- Plante, P., 1985, Manuel de programmation en DEREDEC, Centre d'Analyse des Textes par Ordinateur, Université du Québec à Montréal.
- Plante, P., 1985, "La structure des données et des algorithmes en DEREDEC," *Revue québécoise de linguistique*, vol. 14, n°2, :119-143.
- Putnam, H., 1975, "The Meaning of Meaning," *Minnesota Studies in The Philosophy of Science*, vol. VII: 131-193.
- Reinhart, T., 1982, *Pragmatics and Linguistics: an Analysis of Sentence Topics*, Indiana University Linguistics Club.
- Wittgenstein, L., 1978, *Philosophical Grammar*, U. of California Press.