WP01: Consolidamento e Confronto delle Risorse Linguistiche (Resp Delmonte – Unità 4)

Detecting and Comparing Peculiar Features of Different Dependency Treebanks

The report is organized as follows:

1. Introduction

The following is a list of factors that are of fundamental importance in deciding how a treebank and its underlying corpus should be organized. These factors are at the same time conditions of well-formedness of a treebank and may constitute an obstacle against the usability of the same treebank for machine learning purposes. We believe that a treebank should be endowed with:

- Representativeness in terms of text genres
- Representativeness in terms of linguistic theory adherence
- Coherence in allowing Syntactic-Semantic Mapping
- Ability to highlight distinctive linguistic features of the chosen language.

Each factor can impact negatively on the linguistic texture of a treebank, and may undermine its utility as a general linguistic reference point for studies of the chosen language. More specifically, we assume that the above factors would have to be determined on the basis of the following choices:

- Corpus (balanced) and representative of 6 or 7 different text genres vs. unbalanced/mono genre
- Strictly adherent to linguistic principles vs. loosely/non adherent (e.g. more hierarchical vs. less hierarchical)
- Constituency/Dependency/Functional structures are semantically coherent vs. incoherent
- Language chosen is highly canonical and regular vs. almost free word order language.

The final item is clearly inherent in the language chosen and not to be attributed to responsibilities of the annotators. However, as will be shown and discussed at length below, it may turn out to be the main factor in determining the feasibility of the treebank for grammar induction and probabilistic parsing.

In this report I will be concerned with Italian Dependency Treebanks (IDT for short) and will trace minor/major differences with the purpose that they might be useful in case of need for a "normalization" or simply a wrapper to map from one representation to another.
It is important to note that currently available CoNLL style IDT are just three:
- TUT of Tourin University
- TALN of Pisa University
- VIT of Venice University
We will be concerned mainly with common and generally accepted information structure that is distributed on some 8/10 columns. Only TUT and VIT have a version of the treebank that includes Null Elements, however: this will be discussed in a separate section below. We will not compare tags - POS or morphological features, nor functional labels. We will just highlight macro differences and then will concentrate on some thorny issues and compare different solutions. Different treatments in IDT will then be compared with other treebanks available in CoNLL format for similar languages – the AnCora Dependency Treebank (ADT) for Catalan and the CoNLL-X distributed version of a Portuguese treebank CPT- and also with Penn TB. We will in some cases also look at PARC-700 LFG dependency treebank.

1.1 Amalgams

First of all, it is important to note that Italian has two major types of amalgams: prepositional-article compounds – also called D-to-P incorporation - and cliticized verbs. There are two variables at play here: preserve the original orthography and insert additional information in the feature slot; else modify the orthography and distribute the information in different entries. Differently from what TUT and VIT have decided to do - to decompose both types - TALN has decided to decompose only verbal amalgams, that is cliticized verbs. Prepositional amalgams are left as they appear but their nature is clarified by the additional morphological features:

16   alla   a   E   EA   num=s|gen=f   15   comp _   _   INDCOMPL

Now the other variable is constituted by the running word indices associated to each token in the entry text. TALN introduces a new separate index for the enclitic pronoun, as shown below:

35	di	di	E	E	_	34	comp	_	_	RMOD
36	poter-	poterE	V	VM	mod=f	38	modal	INTRANS	_	ARG
37	li	li	P	PE	num=p|per=3|gen=m	38	obj	LOBJ|CLITIC	_	VISITOR
38	utilizzare	utilizzare	V	V	mod=f	35	prep	TRANS	_	INDCOMPL

As can be easily gathered, in this way the original orthography is not preserved and in order to recover the original token POTERLI, two separate dependency representations have to be concatenated, poter-li.

TUT mantains a homogenous behaviour for both types: it multiplies the amalgamated preposition-article into two entries which also have separate indices, as follows:

10	della	DI	PREP	PREP	MONO	9	INDCOMPL	_	_
11	della	IL	ART	ART	DEF|F|SING	10	ARG	_	_
12	cosa	COSA	NOUN	NOUN	COMMON|F|SING	11	ARG	_	_

the same thing happens with cliticized verbs,

18	venderla	VENDERE	VERB	VERB	MAIN|INFINITE|PRES|TRANS	16	INDCOMPL	_	_
19	venderla	LA	PRON	PRON	PERS|F|SING|3|LOBJ|CLITIC	18	OBJ	_	_

In fact, the original complete version of TUT has a different treatment for duplicate entries: they are differentiated from the original previous entry with the additional of a subindex. As for VIT, the situation is a mixture of both approaches: amalgamated preposition-article entries are duplicated but retain the same index and are differentiated with an additional subindex (as the original TUT, but with a morpheme rather than the same lexical entry):

2 sull su part(preposition_plus_article) sp num=s|per=fm 0 pcomp det
2.1 l il art sn num=s|per=fm 0 det def

On the contrary, enclitics are separated and reintroduced as separate indipendent words, and constitute an additional entry index:

6 vengono venire vprog(verb_progressive) ibar punt 8 ibar aux
7 a a pt(verbal_participle) ibar - 8 ibar nil
8 delinearsi delineare vit(verb_trans_infinitive) cl(main) punt - ibar rifl
9 si si clit(clitic_pronoun) compt per=3|gen=m|num=sp 8 compt acc

The need for the separate entry is justified by the need to specify argument relations: this principle is followed by all three IDT. Differences are only visible at the level of the other amalgam which is less relevant than argument specification, but still important at referential level, to indicate the presence of Definiteness.

1.1.1 Amalgams in other treebanks

In the AnCora Dependency treebank the preposition+article amalgams are not decomposed and are treated as TALN does. As for enclitics, their treatment is partly close to TALN and partly to VIT, as can be seen here below,

19	ser	ser	v	vs	mod=n	8	SF	_	_
20	un	un	d	di	num=s|gen=m	21	ESPEC	_	_
21	dret	dret	n	nc	num=s|gen=m	19	ATR	_	_
22	per	per	s	sp	for=s	19	CC	_	_
23	convertir	convertir	v	vm	mod=n	22	SNF	_	_
24	-se	ell	p	p0	per=3	23	MORF	_	_
25	en	en	s	sp	for=s	23	CREG	_	_
26	una	un	d	di	num=s|gen=f	27	ESPEC	_	_

| 27 | aspiració | aspiració | n | nc | num=s\|gen=f25 | SN | _ | _ |

The enclitic is separated from the original wordform and makes a new separately indexed entry with the explicit indication, - the presence of an hyphen -, that it is a morpheme and not a word. This is further specified by the constituent label MORF.

1.2 Multiwords

All IDTs indicate Multiwords, however they do it differently. We use the term multiword in its general sense, to indicate collocations and locutional combination of words.
Locutions are entered in separate lines but are internally indicated by the lemma:

TUT
4	in	IN@QUANTO	CONJ	CONJ	SUBORD|CAUS|LOCUTION	1	RMOD	_	_
5	quanto	IN@QUANTO		CONJ	CONJ	SUBORD|CAUS|LOCUTION	4	CONTIN+LOCUT
		_	_						

TALN
35	per	per@conto@di	E	E	_	28	comp	LOCUTION	_	RMOD
36	conto	per@conto@di	E	S	_	35	prep	LOCUTION	_	CONTIN+LOCUT
37	di	per@conto@di	E	E	_	36	comp	LOCUTION	_	CONTIN+LOCUT

VIT
36 in in_termini_di php(preposition_locution) sp - 34 adj nil
37 termini in_termini_di php(preposition_locution) sp num=p\|per=3 36 adj nil
38 di in_termini_di php(preposition_locution) sp - 37 adj nil

Additional cases of multiword show different behaviours in treatment. TUT will only indicate abbreviations and acronyms in full form, and some Latin locution as here,

38	res	RES@NULLIUS	NOUN NOUN COMMON|F|SING|LOCUTION	37	PREDCOMPL+OBJ	_
	_					
39	nullius	RES@NULLIUS	NOUN NOUN COMMON|F|SING|LOCUTION	38	CONTIN+LOCUT
	_	_				

TALN on the contrary will also indicate as multiwords geographical compound names as for instance,

13	Santa	santa@maria@delle@grazie	S	SP	num=s|gen=f	12	prep	LOCUTION
	CHURCH	ARG						
14	Maria	santa@maria@delle@grazie	S	SP	num=s|gen=f	13	concat	LOCUTION
	CHURCH	CONTIN+LOCUT						
15	delle	santa@maria@delle@grazie	S	SP	num=s|gen=f	14	concat	LOCUTION
	CHURCH	CONTIN+LOCUT						
16	Grazie	santa@maria@delle@grazie	S	SP	num=s|gen=f	15	concat	LOCUTION
	CHURCH	CONTIN+LOCUT						

This is also partly the position taken by VIT, which acknowledges the presence of compound words, which are usually named entities like geographical and person names, but also collocation with a strong lexical dependency, but also of all those cases in which the meaning of the compound is not compositional, as for instance,

2 il il art(article) sn num=s\|gen=m 3 sn def
3 cessate_il_fuoco [cessate,il,fuoco] n(noun) sn num=m 1 obj invar

38 polo_di_attrazione [polo,di,attrazione] n(noun) sn num=m 37 ncomp invar

5 rischio_pentiti rischio_pentiti n(noun) sn num=s\|gen=m 3 obj-theme_aff com

31 rischio_Italia rischio_Italia n(noun) sn num=s\|gen=m 30 pobj com

24 capi_mafia capi_mafia n(noun) sn num=p|gen=m 23 pobj com

10 top_secret top_secret agn(nominal_adjective/past_participle) sa num=m 8 acomp abst

3 Via_XX_Settembre [via,xx,settembre] np(noun_proper_geographic) sn num=s|gen=f 2 pobj-loc geo

81 Còte_d_Azur Còte_d_Azur np(noun_proper_geographic) sn [] 78 sn geo

44 Stati_Uniti Stati_Uniti np(noun_proper_geographic) sn [] 43 pobj-loc geo

33 New_York New_York np(noun_proper_geographic) sn [] 32 pobj-loc geo

On the contrary, multiwords that need not be compounded to retain their semantic identity are indicated as multiwords but remain separate: this happens typically for person proper names:

5 Lamberto Lamberto mw(multi_word) sn num=s|gen=m 6 sn nil
6 Dini Dini nh(noun_human) sn propr 1 sn hum

1.2.1 Multiwords in other treebanks

Again looking at ADT, we see that it is similar to VIT in introducing multiwords directly as single entries rather than as separate tokens. They also decided to treat as multiwords not only locutions of function words like preposition, adverbials and conjunctions, but also collocations and even appositions, as can be seen here below,

9	a_través_de	a_través_de	s	sp	for=s	19	CC	_	_
36	Capa_i_Espasa	Capa_i_Espasa	n	np	_	35	SN	_	_
14	Colla_Gegantera	Colla_Gegantera	n	np	_	12	SN	_	_
6	Lliurament_de_Premis_del_III_Concurs_de_Còmic	Lliurament_de_Premis_del_III_Concurs_de_Còmic	n	np	_	4	CD	_	_
16	Plaça_de_l'_Univers_del_Passeig_de_Riells	Plaça_de_l'_Univers_del_Passeig_de_Riells	n	np	_	14	SN	_	_
18	Adrià_Blasco	Adrià_Blasco	n	np	_	17	SN	_	_
23	Cocoduro_,_el_conde_de_Calvatiesa	Cocoduro_,_el_conde_de_Calvatiesa	n	np	_	21	SN	_	_

1.3 Functional Heads

Another important issue is constituted by the way in which dependencies are organized in all those cases in which heads have specifiers: verbs have auxiliaries, nouns have determiners, etc. Different strategies have been followed to indicate dependencies: for instance TUT makes lexical verbs dependent on auxiliaries, as follows:

TUT COMPLETE
4 Il (IL ART DEF M SING) [6;VERB-SUBJ]
5 proprietario (PROPRIETARIO NOUN COMMON M SING) [4;DET+DEF-ARG]
6 puo` (POTERE VERB MOD IND PRES INTRANS 3 SING) [0;TOP-VERB]
7 compiere (COMPIERE VERB MAIN INFINITE PRES TRANS) [6;VERB+MODAL-INDCOMPL]

TUT CONLL

6	può	POTERE	VERB	VERB	MOD|IND|PRES|INTRANS|3|SING	0	TOP	_	_
7	essere	ESSERE	VERB	VERB	AUX|INFINITE|PRES|INTRANS	11	AUX+PASSIVE	_	_
8	parzialmente	PARZIALMENTE	ADV	ADV	MANNER	11	RMOD	_	_
9	o	O	CONJ	CONJ	COORD|DISJ	8	COORD+BASE	_	_
10	totalmente	TOTALMENTE	ADV	ADV	MANNER	9	COORD2ND+BASE	_	_

11	sacrificata	SACRIFICARE		VERB	VERB	MAIN|PARTICIPLE|PAST|TRANS|SING|F	6
	INDCOMPL	_	_				

On the contrary, TALN behaves like VIT in considering specifiers dependent on their heads:

1	Piero	piero	S	SP	num=s|gen=m	2	mod	_	NAME	SUBJ
2	della	della@francesca	S	SP	_	5	subj	LOCUTION	_	CONTIN+DENOM
3	Francesca	della@francesca	S	SP	_	2	concat	LOCUTION	_	CONTIN+LOCUT
4	ha	avere	V	VA	num=s|per=3|mod=i|ten=p	5	aux	TRANS	_	AUX+TENSE
5	realizzato	realizzare	V	V	num=s|mod=p|gen=m	0	ROOT	TRANS	_	TOP

In fact, TUT behaves homogeneously by considering other functional heads specifiers as encoding the main dependent relation of argument of a predicate, such as the article, here below, where we see that the article IL is encoded as the SUBJect of the verb PUO', and the lexical head PROPRIETARIO - pedending on the article - as the ARG:

TUT COMPLETE
4 Il (IL ART DEF M SING) [6;VERB-SUBJ]
5 proprietario (PROPRIETARIO NOUN COMMON M SING) [4;DET+DEF-ARG]
6 puo` (POTERE VERB MOD IND PRES INTRANS 3 SING) [0;TOP-VERB]
7 godere (GODERE VERB MAIN INFINITE PRES INTRANS) [6;VERB+MODAL-INDCOMPL]

TUT CONLL
1	Il	IL	ART	ART	DEF|M|SING	9	SUBJ	_	_	
2	codice	CODICE	NOUN	NOUN	COMMON|M|SING	1	ARG	_	_	
3	,	#\,	PUNCT	PUNCT	_	1	SEPARATOR	_	_	
4	e	E	CONJ	CONJ	COORD|COORD	1	COORD+BASE	_	_	
5	alcune	ALCUNO	ADJ	ADJ	INDEF|F|PL	4	COORD2ND+BASE	_	_	
6	leggi	LEGGE	NOUN	NOUN	COMMON|F|PL	5	ARG	_	_	
7	speciali	SPECIALE	ADJ	ADJ	QUALIF|ALLVAL|PL	6	RMOD	_	_	
8	,	#\,	PUNCT	PUNCT	_	9	SEPARATOR	_	_	
9	prevedono	PREVEDERE	VERB	VERB	MAIN|IND|PRES|TRANS|3|PL	0	TOP	_	_	
10	alcuni	ALCUNO	ADJ	ADJ	INDEF|M|PL	9	OBJ	_	_	
11	istituti	ISTITUTO	NOUN	NOUN	COMMON|M|PL	10	ARG	_	_	

The same applies to the quantifier "alcuni/alcune" which are treated as functional heads on which lexical semantic heads are dependent. Again, TALN and VIT behave differently, and functional categories are computed as minor dependents on their lexical heads, as shown here

1	La	il	R	RD	num=s|gen=f	3	det	_	_	SUBJ
2	prima	primo	N	NO	num=s|gen=f	3	mod	_	_	RMOD
3	opera	opera	S	S	num=s|gen=f	5	subj	_	_	ARG
4	è	essere	V	VA	num=s|per=3|mod=i|ten=p	5	aux	INTRANS	_	AUX
5	stata	essere	V	V	num=s|mod=p|gen=f	0	ROOT	INTRANS	_	TOP
6	il	il	R	RD	num=s|gen=m	7	det	_	_	PREDCOMPL+SUBJ
7	Libellus	libellus	S	S	num=s|gen=m	5	pred	_	_	ARG*TITLE

And this is VIT,

19 l il art(article) sn num=s|gen=m 20 sn def
20 ufficio ufficio n(noun) sn num=s|gen=m 25 subj-source_info com
21 tecnico tecnico ag(adjective) sa num=s|gen=m 20 mod nil
22 erariale erariale ag(adjective) sa num=s|per=fm 21 mod nil
23 non non neg(negation) ir_infl - 25 neg nil
24 avrà avere auair(auxiliary_avere_mood_irrealis) ir_infl nil 25 ir_infl [sems=aux,mfeats=futL3s]
25 espresso esprimere vppt(verb_trans_past_participle) ir_infl punt 3 ir_infl refl/inform

where both articles and auxiliaries are dependent and minor ones to their lexical and semantic head, respectively, the noun and the verb.

1.3.1 Functional Heads in other treebanks

Neither auxiliaries nor articles, nor other determiners are treated as semantic heads in ADT. The same applies to the CPT. As for Penn Treebank, as will be discussed further on, it treats auxiliaries as heads of lexical verbs, as shown here below:

```
5     said    _    VBD    _    _    0     ROOT _    _
10    that    _    IN     _    _    5     OBJ  _    _
13    it      _    PRP    _    _    14    SBJ  _    _
14    would   _    MD     _    _    10    SUB  _    _
15    appear       _     VB    _    _    14    VC    _    _
16    that    _    IN     _    _    15    OBJ  _    _
17    nothing      _     NN    _    _    19    SBJ  _    _
18    substantive  _    JJ     _    _    17    APPO _    _
19    has     _    VBZ    _    _    16    SUB  _    _
20    changed      _    VBN    _    _    19    VC    _    _
```

PTB conversion tools propose a treatment of auxiliaries identical to the one used by TUT. However, one needs to consider that conversion tools are a mapping from constituency to dependency structure: to understand the reason why the output of the conversion tool has an auxiliary as head of the lexical verb one simply needs to note the fact that PTB decided to use chomsky-adjunction to compute the structure of verbal compounds. In this way complex verbal compounds have one VP node per verb, - auxiliaries included - disregarding the semantic role the verb eventually will have in the final computation. This fact will condition the working of any conversion tool that will map any abstract syntactic constituent nodes to dependency nodes, and could only be amended by some post-processing, which however does not take place.

1.4 Coordination

Coordination is regarded, rightly, a thorny issue and difficult to solve automatically. Solutions adopted by the three IDT differ slightly. We will start by looking at TALN treatment by quoting a long stretch of text which is however very illuminating:

```
35     per    per    E    E    _    32    comp   _    _    RMOD
36     la     il     R    RD   num=s|gen=f    37    det    _    _    ARG
37     determinatezza    determinatezza    S    S    num=s|gen=f    35    prep    _    _    ARG
38     delle  di     E    EA   num=p|gen=f    37    comp   _    _    RMOD
39     figure figura S    S    num=p|gen=f    38    prep   _    _    ARG
40     trattate    trattare    V    V    num=p|mod=p|gen=f    39    mod    TRANS    _
       RMOD+RELCL+REDUC
41     ,      ,      F    FB   _    42    punc   _    _    OPEN+PARENTHETICAL
42     solide solido A    A    num=p|gen=f    39    mod    _    _    APPOSITION
**43   e      e      C    CC   _    42    con    COORD   _    COORD+BASE
44     poliedriche    poliedrico    A    A    num=p|gen=f    42    conj    _    _    COORD2ND+BASE
45     ,      ,      F    FB   _    42    punc   _    _    CLOSE+PARENTHETICAL
**46   e      e      C    CC   _    35    con    COORD   _    COORD+BASE
47     per    per    E    E    _    35    conj   _    _    COORD2ND+BASE
48     l'     il     R    RD   num=s|gen=n    49    det    _    _    ARG
49     assenza    assenza    S    S    num=s|gen=f    47    prep    INTRANS    _    ARG
50     di     di     E    E    _    49    comp   _    _    SUBJ
51     dimostrazioni    dimostrazione    S    S    num=p|gen=f    50    prep    TRANS    _
       ARG
52     classiche    classico    A    A    num=p|gen=f    51    mod    _    _    RMOD
**53   e      e      C    CC   _    35    con    COORD   _    COORD+BASE
54     per    per    E    E    _    35    conj   _    _    COORD2ND+BASE
55     l'     il     R    RD   num=s|gen=n    56    det    _    _    ARG
56     uso    uso    S    S    num=s|gen=m    54    prep    TRANS    _    ARG
```

57	di	di	E	E	_	56	comp	_	_	OBJ
58	regole	regola	S	S	num=p|gen=f	57	prep	_	_	ARG
59	aritmetiche	aritmetico	A	A	num=p|gen=f	58	mod	_	_	RMOD
**60	e	e	C	CC	_	59	con	COORD	_	COORD+BASE
61	algebriche	algebrico	A	A	num=p|gen=f	59	conj	_	_	COORD2ND+BASE

Coordinating conjunctions like "e"/and, are made dependent on the previous head they are coordinators of: so in case they coordinate two nouns, the head will be the previous noun. In case they coordinate two prepositional phrases, the head will be the previous preposition. I highlighted the conjunction by inserting a double star beginning of the line. The first conjunction 43, is dependent on 42, the previous adjective. The second conjunction, 46, coordinates between two prepositions and it is correctly linked to 35, the previous preposition. The same applies to 53; 60, on the contrary, is locally coordinating between two adjectives and is linked to 59.
If we look at TUT, we see the same behaviour:

8	parzialmente	PARZIALMENTE	ADV	ADV	MANNER	11	RMOD	_	_
9	o	O	CONJ	CONJ	COORD|DISJ	8	COORD+BASE	_	_
10	totalmente	TOTALMENTE	ADV	ADV	MANNER	9	COORD2ND+BASE	_	_

Differences emerge clearly due to the treatment of specifiers: here below, the conjunction is linked to the article and not to the head noun.

1	Il	IL	ART	ART	DEF|M|SING	9	SUBJ	_	_
2	codice	CODICE	NOUN	NOUN	COMMON|M|SING	1	ARG	_	_
3	,	#\,	PUNCT	PUNCT	_	1	SEPARATOR	_	_
4	e	E	CONJ	CONJ	COORD|COORD	1	COORD+BASE	_	_
5	alcune	ALCUNO	ADJ	ADJ	INDEF|F|PL	4	COORD2ND+BASE	_	_
6	leggi	LEGGE	NOUN	NOUN	COMMON|F|PL	5	ARG	_	_
7	speciali	SPECIALE	ADJ	ADJ	QUALIF|ALLVAL|PL	6	RMOD	_	_
8	,	#\,	PUNCT	PUNCT	_	9	SEPARATOR	_	_
9	prevedono	PREVEDERE	VERB	VERB	MAIN|IND|PRES|TRANS|3|PL	0	TOP	_	_

5 di (DI PREP MONO) [3;VERB-INDCOMPL-THEME]
6 godere (GODERE VERB MAIN INFINITE PRES INTRANS) [5;PREP-ARG]
7 ((#\(PUNCT) [6;OPEN+PARENTHETICAL]
8 1 (|1| NUM 1) [6;APPOSITION]
9) (#\) PUNCT) [6;CLOSE+PARENTHETICAL]
10 e (E CONJ COORD COORD) [6;COORD+BASE]
11 disporre (DISPORRE VERB MAIN INFINITE PRES TRANS) [10;COORD2ND+BASE]

4 in (IN_QUANTO CONJ SUBORD CAUS LOCUTION) [1;VERB+FIN-RMOD]
5 quanto (IN_QUANTO CONJ SUBORD CAUS LOCUTION) [4;CONTIN+LOCUT]
6 delineano (DELINEARE VERB MAIN IND PRES TRANS 3 PL) [4;CONJ-ARG]
7 e (E CONJ COORD COORD) [6;COORD+BASE]
8 definiscono (DEFINIRE VERB MAIN IND PRES TRANS 3 PL) [7;COORD2ND+BASE]

Another important difference is the linking of the conjoined head, which can vary between linking it directly to the coordinating conjunction - as TUT does - or linking it to the previous head, as the coordinating conjunction, as TALN does.
VIT treats conjunctions differently, and in case of sentence coordination between clauses makes them usually dependent on the governing verb, as follows:

2 fece fare vt(verb_trans_tensed) cl(main) punt - ibar trans/camb_attiv
3 6 6 num(numeral) sn [] 4 sn card
4 morti morte n(noun) sn num=p\|gen=f 5 obj com
5 e e cong(conjunction) coord [] 2 coord sum
6 mille mille num(numeral) sn [] 7 sn card
7 feriti ferito n(noun) sn num=p\|gen=m 5 obj com

4 era essere ause(auxiliary_essere_tensed) ibar punt 5 ibar aux
5 parsa parere vppc(verb_copulative_past_participle) cl(main) punt - ibar refl/exten
6 scontata scontato ag(adjective) sa num=s|gen=f 5 acomp nil
7 e e cong(conjunction) fc [] 5 fc sum
8 non non neg(negation) ibar - 10 neg nil
9 ha avere ausa(auxiliary_avere_tensed) ibar punt 10 ibar aux
10 sorpreso sorprendere vppt(verb_trans_past_participle) ibar punt 7 ibar psych1/exten
11 neppure neppure neg(negation) savv - 10 adj nil
12 mohammed mohammed mw(multi_word) sn num=w|gen=m 13 sn nil
13 salameh salameh nh(noun_human) sn propr 20 obj hum
14 , , punt(sentence_internal) coord punt 20 coord nil
15 ahmad ahmad mw(multi_word) sn num=w|gen=m 16 sn nil
16 ajaj ajaj nh(noun_human) sn propr 20 obj hum
17 , , punt(sentence_internal) coord punt 20 coord nil
18 mahmud mahmud mw(multi_word) sn num=w|gen=m 19 sn nil
19 abouhalima abouhalima nh(noun_human) sn propr 20 obj hum
20 e e cong(conjunction) coord [] 10 coord sum
21 nidal nidal mw(multi_word) sn num=w|gen=m 22 sn nil
22 ayyad ayyad nh(noun_human) sn propr 20 obj hum

In both cases, there is an abstract COORD node which is linked to the main verb and all coordinated heads are dependent on that node: 5, in the first excerpt, and 20 in the second one. This happens not only with complex structures, like NPs of the two excerpts, but also with ellipsed fragments, single words or others, as shown below:

4 riconosciuti riconoscere vppt(verb_trans_past_participle) cl(main) punt - ibar tr/exten
5 colpevoli colpevole ag(adjective) sa num=p|per=fm 4 acomp nil
6 di di pt(verbal_participle) sv2 - 5 sv2 nil
7 aver avere ausai(auxiliary_avere_infinitival) sv2 punt 5 sv2 aux
8 preparato preparare vppt(verb_trans_past_participle) sv2 punt 9 sv2 tr
9 ed ed cong(conjunction) coord [] 5 coord sum
10 eseguito eseguire vppt(verb_trans_past_participle) sv2 punt 9 sv2 tr

It clearly also applies when sentences are coordinated:

49 il il art(article) sn num=s|gen=m 50 sn def
50 cuore cuora n(noun) sn num=p|gen=f 54 subj-theme_unaff com
51 finanziario finanziario ag(adjective) sa num=s|gen=m 50 mod nil
52 della di partd(preposition_di_plus_article) spd num=s|gen=f 50 mod det
52.1 la il art sn num=s|gen=f 50 det def
53 città città n(noun) sn num=f 52 pobj invar
54 tremò tremare vin(verb_intrans_tensed) ibar punt 55 ibar unerg/exten
55 e e cong(conjunction) fc [] 34 fc sum
56 una uno q(quantifier) sq num=s|gen=f 65 subj-agent ind
57 delle delle ccom(conjunction_come_comparative) sc num=m|per=o|md=c|ts=c 56 sc nil
58 due due num(numeral) sn [] 59 sn card
59 torri torre n(noun) sn num=p|gen=f 57 sn com
60 gemelle gemello ag(adjective) sa num=p|gen=f 59 mod nil
61 del di partd(preposition_di_plus_article) spd num=s|gen=m 59 mod det
61.1 l il art sn num=s|per=fm 59 det def
62 world world mw(multi_word) sn num=w|gen=m 64 sn nil
63 trade trade mw(multi_word) sn num=w|gen=m 64 sn nil
64 center center npro(noun_proper_institution) sn [] 61 pobj inst
65 rimase rimanere vin(verb_intrans_tensed) ibar punt 55 ibar unac/posit

1.4.1 Coordination in other treebanks

If we look at ADT we see that it treats coordination in the same way as TALN does:

```
1    Les    el    d    da    num=p|gen=f    2    ESPEC    _    _
2    doctores    doctora    n    nc    num=p|gen=f    13    SUJ    _    _
3    Anna_Ribas    Anna_Ribas    n    np    _    2    SN    _    _
4    i    i    c    cc    _    3    CO    _    _
5    Montserrat_Ventura    Montserrat_Ventura    n    np    _    3    CONJUNCT    _    _

7    formen_part    formar_part    v    vm    num=p|per=3|mod=i|ten=p    5    SF    _    _
8    del    del    s    sp    num=s|gen=m|for=c    7    CREG    _    _
9    programa    programa    n    nc    num=s|gen=m    8    SN    _    _
10   del    del    s    sp    num=s|gen=m|for=c    9    SP    _    _
11   desè    desè    a    ao    num=s|gen=m    12    SADJ    _    _
12   aniversari    aniversari    n    nc    num=s|gen=m    10    SN    _    _
13   del    del    s    sp    num=s|gen=m|for=c    12    SP    _    _
14   Grup_de_Defensa_del_Ter    Grup_de_Defensa_del_Ter    n    np    _    13    SN    _    _
15   i    i    c    cc    _    10    CO    _    _
16   del    del    s    sp    num=s|gen=m|for=c    10    CONJUNCT    _    _
17   Dia_Mundial_del_Medi_Ambient    Dia_Mundial_del_Medi_Ambient    n    np    _    16    SN    _    _
```

Similarly the CPT treats coordination as ADT:

```
17   continua    continuar    v    v-fin    PR|3S|IND    2    A<    _    _
18   a    a    prp    prp    _    17    PRT-AUX<    _    _
19   manter    manter    v    v-inf    _    17    MV    _    _
20   os    o    art    art    <artd>|M|P    21    >N    _    _
21   traços    traço    n    n    M|P    19    ACC    _    _
22   decorativos    decorativo    adj    adj    M|P    21    N<    _    _
23   e    e    conj    conj-c    <co-acc>    21    CO    _    _
24   as    o    art    art    <artd>|F|P    25    >N    _    _
25   clientelas    clientela    n    n    F|P    21    CJT    _    _
26   de    de    prp    prp    _    25    N<    _    _
27   sempre    sempre    adv    adv    _    26    P<    _    _

1    Alguns    algum    pron    pron-det    <quant>|M|P    4    SUBJ    _    _
2    até    até    adv    adv    _    4    ADVL    _    _
3    já    já    adv    adv    _    4    ADVL    _    _
4    desapareceram    desaparecer    v    v-fin    PS|3P|IND    0    STA    _    _
5    ,    ,    punc    punc    _    4    PUNC    _    _
6    como    como    adv    adv    <rel>|<prp>    1    N<    _    _
7    o    o    pron    pron-det    <artd>|M|S    6    SUBJ    _    _
8    de    de    prp    prp    _    7    N<    _    _
9    Castro_Verde    Castro_Verde    prop    prop    M|S    8    P<    _    _
10   ,    ,    punc    punc    _    4    PUNC    _    _
11   e    e    conj    conj-c    <co-vfin>|<co-fmc>    4    CO    _    _
12   outros    outro    pron    pron-det    <diff>|M|P    13    SUBJ    _    _
13   têm    ter    v    v-fin    PR|3P|IND    4    CJT    _    _
14   vindo    vir    v    v-pcp    _    13    AUX    _    _
```

On the contrary, PTB treats coordination as TUT does: non-clause level coordination is made by linking conjunction to the first coordinated head, then second and following ones are linked to the conjunction or punctuation mark:

```
2    the    _    DT    _    _    4    NMOD    _    _
3    stock    _    NN    _    _    4    NMOD    _    _
4    market    _    NN    _    _    5    SBJ    _    _
5    does    _    VBZ    _    _    1    OBJ    _    _
```

6	n't	_	RB	_	_	5	ADV	_	_
7	crash	_	VB	_	_	5	VC	_	_
8	again	_	RB	_	_	7	TMP	_	_
9	and	_	CC	_	_	7	COORD	_	_
10	completely	_	RB	_	_	11	ADV	_	_
11	discredit	_	VB	_	_	9	CONJ	_	_
12	yuppies	_	NNS	_	_	11	OBJ	_	_
13	and	_	CC	_	_	12	COORD	_	_
14	trading	_	NN	_	_	15	NMOD	_	_
15	rooms	_	NNS	_	_	13	CONJ	_	_

52	genius	_	NN	_	_	50	COORD	_	_
53	and	_	CC	_	_	52	COORD	_	_
54	energy	_	NN	_	_	53	CONJ	_	_

2	with	_	IN	_	_	11	ADV	_	_
3	all	_	PDT	_	_	5	NMOD	_	_
4	the	_	DT	_	_	5	NMOD	_	_
5	money	_	NN	_	_	2	PMOD	_	_
6	and	_	CC	_	_	5	COORD	_	_
7	glamour	_	NN	_	_	6	CONJ	_	_

8	Michelle	_	NNP	_	_	15	SBJ	_	_
9	,	_	,	_	_	8	P	_	_
10	a	_	DT	_	_	13	NMOD	_	_
11	junior	_	JJ	_	_	13	NMOD	_	_
12	bond	_	NN	_	_	13	NMOD	_	_
13	trader	_	NN	_	_	8	APPO	_	_
14	,	_	,	_	_	8	P	_	_

However, clause-level coordination is done by linking conjunction/punctuation to the first governing verb, while second and following coordinated verbs are linked to the conjunction and not to first head:

1	The	_	DT	_	_	3	NMOD	_	_
2	trade	_	NN	_	_	3	NMOD	_	_
3	figures	_	NNS	_	_	4	SBJ	_	_
4	turn	_	VBP	_	_	0	ROOT	_	_
5	out	_	RP	_	_	4	PRT	_	_
6	well	_	RB	_	_	4	ADV	_	_
7	,	_	,	_	_	4	P	_	_
8	and	_	CC	_	_	4	COORD	_	_
9	all	_	PDT	_	_	13	NMOD	_	_
10	those	_	DT	_	_	13	NMOD	_	_
11	recently	_	RB	_	_	12	AMOD	_	_
12	unloaded	_	VBN	_	_	13	NMOD	_	_
13	bonds	_	NNS	_	_	14	SBJ	_	_
14	spurt	_	VBP	_	_	8	CONJ	_	_
15	in	_	IN	_	_	14	ADV	_	_
16	price	_	NN	_	_	15	PMOD	_	_

Pros and cons for one approach versus the VIT approach are:
- theoretically speaking the VIT approach is the one that ensures semantic consistency, any coordination corresponding to a set of individual representations;
- from a practical point of view a COORD constituent or dependency node does not constitute sufficient information by itself of the coordinated elements – being a conjunction or a punctuation mark, and a further node needs to be searched for to know what is being coordinated;

- however a similar problem arises for the other versions, for the second/all nodes attached to the coordinating conjunction. In this case, in order to get the head of the dependent two more nodes need to be searched: beside the coordinating conjunction, also the first conjunct must be searched because it is just this one that will carry the dependent node to the head of the coordination.

The most straighforward coordinate structure then seems to be the one used by TALN, ADT and CPT, in which all coordinated items – including the coordinating head – are made dependent on the first conjunct. However in this way the semantic dependence of the coordinated set is marked only in one place and needs to be recovered from there for the other conjuncts: in particular, we know that a nominal head is SUBJect or OBJect from the first conjunct and not locally.

So, eventually all three versions require some extra specialization in order to make semantic processing work, in particular to build correct predicate-argument structures.

1.5 CHE complementizer

In all romance languages but also in English, the complementizer is a wordform used for many syntactic purposes. This will also allow us to start looking into the problem of Null or Empty Elements and their role in the dependency structure. At first however, we will limit ourselves at looking in detail at the way in which surface and not deep dependency structures deal with CHE (QUE – THAT).

The complementizer is a function word that is used for at least the following purposes:
1. to indicate the beginning of a complement clause or sentential complement
2. to indicate the beginning of a relative clause
3. sometimes, to indicate the beginning of an adjunct clause or sentential adjunct
4. to signal the beginning of an interrogative clause
5. to signal the beginning of an exclamative clause
6. to indicate the second element of a comparative structure

In surface or shallow dependency structures as the ones that will concern us first, the difference between all of these functions will not always be sufficiently signalled. As will become clear, in fact there are only two possible ways of treating CHE, either as head or as dependent, and we will expect TUT to treat it as head, while TALN will do the the opposite. But this is not the case. So we will start by looking at each different type of structure and comment on each one.

1.5.1   CHE as Functional Head for Sentential Complements

The first case is that of a clear functional head in which the only information to be passed down is that the complement clause main VERB is dependent on the governing verb of the complement clause, usually a communication verb like DIRE, RITENERE, etc.

We look at TALN first, which correctly links CHE to the governing verb and the Verb of the Complement Clause to CHE. In this way CHE complementizer is treated as a HEAD and not simply as a dependent, which is what happens in TUT:

14	ritenere	ritenere	V	V	mod=f13	prep	TRANS	_	ARG	
15	che	che	C	CS	_	14	arg	NEUTRAL	_	OBJ
16	queste	questo	P	PD	num=p|gen=f	23	subj	LSUBJ+LOBJ+OBL	_	SUBJ
17	,	,	F	FB	_	18	punc	_	_	OPEN+PARENTHETICAL
18	se	se	C	CS	_	23	mod	COND	_	RMOD
19	ben	bene	B	B	_	20	mod	MANNER	_	RMOD
20	indirizzate	indirizzare	V	V	num=p|mod=p|gen=f	18	sub	TRANS	_	ARG
21	,	,	F	FB	_	18	punc	_	_	CLOSE+PARENTHETICAL
22	potessero	potereV	VM	num=p|per=3|mod=c|ten=i	23	modal	INTRANS	_	ARG	
23	far	fare	V	V	mod=f	15	sub	TRANS	_	INDCOMPL	

14	documenti	documento	S	S	num=p|gen=m	15	subj	_	_	ARG
15	confermano	confermare	V	V	num=p|per=3|mod=i|ten=p	3	conj	TRANS	_	COORD2ND+BASE
16	che	che	C	CS	_	15	arg	NEUTRAL	_	OBJ
17	Piero	piero	S	SP	num=s|gen=m	18	subj	_	NAME	SUBJ
18	fu	essereV	V	num=s|per=3|mod=i|ten=s	16	sub	INTRANS	_	ARG	
19	suo	suo	A	AP	num=s|gen=m	20	mod	P-3|P-SING	_	PREDCOMPL+SUBJ

20	assistente	assistente	S	S	num=s|gen=n	18	pred	_	_	ARG

And this is TUT treatment of CHE complementizer:

21	ricordando	RICORDARE	VERB	VERB	MAIN|GERUND|PRES|TRANS	18	RMOD	_	_
22	che	CHE	CONJ	CONJ	SUBORD|NEUTRAL	21	OBJ	_	_
23	l'	IL	ART	ART	DEF|F|SING	27	SUBJ	_	_
24	Albania	ALBANIA	NOUN	NOUN	PROPER|F|SING|££STATE	23	ARG	_	_
25	e'	ESSERE	VERB	VERB	AUX|IND|PRES|INTRANS|3|SING	27	AUX+TENSE	_	_
26	gia'	GIÀ	ADV	ADV	TIME	27	RMOD	_	_
27	diventata	DIVENTARE	VERB	VERB	MAIN|PARTICIPLE|PAST|INTRANS|SING|F	22	ARG	_	_

TUT treats as head a typical functional word like CHE complementizer, thus perfectly in line with the treatment of the other function words that we saw previously. This treatment is perfectly correct and in line with all the other treebanks, VIT included.

```
0 Si si clit(clitic_pronoun) ibar per=3|gen=m|num=sp 1 ibar nom
1 sa sapere vt(verb_trans_tensed) cl(main) punt - ibar trans/presuppositional
1.11 pro si pro(little_pro) sn per=3|gen=m|num=sp 0 s_impers-actor nom
2 soltanto soltanto cong(conjunction) compt [] 1 cong q
3 che che pk(complementizer) fac - 1 fcomp nil
4 l' il art(article) sn num=s|gen=f 5 sn def
5 offerta offerta n(noun) sn num=s|gen=f 8 subj-agent com
6 italiana italiano ag(adjective) sa num=s|gen=f 5 mod nil
7 sarà essere aueir(auxiliary_essere_mood_irrealis) ir_infl punt 8 ir_infl aux
8 compresa comprendere vppt(verb_trans_past_participle) ir_infl punt 3 ir_infl tr/inclus
```

The same argument will apply to CHE complementizer governed by an epistemic nominal (possibly deverbal) head as shown here below, from TALN:

32	la	il	R	RD	num=s|gen=f	33	det	_	_	SUBJ
33	congettura	congettura	S	S	num=s|gen=f	43	subj	_	_	ARG
34	del	di	E	EA	num=s|gen=m	33	comp	_	_	RMOD
35	Vasari	vasari	S	SP	_	34	prep	_	_	ARG
36	che	che	C	CS	_	33	arg	NEUTRAL	_	RMOD
37	abbia	avere	V	VA	num=s|per=3|mod=c|ten=p	38	aux	TRANS	_	AUX+TENSE
38	preso	prendere	V	V	num=s|mod=p|gen=m	36	sub	TRANS	_	ARG
39	il	il	R	RD	num=s|gen=m	40	det	_	_	OBJ
40	cognome	cognome	S	S	num=s|gen=m	38	obj	_	_	ARG
41	dalla	da	E	EA	num=s|gen=f	38	comp	_	_	RMOD
42	madre	madre	S	S	num=s|gen=f	41	prep	_	_	ARG
43	è	essere	V	V	num=s|per=3|mod=i|ten=p	2	conj	INTRANS	_	COORD2ND+ADVERS
44	inattendibile	inattendibile	A	A	num=s|gen=n	43	pred	_	_	PREDCOMPL+SUBJ

1.5.2   CHE as Functional Head for Sentential Complements in other treebanks

We will look at PTB at first which proposes a dependency structure similar to VIT:

5	said	_	VBD	_	_	0	ROOT	_	_
10	that	_	IN	_	_	5	OBJ	_	_
13	it	_	PRP	_	_	14	SBJ	_	_
14	would	_	MD	_	_	10	SUB	_	_
15	appear	_	VB	_	_	14	VC	_	_
16	that	_	IN	_	_	15	OBJ	_	_
17	nothing	_	NN	_	_	19	SBJ	_	_
18	substantive	_	JJ	_	_	17	APPO	_	_
19	has	_	VBZ	_	_	16	SUB	_	_
20	changed	_	VBN	_	_	19	VC	_	_

This is what AnCORA does with QUE Subord

18	ha	haver	v	va	num=s|per=3|mod=i|ten=p	19	AUX	_	_
19	indicat	indicar	v	vm	num=s|mod=p|gen=m	0	S	_	_
20	que	que	c	cs	_	22	SUBORD	_	_
21	es	es	p	p0	_	22	IMPERS	_	_
22	tracta	tractar	v	vm	num=s|per=3|mod=i|ten=p	19	CD	_	_

25	hem	haver	v	va	num=p|per=1|mod=i|ten=p	26	AUX	_	_
26	coincidit	coincidir	v	vm	num=s|mod=p|gen=m	3	CONJUNCT	_	_
27	plenament	plenament	r	rg	_	26	CC	_	_
28	que	que	c	cs	_	31	SUBORD	_	_
29	ha	haver	v	va	num=s|per=3|mod=i|ten=p	31	AUX	_	_
30	de	de	s	sp	for=s	31	PREP	_	_
31	ser	ser	v	vs	mod=n	26	CREG	_	_

| 23 | comprovarà | comprovar | v | vm | num=s\|per=3\|mod=i\|ten=f | 19 | SF | _ | _ |
| 24 | - | - | F | Fg | _ | 26 | PUNC | _ | _ |

...

28	que	que	c	cs	_	30	SUBORD	_	_
29	Madrid	Madrid	n	np	_	30	SUJ	_	_
30	ofereix	oferir	v	vm	num=s|per=3|mod=i|ten=p	23	CD	_	_
31	més	més	r	rg	_	32	ESPEC	_	_
32	avantatges	avantatge	n	nc	num=p|gen=m	30	CD	_	_
33	que	que	c	cs	_	35	SUBORD	_	_
34	altres	altre	d	di	num=p|gen=c	35	ESPEC	_	_
35	espais	espai	n	nc	num=p|gen=m	32	COMP	_	_
36	espanyols	espanyol	a	aq	num=p|gen=m	35	SADJ	_	_

What we see here is the treatment of QUE complementizer as a functional head linked to the main verb of the sentential complement TRACTAR, SER, COMPROVAR; and in turn these verbs are linked to their governing verb in the main higher clause. The same treatment is done by the Portuguese treebank, that we report below:

15	secretário_de_Estado	secretário_de_Estado	n	n	M|S	13	SUBJ	_	_
16	,	,	punc	punc	_	15	PUNC	_	_
17	que	que	pron	pron-indp	<rel>|M|S	18	SUBJ	_	_
18	considera	considerar	v	v-fin	PR|3S|IND	15	N<PRED	_	_
19	que	que	conj	conj-s	_	25	SUB	_	_
20	a	o	art	art	<artd>|F|S	21	>N	_	_
21	solução	solução	n	n	F|S	25	SUBJ	_	_
22	para	para	prp	prp	_	21	N<	_	_
23	as	o	art	art	<artd>|F|P	24	>N	_	_
24	autarquias	autarquia	n	n	F|P	22	P<	_	_
25	é	ser	v	v-fin	PR|3S|IND	18	ACC	_	_

So eventually, we have come up with three different approaches to the problem of treating CHE/QUE/THAT complementizer
- linked to the governing verb in the main clause
  - o the governed verb linked to the governing verb (TUT, TALN)
  - o the governed verb linked to CHE (VIT)
- linked to the governed verb in the sentential complement (ALL Other TBs)

1.5.3 CHE as beginner of a Relative Clause

In shallow or surface dependency treebank, relative pronouns are only visible if lexically expressed. So the case of implicit relative pronoun signalled by CHE complementizer does not exist – but it does in all deep dependency treebanks as we will show below. What is usually done, is the transformation of CHE itself into a relative pronoun like CHI, CUI or QUALE and others. We will discuss other cases

below. However, even though this is what all shallow treebanks do, the treatment of CHE is not uniform. We look at first into TALN:

```
40   luce    luce    S   S    num=s|gen=f      2    conj   _     _      COORD2ND+BASE
41   non     non     B   BN   _          42    neg    _     _      RMOD
42   fisica  fisico  A   A    num=s|gen=f      40   mod    _     _      RMOD
43   ma      ma      C   CC   _          42    con    ADVERS     _      COORD+ADVERS
44   intellettuale intellettuale A   A    num=s|gen=n      42   conj   _     _
     COORD2ND+ADVERS
45   ,       ,       F   FF   _          47    punc   _     _      SEPARATOR
46   che     che     P   PR   num=n|gen=n      47   subj   LSUBJ+LOBJ _    SUBJ
47   pervade pervadere V   V    num=s|per=3|mod=i|ten=p   40   mod_rel    TRANS
     _       RMOD+RELCL
```


```
1    La      il      R   RD   num=s|gen=f      4    det    _     _      SUBJ
2    prima   primo   N   NO   num=s|gen=f      4    mod    _     _      RMOD
3    sua     suo     A   AP   num=s|gen=f      4    mod    P-3|P-SING  _     RMOD
4    opera   opera   S   S    num=s|gen=f      9    subj   _     _      ARG
5    che     che     P   PR   num=n|gen=n      8    subj_pass   LSUBJ+LOBJ _   OBJ/SUBJ
6    ci      ci      P   PC   num=n|gen=n      8    comp_ind    LOC|CLITIC _   RMOD
7    è       essere  V   VA   num=s|per=3|mod=i|ten=p   8    aux    TRANS    _
     AUX+PASSIVE
8    conservata conservare V   V    num=s|mod=p|gen=f    4    mod_rel    TRANS    _
     RMOD+RELCL
9    è       essere  V   V    num=s|per=3|mod=i|ten=p   0    ROOT   TRANS    _     TOP
10   la      il      R   RD   num=s|gen=f11    det    _     _      PREDCOMPL+SUBJ
11   Madonna madonna S   SP   num=s|gen=f      9    pred   _     NAME ARG*TITLE
12   col     con     E   EA   num=s|gen=m      11   concat _     _      RMOD*TITLE
13   Bambino bambino S   S    num=s|gen=m      12   concat _     _      ARG*TITLE
```

As can be easily noticed, CHE is only treated as functional dependent on the verb of Relative Clause. As functional dependent it shouldn't be associated to any semantic function: on the contrary it is marked as SUBJect of the verb PERVADERE and CONSERVARE. These verbs are then linked as dependents to the head noun governing the relative pronoun – in this case coinciding to CHE – thus introducing a specialization for verb relations in addition to linking to other verbs, or to subordinating/coordinating conjunctions. TUT treatment is identical to TALN:

```
1    Nelle   IN    PREP PREP MONO15    RMOD_     _
2    Nelle   IL    ART  ART  DEF|F|PL  1    ARG  _     _
3    societa'      SOCIETÀ    NOUN NOUN COMMON|F|ALLVAL    2    ARG  _     _
4    di      DI    PREP PREP MONO3     RMOD_     _
5    Tirana  TIRANA     NOUN NOUN PROPER|F|SING|££CITY   4    ARG  _     _
6    che     CHE   PRON PRON RELAT|ALLVAL|ALLVAL|LSUBJ+LOBJ   8    SUBJ _     _
7    hanno   AVERE      VERB VERB AUX|IND|PRES|TRANS|3|PL   8    AUX+TENSE _   _
8    truffato      TRUFFARE   VERB VERB MAIN|PARTICIPLE|PAST|TRANS|SING|M    3
     RMOD+RELCL    _     _
...
14   c'      CI    PRON PRON LOC|ALLVAL|ALLVAL|LOC|CLITIC    15   RMOD_     _
15   e'      ESSERE     VERB VERB MAIN|IND|PRES|INTRANS|3|SING    0    TOP  _     _
```

In this example, the prepositional phrase headed by NELLE is linked to the main verb ESSERE in 15, and the noun SOCIETA' governing the relative clause is linked to NELLE; TRUFFARE the verb of the relative clause is linked to SOCIETA', but the information as to what grammatical function this head noun plays in the relative clause is indicated in the CHE.
So the recovery of grammatical relations for TRUFFARE – and the same will apply to previous cases of TALN – will go through a process of restructuring of the argument subject SOCIETA' with CHE that

works as functional head but does not have any explicit link with it. This is different from what happens in VIT:

13 emergere emergere n(noun) sn num=s|gen=m 12 pobj com
14 di di pd(preposition_di) spd - 13 mod nil
15 una uno art(article) sn num=s|gen=f 17 sn ind
16 crescente crescente ag(adjective) sa num=s|per=fm 17 mod nil
17 concorrenza concorrenza n(noun) sn num=s|gen=f 14 pobj com
18 che che rel(relative) f2 - 17 subj-theme_aff nil
19 si si clit(clitic_pronoun) ibar per=3|gen=m|num=sp 22 ibar acc
20 è essere ause(auxiliary_essere_tensed) ibar punt 22 ibar aux
21 progressivamente progressivamente avv(adverb) ibar [] 22 adjv mn
22 spostata spostare vppin(verb_intrans_past_participle) ibar punt 18 ibar refl_in/posit

In this case we see that CHE is bound to its noun head CONCORRENZA which has a certain role in the sentence to which it belongs, headed by EMERGERE through preposition DI; then CHE is the intermediary between the head noun and verb of the relative SPOSTARE which is linked to it. The role played by the CHE is played, in the case of a deep representation, by the empty category, otherwise it will be the CHE itself to carry the information of both grammatical function and semantic role, as happens in previous treebanks.

1.5.4 CHE as beginner of a Relative Clause in other treebanks

We will look at PTB as it has been converted by Malt conversion tools for CoNLL:

```
10    has     _    VBZ   _    _    0    ROOT _    _
11    proposed    _    VBN   _    _    10   VC    _    _
12    a       _    DT    _    _    14   NMOD _    _
13    sweeping    _    JJ    _    _    14   NMOD _    _
14    restructuring    _    NN    _    _    11   OBJ   _    _
15    that    _    WDT   _    _    16   SBJ   _    _
16    would   _    MD    _    _    14   NMOD _    _
17    pare    _    VB    _    _    16   VC    _    _
18    it      _    PRP   _    _    17   OBJ   _    _
```

here we see that the head noun RESTRUCTURING is the OBJect of its main verb PROPOSE and the CHE/THAT is the SUBJect of the verb of the relative clause PARE through WOULD the modal auxiliary: in turn WOULD is linked to the head noun of the relative RESTRUCTURING. So eventually, what PTB does – or at least its conversion tool – is proposing the same distribution of dependencies as TUT and TALN have done. Here below is what the portuguese treebanks has for relatives where we see that the head noun of the relative is linked and has roles in the main clause, the relative pronoun carried the roles it has in the relative clause and is linked to its verb, the verb of the relative clause is linked to the head noun:

```
7     os    o      art    art    <-sam>|<artd>|M|P 8    >N    _    _
8     tempos    tempo n    n    M|P    6    P<    _    _
9     que    que    pron   pron-indp    <rel>|M|P    10   SUBJ _    _
10    correm    correr v    v-fin    PR|3P|IND    8    N<    _    _


23    os    o      pron   pron-det    <dem>|M|P 22    P<    _    _
24    que    que    pron   pron-indp    <rel>|M|P    25   SUBJ _    _
25    querem    querer    v    v-fin    PR|3P|IND    23   N<    _    _
26    fugir fugir v    v-inf    _    25   ACC   _    _


7     milhões    milhão    n    n    F|P    4    P<    _    _
8     que    que    pron   pron-indp    <rel>|M|P    12   ACC   _    _
9     o    o      art    art    <artd>|M|S    10   >N    _    _
```

10	Ministério_do_Planeamento_e_Administração_do_Território									
	Ministério_do_Planeamento_e_Administração_do_Território	prop	prop	M|S	12	SUBJ	_		_
11	já	já	adv	adv	_	12	ADVL	_		_
12	gasta	gastar	v	v-fin	PR|3S|IND	7	N<	_		_

this is identical to what AnCORA has done,

19	empresa	empresa	n	nc	num=s|gen=f	14	CONJUNCT	_		_
20	,	,	F	Fc	_	23	PUNC	_		_
21	"	"	F	Fe	_	23	PUNC	_		_
22	que	que	p	pr	num=n|gen=c	23	SUJ	_		_
23	comprovarà	comprovar	v	vm	num=s|per=3|mod=i|ten=f	19	SF	_		_
24	-	-	F	Fg	_	26	PUNC	_		_

20	una	un	d	di	num=s|gen=f	21	ESPEC	_		_
21	part	part	n	nc	num=s|gen=f	18	SN	_		_
22	petita	petit	a	aq	num=s|gen=f	21	SADJ	_		_
23	que	que	p	pr	num=n|gen=c	24	SUJ	_		_
24	afecta	afectar	v	vm	num=s|per=3|mod=i|ten=p	21	SF	_		_

26	el	el	d	da	num=s|gen=m	27	ESPEC	_		_
27	treball	treballn	nc		num=s|gen=m	24	CD	_		_
28	que	que	p	pr	num=n|gen=c	30	SUJ	_		_
29	es	es	p	p0	_	30	PASS	_		_
30	desplaça	desplaçar	v	vm	num=s|per=3|mod=i|ten=p	27	SF	_		_

Now, the difference in treatment is clear and can be summarized below. As with the complementizer of sentential complements, we have come up with three different approaches to the problem of treating CHE/QUE/THAT relative pronoun

- linked to the governing Noun head (VIT)
    - o the governed verb of the relative clause linked to CHE
- linked to the governed verb in the relative clause (ALL Other TBs)
    - o the governed verb being an auxiliary if present (TUT, PTB)
    - o the governed verb being the lexical semantic verb (ALL Other TBs)

To substantiate our proposal we now show the output of other DEEP treebanks, like the one proposed by PARC- XEROX and organized on the basis of LFG theoretical framework. This treebank is based on the same set of newspaper articles from Wall Street Journal of PTB and in particular it contains all articles belonging to section 23. We will indicate only relevant dependency nodes to highlight differences from previous treebanks.

- THAT COMPLEMENTIZER OF SENTENTIAL COMPLEMENTS

sentence_form(The Justice Department announced that the FBI has been given the authority to seize U.S. fugitives overseas without the permission of foreign governments.)

  comp(announce~0, give~52)
  subord_form(give~52, that)

As can be seen, THAT is related to the governed verb GIVE and not to governing verb ANNOUNCE, which is then linked to GIVE by a COMP grammatical relation.

- THAT COMPLEMENTIZER OF RELATIVE CLAUSES

sentence_form(But not much money was spent on the shows\, either\, a situation that encouraged cheap-to-make talk and game shows\, while discouraging expensive-to-produce dramas.)

  pron_rel(encourage~12, pro~16)

```
subj(encourage~12, pro~16)
topic_rel(encourage~12, pro~16)
case(pro~16, nom)
num(pro~16, sg)
pers(pro~16, 3)
pron_form(pro~16, that)
pron_type(pro~16, relative)
adjunct(situation~7, encourage~12)
adjunct_type(encourage~12, relative)
```

THAT is treated as complementizer and is linked to PRO relative. PRO is marked as NOMINATIVE and is linked to ENCOURAGE, the verb of the relative. SITUATION is linked to ENCOURAGE, and ENCOURAGE is linked to PRO. This is partially coincident with what VIT has done and partially with TUT and TALN. In the DEEP treebank, it is the implicit pronoun PRO which plays the role of intermediary between the complementizer THAT and the verb of the relative. The relative clause as a whole is then linked as dependent to the head noun SITUATION.

Let's look now at the deep version of VIT:

```
13 emergere emergere n(noun) sn num=s|gen=m 12 pobj com
14 di di pd(preposition_di) spd - 13 mod nil
15 una uno art(article) sn num=s|gen=f 17 sn ind
16 crescente crescente ag(adjective) sa num=s|per=fm 17 mod nil
17 concorrenza concorrenza n(noun) sn num=s|gen=f 14 pobj com
18 che che rel(relative) f2 - 17 binder nil
19 si si clit(clitic_pronoun) ibar per=3|gen=m|num=sp 22 ibar acc
20 è essere ause(auxiliary_essere_tensed) ibar punt 22 ibar aux
21 progressivamente progressivamente avv(adverb) ibar [] 22 adjv mn
22 spostata spostare vppin(verb_intrans_past_participle) ibar punt 18 ibar refl_in/posit
22.11 rel_pro concorrenza rel(relative_pronoun) bindee num=s|gen=f ant=17 subj-theme_aff nil
```

From a computational point of view, a chain allows recovering easily all relations needed in case of further processing of dependency structures for semantic purposes. Here the chain goes from the relative pronoun to its noun head binder, however all relevant information is already encoded in the rel_pro additional entry which as a null element has already undergone head substitution with its binder antecedent. The index carried by the null rel_pro is the same as that of the verb of the relative, in this way partly resembling the linking of the verb to relative head noun. The complementizer can in this case simply be done away with in the semantics, as would happen with the case of sentential complements, being a functional head with no semantic content.

If we look at PTB original constituency structure, we see that relative pronouns are embedded in the NP of the head noun they depend on; then the relative pronoun and its preposition if existent, and following head noun in the case of WHOSE, are all included in a SBAR structure, that is the relative clause, that they are beginners of. Finally, the index associated to the relative pronoun is then "landed" in a position around the verb of the relative clause, either just before or after in case of argument relation, or after the expressed arguments in case of adjunct relation as shown below:

```
(NP (NP three levels)
    (SBAR (WHPP-1 on
            (WHNP which))
        (S (NP-SBJ *)
          (VP to
            (VP treat
              (NP the subject)
              (PP-LOC *T*-1))))))
```


```
(S (NP-SBJ-2 I)
  (VP do n't
    (VP know
```

```
(SBAR (WHNP-1 what)
    (S (NP-SBJ *-2)
       (VP to
          (VP do
             (NP *T*-1)
             (PP-CLR with
                    (NP this sentence)))))))))))
```

In some case as the following one, the trace lands deeper below, inside an infinitival, complement of the main verb of the relative clause:

```
(S (NP-SBJ The following prompts)
    (VP allow
       (S (NP-SBJ you)
          (VP to
             (VP specify
                (SBAR (WHADVP-1 how)
                   (S (NP-SBJ you)
                      (VP want
                         (S (NP-SBJ the printed output)
                            (VP to
                               (VP look
                                  (ADVP-MNR *T*-1))))))))))))))
```

Penn Treebank also signals with traces passive constructions so that in case of a passive relative clause the number of traces is doubled:

```
(NP (NP a factor)
    (SBAR (WHNP-1 which)
       (S (NP-SBJ *T*-1)
          (VP has
             (VP caused
                (S (NP-SBJ the Citizens Group)
                   (VP to
                      (VP obtain
                         (NP signatures)
                         (PP under
                            (SBAR-NOM
                             (WHNP-2 what)
                             (S (NP-3 *T*-2)
                                (VP were
                                   (VP termed
                                      (S (NP-SBJ *-3)
                                         ``
                                         (NP-PRD false pretenses)
                                         ")))))))))))))))
```

There are no principled reasons for not using a chain-like description of the relative clause structure, from what is contained in this annotation. If embedding is used to detect dependency, then the relative pronoun should always be dependent on the head noun it is governed by. The presence of the trace in the following clause should then be used to make the verb of the relative clause dependent on the relative pronoun. Romance languages have a much wider inventory of relative pronouns than German ones, in particular Italian has certainly the most extended one, and we will discuss them in the section below.

1.6 Lexical Relative and Interrogative Pronouns

Lexical pronouns have a different status from CHE complementizer at least in as far as they would contain internally enough information to an independent semantic specification. In fact, relative pronouns can also be subdivided by the traditional categorization of "analytic" (IL QUALE, etc) vs. ""synthetic" (CHE, CUI) pronouns: this subdivision, however, is irrelevant to the discussion of

dependency structure. We will look into CUI and QUALE preceded or not by preposition. From the structures below, we see that the same technique is being used for linking relative pronouns and their prepositions: dependency links are established as before, between the verb of the relative clause which is made dependent upon the nominal head of the relative pronoun; then the preposition is made dependent on the verb of the relative clause, and the relative pronoun on the preposition. In all the examples below, recovering the binder and noun antecedent of the relative pronoun requires at least a search in two steps as will be explained below. We will start by looking at excerpts from TUT deep:

- DAL QUALE
6 shock (SHOCK NOUN COMMON M ALLVAL) [4;DET+INDEF-ARG]
7 dal (DA PREP MONO) [10;PREP-RMOD-LOC+METAPH]
7.1 dal (IL ART DEF M SING) [7;PREP-ARG]
8 quale (QUALE PRON RELAT ALLVAL SING 3 LSUBJ+LOBJ+OBL) [7.1;DET+DEF-ARG]
9 e` (ESSERE VERB AUX IND PRES INTRANS 3 SING) [10;AUX+TENSE]
10 riemersa (RIEMERGERE VERB MAIN PARTICIPLE PAST INTRANS SING F) [6;VERB-RMOD+RELCL]

- IN CUI
31 nei (IN PREP MONO) [30;PREP-RMOD-LOC+IN]
31.1 nei (IL ART DEF M PL) [31;PREP-ARG]
32 luoghi (LUOGO NOUN COMMON M PL) [31.1;DET+DEF-ARG]
33 abituali (ABITUALE ADJ QUALIF ALLVAL PL) [32;ADJC+QUALIF-RMOD]
34 in (IN PREP MONO) [39;PREP-RMOD-LOC+IN]
35 cui (CUI PRON RELAT LIOBJ+OBL) [34;PREP-ARG]
36 di (DI PREP MONO) [39;VERB-INDCOMPL-THEME]
37 Tv (|Tv| NOUN PROPER) [36;PREP-ARG]
38 si (SI PRON REFL-IMPERS ALLVAL ALLVAL 3 LSUBJ+LOBJ+LIOBJ CLITIC) [39;VERB-SUBJ/VERB-SUBJ+IMPERS]
39 parla (PARLARE VERB MAIN IND PRES INTRANS 3 SING) [32;VERB-RMOD+RELCL]

1.6.1 Pied Piping relative pronouns
In this section we present relative pronouns headed by a preposition which in turn are embedded in another prepositional phrase that is then governed by the nominal head of the relative:

- DI CUI
8 e` (ESSERE VERB MAIN IND PRES INTRANS 3 SING) [0;TOP-VERB]
9 il (IL ART DEF M SING) [8;VERB-PREDCOMPL+SUBJ]
10 coronamento (CORONAMENTO NOUN COMMON M SING CORONARE TRANS) [9;DET+DEF-ARG]
11 del (DI PREP MONO) [10;NOUN-OBJ]
11.1 del (IL ART DEF M SING) [11;PREP-ARG]
12 dialogo (DIALOGO NOUN COMMON M SING) [11.1;DET+DEF-ARG]
13 di (DI PREP MONO) [17;VISITOR]
14 cui (CUI PRON RELAT LIOBJ+OBL) [13;PREP-ARG]
15 oggi (OGGI ADV TIME) [17;ADVB-RMOD-TIME]
16 si (SI PRON REFL-IMPERS ALLVAL ALLVAL 3 LSUBJ+LOBJ+LIOBJ CLITIC) [17;VERB-SUBJ/VERB-SUBJ+IMPERS]
17 vedono (VEDERE VERB MAIN IND PRES TRANS 3 PL) [12;VERB-RMOD+RELCL]
18 i (IL ART DEF M PL) [17;VERB-OBJ]
19 risultati (RISULTATO NOUN COMMON M PL RISULTARE INTRANS) [18;DET+DEF-ARG]
19.10 t [13f] (DI PREP MONO) [19;NOUN-SUBJ]

This seems the only case in which a trace is inserted to allow for the genitive DI CUI to be linked approriately as complement of RISULTATI. However here again in order to get the antecedent of CUI, which is the nominal head DIALOGO, one has to search the verb.

- DEL CUI
29 lo (IL ART DEF M SING) [23;APPOSITION]
30 studente (STUDENTE NOUN COMMON M SING) [29;DET+DEF-ARG]
31 di (DI PREP MONO) [30;PREP-RMOD-LOC+ORIGIN]

32 Ancona (ANCONA NOUN PROPER F £CITY) [31;PREP-ARG]
33 scomparso (SCOMPARIRE VERB MAIN PARTICIPLE PAST INTRANS SING M) [30;VERB-RMOD+RELCL+REDUC]
40 e (E CONJ COORD COORD) [33;COORD+BASE]
41 del (DI PREP MONO) [46;VERB-INDCOMPL-THEME]
41.1 del (IL ART DEF M SING) [41;PREP-ARG]
42 cui (CUI PRON RELAT LIOBJ+OBL) [43;PRON-RMOD]
43 caso (CASO NOUN COMMON M SING) [41.1;DET+DEF-ARG]
44 si (SI PRON REFL-IMPERS ALLVAL ALLVAL 3 LSUBJ+LOBJ+LIOBJ CLITIC) [46;EMPTYCOMPL]
45 era (ESSERE VERB AUX IND IMPERF INTRANS 3 SING) [46;AUX+TENSE]
46 occupata (OCCUPARE VERB MAIN PARTICIPLE PAST TRANS SING F) [40;COORD2ND+BASE]

Recovering the antecedent in this case requires climbing the coordinate structure, then from the reduced relative SCOMPARSO finding the nominal head STUDENTE. However this seems to be identical to the previous example where RISULTATI was lacking its complement: CASO perhaps should have been followed by a trace that identified its complement clearly, in this way the genitive would have been explicitated.

- CON CUI
1 La (IL ART DEF F SING) [6;VERB-SUBJ]
2 bella (BELLO ADJ QUALIF F SING) [3;ADJC+QUALIF-RMOD]
3 cantante (CANTANTE NOUN COMMON ALLVAL SING) [1;DET+DEF-ARG]
4 di (DI PREP MONO) [3;PREP-RMOD-LOC+ORIGIN]
5 Filadelfia (FILADELFIA NOUN PROPER F £CITY) [4;PREP-ARG]
6 sara` (ESSERE VERB MAIN IND FUT INTRANS 3 SING) [0;TOP-VERB]
7 ospite (OSPITE NOUN COMMON ALLVAL SING) [6;VERB-PREDCOMPL+SUBJ]
8 di (DI PREP MONO) [7;PREP-RMOD]
9 Massimo (MASSIMO NOUN PROPER M £NAME) [8;PREP-ARG]
10 Ranieri (|Ranieri| NOUN PROPER) [9;CONTIN+DENOM]
11 ((#\(PUNCT) [9;OPEN+PARENTHETICAL]
12 con (CON PREP MONO) [14;PREP-RMOD-TOGETHER]
13 cui (CUI PRON RELAT LIOBJ+OBL) [12;PREP-ARG]
14 cantera` (CANTARE VERB MAIN IND FUT TRANS 3 SING) [9;VERB-RMOD+RELCL]
15 in (IN PREP MONO) [14;PREP-RMOD-MEANSMANNER]
16 duetto (DUETTO NOUN COMMON M SING) [15;PREP-ARG]

In this case, the structure is wrongly built, because the relative verb CANTERA' is left without a SUBJect, empty in this case and bound to CANTANTE: the explicitate sentence should have been something like,
"[la cantante] canterà con cui [Massimo Ranieri]"

- DA CUI
1 Il (IL ART DEF M SING) [12;VERB-SUBJ]
2 punto (PUNTO_DI_VISTA NOUN COMMON M SING LOCUTION) [1;DET+DEF-ARG]
3 di (PUNTO_DI_VISTA NOUN COMMON LOCUTION) [2;CONTIN+LOCUT]
4 vista (PUNTO_DI_VISTA NOUN COMMON LOCUTION) [3;CONTIN+LOCUT]
5 da (DA PREP MONO) [7;VISITOR]
6 cui (CUI PRON RELAT LIOBJ+OBL) [5;PREP-ARG]
7 volevo (VOLERE VERB MOD IND IMPERF TRANS 1 SING) [2;VERB-RMOD+RELCL]
7.10 t [] (DEITT-T PRON PERS ALLVAL SING 1) [7;VERB-SUBJ]
8 raccontare (RACCONTARE VERB MAIN INFINITE PRES TRANS) [7;VERB+MODAL-INDCOMPL]
8.10 t [7.10p] (DEITT-T PRON PERS ALLVAL SING 1) [8;VERB-SUBJ]
8.11 t [5f] (DA PREP MONO) [8;PREP-RMOD-PERSPECT]
9 la (IL ART DEF F SING) [8;VERB-OBJ]
10 storia (STORIA NOUN COMMON F SING) [9;DET+DEF-ARG]
11 non (NON ADV NEG) [12;ADVB-RMOD-NEG]
12 era (ESSERE VERB MAIN IND IMPERF INTRANS 3 SING) [0;TOP-VERB]

In this structure, the trace reappears again to bind the relative DA CUI to RACCONTARE. However, here again in order to recover the binder nominal head PUNTO one has to produce more search steps from VOLEVO the modal inflected verb, then up to PUNTO.

The same remarks can be made if we look at TALN, where the relation intervening between the relative oblique pronoun and its nominal head binder is not available and must be recovered indirectly from the verb:

- IN CUI

17	attimo	attimoS	S	num=s|gen=m	16	prep	_	_	ARG	
18	in	in	E	E	_	22	comp	_	_	RMOD
19	cui	cui	P	PR	_	18	prep	LIOBJ+OBL	_	ARG
20	esso	esso	P	PE	num=s|per=3|gen=m	22	subj	LSUBJ+LOBJ	_	SUBJ
21	può	potereV	VM	num=s|per=3|mod=i|ten=p	22	modal	INTRANS	_	RMOD+RELCL	
22	eternar-	eternare	V	V	mod=f	17	mod_rel	TRANS	_	INDCOMPL

1	Forma	forma S	S	num=s|gen=f	0	ROOT	_	_	
2	in	in	E	E	_	5	comp	_	_
3	cui	cui	P	PR	num=n|gen=n	2	prep	_	_
4	si	si	P	PC	num=n|gen=n	5	clit	_	_
5	presenta	presentare	V	V	num=s|per=3|mod=i|ten=p	1	mod_rel	_	_
6	in	in	E	E	_	5	comp	_	_
7	natura	natura	S	S	num=s|gen=f	6	prep	_	_
8	.	.	F	FS	_	1	punc	_	_

- DURANTE I QUALI

46	anni	anno S	S	num=p|gen=m	45	prep	_	_	ARG	
47	1435	0	N	N	_	46	mod	_	_	APPOSITION
48	-	-	F	FF	_	47	con	_	_	SEPARATOR
49	1440	0	N	N	_	47	conj	_	_	ARG
50	,	,	F	FF	_	51	punc	_	_	SEPARATOR
51	durante	durante	E	E	_	54	comp_temp	TIME	_	RMOD
52	i	il	R	RD	num=p|gen=m	53	det	_	_	ARG
53	quali	quale	P	PR	num=p|per=3|gen=n	51	prep	LSUBJ+LOBJ+OBL	_	ARG
54	era	essereV	V	num=s|per=3|mod=i|ten=i	46	mod_rel	INTRANS	_	RMOD	
55	ancora	ancora	B	B	_	54	mod_temp	TIME	_	RMOD
56	collaboratore	collaboratore	S	S	num=s|gen=m	54	pred	_	_	PREDCOMPL+SUBJ

1.6.1 Lexical Relative Pronouns in other treebanks

TALN and TUT encode relations in the same way in which PTB and other treebanks do, as shown below:

- ABOUT WHICH

27	and	_	CC	_	_	17	COORD	_	_	
28	should	_	MD	_	_	27	CONJ	_	_	
29	honor	_	VB	_	_	28	VC	_	_	
30	local	_	JJ	_	_	31	NMOD	_	_	
31	convictions	_	NNS	_	_	29	OBJ	_	_	
32	about	_	IN	_	_	31	NMOD	_	_	
33	which	_	WDT	_	_	34	NMOD	_	_	
34	tasks	_	NNS	_	_	36	SBJ	_	_	
35	most	_	RBS	_	_	36	ADV	_	_	
36	need	_	VBP	_	_	32	PMOD	_	_	
37	doing	_	VBG	_	_	36	OBJ	_	_	

- AT WHICH

26	the	_	DT	_	_	27	NMOD	_	_
27	level	_	NN	_	_	25	PMOD	_	_
28	at	_	IN	_	_	36	LOC	_	_
29	which	_	WDT	_	_	28	PMOD	_	_
30	the	_	DT	_	_	32	NMOD	_	_
31	stock	_	NN	_	_	32	NMOD	_	_
32	market		_	NN	_	_	34	SBJ	_	_
33	itself	_	PRP	_	_	32	APPO	_	_
34	would	_	MD	_	_	27	NMOD	_	_
35	have	_	VB	_	_	34	VC	_	_
36	closed	_	VBN	_	_	35	VC	_	_

- ON WHICH

26	products	_	NNS	_	_	23	PMOD	_	_
27	,	_	,	_	_	26	P	_	_
28	on	_	IN	_	_	31	ADV	_	_
29	which	_	WDT	_	_	28	PMOD	_	_
30	deliveries		_	NNS	_	_	31	SBJ	_	_
31	began	_	VBD	_	_	26	NMOD	_	_

- IN WHICH

36	climate		_	NN	_	_	34	OBJ	_	_
37	of	_	IN	_	_	36	NMOD	_	_
38	trust	_	NN	_	_	37	PMOD	_	_
39	and	_	CC	_	_	38	COORD	_	_
40	peace	_	NN	_	_	39	CONJ	_	_
41	in	_	IN	_	_	46	LOC	_	_
42	which	_	WDT	_	_	41	PMOD	_	_
43	his	_	PRP$	_	_	44	NMOD	_	_
44	government		_	NN	_	_	45	SBJ	_	_
45	can	_	MD	_	_	36	NMOD	_	_
46	begin	_	VB	_	_	45	VC	_	_

In all these examples, the preposition is linked to the verb of the relative clause, WHICH is linked to the preposition, and the verb of the relative is in turn linked to the head noun of the relative pronoun. To show the difference intervening between WHICH and WHOSE – that is somewhat comparable to CUI - we will now present some examples with the genitive relative pronoun WHOSE – that is always accompanied by at least a noun or a noun with modifiers -, which resembles in some way the sequence (Art) CUI, but without preceding articles.

- WHOSE

4	those	_	DT	_	_	3	PMOD	_	_
5	of	_	IN	_	_	4	NMOD	_	_
6	us	_	PRP	_	_	5	PMOD	_	_
7	whose	_	WP$	_	_	8	NMOD	_	_
8	views	_	NNS	_	_	9	SBJ	_	_
9	are	_	VBP	_	_	4	NMOD	_	_

- WHOSE

6	Nunn-McCurdy		_	NNP	_	_	5	PMOD	_	_
7	,	_	,	_	_	6	P	_	_
8	whose	_	WP$	_	_	9	NMOD	_	_
9	benefits		_	NNS	_	_	12	SBJ	_	_
10	to	_	TO	_	_	9	NMOD	_	_
11	enrollees		_	NNS	_	_	10	PMOD	_	_
12	are	_	VBP	_	_	6	NMOD	_	_

- WHOSE

2	Sony	_	NNP	_	_	18	SBJ	_	_
3	,	_	,	_	_	2	P	_	_
4	whose	_	WP$	_	_	8	NMOD	_	_
5	innovative	_	JJ	_	_	8	NMOD	_	_
6	,	_	,	_	_	8	P	_	_
7	premium-priced	_	JJ	_	_	8	NMOD	_	_
8	products	_	NNS	_	_	9	SBJ	_	_
9	are	_	VBP	_	_	2	NMOD	_	_

- WHOSE

3	Kollmorgen	_	NNP	_	_	2	OBJ	_	_
4	,	_	,	_	_	3	P	_	_
5	whose	_	WP$	_	_	6	NMOD	_	_
6	agreement	_	NN	_	_	17	SBJ	_	_
7	to	_	TO	_	_	6	NMOD	_	_
8	be	_	VB	_	_	7	IM	_	_
9	acquired	_	VBN	_	_	8	VC	_	_
10	for	_	IN	_	_	9	ADV	_	_
11	$	_	$	_	_	10	PMOD	_	_
12	25	_	CD	_	_	11	NMOD	_	_
13	a	_	DT	_	_	14	NMOD	_	_
14	share	_	NN	_	_	11	ADV	_	_
15	by	_	IN	_	_	9	LGS	_	_
16	Vernitron	_	NNP	_	_	15	PMOD	_	_
17	collapsed	_	VBD	_	_	3	NMOD	_	_

- WHOSE

10	area	_	NN	_	_	7	PMOD	_	_
11	whose	_	WP$	_	_	13	NMOD	_	_
12	economic	_	JJ	_	_	13	NMOD	_	_
13	health	_	NN	_	_	14	SBJ	_	_
14	is	_	VBZ	_	_	10	NMOD	_	_

As can be noticed, WHOSE requires a totally different treatment from WHICH: it is linked to the head noun it modifies – it specifies its semantic content - and this noun is then linked to the verb of the relative clause. The verb of the relative is then linked to the head noun but this noun does not modify the verb, in fact it does not have any relation with it being a modifier of one of the arguments of the relative clause. I indicate here below in brackets the position of WHOSE and of its lexical substitute:

innovative products [of Sony] are / Sony [whose] innovative products are
agreement [of Kollmorgen] collapsed / Kollmorgen [whose] agreement collapsed
economic health [of area] is / area [whose] economic health is

For this reason, I don't see why the verb of the relative should be linked to the head noun of the relative pronoun, rather than directly to the relative pronoun, and the latter in turn linked to the head noun.
In the case of WHICH, the relations are different:

deliveries began on [products] / products on [which] deliveries began
his government can begin in [climate] / climate in [which] his government can begin

Relative pronoun WHICH is governed by the preposition which is heading an adjunct or argument of the verb of the relative itself. Very much the same would happen with simple relative pronouns which are arguments of the verb of the relative. So eventually, the treatment of WHOSE/CUI seems inadequate in particular in view of its mapping onto a semantic predicate-argument structure. To see in more depth the ways in which the mapping of oblique/genitive relative pronouns may take place

we look into PARC-700 relevant portions to check how the LFG has decided to encode it. We look at five different examples and we see that the treatment is definitely organized on the basis of the presence of a NULL element, pro. What is important to stress here is the fact that WHOSE expresses a possessive genitive relation with its local head that it modifies, and that this relation is represented by an abstract *"pro"* linked to WHOSE and also linked to the verb of the relative: RESPECT, SING, BE, DETERMINE, KEEP.

sentence_form(But Mr. Davis\, whose views are widely respected by money managers\, says he expects no 1987-style crash.)
  adjunct(Mr. Davis~1, respect~18)
  adjunct(respect~18, widely~24)
  adjunct_type(respect~18, relative)
  obl_ag(respect~18, manager~19)
  pron_rel(respect~18, pro~22)
  subj(respect~18, view~20)
  topic_rel(respect~18, view~20)
  mod(manager~19, money~28)
  pcase(manager~19, by)
  poss(view~20, pro~22)
  pron_form(pro~22, whose)
  pron_type(pro~22, relative)

sentence_form(One of Italy's favorite shows\, ``Fantastico\,'' a tepid variety show\, is so popular that viewers clamored to buy a chocolate product\, ``Cacao Fantastico\,'' whose praises were sung each week by dancing showgirls -- even though the product didn't exist.)
  adjunct_type(Cacao Fantastico~61, parenthetical)
  poss(praise~62, pro~64)
  pron_form(pro~64, whose)
  pron_rel(sing~54, pro~64)
  pron_type(pro~64, relative)
  subj(sing~54, praise~62)
  obl_ag(sing~54, showgirl~43)
  adjunct_type(sing~54, relative)
  topic_rel(sing~54, praise~62)
  adjunct(product~92, Cacao Fantastico~61)

sentence_form(And it has remained there\, as evidenced by its reappearance in a 1972 CBS sitcom called ``Bridget Loves Bernie\,'' whose sole distinction was that it led to the real-life marriage of Meredith Baxter and David Birney.)
  subj(call~18, pro~26)
  subj(Bridget Loves Bernie~25, pro~26)
  xcomp(call~18, Bridget Loves Bernie~25)
  adjunct_type(be~19, relative)
  subj(be~19, distinction~31)
  topic_rel(be~19, distinction~31)
  pron_rel(be~19, pro~32)
  pron_form(pro~32, whose)
  pron_type(pro~32, relative)
  poss(distinction~31, pro~32)

sentence_form(The White House Office of Management and Budget\, whose calculations determine whether the Gramm-Rudman targets are met\, estimated that the House-passed deficit-reduction measure would cut the fiscal 1990 shortfall by $6.2 billion\, almost half of the Congressional Budget Office's estimate of $11.0 billion.)
  adjunct(Office of Management and Budget~4, determine~26)
  adjunct_type(determine~26, relative)
  pron_rel(determine~26, pro~33)
  subj(determine~26, calculation~31)
  poss(calculation~31, pro~33)

pron_form(pro~33, whose)
    pron_type(pro~33, relative)

sentence_form(Her friend Susan\, whose parents kept reminding her she was unwanted\, slept on a narrow bed wedged into her parents' bedroom\, as though she were a temporary
    adjunct(Susan~1, keep~45)
    mod(Susan~1, friend~61)
    adjunct_type(keep~45, relative)
    pron_rel(keep~45, pro~48)
    subj(keep~45, parent~49)
    topic_rel(keep~45, parent~49)
    pron_form(pro~48, whose)
    pron_type(pro~48, relative)
    poss(parent~49, pro~48)

For these reasons, the role of CUI in particular has been given a lot of attention in the deep version of VIT, that we comment here below.

1.6.2 *CUI in VIT*

There at least four different typologies of structure accompanying CUI oblique relative pronoun, and that we have found in VIT:
1. argument/adjunct of relative verb
          - it directly modifies the main verb of the relative clause
2. adjunct modifier of argument of relative verb
          - it modifies an argument of the verb of relative clause
3. adjunct modifier of a noun
4. adjunct modifier of the internal nominal head

1.6.2.1 **ARGUMENT/ADJUNCT OF RELATIVE VERB**

All of the following examples show the variety of cases in which CUI can act as an adjunct but also as an argument with different semantic roles:

- IN CUI (Locative)
38 dell di partd(preposition_di_plus_article) spd num=s|per=fm 49 mod det
38.1 l il art sn num=s|per=fm 49 det def
39 ambiente ambiente n(noun) sn num=s|gen=m 38 pobj com
40 socio_economico socio_economico ag(adjective) sa num=s 39 mod nil
41 in in p(preposition) sp - 39 adj nil
42 cui cui relob(relative_oblique) sn [] 41 binder rel_obl
43 sono essere ause(auxiliary_essere_tensed) ibar punt 44 ibar aux
44 inserite inserire vppt(verb_trans_past_participle) ibar punt 39 ibar refl_in/into_hole
44.11 prep_relob in_ambiente prep_relob(prepositional_rel_oblique) sp num=s|gen=m ant=41_42 bindee com

- IN CUI (Part)
7 nella in part(preposition_plus_article) sp num=s|gen=f 6 obl det
7.1 la il art sn num=s|gen=f 6 det def
8 norma norma n(noun) sn num=s|gen=f 7 pobj com
9 in in p(preposition) sp - 8 adj nil
10 cui cui relob(relative_oblique) sn [] 9 binder rel_obl
11 si si clit(clitic_pronoun) ibar per=3|gen=m|num=sp 12 ibar nom
12 stabilisce stabilire vt(verb_trans_tensed) ibar punt 8 ibar refl/exten
12.10 pro si pro(little_pro) sn per=3|gen=m|num=sp 11 s_impers-agent nom
12.11 prep_relob in_norma prep_relob(prepositional_rel_oblique) sp num=s|gen=f ant=9_10 bindee com

- SU CUI (Argument)
9 proposte proposta n(noun) sn num=p|gen=f 8 pobj com

10 su su p(preposition) sp - 9 adj nil
11 cui cui relob(relative_oblique) sn [] 10 binder rel_obl
12 lavorare lavorare viin(verb_intrans_infinitival) sv2 punt 9 adj intr
12.11 prep_relob su_proposta prep_relob(prepositional_rel_oblique) sp num=p|gen=f ant=10_11 bindee com

- SECONDO CUI (Attribution)
11 confraternita confraternita n(noun) sn num=s|gen=f 10 pobj-recipnt com
12 dei di partd(preposition_di_plus_article) spd num=p|gen=m 11 mod det
12.1 i il art sn num=p|gen=m 11 det def
13 piagnoni piagnone n(noun) sn num=p|gen=m 12 pobj com
14 , , punt(sentence_internal) sn punt 11 sn nil
15 secondo secondo p(preposition) sp - 11 adj nil
16 cui cui relob(relative_oblique) sn [] 15 binder rel_obl
17 manca mancare vin(verb_intrans_tensed) ibar punt 11 ibar tr/not_exten
17.11 prep_relob secondo_confraternita prep_relob(prepositional_rel_oblique) sp num=s|gen=f ant=15_16 bindee com
18 personale personale n(noun) sn num=s|gen=m 17 subj-agent com

- IN CUI (Manner)
5 al a part(preposition_plus_article) sp num=s|gen=m 11 iobj det
5.1 l il art sn num=s|per=fm 11 det def
6 modo modo n(noun) sn num=s|gen=m 5 pobj com
7 in in p(preposition) sp - 6 adj nil
8 cui cui relob(relative_oblique) sn [] 7 binder rel_obl
9 è essere ause(auxiliary_essere_tensed) ibar punt 10 ibar aux
10 interrogato interrogare vppt(verb_trans_past_participle) ibar punt 6 ibar tr/inform
10.10 pro pro pro(little_pro) sn num=s|per=3 ant=1 s_impl-address nil
10.11 prep_relob in_modo prep_relob(prepositional_rel_oblique) sp num=s|gen=m ant=7_8 bindee com

- CON CUI (Instrument)
35 le il art(article) sn num=p|gen=f 36 sn def
36 argomentazioni argomentazione n(noun) sn num=p|gen=f 34 pobj com
37 politiche politico ag(adjective) sa num=p|gen=f 36 mod nil
38 con con p(preposition) sp - 36 adj nil
39 cui cui relob(relative_oblique) sp [] 38 sp rel_obl
40 ha avere ausa(auxiliary_avere_tensed) ibar punt 41 ibar aux
41 motivato motivare vppt(verb_trans_past_participle) ibar punt 36 ibar tr/exten
41.10 pro pro pro(little_pro) sn num=p|per=3 ant=27 s_impl-theme_unaff nil
41.11 prep_relob con_argomentazioni prep_relob(prepositional_rel_oblique) sp num=s|gen=m ant=38_39 bindee com

- PER CUI (Motivation)
4 una uno art(article) sn num=s|gen=f 5 sn ind
5 condizione condizione nf(noun_factive) sn num=s|gen=f 3 s_top-agent com
6 per per p(preposition) sp - 5 adj nil
7 cui cui relob(relative_oblique) sn [] 6 binder rel_obl
8 manca mancare vin(verb_intrans_tensed) ibar punt 5 ibar tr/not_exten
8.11 prep_relob per_condizione prep_relob(prepositional_rel_oblique) sp num=s|gen=f ant=6_7 bindee com

- DI CUI (Theme)
2 cose cosa n(noun) sn num=p|gen=f 1 ncomp com
3 più più in(intensifier) sa [] 4 sa q
4 importanti importante ag(adjective) sa num=p|per=fm 2 mod nil
5 di di pd(preposition_di) spd - 2 adj nil
6 cui cui relob(relative_oblique) sn [] 5 binder rel_obl
7 occuparmi occupare vcl(verb_with_enclitic) sv2 punt 2 adj tr
7.11 prep_relob di_cosa prep_relob(prepositional_rel_oblique) sp num=p|gen=f ant=5_6 bindee com

(here DI CUI is ARGUMENT of main verb of the relative clause, OCCUPARMI)

1.6.2.2 ADJUNCT MODIFIER OF ARGUMENT OF RELATIVE VERB

The samples in this subsection are all referred to the special case of copulative constructions as relative clauses, in which the oblique relative is a modifier of the predicate, usually an adjective.

- AL QUALE - IN CUI
31 il il art(article) sn num=s|gen=m 32 sn def
32 costo costo n(noun) sn num=s|gen=m 39 ncomp com
33 al a part(preposition_plus_article) sp num=s|gen=m 32 mod det
33.1 l il art sn num=s|per=fm 32 det def
34 quale quale rel(relative) f2 num=s|gen=m 33 binder nil
35 la il art(article) sn num=s|gen=f 36 sn def
36 sostituzione sostituzione n(noun) sn num=s|gen=f 37 subj-tema_bound com
37 è essere vc(verb_copulative) ibar punt 34 ibar cop/esistenza
38 possibile possibile ag(adjective) sa num=s|per=fm 37 acomp nil
39 e e cong(conjunction) coord [] 30 coord sum
40 la il art(article) sn num=s|gen=f 41 sn def
41 misura misura n(noun) sn num=s|gen=f 39 ncomp com
42 in in p(preposition) sp - 41 adj nil
43 cui cui relob(relative_oblique) sn [] 42 binder rel_obl
44 è essere vc(verb_copulative) ibar punt 41 ibar cop/esistenza
44.10 pro sostituzione pro(little_pro) sn num=s|per=3 ant=36 s_impl-tema_bound nil
45 fattibile fattibile ag(adjective) sa num=s|per=fm 44 acomp nil
45.11 prep_relob in_misura prep_relob(prepositional_rel_oblique) sp num=s|gen=f ant=42_43 bindee com

Even though "la_misura_in_cui" may sometimes be used as adverbial locution, in this case it is just the SUBJect of BE and consequently CUI is head of relative clause that modifies FATTIBILE - in ... misura. The same applies to the example below, where the relative pronoun is a modifer of RESPONSABILE.

- DI CUI
36 sulle su part(preposition_plus_article) sp num=p|gen=f 32 pcomp det
36.1 le il art sn num=p|gen=f 32 det def
37 due due num(numeral) sn [] 38 sn card
38 branche branca n(noun) sn num=p|gen=f 36 pobj com
39 operative operativo ag(adjective) sa num=p|gen=f 38 mod nil
40 , , punt(sentence_internal) sn punt 38 sn nil
41 di di pd(preposition_di) spd - 38 adj nil
42 cui cui relob(relative_oblique) sn [] 41 binder rel_obl
43 pure pure cong(conjunction) f2 [] 38 cong sum
44 è essere vc(verb_copulative) ibar punt 38 ibar cop/esistenza
45 nominalmente nominalmente avv(adverb) savv [] 46 adjm mn
46 responsabile responsabile ag(adjective) sa num=s|per=fm 44 acomp nil
46.11 prep_relob di_branca prep_relob(prepositional_rel_oblique) sp num=p|gen=f ant=41_42 bindee com

1.6.2.2.1 ADJUNCT MODIFIER OF ARGUMENT OF RELATIVE VERB

In this subsection, the oblique relative is a modifer of a nominal predicate, BISOGNO and further on CANDIDATO.

- DI CUI
21 fondi fondo n(noun) sn num=p|gen=m 19 obj com
22 di di pd(preposition_di) spd - 21 adj nil
23 cui cui relob(relative_oblique) sn [] 22 binder rel_obl
24 abbiamo avere vc(verb_copulative) ibar nil 21 ibar cop/stato
24.10 pro pro pro(little_pro) sn nil ant=7 s_impl-esperiente impL1p
24.11 prep_relob di_fondo prep_relob(prepositional_rel_oblique) sp num=p|gen=m ant=22_23 bindee com
25 bisogno bisogno n(noun) sn num=s|gen=m 24 ncomp com

- ALLA CUI

8 commissione commissione n(noun) sn num=s|gen=f 6 pobj com
9 esteri estero ag(adjective) sa num=p|gen=m 8 mod nil
10 alla a part(preposition_plus_article) sp num=s|gen=f 8 adj det
10.1 la il art sn num=s|gen=f 8 det def
11 cui cui relob(relative_oblique) sp [] 10 sp rel_obl
12 presidenza presidenza n(noun) sn num=s|gen=f 10 pobj com
13 è essere vc(verb_copulative) ibar punt 8 ibar cop/esistenza
14 candidato candidato n(noun) sn num=s|gen=m 13 ncomp com
14.11 prep_relob alla_commissione prep_relob(prepositional_rel_oblique) sp num=s|gen=m ant=10_11 bindee com

1.6.2.2.2 ADJUNCT MODIFIER OF ARGUMENT OF RELATIVE VERB

This example shows a case of oblique relative which is a modifier of an argument embedded in an infinitival complement of a process verb CONTINUARE, very much like the example we saw from PTB in the section above:

- DI CUI

0 Una uno art(article) sn num=s|gen=f 1 sn ind
1 strategia strategia n(noun) sn num=s|gen=f 13 sn com
2 di di pd(preposition_di) spd - 1 adj nil
3 cui cui relob(relative_oblique) sn [] 2 binder rel_obl
4 tutti tutto qc(quantifier_collective) sq num=p|gen=m 6 sq nil
5 i il art(article) sn num=p|gen=m 6 sn def
6 ministri ministro n(noun) sn num=p|gen=m 8 subj-exper com
7 interessati interessato ppas(past_participle_absolute) sa num=p|gen=m 6 mod nil
8 continuano continuare vt(verb_trans_tensed) cl(main) punt - ibar raisn/process
9 a a pt(verbal_participle) sv2 - 10 sv2 nil
10 sottolineare sottolineare vit(verb_trans_infinitive) sv2 punt 8 vcomp tr
10.10 pPro pPro pPro(big_pro) sn nil ant='6' s_impl-causer ministro
11 la il art(article) sn num=s|gen=f 13 sn def
12 < < par(parenthetical) sn - 13 sn nil
13 collegialità collegialità n(noun) sn num=f 10 obj invar
13.11 prep_relob di_strategia prep_relob(prepositional_rel_oblique) sp num=s|gen=f ant=1_2 bindee com

In this example, we want to say that the relative pronoun modifies COLLEGIALITA', and the semantics should compose the following pseudo-structure:

una strategia [di cui] tutti i ministri interessati continuano a sottolineare la collegialità [t] → la collegialità [della strategia]

1.6.2.3 ADJUNCT MODIFIER OF AN ELLIPSED NOMINAL HEAD

Not all cases of relative pronouns are connected to a fully lexicalized relative clause: there are cases in which the clause is unexpressed – as would happen with reduced relatives – but also ellipsed as shown in the following examples:

- TRA CUI

12 nomi nome n(noun) sn num=p|gen=m 11 obj-theme_unaff com
13 di di pd(preposition_di) spd - 12 mod nil
14 rilievo rilievo n(noun) sn num=s|gen=m 13 pobj com
15 , , punt(sentence_internal) sn punt 12 sn nil
16 tra tra p(preposition) sp - 12 adj nil
17 cui cui relob(relative_oblique) sn [] 16 binder rel_obl
18 l il art(article) sn num=s|gen=m 20 sn def

19 ex ex ag(adjective) sa num=f|gen=m 20 mod invar
20 ministro ministro n(noun) sn num=s|gen=m 17 subj com
20.11 prep_relob tra_nome prep_relob(prepositional_rel_oblique) sp num=p|gen=m ant=16_17 bindee com
21 della di partd(preposition_di_plus_article) spd num=s|gen=f 20 mod det
21.1 la il art sn num=s|gen=f 20 det def
22 difesa difesa n(noun) sn num=s|gen=f 21 pobj com

- TRA CUI
5 variabili variabile n(noun) sn num=p|gen=f 4 pobj com
6 strutturali strutturale ag(adjective) sa num=p|per=fm 5 mod nil
7 tra tra p(preposition) sp - 5 adj nil
8 cui cui relob(relative_oblique) sn [] 7 binder rel_obl
9 le il art(article) sn num=p|gen=f 10 sn def
10 istituzioni istituzione n(noun) sn num=p|gen=f 11 sn com
…
32 che che rel(relative) f2 - 11 binder nil
33 si si clit(clitic_pronoun) ibar per=3|gen=m|num=sp 34 ibar acc
34 modificano modificare vt(verb_trans_tensed) ibar punt 32 ibar refl_in/exten
34.11 rel_pro istituzioni rel_pro(relative_pronoun) bindee num=p|per=3 ant=11 subj-theme_aff nil

The specialty of this structure is the fact that it is a fragment which however has a main nominal head: to complete the semantics it could be enriched by the presence of a "dummy BE" verb, or perhaps a dummy THERE_BE, so that the head noun MINISTRO is computed as subject of predication. The oblique relative modifies directly the subject nominal MINISTRO or indirectly, in case of presence of dummy BE, through the predication:

→ l'ex ministro … *E' tra i nomi*

The same applies to the example below:

- TRA CUI
29 collaboratori collaboratore n(noun) sn num=p|gen=m 28 pobj com
…
37 tra tra p(preposition) sp - 29 adj nil
38 cui cui relob(relative_oblique) sn [] 37 binder rel_obl
39 il il art(article) sn num=s|gen=m 40 sn def
40 capo capo n(noun) sn num=s|gen=m 38 sn com
40.11 prep_relob tra_collaboratori prep_relob(prepositional_rel_oblique) sp num=p|gen=m ant=37_38 bindee com
41 della di partd(preposition_di_plus_article) spd num=s|gen=f 40 mod det
41.1 la il art sn num=s|gen=f 40 det def
42 polizia polizia n(noun) sn num=s|gen=f 41 pobj com

1.6.2.4 **ADJUNCT OF THE SUBJECT/OBJECT NOMINAL HEAD OF THE RELATIVE**

Eventually, we also found cases in which the relative CUI modifies the SUBJect head noun it depends on, as is the case in the example below:

- LA CUI
0 Non non neg(negation) ir_infl - 1 neg nil
1 sarà essere vcir(verb_copulative_mood_irrealis) cl(main) punt - ir_infl cop/esistenza
2 presente presente ag(adjective) sa num=s|per=fm 1 acomp nil
3 , , punt(sentence_internal) compc punt 1 compc nil
4 invece invece congf(conjunction_sentential) compc [] 1 cong av
5 , , punt(sentence_internal) compc punt 1 compc nil
6 l il art(article) sn num=s|gen=m 7 sn def
7 uomo uomo n(noun) sn num=m 1 s_top-tema_bound invar
…

19 la il art(article) sn num=s|gen=f 21 sn def
20 cui cui relob(relative_oblique) sn [] 7 sn rel_obl
21 posizione posizione n(noun) sn num=s|gen=f 24 subj-theme_unaff com
21.11 prep_relob di_uomo prep_relob(prepositional_rel_oblique) sp num=s|gen=m ant=7 bindee com
22 è essere ause(auxiliary_essere_tensed) ibar punt 24 ibar aux
23 stata essere ausep(auxiliary_essere_past_participle) ibar punt 24 ibar aux
24 stralciata stralciare vppt(verb_trans_past_participle) ibar punt 21 ibar tr/possess
…

In this structure the oblique is only active locally even though the main verb would occur in the following portion of the sentence, it does not contribute to the following relative clause structure, neither as argument nor as adjunct nor as modifiers of some argument.

1.6.3 **OBLIQUE RELATIVE in other treebanks**

We already saw in the previous section the treatment of WHOSE in PTB. We will now document the treatment of oblique relatives in the Portugues and Catalan treebank. As is clearly shown, the relative pronoun is linked to the local nominal head which is linked to the relative verb, the relative verb in turn is linked to the governing nominal head of the relative, which is what we already saw in previous examples and treebanks,

CUJA

1	A	o	art	art	<artd>|F|S	2	>N	_	_
2	Conferência_Haitiana_de_Religiosos	Conferência_Haitiana_de_Religiosos	prop	prop	F|S	11			
	SUBJ	_	_						
3	,	,	punc	punc	_	2	PUNC	_	_
4	cuja	cujo	pron	pron-det	<rel>|F|S	5	>N	_	_
5	direcção	direcção	n	n	F|S	6	SUBJ	_	_
6	é	ser	v	v-fin	PR|3S|IND	2	N<PRED	_	_
7	tida	ter	v	v-pcp	F|S	6	MV	_	_
8	por	por	prp	prp	_	7	SC	_	_
9	moderada	moderar	v	v-pcp	F|S	8	P<	_	_
10	,	,	punc	punc	_	2	PUNC	_	_
11	vem	vir	v	v-fin	PR|3S|IND	0	STA	_	_
12	respondendoresponder		v	v-ger	_	11	MV	_	_

10	a	o	art	art	<-sam>|<artd>|S	11	>N	_	_
11	Igreja	Igreja	prop	prop	F|S	9	P<	_	_
12	,	,	punc	punc	_	11	PUNC	_	_
13	cuja	cujo	pron	pron-det	<rel>|F|S	14	>N	_	_
14	hierarquia	hierarquia	n	n	F|S	24	SUBJ	_	_
15	,	,	punc	punc	_	14	PUNC	_	_
…									
23	,	,	punc	punc	_	14	PUNC	_	_
24	prefere	preferir	v	v-fin	PR|3S|IND	11	N<PRED	_	_

CUJAS

17	o	o	art	art	<artd>|M|S	18	>N	_	_
18	SIED	SIED	prop	prop	M|S	16	ACC	_	_
19	,	,	punc	punc	_	18	PUNC	_	_
20	cujas	cujo	pron	pron-det	<rel>|F|P	21	>N	_	_
21	competências	competência	n	n	F|P	22	SUBJ	_	_
22	foram	ser	v	v-fin	PS|3P|IND	18	N<PRED	_	_
23	,	,	punc	punc	_	22	PUNC	_	_
24	entretanto	entretanto	adv	adv	<kc>	26	ADVL	_	_
25	,	,	punc	punc	_	22	PUNC	_	_
26	transferidas	transferir	v	v-pcp	F|P	22	MV	_	_

1.6.3.1 OBLIQUE RELATIVE in other treebanks

As an experiment I tried out a sentence which contained a pied piped oblique genitive in English, with both CONNEXOR and STANFORD parsers to see the relations they manage to encode in the output. However none of the output is able to show differences in treatment from previous examples.

```
1     John   john          @OBJ %NH N NOM SG
2     ,      ,
3     in     in            @ADVL %EH PREP
4     whose  who    attr:>5      @A> %>N <Rel> PRON WH GEN
5     house  house         @<P %NH N NOM SG
6     the    the    det:>7 @DN> %>N DET
7     accident      accident      subj:>8      @SUBJ %NH N NOM SG
8     took   take   pcomp:>3     @+FMAINV %VA V PAST
9     place  place  obj:>8       @OBJ %NH N NOM SG
10    ,      ,
11    is     be     v-ch:>12     @+FAUXV %AUX V PRES SG3
12    leaving       leave         @-FMAINV %VA ING
13    home   home   goa:>12      @ADVL %EH N NOM SG
14    now    now    tmp:>12      @ADVL %EH ADV
15    .      .
16    <s>    <s>
```

And this is the STANFORD parser output:

John in whose house the accident took place is leaving home now.

```
(ROOT
  (S
    (NP
      (NP (NNP John))
      (PP (IN in)
        (SBAR
          (WHNP (WP$ whose)
            (NP (NN house)))
          (S
            (NP (DT the) (NN accident))
            (VP (VBD took)
              (NP (NN place)))))))
    (VP (VBZ is)
      (VP (VBG leaving)
        (NP (NN home))
        (ADVP (RB now))))
    (. .)))
```

Typed dependencies

```
nsubj(leaving-10, John-1)
prep(John-1, in-2)
poss(house-4, whose-3)
dobj(took-7, house-4)
det(accident-6, the-5)
nsubj(took-7, accident-6)
pcomp(in-2, took-7)
dobj(took-7, place-8)
aux(leaving-10, is-9)
root(ROOT-0, leaving-10)
dobj(leaving-10, home-11)
advmod(leaving-10, now-12)
```

Typed dependencies, collapsed

```
nsubj(leaving-10, John-1)
poss(house-4, whose-3)
dobj(took-7, house-4)
det(accident-6, the-5)
nsubj(took-7, accident-6)
prepc_in(John-1, took-7)
dobj(took-7, place-8)
aux(leaving-10, is-9)
root(ROOT-0, leaving-10)
dobj(leaving-10, home-11)
advmod(leaving-10, now-12)
```

What is missing, then here, is the information that "the house" belongs to John, and the role of WHOSE is left unexplained.

I also tried the Italian version of the same sentence with TALN online parser called DeSR. As can be noticed, positioning the SUBJect NP "l'incidente" in postverbal position causes the system to label it as OBJect, while it correctly labels the NP as subject if in preverbal position. This could be obviated easily if the parser was endowed with a syntactico-semantic classification of verbs, limited to exceptional classes like AVVENIRE, which are impersonal verbs, disallowing semantic subject but allowing syntactic surface subjects.

ID	FORM	LEMMA	CPOS	POS	MORPH	HEAD	DEP	PH	PD
1	Gino	Gino	S	SP	_	11	subj	_	_
2	,	,	F	FF	_	1	punc	_	_
3	nella	in	E	EA	num=s|gen=f	9	comp	_	_
4	cui	cui	D	DR	num=n|gen=n	5	mod	_	_
5	casa	casa	S	S	num=s|gen=f	3	prep	_	_
6	l'	il	R	RD	num=s|gen=n	7	det	_	_
7	incidente	incidente	S	S	num=s|gen=m	9	subj	_	_
8	è	essere	V	VA	num=s|per=3|mod=i|ten=p	9	aux	_	_
9	avvenuto	avvenire	V	V	num=s|mod=p|gen=m	1	mod_rel	_	_
10	,	,	F	FF	_	1	punc	_	_
11	sta	stare	V	V	num=s|per=3|mod=i|ten=p	0	ROOT	_	_
12	partendo	partire	V	V	mod=g	11	mod	_	_
13	da	da	E	E	_	12	comp	_	_
14	casa	casa	S	S	num=s|gen=f	13	prep	_	_
15	ora	ora	B	B	_	12	mod	_	_
16	.	.	F	FS	_	11	punc	_	_

ID	FORM	LEMMA	CPOS	POS	MORPH	HEAD	DEP	PH	PD
1	Gino	Gino	S	SP	_	11	subj	_	_
2	,	,	F	FF	_	1	punc	_	_
3	nella	in	E	EA	num=s|gen=f	7	comp	_	_
4	cui	cui	D	DR	num=n|gen=n	5	mod	_	_
5	casa	casa	S	S	num=s|gen=f	3	prep	_	_
6	è	essere	V	VA	num=s|per=3|mod=i|ten=p	7	aux	_	_
7	avvenuto	avvenire	V	V	num=s|mod=p|gen=m	1	mod_rel	_	_
8	l'	il	R	RD	num=s|gen=n	9	det	_	_
9	incidente	incidente	S	S	num=s|gen=m	7	obj	_	_
10	,	,	F	FF	_	1	punc	_	_
11	sta	stare	V	V	num=s|per=3|mod=i|ten=p	0	ROOT	_	_
12	partendo	partire	V	V	mod=g	11	mod	_	_
13	da	da	E	E	_	12	comp	_	_
14	casa	casa	S	S	num=s|gen=f	13	prep	_	_
15	ora	ora	B	B	_	12	mod	_	_
16	.	.	F	FS	_	11	punc	_	_

1.7 QUESTIONS

Questions are hard to parse for statistical parsers, given their sparsity in available treebanks. We will now look at some questions and the way in which they are encoded in dependency structures. Apart from some mistakes present in the analysis, the major difficulty seems interpreting the grammatical function of the interrogative pronouns. TALN parser is sensitive to the position of the pronoun, that being anteposed to the verb is interpreted as SUBJect. Also the case of a prepositional interrogative, "A CHE" is interpreted as such.

1	Che	che	D	DQ	num=n|gen=n	2	mod	_	_
2	cosa	cosa	S	S	num=s|gen=f	3	subj	_	_
3	significa	significare	V	V	num=s|per=3|mod=i|ten=p	0	ROOT	_	_
4	quest'	quest'	V	V	num=s|per=2|mod=m|ten=p	0	ROOT	_	_

| 5 | asserzione | asserzione | S | S | num=s\|gen=f4 | | obj | _ | _ |
| 6 | ? | ? | F | FS | _ | 5 | punc | _ | _ |

1	A	a	E	E	_	0	ROOT	_	_
2	che	che	P	PR	num=n|gen=n	3	subj	_	_
3	pensa	pensare	V	V	num=s|per=3|mod=i|ten=p	1	prep	_	_
4	costui	costui	P	PD	num=s|per=3|gen=m	3	obj	_	_
5	?	?	F	FS	_	4	punc	_	_

TALN, as TUT does, encode the relation intervening between the interrogative pronoun and the verb of the relative directly by linking it to the verb,

8 decidere (DECIDERE VERB MAIN INFINITE PRES TRANS) [7;PREP-ARG]
8.10 t [6.10f] (IL ART DEF M SING) [8;VERB-SUBJ]
9 quali (QUALE ADJ INTERR ALLVAL PL) [13;VERB-OBJ]
10 misure (MISURA NOUN COMMON F PL) [9;DET+INTERR-ARG]
11 di (DI PREP MONO) [10;PREP-RMOD]
12 sicurezza (SICUREZZA NOUN COMMON F SING) [11;PREP-ARG]
13 adottare (ADOTTARE VERB MAIN INFINITE PRES TRANS) [8;VERB-OBJ]
13.10 t [8.10f] (IL ART DEF M SING) [13;VERB-SUBJ]

1 Cosa (COSA PRON INTERR ALLVAL SING LSUBJ+LOBJ) [2;VERB-OBJ]
2 faranno (FARE VERB MAIN IND FUT TRANS 3 PL) [0;TOP-VERB]
3 adesso (ADESSO ADV TIME) [2;ADVB-RMOD-TIME]
4 il (IL ART DEF M SING) [2;VERB-SUBJ]
5 Governo (GOVERNO NOUN COMMON M SING) [4;DET+DEF-ARG]

1 Perche' (|perche'| ADV INTERR) [3;ADVB+INTERR-RMOD]
2 avete (AVERE VERB AUX IND PRES TRANS 2 PL) [3;AUX+TENSE]
3 ucciso (UCCIDERE VERB MAIN PARTICIPLE PAST TRANS SING M) [0;TOP-VERB]
3.10 t [] (DEITT-T PRON PERS M PL 2) [3;VERB-SUBJ]
4 altri (ALTRO ADJ DEITT M PL) [3;VERB-OBJ]
5 albanesi (ALBANESE NOUN COMMON ALLVAL PL) [4;DET+DEF-ARG]
6 ? (#\? PUNCT) [3;END]

1 Ma (MA CONJ COORD ADVERS) [0;TOP-CONJ]
2 cosa (COSA PRON INTERR ALLVAL SING LSUBJ+LOBJ) [3;VERB-SUBJ]
3 accadra' (ACCADERE VERB MAIN IND FUT INTRANS 3 SING) [1;COORD2ND+ADVERS]

6 Vediamo (VEDERE VERB MAIN IND PRES TRANS 1 PL) [0;TOP-VERB]
6.10 t [] (DEITT-T PRON PERS ALLVAL PL 1) [6;VERB-SUBJ]
7 cosa (COSA PRON INTERR ALLVAL SING LSUBJ+LOBJ) [9;VISITOR]
8 si (SI PRON REFL-IMPERS ALLVAL ALLVAL 3 LSUBJ+LOBJ+LIOBJ CLITIC) [9;VERB-SUBJ/VERB-SUBJ+IMPERS]
9 puo` (POTERE VERB MOD IND PRES INTRANS 3 SING) [6;VERB-OBJ]
10 fare (FARE VERB MAIN INFINITE PRES TRANS) [9;VERB+MODAL-INDCOMPL]
10.10 t [8f] (SI PRON REFL-IMPERS ALLVAL ALLVAL 3 LSUBJ+LOBJ+LIOBJ CLITIC) [10;VERB-SUBJ]
10.11 t [7f] (COSA PRON PERS INTERR) [10;VERB-OBJ]

2 tutti (TUTTO PREDET M PL) [3;PDET+QUANTIF-RMOD]
3 gli (IL ART DEF M PL) [17;VERB-SUBJ]
4 scienziati (SCIENZIATO NOUN COMMON M PL) [3;DET+DEF-ARG]
...
21 capire (CAPIRE VERB MAIN INFINITE PRES TRANS) [20;PREP-ARG]
21.10 t [3f] (IL ART DEF M PL) [21;VERB-SUBJ]
22 cosa (COSA ADV INTERR) [24;VERB-SUBJ]

23 stia (STARE VERB AUX CONG PRES INTRANS 3 SING) [24;AUX+TENSE]
24 succedendo (SUCCEDERE VERB MAIN GERUND PRES INTRANS) [21;VERB-OBJ]

The following case is very interesting: we have a purpose infinitival clause governed by VEDERE which has as complement an indirect interrogative clause headed by a pronoun which has an anaphoric link with an antecedent (BANCHE) that is placed in the main clause. This is correctly marked but with a trace, as if it were a syntactically governed relation.

4 il (IL ART DEF M SING) [3;VERB-OBJ]
5 privato (PRIVATO ADJ QUALIF M SING) [6;ADJC+QUALIF-RMOD]
6 cittadino (CITTADINO NOUN COMMON M SING) [4;DET+DEF-ARG]
7 che (CHE PRON RELAT ALLVAL ALLVAL LSUBJ+LOBJ) [8;VERB-SUBJ]
8 fa (FARE VERB MAIN IND PRES TRANS 3 SING) [6;VERB-RMOD+RELCL]
9 il (IL ART DEF M SING) [8;VERB-OBJ]
10 giro (GIRO NOUN COMMON M SING) [9;DET+DEF-ARG]
11 delle (DI PREP MONO) [10;PREP-RMOD]
11.1 delle (IL ART DEF F PL) [11;PREP-ARG]
12 banche (BANCA NOUN COMMON F PL) [11.1;DET+DEF-ARG]
13 per (PER PREP MONO) [8;PREP-RMOD-REASONCAUSE]
14 vedere (VEDERE VERB MAIN INFINITE PRES TRANS) [13;PREP-ARG]
14.10 t [7p] (CHE PRON RELAT ALLVAL ALLVAL LSUBJ+LOBJ) [14;VERB-SUBJ]
15 in (IN PREP MONO) [17;PREP-RMOD-LOC+METAPH]
16 quale (QUALE ADJ INTERR ALLVAL SING LSUBJ+LOBJ) [15;PREP-ARG]
16.10 t [12f] (BANCA NOUN COMMON F PL) [16;DET+INTERR-ARG]
17 guadagnerebbe (GUADAGNARE VERB MAIN CONDIZ PRES TRANS 3 SING) [14;VERB-OBJ]
17.10 t [14.10p] (CHE PRON RELAT ALLVAL ALLVAL LSUBJ+LOBJ) [17;VERB-SUBJ]

In the following example the relation between the interrogative pronoun and the verb of governed clause is marked as OBJ – seen that the verb is in passive mood. However, the grammatical relation is usually referred to the surface syntactic structure and not the deep structure.

13 scoprire (SCOPRIRE VERB MAIN INFINITE PRES TRANS) [12;PREP-ARG]
13.10 t [] (GENERIC-T PRON PERS ALLVAL ALLVAL ALLVAL) [13;VERB-SUBJ]
14 quale (QUALE ADJ INTERR ALLVAL SING) [29;VERB-OBJ]
15 elemento (ELEMENTO NOUN COMMON M SING) [14;DET+INTERR-ARG]
16 nella (IN PREP MONO) [29;PREP-RMOD-LOC+METAPH]
16.1 nella (IL ART DEF F SING) [16;PREP-ARG]
17 struttura (STRUTTURA NOUN COMMON F SING) [16.1;DET+DEF-ARG]
18 ((#\(PUNCT) [15;OPEN+PARENTHETICAL]
19 sia (ESSERE VERB MAIN CONG PRES INTRANS 3 SING) [15;APPOSITION]
20 esso (ESSO PRON PERS M SING 3 LSUBJ+LOBJ) [19;VERB-SUBJ]
21 un (UN ART INDEF M SING) [19;VERB-PREDCOMPL+SUBJ]
22 arco (ARCO NOUN COMMON M SING) [21;DET+INDEF-ARG]
23 o (O CONJ COORD DISJ) [21;COORD+BASE]
24 un (UN ART INDEF M SING) [23;COORD2ND+BASE]
25 nodo (NODO NOUN COMMON M SING) [24;DET+INDEF-ARG]
26) (#\) PUNCT) [15;CLOSE+PARENTHETICAL]
27 era (ESSERE VERB AUX IND IMPERF INTRANS 3 SING) [28;AUX+TENSE]
28 stato (ESSERE VERB AUX PARTICIPLE PAST INTRANS SING M) [29;AUX+PASSIVE]
29 introdotto (INTRODURRE VERB MAIN PARTICIPLE PAST TRANS SING M) [13;VERB-OBJ]

And in the following sentence there are two SUBJect relations: the one headed by the interrogative pronoun, and the one headed by "I POLITICI" which should be the legitimate SUBJect.

1 Chi (CHI PRON INTERR ALLVAL ALLVAL LSUBJ+LOBJ) [2;VERB-SUBJ]
2 sono (ESSERE VERB MAIN IND PRES INTRANS 3 PL) [0;TOP-VERB]
3 i (IL ART DEF M PL) [2;VERB-PREDCOMPL+SUBJ]
4 politici (POLITICO NOUN COMMON M PL) [3;DET+DEF-ARG]

And now we will look into VIT:

0 Cosa cosa int(interrogative_pronoun) fint num=s 3 fint int
1 risponde rispondere vt(verb_trans_tensed) cl(main) punt - ibar trans/dir_speech
1.10 pro pro pro(little_pro) sn num=s|per=3 ant=sent_00195/10 s_impl-agente nil
1.11 rel_pro cosa rel_pro(relative_pronoun) bindee num=s|gen=m ant=0 obj-info com
2 loro egli pron(pronoun) sn num=p|per=fm 1 iobj pers
3 ? ? puntint(punctuation_non_declarative) fint nil 1 fint puntint

0 Ed ed cong(conjunction) fc [] 1 fc sum
1 è essere vc(verb_copulative) cl(main) punt - ibar cop/esistenza
1.11 pro pro pro(little_pro) nil num=s|per=3 9 s_expl nil
2 su su p(preposition) sp - 1 pcomp nil
3 quali quale int(interrogative_pronoun) fint num=p|per=fm 2 binder nil
4 intendere intendere vit(verb_trans_infinitive) sv2 punt 3 adj intr
5 come come ccom(conjunction_come_comparative) sc [] 4 sc comp
6 commissioni commissione n(noun) sn num=p|gen=f 5 sn com
7 di di pd(preposition_di) spd - 6 mod nil
8 controllo controllo n(noun) sn num=s|gen=m 7 pobj com
9 che che pk(complementizer) fac - 1 fcomp nil
10 maggioranza maggioranza n(noun) sn num=s|gen=f 11 subj com
11 e e cong(conjunction) coord [] 9 coord sum
12 opposizione opposizione n(noun) sn num=s|gen=f 11 subj com
13 parlano parlare vin(verb_intrans_tensed) ibar punt 9 ibar intr/attivita
13.11 rel_pro su_quali rel_pro(relative_pronoun) bindee num=p|gen=fm ant=3 obl-info com
14 un uno art(article) sn num=s|gen=m 15 sn ind
15 linguaggio linguaggio n(noun) sn num=s|gen=m 13 obj-informtn com
16 diverso diverso ag(adjective) sa num=s|gen=m 15 mod nil
17 . . punto(sentence_final) fc - 1 fc nil

9 ricordandomi ricordare vgt(verb_trans_gerundive) sv5 punt 3 adj tr
10 mi mi clit(clitic_pronoun) compt per=1|gen=f|num=sm 9 compt nom
11 con con p(preposition) sp - 9 adj nil
12 quale quale int(interrogative_pronoun) sp num=s|gen=m 13 binder nil
13 decisione decisione n(noun) sn num=s|gen=f 11 pobj com
14 sono essere ause(auxiliary_essere_tensed) ibar punt 16 ibar aux
15 stati essere ausep(auxiliary_essere_past_participle) ibar punt 16 ibar aux
16 presi prendere vppt(verb_trans_past_participle) ibar punt 10 ibar trans/attivita
16.11 rel_pro con_quale rel_pro(relative_pronoun) bindee num=p|gen=fm ant=12 obl-info com
17 provvedimenti provvedimento n(noun) sn num=p|gen=m 16 s_top-actor com
18 a a p(preposition) sp - 16 iobj nil
19 vicenza vicenza np(noun_proper_geographic) sn [] 18 pobj-loc geo

1.7.1 Interrogative Pronoun in other treebanks

We look now at CPT and ADT to see how they treat interrogative clauses, ADT first:

1	¿	¿	F	Fi	_	3	PUNC	_	_
2	Qui	qui	p	pt	num=s|gen=c	3	SUJ	_	_
3	és	ser	v	vs	num=s|per=3|mod=i|ten=p	0	S	_	_
4	l'	el	d	da	num=s|gen=c	5	ESPEC	_	_
5	enemic	enemic	n	nc	num=s|gen=m	3	ATR	_	_
6	del	del	s	sp	num=s|gen=m|for=c	5	SP	_	_
7	Barça	Barça	n	np	_	6	SN	_	_
8	?	?	F	Fi	_	3	PUNC	_	_

6	¿	¿	F	Fi	_	8	PUNC	_	_
7	què	què	p	pt	num=s|gen=c	8	CD	_	_
8	feia	fer	v	vm	num=s|per=3|mod=i|ten=i	0	S	_	_
9	el	el	d	da	num=s|gen=m	10	ESPEC	_	_

10	directiu	directiu	n	nc	num=s|gen=m	8	SUJ	_	_
11	de	de	s	sp	for=s	10	SP	_	_
12	l'	el	d	da	num=s|gen=c	13	ESPEC	_	_
13	estudi	estudi	n	nc	num=s|gen=m	11	SN	_	_
14	?	?	F	Fi	_	8	PUNC	_	_

1	¿	¿	F	Fi	_	5	PUNC	_	_
2	Quantes	quant	d	di	num=p|gen=f	3	ESPEC	_	_
3	vegades	vegada	n	nc	num=p|gen=f	5	CC	_	_
4	hem	haver	v	va	num=p|per=1|mod=i|ten=p	5	AUX	_	_
5	sentit	sentir	v	vm	num=s|mod=p|gen=m	0	S	_	_
6	que	que	c	cs	_	15	SUBORD	_	_
7	per	per	s	sp	for=s	15	CI	_	_
8	als	al	s	sp	num=p|gen=m|for=c	7	SP	_	_
9	restaurants	restaurant	n	nc	num=p|gen=m	8	SN	_	_
10	i	i	c	cc	_	9	CO	_	_
11	les	el	d	da	num=p|gen=f	12	ESPEC	_	_
12	esglésies	església	n	nc	num=p|gen=f	9	CONJUNCT	_	_
13	els	el	d	da	num=p|gen=m	14	ESPEC	_	_
14	casaments	casament	n	nc	num=p|gen=m	15	SUJ	_	_
15	són	ser	v	vs	num=p|per=3|mod=i|ten=p	5	CD	_	_
16	un	un	d	di	num=s|gen=m	17	ESPEC	_	_
17	negoci	negoci	n	nc	num=s|gen=m	15	ATR	_	_
18	?	?	F	Fi	_	5	PUNC	_	_

2	¿	¿	F	Fi	_	11	PUNC	_	_
3	Per	per	s	sp	for=s	11	CC	_	_
4	què	què	p	pt	num=s|gen=c	3	SN	_	_
5	per	per	s	sp	for=s	11	CCT	_	_
6	primera	primer	a	ao	num=s|gen=f	7	SADJ	_	_
7	vegada	vegada	n	nc	num=s|gen=f	5	SN	_	_
8	el	el	d	da	num=s|gen=m	9	ESPEC	_	_
9	Vaticà	Vaticà	n	np	_	11	SUJ	_	_
10	ha	haver	v	va	num=s|per=3|mod=i|ten=p	11	AUX	_	_
11	desmentit	desmentir	v	vm	num=s|mod=p|gen=m	0	S	_	_
12	una	un	d	di	num=s|gen=f	13	ESPEC	_	_
13	notícia	notícia	n	nc	num=s|gen=f	11	CD	_	_
14	?	?	F	Fi	_	11	PUNC	_	_

And this is CPT:

1	que	que	pron	pron-det	<interr>|M|S	2	>N	_	_
2	tipo	tipo	n	n	M|S	8	ACC	_	_
3	de	de	prp	prp	_	2	N<	_	_
4	relação	relação	n	n	F|S	3	P<	_	_
5	económica	económico	adj	adj	F|S	4	N<	_	_
6	se	se	pron	pron-pers	M/F|3S/P|ACC	7	SUBJ	_	_
7	poderá	poder	v	v-fin	FUT|3S|IND	0	QUE	_	_
8	estabelecer	estabelecer	v	v-inf	_	7	MV	_	_
9	entre	entre	prp	prp	_	8	ADVL	_	_
10	Israel	Israel	prop	prop	M|S	9	P<	_	_
11	e	e	conj	conj-c	<co-prparg>	10	CO	_	_
12	os	o	art	art	<artd>|M|P	14	>N	_	_
13	seus	seu	pron	pron-det	<poss|3S>|<si>|M|P	14	>N	_	_
14	vizinhos	vizinho	n	n	M|P	10	CJT	_	_
15	árabes	árabe	adj	adj	M|P	14	N<	_	_
16	?	?	punc	punc	_	7	PUNC	_	_

3	Como	como	adv	adv	<interr>	4	SC	_	_	
4	é	ser	v	v-fin	PR|3S|IND	1	QUE	_	_	
5	a	o	art	art	<artd>|F|S	7	>N	_	_	
6	sua	seu	pron	pron-det	<poss|3S>|F|S	7	>N	_	_	
7	relação	relação	n	n	F|S	4	SUBJ	_	_	
8	com	com	prp	prp	_	7	N<	_	_	
9	o	o	art	art	<artd>|M|S	10	>N	_	_	
10	piano	piano	n	n	M|S	8	P<	_	_	
11	?	?	punc	punc	_	4	PUNC	_	_	

6	quem	quem	pron	pron-indp	<interr>|M|P	7	SC	_	_
7	eram	ser	v	v-fin	IMPF|3P|IND	0	QUE	_	_
8	os	o	art	art	<artd>|M|P	9	>N	_	_
9	corruptores	corruptor	n	n	M|P	7	SUBJ	_	_
10	e	e	conj	conj-c	<co-vfin>	7	CO	_	_
11	por	por	pp	pp	<new>	15	ADVL	_	_
12	que	que	pron	pron-det	<new>|M|S	13	>N	_	_
13	motivo	motivo	n	n	M|S	11	P<	_	_
14	não	não	adv	adv	_	15	ADVL	_	_
15	foram	ser	v	v-fin	PS|3P|IND	7	CJT	_	_
...									
24	?	?	punc	punc	_	7	PUNC	_	_

The treatment is identical to TALN and TUT, in that the interrogative pronoun is directly linked to the following verb, and in case it is headed by a preposition, it is linked to the preposition which in turn is linked to the following verb.

1.8 CONCLUSIONS and RECOMMENDATIONS

We conclude by listing in a succint form the items we reviewed in the previous pages, indicating for each item the way in which the various treebanks have produced a dependency link. At the end we try to propose some recommendations.

1. Amalgams
1.1 Verbs and prepositions
- VIT and TUT decompose both preposition+article amalgams and criticized verbs
- TALN decompose only verbal amalgams
1.2 Running indexes
- TALN introduces a new separate index for the enclitics
- TUT CoLNN shallow version, introduces a new index for both amalgams
    - TUT deep version, only duplicates entries but no new index
- VIT introduces a new index only for enclitics
1.3 Other treebanks
- AnCora as TALN does not decompose preposition+article amalgams
    - it decomposes verb+clitics and introduces a new index, the clitic is treated as a morpheme

2. Multiwords
2.1 Locutions
- in both TUT and TALN locutions are entered in separate lines and are internally indicated as such by the lemma
- VIT only treats locutions as single entries - no decomposition
2.2 Named Entities
- TUT only indicates abbreviations, acronyms in full form and some formulaic Latin expression, by entering each part in separate lines and indicating their full form as lemma
- TALN on the contrary indicates NEs in the same way as locutions
- VIT keeps them as single entries but indicates they are multiwords in the lemma

2.3 Other treebanks
- AnCora is similar to VIT as all multiwords are entered as single entries.

3. Functional heads
3.1 Specifiers
- TUT - both deep and shallow - makes lexical verbs dependent on auxiliaries, nouns dependent on determiners
- TALN and VIT on the contrary treat minor categories contained in specifier position as dependent on the head, verb or noun
3.2 Other treebanks
- ADT treats determiners as dependent
- PTB however treats auxiliaries as heads of the lexical verb

4. Coordination
- TUT makes conjunctions dependent on first head they are coordinators of: second and following coordinated heads are dependent on the conjunction/punctuation
- TALN, ADT and CPT make conjunctions dependent on first head and sequent heads also dependent on first head;
- PTB distinguishes two cases of coordination: clause level and other cases. Clause level coordination is done as TUT does by linking second and following conjuncts to the conjunction/punctuation. Other cases are treated as TALN does.
- VIT makes conjunctions dependent on the governing verb if coordination is between clauses, otherwise it builds an abstract COORD node and constituent heads are made dependent on it

5.0. CHE complementizer
- TUT, TALN and VIT link the complementizer to the governing verb, and the verb of the sentential complement to CHE, thus realizing a chain. The same applies to PTB.
- In ADT and CPT, on the contrary, treat the complementizer as a minor constituent and functional head and do not treat as intermediate head: they link QUE to the verb of the sentential complement, which in turn is linked to matrix clause main verb.

5.1. CHE relative
- TUT and TALN treat CHE complementizer as functional head dependent on the verb of the Relative Clause. In this way, they treat it differently from the previous case.
- VIT treats it as we saw before in section 5.0.

6. Lexical (analytic) relative pronouns
- TUT and TALN link the preposition of the relative pronoun, and also the nominal head of the relative to the verb of the relative; then it links the relative pronoun to the preposition.
- Also PTB treats WHOSE in the same way as CUI, and WHICH as TUT and TALN do.
- PARC-700 LFG dependency treebank introduces an abstract PRO that is linked to WHOSE and to the verb of the relative, thus working as intermediate element of a chain
- VIT treats it differently according to the role of CUI/QUALE
6.1 VIT treatment according to different syntactic roles
a. argument/adjunct of relative verb
          - it directly modifies the main verb of the relative clause
b. adjunct modifier of argument of relative verb
          - it modifies an argument of the verb of relative clause
c. adjunct modifier of a noun
d. adjunct modifier of the internal nominal head
e. adjunct modifier of an ellipsed nominal head
f. adjunct of the subject/object nominal head of the relative
6.2 Relative pronouns in other treebanks
- are treated as previously shown with TUT, TALN and PTB
6.3 WHOSE in structures built by online parsers

- TUT and TALN link the interrogative pronoun (direct and indirect) to the verb of the following clause
- The same happens in CPT and ADT
- VIT introduce traces that are bound at the verb index: so eventually it links the interrogative pronoun to the following verb, in all cases, even when there is a preposition heading it.

1.8.1 Recommendations

As appears, there are differences between the treebanks considered in this report. However, it is important to remind that the main difference lies in the decision to produce a deep vs. a shallow structure. The deep structure may well encode dependency relations with the auxiliary help of null elements. I would say that major differences are in fact only minor problems that can be easily mended by a ad hoc script. What is not easy to produce is the presence of Null Elements which require a lot of additional computation and of manual checking.

The main difference is in the treatment of relative pronouns and relative clauses. In particular, we think it badly needing some amendment, the way in which CUI and other pied piping constructions have been treated. These structure need to be represented differently from relative clauses headed by relative pronouns which act as direct Argument/Adjunct of the relative verb. Of course this is something that can be done at best by inserting some empty element. However in some cases, there is a need to check dependencies which are not directly to the verb but to an argument/adjunct of the verb of the relative.

Luckily, these structure seem to be fairly uncommon. So eventually the net advantage in modifying a parser or some automatic procedure for the treebank annotation, is very small.

Other questions regard minor items, as said above, and they can be interpreted in terms of overall treebank consistency/coherence, and/or its strict/loose adherence to a linguistic theory. Treebank conversion tools made available for the CoNLL international challenge have determined a "de facto" standard in the way in which dependency relations are encoded. And this is obviously reflected in the fact that Penn Treebank has become the "de facto" standard of all syntactic treebanks. But it is clear that mapping constituency to dependency is not always easy and may require difficult decisions to be taken. Uniformity in the mapping encoded in a script is not always easy to guarantee, as we saw above. Also decisions as to what constitute a HEAD in dependency terms is not an easy decision in some cases, even though the theoretical background of linguistic theories should be helpful if properly used. In particular, if functional heads are treated as dependents they should always be treated as such; the same applies to the opposite case. However, this might become ambiguous in case there is a need to represent an implicit category like PRO for unexpressed relative pronouns.

Eventually what is needed is Semantic Transparency.
In other words, annotations in treebanks should be as much as possible transparent to semantic mapping procedures if they are to be of any use at all. For this reason we are convinced of the following:
- Minor categories and functional heads should always be treated as dependent, or if needed, be part of a chain with a semantic head; this is particularly true for the case in which negation is linked to the auxiliary rather than the lexical verb.
- Preserving the original orthography is not a major issue of a dependency treebank; multiwords should be treated as one unit if that is semantically justified; amalgams should be decomposed if needed for semantic opportunity – enclitics constitute arguments that will undergo anaphoric processes, but incorporated articles don't need to be assigned a separate index.
- Positing the existence of an abstract category like COORD which may serve for semantic purposes might be allowed even if it is linked to punctuation.

References

Bikel, Daniel M., Intricacies of Collins' parsing model. Computational Linguistics, 30 (4), (2003), 479-511.

Black, E. , S. Abney, D. Flickinger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini, and T. Strzalkowski: A Procedure for Quantitatively Comparing the Syntactic Coverage of English Grammars, Proceedings of the DARPA Speech and Natural Language Workshop, (1991), 306-311.

Brants,T. TnT: A statistical part-of-speech tagger. ANLP 2000.

Brill, E.. 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. Computational Linguistics 21, (1995), 543-565.

Carroll, J., T. Briscoe, A. Sanfilippo, Parser Evaluation: a Survey and a New Proposal, Proceedings of the [First] International Conference on Language Resources and Evaluation, (1998), 447-454.

Collins, Michael, A new statistical parser based on bigram lexical dependencies, Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, (1996), 184–191.

Corazza A., Lavelli A., Satta G., Zanoli R., Analyzing an Italian Treebank with State-of-the-Art Statistical Parsers, Proceedings of the 3rd Workshop on Treebanks and Linguistic Theories (TLT-2004), Tübingen, Germany, (2004), 39-50.

Delmonte R, Luminita Chiran, Ciprian Bacalu, Elementary Trees For Syntactic And Statistical Disambiguation, in Proc. TAG+5, Paris, (2000), 237-240.

Delmonte R., How to Annotate Linguistic Information in FILES and SCAT, in Atti del Workshop "La Treebank Sintattico-Semantica dell'Italiano di SI-TAL", Bari, (2001), 75-84.

Delmonte R., Shallow Parsing And Functional Structure In Italian Corpora, LREC, Atene, (2000),113-119.

Delmonte R., Strutture Sintattiche dall'Analisi Computazionale di Corpora di Italiano, in Anna Cardinaletti(a cura di), Intorno all'Italiano Contemporaneo, Franco Angeli, Milano, (2004), 187-220.

Delmonte R., Treebanking in VIT: from Phrase Structure to Dependency Representation, in Sergei Nirenburg (ed.), Language Engineering for Lesser-Studied Languages, IOS Press, Amsterdam, The Netherlands, pp.51-80, ISBN 978-1-58603-954-7.

Delmonte R., Antonella Bristot, Luminita Chiran, Ciprian Bacalu, Sara Tonelli, Parsing the Oral Corpus AVIP/API, in Albano Leoni A., Cutugno F., Pettorino M., Savy R.(a cura di), Atti del Convegno "Il Parlato Italiano", M.D'Auria Editore, N08, (2004), 1-20.

Delmonte R., Bristot A., Tonelli S. VIT - Venice Italian Treebank: Syntactic and Quantitative Features, in K. De Smedt, Jan Hajic, Sandra Kübler(Eds.), Proc. Sixth International Workshop on Treebanks and Linguistic Theories, Nealt Proc. Series Vol.1, (2007), 43-54.

Delmonte R., E.Pianta, IMMORTALE - Analizzatore Morfologico, Tagger e Lemmatizzatore per l'Italiano, in Atti V Convegno AI*IA "Cibernetica e Machine Learning", Napoli, (1996), 19-22

Delmonte R., E.Pianta,Tag Disambiguation in Italian, in Proc. Treebank Workshop ATALA, Paris, (1999), 43-49.

Delmonte R., From Shallow Parsing to Functional Structure, in Atti del Workshop AI*IA - "Elaborazione del Linguaggio e Riconoscimento del Parlato", IRST Trento, (1999), 8-19.

Delmonte R., Parsing Spontaneous Speech, in Proc. EUROSPEECH2003, Pallotta Vincenzo, Popescu-Belis Andrei, Rajman Martin "Robust Methods in Processing of Natural Language Dialogues" , Genève, (2003), 1-6.

Delmonte R., R.Dolci, Parsing Italian with a Context-Free Recognizer, Annali di Ca' Foscari XXVIII, 1-2, (1989), 123-161.

Duchier, D. and Debusmann, R., Topological dependency trees: A constraint-based account of linear precedence, Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL), (2001), 180-187.

Gildea D. 2001. Corpus variation and parser performance. In Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP 2001), (2001), Pittsburgh, PA, 167–202.

Grosz B., Sidner C., Attention, Intentions, and the Structure of Discourse, Computational Linguistics 12 (3), (1986), 175-204.

Harper, M. P. and Helzerman, R. A., Extensions to constraint dependency parsing for spoken language processing, Computer Speech and Language 9, (1995), 187-234.

Hellwig, Peter, "Dependency unification grammar", Proceedings COLING--86, (1986), 195--198.

http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/

Hudson R. English Word Grammar, Blackwell, London, (1990).

Hudson R. Word Grammar, Blackwell, London, (1984).

Jackendoff, R., X-Bar Syntax, The MIT Press, Cambridge, MA., 1977.

Jaervinen, T. and Tapanainen, P., Towards an implementable dependency grammar, in Kahane, S. and Polguère, A. (eds), Proceedings of the Workshop on Processing of Dependency-Based Grammars, (1998), 1-10.

Carletta, Jean , Amy Isard, Stephen Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon, and Anne Anderson, The reliability of a dialogue structure coding scheme. Computational Linguistics 23 (1), (1997), 13-32.

Mann, W.C. & Thompson, S.A., Rhetorical Structure Theory: A theory of text organization, Technical Report ISI/RS-87-190, ISI, 1987.

Marcus M. et al., Building a Large Annotated Corpus of English: The PennTreebank, Computational Linguistics, 19, 1993.

Maruyama, H. Structural disambiguation with constraint propagation. Proceedings of the 28th Meeting of the Association for Computational Linguistics (ACL), Pittsburgh, PA, (1990), 31-38.

Mel'cuk, I., Dependency Syntax: Theory and Practice, State University of New York Press, (1988).

Menzel, W. and Schroeder, I., Decision procedures for dependency parsing using graded constraints. In Kahane, S. and Polguère, A. (eds), Proceedings of the Workshop on Processing of Dependency-Based Grammars, (1998), 78-87.

Montemagni et al., The Italian Syntactic-Semantic Treebank: Architecture, Annotation, Tools and Evaluation, LINC, ACL, Luxembourg, (2000), 18-27.

Montemagni, S. F. Barsotti, M. Battista, N. Calzolari, A. Lenci O. Corazzari, A. Zampolli, F. Fanciulli, M. Massetani, R. Basili R. Raffaelli, M.T. Pazienza, D. Saracino, F. Zanzotto, F. Pianesi N. Mana, and R. Delmonte, Building the Italian Syntactic-Semantic Treebank. In Anne Abeillé, (editor), Building and Using syntactically annotated corpora, Kluwer, Dordrecht, (2003), 189–210.

Musillo G. & Khalil Sima'an, Towards comparing parsers from different linguistic frameworks. An information theoretic approach. In Proceedings of the LREC-2002 workshop Beyond PARSEVAL. Towards Improved Evaluation Measures for Parsing Systems, Las Palmas, Spain, (2002), 44-51.

Nivre, Joakim , Koenraad de Smedt and Martin Volk: Treebanking in Northern Europe: A White Paper, Nordisk Sprogteknologi. Nordic Language Technology. Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004. Editor: Henrik Holmboe. Copenhagen. 2005.

Petr Sgall, Eva Hajiçova, and Jarmila Panevova, The Meaning of the Sentence in Its Semantic and Pragmatic Aspects, D. Reidel & Academia, Dordrecht & Prague, (1986).

Rizzi L., Issues in Italian Syntax, Foris Publications, Dordrecht, 1982.

Tapanainen, P. and Jaervinen, T., A non-projective dependency parser, Proceedings of the 5th Conference on Applied Natural Language Processing, (1997), 64-71.

Tesnière, Lucien. Éléments de syntaxe structurale, Paris, Klincksieck, (1959).

Tonelli, Sara, Rodolfo Delmonte, Antonella Bristot (2008), Enriching the Venice Italian Treebank with dependency and grammatical relations, LREC 2008.