DANDELION

Discourse functions and Discourse Representation:
An Empirically and Linguistically Motivated,
Interdisciplinary-Oriented Approach to Natural Language Texts

ESPRIT Basic Research Project 6665

Deliverable R1.2.2b

Theme, Discourse Topic and Information Structuring

Julia Lavid

September, 1994

CLS Tilburg/ Nijmegen

University of Edinburgh GMD-IPSI Darmstadt Universität Saarbrücken

Universidad Complutense de Madrid

Document ID : R1.2.2b
Issue date : September 1994
Version : final
Status : public
WP number : WP1: Empirical Studies
Distribution : Dandelion
Location : Universidad Complutense de Madrid
Keywords : Theme, information structuring, text types

Theme, Discourse Topic and Information Structuring

Julia Lavid

Universidad Complutense de Madrid

Abstract

The present deliverable is an investigation into the role of various phenomena responsible for the structuring and development of information in discourse and the contextual factors which constrain their deployment. These phenomena, studied and redefined in Deliverable R1.2.1, include Theme, Topic, Focus and Focus-shift rules. This paper refines and tests the hyothesis presented in our R1.2.1, according to which certain features of the communicative situation - i.e. those that determine the text type - prespecify the selection of particular lexico-grammatical and textual realization options. The result of the empirical analysis presented here is a specification of the factors governing information structuring, empirically tested on a variety of text type examples. The paper also outlines a dynamic prespecification of the process of information development with a view towards computational implementation.

Contents

1 Introduction

As studied in Deliverable R1.2.1, phenomena like Theme, Given, Topic and Focus have been the object of research in the fields of Linguistics and Computational Linguistics where different authors have used the terms sometimes interchangeably and with such ranges of overlap and confusion that make their study a hard task for the researcher. The aim of Deliverable R1.2.1 was to clarify these notions by mapping out their functionalities and extracting their commonalities in such a way that they are useful for both the linguistic and the computational communities. In particular, it attempted to provide operational redefinitions for these notions and explored the constraints imposed by several contextual factors on the discourse phenomena.

The aim of this deliverable is to refine and to test through empirical text analysis the hypothesis presented in Deliverable R1.2.1, according to which several factors from the communicative situation responsible for the definition of text types prespecify the selection of specific options in the phenomena studied. For that purpose, we have analysed a corpus of seventy texts -grouped according to discourse purpose and subject-matter- in terms of chaining strategies, theme and focus selection and focus-shift candidates to validate the hypothesis for their use illustrated in Figure 7 of Deliverable R1.2.1.

This paper is organized as follows: first, we summarize the definitions provided in the literature for the notions of theme, given, topic and focus, and present the operational redefinitions provided for theme, chaining strategies, focus and focus-shift candidates as originally proposed in R1.2.1. Second, we study the behaviour of these phenomena in specific contexts of use, thus arriving at a preliminary specification of factors. Third, we present the results of empirical text analysis of a corpus of seventy texts distributed in groups of ten according to discourse purpose and subject-matter. Finally, we outline a prespecification of the process of information development with a view towards computational implementation.

2 Summary of terms and definitions

As explained in R1.2.1, it appears that the concepts of focus, given-new, theme-rheme and topic describe related, albeit different phenomena, since they are the result of different processes.

Focus in Artificial Intelligence is a constructive category, resulting from the process of content selection by which the discourse participants center their attention on some subpart of the information contained in the knowledge base. In Linguistics, however, focus is a sentence-level category, which captures contextually non-bound (irrecoverable) information predicating something about the "given", recoverable, contextually bound part, that is the *topic* or the *theme* of the sentence. Therefore, it is the result of the process of information distribution which reflects the speaker's estimates of the listener's familiarity with the subject matter.

Theme is that element from which the clause "departs", according to Halliday, and, as such, it is the result of a process by which the speaker presents his material from a particular perspective or point of view. Theme normally coincides with Given, but not necessarily, since they belong to two different types of structuring. However, in the Prague School treatment of the Theme-Rheme structure, thematic elements are those which carry the lowest degrees of communicative dynamism (i.e. which are 'given') and rhematic elements are those which carry the communication forward, thus conveying new information.

Topic has been traditionally treated as a category which refers to the recoverable, contextually-bound part of the sentence, distinguishing thus a dichotomy of topic and focus. However, as a discourse category, the notion of topic captures the same notion studied under the label of focus in computational linguistics. Table 1 below summarizes these points:

Table 1: Summary of terms and definitions

COMPUTATIONAL LINGUISTICS	LINGUISTICS	DEFINITION
Focus	Topic	recoverable information
	Theme1	perspectival departure
	Theme2 (= Given)	low communicative dynamism
	Focus	irrecoverable information

In an attempt to integrate the findings in both research fields and make them useful for a theory of discourse structure, we proposed in R1.2.1 to offer operational redefinitions for these notions adopting a discourse perspective which would allow us to extract their functionalities from their behaviour in real texts. We present these redefinitions here again briefly in the following subsections:

2.1 Theme and chaining strategies

The disparity in the treatment of theme in the literature led us to propose a redefinition of this category which captured the two main discourse functions which we have identified so far: the **bridging** function and the **guiding** function. We, therefore, provisionally proposed the following discourse characterization of theme (Lavid & Hovy, 1993):

> **DEF**: Theme is that element that informs the listener as discourse unfolds how to relate the incoming information to what is already known, providing the attachment point to new information pieces from the discourse. More specifically, theme acts as a signpost for the reader of a specific text structuring or chaining strategy.

The first part of the definition illustrates the bridging function of theme and explains why it usually co-occurs with Given and why it is considered the "point of departure" of the clause. The second part captures the guiding function of theme in the listener's journey through the text, which tends to be organized by means of specific lines of development. Theme acts in these developments as a signpost for the reader of a specific chaining strategy selected by the writer to organize his/her text. We proposed in R1.2.1 to call these lines of development *chaining strategies*, and we proposed the following definition for them:

> **DEF**: Chaining strategies are devices used by the writer to steer his/her text along a specific line of development or frame with the purpose of achieving a maximally profitable text organization, in view of the discourse purpose and the subject matter

Chaining strategies are like maps to show the reader a destination and to suggest to him/her a route throughout the text. Chaining strategies include temporal and spatial developments, general to specific, characterization, through- or counter-argument patterns, etc.

The textual role of theme in these developments is to act as a signpost for the reader, to mark the temporal or spatial development, to bring out the point of elaboration, of contrast. Themes identify these developments so that the reader can decode the chaining strategy used by the writer, acting, at the same time, as the bridge between the information already presented and the new information portions, thus helping to segment the text into units of various sizes. Among the chaining strategies identified so far through empirical text analysis, we include the following: characterization, spatial, temporal, participant, sequential, through- or counter-argument.

Table 2 below summarizes the main chaining strategies found in the texts analysed together with the most frequent thematic selection options, according to our initial hypothesis later confirmed by empirical text analysis.

Table 2: Chaining strategies and thematic selection

CHAINING STRATEGY	THEMATIC SELECTION
Characterization	Topical
Spatial	Location
Temporal	Time
Participant	Topical
Sequential	Time Process
Through- or counter-argument	Conjunctive

2.2 Focus and focus-shift candidates

With respect to focus, whereas in Linguistics this notion has been traditionally associated with "irrecoverable information" (Halliday, 1967; Chomsky, 1971; Sgall, 1979; Sgall et al., 1973; Rochemont and Culicover, 1990), in Artificial Intelligence it has been used as a constructive category which intuitively captures what a text, or part of a text is about, thus being similar to the notion of discourse topic.

The progression of focus, however, has been studied only in Artificial Intelligence mainly because one cannot simply string together sentences arbitrarily and expect smooth text flow. Therefore, some research efforts in AI with respect to focus have concentrated upon the design of a set of rules which dictate the legal or permissible focus moves during text generation. McKeown's proposal was later modified and enhanced by McCoy and Cheng's notion of a *discourse focus tree* as a mechanism for controlling focus shifts in discourse (McCoy and Cheng, 1990). Their hypothesis is that a focus tree is constructed and traversed by the participants as the discourse progresses, one node being visited at a time. Each node in the tree is something which is being talked about in the discourse and points to an entity from the knowledge base. The type of the currently visited node -*object, setting, action*, and *event* - determines what entities from the knowledge base are in focus, and generates expectations about what may be said next in a discourse, which must generally come from the highlighted set of knowledge. A change in the focus of attention corresponds to changing the currently visited node.

In an attempt to integrate this notion of focus with the one proposed in Linguistics, we repeat here the provisional definition proposed in R1.2.1 [1]:

> **DEF:** Focus is the locus, within each clause, of principal inferential effort. It is those concept(s) on which the speaker wants the hearer to spend the most thought.

The speaker determines the focus at a given point in discourse by identifying the concept(s) contained in it which are most central to the mental processing (the inferences) required by the listener to understand the clause, and to follow a coherent progression of information. (Of course, the speaker has to take into account what information has already appeared in the discourse, and also his or her overall communicative intentions). By signalling these central concepts as the focus, the speaker is saying: “Concentrate your thoughts around this. This is where I want your thinking to start.”

Focus selection is taken for granted at the beginning of a discourse, and from there the inferential mechanism proceeds by expanding the currently visited node, or one of its previously unexpanded children, or a new expansion of its ancestors. However, this mechanism does not explain the preference of one focus node type over another at the beginning of a discourse. The question is: why should we focus on an object, or an action, or an event in the first place?

It is our feeling that contextual factors such as the *discourse purpose* and the *subject-matter* , among others, contribute to the focus selection process. We explain this influence in the next section.

[1] The definitions of Focus and Theme presented in this paper were originally proposed in Lavid & Hovy (1993). Focus, Theme, Given and Other Dangerous Things. Working Paper.

3 Contextualizing phenomena: a preliminary specification of factors

Our initial hypothesis -proposed in R1.2.1- was that several factors from the communicative situation would constrain the selection of the discourse phenomena which govern the structuring and development of information in discourse. We identified two main factors: *discourse purpose* (and its congruent realization into a *text type*) and *subject-matter* . The figure which we enclose at the end of this document (Figure 7 at the end of Deliverable R1.2.1) illustrates graphically the correlation between contextual factors and discourse phenomena. This can be explained as follows:

If the discourse purpose is expository and the subject-matter deals with whole classes of objects or generic concepts, then these two factors typically determine: a) the selection of a specific (global) chaining strategy (one of the most typical chaining strategies is characterization, although there are other possibilities) realized by the selection of topical themes as markers of that chaining strategy, b) the selection of an object focus node and of an object focus-shift candidate.

If the discourse purpose is descriptive and the subject-matter deals with place relations, these two factors will constrain: a) the selection of a (global) spatial chaining strategy, signalled by the selection of (global) locative themes, b) the selection of a setting focus node and of an object focus node as the most likely focus-shift candidate. If the subject matter deals with objects and people, e.g., the chaining strategy selected will be characterization, the focus node will be an object and the most likely focus-shift candidate will be attributes of the object.

If the discourse purpose is narrative and the subject-matter deals with events and participants, the most typical global strategy selected by the writer will be a temporal one. At a more local level of analysis, a participant strategy is often used as a minor chaining strategy. The temporal strategy is typically signalled by the selection of temporal themes, and the participant strategy by topical ones. Focus nodes tend to be event or object focus nodes and focus-shift candidates tend to be actions.

If the discourse purpose is instructive and the subject-matter deals with steps in a procedure, then these two factors will determine the selection of a sequential chaining strategy, signalled by temporal and/or process themes, depending on the genre. The focus nodes will probably be actions, and the focus-shift candidates will be either actions in a sequence or the purpose of the action, or the object of the action, to name a few.

If the discourse purpose is argumentative, and the subject-matter deals with complex facts and ideas, these will determine the selection of two typical chaining strategies: through-argument and counter-argument, both signalled by the selection of conjunctive themes as markers. The most typical focus node category will be facts and the focus-shift candidates will be relations between facts.

In order to validate this initial hypothesis, we performed empirical text analysis on a corpus of seventy texts. We explain our analysis and the results obtained in the next section.

4 Results of empirical text analysis

The aim of this section is to present the methodology and results of empirical text analysis carried out on a corpus of seventy texts. Ten texts were collected for each of seven 'text types'. The 'text types' were in priciple defined by the discourse purpose, but were further subclassified by subject matter. The selection of each specific group was done as follows:

Expository texts:

Ten encyclopedia entries. We initially selected a more varied sample of texts, including definitions found in encyclopedia entries, essays and abstracts with an expository purpose and dealing with generic concepts and whole classes of objects. We found that there was a wide variety of possible chaining strategies, including characterization, compare and contrast, general/specific, etc. The chaining strategy found in all encyclopedia entries was that of characterization. Therefore, and for the purpose of presenting a systematized and clear correlation between contextual factors and discourse phenomena, we restricted our sample to ten encyclopedia entries.

Descriptive texts:

Ten texts which described place relations of some kind; these included mainly geographical descriptions of countries and cities. All of these texts presented a spatial chaining strategy.

Ten texts describing objects and people, including pen portraits. The chaining strategy selected here was characterization. The difference between these texts and the expository ones lies in the subject matter: the texts selected which reflected an expository discourse purpose dealt with whole classes of objects, whereas the texts selected which reflected a descriptive discourse purpose dealt with individual objects or people.

Narrative texts:

Ten historical passages including biographies and the history of two cities. The ten texts selected displayed a major (global) temporal strategy, and some texts also selected a strategy centering around participants on a minor (local) level of analysis.

Instructive texts:

Ten instructional passages from instruction manuals and ten recipes. The results of an initial analysis -where we had selected ten texts including recipes and instruction manuals at random- showed disparity in the selection of themes to signal the same sequential chaining strategy. For this reason, we decided to extend the sample to twenty texts (ten instructive sections from recipes and ten from instruction manuals) to test the hypothesis of different theme selection options to signal the same chaining strategy.

Argumentative texts:

Ten texts with an argumentative discourse purpose, including essays, discussion sections from scientific articles, and declarations. All the texts selected presented a clear chaining strategy either of through-argument or of counter-argument. These two strategies are not the only ones displayed by argumentative texts, but they tend to be the most frequently used on a global level of analysis.

In the following sections, we describe in detail the results of empirical text analysis for each of the ten groups of texts selected according to discourse purpose and subject-matter, and explain the behaviour of each of the phenomena studied above. The results of these analyses will serve as validation of the initial hypothesis presented in section 3 above.

4.1 Expository texts

We initially started with twenty texts reflecting an expository discourse purpose and dealing with whole classes of concepts. Our hypothesis was that these two factors would determine the selection of a chaining strategy and the most likely category of focus node. With respect to the chaining strategy we found that out of all the possible chaining strategies which organize a text on a global level (e.g. general/specific, exemplification, characterization, compare and contrast, etc.), the strategy of characterization was a very common one. Therefore we restricted our sample of expository texts to ten encyclopedia entries organized on a global level by means of a characterization strategy. We describe below the results of the analysis with respect to these phenomena:

Chaining strategy: our initial hypothesis was that the selected texts would signal the characterization strategy by selecting unmarked topical themes. In order to test this hypothesis, we searched for references to the concepts or to the characteristics of the class of concept or object (including attributes and parts), and checked whether each reference was marked by the selection of a topical theme. The results of our analysis are shown in table 3 below. As the figures show, the correlation between the total number of references to the concepts or to the characteristics of the class of concept or object in the sample (total: 102) - which constitutes the strategy of characterization- and the number of topical themes as markers of that strategy (total: 86) is very high. This fact confirms our hypothesis according to which unmarked topical themes act as a signal of a characterization chaining strategy.

Table 3: Distribution of thematic options in expository texts

		THEMATIC SELECTION					
TEXT Nº	Nº of references	TOPICAL	LOCAT.	TIME	PROCESS	CONJUN.	OTHERS
Nº 1	7	6	0	0	0	1	1
N º2	11	9	0	0	0	2	2
Nº 3	8	4	0	0	0	0	4
Nº 4	13	13	0	0	0	0	0
Nº5	14	12	0	0	0	0	3
Nº6	12	12	0	0	0	1	0
Nº7	7	5	0	0	0	0	2
Nº 8	16	13	0	0	0	0	3
Nº 9	6	5	1	0	0	1	0
Nº 10	8	7	0	0	0	1	1
TOTAL	**102**	**86**	**1**	**0**	**0**	**6**	**16**

Focus nodes: all the expository texts we analysed selected an object focus node category to highlight a particular set of knowledge base entities. Object nodes cause an object to be focused in terms of the specific set of properties that object may have. An object node is usually directly given in the discourse in the sense that it is most often inferred because of the explicit noun mention of an object in the discourse. Here focus node and theme selection tend to conflate in their linguistic realization when first mentioned in the discourse. As to the focus-shift candidates, all the texts analysed presented attribute nodes as candidates to develop the information.

Example (1) below illustrates the phenomena studied above:

> DRONGO, family name DIRCRUDIDAE, any of 20 species of Olf World woodland birds (order Passeriformes), often with forked tails. Drongos frequently attack much bigger birds (e.g. hawks and crows) that might hurt their eggs or young; innocuous birds (such as doves and orioles) nest near drongos and thereby gain protection.
>
> Most drongos are 18 to 63,5 centimetres (7 to 25 inches) long and glossy black, sometimes with white on the head or underparts (sexes alike); the eyes, in most, are fiery red. Some are crested or have head plumes, and the tail is usually long and forked, with out-turned corners. The tail of the South-east Asian racket-tailed drongo (dicrurus paradiseus) bears 30 cm (12-in) "wires", outer feathers that are unbranched for most of their length and carry rather large vanes at the ends.
>
> Drongos range from Africa to central Asia, Australia, and western Pacific islands, inhabiting forest, open country, and gardens. They feed like flycatchers or shrikes, taking large insects and termites. Their voices are lound mixtures of harsh and sweet sounds; some species, like the racket-tail, are good mimics. The nest is a flimsy basket that seems too small for the brooding bird.
>
> (Drongo. The Encyclopaedia Britannica. Micropedia, vol. 4. 445. 1991)

The chaining strategy of characterization is signalled by the selection of topical themes which refer to drongos as a class of birds, or to their characteristics or attributes. There are eleven references to drongos or their characteristics, and they are all marked by topical themes.

The focus node would be object (*drongos*) and the focus-shift candidates are attribute nodes which occur as subordinate to the object node. We can mention here attribute nodes such as activities, physical appearance and natural habitat as the main ones under which other subordinate attributes can be grouped.

4.2 Descriptive texts

We studied twenty texts reflecting a descriptive discourse purpose and dealing with two main kinds of subject matter: place relations, on the one hand, and objects and people, on the other. Our hypothesis was that these two factors would determine the selection of a specific chaining strategy (signalled by thematic selection options), and of the most likely type of focus node. Accordingly, we selected ten texts dealing with place relations and ten texts dealing with objects and people.

4.2.1 Place relations

Ten texts were selected describing place relations of some kind. We found that the texts presented spatial references to a central place of observation, or relations between different location points with respect to each other. Our hypothesis was that both the discourse purpose (in this case, descriptive) and the subject-matter (place relations) would determine the selection of a specific global chaining strategy (signalled by specific thematic options) and the selection of a type of focus node. We describe below the results of the analysis with respect to these phenomena:

Chaining strategy: our initial hypothesis was that these texts would be organized according to a spatial strategy which would reflect the writers' tendency to steer their texts along a spatial frame when having to describe place relations of some kind. In order to test our hypothesis, we searched for references in the texts to different points along a spatial line (spatial scopes), and checked whether each spatial scope (defined as new locations along a spatial line of development) was marked by the selection of a locative theme or not. The results of our analysis are shown in table 4 below. The table captures the number of spatial scopes present in the ten texts selected, and the selection of different kinds of themes to mark those spatial scopes. For example, text 2 in the table -reproduced in example (2) below- presents nine spatial scopes all of which are signalled by a locative theme.

Table 4: Distribution of thematic options in descriptive texts (spatial relations)

THEMATIC SELECTION

TEXT Nº	Nº spatial scopes	LOCAT.	TOPICAL	TIME	PROCESS	CONJUN.	OTHERS
Nº1	5	5	0	0	0	0	0
Nº2	9	9	0	0	0	0	0
Nº3	2	2	0	0	0	0	0
Nº 4	4	4	0	0	0	0	0
Nº5	6	5	1	0	0	0	0
Nº6	5	5	0	0	0	0	0
Nº7	9	8	1	0	0	0	0
Nº8	9	8	1	0	0	0	0
Nº9	3	3	0	0	0	0	0
Nº10	4	4	0	0	0	0	0
TOTAL	**56**	**53**	**3**	**0**	**0**	**0**	**0**

The total number of of locative themes (53) selected to signal the spatial strategy in the ten texts analysed almost always correlates with the number of spatial scopes (56). This means that each time there is a movement in space in the texts, locative themes are the most likely candidates to act as signposts to carry the reader through the text.

Focus node: all the spatial descriptive texts we analysed selected a setting focus node category to highlight a particular set of knowledge base entities. In its original proposal this category of node is usually inferred via bottom-up inference from the presence of a number of seemingly unrelated objects whose common thread is often that they are part of a setting (McCoy and Cheng, 1990). However, we have extended this definition to refer to the highlighting of different points in a scenery. In this latter case, the focus node can also be inferred in a top-down fashion from the presence of a series of objects (places) which together configure a spatial location. Once the setting has been inferred, it constrains what further objects may be introduced. The expected children of this type of node are usually objects (including places). All the texts analysed selected this kind of focus node and shifted to object focus nodes to make the information progress.

Example (2) below illustrates the phenomena discussed above:

TOURING FROM CARDIFF

Here we shall concentrate on the fascinating variety of attractions withing a hour' s drive of Cardiff -- too numerous to list in full so we offer you a sample:

Immediately to the NORTH of Cardiff lie the industrial valleys of South Wales made famous throughout the world by films such as How Green was my valley. Here small working communities lie hemmed in by green hills. Here the archeology of the industrial revolution abounds in vivid contrast to the breath-taking natural beauty you enjoy when negotiating the steep winding roads that go "over the top" from one valley to the next, such as the road over the Rhigos, from Maerdy to Aberdare. Beyond the valleys lie the mountains of the Brecon Beacons National Park. Here the sheep predominate. Man-made lakes that supply water to the population of South East Wales lie between the mountains. Stopping in the car park at Storey Arms one can relax and enjoy the beautiful surroundings or, more energetically, follow the footpath to the top of 2,907 ft high Penyfan. Visit the mountain centre at Libanus or the interpretive centre in the Garwnant Forest, for an insight into the wildlife of the area.

Immediate to the WEST of Cardiff lies the Vale of Glamorgan with its villages, leafy lanes, farmland and heritage coastline. Seaside resorts start just beyond the city boundary with the Edwardian-style resort of Penarth, complete with pebble beach and pier. A few miles west lies the livelier type of resort, Barry Island, with all the joys of the pleasure park and a wide, sandy beach. And beyond lie more beaches-Fontygary, Llantwit Major, Southerndown, Ogmore. In the Vale of Glamorgan you will find a variety of seaside catering for most tastes. Behind those beaches lie unspoilt country with historic villages and delightful pubs. Well within the hour' s drive lies Porthcawl another lively resort complete with funfair.
Just about that hour' s drive away to the west lie some of Britain' s loveliest coastline, the Gower Peninsula, with its wide sandy beaches and great headlands.

> To the EAST we have the beautiful wooded Wye Valley, with the remains of a famous abbey on the banks of the river at Tintern, the Roman remains of Caerleon, Caerwent, Raglan Castle, Monmouth and the Forest of Dean.
>
> (The Cardiff Guide, Cardiff City Council)

The spatial chaining strategy is signalled most of the times by the selection of locative themes (underlined in the text) to mark the reference to a new point in a spatial line: *Immediately to the NORTH of Cardiff; Beyond the valleys; Immediately to the WEST of Cardiff; A few miles west; and beyond; Behind those beaches; Well within the hour's drive; Just about that hour's drive away to the west; To the EAST.*

The focus node would be a setting of Cardiff, and its focus-shift candidates are the different objects (*the industrial valleys of South Wales, the mountains of the Brecon Beacons National Park, the Vale of Glamorgan, Barry Island, the beaches of Fontygary, Llantwit Major, Southerndown, Ogmore, unspoilt cuntry, Porthcawl, the Gower Peninsula, Wye Valley*) which together configure the description of Cardiff.

4.2.2 Objects and people

With respect to the description of objects and people, we selected ten texts describing objects and people. As above, our hypothesis with respect to this kind of texts was that the discourse purpose (descriptive) and the subject-matter (object and people) would determine the selection of a specific global chaining strategy (characterization) -signalled by topical themes- and the selection of an object focus node category. We describe below the results of the analysis with respect to these phenomena:

Chaining strategy: in order to test this hypothesis we searched for references to the object itself or to parts or characteristics of the objects being described. Each time a new reference appeared in the text we checked whether the thematic selection signalled that strategy by selecting topical themes. The results of our analysis are shown in table 5 below:

Table 5: Distribution of thematic options in descriptive texts (object, people)

		THEMATIC SELECTION					
TEXT Nº	Nº references	LOCAT.	TOPICAL	TIME	PROCESS	CONJUN.	OTHERS
Nº1	4	0	4	0	0	0	0
Nº2	11	0	11	0	0	1	0
Nº3	9	1	8	0	0	0	0
Nº 4	3	0	3	0	0	0	0
Nº5	3	0	3	0	0	0	0
Nº6	8	0	8	0	0	0	1
Nº7	2	0	2	0	0	0	0
Nº8	12	0	9	0	0	0	2
Nº9	9	0	7	0	0	0	2
Nº10	9	0	7	0	0	0	2
TOTAL	**70**	**2**	**62**	**0**	**0**	**1**	**7**

As we can see, the number of references to the object being described, or to parts or aspects of it (70) is signalled most of the times by topical themes (62). This means that topical themes are the most likely candidates to signal the chaining strategy of characterization in this kind of texts.

Focus nodes: all the texts in this category selected an object focus node category to highlight a particular set of knowledge base entities. Object nodes cause an object to be focused on in terms of the specific set of properties that object may have. An object node is usually directly given in the discourse in the sense that it is most often inferred because of the explicit noun mention of an object in the discourse. Here focus node and theme selection tend to conflate in their linguistic realization when first mentioned in the discourse. As to the focus-shift candidates, the expectation is that attribute nodes or action nodes are selected to develop the information.

Example (3) below illustrates the phenomena studied above:

(Chris Morley, Concorde Captain).

> He is one of the original seven British Airways Concorde pilots, and only the week before he has taken Mr. Callaghan to Washington to meet President Carter. He is Central Casting´s idea of a senior airline captain, 47,

> glamorously grey, firm of jaw, calm of voice, with a lean athletic build and a mien of command and reassurance. He earns more than 16,000 a year which still puts him low in the league in comparison with foreign captains. He has four children, a pretty wife, a spacious house, and an 11-year-old Renault 16, which he now gets into drive to Heathrow.

(Ronald V. White. *3 Teaching Written English. Practical Language Teaching*. Heinemann Educational Books, London, 1980)

The characterization strategy is signalled most of the times by the selection of unmarked topical themes (underlined in the text) whenever a reference is made to the object being described or characteristics of it . However, this does not always happens, as in the second clause where the theme is "only the week before".

The focus node category is object (*Chris Morley, Concorde Captain*), and the focus-shift candidates are attribute nodes which reflect different characteristics of the object (*profession, physical appearance, earnings and private life*).

4.3 Narrative texts

We selected ten historical passages including biographies and the history of two cities. All of them reflected a narrative discourse purpose and concentrated on events and participants involved in those events. Our hypothesis was that the discourse purpose (narrative) and the subject-matter (events and participants) would determine the selection of a specific chaining strategy (temporal or participant) -signalled by topical themes- and the selection of an object focus node category.

All the texts were organized by means of a *temporal chaining strategy* on a global level, which we defined as a continuous line of temporal development which is normally signalled on the surface structure by the thematization of temporal themes. Some texts, however, also present a *participant strategy* -defined as the reference to different participants involved in the events- which steers the text on a more local level. Example (5) below displays this local strategy together with the more global temporal strategy. The participant strategy is a very common local strategy in children stories, for example, where the narration of events along a temporal line is also linked to the introduction of different participants along a participant line. We describe below the results of the analysis with respect to these phenomena:

Chaining strategy: in order to test the hypothesis according to which a temporal strategy is the responsible for text organization on a global level, we searched for temporal scopes along a temporal line of development. Each time a new temporal scope was introduced in the text, we checked whether that scope was signalled by the selection of a temporal theme. The results of our analysis are shown in table 6 below.

Table 6: Distribution of thematic options in narrative texts

TEXT Nº	Nº temporal scopes	LOCAT.	TOPICAL	TIME	PROCESS	CONJUN.	OTHERS
Nº1	8	0	2	6	0	0	0
Nº2	6	0	0	6	0	0	0
Nº3	10	0	1	9	0	0	0
Nº 4	5	0	0	5	0	0	0
Nº5	8	0	2	6	0	0	0
Nº6	5	0	1	4	0	0	0
Nº7	8	0	2	6	0	0	0
Nº8	10	0	2	7	0	0	1
Nº9	4	0	0	4	0	0	0
Nº10	16	0	4	12	0	0	0
TOTAL	**80**	**0**	**14**	**65**	**0**	**0**	**1**

Interestingly, the results of the analysis showed that the global temporal strategy is not always signalled by temporal themes. In seven of the ten texts analysed there were occasions when topical themes were selected when a new temporal scope was opened. Text 1 -reproduced below as example (4) below- illustrates this lack of parallellism. From the eight temporal scopes present in the text, six are explicitly signalled by a temporal theme, but the remaining two are not, and the themes selected in these two occasions are topical: the first occurs at the beginning of the passage (*Berne was founded in 1191*) and the second at the end (*Berne was declared World Landmark by the United Nations and was given the title of "Europe's most beautiful floral city*"); here the temporal reference is not explicit but a new temporal scope can be inferred from the reference to the United Nations.

Focus node: most of the texts analysed selected an event focus node category to highlight a particular set of knowledge base entities. Event nodes cause a higher level action to be focused on and can be thought of as a "super action" which captures something common about a group or sequence of actions sharing the same participant in the actions.

As to the focus-shift candidates, the expectation is that action nodes are selected to develop the information.

Example (4) below illustrates the phenomena studied above:

Example (4)

BERN

> Berne was founded in 1191 by Berchtold V. Duke of Zähringen, for strategic reasons and enlarged by stages (Clock Tower before 1218, Prison Tower 1256, St. Christopher' s Tower 1344). <u>In 1353</u> Berne joined the Swiss Confederation. <u>After the fire of 1405</u>, which almost completely destroyed the wooden-built town, the houses were rebuilt of sandstone. The medieval structure of the city originating from that time has remained unchanged up to the present day. <u>From the 14th to the 16th century</u>, Berne reached the zenith of its power by enlarging its territory and gaining great political influence. <u>In 1798</u> Napoleon' s troops invaded berne and thus the collapse of the Ancien Régime was witnessed. <u>In 1834</u> Berne became a university town and <u>in 1848</u> the Federal Capital of Switzerland as well as the capital of the Canton of Berne (second largest Swiss canton). Berne was declared a World Landmark by the United Nations and was given the title of "Europe's Most Beautiful Floral City" .
>
> (Touristic Brochure printed in Switzerland in 1989)

The temporal strategy is signalled most of the times by the selection of temporal themes (underlined in the text) whenever a new temporal scope is opened. However, some temporal scopes are not marked by temporal themes, as the one at the beginning of the discourse and the one at the end of the example text.

An event focus node (the socio-political development of the city of Bern) can be inferred as the "superaction" in the sense that it captures what is common about a sequence of actions where Berne is the participant involved (*Berne was founded in 1191; Berne joined the Swiss Confederation; its houses were rebuilt of sandstone; its medieval structure remained unchanged; Berne reached the zenith of its power; Napoleon's troops invaded Berne; Berne became a University town; Berne became the Federal Capital of Switzerland; Berne was declared a World Landmark*).

Example (5) below illustrates both a temporal strategy operating at a global level and a participant strategy operating at a more local level; the temporal strategy is signalled in most cases by temporal marked themes (underlined in the text); the participant strategy - defined as the reference to different participants involved in the events- is only signalled by topical themes in some occasions (italics in the text). Other references to participants are not signalled by topical themes.

Example (5)

> The search for a short cut between the Pacific and Atlantic oceans brought the first Europeans to the waters of British Columbia. In the late 16th century Sir Frances Drake reportedly sailed as far north as Vancouver Island looking for this passage. In 1778, Capt. James Cook landed on the island. His journals sparked interest in the North Pacific trade.
>
> *Spain* responded by capturing an English trading settlement on Nootka Sound -an act it justified by right of prior discovery. *Britain* threatened war, and Spain relinquished its claim. The next year the admiralty sent Capt. George Vancouver to take possession of the island -which was later named for him- and instructed him to map the Pacific coastline in the hope of discovering a passage to the Atlantic.
>
> Despite the allure of the Northwest Passage, it was the fur trade that became the dominant force in this region; *rival companies and countries* competed for control of this lucrative business.*The two major rivals in Canada* were the Hudson's Bay Co. and the North West Co.
>
> *The North West Co.*, based in Montréal, sent expeditions across Canada to the Pacific in search of a more profitable trade route for its furs. In 1793, Alexander Mackenzie made the first overland journey across the continent north of the Spanish possessions. *Simon Fraser* followed, landing at the mouth of the Fraser River in 1808. Finally David Thompson discovered the source of the Columbia River and followed it to the Pacific.
>
> (Western Canada and Alaska Tourbook, 1993)

Table 7 below specifies both the temporal and the participant scopes in the text. As we can see, there are six temporal scopes in the text, only five of which are marked by marked temporal themes (*In the late XVIth c., in 1778, the next year, in 1793, finally*) and one is not (*In 1808*). Likewise, there are ten participant scopes in the text (*Sir Francis*

Drake, Capt. James Cook, Spain, Britain, Capt. G. Vancouver, Hudson's Bay, North West Co., A. Mackenzie, Simon Fraser, David Thompson), only six of which are signalled by topical themes (*Spain, Britain, rival companies and countries, the two major rivals in Canada, the North West Co., Simon Fraser*).

Table 7: Specification of thematic candidates in example (5)

TEMPORAL SCOPE (global)	**PARTICIPANT SCOPE (local)**		
	(a) people	**(b) countries**	**(c) companies**
In the late XVIth c.	Sir Francis Drake		
In 1778	Capt. James Cook	Spain	
		Britain	
The next year	Capt. G. Vancouver		Hudson's Bay Co.
			North West Co.
In 1793	A. Mackenzie		
In 1808	Simon Fraser		
Finally	David Thompson		

(i) major strategy: temporal scoping sequence
(ii) minor strategy: participant scoping sequence
(a) people
(b) countries
(c) companies

4.4 Instructive texts

As in previous occasions, we selected ten texts which reflected an instructive discourse purpose and where a task of some kind had to be performed by the reader. The texts were method sections from recipes and instructional sections from instruction manuals. This selection was done on the basis that not all the sections from instructional manuals or from recipes have an instructive purpose, but can be descriptive or narrative. In this sense, we follow Wikberg (1992, p. 104) in capturing the essential character of instructive or procedural discourse with the formula:

X PRESCRIBE HOW Y DO Z

where X is the knowledgeable text producer, Y is the ignorant addressee, and DO Z is the goal. Both genres shared the same discourse purpose and the same subject-matter (steps in a procedure). Our hypothesis was that these contextual factors would determine the selection of a specific global chaining strategy (sequential) and the selection of a focus node category (action). This hypothesis was confirmed in all the texts analysed, since they were all organized by means of a *sequential strategy* which iconically reflects the task to be performed by the reader.

With respect to theme selection, we hypothesized that process themes would predominate to reflect that sequential strategy. However, the selection of themes differed depending on the genre: instruction manuals tended to signal the chaining strategy with temporal themes and process themes simultaneously, whereas recipes used only process themes in most of the cases. Therefore we decided to extend our sample and to analyse ten texts per genre to check whether our initial analysis was valid for a wider sample. The results of our analysis are shown in tables 8 and 9 below.

As we can see, both process and temporal themes -they are selected together in our sample- predominate as thematic options in instruction manuals, whereas in recipes the most preferred thematic option is process theme. It is also interesting to note that instruction manuals also select preconditions and goals as thematic options in some occasions. We describe below the results of the analysis with respect to these phenomena.

4.4.1 Instruction manuals

Chaining strategy: in order to test the hypothesis according to which a sequential strategy is the responsible for text organization on a global level, we searched for steps in the procedure to be performed by the reader. Each time a new step was introduced in the text, we checked whether it was signalled by the selection of a process theme which would express the action to be performed. In the case of instruction manuals process themes and temporal themes (usually presented typographically by numbers) were selected in most of the cases. As the results of table 8 show, the correlation between the total number of steps which constitutes the sequential strategy, and the total number of process themes (39) to signal that strategy is high. It is also interesting to note that instruction manuals show a tendency to signal the sequential strategy also by selecting temporal themes; and that in some occasions (4) either the goal of the action or the precondition were thematized. This is illustrated in table 8 below:

Table 8: Distribution of thematic options in instruction manuals

THEMATIC SELECTION

TEXT Nº	Nº STEPS	LOCAT.	TOPICAL	TIME	PROCESS	CONJUN.	OTHERS
Nº1	4	0	0	4	4	1	0
Nº2	5	0	0	5	5	0	0
Nº3	5	0	0	5	4	0	2
Nº 4	3	0	0	3	3	0	0
Nº5	7	0	0	7	6	1	0
Nº6	4	0	0	4	4	0	0
Nº7	7	0	0	7	5	0	2
Nº8	4	0	0	4	4	0	0
Nº9	2	0	0	2	2	0	0
Nº10	2	0	0	2	2	0	0
TOTAL	**43**	**0**	**0**	**43**	**39**	**2**	**4**

For example, the sequential strategy in text (7) in the table -reproduced below as example (6)- consists of seven steps all of which are signalled by a temporal theme (1, 2, 3, 4, 5, 6, 7); five are signalled by a process theme (*press, set, set, set, adjust*), and two of them have selected two goals as clausal themes (*to playback the tape through your amplifier; to stop the tape at any position on the tape*). Within each of the steps it is possible to encounter substeps marked by process themes (*insert, close*) on a more local level of analysis.

Example (6), an extract from an instruction manual, illustrates the phenomena studied above:

Example (6)

PLAYBACK

1. Press the STOP/SELECT button and insert the recorded cassette into its compartment and close the cassette compartment lid.
2. Set the selector switch to the TAPE position.
3. Set the MODE switch to the proper position.

4. Set the MODE switch to the proper position.
5. Adjust the VOLUME and TONE controls for your desired sound.
6. To playback the tape through your amplifier, connect optional REC/PB cord from the REC/PB socket of its unit to your amplifier. The output power is fixed regardless of VOLUME position.
7. To stop the tape at any position on the tape, press the STOP/EJECT button. When the tape is wound completely, the PLAY button will be released automatically.

(The Portable Cassette Recorder with Radio Instruction Manual Sanyo M9 930 K, Sanyo Electric Co. Ltd. Japan, 6.)

The sequential strategy is signalled most of the times by the selection of temporal themes and process themes simultaneously (underlined in the text) whenever a new step in the procedure has to be performed. Substeps in each of the subsequent steps tend also to be signalled by process themes in most texts. In some occasions, however, goals of the actions reflecting different steps are thematized, as in steps 6 and 7 above.

The focus nodes are actions corresponding to the subsequent steps in the procedure. The first step consists of three substeps (*press, insert, and close*). Focus-shift candidates are next actions capturing the subsequent steps (*set, set, set, adjust, connect, and press*); in the case of the two last steps captured by the actions of *connect* and *press*, the focus-shift candidates would be the goals of the actions.

4.4.2 Recipes

With respect to recipes, we obtained similar results to the ones obtained for instruction manuals: there is a high correlation between the total number of steps (105) and the selection of process themes (105) to signal the sequential strategy. However, in contrast to instruction manuals, temporal themes are only selected in two occasions together with process themes. These results are shown in table 9 below:

Table 9: Distribution of thematic options in recipes

		THEMATIC SELECTION					
TEXT Nº	Nº STEPS	LOCAT.	TOPICAL	TIME	PROCESS	CONJUN.	OTHERS
Nº1	8	0	0	0	8	0	0
Nº2	6	0	0	0	6	0	0
Nº3	20	0	0	0	20	0	0
Nº 4	10	0	0	0	10	0	0
Nº5	9	0	0	1	9	0	1
Nº6	7	0	0	1	6	0	1
Nº7	10	0	0	0	10	0	0
Nº8	7	0	0	0	7	0	0
Nº9	7	0	0	0	7	0	0
Nº10	21	0	0	0	21	0	0
TOTAL	**105**	**0**	**0**	**2**	**105**	**0**	**2**

For example, the sequential strategy in text (3) in the table -reproduced below as example (7)- consists of twenty steps all of which are signalled by process themes.

Example (7)

SALMON CUTLETS WITH AN AVOCADO MAYONNAISE

<u>Peel</u> and <u>slice</u> both the onion and carrot and <u>put</u> into a wide, flattish pot with a lid. <u>Stir</u> in the wine, bay leaf, salt and pepper. <u>Bring</u> to a boil and <u>simmer</u> for 10 mins. <u>Add</u> the salmon cutlets, <u>cover</u> and <u>simmer</u> on a low heat for about 7/8 mins.

Sauce:

<u>Peel</u> and <u>stone</u> the avocado and <u>slice</u> the flesh. <u>Put</u> the mayonaise, brandy, avocado and lemon juice in a food processor until smooth. <u>Remove</u> the salmon cutlets from the pot and <u>set</u> one on each plate and <u>allow</u> to cool. <u>Discard</u> the vegetables and <u>simmer</u> the juices left in the pot until they reduce to about two tablespoons. <u>Stir</u> the juices into the mayonnaise and <u>serve</u> with a plain, green lettuce salad and steamed new potatoes.

The sequential strategy is always signalled by the selection of process themes (underlined in the text) whenever a new step in the procedure has to be performed. As in the case of instruction manuals, the focus nodes are actions corresponding to the subsequent steps in the procedure. Focus-shift candidates are the next actions in the process of preparing the dish. As in the case of expository texts and descriptions of objects and people, the selection of process themes and focus nodes may conflate in the linguistic realization.

4.5 Argumentative texts

We selected ten texts with an argumentative or persuasive purpose where the writer presented facts and ideas to convince readers or the truth or the falsity of some fact. Our hypothesis was that the discourse purpose (persuasive or argumentative) and the subject-matter (facts and ideas) would determine the selection of two prototypical chaining strategies: through-argument or counter-argument. All the texts were organized by means of one of these strategies on a global level, which we defined as a the presentation of a series of arguments in favour or against the thesis defended by the writer. Moreover, our hypothesis was that most texts following one of these strategies would select conjunctives as the most likely thematic option to signal the chaining strategy, and that facts focus nodes would be the most common focus selection for information development. We describe below the results of the analysis with respect to these phenomena:

Chaining strategy: in order to test the hypothesis according to which a through-argument or a counter-argument strategy is responsible for text organization on a global level, we counted the arguments in favour or against the thesis presented by the author. Each time a new argument or a new step in an argument was introduced in the text, we checked whether it was signalled by a conjunctive theme or by any other theme. The results of our analysis are shown in table 10 below.

Table 10: Distribution of thematic options in argumentative texts

THEMATIC SELECTION

TEXT Nº	Nº ARGU-MENTS	LOCAT.	TOPICAL	TIME	PROCESS	CONJUN.	OTHERS
Nº1	2	0	0	0	0	2	0
Nº2	5	0	5	0	0	0	0
Nº3	6	0	0	0	0	6	0
Nº 4	6	0	1	0	0	5	0
Nº5	4	0	0	0	0	4	0
Nº6	3	0	0	0	0	3	0
Nº7	2	0	0	0	0	2	0
Nº8	3	0	0	0	0	3	0
Nº9	2	0	2	0	0	0	0
Nº10	2	0	0	0	0	2	0
TOTAL	**35**	**0**	**8**	**0**	**0**	**27**	**0**

In our sample conjunctive themes were the most frequently selected types to signal the chaining strategy; however, we found instances -like text 2- where each new argument was not signalled by a conjunctive theme, but by a topical theme.

Focus nodes: in order to account for the kind of focusing present in most of the texts analysed, we introduced a new category, which we call "fact" focus node, which captures the focusing mechanism typical of argumentative texts. Under a fact focus node, the expectation is to find relations of different kind as focus-shift candidates, more specifically evidences for the facts presented in the thesis part of the text. All the texts analysed selected this type of focus node category.

Example (8) below illustrates the phenomena studied above:

Example (8)

OWNING A CAR

Should a person own a car? This is an important question. In a large urban area, there are some good reasons for owning a car. First, a car allows a person to move around freely. With a car, there is no need to check a bus schedule or wait for a train. Second, a car is a comfortable way to travel, especially in the winter time. In bad weather, the driver stays warm and dry, while the poor bus or train driver might have to stand in the rain. Finally, a driver is usually safe in a car at night. The rider might need to walk down a dark street to get to the stop or wait in a dark corner.

(Linda Lonon Blanton. 1981). Linda Lonon Blanton. *A Text for Students of English as L2.* Intermediate Composition Practice. Book 1. Newbury House. 31. 1981.)

The chaining strategy in this example is through-argument, presenting three arguments in favour of the thesis according to which there are good reasons for owning a car in a large urban area. These three arguments are signalled by the selection of conjunctive themes (*first, second, finally*) underlined in the text.

The focus node here is the fact that there are good reasons for owning a car in a large urban area. This is the thesis whose truth the writer wants to assert. The expected children of this node are usually evidences in the form of arguments supporting the thesis presented in the fact node: *a car allows person to move around freely; a car is a comfortable way to travel, especially in winter time; a driver is usually safe in a car at night* . Each of the focus-shift candidates raises expectations about how the discourse should progress. Thus, under an evidence node in the tree we would expect to see relation nodes, like elaboration nodes, as in example (8).

5 Towards a specification of the process of information development

In this section we would like to present a prespecification of the process of information development illustrated in the figure at the end of this document. The purpose of this prespecification is to go beyond a pure static representation of the resources which control the structure and development of information in discourse, but to interpret them in dynamic terms so that, for example, a text generation system can process them accordingly. With this aim in mind, we present here some tentative informal rules which might guide a future computational implementation of the results obtained in this empirical research. These rules are an attempt to capture in procedural terms the behavioural tendency of the phenomena studied under the identified contextual conditions, and, in this sense, they can be considered as empirically motivated rules for the production of pragmatically appropriate texts. The informal rules are the following:

1. Decide on discourse purpose and subject matter
2. Find chaining strategy; if there are more than one, select which chaining strategy is major or global, and which one(s) are minor or local, that is, rank them in order.
3. a) For each (global) strategy, list, in order, the sequence of themes for the whole text, to build a table like table A.
   b) For next highest (local) strategy -if there is one, within the context of each global theme- list in the table, in order, the sequence of local themes.
   c) Repeat for each global context, if any, all down the list.
   d) Repeat (b) for next lower strategy, if any, getting more and more local.

RESULT: a table of theme candidates, like table A below.

Table A: Specification of thematic candidates in example (5)

TEMPORAL SCOPE (global)	**PARTICIPANT SCOPE (local)**		
	(a) people	**(b) countries**	**(c) companies**
In the late XVIth c.	Sir Francis Drake		
In 1778	Capt. James Cook	Spain	
		Britain	
The next year	Capt. G. Vancouver		Hudson's Bay Co.
			North West Co.
In 1793	A. Mackenzie		
In 1808	Simon Fraser		
Finally	David Thompson		

(i) major strategy: temporal scoping sequence
(ii) minor strategy: participant scoping sequence
(a) people
(b) countries
(c) companies

4. For each message, build a clause, and thematize the different theme candidates in the tables, according to the following rules:
   a) If the chaining strategy is characterization, whenever a reference to the object, or the concept appears in the text, thematize the unmarked topical themes.
   b) If the major chaining strategy is spatial, whenever a new spatial scope is opened, thematize the locative themes.
   d) If the major chaining strategy is temporal, whenever a new temporal scope is opened, thematize the marked temporal themes in the table.
   f) If the chaining strategy is sequential, thematize the processes. Depending on genre, thematize also as follows:
      - in instruction manuals, thematize the temporal themes; occasionally, thematize the goals or the preconditions at random
      - in recipes, thematize the processes only; occasionally, thematize the temporal themes.
   d) if the chaining strategy is through-argument or counter-argument, whenever a new argument or a step in an argument is presented, thematize the conjunctive textual themes.
5. To build the discourse focus tree, follow the following rules:
   (i) Decide on discourse purpose and subject matter.
   (ii) Find category of focus node.
   (iii) For each focus node selected, list, in order, the focus-shift candidates, to build a table (like table A1)

RESULT: a table of focus nodes with their corresponding focus-shift candidates.

Table A1: Specification of focus node and focus-shift candidates in example (4)

FOCUS NODE TYPES	FOCUS-SHIFT CANDIDATES
EVENT	Berne was founded in 1191
Socio-political development of the city of Bern	Berne joined the Swiss Confederation
	the houses were rebuilt of sandstone
	the medieval structure remained...
	Berne reached the zenith of its power
	Napoleon's troops invaded Berne
	Berne became a university town
	Berne became the Federal capital of S.
	Berne was declared a World Landmark

6 Concluding remarks

In this paper we have investigated the role of related phenomena such as theme and chaining strategies, focus nodes and focus-shift candidates in the structuring and development of information in discourse and the contextual factors which influence their behaviour. The contextual factors from the communicative situation which we have identified so far include the purpose and the subject-matter of discourse. The influence of these two main factors -responsible for the definition of text types- has been empirically validated through text analysis of a corpus of seventy texts reflecting variation in both factors. The results of the empirical text analysis have shown that both discourse purpose and subject-matter are determining constraints in the selection of a specific chaining strategy along which writers organize their texts, and in the selection of the focus node type. They have also illustrated the discourse function of theme as a signal of a chaining strategy and of focus as the locus of principal inferential effort which governs the selection of entities from the knowledge base. Finally, the paper has outlined a dynamic prespecification of the process of information development with a view towards computational implementation.

7 Acknowledgements

We thank the Ph. D. students Isabel Alonso and Francisco Ballesteros for their contribution in the collection and analysis of the texts used for this deliverable. We also thank Dr. Gisela Redeker for her useful comments on an earlier draft of this paper.

References

(Chomsky, 1971) Noam Chomsky. Deep Structure, Surface Structure and Semantic Interpretation. In D.D. Steinberg and L.A. Jakobovits, eds. *Semantics: an I nterdisciplinary Reader in Philosophy, Linguistics and Psychology.* Cambridge, 1971.

(Halliday, 1967) M.A.K. Halliday. Notes on Transitivity and Theme in English. *Journal of Linguistics* 3: 37-81, 1967.

(Lavid & Hovy, 1993). Lavid, J. and Hovy E.H. 1993. Focus, Theme, Given, and Other Dangerous Things. Working paper.

(McCoy and Cheng, 1990) K McCoy and J. Cheng. Focus of attention: constraining what can be said next. In C. Paris, W. Swartout and W. Mann, editors. *Natural language generation in artificial intelligence and computational linguistics*. Kluwer, Dordrecht, 1990.

(McKeown, 1985) Kathleen R. McKeown. *Text generation: Using discourse strategies and focus constraints to generate natural language text.* Cambridge University Press, Cambridge, 1985.

(Rochemont and Culicover, 1990) M.S. Rochemont and P. W. Culicover. *English focus constructions and the theory of grammar.* Cambridge University Press, 1990.

(Sgall, 1979) Peter Sgall. Towards a definition of focus and topic. *Prague Bulletin of Mathematical Linguistics* 31: 3-26, 1979.

(Sgall et al, 1973) P. Sgall, E. Hajicova, and J. Panevova. *Topic, Focus, and Generative Semantics*. Kronberg/Taunus, 1973.

(Wikberg, 1992) K.Wikberg. Diversifying procedural discourse. In A-C. Lindberg, N.E.Enkvis and K.Wikberg, editors. Nordic Research on Text and Discourse: NORDTEXT Symposium 1990. Åbo: Åbo Academy Press.

Texts:

Expository

The Encyclopaedia Americana. Vol. 3. Americana Corporation, 508, 1963. (Text 1)

The Encyclopaedia Americana. Vol. 3. Americana Corporation, 508, 1963. (Text 2)

(Chesterton, 1969) G.K. Chesterton. Prohibition. In Kane, T.S & Peters, L. J. *Writing Prose. Techniques and Purposes*. 20, 1969 (Text 3).

The Encyclopaedia Americana. Vol. 3. Americana Corporation, 288, 1963. (Text 4)

The Encyclopaedia Americana. Vol. 14. Grolier Incorporated, 46, 1991. (Text 5)

The Encyclopaedia Britannica. Vol. 14. Micropedia, Encyclopaedia Britannica Inc. University of Chicago, 445, 1991. (Text 6)

The New Columbia Encyclopedia Illustrated. Vol. A. Columbia University Press, 336, 1975. (Text 7)

The Encyclopaedia Britannica. Vol. 4. Micropedia, Encyclopaedia Britannica Inc. University of Chicago, 560, 1991. (Text 8)

The Encyclopaedia Britanica. Vol. 4. Micropedia, Encyclopaedia Britannica Inc. University of Chicago, 445, 1991. (Text 9)

The New Columbia Encyclopedia Illustrated. Vol. A. Columbia University Press, 1, 1975**.** (Text 10)

Descriptive (spatial relations)

(Tomkinson, 1985) Michael Tomkinson. *TUNISIA*. Société Tunisienne de Diffusion, Tunis, 7, 1985. (Text 1)

The Cardiff Guide. Cardiff City Council, 32-33. (Text 2)

Western Canada and Alaska Tourbook, 1993. (Text 3)

The Queen' s House, A Royal Palace by the Thames. Centurion Press Ltd. London, 36-37. (Text 4)

(Tomkinson, 1985) Michael Tomkinson. *TUNISIA*. Société Tunisienne de Diffusion, Tunis, 40, 1985. (Text 5)

(Tomkinson, 1985) Michael Tomkinson. *TUNISIA*. Société Tunisienne de Diffusion, Tunis, 41, 1985. (Text 6)

Welcome to Britain (touristic brochure). BTA, 10-11. (Text 7)

The Royal Naval College Greenwich. Paul Woodmansterne Publications Ltd. Watford, 1987. (Text 8)

Guide to the Rijksmuseum Amsterdam. Educational Service, Rijksmuseum, Utrecht, 19, 1985. (Text 9)

Mid Wales, Escape to our Beautiful Landscapes. Mid Wales Tourism, 39. (Text 10)

Descriptive (objects, people)

(White, 1980) Ronald V. White. *3 Teaching Written English.* Practical Language Teaching. Heinneman Educational Books, London, 53, 1980. (Text 1)

(Jordan, 1980) R.R. Jordan. *Academic Writing Course.* Collis Study Skills in English, Glasgow, 31, 1980. (Text 2).

(White, 1980) Ronald V. White. *3 Teaching Written English.* Practical Language Teaching. Heinneman Educational Books, London, 43, 1980. (Text 3)

(Pincas and Hadfield, 1982) Anita Pincas and Charles Hadfield. *Writing in English.* Book 3. Mac Press, London, 40, 1982. (Text 4)

(Pincas and Hadfield, 1982) Anita Pincas and Gillian and Charles Hadfield. *Writing in English.* Book 3. Mac Press, London, 41-2, 1982. (Text 5, 7)

Charles Dickens. *Hard Times* (adapted version). Ed. Collins ELT, 15, 1979. (Text 6)

(Foll and Kelly, 1994) David Foll and Anne Kelly. *First Certificate: Avenues.* Cambridge University Press, 172, 1994. (Text 8).

(Soars, 1987) Maughan Somerset. The Lotus Eater. In John and Liz Soars, eds. *Headway, Upper-Intermediate*. Oxford University Press, 68, 1987. (Text 9)

(Foll and Kelly, 1994) David Foll and Anne Kelly. *First Certificate: Avenues.* Cambridge University Press, 43, 1994. (Text 10).

Narrative

Bern (touristic brochure). Switzerland, 1989. (Text 1)

(Berkin et al., 1992) Carol Berkin, Alan Brinkley, Clayborne Carson, Robert W. Cherny, Robert A. Divine, Eric Foner, Jeffrey B.Morris, Arthur Wheeler, Leonard Wood. *AMERICAN VOICES, A History of The United States.* Scott, Foresman and Company, a Division of Harper Collins Publishers, Glenview, Illinois, 550-51, 1992. (Text 2)

The Encyclopedia Britannica. Vol. 14. Macropedia, 15th Edition by Encyclopaedia Britannica Inc. University of Chicago, 392, 446, 447, 1988. (Text 3, 4, 5).

The New Columbia Encyclopedia Illustrated. Columbia University Press, 182, 1975. (Text 6)

The Encyclopedia Britannica. Vol. 14. Macropedia, 15th Edition by Encyclopaedia Britannica Inc. University of Chicago, 733-34, 1988. (Texts 7 and 8).

Western Canada and Alaska Tourbook, 1993. (Text 9)

The New Columbia Encyclopedia Illustrated. Columbia University Press, 1334, 1975. (Text 10)

Instructional

The Sony Video Cassette Recorder Beta SL - C30E Instructions. Sony Corporation, 3, 4 (Text 1, 3)

The Sharp Video Cassette Recorder VHS, Operation Manual. Sharp Manufacturing, U.K. 9, 1991. (Text 2)

The Portable Cassette Recorder with Radio Instruction Manual Sanyo M9 930 K. Sanyo Electric Co. Ltd. Osaka, Japan, 3-6. (Text 4- 7)

The Random House Webster' s Electronic Dictionary and Thesaurus College Edition Instruction Manual. Software International, San Francisco, 30, 1992. (Text 8)

The Operating Instructions of Room Air Conditioner CS/CU 703 KE Panasonic. Matsushita Electric Industrial Co. Ltd. Japan, 10, 12, 1991. (Text 9, 10)

Argumentative

(Muscatine and Griffith, 1988) Thomas S. Szasz. Drug Prohibition. Reason. In Charles Muscatine and Marlene Griffith. *The Borzoi College Reader*. Sixth Edition. Alfred A. Knopf, New York, 616, 1988. (Text 1)

(Muscatine and Griffith, 1988) Thomas Jefferson. Declaration of Independence. In Charles Muscatine and Marlene Griffith. *The Borzoi College Reader*. Sixth Edition. Alfred A. Knopf, New York, 579, 1988. (Text 2)

(Muscatine and Griffith, 1988) Jonathan Swift. A Modest Proposal. In Charles Muscatine and Marlene Griffith. *The Borzoi College Reader*. Sixth Edition. Alfred A. Knopf, New York, 94-95, 1988. (Text 3)

(Muscatine and Griffith, 1988) Sissela Bok. Lies for the Public Good. In Charles Muscatine and Marlene Griffith. *The Borzoi College Reader*. Sixth Edition. Alfred A. Knopf, New York, 496, 1988. (Text 4)

(Muscatine and Griffith, 1988) Lewis H. Van Dusen. Jr. Civil Disobedience: Destroyer of Democracy. In Charles Muscatine and Marlene Griffith. *The Borzoi College Reader*. Sixth Edition. Alfred A. Knopf, New York, 630-631, 1988. (Text 5)

(Muscatine and Griffith, 1988) Susan Wright. The Recombinant DNA Controversy Revisited. In Charles Muscatine and Marlene Griffith.*The Borzoi College Reader.* Sixth Edition. Alfred A. Knopf, New York, 781-784, 1988. (Texts 6 and 7)

(Blanton, 1981) Linda Lonon Blanton. *A Text for Students of English as L2.* Intermediate Composition Practice, Book 1. Newbury House, 1981. (Text 8)

(Bregazzi and Downing, 1979) John Carey. The Observer. January 12, 1977. In Bregazzi and Downing. *A Mosaic of Modern English.* S.G.E.L. 95, 1979. (Text 9)

(Muscatine and Griffith, 1988) Susan Wright. The Recombinant DNA Controversy Revisited. Charles Muscatine and Marlene Griffith. *The Borzoi College Reader.* Sixth Edition. Alfred A. Knopf, New York, 782-783, 1988. (Text 10)