

**Automated syntactic text description enhancement :  
determination analysis<sup>1</sup>**

Jules Duchastel, Louis-Claude Paquin et Jacques Beauchemin  
Université du Québec à Montréal

## **1. The Context**

This paper seeks to make a pragmatic contribution to computer-assisted discourse analysis by showing how a syntactical description could be used to gain a deeper understanding of the texts. During the '80s, a large corpus (5,000 pages) of varied political discourses covering twenty five years (1934-1960) was collected and a double description of it was carried out. On the one hand, to deal with the occurrence of terms referring to the same notions in the lexicon, a set of 144 sociological categories was assigned to relevant expressions. For example, the following terms were tagged as 'financial notions' : bank, credit, savings, dividend, etc. On the other hand, a surface syntactical description of each sentence of the corpus was produced by means of a parser (Plante 1979) controlled by an heuristic strategy and programmed in Déredec (a LISP sub-language). The use of a syntactical description was founded on the hypothesis that it would reduce the random nature of the lexical distribution of the words in the texts. A lexicon of the clause subjects is more 'qualified' than a lexicon with no criteria. This particular description identifies three contextual dependency relations for every clause, namely: theme/rheme, determination and verbal arguments.

---

<sup>1</sup> To appear in the *Proceedings of the joint meeting of the Association for Literary and Linguistic Computing (ALLC) and Association for Computing in the Humanities*, Oxford University Press, 1991.

An analysis of the two first descriptions applied to an extract of the corpus, the budget discourses, has already been published (Bourque, Duchastel 1988). In that analysis the authors used the nominal determination to delimit the referential scope of some notions of political discourse. Above all, they stressed the relation between the determination patterns and the thematic structure. The theme was then interpreted as the discourse object in itself and the determination as the work of delimitation of meaning. Since then, a diversification in our interests has occurred: from a specific computer-assisted discourse analysis - the Duplessis era - to a more global one - a computer-assisted textual data analysis workbench. Over two years (1987-89), we participated in a research project the aim of which was to design a computer-assisted content analysis system (Duchastel, Dupuy, Paquin, Beauchemin, Daoust. 1991). The main result of this research is twofold: a software development project (Daoust, Dupuy , Paquin. 1989) and the introduction of validating measures and processes into the analysis practice.

This methodological preoccupation has brought us to question certain aspects of the public discourse analysis already done: first, the syntactical description provided by the parser; second, the theoretical background that allows us to extract and qualify data; and, third, the analysis and interpretation guidelines. Having already presented a revised analytical frame of the discourse utterances' thematic structure (Duchastel et al. 1989), we now propose the same for the nominal determination. Before setting out some preliminary analysis results which relate certain determination patterns to semantic classes of words, we will describe our chosen theoretical background of determination and the interpretation guidelines ensuing from it. We will also evaluate the description of determination provided by the parser and describe the extraction and presentation algorithms. In short, this paper constitutes an effort to illustrate the enhancement of research work already performed at high cost and evaluates the contribution of syntactical descriptions to the interpretation process.

## 2. The Theoretical Background

Among all the syntactical approaches, we have chosen the functionalist one because of its orientation towards textual analysis. Unlike other approaches which focus on structural description, the functional grammar leans to a description of the linguistic configuration only if it contributes to the meaning of the text. M. A. K. Halliday, from whom we have borrowed our theoretical background, expresses the axiomatics underlying this approach as follows:

'In general, therefore, the approach leans towards the applied rather than the pure, the rhetorical rather than the logical, the actual rather than the ideal, the functional rather than the formal, the text rather than the sentence. The emphasis is on text analysis as a mode of action, a theory of language as a means of getting things done.' (Halliday 1987, XXVIII)

The author (Halliday 1987, 159 - 175) sets the number of functional elements which could be found in a nominal group specifying the *thing* (the semantic core) at four. First, the element which indicates whether or not some specific subset of the *thing* is intended is labelled *deictic*. Second, the element which indicates some numerical feature of the subset, either quantity or order, exact or inexact, is labelled *numerative*. Third, the element which indicates some quality of the subset or the expression of the speaker's subjective attitude towards it is labeled *epithet*. Fourth, the element which indicates a particular subclass of the thing is labelled *classifier*. These four elements precede the *thing* and are words or sometimes word complexes. What follows the *thing* is either a prepositional phrase or a relative clause embedded in the nominal group and is called a *qualifier*.

All these functional elements constitute modifications or determinations of the *thing*. The determination delimits the meaning extension of the *thing* by defining or specifying it. Determination by a relative clause differs from the other types because 'the characterization here is in terms of some process within which the *thing* is, directly or indirectly, a participant' (Halliday

1987, 167). For our purpose, that is, to characterize the work on the basis of the meaning of an expression produced by other expressions, we will not consider the *deictic* specification. All the other functional elements, including the prepositional phrase, will be grouped together to form the contiguous determination. Determination by the relative clauses is, however, kept separate in order to be compared with the contiguous determination. We will not be able to include in our analysis the distinctions Halliday made among the functional elements because we do not have the benefit of a words list to sub-categorize these elements.

### **3. The Syntactical Description**

The syntactical description of our corpus was obtained by a parser with a bottom-up strategy. First, a morphological category is assigned in context to the words for which a decision cannot be taken out of context; for example *été* could be a noun meaning 'summer' but could also be the past participle of the verb 'to be'. Second, groups are built and their contextual dependencies are assigned. Third, the clauses are identified and their constituents are related. Fourth, the clauses are related. The complexes or phrases, called 'groups' are built with parentheses and prefixed by their tag (GN for nominal group; GD13 for adjectival group). The contextual dependency relations are indexed on both constituents with numbers showing the position of the others and a + or - sign showing the direction; they are also prefixed with a tag (TP for theme and DET for determination).

Figure 1. Illustration of GDSF Surface Description

```
(GN ((TP 1 +))
  (GN ((DET 2 -))
    (D12 () "les" ) (N1 () "résultats" )))
    (C211 () "de" )
  (GN ((DET -2 +))
    (GN ((DET 1 -))
      (GN ((DET 1 -))
        (D12 () "la")
        (N1 () ("conférence" "in0" )))
      (GD13 ((DET -1 +)) (D13 () ("économique" "ec0"))))
    (GD13 ((DET -1 +)) (D13()("mondiale" "us7"))))
```

*'les résultats de la conférence économique mondiale'*

*Results of the world economic conference*

This extract from a nominal group, involved in a thematic relation with the rest of the clause, shows the recursive building of the groups. The description never obscures the text itself, enabling a process of validation and adjustment to be carried on even if the parser is no longer maintained by its author. Given the high cost involved in the development of an improved parser, we thought it would be more 'ecological' and useful to validate and adjust what had been done. It will enable us to work on the analysis itself to produce interpretation guidelines that could be useful to anyone in the community whose text could be parsed.

Pencil in hand, we annotated a printed listing of two hundred pages composed of ten extracts randomly selected from the corpus. A meticulous reading led us to consider the nominal group analysis exact enough to proceed with our project. Only three problems showed up. The first was an incorrect analysis due to initial mistakes in the morphological categorization. In that case, nothing could be done, but the noise introduced in the results was not significant since this kind of mistake was infrequent. Secondly, the relative clauses introduced by the pronoun *que* were misidentified by the parser as completive clauses introduced by the same conjunction. We found that when *que* is immediately preceded by a noun, word or group, it is usually a pronoun. The only flaw in this heuristic is the use of conjunctive locutions built with a noun such as *de sorte que* or *le fait que*;

those nouns are not usually meaningful and will be so interpreted. Thirdly, the parser indicated the alternative between the direct object noun complement and the indirect object itself for the prepositional group. Those occurrences represented twenty seven per cent of the total determination relations assigned. A verbal sub-categorization identifying the maximal number of complements for each verb would be necessary to solve the problem. Since such a list was unavailable, we studied a printed listing of more than a thousand occurrences and we found the following heuristic: when a verb occurs at one of the first three positions before the prepositional group, it is considered an indirect object.

Since the parser Déredec code is orphan and not documented, we opted to work from the outside in LISP.

#### **4. The Data Extraction and Presentation of Algorithms**

To analyze the determination as work on the meaning of an expression produced by other expressions, two operations are needed: the computation of a determination rate for each word by dividing the total frequency of words by the number of times it is determined; and the identification, for each word of the text, of all the words by which it is determined and all the words it determines. Furthermore, we need to differentiate between contiguous and relative determinations. This data enables us to compute a determining/determined rate and a relative/contiguous rate for each word.

For our previous analysis, we extracted the determination data with the extraction mechanism provided by the Déredec package. This mechanism, by means of structural pattern matching, proved to be of great use because of its generality. To make a list of the determined words, a pattern like the following was needed:

(GN ((X) (DET +) (GN ((N1 ((=X)))))))

This pattern should be read as follows: an 'x' which has been categorized N1 (noun) which is determined by an 'x'. Projecting this on the previous example would result in the following list: *conférence, résultats*. We only analyzed this result by making a lexicon out of the list and dividing the frequencies of each word by the total frequency to obtain a determination rate which we correlated with a thematization rate.

To distinguish between contiguous and relative determinations as our new theoretical framework required, the suitable Déredec pattern would be more complex but feasible. The extraction of the related determining part would, however, be difficult if not entirely unsuitable, because the Déredec device was not conceived to extract related information; the only way to overcome this limitation is to project a specific pattern for each word determined as in the following example:

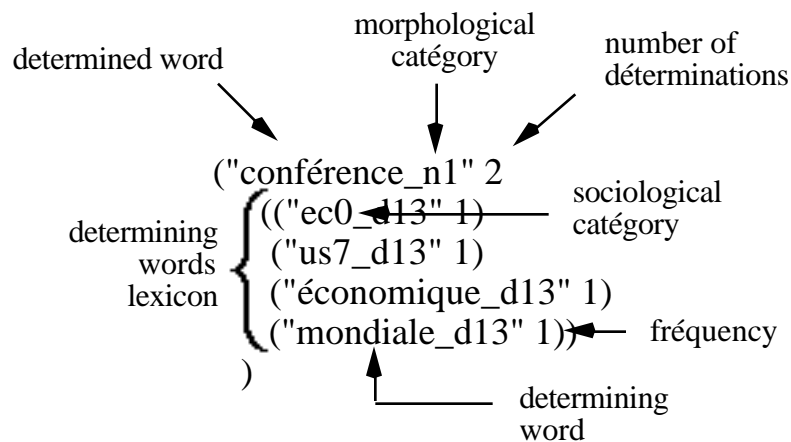
(=GN ((X) (DET +) (GN ((N1 ("conférence"))))))

The list resulting from this pattern projection would be *économique, mondiale*, while the list of words determining the noun *résultats* would be the following: *de la conférence économique mondiale*. The last two words should not be part of the list, but since the Déredec extraction device works recursively on similar embedded structures, it is impossible to indicate at which level of the structure the extraction should be terminated. This is why we decided to construct our own extraction algorithm from scratch, one which also proceeds to the description adjustments we mentioned in the last section.

Two algorithms were constructed : one for the relative clauses and the other for the contiguous determination. The relative clauses algorithm looks first for relative pronouns categories. When it finds one, it proceeds to identify the nominal head of the antecedent nominal group. It then extracts

all the nouns and adjectives in the clause; the embedded relative clauses are not considered. The contiguous algorithm looks first for words determined, that is, identified by a determination dependency relation tag with the minus sign (-). When it finds one, it verifies whether there is a conjunction, a relative pronoun or a verb at the first position before an adverb. If such is the case, the occurrence is rejected. Then, the contiguous algorithm tests the verb for two positions onward. If it is still negative, it jumps the number indicated on the tag and extracts all the nouns and adjectives in the nominal group; the embedded relative clauses and prepositional phrases are not considered. In both cases, the extraction is the same; a database of determined words is built. The illustration below shows one record:

Figure 2. Illustration of a Determination Database Record



For each record, that is, each of the determined words, the following information is given: the morphological category, the number of determinations and a lexicon of the determining words. Each entry of that lexicon presents the determining word, its morphological category and its frequency. When the determined word was assigned a sociological category, another record was created; when the determining word was assigned a sociological category, another entry on the lexicon was created.



## Automated syntactic text description enhancement : determination analysis

We merged the two databases to produce two outputs: a lexicon with certain numerical information; and for each determination, an overview. The lexicon presents the following information:

Figure 3. Example of the lexicon with numerical information

<i>words</i>	<i>i</i>	<i>ii</i>	<i>iii</i>	<i>iv</i>	<i>v</i>	<i>vi</i>	<i>vii</i>	<i>viii</i>	<i>ix</i>
conférences	7	6	7	1	1	0	0	0	0
résultats	42	32	47	7	8	1	1	6	6

- i. the total frequency of the word,
- ii. the number of types determining it contiguously
- iii. the number of tokens determining it contiguously
- iv. the number of types it determines contiguously
- v. the number of tokens it determines contiguously
- vi. the number of types determining it in relative clauses
- vii. the number of tokens determining it in relative clauses
- viii. the number of types it determines by relative clauses
- ix. the number of tokens it determines by relative clauses

With these basic figures, it is possible to calculate several indices: the redundancy rates, the determined/determining rate, the relative/contiguous rate. The determination printed overview combines four lexicons:

Figure 4. Example of an overview for the word 'résultats'

<b>résultats</b>	<b>&lt;- co net: 42</b>	<b>&lt;- rl net: 1</b>	
47<- co	8 -> co	1 <- rl	2 -> rl
1 administration	1 analyse	1 1955	1 cas
2 année-financière	1 colonisation		1 solution
1 année-fiscale	1 comparaison		
6 année	2 compte		
1 années	1 ententes		
1 asbestos	1 évaluation		
1 bons	1 indications		
1 calcul			
1 conférence			
1 considérables			
1 définitifs			
1 économiques			
...			

At this stage, the listing is quite burdensome and does not lend itself to easy consultation: 63,000 entries spread over 14,678 pages. We foresee using the hypertext technology to consult the overview electronically: when a word from one of the four lexicons is selected, the determination overview of that word would appear on the screen and so on.

## 5. The Determination Patterns as Analytical Framework

There is no unilateral interpretation of the semantic work of determination. Determination specifies or modifies the semantic content of words, but we assume, however, that it does so in a different manner depending on the different referential classes of words, for example, functional semantic words as opposed to full semantic words. Our algorithm will produce a great amount of so-called 'relative determination' in the case of conjunctive locutions built on a noun such as 'le fait que' (The fact that), 'en sorte que' (such that). We will not interpret this as determination. Other functional semantic words will be characterized by contiguous determination as should be expected, such as 'augmentation de...' (increase in), 'diminution de...' (decrease in). On the other hand, in the case of full semantic words, the existence and type of determination will be more meaningful in interpreting

the semantic closure of those words. The question raised here is what more or less determination of a certain type means for a given word? An analysis of the Budget speech permitted us to answer this question in part.

We constructed three determination indexes with the information contained in the database. We then identified some particular patterns of determination on the basis of these indexes. Our hypothesis was that those patterns were producing different effects on the word determined. Finally, we correlated these patterns with different classes of words or conceptual categories.

### 5.1. Index Definitions

Three parameters were selected to construct the determination indexes: the relation between relative and contiguous determination; the relation between types and tokens in the determination process; and the relation between determining and being determined.

#### 5.1.1. Relative versus Contiguous Determination

The index is the product of the number of words determining in contiguity divided by the number of words determining in the relative clause. This index permits us to evaluate the predominant type of determination for one word or one class of words. The contiguous determination defines and specifies the *thing*, while the relative determination expands the *thing's* meaning. In the latter case, there is a presupposition that the thing is a participant in some kind of a process. We then presume that the words with a high rate of relative determination have more significance in terms of discourse activity.

### 5.1.2. Redundancy Index

The total number of words (tokens) is divided by the total number of forms (types). This index measures the relative referential stability of words determining other words. An overwhelming redundancy rate would signify that the meaning of one word is rather settled. In contrast, a low rate indicates that the word is still in the process of being defined.

### 5.1.3. Determining versus Being Determined Relation

This index answers the question of the dominant propensivity for a word to determine or be determined. The total number of words receiving a contiguous or relative determination is divided by the total number of words being determined by the former in either a contiguous or relative relation. A tendency to determine rather than to be determined will be interpreted as an indication of greater semantic stability.

## 5.2. Definition of Patterns of Determination

Empirical analysis of the results produced with the indexes previously defined permitted us to define ideal type patterns of determination. These patterns were conceived on the basis of a particular configuration of the indexes previously defined. We must stress that the three proposed patterns are not exhaustive of our empirical observations. As well, the word or categories compiling each pattern do not necessarily match its configuration totally. From the perspective of discourse analysis, we judge that these three patterns were rich in interpretation.

### 5.3. Ideal Type Patterns

#### 5.3.1. The Passive referential Pattern

This pattern is characterized by a high redundancy rate, mostly determined, and a low relative determination rate. These referents are expected objects of discourse which do not require great elaboration. They are passive referents in the sense of semantic sedimentation.

#### 5.3.2. The Structuring Referential Pattern

This pattern reveals a low redundancy rate, mostly determined, with a high relative determination rate. These referents are semantically unstable and thus need to be specified. They are directly or indirectly participants in a process.

#### 5.3.3. The Active Referential Pattern

This pattern also has a low redundancy rate, mostly determining with a low relative determination rate. These referents tend to be stable and are capable of specifying other objects. At the same time, they represent the active elements of discourse.

## **6.- Application of the Patterns of Determination to a Corpus of Political Discourse**

The patterns of determination were applied to a corpus of the Budget speeches in Quebec from 1934 to 1960. Such discourse implies much more than economic or budget considerations. The Government's intentions and orientations are largely discussed throughout. The period under investigation is very rich in transformation from the standpoint of the role of the State and evolution of the constitutional debate, while the political discourse reflects the social alliances typical of Quebec during those years. We were particularly interested in discourse strategies referring to

those alliances and believe that an examination of the determination structure reveals more about them.

The text description mentioned above permitted us to produce a database containing the main features of the Budget discourse's determination structure. The database contains information not only on the nouns and adjectives of the corpus, but also on the sociological categories regrouping those same words. We will be able to compare our present analysis of the determination structure with the one made in our previously cited book. We will proceed in two steps: firstly, we will try to correlate the determination patterns with certain classes of words; and secondly, we will evaluate the possibility of correlating the patterns and certain groupings of sociological categories.

### 6.1. Classes of Words

The database provided us with a lexicon of all the nouns and adjectives with nine types of numerical information, as shown in figure 3. Our three indexes were computed from this information. We can see in figure 5 a short list of words and the actual value for each index. For our purposes, we have worked on a version of the lexicon containing words having twenty or more occurrences.

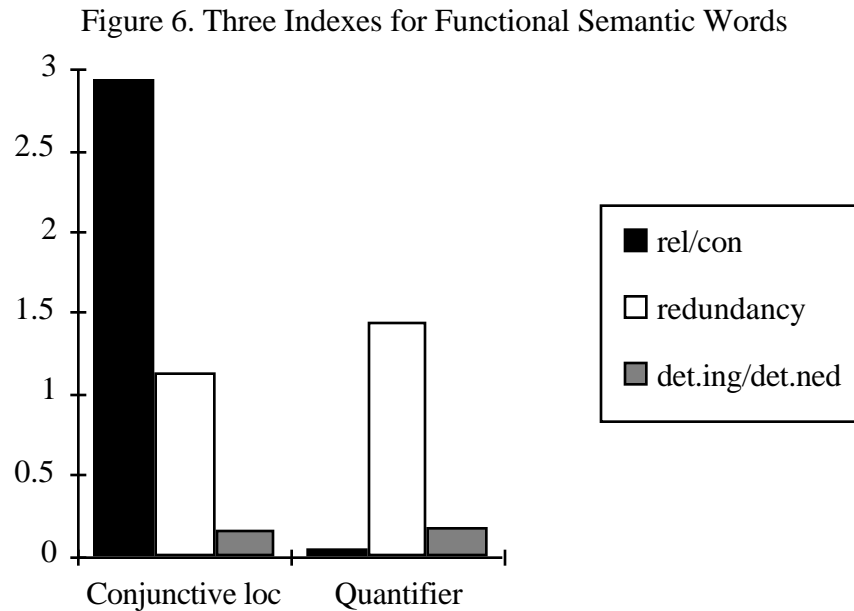
Figure 5. Partial List of Words with their Indexes Values

<b>Words</b>	<b>Freq.</b>	<b>relative/ contiguous</b>	<b>determining/ determined</b>	<b>redundancy</b>
jeunes	52	5.00	2.67	1.07
industriels	26	1.40	1.67	1.00
cultivateurs	103	0.86	0.82	1.00
dette-nette	50	0	0.40	0.00
dépenses	413	0.11	0.41	1.15
revenus	278	0.22	0.68	1.06
dette	101	0.06	0.52	1.00

On one hand we distinguished what we call the functional semantic words from the full semantic words. In the first category, we were again able to distinguish many sub-classes of words. For

instance, let us mention two types of semantic functional words: the noun operating in a conjunctive locution ('de sorte que', 'd'avis que', 'nul doute que') and the quantifier ('accroissement', 'diminution', 'augmentation',...). In the full semantic category, we will distinguish between one group of words referring to actors from another group of words referring to objects. In the first instance, we will find national actors ('canadiens', 'québécois',...), political actors ('gouvernement', 'État',...) or social actors ('jeunes', 'classes', 'professions',...). The other group designates the objects of the Budget speech ('paiement', 'intérêts', 'revenus', 'dépenses',...). We must stress the fact that these categories did not permit us to exhaustively classify all occurrences of nouns or adjectives. They were chosen as significant examples.

Figure 6 shows the behavior of the three indexes for the two sub-classes of functional semantic words. Our first observation is that it is not possible to interpret the results concerning functional semantic words and full semantic words in the same manner. The concordance of a functional semantic word with one predefined pattern of determination will have to be considered differently from the realization of the same pattern in the case of a full semantic word.



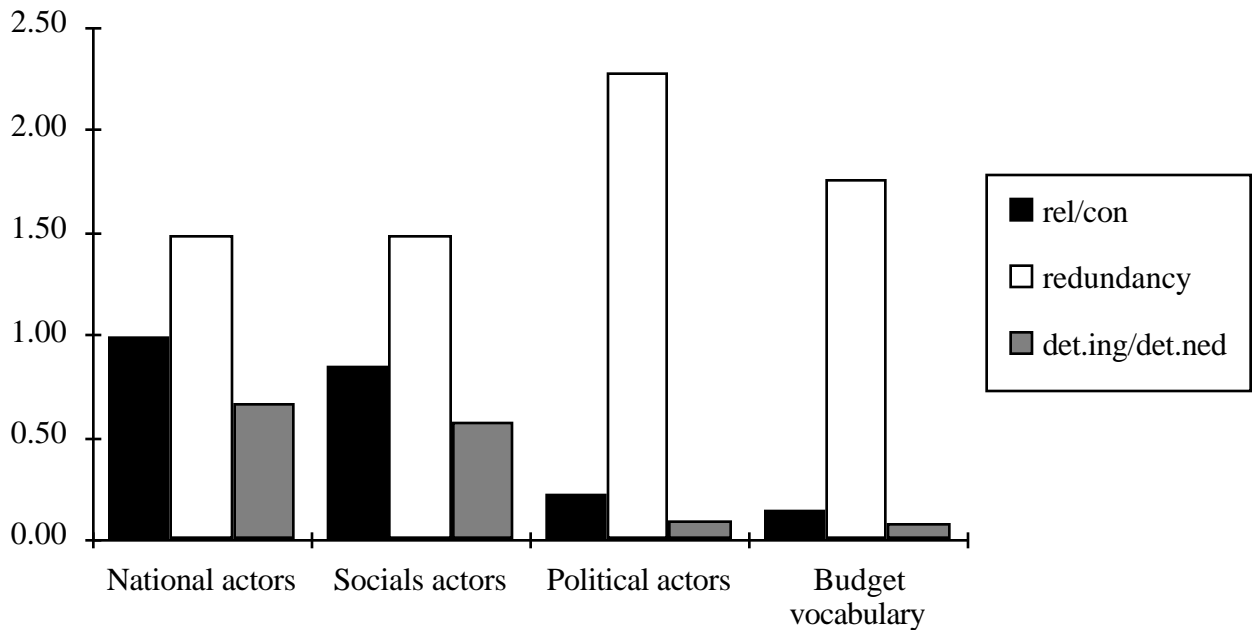
Both classes of functional semantic words have a high general rate of determination. For the conjunction locution's noun, it is expected that the relative determination rate will be very high, the algorithm being incapable of discriminating between this type of construction and a determination structure. In the case of the quantifier, the relative determination rate is very low, meaning that the contiguous one is high. The redundancy rate is also extremely high, which means that the number of specifiers is very limited. This behaviour fits the passive referential pattern, but in a more universal manner than the semantic objects would. In other words, the very existence of these functional semantic words is related to the existence of their contiguous determination ('l'augmentation de quelque chose'). We must also mention that functional words of this type contribute more to the specification of objects than their syntactical structure would suggest. It is their syntactical determinee which is semantically determined by them ('quelque chose qui augmente'). Nevertheless, we may speak of semantic sedimentation with respect to these occurrences.

We will now turn to richer material, the full semantic words. The figure 7 will enable us to identify three different patterns of determination. The semantic class which we named 'actors' tends to realize the active referential pattern with the exception of one sub-category of actors, the political



ones. The actor category has quite a high relative determination rate suggesting that these entities are participating in some action or process. Their redundancy rate is lower than average, which means that the semantic environment is richer. Finally, their ability to determine more than to be determined seems to confirm their full activity in discourse strategies. The sub-class of political actors does not seem to work in line with this pattern but more like the passive referential pattern. The high rate of redundancy is as expected here, the political actors being relatively scarce ('le premier ministre du Québec', 'le gouvernement du Canada',...) but the other two indexes differ greatly from the other types of actors. Is this because the political actors in a political discourse are taken for granted, not requiring greater specification in expansion structures?

Figure 7. Three Indexes for Full Semantic Words



We chose the budget vocabulary to exemplify the passive referential pattern. These words refer to the expected object of this very systematic type of discourse; they have very stabilized referents. We can observe that they have neither a high relative determination rate, nor a high determination

rate. As for redundancy, the pattern is realized as expected; these objects are part of a well-defined world.

Our observations confirm what was already thought with respect to this particular discourse. The more typical economic or budget notions, which were, for the most part, anticipated, are the ones which seem neither to be thematized, nor to be the object of a complex determination pattern. On the other hand, the non-economic categories function as the real agents of the discourse. This conclusion lead us to affirm that this discourse was not only economic, but also political. We will now attempt to carry this further by exploring an analysis of the sociological categories in concordance with their determination patterns.

## 6.2. Analysis of Families of Categories

The preceding analysis of the correlation between determination patterns and classes of words was only partial and designed to show a promising direction for further research. What we wish to do now is still more ambitious and thus requires a thorough explanation. Moving one step farther, we now wish to examine the correlation between determination patterns and groupings of words in families of sociological categories. If determination patterns vary according to semantic classes of words, it should be interesting to study equivalent variations in categories grouping many words on a socio-semantic basis? Our experience tends to confirm the potential in this approach.

The system of sociological categories defines conceptual objects in which different semantic words will be grouped and tagged. Consequently, the statistical behaviour of categories can be explained by intertwined effects pertaining to different classes of words. On the other hand, to the extent that the system has been well-conceived, we may hypothesize that the objects receiving the same tag should be relatively close, semantically speaking.

## Automated syntactic text description enhancement : determination analysis

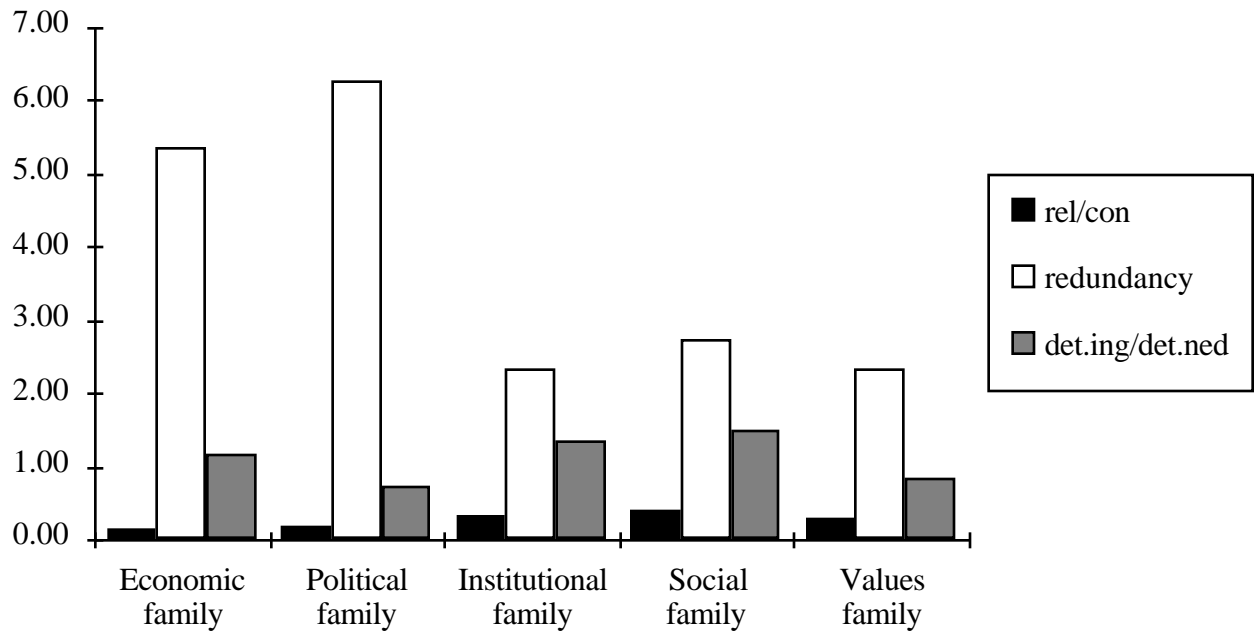
The system is composed of five main families of categories, plus one group of semantic functional words; we shall concentrate on the full semantic categories. The first group refers to economic categories: technical, institutional and managerial aspects of the State Budget (Agriculture, Industry, Finance,...). The political family refers to political institutions and actors (Government, Bureaucracy, Power, Law,...). The next family refers to the civil institutions (Family, Church, Education, Media,...). The social family concerns the social actors and social dimensions of life in society (Classes, Age Groups, Language, Professions,...). The fifth family is value-oriented, defining the main values of discourse (Progress, Liberty, Morality, Tradition,...).

We have identified correlations between, on the one hand, certain families of categories or certain specific categories and, on the other hand, two determination patterns: passive and active referential patterns. Figure 8 indicates that the economic and political categories behave according to the passive referential pattern. The redundancy rates are very high, the relative determination rates are the lowest and the categories tend to be more determined than determining. This is expected considering the fact that these categories refer to the natural objects of a Budget speech. The meaning is stabilized and is manifested through a strong contiguous determination without great variation in the determining elements. Our hypothesis is that the more problematical objects of discourse will receive a more complex determination. This would mean that economic and political objects, even if numerically very significant, do not represent the strategic elements of this discourse. This confirms our previous analysis of the corpus.

If we look at the more specific economic categories, we find that this general pattern is still reinforced. The specific categories (Budget, Economy, Natural Resources, Industry) are largely responsible for the realization of the passive referential pattern. They are the most expected objects of a Budget speech. Being so, their specification or definition has a very limited scope. We may also add that being the objects of discourse, they do not seem to be associated with the process aspect of determination. In the political family of categories, we note that the categories most typical

of the passive referential pattern are Government and State. Again, these categories fall into the class of expected and passive objects of discourse.

Figure 8: Three Indexes for Families of Categories



The two families of categories referring to social institutions and social actors are more in line with the active referential pattern. They have a greater relative determination rate, which means a greater expansion of determination. Their redundancy rate is much lower indicating a wider range of determining elements. Their capacity to determine is much greater, evidencing an active role in constructing the meaning of discourse. If we look at some specific categories, we find that the central social actors of the Duplessis Regime are Party, Family and Relation of Work. Social categories such as Class and Profession are also be identified as participating in the active referential pattern.

The family of value categories seems to behave in the same manner as the institutions and actors. The relative determination rate is high and the redundancy rate is low, but the propensity to determine is lower than in the active referential pattern. These categories would be more in

accordance with the structuring referential pattern. From this we are able to conclude that the values in the discourse are not static. They are active elements of discourse strategy, but they do not have the stability needed to determine other objects.

### 6.3. Analysis of specific categories

To conclude the preceding analysis, it seems that two main groups of categories can be defined: one that refers to objects of discourse and one referring to actors. Some families of categories contain more objects than actors and vice versa. The objects tend to be treated as passive referents and actors as active referents from the standpoint of determination. This section will examine the behavior of some specific categories which either reinforce the general pattern of one family or which oppose it. Figure 9 gives the indexes for each category that presented here.

Figure 9. Three Indexes for Specific Categories

	<b>rel/con</b>	<b>det.ing/det.ned</b>	<b>redundancy</b>
<b>Economic average</b>	<b>0.16</b>	<b>1.15</b>	<b>5.37</b>
Agriculture	0.32	2.80	3.19
Transport and Communication	0.28	0.84	5.39
<b>Political average</b>	<b>0.17</b>	<b>0.72</b>	<b>6.25</b>
Law	0.21	0.49	3.56
Power	0.21	0.50	3.90
Government	0.11	0.56	12.49
Elected Members	0.13	0.27	4.42
<b>Institutional average</b>	<b>0.33</b>	<b>1.34</b>	<b>2.34</b>
Relations of Work	0.48	1.19	1.41
Education	0.17	1.05	4.51
Culture	0.13	0.44	4.16
Political Party	1.44	3.05	1.77
Family	0.31	1.02	2.00
<b>Social average</b>	<b>0.42</b>	<b>1.51</b>	<b>2.72</b>
Social classes	0.35	0.63	2.83
Language	0.02	2.37	4.18
<b>Values average</b>	<b>0.28</b>	<b>0.85</b>	<b>2.34</b>
Progress	0.06	0.36	6.61
Morality	0.63	0.70	1.59
Religion	0.38	1.14	1.50
Effort	0.40	0.61	2.31
Tradition	0.30	0.96	1.62
Justice	0.09	0.86	2.90
Liberty	0.22	1.15	1.43

We have seen that the economic categories are mainly passive, but there are two exceptions which conform more to the active referential pattern: Agriculture and Transport and Communication. These categories represent the principal topics of the Duplessis discourse. The Duplessis Regime was founded on a strong alliance with the agricultural class and local bourgeoisie. As a result, many economic policies were developed in order to please those classes. One such policy was the development of natural resources necessitating the expansion of transport and the construction of means of communication. These elements were highly strategic in Duplessis discourses. We can thus understand why Agriculture and Transport and communication were treated as active elements rather than passive. The discourse operates on these elements more than it does on other economic entities.

We have already mentioned that the political category Government constitutes a good example of the passive referential pattern. Although of great importance as an element of discourse, this category seems to be objectified. It is a given object which requires no further elaboration and has a strong contiguous and redundant determination pattern. We speak of the Government of Québec or Canada; no more expansion is given to the notion. The same may be said with respect to Elected members of Parliament.

On the other hand, if we look at categories such as Power and Law, we will find a different pattern: they are highly developed in relative determination. This may be because of the general nature of these categories which favours broader explanation. In addition, the action of Government is founded on these notions, which would explain their active behaviour pattern.

The social institution categories are correlated to the active referential pattern. We hypothesized that these categories were closely associated to the action of the Regime. There are three institutions that realize this pattern: Political party ('l'Union Nationale' de Maurice Duplessis), Relations of Work (the Regime is very preoccupied with the relation between Capital and Work) and Family (which represents the fundamental institution of traditional society). Church is also, to a certain extent, a good example of this pattern. The categories all refer to dynamic actors of the system: the Party, the Capital, the Workers and the Family.

In contrast, Education and Culture, both social institutions, do not conform to the active pattern, but rather to the passive one. They appear as passive objects of discourse and their determination is mostly contiguous and redundant. They do not appear to participate in a process from the standpoint of determination. Our interpretation of this phenomena is that these objects are stabilized in this discourse and, even if they are largely present in discourse, they do not give way to modification.

The social family of categories is mainly correlated with the active referential pattern. This is due to the very strong participation in an associated pattern of three categories: Persons, Classes and Liberal Professions. This associated pattern is the structuring referential pattern. The difference from the active referential pattern is the propensity to be more determined than determining, although the active dimension is very present. These findings confirm the very important status of the petty bourgeoisie in the Duplessis Regime and also indicate that the class question was very central in this political discourse.

In contrast, two categories in the same family (Community and Language) are very poor in relative determination and tend to act like passive referents. Again, these elements of discourse seem to be very stable. They have a strong contiguous and redundant determination. Although they are important objects of discourse, they are not treated as active participants in a process.

The family of Value categories cannot be considered on the same level as the other families, being neither objects nor actors of discourse. They are part of the evaluative process of every discourse. The fact that the values are highly developed in relative determination and that the redundancy rate is low indicates that these objects are in the process of being defined. They receive much more determination than they give to other objects. The values of discourse seem to be in the process of construction.

Two sub-categories of values conforms completely to this pattern: Traditional values (Tradition, Religion), Disciplinary values (Effort, Sacrifice, Morality). The Duplessis discourse is very preoccupied by class relations and social order; the preceding values are the ones that are invoked to deal with these problems. Not surprisingly, they are the most active values. On the other hand, some values like Progress, Liberty and Justice tend to be of the more passive type. These values are objectified and operate as stabilized elements of discourse.



## **7. Conclusion**

The work we have done to date has been preliminary in nature. We are now able to identify certain limits and possibilities. Firstly, the necessity of simplifying the model of determination in order to deal with the phenomena of large textual bases introduces some impoverishment of the functional approach. Secondly, we will need a much more complex theory of semantic classes of words. We have only illustrated the kind of distinctions which could be made between some of these classes. Thirdly, jumping to sociological categories takes us farther away from the original texture of discourse. We must then modify our reasoning of the problem. The patterns defined here have served to let us identify general patterns of functioning from one family to the other. Variations from the general pattern should now be more deeply analyzed to evaluate the effectiveness of this method.

Overall, we can conclude by stating that this determination analysis has tended to confirm findings made in an earlier stage of our research. The general behaviour of our categories is easily interpreted within the principal currents identified at that stage.

References

Bourque, G., Duchastel, J. (1988), *Restons traditionnels et progressifs, Pour une nouvelle analyse du discours politique: Le cas du régime Duplessis au Québec*, (Montréal).

Daoust, F., Dupuy, L., Paquin, L.-C. (1989) 'ACTE : workbench for knowledge engineering and textual data analysis in the social sciences', *Proceedings of the Fourth International Conference on Symbolic and Logical Computing (ICEBOLA)*, Dakota State University, (Madison), 122-136.

Duchastel, J., Dupuy, L., Paquin, L.-C., Beauchemin, J., Daoust, F. (1991), 'The SACAO Project: using computation toward textual data analysis in the social sciences', to be published in *Content, Concepts, Meaning. Advances in Computing and the Humanities*, (Greenwich Conn.), vol 3-4.

Duchastel, J., Paquin, L.-C., Beauchemin, J. (1989), 'Automated syntactic text description enhancement : thematic structure of the discourse utterances analysis', *The Dynamic Text, 16th International Congress of the Association for Literary and Linguistic Computing (ALLC)*, Centre for Computing in the Humanities, (University of Toronto), June the 8th.

Halliday, M. A. K. (1987), *An Introduction to Functional Grammar*, (London).

Plante, P. (1979), 'GDSF, Grammaire de surface du français', Doctoral dissertation (Université du Québec).