



**HAL**  
open science

## Contextual phenomena in dialogue

Karine Baschung, Gabriel G. Bès, Rachel Panckhurst, Henk Zeevat

► **To cite this version:**

Karine Baschung, Gabriel G. Bès, Rachel Panckhurst, Henk Zeevat. Contextual phenomena in dialogue. [Research Report] T2.3, Laboratoires de Marcoussis; Centre for Cognitive Science, University of Edinburgh; Université Blaise-Pascal, Clermont-Ferrand. 1986. hal-01143099

**HAL Id: hal-01143099**

**<https://hal.science/hal-01143099>**

Submitted on 16 Apr 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Contextual phenomena in dialogue**

Karine Baschung<sup>1</sup>, Gabriel G. Bès<sup>1</sup>, Rachel Panckhurst<sup>1</sup>, Henk Zeevat<sup>2</sup>

1. Formation doctorale Linguistique et informatique, Université Blaise-Pascal, Clermont-Ferrand

2. Centre for Cognitive Science, University of Edinburgh

*ACORD ESPRIT Project 393, Task 2.3, 1986*

### **Abstract**

This document is an intermediate deliverable for the ACORD Esprit Project 393, on the construction and interrogation of knowledge bases using natural language text and graphics. It examines a series of linguistic phenomena related to a dialogue situation. Ellipsis and anaphora are treated in chapter 1 and partly in chapter 2, under the heading “Notation of different types of anaphora”. The rest of chapter 2 is dedicated to a definition a “Minimal Dialogue” and a detailed description of French interrogative structures. Chapter 3 deals with descriptive problems met when trying to implement the targeted structures in the UCG formalism. Chapter 4 deals with the semantic representation of questions. Chapter 5 ends the survey by presenting some suggestions for implementation standards for the ACORD prototype Dialogue Parsers.

### **See also**

Gabriel G. Bès (dir.), *Technical Documentation*, ACORD ESPRIT Project 393, Laboratoires de Marcoussis, December 8 1989.

Gabriel G. Bès, Thierry Guillotin (eds), *A Natural Language and Graphics Interface: Results and Perspectives from the ACORD Project*, Springer, 1992.

July, 1986

ESPRIT PROJECT 393

ACORD

Construction and Interrogation of Knowledge Bases  
using Natural Language Text and Graphics

Draft

Deliverable of task T2.3

**Contextual Phenomena in Dialogue**

*prepared by :*  
*Laboratoires de Marcoussis*  
*Centre for Cognitive Science*  
*Université de Clermont II*



## 0. INTRODUCTION

As can be inferred from the title, the main concern of this deliverable was to examine phenomena, currently thought of as being related to a dialogue situation. These include, according to the Technical Annex : ellipsis and anaphora. Other phenomena such as, interrogative structures and their relation to answers, temporal reference, speech acts, discourse structure, etc., are at least as important to a succesful implementation of a dialogue system.

It is obvious that not all of these problems could be treated in detail. Moreover, some of them had already been touched upon by the chapter on *Further Syntactic Phenomena* of the T2.1 deliverable. As a reminder, phenomena treated covered : Unbounded Dependencies, Coordination, Ellipsis and Anaphora, Temporal Phenomena, Non-declarative Clause Types. For more detail on speech acts and discourse structure we refer to the deliverable T4.2 devoted to the *Investigation of the Dialogue Manager*. The deliverable related to task T2.5 - *Dialogue in Dynamic Environment* will be entirely dedicated to temporal reference. Phenomena related to the dialogue context as concieved in ACORD - i.e. interaction between graphics and textual means - should be treated in the text on the *Investigation on Direct Manipulation* (T4.3).

As to the present document, it has been given the following contents :

Anaphora and Ellipsis are treated in the first chapter, "Anaphora and Resolution", and partly in the second chapter under the heading "Notation of Different Types of Anaphora".

The rest of the second chapter is dedicated to a definition of a "Minimal Dialogue" and a detailed description of French interrogatives. For some of the question types, the problem of the conformity of an answer to a question is also raised. However, time constraints didn't allow for an exhaustive treatment of the problem. The ideas of this chapter should nevertheless be examined by the team involved in Dialogue Management. As far as Answer Generation is concerned, this item will certainly be picked up again in the deliverable related to this task (T2.10).

The teams actually involved in dialogue parsing are both working with the UCG formalism. This gave rise to the third chapter : "Descriptive problems related to the present status of the UCG formalism".

The final aim of parsing is to give a semantic representation of what has been parsed. For dialogue parsing this involves giving a "Semantic Representation of Questions". This is accounted for by the fourth chapter.

Last, some suggestions are presented for implementation standards for the ACORD prototype Dialogue Parsers.

Contributions in chapter I and IV are due to Henk Zeevat (Centre for Cognitive Science).

Work on interrogatives has been done by the team of Clermont II University, more precisely, by G.G. Bès and R. Pankhurst for chapter II and by K. Baschung for chapter III. Section 4 of chapter III is largely indebted to discussions between this team and A. Corluy and Th. Guillotin at Laboratoires De Marcoussis.

Suggestions in chapter V resulted from plenary discussion between all participants to the work on this deliverable.

## 1. Anaphora and Resolution

### 1.1. Resolution

Anaphoric elements in a language are those expressions, that need information from the linguistic context in order to be fully interpretable. The most studied elements are the anaphoric pronouns, but it is generally accepted that for interpreting phenomena like ellipsis and contextual restrictions one also needs to pick up material that was introduced by explicit mention before.

Resolution can be loosely described as the process by which the linguistic material, or its semantical representation, that is needed for adequate interpretation of anaphoric elements, is found in the context. A preliminary step, that does not belong to resolution itself is the recognition of the fact that a linguistic element requires information and identification of the manner in which this requirement is indicated and the type of the information that is wanted. Resolution as such involves the way in which the information is found, the structure of the contextual information and the way in which the identified information is brought to bear on the representation of the structure to which the element belongs.

The first section is a short overview of anaphoric ele-

ments in the sense of the definition. The next sections deal, equally briefly, with some of the approaches taken to the resolution problem in terms of restrictions on the search procedure, and with a general procedure for dealing with resolution.

## 1.2. Pronouns

Almost all pronouns need a form of resolution: it is necessary in order to interpret them to find out what entity they stand for in the context. For all these it holds that their reference varies with the occasions of use of the pronoun. But the resolution proceeds differently for different types of pronouns. Relatively unproblematic are the so-called indexicals. These pronouns ( I, now, here ) stand for objects that stand in a particular relation to the utterance of which they are a syntactical part: I for the person who utters it, now for the moment at which it is uttered and here for its location. Reflexive pronouns require the identification of the subject of the clause, or the local subject as in (1).

(1) Harry persuaded John to take care of himself.

Another type are the deictic pronouns. These are to be identified by taking note of the object the speaker is pointing to at roughly the time of the utterance. It is



not always possible to distinguish deictic use from reference to things which are prominent in the non-linguistic context of the utterance, which can be the function of both personal and demonstrative pronouns. Lastly, there are anaphoric uses: here the referent of the pronoun is identical to that of another term that appears in the linguistic context, or closely related to it. It is possible to subdivide anaphora in further classes, but since the subdivisions already involve theory, it does not seem to be appropriate at this point. The following scheme covers the pronouns and some other words whose meaning involves resolution.

### 1.2.1. Personal pronouns.

I me my mine	indexical
you your yours	indexical/indexical cum anaphoric
he him his	deictic/indexical/anaphoric
she her hers	deictic/indexical/anaphoric
it its	deictic/indexical/anaphoric
we us our ours	indexical/indexical cum "anaphoric"
they them their theirs	deictic/indexical/anaphoric

### 1.2.2. Demonstrative Pronouns.

that/those	deictic/anaphoric(to text/sentence/situation)
this/these	deictic/anaphoric(to text/sentence/situation)
that/those CN	deictic/anaphoric(to text/sentence/situation)
this/these CN	deictic/anaphoric(to text/sentence/situation)

### 1.2.3. Reflexives

myself/yourself/himself/ourselves/themselves

### 1.2.4. Temporal

Indexical:                   now  
                              yesterday/today/tomorrow  
                              last/next/ago  
                              Tuesday

Contextual:                   now/then

Anaphoric/Contextual:       previously  
                                  afterwards  
                                  later  
                                  Tuesday  
                                  since  
                                  until  
                                  before

### 1.2.5. Local

here  
there  
there/elsewhere

### 1.2.6. Other

One       anaphoric or contextual reference to classes, expressed by a  
          (complex) CN  
So        anaphoric or contextual reference to a reason or cause or action  
          expressed before  
Such     anaphoric or contextual reference to something expressible  
          by an adjective

### 1.3. VP and NP ellipsis

Opinion can diverge as to the precise definition of ellipsis. Here it is reserved for cases like (2a-b):

- (2) a. John will do his homework but Bill hasn't.  
b. Mary bought peanuts. John bought some [] too.

It has been attempted to explain ellipsis as a purely syntactical phenomenon. Certainly it is the case that the permissibility of ellipsis is a phenomenon that varies between even closely related languages. In Dutch e.g. one has cases of ellipsed nouns like (3) where English must use an anaphoric pronoun.

- (3) Marie kocht een gele trui en Jan [] een rode [].  
Mary bought a yellow sweater and John [] a red one.

Though such peculiarities are relevant for parsing, it does not follow that the semantic treatment of ellipsis should be governed by syntactical analysis. The resolution of ellipsis seems to work rather like the resolution of anaphoric pronouns, though it may be that there are special restrictions on the position of the antecedent that do not hold in the case of anaphoric pronouns.

### 1.4. Descriptions

According to the familiarity theory of definiteness, all definites refer to something the speaker assumes his

audience to be familiar with. Familiarity must be taken in a very broad sense however. It includes: acquaintance, having a definition of, being referred to by a preceding term, being salient in the non-linguistic context, or being functionally related to an item that is familiar in one of these senses.

Only the last three cases belong to the realm of resolution. This does not mean that the Acord system as a whole does not need to deal with the other uses of definites, that presuppose acquaintance, or the availability of a definition. The knowledge base presumably needs to identify all uniquely described entities, since the speaker indicates by his use of a definite term, that he expects his interlocutor to do so. But this identification is a matter of searching through the knowledge base rather than through the context.

Even within the context of man machine dialogue one can imagine an example of referring to the non linguistic context as in an utterance like (4)

(4) There is something wrong with the terminal

after the last machine action led to an incomprehensible display on the screen.

But the more relevant cases are the ones where the

description is used to pick up an element that was under discussion before. These come in a number of forms:

- (5)                   the former  
                          the plumber  
                          the terminal

The first of these is typically used in a situation where there is more than one item equally prominent in the preceding sentence, but where normally one would have used a pronoun. The second would normally be a so-called short description that is used to resume discussion of a plumber mentioned before, but too far away to be accessible for a pronoun. The last, finally, (an associative reference) refers to something not mentioned before but functionally related to something that has been mentioned previously and is accessible for reference by means of a pronoun.

### 1.5. Tense

It is standard within the DRT literature to treat tense as a special case of anaphoric/indexical resolution. Normally, present tense, like now, makes reference to the time of speech, whereas the past tense refers to either a moment related to the event mentioned in the preceding sentence or to a moment referred to by a temporal modifier in the same clause. Contextual binding is also possible here. Other temporal distinctions expressed by auxiliaries

in English, like future or perfect, should not be confused with tense in this respect. For discussion on this issue see (Hinrichs 1982), (Kamp 1979), (Kamp 1981), (Kamp and Rohrer 1983) and (Partee 1984).

An important point here is that the regularities noted by Hinrichs have been derived from consideration of a very particular type of texts: simple monologic stories. It has been repeatedly observed that other texts, notably the type considered in the Acord project, do not necessarily conform to this pattern. There are other means available for organising information than the one predominant in narrative text, where the sequence of events and accompanying states in the story corresponds with the sequence of sentences in the text. Studies of discourse structure -see also below- suggest that dominance relations are at least as important as temporal precedence in organising texts, also where the role of the past tense is concerned. The problem discussed in (Partee 1984) of dealing with successive events from a model theoretic point of view may perhaps be solved using dominance. If a certain text is elaborating a common dominating element, the next event is naturally the next event that is dominated by the element that is elaborated. Of course this does not mean that we thereby obtain a mathematically precise notion.

## 1.6. Further resolution

### 1.6.1. Contextual restriction

It is normal to assume in philosophical logic that quantification -even if it is unrestricted- is only in the rarest of circumstances about every object there is. In most contexts, there are certain stronger restrictions available ("the domain of discourse") that limit the quantifications (including the restricted ones) to some contextually determined set of objects. It seems that we find here a phenomenon that is closely related with both associative and short definite descriptions. Locatives, both of the local and the temporal kind, may also play an important role.

- (6) We were in Germany last week. Every city was covered by snow. The roads were slippery, and we had to take care, until we crossed the French border. There, no snow had fallen.

In (6) the extension of city in the quantification is limited to German cities, the roads to German roads, and the relevant borders to the franco-german ones. In the last clause, the extension of snow is restricted by there.

### 1.6.2. Disambiguation

Finding the right reading in a syntactically or lexically ambiguous expression may be thought of as a kind of resolution. There is a clear difference however with the kind of resolution involved in the cases above: it is not the context, but the lexicon that specifies the possibilities. Some of the elements that are involved in resolution, like the register of the text, the global topic, and other things that go to determine the intention of the utterance, may however be decisive in disambiguating.

### 1.6.3. Speech act recognition

The same holds in this area as for disambiguation. The choices are not given from the context but rather from the types of speech act that the language contains. Global factors concerning the context may be instrumental in determining what speech act one is dealing with.

It seems best to treat these two problems as falling outside a resolution component.

## 1.7. Theory

In this chapter, some of the restrictions found in the literature on resolution are reviewed. Some discussion, mostly on syntactical and logical restrictions, can also



be found in Bull's contribution to task 1.3. A useful, more extended survey, that has been of great use in preparing these pages is (Hirst 1981).

### 1.7.1. Syntactical Restrictions

The antecedent precedes the incomplete element.

This holds for sentences, and for coordination within sentences: it is not possible to have a pronoun in a sentence that precedes the sentence containing its antecedent. Exceptions are rhetorical exploitations of this restriction:

- (7) We have waited for him for seven years. But here he is at last. Ladies and gentlemen, your applause please for Joe Dippee!

Within a sentence, it is not impossible, but nevertheless restricted. Compare (8a-1).

- (8)
- a. If he comes, John will be surprised
  - b. His father will be angry at John.
  - c. \*His father will be angry at every boy.
  - d. ?If he wins, every athlete is happy.
  - e. ?If he wins, no man is unhappy.
  - f. If he comes, Mary will tell him why she hates John.
  - g. Because he does not feel well, Mary will tell his boss that John cannot come.
  - h. ?When he never calls her, every girl is angry at her boyfriend.
  - i. ?If they are rich, every farmer cares for his donkey.
  - j. ?If they are ill, every farmer cares for his donkey.
  - or
  - k. \*He came in and John went to the window  
but

- l. \*He came in. John went to the window.
- m. ?When you see it, shoot a rabbit.
- n. When you see it, shoot the rabbit.

The restrictions on cataphora seem straightforward: the pronoun must be in a syntactically subordinate position, and the antecedent is a constituent of the focal clause, i.e. the clause that carries the central message. On the other hand, if this restriction is satisfied, cataphora always possible, though opinions tend to diverge on the acceptability of (8d, e, i and j). Coordination and juxtaposition of sentences destroys cataphoric possibilities.

A real problem is explaining the irregularity of (8c)

- (8c) \*His father will be angry at every boy

The explanation cannot be fully semantical (see the next section) since (9)

- (9) A teacher will be angry at every boy

seems to allow a reading where 'every boy' has widest scope.

The antecedent cannot be a NP in the same clause.

A second limitation on the possibility of anaphora that has been often noted is the impossibility of having a

pronoun whose antecedent is part of the same main clause. One of the explanations of this phenomenon is that languages like English, German or French require reflexivisation in such cases. So we do not have (10), with him coreferential with John

(10) John likes him

but (11).

(11) John likes himself

And similarly not (12) with him and Peter coreferential,

(12) John described Peter to him

but (13).

(13) John described Peter to himself.

It does not seem that this constraint holds for resolution in general. Notably there are cases like (14),

(14) At 3.30, John made the phone call.

where the time of the event is decided by the temporal modifier in the same clause, and not by a preceding clause as is normally the case. Similarly in expressions like (15)

(15) In the shop, the attendant was telephoning his mother.

the initial modifier seems to set up the implicit genitive that allows us to understand who the attendant is. French topicalisation similarly may be taken as a related phenomenon, though in (16)

(16) Jean, Marie parlait souvent de lui.

it may be questioned whether 'Jean' is a constituent in the main clause. Compare also the coreferential readings of (17a-b).

(17) a. The hairdresser needs one too.  
b. Every bald man has such a brother.

### 1.7.2. Logical Restrictions

DRT is a semantical theory that explains both coreference with an indefinite antecedent, and the impossibility of referring back to indefinite (and some definite) antecedents inside a quantification in the sense of their being an identity between the referents of the pronoun and the antecedent. We do get (18),

(18) If John owns every donkey, he feeds them.

but here we have no identity. It is the second phenomenon that concerns us here: the first is not a restriction; it is rather a lacuna of other semantical theories that they cannot account for it in a satisfactory manner.

As (Chierchia and Rooth 1984) remark, the real explanation ( that is also expressed in a more syntactical way by the notion of accessibility) is that (contrary to definites and indefinites) quantifiers ( every, if then, not, or, no) do not cause a discourse variable introduced in their scope to have the same value outside the scope of the quantification.

So identification of a discourse referent involved in a quantification with a variable introduced by a pronoun does not have any effect on the truth value: the values of the discourse referent inside and outside the quantification are unrelated.

In DRT, all quantification can be reduced to the use of  $\Rightarrow$ :

every	$[s](A \Rightarrow B)$
or	$[s]([s'](A \Rightarrow I) \Rightarrow B)$
not	$[s](A \Rightarrow I)$
if then	$[s](A \Rightarrow B)$
whenever	$[s](A \Rightarrow B)$
no	$[s]([s'](A \Rightarrow [B \Rightarrow I]))$

The DRT restriction can be formulated in terms of these semantical representations. A pronoun can be coreferential with an indefinite antecedent or a dependent definite unless the antecedent occurs in an implication to which the pronoun does not belong or the antecedent belongs to

the consequence of an implication and the pronoun belongs to its condition.

It is possible to state this constraint directly on English syntax. This is however complicated and it seems better to stick to the level of semantic representation.

Similarly a pronoun can be coreferential with a term of the form 'every CN', 'any CN' or 'each CN' iff it belongs to the implication introduced by that term.

A considerable number of (sometimes only apparent) exceptions have been found against the generalisations on coreference in DRT. These are, somewhat misleadingly, known as cases of "nodal subordination". It seems premature to introduce refinements into the theory before these cases are better understood.

### 1.8. Discourse Structure

Notions of discourse structure should not be confused with DRT, even though eventually one would like to merge both approaches. In DRT the goal of the enterprise is to be able to interpret larger units than the sentence as one semantic whole; in discourse structure the goal is the development of a notion of structure within a text or dialogue which is more complex than just the linear

sequence of sentences, and which contributes to the assignment of meaning to the text as a whole. We are here interested in the claim that these structural notions have a bearing on resolution, particularly of personal pronouns, and on the choice of referential expressions at a certain stage in a text or a dialogue. The treatment below is close to (Polanyi and Scha 1984).

I will use the expression connection of a sentence in a discourse for its relation to the preceding discourse. An assignment of a connection to each sentence in a text or dialogue imposes a structure on a text or dialogue that may be compared with the structure of a sentence, and can likewise be described by a grammatical formalism that generates the text. The following is a list of possible connections:

Opening:

Sets the scene, both as beginning of text/dialogue or as first element in an interruption.

Elaboration:

An elaboration takes up a theme introduced in the previous sentence and describes it. After the elaboration one may "pop" back to the original level.

Continuation:

An utterance of this kind continues the business of the last sentence. If the last was elaborating, or opening or answering, this one also elaborates the same theme, sets the same scene, or answers the same question.

Closing:

Ends a subtask in a text.

Answering:

Answers a question.

Popping:

Moves back to an earlier (subordinate) part of the text/dialogue.

Contrasting:

Resets expectations raised by the preceding sentence.

Interrupting:

Starts a new scene, after which the old one may be popped up.

At certain structural positions, not all of the connections are possible.

This notion of discourse structure has been claimed to have many consequences for coreference ((Reichman-Adar 1984a), (Reichman-Adar 1984b), (Linde 1979), (Polanyi 1985), (Polanyi and Scha 1984)). Some of these are collected by the following schema:

Personal Pronouns:

- continuation: pronoun and antecedent in the preceding sentence have the same syntactical function in their respective sentences.
- elaboration: Coreference in the present sentence is to a non thematic element in previous sentence
- popping: like continuation, but the last sentence on the level popped to functions as the preceding sentence.
- contrasting or interrupting: coreference to antecedents in the preceding sentence is possible, but without restrictions



on the syntactical function.

**Definite reference:**

**associative:** The associated element must be available for reference by a pronoun.

**short:** 1. Mentioned before, but not available for reference by a pronoun.  
2. Standing in for a pronoun, e.g. after a contrastive pair that puts two different elements in the same syntactic function so that pronominal reference would not be resolvable.

**full:** the element has not been discussed before

**Demonstrative:**

**this:** continues or sets the theme.

**that:** closes off a theme, or refers to something in the background.

There are numerous problems in attempting to use discourse structure for automatic language processing. One of the main problems is that connection is not always syntactically marked. There are however many devices that indicate connection. In fact, reference to previously introduced elements is one of the means to indicate connection. From the point of view of resolution this appears to be circular: one needs discourse structure to resolve, and resolving to find discourse structure. But this is only the case, if we think of the two processes, resolving and building discourse structure, as isolated from each other. It is more reasonable however to think of both processes... as generating hypotheses that can mutually confirm or refute each other. In that case, failure of resolution on a particular hypothesis concerning connection may generate

a new hypothesis, that leads to a successful resolution. Confirmation of the connection hypothesis may lead to further successful resolutions.

Moreover, it is not entirely clear how to explain discourse structural notions. Apart from fairly clear intuitions, there is not much to go on. It may be that further developments in semantics will lead to a better understanding of intrasentential phenomena.

Lastly, there are many matters of detail that deserve further investigation in this area. For example, the notion of connection between clauses in a single sentence is not well understood. Work in the whole area is still fragmentary.

### 1.2. Filtering

At least as important as all the structural restrictions discussed above are more general demands on interpretation. An utterance can only make a contribution to a text or dialogue if it is pragmatically correct in the sense of (Grice 1975). This means that it must be interpretable as a true, informative and relevant utterance. Certain resolutions that conform to the conditions discussed above do not conform to this correctness notion. They may lead to statements that are nonsensical, obviously false or in

contradiction with what has been said before, and are thereby false with respect to the context. Similarly, a statement may result that is wholly irrelevant or uninformative in the present discourse. Of course, this may also happen for different reasons: faulty disambiguation or a misrecognition of illocutionary force.

The state of the art does not allow a reconstruction of the whole picture of interpretation by an automatic interpreter. Therefore, it seems inevitable that our resolvers will sometimes perform incorrectly, just as they will with respect to disambiguation and speech act recognition. It is possible to mimic some aspects of real interpretation however. Consistency checks, both with the general knowledge and the previously communicated information, may be useful eliminating certain readings. Type checks may be carried out to detect category mistakes and similar misfits. So it seems that deduction and information storage can be used to enhance the performance of the resolver.

#### 1.10. World Knowledge

Deduction and information storage are essential when it comes to resolving definite descriptions. Excepting the associated use (that may use a directly accessible constituent) to acquire a referent, or the pseudo pronominal use, descriptions are generally used to refer to items

that have been mentioned before but are currently out of local focus. The search space for antecedents is therefore larger and the information provided in the class name must be used to find the antecedent. This involves both deduction and (local and global) world knowledge. So-called full definite descriptions involve the identification of an item that belongs to common knowledge, and this, if it is to be identified at all, must be contained in the knowledge base. Associative reference, lastly, involves the coding up in the knowledge base of certain systematic relations: a cargo with a price, a truck with a location, and a route and so on. Usage of frames, that associate various functions with a concept, and instantiate the functions, whenever an instance of the concept is created, such as envisaged for the Acord knowledge base, is crucial here.

## 2. Towards a resolver

The last section has left us with a kind of theory for resolving anaphoric singular pronouns:

An anaphoric pronoun has an antecedent  $A$ , such that (i)  $A$  occurs before or in a syntactically superordinate position of the pronoun in a different clause, (ii) the variable introduced by  $A$  is accessible from the pronoun in terms of

the DRT definitions, (iii) the clause in which the pronoun occurs is connected in the appropriate manner to the clause of the antecedent and (iv) the utterance under the assumption of coreference becomes pragmatically correct and makes sense.

Mutatis mutandis, similar theories could be formulated for plural pronouns, ellipsed VP's, short descriptions and the like. It seems more useful, however, to examine the consequences of such theories for resolution algorithms.

One of the conclusions concerning an algorithm is that it needs to have access to many levels of representation: morphology, syntax, semantic representation, discourse structure, pragmatics, contextual and general knowledge.

This means that the resolution component needs access to all the other relevant components, linguistic components and dialogue structure and knowledge base. From the linguistic point of view it means a choice between having either a single level on which all grammatical and semantic information concerning a sentence is available, or having the resolution work on on a combination of structures.

Keeping track of the dialogue history, a task of the dialogue manager in the system, can be usefully exploited for dealing with dialogue structure.. The integration of

new knowledge in the temporary database, enriched by knowledge from the permanent one, makes consistency testing a possibility.

The task of the resolver is to come up with a hypothesis for a resolution for an anaphoric element in its context of occurrence. There are two possibilities for finding such candidates. The first is keeping a list of everything that can, at this point in the text, be a potential referent in a stack or another convenient data structure. As parsing proceeds this list grows under the influx of new material referred to in the sentence. The other possibility is to regard the semantical representation itself as a data structure that can be searched for potential referents in a more complicated way. This has the advantage that one does not need to build up an extra structure, an advantage that may however be undermined by the more complex search procedures. Since the resolver deals with many types of "anaphora", such a list would necessarily grow very fast, and generally most of the information would be superfluous. For example, the sentence (19)

(19) When John came home, Mary had baked a cake.

contains the following potential antecedents (the line under the linguistic representation is the corresponding

element in (20); the \$-sign indicates that the variable is not to be identified with an occurrence of it in the representation).

antecedents:

John  
john  
Mary  
mary  
John and Mary  
{john,mary}  
a cake  
x  
cake  
cake  
come home  
[\$e]come(\$e,home,\$x)  
bake a cake  
[\$e]{cake(\$x),bake(\$e,\$y,\$x)}  
cake  
cake  
home  
home  
the event of John coming home  
e  
the event of Mary baking a cake  
e''  
the state of Mary having baked a cake  
s

possible continuations:

He was hungry.  
She cut two slices.  
They ate it all in two hours.  
He liked it.  
He never eats them.  
When Bill did, Sue hadn't.  
idem  
He had had one the day before.  
He liked being there.  
Earlier, he had been sleeping in the park  
She had burnt her fingers.  
She was exhausted.

I do not claim that it is always easy to do so, but it seems possible to retrieve or reconstruct most of the antecedents from the semantic representation (20).

(20) [s][  
[e]{at(e,t),before(e,now),come(e,home,john)},  
at(s,e),before(s,now),post(e'',s),  
[e'']{cake(x),bake(e'',mary,x)}]

## 2.1. An algorithm.

The basic theory that was the result of the last paragraph can be exploited in the following way. It is essentially a search algorithm.

Once an anaphoric element is recognised, a number of structural and other conditions are identified on the basis of the structure of the element and the connection of the sentence, and a search backwards through the semantic representation is started to identify an element that meets these conditions. It is possible to cause backtracking if the incorporation in the temporary knowledge base is not successful.

The real work involved in implementing such an algorithm is in the various subroutines required. For example, "get me a predicate", necessary for VP-ellipsis, involves identifying part of a representation that is a proper predicate translation, and undoing the binding of the subject variable. One-anaphora requires one to search back for a common noun translation. They-anaphora requires the searching for a group that may be introduced by various terms in the preceding discourse.

To arrive at a satisfactory implementation it will be necessary to do a good deal of detailed descriptive work. It is not known in many cases (singular third person



pronouns are the exceptions) if any special restrictions obtain for one, so and such anaphora. It has turned out that plural pronouns are different from singular ones, but they are still badly understood.

Simultaneously, an algorithm like the one sketched above puts a number of constraints on various aspects of the system. If searching takes place in the finished semantical representation, it must be possible to identify the correlates of predicates, common nouns and adjectives in it. Similarly, the structure must reflect a good deal of discourse structure, generated by assumptions about connection. It requires a good deal of sortal information on the objects and temporalities since anaphoric elements are often marked for gender, number, state or eventhood.

From a parsing component it must be expected that it identifies anaphoric elements and generates a description of the element to be identified, e.g. its type, and sort, and maybe the list of things it cannot be identical with. Moreover, to keep track of discourse structure, it must recognise connection markers and represent them for further processing.

Chierchia, G. and Rooth, M. (1984) Configurational notions in discourse representation theory. In NELS 14, Graduate Linguistic Student Association, University of Massachusetts, Amherst, 1984.

- Grice, H. P. (1975) Logic and Conversation. In Cole, P. and Morgan, J. L. (eds.) Syntax and Semantics, Volume 3: Speech Acts, pp41-58. New York: Academic Press.
- Hinrichs, E. (1982) Temporal Anaphora in Discourses of English. Paper presented at the Ohio State University Conference on the Semantics of Tense and Aspect in Discourse, May 16-17, 1982. To appear in Linguistics and Philosophy 9 (1986).
- Hirst, G. (1981) Anaphora in Natural Language Understanding. Berlin: Springer-Verlag.
- Kamp, H. (1979) Events, Instants and Temporal Reference. In Bauerle, R., Egli, U. and von Stechow, A. (eds.) Semantics from different points of view, pp376-417. Berlin: de Gruyter.
- Kamp, H. (1981) Evenements, representations discursives et reference temporelle. Langages, 64, 39-64.
- Kamp, H. and Rohrer, C. (1983) Tense in Texts. In Bauerle, Schwarze and von Stechow (eds.) Meaning, Use and Interpretation of Language, pp250-269. Berlin: de Gruyter.
- Linde, C. (1979) Focus of Attention and the Choice of Pronouns in Discourse. In Givon, T. (ed.) Syntax and Semantics, Volume 12: Discourse and Syntax. New York: Academic Press.
- Partee, B. H. (1984) Nominal and Temporal Anaphora. Linguistics and Philosophy, 7, 243-286.
- Polanyi, L. and Scha, R. (1984) A Syntactic Approach to Discourse Semantics. In Proceedings of the 10th International Conference on Computational Linguistics and the 22nd Annual Meeting of the Association for Computational Linguistics, Stanford University, Stanford, Ca., 2-6 July, 1984, pp413-419.
- Polanyi, L. (1985) A Theory of Discourse Structure and Discourse Coherence. In 21st Meeting of the Chicago Linguistics Society, University of Chicago, Chicago, April, 1985.

Reichman-Adar, R. (1984) Extended Person-Machine Interface. Artificial Intelligence, 22, 157-218.

Reichman-Adar, R. (1984) Technical Discourse: The Present Progressive Tense, The Deictic "That", and Pronominalization. Discourse Processes, 7, 337-369.



## Chapter II

### French interrogative structures

1 The purpose of this chapter is to present an overall view of descriptive problems related to French interrogative structures. The question has revealed to be a rather intricate one and a previous clarification is necessary.

#### 1.1 Dialogue and minimal dialogue

The questions raised in chapter 1 §1.8 are characteristic of a complete and extended dialogue. They are somewhat simpler if only a minimal dialogue is envisaged, i.e a dialogue which is restricted to :

- (a) Question(s) associated with only one grammatical sentence
- (b) Answers associated with only one grammatical sentence and which answer questions as in (a)
- (c) Question(s) in (a) don't concern a previous text
- (d) The dialogue is closed after an answer such as in (b)

The following situations are thus all excluded from the minimal dialogue :

- (1) Où et à qui Jacques a donné ce livre ? A Marie ? Aux enfants ? A la campagne ? Dans la voiture ?
- (2)(a) Pierre et Marie se sont mariés  
Qui et qui se sont mariés ?
- (b) L'inflation a fait monter l'or  
Qu'est-ce que tu me racontes par là ?
- (3) On a livré les machines à écrire à mon entrepôt de Hanover. Est-il possible d'y aller pour les transporter à Paris avant la fin du mois ?

Instead, (4) will be a possible structure of a minimal dialogue :

- (4) Les machines à écrire de mon entrepôt de Hanover, est-il possible d'aller les chercher pour les transporter à Paris avant la fin du mois ?  
Non, pas avant début juin

In (1) to (4) there are anaphoric relations but the restrictions on the minimal dialogue leave out some peculiar ones in (1) to (3).

## 1.2 The (ana)phoric relation

Chapter 1 presented an overall view of phoric relations (anaphoric and cataphoric) to which ellipsis and all the forms of gapping belong : these are particular cases where the phoric element is a nil string. As stated in Chap. 1, there is some sort of resolution wanted for the phoric element, that is, its semantics is, at least partially, determined by the semantic representation of the antecedent.

In the following typical examples, there is a semantic covariation between the string  $\alpha$  (the antecedent) and the string  $\alpha'$  the (ana)phoric element :

- (1) Jacques  $\alpha$  aime les romans et Marie  $\alpha'$  ,e la poésie
- (2) Jacques aime  $\alpha$  les romans d'espionnage et Marie  $\alpha'$  , ceux d'aventure
- (3) Sa  $\alpha$  maison est plus grande que la  $\alpha'$  ,e tienne
- (4) Jacques  $\alpha$  a embrassé Marie dans le parc alors que Paul  $\alpha'$  , l'avait déjà fait à côté de la piscine
- (5) Avec  $\alpha$  le vol de cette voiture Paul a recommencé à faire  $\alpha'$  , ce qu'il avait déjà fait dans le parc
- (6) Paul  $\alpha$  a compris l'Odyssée mieux que Jacques  $\alpha'$  ,e
- (7) Si Jacques s'occupe d' $\alpha$  un chat, il  $\alpha'$  , le soigne bien
- (8)  $\alpha$  Tout homme sait qu' $\alpha'$  , il est seul
- (9) Paul a ordonné à  $\alpha$  Marie de  $\alpha'$  ,e partir ce soir
- (10)  $\alpha$  Pierre et Paul disent qu' $\alpha'$  , ils sont heureux
- (11)  $\alpha$  Paul  $\alpha'$  , se lave
- (12)  $\alpha$  Le livre que Jacques dit que Paul a lu  $\alpha'$  ,e m'ennuie
- (13)  $\alpha$  Jacques Chirac a parlé à la télévision.  $\alpha'$  , Le premier ministre a annoncé des mesures nouvelles
- (14) Jacques a fait cadeau d' $\alpha$  un livre à Pierre et celui-ci est venu avec  $\alpha'$  ,e
- (15) Jacques  $\alpha$  cria et frappa à la porte mais  $\alpha'$  , ces actions furent vaines

- (16) Jacques *cria* et Marie *,aussi*
- (17) Jacques *cria* et Marie *,fit de même*
- (18) Jacques pensait que *Pierre* était stupide alors que Marie pensait que *,le même homme* était très intelligent

The explicitation of a phoric relation requires the explicitation of :

- (19) (a) the syntactic structures associated with the antecedent and the phoric element(s)  
 (b) the syntactic conditions - if any - relating the antecedent to the phoric element(s)  
 (c) the strings representing the antecedent and the phoric element(s)  
 (d) the semantic representations associated with the antecedent and the phoric element(s)  
 (e) the kind of identity between the semantic representations of the antecedent and the phoric element(s)

We intend to develop the last point (above) further. What seems intrinsic to the phoric relation is that there must be some kind of "sameness" in the semantic representations of the antecedents and in their phoric elements. But this kind of sameness may differ from one case to another. In, for example (2), the related elements have in common the same predicates but, in a Kamp framework, they differ in the discourse referents (and the objects in the model). In (13), sameness concerns the discourse referents but not the predicates. In (3), no discourse referent is involved, sameness concerns only the semantic predicates.

In (2), (3) and (13), time is unexpressed but in other cases, whether there is sameness or not in relation to time, it must be expressed : in (4), the moment of time associated with the antecedent is necessarily different from the case for (1).

The sameness between semantic predicates, discourse referents and time are three distinct (and to some extent) independent identities : in (4) there is identity between semantic predicates and discourse referents but no identity in time ; in (7) identity in the three aspects ; in (2) only in the semantic predicates ; in (13) only in the discourse referents.

The same type of identity can be expressed by different syntactic constructions. Sameness in predicates but not in discourse referents can be associated with the nil string (Cf. (1) and (3)), or with an overt form, pronominal (Cf. (2)), or verbal (Cf. (5)). A nil string may express identity in the

three aspects (Cf. (6)), or in relation only with predicates and discourse referents, time not being concerned (Cf. (9)), or only with predicates (Cf. (3)). Pronouns are typically phoric elements which are grammatical, but there are also lexical ones, (Cf. (17) and (18)).

As identity in three different aspects seems to be the basic foundation of the phoric relation, each phoric element can be characterized by : (a) the string that represents it, with, if necessary, its syntactic structure or categorization ; (b) its antecedent ; (c) its identity or not with the three aspects discussed before : semantic predicates (SEM), discourse referents (DR), and temporal moments (TEMP) thus it is possible to note the phoric element of (1) by the subsequent formulas (one in extenso, the other in abbreviated form) :

*i*,e/,aime;=SEM,Ø DR,= TEMP/  
e/=,Ø,=/

The phoric elements of (1) to (18) can thus be represented :

- (1) *e*/=,Ø,=/
- (2) *ceux*/=,≠,Ø/
- (3) *e*/=,Ø,Ø/
- (4) *l'avait déjà fait*/=,≠,≠/
- (5) *ce qu'... avait déjà fait*/=,≠,≠/
- (6) *e*/=,=,=/
- (7) *le*/=,=,Ø/
- (8) *il*/=,=,Ø/
- (9) *e*/=,=,Ø/
- (10) *ils*/=,=,Ø/
- (11) *e*/=,=,Ø/
- (12) *e*/=,=,Ø/
- (13) *le premier ministre*/≠,=,Ø/
- (14) *e*/=,=,Ø/
- (15) *ces actions*/=,=,=/
- (16) *aussi*/=,Ø,=/
- (17) *fit de même*/=,Ø,=/



(18) *le même homme*/≠,=,∅/

In all the preceding cases we have complete identity between discourse referents and/or moments in time. This is also the case in, for example (7) and (8), where we have a universal. But quantity is a notion involved in the semantics of natural language and phoric elements can express only partial identity with the antecedent :

(20) Les enfants sont rentrés contents. Beaucoup avaient joué au foot.

In (20) *beaucoup* must be interpreted as *beaucoup parmi les enfants qui sont rentrés*, that is, there is a partial overlapping between the discourse referents of *les enfants* and those of *beaucoup*. This can be noted by "p"; so (20) can be noted as (21), or more analytically, as in (22) :

(21) <sub>1</sub>Les enfants sont rentrés contents. *Beaucoup* avaient joué au foot (*beaucoup*)/=,p=,∅/)

(22) Les <sub>1</sub>enfants sont rentrés contents. <sub>2</sub>*Beaucoup* <sub>1</sub>*e*  
<sub>2</sub>avaient joué au foot (<<sub>1</sub>*e*/=,∅,∅/; <sub>2</sub>*beaucoup*/=,p=,∅, >

The preceding cases don't take into account all the types of (ana)phoric relation in natural languages. Typically, there are also the cases related to a metalanguage use of language, as in (23) and cases where the phoric element bears a negation of some aspect of its antecedent, as in (24) :

(23)(a) <sub>1</sub>*Giorgione* était <sub>1</sub>*ainsi* nommé parce que son père aimait les rimes en one

(b) Marie ne croyait pas à <sub>1</sub>*l'amour* bien que <sub>1</sub>*ce mot* lui plût

(c) Jacques écrivit <sub>1</sub>*à Marie et à Jeanne* dans <sub>1</sub>*cet ordre*

(d) <sub>1</sub>*Pierre et Jacques* ont donné un livre à <sub>1</sub>*respectivement Marie et Suzanne*

(24)(a) Les <sub>1</sub>*fleurs* <sub>2</sub>*blanches* plaisent à Marie et les <sub>2</sub>*autres* <sub>1</sub>*e* à Jacques

(b) Jacques est <sub>1</sub>*très attaché à son pays* mais Pierre est <sub>1</sub>*différent*

## 1.3 Interrogative structures, questions and answers

Of the following sentences, only (25 a) presents an interrogative structure associated with a question, (25 b) may be semantically classified as a question - it is in any case a possible input to a KB - but with no interrogative structure and (25 c) is an interrogative structure with no question

- (25)(a) A quelle heure viendra le camion?  
 (b) Dites-moi les prix de location des voitures  
 (c) Pouvez-vous ouvrir la fenêtre ?

The following sections will be dedicated to interrogative structures, independently of their mappings with questions. Structures other than interrogative ones, even if mapped with questions will not be treated.

Both questions in (26) and (27) must be answered semantically by one and the same answer

- (26)(a) Marie est-elle aimée par Pierre ?  
 (b) Pierre aime-t-il Marie ?  
 (27)(a) Par qui Marie est aimée?  
 (b) Qui aime Marie?

But syntactically (28 c) is a rather dubious answer to (26 a)), or at least much less natural than (28 a, b), and (28 g) is excluded, however they are perfectly correct answers to (26 b). On the other hand, (29 a) answers (27 a) and (29 b) answers (27 b) and not the other way round.

- (28)(a) Oui, Marie est aimée par Pierre  
 (b) Oui, Marie elle est aimée par Pierre  
 (c) Oui, Pierre aime Marie  
 (d) Non, Marie n'est pas aimée par Pierre  
 (e) Non, Marie n'est pas aimée par Pierre mais par Jacques  
 (f) Non, par Jacques  
 (g) Non, Jacques  
 (29)(a) Par Pierre  
 (b) Pierre

From the preceding examples, it is apparent that the anaphoric *e* will be strongly represented in answers.

Naturalness in dialogue, even in the minimal one, requires the respect of syntactic conditions on the answer determined by the syntactic structure of the question; in this respect, specifications of e anaphoric elements in answers seem crucial and analogous to similar cases in coordination and comparatives.

## 2. A survey of interrogative French structures

In this §2 we attempt to present an overall view of descriptive question related to French interrogative structures. This will be presented in the limits of a minimal dialogue, in a descriptive language as neutral as possible in relation to competing linguistic models, and in an organized pattern.

### Notational conventions

In the following, a GPSG notation is adopted, but all the entities in formulas are purely descriptive ones. They are defined in an ostensive way, so, for example, "NP" means "those constructions which linguists call "NP" with possible examples added.

The observations are organized in formulas, each one being a mapping between input structures and output ones. Each formula expresses the following general relation : if there is an input declarative structure covered by the conditions on the input, there is also a corresponding grammatical interrogative structure in the output. Note that the input does not define the grammaticality conditions of declarative sentences: these are supposed rather than stated. The general aim is to point out difficulties added by interrogatives. Optional constituents are noted in parenthesis, and it is assumed that if an optional constituent is in the input, it must also be in the output. An index to the left notes phoric relations; to the right, it simply identifies a constituent.

### Constituents :

S	:	sentence
NP	:	nominal phrase
Prep	:	preposition
PP	:	prepositional phrase
Adv	:	adverb
Adj	:	adjective
Pron	:	pronoun; je, tu, il and the related fem. and plural forms
Aux	:	auxiliary
VR	:	verbal root
XP	:	variable over constituents
que[PRO]	:	constant <i>que</i> , initial of a sentence with a trace related to <i>que</i> as in : <i>les enfants que Marie regarde</i>
que[Comp]	:	constant <i>que</i> , initial of sentence with no trace related to : <i>que</i> as in <i>il dit que Jean est ici</i>

### Functional qualifications of constituents :

ARG	:	argumental (all verb arguments except the subject; topicalised constituents are not argumental)
Sm	:	sentence modifier
NOM	:	nominative
OBJ	:	objective
DAT	:	dative
OBL	:	oblique
,NOMP	:	referential nominal predicate, as in <i>Jacques est le président</i> , with the reading : the object associated with <i>Jacques</i> is the same object as the one associated with <i>le président</i>
NOMP	:	non referential nominal predicate, as in <i>Jacques est ingénieur</i>

Evaluation argument : as in *il coûte dix francs, il s'éloigne de dix pas.*

Qualifications of constituents :

Sentences :

r	:	root
Af	:	affirmative
Neg	:	negative
emb	:	embedded
Cond.S	:	conditional sentence

Nouns :

num	:	numerable
timen	:	time nouns : <i>mois, heures</i>
Money	:	
units	:	francs, dollars ...

Adjectives :

numer	:	numeral
Indef	:	indefinites

Adverbs :

Q1	:	environ, à peine
Q2	:	beaucoup, assez

Morphological entities :

+Pl	:	plural
-Pl	:	singular
fl	:	inflected
inf	:	infinitive
pres	:	present
imp	:	imperfect
Cond	:	conditional

The anaphoric elements, when necessary, are noted as indicated in §1.2. Moreover :

f maps from (a) to (b). Examples :

f<sub>aff/int</sub> : (a) sait-il ouvrir la porte?  
(b) il sait ouvrir la porte

f<sub>aff/int-neg</sub> : (a) ne sait-il pas ouvrir la porte?  
(b) il sait ouvrir la porte

f<sub>neg/int-neg</sub> : (a) ne sait-il pas ouvrir la porte?  
(b) il ne sait pas ouvrir la porte

Formulas proposed in the following can be true or false in a flat way. Needless to say, the ideal situation of exhaustivity is not attained. Perhaps one of their main advantages is to point out the intricacy of descriptive problems related to French interrogatives. We think that from now on it will be possible to take advantage of a structural pattern and that it will be possible to cover remaining gaps - even if they are large ones, rather than continue to juxtapose observations or to present formalised solutions of (very) restricted subsets of problems.

Five major types of interrogative structures are distinguished : INV-1, INV-2, INV-3, IE, DIE. Each one presents several types and sub-types. For each sub-type, we present in the following a formula which is intended to cover the main descriptive conditions of the corresponding interrogative structure. In several cases (INV-1, INV-3i-a, *qui*) it was also possible to extend the observations covered by the formula and to point out problems in more complicated structures, such as negation, coordination and disjunction. In these cases, the conditions on answers were also explored.

## 2.1 INV-1

## 2.1.1 Syntactic description

The general formula of INV-1 is :

(1) [ (<sub>1</sub>NP<sub>1</sub>)<sub>1</sub> · NP<sub>1</sub> ... β-fl ... ]  
α

NP<sub>1</sub> = NP [NOM, PRON]

α = r-Af-S [fl]

β = Aux or VR

⇒ [ (<sub>1</sub>NP<sub>1</sub>) β-fl <sub>1</sub> · NP<sub>1</sub> ... ]  
α

The present formula covers all the following typical examples :

- (2) (a) Voyage-t-il par avion?  
 (b) L'entreprise Durand envoie-t-elle la marchandise par bateau?  
 (c) A-t-il conduit des poids lourds?  
 (d) Est-il question d'envoyer des machines par le train?

The core defining feature of INV-1 is :

A pronominal subject - specifically *je tu il* and the corresponding plural and/or feminine forms, but no other pronouns - is placed immediately to the right of the word that has the morpheme of inflection as an internal constituent; the following strings are thus all excluded :

- (3) (a) \* Voyager-il par avion?  
 (b) \* A-celui-ci conduit des poids lourds?  
 (c) \* A-lui conduit des poids lourds?  
 (d) \* Est sûr il d'envoyer des machines par le train?

## 2.1.2 The phoric relation

In a formulated INV-1 sentence the phoric relations seem to be the same as those of the corresponding affirmative ones.

- (4) (a) <sub>1</sub>Il sait <sub>1</sub>'e conduire des poids lourds  
           Sait-<sub>1</sub>il <sub>1</sub>'e conduire des poids lourds?  
 (b) <sub>1</sub>Pierre <sub>1</sub>'<sub>1</sub>il sait <sub>1</sub>'<sub>2</sub>e conduire des poids lourds

1 Pierre sait-il conduire des poids lourds?

- (c) 1 Il est sûr d'envoyer ces marchandises par train  
Est-il sûr d'envoyer ces marchandises par train?

### 2.1.3 Négation

Phrase negation is covered by (1), if the constraints on  $\alpha$  are relaxed; with  $\alpha$  as in (5) :

- (5)  $\alpha = r\text{-Af}$  or  $\text{Neg-S}[f]$

the following typical cases are all allowed :

- (6) (a) Ne voyage-t-il pas par avion?  
(b) L'entreprise Durand n'envoie-t-elle pas la  
marchandise par bateau?  
(c) N'a-t-il pas conduit des poids lourds?  
(d) N'est-il pas sûr d'envoyer des marchandises par le  
train?

Constituent negations is excluded from the subject by the constraints on  $\text{NP}_i$  of (1) :

- (7) (a) Aucun ne voyagera par avion  
(b) \* Ne voyagera aucun par avion

### 2.1.4 Coordination

Coordination of INV-1 sentences is grammatical, as is coordination of the corresponding affirmative sentences :

- (8) (a) Il sait conduire des poids-lourds et il connaît la  
route de Lyon à Milan  
(b) Sait-il conduire des poids-lourds et connaît-il la  
route de Lyon à Milan?  
(c) Il est bon d'envoyer les marchandises par train et  
il est facile de le faire  
(d) Est-il bon d'envoyer les marchandises par train et  
est-il facile de le faire?

When there are the same pronouns in the two conjuncts, and if there is no particular situation with a particular gesture, there is a phoric relation of type / $\text{=,=},\emptyset$ / between them (cf. (11)), though it is also possible to have different pronouns with different discourse referents :

- (9) Sait-il conduire des poids lourds et connaît-elle la  
route de Lyon à Milan?



When <sub>1</sub>NP of (1) is not nil, it can be omitted from the second conjunct - not from the first one - but, even if omitted, the NP of the first conjunct will be in phoric relation with the inverted <sub>1</sub>NP of the second sentence :

- (10) <sub>1</sub>L'entreprise Durand envoie-t-<sub>1</sub>elle ses marchandises par train et paye-t-<sub>1</sub>elle comptant ?
- (11) \* Envoie-t-<sub>1</sub>elle ses marchandises par train et <sub>1</sub>l'entreprise Durand paye-t-<sub>1</sub>elle comptant?

The conjunction of an INV-1 sentence with a negative INV-1 sentence seems very doubtful.

- (12) ? Ne connaît-il pas la route et sait-il conduire des poids-lourds?

Furthermore there are severe restrictions in the possible conjuncts of an INV-1 sentence. An INV-1 sentence is the conjunct of neither (cf. (13)) an affirmative sentence, nor of a negative sentence (cf. (14)), nor of an interrogative sentence other than an INV-1 one (cf. (15)) even if it is a confirmative one (cf. (16)) :

- (13) \* Sait-il conduire des poids-lourds et (il, Pierre) connaît la route de Lyon à Milan?
- (14) \* Sait-il conduire des poids-lourds et (il, Pierre) ne connaît pas la route de Lyon à Milan?
- (15)(a) \* Sait-il conduire des poids-lourds et qui connaît la route de Lyon à Milan?
- (b) \* Sait-il conduire des poids-lourds et c'est qui qui connaît la route de Lyon à Milan ?
- (16) \* Sait-il conduire des poids lourds et est-ce qu'il (Pierre) connaît la route de Lyon à Milan?

Constituent conjunction, with exception to the subject, works in INV-1 sentences as in the corresponding affirmative ones :

- (17)(a) Il sait conduire des motos et des poids-lourds  
 (b) Sait-il conduire des motos et des poids-lourds?  
 (c) L'entreprise Durand possède des entrepôts à Lyon et à Milan  
 (d) L'entreprise Durand possède-t-elle des entrepôts à Lyon et à Milan?

Coordination in the inverted pronominal subject is excluded, though the <sub>1</sub>NP of (1) admits it with the corresponding consequences on the phoric <sub>1</sub>NP :

- (18) (a) Lui et elle savent conduire des motos  
 (b) \*Savent il (lui) et elle conduire des motos?  
 (c) *1* Pierre et Marie *1* ils savent conduire des motos  
 (d) *1* Pierre et Marie savent-*1* ils conduire des motos?

### 2.1.5 Disjunction

Disjunction INV-1 sentences obey by and large the same restrictions as coordinated ones. Pragmatic factors rather than syntactic ones will impose a *vel* or an *and/or* interpretation. The same factors seem to impose some differences in relation to coordination of INV-1 sentences (cf. (19 c and d)) :

- (19) (a) Sait-il conduire des poids-lourds ou connaît-il quelqu'un qui sache le faire?  
 (b) Est-il bon d'envoyer les marchandises par train ou est-il facile de le faire?  
 (c) ? Sait-il conduire des poids-lourds ou connaît-il la route de Lyon à Milan?  
 (d) *1* L'entreprise Durand envoie-t-*1* elle ses marchandises par train ou a-t-*1* elle décidé d'utiliser l'avion?  
 (e) ? Ne connaît-il pas la route ou sait-il conduire des poids-lourds?  
 (f) \* Sait-il conduire des poids-lourds ou (il, Pierre) connaît la route de Lyon à Milan?  
 (g) Sait-il conduire des motos ou des poids-lourds?  
 (h) \* Savent il ou elle conduire des motos?  
 (i) *1* Pierre ou Marie savent *1* ils conduire des motos?

### 2.1.6 The mapping of INV-1 structures with questions

All INV-1 structures are not confirmation questions. The expected answers of (20) and (21) are not of the *yes or no* type, but rather the indications of how to proceed in doing something, and the name of the person who is going to deliver the oranges, respectively.

- (20) Puis-je savoir comment envoyer cette marchandise de Lyon à Milan?  
 (21) Savez-vous qui viendra livrer les oranges?

The main factors which seem to determine if not all, at least a wide range of mappings from INV-1 structures to questions are :

- (22) (a) The person of the inverted pronoun  
 (b) The class of the root verb and of its embedded verbal phrase(s)

For establishing the pertinent regularities the verbs may be classed as :

- (23) (a) *pouvoir*, (b) Cl I : *dire, expliquer ...* ; (c) Cl II : *savoir, connaître ...*

The following regularities seem at work :

- (24) (a) If the inverted pronoun is [PER 3], the INV-1 structures maps to a confirmative question  
 (b) If the inverted pronoun is [PER 1] or [PER 2] & - if the INV-1 structure is covered by one of the formulas in (6), the INV-1 structure maps to the type of question which must be associated with  $\sigma$  - otherwise, the INV-1 structure maps to a confirmative question  
 (c-1) VR [pouvoir] NP, [PER 1] VR [Cl II, INF]  $\sigma$   
 (c-2) VR [pouvoir] NP, [PER 2] (pronoun) VR [Cl I, INF]  $\sigma$   
 (c-3) VR [Cl III] NP, [PER 2]  $\sigma$  [VP...]

The following examples illustrate (24 a) and (24 b) :

- (25) (a) Peut-il savoir comment transporter des oranges?  
 (b) Sait-il qui est venu?  
 (c) Puis-je savoir comment envoyer cette marchandise?  
 (d) Puis-je savoir qui a fait cet emballage?  
 (e) Pouvez-vous me dire comment transporter des oranges?  
 (f) Pouvez-vous m'expliquer qui m'a envoyé cette caisse?  
 (g) Savez-vous qui m'a envoyé cette caisse?  
 (h) Savez-vous où dédouaner cette marchandise?  
 (i) Puis-je envoyer cette marchandise à Florence?  
 (j) Connaissez-vous la route?

### 2.1.7 Answers

An INV-1 structure admits different types of answers. Between the phoric(s) of the answer and the antecedent(s) in the question there are particular relations in each type. We review in the following what seem to be the principal patterns.

The possible string answers to (28) are indicated in (29). *Oui* and *non* are typical instances of larger classes indicated in (30) and (31) respectively.

(26) Voyage-t-il par avion?

- (27) (a) Oui  
 (b) Non  
 (c) Oui, il voyage par avion  
 (d) Non, il ne voyage pas par avion

(28) (Certainement, sans doute, affirmatif, tout-à-fait)

(29) (Non, sûrement pas, certainement pas, négatif ...)

(27 a) and (27 b) may be preceded by a modal and in (27 c) and (27 d), a proper noun or a definite description may take the place of *il* :

- (30) (a) Probablement oui  
 (b) Probablement non  
 (c) Oui, Jacques voyage par avion  
 (d) Non, le père de Martine ne voyage pas par avion

Questions related to (30) will not be developed here. It is assumed that there must be a nil phoric to the right of *oui* and *non* in (27 a) and (27 b). It will express the scope of these words and it must be related to (26).

The *e* of (27 b) cannot be associated with the semantics of the overt phrase in (27 d); if it were, (27 b) would be interpreted as :

(31) Oui, il voyage par avion

Thus, for obtaining (27 a) to (27 d) we propose the formulas (32 b) and (32 c) :

- (32) (a)  $\text{!Voyage-t-il par avion ?}$   
 (b)  $\text{Oui} \left\{ \begin{array}{l} \text{!} \cdot e/f_{af}/int(i\theta) ; =, =, =/ \\ ; f_{af}/int(i\theta) \end{array} \right\}$   
 (c)  $\text{Non} \left\{ \begin{array}{l} \text{!} \cdot e/f_{af}/int(i\theta) ; =, =, =/ \\ ; f_{neg}/int(i\theta) \end{array} \right\}$

The possible string answers associated with (32) are indicated in (33) :

(33)  $\text{!Ne voyage-t-il pas par avion ?}$

- (34) (a) Si  
 (b) Si, il voyage par avion  
 (c) Non

(d) Non, il ne voyage pas par avion

The following formulas are thus proposed (in parallel to (32 b) and (32 c)) :

(35) (a) 
$$\text{Si} \left\{ \begin{array}{l} 1. e/far/int-neg (1\sigma) ; =, =, =/ \\ ; far/int-neg (1\sigma) \end{array} \right\}$$

(b) 
$$\text{Non} \left\{ \begin{array}{l} 1. e/far/int-neg (1\sigma) ; =, =, =/ \\ ; fneg/int-neg (1\sigma) \end{array} \right\}$$

3.3 Possible string answers related to (36) are indicated in (37) :

(36)  $_1$  Sait-il conduire des poids-lourds et  $_2$  connaît-il la route de Lyon à Milan ?

- (37) (a) Oui  
 (b) Oui, il sait conduire des poids-lourds et il connaît la route de Lyon à Milan  
 (c) Non  
 (d) Non, il ne sait pas conduire des poids-lourds et il ne connaît pas la route de Lyon à Milan  
 (e) Non, il sait conduire des poids-lourds, mais il ne connaît pas la route de Lyon à Milan  
 (f) Non, il connaît la route de Lyon à Milan, mais il ne sait pas conduire des poids-lourds  
 (g) Il sait conduire des poids-lourds mais il ne connaît pas la route de Lyon à Milan (and in a different order)  
 (h) Oui à la première question, non à la deuxième ( and in a different order)

We proposed the following for the phoric element of (37 a) and (37 c) :

(38)  $_1 e/far/int (1\sigma) ; =, =, =/ \& _2 e/far/int (2\sigma) ; =, =, =/$

The pattern of answers to disjunctive questions is not at all clear. Tentatively, we propose the following one :

- (39) (a) If the disjunctive question has a clear *vel* status, the answers by *oui* or *non* are not accepted (cf. (40))  
 (b) If the disjunctive question has an *and/or* status, answers by *oui* and *non* must be interpreted as if the *ou* of the question was an *et* (cf. (41)).
- (40) (a) A-t-il décidé de partir ou a-t-il préféré rester?  
 (b) \* Oui; \* Non  
 (c) Voudra-t-il envoyer cette marchandise par train ou préférera-t-il l'avion?  
 (d) \* Oui; \* non

- (41)(a) Sait-il conduire des poids-lourds ou connaît-il la route de Lyon à Milan?  
 (b) Oui, il sait conduire des poids-lourds mais il ne connaît pas la route de Lyon à Milan

The semantics of an affirmative answer to a confirmative question are essentially the information that all the truth conditions of the questions are verified, otherwise, the answer must be *non*.

There is thus some radical ambiguity associated with a negative answer, because it can be justified by the failure of any of the truth conditions of the union of truth conditions that the question must verify for being true. If, for example, the answer to (42) is *non*, this may be because one or all of the separate constituents of (42) or any combination of them is not true. So (42) may have an answer *non* if any of the situations in (43) are true :

(42) Pierre a-t-il donné hier un livre à Marie et à Suzanne?

- (43)(a) Jacques a donné le livre  
 (b) Pierre a emprunté le livre  
 (c) Pierre a donné le livre avant-hier  
 (d) Jacques a emprunté le livre avant-hier  
 (e) Pierre a emprunté une revue à Jeanne  
 (f) Personne n'a rien fait à personne

The ambiguity of negative answers by *non* is in conflict with exhaustivity, and is frequently continued with an expression that makes clear which truth conditions failed; so expected negative answers of (42) will be :

- (44)(a) Non, un cahier  
 (b) Non, (il l'a fait) avant-hier  
 (c) Non, c'est Jacques (qui l'a fait)  
 (d) Non, seulement à Suzanne  
 (e) Non, il le leur a (seulement) prêté

Note that the following are all excluded, even if the corresponding situations imply a failure of the truth conditions :

- (45)(a) \*Non, c'est Jacques à Thérèse  
 (b) \*Non, avant-hier à Suzanne

Thus there are severe syntactic restrictions on the way of satisfying exhaustivity in a negative answer to a confirmative question. It seems that the following regularities are at work here :

- (46)(a) The overt string to the right of *non* is syntactically associated with one surface

constituent of the question (and in some non clear cases to two or more contiguous ones).

- (b) The phoric elements associated with this string are of the / = , = , Ø / or of the / = , = , / types and conjointly with the next string they must "cover" all the affirmative sentences corresponding to the question.

These conditions are far from determining all aspects related to desambiguating strings in negative answers. At least two more factors must be added, illustrated in (48) :

- (47) (a) Pronominalisation is accepted for other constituents than the subject  
 (b) The string in the answer cannot be related to any constituent of the interrogative structure
- (48) (a) Pierre a-t-il donné hier un livre à Marie ?  
 (b) Non, il l'a fait avant-hier  
 (c) Non, Pierre l'a fait avant-hier  
 (d) Non, il lui a donné un cahier  
 (e) Non, il a donné un cahier à Marie  
 (f) Pierre a-t-il dit que Marie avait acheté un cahier rouge ?  
 (g) \* Non, bleu

## 2.2 INV-2

The general formula of INV-2 is

- (1)  $\alpha$ [NP<sub>i</sub>...VR- $\beta$  ...]  
 $\beta$  = fl or pp  
 NP<sub>i</sub> = NP<sub>i</sub> [NOM, ≠ Pron]  
 $\alpha$  = r-Af-S[fl] or emb-S[fl, ≠ que [Comp]]  
 $\Rightarrow$  \*  $\alpha$ [...VR- $\beta$  NP<sub>i</sub> ...]

In a root sentence, (1) covers the typical examples of (2) :

- (2)(a) \* Connait Marie la route  
 (b) \* A connu Marie la route  
 (c) \* Est aimé Marie par Pierre  
 (d) \* A été donné un livre à Marie  
 (e) \* A téléphoné Marie à Pierre  
 (f) \* S'est réveillée Marie à cinq heures

In an embedded sentence we have results as follows :

- (3)(a) \* Il sait a donné Marie les livres à Paul  
 (b) \* Je me demande va ta mère bien

(1) excludes sentences such as :

- (4)(a) \* A connu il la route  
 (b) \* A été donné il à Marie  
 (c) \* Il sait a donné il les livres à Marie

If an interrogative element is displaced initially, the sentence becomes grammatical :

- (5)(a) Que connaît Marie?  
 (b) A quelle heure s'est réveillé Marie?  
 (c) Il sait quel livre a lu Marie



## 2.3 INV-3

INV-3 covers a wide range of French interrogative structures which, for the most part, are very common. They all have in common the inversion of the form C(e) that functions as a subject of a 3rd person simple form of the verb être. This sort of INV-3 common use is implemented with variations depending on the topicalised constructions admitted to the left of C(e) and on the constituent - of a predicate nominal type - to the right of the form of être. It is thus possible to distinguish the subsequent types and sub-types of INV-3, 3i-a, 3i-b, 3ii, 3iii-a, 3iii-b. Their affirmative counterparts are illustrated by (1) to (5), respectively.

- (1) Pierre, c'est mon ami  
Un livre, c'est agréable  
Chez moi, c'est calme  
Que Jacques soit sorti, c'est bien  
La maison de Jacques, c'est ici
- (2) C'est mon ami, Pierre  
C'est agréable, un livre  
C'est calme, chez toi  
C'est bien, que Jacques soit ici  
C'est ici, la maison de Jacques
- (3) Si Marie ne vient pas demain, c'est que Jacques est malade  
La vérité, c'est que Marie ne viendra pas
- (4) C'est Marie que Pierre aime  
C'est bien que Pierre ait vu le film
- (5) C'est Marie qui est venue hier

In respect to mappings onto interrogative counterparts, the construction *c'est* obeys restrictions, related to *c'(e)* and/or the form of the verb être, which are absent from analogous constructions with equivalent semantics and similar syntactics. The two forms of (6) are grammatical, but INV-3, with restrictions on être is a possible mapping of (6 a), where there is no analogous inversion of (6 b), (cf. (7)) :

- (6)(a) Les enfants, la fête, les amis, c'est merveilleux.  
(b) Les enfants, la fête, les amis, cela est merveilleux.
- (7)(a-1) Les enfants, la fête, les amis, a été ce merveilleux?  
(a-2) \* Les enfants, la fête, les amis, a ce été merveilleux?

- (a-3) \* Les enfants, la fête, les amis, a été ce merveilleux?  
 (b-1) \* Les enfants la fête, les amis, est-ce merveilleux?  
 (b-2) \* Les enfants, la fête, les amis, a cela été merveilleux?  
 (b-3) \* Les enfants, la fête, les amis, a été cela merveilleux?

The form C(e) has particular agreement relations : (8 a) and (8 b) are both acceptable, Note that c in (8 c) though a phoric of *moi*, does not respect person agreement. The interrogative (8 d) seems less natural than (8 e) :

- (8)(a) Les auteurs du livre, c'est eux  
 (b) Les auteurs du livre, ce sont eux  
 (c) Moi, c'est moi  
 (d) Les auteurs du livre, sont-ce eux?  
 (e) Les auteurs du livre, est-ce eux?

In the following, we shall explicit what seem to be the principal patterns concerning the inversion of the C(e) element. Because, as indicated the aim is to have an overall view of interrogative French structures, the inputs to the mapping are not intended to cover all affirmative structures with a C(e) like element preceded by a topicalised construction. Constructions like those in (6 b) or (9 a), owing to the impossibilities of (7 b-1) to 7 b-3) and (9 b) will not be considered. In general, a rather permissive view is adopted concerning the conditions on être, so samples as those in (8 d) will be covered but (9 d) will be excluded.

- (9)(a) A l'école, ça obéit  
 (b) \* A l'école, obéit ça ?  
 (c) L'auteur du rapport, ce sera Jacques  
 (d) \* L'auteur du rapport, sera ce Jacques ?

### 2.3.1 INV-3i-a

#### 2.3.1.1 Syntactic description

The general formula of INV-3i-a is

- (10)  $\alpha [ ({}_1XP_1) {}_c\text{'être}[B] XP_2 ]$

$\beta$  = Pres or Imp or Cond  
 $XP_2 \neq$  que S  
 $\alpha$  = r-Af-S[f1]

- $\nexists \alpha [ ({}_1XP_1) \text{être}[B] \wedge XP_2 ]$

*a?*

(10) covers the samples in (1) and also those in (11), where, thanks to the absence of  $XP_i$ ,  $c'(e)$  is a deictic pronoun. There is no phoric relation with a antecedent string :

- (11) Est-ce Jacques?  
Est-ce loin de Paris?

### 2.3.1.2 The phoric relation

When  $XP_i$  is not nil, it is the antecedent of the phoric,  $ce$ , which is of type  $/=, =, =$  or  $\emptyset/$ . The relation holds for all persons and all selectional features of  $XP_i$  despite the fact that  $ce$  requires  $[F3, -P1]$  as syntactic features.

- (12) Un camion, est-ce bien solide ?  
Ce conducteur, est-ce ton cousin?  
Par là, est-ce rapide?

$ce$  is the subject of  $XP_j$  and this character shares the same phoric relations as ordinary subjects :

- (13) (a) *Jacques est l'ami de Pierre dont je t'avais parlé*  
(b) *Jacques, est-ce l'ami de Pierre dont je t'avais parlé ?*

### 2.3.1.3. Negation

Sentence negation is allowed but constituent negation is excluded.

- (14)(a) Jacques, n'est-ce pas ton ami?  
(b) Jacques, ce n'est pas ton ami?  
(c) \* Aucune maison, n'est-ce pas ici?  
(d) \* Jacques, n'est-ce aucun ami?

But there is a confirmation negative question, with no affirmative grammatical counterpart, represented in

- (15)(a) L'entreprise Durand est très efficace, n'est-ce pas?  
(b) L'entreprise Durand n'est pas très efficace, n'est-ce pas?  
(c) \* L'entreprise Durand est très efficace, c'est?

The general formula may be represented by the following :

- (16)  $r-(Af, Neg)-_iS[f1], n'est-ce/=, =, =/ pas$

## 2.3.1.4 Coordination

Coordination of INV-3i-a sentences is possible (Cf. 17 a and b) but it is excluded when XP<sub>1</sub> is represented by the same string in the two conjuncts (cf 17 c) or when the first one is deleted (17 d). Constituent coordination of XP<sub>1</sub> and XP<sub>2</sub> is possible, but excluded for the ce element

- (17)(a) Le transport fluvial, est-ce bon marché et le transport aérien, est-ce rapide?
- (b) Le transport fluvial, est-ce bon marché et est-ce efficace?
- (c) \* Le transport fluvial, est-ce bon marché et le transport fluvial, est-ce efficace?
- (d) \* Est-ce bon marché et l'entreprise Durand, est-ce efficace?
- (e) L'entreprise Durand et le transporteur Schiffer, est-ce en France?
- (f) Le transport par train, est-ce autorisé et bon marché?
- (g) \* Le transport par train et le transport par route, est-ce et ce autorisé?

Coordination of an INV-3i-a sentence with an affirmative or negative one is excluded (cf (18)). If the conjunct is an interrogative one, coordination is in principle also excluded (cf. (19 a, b)) except in the following cases : the conjunct is an INV-3 sentence (cf (19 c and f)) or it is an INV-1 sentence with a topicalised constituent which is the same as the one in the INV-3i-a sentence conjunct; this constituent will be the antecedent of a phoric other than the inverted pronominal subject of the second conjunct (cf. (17 g et h) :

- (18)(a) \* La distance d'ici à Paris, est-ce grande et le prix c'est élevé ?
- (b) \* La distance d'ici à Paris, est-ce grande et le prix n'est pas élevé ?
- (19)(a) \* La distance d'ici à Paris, est-ce grande et qui peut y aller?
- (b) \* La distance d'ici à Paris, est-ce grande et sait-il y aller?
- (c) \* Le transport fluvial, est-ce bon marché et est-ce bien que l'entreprise l'ait utilisé? (INV 3-i-b)

- (d) Le transport fluvial, est-ce bon marché et est-ce qu'on va l'adopter? (INV-3ii)
- (e) Le transport fluvial, est-ce bon marché et est-ce en France qu'on l'a réglementé récemment? (INV-3iii-a)
- (f) Le transport fluvial, est-ce bon marché et est-ce Paul qui l'a recommandé? (INV-3iii-b)
- (g) \* Le transport fluvial, est-ce bon marché et est-il rapide?
- (h) Le transport fluvial, est-ce bon marché et peut-on l'utiliser couramment?

#### 2.3.1.5 Disjunction

Disjunction of INV-3i-a sentences obey by and large the same restrictions as coordinated ones. Pragmatic conditions rather than syntactic ones will impose a *vel* or an *and/or* interpretation. Examples (20 a) to (20 f) are parallel to those of (19 c) to (19 h) :

- (20)(a) Le transport fluvial, est-ce efficace ou est-ce nécessaire que l'entreprise fasse un autre choix?
- (b) Le transport fluvial, est-ce suffisant ou est-ce qu'on aura recours au train?
- (c) Le départ de ce camion, est-ce pour demain, ou est-ce à après-demain qu'on le fixera?
- (d) Le départ de ce camion, est-ce pour aujourd'hui ou est-ce Paul qui le conduira demain?
- (e) \* Le départ de ce camion, est-ce pour aujourd'hui ou peut-il être laissé pour demain.
- (f) Le départ de ce camion, est-ce pour aujourd'hui ou peut-on le laisser pour demain?

#### 2.3.1.6 The mapping of INV-3i-a structures with questions

We haven't found any INV-3i-a interrogative structures which must not be mapped onto a confirmation question, the same is true for structures obeying to the formula in (1 c).

#### 2.3.1.7 Answers

The more generally expected answers to INV-3i-a sentences are naturally *oui* or *non*, but expressions of (28) and (29) of § 1 are also possible. The general formulas representing

answers to single INV-3i-a sentences (i.e. without coordination or disjunction) are the following :

(21)(a) INV-3i-a (=  $i\sigma$ )

(b) Oui  $\left\{ \begin{array}{l} i \cdot e / f_{ar} / int (i\sigma); =, =, = / \\ , f_{ar} / int (i\sigma) \end{array} \right\}$

(c) Non  $\left\{ \begin{array}{l} i \cdot e / f_{ar} / int (i\sigma); =, =, = / \\ , f_{Neg} / int (i\sigma) \end{array} \right\}$

It is thus possible to answer affirmatively the question in (22) by (23) and negatively by (24)

(22) La maison Durand, est-ce en France?

(23)(a) Oui

(b) Oui, la maison Durand c'est en France

(24)(a) Non

(b) Non, la maison Durand ce n'est pas en France

Interrogative structures specified by (10) must also be mapped onto confirmation question, observe that (25) is not acceptable :

(25) \* Puis-je savoir comment envoyer cette marchandise, n'est-ce pas?

In this case, the general formulas representing the answer are the following :

(26)(a) r-Af- $i$ S[f1], n'est- $i$ .ce pas?

r-Af- $i$ S[f1] =  $i\sigma$

(b) (oui, si)  $\left\{ \begin{array}{l} i \cdot e / i\sigma; =, =, = / \\ i\sigma \end{array} \right\}$

(c) non  $\left\{ \begin{array}{l} i \cdot e / i\sigma; =, =, = / \\ , f_{Neg} / ar (i\sigma) \end{array} \right\}$

It is thus possible to have for the question in (27), the affirmative answers of (28) and the negative ones of (29)

(27) La maison Durand est en France, n'est-ce pas?

(28)(a) Oui

(b) Oui, la maison Durand est en France

- (29)(a) Non  
 (b) Non, la maison Durand n'est pas en France

When the expression *n'est-ce pas* is to the right of a negative sentence, no *oui* string in the answer is expected, so we have

- (30)(a) r-Neg-<sub>1</sub>S[f1], n'est-<sub>1</sub>·ce pas?  
 r-Neg-, S[f1] = <sub>1</sub>σ

- (b) non  $\left\{ \begin{array}{l} \text{<sub>1</sub>·e/far/Neg (<sub>1</sub>σ); =, =, =/} \\ \text{<sub>1</sub>σ} \end{array} \right\}$

- (c) en effet  $\left\{ \begin{array}{l} \text{, <sub>1</sub>σ} \\ \text{<sub>1</sub>·e/<sub>1</sub>σ<sub>1</sub>=, =, =/} \end{array} \right\}$

Possible answers to (15 b) are as in (31) which express a positive confirmation of the question (the expression of disagreement to structures such as (15 b) will not be treated here)

- (31)(a) Non  
 (b) Non, l'entreprise Durand n'est pas très efficace  
 (c) En effet, l'entreprise Durand n'est pas très efficace

The answers (23) and (24) to the question (22) are not the only possible ones; those in (32) are also acceptable, but observe that in (33) all answers are unacceptable :

- (32)(a) Oui, la maison Durand est en France  
 (b) Non, la maison Durand n'est pas en France  
 (33)(a) Chez toi, est-ce agréable?  
 (b) \* Oui, chez toi est agréable  
 (c) La maison Durand et son service livraison, est-ce ici?  
 (d) \* Oui, la maison Durand et son service livraison est ici

It is thus necessary to specify two mappings from the INV-3i-a sentence onto the affirmative and negative structures in the answers. They are formulated in (34),

- (34)(a) far-etre/int est-ce ;  
 iNP<sub>1</sub>[P3] être-ice XP<sub>1</sub>  
 \*NP<sub>1</sub>[P3, α P1] être [P3, αP1] XP<sub>1</sub>

- (b)  $f_{\text{Neg-}\alpha\text{etre/Int est-ce}}$  :
- ${}_1\text{NP}_i[\text{P3}] \text{ n'etre-ice pas } \text{XP}_j$
- $\rightarrow \text{NP}_i[\text{P3}, \alpha\text{P1}] \text{ n'etre } [\text{P3}, \alpha\text{P1}] \text{ XP}_j$

(21) is completed by (35)

(35) If  $f_{\alpha\text{-}\alpha\text{etre/Int est-ce}}$  can apply to INV-3i-a (=, S), then the following are also possible answers :

- (a) Oui,  $f_{\alpha\text{-}\alpha\text{etre/Int est-ce}}$
- (b) Non  $f_{\text{Neg-}\alpha\text{etre/Int est-ce}}$

Given (34), (33 a) will be excluded and (33 c) will admit the answers of (36) :

- (36)(a-1) Oui
- (a-2) Oui, la maison Durand et son service livraison c'est ici
- (a-3) Oui, la maison Durand et son service livraison sont ici
- (b-1) Non
- (b-2) Non, la maison Durand et son service livraison, ce n'est pas ici
- (b-3) Non, la maison Durand et son service livraison ne sont pas ici

### 2.3.2 INV-3i-b

The general formula of INV-3i-b is :

- (1)  ${}_1\alpha[\text{c'etre}[\beta], \text{XP}_i[\text{NOMP}] \text{ XP}_j]$
- $\beta = \text{Pres or Imp or Cond}$
- $\alpha = \text{r-Af-S[f1]}$
- $\text{XP}_j \neq \text{que}[\text{PRO}] \text{ S}$
- $\rightarrow \alpha\text{c'etre } [\beta]\text{-ce } \text{XP}_i[\text{NOMP}] \text{ XP}_j$

This formula covers the following typical examples :

- (2)(a) Est-ce mon ami Pierre?
- (b) Est-ce agréable un livre?
- (c) Est-ce bien que Jacques soit ici?
- (d) Est-ce bien d'aller au cinéma?

A sentence such as in (3) is not covered by INV-3i-b but by INV-3iii-a :



(3) Est-ce le président que Jacques a vu?

### 2.3.3 INV-3ii

The general formula of INV-3ii is :

(1)  $\alpha$ [ \*(<sub>1</sub>Cond-S) <sub>1</sub>.c'être [B] que[Comp] S]

$\alpha$  = r-Af-S[f1]

B = Pres or Imp or Cond

$\nrightarrow$  $\alpha$ [ (<sub>1</sub>Cond-S) être[B]-<sub>1</sub>.ce que[Comp] S]

The optional conditional sentence in (1) covers the sample

(2) Si Jacques vient demain, est-ce que Marie le saura?

But (1) is much more productive when the input is an ungrammatical sentence (sentences (3 a) and (3 c) are acceptable as answers to causal questions (cf. ) and in particular speech acts) :

(3)(a) \* C'est que Paul viendra

(b) Est-ce que Paul viendra?

(c) \* C'est que Marie a acheté un livre à Paul

(d) Est-ce que Marie a acheté un livre à Paul?

### 2.3.4 INV-3iii-a and INV-3iii-b

INV-3iii concerns the inversion of *ce* in cleft sentences. The two proposed formulas are :

(1)  $\alpha$ [c'être [B] <sub>1</sub>XP  $\alpha$ [que[PRO] ...,it ...]

$\alpha$  = r-Af-S[f1]

B = Pres or Imp or Cond

$\nrightarrow$  $\alpha$ [être[B]-ce <sub>1</sub>XP<sub>s</sub>[<sub>1</sub>que[PRO] ...,it ...]

(2)  $\alpha$ [c'être [B] <sub>1</sub>NP  $\alpha$ [qui,...]

$\alpha$  = r-Af-S[f1]

B = Pres or Imp or Cond

$\nrightarrow$  $\alpha$ [être[B]-ce <sub>1</sub>NP<sub>s</sub>[<sub>1</sub>qui,...]

The following samples of (3) and (4) are covered by (1) and (2) respectively :

- (3)(a) Est-ce par bateau que cette marchandise doit être envoyée
- (b) Est-ce à Londres que cette marchandise doit être envoyée?
- (c) Est-ce des cacahuètes que l'entreprise Durand envoie à Naples?
- (4)(a) Est-ce l'entreprise Durand qui a pris la commande?
- (b) Est-ce le conducteur qui a été pénalisé?

## 2.4 Interrogative elements

IE is a label for a set of mappings which specify interrogative constituents. In the following, we review three particular IEs, namely ; qui, quoi, combien. They will be observed in relation to the syntactic conditions associated with them and particularly, in relation to the prepositions à, de, en, dans, avec, pour, par. Obviously we aim at exhaustivity, but in the first instance a limited (but representative) number of prepositions have been analysed in order to formulate coherent generalisations. We will assume in this section that there is no more than one IE by sentence, for questions relating to the presence of two or more IEs in the same sentence, Cf. 86.

## 2.4.1 QUI

The general formula is :

- (1) ... (Prep) NP<sub>x</sub>[NOM [-PI] or OBJ or OBL[ARG or Sm] or ;NOMP;  
+HUM; -CLIT] ...  
⇒... (Prep) qui[NP <sub>x</sub>[r]] ...

It is apparent from (1) that qui has a quite wide distribution :

- (2)(a) Qui est venu?  
(b) Pierre aime qui?  
(c) Il ressemble à qui?  
(d) Vous parlez de qui?  
(e) On commence par qui?  
(f) Le bureau a été fait par qui?  
(g) Il est haï de qui?  
(h) Il a acheté la voiture de qui?  
(i) Il travaille avec qui?

Typical cases excluded by (1) are :

- (3)(a) \* Qui sont venus?  
(b) \* Il est qui? (with an expected answer of the type *Il est gardien ou Il est le gardien* [PRN])  
(c) \*Il distribue un chocolat par qui? (cf. *Il distribue un chocolat par enfant*)  
(d) \* Il a acheté un manteau de qui? (With an expected answer of the type *Il a acheté un manteau d'enfant*)  
(e) \*Jacques qui lui envoie? (cf. *Jacques le lui envoie*)

Within the interrogative sentence, the constituent *qui* has the same phoric relations as the corresponding non interrogative one.

- (4)(a) Jacques a promis de voyager
- (b) Qui a promis de voyager?
- (c) Jacques a ordonné à Pierre de voyager
- (d) Jacques a ordonné à qui de voyager?

*Qui* accepts phrase negation and constituent negation with the exception of the subject.

- (5)(a) Qui n'a pas travaillé?
- (b) Jacques n'a donné aucun livre à qui?
- (c) Qui n'a apporté aucun livre?
- (d) \* Aucun enfant n'aime qui?

Coordination of a *qui* sentence with a non *qui* one seems excluded :

- (6)(a) \* Qui aime Jean et Marie aime Pierre?
- (b) \* Jean aime qui et Marie aime Pierre?
- (c) \* Jean aime qui et Marie n'aime pas Pierre?

But coordination of *qui* sentences are much more acceptable :

- (7)(a) Qui aime Jean et qui aime Marie?
- (b) Jean aime qui et qui aime Marie?

Constituent coordination in other constituents than the *qui* one functions as ordinary constituent coordination (cf. (8)), but constituent coordination with a *qui* constituent is totally excluded (cf. (9)) :

- (8)(a) Qui aime Marie et Jeanne?
- (b) Pierre et Marie aiment qui?
- (9)(a) \* Qui et Jean aiment Marie?
- (b) \* Marie qui et Jean?
- (c) \* Qui et qui aime(nt) Marie?
- (d) \* Jean donne un livre à qui et à qui?
- (e) \* Qui et aucun ami n'aime Jean?

Both disjunctions, the phrasal and the constituent one, are excluded :

- (10)(a) \* Pierre ou qui aime(nt) Marie?
- (b) \* Pierre aime Marie ou qui aime Marie?

An answer to a *qui*-sentence represented in (11 a) is a sentence represented in (11 b-1) or (11 b-2)

(11)(a)  $(\text{1}\sigma)$  (Prep) qui[NP [NOM[-P1]  $\alpha$ [ or OBJ or OBL[ARG or Sm] or  
 $\text{1NOMP}$  ; +HUM] ( $\text{2}\sigma$ )

(b-1)  $\text{1}\cdot\text{e}/=,=\emptyset$  or = / (Prep) NP[NOM [ $\pm$ P1]  $\alpha$ [ $\sigma$ ]  
 $\text{2}\cdot\text{e}/=,=\emptyset$  or = /

(b-2)  $\text{1}\sigma$  (Prep) NP[NOM [ $\pm$ P1]  $\alpha$ [ $\sigma$ ]  $\text{2}\cdot\text{e}/=,=\emptyset$  or = /

if  $\text{1}\sigma = \emptyset$  :

(b-3) NP[NOM [ $\pm$ P1]  $\alpha$ [ $\sigma$ ]  $\text{2}\sigma$

(11) is intended to cover adequately all the following cases :

- (12)(a) Qui a ouvert la porte?  
 (b-1) {Pierre, un garçon} a ouvert la porte  
 (b-2) Pierre et Paul ont ouvert la porte  
 (b-3) Pierre  
 (b-4) Les enfants  
 (c) Pierre a donné un livre à qui hier?  
 (c-1) A Suzanne  
 (c-2) A Marie et à Suzanne  
 (c-3) Aux enfants  
 (c-4) Pierre a donné un livre aux enfants  
 (c-5) Pierre a donné un livre aux enfants hier  
 (c-6) \* Aux enfants hier

#### 2.4.2 QUOI

The general formulas are :

(1)

... (Prep) NP[ $\alpha$ [OBJ or OBL[ARG or Sm] -HUM, -CLIT] or  $\beta$ [NOMP[+HUM]],...

Prep NP  $\neq$  Sm of place, time or manner

$\rightarrow$  ... (Prep) quoi [ $\alpha$ [ $\sigma$ ] or  $\beta$ [ $\sigma$ ]],...





- (g) Il l'a embauché pour finir le travail
- (h) Il l'a embauché pour faire quoi ?
- (i) Il a oublié de téléphoner à Marie
- (j) Il a oublié de faire quoi ?
- (k) Il cherche à joindre Paul
- (l) Il cherche à faire quoi ?

(3)(a) typically excludes examples like :

- (10)(a) Il est intelligent
- (b) \* Il fait quoi ?
- (c) Il lit un livre à Marie
- (d) \* Il fait quoi à Marie ?
- (e) Il donne un livre à Marie dans la librairie
- (f) \* Il fait quoi à Marie dans la librairie ?

Formula (2)(b) brings out the inverted form *quoi faire*. As concerns prepositional usage, *pour* seems relatively more flexible than *à* or *de*. Both (2)(a) and (2)(b) can apply to *pour* :

- (11)(a) Il l'a embauché pour finir le travail
- (b) Il l'a embauché pour faire quoi ?
- (c) Il l'a embauché pour quoi faire ?
- (d) Il l'a embauché pour finir le travail dans la grange
- (e) Il l'a embauché pour faire quoi ?
- (f) Il l'a embauché pour faire quoi dans la grange ?
- (g) Il l'a embauché pour quoi faire ?
- (h) Il l'a embauché pour quoi faire dans la grange ?

Inverting the infinitive and *quoi* for *à* and *de* seems slightly more doubtful :

- (12)(a) Il a oublié de venir
- (b) Il a oublié de faire quoi ?
- (c) ? Il a oublié de quoi faire ?
- (13)(a) Il cherche à te joindre
- (b) Il cherche à faire quoi ?
- (c) ? Il cherche à quoi faire ?

If we pursue this application one step further, we notice that an elliptical usage of *faire* is possible with *pour*, but this quite obviously has something to do with written french differences between *pour quoi* (for what) and *pourquoi* (why), as the ellipsis can only apply to *pourquoi* (Cf.(3)(c)). A and *de* forbid this :



- (14)(a) \* Il l'a embauché pour quoi ?  
 (b) Il l'a embauché pourquoi ?
- (15)(a) \* Il a oublié de quoi ?  
 (b) \* Il cherche à quoi ?

Deleting *quoi* in all three cases remains impossible :

- (16)(a) \* Il l'a embauché pour faire ?  
 (b) \* Il a oublié de faire ?  
 (c) \* Il cherche à faire ?

Finally, in the formula (2)(d) the deletion of both the preposition and the infinitive seems feasible and natural for *à* and *de* (in the situation where the listener misses the whole last part of the question), but totally excluded for *pour* :

- (17)(a) Il a oublié | de venir  
                                   | quoi ?
- (b) Il cherche | à te joindre  
                                   | quoi ?
- (c) Il l'a embauché | pour finir le travail  
       \*                                  | quoi ?

### 2.4.3 Combien

(1)(a)

$$\dots \alpha[(\text{Prep}_1) (\text{Adv}_1[\text{Q1}]) \left\{ \begin{array}{l} \text{Adj} [\text{num}] \\ \text{Adv} [\text{Q2}] \\ \text{Adj} [\text{indef}] \end{array} \right\} (\text{de}) N_1[\text{num}, \pm\text{Pl}]]$$

$$\alpha[\sigma] \neq \beta[\sigma], \gamma[\sigma]$$

†... (Prep<sub>1</sub>) (Adv<sub>1</sub> [Q1]) combien de N<sub>1</sub>[num, +Pl]

If N<sub>1</sub> = N<sub>1</sub>[timen, num, ±Pl]

$$\dagger\dots (\text{Prep}_1) (\text{Adv}_1[\text{Q1}]) \text{ combien } \left\{ \begin{array}{l} \text{de temps} \\ \text{de } N_1[\text{timen}, \text{num}, \pm\text{Pl}] \end{array} \right\}$$

(1)(b)

...<sub>B</sub>[(Prep<sub>1</sub>) (Adv<sub>1</sub>[Q1]) Adj[Numer] (de) (N<sub>1</sub>[Num, ±P1])

If <sub>B</sub>[σ] = NOMF

or if Prep<sub>1</sub> = (pour, avec) & N<sub>1</sub> = [Money Units]

or if Prep<sub>1</sub> = à & N<sub>1</sub> = [Money Units] & <sub>B</sub>[σ] = NomAdj

‡...[(Prep<sub>1</sub>) (Adv<sub>1</sub>[Q1]) combien (de N<sub>1</sub>[num, +P1])

(1)(c)

...V<sub>1</sub>[α sf],...<sub>x</sub>[(Prep<sub>1</sub>) (Adv<sub>1</sub>[Q1]) { Adj[Numer] } (de) N<sub>1</sub>[num, ±P1, α sf]  
 { Adv[Q2] }  
 { Adj[Indef] }

<sub>x</sub>[σ] = Evaluation argument

‡...V<sub>1</sub>[α sf],...<sub>x</sub>[(Prep<sub>1</sub>) (Adv<sub>1</sub>[Q1]) combien (de N<sub>1</sub>[num, +P1])

(2)

...tous Art[+P1] N[num, +P1],...

‡...tous Art [+P1] combien

(3)

...V[être, 3-P1, 1+P1] le Adj[Num, -p1]

‡...V[être, 3-P1, 1+P1]] le combien

(4)

...V[être, ±p1] Art N [num, ±p1],...

‡...V[être, ±p1] Art { combien...  
 quatrième  
 combienième  
 combienième }



- (h) Il a vu combien de personnes ?
- (i) Il part dans deux jours
- (j) Il part dans combien de jours ?
- (k) 10 personnes étaient là
- (l) Peu de personnes étaient là
- (m) Plusieurs personnes étaient là
- (n) Combien de personnes étaient là ?
- (o) Le bureau a été fait par 2 personnes
- (p) Le bureau a été fait par combien de personnes ?

Since in formula (1)(a) we have  $\alpha[\sigma] \neq \beta[\sigma], \gamma[\sigma]$ , then the following typical examples are excluded, because **combien** + string is optional in those cases :

- (7)(a) \* Il pense à (environ) combien ?
- (b) \* Il l'a vu combien ?
- (c) \* Combien étaient là ?
- (d) \* Le bureau a été fait par combien ?

(when expected answers to the above examples have to do with people)

- (e) \* Il part dans combien ?
- (f) Le film dure combien de temps ?
- (g) Le film dure combien ?

(6)(f) & (g) are grammatical structures, but do not belong to formula (1)(a) as the string is optional. They are covered by formula (1)(c) which takes into account the verb and various corresponding selectional features. (See below)

Examples containing an N to do with TIME, as in (6)(i) can be transformed into two question forms :

- (8)(a) Il part dans combien de jours ?
- (b) Il part dans combien de temps ?

Prepositions in some cases may be optional for Sm of time, but in (8) above this would cause a change of meaning :

- (9)(a) Il part dans combien de temps ?  
(In how much time is he leaving ?)
- (b) Il part combien de temps ?  
(How long is he going away for ?)

(9)(c) & (d) however, are semantically identical :

- (9)(c) Il l'a loué 2 mois



- (d) Ça coûte dix francs
- (e) Ça coûte combien de francs ?
- (f) Ça coûte combien ?
  
- (g) Il a reculé de trois pas
- (h) Il a reculé de combien de pas ?
- (i) Il a reculé de combien ?

As the remaining three formulas (2),(3) & (4) concern restricted cases, we explicit examples here very briefly ;

- (13)(a) Ça arrivait tous les 3 mois
- (b) Ça arrivait tous les combien ?
  
- (c) { Nous sommes } le 20 juillet  
       { On est }
- (d) { Nous sommes } le combien ?  
       { On est }
  
- (e) Elle est la dixième (de la classe)
- (f) Elle est la combien (de la classe) ?  
       quantième

## 2.5 DIE

The interrogative elements of 2,4 can be displaced to an initial position of a root in embedded sentence ; we have than a D(isplaced) I(nterrogative) E(lement), that is, a particular case of unbounded dependencies,

In this section we attempt to give a noted statement of descriptive conditions concerning DIE.

Verbs can be categorized in the five classes of (1); each one is illustrated in (2) to (6) (observe in (7) a verb which does not belong to  $V_{xxx}$ );

(1) Verb categorisation :

$V_1$  : ...NP Verb[fl or inf]  $\sigma_1$  [V[inf],...]

$\sigma_1 = ((\text{Prep}) \text{NP}) (\text{Prep})$

$V_{11}$  : ...NP Verb[fl or inf] [<sub>1</sub>DIE V[inf],...]<sub>1</sub>t...

$V_{III}$  : ...<sub>i</sub>DIE NP Verb[f1]  $\sigma_j$  [V[Inf]...it...]

$\sigma_j$  = ((Prep) NP)

$V_{IV}$  : ...NP Verb[f1]  $\sigma_j$  [que[Comp]...]

$V_V$  : ...NP Verb[f1 or inf] [<sub>i</sub>DIE ...V[f1]...it...]

- (2) Jacques veut acheter un livre  
Jacques dit vouloir acheter un livre
- (3) Jacques sait sur qui compter  
Jacques dit savoir sur qui compter
- (4) Où Jacques souhaite partir?  
Où Jacques a ordonné à Marie d'aller?
- (5) Jacques a dit à Marie que la vie est belle  
Jacques souhaite que Marie s'en aille
- (6) Jacques ignore avec qui Marie est partie
- (7) Pierre s'engouffre dans le couloir acheter un livre  
\* Quel livre Pierre s'engouffre dans le couloir acheter?

With (1) it is possible to define embedded sequence, in (8) :

- (8) Emb-seg<sup>g</sup> : A sequence of any number of VP[Inf] and/or S[f1] such that each one is embedded in a  $V_i$  or  $V_{iv}$ , respectively.

Embedded sequences are illustrated in (9) :

- (9)(a) J. dit souhaiter exiger à Marie d'acheter un livre  
(b) J. affirme avoir promis d'accepter de jouer aux boules  
(c) J. dit que P. affirme que M. pense que Ch veut que S. vienne  
(d) J. dit penser que P. souhaite acheter un livre  
(e) J. affirme avoir promis d'accepter que M. joue aux boules  
(f) J. dit penser que P. souhaite demander à Marie qu'elle achète un livre.

The general formula for DIE is :

$\alpha$  [ ...  $V_j$  ... (1) ...  $XP_1$  ... ]

$\exists \alpha$  [  ${}_1$ DIE  $\beta$  ...  $V_j$  ... (1) ...  $t_1$  ... ]

(a) If  ${}_1$ DIE = *que*[PRO], then  $\beta$  = (Clit)  $\gamma$ -(f1 or inf)

$\gamma$  = (Aux or VR)

(b)  $XP_1 \neq$  NP[ENOMP]

(c) Open coordination excepted,  ${}_1$ DIE is unique

(d)  $XP_1$  is not a constituent of a relative sentence

& (e) If  $\alpha$  =  $VP_1$ [inf] (Emb-seg)

(e,1)  $\alpha$  is embedded in a  $V_{11}$

& (e,1)  $t_1$  is any constituent of  $\alpha$

& (f) If  $\alpha$  = S[f1] (Emb-seg)

(f,1)  ${}_1$ DIE  $\neq$  *quoi*

& (f,2,1)  $\alpha$  = r- $S_1$ [f1]

&  $t_1$  is in r- $S_1$ [f1]

or (f,2,2)  $V_j$  is a  $V_{111}$

&  $t_1$  is in a  $V$ [inf] (Emb-seg) embedded in  $V_j$

&  $\alpha$  = r-DIE-S[f1]

or

$\alpha$  is embedded in  $V_{\downarrow}$

or (f,2,3)  $V_j$  is a  $V_{1\downarrow}$

&  $t_1$  is in an S[f1] (Emb-seg) embedded in  $V_j$

&  $\alpha$  = r-DIE-S[f1]

or

$\alpha$  is embedded in  $V_{\downarrow}$



(10) covers the following examples :

- (11)(a) J, sait avec qui partir
- (b) A qui J, a téléphoné?
- (c) J, sait avec qui et où partir
- (d) \* J, sait avec qui où partir
- (e) \* A qui les livres que J, a donnés sont beaux?
- (f) \* Que J, a regardé?
- (g) Qu'a regardé Jacques?
- (h) \* Qui Jacques ne sait pas parler?
- (i) Il sait quoi faire
- (j) \* Quoi J, sait?
- (k) A qui Pierre a téléphoné?
- (l) A qui c'est que P, a téléphoné?
- (m) Avec qui J, a promis à P, d'essayer de jouer?
- (n) J, Ignore avec qui P, a promis d'essayer de jouer
- (o) Quel livre J, pense que P, a dit que S, a lu?
- (p) Quel livre J, dit penser que P, souhaite demander à Marie d'acheter?

Other DIE as such in (12) are not covered by precedent formulas :

- (12)(a) A qui est-ce que Marie a donné un livre?
- (b) Combien est-ce, le n-ième?

## 2.6 Interactions between the different types of interrogative structures

The interrogative structures described from 2.1 to 2.5 cover the following examples :

- (1)(a) Il a téléphoné à qui?
- (b) Qui a téléphoné?
- (c) A qui il a téléphoné?
- (d) A-t-il téléphoné?
- (e) Est-ce qu'il a téléphoné?
- (f) Jacques a-t-il téléphoné?
- (g) Je me demande à qui il a téléphoné?
- (h) Je me demande à qui téléphoner?
- (i) Est-ce qu'il a téléphoné?
- (j) C'est à qui qu'il a téléphoné?
- (k) Est-ce à Jacques qu'il a téléphoné?

Two of the preceding structures can be at work in the same sentence :

- (2)(a) A qui a-t-il téléphoné?
- (b) Qui a téléphoné à qui?
- (c) A qui est-ce qu'il a téléphoné?

But it is not true that any two interrogative structures can participate in the construction of one sentence :

- (3)(a) \* Est-ce à qui qu'il a téléphoné?  
 (b) \* Sait-il que Jacques a téléphoné à qui?

Observe also the following situation ; (4 b) and (4 c) are possible interpretations of (4 a). Though (4 d) is a grammatical sentence, it does not admit (4 f) as a possible interpretation so in this case there is no ambiguity.

- (4)(a) J, ignore à qui P, a dit que M, a téléphoné  
 (b) J, ignore à <sub>1</sub>qui P, a dit <sub>1</sub>t que M, a téléphoné  
 (c) J, ignore à <sub>1</sub>qui P, a dit que M, a téléphoné <sub>1</sub>t  
 (d) J, ignore à qui on a dit quel livre acheter  
 (e) J, ignore à <sub>1</sub>qui on a dit <sub>1</sub>t quel livre acheter  
 (f) \* J, ignore à <sub>1</sub>qui on a dit quel livre acheter <sub>1</sub>t

The interactions between the different interrogative structures are expressed in (5). The table of (5) must be read from left upwards. The possible values of a case are : \*,+,-,±. These symbols occur respectively :

- \* The formulas are so defined, that the two structures can not co-occur in the same sentence (Example ; an INV-1 sentence cannot be the input of the formula of INV-3i-a sentences).
- + The two structures must occur in the same sentence (Example ; an INV-2 root sentence must have a DIE).
- ± The two structures can occur in the same sentence (Example ; an INV-1 sentence can have a DIE).

T(otal) ; the scope of the value is the whole sentence and not only the one where the observed structures appear; the default value is the last one. Example : it is impossible to have an IE in the whole sentence where an INV-1 sentence appears, and not only in this one.

Local ; DIE and its trace are in the same sentence.

Extr ; DIE is in one sentence, a <sub>1</sub>t is in another embedded one.

Examples of (5) can be found in (6); numbers refer to the corresponding cases in (5).

	IE (1)	DIE		INV-1	INV-2		INV-3				
		Local	Extr.		n-S	emb-S	3i-a	3i-b	3ii	3iii-a	3iii-b
INV-1	1 -T	2 ±	3 ±	4 *	5 *	6 ±	7 *	8 *	9 *	10 *	11 *
	12 -T	13 ±	14 -	15 *	16 *	17 -	18 *	19 *	20 *	21 *	22 *
	Emb-S	24 +	25 -	26 ±	27 ±	28 *	29 *	30 ±	31 ±	32 ±	33 ±
INV-2	34 -T	35 -	36 *	37 *	38 *	39 *	40 *	41 *	42 *	43 *	44 *
	45 -T	46 ±	47 ±	48 *	49 *	50 ±	51 *	52 *	53 *	54 *	55 *
	56 -T	57 -	58 -	59 *	60 *	61 ±	62 *	63 *	64 *	65 *	66 *
INV-3	67 -T	68 ±	69 -	70 *	71 *	72 ±	73 *	74 *	75 *	76 *	77 *
	78 -T	79 ±	80 -	81 *	82 *	83 ±	84 *	85 *	86 *	87 *	88 *
	89 ±	90 ±	91 ±	92 = 1	93 = 12	94 = 23	95 = 34	96 = 45	97 = 56	98 = 67	99 = 78
DIE	100 II	101	102	103 ±	104 ±	105 -	106 *	107 -	108 *	109 +	110 *
	111 II	112 III	113 III	114 ±	115 -	116 -	117 *	118	119	120	121

(5)

- (1) IE ≠ [NP[NOM]]<sub>i</sub>; α = emb-S  
 I : A IE forbids a DIE to its left.  
 II : A DIE forbids a IE to its right.  
 III: A ;DIE<sub>i</sub> forbids a DIE<sub>j</sub> to the left of <sub>i</sub>;t<sub>j</sub>.

- 1 \* Regarde-t-il quoi ?  
\* Sait-il que Jacques regarde quoi?
- 2 Avec qui viendra-t-il?  
Viendra-t-il avec Pierre?
- 3 Avec qui a-t-il dit que J. viendra?  
A-t-il dit que J. viendra avec Marie?
- 6 A-t-il dit quels livre a regardé Marie?  
A-t-il dit quels livres Marie a regardé?
- 12 \* A regardé Marie quel livre?  
\* A dit Marie que Pierre a regardé quel livre?  
\* Quel livre a dit Marie que Pierre a regardé où?
- 13 \* A regardé Marie quel livre?  
Quel livre a regardé Marie?
- 14 \* A dit Marie que Pierre a regardé quel livre?  
Quel livre a dit Marie que Pierre a regardé?
- 17 \* Où a dit Marie quel livre Pierre a regardé?  
\* Où a dit Marie quel livre a regardé Pierre?  
cf. Où Marie a dit quel livre Pierre a regardé?
- 23 ?\* Il a dit à qui quel livre a regardé Marie?  
\* Il a dit quel livre a regardé Marie à qui?
- 24 Il a dit quel livre a regardé Pierre.  
\* Il a dit à regardé Pierre quel livre
- 25 \* Il a expliqué à qui a dit Pierre que Marie a téléphoné  
\* Il a expliqué a dit Pierre que Marie a téléphoné à qui
- 26 A-t-il à qui a téléphoné Marie?  
Il a dit à qui a téléphoné Marie
- 27 \* A qui a dit Marie avec qui a joué Jacques?  
Marie a dit à Pierre avec qui a joué Jacques?
- 30 Est-ce vrai que Marie a dit avec qui a joué Pierre?
- 31 Est-ce que Marie a dit avec qui a joué Pierre?
- 32 Est-ce sur la place que Marie a dit qu'a joué Pierre?  
Est-ce sur la place que Marie a dit que Pierre a joué?
- 33 Est-ce Pierre qui a dit avec qui Pierre a joué?  
Est-ce Pierre qui a dit avec qui a joué Pierre?
- 34 \* La maison, est-ce où?  
\* Jacques, est-ce qui?

- 35 \* Où la maison est-ce?  
\* Qui Jacques est-ce?
- 45 \* Est-ce comment Jacques?  
\* Est-ce bien que Jacques soit où?
- 46 Est-ce ici la maison de Jacques?  
Où est-ce la maison de Jacques?
- 47 A qui est-ce bien que Jacques ait téléphoné?
- 50 Est-ce bien que Jacques ait dit avec qui a joué Pierre?  
Est-ce bien que Jacques ait dit avec qui Pierre a joué?
- 56 \* Est-ce que Jacques a téléphoné à qui?  
\* Est-ce que Jacques a dit que Pierre a téléphoné à qui?
- 57 \* A qui est-ce que Jacques a téléphoné? (with que [Comp])
- 58 \* A qui est-ce que Jacques a dit que Pierre a téléphoné? (with que [Comp])
- 61 Est-ce que Jacques sait quel livre regarde Marie?  
Est-ce que Jacques sait quel livre Marie regarde?
- 67 \* Est-ce à qui que Marie a téléphoné?  
\* Est-ce à Pierre que Marie a donné quel livre?
- 68 Est-ce à Pierre que Marie a téléphoné?  
A qui est-ce que Marie a téléphoné?
- 69 \* Quel livre est-ce à Pierre que M. a donné?
- 72 Est-ce à Pierre que L. a dit quel livre Marie regarde?  
Est-ce à Pierre que M. a dit quel livre regarde Marie?
- 78 \* Est-ce qui qui téléphonera ce soir?  
\* Est-ce Paul qui téléphonera ce soir?
- 79 Qui est-ce qui téléphonera ce soir?  
Est-ce Paul qui téléphonera ce soir?
- 80 \* Quel livre est-ce Paul qui donnera?
- 83 Est-ce Paul qui a dit à qui a téléphoné Pierre?  
Est-ce Paul qui a dit à qui P. a téléphoné?
- 89 Qui a donné quel livre à qui?  
But ;  
? Qui a dit que Jacques a téléphoné à qui?
- 90 \* Quel livre qui a donné?

- ?\* Quel livre Pierre a donné à qui?  
?\* A qui Pierre a donné quel livre?
- 91 \* Quel livre Pierre a dit à qui que M. a acheté?  
cf. Quel livre Pierre a dit à Paul que M. a acheté?  
\* Avec qui Pierre a dit où que M. avait joué?  
cf. Avec qui Pierre a dit à la réunion que M. avait joué?
- 100 \* Que donne qui à Pierre?  
\* Que donne Pierre à qui?
- 103 \* Que il a dit?  
Qu'a-t-il dit?
- 104 \* Que Marie a regardé?  
Qu'a regardé Marie?
- 105 \* J. ne sait pas que Marie a regardé  
\* J. ne sait pas qu'a regardé Marie  
cf. J. ne sait pas ce qu'a regardé Marie
- 107 \* Que c'est Jacques?  
\* Qu'est-ce Jacques?
- 109 \* Que c'est que Jacques a vu?  
Qu'est-ce que Jacques a vu?
- 111 \*<sub>1</sub> Quel livre Pierre a dit que M. a donné et à qui?  
\*<sub>1</sub> Avec qui P. a dit quand que M. a joué et?
- 112 & 113 (III is intended to apply to any two possible DIE)
- \* J. ignore <sub>1</sub> à qui on a dit <sub>1</sub> quel livre acheter et et  
cf. J. ignore <sub>1</sub> à qui on a dit et <sub>1</sub> quel livre acheter et  
J. a dit <sub>1</sub> à qui P. avait promis et de jouer avec Marie  
<sub>1</sub> Avec qui J. a dit <sub>1</sub> à qui P. avait promis et de jouer et?
- 114 Avec qui a-t-il joué?  
A-t-il joué avec Pierre?
- 115 \* A qui a donné Marie un livre?  
A qui Marie a donné un livre?
- 116 Il ne sait pas à qui Marie a donné un livre  
\* Il ne sait pas à qui a donné Marie un livre

## 2.7 Interrogative structures not investigated

The following examples illustrate other interrogative structures than the preceding ones which have not yet been investigated ;

- (1) Pourquoi Jacques lit un livre?
- (2) Comment est-ce que Jacques lit un livre?
- (3) Comment envoyer des périodiques?
- (4) Des journaux, combien on en reçoit?
- (5) Il veut savoir ce que Pierre a dit

Cf. in chapter 3 a UCG treatment of infinitival indirect interrogatives.





## Chapter III

### Descriptive questions on interrogative structures in the present state of the UCG formalism

#### 1 Auxiliaries

The UCG-description of auxiliaries assumed so far (Cf. A UCG French Grammar, Linguistic Specification, April 10, 1986) was intended to cover the following typical constructions :

- (1) {a, ont} {donné, regardé, choisi, ...}  
{est, sont} {sorti(s), sortie(s), ...}  
{est, a été} {donné(e), regardé(e), ...}  
{sont, ont été} {donné(e)s, regardé(e)s, ...}

The agreement questions in these constructions and those related to objective clitics in perfect forms, e.g. in :

- (2) G Les livres ont été donnés  
\* Les revues ont été donnés  
G Jacques l(a) a vue  
\* Jacques l(a) a vu  
G Jacques les a vus  
\* Jacques les a vu

were correctly accounted for by means - among others - of the following lexical entries, where the auxiliaries in (4) work as functors of a verb (PSPA, PSPE, PAS), of a nominal predicate (PNOM) or of été (Cf. (3)) :

- (3) regardé: PSPA  
<\$:np:[NBRE(x)]:\$,OBJ:np:[lex(y)]:post,NOM:np:[NBRE(x)]:pre>  
[e]REGARDE(e, x, y)
- sorti: PSPE  
<\$:np:[sg(x)]:\$,NOM:np:[masc,sg(x)]:pre>  
[e]SORT(e, x)
- sortie: PSPE  
<\$:np:[sg(x)]:\$,NOM:np:[fem,sg(x)]:pre>  
[e]SORT(e, x)
- regardés: PSPA  
<\$:np:[NBRE(x)]:\$,OBJ:np:[pro,masc,pl(y)]:pre,NOM:np:[NBRE(x)]:pre>  
[e]REGARDE(e, x, y)
- regardées: PSPA  
<\$:np:[NBRE(x)]:\$,OBJ:np:[pro,fem,pl(y)]:pre,NOM:np:[NBRE(x)]:pre>  
[e]REGARDE(e, x, y)

regardé: PAS  
 <\$:np:[sg(y)]:\$,PAR:np:[lex(x)]:post,NOM:np:[masc,sg(y)]:pre>  
 [e]REGARDE(e, x, y)

regardées: PAS  
 <\$:np:[pl(y)]:\$,PAR:np:[lex(x)]:post,NOM:np:[fem,pl(y)]:pre>  
 [e]REGARDE(e, x, y)

donnés: PAS  
 <\$:np:[pl(y)]:\$,A:np:[lex(z)]:post,PAR:np:[lex(x)]:post,  
 NOM:np:[masc,pl(y)]:pre>  
 [e]DONNE(e, x, y, z)

(4) est: FIN  
 <PSPE: <\$:np:[sg(x)]:\$|L>:[a]S:pre|L>  
 [s]l[s]a, AT(s, NOW), [a]S|  
*Jacques est sorti. Marie est sortie*

sont: FIN  
 <PSPE: <\$:np:[pl(x)]:\$|L>:[a]S:pre|L>  
 [s]l[s]a, AT(s, NOW), [a]S|  
*Les garçons sont sortis. Les filles sont sorties*

est: FIN  
 <PAS or PNOM: <\$:np:[sg(y)]:\$|L>:[a]S:pre|L>  
 [a]l[AT(a, NOW), S]|  
*Jacques est { regardé. } Marie est { regardée }*  
*{ européen } { européenne }*

sont: FIN  
 <PAS or PNOM: <\$:np:[pl(y)]:\$|L>:[a]S:pre|L>  
 [a]l[AT(a, NOW), S]|  
*Les garçons sont { regardés } . Les filles sont { regardées }*  
*{ européens } { européennes }*

a: FIN  
 <PSPA: <\$:np:[sg(x)]:\$|L>:[a]S:pre|L>  
 [s]l[s]a, AT(s, NOW), [a]S|  
*{ Le garçon } a donné { un livre }*  
*{ Jacques } { des livres }*  
*{ Marie }*

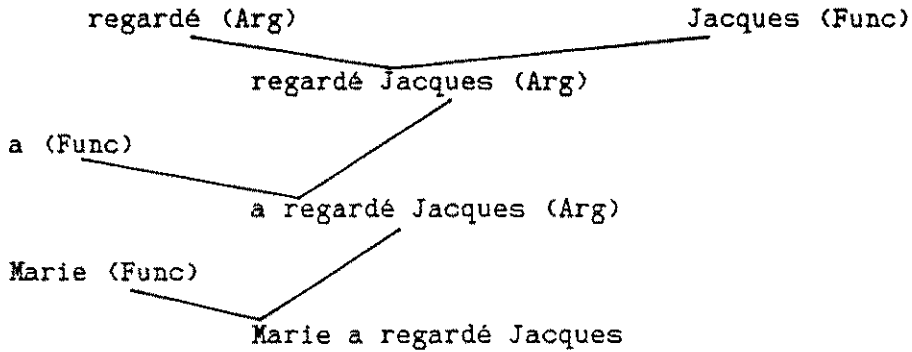
ont: FIN  
 <PSPA: <\$:np:[pl(x)]:\$|L>:[a]S:pre|L>  
 [s]l[s]a, AT(s, NOW), [a]S|  
*Les garçons ont donné { un livre }*  
*{ des livres }*

été: PSPA  
 <\$:np:[NBRE(x)]:\$, PAS or PNOM: <\$:np:[NBRE(x)]:\$|L>:[a]S:pre|L>  
 [a]S|  
*Le garçon a été { regardé }*  
*{ européen }*  
*La fille a été { regardée }*  
*{ européenne }*

Les garçons ont été { regardés }  
 { européens }  
 La filles ont été { regardées }  
 { européennes }

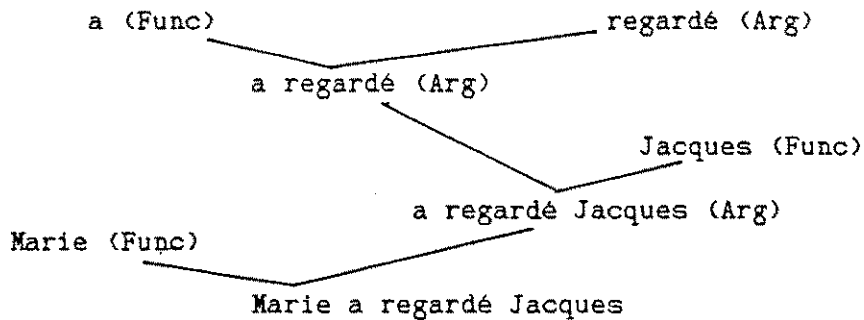
Instead of a "downward auxiliary phrase", as the following one (or as in the specification of UCG in T2.1) :

(5)



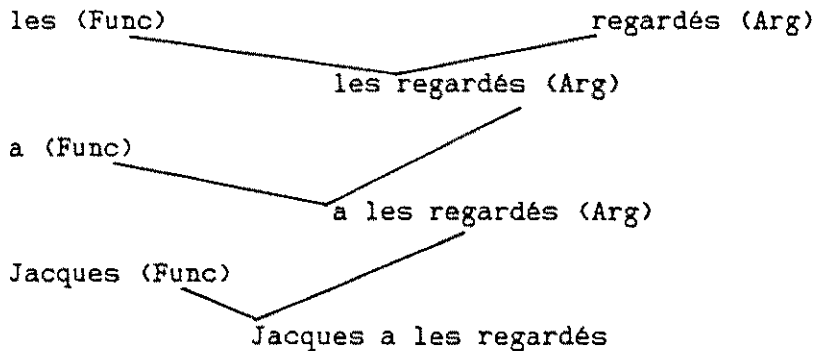
the grammar is intended to specify an "upward" one :

(6)



The main justification for the proposed analysis is clitic placement. With an analysis along the lines of (5) we should obtain incorrectly :

(7)





can be handled modulo the relevant order specification for the qu-item (in a way which is analogous to the treatment of relative sentences) :

- (2) dit :FIN  
 <QU-OBJ: np:[pro(y)]:pre, NOM: np:x:post>  
 [e] DIT (e,x,y)
- pense :FIN  
 <QU-A: np:[pro(y)]:pre, NOM: np:x:post>  
 [s] PENSE (s,x,y)

The matter is slightly complicated by the possibility of having an anaphoric pronoun to the right of the verb in INV-1 as in :

- (3) A qui Marie pense-t-elle?

A lexical entry for this pronoun could look like :

- (4) (-t)-elle : F  
 <F: <PRO: np:[pro, fem, sg(x)]:O|L>:[a]S:O|L>  
 [a]S

(4) is intended to make no semantic contribution. The corresponding entry for *penser* could be something like :

- (5) pense :FIN  
 <NOM: np:[flex(x)]:pre, QU-A: np:[pro(y)]:pre, PRO: np:[pro(x)]:post>  
 [s] PENSE (s,x,y)

After unification of the first element of the cat-list with *Marie*, the pronoun PRO will obligatorily get the specifications [fem, sg(x)]. Thus agreement in such constructions is correctly accounted for.

#### 4 Topicalisation

French topicalisation obeys few syntactical constraints. There can be more than one topicalised element at once :

- (1) Jean, Marie, il l'aime

Thus we get ambiguities as in :

- (2) La fille, elle la déteste  
 Jeanne, Marie, elle la déteste

where both the following readings are allowed, respectively :

- (3) La fille<sub>i</sub> , elle<sub>i</sub> la<sub>j</sub> déteste  
 La fille<sub>i</sub> , elle<sub>j</sub> la<sub>i</sub> déteste

Jeanne<sub>i</sub> , Marie<sub>j</sub> , elle<sub>i</sub> la<sub>j</sub> déteste  
 Jeanne<sub>i</sub> , Marie<sub>j</sub> , elle<sub>j</sub> la<sub>i</sub> déteste

In the main sentence, there must always be a pronoun recovering the topicalised item, which can be placed everywhere in a constituent, e.g. :

(4) Jean<sub>i</sub> , Marie<sub>j</sub> aime sa<sub>i</sub> maison

The principal constraint is of course gender and number agreement between the topicalised NP and the following pronoun(s), an agreement which may react on the verb itself :

(5) Jean et Marie, Pierre les a vus hier  
 vs. Jeanne et Marie, Pierre les a vues hier

One way to handle sentences like (1) - (2) in UCG could be to derive - via a lexical rule - for each verbal entry containing a pro-NP, a lexical entry where a lex-NP is added at the end of the cat-list. For example from *elle la déteste*, e.g.

(6) déteste : FIN  
 <OBJ: np:[pro(y)]:pre, NOM: np:[pro(x)]:pre>  
 [e] DETESTE (e, x, y)

we thus obtain *la fille<sub>i</sub>, elle<sub>i</sub> la<sub>j</sub> déteste* :

(7) déteste : FIN  
 <OBJ: np:[pro(y)]:pre, NOM: np:[pro(x)]:pre, NOM: np:[lex(x)]:pre>  
 [e] DETESTE (e, x, y)

and *la fille<sub>i</sub>, -elle<sub>j</sub> la<sub>i</sub> déteste*

(8) déteste : FIN  
 <OBJ: np:[pro(y)]:pre, NOM: np:[pro(x)]:pre, NOM: np:[lex(y)]:pre>  
 [e] DETESTE (e, x, y)

By unification, the two occurrences of the variables x in (7) and y in (8) get the same agreement specifications. By adding one more lex-NP, we also obtain *Marie<sub>i</sub>, Jeanne<sub>j</sub>, elle<sub>i</sub> la<sub>j</sub> déteste* and *Marie<sub>i</sub>, Jeanne<sub>j</sub>, elle<sub>j</sub> la<sub>i</sub> déteste*, respectively :

(9) déteste : FIN  
 <OBJ: np:[pro(y)]:pre, NOM: np:[pro(x)]:pre, OBJ: np:[lex(y)]:pre,  
 NOM: np:[lex(x)]:pre>  
 [e] DETESTE (e, x, y)

(10) déteste : FIN  
 <OBJ: np:[pro(y)]:pre, NOM: np:[pro(x)]:pre, NOM: np:[lex(x)]:pre,  
 OBJ: np:[lex(y)]:pre>  
 [e] DETESTE (e, x, y)

However, a treatment along these lines encounters two main difficulties. First, since it works only on the arguments of one verb, it cannot be extended to "long-distance" topicalisation, as in

(11) Marie, Jeanne dit à Pierre qu'elle la déteste

Second, it greatly increases the number of the required lexical entries corresponding to a particular verb. A preferable solution seems to be to consider topicalisation as the combination of an NP with a whole FIN sentence. Working somewhat outside of the UCG framework, a possible entry for the topicalised NP could thus look like :

(11) (la fille : F  
[-:FIN:[ ]:S: -]  
[a, 'FILLE' (Y), set V (pro`fem`sg in S,Y), S]  
- )

The predicate set *V* defines the set of pro-NPs which satisfy the condition of bearing the specifications fem and sg, and allows 'FILLE' to "bind" only one of these pro at once (therefore the disjunction). Thus we have only one lexical entry for all possible readings. On the other hand, we can express the relation holding between the topicalised lex-NP and a following pro-NP over one or several sentences.

### 5 Indirect interrogation in infinitival complements

Indirect interrogation is a very limited phenomenon concerning only a few verbs subcategorized for infinitival complements. To account for them in UCG, we can thus list directly the corresponding lexical entries. Moreover, if there seems to be actually an interrogation in sentences like

(1) Je demande à Jean que faire  
Je cherche quel chemin prendre  
Je veux savoir où aller

it is not true that all the so-called 'indirect interrogatives' have the semantic value of a question :

(2) J'indique à Jean que faire  
J'explique à Jean quelle réponse donner à Marie  
Je montre à Jean comment faire

Although the topic is not ever clear, let us admit that the qu-item is rather an interrogative one in example (1) and something like a relative one in example (2). We leave open the question as how these differences are expressed within semantics.





- (12) donner : INF  
 <OBJ:np:[lex(y)]:post, A:np:[lex(z)]:post, NOM:np:x:pre>  
 [elDONNE (e,,x,y,z)]

we assume the following ones (which could be obtained via a lexical rule operating on the standard entry, but we don't focus on this matter here).

For sentence (4) :

- (13) donner : INFINT  
 <A:np:[lex(z)]:post, QU-OBJ or QU-OBJ':np:[pro(y)]:pre, NOM:np:x:pre>  
 [elDONNE (e,x,y,z)]

where QU-OBJ is the feature marking the interrogative item *quoi* and QU-OBJ' the item *que*. We need to make this distinction because *quoi* and *que* do behave differently, in particular with respect to prepositions e.g. *à quoi* vs. \* *à que*, *sur quoi* vs. \* *sur que*, etc. (see below).

For sentence (9) we have :

- (14) donner : INFINT  
 <OBJ:np:[lex(y)]:post, QU-A:np:[pro(z)]:pre, NOM:np:x:pre>  
 [elDONNE (e,x,y,z)]

And for the sentences in (10), respectively :

- (15) donner : INFINT  
 <QU-OBJ:np:[pro(y)]:pre, QU-A:np:[pro(z)]:post, NOM:np:x:pre>  
 [elDONNE (e,x,y,z)]
- (16) donner : INFINT  
 <QU-OBJ:np:[pro(y)]:post, QU-A:np:[pro(z)]:pre, NOM:np:x:pre>  
 [elDONNE (e,x,y,z)]

The lexical entries for the qu-items are as follows. For sentences like

- (17) Jean se demande qui rencontrer

we have :

- (18) qui : F  
 <F<QU-OBJ:np:[hum, pro(x)]:OIL>:[b]S:OIL>  
 [b][hum,pro (?x), S]

And for example (4), respectively :

- (19) quoi : F  
 <F<QU-OBJ:np:[-hum, pro(x)]:OIL>:[b]S:OIL>  
 [b][[-hum,pro (?x), S]

- (20) que : F  
 <F:QU-OBJ':np:[-hum, pro(x)]:O|L>:[b]S:O|L>  
 [b] [-hum, pro (?x), S]

We want to allow them to combine with a preposition as in (9) or as in (21) :

- (21) Jean cherche à quoi penser

Therefore we have prepositions like

- (22) à : F  
 < QU-OBJ:np:[pro(x)]:post, <F:QU-A:np:[pro(x)]:O|L>:S:O|L>>  
 [b]S

The combination of (22) with (18) for instance gives :

- (23) a qui : F  
 <F:QU-A:np:[hum, pro(x)]:O|L>:[b]S:O|L>  
 [b] [hum, pro (?x), S]

which can recombine with (14) or (15).

Because the relevant feature here is QU-OBJ and not QU-OBJ' we don't generate :

- (24) \* Jean cherche à que penser

Pending a general treatment of interrogatives and relatives, we have expressed the semantic contribution made by qu-items simply by means of the notation (?x), which corresponds to the intuitive reading 'for which x' (i.e. x is indeterminate).

Another type of indirect interrogation involves determiners as in :

- (25) Jean cherche quel livre donner à Marie

- (26) Jean cherche à quelle fille donner le livre

The lexical entries for these qu-determiners are :

- (27) quel  
 <n:[masc, sg(x)]R: pre, <F:QU-OBJ:np:[masc, sg(x)]:O|L>:[b]S:O|L>>  
 [b] [a]R (?x), S]
- (28) quelle  
 <n:[fem, sg(x)]R: pre, <F:QU-OBJ:np:[fem, sg(x)]:O|L>:[b]S:O|L>>  
 [b] [a]R (?x), S]

By combination of (27) with

- (29) livre  
 n  
 [lex, mas, sg(x)]LIVRE(x)

we obtain :

- (30) quel livre : F  
 <F:QU-OBJ:[masc,sg(x)]:O|L>:[b]S:O|L>  
 [b]l[[masc,sg(x)]LIVRE (?x), S]

which can combine with (13) or (15).

(30) can also combine with (22) to account for (26).

So far the interrogated element was an argument of the infinitival verb. In some cases we find indirect interrogation on modifiers :

- (31) Jean cherche où aller  
 Jean ignore comment faire
- (32) Jean cherche avec qui partir  
 Jean ne sait pas sur quoi s'asseoir

For the first sentence of (31) we assume the following entries:

- (33) où : INFINT  
 <F:X(<np>):[a]S:pre|X>  
 [a]l[LIEU(a,?z),S]

- (34) aller : INF  
 <x>  
 [e]VA (e,x)

(34) is of course the standard entry for the infinitival form. The combination of (33) with (34) gives :

- (35) où aller : INFINT  
 <x>  
 [e]l[LIEU(e,?z),VA (e,x)]

And (35) can combine normally with (8). (The same holds for *comment* modulo the relevant modification of phonology and semantics in (33)).

For the first sentence of (32) we have :

- (36) avec : F  
 <QU-OBJ:np:[pro(x)]:post,<F:X(<np>):[a]S:pre|X>>  
 [a]l[AVEC (a,x), S]

Like (22) before, (36) can combine with (18) or (19). With (18) we obtain for example :

- (37) avec qui : F  
 <F:X(<np>):[a]S:pre|X>  
 [a]l[hum,pro (?x), AVEC (a,x), S]

(37) can combine with an infinitival form and the result with (8). (The same for *sur*, etc. modulo changes in phonology and semantics).

As it stands, the UCG formalism accounts without difficulties for indirect interrogation in infinitival complements, for displacements here are only local, not involving UDCs.

An infinitival verb being a category still waiting for a subject, direct interrogation with infinitives (not infinitival complement) as in :

(38) Que faire?  
Où aller?  
Sur qui compter?

cannot however be handled according to the present framework.

### 3.6 Unbounded dependencies

Extraction and UDCs cannot be handled according to UCG February 86. This is of course a crucial topic to work on if one wants to treat interrogation.

## 1. THE SEMANTICS OF QUESTIONS AND ANSWERS IN DRT.

### 1.1. SEMANTICAL THEORIES OF QUESTIONS.

There are two points of departure for those who want to investigate the logic of questions. Whereas for assertions it is possible to have direct intuitions about the entailment relations they enter into, for questions this is not the case: the function of a question is not to give information that may entail other formula's but rather to signal the absence of information. So it is necessary to restrict oneself to more indirect evidence: intuitions about the relation between questions and correct answers to those questions, and the entailment relations of sentences containing indirect questions.

It is probably useful to review some of the more common observations in this area. Let (1) be the question:

(1) Who won ?

Then this can be answered by (2)

(2) John.  
Maybe Harry.  
A boy.  
John won.

Two distinctions may be drawn: constituent versus senten-

tial answers, and complete versus partial answers. (2d) is an example of a sentential answer, (2a), (2b) and (2c) are examples of constituent answers. Moreover, (2a) and (2d) are examples of complete answers, whereas (2b) and (2c) indicate that the speaker in answering has less than the information required to give a complete answer.

A complete answer is characterised by the properties:

truth  
 exhaustivity  
 rigidity

That an answer must be true to be the answer to a question goes without saying. Exhaustivity stands for the property that the speaker has exhausted all the people who won in his answer "John": it is implied that the speaker in answering by (2a) assumes that nobody else won but John. Rigidity finally stands for the choice of the expressions in the answer: one wants the identification of the answer to be independent of contingent facts that the questioner may not know or be mistaken about.

A theory about the status of complete answers is the basis for most theories of questions. The second group of observations stems from indirect questions such as

- (3)      a. John knows whether Mary walks.  
           b. John says who won the game.  
           c. John wonders which boy gave the party.

(3a) together with the premiss

(4) Mary walks .

entails (5)

(5) John knows that Mary walks.

Similarly from (3b) and

(6) Mary won the game

it follows that

(7) John says that Mary won the game.

Similar inferences can not be made from verbs like "wonder" as in (3c).

## 1.2. GROENENDIJK & STOKHOF

The most extensive attempt to deal with questions in Montague Grammar is (Groenendijk 1984).

The main features of their theory are the following:

1. A question denotes the concept of its true exhaustive answer.

Note that under the assumptions of truth and exhaustivity, in every situation, a question has an unique answer, modulo logical equivalence. The propositional concepts are derived in two steps:

An abstract is formed which translates a phrases like (8a-b):

- (8) a. who walks  
b. which girl dances

into (9a-b).

- (9) a.  $\lambda x \text{ walk}(a)(x)$   
b.  $\lambda x \{ \text{girl}(a)(x) \ \& \ \text{dance}(a)(x) \}$

Here the variables  $a$  and  $i$  stand for indices (world time pairs). To these abstracts an operation is applied to yield:

- (10) a.  $\lambda i \{ \lambda x \text{ walk}(a)(x) = \lambda x \text{ walk}(i)(x) \}$   
b.  $\lambda i \{ \lambda x \{ \text{girl}(a)(x) \ \& \ \text{dance}(a)(x) \} = \lambda x \{ \text{girl}(i)(x) \ \& \ \text{dance}(i)(x) \} \}$

((10a) denotes the set of possible worlds in which the set of walkers is the same as in the actual world. So this expresses the proposition that the set of walkers is precisely the set of the actual walkers.

One can deal with indirect questions in a very straightforward way. A verb like "wonder" expresses a relation with the meaning (the intension) of the question, while verbs like "know" or "say" express a relation with the denotation of the question. This deals with the entailments in a simple and economical way.

The notion of a canonical answer that the theory gives is



sound but perhaps not very natural: a question is canonically answered by an assertion whose meaning is the denotation of the question: i.e. by an assertion that states explicitly and rigidly the answer in the form of a sentence. E.g. an answer to (I) would be

(II) Only Bill and Tom walk.

This is not in line with what occurs in normal discourse. The exhaustivity of an answer is generally implicit: it seems that one has to indicate that the answer is not exhaustive if that is the case, and not the other way round. Moreover there seems to be a preference for the shorter constituent answers (No!, John and Bill, etc.). These problems can however be dealt with after a fashion: it is possible to treat the syntax and semantics of linguistic answers by giving rules that generate question-answer pairs together in a single generation.

Where the theory is perhaps most impressive<sup>1</sup> is in its smooth integration with the definitions of pragmatic answerhood given an information set. As such it can be shown to be quite adequate within the limits set by the framework.

-----  
<sup>1</sup> It shows quite unexpected possibilities for the possible worlds framework in which it is formulated.

### 1.3. SCHA

A very similar theory, that is rather close to my own proposals, is Remko Scha's, developed in the course of the PHLIOA project, and explained in his [Scha 1983]. Like Groenendijk & Stokhof, Scha starts from the notion of a complete (exhaustive) answer, but, unlike them, takes the notion of a constituent answer to be more basic. (It should be admitted that Scha makes no attempt to deal with indirect questions.) A question in his theory denotes the abstract from which Groenendijk & Stokhof derive their propositional answer, and the meaning of a question is the function that assigns to a possible world the denotation of the abstract. This allows for representations of questions like:

Which boys run ?                    {x ∈ BOYS | run(x) }

An advantage of Scha's theory is that the relation between questions and constituent answers becomes very simple, which is of course as it should be in a question-answering system that answers all questions by constituent answers. Within the MG framework this approach is less adequate for dealing with sentential answers and indirect questions without further ado.

#### 0.4. A QUESTION THEORY IN DRT

It would be easy and pleasant to be able to transport the above theory to DRT without further ado. In this way we might conceivably be able to do away with some of the limitations of the MG framework as well. Unfortunately, there are at least three reasons why this is not possible. The first two of these are less essential: DRT does not have intensional types and lambda abstraction. It might, as some authors have suggested, be supplied with such devices. Others on the contrary would maintain that the absence of such devices is a merit rather than a shortcoming of DRT: the simplicity that seems to be a characteristic of DR-approaches, and the ontological parsimony in its interpretation, has a lot to do with the absence of a full blown type theory that would accompany the introduction of lambdas and intensions. Intensionality can (though there is no agreement as to how this should be done) be dealt with in a more syntactical way, and the usefulness of lambda abstraction is reduced as soon as one gives up the idea that all combination of meanings should be done by function application.

The other problem is more serious, and seems to be something that DRT has in common with so called Russellian semantics, namely the impossibility of iterating intensionality that is typical of the Frege-Church-Montague

school in semantics. Actually, the G&S treatment of questions is the only case I know of where this feature is really exploited. In a Russellian approach an expression has only a sense and a denotation, and there are no such things as intensions of intensions. Typically also, syntactical theories of intensionality can mimick one level of intension but not two. Once one has gone over from the denotation to the syntax the possibilities are exhausted: going to the syntax again does not bring one anywhere else. So the whole notion of a concept of propositions cannot be handled in a Russellian or DR theory.

### 1.5. WHAT IS A DRS

A formula in DR theory can be interpreted in general (not all interpretations are always available) as standing for the following kind of entities.

#### 1. A proposition.

The formula, both taken in isolation and in a context, makes a truth claim. Since this truth claim may be successful in some models, (or parts of models) and fail in others, the formula can be regarded as a criterion from possible worlds to truth values. This corresponds to an expression of type  $\langle s,t \rangle$  in Montague grammar.

#### 2. A "definition".

The formula, if it has discourse referents of its own, contributes values for the discourse referents of the succeeding text. This process is mostly referred to as "setting up a value" or "making an indefinite reference". In certain cases, if the formula allows only one object to be the value of a discourse referent, the formula defines a value for the discourse referent in a strict sense.

As a definition the formula may correspond to objects of type  $\langle s, e^n \rangle$  that is as a concept of (a sequence of) objects.

### 3. A property.

Since a formula may have both free variables and discourse referents it can be exploited in various ways as a predicate, and can therefore also be seen as a property. The context may provide values for certain variables and discourse referents and the formula may serve as a predication on them. In this way it corresponds also to Montague's  $\langle s, \langle e^n, t \rangle \rangle$ .

## 0.6. AN ABSTRACTION FORMULA

Most people that have tried to deal with plural in LOGO have found it useful to introduce special variables for plural entities. There may be a certain latitude when one

is dealing with plural as to whether one prefers a set theoretical interpretation for plural variables or -along the lines of Link- a mereological or lattice theoretical approach to the interpretation of these variables. From the point of view of questions there is a lot to say for a Link type approach since we are dealing with a number of different categories of entities that questions can ask for: plural objects (whn. which), quantities of mass (which water is contaminated), temporal entities (when), measure entities (how much), booleans (yes/no questions), reasons (why). For all these we would like to have notions of "membership" and "part" without necessarily wanting to reduce all of them to sets.

Essentially what one wants is a notion of the maximum of a group of elements of a particular sort. This can be a DFT-formula like

$$\max(a, b, A)$$

which is defined for all the sorts, in terms of the lattice structure of the extension of the sorts, as true whenever  $\underline{a}$  is the supremum of  $\{b|A\}$ , the set of those  $b$ 's for which  $\underline{a}$  holds.

Over objects,  $\underline{a}$  generally is a plurality of objects, and similarly over events, one would expect the suprema to be pluralities of events rather than events again. But for

masses, states, quantities and pluralities the supremum has the same sort as the elements it maximises.

[Link 1983] has advocated, in treating the plural, the use of so-called complete join semilattices. For plural it becomes possible to capture certain generalisations over plural objects and mass terms in this way. For questions such generalisations are even more important since one can ask for specification of objects of all types. Join semilattices are structures  $(A, \cup)$  where  $A$  is a nonempty set, closed under the binary operation  $\cup$ , that satisfies the following axioms:

1.  $a \cup a = a$
2.  $a \cup b = b \cup a$
3.  $a \cup (b \cup c) = (a \cup b) \cup c$

On a structure of this kind it is easy to define a partial order:

$$1. \quad x \leq y \text{ iff } x \cup y = y$$

$\leq$  is a semi-connected partial ordering, i.e., it is reflexive, transitive, antisymmetric, and for any two elements there is an element that is  $\leq$  both of them. (1) ensures reflexivity, transitivity follows 1; noting that if  $a \leq b$  and  $b \leq c$  then

$$a \cup c = a \cup (b \cup c) = (a \cup b) \cup c = b \cup c = c$$

For asymmetry note that if  $a \leq b$  and  $b < a$

$$a \cup b = b \cup a = a \cup b$$

For semiconnectedness: just take  $a \cup b$ .  $a \subseteq a \cup b$  because

$$a \cup (a \cup b) = (a \cup a) \cup b = a \cup b,$$

and similarly  $b \subseteq a \cup b$ .

It is this relation with regard to which we take the maximum of certain elements. For an infinite set of such elements, it is not guaranteed in general that such maxima exist in the structure. For this reason we demand that the semilattices are complete, i.e.:

$$5. X \subseteq A$$

$$\exists y \forall x (x \in A \rightarrow y \leq x)$$

This ensures that any collection of objects has a maximum. Some elements may not have proper parts. In this case they are called atomic.

$$6. \text{atomic}(x) \text{ iff } y \leq x \Rightarrow y = x$$

Atomic elements are elements without real parts. We can use them to understand the difference between a singular and plural object, or a single event and a collection of them. Not all the sorts we are considering need to have an atomic elements: there need not be atomic states, propositions, bits of cheese, quantities and the like. But spatiotemporal continuants, events and perhaps other kinds exhibit a clear distinction between plural and singular.



which can be expressed by saying that that plural objects are different from singular ones, but are constituted by singular ones. Structures where every element is constituted by atomic elements are called atomic:

$\gamma, A$  is atomic iff  $\forall x \in A \Rightarrow \exists B \subset A \ x = \cup B$  and  $\forall y \in B \Rightarrow$   
atomic( $y$ )

Let us first consider how a formula like maximum can be built up. The semilattices provide us with a notion of part of, for which we can introduce a constant:

part( $y, X$ ), is true iff  $y$  is a part of  $X$  ( $y$  is  $\cap$  of  
part( $y, X$ ),  $X$  is not) in terms of the join  
semilattice that interprets the sort of  $X$   
and  $y$ .

A notion of distributivity can be defined that says, that all the parts of a given entity have a certain property:

dist( $X, y, A$ ):= [part( $y, X$ )  $\Rightarrow$  A]

Similarly, we can invert distributivity to get a notion of subsumption.  $X$  subsumes all the things that have the property  $A$ .

sub( $X, y, A$ ):= [A  $\Rightarrow$  part( $y, X$ )]

These two notions suffice for the definition of maximum:

$$\max(X,y,A) := [\text{true}(X), \text{dist}(X,y,A), \text{sub}(X,y,A)]$$

The formula  $\max(X,y,A)$  says: there is an  $X$  such that  $X$  collects the  $y$ 's such that  $A$ . If we assume some kind of collection axiom to hold it is a trivial proposition.

As a definition it is a definition of  $X$ , the collection of the things such that  $A$ . Notice that it has uniqueness built in by the algebraic construction of the objects.

As a property (necessarily of some value for  $X$ ) it says that this object collects all the objects  $y$  such that  $A$ .

I will reserve the term abstraction for formula's of the form  $\max(x,y,A)$  in the sequel.

A special case are the *whether* questions. It is possible, if we want to, to reduce them to the standard case by adding variables for booleans ( $B,C,\dots$ ) (the booleans form a complete join semilattice, taking disjunction as join), and giving the translation:

$$\max(B,C,[C \Rightarrow A])$$

E.g.

(12) Does John walk ?

would correspond to

(13)  $\max(B, C, [C \Rightarrow \text{walk}(\text{john})])$

which would denote "true" if john walks and "false" otherwise. Suppose John walks. Then both formula's:

```
{14}  [1 => walk(john)]
      [ T => walk(john)]
```

are true. So the maximum is T, and the answer is "yes". If on the other hand, John does not walk, only

```
{15}  [1 => walk(john)]
```

is true, so that the maximum is 1 and the answer is "no". If this reduction is desirable, is a question that we will not try to answer here.

Another special case is formed by the alternative questions like:

```
{16}  Do you want tea or coffee?
      Is Harry in Paris or has the firm gone bankrupt?
```

These questions prod one for a choice from the alternatives offered; a simple yes or no is not sufficient. It is tempting to attempt a reduction to the standard case, but one might wonder, as with *whether* if this will really lead to the simplest possible solution. Simpler treatments are possible if we introduce new primitives. E.g. a formula

```
whether(X,A)
```

that is true whenever X denotes yes and A is true, or X denotes no and A is not true.

Similarly  $n+1$ -place predicates

$or_n$

such that e.g.

$or_3(X, A_1, A_2, A_3)$

is true whenever  $X$  denotes  $A_i$  and  $A_i$  is true, might be considered.

### 1.7. WHY ARE ABSTRACTIONS MEANINGS OF QUESTIONS.

The abstracts that were introduced in the brief sketch of the Stokhof & Groenendijk theory of questions correspond to sets or extensional relations. The corresponding abstractions are clearly definitions of the same set. At the same time however, they can be read as the proposition that says of the set that it is the set with this definition, or as the property expressed by the definition.

Let's see if we can define the relation between questions and answers on the basis of what we have until now. The following can be a first attempt.

1.  $A$  is the answer to  $B$  iff  $A$  is an assertion or isolated constituent that defines the same entity as the question  $B$ .

2.  $A$  defines  $e$  on  $M$  iff  $A$  has a discourse referent  $D$  and

for all  $f$  such that  $H, f \models A \ f(D) = e$ .

Let us now consider the question:

(17) Who walks?

with translation:

(18)  $\max(X, x, \text{walk}(x))$

and some possible answers to this question:

(19)	John.	$A(j)$
	John walks.	$\text{walk}(j)$
	John and Harry.	$A(\{j, h\})$

The first three answers, give reason to assume that we must extend our definition of "defines" to include being a referent of a proper name, or the referent of a complex proper name mentioned in the sentence. It is odd perhaps that we allow any syntactical environment of such a proper name to be an answer.

(20)	The schoolmaster	$\{x=y, \text{schoolmaster}(x), A\}$
	The boys.	$\{\max(\gamma, x, \text{boy}(x)), A\}$

Since both answers are definite they define an object, and this can be identical with the object defined by the question. So for definite constituent answers the relation is well defined.

It is however more elegant to treat definite answers in a more syntactical way. The appearance of an unbound vari-

able for a formula in the representation of the answer is not desirable, since it does not stand for anything. But there is something more serious: though the relation of answerhood can be defined, the representation of the answer does not contribute anything to the meaning of the preceding representations. That is, the new information provided by the answer is not represented. Both problems can be accommodated by unifying the sentence variable A in the definite answer, with the formula  $Y=X$  where Y is the object defined by the answer, and X the object defined by the question. Compare [21].

[21] Who walks?  
 $\max(X,y,\text{walk}(y))$

The boys.  
 $\{\max(Y,x,\text{boy}(x)),A\}$  (as a term translation)  
 $\{\max(Y,x,\text{boy}(x)),X=Y\}$  (as a answer translation)

Together with the question, [21] yields [22].

[22]  $\{\max(X,y,\text{walk}(y)),\{\max(Y,x,\text{boy}(x)),X=Y\}\}$

This still does not enable us to deal with partial answers, and with indefinites. Stokhof & Groenendijk deal with those as essentially involving pragmatic relations. Even if this is so, it is necessary to be able to deal with syntactically incomplete indefinite and partial answers. Notice, that the solution for definites does not work:

(23)    some boys  
           $\max\{Y, x, [\text{boy}(x), A], \text{some}\{Y\}\}$

If we unify A with  $X=Y$ , Y will denote the set of all boys. What we seem to want is to unify A with the third argument of the question, and identify both second arguments. What results for the example is (24),

(24)    some boys  
           $\{\max\{Y, y, [\text{boy}(y), \text{walk}(y)]\}, \text{some}\{Y\}\}$

which, as it should be, is the same as the translation of (25).

(25)    Some boys walk.

Dependent and Multiple WH.

Under the assumption that the WH introduces the abstraction formula, multiple WH questions assume the aspect of a question with dependent questions. E.g.

(26)    Which boy likes which girl.

would translate as

(26b)    $\max\{X, y, [\text{boy}(y), \max\{Y, x, [\text{girl}(x), \text{like}(y, x)]\}]\}$

Under the assumption that max behaves as a abstraction operator, X denotes the set of all boys, while Y has a denotation that depends on each of the members of X. An answer to the question correspondingly should list all of

the boys and specify the denotation of Y for that choice of boy.

(26c) John Suzy and Alexandra, Bill Harriet, and Tom none at all.

These are similar to so-called dependent questions like (27):

(27) Which student does every professor like.  
 $[\text{professor}(x) \Rightarrow \max(X,y, [\text{student}(y), \text{like}(x,y)])]$

Most of the things said in this paper about questions can only be transported with great difficulty to cover these cases. To deal with these cases it is probably best to look for a different solution than we would get by simply iterating and quantifying in.

Something which would conform better with the rest of our treatment is to treat the formula as somehow setting up a skolem function type, as in

(28)  $\max\{f(X, f), [y, [y, x]], [\text{boy}(y), \text{girl}(x), \text{like}(y, x)]\}$   
 $\max\{g, f, [\text{professor}(x) \Rightarrow \{\text{student}(f(x)), \text{like}(x, f(x))\}]\}$

In this way both the treatment of indirect questions, and of the question answering relation would go on. These questions are related to questions with plural's such as coreference of plural pronouns that seems to maintain a functional link.



## 1.8. INDIRECT QUESTIONS

In contrast to what happens in Montague Grammar, it seems possible to defend that the semantics given here for questions can be adapted directly to indirect questions.

The formula that corresponds to the abstraction is a tautology as a proposition, if taken in isolation.

$$\exists x(X, x, A)$$

does not say anything more than that there is an object that collects everything with the property A. This is, in normal cases, the content of set abstraction, and thereby always true. This is however no longer the case if the main variable is bound by the context. In this case it functions like an attribution of a property to the value of X in the context. So in such cases it is not a triviality that somebody knows that  $\exists(X)$ , but something that depends on his acquaintance with and knowledge about the value of X. Similarly, saying  $\exists$  is in this case not claiming that set abstraction holds once more but making a claim about a particular object: the object individuated by the question. We can make X stand for this object in a very simple way: namely by adding the  $\exists$  once more, outside the indirect context. (something which seems useful in the case of factive propositional attitudes, and vague modifiers as well)

In the other cases, it may seem strange to assume that a relation like "wonder" holds of tautological propositions and of nothing else. But in DRT we are not forced to regard all logically true propositions as the same: the syntactic approach to intensionality makes questions different, be it logically true, propositions. Thereby, the objection is not good in this context. Under the fairly plausible analysis that "wonder" means "want to know" we get:

$$\text{want}(x, [\text{Q}, \text{know}(x, \text{Q})])$$

which expresses a desire to a proposition that is no longer logically true.

In a case like "decide" perhaps we can say something like the following.

(28) John decides who is appointed at the department.

The variable set up in the representation of the indirect question perhaps just does not have a value. Our usage of words like decide suggests that there is not, independently of John's making up his mind, a set X that consists of the persons that will really be appointed. This suggests that we should apply the property interpretation. John here has a relation to the property, rather than to the set or the proposition.

To conclude: it seems possible to exploit the functional ambiguity of DRS's to maintain that there are representations that are simultaneously concepts of answers, tautological propositions and contingent propositions, so that it is possible to give an integrated theory by means of them.

## APPENDIX.

Some UCG.

```
w knows w1
sent[fin]/w:np[nom]:x/w1:sent:A
[A, know(x,A)]
```

```
w know w1
sent[fin]/w:np[nom]:x/w1:sent[query]:0
{0, know(x,0)}
```

```
w1 wonders w2
sent[fin]/w1:np[nom]:x/w2:sent[query]:0
want(x, {0, know(x,0)})
```

```
does w1 w2
sent[que]/w2:{sent[inf]/np:x}:A/w1:np[nom]:x
max(bool(a), bool(b), {b => A})
```

```
w1 does w2 w3
sent[que]/w1:np[wh]:y/w1:{sent[inf]/np:x/np:y}:A/w2:np[nom]:x
A
```

whom does jane like

```
w1 does jane w2
sent[que]/w1:np[wh]:y/w3:{sent[inf]/np:jane/np:y}:A
A
```

```
w1 does jane like
sent[que]/w1:np[wh]:y
like(jane,y)
```

whom  
 $X/(X/np[wh]:x):A$   
 $max(X,x,A)$

whom does jane like  
 $sent[que]$   
 $max(X,x,like(jane,x))$

w1 does w2  
 $sent[que]/w1:np[nom,wh]:x/w2:(sent[inf]/np:x):A$   
 $A$

who does walk

does walk  
 $sent[que]/w1:np[nom,wh]:x$   
 $walk(x)$

who does walk  
 $sent[que]$   
 $max(X,x,walk(x))$

Groenendijk (1984) Studies on the Semantics of Questions and the Pragmatics of Answers. PhD Thesis, Centrale Interfaculteit Amsterdam. private.

Link, G. (1983) The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretical Approach. In Bauerle, R., Schwarze, C. and Stechow, A. (eds.) Meaning, Use and Interpretation of Language, pp102-123. Berlin: de Gruyter.

Scha, R. J. H. (1983) Logical Foundations for Question Answering. Eindhoven, The Netherlands: Philips

## 5. Suggestions

The above reflects some work that has been done in Edinburgh, Clermont-Ferrand and Marcoussis on contextual phenomena in dialogue. Other deliverables are equally relevant :

Problems of Dialogue Parsing - T2.1 (EPI)  
 Investigation on Direct Manipulation - T4.3 (FHG)  
 Deductive Natural Language Processing for DRSs - T1.2 (BULL)

In this conclusion some issues will be taken up that are relevant for the work on the demonstrator, -and in some cases- for the prototype of the project. Some of these are oriented towards a specification of the architecture, others are oriented towards concrete implementational goals to be achieved in the coming years.

### 5.1. The semantics of questions

The DRT-version of the MG-style semantics of questions in chapter IV may not be immediately usable in a question-answering environment for two reasons. The first is the form of the semantics of a question like :

*Who walks* :

$$\max(X, x, \text{walk}(x))$$

Literally, it would ask for a deduction process that delivers a set A of people, proves of the members of A that they walk, and proves that nothing but the members of A walk. Such a process is different from the one usually assumed in PROLOG, where the form of a query would be

$$\text{walk}(x).$$

and the deduction process would, maybe as often as it can, instantiate the variable x to something that walks. It may have advantages to remain close to PROLOG.

The second problem is that of dealing with negative information. In the context of the ACORD project, we are dealing typically with incomplete information, so that negation by failure means, " I do not know ". It is thereby hard to assume that - in the short run - a solution will be found that can prove exhaustivity in a satisfactory way.

For both these reasons, and for dealing with multiple wh-questions after a fashion (we obtain the so-called cumulative readings) it is interesting to look at an alternative, that essentially extends DRT-syntax, but leads to the same results for single wh-questions.

The alternative takes up the idea of discourse markers. These are those variables of a representation, that are instantiated by any embedding that satisfies the representation. In a similar way, one can, in a question, distinguish wh-markers. These receive a maximal instantiation from those embeddings that satisfy the

question.

To make these ideas more precise, consider the following question :  
*Who gives whom a book ?*

wh <sub>1</sub> wh <sub>2</sub>
y
book(y) wh <sub>1</sub> gives wh <sub>2</sub> y

An embedding  $g$  satisfies such a marker iff

- (1)  $g(\text{wh}_1)$  and  $g(\text{wh}_2)$  are defined
- (2) for all  $h \in \text{h}^-_{\{\text{wh}_1, \text{wh}_2\}}$   $g$  such that  $h(\text{wh}_1) \in g(\text{wh}_1)$  and  $h(\text{wh}_2) \in g(\text{wh}_2)$  there is an  $\{y\}$ -extension  $f$  of  $h$  that satisfies each of the conditions
- (3) there is no  $\{\text{wh}_1, \text{wh}_2\}$ -variant  $j$  of  $g$  that also fulfills (2) such that  $g(\text{wh}_1) \subset j(\text{wh}_1)$  and  $g(\text{wh}_2) \subset j(\text{wh}_2)$

Like the treatment in chapter III, section 5 (p.7), this treatment guarantees a maximal interpretation for the wh-elements. Its behaviour is however more consistent with the exhaustive PROLOG search. The two wh-elements are instantiated by the list of the  $x_1, x_2$ , instances of

$\text{give}(x_1, x_2, y), \text{book}(y).$

In InL a question marker is rendered by two devices. The first would be wh-sorted variables for the relevant sorts. This could be achieved by adding a wh factor to the definition of sort. The second would be the addition of an operator *question* to the syntax. We would then translate questions as follows :

*Who walks ?* :  $\text{question}([\text{wh}(x), \text{walk}(x)])$

*Which boy does Mary like ?* :  $\text{question}([\text{wh}(x), \text{boy}(x), \text{like}(\text{Mary}, x)])$

*How many books are stolen ?* :  
 $\text{question}([\text{number}(n, X), \text{wh}(n), \text{max}(X, x, \text{book}(x), \text{stolen}(x))])$

*Who gives Peter what ?* : question([wh(x),wh(y), give(x, y, peter)])

*Does Peter walk ?* : question([walk(peter)])

Pending a better understanding of multiple wh-questions, dependent questions and of the proof theory of  $\max(\_, \_)$  it may be that the above treatment is to be preferred for implementation

## 5.2. Interface with the Dialogue Manager.

The latest version of the ACORD architecture shows the following relevant scheme for the interface between Dialogue Manager and Dialogue Parsers.



There are at least three reasons for thinking that the picture needs further clarification.

- a. The various parsers have, if they have a resolution capacity at all, no real access to the resources for dealing with full resolution. The DRS that may pass over the arrow must therefore contain a number of unresolved pronouns, and features on those. That can be settled only when the full resources are available in the Dialogue Manager. The DRS therefore - though it may be readable as a formula - will not be a finished representation.
- b. The question-answer relation  
Some of the material in chapter II sections 1.3, 2.1.7 and 2.3.1.7 gives reason to think that an answer generator needs to have access to the syntactic analysis of the question. It is not clear how such information will be passed over. It might be coded into the DRS or the syntactic analysis may be passed over entirely. In view of point (c.) it may be useful to set up a pragmatic analysis, which would deal with both the passing of the relevant information to the generator and part of the information for setting up the dialogue model and helping resolution.
- c. There are a number of aspects in NL-sentences that do not contribute to the truth conditions of a sentence. Frege used the terms "Farbe " (colour) or "Beleuchtung" (lighting) for these aspects. A sentence of the form

A but B

has the same reading as

## A and B

but a different "Beleuchtung". There are many such expressions, that seem to set up little or no extra truth condition.

*However, on the contrary, finitely, though, by the way, on the other hand, also, again, ..*

All these seem to be explainable using the rich notion of *discours structure* referred to in Section I. These markers tell one how to integrate the sentence in the context built up until then.

Other aspects of "Beleuchtung" may be taken as relating to speech act markation. Negative polarity items indicate questionhood, in a non-negative setting, inversion markers questionhood, deleted subject the imperative, though truthconditionally there is little that changes.

Again it seems wrong to code it all up in the semantic representation. Rather it seems that a pragmatic "representation", maybe just consisting in a feature list, should be part of what flows from the parser to the Dialogue Manager, or the resolution component.

### 5.3. Ellipsis, long distance dependencies and questions.

The three subjects mentioned in the title form a special problem for a parser based on categorial grammar. In all three cases one wants to have category changes of a single string : in ellipsis to make an incomplete string complete by adding a pronominal element in the semantics, in long distance dependencies to allow the constituent that normally combines with a neighbour to combine with a constituent over a distance, in questions to allow the semantics of a protoquestion to become the semantics of a question (in the sense of the new proposal).

In categorial grammar there seem to be two ways to deal with the problem. the first is the introduction of elements with zero phonological realisation of the appropriate types and with the appropriate semantics. The normal combination rules take care of the application. It is however not trivial to let the parser infer the existence of such zero elements when appropriate : it involves changing the input string. Perhaps one can think of this solution as a justification of the second solution : the introduction of special rules for cases of this kind.

The second solution - rules - shares a problem with the first : the special rules considerably increase the number of possibilities that have to be investigated in an exhaustive parse. Having such rules, or empty elements for that matter, obviates the need for a control regime that takes care they are not unnecessarily involved.

#### 5.3.1. Ellipsis

In order for the parsing to succeed, it is necessary to reconstruct the *missing constituents* in examples like the following :



- (a) John ate many.
- (b) John did too.
- (c) John Suzy.
- (d) John too.

For (a) it is necessary to be able to parse *many* not as a determiner but as a full np. This calls for a rule that takes *many* (and other determiners) and makes it into a homophonous np, changing the semantics to something that contains an element PROCN, where otherwise the noun would have appeared. Resolving the PROCN by making it identical with a contextually prominent 'kind'-expression then leads to its full representation.

For (b) similarly we want to generate from *did* (that is subcategorized for subject and infinitival clause) a full, phonologically identical VP that contains a PROVP element in its semantics.

Similarly (c) calls for a rule turning "Suzy" into a VP (or equally "John Suzy" into a sentence) introducing a PROVERB element in the semantics. (d) can be completed by adding PROVP. If resolution rules work on the preceding contexts (a'-d')

- (a') Mary ate some doughnuts.
- (b') Harry worked in the garden.
- (c') Bill likes Sue.
- (d') Bill likes Sue.

It must be reasonably clear what goes on.

An adequate treatment would be to have a new rule for each of these cases (with the relevant restrictions). Such rules have the problem that one does not want to involve them, unless they are really needed. If this problem could be dealt with, only resolution rules needed for a complete treatment.

### 5.3.2. Long distance dependencies

Questions, topicalisation and relativisation are the main points where these dependencies occur. There are many ways to deal with them. All solutions require a solution to the problem of not letting the search space grow too rapidly. It requires parsing rules that only involve the threading, or stacking or composition to work when one needs to, and not otherwise. Given such a control mechanism, it is fairly straight forward to use extraposition lists (cf. Pereira), threading (cf. Karttunen) or composition (cf. Steedman), or to use a different idea to deal with long distance dependencies.

### 5.3.3. Transformations

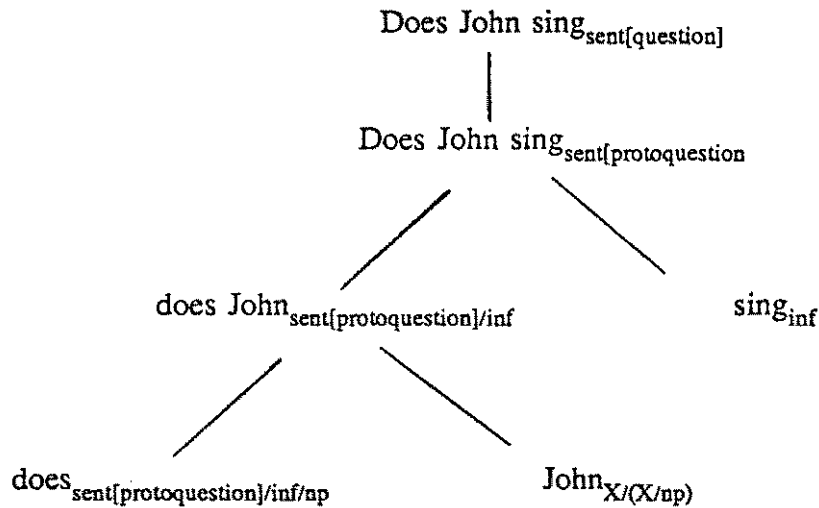
Also when one is not dealing with ellipsis, it is sometimes necessary to transform the categorisation of an expression and its semantics. This seems

needed in bare plurals, and also for questions. A question is semantically a whole, that introduces an operation on the semantics of the sentence that constitutes it. But this operation can not be localised, or attributed to any particular component. Therefore it seems necessary to have a rule (like in other treatments of questions) that takes a sentence that can be made into a question (this can be a feature) and makes it into a question. We will just give an example based on the treatment of questions. In this treatment a question is derived from a protoquestion. Both questions and protoquestions are best treated as features in the verb.

Ignoring time, as elsewhere, a question like :

Does John sing ?

is derived as follows :



where the last rule that turns a protoquestion into a question can be given as :

W		W
sent[protoquestion]	==>	sent[question]
A		question(A)