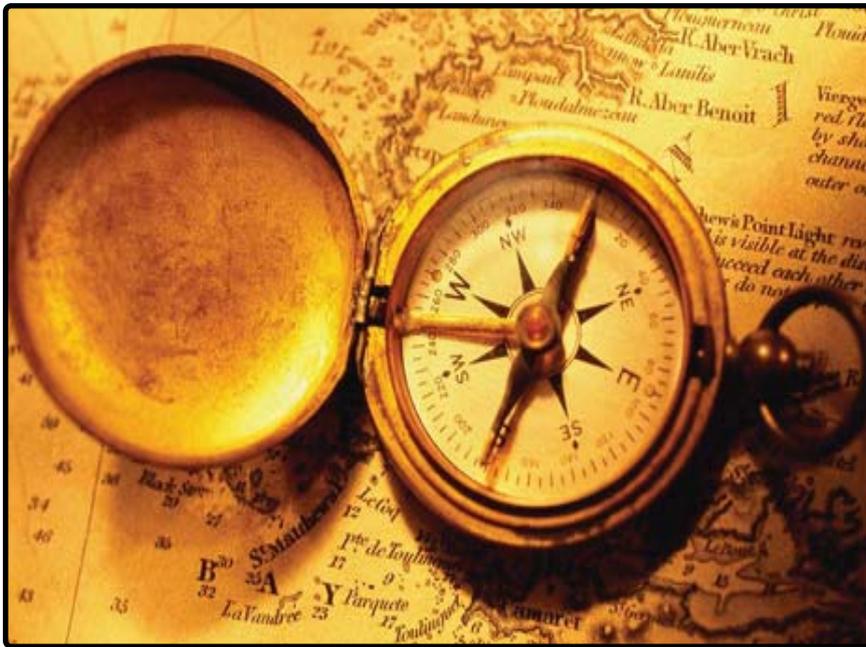


# STiL

## Studies in Linguistics

CISCL working papers on language and cognition

Vol 2.  
November 2008



**CISCL**  
**CENTRO INTERDIPARTIMENTALE DI STUDI COGNITIVI SUL LINGUAGGIO**  
Interdepartmental Centre for Cognitive Studies on Language  
Facoltà di Lettere e Filosofia

UNIVERSITÀ DEGLI STUDI DI SIENA



STiL - Studies in Linguistics  
Vol2, 2008

ISSN: 2281-3128

ISBN: 9788890794377

# STiL

## Studies in Linguistics

CISCL Working Papers on Language and Cognition  
Vol2, 2008

Edited by  
Vincenzo Moscati

board:  
Adriana Belletti  
Valentina Bianchi  
Cristiano Chesi  
Elisa Di Domenico  
Vincenzo Moscati  
Luigi Rizzi

CISCL WPs are open to submission by researchers who contribute or have contributed to the activities of the Interdepartmental Centre of Cognitive Studies on Language.

Correspondence can be addressed to:

CISCL – Centro Interdipartimentale di Studi Cognitivi sul Linguaggio  
Dipartimento di Scienze della Comunicazione  
Complesso S. Niccolò, Via Roma, 56  
I-53100 Siena, Italy

or by email at:  
[moscati@unisi.it](mailto:moscati@unisi.it)

#### NOTICE

CISCL Working Papers on Language and Cognition are meant to host unpublished work in progress. The copyrights remain with the authors and inclusion in an issue of the CISCL WP does not preclude publication elsewhere.

## Contents

*Adriana Belletti*

The CP of Clefts 7

*Linda Badan*

Preposed Object and Low Periphery in Mandarin Chinese 19

*Guglielmo Cinque & Luigi Rizzi*

The cartography of Syntactic Structures 43

*Jacopo Garzonio & Cecilia Poletto*

Minimizers and quantifiers:  
a window on the development of negative markers 60

*Valentian Bianchi & Cristiano Chesì*

Phases, Strong Islands, and Computational Nesting 82

*Maria Teresa Guasti, Stavroula Stavrakaki & Fabrizio Arosio*

Number and Case in the comprehension of relative clauses:  
Evidence from Italian and Greek 102

*Ida Ferrari & Simona Matteini*

On the residual omission of Determiners in Italian L1: a case study 120

*Irene Franco*

V1, V2 and criterial movement in Icelandic 142

*Cristiano Chesì, Gianluca Lebani & Margherita Pallottino*

A Bilingual Treebank (ITA-LIS) suitable for Machine Translation:  
what Cartography and Minimalism teach us 166



# The CP of *clefts*

ADRIANA BELLETTI

Ciscl – University of Siena  
belletti@unisi.it

## 1. Introduction

I have argued in recent work (Belletti 2008) that the CP of cleft sentences has some peculiar properties, the crucial ones being the following two:

- i. it is a reduced CP;
- ii. the reduced CP may or may not be endowed with an EPP feature.

In the pages that follow I would like to review the main arguments for i.-ii. in light of the different kinds of focalization that the two CP structures lead to in clefts. The main emphasis of the discussion here will be on the particular shape of the articulated CP projection in clefts. Throughout I will limit my attention to structures where the clefted constituent is a DP (or a PP), either a subject or a direct /indirect object.

## 2. The CP small clause

Consider the basic shape of the split CP projection. According to current analyses (Rizzi (1997), Benincà & Poletto (2004), Haegeman (2006), Bocci (2004) Grewendorf (2005), Miotto (2003), and related work), the Fin head is found at the bottom of the projection selecting the inflectional functional system of the following clause, while the Force head sits at the top of the projection expressing the illocutionary content of the clause, e.g. whether it is a declarative or an interrogative; the Force head is selected by the matrix verb when CP is embedded. Given this familiar background of assumptions, the question of the status of the CP of clefts naturally arises. Take the following two cleft sentences in (1) from Italian. I will use Italian throughout to illustrate different properties, unless other languages, in particular French, are needed to draw relevant distinctions.

- (1) a (subject cleft) E' Gianni che ha parlato  
it is Gianni that has spoken
- b (object cleft) E' Gianni che i ragazzi hanno salutato  
it is Gianni that the boys have greeted

Two main questions should be asked: i. where in the clause structure is the clefted constituent located? ii. Is the shape of the CP the same in the two cases?

Assume a vP periphery along the lines I have argued for in previous work (Belletti 2004, 2005); assume the classical hypothesis according to which the copula - *be* as a

shortcut – takes as its complement a small clause (a long standing hypothesis, dating back at least to Burzio (1986), Stowell (1983), and thoroughly developed in Moro (1997), Rothstein (2000)). It can be proposed that, in the case of clefts, the small clause of the copula is a CP, as schematized in (2) (Belletti (2008) and references cited there):

- (2) ..... *Be* [ CP .....]

The dots above (left of) *be* contain a vP periphery, where a new information focus head is present whose specifier is ready to host a new information constituent (see the references quoted for details). We have now to make explicit what the dots in CP correspond to. It seems correct to assume that they differ in part in subject vs non subject clefts.

### *2.1 The small CP of subject clefts*

As clearly evidenced by different languages, e.g. French, the postcopular subject DP of subject clefts can be the focus of new information. Typically, a subject cleft (with an often deleted/unpronounced predicate) can provide the answer to a question on the identification of the subject of the clause:

- (3) a      Qui (est-ce que qui) a parlé?  
              who spoke
- b      C'est Jean (qui a parlé)  
              it is Jean (who spoke)

As I have discussed in detail in the references quoted, this characteristic answering strategy of French share a crucial property with the inversion strategy characteristically adopted in similar contexts in languages allowing for post verbal new information subjects, such as, e.g., Italian:

- (4) a      Chi ha parlato?  
              who spoke
- b      Ha parlato Gianni  
              has spoken Gianni

In a cartographic perspective, in both cases the subject fills the same position: the specifier of the low vP peripheral new information focus position. It is in this position that it is interpreted as the constituent carrying the required new information. According to this analysis, the concealed/disguised inversion of subject clefts like (3)b is attributed the analysis in (5), details omitted<sup>1</sup>:

---

<sup>1</sup> See Belletti (2008). The vP periphery of the copula may also contain Top-type positions (within the dots in (5)) as is the case with the vP periphery of lexical verbs. However, a peculiarity of the copula, in particular in clefts, seems to be that it necessarily involves focalization. This could be expressed by the idea that the vP periphery of the copula is reduced and it solely contains the focus head. I will not develop this idea in detail here, but just note that one could go further and assume that the copula itself is a realization of the focus head in clefts. This idea would at the same time account for the necessary focalization involved in clefts and provide a natural characterization of the frequently observed fact across languages according to which the copula tends to grammaticalize into a focus



much as the subject position of regular TP clauses where a predication relation is established with the verbal predicate. The same analysis can be attributed to the small CP complement of perception verbs in pseudorelatives, of the type illustrated in (7)a, b in Italian, thus essentially updating the proposal originally due to Guasti (1993), keeping the main insight unchanged:

- (7) a Ho visto Maria che parlava con Gianni  
I have seen Maria that spoke to Gianni
- b Ho visto [CP...[EPPMaria [ FinPche [(*pro*) parlava (-) con Gianni]]]]

In (7), “Maria” is either directly merged in the EPP position of the small CP and a small *pro* related to it is present in the subject position of the following clause, or it is moved to the EPP position from the position where it is merged in the clause. In the latter derivation, extraction should take place from the vP-internal postverbal position indicated as “-“ in (7)b, and an expletive *pro* should sit in the preverbal high subject position (Rizzi & Shlonsky 2006, Cardinaletti 2004)). The former derivation is the one which most directly represents an update of the one assumed in Guasti (1993).<sup>3</sup> This issue aside, the small CP of (7)b has exactly the same shape as the one of a subject cleft in (6).

Pseudorelatives of the type in (7) differ from a subject cleft in one respect: while the cleft requires focalization of the small clause subject, focalization in the pseudorelative can either affect the subject or the entire small clause. This is witnessed by the possibility of using the same sentence (7)a as an answer to the following two questions:

- (8) a Chi hai visto (che parlava con Gianni)?  
whom have you seen (that spoke to Gianni)
- b Che cosa hai visto?  
what have you seen

As discussed in Guasti (1993) and Rizzi (2000), in both (8)a and b there is direct perception of “Maria”, but in (8)b it is the whole small CP which the question focuses on. Clefts, on the other hand, imply a peculiar semantics which provides a unique identification explicitly expressed by the focussed argument. In subject clefts, the (uniquely) identified argument is precisely the subject.<sup>4</sup>

---

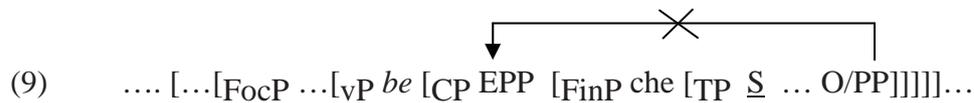
<sup>3</sup> A third derivation is possible, where “Maria” may be moved to the EPP position within CP from a “big DP” inside the clause, stranding a silent (referential, Belletti (2005)) small *pro* in the relevant EPP position within the clause. I leave a resolution of the various technical implementations open, which should optimally be decided on both empirical and theoretical independent grounds.

<sup>4</sup> Exchanges like the following, discussed in Rialland, Doetjes & Rebuschi (2002) are possible in French:

- i. Q. Qu’est-ce qui se passe?  
what happens  
A. C’est le petit qui est tombé dans l’escalier (qui se passe)  
it is the kid who has fallen on the stairs (that happens)

In cases like this, as suggested by the possibly unpronounced intended predicate in i.A, the uniquely identified argument is present in the cleft answer to the general question of information, modified by a (restrictive) relative clause.

Given the A status of the EPP position of the small CP complement of the copula that we have assumed, a prediction is directly derived: this position can be filled by the subject of the following clause, but it cannot be filled by a DP corresponding to the direct or indirect object of the following clause. This is so for locality reasons: Relativized Minimality (RM) would be violated in moving an object to the EPP position of the small CP, crossing over the intervening subject. The relevant part of the derivation is illustrated in the following schema in (9):



Hence, a direct consequence of the proposed analysis of subject clefts is that only the subject of the (TP) clause can reach the EPP position - or be directly merged there - in the small CP, for principled reasons.<sup>5</sup> Indeed, this is precisely what happens in pseudorelatives. Only the subject of the clause can be the head of a pseudorelative. This is a well known fact accounted for in similar locality terms in Guasti (1993). The ungrammaticality of (10)a,b, minimally contrasting with the wellformedness of (7)a, repeated in (10)c, illustrates the relevant contrast:

- (10)      a \*Ho visto Maria che Gianni/i ragazzi salutava/salutavano  
                  I have seen Maria that Gianni /the boys greeted
- b \*Ho visto con Maria che Gianni parlava  
                  I have seen with Maria that Gianni spoke
- c Ho visto Maria che parlava con Gianni  
                  I have seen Maria that spoke with Gianni

In contrast, in non subject clefts, to which we turn in the following section, the unique identification implied by the semantics of clefts may also be brought about by a focussed non subject argument. The kind of focalization, however, is not the same in subject and non-subject clefts as we argue in 2.2. And this is the key of the contrast with pseudorelatives.

### 2.2 *The reduced/truncated CP of non subject clefts*

Suppose that the CP complement of the copula has the same shape as in (9), but that no active EPP feature is present. In this case the CP is not a small CP, in the technical sense defined above. Nevertheless, it is a similarly reduced CP, where the Force head is lacking. We may see this CP as a truncated CP, in Rizzi's (2005) sense. We can propose that the CP complement of the copula in non subject clefts is precisely a reduced/truncated CP of this sort. If there is no EPP to be satisfied, this has the consequence that there should be no restriction for non-subject arguments to move into the reduced/truncated CP crossing over the intervening subject argument in TP;

<sup>5</sup> This in turn has the consequence that only the subject can then reach the new information focus position in the vP periphery of the copula. See the discussion in 2.2 .

Note that for a direct/indirect object the option of being directly merged in the EPP position with a related *pro* sitting in the argument position of the clause, is not an available option (due to the licensing constraints on object *pro*). Hence, the intervention problem necessarily arises in this case. Furthermore, the movement of a PP to the EPP position would be different from a PP pre-posing operation, which is an A' type operation.

the movement implemented in this case would be an A' type movement, hence no intervention effect should be produced in this derivation.

To the extent that clefts involve a form of focalization, the natural proposal can then be made that non subject arguments move to the focus position of the reduced/truncated CP. Indeed, this kind of left peripheral focalization within the CP complement of the copula is the only focalization admitted for non subject arguments. Specifically, an alternative direct long movement of an (direct or indirect) object from the embedded TP into the new information focus position in the vP-periphery of the matrix copula is excluded on principled locality grounds. Phase theory (Chomsky (2005)) explicitly rules out the possibility of such long direct movement with no intermediate steps (within the CP), with the embedded CP sent to spell out. But no intermediate step is possible in this case as the reduced/truncated CP complement of the copula does not contain any escape hatch *edge* position, given its reduced/truncated nature. In particular, it does not contain any position different from the criterial interpretable ones, such as e.g., the focus position, from which movement is excluded in principle, through any version of criterial freezing (Rizzi (2006)). The consequence of all this is that while the focalization of subject clefts can occur in the vP periphery of the copula and correspond to the new information focalization expressed by this position, the focalization of non subject arguments necessarily corresponds to left peripheral focalization.

As discussed in detail in various works (Belletti (2004, 2008; Bocci (2004), Rizzi (1997), a.o.), left peripheral focalization involves more than just new information. It typically is contrastive/corrective focalization. If this is the case, then, one direct consequence of the proposal is that although clefts constitute a form of focalization in general, the focalization of subjects can be new information focalization, while the focalization of non subject arguments is contrastive/corrective focalization<sup>6</sup>. I have proposed in the quoted references that a direct reflex of this different way of focalization may be found in the fact that although a cleft (with an often deleted/unpronounced predicate) can be used as a felicitous answer to a question of information on the identification of the subject, the same possibility is not available for a question of information on the identification of the object. This is visible in those languages where clefts are used as a most suitable answering strategy, as in the case of French. The contrast in (11) in French can be taken as an illustration of this important distinction:

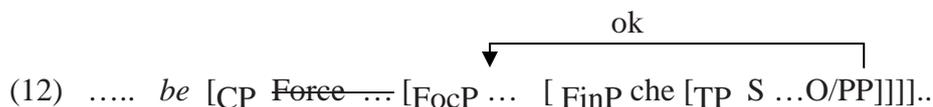
- (11)           Q Qu'est-ce-que t'as acheté (/Qu'as-tu acheté)?  
                  what have you bought  
                  A \*C'est un livre  
                      it is a book
- Q Avec qui es-tu sorti?  
                      with whom did you get out  
                  A \*C'est avec Jean  
                      it is with Jean

---

<sup>6</sup> Of course, a subject cleft can also instantiate contrastive/corrective focalization, implementing the same movement in the reduced/truncated CP left periphery as non subject arguments. Nothing excludes this possibility. "S" can thus be involved in the same derivation illustrated in (12) following, for non subject clefts the only available derivational option.

On the other hand, it may be speculated that this clear distinction should not hold in languages where both new information focus and contrastive focus are realized in the left periphery of the clause. Should languages of this type exist, all other things being equal, in these languages both subject and non subject clefts should qualify as possible answering strategies to questions of information. I leave the development of this parametrical option open for further study.<sup>7</sup>

The proposed derivation of non subject clefts is schematized in (12):



For the sake of explicitness, in (12) the reduced status of the CP complement of the copula is illustrated in terms of the truncation idea.

Summarizing, in both subject and non subject clefts the copula *be* takes a reduced/truncated CP as complement. The reduced/truncated CP may or may not contain an active EPP feature. When it does, it is a small clause CP (a small CP, as we have called it) where a predication type relation holds between the subject of the small clause and the (rest of the) CP predicate, the same kind of relation instantiated in the pseudorelative complement of perception verbs. For principled locality reasons, only the subject of the CP predicate can check the EPP feature. One consequence of this is that only the subject can realize the new information focus in the vP periphery of the matrix copula. If, in contrast, the reduced/truncated CP complement of the copula does not contain any EPP feature, then the focalization implemented by the cleft is left peripheral focalization within the reduced CP complement of the copula. Such focalization is thus contrastive/corrective focalization and not simple new information focalization and, crucially, may also affect non-subject arguments.

### 3. The position of “*che*” and the nature of the CP complement of “*be*”

Given the general analysis presented in the preceding paragraphs, I would now like to look more closely at the shape of the reduced CP complement of the copula, concentrating more specifically on the position and nature of the complementizer present in clefts, *che* in Italian in the examples we will consider<sup>8</sup>.

According to the proposed analysis *che* is not the realization of the Force head in clefts; rather, it is the realization of finiteness, the lowest Fin head in the articulated CP. This amounts to claiming that the (reduced) clausal complement of the copula is not a full fledged declarative sentence in any case.

<sup>7</sup> Hungarian and Sicilian (Belletti (2008)) may be two languages worth looking at in this perspective. Furthermore, other languages may also more or less parasitically exploit this UG option. One first thing to determine, however, is whether a cleft is an answering strategy normally adopted in the relevant language. I am not in a position to provide structured data in this domain for the moment, so I leave open to future investigation this intriguing comparative issue.

<sup>8</sup> A regular restrictive relative introduced by a clear relative pronoun does not seem to be possible in Italian, as witnessed by the strong marginality of sentences like i.a which contrast with i.b:

- i. a\*? E' (a) Gianni a cui parlerò di questo problema  
it is (to) Gianni to whom I will speak of this problem
- b E' a Gianni che parlerò di questo problema  
it is to Gianni that I will speak of this problem

Clear distributional evidence in favour of this distinction comes from the contrast in (13). In (13)a left peripheral focalization is implemented within the declarative CP complement of the verb “dire” (say); as the ungrammaticality of (13)b shows, the mandatory respective order of the complementizer and the focalized argument, here a direct object, is C – Foc and cannot be Foc – C. This is expected given the shape of the articulated CP, the very nature of the Force head, and the respective order of the Force and Focus heads, with Force the highest head of the articulated CP.

- (13) a Ho detto che GIANNI avrebbero assunto (non Maria)  
I have said that GIANNI they would have hired (not Maria)
- b \*Ho detto GIANNI che avrebbero assunto (non Maria)  
I have said GIANNI that they would have hired (not Maria)

Similarly, whenever a perception verb like “vedere”(see) is used in its epistemic reading (and not in its perception reading) the order, as expected, is once again C – Foc and not Foc – C, as illustrated in (14).

- (14) a Ho visto che GIANNI avrebbero assunto (non Maria)  
I have seen that GIANNI they would have hired (not Maria)
- b \*Ho visto GIANNI che avrebbero assunto (non Maria)  
I have seen GIANNI that they would have hired (not Maria)

In the CP complement of the copula in clefts, which in the proposal we have developed is reduced/truncated under Force, the order is rather Foc – C, as (15) reminds.

- (15) a E' GIANNI che assumeranno (non Maria)  
it is GIANNI that they will hire (not Maria)
- b E' con GIANNI che parleranno del problema (non con Maria)  
it is with GIANNI that they will speak of the problem (not with Maria)
- c E' GIANNI che ha parlato (non Maria)  
it is GIANNI that spoke (not Maria)

The respective order of Foc and C in (15) is directly obtained if *che* is not here the realization of Force, but rather the realization of Fin, as assumed. (15)c is an instance of left peripheral focalization of the subject in the reduced CP complement of the copula, an option available for all kinds of arguments, direct and PP complements included as in (15)a, b (see footnote 6, and the discussion in 2.2).

If *che* is not the expression of the declarative Force of the clause in clefts, this comes close to claiming that clefts like those in (15), which instantiate left peripheral focalization, are not that different from root left peripheral focalization in sentences like (16).

- (16) a GIANNI assumeranno (non Maria)  
GIANNI they will hire (not Maria)

- b Con GIANNI parleranno del problema (non con Maria)  
with GIANNI they will speak of the problem (not with Maria)
- c GIANNI ha parlato (non Maria)  
GIANNI spoke (not Maria)

Under cartographic assumptions, in (16) the contrastively/correctively focalized phrase fills the specifier of the high focus position within the articulated CP; this is exactly the same position occupied by the embedded focalized argument in the clefts of (15).

However, even though no declarative force is expressed by the complementizer in clefts like (15), this is not equivalent to saying that focalizing by means of a cleft as in (15) amounts to exactly the same kind of focalization as root left peripheral focalization of the kind in (16). As noted above, a cleft is not just a way of focalizing a phrase. Even if this may be (one of) the most salient property of clefts in general, other semantic/discourse values are implied by use of a cleft. In particular, a cleft also implies a unique identification of the focussed element (Kiss (1988); Abels & Muriungi (2005) for more recent discussion). Furthermore, there is in clefts what we may call a presupposition of existence, likely to be induced by the very presence of the copula.<sup>9</sup> A similar presupposition is not necessarily implied in (root) left peripheral focalization. The following contrast in (17), brought to my attention by Paola Benincà, suggests exactly this kind of distinction between the two focalization procedures. The contrast identifies one context where use of one of the two structures is not just infelicitous, it is plainly ungrammatical.

- (17) a NESSUNO ho incontrato (non tutti)  
nobody I met (not everybody)
- b. \*(Non) E' NESSUNO che ho incontrato (non tutti)  
it is (not) nobody that I met (not everybody)

In (17) the indefinite negative quantifier “nessuno” (nobody), corresponding to the direct object of the clause, is (contrastively) focalized in the left periphery. The clear ungrammaticality of (17)b in contrast with the possibility of (17)a indicates that such focalization can be done by means of a plain left peripheral operation, but not through a cleft.<sup>10</sup> We interpret this as due to the special semantic-discourse value implied by use of a cleft, which, as a first approximation, we have identified in a presupposition of existence of the uniquely identified argument, linked to the very presence of the copula. Thus, although the kind of focalization can be the same in the two structures, involving the left peripheral focus position in both cases, the two structures are not equivalent in their overall discourse value, with the consequence that, in pairs like (17), one is just plainly ungrammatical.

<sup>9</sup> An interesting comparative question that I am not in a position to properly address here, concerns the status of clefts (or rather their equivalents) in languages without the copula. This is left open to future investigation.

<sup>10</sup> The left peripheral focalization of the indefinite quantifier can also be located in an embedded CP, with the declarative complementizer preceding the focalized phrase, as always. This is illustrated in

i. Ho detto che NESSUNO assumeranno (non tutti)  
I have said that NOBODY they will hire (not everybody)

It thus seems that there are good reasons to believe that the complementizer present in clefts is not the one found in embedded declarative clauses expressing the declarative illocutionary force of the sentence. Since, however, it is the same word, at least in Italian (*che*) and several other languages, it would be most welcome if this coincidence could be expressed in some form. Let us then hypothesize in this connection that a complementizer like *che* always originates in Fin. Let us further assume that in a full fledged selected subordinate CP it raises up to the Force head to check the interpretive illocutionary declarative force of the clause. It would be through a mechanism of this sort that *che* at the same time expresses the finite nature of the clause and its declarative force. Given these assumptions, should the Force head not be present, *che* could solely express finiteness. Our proposal has been that this is precisely the case of clefts, where *che* remains in Fin and there expresses the finite nature of the embedded clause.

It is known that there are languages where more than one complementizer is/can be expressed in complementation.<sup>11</sup> If the derivational mechanism just described is on the right track and if it has a general application, it would provide a direct reason why this possibility should arise: given the view that movement is copying (Chomsky (1995) and subsequent work), the two instances of the complementizer could just be seen as two spell-outs of different copies. The copies would be located in the distinct Force and Fin heads, with the complementizer originally externally merged in Fin, and subsequently internally re-merged in Force.<sup>12</sup> While it is generally just the highest copy the one which is sent to spell out, it is in principle conceivable that, under defined conditions whose identification is beyond the aims of the present work, both copies be phonetically realized. This may be a non trivial general consequence of the idea that *che* is primarily the realization of Fin. A property that cleft sentences appear to overtly realize in a reduced/truncated CP, where *che* remains in Fin.

## References

- Abels, K. ad Muriungi, P. (2005) 'The Focus particle in Kîtharaka', *ZAS Papers in Linguistics*, 42: 219-239.
- Belletti, A. (2008) "Answering strategies: New information subjects and the nature of clefts", chapter 10 of *Structures and Strategies* New York: Routledge, in press.
- Belletti, A. (2004) 'Aspects of the low IP area', in L. Rizzi (ed.) *The structure of CP and IP. The cartography of Syntactic Structures*, vol. 2, Oxford: Oxford University Press, 16-51.
- Belletti, A. (2005) 'Extended Doubling and the vP periphery', *Probus*, 17(1): 1-35.
- Benincà, P. and Poletto, C. (2004) 'Topic, Focus and V2: Defining the CP Sublayers', in Rizzi (ed.) *The structure of CP and IP. The Cartography of Syntactic Structures*, vol. 2, New York: Oxford University Press, 52-75.

---

<sup>11</sup> See Paoli (2007) for recent discussion of different Romance varieties manifesting this possibility, which is interpreted along lines related to ours, with the complementizers exploiting the two heads Force and Fin. Our proposal capitalizes on the nature of movement as copying as is discussed momentarily in the text.

<sup>12</sup> As pointed out in Belletti (2008), the proposal also opens up the possibility that some language may have two different complementizers realizing Force and Fin respectively, or else that some language realizes the complementizer of clefts in a way different from the declarative complementizer. A way of characterizing these (hypothetical) languages could be that the complementizer in Fin would have the property of not moving to Force, where the relevant feature would be expressed by a complementizer directly merged in Force.

- Bocci, G. (2004) 'Contrastive focalization on topics and preverbal subjects in Italian', *Rivista di Grammatica Generativa*, 29, 3-59
- Burzio, L. (1986) *Italian Syntax. A Government-Binding Approach*, Dordrecht: Reidel.
- Cardinaletti, A. (2004) 'Towards a Cartography of Subject Positions', in L.Rizzi (ed.) *The structure of IP and CP. The Cartography of Syntactic Structures*, vol. 2, New York: Oxford University Press, 115-165.
- Chomsky, N. (1995) *The Minimalist Program*, Cambridge Mass: MIT Press.
- Chomsky, N. (2005) 'On Phases', in R.Freidin, C. Otero and M.L.Zubizarreta (eds) *Foundational Issues in Linguistic Theory*, Cambridge, Mass: MIT Press.
- Frascarelli, M. & A.R. Puglielli (2005), "The Focus System in Cushitic Languages", in P. Fronzaroli e P. Marrassini (eds.), *Proceedings of the 10th Hamito-Semitic Congress. Afroasiatic Linguistics*
- Grewendorf, G. (2005) "The discourse configurability of scrambling", in J.Sabel and M. Saito (eds) *The Free Word Order Phenomena*, Berlin:Mouton de Gruyter, 75-136.
- Guasti, M.T. (1993) *Causative and Perception Verbs. A Comparative Study*, Torino: Rosenberg and Sellier.
- Haegeman, L. (2004) 'Topicalization, CLLD and the Left Periphery', in B.Shaer, W.Frey and C.Maienborn (eds.) *Proceedings of the Dislocated Elements Workshop ZAS Berlin, November 2003*. 35:12 Berlin: ZAS
- Hiraiwa, K. & S Ishihara (2002) "Missing Links: Cleft, Sluicing, and "No da" Construction in Japanese", *MIT Working Papers in Linguistics*
- Kiss, K.E. (1998) 'Identificational Focus versus Information Focus', *Language*, 74,2: 245- 273.
- Mioto, C. (2003) 'Focalização e quantificação', *Revista Letras*, 61, Curitiba, 169-189.
- Moro, A. (1997) *The Raising of Predicates*, Cambridge: Cambridge University Press.
- Paoli, S. (2007) "The fine structure of the left periphery: COMPs and subjects. Evidence from Romance", *Lingua* 117: 1057-1079
- Rialland, A., J. Doetjes & G. Rebuschi (2002) "What is focussed in *C'est XP qui/que* Cleft sentences in French", *Speech Prosody 2002*, ISCA Archive
- Rizzi, L. (1997) 'The Fine Structure of the Left Periphery', in L.Haegeman (ed.) *Elements of Grammar*, Dordrecht: Kluwer, 281-337.
- Rizzi, L. (2000) 'Direct perception, government and thematic sharing', in L.Rizzi *Comparative Syntax and Language Acquisition*, London/New York:Routledge, 189-210.
- Rizzi, L. (2005) 'On the grammatical basis of language development: a case study', in G. Cinque and Kayne R. (eds) *The Oxford Handbook of Comparative Syntax*. New York: Oxford University Press.
- Rizzi, L. (2006) 'On the Form of Chains: Criterial Positions and ECP Effects' in S.Cheng and N.Corver (eds), *WH-Movement Moving on*, MIT Press, 97-134.
- Rizzi, L. and Shlonsky, U. (2007) 'Strategies of Subject Extraction', in U. Sauerland and H.M. Gärtner (eds) *Interfaces + Recursion = Language?*, Mouton De Gruyter, 115-160.
- Rothstein, S. (2000) *Predicates and Their Subjects*, Kluwer.
- Starke, M. (1995) 'On the Format of Small Clauses', in A.Cardinaletti and M.T.Guasti (eds.) *Small Clauses, Syntax and Semantics*, vol.28, New York: Academic Press 237-269.

Stowell, T. (1983) 'Subjects across categories', *The Linguistic Review*, 2.3, 285-312.

# Preposed Object and Low Periphery in Mandarin Chinese\*

LINDA BADAN

City University of Hong Kong  
lbadan@cityu.edu.hk

In this paper I analyse the Low Periphery in Chinese, following the basic lines of Belletti (2001, 2004) and Paul (2005). Like Italian, I show that Chinese displays Topic and Focus projections within IP. I individuate two different Functional Projections occupied by two distinct elements: the bare preposed Object (between Subject and verb) and the sentence-internal *lian* “even”+XP. Moreover I show that both have moved with A-movement. Contrary to the traditional analysis (Ernst & Wang 1995; Shyu 1995, 2001 among others), I finally argue that the bare preposed Object is not a Focus, but a Topic-like element with a Focus stress and it can be analyzed as a Contrastive Topic.

## 1. Introduction

Belletti (2001, 2004) proposes that the architecture of the domain below IP and above VP is parallel to the clause-external Left Periphery, i.e. in the CP area (see also Poletto 2006). In this article I follow Paul (2005), who applies Belletti’s proposal to Mandarin Chinese, confirming the parallelism between CP and IP peripheries.

In the first part, I illustrate some tests to prove the existence of the Low Periphery in Chinese. In section 4, I study the two kinds of Object items that can occupy the position between Subject and Verb: the *lian* “even”+XP and the direct Object (without any additional marker) moved in a position between Subject and Verb. Following Shyu (1995, 2001), Ting (1995), Zhang (1996) a.o., I discuss the fact that the two preposed elements within IP are dislocated by A-movement (section 4.2). Furthermore, I will investigate the nature of the projections activated in the Low Periphery in Chinese. Leaving out the *lian...dou* construction, in the last part of the paper I analyze the SOV order. Contrary to the traditional analysis as a Focus item (Ernst & Wang 1995; Shyu 1995, 2001; Tsai 1994; Zhang 1996), I argue that it can be considered a Contrastive Topic, i.e. a syntactic Topic that can get contrastive stress, on the basis of its syntactic behavior and its pragmatic/semantic interpretation.

---

\* I would like to thank Adriana Belletti, Pan Haihua, Waltraud Paul and the anonymous reviewers for their helpful comments and suggestions. All errors remain my responsibility.

## 2. Belletti (2001, 2004)

Belletti (2001, 2004) proposes an analysis of the fine-grained structural cartography of the clause (IP)-internal Low Periphery<sup>1</sup>. She shows that the area immediately above VP is parallel - to some extent - to the (clause-external) Left Periphery of the clause. She refers to this “internal” area as the “clause internal periphery” or “Low periphery”. The Low Periphery contains different positions associated to the corresponding interpretations and partly to different intonations, as opposed to the projections located in the High Periphery of the CP area. Chomsky (2000) (quoted in Belletti 2004) in a recent version of the Minimalist Program reached a similar conclusion, arguing for the consideration of CP and VP as two “strong Phases”, i.e. two syntactic units, independent from each other, which are the domains of syntactic operations<sup>2</sup>. This idea suggests a parallelism between CP and VP internal structures and properties. Considering such a resemblance, Belletti (2001, 2004) proposes that in Italian there are two positions dedicated to Focus in the clause: a structurally high one, in the CP area, and a structurally low one, in the “clause internal periphery”. She aims at showing that these two Focus Projections are different: the low Focus is restricted to Information Focus and the high Focus in the Left Periphery is a Contrastive Focus, and carries a special stress. After analyzing Subject inversion in Romance languages<sup>3</sup> and following Calabrese (1992), who proposed that the post-verbal Subject in Italian is Focalized, Belletti argues that the Spec of the low (Info)FocusP (a clause-internal Projection, above VP) is the landing site for a post-verbal Focalized Subject<sup>4</sup>. The Subject moves to the Spec of (Info) FocusP and the verb raises higher up, producing the order Verb-Subject:

(1) ... [I Verb [TopP [FocP **Subj** [TopP [ *t*<sub>subj</sub>



(2) Q: Chi ha parlato?<sup>5</sup>

(Belletti 2001: 3)

Who has spoken

“Who spoke?”

A: Ha parlato Gianni<sub>InfoFocus</sub>

Has spoken Gianni

“Gianni spoke.”

B: # GIANNI ha parlato.

Gianni has spoken

“Gianni spoke.”

<sup>1</sup> See also Jayaseelan (2001), Belletti & Shlonsky (1995) and Poletto (2006).

<sup>2</sup> VP/vP are assimilated to the general format of the small clauses, which have been analyzed as full clauses that include a peripheral C Projection (Starke 1995; Sportiche 1995 quoted in Belletti 2004).

<sup>3</sup> Free Subject Inversion is a property of Null Subject languages, which allow the Subject to be phonetically unrealized (Kayne 1984; Belletti 2004).

<sup>4</sup> Notice that with appropriate pragmatic condition and the proper intonation the postverbal Subject can be interpreted as a Topic:

(i) Q: Che cosa ha poi fatto Gianni?

(Belletti 2004: 10)

What has then done Gianni

“What (then) did Gianni do?”

A: Ha (poi) parlato, Gianni.

Has then spoken Gianni

“He spoke, Gianni.”

<sup>5</sup> The following abbreviations are used in glossing examples: CL classifier; .Cl clitic; DE determination particle; EXP experiential aspect; FP final particle; PERF perfective aspect; Q question marker; TOP topic marker; SHI...DE cleft construction.

As for the low Contrastive Focus, she proposes that the Subject moves to the Spec of the (Contrastive)FocusP in the CP area and the Object moves up to TopicP lower than (Contrastive)FocusP (Rizzi 1997)<sup>6</sup>. Their post-verbal position results from movement of the remnant IP to the Spec of a higher TopP, past the peripheral focalized Subject and topicalized direct Object (see (3) and the schematized movements in (4)):

(3) Ha comprato MARIA, il giornale. (Belletti 2004: 24Bb-27)  
Has bought MARIA the newspaper

(4) [[IP<sub>k</sub> e<sub>i</sub> ha comprato e<sub>j</sub>]<sub>TOP</sub> [[MARIA]<sub>FOC</sub>] [[il giornale]<sub>TOP</sub>]<sub>TOP</sub>...IP<sub>k</sub>

As for Topics, consider the following sentences:

(5) a. L' ha comprato Maria, il giornale  
It.*Cl* has bought Maria the newspaper  
b. Ha comprato Maria, il giornale  
has bought Maria the newspaper

(5)a is a case of Clitic Right Dislocation; (5)b is a case of so-called *emarginazione* “marginalization” in Antinucci & Cinque’s (1977) sense<sup>7</sup>. Following Cecchetto (1999), Belletti assumes that the right dislocated phrase is located in a clause-internal low Topic position (below the clause-internal Focus): the clitic is raised to the high position in the clause, for Case requirements, leaving behind the topicalized Object. The fact that in (5)b there is not the clitic, means that the Object is related to its Case assigning Head directly, without the mediation of a clitic.

In summary, Belletti’s proposal is that the Low Periphery is symmetric to the Left Periphery as concerns Focus and Topics Heads: there is a low FocusP and also two TopicPs that surround it<sup>8</sup>.

(6) [[IP [TopicP\* [FocusP [TopicP\* [VP]]]]]]

Here I follow Benincà (2001) and Benincà & Poletto’s (2004) more restrictive theory, and I assume that it is not possible to have a Topic Projection lower than FocusP in the CP area. I maintain the same idea too as far as it concerns the Low periphery.

### 3. The Low Periphery in Chinese

#### 3.1 Previous proposals

The Chinese language displays the possibility to have the “bare” direct Object (without any additional marking) not in its canonical post-verbal position (SVO word order), but raised to the left of verb<sup>9</sup> and below Subject, yielding the SOV order (henceforth I indicate the bare preposed Object within IP also with “SOV”):

<sup>6</sup> The hierarchy of the Left Periphery in the CP area proposed by Rizzi (1997) argues for a FocusP surrounded by Topic Projections.

<sup>7</sup> Both of these Topics are pronounced, after a pause, with a downgrading intonation.

<sup>8</sup> She follows Rizzi (1997), who hypothesizes that Topic is a set of recursive projections (he indicates recursion with a \*) occurring both higher and lower than a single Focus projection.

<sup>9</sup> For the moment I leave apart its pragmatic/semantic interpretation (see section 5.1)

- (7) a. Lisi mei kanguo [zhe ben shu]. (canonical SVO order)  
 Lisi not read this CL book  
 “Lisi did not read this book.”  
 b. Lisi [zhe ben shu]<sub>i</sub> mei kanguo *e*<sub>i</sub>. (bare preposed Object (SOV order))  
 Lisi this CL book not read

Also consider the cases of *even*-construction in Chinese (see Paris 1979, 1998, 1999; Shyu 1995, 2004; Gao 1994; Tsai 1994; Paul 2005, 2006; Hole 2004 among others). The construction is formed by two elements: *lian* and *dou*. *Lian* is traditionally associated with the meaning of “even” in English. *Lian* precedes the focalized element and its presence is optional (see Badan 2007 for further discussions). The Object preposed by *lian* is given emphasis, “the major stress” (Paris 1979). Literary *dou* means “all” and it must always be present, but is never literally translated in this context<sup>10</sup>. When a sentence contains the *lian...dou* construction, the Object is always obligatorily preposed<sup>11</sup>. A possible landing site is between Subject and Verb:

- (8) a. Wo kanwan [zhe ben shu] le. (unmarked sentence (SVO order))  
 I read this CL book FP  
 “I read this book.”  
 b. Lisi [*lian* zhe ben shu] *dou* yijing kanwan le.  
 Lisi even this CL book all already read FP  
 “Lisi have already read even this book.”

In (8)b *lian* followed by the focalized Object appears on the left of *dou* and the verb. We argue that this order is the effect of an obligatory movement of the phrase *lian*+XP to the left of *dou*<sup>12</sup>. This movement in the *even*-construction is always obligatory:

- (9) \*Wo *dou* kanwan *lian* zhe ben shu le.  
 I all read even this CL book FP

The position of *lian*+XP...*dou* between the Subject and the verb is traditionally defined as a “sentence-internal” position (see (8b)). The whole “sentence-initial” position represents the case where *lian* and the XP move to the Left Periphery, namely to the left of the Subject. *Dou* never moves, but obligatorily stays in its position preceding the verb<sup>13</sup>:

<sup>10</sup> *Dou* is interchangeable with *ye* “also”. Hole (2004) provides evidence for the quasi-fully interchangeability between these two elements; however I will use only on *dou*.

<sup>11</sup> Notice that the Subject can occur in the *lian...dou* construction:

(i) Lian Zhangsan dou kanwan zhe ben shu.  
 Even Zhangsan all read this CL book  
 “Even Zhangsan read this book.”

<sup>12</sup> Notice that this is the same as the movement you see in other sentences with the quantificational *dou* related to an object:

(i) Wo zhe xie shu dou kanwan le.  
 I this CL book all read FP  
 “I read all these books.”

<sup>13</sup> I assume that *dou* have to precede the VP.

- (10) (*Lian*) zhe ben shu, wo *dou* kanwan le.  
 Even this CL book I all read FP  
 “I read even this book.”

In this paper, I concentrate only on *lian*+Object in the sentence-internal position. Considering the sentences above, are we dealing with a Double Topicalization of Subject and Object or with internal Projections? Several previous studies have proposed different analyses for these structures. Xu & Langendoen (1985), Tang (1990), Lin (1992) propose the Double Topicalization Hypothesis (DT). DT consists of two steps: (i) Topicalization of the Object that adjoins to IP; (ii) Topicalization of the Subject across the Object.

Here I reject such a hypothesis and, following Paul (2005), I provide further tests in favor of the idea that the preposed Object in Chinese is located above VP and below IP, in a Low Periphery.

First of all, consider more recent studies that refuse the DT Hypothesis, arguing for two different approaches that support the idea of the existence of a Periphery within the IP: Adjunction (Ernst & Wang 1995, Lu 1994, among others) and Substitution (Qu 1994, Shyu 2001). Both approaches exclude the idea that the Subject moves out of the IP to a Topic position; they propose that the Subject is located in IP and that the landing site for the preposed Object is IP-internal. Ernst & Wang (1995) argue that bare preposed Object undergoes VP (or ModalP)-adjunction and they distinguish it from preposed *lian*-Object<sup>14</sup>. Preposed Object is adjoined to VP with the verb Head bearing [+ Focus] features, while *lian*-Object is raised up to Spec, FunctionalP.

Qu (1994) argues that in Chinese Subject and Object can move covertly or overtly to the Functional AgrSP or AgrOP for features and Case checking. In this way he aims to explain different possible word orders in Mandarin Chinese.

Shyu (1995, 2001) argues that the SOV order is not related to Case checking and that it derives from the Object movement on par with *lian*-Object. Thus she proposes an uniform movement approach, triggered by the [+Focus] feature to a FocusP, which is either covert, in the case of bare preposed Object, or lexically realized, in the case of *lian...dou* structures.

In my paper I adopt Paul’s (2005) analysis on Mandarin Chinese, which applies Belletti’s (2001, 2004) proposal on the Low Periphery (presented here in section 2). As I have illustrated above (section 2), Belletti examines the position between IP and VP occupied by the preposed Object (SOV order) and she argues that it is a clause-internal position. Paul confirms the parallelism between CP and the low IP area. Her final hierarchy for the Low Periphery in Chinese is the following:

- (11) IP > inner TopicP > *even*-Focus > vP

(11) corresponds only partially to the low hierarchy proposed by Belletti (2004); Paul shows that in Chinese no additional TopicP is allowed below *even*-Focus. Such a hierarchy corresponds to the more restricted structure adopted for the external periphery by Benincà (2001) and Benincà & Poletto (2004), excluding TopicP below FocusP, which I also adopt here, as already mentioned.

<sup>14</sup> Lu (1994) also shows a similar VP-adjunction analysis.

### 3.2 Diagnostic tests

With my diagnostic tests I aim at proving the hypothesis that Chinese, like Italian, displays a Low Periphery in the IP area, i.e. below the Subject and above the VP<sup>15</sup>. As mentioned above, Paul (2005) argues for the status of the preposed Object as a clause-internal Topic position. She shows some differences between the internal *versus* the external Topic. For example, only DPs, but no clauses are acceptable in the internal Topic position:

- (12) a. Ta wang le [<sub>s</sub> ji-dianzhong kai hui] (Paul 2005, 55)  
 He forget PERF what time hold meeting  
 “He forgot at what time the meeting is.”  
 b. \*[<sub>IP</sub> Ta [<sub>s</sub> ji dianzhong kai hui wang le]  
 he what time hold meeting forget PERF  
 c. [<sub>TopP</sub> [<sub>s</sub> Ji dianzhong kai hui ] [<sub>IP</sub> ta wang le]], [<sub>TopP</sub> [<sub>s</sub> ji dianzhong chi fan] [<sub>IP</sub> ta mei wang]  
 what time hold meeting he forget PERF what time eat food he  
 not forget  
 “At what time the meeting is, he forgot; at what time the meal is served, he did not forget.”

Moreover Paul shows that multiple topics are allowed in external Topic position, but are excluded for the internal topic position:

- (13) a. \*Ni [<sub>DP</sub> huiyuan dahui] [<sub>DP</sub> mingtian de richeng ] anpai hao le meiyou?  
 you member meeting tomorrow DE program plan finish PERF not  
 b. [<sub>DP</sub> Huiyuan dahui], ni [<sub>DP</sub> mingtian de richeng ] anpai hao le meiyou?  
 member meeting you tomorrow DE program plan finish PERF not  
 “The general membership meeting, have you fixed tomorrow’s program?”  
 (Paul 2005 ex 47)

The following sentences are additional tests of the presence of multiple Topics inside IP:

- (14) a. Hua (a), Zhangsan zui xihuan meiguihua.  
 Flowers TOP Zhangsan most like roses  
 b. Hua (a), Zhangsan [meiguihua] zui xihuan.  
 Flowers TOP Zhangsan roses most like  
 c. Hua (a), meiguihua, Zhangsan zui xihuan.  
 Flowers TOP roses Zhangsan most like  
 d. \*Zhangsan [hua] [meiguihua] zui xihuan.  
 Zhangsan flowers roses most like  
 “Among flowers, I like roses very much.”

In (14)a there is only one Topic in the CP area, (14)b displays a Topic in the Left Periphery and a bare preposed Object; in (14)c there are two high Topics, but in (14)d the sentence is ungrammatical, due to the two bare internal Topics, which are not

<sup>15</sup> Cheng & Downing (2007) show that also in Durban Zulu there are two preverbal Topic positions, one preceding and one following the Subject.

allowed. This shows that the area on the left and on the right have different characteristics.

The subject position can be occupied by an indefinite DP; on the contrary, Topic position cannot: a Topic has to be either definite or generic<sup>16</sup>. In (15) the first DP is clearly indefinite *yi qun* “a couple”, thus it can be analyzed as located in the Subject position, but not in Topic position, which always needs a definite DP.

- (15) Wanshang de shihou wo kandao yi qun ren sha le Lisi de gou.  
 Evening DE when I saw a couple persons kill PERF Lisi DE dog.  
 “During the night I saw that a couple of persons killed Lisi’s dog.”

A further difference between the positions on the left and on the right of the Subject position is evidenced by the presence *versus* the absence of a Topic marker (*a*) following *lian*+XP:

- (16) a. Zhangsan<sub>i</sub>, *lian zhe ben shu* (**a**), ta<sub>i</sub> dou yijing mai le.  
 Zhangsan even this CL book TOP he all already buy FP  
 b. \*Zhangsan<sub>i</sub>, ta<sub>i</sub> *lian zhe ben shu a* dou yijing mai le<sup>17</sup>.  
 Zhangsan he even this CL book TOP all already buy FP  
 c. Zhangsan<sub>i</sub>, ta<sub>i</sub> *lian zhe ben shu* dou yijing mai le.  
 Zhangsan he even this CL book all already buy FP  
 d. \*Zhangsan *lian zhe ben shu a* dou yijing mai le.  
 Zhangsan even this CL book TOP all already buy FP

<sup>16</sup> Huang, A. Li & Y. Li (forthcoming: ch. 7: 3-4): “the Object in the SOV and OSV patterns (preverbal Object) generally does not allow an indefinite non-specific expression; but the Object of SVO (postverbal Object) easily allows it.

- (i) a. wo zai zhao yi ben xiaoshuo.  
 I at seek one CL novel  
 “I am looking for a novel.”  
 b. \*wo yi ben xiaoshuo zai zhao.  
 I one CL novel at seek  
 c. \*yi-ben xiaoshuo, wo zai zhao.  
 one-CL novel I at seek

The use of an indefinite expression *a novel* is not possible preverbally. When a bare nominal appears preverbally, it is generally interpreted as definite.

- (ii) a. shu, wo hui kan.  
 book, I will read  
 “The book(s), I will read.”  
 b. wo shu hui kan.  
 I book will read  
 “I, the book(s), will read.”  
 c. wo hui kan shu.  
 I will read book  
 “I will read books.”

(ii a-b) contrast with (iic). Only the latter allows the Object *shu* ‘book’ to be interpreted as indefinite...If an expression denotes quantity, such as ‘a novel’ below, it is possible in the preverbal position:

- (iii) Yi ben xiaoshuo, ta yi ge wanshang jiu kan wan le.  
 One CL novel he one CL evening then read finish FP  
 “A novel, he finished reading in an evening.”  
 (iv) Ta yi ben xiaoshuo yi ge wanshang jiu kan wan le.  
 he one CL novel one CL evening then read finish FP  
 “He, a novel, finished reading in an evening.”

<sup>17</sup> Notice that the sentences (16b and d) are acceptable only with a comma or a pause after the Topic particle *a*, but this indicates a completely different structure.

(16)a shows *lian*+XP in initial-position, on the left of Subject, that may be followed by the Topic marker; in (16)b *lian*+XP is in clause-internal position, thus the Topic marker is not allowed; (16)c is perfectly grammatical, since the *lian*+XP is in low position, but without Topic marker; finally (16)d shows that *lian*+XP cannot be followed by a Topic marker, this means that it is located in sentence-internal position, thus *Zhangsan* is in Subject position within IP and it is not topicalized to the CP area (as, on the contrary, the Double Topicalization Hypothesis predicts).

Now consider the structure of the Left Periphery in Chinese sketched by Paul (2005) and Badan & Del Gobbo (in press). They show that *lian*+XP always occupies the lowest position of the Left Periphery, i.e. below (different kinds of) Topics and above Subject:

(17) [<sub>CP</sub> Topics > *lian*+XP] > [<sub>IP</sub> Subject...]

Thus consider the following sentence displaying *lian*+XP on the left of a co-indexed resumptive pronoun *ta* “him//her”:

(18) Lian Zhangsan<sub>i</sub>, ta<sub>i</sub> zhe ben shu dou yijing kanwan le.  
Lian Zhangsan he this CL book all already read FP  
“Even Zhangsan, he read this book.”

Following the idea that *lian*+XP occupies the lowest position of the CP and cannot be followed by other Topic or Focus projections, the resumptive pronoun *ta* “him” cannot be considered in a Topic position in the Left Periphery, but only in the Subject position within IP.

On the basis of the tests above, I argue that the bare preposed Object and sentence-internal *lian*+XP are located in a Low Periphery below IP and above VP, parallel to the Left Periphery in the CP area.

#### **4. Preposed Object (SOV) and sentence-internal *lian*+XP**

Shyu (1995, 2001) proposes a uniform Object movement approach for both bare preposed Objects and sentence-internal *lian*+XP. She analyzes them as derived by a substitution mechanism, triggered by the [+Focus] feature, which is either phonologically null or lexically realized in *dou*-sentences or *lian...dou* structures. Remember that she considers *dou* the Head of the FocusP that can be overtly expressed (in the case of *lian*+XP) or covert (in the case of the preposed Object). As I mentioned earlier, I do not consider *dou* as Head of FocusP and following Paul (2002, 2005), I analyze the SOV and *lian*+XP as two different items that have moved up into two different landing sites, as they have two different semantic/pragmatic interpretations.

##### *4.1. Two different positions*

Paul (2002) suggests that the bare preposed Object SOV is higher than the *lian*+XP in the Low Periphery. With the following tests I show that SOV and the sentence-internal *lian*+XP cannot be analyzed in a uniform way: they occupy two distinct positions in the Low periphery, corresponding to two different Functional Projections, and the former is higher than the latter.

1. The preposed Object must precede the Aspectual (repetitive) adverbs<sup>18</sup> like *you* “again”, while *lian*+XP must follow it.

- (20) a. Ta (\**you*) [nei ben shu] *you* kan le yibian. (Paul 2002: 22 a-b)  
 He again that CL book again read PERF once  
 “He has read that book one more time.”  
 b. Wo *you* [lian yi fen qian ye] dou mei you le.  
 I again even one CL money also all not have FP  
 “Once again I don’t have a cent.”

2. SOV order and sentence-internal *lian*+XP can co-occur; the resumptive pronoun in Subject position shows that we are dealing with the Low Periphery and two different internal Projections.

- (21) Zhangsan<sub>1</sub>, ta<sub>1</sub> [zhe ge tang] *lian* wo de xiaohaizi dou song le!<sup>19</sup>  
 Zhangsan he this CL sweet even I DE children all give FP  
 “As for Zhangsan, he gave the sweets even to my children!”

- (22) [<sub>IP</sub> Lisi, [<sub>ta<sub>i</sub></sub> [<sub>int.TopP</sub> yingyu [<sub>FocP</sub> *lian* liushi fen [<sub>VP</sub> dou mei nadao ]]]]  
 Lisi he English even 60 point all not obtain  
 “Lisi didn’t even obtain 60 points in English.”  
 (Paul 2006: 60)

If sentence-internal *lian*+XP is in a higher position with respect to the bare preposed Object, the clause is ungrammatical (see also Paul 2002, 2005):

- (23)\*Zhangsan<sub>i</sub>, ta<sub>i</sub> *lian* wo de xiaohaizi dou [zhe tang] *gei* le!  
 Zhangsan he even I DE children all this sweet give FP

- (24)\*[<sub>IP</sub> Lisi<sub>i</sub>, [<sub>ta<sub>i</sub></sub> [<sub>FocP</sub> *lian* liushi fen [<sub>intTopP</sub> yingyu [<sub>VP</sub> dou mei nadao ]]]]  
 Lisi he even 60 point English all not obtain

3. Furthermore, another issue to defend the idea that the bare preposed Object occupies a different position from sentence-internal *lian*+XP is the fact that the SOV can be followed by a Topic marker (25)a, while *lian*+XP cannot (25)b. Notice that in order for (25)a to be acceptable, the preposed Object must be stressed.

- (25) a. Zhangsan<sub>1</sub>, ta<sub>1</sub> [zhe ben shu] **a** yijing kan wan le.  
 Zhangsan he this CL book TOP already read finish FP  
 “As for Zhangsan, he already read this book.”  
 b. \*Zhangsan<sub>1</sub>, ta<sub>1</sub> [lian zhe ben shu] **a** dou yijing kan wan le.  
 Zhangsan he even this CL book TOP all already read finish FP

4. The bare preposed Object displays a characteristic proper of a Topic-like item in Chinese: it cannot be indefinite, while the element following sentence-internal *lian* may be:

<sup>18</sup> These kinds of adverbs are in low positions in Cinque’s (1999) hierarchy. Traditionally they are called “VP adverbs”.

<sup>19</sup> I owe this example to Lisa Cheng.

- (26) a. \*Ta [yixie jiu xinfeng] baocun zhe. (Zhang 1996: 15-16)  
He some old envelope kept FP  
“He has kept some old envelopes.”  
b. Ta [lian yixie jiu xinfeng] dou baocun zhe.  
He even some old envelope all kept FP  
“He has kept even some old envelopes.”

5. A bare pronoun can be preposed within *lian...dou* construction, while without any marking it cannot (Paul 2002):

- (27) a. Zhangsan [lian wo] ye piping le.  
Zhangsan even I also criticize FP  
“Zhangsan criticized even me.”  
b. \*Zhangsan [wo] piping le.  
Zhangsan I criticize FP  
“Zhangsan criticized me.”

6. A bare preposed Object cannot be in a cleft configuration by means of *shi...de* (see Paul & Whitman 2001), which is different from *lian+XP* constituent:

- (28) a. \* Zhangsan *shi* [zhe ben shu] kanwan *de*.  
Zhangsan SHI this CL book read ...DE  
Lit: “Zhangsan, it’s this book (that) he read.”  
b. Zhangsan *shi* [lian zhe ben shu] dou kanwan *de*.  
Zhangsan SHI even this CL book all read ...DE  
“It’s even this book that Zhangsan read.”

Through the tests above I provide evidence for the following facts: the bare preposed Object above VP and the preposed *lian+XP* are not the same kind of element. They occupy two different Functional Projections: they display distinct behaviours with respect to some adverbs, the presence of the Topic marker, the possibility to be in a cleft sentence. Moreover, they can co-occur and the bare preposed Object has to be placed in a higher position with respect to *lian+XP*.

#### 4.2. A-movements

It is generally assumed that the SOV and the sentence-internal *lian+XP* are derived by movement. (29)a, a case with a bare preposed object and (29)b, a sentence-internal *lian+XP*, are grammatical only with a gap in the object (base-generated) position. Thus, on the basis of what I said for Topics, I argue that both structures are derived by movement.

- (29) a. Zhe zhi gou [ziji de zhuren]<sub>i</sub> yao le (\*ta<sub>i</sub>), bieren que bu yao. (Shyu 2001: 50)  
this CL dog self DE master bite PERF him, others but not bite  
“This dog bit its own master, but not others.”  
b. Zhe zhi gou [lian ziji de zhuren]<sub>i</sub> dou yao le (\*ta<sub>i</sub>), bieren que bu yao.  
this CL dog even self DE master all bite PERF him others but not bite  
“This dog bit even its master, but not others.”

It seems that the empty element on the right of the verb is A-bound, since the two movements display several A-properties (see Fu 1994; Qu 1994; Ting 1995; Shyu

1995, 2001; Zhang 1996). I consider the landing-site for sentence-internal *lian+XP* as a Focus position derived by A-movement. A-chain Focalization is not a new idea; as Belletti & Shlonsky (1995) point out, Focalization is not a unitary phenomenon. In Italian and Hebrew it seems to be (in part)<sup>20</sup> part of the A- and not the Abar-system; on the contrary, Focus in Hungarian involves an Abar-chain. In this section I show the A-properties of SOV and sentence-internal *lian+XP*: clause-boundness, absence of Reconstruction for Principle C, absence of resumption.

### 1. Clause-boundness.<sup>21,22</sup>

The embedded Object cannot be preposed across a tensed clause boundary to matrix post-Subject/ pre-Verb position (Focus is Subject only to local movement):

(30) \*Zhangsan pingguo<sub>i</sub> zhidao [<sub>CP</sub>Lisi chidiao le  $e_i$ ] (Ting 1995: 7)  
 Zhangsan apple know Lisi ate FP  
 “Zhangsan knows that Lisi ate the apples”

(31) a. Zhangsan renwei [<sub>CP</sub>Lisi hen xihuan Mali] (Shyu 2001: 3-4)  
 Zhangsan think Lisi very like Mali  
 “Zhangsan thinks that Lisi likes Mali.”

b. \* Zhangsan Mali<sub>i</sub> renwei [<sub>CP</sub> Lisi hen xihuan  $t_i$ ].  
 Zhangsan Mali think Lisi very like  
 “Zhangsan thinks that Lisi likes Mali.”

I can refer to this phenomenon as *adjacency requirement*, following Belletti & Shlonsky (1995: 501), who show that in Italian (and in Hebrew) the postverbal Subject (in Spec, FocusP) is more acceptable when it is adjacent to the verb<sup>23</sup>.

Notice, on the contrary, that OSV word order displays long-distance dependency:

(32) Pingguo, Zhangsan zhidao [<sub>CP</sub>Lisi chidiao le  $e$ ]. (Ting 1995:6)  
 apple Zhangsan know Lisi ate FP

(33) Mali<sub>i</sub>, Zhangsan renwei [<sub>CP</sub>Lisi hen xihuan  $e_i$ ]  
 Mali Zhangsan think Lisi very like

Sentence-internal *lian+XP* (34)a *versus* sentence-external *lian+XP* (34)b:

<sup>20</sup> In Italian there is also the focalization to the left periphery through an Abar movement.

<sup>21</sup> “It has been often observed when A-movement applies, for example, in the case of super-raising:

(i) \*John<sub>i</sub> seems [that it is likely [ $t_i$  to win]

The NP John raises across a tensed clause boundary and the sentence is ungrammatical. On the other hand, Abar- movement can freely take place out of a tensed clause, if no barrier is crossed:

(ii) What<sub>i</sub> do you think [that John fixed  $t_i$ ].

(Ting, 1995: 292).

<sup>22</sup> See Fu (1994), Qu (1994), Shyu (2001).

<sup>23</sup> The examples analyzed by Belletti & Shlonsky (1995) for Italian are the followings:

(i) a. ?Ha dato un libro a Maria Gianni.

has given a book to Maria Gianni.

b. \*(?)Ha dato a Maria un libro Gianni.

has given to Maria a book Gianni.

c. ?Ha messo il libro sul tavolo Maria.

has put the book on-the table Maria

d. \*(?)Ha messo sul tavolo il libro Maria.

has put on-the table the book Maria.

e. \*?Ha dato a Maria Gianni un libro.

has given to Maria Gianni a book.

- (34) a. \*Zhangsan *lian* Mali<sub>i</sub> renwei [<sub>CP</sub>Lisi *dou* bu xihuan *e<sub>i</sub>*]. (Shyu 2001: 3-5)  
 Zhangsan even Mali think Lisi all not like  
 b. *Lian* Mali<sub>i</sub>, Zhangsan renwei [<sub>CP</sub>Lisi *dou* bu xihuan *e<sub>i</sub>*].  
 Even Mali Zhangsan think Lisi all not like  
 “Zhangsan thinks that Lisi doesn’t like even Mali.”

2. No Reconstruction effects for Principle C of the Binding Theory.

“Though coreference between the pronoun *ta* and its antecedent *Zhangsan* in sentence (35) impossible, it becomes possible when the indirect Object containing *Zhangsan* has undergone bare Object Movement (in (36)a) and Focalization (in (36)b) (Shyu 2001).

- (35) \*Wo bei ta<sub>i</sub> qiang-zou le [yi ben Zhangsan<sub>i</sub> de shu]. (Shyu 2001: 4)  
 I by him rob-away PERF one CL Zhangsan DE book  
 Lit. ‘I was robbed by him<sub>i</sub> of a book of Zhangsan<sub>i</sub>.’

- (36) a. Wo [Zhangsan<sub>i</sub> de shu]<sub>j</sub> jiao ta<sub>i</sub> na-zou le *e<sub>j</sub>* (Shyu 1995:105, 83)  
 I Zhangsan DE book let him take-away FP  
 “I asked him to take away Zhangsan’s books.”  
 b.? Wo *lian* [Zhangsan<sub>i</sub> de shu]<sub>j</sub> dou bei ta<sub>i</sub> qiang-zou le *e<sub>j</sub>*  
 I even Zhangsan DE book all by him rob-away FP  
 “I was robbed of [even Zhangsan<sub>i</sub>’s book] by him<sub>i</sub>.”

3. No resumption.

“It is generally assumed that the gap left by A-movement cannot be filled with an overt pronominal” (Ting 1995: 295).

- (37) \*Lisi [nei ge ren<sub>i</sub>] ji bu de ta<sub>i</sub> le. (Ting 1995: 17 s. m.) (SOV)  
 Lisi that CL person remember not be-able him FP  
 Lit: “Lisi that person cannot remember *her/him*.”  
 (38) \*Lisi [*lian* Mali]<sub>i</sub> dou hen xihuan ta<sub>i</sub>. (Sentence-internal *lian*+XP)  
 Lisi even Mali all very like him  
 Lit: “Lisi even Mali likes very much her.”

Could the impossibility of the presence of the resumptive pronoun be derived from the violation of Principle B? Consider the following examples:

- (39)\*Wo [nei ge ren]<sub>i</sub> renwei Lisi genben ji bu de ta<sub>i</sub> le.(Ting 1995: 17)  
 I that CL person think Lisi totally remember not be-able him FP  
 Lit: “I that person think Lisi totally can’t remember him.”

- (40) \*Zhangsan *lian* Mali<sub>i</sub> ***dou*** renwei [<sub>CP</sub> Lisi hen xihuan (ta<sub>i</sub>)]. (Shyu 1995)  
 Zhangsan even Mali all think Lisi very like (her)  
 “Zhangsan thinks that Lisi likes even Mali.”

The ungrammaticality of (39) and (40) indicate that SOV and sentence-internal *lian*+XP are clause-bound, which is considered a property of A-movement<sup>24</sup>. Ting (1995): “the ungrammaticality of (39) can no longer be attributed to the binding condition B, since the binding domain for the pronominal *ta* ‘he’ is free in the embedded clause, satisfying the binding condition B, so there must be some other reasons for the ill-formedness of (39). Given the A-movement analysis, the ungrammaticality of (40) naturally follows, since it is generally assumed that the gap left by A-movement can not be filled with an overt pronominal.”

As Ernst & Wang (1995) point out, the only case in which a bare preposed Object or *lian*+XP merged in the embedded clause has the position between the Subject and the matrix verb as its landing site, is when the Object is preposed from a nonfinite embedded Object position: “it is well known that nonfinite complements are Subject to clause union phenomena, in which matrix and embedded complement together display some properties of a single clause” (Ernst & Wang 1995: 245). Shyu (2001: fn27) shows that also with infinitive the resumption is still not allowed:

- (41) Lisi bi [<sub>IP</sub> Zhangsan ma Mali] (Shyu 2001: fn 27) (base sent.)  
 Lisi force Zhangsan scold Mali  
 “Lisi forces Zhangsan to scold Mali.”
- (42) Lisi Mali<sub>i</sub> bi [<sub>IP</sub> Zhangsan ma (\*ta<sub>i</sub>)] (bare preposed Object)  
 Lisi Mali force Zhangsan scold her  
 Lit: “Lisi Mali forces Zhangsan to scold her.”
- (43) Lisi *lian* Mali<sub>i</sub> dou bi [<sub>IP</sub> Zhangsan ma (\*ta<sub>i</sub>)] (Sentence-internal *lian*+XP)  
 Lisi even Mali all force Zhangsan scold her

Thus SOV and sentence-internal *lian*+XP are A-moved. Consider that their movements also display Abar-properties: the site from which the XP moves is a position to which Case is assigned. I assume that Object Case is checked by verb government (Ernst 1998). On the contrary, A-movement forms a chain between the original position which is assigned a  $\theta$ -role, but not Case. The landing site is a position where no Case and no  $\theta$ -role are assigned.

Following Shyu (2001), I argue that the bare preposed Object/*lian*+XP-movements must have a sort of trigger, rather than Case assignment. The bare preposed Object is attracted by “selected” properties, following the Spec-Head checking relation within the maximal Projection of a FP. As I will show in the following section, the preposed object yields a contrastive Topic reading. Thus I propose that the bare object within IP moves up to check its Topic feature, in a Spec-head agreement configuration. I do not need to stipulate the optional Case checking for Chinese<sup>25</sup>.

<sup>24</sup> Notice that both of them can stay in embedded position, for instance in relative clauses:

(i) Qing zai [[ta **nei ben shu** kanwan] de shihou] (Ernst & Wang 1995: 29)

please at he that CL book read of time

“Please come see him when that book, he finishes reading.”

<sup>25</sup> Qu (1994) proposes Functional AgrPs to derive Subject and Object Case agreement in Chinese. Shyu (2001) argues that SOV is not triggered by Case assignment nor is Case related. She assumes that a Subject is base-generated in the Spec, VP position, following the Internal Subject Hypothesis (Kuroda 1988; Koopman & Sportiche 1990). She assumes that Subject raising to [Spec, IP] is obligatory, even though INFL is defective in Chinese. This Subject raising is for assigning abstract nominative Case. As I mentioned earlier, Object abstract Case is checked by Verb government.

### 5. Bare Preposed Object (SOV): Topic or Focus?

In this section I concentrate on the syntactic properties of the SOV in the Low Periphery. The SOV shows clear Topic-like properties: presence of Topic markers, impossibility to be clefted by means of *shi...de* “be...DE”, co-occurrence with a Focus *in situ*, definiteness requirement. From a pragmatic/semantic point of view, SOV requires a contrastive reading, i.e. it is always an emphasized element in the sentence. As mentioned earlier, the contrastive stress does not indicate by itself that an item is focalized, thus I can argue that the Chinese bare preposed Object moves up to the Low Periphery in order to occupy the Spec of a Contrastive Topic Projection. At a first sight SOV seems to be a focalized item, since, as I will illustrate below, it generally needs a context in which it gets emphasis. Indeed, in the literature it is generally assumed to involve Focalization (Ernst & Wang 1995; Shyu 1995, 2001; Tsai 1994; Zhang 1996), even if the role of such an emphasis is not always clear. Actually, from a syntactic point of view it displays only two Focus properties, while most of its characteristics are typical of Topic-like elements<sup>26</sup>.

Focus properties:

1. The resumptive pronoun is not allowed. This fact indicates that the SOV is subject to an A-type movement (see section 4) and not to the typical Topicalization Abar-movement.

(44) \*Zhangsan **Mali**<sub>i</sub> hen xihuan ta<sub>i</sub>.  
Zhangsan Mali very like her

2. SOV cannot be multiple. The impossibility to be multiple can be derived from the fact that the Low Periphery seems to be “more restricted” than the CP area, thus it does not admit more than one Topic.

Notice that SOV can co-occur with sentence-internal *lian*+XP. As mentioned above, multiple Foci are not allowed, thus: (i) one of them is a Focus and the other is a Topic; (ii) none of them is a Focus. Furthermore, when they co-occur, the main stress is on *lian*+XP and not on the bare preposed Object.

---

<sup>26</sup> First of all, if I follow Rizzi’s (1997) tests in order to distinguish Topic from Focus, I have to take into consideration also the WCO, as I do for the elements in the CP area. In the case of the SOV, the results are not so clear. Qu (1994) and Shyu (1995) have both noted that Chinese SOV does not show WCO effects. SOV in the Low Periphery can be coreferent to the corresponding pronoun *ta*:

(i) Wo mei ge haizi dou bei [youguai ta<sub>i</sub> de ren] pian-zou le e<sub>i</sub>.

I every CL child all by abduct him DE person kidnap-away FP

Lit.: “I was affected by every<sub>i</sub> child being kidnapped by the person who abducted him<sub>i</sub>.”

(Shyu 1995: 105, 84)

However, the result is not so clear: my Chinese informants have too many dissenting opinions about the grammaticality of the sentences showing SOV within WCO structure. See, for instance, another clause displaying WCO context, the result is ungrammatical:

(ii) \*Zhangsan [Mali<sub>i</sub>] zai ta<sub>i</sub> de jia jiandao le.

Zhangsan Mali in her DE home met FP

“Zhangsan met MALI at her home.”

I think that the non-conforming judgments are probably due to some phenomena that interact with each other, thus they cannot be used as a valid WCO test in order to distinguish Topic from Focus.

- (45) Zhangsan zhe zhong tang lian WO DE XIAOHAIZI dou song le...  
 Zhangsan this CL sweet even I DE child all give FP  
 “Zhangsan gave this sweet even to my child.”

Most of the properties of the SOV are Topic-like.

Topic properties:

1. Compatibility with a *wh*-element. Bare preposed Object does not interact with the *wh*-item.

- (46) Zhangsan [zhe ben shu] huan gei le shei?  
 Zhangsan this CL book give-back to PERF who  
 Lit. “Zhangsan gave back this book to whom?”

On the contrary the focalized item *lian*+XP interacts with a *wh*:

- (47) \*Zhangsan lian zhe ben shu dou huan gei le shei?  
 Zhangsan even this CL book all give-back to PERF who  
 Lit. “Zhangsan gave back this book even to whom?”

2. The preposed Object can be followed by Topic markers.

- (48) Zhangsan [zhe ben shu] (a) yijing mai le.  
 Zhangsan this CL book TOP already buy FP  
 “Zhangsan this book already bought.”

On the contrary, as showed in (16), the focalized item *lian*+XP cannot be followed by a Topic marker *a*:

- (49) \*Zhangsan lian zhe ben shu a yijing dou mai le<sup>27</sup>.  
 Zhangsan even this CL book PART already all buy FP

3. Bare preposed Object cannot be clefted by means of *shi...de* pattern, which would, however, be expected if it were really a Focus (Paul & Whitman 2001).

- (49) a. Women [gugong] qu guo le. (Paul 2002: 21)  
 We imperial-palace go EXP FP  
 “We have been to the imperial palace.”  
 b. \*Women *shi* [gugong] qu guo *de*.  
 We SHI imperial-palace go EXP ...DE

- (50) \*Zhangsan shi [zhe ben shu] kanwan de.  
 Zhangsan SHI this CL book read ...DE  
 “It’s this book that Zhangsan read.”

4. It can co-occur with a Focus *in situ*. Having in mind the impossibility of multiple Foci, it derives that the Object in a SOV sentence is not a Focus.

---

<sup>27</sup> This sentence is acceptable only with a comma or a pause after the Topic marker *a*.

- (51) Mali [zhe ben shu] huan gei LISI (bu gei Zhangsan)  
Mali this CL book give-back to Lisi not to Zhangsan  
Lit: “Mali, this book, gave back to Lisi (not to Zhangsan!).”

5. Like the topicalized elements in the CP area (OSV), bare preposed Object generally cannot be an indefinite non specific expression.

- (52)a. Shu, wo hui kan. (Huang, A. Li & Y. Li forth.: 16) (Topic: OSV)  
Book I can read  
“THE books, I will read.”  
b. Wo shu hui kan. (preposed Object: SOV)  
I book can read  
“I THE books will read”  
c. Wo hui kan shu. (canonical word order: SVO)  
I can read book  
“I will read (some) BOOKS”

Shyu (2001: 16) claims that, differently from a Topic in the CP area, a bare preposed Object in the IP can be indefinite. In order to indicate indefiniteness, she uses the numeral *yi* “one” (followed by the Classifier). Yet notice that an element introduced by the numeral *yi* “one” in Topic position and in sentence-internal position (the preposed Object position) is acceptable only if it is contrasted with another numeral item (53b). This means that in Topic position its interpretation is always definite:

- (53) a. \*Yi pian lunwen, wo hen xihuan. (Tsai 1994: 31) (Topic: OSV)  
one CL paper I very like  
“A paper I like very much.”  
b. [Yi pian lunwen], wo hai keyi yingfu, [liang pian na] jiu tai duo le.  
One CL paper I still can handle two CL that then too much FP  
“One paper, I can handle, but two papers, that’s too much.”

With the preposed Object, the contrastive construal of the sentence is obligatory, i.e. the clause with a preposed Object requires a conjunct with which to put it in contrast:

- (54) Wo yi pian lunwen keyi yingfu \*(lian pian jiu bu xing le). (Tsai 1994: 32)  
I one CL article can handle two CL then not possible FP  
“A paper, I can handle (but two papers, I can’t).”

### 5.1. SOV: semantics / pragmatics

As mentioned earlier, Chinese Object preposing (SOV) is commonly assumed to involve Focalization (Ernst & Wang 1995; Shyu 1995, 2001; Tsai 1994; Zhang 1996). It normally has an emphatic function, but such an emphatic effect is not always clear. Some linguists have doubts about its Focus function and propose to treat it as a kind of Topic endowed with some Focus properties. For instance, Ernst & Wang (1995) show the pragmatic differences between the Topic in initial position (OSV), which they call “discourse Topic”, and the preposed Object (SOV), called “Focus Topic”. Ting (1995), borrowing the term introduced by Tsao (1997) for the *ba*-NP<sup>28</sup>, defines the

<sup>28</sup> In Chinese the direct Object moved to a preverbal position can be preceded, obligatorily or optionally, by the morpheme *ba*. The exact function of *ba* is a widely discussed topic among linguists: it is treated either as a verb (Hashimoto 1971), a preposition (Travis 1984, Li 2001) or as a Case marker

bare preposed Object as a “secondary Topic”, in opposition to the “primary Topic” OSV, i.e. a Topic in the CP area, and Paul (2002, 2005) analyzes it as a sentence-internal Topic preceding the Focus position occupied by *lian*+XP. Following the authors cited above, I adopt the proposal that Chinese bare preposed Object occupies the Spec of a Topic position, more precisely of a Contrastive Topic position.

First of all, there is a different pragmatic (and syntactic) requirement connecting sentence-initial Topic and the preposed Object in the IP (Ernst & Wang 1995; Tsai 1994; Huang, A. Li & Y. Li forthcoming among others).

The Object in SOV clauses must display some sort of contrastive reading, while the Object in OSV clauses does not need to, though it may be contrastive<sup>29</sup>:

(55) a. [Zoumingqu], Zhangsan hen xihuan tan, dajia ye hen xihuan ting.  
 Sonata Zhangsan very like play all also very like listen  
 “As for sonatas, Zhangsan likes to play it and everyone also likes to listen to it very much.”

b. (Wo dui lanqiu hen shou, danshi) [zuqiu], wo yi qiao bu tong.  
 I to basketball very familiar but soccer I one intelligence not understand  
 “I’m familiar with basketball, but soccer, I have no idea at all.”  
 (Ting 1995: 3)

Which kind of contrast does bare preposed Object in the IP imply? The following diagnostic tests show that it can be semantically/pragmatically considered neither a Contrastive Focus nor an Informational Focus.

1. Bare preposed Object in the IP area is not an Informational Focus. The reply to a *wh*- question implies new information, i.e. Informational Focus:

(56) Q: Zhangsan mai le shenme?  
 Zhangsan buy PERF what  
 “What did Zhangsan buy?”  
 A1: Zhangsan mai le [zhe ben shu]. (SVO)  
 Zhangsan buy PERF this CL book  
 A2: \*[Zhe ben shu], Zhangsan mai le. (\*OSV)  
 This CL book Zhangsan buy FP  
 A3: \*Zhangsan [zhe ben shu] mai le. (\*SOV)  
 Zhangsan this CL book buy FP  
 “Zhangsan bought this book.”

Only the answer (56)A1 is acceptable; its word order is unmarked and – as have already seen before- that the Informational Focus in Chinese is realized *in situ*<sup>30</sup>. In

---

(Huang 1982, Goodall 1987) or as a higher verbal Head by Paul & Whitman (2005). For an analysis of functions and optionality/obligatoriness of *ba* see also Li (2006) and van Bergen (2006).

<sup>29</sup> Shyu (1995) makes a structural distinction between “Focused” OSV and unmarked OSV. The former is in IP-adjoined position, while the latter occupies the Spec, TopicP. I do not agree with this proposal, but, as I have shown, I propose that every kind of Topic in the CP area can optionally have a contrastive reading.

<sup>30</sup> A reviewer suggests that we could consider the new informational focus object as not located *in situ* position, but moved up to a focus projection of the low periphery. As a consequence, in order to obtain the surface order SVO, we should argue for the subsequent remnant movement of the VP containing the verb and the trace of the object. The landing site position of the of the VP should be the topic position higher than the focus in the low periphery. In the case of the object marked with *lian*, the remnant

contrast, neither (56)A2 nor (56)A3 is a proper answer. The former displays an element in sentence-initial position that cannot function as an Informational Focus, the latter is a case of Object preposing, which cannot be used as an Informational Focus<sup>31</sup> either.

2. Bare preposed Object in the IP area is not a Contrastive Focus. Considering that the bare preposed Object is pragmatically/semantically defined as a Focus-Topic, i.e. a Topic with a Contrastive reading, the next test aims at checking if it can be used as a Contrastive Focus. With Contrastive Focus I mean a stressed item that makes a correction to an information/assertion<sup>32</sup>.

- (57) Q: Zhangsan mai le zhe zhang chuang ma?  
 Zhangsan buy PERF this CL bed FP  
 “Zhangsan bought this bed?” (for his new room?)  
 A1: Bu shi, Zhangsan mai le ZHE ZHANG ZHUOZI. (Focus *in situ*)  
 Not be Zhangsan buy PERF this CL table  
 A2: \* Bu shi, ZHE ZHANG ZHUOZI Zhangsan mai le. (\*OSV)  
 Not be this CL table Zhangsan buy FP  
 A3: \* Bu shi, Zhangsan ZHE ZHANG ZHUOZI mai le. (\*SOV)  
 Not be Zhangsan this CL table buy FP  
 “No, Zhangsan bought this table!”

Compare (57) with Italian sentences:

- (58) Q: Per la sua nuova camera, Gianni ha comprato il letto?  
 For the his new room Gianni has bought the bed  
 “For his new room, did Gianni buy the bed?”  
 A: No, Gianni ha comprato IL TAVOLO! (Focus *in situ*)  
 No Gianni has bought the table  
 “No, Gianni bought THE TABLE!”  
 A1: No, IL TAVOLO Gianni ha comprato. (OSV)  
 No the table Gianni has bought  
 “No, THE TABLE Gianni bought.”

Chinese SOV cannot be defined as a Contrastive Focus since it cannot be used as a correction, even if it bears a sort of “Focus” stress.

---

movement does not take place, yielding the surface order S lian+XP V. It is an interesting idea which requires further work.

<sup>31</sup> Notice that OSV, generally being a Topic without a special stress, should be possible in an answer to a question in which it has been previously mentioned, while in this case SOV is infelicitous:

- (i) Q: Shei mai le zhe ben shu?  
 Who buy PERF this CL book  
 “Who bought this book?”  
 A1: [Zhe ben shu], Zhangsan mai le.  
 This CL book Zhangsan buy FP  
 A2:?? Zhangsan [zhe ben shu] mai le.  
 Zhangsan this CL book buy FP  
 “As for this book, Zhangsan bought.”

In (A1) *zhe ben shu* “this book” is in an external Topic position and the sentence stress has to be on the Subject *Zhangsan*, since it is the Informational Focus of the clause. In (A2) the preposed Object needs a contrastive reading that in this case is infelicitous.

<sup>32</sup> In Chinese the Contrastive Focus cannot (overtly) move up to the Left Periphery, and it is always *in situ* (see Gao 1994, Badan 2007, Badan & Del Gobbo in press).

In summary, we can consider the preposed Object as neither an Info Focus nor a Contrastive Focus.

I noticed that, in every proposal regarding the contrastive stress given to the SOV, it is implied that the sentences in which such SOV appears always require a contrasted context of some sort. I would say that the SOV must be in comparison with two or more items of a set, as a contrasted element in a list. This kind of Topic appearing in analogous contexts in Italian is called List Interpretation Topic by Benincà & Poletto (2004), and more traditionally, Contrastive Topic.

When SOV appears in a simple sentence, this is interpreted as an “open sentence”, i.e. a sentence that implies a conjunction or a contrast, either overtly expressed or not.

(59) Ta yingwen bao kan de dong, danshi dewen bao kan bu dong<sup>33</sup>.

He English newspaper read be-able understand but German newspaper read not understand

“He can read English newspapers, not the German ones.”

(60) Wo zhe pian lunwen xihuan \*(na pian lunwen bu xihuan). (Tsai 1994: 32)

I this CL paper like that CL paper not like

“This paper, I like (but that paper I don’t).”

Compare OSV with SOV: (61)a with the external Object is felicitous on its own, while the simple sentence (61)b containing a SOV cannot be pronounced out of the blue, but it requires a contrastive context or a conjunction (for instance that one in brackets).

(61) a. Yu a, Zhangsan gan chi. (Shyu 2001: 43-44) (OSV)  
fish TOP Zhangsan dare eat

“As for fish, Zhangsan dares to eat.”

b. Zhangsan [yu] gan chi, ([niurou] bu gan chi). (SOV)

Zhangsan fish dare eat beef not dare eat

“Zhangsan dares to eat fish, but wouldn’t dare to eat beef.”

Ernst & Wang (1995: 22) point out that (62)a requires a strong stress on the SOV or the use of the parenthesized clause. On the contrary, (62)b does not need any special stress on the SOV or any kind of contrast in order to be grammatical.

(62) a. Wo [jiu] he (kele bu he). (Ernst & Wang 1995: 22)

I liquor drink Coke not drink

“Liquor I drink (but Coke I don’t drink).”

b. [Jiu], wo he.

Liquor I drink

“(As for) liquor, I drink.”

Other examples are from Shyu (2001): (63)a with an intonationally unmarked external Topic is perfectly grammatical; on the contrary, (63)b is infelicitous if uttered

---

<sup>33</sup> From Abbiati (1998: 164 slightly modified).

out of the blue, but it is improved when uttered in a contrastive context: *yidaliwen* “Italian” is compared with *ladinwen* “Latin”<sup>34</sup>.

- (63) a. [Yidaliwen], geju yanyuan zhidao. (Shyu 2001: 40)  
Italian opera performer know  
“Italian, opera performers know.”  
b. # Geju yanyuan [yidaliwen] zhidao  
opera performer Italian know  
“Opera performers Italian, know.”  
c. Geju yanyuan [yidaliwen] zhidao, (danshi) [ladinwen] jiu bu dong le  
opera performer Italian know but Latin then not understand FP  
“Opera performers know Italian, but they don’t understand Latin.”

Actually, it is possible that a SOV can appear in a sentence without any strong stress, but in that case an emphatic element is obligatorily required, for instance the negation *bu* “not” or the adverb *ye* “also” (Ernst & Wang 1995):

- (64) Wo [jiu] bu he le<sup>35</sup>. (Ernst & Wang 1995: 1)  
I liquor not drink FP  
“I won’t drink liquor any more.”

- (65) Wo wenti hai mei xiangqing chu lai, bu neng wen ni. (Shyu 2001: 30)  
I question still not think go-out come not can ask you  
“I haven’t come up with questions, so I cannot ask you.”

Moreover, Ting (1995) points out that Focus interpretation of the SOV is not the only interpretation available, if there is a “real Focus present in the sentence”:

- (66) Q: Zhangsan zui xihuan zai nali chi pingguo? (Ting 1995: 5)  
Zhangsan most like at where eat apple  
“Where does Zhangsan like to eat apples most?”  
A: Zhangsan [pingguo] zui xihuan ZAI CHUANG SHANG chi.  
Zhangsan apple most like at bed on eat  
“Zhangsan as for apples likes to eat AT BED most.”

In this case the Focus in the clause is *zai chuanshang* “at bed”, which constitutes the Info Focus (the answer to the *wh*-question), while the SOV is simply a piece of old information, already mentioned in the question.

The last case in which SOV seems to lose its strong stress is when it co-occurs with the *lian*+XP:

- (67) Zhangsan [zhe ge tang] lian (gei) wo de xiaohaizi dou song le...  
Zhangsan this CL sweet even (to) I DE child all give FP  
“Zhangsan gave this sweet even to my child

<sup>34</sup> Shyu (2001), following Kratzer’s (1989) distinction between “stage level” predicate, which expresses a specific situation or event, from “individual level” predicate (generic sentences), claims that SOV order can appear in “individual level” clause only when the sentence has contrasting function.

<sup>35</sup> As a reviewer suggested to me, the negation by itself implies a sort of contrast.

In this sentence my Chinese informants point out that the main stress is always on the XP following *lian* and not on the preposed Object<sup>36</sup>.

Many linguists (Tsao 1977; Qu 1994; Shyu 1995) noted that two [+animate] NPs can switch their Theta-roles: [NP1 NP2 V]. In this case it is natural to interpret NP2 as the Subject and NP1 as the Topic. But if NP2 is uttered with a contrastive stress, NP1 functions as the Subject and the NP2 as the Object.

- (68) Ta [Zhang xiaojie]<sub>i</sub> bu xihuan *t<sub>i</sub>*. (Huang, A. Li & Y. Li forthcoming: 18)  
 he Zhang miss not like  
 “Miss Zhang does not like him.”  
 ?? “He does not like Miss Zhang.”

The reading is clearer with a clause highlighting the contrastive use of the preposed Object:

- (69) Q: Ta hui zhui Zhang xiaojie ma? (Huang, A. Li & Y. Li forth.: 19)  
 he will court Zhang Miss Q  
 “Will he court Miss Zhang?”  
 A: Ta [Zhang xiaojie]<sub>i</sub> bu xiang zhui *e<sub>i</sub>*, [Li xiaojie]<sub>j</sub> cai hui zhui *e<sub>j</sub>*  
 he Zhang Miss not want court Li Miss only will court  
 “He does not want to court Miss Zhang; (he) only will court Miss Li.”

Furthermore, consider a typical “Aboutness Topic” in the CP area like the following:

- (70) a. [Zoumingqu], Zhangsan xihuan tan, dajia ye xihuan ting. (Ting 1995:3)  
 Sonata Zhangsan like play everyone also like listen  
 “As for sonatas, Zhangsan likes to play them and everyone also likes to listen to them.”  
 b. # Zhangsan [zoumingqu] xihuan tan, (dajia ye xihuan ting).  
 Zhangsan sonata like play (everyone also like listen)  
 Lit.: # “Zhangsan, sonatas, likes to play them and everyone also likes to listen to them.”

A similar interpretation, i.e. as an “Aboutness Topic” for SOV is not possible. This is further evidence showing that SOV Object is a Topic with a contrastive reading. After the considerations above, I conclude that the SOV occupies a Contrastive Topic position. I also conclude that the Low Periphery in Chinese disposes of only one Topic position, with a Contrastive interpretation. Differently from the CP area, where any kind of Topic may be contrastively stressed, within the IP there is a dedicated position yielding contrastive interpretation (see Badan 2007). With the evidence that in Chinese the landing site of the bare preposed Object within IP is a Contrastive Topic Projection, I have argued that the Object moves up to check its Topic property, in a Spec-Head agreement configuration.

<sup>36</sup> *lian* functions like a Focus stress for the XP that it selects. For this reason, when it co-occurs with another item, it always gets the Focus accent (see Badan 2007).

## 6. Conclusions

In this paper I have applied Belletti's (2001, 2004) proposal for the existence of a Low Periphery. Following Paul (2005), I have shown that Chinese also shows a Low Periphery consisting of two kind of Functional Projections, occupied by the bare preposed Object (SOV) and the *lian*+XP. Finally I have concentrated on the SOV position. Contrary to traditional analyses, I have demonstrated that SOV is not a Focus syntactically speaking, but a Topic that gets Focus stress. I have argued that it can be defined as a Contrastive Topic.

## References

- Abbiati, M. (1998) *Grammatica di Cinese Moderno*. Venezia: Cafoscarina.
- Antinucci, F. & G. Cinque (1977) Sull'ordine delle parole in italiano: l'emarginazione. In *Studi di Grammatica Italiana* 6: 121-146.
- Badan, L. (2007) *High and Low Periphery: a Comparison between Italian and Chinese*. PhD Dissertation, Università degli Studi di Padova.
- Badan, L. & F. Del Gobbo (in press) On the Syntax of Topic and Focus in Chinese. In Benincà P. & Munaro N. (eds.), *Mapping the left periphery*, Oxford-New York: Oxford University Press.
- Belletti, A. (1990) *Generalized verb movement*. Turin: Rosenberg & Sellier.
- Belletti, A. (1999) Italian/Romance clitics: Structure and derivation. In Van Riemsdijk, H. (ed.), *Clitics in the Languages of Europe*, Berlin: Mouton de Gruyter: 543-580.
- Belletti, A. (2001) "Inversion" as Focalization. In Hulk A. & Pollock J.Y. (eds.), *Inversion in Romance and the theory of Universal Grammar*, Oxford-New York: Oxford University Press.
- Belletti, A. (2004) Aspects of the low IP area. In Rizzi L. (ed.), *The Structure of IP and CP. The Cartography of Syntactic Structures, vol. 2*, Oxford-New York: Oxford University Press.
- Belletti, A. ed. (2004) *Structures and Beyond. The Cartography of Syntactic Structures, vol 3*. New York & Oxford: Oxford University Press.
- Belletti, A. & U. Shlonsky (1995) The Order of Verbal Complements. A Comparative Study. *Natural Language and Linguistic Theory* 13: 489-526.
- Benincà, P. (2001) The position of Topic and Focus in the left periphery. In Cinque G. & Salvi G. (eds.), *Current Studies in Italian Syntax. Essays Offered to Lorenzo Renzi*, Amsterdam: Elsevier-North Holland: 39-64.
- Benincà, P. & C. Poletto (2004) Topic, Focus and V2: Defining the CP sublayers. In Rizzi L. (ed.), *The Structure of CP and IP. The Cartography of Syntactic Structures, vol. 2*, New York-Oxford: Oxford University Press: 52-75.
- van Bergen, G. (2006) *To BA or not to BA. Differential Object Marking in Chinese*. MA Thesis, Radboud University Nijmegen.
- Calabrese, A. (1992) Some Remarks on Focus and Logical Structures in Italian. *Harvard Working Papers in Linguistics* 1: 19-27.
- Cecchetto, C. (1999) A comparative analysis of left and right dislocation in Romance. *Studia Linguistica* 53, 1: 40-67.
- Cheng, L.-S. L. & L.J. Downing (2007) Phonology and Syntax at the Left Edge in Zulu. Paper presented at the *What's the TOPIC?-Workshop*, Centre for Language Studies (CLS), 23-24.I, Radboud University Nijmegen.

- Chomsky, N. (2000) Minimalist Inquiries: The Framework. In Martin R., Michaels D., & Uriagereka J. (eds.), *Step by Step. Essays on Minimalist Syntax in Honour of H. Lasnik*, Cambridge, MA: MIT Press: 89-155.
- Cinque, G. (1999) *Adverbs and Functional Heads. A Cross-Linguistic Perspective*. New York-Oxford: Oxford University Press.
- Ernst, T. (1998) Case and parameterization of scope ambiguities. *Natural Languages and Linguistic Theory* 16.1: 101-148.
- Ernst, T. & C. Wang (1995) Object Preposing in Mandarin Chinese. *Journal of East Asian Linguistics* 4, 3: 235-260.
- Fu, J. (1994) SOV order in Chinese and IP specifier. Paper presented at the 6<sup>th</sup> North American Conference on Chinese Linguistics, University of Southern California.
- Gao, Q. (1994) Focus Criterion: Evidence from Chinese. *Proceedings of the 6<sup>th</sup> North American Conference on Chinese Linguistics*, University of Southern California.
- Goodall, G. (1987) On the argument structure and L-marking with Mandarin Chinese *ba*. In McDonough J. & Plunkett B. (eds.), *Proceedings of NELS 17*, Amherst, MA: GSLA: 232-242.
- Hashimoto, A. (1971), Mandarin Syntactic Structures. *Unicorn* 5.
- Hole, D. (2004) *Focus and background marking in Mandarin Chinese: System and theory behind cai, jiu, dou and ye*. London-New York: Routledge Curzon.
- Huang, C.-T. J. (1982) *Logical relations in Chinese and the theory of grammar*. PhD Dissertation, MIT. Published by Garland Publishing, Inc., New York, in 1989.
- Huang, C.-T. J., A. Li & Y. Li (forthcoming) *The Syntax of Chinese*. Cambridge University Press.
- Jayaseelan, K. (2001) IP-internal Topic and Focus Phrases. *Studia Linguistica*, 55, 1: 39-75.
- Kayne, R. (1984) *Connectedness and binary branching*. Dordrecht: Foris.
- Koopman, H. & D. Sportiche (1990) *The Position of Subject*. Ms., UCLA.
- Kratzer, A. (1989) Stage-Level and Individual-Level Predicates. In *Papers on Quantification*, Amherst: University of Massachusetts.
- Kuroda, S.-Y. (1988) Whether We Agree or Not: A Comparative Syntax of English and Japanese. In Poser W.J. (ed.), *Japanese Syntax*, Stanford: CSLI.
- Li, Y.-H. A. (2001) The *ba*-construction. Ms, <http://www.usc.edu/dept/LAS/eacl/chinling/audreyli.htm>.
- Li, Y.-H. A. (2006) Chinese *Ba*. In Everaert M. & van Riemsdijk H. (eds.), *The Blackwell Companion to Syntax, vol.1*, Oxford: Blackwell Publishing.
- Lin, J. (1992) The syntax of *zenmeyang* “how” and *weishenme* “why” in Mandarin Chinese. *Journal of East Asian Linguistics* 1: 293-331.
- Lu, H.-C. (1994) Second Preverbal NPs in Chinese. *Proceeding of the 6<sup>th</sup> North American Conference on Chinese Linguistics*, University of Southern California.
- Paris, M.-C. (1979) Some aspects of the syntax and semantics of the *lian...ye/dou* construction in Mandarin. *Cahiers de linguistique – Asie orientale* 5: 47-70.
- Paris, M.-C. (1998) Focus Operators and Types of Predication in Mandarin. *Cahiers de Linguistique-Asie Orientale* 27, 2 : 139-159.
- Paris, M.-C. (1999) Ordre des mots, topique et focus en chinois contemporain. In Guimier C. (ed.), *La thématization dans les langues. Actes du Colloque de Caen, 9/11-10-1997*.
- Paul, W. (2002) Sentence-internal topics in Mandarin Chinese: the case of object preposing. *Language and Linguistics* 3-4: 695-714.
- Paul, W. (2005) Low IP and left periphery in Mandarin Chinese. *Recherches Linguistiques de Vincennes* 33: 111-134.

- Paul, W. (2006) *What the topic is (not) about: the Case of Mandarin Chinese*. Ms., CRLAO, EHESS, Paris.
- Paul, W. & J. Whitman (2001) The complement structure of *shi...de* clefts in Mandarin Chinese. Paper presented at the *IACL-10/NAACL-13*, 22-24.VI, University of California, Irvine.
- Poletto, C. (2006) Parallel Phases: a study on the high and low left periphery of Old Italian. In Frascarelli M. (ed.), *Phases of Interpretation, Studies in Generative Grammar* 91, Berlin: Mouton de Gruyter: 261-294.
- Qu, Y. (1994) *Object Noun Phrase Dislocation in Mandarin Chinese*. Ph.D. Dissertation, University of British Columbia.
- Rizzi, L. (1997) The fine structure of the left periphery. In Haegeman L. (ed.), *Elements of Grammar. Handbook of Generative Syntax*, Dordrecht: Kluwer: 281-337.
- Shyu, S. (1995) *The Syntax of Focus and Topic in Mandarin Chinese*. PhD Dissertation, University of Southern California.
- Shyu, S. (2001) Remarks on Object Movement in Mandarin SOV Order. *Language and Linguistics* 2, 1: 93-124.
- Shyu, S. (2004) (A)symmetries between Mandarin Chinese *Lian-dou* and *Shenzhi*. *Journal of Chinese Linguistics* 32, 1: 71-128.
- Sportiche, D. (1995) French Predicate Clitics and Clause Structure. In Cardinaletti A. & M.T.Guasti (eds.), *Small Clauses, Syntax and Semantics vol.28*, 287-324.
- Starke, M. (1995) On the Format of Small Clauses. In Cardinaletti A. & M.T.Guasti (eds.), *Small Clauses, Syntax and Semantics vol.28*, 237-69.
- Tang, C.-C. J. (1990) *Chinese Phrase Structure and the extended X-bar Theory*. PhD Dissertation, Cornell University.
- Ting, J. (1995) Deriving the secondary topic construction in Mandarin Chinese. *Proceedings of the 7<sup>th</sup> North American Conference on Chinese Linguistics*, University of Southern California: GSIL: 289-302.
- Tsai, W.-T. D. (1994) *On Economizing the theory of A-bar dependencies*. PhD Dissertation, MIT, Cambridge, MA.
- Tsao, F. (1977) *A functional study of topic in Chinese: the first step toward discourse analysis*. PhD Dissertation USC, Los Angeles, California.
- Travis, L. (1984) *Parameters and Effects of Word Order Variation*. PhD Dissertation, MIT, Cambridge, MA.
- Xu L. & D. T. Langendoen (1985) Topic Structures in Chinese. *Language* 61: 1, 1-27.
- Zhang, N. (1996) Three types of Object Shift in Mandarin Chinese. *Proceeding of the 8th North American Conference on Chinese Linguistics*, University of Illinois, Urbana-Champaign: 413-425.

# The Cartography of Syntactic Structures

GUGLIELMO CINQUE

University of Venice  
cinque@unive.it

LUIGI RIZZI

University of Siena  
rizzil@unisi.it

## 1. Introduction

Syntactic structures are complex objects, whose subtle properties have been highlighted and elucidated by half a century of formal syntactic studies, building on a much older tradition. Structures are interesting objects of their own, both in their internal constitution and in their interactions with various grammatical principles and processes. The cartography of syntactic structures is the line of research which addresses this topic: it is the attempt to draw maps as precise and detailed as possible of syntactic configurations. Broadly construed in this way, cartography is not an approach or a hypothesis: it is a research topic asking the question: what are the right structural maps for natural language syntax? Answers may differ, and very different maps may be, and have been, proposed, but the question as such inevitably arises as a legitimate and central question for syntactic theory. If it is a virtual truism that cartography can be construed as a topic and not as a framework, it is also the case that cartographic studies have often adopted certain methodological and heuristic guidelines, and also certain substantive hypotheses on the nature of syntactic structures, which form a coherent body of assumptions and a rather well-defined research direction; we will try to illustrate some ideas and results of this direction in the present chapter.

If structures have, in a sense, always been central in generative grammar, the idea of focusing on structural maps arose around the early nineties, following a track parallel to and interacting with the Minimalist Program. Perhaps the main triggering factor was the explosion of functional heads identified and implied in syntactic analyses in the first ten years of the Principles and Parameters framework. One critical step was the full-fledged extension of X-bar theory to the functional elements of the clause (Chomsky 1986) as a CP – IP – VP structure; and the observation that other configurations, e.g. nominal expressions, were amenable to a hierarchical structure with a lexical projection embedded within a functional structure (such as Abney's DP hypothesis, Abney 1987). These advances provided a natural format for the study of the structure of phrases and clauses as hierarchical sequences of the same building block, the fundamental X-bar schema (or, later, elementary applications of Merge); the lowest occurrence of the building block typically is the projection of a lexical category, e.g. a noun or a verb, and this element is typically completed by a series of

building blocks headed by functional elements, providing more abstract semantic specifications to the descriptive content of the lexical head: tense, mood, aspect for the verb, definiteness, specificity, number for the noun, etc.

If the first step was the idea that clauses and phrases are formed by a lexical structure and a higher functional structure, both corresponding to elementary building blocks hierarchically organized, the second crucial step was the observation that the functional structure typically consists of more than one head. In fact, a Complementizer Phrase (CP) and an Inflectional Phrase (IP) zone were isolated from the outset, but it became clear very soon that the same kinds of evidence which supported the analysis of inflected verbs in terms of the distinction between I and V would lead to the splitting of I into more elementary components. The same logic led to a later splitting of the CP and DP zones into more articulated hierarchical sequences of functional projections.

The initial impulse for splitting the IP was provided by Pollock's seminal paper on verb movement in French and English (Pollock 1989, versions of which circulated already around the mid eighties). Pollock showed that assuming a single I position did not provide enough space to account for the different positions which can be occupied by different morphological forms of the verb in French: infinitival verbs may remain in the VP, as in (1)a, or be moved to a higher position across lower adverbs like *complètement* (completely), as in (1)b; finite verbs move to an even higher position across negative *pas*, as in (1)c:

- (1) a. ne X<sub>1</sub> pas X<sub>2</sub> complètement [X<sub>3</sub> comprendre] la théorie ...  
      neg not completely understand the theory
- b. ne X<sub>1</sub> pas [X<sub>2</sub> comprendre] complètement X<sub>3</sub> la théorie ...
- c. Il ne [X<sub>1</sub> comprend] pas X<sub>2</sub> complètement X<sub>3</sub> la théorie

If I splits into at least two heads X<sub>1</sub> and X<sub>2</sub>, Pollock argued, the three positions of (1) can be naturally accommodated by assuming optional movement of the infinitival verb from its VP-internal position X<sub>3</sub> to X<sub>2</sub>, and obligatory verb movement of the finite verb to X<sub>1</sub>. This analysis, also building on Emonds (1978), introduced a fundamental insight: adverbs basically don't move, except in the cases in which they are displaced for scope-discourse reasons, focalized, and the like; variations within a language and across languages of verb-adverb orders are due to verb movement in the inflectional space, a particular instance of head movement. This approach in fact united two lines of research which have become integral components of the cartographic studies: on the one hand, the analysis of the word order properties of verbs with respect to adverbial and arguments in terms of head movement, as mentioned; on the other hand, the idea that inflectional morphology is done in the syntax and is the result of movement rules involving roots and affixes, an idea going back to the analysis of verb affixation in English in *Syntactic Structures* (Chomsky 1957). The Emonds-Pollock approach united the two trends by proposing that the verb could be attracted to different functional positions to pick up affixes and get properly inflected, thus changing its position with respect to adverbs and other elements, which made it possible to capture many important form-position correlations.

The question then arose of the proper labeling of  $X_1$  and  $X_2$ . Belletti's (1990) proposal was that the higher functional projection of the clause is the one responsible for subject-verb agreement (AgrS in the traditional terminology), and the lower one expresses tense. This order AgrS – T is immediately reflected in the order of prefixes or particles in e.g., the Bantu languages; while in languages in which these properties are expressed by suffixes, i.e. the Romance languages, the order is the mirror image (see Italian *parl-av-ano*, root-T-AgrS, “(they) spoke”), as is to be expected under Baker's (1988) Mirror Principle: the verb moves to pick up the closest suffix, which therefore appears as the one immediately attached to the root, etc.).

The logic of this argumentation, combining the syntactic make-up of inflectional morphology via head-movement and the study of the order of arguments and adjuncts with respect to different verbal forms, quickly led to a finer splitting of the inflectional space into a sequence of functional heads expressing properties of mood and modality, tense, aspect, voice. For a few years, around the late eighties, this methodology led to the discovery and postulation of a variety of functional heads driven by the analytic necessities of particular morphosyntactic problems, a trend which sometimes gave the impression that the process would lead to an ever increasing complexity of the syntactic representations. How rich could be the “right” functional structure of clauses and phrases? One of the driving ideas of the cartographic projects was precisely to complement this trend of bottom-up, problem-related discovery with a more top-down, global perspective, trying to make a rough estimate of the upper limit of the structural complexity. Instrumental to this endeavor was the working assumption that each morphosyntactic feature would correspond to an independent syntactic head with a specific slot in the functional hierarchy (cf. also Kayne 2005a,15). Much of the cartographic work has consisted in the attempt, in various forms, to use this working hypothesis as a heuristic guideline, thus spelling out empirical arguments supporting or disconfirming its validity across languages.

## 2. Methodology and evidence

In the first half of last century, in part as a reaction to what was then felt as an unwarranted application of European grammatical categories and constructions to non-European languages, the common wisdom in American structuralism (epitomized in Joos 1957,96) was that “languages could differ from each other without limit and in unpredictable ways” so that each language should be studied “without any preexistent scheme of what a language must be”. The rejection of these assumptions, which are still adopted today by many functionalists<sup>1</sup>, was implicit throughout the history of generative grammar<sup>2</sup>, and is made explicit in Chomsky's (2001, ) “Uniformity Principle” (“In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances.”). The cartographic approach follows this idea in assuming that all languages share the

---

<sup>1</sup> See for example LaPolla and Poa (2002,2): “Each language is a unique set of language-specific conventions, and so each language should be described in its own terms”, or Haspelmath (2007, ) “descriptive linguists still have no choice but to adopt the Boasian approach of positing special language particular categories for each language. Theorists often resist it, but the crosslinguistic evidence is not converging on a smallish set of possibly innate categories. On the contrary, almost every newly described language presents us with some “crazy” new category that hardly fits existing taxonomies.”

<sup>2</sup> See, for example, Koopman and Sportiche (1991,218f): “[W]e suppose that the null assumption concerning language variation is that it does not exist.”

same principles of phrase and clause composition and the same functional make-up of the clause and its phrases.<sup>3</sup>

More precisely the cartographic approach assumes, as the evidence of the last several years seems to indicate, that the distinct hierarchies of functional projections dominating VP, NP, AP, PP, IP, etc., may be universal in the type of heads and specifiers that they involve, in their number, and in their relative order, even if languages differ in the type of movements that they admit or in the extent to which they overtly realize each head and specifier (Rizzi 1997; Cinque 1999; 2002,3f).

This is the strongest position one could take; one which implies that if some language provides evidence for the existence of a particular functional head (and projection), then that head (and projection) must be present in every other language, whether the language offers overt evidence for it or not (cf. Kayne 2005,12; Cinque 2006a,4).<sup>4</sup>

A weaker position would consist in assuming that languages may differ in the type or number of functional projections they select from a universal inventory, or in their order.<sup>5</sup>

Although the choice between these two positions will ultimately be decided by the nature of things, methodologically it would be wrong, it seems, to adopt the weaker position as a first working hypothesis. That would only make us less demanding with respect to the facts and could lead us to miss more subtle evidence supporting the stronger position, a risk not present under the other option (Cinque 2002,4).

The question whether such universal hierarchies of functional projections are primitive objects of UG, or can be derived from interface or more general external conditions is important, but fundamentally orthogonal to the prior task of drawing their precise map, and perhaps not easily determinable at the present state of our knowledge.

The evidence brought to bear in the literature on the mapping of universal hierarchies of functional projections comes from a variety of sources.

An early source for postulating (abstract) functional projections was the existence of certain systematic word order differences among languages, like Pollock's (1989) classical argument for positing a non-lexical head higher than VP and lower than I (or T), to which finite verbs raise in French (but not in English), along the lines discussed in the introductory section.

Another important source of evidence is the relative order of the functional morphemes overtly realized in the languages of the world (to the extent that one can establish reasonable correspondences among the functional morphemes of different languages). Though languages differ as to what functional categories they overtly realize, the partial orders displayed by different languages seem to fit in a unique

---

<sup>3</sup> This is not to say that it is always easy to establish precise correspondences between the functional categories overtly displayed by different languages. Caution must be exercised, but there is no a priori reason to rule out the possibility that such correspondences can ultimately be established. In fact, this has turned out to be possible in a number of cases through in-depth investigation. See, for example, the case of French *peu* and English *bit* (rather than *little*) discussed in Kayne (2005, §4.2).

<sup>4</sup> The literature offers a number of cases supporting this general hypothesis. See, for example, the discovery of more subtle evidence for the presence of a DP projection in languages like Serbo-Croatian, Russian, and Japanese, which lack overt determiners (Progovac 1998, Pereltsvaig 2007, Furuya 2008); or the indirect evidence discussed in Kayne (2003,219) and Cinque (2006b) for the presence of numeral classifiers in languages like English and Italian, which are traditionally taken not to be numeral classifier languages.

<sup>5</sup> This is the position taken, for example, by Thráinsson (1996) and Bobaljik and Thráinsson (1998), among others. See also Fukui (1995).

macro-hierarchy despite occasional inconsistencies, which have proved (and hopefully will prove, as our knowledge progresses) solvable.

Preliminary inquiries on the functional hierarchies of the left periphery of the clause (Rizzi 1997, 2001, 2004a,b; Benincà 2001,2006, Benincà and Poletto 2004; Bocci 2004; Benincà and Munaro to appear; Cruschina 2006; Frascarelli and Hinterhölzl 2007, Frascarelli and Puglielli to appear, among others), of the core functional structure of the clause (Cinque 1999, 2006; Shlonsky 1997, 2000; Sigurðsson 2000; Cardinaletti 2004; Schweikert 2005; Bianchi 2006; and, for its relevance for computational linguistics, Chesi 2005), of the DP (Cinque 1994, Scott 2002, Brugè 2002, Giusti 2002, Nicolis 2008, Svenonius 2008a), and of PPs (see the contributions in Asbury, Dotlačil, Gehrke, Nouwen 2008 and Cinque and Rizzi to appear), have largely confirmed the working hypothesis that there may be a universal functional design for the clause and its major phrases holding across languages.<sup>6</sup>

Of course, to determine the relative order of functional morphemes one has to have an idea of what the classes of such elements are as opposed to the lexical ones (see section 3 below for some discussion), and this task often requires “regularizing” the orders found across languages, as they can be obscured to various degrees by various types of syntactic movements. So for example the relative order of functional morphemes that appear to the right of a certain lexical category, as suffixes or free morphemes, is most often (though by no means always) the mirror image of the same functional morphemes that appear on the left of the same lexical category in other languages, arguably a consequence of the lexical category moving across the functional morphemes in the former type of languages (see Baker 1985 for the original formulation of the Mirror Principle and, for recent discussion, see Cinque 2008).

Analogously, as noted in Carlson (1983,73), one of the earliest and most enlightening discussions of functional categories in the generative tradition, the Latin coordinating enclitic conjunction *-que* exemplified in (2) is not interpreted as conjoining with a like constituent what precedes it ( i.e. the unit [*ob eās*]), but the entire higher unit [*ob eās rēs*] (as in English). This again can be “regularized” if the movements that created (2) (from ...*ob eās rēs -que*) are undone.

- (2) *ob eās-que rēs*  
 because.of these-**and** things  
 ‘and because of these things’

---

<sup>6</sup> Some authors have argued that this particular assumption of the cartographic approach is incorrect because it rests on transitivity (if  $A > B$  and  $B > C$ , then  $A > C$ ), which appears to fail in certain cases (see Bobaljik 1999, Nilsen 2003, and also Zwart 2006). Caution however is in order given the otherwise general validity of transitivity, and the possibility that some account exists which renders these cases irrelevant for transitivity issues. See in fact Cinque (2004, footnotes 22 and 43 for evidence to this effect). Van Craenenbroeck’s (2006) analogous argument from an apparent transitivity failure in the left periphery also ignores the possibility that the complementizer may occupy more than one position, thus rendering his case irrelevant to the transitivity issue. That an element like *that* may appear more than once and in different positions in the left periphery of a clause is straightforwardly shown by many cases of multiple occurrences of *that*, e.g., in Brazilian Portuguese, Gascon and Piedmontese, structures with orders like *I think that JOHN that you should meet*, with the first *that* functioning as declarative force marker, and the second as a focus marker (see Mioto 1998, Poletto 2000, 148-50 for relevant discussion).

These are two out of the many cases where care must be taken to render things comparable and to expose the deeper regularities that underlie the functional make-up of the clause and its phrases.

### **3. Inventory of functional categories**

A guiding idea of much current cartographic work is that the inventory of functional elements (heads or specifiers of functional projections) is much larger than is generally thought. In all grammatical traditions it is customary to make, in one way or another, some distinction between lexical categories (like Nouns and Verbs: see Baker 2003) and functional, or grammatical, ones (like Determiners and Complementizers). If we take membership in an open vs. closed class of items as a diagnostic distinguishing lexical from functional elements, then the candidates for the functional lexicon of languages become very numerous. Not only Determiners and Complementizers are functional, but also conjunctions, (functional) adpositions like *of, for, from, at, to, with* (as well as spatial adpositions - see Cinque and Rizzi to appear, and references cited there), mood, modal, tense, aspect, polarity, and voice morphemes<sup>7</sup>, auxiliaries, copulas and other verbs lacking a clear argument structure, (strong, weak, and clitic) pronouns, demonstratives, quantifiers, numerals (see Kayne 2005,13), classifiers, number (plural, dual, etc.) morphemes, gender or class morphemes, diminutive/augmentative morphemes, degree words, indefinite/*wh*-words, Case morphemes, focusing adverbs (like ‘only’ and ‘also’), comparative and superlative morphemes, and many many more (see Kayne 2005, section 2.1). To judge from Heine and Kuteva’s (2002) four hundred, or so, independent grammaticalization targets, the number of functional elements must at least be of that order of magnitude. It is in fact quicker to consider which elements are lexical (belong to an open class). Nouns in all languages appear to be an open class; perhaps the only genuinely open class, as the considerations that follow may indicate. The situation is certainly far less clear for adjectives, adverbs, and verbs (which are often taken to be lexical, open, classes). In many languages, adjectives constitute a closed, often quite small, class of elements. This is especially clear in those languages, like Yoruba (see Cinque 2006a,5 and references cited there), whose adjectives cannot be used predicatively. In such languages the attributive-only adjectives form a closed (generally small) class; a clear sign of their functional status. For discussion and exemplification, see Dixon (1982,2004), Cinque (2006a,4f, to appear). The fact that they appear to form an open class in other languages may be due to the existence of a parallel predicative class of adjectives (which enlarges the set of adnominal adjectives by adding a reduced relative clause source), as well as to possible productive morphological derivations of adjectives from nouns or verbs (e.g. *-al, -ous, -ed*, etc. in English).

A similar situation is encountered with adverbs, which also constitute a clear closed class of elements in some languages (see Dixon 1982,40; Schachter 1985,21ff; Stutzman 1997,75; Cinque 1999,213fn79, 2006,9fn.22, and references cited there). Furthermore, the fact that they are coded as rigidly ordered affixes in certain languages while they are coded as independent words in others (also in a fixed order) may suggest that generation in head or specifier position of a dedicated functional projection is an option left open to languages by UG.

---

<sup>7</sup> Whether bound or free. On the (functional) syntactic import of bound morphemes, see the recent discussion in Kayne (2005,11f).

If Hale and Keyser's (1993) idea that most transitive and intransitive verbs are not primitive but result from the incorporation of a noun into a limited class of light/general purpose verbs ('do', 'give', 'take', 'put', 'hit', etc.), then even the class of *primitive* verbs may turn out to be closed and relatively small. This seems confirmed by the fact that some languages typically fail to incorporate the noun into the light verb so that most 'verbal meanings' are expressed as V + N periphrases. This is for example the case of Persian.<sup>8</sup> The typological literature also reports the case of a number of languages from Australia and New Guinea with closed classes of main verbs (see Dixon 1982,225; Pawley 2006).<sup>9</sup>

#### 4. Comparative syntax and typology

Crucial to the cartographic approach is the evidence coming from comparative and, more broadly, typological studies. These alone may help singling out the variety (and the limits) of the functional lexicon of UG. In-depth studies of a single, or of few languages, however deep they may be, fall short of revealing the actual richness of the functional/grammatical structure of UG owing to the often silent character of a certain functional category in a certain language (see Kayne 2005a, 2006). More importantly still, as noted, comparison of many different languages may provide evidence for determining the precise relative order of the different functional projections by combining the partial orders overtly manifested by different languages into what, in principle, should be a unique consistent order/hierarchy, imposed by UG. This presupposes that the order of functional projections is fixed within one language, and, more crucially, across languages; hardly an obvious assumption.

Comparative evidence is also crucial in exposing how certain ordering properties are strictly impossible across languages. Even in cases in which variation is permitted by UG, it is never the case that "anything goes". There are precise limits to the observed cross linguistic variation, a fact which calls for a principled explanation. Consider for example the order of demonstratives, numerals and adjectives with respect to the N (Greenberg's 1963 Universal 20). Even if variation in their relative ordering is extensive, of the 24 mathematically possible orders of the four elements, only 13 are clearly attested in the languages of the world. Apparently only those orders which are obtainable from a unique base order (Dem Num A N) by moving the N (or NP) leftward to higher functional positions in one of the ways independently admitted by the syntax of natural languages (see Cinque 2005 for discussion).

---

<sup>8</sup> "Most verbal constructions in Persian are formed using a light verb such as *kardan* ('do', 'make'), *dâdan* ('give'), *zadan* ('hit', 'strike'). The number of verbs that can be used as light verbs is limited, but these constructions are extremely productive in Persian." (Megerdooomian n.d.). Also see Karimi-Doostan (1997).

<sup>9</sup> Interestingly, the literature on agrammatism reports the fact that even main verbs are impaired. See Miceli, Silveri, Villa and Caramazza (1984) (thanks to Franco Denes for pointing out this article to us; there are also cases of selective impairment of the nominal system with verbs relatively spared (Caramazza & Shapiro 2004), but these are much rarer than cases of selective V impairment). If main verbs are the morphological merge of a noun plus one of a closed class of 'grammatical' verbs, their conclusion that "agrammatism is a heterogeneous disorder that implicates damage of both lexical and syntactic mechanisms" (p.220) may have to be reassessed, and perhaps reduced to a disorder of (different types of) purely grammatical mechanisms.

## **5. Cartography and Minimalism**

The cartographic projects have been developed roughly at the same time as the rise and development of the Minimalist Program (Chomsky 1995 and much subsequent work). There is, at first sight, an inherent tension between the complexity of the cartographic representations and the simplicity of the generative devices that minimalist syntax assumes, somehow reflected in the structural poverty of the representations typically found in the minimalist literature. We believe that there is no contradiction between these two directions of research, and the tension, where real, is the sign of a fruitful division of labor. Minimalism focuses on the elementary mechanisms which are involved in syntactic computations, and claims that they can be reduced to extremely simple combinatorial operations, ultimately external and internal Merge, completed by some kind of search operation (Chomsky's Agree) to identify the candidates of Merge. An impoverished computational mechanism does not imply the generation of an impoverished structure: a very simple recursive operation can give rise to a very rich and complex structure, as a function of the inventory of elements it operates on, and, first and foremost, of its very recursive nature. The very simplified structural representations often assumed in the minimalist literature, expressed by the C-T-v-V system, are sometimes taken literally, as substantive hypotheses on the nature of clausal configurations, but the structure of the arguments rarely implies a literal interpretation, and often is compatible with an interpretation of C-T-v-V as a shorthand for more complex cartographic structures (a fact explicitly acknowledged, e.g., in Chomsky 2001, fn. 8), with C, T, and v taken as "abbreviations" standing for complex zones of the functional structure. The division of labor here is that Minimalism focuses on the generating devices, and cartography focuses on the fine details of the generated structures, two research topics which can be pursued in parallel in a fully consistent manner, and along lines which can fruitfully interact (see Cinque 1999, section 6.2, Rizzi, 2004a, introduction and Belletti 2008, introduction, for relevant discussion).

In fact, cartographic studies are based on general guidelines which are at the heart of the minimalist program. Minimalism has introduced a principled typology of UG principles, which are traced back to only two kinds of broad categories: principles dictated by the needs of the interface systems, determining the proper legibility and usability of the interface representations, and economy/locality principles, constraining the functioning of the computing machine.

The first class includes principles determining the mapping of a hierarchical structure into a linear sequence expressible by the human articulatory system, such as Kayne's (1994) Linear Correspondence Axiom and its variants; and principles ensuring the expressibility of properties required by the human conceptual-intentional systems and by the needs of an efficient communication: properties of argument structure, referential dependencies, scope, and informational packaging in discourse and dialogue. All these aspects play a critical role in cartographic studies. Much work on the reordering of elements generating superficial exceptions to the hierarchical order crucially makes extensive use of remnant movement (e.g. Cinque 1999, Koopman and Szabolcsi 2000, but also much work on the left peripheral positions of *wh* operators in Romance languages and dialects by Munaro, Obenauer, Poletto, Pollock), a direct offspring of the antisymmetric approach. Work on the cartography of the verbal system (Ramchand 2008) and of prepositions (Svenonius 2008b and the contributions collected in Cinque and Rizzi to appear) investigate the syntactic correlates of argument structure in structural approaches to the lexicon-syntax interface inspired by Hale and Keyser's (1993) perspective. Much work on the

fine structure of the left periphery investigates the syntax of dedicated scope-discourse positions in various languages: Romance (Rizzi 1997, Benincà & Poletto 2004 and many other contributions in Rizzi 2004a and, on Romance dialects, Manzini and Savoia 2005), Germanic (Grewendorf 2002, Haegeman 2006), West African languages (Aboh 2004, 2007), Creole languages (Durrleman 2007), East Asian languages (Endo 2007, Tsai 2007).

The study of locality/economy is also central to the cartographic endeavor, in that the positional articulation uncovered by cartographic studies offers a sound basis for establishing a principled typology of positions which is required by the analysis of intervention locality: within the Relativized Minimality tradition (Rizzi 1990), an intervener of “the same kind” as the target of movement blocks a movement chain; the typology of positions cannot be established in the traditional terms of the A/A’ distinction, too coarse, nor in terms of a featural identity between the target and the intervener (too selective), and seems to require a feature-driven typology of an intermediate level of granularity, which can be directly related to the cartographic structures (Rizzi 2004, Starke 2001, Grillo 2008).

One point in which cartographic studies seem to us to fruitfully implement general simplicity guidelines which are proper of minimalism is the study of the elements of syntactic computations. One useful heuristic principle which has guided much cartographic work is the maxim “one (morphosyntactic) property – one feature – one head”. This guideline does not exclude the possibility that featurally complex heads may arise in syntax, but they cannot be “atoms” of the syntactic computations, they can only arise through derivational procedures, namely head movement, which may create a complex conglomerate of features by moving featurally simple heads into other heads (it does not matter here whether head movement literally extracts a head from its projection, or is a kind of phrasal movement “in disguise”). It is this kind of intuition which guided the “unpacking” of the Infl node of early P&P analyses into its elementary component. Of course, a single surface position may express both the lexical content, tense, mood and subject agreement (as Italian present subjunctive *part-a-no* “that they leave”), but this is done through movement of the verbal head picking up the various elementary specifications. Similar considerations hold for the unpacking of the C node, of the determiner system, etc.

The basic intuition that cartographic studies try to validate empirically is that natural language design opts for local simplicity whenever possible: each syntactic head has a simple featural specification and can enter into few simple relations with its associates. Preservation of local simplicity is the effect massively produced by the pervasive presence of movement in natural language syntax. Consider for instance A’ movement chains, configurations which transparently arise to associate two kinds of interpretive properties to certain expressions. So, the expression *this book* must be interpreted as the thematic argument of the verb *read*, and as the topic of the structure in (3):

(3) This book, I will read \_\_\_ tomorrow

Natural languages express this state of affairs by having the element occur twice, once in the thematic position and once in the left peripheral position dedicated to topicality. The assignment of argumental thematic properties is, uncontroversially, a matter of head-dependent relation: the verb assigns a certain thematic role to its immediate dependent. What about a scope-discourse property like topicality? The line pursued by cartographic studies is that scope-discourse properties are assigned to

elements in a configurationally uniform way, *mutatis mutandis*: there is a dedicated head, Top, normally occurring in the left periphery of the clause, which activates the interpretive instruction “my specifier is to be interpreted as the topic, and my complement as the comment”. Under the copy theory of traces the full representation of (3) is

- (4) This book [ Top [I will read <this book> tomorrow ] ]  
Topic Comment

with the silent copy in object position notated within angled brackets (on traces as silent copies see Chomsky 1995, Sportiche 2007, a.o.). Each head expresses a single property, we do not have complex heads simultaneously assigning to their dependents the complex of properties “patient of the verb *and* topic of the clause”: natural languages opt for local simplicity, simple featural specifications on heads and local attribution of simple interpretive properties, even though the price to pay is a certain increase of global complexity, a richer functional structure and the multiple occurrence (or “movement”) of an element in distinct structural positions. Similar considerations hold for other types of A’ constructions such as focus, questions, relatives, exclamatives, comparatives, etc.

A brief comment on representations like (4). The postulation of a Top head is immediately supported by the fact that in many languages a Top marker is in fact morphologically realized, i.e. Gungbe *yà* (Aboh 2004, 2007), Japanese *wa* (for a particular kind of topic), etc. A partial analogy can be drawn between such left-peripheral markers for scope-discourse semantic properties (topic, focus, Q, etc.) and inherent case for argumental properties (instrumental, locative, benefactive,...): both morphosyntactic entities mark certain interpretive properties of one or the other kind, and both may superficially vary across languages in that they may or may not have a morphophonological realization.

This conception of A’ configurations implements in a very straightforward way the minimalist guideline according to which movement is a device to express an interface effect, and, more generally, that linguistic computations are driven by the satisfaction of certain expressive needs of the interface systems (Fox 2000, Reinhart 2006). Among the advantages of this way of looking at things is the fact that A’ movement conforms to the general fact that movement is formally triggered by the featural constitution of a c-commanding head. More importantly, this conception makes possible a very transparent approach to the interface between syntax and semantics-pragmatics: peripheral functional heads can be seen as overt “flags” carrying very transparent instructions to the interface systems on how their immediate dependents are to be interpreted.

An objection which is sometimes raised against this view is that it seems to threaten the thesis of the autonomy of syntax. Granting the historical importance of the autonomy thesis in the process of properly structuring a rigorous and well-defined theory of syntax, we fail to see the force of this objection. First of all, we do not see why this conception should be perceived as more of a threat to the autonomy of syntax than the Theta Criterion, or the Projection Principle, or the theta-related character of inherent case assignment, or any other principle aiming at illustrating the transparency (ultimately, the simplicity) of the mapping between form and interpretation. Secondly, we fail to see any empirical or conceptual advantage in a system of syntactic heads solely using interpretively opaque elements such as Inflection rather than Tense or Aspect, Complementizer rather than Focus, Topic or Q marker, and so

on. Conceptually, a transparent mapping surely is the null hypothesis, any deviation from which would require clear supporting evidence. Empirically, the transparent view is supported by much overt morphological evidence found across languages. Our own feeling is that the issue of cartography and the autonomy thesis should be looked at in the diametrically opposite perspective. The cartographic studies can be seen as an attempt to “syntacticize” as much as possible the interpretive domains, tracing back interpretive algorithms for such properties as argument structure (Hale and Keyser 1993 and much related work), scope, and informational structure (the “criterial” approach defended in Rizzi 1997 and much related work) to the familiar ingredients uncovered and refined in half a century of formal syntax. To the extent to which these efforts are empirically supported, they may shed light not only on syntax proper, but also on the structure and functioning of the cognitive systems at the interface with the syntactic module.

## 6. Hierarchies, Syntax and Semantics

Cartographic studies have drawn detailed structural maps holding across languages, and have made it plausible that core aspects of the functional structure may be universal. One important question which arises is: where does the hierarchy, and its universal properties, come from? It is hard to imagine that the hierarchy may be an irreducible property of UG, disconnected from any other aspect of human cognition; it is also hard to believe that the hierarchy may be a purely arbitrary “cultural” property, rediscovered by every language learner in the same form, language after language, on the basis of pure inductive learning. So, there must be some principles determining the hierarchical sequence, and guiding the child to “rediscover” it in the course of language acquisition.

In some cases, it is very plausible that certain aspects of the hierarchy (like the relative height, or scope, of the elements that constitute it) depend on independent properties of their semantics, even though precisely what elements make up the hierarchy may simply be the result of the linguistic crystallization of a particular set of cognitive categories among the many more that simply do not find a grammatical encoding in UG. Consider for instance the fact that many languages allow a proliferation of left peripheral topics, while the left-peripheral focus position (if a language uses it at all) appears to be invariably unique. It is plausible that this difference may be derivable from the very interpretive properties of topic and focus (Rizzi 1997). If the left-peripheral focal head assigns the focus interpretation to its specifier, and the presupposition interpretation to its complement,

- (5) [ XP [ Foc YP ] ]  
       Focus                   Presupp.

then a recursion of (5), e.g. with YP headed by a Foc head, would yield an interpretive clash: YP would be presupposed, but would contain a focal constituent. So, the recursion is barred. On the other hand nothing blocks the recursion of a topic phrase: no interpretive property of the comment excludes that it may in turn have a topic-comment structure. Individual languages may opt for a unique topic position as a matter of parametric choice, e.g., in V-2 languages, but there is no universal prohibition stemming from a plausible interpretive constraint in this case. Another example may be the fact that, in the structure of the IP, the element expressing epistemic modality typically is higher than tense: presumably the modality must be

evaluated over a complete proposition, including the tense specification. Similar considerations may hold for the universal order epistemic modality > root modality, tense > aspects, etc.

In other cases, aspects of the hierarchy may be determined by syntactic constraints on movement. Consider for instance the fact that in many languages left-peripheral topic and focus can cooccur in the fixed order Topic – Focus (e.g., Hungarian: Kiss 1995). This may be due to the fact that Focus often requires movement of the inflected verb to C (possibly a property related to the quantificational character of Focus), while Topic does not. In a language requiring inversion with Focus, the order Focus Topic would then be blocked by the impossibility of moving the inflected verb past the Topic head, ultimately a case of the Head Movement Constraint (Travis 1984). The validity of a syntactic account of this sort is supported by the fact that the order Focus Topic seems indeed to be possible in a language like Italian, which does not require verb movement with focus. This strongly supports the view that in this case there is no general scope property enforcing a particular order. Along similar lines, one can observe that if a position has island-creating properties, it must be higher than other positions filled by movement: so, for instance, the Hanging Topic (which has island creating properties) must precede the ordinary topic expressed in Romance by Clitic Left Dislocation (Cinque 1990, Benincà & Poletto 2004). On certain connections between the theory of movement and the hierarchy see Abels (2007).

Going back to the constraining effects of semantics, a qualification is needed. Clearly, it is not the case that any imaginable semantic property or distinction can be grammaticalized, expressed by a functional element, a special morphology, a morphosyntactic feature<sup>10</sup>: there is a fairly restrictive universal set of properties that can be expressed by the functional elements entering into the different hierarchies associated to clauses and phrases. So, syntax is heavily constrained by semantics, but is not totally malleable: on the one hand, it respects purely syntactic constraints (such as locality effects); on the other hand, it is often the case that a syntactic device has a core semantic function, but it often acquires an independent life of its own, as it were, extending its scope well beyond its core semantic function. Consider, for instance, grammatical gender, whose core function is the expression of natural gender, but which gets extended to express an arbitrary classification in the nominal lexicon; the expression of tense, situating the event in time with respect the utterance time, but extending to become an obligatory property of the clausal hierarchy, so that also a tenseless mathematical or logical truth must be expressed via a tensed sentence; the subject – predicate articulation expressing the “aboutness” relation, but becoming a general, obligatory property of clausal structures, which forces the use of expletives if the event is not presented about a particular argument; etc. Syntax is organized to express meaning, but does not dissolve into the mere organization of meaningful units: UG expresses the possible items of the functional lexicon and the way in which they are organized into hierarchies, tailored on the needs of the expression of meanings, but not reducing to them.

---

<sup>10</sup> For example, in the extended projection of a NP, we find evidence for different types of quantifiers, demonstratives, numerals, for functional categories of diminutivization, numerical approximation, etc., but we never find expressed, it seems, distinctions relating to the magical or non magical character of a number (as opposed to its approximation), nor specialized forms meaning dear-to-me, (dear-to-you,) not-dear-to-me-and-you parallel to the universal demonstrative distinctions close-to-me, (close-to-you,) not-close-to-me-and-you. One could easily multiply such theoretically possible, yet non-existing, functional distinctions (also see Cinque 1999,224fn.10, and related text).

## References

- Abels, Klaus. 2007. Towards a Restrictive Theory of (Remnant) Movement. CASTL, Tromsø.
- Abney, Steven P. 1987. *The English noun phrase in its sentential aspect*. Ph.D. Dissertation, MIT
- Aboh, Enoch. 2004. *The Morphosyntax of Complement-Head Sequences*. New York: Oxford University Press
- Aboh, Enoch. 2007. Information structuring begins with numeration. ACLC, University of Amsterdam.
- Asbury, Anna, Jakub Dotlačil, Berit Gehrke, Rick Nouwen, eds., 2008. *Syntax and Semantics of Spatial P*. Amsterdam: Benjamins
- Baker, Mark. 1985. The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry* 16.373-416
- Baker, Mark. 2003. *Lexical Categories*. Cambridge: Cambridge University Press.
- Belletti, Adriana. 1990. *Generalized Verb Movement*. Turin: Rosenberg and Sellier
- Belletti, Adriana. ed., 2004. *Structures and Beyond. The Cartography of Syntactic Structures, vol.3*, New York: Oxford University Press
- Belletti, Adriana. 2008. *Structures and Strategies*. London: Routledge
- Benincà, Paola. 2001. The Position of Topic and Focus in the Left Periphery. In G.Cinque and G.Salvi, eds., *Current Studies in Italian Syntax. Essays offered to Lorenzo Renzi*. 39-64. Amsterdam: Elsevier North-Holland
- Benincà, Paola. 2006. A detailed map of the left periphery of medieval Romance. In R.Zanuttini, H.Campos, E.Herburger and P.Portner, eds., *Crosslinguistic research in syntax and semantics: Negation, tense and clausal architecture*. 53-86. Washington (D.C.): Georgetown University Press
- Benincà, Paola and Cecilia Poletto. 2004. Topic, Focus, and V2. Defining the CP Sublayers. In L.Rizzi, ed., *The Structure of CP and IP. The Cartography of Syntactic Structures, vol.2*. 52-75. New York: Oxford University Press.
- Benincà, Paola and Nicola Munaro, eds., forthcoming. *Mapping the Left Periphery. The Cartography of Syntactic Structures, vol.5*, New York: Oxford University Press
- Bianchi Valentina. 2006. On the syntax of personal arguments. *Lingua* 116, 2023-2067.
- Bobaljik, J.D. 1999. Adverbs: the hierarchy paradox. *Glott International* 4: 27-28.
- Bobaljik, Jonathan D. and Höskuldur Thráinsson. 1998. Two heads aren't always better than one *Syntax* 1: 37-71
- Bocci, Giuliano. 2004. Contrastive Focalisation on Topics and Preverbal Subjects in Italian. *Rivista di Grammatica Generativa* 29.3-60
- Brugè, Laura. 2002. The Position of Demonstratives in the Extended Nominal Projection. In G.Cinque, ed., *Functional Structure in DP and IP. The Cartography of Syntactic Structures. Vol.1*.15-53. New York: Oxford University Press.
- Caramazza, Alfonso and Kevin Shapiro. 2004. Language Categories in the Brain. In Belletti, ed., *Structures and Beyond. The Cartography of Syntactic Structures, vol.3*. 15-38. New York: Oxford University Press
- Cardinaletti, Anna. 2004. Towards a Cartography of Syntactic Positions. In L.Rizzi, ed., *The Structure of CP and IP. The Cartography of Syntactic Structures, vol.2*. 115-165. New York: Oxford University Press
- Carlson, Greg N. 1983. Marking Constituents. In F.Heny and B.Richards, eds., *Linguistic Categories: Auxiliaries and Related Puzzles. Vol.1*. 69-98. Dordrecht: Reidel

- Chesi, Cristiano. 2005. *Phases and cartography in linguistic computation: toward a cognitively motivated computational model of linguistic competence*. PhD Dissertation, University of Siena.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton
- Chomsky, Noam. 1986. *Barriers*. Cambridge, Mass.: MIT Press
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press
- Chomsky, Noam. 2001. Derivation by Phase. In M.Kenstowicz (ed.) *Ken Hale: A Life in Language*. 1-52. Cambridge, Mass.: MIT Press
- Cinque, Guglielmo. 1990. *Types of A-bar Dependencies*. Cambridge, Mass.: MIT Press
- Cinque, Guglielmo. 1994. On the Evidence for Partial N-movement in the Romance DP. In G.Cinque, J.Koster, J.-Y.Pollock, L.Rizzi and R.Zanuttini (eds.) *Paths Towards Universal Grammar. Studies in Honor of Richard S. Kayne*. 85-110. Washington (D.C.): Georgetown University Press
- Cinque, Guglielmo. 1999. *Adverbs and Functional Heads. A Cross-linguistic Perspective*. New York: Oxford University Press
- Cinque, Guglielmo, ed., 2002. *Functional Structure in DP and IP. The Cartography of Syntactic Structures, vol.1*, New York: Oxford University Press
- Cinque, Guglielmo. 2004. Issues in adverbial syntax. *Lingua* 114.683-710
- Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its Exceptions. *Linguistic Inquiry* 36. 315-332
- Cinque, Guglielmo. 2006a. *Restructuring and Functional Heads. The Cartography of Syntactic Structures, vol.4*, New York: Oxford University Press
- Cinque, Guglielmo. 2006b. Are all languages 'Numeral Classifier Languages'? *Rivista di grammatica generativa* 31.119-122
- Cinque, Guglielmo. 2008. The Fundamental Left-Right Asymmetry of Natural Languages. In S.Scalise et al., eds., *Universals of Language Today*. Dordrecht: Springer  
(<http://dspace-unive.cilea.it/handle/10278/214>)
- Cinque, Guglielmo. To appear. *The Syntax of Adjectives. A Comparative Study*.
- Cinque, Guglielmo and Luigi Rizzi. To appear. *Mapping Spatial PPs. The Cartography of Syntactic Structures, vol.6*. New York: Oxford University Press
- Cruschina, Silvio. 2006. Informational focus in Sicilian and the left periphery. In M.Frascarelli, ed., *Phases of Interpretation*. 363-385. Berlin: Mouton de Gruyter
- Dixon, Robert M.W. 1982. *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton Publishers
- Dixon, Robert M.W. 2004. Adjective classes in typological perspective. In R.M.W., Dixon, A. Aikhenvald, eds., *Adjective Classes: A Cross-linguistic Typology*. 1-49. Oxford: Oxford University Press
- Durrleman, Stephanie. 2007. *The Syntax of Jamaican Creole – A Cartographic Perspective*. PhD Dissertation, University of Geneva.
- Emonds, J. (1978) "The verbal complex V'-V in French". *Linguistic Inquiry*, 9:151-175.
- Endo, Yoshio. 2007. *Locality and Information Structure – A Cartographic Approach to Japanese*. Amsterdam: Benjamins.
- Fox, Danny. 2000. *Economy and Semantic Interpretation*. Cambridge, Mass.: MIT Press.
- Frascarelli, Mara and Roland Hinterhölzl. 2007. Types of Topics in German and Italian. In S. Winkler and K. Schwabe, eds., *On Information Structure, Meaning and Form*. 87-116. Amsterdam: Benjamins.

- Frascarelli, Mara and Annarita Puglielli. To appear. Focus in the Force-Fin System. Information Structure in Cushitic Languages. In Enoch Aboh, K. Hartmann and M. Zimmermann, eds., *Focus Strategies: Evidence from African languages*. Berlin: Mouton de Gruyter.
- Fukui, Naoki. 1995. *Theory of Projection in Syntax*. Stanford, CA: CSLI Publications.
- Furuya, Kaori. 2008. DP hypothesis for Japanese “bare” noun phrases. *University of Pennsylvania Working Papers in Linguistics* 14.149-162
- Giusti, Giuliana. 2002. The Functional Structure of Noun Phrases. A Bare Phrase Structure Approach. In G.Cinque, ed., *Functional Structure in DP and IP. The Cartography of Syntactic Structures, vol.1*. 54-90. New York: Oxford University Press
- Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J.Greenberg, ed., *Universals of language*. 73–113. Cambridge, Mass.: MIT Press.
- Grewendorf, Günther. 2002. Left Dislocation as Movement. In S. Manck & J. Mittelstaedt, eds., *Georgetown University Working Papers in Theoretical Linguistics* 31-81
- Grillo, Nino. 2008. Generalized Minimality – Syntactic Underspecification in Broca’s Aphasia. PhD Dissertation, University of Utrecht, University of Siena. Distributed by LOT
- Haegeman, Liliane. 2006. Argument Fronting in English, Romance CILD and Left Periphery. In R.Zanuttini, H.Campos, E.Herburger and P.Portner, eds., *Crosslinguistic research in syntax and semantics: Negation, tense and clausal architecture*. 27-52. Washington (D.C.): Georgetown University Press
- Hale, Kenneth and Jay S. Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In K. L. Hale and S. J. Keyser eds. *The View from Building 20: A Festschrift for Sylvain Bromberger*. 53-109. Cambridge, Mass.: MIT Press.
- Heine, Bernd and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press
- Joos, Martin, ed., 1957. *Readings in Linguistics I*. New York: American Council of Learned Society
- Karimi-Doostan, G. 1997. *Light Verb Constructions in Persian and Kurdish*. Ph.D.Dissertation, University of Essex.
- Kayne, Richard S. 2003. Silent years, silent hours. In L.-O.Delsing, C.Falk, G.Josefsson, and H.Á. Sigurðsson, eds., *Grammatik i fokus. Festschrift till Christer Platzack den 18 november 2003*. 209-226. Lund: Wallin & Dalholm
- Kayne, Richard S. 2005a. Some notes on comparative syntax, with special reference to English and French. In G.Cinque and R.S.Kayne, eds., *The Oxford Handbook of Comparative Syntax*. 3-69. New York: Oxford University Press
- Kayne, Richard S. 2005b. *Movement and Silence*. New York: Oxford University Press
- Kayne, Richard S. 2006. On Parameters and on Principles of Pronunciation. In H.Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz and J. Koster, eds., *Organizing Grammar. Linguistic Studies in Honor of Henk van Riemsdijk*. 289-299. Berlin: Mouton de Gruyter
- Kiss, Katalin, ed., 1995. *Discourse–Configurational Languages*. New York: Oxford University Press.
- Koopman, Hilda and Dominique Sportiche. 1991. The Position of Subjects. *Lingua* 85.211-258

- Koopman, Hilda and Anna Szabolcsi. 2000. *Verbal Complexes*. Cambridge, Mass.: MIT Press
- LaPolla, Randy and Dory Poa (2002) "On Describing Word Order", ms., City Univ. of Hong Kong (downloadable from <http://personal.cityu.edu.hk/%7Ectrandy/describingwo.pdf>).
- Megerdooimian, Karine. n.d. <http://crl.nmsu.edu/Research/Projects/shiraz/ling/morph.html>
- Manzini, M. Rita and Leonardo Savoia. 2005. *I dialetti italiani e romanzi*. Alessandria: Edizioni dell'Orso
- Miceli, Gabriele, M.Caterina Silveri, Giampiero Villa and Alfonso Caramazza. 1984. On the Basis for the Agrammatic's Difficulty in Producing Main Verbs. *Cortex* 20.207-220.
- Mioto, Carlos. 1999. A periferia esquerda no português brasileiro. Ms., Universidade Federal de Santa Catarina, University of Siena.
- Nicolis, Marco. 2008. Reinforcers in Italian: DP-internal and Left Peripheral. To appear in Proceedings of CLS 41.
- Nilsen, Øystein. 2003. *Eliminating positions: syntax and semantics of sentential modification*. PhD dissertation, Utrecht University.
- Pawley, Andrew. 2006. Where have all the verbs gone? Remarks on the organisation of languages with small, closed verb classes. Paper presented at the *11th Biennial Rice University Linguistics Symposium, 16-18 March 2006* ([http://www.ruf.rice.edu/~lingsymp/Pawley\\_paper.pdf](http://www.ruf.rice.edu/~lingsymp/Pawley_paper.pdf)).
- Pereltsvaig, Asya. 2007. On the Universality of DP: A View from Russian". *Studia Linguistica* 61.59-94
- Poletto, Cecilia. 2000. *The Higher Functional Field. Evidence from Northern Italian Dialects*. New York: Oxford University Press
- Pollock, Jean-Yves. 1989. Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20.365-424
- Progovac, Ljiljana. 1998. Determiner Phrase in a language without determiners. *Journal of Linguistics* 34.165-179
- Ramchand, Jillian. 2008. *Verb Meaning and the Lexicon: A first Phase Syntax*. Cambridge: Cambridge University Press
- Reinhart, Tanya. 2005. *Interface Strategies – Reference-set Computation*. Cambridge, Mass.: MIT Press
- Rizzi, Luigi. 1985. Two notes on the linguistic interpretation of Broca's aphasia. In M.-L. Kean, ed., *Agrammatism*. 153-164. New York: Academic Press.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Rizzi, Luigi. 1997. "The Fine Structure of the Left Periphery", in L.Haegeman (ed.) *Elements of Grammar*. 281-337. Amsterdam: Kluwer
- Rizzi, Luigi. 2001. On the position "Int(errogative)" in the left periphery of the clause. In G.Cinque and G.Salvi, eds., *Current Studies in Italian Syntax. Essays offered to Lorenzo Renzi*. 287-296. Amsterdam: Elsevier North-Holland
- Rizzi, Luigi. ed. 2004a. *The Structure of CP and IP. The Cartography of Syntactic Structures, vol.2*, New York: Oxford University Press
- Rizzi, Luigi. 2004b. Locality and Left Periphery. In A.Belletti, ed., *Structures and Beyond. The Cartography of Syntactic Structures, vol.3*, New York: Oxford University Press
- Rizzi, Luigi and Ur Shlonsky. 2007. "Strategies of Subject Extraction", in H.-M.Gärtner and U. Sauerland (eds). *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. 115-16. Berlin: Mouton de Gruyter

- Schachter, Paul. 1985. Parts-of-speech Systems. In T.Shopen, ed., *Language Typology. A Syntactic Description. Vol.I: Clause Structure*. 3-61. Cambridge: Cambridge University Press
- Schachter, Paul and Timothy Shopen. 2007. Parts-of-speech Systems. In T.Shopen, ed., *Language Typology. A Syntactic Description. Vol. I: Clause Structure. Second Edition*. 1-60. Cambridge: Cambridge University Press
- Schweikert, Walter. 2005. The Order of Prepositional Phrases in the Structure of the Clause. Amsterdam: Benjamins
- Scott, Gary-John. 2002. Stacked Adjectival Modification and the Structure of Nominal Phrases. In G.Cinque, ed., *Functional Structure in DP and IP. The Cartography of Syntactic Structures. Vol.1*.91-120. New York: Oxford University Press.
- Shlonsky, Ur. 1997. *Clause Structure and Word Order in Hebrew and Arabic*. New York: Oxford University Press
- Shlonsky, Ur. 2000. Subject positions and copular constructions. In H.Bennis, M. Everaert and E. Reuland, eds., *Interface strategies*. 325-347. Amsterdam: Royal Netherlands Academy of Arts and Sciences
- Sigurðsson, Halldór Á. 2000. The Locus of Case and Agreement. *Working Papers in Scandinavian Syntax* 65.65-108.
- Sportiche, Dominique. 2007. Division of Labor between Merge and Move: Strict Locality of Selection and Apparent Reconstruction Paradoxes. Ms., UCLA.
- Starke, Michal. 2001. Move Dissolves into Merge. PhD Dissertation, University of Geneva.
- Stutzman, Verna. 1997. A study of the Lou verb phrase. M.A. Dissertation, University of Manitoba (<http://mspace.lib.umanitoba.ca/handle/1993/1324>)
- Svenonius, Peter. 2008a. The position of adjectives and other phrasal modifiers in the decomposition of DP. In L.McNally and C.Kennedy, eds., *Adjectives and Adverbs: Syntax, Semantics, and Discourse*. Oxford: Oxford University Press
- Svenonius, Peter. 2008b Projections of P. In Asbury,A., J.Dotlačil, B.Gehrke and R.Nouwen, eds., *Syntax and Semantics of Spatial P*. 63-84. Amsterdam: Benjamins (<http://ling.auf.net/lingBuzz/>).
- Thráinsson, Höskuldur. 1996. On the (non)-universality of functional categories. In W.Abraham, S. D. Epstein, H. Thráinsson and C. J-W. Zwart (eds.), *Minimal Ideas: Syntactic Studies in the Minimalist Framework*, 253–281. Amsterdam: John Benjamins
- Travis, Lisa 1984. Parameters and Effects of Word Order Variation. PhD Dissertation, MIT.
- Tsai, Dylan. 2007. Left Periphery and Why-How Alternations. National Tsing-Hua University, Taiwan.
- Van Craenenbroeck, Jeroen. 2006. Transitivity failures in the left periphery and foot-driven movement operations. *Linguistics in the Netherlands* 23: 52-64
- Zwart, Jan-Wouter. 2006. Uncharted territory? Towards a non-cartographic account of Germanic syntax. Paper presented at the 21<sup>st</sup> Comparative Germanic Syntax Workshop held at Santa Cruz. To appear in the Proceedings to be published by John Benjamins (<http://www.let.rug.nl/~zwart/docs/cgsw21proc.pdf>).

# Minimizers and quantifiers: a window on the development of negative markers

JACOPO GARZONIO

Università di Padova  
jacopo.garzonio@unipd.it

CECILIA POLETTO

Università di Venezia  
poletto@unive.it

Abstract: In this work we examine the diachronic development of two types of negative markers showing that their semantic evolution is similar, as they have both undergone a stage when they become monotone decreasing quantifiers. However, their position is different: the first type, called minimizers, are part of a complex DP, while the second is a bare quantifier. Their original syntactic position determines the final position where they are merged when they are reanalyzed as standard negative markers: the bare quantifier is located in a position dedicated to bare QPs, while minimizers move to a higher position in the low IP area.

## 1. Introduction

In her seminal work on negation in Italo-Romance, Zanuttini (1997) has shown that negative morphemes in Italo-Romance varieties can be subdivided into four main classes according to the position they occupy in the clause structure. These positions, labelled as Neg1, Neg2, etc. are represented in (1)

(1) [NegP 1 [TP 1 [NegP 2 [TP 2 [NegP 3 [Asp Perfective [ NegP 4 ]...]

Neg1 is the position of preverbal negations like standard Italian *non*; Neg2 and Neg3 are postverbal negations which are differentiated by the fact that the former (for instance Piedmontese *pa*) generally precedes Tense Anterior adverbs like ‘already’, while the latter (for instance Piedmontese *nen*) appears inside the field of aspectual adverbs (Cinque 1999).

- (2) a. A l'è pa gia andait a ca'. (Piedmontese, Zanuttini 1997, 70)  
SCL SCL is NEG already gone to home  
'He has not already gone home.'  
b. \*?A l'è nen gia andait a ca'.  
SCL SCL is NEG already gone to home  
'He has not already gone home.'

Neg4 are sentence-final negations like Milanese *no*. Neg2 and Neg3 elements share an interesting property: both classes are made of elements derived from quantifiers originally merged in object position. These elements have been grammaticalized and in many cases are not identical to the corresponding quantifier in synchrony. In other cases the quantifier has totally disappeared in the modern variety (it is the case of standard Italian *mica*).

In this paper we observe that, even if both Neg2 and Neg3 items were in origin object quantifiers, the two classes derive from two distinct types of quantifiers. In fact, Neg2 negations derive from grammaticalized minimizers, which in origin appeared in complex DP structures, followed by an obligatory PP complement, while Neg3 negations derive from the bare negative inanimate quantifier corresponding to English ‘nothing’. Thus, the different etymological and syntactic origin of the two quantifiers is somehow related to a different syntactic position in synchrony when they are reanalyzed as negative markers.

As we will show further more in detail, a relevant factor for the syntactic properties of postverbal negations is whether they are the standard negative morpheme or have a more complex interpretation generally referred to as “emphatic negation” (here, we refer to it as non standard negation because we do not think that emphasis play any role in the distribution of these elements). The status of standard negation can be considered the final stage of the grammaticalization process. Analogue elements have reached different stages of grammaticalization in different varieties. Thus, a comparative analysis of the same etymological type of negative marker in different dialects and of different types in the same variety can shed light on the diachronic development of these elements. In this work we exploit exactly the parallel between geographic and diachronic variation to investigate the development of the various negative markers by means of living dialects.

Standard negations are compatible with any type of sentence or predicate<sup>1</sup>, while non-standard negations need specific contexts. These appropriate contexts vary for different negations, but we provide evidence for the fact that they all undergo a stage in which they require a scale. We will argue that scalarity is the property that triggers the grammaticalization of quantifiers into negative markers. In other words, both types of elements we discuss here become scalar negative markers before becoming polar negation.

The paper has the following structure: in section 2 we examine the distribution of minimizer negation tracing its diachronic path from Old Italian to Modern Italian; we will show that, before becoming a non standard negative marker, this element was a quantifier requiring a scale. The same is true of the second type of negative marker, *gnente*, which is so to speak “caught in the act” of evolving into a negative marker from a scalar quantifier. The analysis of elements like *gnente* ‘nothing’ in a living dialect will provide empirical evidence in favour of our main claim, namely that scalarity is relevant in the grammaticalization process. This is what we discuss in section 4.

## 2. Minimizer negation

Many Italo-romance varieties display negative markers which derive from nouns expressing a small quantity, which were originally lexically related to their

---

<sup>1</sup> We will not deal with constituent negations in this paper.

complement PP. We adopt here the term “Minimizer Negation” to indicate these elements (from now on simply m-negation). A partial list of m-negations, with their original referential meaning, is given in (3):

- (3) ‘step’: Piedmontese/Valdotain *pa*;  
‘crumble’: Emilian *brisa*; Milanese *miga/minga*; Veneto *mina/mia*; Italian *mica*;  
‘bite’: Romansh *bucca*;  
‘point’: Florentine *punto*;  
‘thread’: Salentino *filu*;  
‘flower’: Old Florentine *fiore*;  
‘drop’: Old Venetian *gozo*.

Some of these elements have become the marker of standard negation (like *pa* in Piedmontese, or *bucca* in Romansh), while others have specialized as non-standard negative markers.<sup>2</sup> There are two main groups of non-standard negations: on the one hand, adverbs which reinforce preverbal negation and can be considered as equivalent to ‘at all’ in English - on the other hand, negative morphemes that are used to express the fact that an explicit or implicit assumption made by the interlocutor is wrong<sup>3</sup>. The difference between these two types is shown in (4).

- (4) a. Mario un ha punto mangiato. (Florentine)  
M. NEG has NEG eaten  
‘M. has not eaten at all.’  
b. Mario un ha mica mangiato. (Florentine)  
M. NEG has NEG eaten  
‘M. has not eaten (as you have said/think).’

M-negations behave like a natural class of elements, since they display common syntactic properties. Usually they appear higher than adverbs encoding Aspect and Tense Anterior. This position is occupied by both standard and non standard negative markers. Zanutini (1997) proposes that this position is the specifier of a functional projection which she labels Neg2. Some examples which show that m-negations precede adverbs corresponding to ‘already’ are given in (5), while the relevant part of the clause structure of Zanutini’s analysis (based on Cinque’s (1999) hierarchy of adverbials) is provided in (6):

- (5) a. A l’ha pa gia ciamà. (Piedmontese, from Zanutini 1997)  
SCL SCL has NEG already called  
‘He has not already called.’  
b. I n’an briza beli ciamà. (Emilian, from Colombini 2007, § 5.6.1)  
SCL NEG have NEG already called  
‘They have not already called.’

---

<sup>2</sup> As mentioned in the introduction, with the term non-standard negative markers we indicate negative elements which have a more complex interpretation with respect to the unmarked sentential negation.

<sup>3</sup> Italian *mica* is one of these elements, and it has been called “presuppositional negation” by Cinque (1976). In more recent work Penello-Pescarini (2008) use an updated terminology and analyze the distribution of *mica* in terms of implicatures.

- c. Non hanno mica già chiamato. (Italian, from Cinque 1999)  
 NEG have NEG already called  
 ‘They have not already called.’
- d. Elts an buca magliau trasora. (Romansh, from Manzini-Savoia 2005)  
 SCL have NEG eaten already  
 ‘They have not already eaten.’
- (6) [...[FP *neg* [T Anterior *already* [Asp Terminative *anymore* [Asp Perfective *always* [Asp Completive *tutto*]]]]]]

Contrary to q-negations that we will analyze in the following section, m-negations can cooccur with negative quantifiers, even if adjacent:

- (7) a. A’m dis pa gnente. (Piedmontese, from Zanuttini 1997)  
 SCL me tells NEG nothing  
 ‘She does not tell me anything.’
- b. A veddu pa gnun. (Piedmontese, from Zanuttini 1997)  
 SCL see neg nobody  
 ‘I do not see anyone.’
- c. Al n’i briza arivà endsun. (Emilian, from Colombini 2007, § 2.3)  
 SCL NEG is NEG come nobody  
 ‘Nobody has come.’
- d. A n vegn mia ninsün. (Mantuan, from Manzini-Savoia 2005)  
 SCL NEG comes NEG nobody  
 ‘Nobody comes.’

Usually m-negations appear in postverbal position, both in varieties which have (8a-b) and varieties which lack (8c-d) a preverbal negative marker:

- (8) a. Non sente mica. (Italian)  
 NEG hears NEG  
 ‘He cannot hear.’
- b. No ssienti filu? (Salentino, from Rohlf 1969)  
 NEG hear NEG  
 ‘Don’t you hear?’
- c. Lo film l’era pa dzen. (Valdotain, from Zanuttini 1997)  
 the movie SCL was NEG beautiful  
 ‘The movie wasn’t good.’
- d. El l’ha minga scrivuu. (Milanese, from Zanuttini 1997)  
 He SCL has NEG written  
 ‘He has not written.’

However, in some varieties where the marker of standard negation is preverbal, m-negations can optionally appear in preverbal position. In this case, m-negation is the only negative element in the sentence (leaving aside additional n-words, as in (9c)), although it most probably does not occupy the same position of the preverbal negative marker, which is a head:

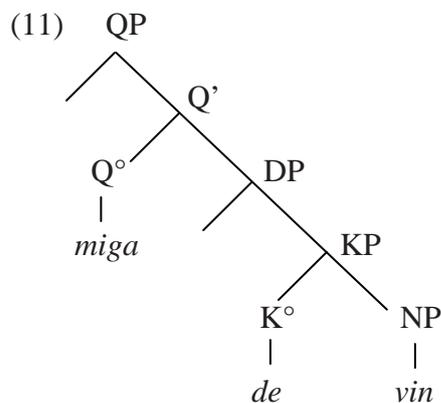
- (9) a. Mica sente quello che dici. (Italian)  
 NEG hears that that say  
 ‘He does not hear what you are saying.’  
 b. Filu sapimu la libbirtà. (Salentino, from Rohlf’s 1969)  
 NEG know the freedom  
 ‘We don’t know what freedom is.’  
 c. Mica abbiamo visto nessuno. (Italian)  
 NEG have seen nobody  
 ‘We have not seen anybody.’

2.1 Development of M-Negations

All m-negations we deal with in this paper derive from nominal minimizers, that is nouns denoting “a negligible number, amount, or part of something” (Kiparsky-Condoravdi 2006, 2). In origin they appeared in semantically restricted predicates (as, for example, ‘not eat a crumble of bread’, ‘not drink a drop of water’, not move an inch (a step), etc.). In these contexts they were indefinite DPs with a prepositional complement containing another DP (the noun referring to the minimally quantified thing). These expressions can be described as emphatic negations, where emphasis is obtained negating the smallest grade of a scale. However, the minimizer retained its referential value, as shown by the lexical restriction imposed on it by the verb. In the following step of the grammaticalization, these elements have become quantifiers. Adopting Roberts-Roussou’s (2003) proposal that grammaticalization is leftward movement in the syntactic structure, we assume that in a first stage of the process, the minimizer N moves to a higher position in the DP structure, which is NumP (in a way parallel to DPs that have become n-words in French):

- (10) miga de vin ‘a little bit of wine’ (Old Milanese)  
 [DP [D Ø] [NumP [Num *miga*] [NP ~~*miga*~~ [PP *de vin*]]]]  
 (see Roberts-Roussou 2003, 153)

Then the whole structure is reinterpreted as a unique DP, with the noun inside the PP as the head of the NP, the preposition as a partitive case marker (that is a K°) and the minimizer as a classifier-like quantifier governing the DP which provides its restriction. Structural changes where XPs are reinterpreted as heads are common in diachrony (Van Gelderen 2004). The second stage of the development is represented in (11):



At this point the minimizer has become a functional element and loses the typical properties of lexical nouns: its phi-features, the possibility of being modified, the possibility of taking PP complements and, more in general, any referential content. Italian *mica*, which is a non-standard m-negation and derives from the Latin word for ‘crumble’ (*micam*) has lost all these properties. It cannot be modified (12a), cannot have a PP complement (12b) and cannot be used to express its original referential meaning (12c):

- (12) a. Non leggono (\*la minima) mica i libri gialli.  
 NEG read the minimal NEG the books yellow  
 ‘They do not read police novels.’  
 b. Non vedo mica (\*di) Mario questa sera.  
 NEG see NEG of Mario this evening  
 ‘I will not meet Mario this evening.’  
 c. \*...una mica di pane.  
 a MICA of bread

In some dialects traces of the original structure can still be found. For instance, in the Piedmontese variety of Quarna Sotto, when non-standard m-negation *mia* is used, the object can appear in the partitive case, even if it expresses a singular non-quantifiable entity:

- (13) Nə caman mia d əu te frial. (Quarna Sotto, from Manzini-Savoia 2005)  
 NEG-SCL call NEG of the your brother  
 ‘They do not call your brother.’

After having become a functional element, the minimizer can undergo a further change: it is moved outside the object position where it was merged in origin. As we have seen, m-negations occupy a precise position in the adverbial hierarchy, which is higher than Tense Anterior. At this point, the element originally quantifying over a DP has become a sentential element. Both standard and non-standard m-negations appear in this position, which, therefore, seems to be irrelevant in order to distinguish between different semantic types of negation. It is worth pointing out that - some m-negations, which, for independent reasons, have not completed the process outlined in this section, still seem to occupy a different position. For instance, the negative adverb *punto* in Florentine, which is a m-negation, follows the adverb corresponding to ‘never’:

- (14) a. \*Un m’è punto mai piaciuto. (Florentine)  
 NEG to-me is PUNTO never liked  
 b. Un m’è mai punto piaciuto.  
 ‘I never liked it at all.’

On the other hand, however, *punto* has developed a special agreeing morphology and is still used to express a minimal quantity (it can be used as a polarity quantifier, also in positive contexts, where it corresponds to ‘any’):

- (15) a. Un ho punte sorelle.  
 NEG have PUNTO.F.PL sisters  
 ‘I have no sisters.’

- b. Hai punte sorelle?  
have PUNTO.F.PL sisters  
'Do you have any sister?'

Thus, it comes as no surprise that m-negations can reach diachronically Neg2 only if they lose all their original lexical properties. We have seen that they do so in two steps: they first become functional elements quantifying over a DP and later they are reanalyzed as sentential elements.

## 2.2 *Vulgar Minimizers as M-Negations*

In this section we present some data regarding a peculiar class of elements, which share many properties with m-negations, the so-called vulgar minimizers (Postal 2003). The analysis of these special minimizers in Italo-Romance can shed light on the development of m-negations.

Vulgar minimizers differ from standard minimizers in the fact that the latter "denote minimal elements on some scale", while the former "are not narrowly restricted to particular dimensions, but can express minimality along many dimensions" (Postal 2003). Usually vulgar minimizers derive from tabooed nouns, all denoting inanimates. The more common vulgar minimizers in standard Italian are *tubo* 'pipe', *cavolo* 'cabbage', *cazzo* 'dick', etc., with many regional or dialectal variants. They are similar to nouns that have become m-negations because they are scalar: they express that the action or the property denoted by the predicate does not reach a minimal degree of acceptability:

- (16) a. Non ho capito un tubo.  
NEG have understood a pipe  
'I have understood nothing.'
- b. Sono andati via senza aver fatto un cavolo.  
are gone away without have done a cabbage  
'They left having done nothing.'
- c. \*Non abbiamo incontrato un tubo.  
NEG have met a pipe  
'We met nobody.'

(16c) shows that these elements can be used appropriately only in the object position of predicates not selecting animate objects. If an animate is required, it is possible to use some nouns (with strict semantic restrictions), but animate nouns lack many of the properties of vulgar minimizers. Some of such properties which are relevant for our analysis are: vulgar minimizers can be modified, for example by adjectives, and can have PP complements (in this case, they have quantificational function over the noun inside the PP):

- (17) a. Non ho capito un emerito tubo.  
NEG have understood a emeritous pipe  
'I have understood nothing.'
- b. Non capiscono un tubo di economia.  
NEG understand a pipe of economics  
'They understand nothing of economics.'

Furthermore, in varieties of Central Italy, it is possible to use vulgar minimizers as the unique negative element of a clause:

- (18) a. So un cavolo chi viene stasera.  
 know a cabbage who comes this evening  
 ‘I do not know who will come this evening.’  
 b. Sono indipendenti un cavolo.  
 are independent a cabbage  
 ‘They are not independent.’

Thus, vulgar minimizers in Italo-Romance are a very peculiar class of elements: on the one hand, they are in some sense more similar to nouns, since they can be modified and can take PP complements, what is not possible anymore for m-negations; on the other hand, they can be the unique negative element in a sentence, that is a purely functional item. As we have seen in the previous paragraph, m-negations can be the unique negation only in some varieties and only when they appear in preverbal position. Vulgar minimizers, on the contrary, as the examples in (16) show, remain in postverbal position. We leave a deeper analysis of these elements to future research, but it is clear that they must be considered together with m-negations and in some sense they appear to be in one of the precedent stages of the historical development of minimizers into m-negations<sup>4</sup>.

### 2.3 On the scalar property of M-Negations

As we have seen m-negations derive from nouns merged in object position. At first these objects had the function of scalar elements. To say that someone has not eaten even a crumble of bread or a bight of food means that the act of eating has not reached the minimal degree to be considered as fulfilled. In some sense the following DP provides a scale and the minimizer object expresses the minimal degree of this scale. This fact explains why in old varieties these elements were possible also in positive contexts:

- (19) On sté de scisceri e miga de vin d'intrà. (Old Milanese, from Vai 1996)  
 a *staio* (20l) of chickpeas and MICA of wine of income

As we have seen, during the development of m-negations, the minimizers are not inserted as object nouns anymore. At the end of the diachronic process, that is when m-negations become the morpheme of standard negation in a given variety, these elements lose the property of expressing a minimal degree. In other words, standard m-negations are not themselves scalar. However, in the case of non-standard m-negations, scalarity seems still to be a relevant factor. M-negations used to reinforce standard negation have a function similar to adverbs like ‘at all’. In this case, even if m-negation does not occupy the object position, the predicate is denoted as not fulfilled since a minimal “amount” of the act or the property expressed by it has not been reached. Vulgar minimizers have a similar semantics, with the difference that they express the fact that a minimal degree of acceptability or convenience has not been reached. “Presuppositional” m-negations are different, but we argue that scalarity

<sup>4</sup> Notice that Postal (2003) argues that vulgar minimizers (at least in American English) are not negations. Furthermore, they never become the standard negation, while many varieties in Northern Italy have a standard m-negation. This fact could be related to their connotative meaning, which seems to be incompatible with pure functional items.

is relevant also for this type of negation. As we have seen, these negations are used to deny some assumption considered to be wrong by the speaker. If we postulate that the propositional content of a sentence can be measured according to a scale of plausibility (or acceptability), m-negations of this type express the fact that the minimal degree of this scale has been overstepped. Thus, it appears that scalarity is a property that m-negations keep till they become standard negation morphemes.

### 3. Q-negation

The other type of negative marker which can be either a non standard negative marker or a standard one, and which has received little attention in the literature, (apart from Zanuttini's work), is q-negation (from now on q-negation).

Q-negation has developed into the standard negative marker in dialects like Piedmontese and some Badia Rhaetoromance dialects:

- (20) a. A parla nen. (Turin)  
SCL speaks NEG  
'He/she does not speak.'
- b. Al ven nia. (Corvara<sup>5</sup>)  
he comes NEG  
'He does not come.'

In some other Rhaetoromance dialects, the negative quantifier is combined with a preverbal negative marker *n*, which alone cannot express negation:

- (21) Dytaurela n el \*(nia) gny. (S. Leonardo di Badia)  
yet NEG is- SCL NEG come  
'He has not come yet.'

Notice that the element in some dialects is still homophonous with the negative quantifier meaning 'nothing', as in Rhaetoromance, while in Piedmontese, though etymologically related, the two words for the negative marker and the negative quantifier are nowadays different: *nen* is the negative marker, while *gnente* is the word for 'nothing'.

Generally, this type of negative marker is either not compatible with negative quantifiers, as in Rhaetoromance:

- (22) a. \*I n a nia ody degugn.  
I NEG have not seen no one  
'I have not seen anybody.'
- b. I n a ody degugn.  
I NEG have seen no one  
'I have not seen anybody.'

Otherwise, there are strong restriction on the co-occurrence between the two, which cannot be adjacent:

---

<sup>5</sup> Notice that this structure is only found for younger speakers, older speakers still use the same type of discontinuous negative marker exemplified by the example of S. Leonardo.

- (23) a. \*A'm dis nen gnente. (Piedmontese, from Zanuttini 1997)  
 SCL me tells NEG nothing  
 'He does not tell me anything.'
- b. \*A veddu nen gnun.  
 I see NEG nobody  
 'I do not see anybody.'

Zanuttini (1997) already notes that this type of negative marker occupies a very low position in the clause, as shown by the following examples which exploit the position of low adverbs to show the point:

- (24) a. A l'avia già nen volu 'ntlura. (Piedmontese, Zanuttini 1997)  
 SCL it had already NEG wanted then  
 'He hadn't already wanted at that time.'
- b. A l'ha nen dine sempre tut.  
 SCL it has NEG said always everything  
 'He has not always said everything.'

The following structure is the one proposed by Zanuttini on the basis of Cinque's hierarchy of low adverbs.

- (25) [...[T Anterior *already* [Asp Terminative *anymore* [FP *neg* [Asp Perfective *always* [Asp Completive *tutto*]]]]]]]

An additional argument in favour of her hypothesis is the fact that in V2 Rhaetoromance dialects, the negative marker is topicalized to (some) SpecCP with the verb:

- (26) Nia desmentié ne podun-se dōta chë jënt che...  
 NEG forget NEG can-we all those people who...  
 'We cannot forget those people who...'

In these dialects, q-negation is treated as the negative morpheme which does not trigger any special pragmatics, and is compatible with any verb type or structure and possible with any sentence type in main as well as in embedded domains<sup>6</sup>.

Notice that an example like (26) shows that the presence (or its absence see (24)) of the preverbal negative marker is not due to the fact that 'nothing' is a polarity item. If it were, it could not be placed in a position higher than the negative marker. We believe that doubling phenomena between negative markers are different from negative concord between what looks like a negative quantifier (we will refer to this class with the theory-neutral term "n-words") and negation. One strong argument in favour of this split is that there is dissociation between the two properties: not all dialects where negative doubling is possible tolerate negative concord with n-words. Therefore, the occurrence or absence of a negative marker with the q-negation is not

<sup>6</sup> One exception is the imperative form in S. Leonardo Rhaetoromance, which is not compatible with *n...* *nia*, but requires a different negative markers. Those speakers who only use *nia* for sentential negation and not the discontinuous form, can have *nia* in imperative contexts. See Poletto and Zanuttini (2003) for a detailed description of this phenomenon.

to be assimilated to the standard negative concord phenomenon, and q-negation will be not analyzed as a polarity item as often n-words are.

An observation which shows that *niente* is “special” within the domain of n-words also in standard Italian are cases of constituent negation of elliptical constructions of the following type:

- (27) a. niente storie, eh?  
no stories PRT  
‘Don’t make a fuss.’  
b. niente patate, per piacere sono a dieta  
no potatoes please am at diet  
‘No potatoes please, I am on a diet.’

In these examples *niente* behaves in a parallel way to the English negative article *no*. (see Moscati (2006) on this). However, notice that the Italian construction is much more restricted, as it can only occur in the elliptical construction above, which has the flavour of a negative imperative where the verbal complex has been deleted. In any other context, the usage of *niente* as a negative article, or as constituent negation, is banned:

- (28) a. \*Ha niente vino.  
has no wine  
‘He has no wine.’  
b. \*Ho visto niente gatti, ma cani.  
Have seen no cats, but dogs  
‘I have not seen any cats, but dogs.’

One further interesting property about Italian *niente* is that it has non-negative usages in contexts like the following:

- (29) Waiter: Cercava qualcosa?  
‘Where you looking for something?’  
Customer: niente, cercavo il grana.  
‘Nothing, I was looking for parmesan cheese.’

Here the customer is not saying he was not looking for anything, he is “minimizing his action”, expressing the fact that he does not need any help, although he was indeed looking for something. This type of usage is probably scalar, which is, as we will see, a core property of those elements that can be reanalyzed as standard negative markers. In view of the diachronic path we are trying to establish, it is more interesting to examine the negative marker originating from a quantifier in those dialects where it is not (yet) the standard negative marker, but is a negative ‘reinforcer’ roughly meaning ‘at all’. The standard Italian expression translating ‘at all’ is *per niente*. However, the distribution of bare ‘nothing’ in Veneto dialects is more restricted, and we think revealing, with respect to the corresponding complex AdvP *per gnente*, which also means ‘at all’. Apparently standard Italian has not yet begun the process which might take ‘per niente’ to become a negative marker, while Veneto has started, but not completed the process. This is the reason why this can be more telling in view of the diachronic process examined in section 2.

### 3.1. The Distribution of *q*-negation standard negation

In what follows we describe the distribution of *q*-negation in Venetian, a dialect where bare ‘nothing’ is so to speak caught in the act of becoming a negative marker.<sup>7</sup> The element *gnente* ‘nothing’ seems *prima facie* incompatible with a direct object of transitive verbs, with unaccusative and passive subjects:

- (30) a. Nol lavora gnente.  
NEG-SCL works nothing  
b. Nol dorme gnente.  
NEG-SCL sleeps nothing  
c. \*Nol leze gnente i libri.  
NEG-SCL reads nothing the books  
d. \*Nol magna gnente la me torta.  
NEG-SCL eats nothing my cake  
e. \*Nol riva gnente.  
NEG-SCL arrives nothing  
f. \*Nol ze sta arestà gnente.  
NEG-SCL is been arrested nothing

*Q*-negation would thus be possible only with real intransitive verbs, which have no object as shown in (30). This set of data might at first sight lead the observer to the conclusion that, though *q*-negation is not an object but an adverbial element, it is still merged in object position (and then moved). This would be the reason why it is incompatible with anything else occupying the object position (either object of transitive verb, unaccusative subjects or passives). This is actually the hypothesis formulated by Bayer (2008) for *nichts* in German varieties or *nothing* in some colloquial varieties of English.

However, a closer look at the phenomenon reveals a more complex picture in the NIDs.

Other unaccusative verbs are indeed compatible with *q*-negation, and the same is true of subjects of psych-verbs which, according to Belletti-Rizzi (1988), should be parallel to unaccusative in having a subject generated in the object position:

- (31) a. No la crese gnente.  
NEG SCL grows nothing  
b. Nol me piaze gnente.  
NEG-SCL me likes nothing

The distinction internal to the class of unaccusative verbs is the one proposed by Tortora (1997): inherently directed motion verbs are incompatible with *q*-negation, while non inherently directed motion verbs are indeed compatible with it.

Moreover, *q*-negation is not per se incompatible with an element in object position, if the object is a bare plural, (defining some sort of activity, like ‘read books’) the combination is indeed possible. Consider the following contrast:

- (32) a. \*Nol leze gnente i libri, sto fio.  
NEG-SCL reads nothing the books this boy

<sup>7</sup> The phenomenon is rather widespread in the Veneto area, in some dialects the item can also occur without the preverbal negative marker, showing that *gnente* cannot be treated as a negative polarity item occurring in a negative concord structure.

- b. Nol leze gnente libri, sto fio.  
NEG-SCL reads nothing books this boy

One again might be lead to think that there is a distinction between definite and indefinite objects, but consider the following example:

- (33) a. %Nol salta gnente.  
NEG-SCL jumps nothing  
\* 'It does not jump'.  
OK (said of a long jump athlete) 'He does not jump much.'

A verb like *saltar* 'jump' is intransitive, hence it should be compatible with q-negation. However, it is not unless the interpretation is one of activity (for instance in the context of a professional long-jump athlete, whose job is to jump).

The following contrast is even more illuminating:

- (34) a. No-l zola gnente, sto aereo di carta.  
NEG-SCL flies nothing this plane of paper  
'This paper plane cannot fly at all.'  
b. \*Nol zola via gnente, sto aereo de carta.  
NEG-SCL flies away nothing this plane of paper  
'This paper plane cannot fly away at all.'

While *zolar* 'fly' is an activity, *zolar via* 'fly away' is telic, and q-negation is only compatible with the first verb, though in neither of the two cases is there an object.

Hence, we cannot conclude that the relevant property in banning q-negation is the presence of an object. Rather, it must be related to type of aspectual distinction (or better Aktionsart), which can be activated by the presence of a definite object, or be intrinsic to the type of verb or required by the presence of some verb modifiers.

That Aktionsart is involved is also shown by cases of activity verbs which can be turned into accomplishment simply by adding a preposition and forming a phrasal verb. Venetian, just like English has a couple of verbs like 'eat' and 'eat up' *magnar* and *magnar fora* (literally 'eat out'):

- (35) a. Nol magna gnente<sup>8</sup>.  
NEG-SCL eats nothing  
'He does not eat at all.'  
b. %Nol magna fora gnente.  
NEG-SCL eats out nothing  
'He does not eat up at all.'

The second sentence is impossible in the relevant reading, unless the sentence is interpreted as considering 'eat up' a habit, hence turning again the verb into an activity one.

Notice however that the distinction cannot simply be one of telicity, as q-negation is not automatically compatible with any atelic verb: *sercar* 'to look for' is for instance atelic, but it is still incompatible with the negative quantifier.

---

<sup>8</sup> This sentence is ambiguous between a reading in which *gnente* is the object of the verb *magnar* and a reading in which *gnente* means 'at all'. Obviously, we will consider only the second reading.

- (36) \*Nol serca gnente libri.  
 NEG-SCL looks-for nothing books  
 ‘He does not look for books at all.’

Hence, the fact that *gnente* is incompatible with telic verbs seems to be only a byproduct of a deeper property of the negative marker.

Moreover, although all cases of telic verbs, accomplishment and achievement verbs are not compatible with q-negation, notice that also stative verbs, like ‘live’, ‘be’, ‘stay’ are banned with *gnente*:

- (37) \*Nol vive gnente a Venessia.  
 NEG-SCL lives nothing in Venice  
 ‘He does not live in Venice at all.’

Therefore, telicity cannot be the core property here, although the property must have some connection to Aktionsart. For the moment, we propose the following empirical generalization:

- (38) q-negation is only compatible with activities.

In what follows we try to disentangle the problem.

### 3.2. *Q-negation as a monotone decreasing quantifier*

We would like to propose that the reason why q-negation is compatible only with activity verbs is that it is a scalar operator, which requires that each step of the scale be homogeneous.

The argument in favour of this is that other scalar elements like ‘poco’ ‘little’ have the same distribution:

- (39) a. Nol ze gnente simpatico.  
       Not-he is nothing nice  
       b. \*Nol ze gnente malà.  
       Not-he is nothing sick  
       c. El ze poco simpatico.  
       He is little nice  
       d. \*El ze poco malà.  
       He is little sick

Therefore, *gnente* is a scalar element quantifying over the scale being provided by the predicate itself. If *gnente* is parasitic on the predicate for a scale to quantify over, only predicates which can be in any way interpreted as scalar are compatible with *gnente*. In informal terms we state that in order to be compatible with the type of scale required by *gnente*, the predicate must first be apt to be split into a set of discrete smaller events, which can then be placed onto the scale. Activity verbs can be interpreted as a set of similar events occurring at different points in time and thus are indeed compatible with a scalar interpretation<sup>9</sup>:

<sup>9</sup> Notice furthermore that there is a distinction between *gnente* ‘nothing’ and *per gnente* ‘at all’, though they seem at first sight to be very similar, as *per gnente* is compatible also with non scalar predicates.

- (40) Nol lavora gnente.  
NEG-SCL works nothing  
'He does not work at all.'

The example above means that he does not work at moment A, B, C etc. Hence, the scale seems to be uniform in term of the type of event and can only be a temporal one. Punctual verbs or verbs indicating a single process which cannot be split into smaller events are not compatible with *gnente* because the quantifier has no scale to apply over.

Therefore, the link between q-negation and the direct object needs not be a direct link in the sense that q-negation and the object compete for the same position, but an indirect one, due to the fact that the presence of a referential object prevents the predicate from being interpreted as a scale of similar events: if the predicate is 'eat the apple' there cannot be many instances of eating the same apple. Therefore, the interpretation of the predicate as a uniform activity necessary to construe a scale on which the quantifier can apply.

Notice however, that if a temporal scale of the type seen above were the only possible one, cases like the following should be ungrammatical:

- (41) Nol ga dormio gnente.  
NEG-SCL has slept nothing  
'He did not sleep at all.'

Given that this is a past form,<sup>10</sup> it cannot be read as a sequence of different events of sleeping which are negated. In this case the scale is provided by the type of predicate itself, which can be put onto a scale, so that each event that is negated corresponds to a scale like the following: he did not sleep for eight hours, he did not sleep for seven hours,... he did not sleep for a single minute. Notice that this is a temporal scale too, as the one seen above, but a temporal scale which is established on a single instance of the event, because of the type of event, which is intrinsically prolonged.

We can conclude that *gnente* is sensitive to the type of scale used, and that not any scale is possible, but only a temporal one, which either quantifies over several instances of similar occurrences across time or on a single instance which can be split in time intervals.

Therefore, saying that *gnente* is scalar is not enough, the scale must uniform in terms of type of event it quantifies on.

- 
- (i) Non è malato per gnente. (standard Italian)  
Not is sick at all  
'He is not sick at all.'

The distinction between dialects where q-negation is the standard marker for negation and those where it is a special marker has to do exactly with the fact that in those dialects (and languages) where q-negation has a non standard value, it is still a quantifier, whose properties require a scalar reading. The scalar reading is probably the connection between the next step in the grammaticalization process which reanalyzes the quantifier as a negative marker. For a recent overview of Negative quantifiers and negative concord see Moscati (2006) chapter 5.

<sup>10</sup> in the Northern Italian dialects there is no difference between present perfect and simple past, the only form is the one above)

The hypothesis we would like to put forth here is that *gnente* is a monotone decreasing quantifier<sup>11</sup> similar to, *no*, *few* and *only*, and that its apparently bizarre properties are derived from the fact that it belongs to this class of quantifiers.

Monotone decreasing quantifiers (also called downward entailing) have the following property: if the first sentence is true, the second also is:

- (42) a. No boy runs.  
b. No boy runs fast.

The same property can be shown to be valid for *gnente*:

- (43) a. Nol core gnente.  
not runs nothing  
'He does not run at all.'  
b. Nol core gnente forte.  
Not runs nothing fast  
'He does not run fast at all.'

The basic mathematical property of a monotonic function, is that it has to go always in the same direction. Being *gnente* an adverb, its monotonicity has to apply to the predicate. Given that in mathematics, the function expressed by the monotonic quantifier has to be defined on a subset of the real numbers with real values, this implies that it must be possible to create a subset of instances of the predicate which are discrete, in other words, it must be possible to create a subset of similar events, for each of which the function is applied, hence, each of which is then negated. The fact that the instances of the event negated are of the same type thus derives from the monotonicity of the quantifier. Suppose furthermore that *gnente* is a strictly decreasing monotonic quantifier, this means that the set of events on which the quantifier applies will have to order on a (decreasing) scale and each of them is associated with the function of the quantifier, hence negated. This property derives the fact that *gnente* needs a scale, as each instance of the event has to be put on a decreasing scale in order to be associated to the function of the quantifier (which is negation).

Hence, the property of requiring a set of discrete events comes from the monotonic property and the fact that they have to be ordered onto a strictly decreasing scale comes from it being a strictly decreasing quantifier. We can conclude that in the dialect examined the n-word<sup>12</sup> has become an adverbial element with the properties of a strict monotone decreasing quantifier.

### 3.3. Some diachronic observation

Once we have seen that the n-word corresponding to 'nothing' can be reanalyzed as an adverbial with the properties of a monotone decreasing quantifier scoping over the

<sup>11</sup> Monotone decreasing GQs are among the expressions that can license a negative polarity item, such as *any*.

1. Good: No boy has **any** money.
2. Bad: \*Every boy has **any** money.

Probably *gnente* is itself formed by a quantifier similar to the determiner *no* and an NPI in its internal structure. We will not pursue this here any further.

<sup>12</sup> We use the term n-words here to refer to all those elements which in the Romance languages can either have the properties of negative quantifiers or of negative polarity items.

event in Venetian, we can go back to the original problem, namely the one of the diachronic evolution of different elements which already appear to have undergone Jespersen's cycle and have become standard negative markers. Veneto dialects seem to constitute the first step of the grammaticalization process of the n-word into a negative marker passing through a stage in which the element, originally used only in argumental position, can be reanalyzed as an adverb. The first question is then why among all n-words only the one corresponding to 'nothing' has undergone this process. In the Northern Italian dialects there is no other n-word which behaves so: we do not find cases of 'nobody' or other n-words which can be reanalyzed as the sentential negative marker<sup>13</sup>. On the one hand, the fact that 'nothing' is selected among the negative quantifiers to become the negative marker is part of a more general process which has to do with grammaticalization as loss of lexical features by the element becoming functional. On the other, this n-word must have some syntactic/semantic special property which singles it out in the domain of all n-words. Notice that the same type of process can be seen with wh-items, where the element becoming a wh-clitic (as the interrogative wh-item *que* in French, see Poletto and Pollock, 2004) or a complementizer is always the semantically (and syntactically) barest operator; in the sense that it has the smallest set of features because its lexical restrictor is virtually non-existent (see Obenauer, 1994)<sup>14</sup>. So, while a wh-item like 'who' or 'where' contain a lexical restrictor which is respectively [+human] and [+place] the element corresponding to 'what' has no lexical restrictor, and thus it is the barest and more functional element, in a sense the 'purest' operator and most functional one due to lack of semantic features and consequently to a different internal structure which lacks a lexical restrictor. Hence, while elements like 'noone' etc. have a Q projection inside, followed by an existential one and a restrictor, *gnente* also lacks the restrictor on a par with 'what'<sup>15</sup>.

But, if *gnente* has no lexical restrictor to quantify over in its internal structure, this means that it is the most prone element to become an adverbial form parasitic on the external structure (namely IP) to find a restrictor. We have seen that the first step of the grammaticalization is precisely the one which transforms the n-word in argumental position into an adverbial quantifier which takes the predicate in its scope. This is expected if, as already shown by Zanuttini (1997), the position of *gnente* is inside the aspectual field. Zanuttini uses Cinque's theory of lower adverbs to place q-negation in the IP domain and shows, as already mentioned, that it is located in the following position:

- (44) [...[T Anterior *already* [Asp Terminative *anymore* [FP *neg* [Asp Perfective *always* [Asp Completive *tutto*]]]]]]]

<sup>13</sup> There seems to be cases of sentential negation formed with no+ever in the Germanic languages. It is however a fact that in Romance the corresponding element *mai* 'never' has not become a negative marker in any of the dialects present in the ASIS database. The difference between the Germanic and its Romance counterpart, if a true one, still remains obscure.

<sup>14</sup> Notice that however, the element corresponding to the inanimate object is not the only one which can become a clitic, or a complementizer (see for instance cases of clitic whs corresponding to 'where' or complementizers like Bavarian *wo* 'where'), it is only the most frequent one.

<sup>15</sup> A similar type of process is also reported for the formation of object clitics in Benincà and Poletto (2005): direct object clitics are always the most frequent clitics even in languages which do not have any other type of (dative, nominative, partitive or locative) clitic.

Can we determine what the position of q-negation is in Veneto, which represents the intermediate diachronic stage? The element is compatible with *più*, and it occurs to its right, as shown by the following example:

- (45) Nol ga più dormio gnente, da chela volta.  
 NEG-SCL has anymore slept nothing since that time  
 ‘Since then, he did not sleep at all anymore.’

However, q-negation is not compatible either with the adverb *sempre* ‘always’ or with the quantifier *tutto* ‘everything’, so the test is not complete and we cannot really determine what the exact position of q-negation is.

However, there are indirect clues of the position of q-negation: Old Italian also had cases of q-negation with properties that are strikingly similar to the one discussed here for Venetian. Differently from Venetian, Old Italian had overt quantifier raising across the past participle for both the quantifiers *tutto* and *niente* (see Poletto, 2008, for a detailed analysis of the position of *niente* and *tutto* in Old Italian) with *niente* occurring higher than *tutto*.

Similar clues come from an analysis of negative quantifiers put forth by Vecchiato (1999) which gives the following hierarchy in modern French:

- (46) guère > trop > rien > complètement > tout > beaucoup/peu > bien >...

This shows that the standard negative marker derived from q-negation and the adverbial quantifier are in the same position, a position dedicated to (bare) quantifiers in the low IP area, which, as shown in Cinque (1999) is the same position occupied by French *rien* ‘nothing’ even when it is an argument:

- (47) a. Il a rien vu.  
 he has nothing seen  
 b. \*Il a vu rien.  
 he has seen nothing  
 ‘He has seen nothing.’  
 c. Il a vu personne.  
 he has seen nobody  
 d. \*Il a personne vu.  
 he has nobody seen

Does *gnente* in Venetian also occupy the quantifier position when it is an argument as it does in French? Unfortunately, the test of French is not available, as the Venetian past participle raises too high to give rise to the order *gnente*-past participle. If we apply Cinque’s tests on adverbials the results are ambiguous: if *gnente* were in the quantifier IP space, it should precede the lowest adverb, namely ‘well’, in Venetian *ben*:

- (48) a. Nol ga visto gnente ben.  
 Not.he has seen nothing well  
 b. Nol ga visto ben gnente.  
 Not.he has seen well nothing

Both sentences are correct, with the difference that in the second one *gnente* seems to have focus. Focus is however a disturbing factor, as the focalized XP might have moved to a low left peripheral Focus position internal to the low vP phase (see Poletto, 2007, on this). If we contrast object *gnente* with other objects we see that *gnente* needs to precede other arguments, while this is not the case for DPs:

- (49) a. Nol ghe ga dito gnente a Nane.  
NEG-SCL him has said nothing to N.  
'He did not say anything to N.'
- b. \*Nol ghe ga dito a Nane gnente.  
NEG-SCL him has said to N nothing  
'He did not say anything to N.'
- c. El ghe ga dito a Nane sta storia.  
SCL him has said to N. this story  
'He told this story to N.'
- d. El ghe ga dito sta storia a Nane.  
SCL him has told this story to N.

This contrast can only be explained by assuming that *gnente* obligatorily moves to the quantifier position, as it does in French.

Hence, given that *gnente* also moves in Venetian as it does in French, we can conclude:

- a) that the argumental, the non standard and the standard q-negative marker all occupy the same position dedicated to bare quantifiers in the low IP area.
- b) that reanalysis of the bare quantifier 'nothing' into a monotone decreasing quantifier and then into the standard negative marker is not ensued by moving the reanalyzed element further up in the structure but simply by turning a movement process (the object quantifier raises to its dedicated position) into a merge one (the adverbial quantifier is directly merged in the position where it surfaces).

#### **4. The negative field**

Having analyzed two distinct types of elements which both end up as a standard negative marker, we now have to reconsider their differences and similarities in order to arrive at a conclusion concerning what the core property is which enables to reanalyze a given item as a negative marker.

Starting from the differences, we have observed that minimizers are real DPs which take a complement and can have adjectives and an article. The first step of the reanalysis is assumed to be an internal one, in which the minimizer is not the head noun anymore, but a classifier belonging to the functional domain of the DP whose lexical N is the original complement of the minimizer. Being minimizers reanalyzed as classifiers of measure, they imply the presence of a scale, which enables them to be once more reanalyzed as FPs which are part of the IP. Once they have been reanalyzed as classifiers, they are treated as measure classifiers because of their meaning, which implies the presence of a scale. This in turn enables them to be further reanalyzed as functional XPs located in the domain of IP, with the scale being provided by the predicate itself.

Q-negation is different, because it cannot be reinterpreted as a classifier of its original complement, being the whole complement itself. However, being a quantifier it raises to a dedicated position for quantifiers in the domain of lower adverbs (see Poletto, 2007, for an analysis of bare quantifiers in these terms). As *gnente* is the barest

quantifier in terms of lexical restriction, it can be reanalyzed as an adverbial quantifier whose scope does not simply range over the object, but over the entire event (which must then have the right properties in order to provide the uniform type of scale required by the quantifier, as seen above). Given that the object quantifier already moves to a position inside the Aspectual field, reanalysis simply cancels the trace in object position and q-negation is merged in the FP where QPs move.

Therefore, the diachronic path followed by the two elements seems to be rather different but then the original problem remains: is there any common property which enables those items to be reanalyzed as negative marker and if so, what is it?

We would like to propose that the common property the two types of items have in common is the one already discussed for both cases, namely scalarity.

Minimizers which become measure classifiers and the n-word becoming a uniform monotone decreasing quantifier both have scalar properties. The scale is at the beginning of the reanalysis process simply “accessory”, it is implied by the process: a noun becoming a classifier acquires scalar properties and the same does an n-word becoming a monotone decreasing quantifier scoping on the event. If this is the core property which turns the two elements into negative items, why is scalarity so important? At the moment we have no precise answer to this question, but we would like to point out that another type of sentential negation in Romance, namely the preverbal negative marker *ne* is analyzed as a scalar element by Roorick (2008). He clearly shows that *ne* is a minimizer and that it is not polarity negation, but scalar one, which takes the minimal amount of a scale. One could further speculate whether this is so in other Romance languages as well. Although Roorick explicitly denies that the corresponding negative marker in Catalan is scalar (although it has scalar uses, as he observes), he leaves the possibility open that in Romance languages having only a preverbal negative marker, there is a null element corresponding to *pas*, which is (according to Roorick) the real negation. Italian is also a candidate to this, as the scalar uses of French *ne* can be replicated for Italian *non*.

At this point we do not know whether Roorick is correct in assuming that the position where French *pas* occurs is the real negative one. Italian counterpart of *pas*, *mica*, is not a negative marker *per se*, but a minimizer as the ones discussed in this work.

If preverbal negation is a minimizer and it combines with a minimizer, then there is no polar negation around whatsoever, and this opens up the possibility that at least in some Romance language, negation is not the linguistic counterpart of logic negation, it is not an operator denying the truth value of the entire proposition, but it is a sort of inference produced by opening a scale and taking the minimal possible value. However, this topic would lead us too far from the scope of this article, which is to show what kind of property different types of items becoming negative markers have in common. Therefore, we conclude the work leaving this perspective for future research.

A further problem which opens us and which we leave to future research (see Poletto, 2008, on this) is due to the position of these negative markers: it seems to be a general fact that non-standard and standard negation occupy the same position, so the position is not determined by its status (standard/non-standard). If it were so, we would expect to find a change in the position of items which are realized as standard negative markers. On the contrary, the position of negative markers always coincides with their etymological type. This, as originally pointed out by Zanuttini (1997), shows that sentence negation is not related to a single position in the sentence. However, it forcedly leads us to the following question: how can elements located in different points of the structural tree still license the same reading? The answer we put forth

here is again the one mentioned before: negation is not a polarity operator which denies the truth value of the CP, negation is a complex phenomenon and involves the activation of several projections in the clause structure, none of which in the end can probably be labelled as NegP. Each of them is checked by a different type of element, a scalar, a minimizer, a quantifier (or even a focus one, see again Poletto, 2008). The presence of only one of these elements can “re-construe” all the others simply by signalling that they are present. In other words, what we call NegP is a complex set of projections, the lexicalization of just one element is enough to activate the whole NegP. This in turn implies that, although according to the analysis put forth here there are at least three projections in the clause (the preverbal negative marker, m-negation and q-negation) which have to be checked by elements located in the four corresponding projections. Only one of these projections needs to be lexicalized, the others can be phonologically silent, but still be there.

## **5. Conclusion**

In this work we have taken into account minimizers which become (non standard) negative markers and have traced their diachronic path from a lexical noun, to a classifier internal to the DP, to a minimizer quantifying over the entire event. Although m-negation is not per se a scalar element anymore in modern Italian, its etymology and older stages of Italian clearly show that it must have been during the grammaticalization process. Therefore, we have taken into account a second type of element, q-negation, which is still within the process of grammaticalization in some Italian dialects and we have shown that it is indeed a scalar element. We can conclude that in order to become a negative marker, a lexical item has to undergo a process in which it becomes scalar (and therefore, only element which can be reinterpreted as scalar ones, can become negative markers). Another common property of the two types of elements is that they both start out as a negative object, whose scope is then extended to the entire predicate. In the case of m-negation, the element has moved from its original object position to a position internal to the functional space of the past participle. In the case of q-negation, the element, being a negative quantifier, is already located in the functional space of the past participle and remains where it is. We are aware of the fact that our line of research opens more problems than it solves, but we hope to have at least shown that scalarity is a crucial property in the process of development of negative markers.

## **References**

- Bayer, J. (2008): Nominal negative quantifiers as adjuncts. Ms. University of Konstanz.
- Belletti, A. & L. Rizzi (1988) Belletti, A. and L. Rizzi, (1988). Psych-verbs and Theta Theory. *Natural Language and Linguistic Theory*, 6:291-352.
- Benincà, P. & C. Poletto (2005) "On some descriptive generalizations in Romance" in Handbook of Comparative Syntax, R. Kayne e G. Cinque (eds.), Oxford University Press, New York & Oxford. P221-258.
- Cinque, G. (1976) Mica. *Annali della Facoltà di Lettere e Filosofia dell'Università di Padova 1*, 101-112.
- Cinque, G. (1999) *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. New York & Oxford: Oxford University Press.

- Colombini, F. (2007) *La negazione nei dialetti emiliani: microvariazione nell'area modenese*. University of Padua: MA Thesis.
- Gelderen, E. van (2004) *Grammaticalization as Economy*. Amsterdam: John Benjamins.
- Kiparsky, P. & C. Condoravdi (2006) Tracking Jespersen's Cycle. In M. Janse, B. D. Joseph & A. Ralli (Eds.) *Proceedings of the 2nd International Conference of Modern Greek Dialects and Linguistic Theory*, Mytilene: Doukas.
- Manzini, M.R. & L.M. Savoia (2005) *I dialetti italiani e romanci: morfosintassi generativa*. Alessandria: Edizioni dell'Orso.
- Moscato, V. (2006) *The Scope of Negation*. PhD: Thesis, University of Siena
- Obenauer, H.G. (1994) *Aspects de la syntaxe A-barre - Effets d'intervention et mouvements des quantifieurs*, unpublished thèse d'Etat, Université de Paris VIII.
- Penello, N. & D. Pescarini (2008) Osservazioni su mica in italiano e alcuni dialetti veneti. *Quaderni di lavoro dell'ASIt* 8, 43-56.
- Poletto, C. (2007) "Doubling as economy" on line publication at the following website: <http://www.meertens.knaw.nl/projecten/edisyn>
- Poletto, C. (2008) "On negative doubling" ms. University of Venice
- Poletto C. & Pollock J-Y- (2004) "On Wh-clitics, Wh-doubling in French and some North Eastern Italian Dialects", *Probus*, n° 16.2, 241-272.
- Poletto, C. & R. Zanuttini (2003) Marking Imperatives: Evidence from Central Rhaetoromance. In C. Tortora (Ed.) *The Syntax of Italian Dialects*, (pp. 175-206). New York & Oxford: Oxford University Press.
- Postal, P.M. (2003) The Structure of One Type of American English Vulgar Minimizer. In *Skeptical Linguistic Essays*, (§ 5). New York & Oxford: Oxford University Press.
- Roberts, I. & A. Roussou (2003) *Syntactic Change: A minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.
- Rohlf, G. (1969) *Grammatica storica della lingua italiana e dei suoi dialetti III: Sintassi*. Torino: Einaudi.
- Roorick, J. (2008) On the scalar nature of syntactic negation in French. Submitted to NLLT.
- Vai, M. (1996) Per una storia della negazione in Milanese in comparazione con altre varietà altoitaliane. *ACME* 40, 1. Milano.
- Vecchiato (1999) *Sur l'ordre de certains adverbes de quantification, de temps et d'itération en français Tesi di laurea*. University of Venice.
- Tortora, Ch. (1997) "The Syntax and Semantics of the Weak Locative". PhD Thesis, University of Delaware.
- Zanuttini, R. (1997) *Negation and Clausal Structure*. New York & Oxford: Oxford University Press.

# Phases, Strong Islands, and Computational Nesting\*

VALENTINA BIANCHI,

University of Siena  
bianchi10@unisi.it

Cristiano Chesi

University of Siena  
chesi@media.unisi.it

This paper is an attempt to recast the connectedness condition (Kayne 1983) in derivational terms: we will show that a Top-Down derivation is adequate to describe strong island conditions (as in Huang's original proposal), without losing the ability to discriminate among distinct phenomena (preverbal subject islandhood, complex-NP constraint, special properties of the first argument in double object constructions, intermediate status w.r.t. extraction/gapping, for right adjuncts), and predicting, moreover, the grammatical distribution of parasitic/licensed gaps in the derived structure.

## 1. Left branch islands and the connectedness effect

In this paper we reconsider the connectedness effect discussed by Kayne (1983), and illustrated in the examples (1)-(3). Kayne observed that in VO languages, left branch constituents are strong islands for extraction;<sup>1</sup> however, an illegitimate gap inside a left branch island can be rescued by another gap embedded in a lower right branch constituent. The examples in (1)-(2) illustrate preverbal subject islands, and (3) a small clause subject island: while in the (a) examples, extraction from the left-branch subject is impossible, in the (b) examples, the illegitimate gap is followed by a legitimate gap on a right branch and this creates a grammatical configuration.

---

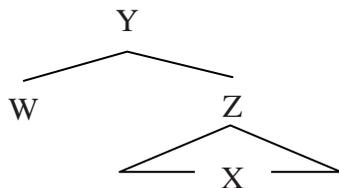
\* This paper is an expanded and revised version of the talk presented at the 28<sup>th</sup> Glow Colloquium (Geneva, March-April 2005). We wish to thank Adriana Belletti, Cedric Boeckx, Alec Marantz, Andrew Nevins, Luigi Rizzi, Uri Shlonsky and Michal Starke for discussion of the material presented in this paper, as well as the audiences at the 29<sup>th</sup> Penn Linguistic Colloquium (February 2005), the Harvard Grammatical Locality reading group and at the 28<sup>th</sup> Glow Colloquium (Geneva, March-April 2005).

<sup>1</sup> Throughout the paper, by “strong islands” we mean *nonselective* islands, which do not give rise to argument/adjunct asymmetries in extraction, as opposed to weak (Relativized Minimality) islands, which selectively affect the extraction of certain constituents: see Rizzi (1990, 1994, 2002). See Starke (2001) for an interesting attempt at unification of the islands phenomena which however does not account for connectedness effects.

- (1) a. \*[Which famous playwright]<sub>i</sub> did [close friends of  $e_i$ ] become famous?  
 b. ?[Which famous playwright]<sub>i</sub> did [close friends of  $e_i$ ] admire  $e_i$  ?
- (2) a. \*Who<sub>i</sub> did [my talking to  $e_i$ ] bother Hilary?  
 b.  $\checkmark$ Who<sub>i</sub> did [my talking to  $e_i$ ] bother  $e_i$ ?
- (3) a. \*Who<sub>i</sub> did you consider [friends of  $e_i$ ] angry at Sandy?  
 b.  $\checkmark$ Who<sub>i</sub> did you consider [friends of  $e_i$ ] angry at  $e_i$  ?

Kayne (1983) proposed an essentially representational constraint to account for these data, the Connectedness Condition. The central notion is that of a g-projection, which is defined in (4)-(5). In a VO language like English, every right branch is in a canonical government configuration, by definition (4); the recursive definition in (5) ensures that all the maximal projections dominating a structural governor X and lying on a right branch are g-projections of X.

- (4) W and Z (Z a maximal projection, and W and Z immediately dominated by some Y) are in a canonical government configuration iff
- V governs NP to its right in the grammar of the language and W precedes Z
  - V governs NP to its left in the grammar of the language and Z precedes W
- (5) Y is a g-projection of X iff
- Y is an (X') projection of X or of a g-projection of X, or
  - X is a structural governor and Y immediately dominates W and Z, where Z is a maximal projection of a g-projection of X, and W and Z are in a canonical government configuration



Thus, the g-projections of X can extend upward as long as any dominating maximal projection is on a right branch.

The Connectedness Condition (henceforth CC) requires that the set of the g-projections of (the governor(s) of) the empty category(ies) bound by a given binder and the binder itself form a connected subtree:

- (6) The g-projection set  $G_\beta$  of a category  $\beta$  is defined as follows ( $\gamma$  governs  $\beta$ ):
- $\forall \pi, \pi = \text{a g-projection of } \gamma \rightarrow \pi \in G_\beta$
  - $\beta \in G_\beta$  and
  - $\delta$  dominates  $\beta$  and  $\delta$  does not dominate  $\gamma \rightarrow \delta \in G_\beta$ <sup>2</sup>
- (7) Connectedness Condition<sup>3</sup>  
 Let  $\beta_1 \dots \beta_j, \beta_{j+1} \dots \beta_n$  be a maximal set of empty categories in a tree T such that  $\exists \alpha \forall j, \beta_j$  is locally bound by  $\alpha$ . Then  $\{\alpha\} \cup (\bigcup_{1=j=n} G_{\beta_j})$  must constitute a subtree of T.

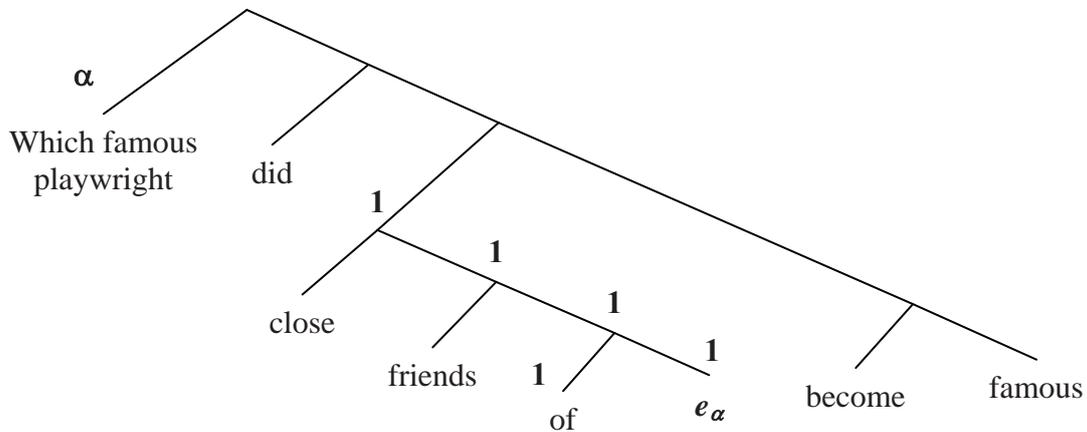
In case of a single gap, the CC requires that all the maximal projections in the path between the gap and its binder be on a right branch. Consider for instance the

<sup>2</sup> Clause b' takes care of government across a clausal boundary  $\delta$ .

<sup>3</sup> For our present purposes, we ignore the modification of the CC proposed by Kayne to account also for multiple wh structures and for negative elements licensed by clausal negation in Italian.

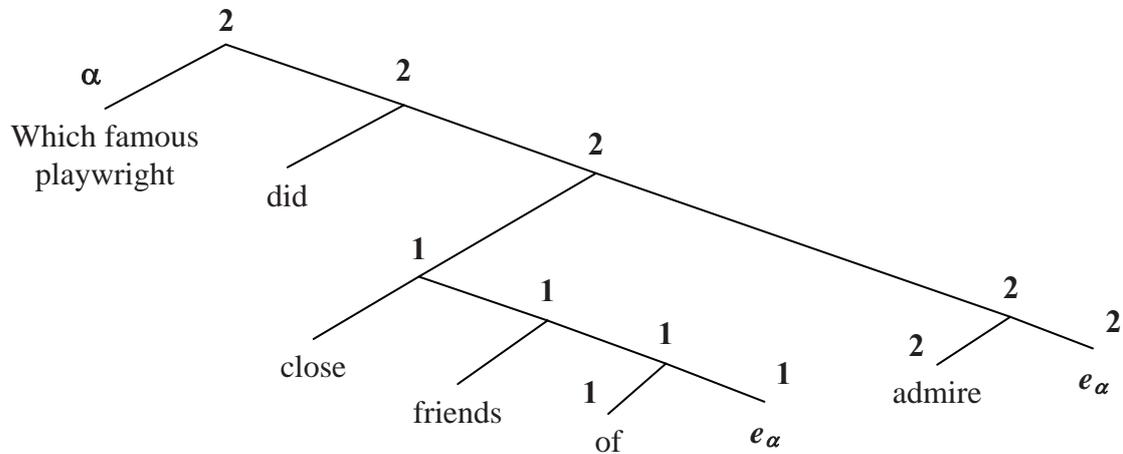
ungrammatical example in (1a): as the tree graph below makes clear, the g-projections of the gap stop at the level of the preverbal subject, which is a left branch and hence not in a canonical government configuration. Therefore, the g-projections cannot extend upward to reach the binder, and the CC is violated:<sup>4</sup>

(1.a) \*



The rescuing effect in (1b) is due to the fact that the g-projections of the lower gap in the object position extend upward and connect to the g-projections of the illegitimate gap embedded in the subject, as shown in the tree below. As a result, the two g-projection sets form a connected subtree including the binder, and the CC is satisfied:

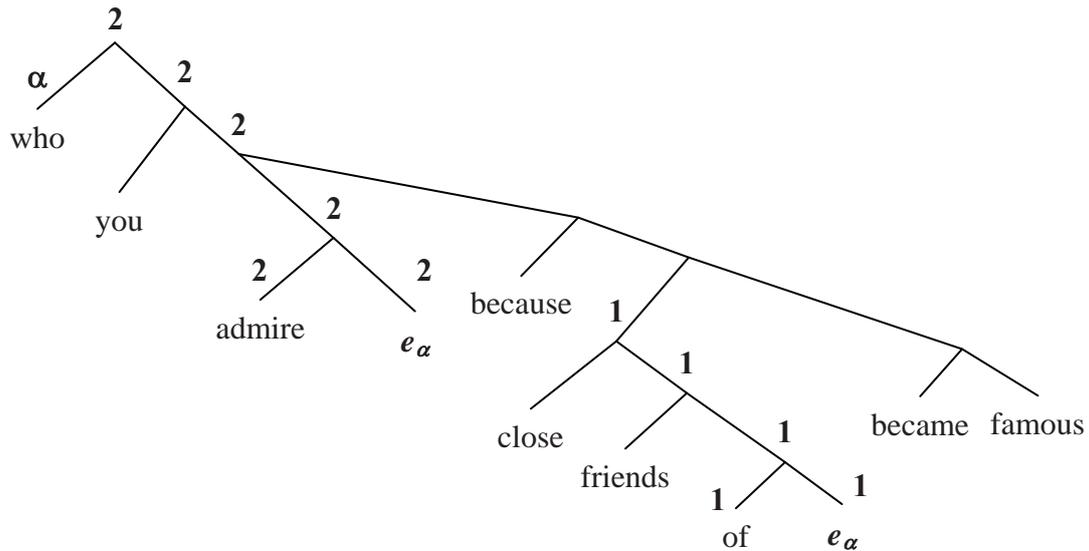
(1.b)



On the contrary, no rescuing effect arises if the legitimate gap is too high in the tree for its g-projection set to connect to that of the illegitimate gap, as in the following example:

(8) \* a person who you admire e because [close friends of e] became famous

<sup>4</sup> As in Kayne (1983), the numerical indices are introduced for expository purposes to mark the g-projection paths of the empty categories, and have no theoretical significance.



The CC differs in various respects from other approaches to parasitic gaps in the GB framework. Firstly, even though there is a clear sense in which in the (b) examples of paradigms (1)-(2) the gap inside the left branch is "illegitimate" or parasitic, and the other one is legitimate, there is no other assumed difference between them, either with respect to the nature of the empty category or of its relation to the binder.

But the status of the parasitic gap and of its relation to the binder is actually debated, as can be seen in the collection of papers edited by Culicover & Postal (2001). Cinque (1990) and Postal (1994) have pointed out various types of evidence which suggest that the parasitic gap is a null resumptive pronoun rather than an ordinary extraction gap. The evidence comes from the lack of reconstruction effects in the parasitic gap position, the impossibility for parasitic gaps to occur in Postal's antipronominal contexts, and the restriction of parasitic gaps to the NP category. However, all these types of evidence have been called into question by other authors (cf. e.g. Hukari et al. 2001, Levine & Sag 2003); it seems fair to say that the issue is still open.

Secondly, note that the CC is designed to capture left branch islands only. Other strong island types, like e.g. right-hand adjuncts and relative clauses, are not subsumed under this condition. It is empirically debated whether strong islands are a uniform class falling under a single principle (as proposed for instance in Cinque 1990). As we will discuss in §4, even Longobardi's (1985) extension of the CC cannot subsume all the island effects that are usually classified as strong (unselective) islands. Despite these problems, we believe that the CC incorporates an important insight, which we will formulate as follows:

- (9) Generalization on legitimate recursion and gap licensing  
 Legitimate gaps lie on the main recursive branch of the tree, whereas illegitimate gaps lie on "secondary" branches, which do not allow for unlimited recursion (in that such a secondary branch cannot be the lowest one in a tree).

It is this insight that we will try to capture in our approach, though in an essentially derivational perspective.

As a first pass, we will propose a derivational hypothesis that has the same empirical scope as Kayne's original CC, and only accounts for left branch islands (§3). In §4 we will come back to the problem of right-hand adjuncts.

As to the question of the (a)symmetry between legitimate and parasitic gap-antecedent dependency to a standard antecedent-gap dependency, and treat both in terms of copy-remerging. However, we believe that the constraints on the structure of the computation that we are going to highlight are also consistent with an analysis in terms of a null resumptive pronoun.

Our proposal will be implemented in the computational model of a top-to-bottom oriented Minimalist Grammar proposed in Chesi (2004, 2007). Although limitations of space prevent us from fully justifying the proposed model, we will now give a brief sketch, which will constitute the background of our proposal.

## 2. The computational model

### 2.1. The general architecture

Chesi (2004, 2007) proposes a formalization of a minimalist grammar (adapting the formalism discussed in Stabler 1997) with two main components:

- a. a lexicon consisting of feature structures composed of semantic, syntactic and phonetic features<sup>5</sup>;
- b. three structure building operations (*Merge*, *Move* and *Phase Projection*).

Chesi argues that for reasons of computational efficiency and cognitive plausibility<sup>6</sup>, the grammar should have the property of *flexibility*: namely, it should be directly usable both in a parsing and in a generation context. The flexibility requirement leads Chesi to abandon the bottom-to-top orientation of the standard minimalist derivation, and to assume instead a top-to-bottom orientation (as in Phillips 1996).

Assume a Structural Description (SD) to be definable simply in terms of immediate relations (immediate dominance<sup>7</sup> and immediate precedence); assume, moreover, that any item is licensed within a SD (leading then to grammaticality) if and only if it is *selected*<sup>8</sup> or it is a possible *functional specification* of a lexical head. Accordingly, a lexical head is specified for two types of features: the SELECT features specify its argumental valency, and license the head's arguments (they correspond to the standard theta-grid or argument structure); the LICENSOR features instead specify the possible functional specifications that can be associated with the head: these correspond to the standard functional heads (FPs) in the lexical head's extended projection (Grimshaw 1991). Importantly, the LICENSOR features associated to a given lexical head are limited in number and are hierarchically ordered, much as in the cartographic approach

---

<sup>5</sup> Nothing in this paper hinges on the feature theory we will use. For sake of simplicity, let us assume a simple (non recursive) privative system (Adger 2003, 2007). In the original version (Chesi 2004) complete attribute-value matrices were allowed, as in most unification grammars (e.g. HPSG, Pollard and Sag 1994).

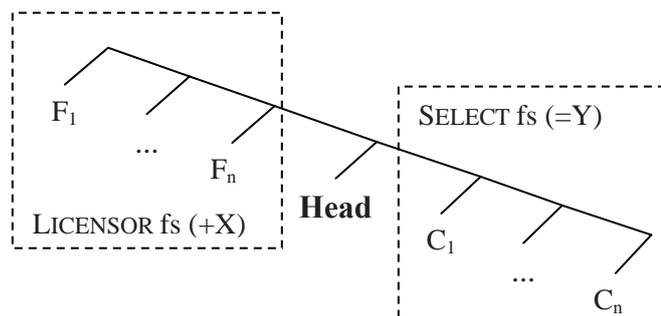
<sup>6</sup> It is hardly plausible that we would speak a language using a particular grammatical competence and that we would produce the very same language using a different knowledge.

<sup>7</sup> The statement "A *immediately dominates* B" would correspond to the result of a merge operation where A projects over B: [<sub>A</sub> A B].

<sup>8</sup> Here *selection* means both *C(ategorial)-selection* and *S(emantic)-selection* (Pesetsky 1982).

proposed by Cinque (1999), Rizzi (1997, 2004). The general schema is then the following:

(10)



Chesi (2004) then defines a general top-to-bottom algorithm which can be exploited both in generation and parsing; specifically, in generation, the algorithm converts a set of immediate dominance relations among semantic/formal feature structures into a set of immediate precedence relations among lexicalized phonological feature structures; vice versa, in parsing, it converts a set of immediate precedence relations among phonological structures into a set of immediate dominance relations among lexicalized semantic/formal structures).

From this perspective the structure building operations can be redefined as follows<sup>9</sup>:

- (11) **Merge** is a binary function (sensitive to temporal order) which takes two feature structures and unifies them (in the sense of unification grammars, Shieber 1983)
- (12) **Phase Projection** is the minimal set of dominance relations introduced in the SD based on the expectations triggered by the SELECT features of the currently processed lexical head.
- (13) **Move** is a top-down oriented function which stores an unselected element in a memory buffer<sup>10</sup> and re-merges it at the point of the computation where the element is selected by a lexical head.

An *unselected* element is any element that is processed before the lexical head is found, and hence temporally and linearly precedes the head itself, according to the following Linearization Principle (inspired by Kayne's (1994) LCA):

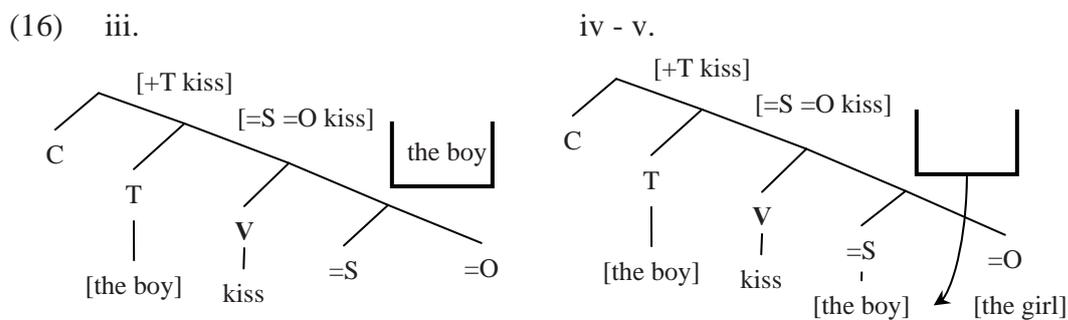
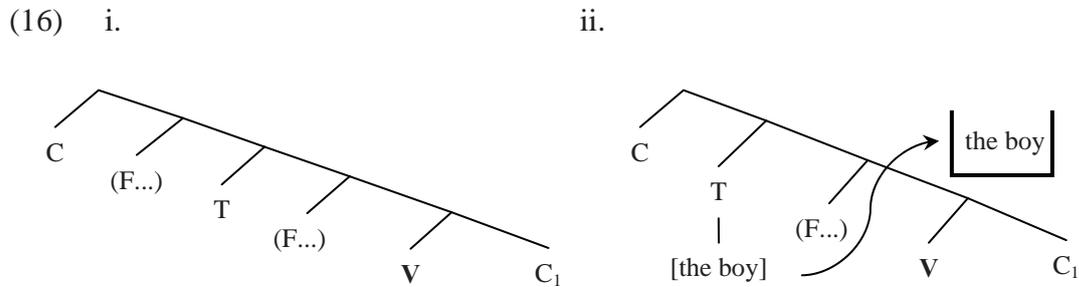
- (14) **Linearization Principle**
- a.  $\langle A, B \rangle$  if A (is a lexical head and) selects B as an argument
  - b.  $\langle B, A \rangle$  if B is a functional specification of A.

Though limitations of space prevent us from fully describing the proposed model, we will illustrate how the structure building operations work in a simple example, where all the basic ingredients are involved:

<sup>9</sup> See Chesi (2004), Ch. 3.3.2 for an explicit and thorough formalization.

<sup>10</sup> Limitations of space do not allow us to fully characterize the memory buffer (the reader is referred to Chesi 2004). Let us simply emphasize two points. First, the memory buffer must be *multidimensional*, i.e. different kinds of elements are stored in separate lists; this will account for the selectivity of intervention (Relativized Minimality) effects, cf. Rizzi (1997, 2001). Second, the minimality effect itself can be captured by assuming a Last In First Out memory, so that at a given point of the computation only the last element that was inserted in the buffer can be retrieved, and the previously inserted ones cannot.

- (15) The boy kissed the girl.
- i. As the initial step, the system projects a top-down expectation of a verbal phase (i.e. a CP),<sup>11</sup> whose lexical head will have to be a verb.
  - ii. The constituent [*the boy*] is processed<sup>12</sup> and, being compatible with the functional Tense-related specification<sup>13</sup>, it is inserted at the corresponding functional level. Since the element is not selected in this position, it is also stored in the memory buffer.



- iii. The lexical item *kissed* (analysed as *kiss* +T) is processed; this introduces in the derivation the verb's SELECT features, here abbreviated as =S (external argument) and =O (internal argument) which are projected, according to Phase Projection, starting from the most external one.<sup>14</sup>
- iv. The constituent [*the boy*] previously stored in the memory buffer is re-merged as a sister to the verb to satisfy the verb's =S feature.
- v. As a final step, the computation proceeds by processing the direct object.

We return immediately to the special status of the lowest selected complement, which follows from a novel definition of phase.

## 2.2. Phases

Chesi (2004) argues that in order to gain computational tractability, the derivation must be broken up into *phases*, i.e. subparts of the computational process with a fixed upper bound in complexity. The phase can be roughly defined as follows:

<sup>11</sup> This root application of Phase Projection is obviously not triggered by any SELECT feature.

<sup>12</sup> This actually constitutes a separate and “nested” computational phase, as will become clear in §2.2.

<sup>13</sup> Let us avoid complications with respect to the exact position of the subject, which is irrelevant for the present discussion.

<sup>14</sup> This is because we want to preserve scope relations and, as Phillip’s (1996) Merge Right, from a derivational perspective, we expect these intermediate constituents to be built in the following order: [<sub>v</sub> S V] → [<sub>v</sub> S [<sub>v</sub> V O]]

(17) **Phase**

A phase is the minimal part of a top-to bottom computational process in which all the functional and selectional specifications associated to a given lexical head are satisfied.

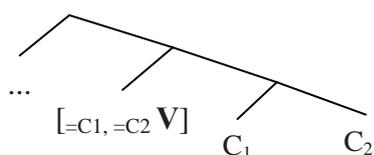
Intuitively, each phase corresponds to the computation of a "minimal chunk" of syntactic structure like (10) above. Importantly, each phase will have a fixed upper bound in depth, determined by a limited number of possible functional specifications (Cinque 1999, 2002) and of selected arguments (Pesetsky 1982). Note however that, contrary to the standard bottom-to-top derivation, here a phase does not correspond to a complete subtree. In fact, when Phase Projection is triggered by the last SELECT feature of the lexical head, the current phase gets closed, and the computation of the complement constitutes the next phase. Thus, a phase corresponds to a subtree whose lowest selected element is not yet expanded.<sup>15</sup> For the sake of simplicity, we assume here that only V and N can head a phase, and accordingly, phases correspond to the computation of a CP or DP chunk.<sup>16</sup>

Crucial to our argument is the distinction between *sequential* and *nested* phases.<sup>17</sup> As we have just said, when a phase reaches the lowest position selected by the lexical head, it is closed off: the expansion of the complement constitutes the next, *sequential* phase. A sequential phase thus follows the phase of the selecting head, and is separated from it.

On the other hand, any DP or CP within a phase  $P_n$  that does not occur in the lowest position selected by the lexical head of  $P_n$  constitutes a *nested* phase, which must be processed while  $P_n$  is still incomplete. Hence, all unselected DPs or CPs preceding the lexical head of  $P_n$  are necessarily nested phases: a preverbal subject, a fronted wh- or topical phrase can only be a nested phase (and additionally, when its computation is completed it is stored in the memory buffer of  $P_n$ ). In (15), for instance, the subject DP [*the boy*] constitutes a nested phase within the matrix CP phase.

Consider now a situation where a lexical head selects two complements. Since both are selected, in principle it is possible to apply Phase Projection for both:

## (18)



One possibility is to allow for both  $C_1$  and  $C_2$  to be computed as sequential phases (with  $C_2$  sequential to  $C_1$ , which is in turn sequential to the selecting V's phase).

Alternatively, we can make the more restrictive assumption that only  $C_2$  can be sequential to the V's phase, and  $C_1$  constitutes instead a nested phase. These assumptions will have different consequences for the islandhood of double complement structures (see below the discussion around (22)-(24)).

<sup>15</sup> Alternatively, the last SELECT feature can be satisfied by discharging an already processed constituent stored in the memory buffer by a previous application of Move.

<sup>16</sup> From our perspective, vP is not a separate phase from CP.

<sup>17</sup> The distinction between sequential and nested phases is independently justified by their different effects on the computational complexity function (see Chesi 2004 for thorough discussion).

If phases are minimal chunks of the syntactic computation, it is reasonable to assume that each phase has its own local memory buffer for Move.<sup>18</sup> However, since long-distance movement can cross phase boundaries, it is necessary to devise a way to transmit the content of a phase's memory buffer to that of another phase. For this purpose we adopt the following Success Condition:

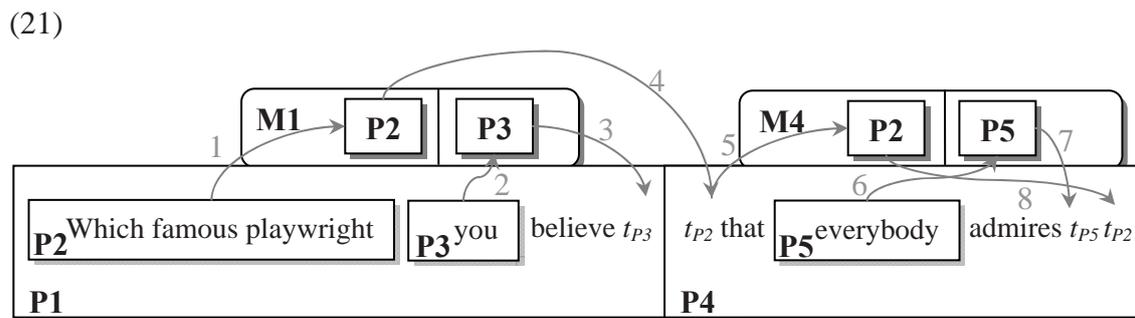
(19) **Success Condition**

At the end of each phase the local buffer is empty, or else its content is inherited by the memory buffer of the next sequential phase (if any).

Crucially, this condition only allows for communication between the memory buffer of two adjacent sequential phases. (Obviously, at the end of the last phase of all the local buffer will have to be empty). This accounts for the transparency of the lowest recursive branch of the tree.

To see this, consider for instance a computation for (20), as schematically represented in (21) (where the boxes identify phase boundaries):<sup>19</sup>

- (20) Which famous playwright do you believe that everybody admires?  
 [CP [DP Which famous playwright]<sub>i</sub> do you believe  
 [CP *whP<sub>i</sub>* [that everybody admires *whP<sub>i</sub>* ]?



The algorithm initializes a CP phase 1 (P1). Then it computes the *wh*-phrase, which constitutes a separate nominal phase 2 (P2). Since the *wh*-phrase is not selected, it is stored in the local memory buffer (M1) of P1 by Move (step 1). Then, the computation of P1 proceeds, down to the complement position of the matrix verb *believe* (we disregard the computation, storage and retrieval of the subject phase P3: step 2, 3). At this point P1 is closed and the *wh*-phrase (P2) in its memory buffer is discharged into the complement CP phase 4 (P4), since the latter is sequential and selected. We propose that this takes place by re-merging the content of the memory buffer of P1 in the left periphery of the complement CP, P4 (step 4); since this position is unselected, the *wh*-phrase is re-stored in the local memory buffer of P4 (step 5). As a result, the "inheritance" mechanism leaves an intermediate copy/trace in the edge of the complement CP phase.<sup>20</sup> The computation proceeds down to the object position of the

<sup>18</sup> This is our way to reconstruct Chomsky's "Phase Impenetrability Condition" for movement.

<sup>19</sup> The graphic representation (21) and the following ones are not very perspicuous, but it is the best possible representation on a bi-dimensional sheet that we could figure out. Intuitively, the left-to-right orientation of the written line corresponds to the progress of the computation in time; the boxes represent phase boundaries; the arrows represent steps involving storage into / retrieval from a memory buffer; the "sectors" in the memory buffer represent the different cells in which different types of moved elements are stored (cf. note 8).

<sup>20</sup> Although this assumption is not strictly necessary for the algorithm to work, it seems fairly natural and it allows us to capture various successive cyclicity effects, like e.g. Irish complementizer

verb *admires*, where the wh-phrase P2 is discharged from the local memory buffer of P4 and re-merged (step 8): the Success Condition is thus satisfied at the end of the computation.

Going back to the problem of double complements as in (18), if both are computed as sequential phases we expect both to be transparent for successive cyclic “extraction” or – to state the same thing from our current perspective – to be able to inherit the content of the memory buffer of the matrix V's phase. In fact, in a system like Chomsky (1986) or Cinque (1990) both complements are expected to be transparent, since both are selected (theta-marked) by the V head. However, the empirical evidence is not that simple. Kuno (1973, 380 ff.) pointed out long ago that for certain speakers, the first complement in a double complement structure resists subextraction:

- (22) a. John gave a picture of Mary a finishing touch.  
 b. <sup>??</sup> Who did John give [a picture of *t*] a finishing touch?
- (23) a. John handed a picture of Mary to Bill.  
 b. <sup>??</sup> Who did John hand [a picture of *t*] to Bill?
- (24) a. John gave moving to Florida serious consideration.  
 b. <sup>??</sup> Where did John give [moving to *t*] serious consideration? <sup>21,22</sup>

These data led Kuno to propose the “clause nonfinal incomplete constituent constraint” (cf. Pollard & Sag 1994:190), also subsuming subject islands. According to Kuno, this constraint only holds for some speakers. On the other hand, judging from the literature, subject islands of the type exemplified in (1)-(3) seem to be much more robust than the island effects in (22)-(24): we suspect that the two phenomena should not be collapsed under one and the same constraint. From our perspective, preverbal subjects can never be sequential phases (see §3 below for more discussion); however, the first complement in a double complement structure might be sequential or not, depending on the choice we make for the configuration (18). It is at least conceivable that certain speakers might be more restrictive, computing only the last complement as a sequential phase, while others might be less restrictive and allow for both complements to be computed as sequential phases. In our own native language, we do perceive a contrast in (25), where the first complement is a finite clause undergoing extraction in the (b) example:

- (25) a. <sup>(?)</sup> Ho annunciato [che licenzierò Maria] [a tutti i miei colleghi].  
 (I) have announced that (I) will-fire Mary to all my colleagues  
 b. <sup>?\*</sup> Chi hai annunciato [che licenzierai *t*] [a tutti i tuoi colleghi]?  
 Whom have (you) announced that (you) will-fire to all your colleagues

The data are not clear enough yet to draw a firm conclusion: we leave the issue at that. In the following discussion we will not consider double complement structure. To

alternations (McCloskey 1990) or French stylistic inversion (Kayne & Pollock 1978). We thank Luigi Rizzi for discussion of this point.

<sup>21</sup> According to Pollard & Sag (1994:182ff.), this configuration too gives rise to connectedness effects:

(i) Which of our relatives should we send [snapshots of *e*] [to *e*]?

<sup>22</sup> According to Kayne (1983), ECM subjects constitute left branch islands:

(i) \* [Which book]<sub>i</sub> do you believe [the first chapter of *e*<sub>i</sub>] to be full of lies?

This is exactly what the CC (and our reconstruction of it) would predict on the basis of the classical GB analysis of ECM complements. However, three informants whom we consulted found this sentence only mildly deviant. We leave the problem open for future research.

summarize, the following points of Chesi's (2004) model will be crucial for the development of our analysis:

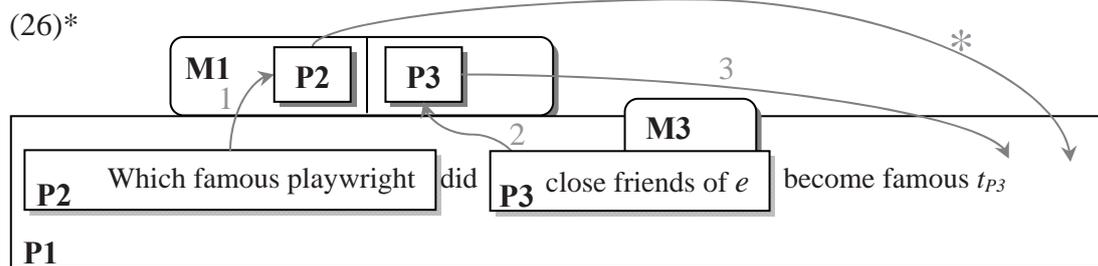
- a. Every computation is a top-down process divided into phases of fixed maximal size.
- b. A phase gets closed when the lowest selected position of its head is processed; the lowest selected complement constitutes the next sequential phase.
- c. All unselected constituents are instead nested phases: they are processed while the superordinate phase has not been closed yet.
- d. The Move operation stores an unselected element found before (i.e. to the left of) the head in the local memory buffer of the current phase, and discharges it in a selected position if possible; if not, when the phase is closed the content of the memory buffer is inherited by the next sequential phase. The memory buffer of the last phase must be empty at the end of the computation.

### 3. Left-branch islands are computationally nested phases

With this background, we can now go back to our initial problem, namely, left branch islands and the connectedness effect. The contrast in (1) is repeated here for convenience:

- (1) a. \*[Which famous playwright]<sub>i</sub> did [close friends of  $e_i$ ] become famous?  
 b. ?[Which famous playwright]<sub>i</sub> did [close friends of  $e_i$ ] admire  $e_i$  ?

Consider now a computation for (1b), as schematically represented in (26):



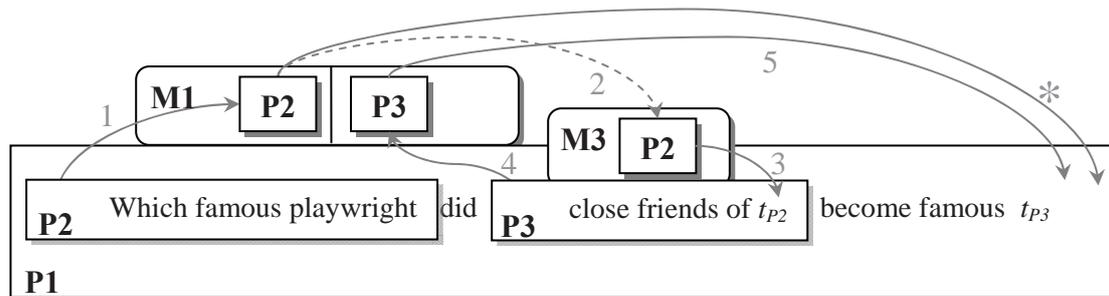
Once again, the algorithm initializes a CP phase 1; then it computes the wh-phrase in a separate nominal phase 2, and stores it in the local memory buffer of phase 1 (M1, step 1). The computation of phase 1 proceeds, inserting *did* in C. As a next step, a distinct nominal phase 3 for the subject DP must be opened, while the clausal phase 1 is still incomplete. The DP phase 3 is thus a nested phase, and its local memory buffer (M3) does not contain the wh-phrase which was stored in the memory buffer of phase 1 (M1): hence, the wh-phrase cannot be discharged in the selected gap position within the subject DP. The wh-phrase also remains undischarged at the end of the computation of phase 1, violating the Success Condition (19). This accounts for the strong island effect.

Suppose now that we optionally allow the memory buffer of the nested subject DP phase 3 to "copy" the buffer of the immediately superordinate phase 1, which contains the wh-phrase (this "parasitic copying" is represented by a dotted line in (27), step 2).<sup>23</sup> Then, the wh-phrase can be discharged in the gap position within the DP phase 3

<sup>23</sup> Optional copying would actually introduce non-determinism in the computation. In order to avoid this, we could assume either the possibility of using (a.) backtracking or (b.) parasitic usage (without discharging its content) of the first active memory buffer of a super-ordinate phase: when the computation of phase 3 reaches the position selected by the noun *friends*, since there is no more lexical

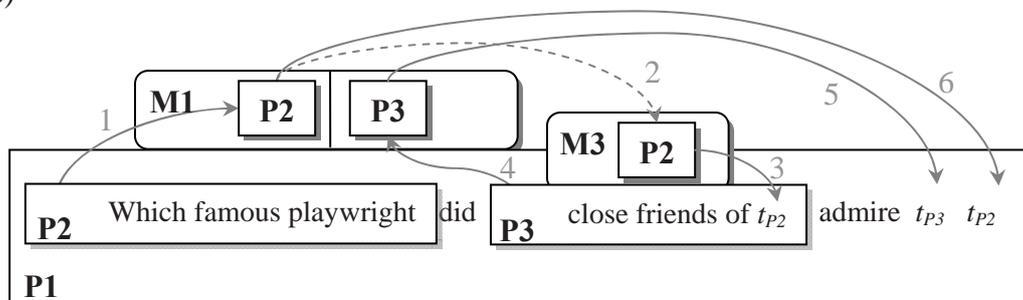
(step 3). However, this step will only empty the local memory buffer of phase 3. We crucially assume that parasitic copying into the memory buffer of the nested phase cannot discharge the memory buffer of the superordinate phase. As a result, even after the "parasitic gap" is computed, the local memory buffer M1 of the yet incomplete matrix phase P1 still contains the wh-phrase. This remains undischarged at the end of the computation, violating the Success Condition (19).

(27)\*



On the other hand, the copying mechanism does lead to a successful computation in the case of (1b). As in (27), the "parasitic" copy of the wh-phrase in the memory buffer of the subject DP phase 3 is discharged in the first gap position (step 3); however, the matrix CP phase 1 contains another selected position where the wh-phrase can also be discharged from the memory buffer of phase 1 (step 6). This derivation complies with the Success Condition, as shown in (28). This accounts for the connectedness effect.

(28)



Consider also the more complex configurations in (29) and (30) (from Kayne 1983):

- (29) a. ?a person who [cousins of [friends of  $e$ ]] usually end up hating  $e$   
 b. \*a person who [friends of  $e$ ]'s parents] usually end up hating  $e$

In these examples, the parasitic gap is embedded in a phase within the subject DP phase. However, in the (a) example the embedded phase is selected, hence sequential

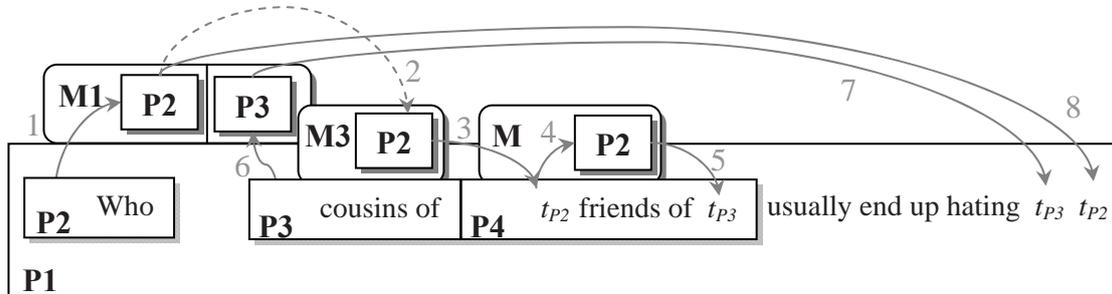
---

material available in phase 3 and the local memory buffer is empty, the system (a.) could backtrack and copy in the memory buffer of phase 3 the content of the buffer of the immediately super-ordinate phase 1; alternatively, (b.) the computation could simply access the first active memory buffer of a super-ordinate phase (i.e. a memory buffer with at least one element inside) and use the first available item without copying it in its memory buffer. The crucial point is that in both solutions "parasitic copying/usage" do not discharge the local buffer of phase 1. Data in (31) suggest the second solution (b.) is not restrictive enough and is empirically inadequate.

to the subject DP phase; in the (b) example, instead, the embedded phase is nested within the subject DP phase (which is itself nested in the relative CP phase).

Consider first a schematic computation of (29a), as represented in (30):

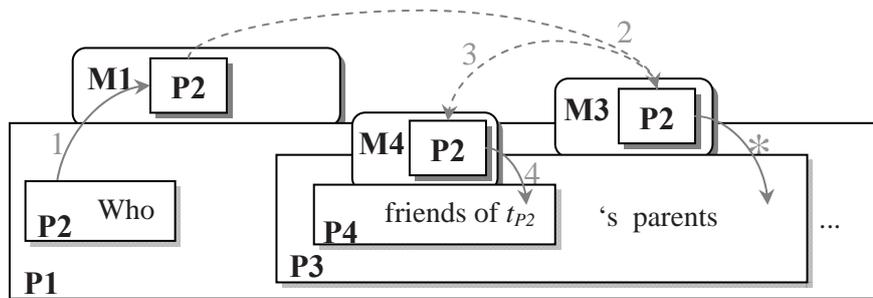
(30)



The relative phrase *who* is first stored in the memory buffer (M1) of the relative CP phase 1 (P1). Then, it is “parasitically” copied in the memory buffer (M3) of the subject DP phase 3 (step 2). When this phase gets closed, the content of M3 is inherited by the next phase sequential P4, that is, the complement of the lexical head *cousins*, via intermediate trace re-merge (step 3); M3 thus remains empty. The relative wh-phrase is then re-stored in the memory buffer of P4 (step 4) and locally discharged in the lowest position selected by the lexical head *friends* (step 5), so that M4 remains empty as well.<sup>24</sup> The computation goes back to the phase P1, and the wh-phrase is also discharged from M1 in the position selected by the matrix predicate (*hating*). All the memory buffers have been emptied by the end of the computation, so that the Success Condition is satisfied.

Consider now a schematic computation for (29b):

(31)



As the dotted lines make clear, here “parasitic copying” must apply twice: first, the wh-phrase is copied from the memory buffer M1 of the relative CP phase onto the memory buffer M3 of the subject DP phase P3 (step 2); second, the wh-phrase is parasitically copied from the memory buffer of P3 into the memory buffer M4 of the unselected phase P4 in pre-nominal position (step 3). The lexical head of P4 (*friends*) selects a complement position where the wh-phrase can be re-merged (step 4); this will empty the memory buffer of P4, but crucially, the buffer of P3 (M3) remains undischarged (since it has been parasitically copied, but not inherited by a phase sequential to P3). Since the head of P3 (*parents*) has no other selected position where

<sup>24</sup> Note that under the “backtracking” view (cf. the preceding note), the system should here backtrack from the doubly embedded phase which contains the parasitic gap to the CP phase 1.

the wh-phrase could possibly be discharged, at the end of the computation of P3 the local memory buffer remains non-empty, in violation of the Success Condition, so that the derivation fails.

To conclude this discussion of left branch islands, let us summarize the main aspects of the proposed analysis. The Connectedness Condition has been recast in derivational terms, by assuming:

- a. a top-to-bottom derivation divided in phases
- b. a “storage” conception of the Move operation, which stores an unselected element in the local memory buffer of the current phase and re-merges it in a selected position;
- c. a distinction between sequential phases (corresponding to the “canonically governed” branches on the recursive side of the tree) and nested phases (corresponding to the “non canonically governed” branches on the non-recursive side of the tree).

The crucial element in our account of left branch islands is the idea that the content of the memory buffer of a phase can only be inherited by the next sequential phase, and not by a nested phase; in other terms, the content of the memory buffer can be “bequeathed” only after the relevant phase has been completed. In order to account for parasitic gaps licensed under connectedness, we have allowed for the possibility of parasitically copying the content of the buffer of a matrix phase into the buffer of a nested phase; this parasitic copy, however, cannot empty the matrix memory buffer, whence the necessity of *another* (selected) gap within the matrix phase itself (or within a phase that is sequential to the matrix one).

At this point, we return to the problem of right-hand strong islands, which apparently lie on the recursive side of the tree.

#### 4. Right-hand modifiers

As we noted in the introduction, Kayne’s original CC only accounts for left branch islands; it does not account for the islandhood of right-hand modifiers, which, however, also show connectedness effects, as was pointed out by Longobardi (1985):

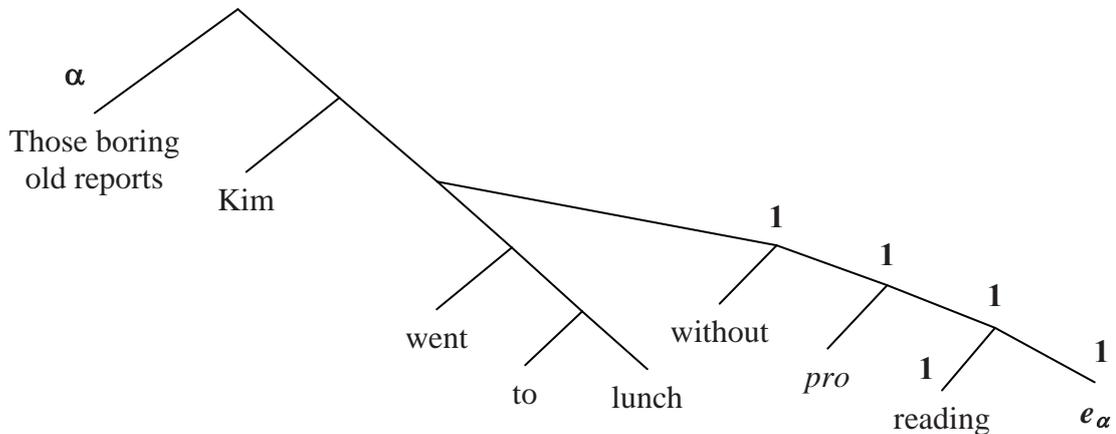
- (32) a. <sup>??</sup>Those boring old reports, Kim went to lunch [without reading  $e_i$ ].  
 b. <sup>√</sup>Those boring old reports, Kim filed  $e_i$  [without reading  $e_i$ ].

In order to subsume this kind of data under the CC, Longobardi (1985) proposed to strengthen the notion of g-projection, by requiring that each maximal g-projection be properly governed. We report below the relevant definitions.

- (33)  $\alpha$  **governs**  $\beta$  iff  $\gamma$
- i.  $\alpha$  is lexical or  $\alpha$  and  $\beta$  are coindexed and
  - ii.  $\beta$  is minimally contained in the maximal projection of  $\alpha$ , or  $\beta$  is in the Spec or the head of  $\gamma$ ,  $\gamma$  is minimally contained in the maximal projection of  $\alpha$ .
- (34)  $\alpha$  **properly governs**  $\beta$  iff
- i.  $\alpha$  governs  $\beta$  and
  - ii.  $\beta$  is (in) a complement (or predicate) of  $\alpha$ .
- (35) A **non properly governed** maximal projection is a boundary to the extension of g-projections.

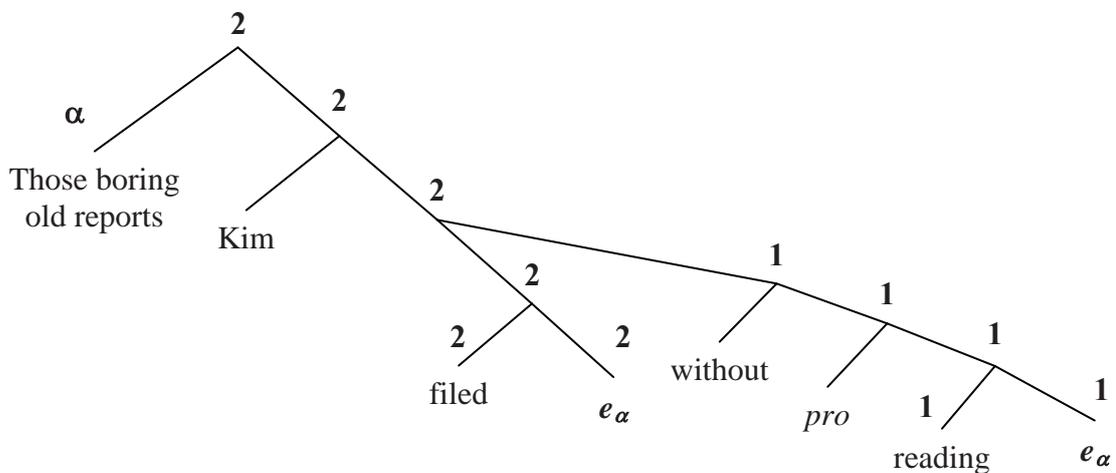
It is easy to see that under these definitions, a right-hand adjunct clause blocks the upward extension of g-projections even though it lies on the canonically governed side of the tree in a VO language like English:

(32)a.\*



In (32b), on the other hand, whatever the precise level of adjunction of the adverbial clause, the g-projection path stopping at the top of this clause can connect to that of the VP-internal gap, as shown in the following tree representation:

(32)b.



The adjunct island is thus assimilated to the subject island, much as in Huang's (1982) Condition on Extraction Domains.<sup>25</sup>

There are some problems with this move, though. First, many authors have pointed out that not all right-hand adjuncts give rise to strong island effects, whereas subject islands are much more robust (see for instance Pollard & Sag 1994, 191; Levine & Sag 2003, §3.2; Haider 2003, among many others):

- (36) a. Who did you go to Girona [in order to meet *e*]?
  - b. This is the blanket that Rebecca refuses to sleep [without *e*].
  - c. How many of the book reports did the teacher smile [after reading *e*]?
- (Pollard & Sag 1994)

<sup>25</sup> For recent reformulations of the CED, see Saito & Fukui (1998) and Nunes & Uriagereka (2000).

- (37) a. the car that he left his coat [in  $e$ ]  
 b. the day that she was born [on  $e$ ]  
 c. \*the day that she was born in England [on  $e$ ]

(Haider 2003, 3)

On the basis of this difference, Pollard & Sag (1994) decide not to unify adjunct islands with left branch (subject) islands. Another possibility is to divide the class of right-hand modifiers in two subclasses: true adjuncts (which give rise to strong islands) and “oblique complements”, inserted in a complement position (cf. e.g. Larson 1988, 1990), which may be transparent for extraction, or at least, do not constitute real adjunct islands.<sup>26</sup>

Another problem with Longobardi’s extension of the CC arises with respect to complex NP islands of the relative clause type.<sup>27</sup> Consider the following examples (from Kayne 1983):

- (38) a. \*A person who [people [<sub>CP</sub> that talk to  $e_i$ ]] usually have money in mind  
 b. ?A person who [people [<sub>CP</sub> that talk to  $e_i$ ]] usually end up fascinated with  $e_i$

The connectedness effect in (38b) implies that the relative clause does not block the extension of g-projections up to the NP/DP node which itself constitutes a left branch.

This means that the complex NP island effect on extraction does *not* follow from the blocking of g-projections at the relative CP level, but it must be dealt with by an independent constraint. This was actually an immediate consequence of Kayne’s (1983) approach, since a right-hand relative CP is in a canonical government configuration w.r.t. the modified NP. On the other hand, Longobardi (1985) assumes that the relative clause is properly governed because it is predicated of its sister node (cf. definition (34)): this accounts for the possibility of a connectedness effect in (38b), but leaves the complex NP island unaccounted for, as Longobardi himself points out.<sup>28</sup> Thus, even Longobardi’s (1985) extension fails to cover all the island effects that have been descriptively classified as “strong islands”.

In sum, the variable island effects of right-hand modifiers and the unresolved status of the relative clause complex NP island in the connectedness approach cast some doubt on the idea that right-hand modifiers should be assimilated to left branch

<sup>26</sup> Our approach, exactly like Kayne’s (1983) original CC, cannot account for “symbiotic gaps” of the kind discussed by Levine & Sag (2003, §§ 2.3 and 3.2):

(i) What kind of books do [authors of  $\_$ ] argue about royalties [after writing  $\_$ ]  
 unless the right-hand adverbial clause is of the “transparent” type (cf.(36), (37)); this is actually close to the conclusion drawn by Levine & Sag (§3.2).

<sup>27</sup> As for complex NP islands of the N-complement type, it has been repeatedly noted in the literature that they are weaker islands than the relative clause type. From our perspective, if the CP complement is selected by the N head, then it constitutes a sequential rather than a nested phase (and the island effect does not follow from computational nesting of the CP complement). We leave aside this kind of island effect.

<sup>28</sup> Actually, we suspect that there might be something more going on in examples like (38b). Chomsky (1986, 48 ff.), quoting Chung & McCloskey (1983), discusses the fact that the complex NP island effect is sometimes suspended when the subject position is relativized:

(i) ? This is a paper<sub>1</sub> that we need to find [someone who understands  $e_1$ ]  
 (ii) \* This is a paper<sub>1</sub> that we need to find [someone that we can intimidate  $e_2$  with  $e_1$ ]

Chomsky proposes the Vacuous Movement Hypothesis, whereby in (i) the relative operator does not move from the subject position to Spec,CP, and the latter is available as an escape hatch for the extracted phrase. (This is actually equivalent to the GPSG analysis proposed by Chung & McCloskey.) Notice that the connectedness effect in (38b) involves a complex NP island whose relative clause has a relativized subject, much as in (i). We suspect that this might play a role in the acceptability of (38b), but we leave the problem open for the time being.

islands, as in Longobardi's approach (cf. Pollard & Sag 1994, Levine & Sag 2003 for a similar conclusion).

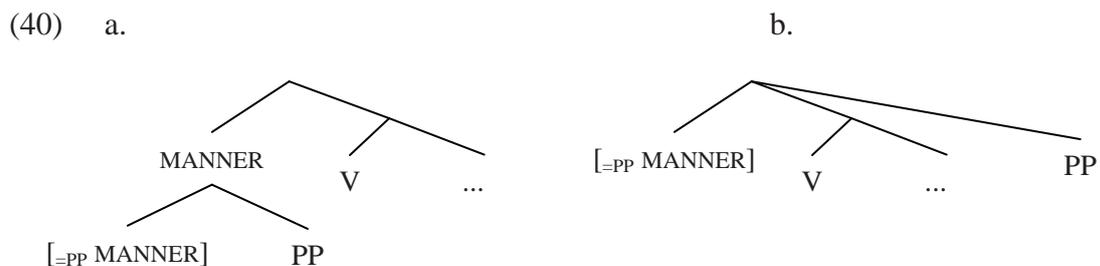
From our present perspective, right-hand modifiers have a somewhat intermediate status w.r.t. our classification of nested Vs. sequential phases. On the one hand, they follow the lexical head of the phase and its complements, so that the superordinate phase is potentially complete; on the other hand, since they are not selected by the lexical head, they cannot be inserted in the computation by an application of Phase Projection (12). In the following discussion we tentatively sketch out a possible treatment of right-hand adverbials as nested phases.

First we will assume, as in the strictly cartographic approach proposed by Cinque (1999, 2002), that each modifier corresponds to a functional specification of the lexical head: namely, in our terms, to a licenser feature. In particular, we assume that a subset of the licenser features (which we can dub Mod(ifier) features) are intrinsically relational, in that they license a relation between a subpart of the lexical head's extended projection and a modifying constituent (e.g. a prepositional phrase). For instance, a MANNER feature placed at a given point of the licenser hierarchy will establish a link between the lower structure and a manner-modifying phrase; a TEMPORAL LOCATION feature will establish a link between an event (or time) denoting portion of the verb's extended projection and a temporal adverbial.

The mediating role of the Mod features is rendered necessary by the fact that, in our system, the modifier cannot directly select the modified portion of clausal structure (as proposed by Gonzalez Escribano 2004): this would in fact interrupt the continuity of the modified phase. It is fair to say that the stipulated Mod features are equivalent to silent functional heads, which mediate the relation between the modifier and the VP. As a concrete implementation of this idea, we assume that a Mod feature, e.g. MANNER, has a selectional specification associated with its (empty) head, e.g.:

(39) [<sub>=PP</sub> H<sub>manner</sub>]

The Mod feature effectively acts as a head: it selects an argument, the PP, giving rise to the configuration presented in (40.a):



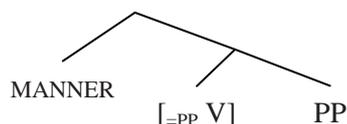
Since manner is a licenser feature of the verb, it has to be computed while this verbal phase is still open: hence, the selected PP constitutes a nested phase.

The right-hand position of the selected PP can be attributed to the special status of the Mod feature. We stipulate that the selected PP can be actually inserted/expanded (by Phase Projection) in the selected position only after the matrix phase select features have been projected (as shown in (40.b)). Note that although the right-hand modifier is selected, it is not selected by the verbal lexical head (unlike Larson 1988, 1990)<sup>29</sup> hence, it is not a sequential phase but a nested one.

<sup>29</sup> In this way the modifier is structurally superior to the VP-internal constituents; this avoids a number of problems with a generalized Larsonian "adjunct as complement" analysis (see Bianchi 1997, 2000, 2001 for detailed discussion).

Moreover, we could capture the unexpected behavior of some right-hand modifiers which seem to be transparent for extraction (cf. (36)-(37)) by assuming a minimal difference between (40.b) above, where the island PP is “selected” by a Mod feature, and (40.c), where the select feature is specified on the verbal head rather than on the Mod feature, so that the PP constitutes a sequential phase.<sup>30</sup>

(40) c.



This is just a sketch of a possible analysis, whose development we leave for further research.

### 5. Further prospects and conclusions

To summarize, in this paper we have proposed an approach to strong islands and to the connectedness effect within a top-to-bottom derivational framework (formalized in Chesi 2004, 2007).

Among the problems we cannot discuss in these few pages, a major one that immediately springs to mind is the status of phases in a strictly head-final language like Japanese. In such a language all phases, whether selected or not, seem to (linearly and) temporally precede the processing of the superordinate phase’s head; thus, at first sight, they all appear to be “nested”. Nevertheless, head-final languages do not block extraction from “selected” pre-head phases; on the contrary, it has been claimed that they even allow extraction from subjects<sup>31</sup> (as is actually predicted by Kayne’s original CC, since in these languages left branches are in a canonical government configuration):

- (41) a. ?Nani-o<sub>i</sub> [John-ga [NP [IP Mary-ga t<sub>i</sub> katta] koto]-o mondai-ni siteru] no.  
 what-ACC John-NOM Mary-NOM bought fact-ACC problem-into making Q  
 ‘What<sub>i</sub>, John is making an issue out of [the fact that Mary bought t<sub>i</sub>].’
- b. ?Nani-o<sub>i</sub>[ John-ga [CP [NP [IP Mary-ga t<sub>i</sub> katta] koto]-ga mondai-da to]  
 what-ACC John-NOM Mary-NOM bought fact-NOM problem-is that  
 omotteru] no.  
 think Q  
 ‘What<sub>i</sub>, John thinks that [the fact that Mary bought t<sub>i</sub>] is a problem.’  
 (Saito & Fukui 1998)

This fact could be captured within the proposed model in three alternative ways. One possibility is to propose a parameterization of the Linearization Principle that would allow the selected phases to be linearized to the left of the head; this solution is however not fully satisfactory, since it would predict a generalized transparency of all left-hand constituents (e.g. adjuncts), which is incorrect (Saito & Fukui 1998).

<sup>30</sup> The idea that a selectional specification for a manner PP can be associated directly to a lexical head is made plausible by the existence of “selected adjuncts” (cf. Rizzi 1990): e.g., a verb like *behave* requires a manner specification; a verb like *weight* requires a measure specification of a certain kind; a verb like *be born* requires a locative or temporal specification, etc.

<sup>31</sup> We thank Shoichi Takahashi for discussion of these data.

Another possible approach to pre-head selected phases in head-final languages is to consider them a. properly selected (Chesi 2008) or b. selectors (Choi and Yoon 2006). In the first case (Chesi 2008) the phase head is actually introduced before the arguments even though the phase head is spelled out at the very end of the phase. In the second case (Choi and Yoon 2006) nominal heads select verbal head, against standard assumptions. Obviously all these solutions would deserve more thorough discussion that we cannot carry forward in these pages.

To conclude, we also wish to point out some further consequences of our general approach: first, the top-to-bottom orientation of the computation allows for a relatively straightforward solution to the problem of phase-by-phase linearization, since phases (both nested and sequential ones) are processed in a well defined order, driven by the LICENSOR and SELECT features of the relevant phase heads.<sup>32</sup> Second, the storage conception of Move avoids the "teleological" mechanism of raising to the edge of each phase in case of bottom to top successive-cyclic movement: such moves are teleological in that in the lower phases the final trigger of movement, i.e. the probe/EPP head, has not been inserted yet.<sup>33</sup>

To the extent that our proposal is tenable, it supports a general conception whereby considerations of computational efficiency and cognitive plausibility at the interface with the performance tasks directly constrain the architecture of the grammar itself.

## References

- Adger D. 2003. *Core Syntax*. Oxford University Press.
- Adger D. 2007. *A Minimalist Theory of Feature Structure*. <http://ling.auf.net/lingBuzz/000583>
- Bianchi, V. 1997. On the structural position of time clauses. *Quaderni del Laboratorio di Linguistica* 11, 66-90. Pisa, Scuola Normale Superiore.
- Bianchi, V. 2000. On Time Adverbials. *Italian Journal of Linguistics* 12.1, 77-106 (monographic issue "On adverbs and adverbial modification", eds. D. Delfitto & N. Corver).
- Bianchi, V. 2001. Antisymmetry and the Leftness Condition: Leftness as anti-command. *Studia Linguistica* 55, 1-38.
- Chesi, C. 2004. *Phases and Cartography in Linguistic Computation: toward a Cognitively Motivated Computational Model of Linguistic Competence*. Ph.D. Thesis, University of Siena.
- Chesi, C. 2007. An introduction to Phase-based Minimalist Grammars: why move is Top-Down from Left-to-Right. *STiL - Studies in Linguistics, CISCL Working Papers* Vol.1, 38-75
- Chesi, C. 2008. *Rightward movement from a Top-Down perspective*. Rightward Movement in a Comparative Perspective; Workshop at the DGfS Meeting, February 27-29, 2008. Bamberg

---

<sup>32</sup> See van Riemsdijk (2005) on the linearization problems in a bottom-to-top computation.

<sup>33</sup> As far as we can see, the problem is not really solved in Fox & Pesetsky's (2004) approach: the wh-phrase moves to the edge of the phase (linearization domain) in order to be linearized to the left of the other internal constituents of the phase; however, the necessity to alter the linearization derives from the ultimate goal of reaching the final left-hand landing site. A common solution is to stipulate uninterpretable wh- or EPP features, which are however justified only by internal constraints of the bottom to top derivation. See McCloskey (2001) for a defense of such features and Rizzi (2002) for a critical discussion.

- Choi, Y., Yoon, J. 2006. *Argument Cluster Coordination and Constituency Test (Non)-Conflicts*, NELS 37, University of Illinois at Urbana-Champaign.
- Cinque, G. 1990. *Types of A' dependencies*. Cambridge, MA, MIT Press.
- Cinque, G. 1999. *Adverbs and Functional Heads*. Oxford University Press.
- Cinque, G. 2002 *Complement and adverbial PPs: implications for clause structure*. Abstract, University of Venice.
- Cinque, G. 2003. Issues in adverbial syntax. *Lingua* 114, 683-710.
- Culicover, P. & P. Postal (eds.). 2001. *Parasitic Gaps*. Cambridge, MA, MIT Press.
- Fox, D. & David P. 2004. *Cyclic Linearization of Syntactic Structure*. Ms., MIT.
- Grimshaw, J. 1991. *Extended projection*. Ms., Brandeis University.
- Haider, H. 2003. Pre- and post-verbal adverbials in OV and VO. *Lingua* 114.
- Kayne, R. S. 1983. Connect & M. Calcagno. 2001. Parasitic gaps in English: some overlooked cases and their theoretical implications. In *Parasitic Gaps*, ed. P. Culicover and P. Postal, 181-222. MIT Press.
- Kayne, R. S. 1994. *The Antisymmetry of Syntax*. MIT Press.
- Levine, R., I. A. Sag. 2003. Some empirical issues in the grammar of extraction. In S. Müller (ed.), *Proceedings of the HPSG03 Conference*, Michigan State University, East Lansing. CSLI Publications.
- Longobardi, G. 1985. *The theoretical status of the adjunct condition*. Ms. Scuola Normale Superiore, Pisa.
- Longobardi, G. 1985b. Connectedness, scope, and c-command. *Linguistic Inquiry* 16.
- McCloskey, J. 2001. The morphosyntax of WH-extraction in Irish. *J. Linguistics* 37:67-100.
- Munn, A. 2001. Explaining parasitic gap restrictions. In P. Culicover & P. Postal (eds.), 369-392.
- Nunes, J. & Uriagereka, J. (2000). Cyclicity and extraction domains. *Syntax*, 3,20-43.
- Pesetsky, D. 1982. *Paths and Categories*. PhD thesis, MIT.
- Phillips, C. 1996. *Order and Structure*. PhD thesis, MIT.
- Pollard, C. & I. Sag. 1994. *Head-driven Phrase Structure Grammar*. Stanford: CSLI.
- Postal, P. M. 1994. Contrasting extraction types. *Journal of Linguistics* 30,159-186.
- Rizzi, L. 1990. *Relativized Minimality*. MIT Press.
- Rizzi, L. 1994. Argument/adjunct (a)symmetries. In Cinque et al. (eds.), *Paths towards Universal Grammar*, 361-376. Georgetown University Press.
- Rizzi, L. 1997. The fine structure of the left-periphery. In *Elements of Grammar* ed. Haegeman, L. Dordrecht: Kluwer.
- Rizzi, L. 2002. Locality and left periphery. In Belletti, A., ed. (2002) *Structures and Beyond. The Cartography of Syntactic Structures*, vol. 3, Oxford University Press.
- Saito & Fukui 1998. Order in Phrase Structure and Movement. *Linguistic Inquiry* 29.3,439-474
- Stabler, E. 1997. *Derivational minimalism*. in Retoré, ed. Logical Aspects of Computational Linguistics. Springer
- Starke, M. 2001. *Move dissolves into Merge*. Ph.D. Thesis, University of Genève.

# Number and Case in the comprehension of relative clauses: Evidence from Italian and Greek

MARIA TERESA GUASTI

University of Milano-Bicocca  
Mariateresa.guasti@unimib.it

STAVROULA STAVRAKAKI

University of Thessaloniki  
SVOULA@AUTH.GR

FABRIZIO AROSIO

University of Milano-Bicocca  
fabrizio.ariosio@unimib.it

## 1. Introduction

Relative clauses have been extensively studied from several perspectives. Cross-linguistic research findings indicate a relatively consistent pattern of performance since they show that subject relatives are significantly easier to acquire and process than object relatives (Frauenfelder, Segui & Mehler 1980; King & Kutas 1995; Schriefers, Friederici & Kuehn 1995, among others). These findings have been interpreted in terms of the length of the dependencies required for subject and object relatives. As shown in the example below, the filler gap dependency in subject relatives is shorter than in object ones:

(1) a. The woman [who/that t is kissing the man] SUBJECT RC

b. The woman [who/that the man is kissing t] OBJECT RC

On the assumption that shorter dependencies are less demanding in processing (and are more economical in grammar, Chomsky, 1995) than longer dependencies, the parser is predicted to opt for the shorter dependency and thus to prefer subject relatives over the object ones in the first-pass parse. This preference is well known in the processing literature as the Minimal Chain Principle (MCP, De Vincenzi 1991) or the Active Filler hypothesis (Frazier and D'Arcais, 1989; Frazier & Clifton 1989). Thus, the parser should always start with a subject relative clause analysis, which is nevertheless abandoned when analysis of an object relative clause is required. In other words, the subject relative clause interpretation is the preferred interpretation and the object relative clause interpretation is obtained through reanalysis, which requires

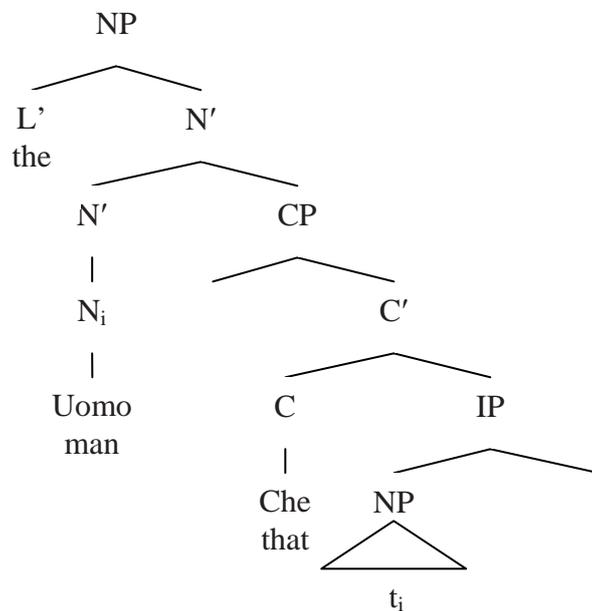
additional time. A fundamental question of the way sentence processing takes place concerns the use of possible cues that contribute to the successful reanalysis of object relative clauses. In Fodor and Inoue (2000) Diagnosis model, reanalysis is easy whenever the information delivered at the disambiguating point clearly indicates which is the error accomplished in the first-pass parse and how to repair it. Whilst for adults this reanalysis is expected to be more or less easy, for children its impact is more dramatic in that reanalysis is performed or not depending on the nature of the informativeness of disambiguating cue. Thus, it may be possible that certain types of object relatives are more difficult to comprehend than other types and that different types of object relatives are acquired at different points of development. What is at a stake is not whether children form relative clauses as adults do (through recursion and wh-movement or head movement), as we believe that there is no difference at this level. Differences exist in the readiness of performing reanalysis, which in turn depends on the informativeness of the disambiguating cues. Following this line of reasoning, acquisition data that show developmental delay in the comprehension of object relatives can be interpreted in terms of difficulty during the diagnoses process that leads to reanalysis. Within this analysis it is expected that language-specific cues, which contribute to disambiguation, have an effect on children's acquisition of the various types of object relative clauses. Furthermore, it is expected that cross-linguistic variation in the acquisition of object relative clauses are significantly determined by the informativeness of the disambiguating cue. As for adults, this view is strongly supported by current research findings. In particular, the processing of subject-verb ambiguities in German (Meng & Bader, 2000) indicated that recovery from a garden path is easier when disambiguation is obtained through case information than through number agreement morphology. What the findings by Meng & Bader show is the differential effects of number and case on successful resolution of a temporary ambiguity. In a similar vein, Arosio, Adani and Guasti (2007) showed that Italian children's comprehension of object relative clauses is modulated by different disambiguating cues. In (2a) we have an example of a subject relative in Italian, while in (2b) and (2c), we have object relatives disambiguated respectively by the position of the embedded subject or by number agreement on the embedded verb.

- (2) a. Fammi vedere l'uomo che saluta le signore  
'Show me the man that is greeting the ladies'
- b. Fammi vedere l'uomo che la signora saluta  
'Show me the man that the lady is greeting'
- c. Fammi vedere l'uomo che salutano le signore  
'Show me the man that are greeting the ladies'  
'Show me the man that the ladies are greeting'

Although the disambiguating information for (2b) and (2c) is found at the same position, just after the complementizer (that), comprehension of these two types of object relatives yielded different results. Typically Developing (TD) Italian speaking children from 5 to 11 years were better at comprehending subject (2a) than object relative clauses (2b,c), but had particular difficulties with the object relative in (2c): while at age 5 comprehension of (2b) was around 70%, comprehension of (2c) was around 25% and it was not until age 11 that the comprehension of (2c) reached adult

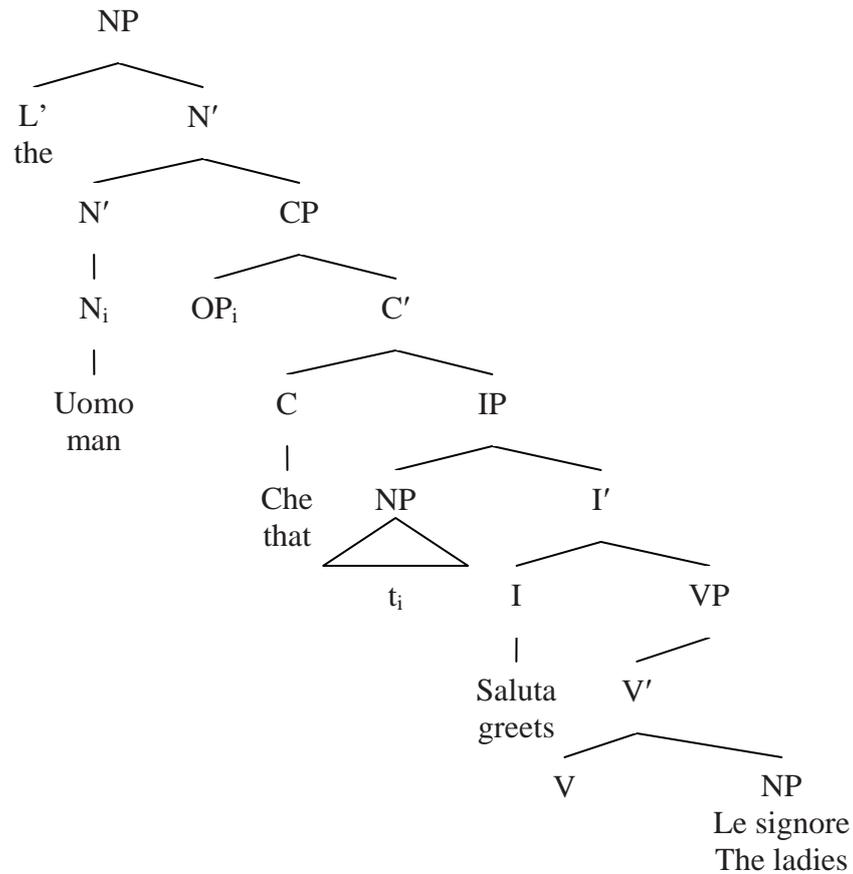
levels. In the framework assumed here this means that whilst disambiguation by syntactic position resolved garden path effects successfully, disambiguation by number agreement on the verb did not until age 11. Let us consider this asymmetry more closely within Fodor and Inoue's Diagnosis model. Assume that the parser works incrementally so that NPs are immediately assigned a grammatical function. Furthermore, when a mismatch (or an error) is encountered that contrasts with the preferred analysis, the parser's action consists in a series of steps, each of which involves the correction of an illegal grammatical function. The initial portion of the three relative clauses in (2) is similar: after the complementizer (that) has been encountered a trace is postulated in Spec IP and the grammatical function Subject is assigned to the chain including the head of the relative clause and the trace, as illustrated in (3).

(3)



For the sentence (2a), the whole representation would be as in (4) and the postverbal NP "le signore" would be assigned the grammatical function object.

(4)



In (2b) and (2c), the plan initiated in (3) goes wrong, when the disambiguating portion of the clause, that is the preverbal subject or the inflected verb, is encountered. Although these two distinct pieces of information arrive at the same point, i.e., after the complementizer, Arosio et al.'s data suggest that the position of the embedded subject is more effective at an earlier point of development than number agreement on the embedded verb. This is because the former directly informs the parser about the solution (it is a positive symptom), while the latter does not, according to Fodor and Inoue, and this affects the process of reanalysis. Let us see why. When the preverbal subject is encountered, the subject trace in Spec IP has to be ousted, so that the NP subject could be put in that position. The subject function must be removed from the chain (head of the relative – t) and linked to the new NP. Thus, the error is remedied and a reanalysis is easily performed: there was a subject and now there is a new subject and the only remaining thing to do is to look for a new gap or trace to associate to the relative head. Things are more complex for (2c). The number on the verb disagrees with the chain (head of the relative – t) that is assigned the grammatical function subject; this means that the trace and the head of the relative must be decoindexed. This eliminates the original error, but leaves the head of the relative in need of a trace and the trace in subject position in need of a licenser.<sup>1</sup> In addition, the

<sup>1</sup> We conjecture that if the verb would be a second or first person verb, like in (i), reanalysis would be easier. The trace could be changed into a null referential subject pro and it would be immediately obvious who the referent of the null subject would be (the hearer or the speaker)

fammi vedere l'uomo che stai/sto salutando  
show me the man that (you/I) are/am greeting

verb argument structure specifies that two arguments are needed, but it is not clear to whom they can be assigned. The subject trace could retain its grammatical function, but it would remain unlicensed. Given that Italian is a null subject language, the parser could postulate a referential pro subject in Spec IP (ousting the trace). This pro would be licensed by agreement on the verb. At this point, it could look for a trace for the head of the relative and the only position available would be the object position. This move would be in line with the MCP (De Vincenzi, 1991). Only one chain between the head of the relative and the trace in object position is built. For the subject no chain is necessary. It is also in line with the assumption of the Diagnosis model: one grammatical function was assigned and it is assigned again; in fact, even the second grammatical function is assigned. The mismatch is solved and a formally legitimate representation (at least up to verb) is built. However, the pro subject fails to be identified, as no antecedent is provided. In addition, when the postverbal NP is encountered a new reanalysis would have to be attempted. It is likely that all this work exceeds children's computational capacities and therefore object relative clauses disambiguated by number agreement are particularly taxing. The right move would be to posit an expletive pro in Spec IP and wait for a postverbal subject; posit a trace in object position and connect it to the head of the relative. In this way two chains are built: pro-NP (expletive pro and postverbal subject) and the chain including the head of the relative and its trace, resulting in a maximal violation of the MCP. Doing this kind of reanalysis seems already hard for adults, as proven by Penolazzi et al. (2005) based on Italian Wh-questions and by Meng & Bader, based on German Wh-questions. It would be no surprise if such kind of reanalysis is beyond the capacities of Italian children at a certain age. Fodor and Inoue call number agreement a negative symptom: number on the verb informs the parser that something went wrong, but does not inform it how the tree has to be reconstructed, i.e. how to reanalyze the sentence. In fact, as we have seen in the case of Italian, more than one option is indeed open. In addition, after the error has been corrected, it leaves two grammatical relations unlinked and two chains (the one headed by the relative head and the one including the subject trace) open and this may require resources that children at certain age do not have.

At first sight, this explanation of the difficulties experienced by Italian children with object relatives does not seem to hold for Greek. Stavrakaki (2001) found that TD Greek speaking children between 3;4 and 5;0 comprehend right branching object relative clauses like in (2c) as well as subject relative clauses (75% correct comprehension) (see also Stavrakaki, 2002 and Stathopoulou, 2007 for production). All these findings are in line with those reported by Varlokosta (1997) who found that Greek children opt for a movement strategy to form object and subject relatives.

A possible interpretation of the difference in the acquisition of relative clauses by Greek and Italian children is related to the specific grammatical characteristics of Greek and Italian. Usually NPs in Greek carry case information and case provides disambiguating information in relative clauses. This is not so in Italian, where NPs are not overtly marked for case. Consequently, whilst the subject relative clauses analysis in both Italian and Greek is a default choice due to the minimal chain principle (De Vincenzi 1991), the interpretation of object relatives requires reanalysis of the default

---

The subject function would be linked to the null subject, i.e., one grammatical function would be corrected. Then, the head of the relative would be connected to a trace in object position. We also predict that reanalysis would be easier, if a context for a third person subject would be provided before the relative clause in (2c). In this case, the trace could be changed into a null referential subject identified by an antecedent in the previous discourse.

subject interpretation and thus becomes highly dependent on the nature of cues involved in the ambiguity resolution.

While the hypothesis of the cue effect on the interpretation of relative clauses by Italian and Greek is quite plausible, a direct comparison between the available Greek and Italian data cannot be offered, as methods and experimental materials were different. The present study is a direct test of this hypothesis, as it offers a comparative investigation of the acquisition of relative clauses by monolingual Greek and Italian children. More specifically, this study addresses the question of the cross-linguistic differences in the acquisition of Greek and Italian relatives by systematically investigating the effect of factors that can possibly affect children's comprehension of relative clauses. In particular, we aim at evaluating the effect of (i) number agreement morphology, (ii) overt morphological case marking on Greek on NPs and (iii) syntactic position of the embedded subject. It includes two experiments. In the first one, we investigate Greek and Italian children's comprehension of subject and object relatives disambiguated by number agreement on the verb or by syntactic position of the embedded subject (plus number agreement). In the second one, we compare Greek children's comprehension of subject and object relative clauses disambiguated by number agreement or case.

## 2. Experiment I

In this experiment we tested comprehension of Greek and Italian relative clauses and compared the results directly. Direct comparison was possible because we neutralized case on Greek NPs through the use of neuter gender that is morphologically underspecified for case and thus ambiguous between Nominative and Accusative. In this condition, the disambiguation is brought about by number agreement on the verb in both languages or by the syntactic position of the embedded subject. On the basis of our previous discussion, we expect that Greek and Italian speaking children behave similarly, i.e., we expect that subject relative clauses are easier to comprehend than object relative clauses. We also expect that disambiguation by number agreement is more difficult than disambiguation by syntactic position of the embedded subject, as it was found by Arosio et al. in Italian.

### *Materials and Method*

*Participants:* Twenty Italian- (M=5;1 SD=0.4 Range: 4;5-5;9) and 37 Greek-speaking children (M=5, SD=0.25 Range: 4;5-5;6) participated to the experiment. Children attended nursery schools in Italy and Greece respectively.

*Materials:* We constructed 6 triples of 3 types of clauses each introduced by the lead-in "Show me". Each triple included a subject-extracted relative clause and an object extracted relative clause with the embedded subject in the post-verbal position, as illustrated in (5). See Appendix A for a complete set of critical sentences.

- (5) a. Il cavallo che sta inseguendo i leoni **OS RC**  
 a'.To alogo pou kiniga ta liontaria  
 'The horse that is chasing the lions'
- b. Il cavallo che i leoni stanno inseguendo velocemente **OO RC Pre-S**  
 b'.To alogo pou ta liontaria kinigoun grigora  
 'The horse that the lions are chasing quickly'

c. Il cavallo che stanno inseguendo i leoni c'.To alogo pou kinigoun ta liontaria 'The horse that are chasing the lions' 'The horse that the lions are chasing'	<b>OO RC Post-S</b>
--	---------------------

Since case on Greek NPs was neutralized through the use of neuter gender, it is only the position of the embedded subject (3b) or number agreement on the embedded verb (3c) that disambiguates between subject and object relative clauses.

*Procedure:* The linguistic materials described above were used for a comprehension task. All participants were tested individually by trained experimenters. The experimental task was preceded by a verb comprehension test aiming at ensuring that participants were able to understand the verbs of the main experiment. The experimental task was also preceded by a training session aiming to familiarize participants with the comprehension of relative clauses. The training session included 3 pictures. In the comprehension experiment an agent selection task was used. The test is an adaptation devised by Adani (in prep) of the De Vincenzi's (1991) test for the comprehension of wh-questions. Participants were presented with pictures including three characters on one sheet of paper and had to point to the one corresponding character. There were 18 pictures in total for the critical items as well as 15 fillers presented in a pseudo-randomized order. An example of a picture is shown in Appendix B. Calculation of accuracy scores included the correct responses, i.e. those responses matching the sentence heard.

*Results:* The children's performance on each sentence type is presented in Table 1. As shown in Table 1, both Greek and Italian children performed better on subject than object relatives, especially object relatives with postverbal subjects.

Table 1. Accuracy in the comprehension of subject and object relatives in Italian and Greek.

Group	OS Mean (SD)	OO Pre-S Mean (SD)	OO Post-S Mean (SD)
Greek children	65.31 (21.65)	52.25 (24.58)	45.97 (24.05)
Italian children	79.99 (19.19)	54.16 (25.86)	40.82 (27.81)

Statistical analysis confirmed these observations. A 2x3 ANOVA (Language X Relative Clause Type (OS, OOPreS, OOPostS)) revealed a significant effect of relative clause type  $F(2,104)=29.849$   $p=.000$ , but no significant effect of language  $F(1,52)=.016$   $p=.9$ . The interaction between Language and Relative clause type was significant  $F(2,104)=4.698$   $p=.011$ . Further analysis using t-tests indicated that Italian children comprehend OS relatives better than their Greek peers [  $t(55)=-2.53$   $p=0.01$ ]. Separate ANOVA showed an effect of type of clause structures (Greek:  $F(2,72)=8.37$ ,  $p<.01$ ; Italian  $F(2,38)=18.63$ ,  $p<.01$ ). Post hoc Scheffè test showed that Greek children performed significantly better on OS than on both OOPreS and OOPostS (OS vs. OOPostS  $p<.01$ ; OS vs. OOPreS  $p<.05$ ). The same holds for Italian children: OS vs. OOPostS  $p<.01$ ; OS vs. OOPreS  $p<.01$ ). Thus, Italian and Greek children showed by and large the same pattern of performance: subject relatives are comprehended better than either types of object relatives, which is presumably due to the fact that in both languages identical morphological cues were available. Thus, when case

information is eliminated, Greek children behave as Italian children. While the first prediction, namely that Italian and Greek children should perform similarly, is borne out, the second prediction is not, although a tendency in the expected direction is observed. No difference was found between object relative clauses disambiguated by syntactic position of the embedded subject or by number agreement. For Italian, this is in contrast with what was found by Arosio et al. In fact, the Italian children in this experiment performed slightly worse than their five-year-old peers in Arosio et al. both on subject relative clauses and on object relative clauses with preverbal subjects (percent correct were about 90% and 70% correct, respectively). The difference may depend on the different methods used in this experiment and in the one carried out by Arosio et al. We used a single picture with three characters involved in two actions simultaneously (two cats chasing a dog that in turn is chasing two cats), while Arosio et al. used two pictures with 2 characters each involved in two separate actions (two cats chasing a dog or a dog chasing two cats). It is possible that in a single picture it is difficult to segregate the two actions. This seems a reasonable explanation given that comprehension of subject relative clauses was also lower, although a difference with object relative clauses was found. This explanation may also be extended to Greek.

In summary, the results of this experiment confirm that when case information is neutralized, Greek speaking children comprehend subject relative better than object relatives and no difference is found between object relatives disambiguated by number agreement in the two languages.

### 3. Experiment 2

The second experiment was conducted only in Greek with the goal of finding out whether object relative clauses disambiguated by number are more difficult to comprehend than object relative clauses disambiguated by case.

*Participants:* Only the 37 Greek-speaking children (M=5, SD=0.25 Range: 4;5-5;6) participated in the second experiment.

*Materials:* We constructed 6 triples of 3 types of clauses each introduced by the lead-in “Show me”, as in (6).

- |  |                     |
|--|---------------------|
| (6) a. ti maimou pou pleni tin arkouda<br>‘The monkey-ACC that is washing the bear-ACC’                                    | <b>OS RC</b>        |
| b. ti maimou pou I arkouda pleni me sampouan<br>‘The monkey-ACC that the bear-NOM is washing with shampoo’                 | <b>OO RC Pre-S</b>  |
| c. ti maimou pou pleni I arkouda<br>‘The monkey-ACC that is washing the bear-NOM’<br>‘The monkey that the bear is washing’ | <b>OO RC Post-S</b> |

Unlike in Experiment 1, NPs displayed unambiguous nominative or accusative case and thus disambiguation of object relatives was brought about by case on the embedded postverbal subject (6c) or by both case and position of the embedded subject (6b). It is expected that object relative clauses disambiguated by Case (and with a postverbal subject) are better understood than object relative clauses disambiguated by number agreement. No difference is expected for object relative

clauses with preverbal subject (whether NPs are case marked as in this experiment or not as in the first experiment).

*Procedure:* The same procedure as in the experiment 1 was followed.

*Results:* The results of the second experiment are presented in Table 2, where we repeat also the results from experiment 1.

Table 2. Accuracy in the comprehension of subject and object relatives in Greek.

Group	OS Mean (SD)	OO Pre-S Mean (SD)	OO Post-S Mean (SD)
Greek children Experiment 1 Number disambiguation	65.31 (21.65)	52.25 (24.58)	45.97 (24.05)
Greek children Experiment 2 CASE disambiguation	72.07 (25.77)	55.4 (25.17)	57.2 (22.4)

We still find an asymmetry between subject and object relative clauses, with subject being comprehended significantly better than both object relative clauses. No difference was found between comprehension of the two object relative clauses. This is confirmed by a one way ANOVA in which the three structures were compared and a significant difference was found,  $F(2, 72)=10,151$ ,  $p < 0,001$ . Post hoc Scheffè test shows that subject relatives are easier than both types of object relative clauses ( $p < 0.01$ ). A further analysis was carried out aiming at comparing the results from experiment 1 and experiment 2. In particular, we wanted to test whether mode of disambiguation influences comprehension of object relative clauses. Thus, we performed several t-tests and found that comprehension of object relative clauses with postverbal subjects disambiguated by number agreement on the verb (5c) is worse than comprehension of object relative clauses disambiguated by case (6c) ( $t(36)=-2.61$ ,  $p = .013$ ) and of object relative clauses disambiguated by position and case (6b) ( $t(36)=-2.05$ ,  $p = .04$ ). No other difference was found, in particular no difference is found between disambiguation by case or by position (and case) (6b vs 6c).

There are three main points that the second experiment revealed. First, the children's performance on object relatives is lower than on subject relatives, as it was in the first experiment. This is in agreement with what is found in other studies and also in adults' sentence processing. Second, the effect of case is significant as shown by the increase of children's performance on object relatives with postverbal subject and disambiguated by case with respect to those disambiguated by number agreement (first experiment). Third, disambiguation by position and case resulted in better comprehension than disambiguation by agreement alone.

## **Discussion**

We discuss our results with respect to the following issues (i) subject-object asymmetry in the acquisition of relative clauses (ii) possible effects of morphological cues (and other cues) on the relative clause comprehension (iii) interplay between cross-linguistic cue exploitation and acquisition of relative clauses.

If we assume that children, as adults, are guided by the Minimal Chain Principle (MCP, De Vincenzi, 1991), then they will postulate a trace in subject position. In other

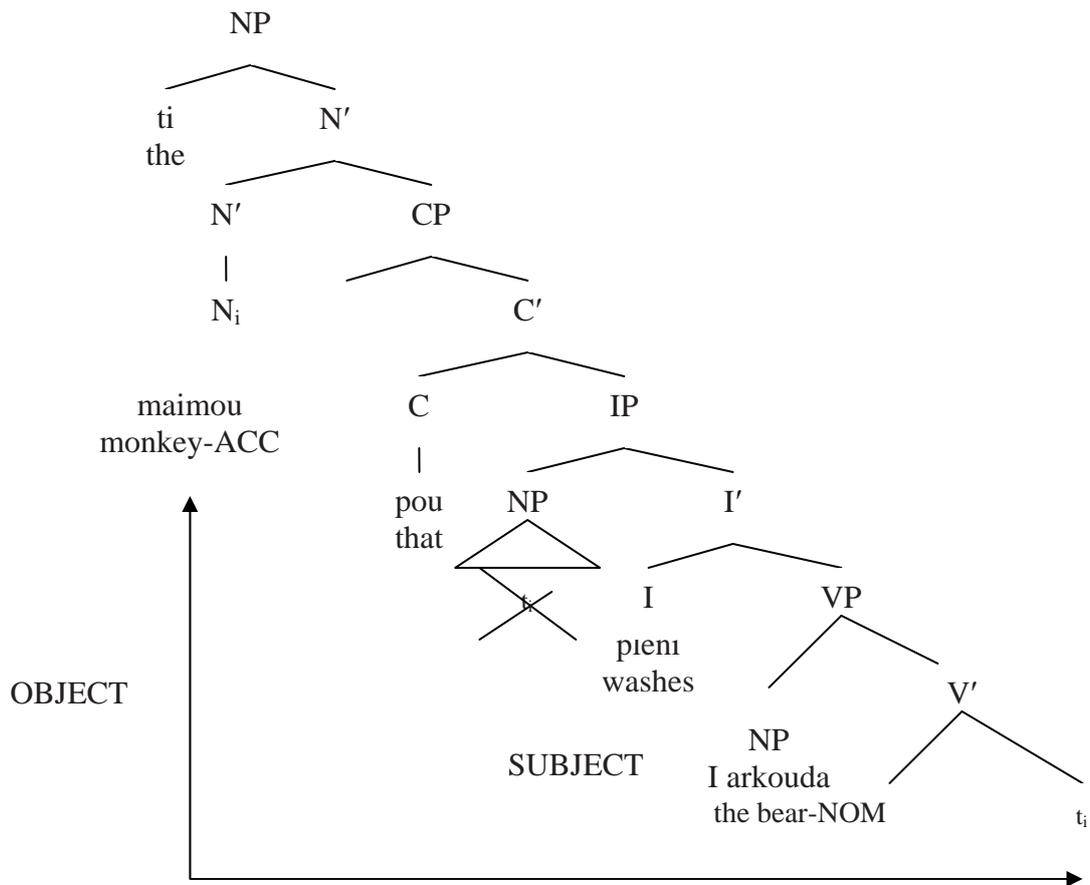
words, the derivation of subject relative clauses is more economic than that of object relative clauses, because the movement in the first case is shorter than in the second (Chomsky, 1995). The results from the Experiment 1 and 2 confirm this dissociation between subject and object relatives in Greek and Italian language acquisition. Recall that due to the use of neuter gender in Greek, Italian and Greek experimental materials had exactly the same morphosyntactic properties. Notice, incidentally, that this result bears on the debate on whether children derive relative clauses by movement (of an empty operator, as in the classical analysis, or of the head, as in Kayne, 1994). If relative clauses were not derived by movement, but by a process of lambda abstraction, as argued by Labelle (1990, 1996), the subject-object asymmetry would be difficult to explain. By contrast, the assumption that they are derived by movement offers an immediate explanation.

Second, results from Experiment 2 indicated that overt morphological case marking contributed significantly to the increase of accuracy scores on the comprehension of object relative clauses with postverbal subject, that is, disambiguation by case on NPs improved children performance with respect to disambiguation by number agreement on verb. This is so, in spite of the fact that disambiguation by case occurs at a later point (at the end of the sentence) than number agreement. This suggests that the particular mode of disambiguation, i.e. case marking, results in a less demanding reanalysis process, likely because the diagnosis process offers clear information about how to repair the structure. This finding is immediately explained within Fodor and Inoue's (2004) Diagnosis model discussed earlier. Assuming that the parser postulates a trace in subject position after the complementizer was heard, it will maintain this analysis until the postverbal nominative NP is encountered. At this point, the subject trace has to be deindexed from the head of the relative, it must be eliminated (or transformed into an expletive *pro*) and the subject function must be assigned to the postverbal NP subject.<sup>2</sup> Thus, one grammatical function was assigned before and a new grammatical function is assigned again. At this point, the parser has to look for a trace for the head of the relative clause. But only one possibility is available, based on the verb's argument structure. As the subject function has already been taken over, the trace will be postulated in the object position, as illustrated in (7).

---

<sup>2</sup> We remain agnostic about whether an expletive *pro* is inserted in preverbal position (Spec IP) and is coindexed with the postverbal subject, as in Rizzi (1982) or if the structure includes only the postverbal subject.

(7)



Reanalysis has been successfully carried out and a new representation for the sentence is built. Thus, overt morphological case marking facilitates children in mapping between arguments and surface syntactic position while learning a language. If we put together the results found in this paper and those found in Arosio et al. (2007) the following picture emerges. When the information delivered by the diagnosis process is part of a process involving checking and assignment of the grammatical function, reanalysis is relatively easy and sentences are relatively well comprehended by children. Let us see why. On the assumption that on the first pass parsing, an analysis is attempted according to the MCP with the assignment of the subject grammatical function to one argument (no other argument is present yet as the verb has not been retrieved). This is illustrated in the first column of table 1. When incoming information is incompatible with the preferred analysis, a diagnosis of the problem and a reanalysis of the structure have to be performed. As seen in the second column of table 1, when after the complementizer an NP is found, the subject grammatical role is reassigned. As the verb has not yet been parsed, no other argument is present at that point. Thus, one argument was assigned and an argument has been reassigned. When the postverbal NP marked nominative is found, i.e., at the end of the clause, the subject function is reassigned. Again, one argument was assigned and an argument is reassigned. In addition, as the verb has already been encountered, also the object role is assigned. In contrast, when the disambiguating number agreement on the verb is encountered, no reassignment of the subject function is possible (and neither of the object function).

Table 3. First pass parsing and reanalysis of relative clauses

Disambiguation cue	First pass	Diagnosis
Position	<a, a: subject	<a, New NP must be subject=a
Case	<a, a: subject	<a,b> Nominative NP must be subject=a And the trace must be object=b
Number agreement 3 <sup>rd</sup> person	<a, a: subject	<a,b> No new subject, no new object

Thus, children, at least at the age of 5 years, can override their preferred analysis, if the diagnosis points to a solution and allows the correction of the illegal grammatical function, as in the first two cases. However, this does not occur in the case of number disambiguation: after the disambiguating information is encountered a grammatical function is not assigned anymore and it is not clear to whom it can be assigned to. In such a situation, children chose not to abandon the preferred analysis and interpret object relative clauses as being subject relative clauses. In other words, children do not engage in an alternative analysis when they have to leave arguments unassigned.

Children, as adults, prefer a subject relative clause, but when prompted by information that is not compatible with this preferred analysis, they abandon it, if the disambiguating information entails a change in the assignment of grammatical functions, but not if these functions remain unassigned, likely because this would require computational resources that exceed children's capacities.

Some problems remain open and deserve further exploration. First, the comprehension scores of Greek relative clauses disambiguated by Case are lower than those found by Stavrakaki (2001) for Greek. This difference may be due to the different methods used. We used pictures, while Stavrakaki used an act out task. Second, the comprehension scores of Italian relative clauses disambiguated by position are lower in this experiment than in the experiment carried out by Arosio et al. (2007). Again different methods were used. Finally, Greek production data seem to indicate that children do not have problems in forming object relative clauses (Stavrakaki 2002, Stathopoulou 2007, Varlokosta 1996). Guasti and Cardinaletti (2003), instead, found that object relative clauses were produced by Italian children (from 5 to 9), but a tendency was observed to transform an object relative clause into a subject relative clause. Nevertheless, children produced both object relative clauses with preverbal subject, as well as object relative clause with postverbal subject, as shown in table 4.

Table 4. Number of object RCs produced by Italian children

Age	Object RCs	
	SV	VS
5	5	4
6	10	8
7	12	17
9	14	9
Adults	5	1

Although an asymmetry exists between subject and object relatives in production, this asymmetry is less evident than in comprehension. This seems to point toward a dissociation between production and comprehension, with the latter being less advanced than the former, a fact that future research will have to address.

## **References**

- Adani, F. (In preparation). The role of grammatical features in the comprehension of TD and SLI children. PH.D. Dissertation, University of Milano-Bicocca.
- Arosio, F., Adani, F. & Guasti, M.T. (2007). Processing grammatical features by Italian children. *Proceedings of GALA 2005*. Cambridge Scholars Press
- Meng, M. & M. Bader (2000). Ungrammaticality detection and garden path strength: Evidence for serial parsing. *Language and Cognitive processes*, 15: 615-666.
- Chomsky, N. (1995) *The Minimalist Program*. Cambridge, Mass: MIT Press
- De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer.
- Fodor, J.D. & Inoue, A. (2000). Syntactic features in reanalysis: Positive and negative symptoms. *Journal of Psycholinguistic Research* 29 (1): 25-36.
- Fraunfelder, U.H., Segui, J., & Mehler, J. (1980). Monitoring around the relative clause. *Journal of Verbal Learning and Verbal Behaviour* 19: 328-337.
- Frazier, L. & Clifton C. J. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4: 93-126.
- Frazier, L. & Flores d'Arcais, G.B. (1989). Filler-driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331-344.
- Guasti, M. T. e A. Cardinaletti (2003) Relative clause formation in Romance child's production. *Probus*. 15:47-88.
- Kayne, R. (1994). *The antisymmetry of syntax*. Cambridge, Mass: MIT Press.
- King, J. & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, Sum Vol 7(3) 376-395.
- Labelle M.(1990). Predication, wh-movement and the development of relative clauses. *Language Acquisition* 1: 95-119.
- Labelle, M.(1996). The acquisition of relative clauses: movement or no movement? *Language Acquisition* 5(2):65-82.
- Meng, M. & Bader, M. (2000). Modes of disambiguation and garden path strength: An investigation of subject-object ambiguities in German. *Language and Speech* 43 (1): 43-74.
- Penolazzi B., De Vincenzi M., Angrilli A. & Job R. (2005). Processing of temporary syntactic ambiguity in Italian "Who" questions: A study with Event-Related potentials. *Neuroscience Letters*: 378.
- Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris.
- Schriefers, H., Friederici, A.D. & Kuehn, K. (1995). The processing of local ambiguous relative clauses in German. *Journal of Memory and Language* 34:499-520.
- Stathopoulou, N. (2007). Producing relative clauses in Greek: Evidence from Down Syndrome. *Essex Graduate Student Paper on Language and Linguistics* 9: 104-125.
- Stavrakaki, S. (2001). Comprehension of reversible relative clauses in Specifically Language Impaired and Normally developing Greek children. *Brain and Language* 77: 419-431.

- Stavrakaki, S. (2002). A-bar movement constructions in Greek children with SLI: Evidence for deficits in the syntactic component of language. In E. Fava (Ed.). *Clinical Linguistics: Theory and applications in speech pathology and therapy. Current Issues in Linguistic Theory*. Amsterdam & Philadelphia: John Benjamins.
- Varlokosta, S. (1997). The acquisition of relative clauses in Modern Greek: Evidence from movement. In A. Sorace, C. Heycock & R. Shillcock (eds). *Proceedings of the GALA 97 conference on Language Acquisition*.

## Appendix A:

### 1<sup>ST</sup> EXPERIMENT (IN GREEK AND ITALIAN)

#### 1. OO (without case, +agreement, post-verbal subject)

Show me:

to alogo pou kinigoun ta liontaria  
the-horse-neuter-that-are chasing-the-lions-neuter  
Il cavallo che i leoni inseguono  
*The horse that the lions are chasing*

To pouli pou kitoun ta gaidouria  
The-bird-neuter-that-are-looking-the-donkeys-neuter  
L'uccellino che stanno guardano gli asini  
*The bird that the donkeys are looking at*

To provato pou travoun ta gaidouria  
The-sheep-neuter-that-are pulling-the-monkeys  
La pecora che le scimmie spingono  
*The sheep that the monkeys are pulling*

To agori pou plenoun ta koritia  
The-boy-neuter-that-are washing-the-girls-neuter  
Il ragazzo che lavano le ragazze  
*The boy that the girls are washing*

To provato pou klotsoun ta aloga  
The-sheep-neuter-that-are kicking-the-horses  
La pecora che calciano i cavalli  
*The sheep that the horses are kicking*

To kouneli pou htipoun ta pontikia  
The-rabbit-neuter-that-are-hitting-the-mice  
Il coniglio che colpiscono i topi  
*The rabbit that the mice are hitting*

*Number and Case in the comprehension of relative clauses*

2. *OO (without case, +agreement, pre-verbal subject (with PP at the end) )*

Show me:

to alogo pou ta liontaria kinigoun me thimo  
the-horse-neuter-that- the-lions-neuter -are chasing-with anger  
il cavallo che i leoni inseguono con rabbia  
The horse that the lions are chasing angrily

To pouli pou ta gaidouria kitoun me iremia  
The-bird-neuter-that- the-donkeys-neuter –are-looking-with-calm  
L'uccellino che gli asini guardano con calma  
*The bird that the donkeys are looking at calmly*

To provato pou ta gaidouria travoun me to shini  
The-sheep-neuter-that-the donkeys-are pulling-with the rope  
La pecora che gli asini tirano con la corda  
*The sheep that the donkeys are pulling with the rope*

To agori pou ta koritsioa plenoun me to sampouan  
The-boy-neuter-that-the girls-neuter-are washing-with-shampoo  
Il ragazzo che le ragazze lavano con lo shampo  
*The boy that the girls are washing with shampoo*

To provato pou ta aloga klotsoun me thimo  
The-sheep-neuter-that-the-horses- are kicking-with anger  
La pecora che i cavalli calciano con rabbia  
*The sheep that the horses are kicking angrily*

To kouneli pou ta pontikia htipoun me ta podia  
The rabbit-neuter-that-the-mice-neuter-are-hitting-with-feet  
Il coniglio che i topi colpiscono coi piedi  
*The rabbit that the mice are kicking*

3. *OS (without case, +agreement)*

Show me:

to alogo pou kiniga ta liontaria  
the-horse-neuter-that-is-chasing-the-lions-neuter  
Il cavallo che insegue i leoni  
*The horse that is chasing the lions*

To pouli pou kita ta gaidouria  
The-bird-neuter-that-is-looking-the-donkeys-neuter  
L'uccellino che guarda gli asini  
*The bird that is looking at the donkeys*

To povato pou trava ta gaidouria  
The-sheep-neuter-that-is-pulling-the-monkeys  
La pecora che spinge le scimmie  
*The sheep that is pulling the donkeys*

To agori pou pleni ta koritsia  
The-boy-neuter-that-is washing-the-girls-neuter  
Il ragazzo che lava le ragazze  
*The boy that is washing the girls*

To provato pou klotsa ta aloga  
The-sheep-neuter-that-is kicking-the-horses-neuter  
La pecora che calcia i cavalli  
*The sheep that is kicking the horses*

To kouneli pou htipa ta pontikia  
The rabbit-neuter-that-is-hitting-the-mice  
Il coniglio che colpisce i topi  
*The rabbit that is hitting the mice*

## 2<sup>ND</sup> EXPERIMENT (ONLY IN GREEK)

### 1. OO (+case, + post-verbal subject)

Show me:

ti maimou pou pleni I arkouda  
the monkey-acc-that-is washing-the bear-nom  
*the monkey that the bear is washing*

ton elefanta pou kiniga I kamila  
the-elephant-acc-that-is chasing-the-camel-nom  
*the elephant that the camel is chasing*

tin agelada pou sprohni o elefantas  
the cow-acc-that-is-pushing-the-elephant-nom  
*the cow that the elephant is pushing*

to rinokero pou htipai I zevra  
the rhino-acc-that-is hitting-the-zebra-nom  
*the rhino that the zebra is hitting*

tin agelada pou kiniga I kamila  
the-cow-acc-that-is chasing the camel  
*the cow that the camel is chasing*

ton pithiko pou trava o rinokeros  
the-monkey-acc-that-is pulling-the-rhino-nom  
*the monkey that the rhino is pulling*

## *Number and Case in the comprehension of relative clauses*

### *2. OO (+case, + pre-verbal subject (with PP at the end))*

Show me:

ti maimou pou I arkouda pleni me sampouan  
the monkey-acc-that-the bear-nom-is washing-with-shampoo  
*the monkey that the bear is washing with shampoo*

ton elefanta pou I kamila kiniga me thimo  
the-elephant-acc-that-the-camel-nom-is chasing with anger  
*the elephant that the camel is chasing angrily*

tin agelada pou o elefantas sprohni me thimo  
the cow-acc-that-the-elephant-nom-is-pushing-with anger  
*the cow that the elephant is pushing angrily*

to rinokero pou I zevra htipai me to podi  
the rhino-acc-that-the zebra-nom-is hitting-with the foot  
*the rhino that the zebra is kicking*

tin agelada pou I kamila kiniga me thimo  
the-cow-acc-that-the-camel-is chasing-with anger  
*the cow that the camel is chasing angrily*

ton pithiko pou trava o rinokeros me shini  
the-monkey-acc-that-the-rhino-nom-is- pulling-with rope  
*the monkey that the rhino is pulling with rope*

### *3. OS +case*

Show me:

ti maimou pou pleni tin arkouda  
the monkey-acc-that-is washing-the bear-acc  
*the monkey that is washing the bear*

ton elefanta pou kiniga tin kamila  
the-elephant-acc-that-is chasing-the-camel-acc  
*the elephant that is chasing the camel*

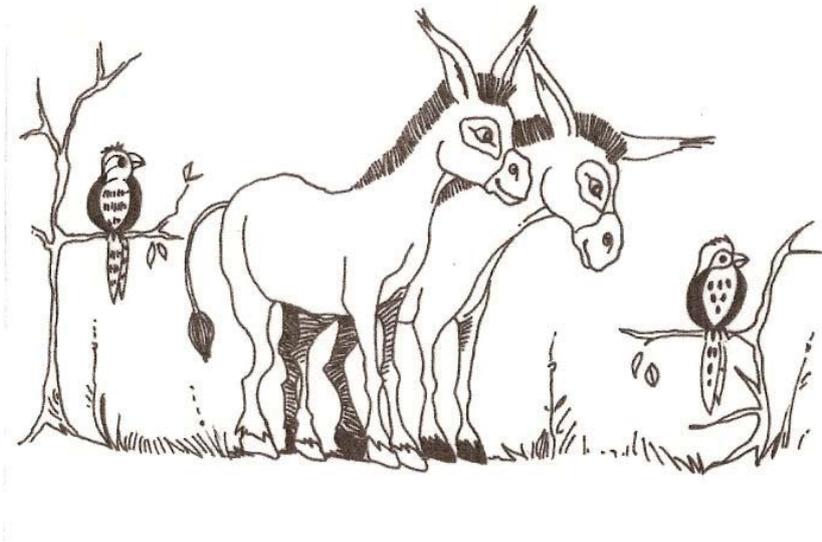
tin agelada pou sprohni ton elefanta  
the cow-acc-that-is-pushing-the-elephant-acc  
*the cow that is pushing the elephant*

to rinokero pou htipai ti zevra  
the rhino-acc-that-is hitting-the-zebra-acc  
*the rhino that is hitting the zebra*

tin agelada pou kiniga tin kamila  
the-cow-acc-that-is chasing the camel-acc  
*the cow that is chasing the camel*

ton pithiko pou trava ton rinokero  
the-monkey-acc-that-is pulling-the-rhino-acc  
*the monkey that is pulling the rhino*

**Appendix B.**



# On the residual omission of Determiners in Italian L1: a case study

IDA FERRARI

University of Florence/Ciscl-University of Siena  
ida\_ferrari@hotmail.com

SIMONA MATTEINI

Ciscl-University of Siena  
si.mat@libero.it

This study deals with the acquisition of functional categories. Specifically, it focuses on determiner omissions in L1, a widespread phenomenon of early grammar which is subject to variation among languages (Chierchia et al., 1999; Guasti et al., 2004). We analyze here the spontaneous production of one Italian monolingual child. On the basis of the distribution of D in this language, we took into consideration semantic/syntactic properties of nominals that are crucial for the licensing of determiners in Italian. Results indicate that, during the period analyzed, D omission is a residual phenomenon more related to the structural configuration in which nominals occur rather than to other properties of Ns. In particular, an interesting selectivity emerges as for the non-target patterns are concerned: items involved in the high DP structures seem to play a crucial role in defining the conditions under which determiners are dropped in L1 grammar.

## 1. Introduction<sup>1</sup>

Previous research on early article omission, based on analysis both of spontaneous production and elicited production, has stressed the crucial role played either by prosodic constraints or syntactic constraints. On one hand, the former group of studies converges in explaining determiner omissions in terms of prosodic constraints on the output of the speech production system (Gerken, 1991 for English L1; Crisma and Tomasutti, 2000 for Italian L1). According to this hypothesis, there is a strict correlation between the prosodic properties of the element preceding or following the determiner and article omissions. On the other hand, the latter group of studies focuses on the correlation between the position occupied by nominals in the sentence and article omissions (Guasti et al., 2004 for Dutch and Italian, Caprin & Yoghà, 2006 for Italian). In particular, the sentence initial position appears to be more sensitive to determiner omission than the sentence internal one. Furthermore, a subject/object asymmetry as for the omission of determiner has been reported; children omit more

---

<sup>1</sup> Previous versions of this work have been presented at Consortium for Linguistics in Taiwan (2007), Galana 3 (2008) and The Romance Turn 3 (2008). We thank the audiences for insightful comments and suggestions. We are particularly indebted to Adriana Belletti, Giuliano Bocci and Luigi Rizzi for discussions on this work. Usual disclaims apply.

determiners in sentence initial subjects rather than in sentence internal objects (Guasti et al., 2004; Caprin & Yoghà, 2006). Moreover, other studies reveal that, once a *First Position Effect* on determiner omissions is excluded, a subject/object asymmetry maintains, but individual variation on the highest rate of omissions in subject/object position is attested (Baauw et al., 2005). Interestingly, these phenomena do not seem to depend on prosodic constraints.

The present research is a contribution to the second group of studies. It aims at reconstructing a fine-grained mapping of syntactic contexts sensitive to D omission, focusing on the spontaneous production of one Italian monolingual child. The main purpose of this study is to investigate the role played by structural configurations in early D omission.

The paper is organized as follows: section 2 contains a brief overview of the morphosyntactic properties of articles in Italian; in section 3, 4 and 5 we present the corpus, the criteria adopted for the identification of the utterances relevant to our analysis and the data collected; section 6, 7, and 8 are devoted to detailed analysis of the different contexts sensitive to determiner omission and to the development of explanatory proposals for non-target patterns. In section 9 we conclude the paper with a general discussion of the findings.

## 2. (Morpho) syntactic properties of the article system in Italian

Italian has a full paradigm<sup>2</sup> of definite and indefinite articles which vary according to gender and number. Furthermore, there are also some allomorphic variants of definite and indefinite masculine articles (*lo, gli, uno*) as well as a reduced form for the definite singular feminine *la* and masculine *lo* (both reduced to *l'*<sup>3</sup>).

With regard to the distribution of determiners, the pattern is quite complex. In general terms, singular count nouns in argumental positions require a determiner:

- (1) Leggo [\* (un) libro]<sub>Obj</sub>  
'I read a book'
- (2) [\* (Il) ragazzo]<sub>Subj</sub> è italiano  
'The boy is Italian'
- (3) Vado in vacanza [PP con [\* (un) amico]<sub>Prepositional Obj</sub>]  
'I go on holiday with a friend of mine'

Bare plurals are allowed as object of a transitive verb (4), and object of a preposition (5). They encode a non-specific (generic) reading:

- (4) Leggo [libri]<sub>Obj</sub>  
'I read books'

<sup>2</sup> See the appendix (Table C and D) for the Italian article paradigm and article choice.

<sup>3</sup> Reduced forms and allomorphs are required in front of nominals beginning with vowels, with clusters of consonant such as, for example, s + consonant or ps, with consonant such as z, x, y:

- (i) l'uomo (the man.ms.sg) vs il libro
- (ii) l'amica (the friend.fm.sg) vs la penna
- (iii) lo/uno specchio (the/a mirror.ms.sg)
- (iv) gli specchi (the mirrors.ms.pl)
- (v) lo xilofono (the xilophone)

- (5) Vado in vacanza [PP con [amici]<sub>Prepositional Obj</sub>]  
'I go on holiday with friends'

Interestingly, the same does not hold true for the preverbal subject position where a bare plural is ungrammatical. In contrast, bare plurals are grammatical as postverbal subjects with, for example, unaccusative verbs (Longobardi 2000). Examples (6) and (7) illustrate the contrast:

- (6) [\***(I)** ragazzi]<sub>Subj</sub> sono italiani  
'The boys are italian'
- (7) Arrivano [(i) ragazzi]<sub>Subj</sub> italiani  
'There arrive Italian boys'

As for mass nouns, they may be licensed as bare nominals in postverbal subject positions (8) and object positions (9 a,b). They receive a non-specific (partitive) reading:

- (8) Viene acqua giù dal tetto  
'The water comes down the roof'
- (9) **a** Bevo [(il) vino]<sub>Obj</sub> tutti i giorni  
'I drink wine every day'
- b** Bevo whiskey [PP con [ghiaccio]<sub>Prepositional Obj</sub>]  
'I drink whiskey with ice''

As for predicative position, only mass nouns and plurals may occur without the determiner <sup>4</sup>:

- (10) Questo è [vino]<sub>Predicative Nominal</sub>  
'This is wine'
- (11) Questi sono [libri]<sub>Predicative Nominal</sub>  
'These are books'

Finally, let us conclude this overview on the morphosyntax of the determiner system in Italian focusing on the distribution of articles with proper names and possessive constructions. While the former is subject to dialectal variation, the latter affects all common nouns (count and mass nouns) preceded by a possessive pronouns, regardless of number and gender. Interestingly, only singular kinship terms introduced by possessive pronouns may be used as bare nominals:

---

<sup>4</sup> Interestingly, once mass nouns and plurals show up with a modifier (AP, sentential modifier, quantifier), the determiner is obligatory:

- (i) **a** Questo è (\*il) mio vino  
'This is my wine'
- b** Questi sono (\*i) libri che vuoi comprare  
'These are the books you want to buy'
- c** Comprò tutto (\*il) vino/ tutti (\*i) libri  
'I buy all wine/all books'

- (12) Cerco la mia borsa.<sub>fm.sg</sub>/i miei libri.<sub>ms.pl</sub>/il mio vino  
‘I am looking for my bag/my books/my wine’
- (13) Vedo tutti i giorni mia sorella / \*(le) mie sorelle  
‘I see my sister/sisters every day’

Summing up, the distribution of determiners in Italian suggests that different properties of the nominals play a role in determining the condition under which articles are obligatory in standard Italian: (i) the distinction singular vs. plural; (ii) the distinction mass vs. count nouns; (iii) the syntactic configuration. In our research we address the question whether and how such properties of nominals may interact in early determiner omission in Italian L1.

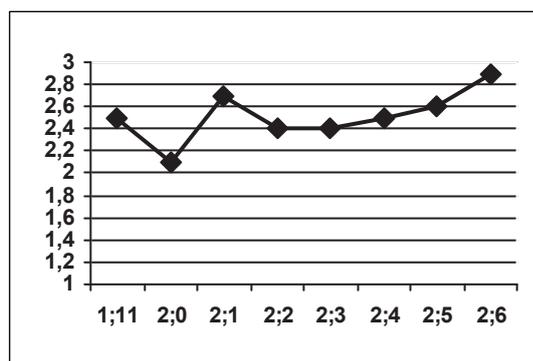
### 3. The Corpus

We base our study on the analysis of an original corpus consisting of 11 recordings of Sabrina, a female Italian monolingual child living in Tuscany (Italy). The corpus was transcribed in CHAT format following the CHILDES criteria and successively double checked. Table 1 and Figure 1 show the MLU and the MLU variation during the period analyzed<sup>5</sup>.

Table 1: MLU

Age	MLU
1;11	2,5
2;0	2,1
2;1	2,7
2;2	2,4
2;3	2,4
2;4	2,5
2;5	2,6
2;6	2,9
<b>Average MLU</b>	2,51

Figure 1: MLU variation



As can be observed, Sabrina’s MLU is high since the first recording. Nevertheless, we identified a difference between a first period (1;11-2;2) during which the MLU considerably fluctuates and the second period (2;3-2;6) during which the MLU variation steadily increases.

### 4. Criteria for the identification of the relevant utterances

We took as ‘determiners’ definite/indefinite articles and their early manifestations as protosyntactic devices (PSD, henceforth). As PSD, we considered the indistinct vocalic morphemes produced by the child in front of nominals which can be taken as morphophonological placeholders according to Bottari et al., (1993/94). On the basis of the morphosyntactic properties outlined in section 2, we considered for our analysis count and mass nouns in argumental/predicative position when they obligatorily

<sup>5</sup> The data have been collected and transcribed by Simona Matteini. They have been further double checked by Valentina Chianciani, Sara Paolucci, and Ida Ferrari.

require a determiner. As for argumental position we included subject/objects of verbs, prepositional objects, nominals produced in isolation as answers to questions about the subject/object of the event. The utterances are exemplified in (14):

- (14) **a** CHI: c' è farfallina.  
there is \_ small butterfly  
'There is a small butterfly.'
- b** CHI: metti a potto libetto?  
put away \_ small book?  
'Can you put away the small book?'
- c** CHI: dai sapone?  
give \_ soap?  
'Can you give me the soap?'
- d** CHI: con matello [martello].  
with \_ hammer  
'with the hammer'
- e** INV: chi arriva?  
'Who is coming?'
- CHI: principe  
\_ prince  
'the prince.'
- f** INV: allora # che disegniamo?  
'What should we draw now?'
- CHI: pinguino.  
\_ penguin  
'A penguin.'

As for predicative position we included singular count nouns in copular constructions:

- (15) CHI: questa è treno.  
this is \_ train  
'This is the train.'
- (16) INV: guarda un po' # chi è questo qui?  
'Look ! Who is this?'
- CHI: drago.  
\_ dragon  
'A dragon.'

Crucially, we included in our analysis also mass noun with a 'specific' reading, as they require a determiner in Italian, as, for example, in possessive constructions:

- (17) CHI: Questo è il latte mio!  
This is the milk my  
'This is my milk!'

We excluded all the contexts not requiring a determiner, such as (i) mass nouns and bare plurals with a 'non specific' reading in argumental or predicative position; (ii) proper names/kinship terms, since they do not require a determiner in the variety of Italian spoken by the child<sup>6</sup>, (iii) all combinations of nominals and prepositions not requiring a determiner in Italian such as, for example, *andare a casa* (to go home). Relevant examples are given in (18) and (19).

- (18) a CHI: questo è ciaccino!  
'This is bun'
- b CHI: vuole mangiare sassolini.  
want.<sub>3.prs.sing</sub> to eat pebbles  
'He wants to eat small pebbles'
- c CHI: quelle so' [=? sono] candele.  
'Those are candles!'
- (19) a CHI: dov'è mamma?  
'Where is mummy?'
- b CHI: dov'è zia Simona?  
'Where is aunt Simona'

Finally, we excluded: (a) idiomatic expressions and routine sentences containing a nominal; (b) unclear sentences, (c) immediately adjacent complete repetitions of the child's own utterances, (d) corrected initial errors.

## 5. The data

On the basis of such criteria, we isolated 661 contexts which required a determiner. On these utterances, the rate of D omissions/occurrences/PSD were calculated. Table 2 and Figure 2 illustrate the pattern we observed:

---

<sup>6</sup> The variety of Italian spoken in Siena (Tuscany). Interestingly, the use of expletive determiners with proper names (and kinship terms) is subject to a high degree of variation among the varieties of Italian spoken in Tuscany as well. Let us take as a case point the contrast between, for example, *Senese* and *Fiorentino*. Proper names and kinship terms are always introduced by a definite article in the latter but not in the former.

Table 2: D occurrences/omissions/PSD

Age	Occ	Omiss	PSD
1;11	(49/101) 48%	(27/101) 27%	(25/101) 25%
2;0	(23/49) 47%	(17/49) 35%	(9/49) 18%
2;1	(76/114) 67%	(35/114) 31%	(3/114) 3%
2;2	(55/81) 68%	(24/81) 30%	(2/81) 2%
2;3	(69/95) 73%	(17/95) 18%	(9/95) 9%
2;4	(82/108) 76%	(24/108) 22%	(2/108) 2%
2;5	(44/55) 80%	(8/55) 15%	(3/55) 5%
2;6	(47/58) 81%	(11/58) 19%	(0/58) 0%
<b>Total</b>	<b>(445/661) 67%</b>	<b>(163/661) 25%</b>	<b>(53/661) 8%</b>

Figure 2: D occurrences/omissions/PSD

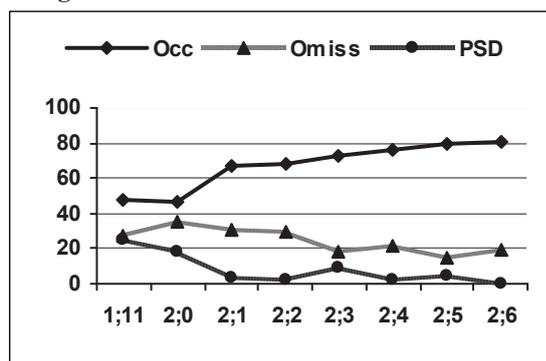
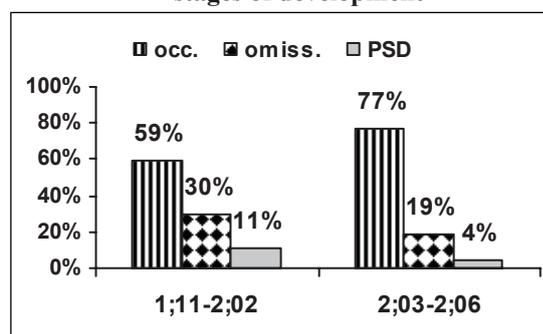


Figure 3: D occurrences/omissions/PSD: stages of development



Overall, determiner omission ranges between a highest rate of 35% (2;0) and a lowest rate 15% (2;5). Hence, this fact allows us to infer that D omission is a residual phenomenon which follows a developmental path during the period analyzed. As suggested by Figure 2, production of PSD and D omission are quite a noticeable phenomena in the first two recordings. Starting from (2;1) the former strategy drastically decreases, while the latter option decreases steadily. As for D occurrences, they increase gradually from 48% (1;11) to 81% (2;6). Considering the decrease of D omission, we identified two stages of development: in the first stage (1;11-2;02) the average rate of D omission is about 30% whereas from 2;03 to 2;06 the average rate of D omission is attested at about 19% (Figure 3).

## 6. Sensitive Contexts for D Omission

In order to identify to which aspect the omission of D may be related, we verified the correlation between the omission of determiners and the following properties of nominals: gender, number, mass distinction. Furthermore, we also considered the position occupied by nominals in the sentence and their functions. Attention has also been paid to the configurations in which nominals are modified by functional/lexical elements (i.e. possessive pronouns and the quantifier *tutto*).

### 6.1 D omission with [+/- Mass] DPs

As for the distinction [+/- Mass], the percentage of article omission with [+ Mass] nominals (22% - 13/58) is similar to the one of [-Mass] nominals (24% - 145/603).

The difference between the two groups is not statistically significant (Fisher's E. P-Value = 0,87)<sup>7</sup>.

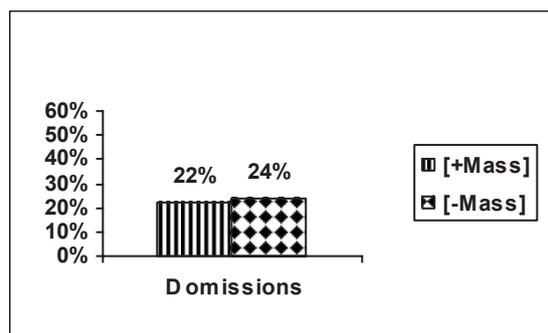
Table 3 and Figure 4 illustrate the low percentage of D omission with respect to D occurrences in both groups of nominals. This fact suggests that, at this stage of acquisition, the [+/- Mass] distinction does not seem to play a crucial role in the phenomenon under investigation. In fact, the child seems to master that bare mass nouns are grammatical only in specific contexts in Italian, such as copular constructions and object position with a non specific reading (*E' brodino!* – This is broth!; *Mangio pollo.* – I eat chicken.).

**Table 3: D omission according to the distinction [+/- Mass]**

	D Omissions	D Occurrences
<b>+ Mass</b>	(13/58) <b>22%</b>	(45/58) <b>78%</b>
<b>- Mass</b>	(145/603) <b>24%</b>	(458/603) <b>76%</b>
<b>p = 0,87</b>		

(20)a CHI: c'ho moccico.	<b>[+ Mass]</b>
'I have snot.'	
b CHI: pulisci pavimento!	<b>[- Mass]</b>
'Clean the floor!'	

**Figure 4: D omission according to the distinction [+/- Mass]**



### 6.2 D omission with [+/- Singular] DPs

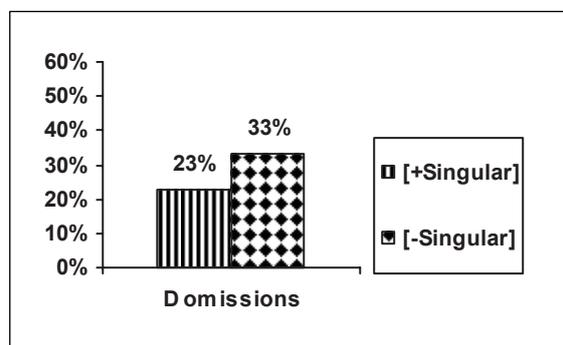
Turning to the [+/-Singular] distinction, Table 4 and Figure 5 show a slightly higher percentage of D omissions within [- Singular] contexts. Determiner omission is attested at 33% (27/83) with plural nominals and at 23% (132/578) with singular ones. This difference is statistically near to the significant threshold: (Fisher's E. P-Value = 0,07).

**Table 4: D omission according to the distinction [+/- Singular]**

	D Omissions	D Occurrences
<b>+ Singular</b>	(132/578) <b>23%</b>	(446/578) <b>77%</b>
<b>- Singular</b>	(27/83) <b>33%</b>	(56/83) <b>67%</b>
<b>p = 0,07</b>		

(21)a CHI: guarda squalo!	<b>[+Sing.]</b>
'Look at the shark!'	
b CHI: pulisci tende!	<b>[- Sing.]</b>
'Clean the curtains!'	

**Figure 5: D omission according to the distinction [+/- Singular]**



Contrary to the tendency emerged within [+/- Mass] nominals, the [+/- Singular] distinction seems to play a role in early determiner omission. However, a further

<sup>7</sup> All the data were statistically analyzed using Fisher test. The significance threshold is 0,05.

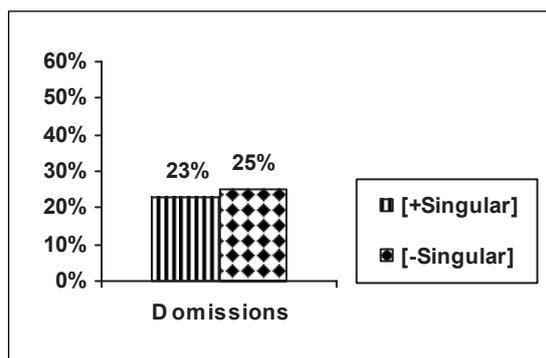
analysis of the data reveals that the rate of D omissions in plural contexts is higher only in the configuration in which the nominals are introduced by the quantifier *tutti/e* as in “*Ho macchiato tutte \*(le) paperine*” – I soiled all ducks. [QP+D+N: + Sing 25% (1/4) vs – Sing. 90% (9/10)]<sup>8</sup>.

Once the nominals introduced by QPs are excluded from the count, the difference between the rate of D omissions with singular nominals (23% - 131/574) and the one with plurals (25% 18/73) is not statistically significant ( $p=0,76$ ) as illustrated in table 5 and Figure 6.

**Table 5: D omissions and occurrences according to +/-Singular distinction of Ns, excluding QPs.**

	D Omissions	D Occurrences
+ Singular	(131/574) 23%	(443/574) 77%
- Singular	(18/73) 25%	(55/73) 75%
$p = 0,76$		

**Figure 6: D omission according to the distinction [+/- singular] excluding QPs.**



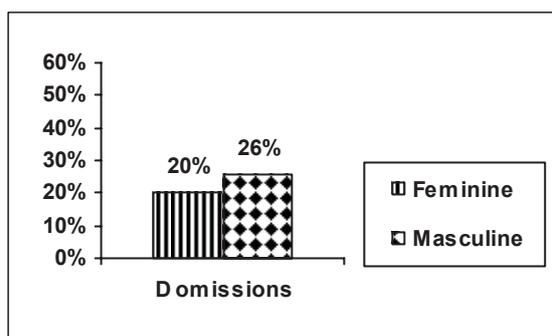
### 6.3 D omission with Feminine/Masculine DPs

As for gender, we observe that the Determiner is omitted 26% (100/381) with masculine nouns and 20% (56/280) with feminine ones. This difference is near to the significant threshold: Fisher's E. P-Value = 0,06.

**Table 6: D omission according to Gender distinction**

	D Omissions	D Occurrences
Feminine	(56/280) 20%	(224/280) 80%
Masculine	(100/381) 26%	(281/381) 74%
$p = 0,06$		

**Figure 7: D omission according to Gender distinction**



(22)a CHI: aspetta sposa. [Fem.] 'He is waiting for the bride.'
b CHI: dov'è tappeto? [Masc.] 'Where is the carpet?'

This result is not unexpected. Taking into consideration the Italian article paradigm, two facts may account for this finding: (i) only the masculine article shows allomorphic variants in Italian; (ii) the definite masculine singular article *il* and the definite masculine plural article *gli* are more complex from a phonetic and a

<sup>8</sup> This fact is particularly interesting in that suggests how article omission may correlate more on DP placement rather than on intrinsic properties of nominals (i.e. number).

phonological point of view. Hence, the child may be induced to omit determiners more frequently in the former context but not in the latter.

#### 6.4 DPs placement and D omission

This section deals with the correlation between article omissions and the position occupied by DPs in the sentence. Specifically, four contexts were analyzed: (i) DP-V; (ii) V-DP; (iii) P-DP<sup>9</sup>; (iv) DP in isolation.

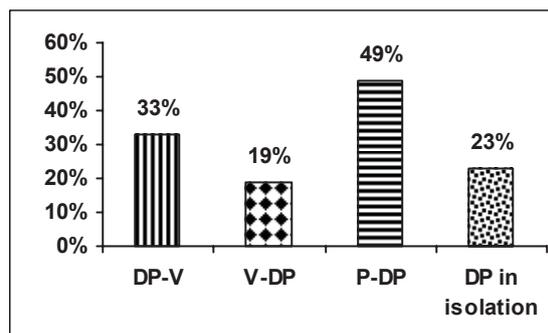
As for (i), we considered all DPs preceding a verb. In this pattern we included preverbal subjects, preposed objects and preposed predicative nominals of copular constructions<sup>10</sup>. As for (ii), we included all DPs following a verb as post verbal subjects, post verbal objects and post verbal predicative nominals. In (iii) we considered all DPs following a monosyllabic preposition, also when the latter has been omitted. Finally, the pattern in (iv) includes: (a) subject and object DPs uttered in isolation as answers to questions about the subject/object of the action; (b) predicative nominals uttered in isolation as answer to questions.

As it emerges from the data analysis reported in the Table 7 and Figure 8 below, the position occupied by nominals in the sentence seems to play a crucial role in D omission.

Table 7: D omission and position of DP

	D Omissions	D Occurrences
<b>DP-V</b>	(3/9) <b>33%</b>	(6/9) <b>66%</b>
<b>V-DP</b>	(67/356) <b>19%</b>	(289/356) <b>81%</b>
<b>P-DP</b>	(42/85) <b>49%</b>	(43/85) <b>51%</b>
<b>DP in isolation</b>	(49/211) <b>23%</b>	(162/211) <b>77%</b>

Figure 8: D omission and position of DP



Indeed, the highest rates of D omission are attested in the contexts P-DP 49% (42/85) and DP-V 33% (3/9)<sup>11</sup>, whereas, it is considerably lower in the contexts V-DP 19% (67/356), and DP in isolation 23% (49/211). The data show that the most sensitive pattern to D omissions is the prepositional context. A significant difference emerges comparing the P-DP values and those of V-DP and DP in isolation respectively (Fisher's E. P-Value < 0,0001 in both cases). On the contrary, we do not observe a statistically significant difference comparing the V-DP values with the DP in isolation values (Fisher's E. P-Value = 0,23).

#### 6.5 DPs functions and D omission

Before focusing on D omissions in prepositional contexts, we analyzed the data on the basis of the function that the DP has in the sentence. Previous studies on this topic have stressed that, besides a first position effect, a subject/object asymmetry is found in children's production data (Baauw et al., 2005). Nevertheless, these studies do not

<sup>9</sup> It has been pointed out that P-DP cannot be considered merely a 'position' as PP can occur preverbally, post-verbally or in isolation. Moreover, P can either be selected by nouns and verbs. Our purpose here was to stress the fact that DP<sub>s</sub> occur after a preposition.

<sup>10</sup> Although ungrammatical in standard Italian, this construction is occasionally produced by the child.

<sup>11</sup> Due to the few occurrences, this pattern will be neither further discussed nor statically analyzed.

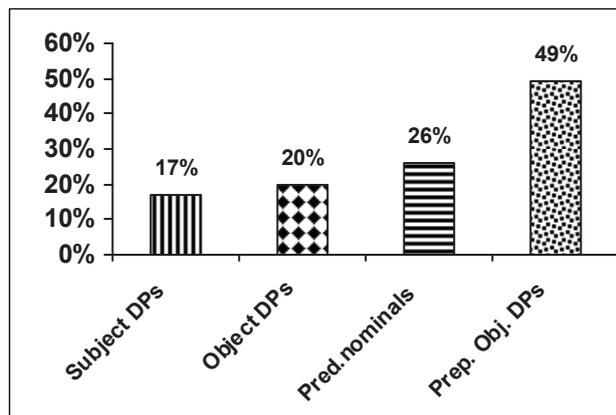
converge in indicating a preference for D omissions in subject or object position. Baauw et al., 2005 found individual variation in Dutch speaking children; Schoenenberger et al., 1997 found a preference for D omission in object position in German children; Caprin & Yoghà, 2006 report a preference of D omission in subject position for Italian speaking children<sup>12</sup>.

In analyzing Sabrina's corpus, we consider the following function: (i) subject, (ii) object, (iii) object of preposition, (iv) predicative nominals. Table 8 and Figure 9 show the results of this analysis:

**Table 8: D omission according to the DP function**

Subject DPs	
D Omissions	D Occurrences
18/108	90/108
17%	83%
Object DPs	
D Omissions	D Occurrences
65/327	262/327
20%	80%
Predicative nominals	
D Omissions	D Occurrences
37/141	104/141
26%	74%
Prepositional Object DPs	
D Omissions	D Occurrences
42/85	43/85
49%	51%

**Figure 9: D omission according to the DP function**



As for D omission, no subject/object asymmetry is found in Sabrina's corpus (D omissions in Subject DPs 17% vs Object DPs 20% - Fisher's Exact P-Value: 0,57). A slightly higher tendency to omit determiners in predicative position, though not statistically significant, emerges in the data analysis (D omission in Pred. DPs 26% vs Subject DPs 17% - Fisher's Exact P-Value: 0,08; D omission in Pred. DPs 26% vs Object DPs 20% - Fisher's Exact P-Value: 0,14). Crucially, the highest rate of omissions in prepositional contexts still remains (49%). A significant difference emerges comparing the values of Prepositional Object DPs with the one of Subject/Object DPs (Fisher's Exact P-Value: <0,0001 in both cases) and with the one of Pred DPs (Fisher's Exact P-Value: 0,0005).

The most striking result prompted by these data, is that the residual phenomenon of D omission seems to be related more to the position occupied by DPs, rather than to other properties of nominals (gender, number, [+/- Mass] distinction)<sup>13</sup>. In particular,

<sup>12</sup> Due to their experimental design, only preverbal subject and post verbal object have been taken in consideration by Caprin and Yoghà, 2006.

<sup>13</sup> The influence of linear order on D omissions has already been reported in the literature (Guasti et al., 2004 for Dutch and Italian, Caprin & Yoghà, 2006 for Italian). Results of these studies converge in indicating the sentence initial position as the most sensitive to D omission. It is worth noticing that the child under consideration produced very few DPs in sentence initial position to have reliable

this study identifies PPs structures as a source of difficulty in determiner provision by the child. This aspect has received little attention in the literature on this topic. In fact, a few studies on language acquisition have focused on D omission in PP contexts. As for Italian L1, Antelmi (1997) has observed that determiners were often omitted when nominals are introduced by a preposition. However, the author does not provide quantitative analysis of the phenomenon. In Leonini (2006), the same tendency has been observed in the acquisition of Italian L2 by German learners (both in a elicited task and in spontaneous production). In the following two sections we will focus on determiner omissions in prepositional contexts and we will formulate some explanatory proposals for the non-target patterns produced by the child.

### 7. Focus on prepositional contexts

This section is devoted to the analysis of the child's production of prepositional phrases requiring a determiner. Only monosyllabic prepositions were considered<sup>14</sup>: 85 PPs obligatorily requiring a determiner out of 226 PPs were identified<sup>15</sup>.

Focusing on these contexts, it emerges that, besides the target form [P+D+N]<sup>16</sup>, three non-target patterns are produced by the child: (a) both the preposition and the determiner are omitted \*[\_P\_D+N]; (b) only the determiner is omitted \*[P\_D+N]; (c) only the preposition is omitted \*[P+D+N]. The relevant patterns are exemplified from (23) to (26):

- |      |   |                     |
|------|---|---------------------|
| (23) | CHI: nella foretta!<br>'in the forest'                            | Target form [P+D+N] |
| (24) | CHI: mette cassetino!<br>put _ drawer<br>'Put it into the drawer' | *[_P_D + N]         |
| (25) | CHI: con principe.<br>with _ prince<br>'with the prince'          | *[P_D+N]            |

---

quantitative and qualitative analysis on this topic. Moreover, none of the previous studies consider D omissions in P contexts separately.

<sup>14</sup> See the appendix (Table E) for Italian monosyllabic prepositions and their syncretic articulated forms.

<sup>15</sup> Considering PPs in general, it emerges that monosyllabic prepositions are attested in the Sabrina Corpus from the first recording (1;11) Specifically, *a*, *di*, *in*, *con*, are used more frequently than *per*, *su* and *da*. Prepositions *fra/tra* are never found in the corpus. A few cases of wrong selections of prepositions are found, as in the following example:

- (i)INV: di chi hai paura?  
      'Who are you afraid of?'  
      CHI: con matigna  
          with \_ stepmother  
          'Of the stepmother'

Moreover, non target possessive constructions are produced by the child (see section 8.2 for a detailed analysis of this pattern and the appendix for a quantitative analysis of the prepositions produced by the child.)

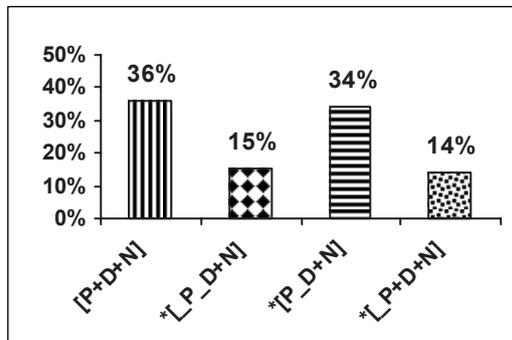
<sup>16</sup> With the the notation [P+D+N] we refer to articulated prepositions produced by the child as: (i) syncretic forms (*nella foresta* – *in the forest*); (ii) non syncretic forms with definite articles (*con la matrigna* – *with the stepmother*) or indefinite ones (*per un bambino* – *for a child*).

- (26) CHI: paura la matigna! \*[\_P+ D+N]  
 fear \_ the stepmother  
 'I am afraid of the stepmother'

Table 9: Different Patterns in P-DP contexts

[P+D+N]	*[_P_D+N]	*[P_D+N]	*[_P+D+N]
(31/85)	(13/85)	(29/85)	(12/85)
36%	15%	34%	14%

Figure 10: Different Patterns in P-DP contexts



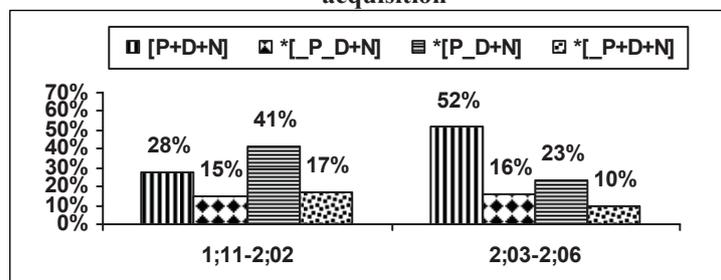
As exemplified in Table 9 and in Figure 10, [P + D] contexts are rather problematic for the child as far as D insertion is concerned. Thus the child resorts to the non-target pattern \*[P\_D+N] to a greater extent (34%).

The non-target patterns involving dropping of the preposition \*[\_P+D+N] or dropping of both the preposition and the article \*[\_P\_D+N] are less attested. Moreover, a developmental path as for the pattern \*[P\_D+N] is attested. In fact, if we consider the two stages of acquisition outlined in section 5, the data show an opposite tendency between the target form [P+D+N] and the \*[P\_D+N] pattern. The former is attested at 28% in the first period and at 52% in the second one. The latter decreases from 41% in the first stage to 23% in the second one. Table 10 and Figure 11 exemplify such contrast.

Table 10: P-DP contexts – Developmental path across two stages of acquisition

	[P+D+N]	*[_P_D+N]	*[P_D+N]	*[_P+D+N]
1;11-2;02	(15/54) 28%	(8/54) 15%	(22/54) 41%	(9/54) 17%
2;03-2;06	(16/31) 52%	(5/31) 16%	(7/31) 23%	(9/31) 10%

Figure 11: P-DP contexts – Developmental path across two stages of acquisition



### 7.1 PPs context as a vulnerable domain for D insertion: proposals

The aim of this section is to provide an explanatory proposal for the non-target pattern emerged within the PP contexts.

Our analysis is based on the assumption that the articulated preposition in Italian undergoes a process of head incorporation, as assumed by Granfeldt (2003) for French, and Giusti (2003) for Italian. According to this view, the determiner in D° adjoins to the head of the PP above the DP projection, as exemplified in (27):



Turning to the child performance observed in this study, we propose that two interacting factors may contribute to make the PPs context a vulnerable domain for determiner omission: (i) articulated prepositions show a syncretic form in Italian; (ii) D and P are in a local configuration and both provide functional structure to the NP.

Let us now focus on the two hypotheses more in detail;

(i) It might be hypothesized that, through D omission, the child is avoiding the extra complex syntactic process at work in the derivation of articulated prepositions in Italian.

Evidence in favour of this hypothesis comes from the fact that determiners are supplied in contexts not requiring a process of head incorporation in Italian, like, for example, preposition with indefinite articles:

- (28) CHI: *stata a uno ballo*  
 been at a dance  
 ‘She went to a dance’

Moreover, the definite article is not omitted when combining with the preposition *per* that does not take a syncretic form in Italian:

- (29) CHI: *pe la mamma.*  
 ‘For the mother’

(ii) Regardless of the head incorporation process, D omission is favoured by the nature of the two heads involved in this configuration. Both P and D are heads of the NP functional extended projection and they are both involved in the NP case assignment. Hence, the child may be induced to omit one in order to prevent overburdening structures for the still immature computational and performance system. The omission of D over P might be preferred in order to avoid a loss of interpretability of the entire PP.

Following Giusti’s (1993, 2003) analysis for Rumanian<sup>17</sup>, we suppose that in \*[P\_D+N] pattern, P is presumably inserted by the child in the highest head (F<sup>max</sup>) of

<sup>17</sup> As reported by Giusti (1993), in Rumanian the enclitic article is ungrammatical with unmodified nominals object of prepositions as exemplified in (i):

- (i) M’am adus la profesor(\*ul)  
 I am gone to professor (\*the)  
 ‘I have been to the professor’

An exception to this pattern is represented by the preposition *cu* (with).

the extended nominal projection, the position usually occupied by the determiner as Case marker as exemplified in (30).

(30) [FP1[CON][NP principe]]

Following this proposal, it might be hypothesized that P and D are in complementary distribution in Sabrina's early grammar.

The two analyses just sketched may be **strictly interrelated**. Further investigation on languages in which prepositions and articles do not show a syncretic form may shed light on the matter. In particular, they may clarify whether children tend to avoid the complexity of head incorporation or, when facing a configuration in which two functional heads in a local relation share similar properties, produce only one for economy reason.

### 8. D omission in other contexts

In the previous section we have analyzed determiner omissions in prepositional contexts. Such a configuration is not the only one involved by this phenomenon in Sabrina's corpus. Interestingly, we found a tendency to omit determiners with the quantifier *tutti/e* and also in possessive constructions. Both configurations require obligatorily an article in standard Italian. Although occurrences of this kind are very few in the corpus, an analysis of the non-target patterns produced in these syntactic domains gives cues on the strategies adopted by the child when new elements are introduced in the DP structure.

#### 8.1 Quantifier *tutti/e*

As for D omission with the quantifier *tutti/e*, we observed an opposite tendency between singular and plural contexts. D omission is attested at a high rate in the latter but not in the former. Such a contrast is exemplified in table (11).

**Table 11: D occurrences/omissions with QPs**

	D Omissions	D Occurrences
<b>Singular QPs</b>	(1/4) <b>25%</b>	(3/4) <b>75%</b>
<b>Plural QPs</b>	(9/10) <b>90%</b>	(1/10) <b>10%</b>

(31) **a** CHI: tutta la torre di Mangiafuoco.  
'Mangiafuoco's whole tower'

**b** CHI: Ho macchiato tutte paperine.  
'I soiled all ducks'

Crucially, nominals introduced by *tutto/tutti* obligatorily require a determiner in Italian regardless of number distinction. The relevant examples are given in (32)a vs b:

(32) **a** Tutta \*(la) famiglia di Maria.  
'Maria's whole family'

---

The enclitic article has to be morphologically realized when the object of preposition is modified by an adjective or by a complement as in (ii):

(ii) M'am adus la profesur \*(ul) tau  
I am gone to professor \*(the) your  
'I have been to your professor'

- b** Tutte \*(le) mele.  
‘all apples’

The asymmetry in (31) recalls the contrast between the ‘definite’/‘indefinite’ quantifier as for the property of selecting a full DP in Italian. Such a property affects only the former but not the latter. Let us focus on the contrast between the ‘definite’ quantifier *tutti* (all) and the indefinite quantifier *molti* (many) in Italian:

- (33) **a** Ho letto tutti \*(i) libri  
‘I read all books’  
**b** Ho letto molti \*(i) libri  
‘I read many books’

As proposed by Giusti (1993) and Giusti & Leko (2001), we assume that both quantifiers in (33)a-b are heads of the functional nominal projection. In (33)a *tutti* selects a full DP while in (33)b *molti* does not require an overt D<sup>18</sup>:

- (34) **a** [Q tutti [D i [F..[NP libri]...]]]  
**b** [Q molti [D  $\emptyset$  [F..[NP libri]...]]]

Child’s tendency to omit D in plural QPs may suggest that she is analyzing *tutti* as an indefinite quantifier which does not require an overt D, as in (34)b<sup>19</sup>.

### 8.2 D omission with Possessive DPs

In this section we deal with D omission and Possessive DPs<sup>20</sup>. As for this particular syntactic domain, we found that the child resorts to the omission of determiners only

<sup>18</sup> The same holds true also for the indefinite quantifiers ‘many’ in English:

(i) I read many (\*the) books.

<sup>19</sup> The interesting question that arises is whether the child is assigning a ‘generic’ reading to the QP *tutte paperine* as in the English example in (i) vs (ii) containing the quantifier ‘all’:

(i) I soiled all ducks<sub>[gen]</sub>

(ii) I soiled all the ducks<sub>[spec]</sub> I have

Unfortunately the few occurrences of QPs of this kind in our corpus prevent us from drawing any conclusion.

<sup>20</sup> Other non-target patterns produced by the child in possessive DPs are concerned with (a) the linear order possessor-noun; (b) the omission/replacement of preposition *di*. As for (a), utterances in (i)-(ii) illustrate the non-target patterns:

(i) INV: Di chi hai paura, Sabrina?  
‘Who are you afraid of?’

(ii) INV: non ti capisco!  
‘I do not understand you’!

CHI: i Cenerentola cappetta!  
of Cinderella shoe  
‘of Cinderella’s shoe’

CHI: ho paura i Cenerentola e cappetta!  
Have<sub>1,ps.sg</sub> fair of Cinderella the shoe  
‘I am afraid of Cinderella’s shoe’

In all cases the possessor precedes the head noun. The construction displays the linear order Poss-N rather than the Italian linear order N-Poss required in Italian possessive constructions containing a non-pronominal possessor. Interestingly, the utterance in (i) and (ii) mirrors the linear order of Germanic possessive construction of the Saxon Genitive-type, where non pronominal possessors show up in

when the possessor occurs prenominally<sup>21</sup>. An example of non-target pattern is given in (35):

- (35) CHI: Dov'è mia penna?  
 'Where's my pen?'

Interestingly, determiner omission is excluded when the possessor is in postnominal position.

Table (12) and examples (36)-(37) show the contrast between insertion/dropping of determiners according to the position of the possessor:

**Table (12): D omission with Possessive DPs**

	D+Poss+N	*[_D+Poss+N]	D+N+Poss	*[_D+N+Poss]
<b>Total</b>	0/3	3/3	6/6	0/6
<b>%</b>	0%	100%	100%	0%

- (36) CHI: oggi era mio compleanno!  
 'Today it was my birthday'

- (37) CHI: dov'è il telefono mio?  
 'Where's my phone?'

The pattern in (36)-(37) suggests that prenominal possessors and determiners are in complementary distribution in Sabrina's early grammar. Such a possibility is subject to variation among languages. In German and English, for example, prenominal possessors do not co-occur with determiners, as in (38)a and b:

- (38) **a** Das ist (\*das) mein Buch  
**b** This is (\* the) my book

On the contrary, in Spanish, only prenominal possessors are in complementary distribution with determiners, as illustrated by the contrast in (39)a vs b:

---

prenominal position. Such option is allowed in Italian only with pronominal possessors (*La sua macchina* vs *\*la di Gianni macchina*).

Concerning (b), we have observed that preposition *di* is sometimes omitted or replaced by the preposition *a* in front of the possessor.

- (iii) CHI: l'albero Babbo Natale  
 the tree \_Santa Claus  
 'Christmas tree'

- (iv) CHI: il gatto Astasia  
 the cat \_Anastasia  
 'Anastasia's cat'

- (v) CHI: c'ho i chiavi a Picasso  
 I have got the keys to Picasso  
 'I have got the keys of the Picasso car'

<sup>21</sup> Similar findings have been reported by Bernardini Roest (2003) in bilingual Italian-Swedish and Italian L1 acquisition

- (39) **a** mi libro (Cardinaletti 1998)  
 ‘my book’

**b** el/este libro mío

According to Cardinaletti (1998), in determiner-less possessive constructions as in (38)a and b and (39)a, an empty D hosts the raised possessive element (a clitic possessive in Cardinaletti’s terms). The derivation in (40) illustrates this fact:

- (40) [D possessive clitic<sub>S<sub>i</sub></sub> [...[<sub>SpecNP</sub> t<sub>j</sub>[<sub>N</sub> N]]]]
- 

Following Cardinaletti’s proposal, we assume that in the ill-formed construction \**Oggi era mio compleanno* (Today it was my birthday) the child is probably adopting the option in (40)<sup>22</sup>. Although this possibility is restricted only to prenominal possessors occurring with singular kinship terms in standard Italian (*Questa è mia madre* – This is my mother), the child seems to extend this option to all DPs with a prenominal possessor. Moreover, the asymmetry in (36)-(37) points out that determiner insertion may be considered by the child a ‘last resort option’ triggered by the necessity to license a full DP when the possessor does not move to D°.

## 9. Conclusion

The results of this study indicate that, during the period analyzed, the omission of D is a residual phenomenon more related to the structural configurations in which the nominals occur than to other factors (i.e. gender, number, +/- mass distinction of nominals). Items involved in the high DP structure (specifically prepositions and prenominal possessors) seem to play a crucial role in determining ‘when’ and ‘how’ D omission applies.

Moreover, results of this study have identified P-DP configurations as one of the most sensitive to D omission in child’s early grammar. A further study on corpora of other Italian monolingual children would be desirable in order to verify whether the phenomenon is attested or other strategies are used. Furthermore, a comparison with corpora from other languages in which prepositions and articles do not show a syncretic form may help in defining possible explanatory proposals for this finding.

<sup>22</sup> This assumption predicts that the child should recognize the ‘functional’ status of possessors with respect to other prenominal modifiers, which do not have to appear in complementary distribution with determiners. In our corpus we only found a few occurrences of complex DPs containing a prenominal modifier other than possessives:

- (i) Un’ atta farfallina  
 a.fm.sg other.fm.sg butterfly  
 ‘another butterfly’
- (ii) Quetta lunga torre ho satto!  
 This.fm.sg long.fm.sg. tower have made  
 ‘I made this long tower!’

In both cases the indefinite modifier *altra* and the attributive adjective *lunga* precede the noun and are introduced by a determiner, as required by standard Italian.

Appendix

Table A: Omissions/occurrences of monosyllabic prepositions in Sabrina corpus

	P omissions	P occurrences
di	21	43
a	6	62
da	4	2
in	3	30
con	11	26
su	2	1
per	//	15
tra/fra	//	//

Table B and Figure A: Omissions/occurrences of monosyllabic prepositions in Sabrina corpus: percentage values

	P omissions %	P occurrences %
di	33	67
a	9	91
da	67	33
in	9	91
con	30	70
su	67	33
per	0	100

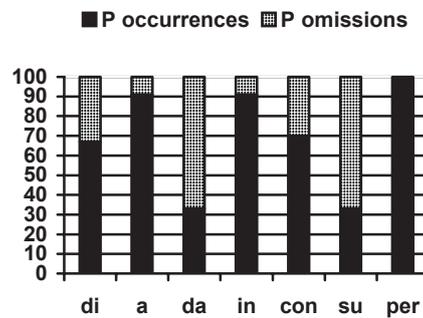


Table C

Italian article paradigm				
	definite		indefinite	
	masculine	feminine	masculine	feminine
singular	il/lo°/l' #	la /l' #	un/uno°	una/un' #
plural	i/gli°	le	Suppletive form: partitive di dei degli° delle	

°allophonic variants

#reduction in front of a vowel

Table D

Italian article choice		
	singular	plural
<b>Definite NP</b> (Known to the speaker and to the hearer-Common ground)	il gatto/la casa	i gatti/le case
<b>Specific Indefinite NP</b> (Known only to the speaker-No common ground)	un gatto/una casa	dei gatti/delle case
<b>Non specific Indefinite NP</b> (Unknown both to the Speaker and to the Hearer-No common ground)	un gatto/una casa	dei gatti/delle case or _gatti/_case

Table E

Monosyllabic Italian Prepositions	Articulated prepositions	Monosyllabic Italian Prepositions	Articulated prepositions
<b>di</b> (of) <sup>23</sup>	+ di + il = del di+ lo = dello di+ l' = dell' di+ la = della di+ i = dei di+ le = delle di+ gli = degli	<b>in</b> (in/at)	+ in + il = nel in + lo = nello in + l' = nell' in + la = nella in + i = nei in + le = nelle in + gli = negli
<b>a</b> (at/to)	+ a + il = al a + lo = allo a + l' = all' a + la = alla a + i = ai a + le = alle a + gli = agli	<b>con</b> (with/by)	+ <b>(in spoken language)</b> con + il = col con + lo = collo con + l' = coll' con + la = colla con + i = coi con + le = colle con + gli = cogli
<b>da</b> (from/to/by)	+ da + il = dal da + lo = dallo da + l' = dall' da + la = dalla da + i = dai da + le = dalle da + gli = dagli	<b>su</b> (on)	+ su + il = sul su + lo = sullo su + l' = sull' su + la = sulla su + i = sui su + le = sulle su + gli = sugli
		<b>per</b> (for/to)	-
		<b>tra</b> (in/between)	-
		<b>fra</b> (in/between)	-

<sup>23</sup> In brackets a roughly corresponding translation is given.

## References

- Abney, S. (1987). *The English Noun Phrase in Its Sentential Aspect*. Phd Diss.: MIT Boston Mass.
- Antelmi, D. (1997). *La prima grammatica dell'italiano*. Bologna: Il Mulino.
- Baauw, S., Avrutin, S., & de Lange, J. (2005). *The omission of D and T in Dutch child language*. ms UiL OTS - Utrecht University.
- Baauw, S., de Roo, E., & Avrutin, S. (2002). Determiner Omission in Language Acquisition and Language Impairment: Syntactic and Discourse Factors. In B. Skarabela, S. Fish, & A. H.-J. Do, *Proceedings of the 25th Boston University Conference on Language Development* (p. 24-35). Boston: Cascadilla Press.
- Belletti, A. (2007). Kinds of Evidence for Linguistic Theory. In E. Agathopoulou, M. Dimitrakopoulou, & D. Papadopoulou, *Selected Papers on Theoretical and Applied Linguistics from the 17th International Symposium on Theoretical and Applied Linguistics. vol.2*, 285-303 Department of Theoretical and Applied Linguistics - University of Thessaloniki.
- Belletti, A. (2004). *Structures and Beyond: The Cartography of Syntactic Structures, vol.3*. New York: Oxford University Press.
- Bernardini Roest, P. (2003). Lo squilibrio nell'acquisizione di due lingue nell'infanzia: indagine longitudinale sullo sviluppo della sintassi nominale. Phd Diss.: Lund University.
- Bottari, P., Cipriani, P., & Chilosi, A. (1993/1994). Protosyntactic Devices in the Acquisition of Italian Free Morphology. *Language Acquisition*, 3, 327-369.
- Caprin, C., & Ioghà, C. (2006). Article omission and the role of the root. In A. Belletti, E. Bennati, C. Chesi, E. Di Domenico, & I. Ferrari, *Language Acquisition and Development Proceedings of GALA 2005* (p. 96-104). Newcastle upon Tyne UK: Cambridge Scholar Publishing.
- Cardinaletti, A. (1998). On the deficient/strong opposition in possessive systems. In A. Alexiadou and C. Wilder, *Possessors, Predicates and Movement in the Determiner Phrase* (p. 17-53). Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Chierchia, G. (1998). References to kinds across languages. *Natural Language Semantics* 6, 339-405.
- Chierchia, G., Guasti, M., & Gualmini, A. (1999). *Nouns and articles in child grammar and the syntax/semantics map*. ms. University of Milano-Bicocca and University of Maryland.
- Chini, M. (1995). *Genere grammaticale e acquisizione*. Milano: Franco Angeli.
- Cipriani, P., Chilosi, A., Bottari, P., & Pfanner, L. (1993). *L'acquisizione della morfosintassi: Fasi e processi*. Padova: Unipress.
- Crisma, P., & Tomasutti, E. (2000). Phonological effects on article omission in the acquisition of Italian. *BUCLD 24 Proceedings* (p. 220-231). Somerville MA: Cascadilla Press.
- De Lange, J., Avrutin, S., & Guasti, M. T. (2006). Cross-Linguistic Differences In Child and Adult Speech Optional Omissions: A comparison of Dutch and Italian. In A. Belletti, E. Bennati, C. Chesi, E. Di Domenico, & I. Ferrari, *Language Acquisition and Development: Proceedings of GALA 2005* (p. 313-343). Newcastle upon Tyne UK: Cambridge Scholar Publishing.
- Gerken, L. A. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language*, 30, 431-451.

- Giorgi, A. & Longobardi, G. (1991). *The syntax of Noun Phrases. Configuration, parameters and empty categories*, Cambridge, Cambridge University Press.
- Giusti, G. (1993). *La sintassi dei determinanti*. Padova: Unipress.
- Giusti, G., & Leko, N. (2001) The categorical status of quantity expressions. In G. Zybatow et al. *Current Issue in Formal Slavic Linguistics*, 96-105, Frankfurt am Main: Ed. Peter Lang.
- Giusti, G. (2003). The Functional Structure of noun phrases: A bare phrase structure approach. In G. Cinque, *Functional Structure in DP and IP* (p. 54-90). Oxford UK: Oxford University Press.
- Granfeldt, J. (2003). *L'acquisition des catégories fonctionnelles*. Phd Diss.: Lund University.
- Guasti, M. T., De Lange, J., Gavarro, A., & Caprin, C. (2004). Article omission: across Child Languages and across Special Registers. In J. van Kampen, & S. Baauw, *Proceedings of GALA 2003 vol.1* (p. 99-110). Utrecht: LOT.
- Kayne, R. (1991). Romance Clitics, Verb Movement and PRO. *Linguistic Inquiry* 22 , p. 647-686.
- Leonini, C. (2006). The acquisition of object clitics and definite articles: Evidence from Italian as L2 and L1. Phd Diss.: University of Florence.
- Longobardi, G. (1999). Bare nouns, proper names, and the syntax-semantics mapping: toward a unified parametric approach. *Rivista di Grammatica Generativa* 24, 45-76.
- Longobardi, G. (2000). Postverbal subjects and the Mapping Hypothesis. *Linguistic Inquiry* 31, 4, 691-702.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pizzuto, E., & Caselli, M. (1992). The acquisition of Italian morphology: implications for models of language development. *Journal of Child Language*, 119 , 491-557.
- Rizzi, L. (1988). Il sintagma preposizionale. In L. Renzi, *Grande grammatica di consultazione*. Bologna: il Mulino.
- Rizzi, L. (2002). *On the Grammatical Basis of Language Development: A Case Study*. ms.: Università di Siena.
- Schoenenberger, M., Penner, Z. & Weissenborn, J. (1997). Object placement in early German Grammar. In A. Hughes, Hughes, M. & Greenhill, A., *Proceedings of the annual Boston University Conference on Language Development* 21, vol.2, 539-549, Cascadilla Press, Somerville, MA.

# V1, V2 and criterial movement in Icelandic\*

Irene Franco

University of Siena - CISCL

[franco12@unisi.it](mailto:franco12@unisi.it)

In this paper I argue that V1 and V2 orders are the outcome of criterial movement to the CP layer of, respectively, a null OP and an XP. Under the assumption that discourse-related properties are encoded in the CP, I propose a cartographic analysis of the mechanisms determining V1 and V2 orders in Icelandic. I show that the labels V2 and V1 are too generic and misleading because they cover constructions with different interpretive properties. On the one hand there are V2 orders found in declarative clauses, which can be distinguished into topic- or subject-initial. On the other hand, V1 orders are marked constructions, where the verb moves to a high position in CP and is preceded by a null OP with a scope-discourse related function. As a consequence, “pure” verb-initial constructions are not found in Icelandic, and, more generally, in Scandinavian. Nonetheless, it is syntactically visible that topic-initial declarative V2 and subject-initial V2 have different pragmatics with respect to what is commonly labelled as V1.

## 1. Introduction

A unitary account of the verb second (V2) phenomena must explain what is its underlying mechanism and how this is related to the imposition of interface conditions on syntax. Whether V2 is the result of V-to-I (specifically to the higher agreement projection in IP) in the so-called symmetric V2 languages (i.e. Insular Scandinavian and Yiddish) or the result of V-to-C movement both in main and subordinate clauses is still under debate.<sup>1</sup> Against the first of these two possibilities, proposed by Holmberg and Platzack (1995) a. o., Hróarsdóttir et al. (2007) provide an argument based on evidence from ECM clauses: “Since ECM infinitives lack the CP domain [...] but contain the IP domain of the clause [...], verb movement in these infinitives should be impossible in Icelandic [...]. This prediction is indeed borne out:[...]”

---

\* This paper is the outcome of research carried out within the graduate program in Cognitive Science at the Interdepartmental Centre of Cognitive Studies on Language (CISCL) of the University of Siena. I want to thank my advisor Luigi Rizzi for his insights and the fruitful discussions; Cecilia Poletto, my coadvisor, for her sharp comments and useful indications; Ian Roberts for his willingness to discuss the issues raised here many times, Liliane Haegeman, Christer Platzack and Anders Holmberg for their precious feed-backs; Hoski Thráinsson, Jóhannes Gísli Jónsson, Tolli Eyþórsson, Sigga Sigurjónsdóttir and the faculty and students at the Institute of Linguistics of Háskóli Íslands for the opportunity to work with them, for their interest towards my research, comments and useful advice.

<sup>1</sup> In the present work I do not address the recent account of V2 as remnant movement, developed by a group of researchers from the University of Tromsø.

- (1) Hann taldi hana (\*hafa) alltaf (hafa) sungið í sturtunni  
he believed her have always have sung in shower.the[...]” (p. 49).

Cecilia Poletto (p.c.) points out that the analysis of control vs. raising and ECM infinitives based on a rough distinction between CP and IP levels is out of date and needs refinements. Indeed, Thráinsson (2007) presents some sets of data whose interpretation is controversial, specifically because what are generally analysed as ECM or raising complements are “rather resistant to modification by sentence adverbs” (see Thráinsson (2007), examples (8.111) and (8.119), pp. 439; 443). According to Thráinsson’s data<sup>2</sup>, the presence of a “higher adverb” (following Cinque’s (1995) hierarchy) degrades this type of clauses regardless the linear order of Adv and V, therefore judging the presence of verb movement becomes difficult in this context.

For the present purposes the exact target position of V in embedded V2 clauses of languages with symmetric V2 is not strictly relevant: either identification with Fin or AgrS allows for a distinction from the target of verb movement in V1 orders.

I present an alternative to Holmberg and Platzack’s (1995) proposal, but I leave the determination of such target position to future research. For ease of exposition, I don’t address the debate throughout the paper, but assume that symmetric V2 can still be considered V-to-C, and focus on the configurational properties of V2 (whose order is XP-V) at a more abstract level. I address what is commonly considered a basic condition for the realization of V2, namely the requirement that one XP is fronted to the first position.

Roberts and Roussou (2002) identify this condition as follows: “... the head containing T must have a filled specifier”, where “the head containing T” is a C-head in the specific V2 case. I will call it “Filled-Spec requirement”:

## (2) Filled-Spec requirement

The Specifier of the position occupied by V in V2 configurations must be filled.

The condition expressed in (2) is purely descriptive, and presupposes that the verb, wherever it ends up, creates a local configuration with the XP in first position. Without a deeper understanding of the syntactic device generating V2, the reasons to have a locality relation at this point of the derivation remain mysterious. Along the lines of Rizzi’s most recent development of the concept of locality, such as criterial movement and freezing effects<sup>3</sup>, discourse-related properties require the satisfaction of certain criteria. The standard way to meet a criterion consists of a head bearing the relevant criterial feature and probing a phrase with a matching feature. The goal is then moved to the specifier position of the criterial projection, where the relevant feature can be interpreted.

The main empirical problem that (2) must face is the presence of productive V1 orders in most Germanic languages. If (2) is a condition applying pervasively to verb movement in Germanic, then the possibility of V1 is unexpected. Otherwise we can assume (2) to hold only in V2 constructions, not in V1. However, (2) does not provide an adequate explanation, but only a description of V2. Following Rizzi’s idea that movement to CP is criterial, I propose to replace the requirement in (2) with that in (3) below:

<sup>2</sup> Whose judgment contrasts with that of (1) in the quotation above...

<sup>3</sup> See Rizzi (2004); (2007) and Rizzi and Shlonsky (2006) a. o.

### **(3) The Criterial-Movement requirement**

In a construction of form XP...V, XP moves to the first position to meet a criterion and ends up in local configuration with the verb.

The assumptions underlying this perspective are that V1 and V2 are distinct phenomena, and that V1 is a spurious set, nonetheless they are both generated by movement to a criterial position.

My argument is based on evidence from Icelandic in comparison with Mainland Scandinavian languages and is developed as follows. Section 2. illustrates the syntactically different distribution of V1 and V2 orders, and disambiguates between V1 orders and constructions that are V2 in disguise, despite looking as V1 orders on the surface. Section 3. gives an interpretation of the data considered, analysing different types of V1 as null OP- V orders. Section 4 is the conclusion.

Two technical remarks are in order: first, in the paper I often use the terms “Topic” and “topicalization”. Although the model assumed is the one proposed by Rizzi (1997)b., it is out of the scope of the paper to distinguish between Topic and Focus in V2 constructions, as further research is needed on this point. I do not exclude that what is called Topic here is indeed a Focus, with the exception of Expletive-V constructions.

Secondly, in order to represent the different configuration of V1 and V2 constructions, I had to face the issue of the reduction of the periphery. A still open question is whether the periphery is always fully projected but some nodes cannot be target of internal/external merge or only necessary nodes are projected<sup>4</sup>. Because I cannot enter this discussion here, I use the terms “active” or “activation” instead of the less neutral “projected” or “present”, with regard to the syntactic structure.

## **2. Different distribution of V1 and V2**

In this section I show how the presence of productive V1 in Icelandic is still compatible with the requirement that verb movement to the CP domain is associated to movement of an XP to the first position. In the previous section, I have argued that the XP movement to the first position is imposed for the satisfaction of a criterion. In order to prove this hypothesis, I present some facts concerning the distribution of V1 and V2 orders. First, I describe the syntactic environment where V1 orders occur. Secondly, I compare it with other clauses where V1 is not possible, but V2 is, distinguishing between real V1 and disguised V2.

### *2.1. V1 orders*

Verb first orders are quite common in Insular Scandinavian among other languages (cf. Maling (1990); Sigurðsson (1990); Thráinsson (2007); Thráinsson et alia (2004)). The first type of V1 order to consider is that of Imperative clauses. Imperative clauses are generally verb-initial in other languages as well, and have a specific syntactic behavior, related to their structural properties. Compare the following facts:

- (4) Far Þú/ Farðu heim! (Icelandic)  
Go you/go-you home  
'Go home!'

---

<sup>4</sup> A position in this debate is offered by Starke (2004) a. o.

- (5) \*Hann sagði [að farðu heim]  
He said that go-you home

[Thráinsson 2007, 28: (2.22)a.:(2.23)]

Imperatives cannot be found in embedded clauses in Icelandic. Platzack (to app.) analyzes Old Icelandic and Old Scandinavian, where embedded imperatives are indeed possible. He observes that the imperative verb cannot raise all the way up to CP, because the projection encoding [Force] is [+ declarative] and realized as a complementizer head, which mismatches with the imperative force of the verb. Embedded imperatives are then possible because the obligatory subject is the rescuing strategy: the overt subject checks the relevant phi-features on T.

- (6) Jak manar thik...At thu sigh mik sannindh (Old Swedish)  
I advise you that you say-IMP me truth-the  
'I advise you to tell me the truth'

[Platzack to app., 64: (28)a.]

Platzack maintains that there is no EPP feature associated with T in this case: in fact imperatives are bare stem and don't need to pick up any morphology in T (thus they cannot check the phi-feature on T, despite being equidistant goal with subjects in Spec,vP). Platzack's hypothesis, then, is that embedded imperatives stay lower<sup>5</sup>. Lack of EPP is proved by the presence of pseudo-OV word order in Old Icelandic and Old Scandinavian, where the finite verb precedes the object both in main clauses and in embedded clauses, but the non-finite verb follows the object as in (7) below:

- (7) Skalt þú jafnan þessu sæti halda (Old Icelandic)  
Shall you always this seat keep  
'You will always keep this seat'

[Platzack to app., 66: (31)b.]

Once the option of OV orders is lost, embedded imperatives are no longer possible. Doing away with the difficulty posed by Old Icelandic and Old Swedish, we can see that the impossibility of being embedded seems a general property of V1 clauses, as Maling (1990) observes.

Another case of productive V1 is found in Y/N questions. In general, Germanic (among other languages) requires subject-verb inversion in that context, whereas inversion is not triggered when the question is indirect:

- (8) a. **Hefur** Jón ekki [lesið bókina]? (Icelandic)  
Has John not read the book

<sup>5</sup> I report Platzack's explanation: "there are two possible ways to delete this EPP-feature: either by the pronoun *thu* or by the imperative verb. Since the structure we are deriving is embedded under a declarative complementiser, we know from the discussion above that it will crash if the imperative verb with its feature [+κIMP] is placed in T<sub>Ev</sub>. Hence, the only way to delete [-φ] in T<sub>Ev</sub> is to raise the pronoun to SpecTP<sub>Ev</sub>. This account predicts both that the subject pronoun must be overt in embedded imperatives, and that it must precede the imperative verb." (See p. 65 of Platzack's paper for further details).

b. **Las** Jón ekki [\_\_\_ bókina]?

Read John not the book

[Thráinsson 2007, 28: (2.21)]

(9) **Heldur** tú, at Zakaris seldi Eivindi tann gamla bilin? (Faroese)

Think you that Zakaris sold Eivind the old car-the

‘Do you think that Zakaris sold Eivind the old car?’

[Thráinsson et alia 2004, 238: (34)c.]

(10) Hann spyr hvort Jón **taki** bækurnar (Icelandic)

He asks if John take(pres.subj.) books-the

[Thráinsson 2007, 397: (8.7)a.]

In the subordinate clause in (10) the verb does not raise to the first position: when a polar question is embedded, V1 order is no longer grammatical. Notice that “inversion” regards the verb and the element occupying the subject position, regardless case-marking. This is shown by Icelandic, where the dative experiencer is the subject of raising predicates:

(11) a. **Virtist** *henni* hesturinn hafa týnt knapanum? (Icelandic)

Seemed her(D) horse-the(N) have lost jockey-the

‘Did it seem to her that the horse had lost its jockey?’

b. \***Virtist** hesturinn *henni* hafa týnt knapanum?

Seemed horse-the to-her have lost jockey-the

[Thráinsson 2007, 440: (8.114)a.; b.]

Another case of V1 is represented by the "narrative style", mainly found in current written Icelandic and earlier written Faroese (it is rare in the modern variety, see Thráinsson et alia (2004), p. 239):

(12) **Koma** Þeir nú að stórum helli og... (Icelandic)

Come they now to big cave and

‘Then they get to a big cave and...’

[Thráinsson 2007, 29: (2.22)b.]

(13) [...] **Hitti** hann har nakrar studentar úr Íslandi og **sótu** teir har leingi (Faroese)

met he there some student from Iceland and sat they there long

[Thráinsson et alia 2004, 240: (39)a.]

According to Thráinsson (2007) narrative V1 clauses can only marginally be embedded "except for the second conjunct of conjoined complement clauses if the complementizer is absent" (Thráinsson (2007), p. 29). However, Thráinsson's judgement on (14) is not accepted uncontroversially by other speakers who do not really like the sentence:

(14) (?) Hann sagði [að **hefðu** Þeir Þá komið að stórum helli og...] (Icelandic)

He said that had(subj.) they then come to big cave and..

- (15) Hann sagði [[að Þeir hefðu haldið áfram]  
 He said that they had(subj) continued  
 a. ... og [að Þeir hefðu Þá komið...]  
 and that they had then come  
 b. ?\* ... og [að hefðu Þeir Þá komið...]  
 c. ... og [hefðu Þeir Þá komið...]  
 ‘He said that they had continued and they had then come...’  
 [Thráinsson 2007, 29: (2.24 a.; b. 1,2,4)]

A refinement comes from the comparison between (14) and (15). Thráinsson (2007) observes that the option in (15)c. is fine and sounds like narrative V1, whereas the one in b. is just bad. The reason of this split may reside in the fact that two different phenomena are at play in (15). The first conjunct in the subordinate clause in (15) is clearly an instance of V2 (cf. with (14) above, not accepted), and the second conjunct is expected to display the same structure. However, a V2 order is not possible in (15)b. because the first position of a V2 structure is left empty, contrary to (15)a., where the subject moves<sup>6</sup> to CP. The structure of the second conjunct in (15)c. is thus different from that of (15)a. and b.

The crucial difference in (15)c. is that the complementizer *að* is missing, and verb movement does not require overt XP-fronting. These syntactic facts reflect the different interpretation of the conjunct as a narrative-style clause, in other words there is syntactic evidence of what characterizes narrative V1 semantically. More specifically, the impossibility to embed V1 clauses is motivated syntactically with the complementary distribution of complementizers and verbs in first position. Maling (1990) describes the distribution of V1 as follows:

“First position for the finite verb (V1) is typically reserved for main clauses, namely for direct questions, imperatives and the so-called narrative style characteristic of Icelandic. In embedded clauses, the finite verb comes first in conditional sentences without the conjunction *ef*” (Maling (1990), p. 72).

The fact that V1 is common in the subordinate clauses of conditional sentences appears as a counterexample to the claim that V1 clauses cannot be embedded. The condition for the realization of V1 is primarily identified by the lack of a complementizer, hence unselected clauses are the optimal environment. Nevertheless, from a syntactic point of view, nothing prevents V1 to occur in complementizerless subordinate clauses: the restrictions at this point depend on the criterial properties of the clause.

Conditional V1 is indeed spread crosslinguistically, and allowed only in absence of the subordinating particle. Luigi Rizzi (p.c.) observes that English (example (20)) allows conditional V1 only in counterfactual contexts. This seems to be a semantic restriction operating at the interface level.

- (16) a. Jón verður góður [*ef* hann **æfir** sig]. (Icelandic)  
 Jon becomes good if he practises self  
 b. Jón verður góður [**æfi** hann sig].  
 Jon becomes good practise(subj.) self  
 ‘Jon will be good if he practices’ [Thráinsson 2007, 30: (2.24)]

<sup>6</sup> On the possibility of a criterial position in CP for subjects triggering V movement further research is needed. Specifically, it is not clear, at this point of the investigation, whether pronominal subjects in Insular Scandinavian pattern like expletives or like lexical subjects.

- (17) a. Bókin kemur út til jóla, *um* tað **gongst** eftir ætlan (Faroese)  
 book-the comes out to Christmas, if it goes after plan  
 b. **Gongst** \_\_ eftir ætlan, kemur bókin út til jóla  
 Goes after plan comes book-the out to Christmas  
 ‘If everything goes according to plan, the book will come out by Christmas.’  
 [Thráinsson et alia 2004, 293: (181)c.]
- (18) a. Sie schaute ihn an, als *ob* er ein großes Verbrechen begangen **hätte** (German)  
 b. Sie schaute ihn an, als **hätte** er ein großes Verbrechen begangen  
 She looked him at as if/had he a big crime committed
- (19) a. \* Sie schaute ihn an, als *ob* **hätte** er ein großes Verbrechen begangen  
 b. \* Sie schaute ihn an, als **hätte** *ob* er ein großes Verbrechen begangen  
 She looked him at as if/had he a big crime committed  
 [Vikner 1995, 44: (20); (22)]
- (20) a. **Had** you studied more, you would have passed the test (English)  
 b. \**If* **had** you studied more, you would have passed the test

Two further remarks are in order. First, notice that English is not a V2 language, but displays only "residual V2". Beside the inversion phenomena triggered by quantificational/negative elements (e.g. "Never would I do this again."); the label "residual V2" has been used to identify other common instances of inversion, such as the one found in conditional clauses, narrative style, etc. I suppose that this classification needs a revision, because what is called "residual V2" may indeed be a case of V1 and the two notions are distinct, as becomes evident in the following sections.

The second remark consists of the fact that the Icelandic sentences in (16)a. and b. are not only different as for the verb position in the subordinate clause, but also for the mood expressed in it. Notice that verb raising to first position in (16)b. is associated with subjunctive morphology, while the verb in a lower position comes in the indicative ((16)a.). Discussing the role played by subjunctive mood in the interpretation of a clause would take much longer than a paragraph in a paper, thus I do not address the issue here. However, it is worth pointing out that the distribution of subjunctive, the "irrealis" mood, is generally more frequent in dependent clauses and related to their criterial properties<sup>7</sup>. From a syntactic perspective, a concrete hypothesis for explaining the facts in (16) is that the presence of subjunctive may obviate the lack of a complementizer typing the clause as conditional. However, the presence of subjunctive and that of certain clause-type features cannot be accounted for as a one-to-one correspondence. I leave this issue for future research, while in the next subsection I concentrate on the differences between V1 and V2 orders.

<sup>7</sup> Thráinsson (2007) himself argues that "if- clauses standardly take the indicative [...] the subjunctive is standardly used in (preposed) conditional clauses if the subordinating conjunction is omitted [...] this indicates further that the semantics of the conditional clauses is related to that of the subjunctive" (pp. 405-406).

## 2.2. Genuine V1 and disguised V2 in Icelandic

With respect to the theoretical framework proposed in section 1., not all cases of V1 orders are so uncontroversial as those just illustrated. In this section I show that some cases of embedded V1 are not problematic, because they are reanalyzed as disguised V2 orders, for which (3) is still satisfied.

Haeberli (2002) presents the facts in (21) below as related to the possibility of having productive V1 in Icelandic<sup>8</sup> embedded clauses, in contrast to Yiddish, which allows V1 only in main clauses, as (22) shows:

- (21) a. Því er sennilegt að **rigni** meira á morgun (Icelandic)  
 Thus is likely that rains more heavily tomorrow  
 ‘Thus it is likely that it will rain more heavily tomorrow’
- b. Ég vissi ekki að **færu** til Grænlands svona mörg skip  
 I knew not that went to Greenland so many ships  
 ‘I didn’t know that so many ships went to Greenland’
- (22) a. \* Er hot gefregt, tsi **iz** varem in shtub (Yiddish)  
 He has asked if is warm in room
- b. \* Ikh hob nit gevust, az **geyn** keyn Grinland azoy fil shifn  
 I have not known that go to Greenland so many ships  
 [Haeberli 2002, 342-343: (58); (59)]

Examples a. are impersonal constructions, examples b. are cases of postponed subjects.

Haeberli explains the difference between (21) and (22) under the assumption that in Icelandic a null expletive is optionally licensed in Spec, CP. If (21) is considered as an instance of embedded V1, then it is not clear why, in this case, the verb and the complementizer *að* are licensed in the same clause. My answer is that (21) is not a case of embedded V1, but of V2, with a null expletive in first position, as proposed by Haeberli. Notice that the embedded clauses in (21)a. and b. are declaratives. Maling (1990) observes that surface V1 in embedded clauses is possible only in presence of a subject gap:

- (23)a. Þetta er handritið sem *hann* **skrifaði** eftir (Icelandic)  
 this is the-manuscript that he copied from
- b. \*Þetta er handritið sem **skrifaði** *hann* eftir  
 [Maling 1990, 84: (42)]

The argument against embedded V1<sup>9</sup> can be summarized as follows:

- a. embedded V1 is possible only when a subject gap is present (no V1 with subject NPs)
- b. main V1 is possible regardless the presence of a subject  
 → main V1 and embedded V1 must be different phenomena.
- Maling describes subject gaps as a characteristic of one of the following constructions:

<sup>8</sup> Declarative V1 is found in Swedish as well, but has a completely different distribution from Icelandic. Specifically, V1 declaratives cannot be embedded (see Mörnjö (2002)). I discuss Swedish in the next subsection.

<sup>9</sup> With the exception of the conditional clauses previously discussed.

## *V1, V2 and criterial movement in Icelandic*

- (i) extraction of the subject NP
- (ii) impersonal passives or lexically impersonal predicates
- (iii) indefinite NP-postposing

The contexts in which subordinate V1 is found in Icelandic are given in the following examples:

Declaratives ((21) is repeated below as (24)):

- (24) a. Því er sennilegt að **rigni** meira á morgun (impersonal construction)  
Thus is likely that rains more heavily tomorrow  
'Thus it is likely that it will rain more heavily tomorrow'
- b. Ég vissi ekki að **færu** til Grænlands svona mörg skip (postponed subject)  
I knew not that went to Greenland so many ships  
'I didn't know that so many ships went to Greenland'
- (25) Þá gæti enginn sagt með vissu, að sva **hefði** verið (with Stylistic Fronting)  
then could no-one say with certainty, that so had been  
[“Tilhugalíf”, Ch. 5, in Maling 1990, 75: (7)d.]

### Indirect questions:

- (26) a. Hann spurði *hvar* \_\_ væri ennþá ekið vinstra megin (impersonal construction)  
He asked where was still driven left side
- b. Hann spurði *hvar* **ekið** væri ennþá vinstra megin  
He asked where driven was still left side
- c. \*Hann spurði *hvar* **Það** væri ennþá ekið vinstra megin  
He asked where it was still driven left side  
'He asked where people still drove on the left side of the road'  
[Maling 1990, 84; 85: (41), (46)]

(25) and (26)b. are a case of Stylistic Fronting (SF): the first position, otherwise empty, is occupied by a fronted element. Overt expletives are generally licensed when non-selected and clause-initial, otherwise the preverbal position can be occupied by a null expletive, as the contrast between (26)a. and c. shows. (27) below shows that SF and null expletives have the same distribution.

### Relative clauses:

- (27) Maðurinn *sem* **farinn** var/ \_\_ var farinn heim heitir Pétur  
The-man that gone had/ \_\_ had gone home is-named Peter  
[Maling 1990, 73: (4)c.]

Notice that in the cases just presented a complementizer or a subordinating particle (in italics) is present despite the initial position of the verb, in contrast to the cases of main clause V1 illustrated in section 2.1. above.

The explanation of this difference is that Icelandic embedded V1 is indeed a case of disguised V2. In fact, in *declarative clauses* the subject gaps “must be filled in order to satisfy V2, either by stylistic fronting or by *Það*-insertion [...]” (Maling (1990), p. 85). This accounts for the following facts, contrasting with (24) above:

- (28) a. ?Þeir segja [að \_\_ verði dansað í brúðkaupinu]

- They say that will-be danced in wedding-the  
 b. Þeir segja [að **Það** verði dansað í brúðkaupinu]  
 They say that there will-be danced in wedding-the

[Thráinsson 2007, 335: (7.35)]

The answer to the dilemma raised by the contrast between (24) and (28)a. comes from Sigurðsson (1990). He provides a different description of subject gaps and argues that the presence of *það* in embedded declaratives with non-topical subjects is not necessary. The fact that *það*-insertion is not required in some embedded declaratives is compatible with the facts in (24) above. Embedded declarative clauses with an unfilled subject gap can be considered a marginal option, strictly related to the presence of a non-topical subject as a syntactic constraint and to discourse related properties (cf. Sigurðsson (1990)).

Let's assume that null expletives in embedded declaratives with non-topical subjects are just an option, whereas *það*-insertion is preferred in certain binding contexts (cf. Sigurðsson (1990)). Furthermore, subject gaps cannot host overt expletives in clauses inherently containing gaps (Maling (1990); Sigurðsson (1990)). Then the two possible embedded structures codifying the first position in the examples (24); (25)-(27) above are:

- (29) a. [decl]: [Matrix [ **XP/ Það** /(<sup>ok</sup>/? \_\_) V (t<sub>XP</sub>)]]  
 b. [rel/Wh-/compar./etc.]: [Matrix [ **XP/ \_\_** /(\***Það**) V (t<sub>XP</sub>)]]

In (29) XP- t<sub>XP</sub> is the chain created by stylistic fronting; \_\_ represents a subject gap, and *það* is the expletive.

Further research is also needed on the distribution of *það* and null expletives, as the judgments seem to vary a lot, especially among speakers of different ages. One possible explanation for the distribution of *það* and null expletives has to do with the way in which Icelandic satisfies the Subject Criterion<sup>10</sup>. Because of the availability of null expletives, in Icelandic it is possible to have certain kinds of subject extraction, in a similar way to null subject languages, where merger of *pro* in the criterial position allows for subject extraction. On the other hand, Icelandic has only non-referential null expletives, thus it may as well adopt a different strategy of subject extraction, similar to the one found in non-null subject languages as French (see Rizzi and Roberts (1989) on the *que/qui* alternation in French). In sum, Icelandic displays a certain optionality in the way to satisfy the Subject Criterion and extract the subject: either non-referential *pro* or *það* is merged, or the relevant features on Fin are valued by movement. In the second case, there is reason to believe<sup>11</sup> that, in default contexts

<sup>10</sup> The Subject Criterion is a reformulation of ECP and EPP principles: the unmovability of subjects and their obligatoriness are in support of the idea that subjects must satisfy a criterion and undergo criterial freezing (cf. Rizzi (2004) and Rizzi and Shlonsky (2006)).

<sup>11</sup> Evidence for the point I am making here comes from the typological differences between Icelandic and other Scandinavian complementizer systems, specifically from the selectional properties of the different complementizer heads as I found them in the data sets in Thráinsson (2007). For instance, in subject indirect questions a subject relative pronoun follows the Wh- pronoun, whereas this is impossible in Icelandic (which admits only Wh- pronouns):

- (i) Jeg ved ikke, hvem der/\*e \_\_ har boet i det hus. (Danish)  
 I know not who that has lived in that house  
 (ii) ég veit ekki hver \*sem/e \_\_ kemur (Icelandic)  
 I know not who that came

in Icelandic, the features specified on Fin, driving intermediate movement to a Criterial position, may be purely formal in the sense of Rizzi (2004, p.10). That means that *any category* can potentially satisfy the subject criterion, which is what happens with stylistic fronting.<sup>12</sup>

Diachronic studies on Old Norse shed some light on the availability of null expletives in Icelandic. Among the characteristics of Old Norse there is a massive presence of V1 which covers a wider range of pragmatic contexts than that of Narrative Inversion, a marked word order. Old Norse V1 is found in sentences with a subject in low position (as there are more positions available for subjects in this language, see Faarlund (2008), p. 230) or in subjectless sentences. Faarlund (2008) divides subjectless sentences into 2 kinds:

- a) Sentences without an external argument (containing impersonal verbs; psych-verbs; double accusative verbs; mental process verbs; *like*-type verbs; passives; gerunds). For instance:

(30) *súrnar í augunum* (Old Norse)  
 becomes-sour in eyes.D-the  
 ‘one’s eyes are smarting’

[Faarlund, 2008, 217: (57)c.]

- b) Sentences with unexpressed external arguments (impersonal use of otherwise regular verbs; ergative constructions; comparatives; non-specific subjects). For instance:

(31) *má þar ekki stórskipum fara* (Old Norse)  
 can there not big-ships.D travel

[Faarlund, 2008, 221: (69)c.]

Notice that both sentence kinds can be rendered with an expletive construction in Icelandic. Furthermore, in Old Norse clausal subjects are usually either extraposed, resulting in V1 constructions, or they are resumed by the demonstrative *þat* of which they are complement:

(32) *satt er þat*, at mjök er niðr fallit ríki Haralds konungs ins hárfagra  
 true.N is that.N,that much is down fallen kingdom.N Harald.G king the hair-  
 fine.DEF

‘It is true that King Harald the Fine-haired’s kingdom has suffered a great decline’

[Faarlund, 2008, 224: (75)a.]

Old Norse lacks the expletive *það*, thus the hypothesis is that the latter is the diachronic evolution of the demonstrative *þat*, which has a very similar distribution, although it may alternate with a null element. Once the lexical entry for the expletive is available in the language, the presence of V1 order decreases dramatically and the form Expl-V is preferred.<sup>13</sup>

---

(see Thráinsson (2007), pp. 448-9 for further details). Such properties are allegedly related to the amount of feature specification on the head, which I assume to be Fin.

<sup>12</sup> An accurate description of subject extraction strategies and the characteristic of Icelandic Fin is currently under investigation.

<sup>13</sup> Þórhallur Eyþórsson (p.c.) confirms that hypothesis.

The fact that, in contemporary Icelandic, unfilled subject gaps in embedded declarative clauses are not always licensed or can alternatively be filled by the overt expletive *það* or by stylistic fronting, proves that embedded V1 is indeed V2 in disguise. We can see that the (purely descriptive) Filled-spec requirement on V2 does not admit a gap “e” in first position. With regard to embedded relative, Wh- or comparative clauses the Criterial Movement Requirement in (3) is expressed with the Subject Criterion, along the lines of Rizzi (2004).<sup>14</sup> In this framework, in order to have a well-formed V2 structure, what is really needed is meeting a criterion, not necessarily the presence of an XP in the Spec, CP position.

### 2.3. A comparison with Mainland Scandinavian: the case of Swedish declarative V1.

In her doctoral dissertation, Mörn sjö (2002) illustrates different cases of V1 declaratives in spoken Swedish. This kind of V1 orders are referred to in the literature as cases of “Topic drop”, although the label “Topic” is quite generic and prone to the attribution of different discourse-related properties such as “aboutness” (i.e. argumental topics), rather than pertinence to the “frame” (e.g. adverbial topics). Mörn sjö (2002) identifies the conditions licensing topic drop with the presence of a presupposition that the speaker estimates as belonging to the communicative common ground he shares with her interlocutor and, as such, he can drop. This pragmatic characterization of Swedish V1 declaratives is substantially different from the syntactic properties licensing embedded V1 in Icelandic. As a matter of fact, Swedish topic drop is impossible in embedded clauses:

(33) Här är pajen. Ø<sub>i</sub> kan du sätta in t<sub>i</sub> direkt i micron, om du vill. (Ø= den)  
 here is pie-the Ø can you put in directly in micro-the if you want (Ø= it)  
 ‘Here is the pie. You can put it directly in the micro if you want’

(34) \*Han pekade på pajen och sa [att Ø kan du sätta in direkt i micron.]  
 He pointed on pie-the and said that Ø can you put it directly in micro-the  
 ‘He pointed at the pie and said that you can put it directly in the micro’  
 [Mörnsjö, 2002, 11: (2:1); (2:2)]

In (33) V1 is the result of topic drop: the object is topicalized to a specifier position in the high left periphery: the V2 configuration is thus created. Then the object is “dropped” at the interface and becomes silent because it is presupposed from previous discourse. This is possible with a full-blown peripheral structure, as is found in root clauses, whereas topic drop is ungrammatical in embedded clauses, as (34) shows. The reason for this ungrammaticality resides in the structural (and interpretive) properties of selected clauses in Mainland Scandinavian. Notice that Swedish does not allow null expletives:

(35) \*På lördag ska åka till Paris (Swedish)  
 on Saturday shall go to Paris  
 [Mörnsjö, 2002, 12: (2:3)]

<sup>14</sup> This is still work in progress. The idea that the feature on Icelandic Fin is purely formal was first expressed by Rizzi and Shlonsky in a brief manuscript on quirky subjecthood: this option is not available for Mainland Scandinavian, as will be clarified later on. Another reason in support of it is the fact that stylistic fronting, despite being identified as head movement to Fin, meets all the interface conditions and rescues the structure.

In (35) the first position is occupied by a topicalised adverbial, and no topic-drop takes place. Whatever the interpretation of the omitted subject could be, the result is still ungrammatical, contrary to Icelandic where a non-referential subject can be a null expletive:

- (36) Svo byrjaði  $\emptyset$  að rigni og rigni (Icelandic)  
 then began to rain and rain

[Example from the web]

Swedish declarative V1 is thus an interface phenomenon which can be still considered V2 syntactically, with a first, empty position.

I suspect that the difference between Icelandic and other Mainland Scandinavian languages depends on the parametric variations regarding the subject criterion. This hypothesis gets visible support from the crosslinguistic variations in V2 structures. The properties that are characteristic of Fin in Icelandic are not equally available in other Scandinavian languages. Consider the contrast between Icelandic and Danish in the following examples:

- (37) **Hvernig** sagði hún **t** að börnin hefðu **t** alltaf lært sögu **t**? (Icelandic)  
 How said she that children-the have(COND) always learned history

- (38) \***Hvordan** sagde hun **t** at børnene havde **t** altid lært historie **t**? (Danish)  
 How said she that children-the have always learned history  
 [Vikner 1995; 112: (118)b.]

The ungrammaticality of (38) is produced by the fact that the topicalised phrase *børnene* intervenes in the A' movement of the Wh- adverbial. This is not the case in (37) and the explanation is that the Icelandic lexical subject *börn* is not A'-moved, but A-moved to a subject position in CP which doesn't seem equally available in Mainland Scandinavian.<sup>15</sup> The availability of an A-position in the Icelandic CP gets further support from the following pair:

- (39) a. **Hvenær** heldur þú [að það verði ball í skólanum \_\_\_]? (Icelandic)  
 When think you that it will-be dance in school-the

- b. ?\***Hvenær** heldur þú [að í skólanum verði ball \_\_\_ \_\_\_]?  
 When think you that in school-the will-be dance

[Thráinsson 2007, 329: (6.44)b.; (6.45)b.]

The ungrammaticality of (39)b. is due to the intervention of the topicalised phrase *í skólanum* with the A'-movement of the Wh- adverb. On the contrary, no minimality effects are triggered by the expletive subject *það* in first position in (39)a. This contrast does not uniquely emerge in extraction contexts but also depends on the selectional properties of matrix clauses on their complement clauses. Consider the contrast between (40) and (41) below:

<sup>15</sup> Platzack (forthcoming) suggests that in Scandinavian there is evidence for one of the two subject positions described in Cardinaletti (2004) in CP. However, there seems to be crucial differences between the availability of such position in Icelandic and Mainland Scandinavian. This issue deserves further investigations.

- (40) a. Hann sagði að **hann** gæti ekki sungið í brúðkaupinu (Icelandic)  
He said that he could not sing in wedding.the
- b. Hann sá eftir að **hann** hafði ekki sungið  
He regretted that he had not sung
- (41) a. Hann sagði að **Þetta lag** gæti hann ekki sungið í brúðkaupinu  
He said that this song could he not sing in wedding.the
- b. \*Hann sá eftir að **Þetta lag** hafði hann ekki sungið  
He regretted that this song had he not sung  
[Hróarsdóttir et alia 2007, 56: (18); (19)]

The ungrammaticality of (41)b. is related to the semantic and selectional properties of the matrix clause. Whether an analysis of the facts considers as the relevant criterion the interpretation of the matrix clause as a whole (see Hróarsdóttir et alia, 2007) rather than the classification of the matrix verb as “non-bridge” (cf. Vikner, 1995), the high left periphery of the selected clause cannot host a topic. A similar contrast is found in Mainland Scandinavian, with the difference that matrix clauses of the b-type (wrt the just mentioned examples) cannot select a V2 clause, and the verb stops in a lower position. With regard to subordinate clauses, the crosslinguistic difference thus concerns the structural properties of the high left periphery in the two classes of languages. In Mainland Scandinavian V2 is only possible with a full-blown CP displaying root properties. In Icelandic it suffices to have an available subject position in order to have a V2 structure, but this is not an option for Mainland Scandinavian<sup>16</sup>, given the different agreement properties<sup>17</sup>.

In this section I have tried to prove that the label V1 is way too generic and groups together different syntactic phenomena.

On the one hand, V1 is found in main clauses: it is in *complementary distribution with subordinating particles and complementizers*, and is generally understood as a pragmatically marked order associated to specific clause types (imperatives, polar questions, narrative style, a.o.). This markedness is also found in certain embedded contexts (conditional clauses), provided that the syntactic conditions for a correct clause-typing are respected, i.e. there is no complementizer introducing the “if-clause”. Therefore the correct interpretation of V1 orders seems to be connected to some discourse-related function, clearly visible in the syntactic structure as linear V-S order.

On the other hand, it is possible to have declarative V1. This is a problem if we agree that [declarative] is the unmarked clause-type, because then it is not possible to find a specific discourse function that distinguishes a V1 from a non-V1 declarative clause.

In order to solve this problem, I have considered the case of Icelandic. By showing that declarative V1 orders, common in Icelandic, are *compatible with complementizers* and are found in embedded clauses, I have argued that they are indeed cases of disguised V2 (with which they have the same distribution).

Notice that in all Scandinavian languages embedded V2 cooccur with complementizers and subordinating particles. This fact proves that, with V2 orders, the target of V in CP must be lower than the projection hosting complementizers. The

<sup>16</sup> For a comparison between Icelandic and Mainland Scandinavian on this topic see tables and data in Hróarsdóttir et al. (2007).

<sup>17</sup> Cf. also Holmberg and Platzack (1995).

assumptions are that analogue Scandinavian complementizers (e.g. *að; at; att*) are merged in the same position, and that V targets a lower position in CP in some complement clauses<sup>18</sup>. Clause-type features are specified on ForceP, which attracts the Scandinavian complementizers<sup>19</sup>. As a consequence, the target of V-movement in V2 clauses will be a head position in the lower CP area.

A comparison with spoken Swedish V1 was brought about to illustrate that declarative verb-initial order in subordinate clauses is found only in Icelandic, among the Scandinavian languages, and that it is related to the particular agreement properties of this language. Notice that the requirement in (3) covers also cases of disguised V2 in Icelandic, as in that case movement is triggered to meet the Subject Criterion.

### 3. Interpretation

In this section I argue that the different syntactic behavior of V1 and V2 constructions depends on interface requirements imposed by discourse-related properties.

The comparative data presented in the last part of the previous section can be interpreted as follows. The Criterial-Movement Requirement determining (declarative) V2 is satisfied in different ways crosslinguistically. In Mainland Scandinavian it results in a Topic-criterion, whereas in Icelandic the Subject Criterion itself is sufficient to obtain a V2 configuration. That doesn't imply that Mainland Scandinavian doesn't have a Subject Criterion, but only that it is not (necessarily) associated with V2<sup>20</sup>.

On the other hand, other criteria satisfied in CP are dissociated from V2 order, as has been described so far. Namely, there are criteria, such as Interr- (for "interrogative") or Imper- (for "imperative"), which are rather associated with a superficial V1 order. In this section it becomes clear that V1 and V2 are indeed two merely descriptive but misleading labels for syntactic phenomena that are instead produced by interpretive requirements.

#### 3.1. Interpreting V1

In his (2006) paper, Allan analyzes the illocutionary force of a clause as dependent on the presence of specific clause-type operators. Allan considers English, but his proposal can be easily extended to other languages. The basic idea is that "each clause-type has a unique **primary illocution** (PI) which provides an initial clue to the pragmatically determined illocutionary point of the utterance containing the clause" (Allan (2006), p. 2). By primary illocution, Allan means the semantics of the clause type and postulates the presence of three main PI operators: declarative (T); interrogative (Q); imperative (I), and other three secondary PI operators: hypothetical (H); expressive (X) and exclamative (!). Main PI operators stand in complementary

---

<sup>18</sup> Depending on whether the language considered has symmetric or asymmetric V2 and on whether Icelandic V2 is analysed along the lines of Holmberg and Platzack (1995) or not.

<sup>19</sup> Western Germanic complementizers (e.g. *dass*), must merge in the lower C-head Fin<sup>o</sup>, and successively move to Force<sup>o</sup> for clause-typing requirements. This would explain why West Germanic allows embedded V2 only in complementizerless clauses: otherwise the periphery is "locked" by Fin<sup>o</sup>-to-Force<sup>o</sup> complementizer movement. Actually one doesn't need to exclude Fin<sup>o</sup>-to-Force<sup>o</sup> movement for Scandinavian complementizers. Following a proposal by Rizzi (to appear and p.c.), Scandinavian may optionally allow for Fin recursion, if Fin is selected by a Mood phrase. The complementizer would merge in the higher projection of Fin whereby it would be free to move to Force, whereas the lower Fin is an available target position for verb movement in V2 constructions. Further investigations are being made in order to verify the plausibility of such a hypothesis.

<sup>20</sup> The satisfaction of the Subject Criterion resulting in a V2 structure analogous to that found in Icelandic seems marginal in Mainland Scandinavian. However this possibility is under investigation.

distribution among each other, namely a Q-clause cannot also be an I-clause and so forth. (T) is the default, unmarked operator. Being declaratives, topic or subject-initial main V2 clauses and subordinate V2 clauses are thus unmarked for clause-type.

Another implication in Allan's (2006) proposal contributes to the present account of V1 orders. As mentioned above, Allan demonstrates that main clause-type operators are in complementary distribution, which is why imperatives and interrogatives cannot be embedded. In section 2. I have presented cases of V1 clauses, and shown that they cannot be found in embedded clauses. From a comparison with Allan's (2006) classification of clause-types, it turns out that marked main PI operators have their syntactic realization in V1 orders in all Germanic languages (included English), as the following paradigm, with Icelandic examples, show:

(42) a. "Farðu!" (Imperative)  
Go-you

b. "Fór Jón til Íslands?" (Interrogative)  
Went John to Iceland

If we assume that the declarative clause-type (T) has a default value on ForceP, no clause-type feature needs to be specified in those cases. Therefore the verb is not required to move to a higher position and be in configuration checking [clause-type] features in declaratives.

The adoption of Allan's (2006) proposal for the identification of the interpretive properties of V1 orders results in a felicitous one-to-one relation with marked clause-type values. The question now is: how is the [clause-type] feature of a clause valued and checked in the syntactic configuration?

Following Allan (2006), a hypothesis is that clause-type features can be identified with different criteria. From a syntactic perspective, what happens is that the Force head selects a clause-type headed by a specific phrase. This phrase must be visible at the interface for full interpretation, therefore it must have phonological realization, and the relevant criterion identifying the clause type must be met. In order to do so, the projection selected by Force must have either an overtly realized head or criterial movement of an overt XP to its specifier position is triggered.

Let us now consider the different clause-types corresponding to V1 orders and how they satisfy the respective criteria.

### 3.2. Imperative clauses

The default imperative form in Icelandic, as well as in Old Norse, consists of the verbal head plus the subject clitic form as in the second option in (4) repeated below as (43):

(43) Far þú/ Farðu heim!  
Go you/go-you home  
'Go home!'

If we compare this subject enclitic form with the studies on Medieval Romance (Benincà (2006)) and Rhaeto-Romance (Poletto (2002)), we may hypothesize that Benincà's analysis applies to Icelandic and Old Norse as well. Benincà observes that enclitics are incompatible with FocP, but are triggered when a higher phrase is activated. From a comparison between Icelandic and Old Norse, the V1 order of

imperative clauses can be analyzed as verb movement to a quite high position in CP, surely higher than the Focus field, and perhaps endowed with the [+ addressee] feature. In this framework, Zanuttini (forthcoming) formulates the hypothesis of a Jussive Phrase characterizing the imperative clauses. Faarlund (2008) observes that in imperative sentences “the topic position is usually empty [...] but it may also be filled by an adverbial”. Dispensing with Faarlund’s generic terminology for identifying the preverbal position, Old Norse facts prove that the first position could have overt realization:

- (44) *þá* tak þú af tvá hluti (Old Norse)  
 Then take.IMP2s you.N off two parts.A  
 ‘Then withdraw two parts’ [Faarlund 2008, 229: (88)]

Given the basic CP structure proposed by Rizzi (1997)b., I assume that the left periphery of some clauses is “defective”. In the case of imperatives, the inflection is reduced to second person, but no Tense is specified. It seems plausible that FinP is equally not projected in this type of clauses, as suggested by Platzack and Rosengren (1998), and Belletti (1999).<sup>21</sup>

The preverbal adverbial *þá* in (44) is replaced by a null OP in contemporary Icelandic, which determines the scope of verb movement. The idea is that V moves to such a high position for scope-discourse related properties, namely as the overt realization of a criterial head.

### 3.3. *Interrogative clauses*

Another marked clause-type displaying V1 order is that of Y/N interrogatives, which can be analyzed in analogy with Wh-questions. There is vast literature on the target position of Wh-movement in questions, and recent cartographic approaches have identified the landing site of Wh-elements in the Focus field (see Rizzi (1997)a. and Benincà (2006) a.o.).

Rizzi (1997)a. argues that Wh-questions implement the Wh-criterion, where a spec-head configuration is created between the verb and a Wh- element. This idea can be extended to polar questions, by assuming that a null, interrogative operator (i.e. Allan’s (2006) (Q) PI) replaces the Wh- element, thus satisfying an “Interr-Criterion”. Straightforward evidence for the presence of an operator in Y/N questions comes from diachronic studies on Old Norse. Old Norse shows an alternation between V1 and V2 orders of the following kind:

- (45) *kantu* n□kkut segja oss til Hákonar jarls? (Old Norse)  
 Can2s.-you some tell us to Hakon earl  
 “Can you tell us about Earl Hakon?”
- (46) *hvárt* grært þú nú, Skarpherðinn?  
 Whether cry.2s you now Skarpherdin  
 “Are you crying now, Skarpherdin?” [Faarlund 2008; 226: (79)a.; (81)a.]

<sup>21</sup> Di Domenico (2004), p. 89, observes that:  
 “Following a suggestion of Valentina Bianchi (p.c.) I assume that the lack of FinP is connected to [...] the fact that they [imperative clauses] cannot be embedded [...]. We can define imperative clauses ‘non-placed expressions’. As such they are identified with (rather than related to) the speech event.”

The question word *hvárt* can be translated with “whether” and is often accompanied by a second option introduced by *eða* (or), but “the second part of the disjunctive can be omitted. This word then comes to function as an introduction to a regular sentence question” (Faarlund (2008), p. 226). Therefore we can assume that when *hvárt* is absent there is a null OP in the specifier of the verb: the alternation thus concerns the overt/covert realization of the operator. In contemporary Icelandic, then, the overt form has been dropped in favor of a V1 order which can be understood as null OP-V.

As for the position targeted by the verb I refer to a recent debate in the cartographic approach. Benincà and Poletto (2004) argue, *contra* Rizzi, that the lower (criterial) position above FinP can be identified with a Focus, not with a Topic. Supposing Benincà and Poletto (2004) are right, the expectation is that a Focus lower than a [+interrogative] V in its criterial position is ruled out, because movement of the interrogative OP and the Focus-OP would trigger minimality effects. The expectation is borne out by facts:

- (47) a. \*Fór **til Íslands** Jón/hann? (Icelandic)  
 Went to Iceland J/he  
 b. Fór Jón/hann til Íslands?  
 Went J/he to Iceland  
 c. \*Hefur **margar bækur** Jón/hann lesið?  
 Has many books J/he read  
 d. Hefur Jón/hann lesið margar bækur?  
 Has J/he read many books

Because the 2<sup>nd</sup> person subject pronoun optionally comes as enclitic on the verb, [+interr]V is incompatible with the activation of Focus (Benincà (2006), cf. above).

- (48) a. Ertu búinn?  
 Are-you ready  
 b. Heitirðu Jón Jónsson?  
 Are-called-you J.J.  
 ‘Is your name J.J.?’

The target position of the verb must be located higher than the Focus position, presumably on the head of an Interr- projection, and is preceded by a null OP undergoing criterial movement.

### 3.4 Conditional clauses

In section 1. I have also considered another instance of V1 order which is also found in embedded context, namely conditional clauses without subordinating particles. The embeddability of conditionals is explained if we agree with Allan (2006). He argues that most hypotheticals (H) “do not contrast with T, Q, and I but fall within dependent clauses in the scope of one of these PI operators whose illocutions they modify to apply to hypothetical worlds [...]” (p. 26). How come, then, that these clauses, which are not equivalent to Q and I (the marked values of main clause-types), may have V1 order? First of all notice that the same clause-typing function can be optionally carried out in a different way. Hypothetical clauses are marked as *irrealis* [-R]. It is worth pointing out that the distinction between the categories of *realis* and *irrealis* is not always clear cut, but there is a lot of crosslinguistic overlapping between the two. According to Allan (2006) a. o., at the opposite sides of the *realis/irrealis* split are



subjunctive to the lack of factivity in a clause<sup>22</sup>. For instance, as imperative is not possible in embedded clauses, subjunctive is instead used with the semantic function of marking the clause as optative (see also Platzack, to app.).

Notice that the presence of a complementizer, in complementary distribution with the verb in first position, is crucial for typing a clause as (H). Therefore we can consider the if-comp in the examples above as the overt PI (H) operator indicated by Allan (2006). Given the interpretive properties of verb initial conditionals, we can postulate the presence of a null OP as for other V1 clauses (see imperatives and interrogatives above). Because of the optionality between the order (i) IF-Subj-V and the order (ii) OP-V, we can assume that the verb targets the high CP head (Force<sup>o</sup>), which can alternatively be specified by the conditional complementizer in order (i). The complementary distribution of V1 and conditional complementizers is then explained with the intervention effect between the A'-OP-movement of the overt subordinating conjunction (=if) and that of the conditional null OP which must end up in preverbal position (as for the other cases of V1 mentioned above).<sup>23</sup>

The presence of OP-movement in V1 conditional clauses, as well as in imperative and interrogative clauses is supported by the incompatibility with topicalization. Haegeman (forthcoming) argues that: “what is needed to licence argument fronting, then, is not simply the presence of Force, but rather [...] ‘declarative’ force” (p.7). In her paper, Haegeman distinguishes between two types of conditional: peripheral conditional (‘echoic’, in the sense that they can “echo Q-propositions about a nonfactual world..” (Declerck and Reed 2001:83)) and central conditionals. She shows that only central conditionals are incompatible with argument fronting because they involve OP movement:

(52) a. \*If water you heat up to 100° C, it will boil.

b. If some precautions they have indeed taken, many other possible measures they have continued to neglect

[Ex. b. is from Haegeman (forthcoming), 22: (44)]

V1 counterfactuals, in Scandinavian as well as in English, belong to the central conditional class, and they do not allow argument fronting:

(53) \*Had some precautions they taken, such consequences would have been avoided

As a consequence the presence of a null OP undergoing A'-movement to the preverbal position seems a correct postulation.

### 3.5 Narrative Inversion

As for conditional, interrogative and imperative clauses, the periphery of narrative style clauses does not allow topicalization. The presence of a topicalized argument or adjunct would make a V1 clause ungrammatical, as shown in the following configuration:

(54) \*V1 Top (subj)...

<sup>22</sup> This correlation is not uncontroversial, see data on indicative/subjunctive alternation in Thráinsson (2007); chap. 8.

<sup>23</sup> I thank Liliane Haegeman (p.c.) for useful insights on this aspect. For a description of intervention effects triggered by null OP movement see her forthcoming paper (see References).

This is indeed borne out by facts:

(55) a. \*Koma að stórum helli þeir og.. (Icelandic)

Come to big cave they and..

b. Koma þeir nú að stórum helli og...

Come they now to big cave and..

[(55)b. is from Thráinsson 2007, 29: (2.22)]

However narrative V1 is compatible with the possibility of an adverb like *þá* (*then*) or *síðan* (*since*) intervening between subject and verb in Old Norse:

(56) Lét *þá* Sigríðr senda eftir þóri hund til Bjarkeyjar (Old Norse)

Let then Sigrid send after Thori hound to Bjarkey

‘Then Sigrid and Thori hound sent for from Bjarkey’

(57) Fóru *síðan* hvárir-tveggju leiðar sinnar

Went since each-two ways their

‘Then both of them went their own way’

[Faarlund 2008; 231: (91)d.]

Icelandic Narrative V1 can be grouped together with those V1 declaratives where all the obligatory elements are present (OEP) accounted for by Mörnshjö (2002). She identifies that class of declaratives with clauses having a “temporal/spatial/logical relation to the preceding linguistic or situational discourse [...] established by means of a connective adverb. [...] OEP should be analysed as XVS sentences [...] having a phonetically non-realized frame topic [...] in Spec-CP” (p. 90). In a narrative inversion context like (58) below, the null frame topic would be the adverb *så* (*so*, *then*):

(58) Så gick han och jag la ner min snöboll så här. ∅ Tog jag upp en snöboll till..

So went he and I put down my snoboll dm took I up one snowball more...

‘So he went and I put down my snowball like this. Then I took up another snowball...’

[(Spoken Swedish), Mörnshjö (2002), 89: (7:2)]

Following Mörnshjö’s (2002) proposal, also Icelandic Narrative inversion could be analysed as an XVS order, with a silent frame topic in first position. Along the lines of Benincà (2005) the “frame field” would be located in the higher part of the CP and host Scene Setting adverbs and Hanging Topics, as in (50) below:

(59) [<sub>Force</sub> C°]{<sub>Frame</sub> [ScSett][HT] C°}{<sub>Topic</sub> ...}{<sub>Focus</sub>...}[<sub>Fin</sub> C°]

In a structure of this kind, the frame-topic identified by Mörnshjö is located in a specific position at the edge of CP where it can remain phonologically silent. According to this perspective, however, sentences like (56) or (57) are problematic because the relevant adverbs (i.e. *þá* (*then*) or *síðan* (*since*)), which are expected to be in the preverbal, scene setting projection, indeed follow the verb.

My proposal is that, as for other marked V1 orders, Narrative V1 clauses have a silent OP in preverbal position, moving to the relevant specifier position in the frame field

for scope reasons. In cases as in (56) and (57) where the frame-topic follows the verb, an null OP is associated to the AdvP and moves to the frame field, with extraction from a lower position. According to this analysis, the OP carries [+scene setting] or [+HT] features, whereas the stranded adverb may just check his feature in ModP and be frozen there<sup>24</sup>. A possible derivation would be:

(60) [<sub>ScSett/HTP</sub> OP] V [<sub>ModP</sub> [<sub>AdvP</sub> t<sub>OP</sub> Adv]] [<sub>Subj2P</sub> Subj]...

Notice that OP movement makes the periphery unavailable to further topicalizations, even in cases like (55)b. where the adverb *nú* follows the subject and OP-movement must take place from a lower position. The association of adverbs and operator is hypothesized along the lines of Haegeman (forthcoming). She argues, for instance, that French temporal clauses displaying Stylistic Inversion are derived by operator fronting.

As for the interpretation of this type of clauses, it is clear that they are much more context-dependent than standard V2 declaratives with a subject or a lower topic in first position.

#### 4. Conclusion

I have argued that the labels V1 and V2 cover inappropriately heterogeneous phenomena. What is known as standard V2 is found mainly in declaratives and is possible in subordinate clauses, but V2 clauses have different pragmatic properties: topic-initial V2 seems to bear the pragmatic value of an assertion, whereas this is not necessarily the case for subject-initial V2, at least in Icelandic.<sup>25</sup>

V1 orders have a different distribution. It is worth remarking that “V1” is a highly generic term which tells little about the underlying structure. Indeed V1 orders can be distinguished as follows:

(a) Disguised V2:  $\emptyset$  Expl/SF – V (in Icelandic subordinate clauses)

(b)  $\emptyset$  OP-V orders: in imperatives, Y/N interrogatives, conditionals and narratives<sup>26</sup>.

The order in (a) is analyzed syntactically as V2, whereas it cannot be unitarily identified semantically as it may lack the assertive character of V2 declarative clauses<sup>27</sup>.

The orders in (b) can be distinguished from other cases of V2 because of their specific discourse related properties, as indicated by verb raising to a higher head in CP and the criterial movement of a null OP in first position. Since the OP has no phonological realization, verb movement is indispensable for satisfying the criterion.

<sup>24</sup> In case the adverb does precede the verb Criterial Freezing (see Rizzi (2004) and (2007)) is still respected. What happens in this case is that the adverb still checks its feature on Mod, but the larger phrase containing it is pied-piped to the frame field in order to check other features.

<sup>25</sup> For an interpretation of V2 clauses see Julien (2007) a.o.

<sup>26</sup> According to Mörnjö (2002), narratives would only be a subset of (spoken Swedish OEP) V1 clauses. A better identification would be “context-dependent” V1.

<sup>27</sup> It is in fact related to morphological and syntactic properties found in Icelandic, but missing in Mainland Scandinavian. Because the  $\emptyset$  Expl-V configuration is found in extraction contexts, and does not depend on the assertivity degree of the clause, I consider it from a purely syntactic perspective.

## References

- Allan K. (2006) "Clause-type, primary illocution, and mood-like operators in English", *Language Sciences* 28, 1-50.
- Belletti A. (1999) "Italian/Romance clitics: Structure and derivation", in Van Riemsdijk H. (ed.) (1999) *Clitics in the languages of Europe*, Mouton de Gruyter, Berlin.
- Benincà P; Poletto C. (2004) "Topic, Focus and V2: defining the CP sublayers", in Rizzi, L. (ed.) (2004) *The Structure of CP and IP. The Cartography of Syntactic Structures*, vol.2, New York, Oxford University Press.
- Benincà P. (2006) "A Detailed Map of the Left Periphery of Medieval Romance". In Zanuttini R.; Campos H.; Herburger H.; Portner P. (eds.) *Crosslinguistic Research in Syntax and Semantics. Negation, Tense and Clausal Architecture*. Georgetown University Press, Washington 2006, pp. 53-86.
- Bianchi V. (2006) "On the syntax of personal arguments", *Lingua* 116: 2023–2067.
- Cardinaletti, Anna (2004), Toward a Cartography of Subject Positions. In *The Structure of CP and IP. The cartography of syntactic structures*, vol. 2, ed. by Luigi Rizzi, 115-165. Oxford: Oxford University Press.
- Cardinaletti A., Roberts I. (2002) "Clause structure and X-second" In Cinque G., *Functional Structure in DP and IP. The Cartography of Syntactic Structures*, volume I, OXFORD, Ed. Oxford University Press, pp. 123-166.
- Cinque G. (1999) *Adverbs and Functional Heads. A Crosslinguistic Perspective*. Oxford University Press, Oxford.
- Di Domenico E. (2004) "Placed, non-placed and anaphorically placed expressions", *Italian Journal of Linguistics*, 16.1, pp. 63-105.
- Faarlund J. T. (2008) *The syntax of Old Norse*, Oxford University Press.
- Haeberli E. (2002) *Features, Categories and the syntax of A-positions. Cross-Linguistic Variation in the Germanic Languages*. Kluwer, Dordrecht.
- Haegeman L. (2004) "Topicalization, CLLD and the left periphery", in Shaer, B., Frey, W. et C.
- Haegeman L. (forthcoming) "The internal syntax of adverbial clauses", Université de Lille 3, ms.
- Maienborn (Réds.), *Proceedings of the Dislocated Elements Workshop, ZAS Berlin, November 2003, ZAS Papers in Linguistics*, 35, Vol 1, 157-192.
- Holmberg A. (2000) "Scandinavian Stylistic Fronting: how any category can become an expletive". Draft.
- Holmberg A. and Platzack C. (1995) *The role of Inflection in Scandinavian Syntax*, Oxford University Press.
- Hooper J. B. (1975) "On assertive predicates", J. P. Kimball (ed.), *Syntax and Semantics*, vol. VI, Academic Press INC. Indiana University, Bloomington, Indiana.
- Hróarsdóttir Þ.; Wiklund A. L.; Bentzen K.; Hrafnbjargarson G. H. (2007) "The afterglow of verb movement" in *Working Papers in Scandinavian Syntax* 80, pp. 45-75, Lund University.
- Julien M. (2007) "Embedded V2 in Norwegian and Swedish" in *Working Papers in Scandinavian Syntax* 80, pp. 103-161, Lund University.
- Maling J. (1990) "Inversion in embedded clauses in modern Icelandic", in *Modern Icelandic Syntax, Syntax and Semantics* 24, pp. 71-91, Academic Press Inc.
- Mörnsjö M. (2002) "V1 declaratives in Spoken Swedish", PhD thesis, Lund University.
- Platzack C. (to app.) "Embedded Imperatives", ms., Lund University.

- Platzack C. (forthcoming) "Backward Binding and the C-T Phase: A Case of Syntactic Haplology", ms. University of Lund.
- Platzack C.; Rosengren I. (1998) "On the subject of imperatives: A minimalist account of the imperative clause", *The Journal of Comparative Germanic Linguistics* 1:177-224.
- Poletto C., (2002) "On V2 and V3 sequences in Rhaetoromance", in *Syntactic Microvariation*, Sjef Barbiers, Leonie Cornips e Susanne van der Kleij (eds).
- Rizzi L. (1997a), "Residual Verb Second and the Wh-Criterion" in Belletti A.; Rizzi L. (eds.), *Parameters and Functional Heads*, Oxford University Press.
- Rizzi L. (1997b), "The fine structure of the left periphery" in L. Haegeman, ed, *Elements of Grammar*. Kluwer, Dordrecht, pp. 281-337.
- Rizzi L. (2004) "On the Form of Chains: Criterial Positions and ECP Effects" ms, University of Siena.
- Rizzi L. (2005) "Grammatically-Based Target-Inconsistencies in Child Language" ms, University of Siena.
- Rizzi L. (2007) "On Some Properties of Criterial Freezing", *STiL - Studies in Linguistics*, Vol.1
- Rizzi L., Roberts I. (1989) "Complex inversion in French", ms., University of Siena.
- Rizzi L., Shlonsky U. (2006) "Strategies of Subject Extraction", ms. University of Siena.
- Roberts I. (2004) "The C-System in Brythonic Celtic Languages, V2, and the EPP" in L. Rizzi (ed) *The Structure of IP and CP*. Oxford/New York: Oxford University Press.
- Roberts I., Roussou A. (2002), "The extended projection principle as a condition on the Tense dependency" in P. Svenonius, ed., *Subjects, Expletives and the EPP*, p. 125-55, Oxford University Press.
- Rögvaldsson E.; Thráinsson H. (1990) "On Icelandic word order once more" in *Modern Icelandic Syntax, Syntax and Semantics* 24, pp. 3-39, Academic Press Inc.
- Sigurðsson H. A. (1990) "V1 declaratives and verb raising in Icelandic" in *Modern Icelandic Syntax, Syntax and Semantics* 24, pp. 41-69, Academic Press Inc.
- Speas P.; Tenny C. (2003) "Configurational properties of point of view roles", in Di Sciullo, (ed.), *Asymmetry in Grammar*, vol I, Benjamins 2003, 315-344.
- Starke M., (2004) "On the inexistence of specifiers and the nature of heads", in Belletti, A. (ed.) *Structures and Beyond. The Cartography of Syntactic Structures, Volume 3*. New York, Oxford University Press, pp. 252-268.
- Thrainsson H., (2007) *The syntax of Icelandic*. Cambridge University Press.
- Thráinsson H.; Petersen H. P.; Jacobsen J.; Hansen Z. S. (2004) *Faroese: an overview and reference grammar*, Tórshavn : Føroya Fróðskaparfelag.
- Vikner S., (1995) *Verb movement and expletive subjects in the Germanic languages*. New York: Oxford University Press.

# A Bilingual Treebank (ITA-LIS) suitable for Machine Translation: what Cartography and Minimalism teach us

Cristiano Chesi<sup>\*</sup>, Gianluca Lebani<sup>+</sup>,  
Margherita Pallottino<sup>\*</sup>

<sup>\*</sup>CISCL - University of Siena, <sup>+</sup>CIMEC - University of Trento  
chesi@media.unisi.it, gianluca.lebani@unitn.it, margherita.pallottino@yahoo.com

In this paper we describe some technical and theoretical aspects related to a manually aligned bilingual treebank Italian (ITA) – Italian Sign Language (LIS) provided with both constituency and dependency annotation (Siena University Treebank, SUT). We briefly discuss the linguistic rationale behind the feature set and the dependency/constituency structure we adopted. Moreover we discuss the tool we used to annotate, semi-automatically, the treebank that, in the end, will be evaluated qualitatively with respect to a specific Transfer-Based Machine Translation (TB-MT) task.

## 1. Introduction

The effectiveness of fine grained grammatical distinctions at morpho-syntactic and semantic level is especially relevant cross-linguistically: the discussion of the aligned, bilingual Treebank, Italian (ITA) - Italian Sign Language (LIS) presented in these pages aims to provide linguistically motivated answers to two main questions:

1. are standard tagsets (e.g. Eagles, Monachini 1995, Tamburini 2007) sufficiently rich to account for (quasi-)deterministic rearrangement of constituents in a Transfer-Based Machine Translation (TB-MT) task?
2. is the phrase structure predicted by current linguistic frameworks (e.g. Minimalism, Chomsky 1995-2005, and Cartography, Belletti 2004, Cinque 2002 and Rizzi 2004) coherent with pervasive corpus-attested syntactic constructions and suitable for massive transformations between two fairly different languages?

In the first part of this paper (§2) we present the structure and the format we adopted to annotate the treebank, briefly discussing the set of features we used to code functional (e.g. topic, focus) and non-manual aspects (e.g. facial expression, movement velocity); in the second part of this paper we justify some radical linguistic assumption (e.g. head-marked, mainly flat tree-structures) on the basis of recent advances of Minimalist and Cartographic approaches. We tried to implement a version of tree structure that productively suits, as efficiently as possible, a TB-MT task: this means that the translation process is based on a structural reordering/pruning

procedure, driven by the leading idea that nothing in the structure must neither be created nor destroyed, but simply rearranged or scattered/collapsed (§3.3). This approach does not guarantee always an High Quality MT, but it results in fairly acceptable translations and it presents appealing computational advantages (§4).

## 2 The bilingual Treebank ITA-LIS

Despite the difficulty in defining a standard for coding a full transcription of a signed language<sup>1</sup>, in building the ITA-LIS Treebank, we faced the problem of accounting, in a compact and meaningful way, for a complex parametric setup in order to exploit the annotated data from a Principles and Parameters (Chomsky 1981) point of view:

<i>ITA</i>	<i>LIS</i>
Head initial	Head final
Verb raising	No verb raising
Obligatory wh-movement	No wh-movement
Poor relative/PPs extraposition	Rich (obligatory?) relative/PPs extraposition
Rich clitic system	No clitics
No classifiers	Rich classifier system
Gender/number agreement	Spatial agreement

**Table 1.** Macro-parametric differences between ITA and LIS

Despite these differences there are also similarities, for instance, they are both *pro-drop* languages, they seem to allow (at least superficially) for a certain degree of variability in word order, they both show some rightward Heavy NP-shifting preferences. These parametric settings require a rich collection of empty elements (e.g. null subjects, traces, ellipses) and a consistent/computable solution to indicate referents and dependencies without losing any relevant linguistic information (e.g. (hanging-)topic/focus, argument doubling etc. Belletti 2004).

### 2.1 Corpus composition

The first release of the corpus is composed by 1018 Italian sentences extracted from public broadcast television news and translated/glossed to Italian Sign Language. 27 editions have been transcribed: 18 special editions written on purpose for LIS translation (shorter sentences, less complex structures<sup>2</sup>) plus 9 standard afternoon editions (standard Italian, without any special attention to the translation task). The ITA section of the Treebank has 17122 tokens (5391 distinct lexical items), while the LIS section counts 11056 tokens (3400 distinct lexical items). The asymmetry is due to the absence of various functional elements (e.g. articles, prepositions, auxiliaries etc.) as distinct lemmas in LIS (these elements are all coded by suprasegmental features, e.g. facial expressions) and to arguments/modifiers incorporation (e.g. the ITA equivalent of “to put a book on the shelf” is translated with the LIS equivalent of “to shelve a book”). The corpus is annotated using XML (Mana and Corazzari 2002), which ensures portability and permits a standard, flexible and human readable multilevel annotation. Structures can however readily (and univocally) be converted

<sup>1</sup> See Bergman et al. (2001).

<sup>2</sup> Roughly speaking, standard arguments order ((S)VO) is fairly maintained, words used are often high frequent lemmas, no relatives are employed and, in general, minimal NP modification is used (always locally); no parenthetical or long/run on sentences object are present.

into PENN (constituency) and TUT (dependency) format (Marcus et al. 1993, Bosco et al. 2000) to the detriment of some relevant linguistic distinction.

## 2.2 Signs, words and features

At the morphosyntactic level any single token (PoS) is enclosed under the tag <word> as follows:

- (1) <word id="6" cat="V.ind.pres" lemma="essere" agree="3.s" role="head" subcat="copula"> è </word>  
(token: è=is, lemma: essere=be)

According to the Document Type Definition (DTD) we adopted (Appendix A), this is the list of attributes that can be specified for the tag <word> and their potential value:

- (2) *id* it is an (unique) identity number for the node (nodes are recursively numbered in each tree from top to bottom, left to right);
- ref* traces, ellipses, pronominal elements and co-referent nodes in general have this attribute filled with the *id* of the reference node (it can be a relative specification: e.g. S-1 means the previous sentence, H-1 means the previous head, according to the numbering scheme just mentioned);
- cat* it is the classical PoS tag. Main tags are *N(ouns)*, *V(ebrs)*, *A(djectivals)*, *ADV(erbials)*, *D(eterminers)*, *Q(uantifiers)*. Each of them can be further subcategorized according to cartographic features (Appendix B): e.g. *V.ind.pres* classifies a *verbal* element, in *indicative* (vs. *subjunctive* vs. *infinitive* etc.) modality, at *present* (vs. *past* vs. *future* etc.) tense; *N.comm.count.inanim* classifies a *common* (vs. *proper*) nominal element, *countable* (vs. *mass*), *inanimate* (vs. *animate* vs. *animate-person*);
- subcat* it expresses obligatory thematic dependencies: *intrans(itive)* requires a subject grammatical position to be filled with an argument associated to the *agent* theta-role; *unacc(usative)* requires the grammatical subject position to be filled with a *patient* theta-role; *trans(itive)* requires two arguments positions subject and object to be filled respectively with an *agent* and a *patient* theta-role etc. (Appendix B); according to the Uniform Theta-role Assignment Hypothesis (Barker 1997) we do not need any further feature to identify univocally any dependency at the argumental level;
- lemma* it is the dictionary form of the token;
- role* three main dependency roles are allowed: (constituent) *head*, *arg(ument)* or *adj(unct)*; *arg* can be *subj(ect)*, *obj(ect)* *pred(ictive object)*, *ind(irect object)* (as in Bosco et al. 2000); *adjs* (and other dependencies) are listed in Appendix C;
- lp* left-peripheral (in the sense of Rizzi 1997) features such as *topic* and *focus* and other “edge” phenomena (Chomsky 2005) such as *expletive*, *doubling*, *extraposition* etc.;
- expr* (only in LIS) it expresses supra-segmental features such as eyebrows position (*eye-up*, *eye-down*), intensity of the sign (*slow*, *fast*, *minimize*, *exaggerate*) and the classifier system (*keep-support-hand*, *gaze-to-sign*, *cl.shape*, *cl.move*, *cl.position* etc. this is when a sign is not signed as reported in the dictionary but it “agrees” in shape, position etc. with another sign, Appendix D);
- agree* (only in ITA) person/gender/number features (e.g. *3.m.s* means third person, masculine, singular); (only in LIS) position features, organized by relevant spatial location such as eyes, mouth, chest etc. (e.g. *body\_contact.mouth.left* means that the token is signed touching the mouth on the left);

*sem* it specifies a reference to the related MultiWordnet sense/synset (Bentivogli et al. 2002; if nothing is specified, the first synset associated to the lemma is picked out; this feature is still under implementation).

*Id* (used for co-reference and alignment), *cat* and *lemma* are obligatory, all the other attributes are optional.

Non-terminal nodes are coded with the tag <node> and they share the same attributes of the tag <word>. For mnemonic (and backward-compatibility issues) we used three standard categories to fulfill the “cat” attribute in <node>: NP (Nominal Phrase), VP (Verbal Phrase), AP (Adjectival/Adverbial Phrase). Such a simplification is not innocent, but it seems to be empirically and computationally tenable (Chesi 2007).

This is a sample of a tagged sentence:

(3) più difficile    la situazione in Senato domani  
 more difficult    the situation in Senate tomorrow  
 “tomorrow the situation in Senate will be more difficult”

```
<node cat="VP" id="2008-01-23.3" role="head">
  <node agree="f.s" cat="AP" id="1" lp="topic" role="arg.predobj">
    <word cat="ADV.streng" id="2" lemma="più">più</word>
    <word agree="f.s" cat="A.qualif" id="3" lemma="difficile">difficile</word>
  </node>
  <word agree="3.s" cat="V.ind.fut" id="4" lemma="essere" role="head" subcat="copula"/>
  <node agree="f.s" cat="NP" id="5" role="arg.subj">
    <word agree="f.s" cat="D.art.def" id="6" lemma="la">la</word>
    <word agree="f.s" cat="N.comm.count.inanim" id="7" lemma="situazione" role="head">
      situazione
    </word>
  </node>
  <node cat="NP" id="8" role="adj.loc">
    <word cat="P.loc" id="9" lemma="in">in</word>
    <word agree="m.s" cat="NE.org" id="10" lemma="senato" role="head">senato</word>
  </node>
  <word cat="ADV.time" id="11" lemma="domani" role="head">domani</word>
  <word cat="END.comma" id="12" lemma=",">,</word>
</node>
```

And this is the LIS translation of the very same sentence:

(4) domani camera-Senato    situazione    difficile    più  
 tomorrow room-Senate    situation    difficult    more

```
<node cat="VP" id="2008-01-23.3" role="head">
  <word cat="ADV.time" id="11" lemma="domani" role="head">domani</word>
  <node cat="NP" id="8" role="adj.loc">
    <word cat="NE.org" id="10" lemma="senato" role="head">camera senato</word>
  </node>
  <node cat="NP" id="5" role="head">
    <word cat="N.comm.count.inanim" id="7" lemma="situazione" role="head">situazione</word>
  </node>
  <node cat="AP" id="1" role="adj.manner">
    <word cat="A.qualif" id="3" lemma="difficile">difficile</word>
    <word cat="ADV.limit" id="2" lemma="più">più</word>
  </node>
  <word agree="3.s" cat="V.ind.fut" id="4" lemma="essere" role="head" subcat="copula"/>
</node>
```

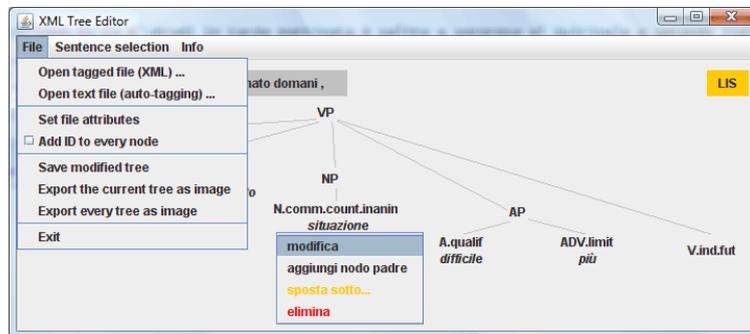
Notice that, despite their equivalence with the corresponding Italian words, characters within the tag <word> are simply indices that points, univocally, to a dictionary entry (that is, in fact, not simply a gloss, but a set of instruction to move an avatar, Bartolini et al. 2006); their relation with the corresponding Italian word is expressed only by the *id* field. In the case of homographs, *sem* is used to retrieve the correct item from the bilingual lexicon.

### 2.3 The annotation procedure

The morphosyntactic annotation consists of assigning to every token a <word> tag with the above mentioned features fully specified; as in other constituent-based annotations, words are grouped under the tags <node> to identify phrases. A well formed tree has one single node at the top. Every <node> has to bear a “cat” and a “role” specification and every well-formed node must be headed, which means it has one, and only one, <node> or <word> child with “role” equals to “head” (this can be phonologically null as the copula in (3)).

The XML structure is manipulated using a Java tool (XML Tree Editor<sup>3</sup>):

(5)



This tool can operate getting in input a text file (one sentence per line): it assigns to every sentence, automatically, a potential structure using a minimalist parser (based on Chesi 2007). Every structure that is automatically created can be graphically edited (nodes can be created, deleted, moved, replaced and every attribute can be modified). To guarantee consistency and reliability during the Treebank building, the grammar used by the parser is enhanced by rooted/terminal and auxiliary trees (as in Tree Adjoining Grammars (TAGs), Frank and Kroch 1995), i.e. previously tagged portions of sentences are ordered by frequency and used to help the parser retrieving the most likely structures.

### 3. Linguistic considerations

Evaluation of computational linguistic resources for Italian (EVALITA 2007) recently proposed a gold standard for the Italian PoS tagset (Tamburini 2007), and for the Constituency/Dependency classes/relations (Bosco et al. 2000) creating a lowest common denominator that includes widely used morphosyntactic/functional classes (e.g. PENN tag set). These standards are sufficiently rich and flexible to account for a wide range of linguistic phenomena, but not for a (quasi-)deterministic MT task between two languages parametrically as different as ITA and LIS (Table 1). The goal of this section is to highlight the major linguistic/computational necessities that induced a refinement (as minimal as possible) of such standards.

<sup>3</sup> The tool is freely available at <http://www.ciscl.unisi.it/ricerca.htm>

### 3.1 Bare Phrase Structure (BPS)

Within the Minimalist framework (Chomsky 1995-2005), many linguists assumed that lexical elements directly create constituency relations without projecting any non-terminal category (*Bare Phrase Structure* hypothesis, *Inclusiveness Condition*, Chomsky 1995); such a grammatical intuition (henceforth BPS) has been shown to be sound (Stabler 1997) and parsable (Harkema 1997) and it would dispense our grammar from using non-terminal (constituency) tags (e.g. NP, PP etc.) at all:

(6) <i>PENN</i> -like	BPS (classic)	BPS (adopting Abney 1987)
(NP (ART the (N dog))	(dog (the) (dog))	(the (the) (dog))

The selecting head (the noun in the classical X-bar generative theory, the determiner, after Abney's influential proposal) projects over the selected element. In this case, we would expect any lexical item to be marked for selection within the lexicon. Since the selecting element is always a head (by definition) we guarantee that the projecting node is the head of the phrase (i.e. an NP node is in fact the projection of a N head, (6).BPS-classic). Notice that while in standard minimalist approaches the projection system results in a binary operation (i.e. *merge*) that strictly produce binary branching trees, we assume that the constituents can have more than one sister. In the following paragraph we will defend the idea that even if we do not assume a binary branching constraint (Kayne 1983), binary branching structures can be readily retrieved from the proposed tree and hence, BPS-related assumptions can be kept. On the other hand, as introduced in §2.2, the fact that we mark nodes as VPs, NPs or APs (*cat* feature in our xml structure), is not against the BPS idea since we can unambiguously track the projecting heads node by node (this is so because every node has exactly one single head).

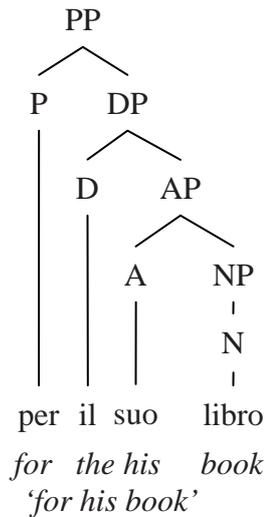
### 3.2 Cartography of functional projections

From an empirical point of view flat trees have often been challenged in literature since non-predictive with respect to many relevant phenomena (e.g. coordination and gapping, binding etc. Kayne 1983); on the other hand, having flat structures reduces the ambiguity in the lexicon<sup>4</sup> and allows us to retrieve, with a minimal search, every relevant feature in a given phrase (Adger 2007, Chesi 2007). This tension seems to be solvable if we accept the cartographic hypothesis (e.g. Cinque 2002): order and hierarchy are in fact tightly related and universally constrained; superficial "free" word order is the epiphenomenon of a sequence of movements that target functional/peripheral (Rizzi 1997) positions. The attribute *lp* (§2.2) exactly expresses these "extra" features and prevents us from implementing a full projection of every functional node, including their potential landing site in the left periphery: about forty distinct positions in the functional VP domain (Cinque 2002) can be collected under the same node (i.e. these forty nodes are optionally present and, when present, all dominated by the same VP node, without requiring any selection mechanism within the lexicon) keeping their cartographic (sub)category (e.g. for the adverbial domain: ADV.manner, ADV.temp, ADV.neg, ADV.asp etc.). The example below shows the tree-translation between standard approaches and the one implemented within the Siena University Treebank (SUT).

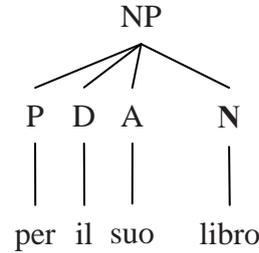
---

<sup>4</sup> This is because of the selection mechanism proposed by Chomsky and formalized by Stabler: if each node has to be marked for selecting its sister category, having or not having an optional adjective, for instance, between the determiner and the noun, would duplicate the number of determiners: one that selects the adjective and another one that does not (Chesi 2007).

(7) Standard tree



SUT simplified tree



The subcategorization of main functional categories (e.g. adverbs) directly expresses the dependency (i.e. the xml attribute “role”) of the <word> element within the phrase with respect to the phrase head (otherwise, within the <node> tag, the “role” attribute is specified as a specific “adjunct” category; this is, for instance, the case of adverbial PPs, that, following our guidelines, are simply tagged as NPs). Building an extensive treebank with such information could then turn out to be a precious tool also for evaluating quantitatively the predictions of the cartographic approach.

#### 4. Evaluation of the Treebank from a TB-MT perspective

The goal of this paper was to present in a fairly intuitive and compact way the process of treebank building and the theoretical assumptions that justified certain choices. In this final section we evaluate in which sense the standard we proposed is different from the alternative Eagles/EVALITA tagset (Monachini 1995, Tamburini 2007) and TUT set of dependencies (Bosco et al. 2000) (§4.1). Then we will verify if such refinements are productive when we try to extract alignment rules from the treebank that should be suitable for a transfer-based MT task (§4.2).

##### 4.1 (Minimally) different standards

Despite main categories such as Verbs, (Pro)Nouns, Articles, Prepositions, Adjectives, Adverbs are consistently adopted following the standard discussed in Monachini (1995) and Tamburini (2007), few differences at sub-categorical and functional level are worth to be reported and justified: as for the functional level, for instance, articles are collected under the PoS D(eterminer) together with quantifiers and demonstratives (see Appendix B for a full list) in order to capture some cartographic intuition (“determiner” vs. “adjectival” field); (subordinating) conjunctions as well as prepositional subordinators are included under the PoS C(omplementizer) again to comply with cartographic ideas (“left-peripheral” Vs. “inflectional” field). On the sub-categorization side, the table below highlights some substantial expansion of the proper name and adjectival classes (again, refer to Appendix B for the whole picture):

Category	Eagles	SUT
Proper Names	SP@NN	N.prop.anim.person.last/first N.prop.inanim.city ...
Adjectival forms	A (adjective), AP (possessive adjective)	Adj.deict, Adj.dem, Adj.nation, Adj.num.ord, Adj.num.card, Adj.poss, Adj.qualif ...

These distinctions are mainly justified by the task we are dealing with: (quasi-)deterministic rearrangement of constituents in a TB-MT task; these distinctions are in fact crucial since names of persons or cities have to be prefixed in LIS by the correct classifier, “person” or “city” respectively. On the other hand, subcategorizing adjectival forms gives us the opportunity to reorder correctly (in standard contexts) these elements in LIS:

(8) Adj.num < Noun (head) < Adj.poss < Adj.dem/deict < Adj.nation < Adj.qualif

On the dependency side, a differently structured set of relations (according to the categories of *head*, *arguments* and *adjuncts*) allows us to correctly predict phenomena such as relative extrapositions or PP clefting in LIS which would be less transparent under the distinction *functional arguments* (e.g. locatives are considered arguments under the label of “indirect complements” much as the subject and the direct object in Bosco et al. 2000).

The necessity for such distinctions becomes clear analyzing the head directionality parameter (table 1, §2): reordering is massive between ITA and LIS and the linguistic assumptions we discussed allows us to deal in a computationally elegant way with this, since relevant constituents are readily accessible within just one level of inspection. This allows us, for instance, to capture in-corporation (9), ex-corporation of arguments/adjuncts (10) analyzing only immediate constituents within a single phrase ( $\theta$  expresses the thematic requirements of the head;  $i\theta$ , indicates an internal, lexical, satisfaction of such requirement):

- (9) (ITA) [VP [head  $\theta_l$ ] [arg.obj]] → [VP [head  $i\theta_l$ ]] (LIS)  
 [VP mettere [arg.obj una firma]] → [VP [head firmare]]  
 put a sign sign
- (10) (ITA) [VP [head  $i\theta_l$ ]] → [VP [arg.obj] [head  $\theta_l$ ]] (LIS)  
 [VP dimettere ] → [VP [arg.obj carica] [head rinunciare]]  
 dismiss position leave

Standard argument/adjuncts reordering (11) as well can be readily decided locally without inspecting further constituents:

- (11) (ITA) [VP [arg.subj] [v-head] [arg.obj][adj.temp]] →  
 (LIS) [VP [adj.temp] [arg.subj] [arg.obj] [v-head]]

Then a richer (cartographic) subcategorization allow us to extract from the corpus non-ambiguous reordering rules of adjuncts:

- (12) (ITA) [VP [v-head] [adj.manner] [adj.matter]] → [VP [adj.manner] [v-head] [adj.matter]] (LIS)

It should be clear then that having more fine grained categories and features allows us to extract more specific transfer-based rules. There is however a drawback in freely multiplying features and categories: the data required to extract statistically reliable information would grow considerably. We attempted to solve this problem using an

hierarchical approach to categories/features expansion: information on the distribution of upper level categories are much more easily retrievable, then using this data guarantees a fairly robust TB-MT approach. On the other hand, accuracy is pursued rewarding sub-categorical distinctions, with respect to main categories, whenever they reach a reasonable frequency threshold.

#### 4.2 *Quality of translation*

To give an example of the real input-output we should expect from a TB-MT system as the one we built using our aligned corpus, we report below an example that shows a simple case:

- (13) Input string:  
il presidente del consiglio parla con un segretario  
*the president of the council talks with a secretary*

Output of the parsing analysis<sup>5</sup>:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<node agree="3.s" cat="VP" role="head">
  <node agree="m.s" cat="NE.per" role="arg.subj">
    <word agree="m.s" cat="D.art.def" id="2" lemma="il">Il</word>
    <word agree="m.s" cat="N.comm.count.person" id="3" lemma="presidente" role="head">Presidente</word>
    <node agree="m.s" cat="NP" id="4" role="adj.matter">
      <word agree="m.s" cat="P.matter" id="5" lemma="di+il">del</word>
      <word agree="m.s" cat="NE.org" id="6" lemma="consiglio" role="head">Consiglio</word>
    </node>
  </node>
  <word agree="3.s" cat="V.ind.pres" lemma="parlare" role="head" subcat="transitive">parla</word>
  <node agree="m.s" cat="NP" role="adj.comitat">
    <word cat="P.comitat" lemma="con">con</word>
    <word agree="m.s" cat="D.art.indef" lemma="un">un</word>
    <word agree="m.s" cat="N.comm.count.anim" lemma="segretario" role="head">segretario</word>
  </node>
</node>
```

Output of the transfer-based MT process:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<node cat="VP" role="head">
  <node cat="NE.per" id="0" role="arg.subj">
    <word cat="N.comm.count.person" id="2" lemma="presidente" role="head">persona presidente</word>
    <node cat="NP" id="3" role="adj.matter">
      <word cat="NE.org" id="5" lemma="consiglio" role="head">Consiglio</word>
    </node>
    <word cat="END.comma" id="7" lemma=",">,</word>
  </node>
  <node cat="NP" role="adj.comitat">
    <word cat="N.comm.count.anim" lemma="segretario" role="head">persona segreteria</word>
  </node>
  <word cat="V.ind.pres" lemma="parlare" role="head" subcat="transitive">dire</word>
</node>
```

Despite this very simple example, such approach allows us to deal with rather subtle phenomena: for instance, extraposition and leftward-movement are constrained in a

---

<sup>5</sup> NE.per are Personal Named Entities, NE.org are Organization Named Entities.

very productive way by flattening the structure: assuming that the attachment point of the extraposed relative/PP is the immediate upper phrase (14), we can capture 96% of extraposed constituents; this is true also for genitive constructions (15):

- (14) (ITA) [VP [arg.subj[adj.rel.restr]] [V-head]] →  
 (LIS) [VP [arg.subj] [V-head] [adj.rel.restr]]

- (15) (ITA) [NP [N-head] [arg.subj[N-head]]] → [NP [N-head]<sub>i</sub> [N-head] [arg.subj[<sub>i</sub>]]] (LIS)  
 la foto di Gianni Gianni foto sua  
 the picture of John John picture his

Moreover, using empty elements (e.g. null-subjects, reduced relatives etc.) and a (relative) referential mechanism allow us to extract rules for re-integrating the referents also in discontinuous dependents:

- (16) [Il rappresentante<sub>i</sub> [di profumi] [che<sub>i</sub> è venuto ieri]] →  
 The perfume salesman that came yesterday

[NP [N-head] ... [NP/RC Relative\_Pro<sub>head\_of\_the\_dominating\_NP</sub> ...]]

In the end we attempted to make a human evaluation of the TB-MT system: a set of 50 sentences (Appendix E) which the system has not been trained on, has been semi-automatically analyzed and then automatically translated according to the rule extracted from the aligned corpus. A native speaker evaluated the provided translations with respect to word order soundness<sup>6</sup>, on a scale from 0 to 3 (0=incomprehensible, 1=comprehensible but sub-standard, 2=comprehensible, 3=good). The translations received an average score of 1.58, which is not a bad result at all for a naïf TB-MT system.

#### 4.3 Concluding remarks

In the end, we showed that the proposed structures/categories, inspired by main current generative frameworks (Minimalism, Chomsky 1995-2005, and Cartography, Belletti 2004, Cinque 2002 and Rizzi 2004) can be coherently implemented in a bilingual aligned treebank ITA-LIS. The TB-MT task seems to take advantage of such a rich structure and the translation provided seems to be fairly acceptable by native speakers. Obviously more tests are required on the word sense disambiguation side and the treebank should be significantly augmented from a quantitative point of view. These first results however seem to show that the undertaken mission is fully promising.

<sup>6</sup> Since some of the lexical items were not present neither in the corpus nor in the aligned bilingual lexicon, we could not expect the system to make the correct lexical choice in these cases.

## **Appendix A – XML DTD**

The Document Type Definition (Siena University Treebank, Version 1.0) is defined as follows (the DTD filename referred by the XML files in the treebank is “SUT.dtd”):

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!ELEMENT text (node|expression)+>
<!ELEMENT node (node|word)*>
<!ELEMENT expression (#PCDATA)>
<!ELEMENT word (#PCDATA)>
<!ATTLIST expression id CDATA #REQUIRED>

<!ATTLIST text id CDATA #REQUIRED>
<!ATTLIST text lang CDATA #REQUIRED>
<!ATTLIST text type CDATA #REQUIRED>

<!ATTLIST node id CDATA #IMPLIED>
<!ATTLIST node cat CDATA #REQUIRED>
<!ATTLIST node subcat CDATA #IMPLIED>
<!ATTLIST node ref CDATA #IMPLIED>
<!ATTLIST node role CDATA #IMPLIED>
<!ATTLIST node agree CDATA #IMPLIED>
<!ATTLIST node expr CDATA #IMPLIED>
<!ATTLIST node lp CDATA #IMPLIED>

<!ATTLIST word id CDATA #IMPLIED>
<!ATTLIST word cat CDATA #REQUIRED>
<!ATTLIST word subcat CDATA #IMPLIED>
<!ATTLIST word ref CDATA #IMPLIED>
<!ATTLIST word agree CDATA #IMPLIED>
<!ATTLIST word role CDATA #IMPLIED>
<!ATTLIST word lemma CDATA #IMPLIED>
<!ATTLIST word expr CDATA #IMPLIED>
<!ATTLIST word lp CDATA #IMPLIED>
<!ATTLIST word sem CDATA #IMPLIED>
```

The following attributes are only used by the XMLTreeViewer tool and are never displayed in the user-accessible XML structure:

```
<!ATTLIST node tmpid CDATA #IMPLIED>
<!ATTLIST node x CDATA #IMPLIED>
<!ATTLIST node y CDATA #IMPLIED>
<!ATTLIST word tmpid CDATA #IMPLIED>
<!ATTLIST word x CDATA #IMPLIED>
<!ATTLIST word y CDATA #IMPLIED>
```

## **Appendix B – Attribute-Value constraints**

This is the list of the main attributes (linguistic features) and their possible values (SUT Version 1.0). The number that precedes the value indicates the absolute order of the features when they are concatenated under the same attribute (concatenation of features is not a linguistically motivated solution, it simply solves a backward compatibility issue; most of the time every row would deserve an independent attribute specification, some other time concatenated features should be grouped in a different way (e.g. ); the next version of the tools should consider this issue).

## Nouns

e.g. “case” (houses): cat=“N.comm.count.inanim”, agree=“f.p”, role=“head” lemma=“casa”

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. N/N.pro[.cl] 2. [ <b>comm</b> /prop] 3. [ <b>count</b> /mass] 4. [anim/[per[.first/.last] /impers/reflex] /inanim/[city/gpe/org]]	noun/pronoun[clitic] common/proper countable/mass animate/[person[first/last name] impersonal/reflexive] /inanimate[city/geo-political entity/company]
Agree	1. [ <b>m</b> /f/n] 2. [ <b>s</b> /p/n]	masc/sing/neut gender sing/plur/null number
Role	<b>head</b> /arg/adj	head / selected argument / unselected adjunct
Sem	[alphanumeric index]	MultiWordnet id
Lemma	[any alphanumeric character]	dictionary uninflected form, if null its value is the token form

*Sem* and *Lemma* (as *Id* and *Ref*, §2.2) will be omitted from the following tables since the same values/constraints discussed here will apply.

## Verbs

e.g. “corre” ((he) runs): cat=“V.ind.pres”, agree=“s”, role=“head” lemma=“correre”)

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. V/V.aux/V.mod/V.asp 2. <b>ind</b> /subj/cond/part/imp/inf  3. pres/past/past+/fut/fut+/impf 4. [state/event[.atelic/.telic[.punct]]]	main/auxiliary/modal/aspectual verb indicative/subjunctive/conditional/ participle/imperative/inifitive mood present/past/remote past/future/ anterior future/imperfect aspectual classes (e.g. “cough” is an event, telic and punctual)
Subcat	transitive/intransitive/ditransitive/ unaccusative/copula/ causative/passive/psych/ control_subj/control_obj	Subcategorization classes
Agree	1. [1/2/3] 2. [ <b>m</b> /f/n] 3. [ <b>s</b> /p/n]	person gender number
Role	<b>head</b> /[adj]	head / unselected adjunct (e.g. auxiliaries, modals)

*Adjectives*

e.g. “forte” (strong): cat=“A.qualif”, agree=“f.s”

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. A 2. deict/dem/excl/indef/interr/nation/ num[.ord/.card]/poss/qualif	adjective deictic/demonstrative/exclamative/ interrogative/geographical specification/numeral[ordinal/cardinal]/ possessive/qualificative
Subcat	super/dimin/compar	superlative/diminutive/comparative form
Agree	as for Nouns	
Role	as for Nouns	

*Adverbs*

e.g. *prima* (before): cat=“ADV.time”

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. ADV 2. adfirm/advers/compar/doubt/ interr/limit/loc[.pro.cl]/manner/neg/ quant/reason/streng/ superl/temp	adverb adfirmirmative/adversative/comparative /doubitative/interrogative/limitative/ locative[.pro.cl]/manner/negative/ quantitative/reason/strength/ superlative/tempoparl
Role	[adj]	adjunct

*Determiners*

e.g. *il gatto* (the cat): cat=“D.art.def”

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. D 2. art[.def/.indef]/demo/ quant[.univ/.exist/.comp/.distr/.neg]	determiner article[definite/indefinite]/demonstrative/ quantifier[universal/exististential/ comparative/distributive/negative]
Agree	as for Nouns	
Role	[adj]	adjunct

*Prepositions*

e.g. “il libro *di* Gianni” (the book of G.): cat=“P.genitive”

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. P 2. advers/benef/comitat/compar /dative/evident/genitive/goal /instr/loc/manner/malefact /material/matter/means/measure /partitive/path/reason/source/temp	adverb adversative/benefactive/comitative/ comparative/dative/evidential/genitive/ goal/instrument/locative/manner/ malefactive/material/matter/means/measure /partitive/path/reason/source/temporal
Role	[adj]	adjunct

*Complementizers*

e.g. “*di*” (to): cat=“C.decl”

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. C 2. coord[.advers]/rel.pro/wh/ subord[.advers/.reason/.goal .conc/.cond/.decl/.fin/.loc/.temp]	complementaizer coordination[.adversative]/relative pronoun/wh-element/ subordinator[adversative/reason/goal concessive/conditional/declarative/ final/locative/temporal]
Role	[adj]	adjunct

*Specials*

e.g. “.” (dot, punctuation): cat=“END.period”

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. END/ABBR/INT/SPECIAL 2. period/comma/colon/scolon/quote	punctuation/abbreviations/interjections/ special characters (e.g. currency, percentage etc.)

*Non terminal nodes*

NPs, VPs and APs

<i>Attribute</i>	<i>Value (default, [optional])</i>	<i>Explanation</i>
Cat	1. NP/VP/AP/FRAG	nominal/verbal/modifier (both adjectival and adverbial) phrases/fragment
Role	adj	adjunct

## Appendix C - Functional Dependencies

The set of dependencies used to annotate the relation between phrases is the following one:

- head phase head
- arg(uments)
  - subj(ect) nominative case-marked argument
  - obj(ect) accusative case-marked argument
  - ind(irect)obj(ect) third argument (e.g. dative)
  - predobj(ect) object in copular constructions
- adj(uncts)
  - advers adversative specification
  - affirm affirmative specification
  - benef benefactive specification
  - cond conditional specification
  - coord coordination specification (second conjunct is marked adj.coord and it is dominated by the previous one)
  - comitat comitative specification
  - compar comparative specification
  - hangtopic extra argument (topic) specification
  - measure measure specification
  - evident evidential specification
  - goal goal specification
  - instr instrument specification
  - loc locative specification
  - malefact malefactive specification
  - manner manner specification
  - matter matter specification
  - means means specification
  - path path specification
  - partitive partitive specification
  - reason reason specification
  - source source specification
  - temp temporal specification
  - rel relative clause
    - restr restrictive relative
    - adpos adpositive relative

We decided to subcategorize prepositions according to the functional specification they introduce (the relation is not always 1-to-1). The following table summarizes the main subcategories briefly explaining them.

Prepositional subcategory	Examples	Brief Explanation [Typically, it can be used to answers a question such as:]
<b>Genitive</b>	<i>il presidente della repubblica</i> (arg.obj - i.e. a specification) [the president of the Republic] <i>la conferma dei socialisti</i> (arg.subj - i.e. subject/owner) [the confirmation of the Socialists] <i>le chiavi di casa</i> (adj.matter) [the keys of the house]	Usually used for animate complements, it introduces a specification or the subject or the owner of something [ <i>of whom?</i> ]
<b>Matter</b>	<i>risultati delle elezioni</i> (arg.obj) [the results of the elections] <i>rinunciare alla carica</i> (indobj) [to give up an office]	Usually used for inanimate complements, it introduces the matter or topic of something [ <i>about/of what?</i> ]
<b>Dative</b>	<i>essere ucciso dai carabinieri</i> (indobj - passive) [being killed by cops]	It introduces the indirect object
<b>Loc</b>	<i>vivo a Roma</i> [I live in Rome]	It introduces the place where the action occurs [ <i>where did it happen?</i> ]
<b>Source</b>	<i>uscire di casa</i> [to leave the house]	It introduces the origin of a movement [ <i>from where does x move?</i> ]
<b>Path</b>	<i>Vado verso la periferia</i> [I'm going towards the outskirts]	It introduces the direction of a movement [ <i>towards what does x move?</i> ]
<b>Benef</b>	<i>mese positivo per l'economia</i> [positive month for the economy]	It introduces the participant who benefits from the action [ <i>for whom?</i> ]
<b>Malefact</b>	<i>dare fuoco al pino</i> [to set fire to the pine tree]	It introduces an opponent, as well as a participant who is penalized by the action [ <i>against whom/what?</i> ]
<b>Manner</b>	<i>corro da solo</i> [I run by myself]	It introduces the manner in which a certain action takes place [ <i>how?</i> ]
<b>Means</b>	<i>vado col treno</i> [I move by train]	It introduces the mean of transportation [ <i>by/with what?</i> ]
<b>Measure</b>	<i>creocere di 3 metri</i> [to grow 3 meters]	It introduces a quantitative description of an action [ <i>how much?</i> ]
<b>Temp</b>	<i>dormo da giorni</i> [I slept for days] <i>pulisco di domenica</i> [I clean up on sunday]	It introduces a temporal characterization of an action [ <i>When? How long? From when? Untill when?...</i> ]

<b>Comitat</b>	<i>l'accordo coi centristi</i> [the deal with the centrists]	It introduces other people that share the role of the subject [with whom?]
<b>Partitive</b>	<i>uno di noi</i> [one of us]	It introduces the set which an object belongs to [of what (set)?]
<b>Instrument</b>	<i>lingua dei segni</i> [sign language - "a language that uses visually transmitted sign pattern"]	It introduces the object used to perform the action [by using what?]
<b>Material</b>	<i>la casa di legno</i> [the house made of wood]	It introduces the substance which an object is made of [made of what?]
<b>Evident</b>	<i>secondo il Presidente</i> [according to the President]	It introduces someone perspective [according to what/whom?]
<b>Compar</b>	<i>più bello di me</i> [more beautiful than me]	It introduces the second term of a comparison [compared to whom/what?]
<b>Reason</b>	<i>accordo per il ballottaggio</i> [the deal for the ballots]	It introduces the cause of a certain action [because of what?]
<b>Goal</b>	<i>corsa per la vittoria</i> [running for victor]	It introduces the goal of an action [why/for what?]

## Appendix D – Special features for tagging Sign Languages

Sign Languages require an enriched set of features to express properties that are not usually present in oral languages (e.g. morpho-syntactic Agreement in Sign Language is on a spatial dimension rather than on a gender dimension as in Oral Languages).

In the table below we report the set of features used to express agreement and other functional features (*lp* attribute in our xml files):

What	Feature	Brief Explanation
		We decide to refer to the space agreement as a “3-dimensional” space in a non-conventional sense:
Agree	body_contact	The first dimension is the contact with the body. By default a sign is not expressed touching a specific part of the body (unless explicitly marked in the lexicon);
	forehead/eyes-nose/mouth/neck/chest/stomach	A second dimension is the height of the sign: by default a sign is expressed in the neutral space, that is in front of the chest; otherwise it can be signed at the forehead level or at the eyes-nose, mouth, neck, chest or stomach levels

---

	left/right	On a horizontal dimension a sign is expressed by default in the neutral space, that is right in front of the chest; otherwise we can specify a left or right position
		No agreement information means that the sign is expressed in the neutral space; otherwise non-default dimensions are concatenated e.g. <i>body_contact. mouth.left</i>
<b>Classifiers</b>	cl.shape/cl.space/ cl.movement	The classifier system indicates when a sign is not expressed as coded within the lexicon; shape, space and movement are the feature that the modified sign inherits from the dependent sign in the context (the head of the phrase if not explicitly marked)
	eye-up eye-down keep-support-hand neg past/fut	Eyebrows up (yes-no question) Eyebrows down (wh-/rhetorical questions) keep-support-hand head shaking expressing negation movement to express past (toward the shoulder) and future (from the shoulder) times
<b>Special functional features</b>	slow/fast exaggerate/minimize now gaze-to-sign labialization	velocity modification of the sign (e.g. depending of the strength adverbial modifiers) exaggerate/minimize the movement of the sign (expresses diminutives, augmentatives features) gaze at the neutral space (it indicates the present time) the gaze directed to the sign marks the familiarity with the object just signed indicates that a specific labialization is required (usually when a word is finger-spelt)

---

### Appendix E – Machine Translation Test

The alignment rules extracted from the corpus have been used to train a MT system. This system has been exposed to 50 new parsed sentences and the pairs original Italian sentence - MT output in LIS has been evaluated by a native speaker. The score ranged from 0 (incomprehensible) to 3 (good) (1 = comprehensible but sub-standard; 2 = acceptable);

This is the raw result of the test: 3 - 22%; 2 - 26%; 1 - 40%; 0 - 12%; mean: 1.58

<i>ID</i>	<i>Original Italian Sentence</i>	<i>Output in LIS<sup>7</sup></i>	<i>Score</i>
1	accade che qualche giovane si accontenti	pro giovane accontentare accadere	3
2	cinquecento studenti sono ospitati in collegi universitari	collegio;PL universitario cinquecento persona studiare ospitare fatto	3
3	una città che non riesce ad accogliere i giovani è destinata a morire	pro accogliere persona giovane riuscire no città morire destinato	3
4	ha deciso , infatti , di non applicare il decreto	proprio decreto applicare no decidere fatto	3
5	così , è stato deciso di battere il bosco	bosco battere decidere fatto	3
6	i commissari di gara lo hanno accusato di avere intralciato Massa	gara commissario;PL accusare Massa intralciare fatto	3
7	quando il vaccino terapeutico risulterà disponibile	data vaccino curare c'è risultare	3
8	riscaldarsi quest' inverno sarà veramente costoso	inverno riscaldare+si costoso veramente	3
9	potrà anche estinguere il mutuo	potere anche mutuo estinguere	3
10	la squadra di Detroit ha quasi sempre vinto la partita	squadra Detroit partita vincere quasi	3
11	io ci ho sempre provato	io provare fatto	3
12	come fate a offrire residenze a prezzo calmierato ?	prezzo calmierato residenza;PL offrire fare	2
13	gli affitti continuano ad essere cari	affitto;PL continuare caro	2
14	gli incidenti sono tutti da attribuire al fattore umano	tutti incidente attribuire fattore umano	2
15	è stata subito sottoposta a terapia malarica	subito terapia malarico fatto	2
16	la pace era già stata raggiunta da Heider	già problema Heider raggiungere fatto	2
17	lui ha deliberatamente frenato troppo tardi	troppo tardi pro frenare fatto deliberatamente	2
18	la fede dipende da Dio , e da Dio solo	fede dipendere Dio + Dio solo	2
19	si tratta di saper distinguere le emozioni	emozioni;PL sapere distinguere	2
20	siamo già in recessione	già recessione	2
21	i fondi per l' Africa si sono drasticamente ridotti	fondo;PL motivo africa ridurre fatto drasticamente	2
22	lo si era capito già in partenza	già partenza capire fatto	2
23	si chiamano nuovi acquisti perché devono portare qualcosa di nuovo	acquisto nuovo chiamare motivo dovere qualcosa nuovo portare	2
24	il contesto sociale in cui si è nati e cresciuti	nato fatto + crescere fatto situazione sociale	2
25	ha deciso di non applicare il decreto	decreto applicare no decidere fatto	1
26	i manager che falliscono saranno messi da parte	pro fallire manager;PL mettere fatto	1

<sup>7</sup> *pro* indicates a deictic sign to a position in the space where the referred object has been previously signed. ;*PL* indicates that the sign that precedes it has to be repeated (according to its plural status). + indicates the sign used for the conjunction of two expression. All suprasegmental features discussed are not included in the simple text transcription.

27	si è così arrivati all' individuazione di numerosi immobili	arrivare fatto immobile;PL numero individuazione	1
28	troverà ad attenderlo una lunga fila di bandiere italiane	bandiera;PL italia fila lungo trovare attendere	1
29	potrebbe essere sciolta la prognosi sulla sopravvivenza	potere sopravvivenza prognosi sciogliere fatto	1
30	sembra che ci fossero anche pietre difficili da individuare	pro anche pietra;PL individuare difficile sembrare	1
31	quando ieri gli ha annunciato che voleva parlare con lui , è rimasto in silenzio	data ieri pro volere parlare pro silenzio annunciare fatto	1
32	una parte della Curia fiorentina si accorgerà che avevamo ragione	curia fiorentino pro ragione avere accorgere	1
33	per stabilire dove stia la ragione e dove stia il torto	motivo ragione stare + torto stare stabilire	1
34	è stato Hamilton a sbagliare	sbagliare Hamilton	1
35	ammetto di aver sbagliato al via	sbagliare fatto ammettere	1
36	tutti voi siete una sola persona in Cristo	tutti voi Cristo persona sola	1
37	il tempo deve diventare la misura della vostra pazienza	dovere periodo pazienza vostro misura diventare	1
38	si consiglia di limitare il consumo di queste verdure pronte	consumo verdura;PL pronto limitare consigliare	1
39	un canarino è evidentemente il migliore rimedio contro le preoccupazioni atomiche	canarino contro preoccupazione;PL atomico rimedio buono evidentemente	1
40	ne abbiamo già parlato anche troppo	già ne parlare fatto anche troppo	1
41	non ho mai pensato di segnare in quel modo	segnare pi modo pensare fatto no contro	1
42	non ha cercato di segnare con la mano	mano segnare cercare fatto no	1
43	ora ci chiediamo se sia giusto questo turn-over massiccio	ora se turn-over massiccio giusto chiedere	1
44	a un gruppo di scrittori emiliani viene assegnato un prodotto tipico	prodotto tipico scrittore;PL emiliano gruppo assegnare fatto	1
45	il locatore chiede all' inquilino di versare ulteriori somme	locatore versare somma;PL ulteriore inquilino chiedere	0
46	gli affitti che gli studenti si trovano a dover pagare	affitto;PL	0
47	è seguito da un tutor che lo aiuta a orientarsi nella scelta dei corsi	pro orientare corso;PL scelta aiutare tutor;PL	0
48	imperversano anche le locazioni in nero	anche nero locazione;PL imperversare	0
49	diventa sempre più difficile venire a studiare nel capoluogo lombardo	venire capoluogo persona lombardo studiare difficile più diventare	0
50	è effettivamente un' azione comune quella che proponiamo	pro proporre diventare azione effettivamente	0

## References

Abney, S. 1987. *The English noun phrase in its nominal aspect*. Doctoral dissertation, MIT Press.

- Adger, D. 2007. *A minimalist theory of feature structure*. <http://ling.auf.net/lingBuzz/000583>
- Baker, M. 1997. "Thematic roles and syntactic structure". In *Elements of grammar: Handbook in generative syntax*, ed. Haegeman, L. Dordrecht: Kluwer: 73-137.
- Bartolini S., Bennati P., Giorgi R. 2006. *Bluesign-2, il nuovo visualizzatore portatile per la Lingua Italiana dei Segni*, in Proceedings of "51<sup>th</sup> Corso Nazionale di Studio, Formazione e Aggiornamento dell'AIES". Siena:Cantagalli:140-145.
- Belletti, A. 2004. *Structures and Beyond*. Oxford, Oxford University Press.
- Bentivogli, L., E. Pianta & C. Girardi, 2002. Multiwordnet: developing an aligned multilingual database. *First International Conference on Global WordNet, Mysore, India*.
- Bergman, B., P. Boyes-Braem, T. Hanke & E. Pizzuto 2001. *Sign Transcription and Database Storage of Sign Information*. Special issue of *Sign Language & Linguistics*, 4:1/2.
- Bosco, C., V. Lombardo, D. Vassallo, & L. Lesmo, 2000. "Building a treebank for Italian: a data-driven annotation schema." *Proceedings of the Second International Conference on Language Resources and Evaluation LREC*, 99-106.
- Chesi, C. 2007. "An introduction to Phase-based Minimalist Grammars: why move is Top-Down from Left-to-Right". *STiL - Studies in Linguistics*, 1.
- Chomsky, N. 1981. "Principles and parameters in syntactic theory." *Explanation in Linguistics: The Logical Problem of Language Acquisition*, 32-75.
- Chomsky, N. 1995. *The Minimalist Program*. MIT Press.
- Chomsky, N. 2005. On Phases. *Manuscript, MIT*.
- Cinque, G. 2002. *Complement and adverbial PPs: implications for clause structure*. Abstract, University of Venice.
- Cinque, G. 2002. *The cartography of syntactic structures. Vol. 1, Functional structure in DP and IP*, Oxford University Press.
- Frank, R. & A. Kroch. 1995. "Generalized transformations and the theory of grammar." *Studia Linguistica* 49:103-151.
- Harkema, H. 2001. "Parsing Minimalist Languages." University of California Los Angeles.
- Kayne, R. S. 1983. Connect & M. Calcagno. 2001. Parasitic gaps in English: some overlooked cases and their theoretical implications. In *Parasitic Gaps*, ed. P. Culicover and P. Postal, 181-222. Cambridge, Mass.: MIT Press.
- Kayne, R. S. 1983. *Connectedness and binary branching*. Foris Publications.
- Mana, N. & O. Corazzari, 2002. "The lexico-semantic annotation of an Italian Treebank." *Proceedings of LREC 2002*.
- Marcus, M. P., M. A. Marcinkiewicz & B. Santorini, (1993). Building a large annotated corpus of English: the penn treebank. *Computational Linguistics*, 19(2), 313-330.
- Monachini, M. 1995. *ELM-IT: An Italian Incarnation of the EAGLES-TS. Definition of Lexicon Specification and Classification Guidelines*. Technical report, Pisa.
- Rizzi, L. 1997. The fine structure of the left-periphery. In *Elements of Grammar: Handbook in generative syntax*, ed. Haegeman, L. Dordrecht: Kluwer. 1997, 281-337.
- Rizzi L. ed. 2004. *The Cartography of Syntactic Structures*, Oxford University Press.
- Stabler, E. 1997. "Derivational minimalism." *LECTURE NOTES IN COMPUTER SCIENCE* 68-95.
- Tamburini, F. 2007. *Evalita 2007: The Part-of-Speech Tagging Task*. Proceedings of EVALITA 2007.